ВЫСШЕЕ ОБРАЗОВАНИЕ

В.Я. Турецкий

МАТЕМАТИКА И ИНФОРМАТИКА

УЧЕБНИК

ВЫСШЕЕ ОБРАЗОВАНИЕ
серия основана в 1996 г.

Министерство образования Российской Федерации
Уральский государственный университет

В.Я. ТУРЕЦКИЙ

МАТЕМАТИКА И ИНФОРМАТИКА

3-е издание, переработанное и дополненное

Допущено
Министерством образования
Российской Федерации
в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по гуманитарным
направлениям и специальностям

Москва
ИНФРА-М
2000

УДК 681.3.06(075.8)
ББК 22.18я73
Т 86

Рецензенты: д-р физ.-мат. наук, профессор *А.В. Арутюнов*,
д-р физ.-мат. наук, профессор *Н.А. Бобылев*

Турецкий В.Я. Математика и информатика. – 3-е изд.,
Т 86 испр. и доп. – М.: ИНФРА-М, 2000. – 560 с. – (Серия «Высшее образование»).

ISBN 5-16-000171-9

Содержит базовые разделы математики и основы информатики. Материал изложен с учетом требований, предъявляемых к студентам гуманитарных направлений и специальностей.

Проект данной книги признан лучшим в конкурсе, проведенном Госкомвузом Российской Федерации в 1998 г.

Предназначен для студентов-гуманитариев.

ББК 22.18я73

ISBN 5-16-000171-9

Уважаемый читатель!

Перед Вами один из учебников нового поколения по дисциплине «Математика и информатика» для студентов высших учебных заведений, обучающихся по гуманитарным направлениям и специальностям профессионального образования, прошедший сложный и длительный путь конкурсного отбора.

Данное учебное пособие является одним из двух победителей по дисциплине «Математика и информатика» Всероссийского конкурса учебников нового поколения по общим фундаментальным естественнонаучным дисциплинам. Этот конкурс впервые в истории высшей школы в России был инициирован Госкомвузом России (в дальнейшем – Минобразованием России) в связи с реформированием структуры и содержания программ высшего образования и проведен в течение 1995–1998 годов на базе Российского университета дружбы народов.

В конкурсе приняли участие свыше 350 авторских коллективов практически из всех регионов России, заявки представлялись по 11 номинациям, а в их оценке участвовало более ста высококвалифицированных экспертов.

В результате двух туров конкурса было отобрано 39 авторских коллективов, чьи заявки, а затем и рукописи более всего соответствовали как новым учебным программам, так и государственным образовательным стандартам по каждой дисциплине.

Конкурсная комиссия выражает надежду, что данное учебное пособие внесет свой полезный вклад в дело дальнейшего совершенствования российского высшего профессионального образования, и желает всем читателям – студентам и преподавателям – больших творческих успехов.

Заместитель министра образования России,
академик Российской академии образования,
председатель конкурсной комиссии профессор
В.Д. Шадриков

Оглавление

Предисловие ректора

Уважаемый читатель! Вы раскрыли новый учебник по математике и информатике, написанный доцентом кафедры информатики и процессов управления Уральского государственного университета Владимиром Яковлевичем Турецким.

В последние годы стремительное проникновение математики и компьютерных технологий в нематематические, гуманитарные сферы стало свершившимся фактом. Все шире математические методы применяются в таких, казалось бы, далеких от математики областях знаний, как психология, социология, лингвистика, педагогика, история. Обработка больших массивов данных становится невозможной без использования компьютеров. Почти неограниченный доступ к информационным ресурсам предоставляет всемирная компьютерная сеть Интернет. В связи с этим встают совершенно новые задачи и перед высшим образованием. Необходимо так организовать процесс обучения на гуманитарных факультетах, чтобы студенты старших курсов не оказались в затруднительном положении, когда им придется применять математические и компьютерные методы в своей работе. Для этого они должны получить основательное базовое образование в области математики и информатики, и чем раньше, тем лучше. Самым разумным представляется введение нового, единого курса «Математика и информатика» для студентов-гуманитариев первого и второго курсов в рамках обучения бакалавров. Такой эксперимент Уральский государственный университет начал в 1992 г. на отделении психологии: была увеличена доля математики в общей сетке учебных часов. С первого дня преподавание математики и информатики для студентов-психологов взял на себя В.Я. Турецкий. Его оригинальный курс лекций и лег в основу предлагаемого вашему вниманию учебника. Особенно важно, что инициатива по внедрению математических методов в учебный процесс исходила от руководителей факультета психологии, понимавших необходимость усиления математической и компьютерной подготовки студентов. В настоящее время курс «Математика и информатика» для гуманитарных факультетов и специальностей официально введен в образовательные стандарты российских университетов.

Качество учебника, безусловно, может быть проверено только временем. Однако лекции и занятия В.Я. Турецкого на отделении психологии УрГУ всегда воспринимались с интересом и пониманием благодаря хорошему языку, доступному стилю изложения, обилию примеров и иллюстраций. Уверен, что эти качества в полной мере присущи данному учебнику и вы получите удовольствие от знакомства с этой книгой.

Несколько слов об авторе учебника. Доцент В.Я. Турецкий — мой ученик. Он специалист в области математической теории управления и дифференциальных игр, автор многих статей и докладов, в том числе на международных научных конференциях. В настоящее время участвует в научном проекте, финансируемом Российским фондом фундаментальных исследований. Проект данного учебника, разработанный В.Я. Турецким, победил в конкурсе Госкомвуза Российской Федерации в 1998 г. Часть, посвященная теории множеств и отношений, опубликована ранее в виде методической разработки.

Таким образом, учебник представляет собой плод пятилетнего труда, параллельного усовершенствования курса лекций и текста книги. Надеюсь, что он облегчит студентам усвоение непростого математического материала и обеспечит качество профессиональной подготовки на уровне государственного стандарта.

С уважением
ректор Уральского
государственного
университета
профессор
В.Е. Третьяков

К читателю

Мне бы хотелось начать учебник словами, которыми я начинаю свою первую лекцию по математике: «Здравствуйте, мои дорогие студенты! Здравствуйте, мои дорогие читатели!» Я вижу перед собой аудиторию, полную студентов-первокурсников. Несколько десятков пар глаз направлены на меня. Эти молодые глаза открыты миру, в них нет ожидания подвоха, в них — готовность понять и научиться. По ходу лекции я увижу в них и иронию, и радость понимания, и недоуменные вопросы. Я весь как на ладони перед этими глазами, и мне так хочется не подвести, оправдать ожидания, ответить на все вопросы, научить...

Это очень трудно — и учиться, и учить. И всегда остается сомнение — а можно ли научить? Научил ли? И после каждой лекции кажется, что не все слова найдены, что можно было бы объяснить проще, нагляднее, короче. С годами приходит опыт, накапливаются приемы, примеры. Однако каждый раз трудна первая лекция. Когда я учил плавать своего младшего сына, мне не было легче, чем со старшим. Точно так же он барахтался, боялся оторваться от меня. Точно так же опускались у меня руки. И точно так же в конце концов он делал первые правильные движения, проплывал первые сантиметры — плыл! Это я его научил или он научился сам? Конечно, и то и другое. Ведь обучение — это взаимный процесс, это путь навстречу, половину которого должен пройти каждый из нас, учитель и ученик. В этой дороге не стоит рассчитывать на быстрый успех. Я понимаю ограниченность своих возможностей. Но мои руки не пусты — у меня есть знания, и я готов поделиться, помочь овладеть ими другому. Остальное — за вами. Смотрите — руки двигаются вот так, ноги — вот так. Ну, что же вы? Вот теперь правильно. Плывите!

Почему учебник?

Быть может, книга, учебник — это способ договорить то, что недосказано на лекциях. Это попытка объяснить все еще раз, расставить по местам, разложить по полочкам. Учебный план ставит нас в жесткие рамки, и, читая лекции, я должен жестко отсеивать материал, отбрасывать все лишнее, не связанное непосредственно с темой. На страницах книги я могу отвлечься, порассуждать об общих (но интересных для аудитории) вопросах. Иной раз философия предмета столь же увлекательна, как и сам предмет, а лекции просто не оставляют на это времени. В книге я могу использовать свой опыт разных лет, опыт других преподавателей, авторов других книг. Наконец, при создании учебника у меня есть уникальная возможность вернуться к уже написанным главам, изложить их по-новому, что-то добавить и что-то вычеркнуть. Преподавание не дает такой возможности. Я могу прочитать эту лекцию по-другому, но передо мной будут уже другие люди. В этом я вижу мои личные мотивы для написания этого учебника. Внешние же обстоятельства вполне понятны: нехватка учебной литературы в университетах, отсутствие адекватного математического курса для студентов гуманитарных специальностей.

Зачем? Два ответа на один вопрос

Итак, мы подходим к главному вопросу, который я читаю в глазах вчерашних (и позавчерашних) школьников на своей первой лекции, а иногда и слышу в непосредственной и простой формулировке. Это вопрос «Зачем нам это нужно?» Речь идет о математике, поскольку необходимость обучения работе на компьютере сейчас мало кем оспаривается. Первокурсники становятся год от года все прагматичнее. Им не хочется тратить время попусту на изучение науки, которая весьма отдаленно связана с их будущей профессией. Ну что ж, их вопрос вполне законный и достоин честного и обстоятельного обсуждения.

Обдумав эту проблему, я пришел для себя к двум возможным ответам. Первый дается с общекультурных позиций. Университетское, а тем более гуманитарное, образование подразумевает освоение базовых фундаментальных достижений человеческой культуры. То, над чем столетиями бились лучшие умы, достойно как минимум уважения, и интеллигентный человек, безусловно, должен иметь представление об этих достижениях. Математике по праву отводится важное место в общечеловеческой культуре. Как способ описания действительности математика занимает промежуточное положение между точными науками (физика, химия, механика и т.д.) и искусством. Математическое мышление сочетает в себе рационализм и эстетические качества, красоту. С одной стороны, математика отталкивается от реальности, ее результаты могут быть применены и применяются на практике. С другой стороны, она развивается по своим внутренним законам, очень близким к законам красоты и соразмерности. Математические теоремы ценны сами по себе,

безотносительно к возможности их практического использования. Они могут доставлять наслаждение так же, как архитектурный ансамбль, картина, музыкальное или литературное произведение. Недаром существуют выражения «красивое доказательство», «красивый результат». По существу, математика представляет собой ту связь между естественными и гуманитарными науками, без которой картина мира распадается на отдельные части. С этой точки зрения качественное гуманитарное образование должно включать в себя основательное изучение математики.

Второй ответ продиктован чисто практическими соображениями. Многие гуманитарные науки в качестве инструмента для своих исследований используют математические методы. Такие методы применяются в психологии, социологии, очень развиты математические подходы в лингвистике. Уверен, что в скором времени и другие гуманитарные исследования будут оснащены математическим аппаратом. Прикладные математические методы опираются на результаты многих математических дисциплин. Эти методы представляют собой лишь верхушку айсберга, их основа, фундамент скрыты в базовых разделах математики – алгебре, математическом анализе, теории вероятностей и др. Поэтому, чтобы овладеть математическими методами, применяемыми в какой-либо области, необходимо иметь представление об их основах, т.е. требуется систематическое изложение математических курсов. С моей точки зрения, это подтверждает целесообразность изучения математики студентами гуманитарных специальностей.

Авторское право

Следующий важный вопрос, нуждающийся в прояснении, – щепетильный вопрос авторского права. Любой учебник в значительной мере представляет собой компиляцию, сводку данных из других книг и пособий. Ведь математические результаты уже получены и описаны другими авторами (если речь идет не о новых достижениях, а о классических разделах математики). Свою задачу я вижу в разумной организации материала, его доступном и четком изложении. При этом не обойтись без заимствований, в том числе и текстуальных. Если какой-либо раздел уже блестяще изложен в другой книге, цитата пойдет лишь на пользу учебнику, поможет лучшему восприятию текста студентами. Цитата может быть дословной, а может быть и смысловой, переработанной в духе и стиле данной книги. В таком случае цитата становится неотъемлемой частью текста и кавычки, ссылки и прочие указания на фактическое авторство, как правило, опускаются. Хотелось бы заранее принести благодарность авторам книг, фрагменты из которых использованы при создании данного учебника. Все соответствующие книги перечислены в списке литературы.

Благодарности

В заключение хочу сердечно поблагодарить моих коллег, беседы и дискуссии с которыми очень помогли мне написать эту книгу. В ходе наших неформальных обсуждений оттачивались подходы, уточнялись принципы расположения материала, план учебника. Я благодарен коллегам за их доброжелательные советы, за время, потраченное на обмен мнений об учебнике. Нельзя не упомянуть о помощи, которую мне оказали В.Е. Третьяков, В.В. Расин, П.М. Аронов, С.А. Пьянзин. Особая благодарность – Ф.А. Шолоховичу, Б.М. Верникову и Ю.М. Важенину, заметившим множество опечаток и неточностей при подготовке второго издания. Кроме того, я благодарен всем сотрудникам кафедры информатики и процессов управления Уральского государственного университета за более свободный режим работы, которым я пользовался в период написания книги.

Я благодарен студентам факультета психологии Уральского государственного университета, которые своим внимательным отношением и вдумчивыми вопросами способствовали ежегодному усовершенствованию читаемого курса. Первые слушатели курса (студенты факультета психологии) сейчас уже закончили университет. Мне известно, что знания, полученные на занятиях по математике и информатике, помогают им в работе.

Особая благодарность – моей семье, жене и сыновьям, за терпение и понимание. Хотелось бы выразить благодарность Госкомвузу Российской Федерации, чья инициатива и поддержка привели к написанию данного учебника, а также руководству Гуманитарного университета, оказавшему помощь в его первом издании.

В.Я. Турецкий

Предисловие автора

Предлагаемый вашему вниманию учебник представляет собой базовый курс математики и информатики для студентов гуманитарных направлений и специальностей на уровне бакалавриата. Содержание математической части учебника, выбор тем и разделов обусловлены двойственным взглядом на математику в системе гуманитарного образования. С одной стороны, это фундаментальные математические понятия как часть общего культурного наследия. С другой стороны, это разделы, необходимые для овладения прикладными математическими методами. В связи с этим курс **математики** в учебнике разбит на пять частей:

1. Основания математики.
2. Основы алгебры и аналитической геометрии.
3. Основы математического анализа.
4. Основы теории вероятностей.
5. Элементы математической статистики.

Не случайно названия почти всех частей содержат слова «основы» или «элементы». Этим подчеркивается общий характер изложения, отсутствие математических тонкостей, не столь важных и интересных для читателя гуманитарного склада.

Курс математики построен по принципу пирамиды: прикладные математические методы невозможно освоить без знания математической статистики. Статистика, в свою очередь, опирается на теорию вероятностей. В теории вероятностей используются результаты математического анализа, основу которого составляют фундаментальные понятия множества и функции. Алгебраические и геометрические представления составляют неотъемлемую часть общематематической культуры, поэтому без них трудно представить себе сколь бы то ни было связный математический курс.

В изложении мы существенно опираемся на школьные знания. Ведь программа средней школы дает весьма полное представление о структуре и основных подходах математики. Это касается, к примеру, эволюции понятия числа – от натурального ряда к действительным числам. Вам знакомы понятия множества и отношения (их сейчас «проходят» учащиеся младших классов), понятие и основные свойства функции и даже такие сложные математические объекты, как вектор, предел, производная и интеграл. Одна из задач учебника – систематизировать известные по школе результаты, ввести их в общематематический контекст, проследить связи и зависимости между ними.

Следует отметить, что изучение математики не должно заканчиваться на этом базовом курсе. Оно может быть естественным образом продолжено на старших курсах изучением специфических математических методов в конкретных областях гуманитарных знаний: «Математические методы в психологии», «Математические методы в лингвистике» и т.д. Этим завершится формирование целостного представления у студента-гуманитария о роли и значении математики в современном мире.

Изложение **информатики** в учебнике подчинено тем же общим принципам, что и изложение математики. С одной стороны, вводятся основные компьютерные понятия и термины, без которых невозможно даже самое непрофессиональное пользование компьютером. С другой стороны, обзор программного обеспечения дает читателю возможность сориентироваться, выбрать подходящий для себя программный продукт, зная и понимая характерные черты различных программ. Учебник, безусловно, не дублирует многочисленные руководства пользователя по различным программам. Его цель – ввести студента в мир компьютера, подсказать пути использования компьютера в повседневной жизни и профессиональной деятельности.

В конце каждой главы приводятся задачи для самостоятельного решения. Они помогут закрепить прочитанный материал, применить теоретические результаты на практике. Типовые задачи разбираются в соответствующих разделах. Задачи могут быть использованы на практических и семинарских занятиях, сопровождающих курс математики и информатики. Некоторые разделы учебника не вошли в основной курс и предназначены для самостоятельного изучения заинтересованным читателем. Такие разделы (и соответствующие им задачи) помечены звездочкой.

Часть 1
ОСНОВАНИЯ МАТЕМАТИКИ

Искусство хитрое цифири
Нас учит: дважды два – четыре,
И помогает нам понять,
Как ложь за истину принять.

И.В. Гете

При изучении оснований математики мы имеем дело с главными математическими понятиями, без которых невозможно изучение любого раздела математики. Такие понятия, как *«множество»*, *«функция»*, *«отношение»* и другие, используются на протяжении всего курса высшей математики. Они представляют собой основу математической культуры, которая является важной частью культуры общечеловеческой.

По иронии судьбы основания математики разрабатывались уже после того, как ее здание, казалось, было полностью построено. Драматичная история математики вообще не похожа на прямую дорогу, на восхождение от фундамента к верхним этажам. Это во многом история ее кризисов, которые в конечном счете оказывались весьма плодотворными и выводили науку на новый уровень развития. Здание математики несколько раз перестраивалось, но раз от раза становилось более строгим и красивым.

Глава 1 данного учебника посвящена *теории множеств*, созданной в конце XIX в. великим немецким математиком Георгом Кантором. Благодаря этой теории были получены ответы на многие вопросы, не дававшие покоя математикам на протяжении нескольких веков. В свою очередь, канторовская теория множеств оказалась не свободной от внутренних противоречий, так называемых парадоксов, что привело в 1897 г. к новому кризису основ математики, который нельзя считать разрешившимся и до сих пор. Тем не менее влияние канторовской теории на дальнейшее развитие математики нельзя переоценить. Другой великий немецкий математик Давид Гильберт (1862–1943) говорил: «Никто не изгонит нас из рая, созданного Кантором».

Главная заслуга Кантора состоит в признании того факта, что бесконечность – это не абстракция, придуманная философами, а реальность; что бесконечные совокупности предметов существуют наравне с конечными. В самом деле, легко построить прямоугольный треугольник с единичными катетами. По теореме Пифагора его гипотенуза равна $\sqrt{2}$, т.е. иррациональному числу, десятичное разложение которого бесконечно и не содержит периода.

Кантор показал, что бесконечными множествами можно оперировать точно так же, как и конечными, он научился определять «размеры» бесконечных множеств, сравнивать их между собой. Одним словом, благодаря его открытиям бесконечные множества стали рабочим аппаратом математики.

При этом Кантор выступал против традиционных представлений, разделяемых великими математиками прошлого. Канторовская теория бесконечных множеств вызвала бурю протестов. Несмотря на то что эта теория нашла применение во многих областях математики, некоторые ученые по-прежнему отказывались принимать актуально бесконечные множества и все, что с ними связано. Леопольд Кронекер считал Кантора шарлатаном и воспрепятствовал его работе в Берлинском университете. Тем не менее теория множеств Кантора нашла и своих горячих сторонников. Бертран Рассел называл Кантора одним из великих мыслителей XIX в. В 1910 г. Рассел писал: «Решение проблем, издавна окутывавших тайной математическую бесконечность, является, вероятно, величайшим достижением, которым должен гордиться наш век».

Между элементами различных множеств могут быть установлены многочисленные связи и зависимости. Эти связи описываются в математике с помощью отношений и функций, что составляет содержание главы 2.

Глава 1
МНОЖЕСТВА

1.1. Канторовское понятие множества

Множество относится к математическим объектам, для которых нет строгого определения. Другим примером неопределяемого понятия служит точка в геометрии. Такие понятия вводятся на интуитивном уровне, тем не менее на их основе даются строгие определения других математических объектов. Мы можем лишь в какой-то мере объяснить такое понятие, т.е. дать описание основных его свойств.

Кантор описывает множество следующим образом.

Множество S есть любое собрание определенных и различимых между собой объектов нашей интуиции и интеллекта, мыслимое как единое целое. Эти объекты называются *элементами* множества S.

Если предмет x является элементом множества S, это обозначается с помощью знака $\in$: $x \in S$. Тогда говорят, что элемент x *принадлежит* множеству S. В противном случае пишут $x \notin S$.

Обсудим основные моменты этого «определения».

«*Мыслимое как единое целое*»: важно, что собрание предметов рассматривается как один предмет. Предметы как бы собираются в мешок, и дальше работают с этим мешком как с единым целым, не задумываясь о содержимом. Такой подход хорошо известен в биологии, где растения и животные классифицируются по группам, видам, классам, отрядам и т.д. При этом внимание переносится с отдельных представителей на общие свойства группы как совокупности. В языке это отражено в таких словах, как «компания», «стая», «стадо» и т.п.

«*Объекты нашей интуиции и интеллекта*»: эта формулировка не накладывает никаких ограничений на природу элементов множества. Это может быть, например, множество студентов-психологов 1-го курса, множество пятен на Солнце, множество зеленых яблок, множество букв «я» в «Евгении Онегине», множество звезд на небе и т.д. Можно привести огромное количество самых разнообразных примеров множеств.

В математике обычно имеют дело со множествами математических объектов: чисел, точек, кривых и т.д. Для числовых множеств имеются общепринятые обозначения: $\boldsymbol{Z}$ – множество целых чисел, $\boldsymbol{Q}$ – рациональных, $\boldsymbol{R}$ – действительных. В школьной геометрии рассматриваются множества точек: например, множество точек на плоскости, равноудаленных от заданной точки, или окружность.

Можно говорить и о множествах, элементы которых нельзя точно указать. Например, множество погибших в Троянской войне, множество «вздохов на скамейке» и т.д.

Группа студентов – множество, элементами которого являются отдельные люди. В свою очередь, студенческая группа есть элемент множества всех групп в университете. Таким образом, элементами множества могут быть множества.

Продолжим обсуждение канторовского «определения».

«*Различимые*»: это слово понимается в следующем смысле. Для любых двух предметов, рассматриваемых в качестве элементов множества, должна иметься возможность решить, различны они или одинаковы. Пусть, например, мы знаем, что среди элементов числового множества есть «наименьшее простое четное число» и «целое число, на единицу меньшее трех». Очевидно, это один и тот же элемент — число 2, хотя и описанный разными способами.

«Определенные»: если даны какое-то множество и некоторый предмет, то можно определить, является этот предмет элементом данного множества или нет. Таким образом, множество полностью определяется своими элементами.

Георг Кантор (1845–1918). Немецкий математик. Родился в Петербурге. Окончил Берлинский университет (1867). С 1869 г. преподавал в университете в Галле (в 1879–1913 – профессор).

Основоположник теории множеств. Сформулировал (1878) общее определение мощности множества, первое определение континуума, ввел понятия счетных и несчетных множеств, пустого множества. Развил принципы сравнения множеств. Систематическое изложение принципов своего учения о бесконечности дал в 1879–1884 гг. Ввел (1883) новое понятие действительных чисел, которое включило как рациональные, так и иррациональные числа.

Настойчивое требование Кантора рассматривать бесконечность как нечто актуально данное было для того времени большой новостью. Предубеждение по отношению к такой точке зрения обусловило непризнание работ Кантора со стороны некоторых математиков, реакция других была более благоприятна. К 1890 г., когда были получены приложения теории множеств к анализу и геометрии, эта теория получила признание в качестве самостоятельного раздела математики.

Кантор тяжело переживал открывшиеся противоречия своей теории (так называемые парадоксы, или антиномии). С 1884 г. он страдал глубокой депрессией и с 1897-го отошел от научной деятельности.

Основатель и первый президент Германского математического общества (1890–1893), инициатор созыва первого Международного математического конгресса в Цюрихе.

1.2. Конечные и бесконечные множества

Приведенные примеры позволяют разбить все множества на две большие группы: конечные и бесконечные. Все элементы *конечного* множества можно перечислить, тогда как элементы *бесконечного* множества даже теоретически нельзя собрать в законченную совокупность.

Запись $x_1, x_2, \dots, x_n \in A$ означает, что все $x_1, x_2, \dots, x_n$ принадлежат множеству A. Конечное множество, состоящее из элементов $x_1, x_2, \dots, x_n$, обычно обозначают $\{x_1, x_2, \dots, x_n\}$. В частности, $\{x\}$ – одноэлементное множество.

Однако перечисление элементов множества в фигурных скобках слишком громоздко для задания больших множеств и совсем неприменимо для бесконечных множеств. Как же задавать такие множества? Эта проблема решается с помощью *характеристического свойства* множества. Пусть $P(x)$ – некоторое предложение, зависящее от x. Например, «x делится на 5», «в книге x встречается буква “ять”», «x любит Иванова», «$x^2 = -1$». Если на место x подставить любой конкретный предмет, мы получим истинное или ложное утверждение. Если, например, $P(x)$ = «x делится на 5», то $P(10)$ – истинное утверждение, а $P(7)$ – ложное.

Такое предложение и называется характеристическим свойством множества. С его помощью можно описывать какие угодно множества в удобном и компактном виде. Запись $A = \{x \mid P(x)\}$ означает, что $a \in A$ тогда и только тогда, когда $P(a)$ – истинное утверждение. Например, множество $A = \{x \mid x$ есть точка плоскости и x находится на расстоянии 1 от начала координат$\}$ есть единичная окружность с центром в точке $(0, 0)$. Это бесконечное множество, но для любого предмета можно точно сказать, принадлежит он этому множеству или нет.

Иногда бывает удобно указать, из какого класса выбираются элементы множества. Тогда пишут $\{x \in A \mid P(x)\}$. Например, $\{x \in \boldsymbol{R} \mid 0 \leqslant x \leqslant 2\}$ есть бесконечное множество действительных чисел, лежащих между 0 и 2 (включительно), а множество $\{x \in \boldsymbol{Z} \mid 0 \leqslant x \leqslant 2\}$ конечно и состоит из трех целых чисел 0, 1, 2. Если, скажем, координатную плоскость обозначить через $\boldsymbol{R}^2$, то описанную выше единичную окружность можно описать короче: $\{(x, y) \in \boldsymbol{R}^2 \mid x^2 + y^2 = 1\}$.

С помощью характеристического свойства удобно задать *пустое множество*, т.е. множество, не содержащее элементов. Например, множество $\{x \in \boldsymbol{R} \mid x^2 < 0\}$ не имеет элементов, т.е. пусто. Точно так же пусто и множество $\{x$ – рыбы $\mid x$ дышит легкими$\}$. Независимо от способа описания, получается одно и то же множество без элементов. Его обозначают знаком $\varnothing$.

1.3. Равенство множеств

Тот факт, что множество определяется своими элементами, можно сформулировать в виде следующего принципа, в котором вводится очень важное понятие равенства множеств.

> Два множества *равны* в том и только в том случае, когда они состоят из одних и тех же элементов.

В самом деле, два множества можно описать совершенно по-разному. Тем не менее при внимательном рассмотрении можно убедиться, что эти множества состоят из одних и тех же элементов. Тогда данные множества равны. Это обозначается так: $X = Y$. В противном случае (когда в одном из множеств найдется элемент, не принадлежащий другому) пишут $X \neq Y$. Пусть, например, A – множество студентов-психологов 2-го курса, а B – множество студентов-психологов, поступивших в университет в прошлом году. Ясно, что эти множества состоят из одних и тех же людей (если, конечно, никто из них не отсеялся), т.е. $A = B$.

Доказательство равенства каких-либо множеств A и B состоит, таким образом, из двух частей:

1) если $x \in A$, то $x \in B$;

2) если $x \in B$, то $x \in A$.

Пусть, например, A – множество всех положительных четных чисел, B – множество положительных целых чисел, представимых в виде суммы двух положительных нечетных чисел. Докажем, что $A = B$.

1. Пусть $x \in A$, т.е. x – положительное четное число. Это значит, что $x = 2k$ при некотором натуральном k. Продолжим равенство: $x = (2k - 1) + 1$, т.е. число x представимо в виде суммы двух положительных нечетных чисел $2k - 1$ и 1. Это как раз и означает, что $x \in B$.

2. Пусть $x \in B$, т.е. $x = m_1 + m_2$, где m_1 и m_2 – положительные нечетные числа. Значит, найдутся натуральные k_1, k_2, такие, что $m_1 = 2k_1 - 1$, $m_2 = 2k_2 - 1$. Таким образом, $x = 2k_1 - 1 + 2k_2 - 1 = 2(k_1 + k_2 - 1) = 2k$, где $k = k_1 + k_2 - 1$ – натуральное число. Значит, x есть положительное четное число и $x \in A$. Равенство доказано.

Приведем еще примеры равенства и неравенства множеств.

Так, $\{2, 4, 6\} = \{2, 6, 4\} = \{2, 4, 4, 6\}$, поскольку эти три множества состоят из одних и тех же элементов.

Множества $\{\{1, 2\}\}$ и $\{1, 2\}$ не равны: первое множество одноэлементно и его элементом является множество; второе, двухэлементное, состоит из чисел. Таким образом, данные множества состоят из элементов различной природы и не могут быть равны.

1.4. Подмножества

Говорят, что множество A есть ***подмножество*** множества B, если каждый элемент A является элементом B. В этом случае также говорят, что множество A *включено* во множество B или множество B *включает* множество A. Это обозначается $A \subseteq B$. Для доказательства включения требуется проверить утверждение: если $x \in A$, то $x \in B$.

Примеры.

1. Пусть A – множество красных яблок, а B – множество всех яблок. Тогда $A \subseteq B$: ведь красное яблоко – это и просто яблоко, поэтому если $x \in A$, то $x \in B$.

2. Множество $\{1, 2\}$ есть подмножество множества $\{1, 2, 3\}$.

3. Множество студентов-психологов 1-го курса есть подмножество множества студентов-психологов, которое, в свою очередь, включено во множество всех студентов университета.

4. Верхняя полуплоскость $\{(x, y) \in \boldsymbol{R}^2 \mid y \geqslant 0\}$ – подмножество всей координатной плоскости.

5. Множество $\{1, 2\}$ не является подмножеством множества $\{\{1\}, 2, 3\}$, так как число 1 не принадлежит последнему.

Сделаем ряд наблюдений о свойствах включения (наблюдения будем обозначать значком ∇ – «набла»).

∇1. *Каждое множество есть подмножество самого себя:* $A \subseteq A$.

∇2. *Если* $A \subseteq B$ *и* $B \subseteq C$, *то* $A \subseteq C$.

На примере этих двух свойств подчеркнем различие между принадлежностью $\in$ и включением $\subseteq$. Аналоги этих свойств включения для принадлежности неверны. Множество может быть собственным элементом (например, каталог всех каталогов есть каталог и принадлежит самому себе), однако это довольно экзотическая ситуация. В то же время множество всегда включено в само себя. Не выполняется для принадлежности и второе свойство (называемое *транзитивностью*). Например, $1 \in \boldsymbol{Z}$ и $\boldsymbol{Z} \in \{\boldsymbol{Z}\}$. Однако $1 \notin \{\boldsymbol{Z}\}$, так как единственный элемент множества $\{\boldsymbol{Z}\}$ – это множество $\boldsymbol{Z}$.

∇3. *Если* $A \subseteq B$ *и* $B \subseteq A$, *то* $A = B$.

Это прямо следует из определений включения и равенства множеств. Таким образом, тот факт, что два множества равны, означает, что каждое из них есть подмножество другого.

Следующее утверждение примем без доказательства.

∇4. *Пустое множество есть подмножество любого множества.*

Суммируя ∇1 и ∇4, получаем

∇5. *Каждое множество* $A \neq \varnothing$ *имеет по крайней мере два различных подмножества: само* A *и пустое множество.*

∇6. *Каждый элемент множества* A *определяет некоторое подмножество множества* A: *если* $a \in A$, *то* $\{a\} \subseteq A$.

Говорят, что множество A *строго включено* в B, если $A \subseteq B$ и $A \neq B$. В этом случае говорят также, что B *строго включает* A или A есть *истинное подмножество* B. Это обозначается $A \subset B$.

∇**7a.** *Если $A \subseteq B$ и $B \subset C$, то $A \subset C$.*

∇**7b.** *Если $A \subset B$ и $B \subseteq C$, то $A \subset C$.*

∇**7c.** *Если $A \subset B$ и $B \subset C$, то $A \subset C$.*

Множество всех подмножеств множества A называется *множеством-степенью* множества A и обозначается через $\mathcal{P}(A)$:

$$\mathcal{P}(A) = \{B \mid B \subseteq A\}.$$

Пусть, например, $A = \{1, 2, 3\}$. Тогда множество-степень состоит из множества A, пустого множества, трех одноэлементных и трех двухэлементных подмножеств множества A:

$$\mathcal{P}(A) = \{A, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \varnothing\}.$$

Можно убедиться, что множество-степень конечного n-элементного множества состоит из 2^n элементов. Этим и объясняется происхождение термина «множество-степень». В приведенном выше примере $n = 3$ и $\mathcal{P}(A)$ содержит $2^3 = 8$ множеств.

Чтобы еще раз подчеркнуть различие и связь между принадлежностью и включением, сформулируем очевидное наблюдение:

∇**8.** *Если $B \subseteq A$, то $B \in \mathcal{P}(A)$; если $a \in A$, то $\{a\} \subseteq A$ и $\{a\} \in \mathcal{P}(A)$.*

1.5. Операции над множествами

Любая математическая дисциплина должна включать не только исходные, неопределяемые, понятия, но и «правила игры», способы работы с этими объектами. Например, числа в арифметике можно складывать и умножать: говорят, что заданы операции сложения и умножения. Ниже вводятся аналогичные операции для множеств.

Объединение множеств A и B есть множество, состоящее из элементов множества A или множества B. Объединение множеств обозначается $A \cup B$. Другими словами,

$$A \cup B = \{x \mid x \in A \text{ или } x \in B\}.$$

Здесь необходимо пояснить смысл союза «или». Дело в том, что в русском языке этот союз имеет два значения. Первое значение: или – или, одно из двух, пан или пропал. Такое «или» называется и с к л ю ч а ю щ и м. В определении объединения множеств, однако, подразумевается второе, н е и с к л ю ч а ю щ е е, значение союза «или». Таким образом, если элемент x принадлежит объединению $A \cup B$, то он может принадлежать только множеству A, только множеству B, а может – одновременно обоим этим множествам. Итак, $x \in A \cup B$ тогда и только тогда, когда x есть элемент хотя бы одного из этих множеств. Например, $\{1, 2, 3\} \cup \{1, 3, 4\} = \{1, 2, 3, 4\}$.

Число 2 принадлежит только первому множеству, число 4 – только второму, а числа 1, 3 – обоим множествам сразу. Операция, которая двум множествам ставит в соответствие их объединение, также называется *объединением* (при сложении чисел, к примеру, применяются разные слова: операция – сложение, ее результат – сумма).

Пересечение множеств A и B есть множество, состоящее из элементов, общих для обоих множеств. Этим же словом называют и соответствующую операцию. Пересечение множеств обозначается $A \cap B$:

$$A \cap B = \{x \mid x \in A \text{ и } x \in B\}.$$

Например, $\{1, 2, 3\} \cap \{1, 3, 4\} = \{1, 3\}$.

Операции объединения и пересечения можно рассматривать как способ образования новых множеств из существующих. Из определений пересечения и объединения легко выводится следующее свойство.

∇**1.** *Для всякой пары множеств A и B имеет место включение*

$$A \cap B \subseteq A \subseteq A \cup B.$$

Два множества называются *непересекающимися* (или *расчлененными*), если $A \cap B = \varnothing$, и *пересекающимися*, если $A \cap B \neq \varnothing$. Система множеств называется *расчлененной*, если любая пара ее элементов является непересекающейся.

Разбиением множества X называется такая расчлененная система $\mathcal{U}$ непустых подмножеств множества X, что каждый элемент X является элементом некоторого (и, значит, единственного) множества системы $\mathcal{U}$. Например, $\mathcal{U} = \{\{1, 2\}, \{3\}, \{4, 5\}\}$ есть разбиение множества $X = \{1, 2, 3, 4, 5\}$.

Следующая операция позволяет образовать новое множество из одного существующего множества. Обычно в ходе какого-либо рассуждения можно выделить такое множество, что все рассматриваемые предметы являются его элементами. Под рассуждением может пониматься и научная теория, и целая книга. Такое широкое множество называется *универсальным* (для данного рассуждения). Например, в зоологии универсальным множеством является вся фауна, в элементарной арифметике – множество целых чисел $\mathbf{Z}$ и т.д. Обычно универсальное множество обозначают U. ***Дополнением*** множества A называется множество $\overline{A}$, состоящее из элементов универсального множества U, не являющихся элементами множества A:

$$\overline{A} = \{x \in U \mid x \notin A\}.$$

Пусть, например, U – множество всех студентов университета, A – множество студентов факультета психологии. Тогда $\overline{A}$ есть множество студентов всех факультетов, кроме психологического.

Для графической иллюстрации операций над подмножествами некоторого универсального множества U используют так называемые диаграммы Венна (их еще называют кругами Эйлера). На диаграммах универсальное множество изображается в виде прямоугольника, а его подмножества – в виде кругов. На рис. 1.1, *а–в* заштрихованные области изображают соответственно множества $A \cup B$, $A \cap B$ и $\overline{A}$.

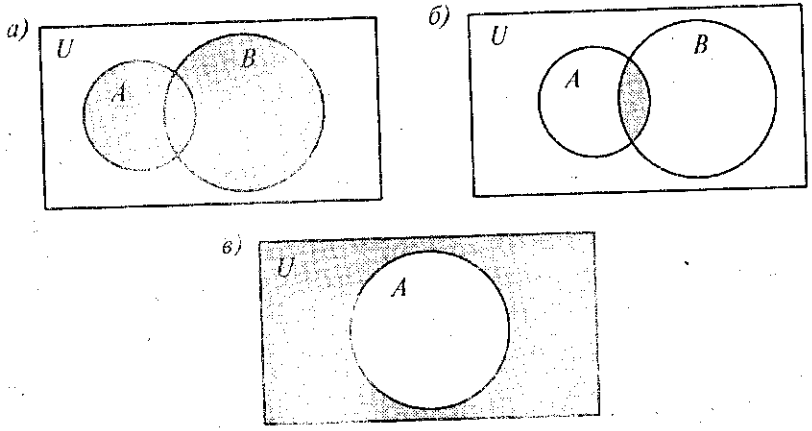


Рис. 1.1

Леонард Эйлер (1707–1783). Великий математик, физик и астроном, академик Петербургской АН (в 1726–1741 и с 1766). Родился в Базеле. Его отец, пастор, был учеником великого швейцарского математика Якоба Бернулли и защитил диссертацию по математике. Первые математические сведения Эйлер получил от отца. В 1726 г. Эйлер был приглашен в Петербургскую АН и в мае 1727 г. прибыл в Петербург, где стал адъюнктом физиологии, позднее – математики. В 1731–1741 гг. – профессор математики. В 1741 г. переехал в Берлин, где прожил 25 лет. Возглавлял Математический класс Берлинской АН. В 1766 г. возвратился в Петербург. Вскоре почти полностью потерял зрение.

Научные интересы Эйлера относились ко всем основным областям естествознания, к которым можно было применить математические методы. В 1727–1741 гг. опубликовал 50 трудов в самых разнообразных областях математики и физики – от теории чисел до гидродинамики. В 1736 г. вышел в свет его трактат по механике, в котором он впервые исследовал динамику точки с помощью математического анализа.

Берлинский период жизни Эйлера был особенно продуктивным. Свои труды он печатал в Берлине и Петербурге. Опубликовал серию работ по астрономии. Его теоретические изыскания послужили основанием для составления таблиц движения Луны. Заложил основы математической физики, механики твердого тела, написал блестящие книги по дифференциальному и интегральному исчислению. Книги Эйлера служили основным руководством для математиков.

По возвращении в Петербург Эйлер продолжал интенсивно работать, подготовив за 17 лет около 400 научных работ. Он диктовал своим ученикам мемуары буквально по всем отраслям математики и механики, издал ряд монографий по теории чисел, навигации, занимался философией («Письма к одной немецкой принцессе»). Список трудов Эйлера содержит около 850 названий, в их числе ряд многотомных монографий. С 1909 г. в Швейцарии издается Полное собрание его сочинений, рассчитанное на 72 тома.

Леонард Эйлер – иностранный почетный член Петербургской АН, член Парижской АН, Берлинской АН, Лондонского королевского общества и многих других академий и научных обществ.

1.6. Алгебраические свойства операций над множествами

Рассмотрим подробнее свойства операций над множествами и связи между ними. Эти свойства, как мы увидим далее, во многом аналогичны знакомым по школе свойствам обычных операций сложения и умножения чисел. Свойства записываются в виде тождеств, т.е. не зависят от того, каково универсальное множество U и какие именно конкретные его подмножества в них фигурируют. В следующей теореме формулируются основные свойства объединения и пересечения.

Теорема 1.1. *Для любых подмножеств A, B и C универсального множества U справедливы следующие тождества*:

1. $A \cup (B \cup C) = (A \cup B) \cup C$.	1′. $A \cap (B \cap C) = (A \cap B) \cap C$.
2. $A \cup B = B \cup A$.	2′. $A \cap B = B \cap A$.
3. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.	3′. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
4. $A \cup \varnothing = A$.	4′. $A \cap U = A$.
5. $A \cup \overline{A} = U$.	5′. $A \cap \overline{A} = \varnothing$.

Справедливость каждого из этих тождеств можно проверить, показав, что множество, стоящее по одну сторону тождества, включено во множество, стоящее по другую сторону. Докажем, например, тождество 3.

Пусть $x \in A \cup (B \cap C)$. Тогда, по определению объединения, $x \in A$ или $x \in B \cap C$. Если $x \in A$, то x принадлежит объединению A с любым множеством. В частности, $x \in A \cup B$ и $x \in A \cup C$. Значит, x принадлежит пересечению этих множеств, т.е. $x \in (A \cup B) \cap (A \cup C)$. Если $x \in B \cap C$, то, по определению пересечения, $x \in B$ и $x \in C$, а значит, $x \in A \cup B$ и $x \in A \cup C$, т.е. снова $x \in (A \cup B) \cap (A \cup C)$. Итак, левая часть включена в правую. Обратное включение доказывается аналогично.

Тождества 1 и 1′ называются *ассоциативными законами* соответственно для объединения и пересечения, тождества 2 и 2′ – *коммутативными законами* для этих операций. Тождества 3 и 3′ – *дистрибутивные законы*.

Если операции $\cup$ объединения множеств поставить в соответствие операцию + сложения чисел, операции $\cap$ – операцию умножения чисел $\cdot$, универсальному множеству U – единицу 1, а пустому множеству $\varnothing$ – нуль 0, возникает почти полная аналогия между множествами и числами. В самом деле, справедливы ассоциативные и коммутативные законы:

$$a + (b + c) = (a + b) + c,$$
$$a \cdot (b \cdot c) = (a \cdot b) \cdot c,$$
$$a + b = b + a,$$
$$a \cdot b = b \cdot a.$$

В школьном курсе эти законы называются сочетательным и переместительным. Кроме того, выполняется и дистрибутивный (распределительный) закон вида 3′:

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

Аналогия нарушается лишь для дистрибутивного закона 3, так как для чисел $a + b \cdot c \neq (a + b) \cdot (a + c)$.

Десять свойств, сформулированных в теореме 1.1, являются фундаментальными в том смысле, что все остальные свойства операций над множествами непосредственно следуют из них. При доказательстве новых свойств мы можем забыть о принадлежности (если $x \in A$, то $x \in B$, и наоборот) и пользоваться только законами 1–5, 1′–5′.

Свойства в теореме 1.1 фигурируют попарно таким образом, что каждое получается из соседнего заменой $\cup$ на $\cap$, U на $\varnothing$ и наоборот. Такие выражения называются *двойственными* друг к другу. Поскольку любое свойство множеств (тождество, теорема) выводится из попарно двойственных свойств 1–5, 1′–5′, справедливо следующее утверждение.

Принцип двойственности. *Для любого тождества множеств двойственное ему выражение также является тождеством.*

В следующей теореме формулируются дополнительные свойства операций над множествами.

Теорема 1.2. *Для любых подмножеств A и B универсального множества U справедливы следующие утверждения:*

6. *Если для всех A имеет место $A \cup B = A$, то $B = \varnothing$.*

6′. *Если для всех A имеет место $A \cap B = A$, то $B = U$.*

7, 7′. *Если $A \cup B = U$ и $A \cap B = \varnothing$, то $B = \overline{A}$.*

8, 8′. $\overline{\overline{A}} = A$.

9. $\overline{\varnothing} = U$.	9′. $\overline{U} = \varnothing$.
10. $A \cup A = A$.	10′. $A \cap A = A$.
11. $A \cup U = U$.	11′. $A \cap \varnothing = \varnothing$.
12. $A \cup (A \cap B) = A$.	12′. $A \cap (A \cup B) = A$.
13. $\overline{A \cup B} = \overline{A} \cap \overline{B}$.	13′. $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

Тождества 12, 12′ называются *законами поглощения*; 13, 13′ – *законами де Моргана*. Законы 7, 7′ и 8, 8′ не меняются при преобразовании в двойственные. Такие законы называются *самодвойственными*.

Теорему 1.2 можно доказать, опираясь только на теорему 1.1. Докажем для примера свойство 7:

B = (по свойству 4) = $B \cup \varnothing$ = (по свойству 5′) = $B \cup (A \cap \overline{A})$ = = (по дистрибутивному закону 3) = $(B \cup A) \cap (B \cup \overline{A})$ = (по коммутативному закону 1 и условию) = $U \cap (B \cup \overline{A})$ = (по свойству 5) = = $(A \cup \overline{A}) \cap (B \cup \overline{A})$ = (по дистрибутивному закону 3) = $\overline{A} \cup (A \cap B)$ = = (по условию) = $\overline{A} \cup \varnothing$ = (по свойству 4) = $\overline{A}$.

Итак, $B = \overline{A}$, что и требовалось доказать.

Остальные утверждения теоремы доказываются аналогично.

Множество – пожалуй, самое универсальное математическое понятие. Все разделы математики так или иначе занимаются множествами: математический анализ – числовыми множествами, геометрия – множествами точек, теория вероятностей – множествами случайных событий и т.д. Знакомство с основами теории множеств – необходимый этап освоения математики.

Задачи

1. Верно ли, что $\{1, 2\} \in \{\{1, 2, 3\}, \{1, 3\}, 1, 2\}$?
2. Описать словесно множества:

а) $\{x \in Z \mid x$ делится на 2 и x делится на 3$\}$;

2 Зак. № 2435 Турецкий

б) $\{(x, y) \in \boldsymbol{R}^2 \mid x^2 + y^2 = 1\}$;

в) $\{(x, y) \in \boldsymbol{R}^2 \mid y = 2x$ и $y = 3x\}$.

3. Привести пример таких множеств A, B, C, что $A \in B$, $B \in C$, но $A \notin C$.

4. Доказать следующие утверждения, пользуясь свойствами чисел:

а) $\{x \in \boldsymbol{Z} \mid$ для некоторого целого числа y $x = 6y\} = \{x \in \boldsymbol{Z} \mid$ для некоторых чисел u и v $x = 2u$ и $x = 3v\}$;

б) $\{x \in \boldsymbol{R} \mid$ для некоторого $y \in \boldsymbol{R}$ $x = y^2\} = \{x \in \boldsymbol{R} \mid x \geqslant 0\}$;

в) $\{x \in \boldsymbol{Z} \mid$ для некоторого целого числа y $x = 6y\} \subseteq \{x \in \boldsymbol{Z} \mid$ для некоторого целого числа y $x = 2y\}$.

5. Пусть A – множество целых чисел, кратных 2; B – множество целых чисел, кратных 3; U – множество целых чисел. Описать множества: $A \cup B$, $A \cap B$, $\overline{A} \cup \overline{B}$, $\overline{A} \cap \overline{B}$.

6. Пусть A – множество всех прямоугольных треугольников на плоскости; B – множество всех равносторонних треугольников; U – множество всех треугольников. Какие треугольники содержатся во множествах $A \cup B$, $A \cap B$, $\overline{A} \cap B$, $A \cap \overline{B}$, $\overline{A} \cup B$, $A \cup \overline{B}$?

7. Перечислить все элементы множества-степени $\mathcal{P}(A)$:

а) $A = \{\{1, 2\}, \{3\}, 1\}$;

б) $A = \{1, \{1\}\}$;

в) $A = \varnothing$;

г) $A = \{\varnothing\}$.

8. Доказать:

а) $A \cup (\overline{A} \cap B) = A \cup B$;

б) $\overline{\overline{A} \cup B} \cup B = A \cup B$;

в) $(A \cap B) \cup (A \cap \overline{B}) = A$;

г) $(\overline{A} \cup \overline{B}) \cap C = \overline{(A \cap B) \cup \overline{C}}$;

д) $(A \cup B) \cap (\overline{A \cap B}) = (A \cap \overline{B}) \cup (B \cap \overline{A})$.

9. Упростить выражения:

а) $\overline{A \cap \overline{B}} \cup B$;

б) $(A \cap B \cap C) \cup (\overline{A} \cap B \cap C) \cup \overline{B} \cup \overline{C}$;

в) $\overline{(A \cap B) \cap (\overline{A} \cup \overline{B})}$;

г) $(A \cap B \cap C \cap \overline{D}) \cup (\overline{A} \cap C) \cup (\overline{B} \cap C) \cup (C \cap D)$.

10. На диаграмме Венна для подмножеств A, B, C универсального множества U (рис. 1.2) прямоугольник, изображающий U, разбивается на 8 неперекрывающихся областей.

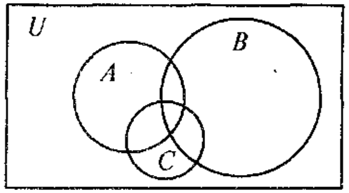


Рис. 1.2

Описать каждую из этих областей с помощью операций объединения, пересечения и дополнения.

Глава 2
ОТНОШЕНИЯ И ФУНКЦИИ

2.1. Понятие отношения

В математике для обозначения какой-либо связи между предметами или понятиями часто пользуются термином «*отношение*». Например, следующие предложения описывают некоторые отношения между предметами:

x меньше, чем y;

x делится на y;

A включено в B;

x является сыном y;

x является начальником y.

В данной главе понятие «отношение» рассматривается в рамках теории множеств.

Из примеров видно, что отношение используется для пар объектов, рассматриваемых в определенном порядке. Пусть, например, $x = 1$, $y = 2$. Тогда пара (x, y) находится в отношении «меньше, чем», а пара (y, x) – нет.

Назовем *упорядоченной парой* (x, y) двухэлементное множество $\{x, y\}$, в котором элемент x находится на первом месте, а y – на втором. Элемент x называется *первой координатой* упорядоченной пары, а элемент y – *второй координатой*. Таким образом, две упорядоченные пары равны, когда совпадают их координаты:

$(x, y) = (u, v)$ тогда и только тогда, когда $x = u$ и $y = v$.

В этом основное отличие упорядоченной пары от двухэлементного множества. В самом деле, множества $\{1, 2\}$ и $\{2, 1\}$ равны, так как состоят из одних и тех же элементов. При этом упорядоченные пары $(1, 2)$ и $(2, 1)$ не равны, так как их координаты не совпадают.

О любой упорядоченной паре можно точно сказать, находится она в данном отношении или нет. Поэтому, чтобы определить отношение, достаточно перечислить все пары, которые находятся в данном отношении. Можно дать следующее определение.

Пусть заданы множества X и Y. Тогда ***отношение*** ρ из X в Y есть некоторое множество упорядоченных пар (x, y), где $x \in X$, $y \in Y$. Поскольку отношение связывает два объекта, его называют *бинарным*. Если $(x, y) \in \rho$, где ρ — некоторое множество упорядоченных пар, говорят, что элемент x находится в отношении ρ с элементом y. Это можно также обозначать $x\rho y$.

Множество всевозможных упорядоченных пар (x, y), таких, что $x \in X$, $y \in Y$, называется ***декартовым произведением*** множеств X и Y и обозначается $X \times Y$. Таким образом, отношение есть подмножество декартова произведения: $\rho \subseteq X \times Y$. Если $X = Y$ и, стало быть, $\rho \subseteq X \times X$, говорят, что отношение ρ есть отношение на множестве X.

Ознакомимся с четырьмя основными свойствами отношений.

Отношение ρ называется *рефлексивным*, если каждый элемент $x \in X$ находится в этом отношении сам с собой: $x\rho x$ для всех $x \in X$.

Примеры.

1. Отношение $\leqslant$ рефлексивно, так как $x \leqslant x$ для любого числа x.
2. Отношение $<$ не является рефлексивным.
3. Отношение делимости «x делится нацело на y» рефлексивно на множествах целых чисел, не содержащих нуля, так как каждое целое число, не равное нулю, делится само на себя.

Отношение ρ называется *симметричным*, если из того, что $x\rho y$, следует, что $y\rho x$.

Примеры.

1. Отношение подобия треугольников симметрично: если треугольник x подобен треугольнику y, то треугольник y подобен треугольнику x.
2. Отношение параллельности прямых симметрично.
3. Отношение делимости не является симметричным: из того, что x делится на y, не следует, что y делится на x.

Отношение ρ называется *транзитивным*, если из того, что $x\rho y$ и $y\rho z$, следует, что $x\rho z$.

Примеры.

1. Отношение $\leqslant$ транзитивно: если $x \leqslant y$ и $y \leqslant z$, то $x \leqslant z$.
2. Отношение $<$ транзитивно.
3. Отношение «быть отцом» не является транзитивным: если x — отец y и y — отец z, то x — не отец z, а дедушка.

Отношение ρ называется *антисимметричным*, если из того, что $x\rho y$ и $y\rho x$, следует, что $x = y$.

Примеры.

1. Отношение делимости антисимметрично на множестве натуральных чисел: если x делится на y и y делится на x, то $x = y$.
2. Отношение $\leqslant$ антисимметрично: если $x \leqslant y$ и $y \leqslant x$, то $x = y$.
3. Отношение перпендикулярности прямых не антисимметрично: если прямая x перпендикулярна прямой y и наоборот, это не значит, что прямые совпадают.

Пусть задано отношение ρ на множестве $X \subseteq \boldsymbol{R}$, т.е. отношение между числами. Удобное графическое представление отношения — график на координатной плоскости. Каждой упорядоченной паре отношения $(x, y) \in \rho$ соответствует точка на плоскости с координатами x, y. Если нанести все эти точки на плоскость, получится график отношения ρ.

Для рефлексивного отношения график будет содержать все точки вида (x, x), у которых абсцисса и ордината совпадают. Такие точки лежат на биссектрисе I и III квадрантов. Итак, график рефлексивного отношения обязательно содержит точки этой биссектрисы.

Если отношение симметрично, то вместе с каждой точкой (x, y) на графике будет присутствовать точка (y, x). Такой график симметричен относительно биссектрисы I и III квадрантов.

Пусть $X = \{1, 2, 3, 4\}$ и $(x, y) \in \rho$, если $|x - y| \leqslant 2$, т.е. расстояние между числами x и y не превосходит 2. Легко выписать все пары, находящиеся в данном отношении: $\rho = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (2, 4), (3, 1), (3, 2), (3, 3), (3, 4), (4, 2), (4, 3), (4, 4)\}$.

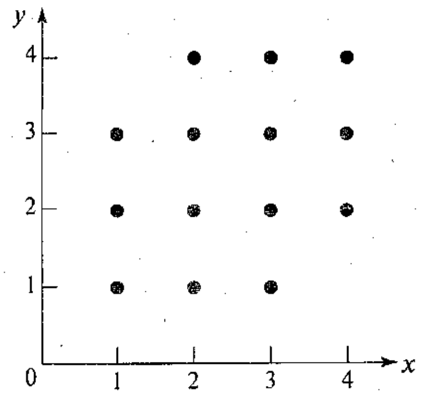


Рис. 2.1

Заметим, кстати, что данное отношение является рефлексивным и симметричным. Значит, его график симметричен относительно биссектрисы I и III квадрантов и содержит точки, лежащие на ней (рис. 2.1).

2.2. Отношение эквивалентности

Отношение на некотором множестве X называется ***отношением эквивалентности***, если оно рефлексивно, симметрично и транзитивно.

Примеры отношений эквивалентности.

1. Отношение подобия на множестве всех треугольников на плоскости.
2. Отношение параллельности на множестве прямых.
3. Отношение «быть одного цвета» на множестве всех предметов.
4. Отношение равенства на произвольной системе множеств.
5. Отношение «проживания в одном доме» на множестве жителей России.

Последний пример иллюстрирует основное свойство любого отношения эквивалентности: с помощью этого отношения можно разбить соответствующее множество на непересекающиеся подмножества. При этом все элементы каждого подмножества находятся между собой в отношении эквивалентности (эквивалентны друг другу) и не находятся в этом отношении ни с одним из элементов другого подмножества. В данном случае все население России разбивается на множества людей, живущих в одном доме.

Пусть ρ — отношение эквивалентности в X. ***Классом эквивалентности*** (ρ-классом эквивалентности), порожденным элементом $x \in X$, называется множество всех элементов y, для которых $x\rho y$. Это множество обозначают $[x]$.

Пример.

Пусть $X = \{1, 2, 3, 4, 5, 6, 7\}$, $x\rho y$ тогда и только тогда, когда $x - y$ делится на 3 (такое отношение называется *отношением сравнимости по модулю* 3).

Легко проверить, что ρ — отношение эквивалентности. Класс эквивалентности, порожденный единицей, состоит из чисел 1, 4, 7: $1 - 1 = 0$, $1 - 4 = -3$, $1 - 7 = -6$, и все эти разности делятся на 3. Ясно, что этот же класс порождается четверкой и семеркой:

$$[1] = [4] = [7] = \{1, 4, 7\}.$$

Кроме того,

$$[2] = [5] = \{2, 5\},$$
$$[3] = [6] = \{3, 6\}.$$

Числа из каждого класса эквивалентности при делении на 3 дают одинаковый остаток. Говорят, что они сравнимы по модулю 3. Числа из разных классов не сравнимы между собой.

Из рассмотренного примера хорошо видны два основных свойства классов эквивалентности: во-первых, каждый элемент множества X попадает в один из классов эквивалентности, и, во-вторых, классы не пересекаются между собой. Математически это формулируется следующим образом.

Теорема 2.1. *Пусть ρ – отношение эквивалентности на X. Тогда система ρ-классов эквивалентности есть разбиение множества X.*

Вспомним, что разбиением множества называется такая система его непересекающихся подмножеств, в которой каждый элемент принадлежит одному из этих подмножеств.

В рассмотренном выше примере с отношением равенства по модулю 3 видно, что классы эквивалентности образуют разбиение множества X.

Наглядный пример разбиения дает отношение «быть одного цвета». Пусть, например, $X = \{$лягушка, уголь, елка, апельсин, крокодил, «Фанта», сажа$\}$. Тогда в первый класс эквивалентности попадут все зеленые предметы (лягушка, елка, крокодил), во второй – все оранжевые (апельсин и «Фанта»), а в третий – все черные (уголь и сажа). Все предметы разбиваются, таким образом, на классы предметов одного цвета.

Итак, отношение эквивалентности выделяет некоторое свойство предметов (например, цвет, знак числа и т.д.), по которому элементы, принадлежащие одному классу эквивалентности, неразличимы между собой. Эти элементы эквивалентны.

2.3. Отношение частичного порядка

Отношение на некотором множестве X называется ***отношением частичного порядка***, если оно рефлексивно, антисимметрично и транзитивно.

Примеры отношений частичного порядка.

1. Отношение делимости «x делится нацело на y» на множестве натуральных чисел.

2. Отношение $\subseteq$, заданное на произвольной системе множеств.

3. Отношение $\leqslant$ на числовых множествах.

Отношение частичного порядка упорядочивает элементы множества X, позволяет расположить их «по возрастанию». Поэтому произвольное отношение частичного порядка также принято обозначать $\leqslant$. Множество X с заданным отношением частичного порядка называют *частично упорядоченным множеством.*

Отношение $\leqslant$ порождает отношение $<$:

$$x < y, \text{ если } x \leqslant y \text{ и } x \neq y.$$

В этом случае говорят, что элемент x *предшествует* элементу y (x меньше y, y больше x). Элемент y *покрывает* элемент x, если $x < y$ и не существует такого элемента u, что $x < u < y$.

Пусть, например, на множестве целых чисел $X = \{1, 2, 3\}$ задано обычное отношение $\leqslant$. В этом случае отношение $<$ есть также обычное числовое отношение «меньше». Тогда двойка покрывает единицу, а тройка – нет, так как $1 < 2 < 3$. Если же отношение $\leqslant$ задано на множестве действительных чисел $X = [1, 3]$, то мы не найдем элемента, покрывающего единицу (или любое другое число), так как между двумя действительными числами всегда существует действительное число, не равное им.

Конечное частично упорядоченное множество удобно изображать в виде д и а г р а м м ы:

- элементы множества изображаются точками;
- точка y ставится выше точки x, если $x < y$;
- если y покрывает x, то точки соединяются отрезком.

Такая диаграмма наглядно показывает, как упорядочено данное множество, и отражает иерархию его элементов – кто кому «подчиняется», кто над кем «начальник».

Примеры.

1. Отношение $\leqslant$ на $X = \{1, 2, 3, 4, 5, 6, 7\}$ (рис. 2.2).
2. Отношение делимости на том же множестве (рис. 2.3).

Рис. 2.2

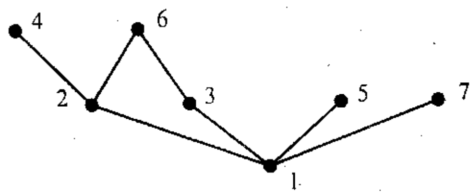


Рис. 2.3

3. Отношение $\subseteq$ на множестве $X = \mathcal{P}(A)$, где $A = \{1, 2\}$ (рис. 2.4).
4. Отношение делимости на множестве $X = \{1, 2, 3, 6\}$ (рис. 2.5).

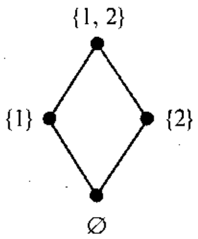


Рис. 2.4

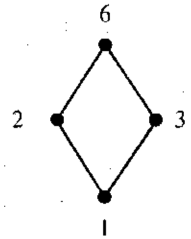


Рис. 2.5

Итак, одно и то же множество, частично упорядоченное разными отношениями, можно изобразить разными диаграммами (см. рис. 2.2 и 2.3). Наоборот, различным частично упорядоченным множествам могут соответствовать одинаковые диаграммы (см. рис. 2.4 и 2.5).

2.4. Функции

Между элементами различных множеств можно установить более тесную связь, так называемую *зависимость*. Самый простой пример зависимости – таблица из двух колонок. Пусть левая колонка содержит фамилии студентов группы, а правая – оценку за экзамен по математике:

Иванов	5
Петров	3
Сидоров	2
Алексеев	4

Такая таблица устанавливает зависимость между студентом и его успеваемостью. Если X – множество студентов из одной группы, а Y – множество целых чисел $\{2, 3, 4, 5\}$, то с помощью таблицы устанавливается соответствие между элементами множества X (студентами) и множества Y (целыми числами). Данная зависимость обладает одним важным свойством: каждый студент имеет только одну оценку по математике. Другими словами, каждому элементу множества X соответствует единственный элемент множества Y.

Рассмотренный пример позволяет ввести математическое понятие функции, которое используется для описания таких зависимостей. Пусть заданы множества X и Y. ***Функцией*** f из X в Y называется правило, с помощью которого каждому элементу $x \in X$ ставится в соответствие е д и н с т в е н н ы й элемент $y \in Y$ (рис. 2.6). Функцию можно называть также ***отображением***.

Функциональная зависимость между множествами обозначается следующим образом:

$$f\colon X \to Y.$$

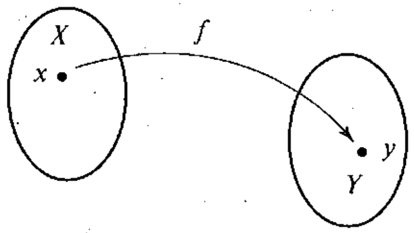


Рис. 2.6

Множество X называется *областью определения* функции f, а множество Y – *областью значений*. Область определения и область значений функции f можно обозначать соответственно D_f и E_f.

Если ясно, какие множества X и Y имеются в виду, пишут просто $y = f(x)$. Элемент $x \in X$ называется *аргументом*, а $y \in Y$ – *значением функции* f. Их можно называть и по-другому: y – *образ элемента* x; x – *прообраз* элемента y.

Примеры.

1. X – множество студентов, Y – множество парт в аудитории. Правило f ставит в соответствие каждому студенту парту, за которой он сидит. Это функция, так как студент не может сидеть за двумя партами сразу. Аргументы функции f – студенты, значения – парты. Парта – образ студента, студент – прообраз парты.

2. Пусть теперь наоборот: X – множество парт, Y – множество студентов. Правило f ставит в соответствие каждой парте студента, который сидит за ней. Данное правило не является функцией, потому что за одной партой могут сидеть два студента. Правило можно уточнить: поставим в соответствие парте студента, сидящего справа. Такое правило уже задает функцию.

Главное свойство функции – единственность образа для каждого прообраза – математически задается так:

$$f(x) = y_1, \quad f(x) = y_2 \;\Rightarrow\; y_1 = y_2.$$

Графиком функции f называется множество упорядоченных пар

$$\Gamma = \{(x,\, y) \mid x \in X,\ y = f(x)\}.$$

Пусть задано множество $A \subset X$ из области определения функции f. Образом множества A называется множество

$$f(A) = \{y \in Y \mid \exists\, x \in A\colon \quad y = f(x)\}$$

всех элементов из Y, имеющих прообразы в A. Образ области определения $f(X)$ называется *множеством значений* функции f.

Две функции f и g *равны* тогда и только тогда, когда они имеют одинаковые области определения и значения их совпадают:

$$f = g \Leftrightarrow D_f = D_g \;\text{ и }\; f(x) = g(x) \;\; \forall\; x \in X = D_f = D_g.$$

Мы использовали два математических знака, которые будут встречаться на протяжении всего курса: $\exists$ означает «существует» (от английского слова Exist – перевернутая буква E); $\forall$ – «для любого» (от английского All, «все», – перевернутая A).

Таким образом, функция однозначно определяется правилом, устанавливающим соответствие между элементами множеств, и своей областью определения.

Пусть, например, функция f ставит в соответствие каждой странице «Евгения Онегина» третью букву пятой сверху строки: D_f – множество номеров страниц, E_f – множество русских букв. Функция g действует по тому же правилу, но для «Отелло». Ясно, что $f \neq g$, так как и области определения, и значения этих функций не совпадают. Пусть теперь функция g действует так же, как f, но не на все страницы «Евгения Онегина», а только на третью главу. Тогда значения функций f и g при одинаковых аргументах совпадают. Однако функции не равны, так как имеют различные области определения (рис. 2.7).

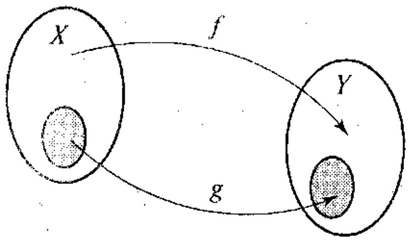


Рис. 2.7

2.5. Числовые функции

Если область определения и область значений функции есть подмножества множества действительных чисел, т.е. $X,\ Y \subseteq R$, то функцию называют ***числовой***. Это самая распространенная в математике разновидность функций. График числовой функции удобно изображать на координатной плоскости (мы уже пользовались ею при изображении отношений). Если по оси абсцисс отложить аргумент x функции f, а по оси ординат – ее значение $y = f(x)$, то совокупность точек (x, y) при $x \in D_f$ соответствует упорядоченным парам графика функции f (рис. 2.8). Такое изображение графика обычно называют просто графиком функции. График дает наглядное представление о свойствах функции.

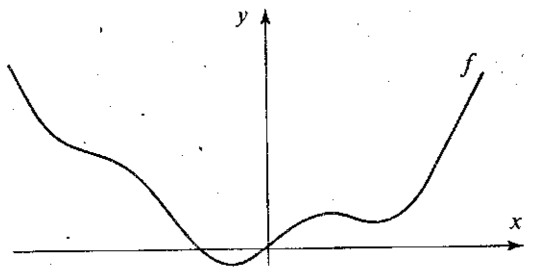


Рис. 2.8

Существуют различные способы задания числовой функции.

- **Аналитический.** Удобнее всего задать числовую функцию с помощью ф о р м у л ы. Когда мы пишем $y = x^2$, понятно, что эта функция ставит в соответствие действительному числу x его квадрат. Не следует забывать, что функция определяется не только правилом (формулой), но и областью определения. Пусть, например, три функции f_1, f_2, f_3 задаются одной и той же формулой $y = x^2$, однако первая определена на всем множестве действительных чисел $\boldsymbol{R}$ (рис. 2.9), вторая – на множестве неотрицательных действительных чисел $\boldsymbol{R}^+$ (рис. 2.10), а третья – на множестве натуральных чисел N (рис. 2.11).

Несмотря на то что функции задаются одной и той же формулой, это р а з н ы е функции, так как у них различные области определения. Обычно, если область определения специально не оговаривается, имеется в виду естественная область определения, т.е. те значения $x \in \boldsymbol{R}$, при которых функция имеет смысл. Например, функция $f(x) = 1/x$ определена при всех действительных x, кроме нуля: $D_f = \boldsymbol{R}\backslash\{0\} = (-\infty,\ 0) \cup (0,\ \infty)$.

$f_1\colon \boldsymbol{R} \to \boldsymbol{R}$

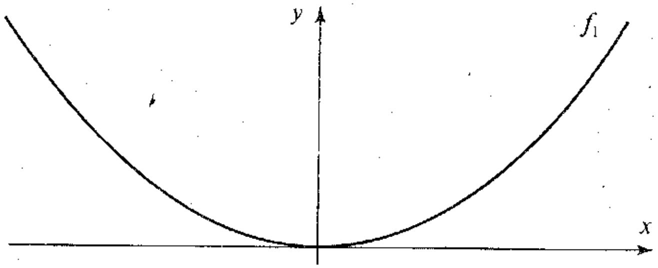


Рис. 2.9

$f_2\colon \boldsymbol{R}^+ \to \boldsymbol{R}$

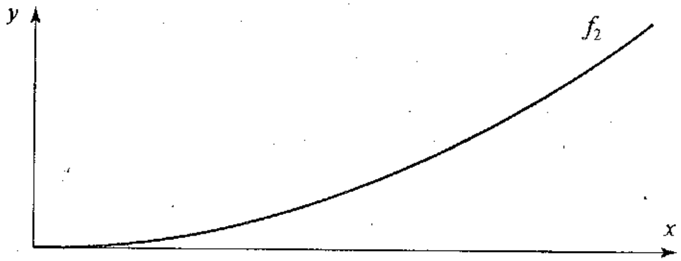


Рис. 2.10

$f_3\colon N \to \boldsymbol{R}$

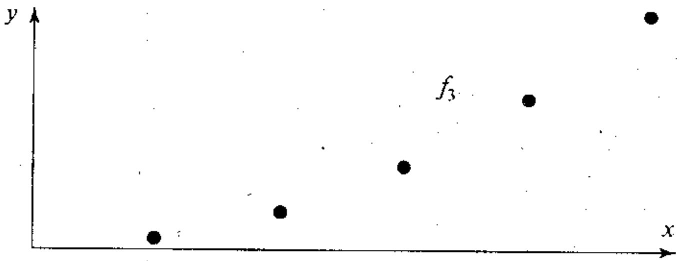


Рис. 2.11

- **Табличный.** Если область определения функции состоит из конечного числа точек, ее бывает удобно задать с помощью т а б л и ц ы: в левой колонке – аргументы, а в правой – соответствующие значения функции. Мы уже знакомы с таким способом задания функции (см. параграф 2.4). Часто с помощью таблицы задаются функции, представляющие собой результаты эксперимента, для которых не существует аналитического (формульного) представления. Типичный пример – температурный лист больного: аргументом является номер дня (порядковый номер измерения), значением функции – измеренная в этот день температура.

• **Графический.** Функция полностью определяется своим графиком. Поэтому вместо того, чтобы задавать числовые значения, можно просто нарисовать г р а ф и к. Чаще всего такие графики вычерчивает самописец при непрерывном измерении какой-то величины. Пример – кардиограмма: аргумент – текущий момент времени, значение функции – напряжение на соответствующем участке сердца.

Ознакомимся с основными свойствами числовых функций.

1. **Четность/нечетность.** Свойства четности и нечетности отражают внутреннюю симметрию функции.

Функция f называется *четной*, если:

1) ее область определения D_f симметрична относительно нуля;

2) $f(-x) = f(x) \quad \forall\ x \in D_f$.

Примеры четных функций: $y = x^2$, $y = \cos x$, $y = |x|$, определенные на всей числовой оси $\boldsymbol{R}$. График четной функции симметричен относительно оси ординат (рис. 2.12). Симметричность области определения важна, так как в противном случае можно найти такую точку $x \in D_f$, что $-x \notin D_f$. Та же функция $y = x^2$, определенная, например, на отрезке $[-2, 3]$, не является четной (рис. 2.13).

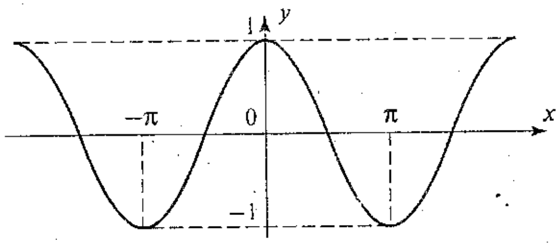


Рис. 2.12

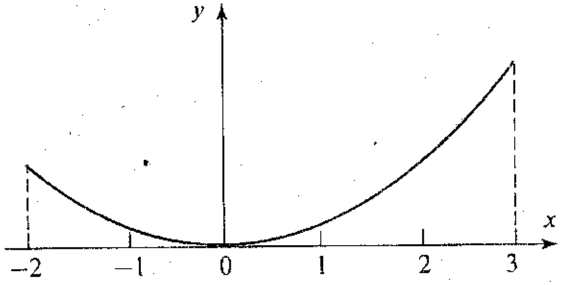


Рис. 2.13

Функция f называется *нечетной*, если:

1) ее область определения D_f симметрична относительно нуля;

2) $f(-x) = -f(x) \quad \forall\ x \in D_f$.

Примеры нечетных функций: $y = x$, $y = x^3$, $y = \sin x$. Для нечетных функций справедливы те же оговорки насчет области определения, что и для четных. График нечетной функции симметричен относительно начала координат (рис. 2.14).

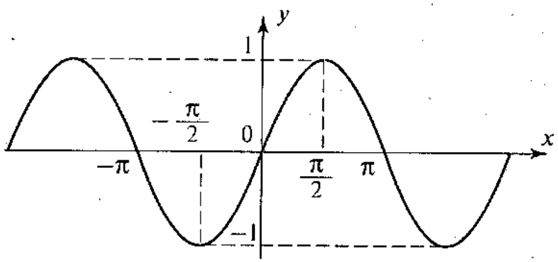


Рис. 2.14

Четные и нечетные функции достаточно определить лишь на одной половине симметричной области определения (скажем, только при $x > 0$). Значения для остальных аргументов легко восстановить.

Разумеется, существуют функции, которые не являются ни четными, ни нечетными: например, $y = x + 1$ (рис. 2.15).

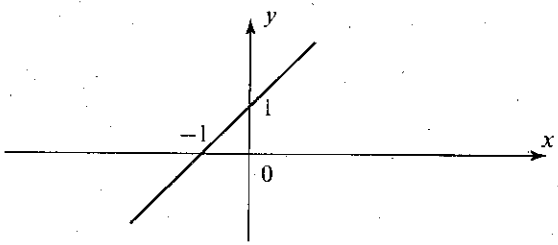


Рис. 2.15

Единственная функция, которая является одновременно четной и нечетной, это тождественный нуль $y \equiv 0$: эта функция ставит в соответствие любому действительному числу x нулевое значение.

2. **Монотонность.** Мы говорим: «Курс доллара на этой неделе растет, а предыдущие десять дней падал». С точки зрения математики мы таким образом описываем зависимость курса доллара от вре-

мени, т.е. поведение функции при возрастании аргумента (номера дня). Выделен участок области определения (предыдущие десять дней), на котором значения функции уменьшаются при увеличении аргумента, и участок, на котором значения функции увеличиваются. Для описания поведения функции при возрастании аргумента вводятся понятия монотонности.

Функция f называется:

неубывающей, если $\forall x_1, x_2 \in D_f$: $x_1 < x_2 \Rightarrow f(x_1) \leqslant f(x_2)$;

возрастающей, если $\forall x_1, x_2 \in D_f$: $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$;

невозрастающей, если $\forall x_1, x_2 \in D_f$: $x_1 < x_2 \Rightarrow f(x_1) \geqslant f(x_2)$;

убывающей, если $\forall x_1, x_2 \in D_f$: $x_1 < x_2 \Rightarrow f(x_1) > f(x_2)$.

Графики неубывающей и возрастающей функций изображены соответственно на рис. 2.16, *а* и *б*, невозрастающей и убывающей — на рис. 2.17, *а* и *б*. Такие функции имеют общее название — *монотонные*. Возрастающие и убывающие функции называются также *строго монотонными*. Единственная функция, одновременно неубывающая и невозрастающая, — тождественная константа $y \equiv c = \text{const}$ (рис. 2.18).

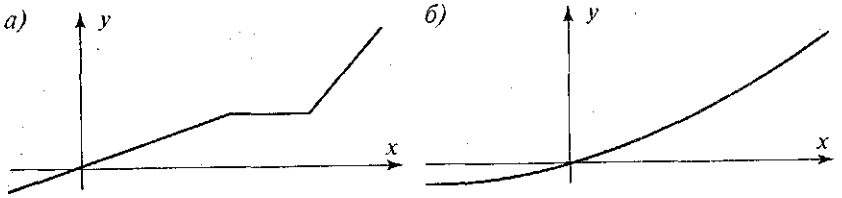


Рис. 2.16

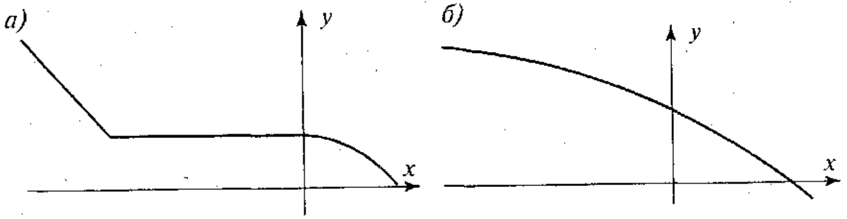


Рис. 2.17

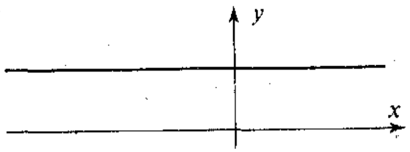


Рис. 2.18

Если в определениях монотонных функций заменить область определения D_f на некоторое множество $A \subseteq D_f$, получатся определения функции, неубывающей (возрастающей и т.д.) на множестве A. Функция, таким образом, может не быть монотонной на всей области определения, однако можно выделить ее *участки монотонности*. Например, функция $y = x^2$ убывает при $x < 0$ и возрастает при $x \geqslant 0$.

Рассмотрим функцию $y = 1/x$ (рис. 2.19). Можно ли утверждать, что эта функция убывает на всей области определения? Нет. В самом деле, возьмем, к примеру, точки $x_1 = -0{,}5$, $x_2 = 0{,}5$. Тогда $f(x_1) = -2$, $f(x_2) = 2$. Видно, что $x_1 < x_2$, однако $f(x_1) < f(x_2)$. Значит, функция $y = 1/x$ не является убывающей на всей области определения, тем не менее она убывает на множествах $(-\infty, 0)$ и $(0, +\infty)$.

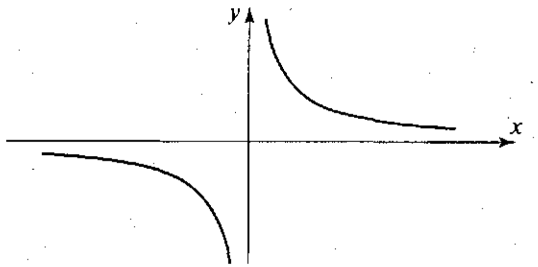


Рис. 2.19

3. **Ограниченность.** Рассматривая графики двух функций $y = x$ и $y = \sin x$ (рис. 2.20), можно заметить их принципиальное отличие друг от друга. Первая функция принимает любые значения от $-\infty$ до $+\infty$: множество ее значений представляет собой все множество $\boldsymbol{R}$. Все значения второй функции лежат между -1 и 1: множество значений синуса есть отрезок $[-1, 1]$. Итак, функции различаются по виду множества значений; у одних это множество можно поместить внутри конечного промежутка, а у других — нет. Для описания этого различия вводится понятие ограниченности функции.

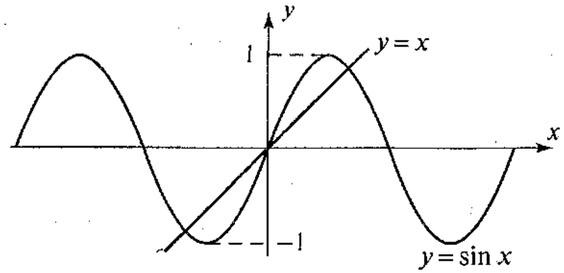


Рис. 2.20

Функция f называется *ограниченной*, если

$$\exists M > 0: |f(x)| < M \quad \forall x \in D_f.$$

Неравенство с модулем означает, что $-M < f(x) < M$, т.е. весь график ограниченной функции лежит в полосе от $-M$ до M. Ясно, что чисел M, удовлетворяющих определению, бесконечно много: если какое-то M_0 подходит, то годятся и все $M > M_0$. Примеры ограниченных функций: $y = \sin x$, $y = \cos x$, определенные на всей числовой оси. В качестве M можно взять любое число больше 1 (рис. 2.21).

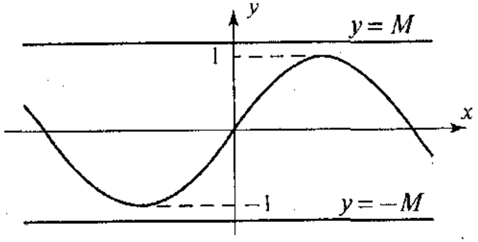


Рис. 2.21

Функция $y = x^2$ не является ограниченной на всей числовой оси (при всех $x \in \boldsymbol{R}$). Однако если рассмотреть ее на более узкой области определения, например на отрезке $[-1, 1]$, функция становится ограниченной (рис. 2.22).

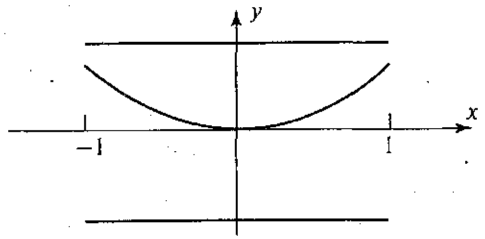


Рис. 2.22

Говорят, что функция ограничена на множестве $A \subseteq D_f$, если $\exists M > 0: |f(x)| < M \ \forall x \in A$. В частности, функция $y = x^2$ ограничена на любом отрезке $[a, b]$. Когда речь идет об ограниченности функции, всегда нужно указывать, на каком множестве она ограничена. Если множество не указано, подразумевается, что функция ограничена на всей области определения.

Функция f называется *ограниченной сверху* на множестве $A \subseteq D_f$, если $\exists M: f(x) < M \ \forall x \in A$, и *ограниченной снизу*, если $\exists M: f(x) > M$ $\forall x \in A$. График функции, ограниченной сверху, лежит ниже некоторой прямой, параллельной оси абсцисс (рис. 2.23), а график функции, ограниченной снизу, – выше некоторой прямой (рис. 2.24).

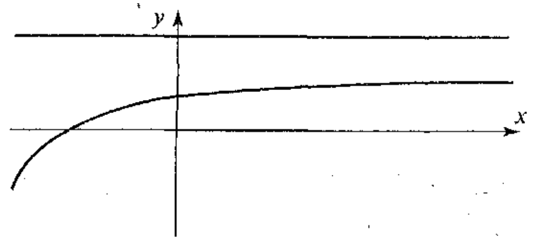


Рис. 2.23

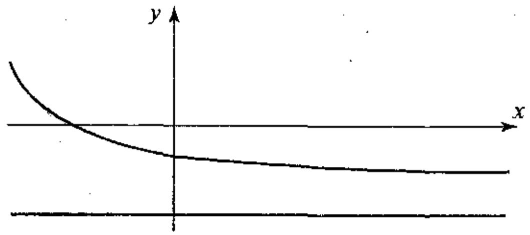


Рис. 2.24

4. **Периодичность.** Давно замечено, что многие явления в жизни циклически повторяются. Сменяются времена года, фазы Луны, на много лет вперед составлено расписание приливов и отливов. Известны циклы солнечной активности, биологические циклы человеческого организма и т.д. Такие повторяющиеся через равные промежутки явления удобно описывать математически с помощью периодических функций.

Функция f называется *периодической*, если существует число $T > 0$, такое, что:

1) $\forall x \in D_f$: $(x \pm T) \in D_f$;
2) $f(x + T) = f(x) \ \forall x \in D_f$.

В этом случае число T называется *периодом* функции f.

Примером периодических функций служат хорошо знакомые по школе тригонометрические функции: синус, косинус, тангенс и др. Рассмотрим повнимательнее функцию $y = \sin x$. Поместим левый конец единичного отрезка в начало координат, а правый – на оси абсцисс, т.е. в точке с координатами (0, 1). Будем вращать отрезок против часовой стрелки. Если угол между отрезком и осью абсцисс обозначить x, то ордината правого конца вращающегося отрезка и будет равна $\sin x$ (рис. 2.25). Легко заметить, что после того, как отрезок сделает полный оборот в 360°, значения синуса начинают повторяться. Ясно, что 360° — период синуса. Обычно углы в математике измеряются в радианах, а 360° составляют 2π радиан. Поэтому $T = 2\pi$ есть период функции $y = \sin x$. Значения будут повторяться и после двух полных оборотов, и после трех; другими словами, любое целое число полных оборотов (т.е. любое число, кратное 2π) также является периодом этой функции (см. рис. 2.14). Это справедливо для любой периодической функции.

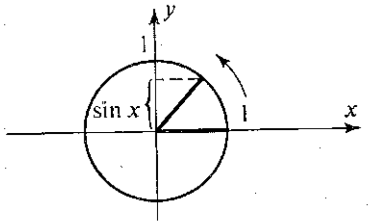


Рис. 2.25

∇**1.** *Пусть f – периодическая функция с периодом T и k – целое число. Тогда число kT также является периодом функции f.*

Например, синус имеет период 4π, 6π и т.д. При этом число, меньшее 2π, не является его периодом.

Рассмотренные выше числовые функции — это функции о д н о й переменной, т.е. их аргументом является одно число. Определение функции не мешает ввести функции двух (трех и т.д.) переменных. В самом деле, пусть задано правило, которое упорядоченной паре $\{(x, y), \quad x, y \in \boldsymbol{R}\}$ ставит в соответствие единственное действительное число z, например, $z = x^2 + y^2$. Как известно, упорядоченные пары чисел есть элементы декартова квадрата $\boldsymbol{R} \times \boldsymbol{R}$. Значит, такое правило задает некоторую функцию $f\colon \boldsymbol{R} \times \boldsymbol{R} \to \boldsymbol{R}$, имеющую два числовых аргумента. Это и есть функция д в у х переменных.

Забегая вперед, можно сказать, что графиком функции двух переменных является поверхность в пространстве (рис. 2.26), или множество точек с координатами $(x, y, f(x, y))$.

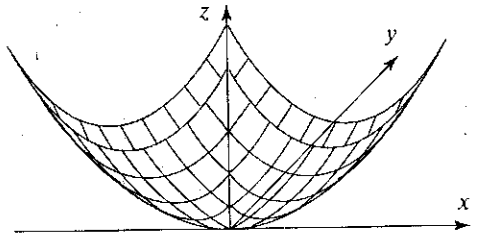


Рис. 2.26

Точно так же определяется функция н е с к о л ь к и х переменных $f(x_1, x_2, \ldots, x_n)$, которая каждому набору чисел $x_1, x_2, \ldots, x_n$ ставит в соответствие действительное число.

2.6. Композиция функций

Пусть заданы две функции:

$$f\colon X \to Y;$$
$$g\colon Y \to Z.$$

Обратите внимание, что область определения функции g совпадает с областью значений функции f. Это значит, что функция g «знает», что делать с любым значением функции f, т.е. в какой элемент $z \in Z$ переводить элемент $y = f(x) \in Y$. Функция f, в свою очередь, позволяет получить такой y из любого элемента $x \in X$. Если сначала воздействовать на x функцией f, а затем на результат этого действия y воздействовать функцией g, получится некоторый, вполне определенный элемент $z \in Z$. Фактически такая процедура определяет новую функцию из множества X во множество Z.

Композицией функций f и g называется функция $h\colon X \to Z$, которая каждому элементу $x \in X$ ставит в соответствие элемент $z = g(f(x)) \in Z$ (рис. 2.27). Композицию обозначают $g \circ f$ и по-другому называют ***суперпозицией*** или ***сложной функцией***.

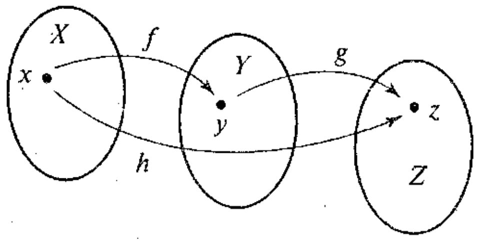


Рис. 2.27

Чтобы построить композицию функций, в качестве аргумента функции g нужно взять значение функции f: $h(x) = g(f(x))$.

Можно представить себе следующий процесс. Писатель f, имея перед собой план x будущей книги, садится за письменный стол и через некоторое время превращает план в рукопись y. После этого вступает в свои права редактор g и доводит рукопись y до состояния z, приемлемого для печати. Если редактор g специализируется на научной фантастике, а рукопись y представляет собой сборник лирических стихотворений, мы не позавидуем автору f. На этом примере отчетливо видно, что функция g должна быть определена на значениях функции f. В противном случае композицию $g \circ f$ построить не удастся.

Пусть исходная область значений функции f – множество Y_1, а область определения функции g – множество Y_2. Обычно, чтобы корректно построить композицию этих функций, соответствующие области сужают до пересечения $Y = Y_1 \cap Y_2$. Это достигается сужением области определения функции f. Новая область определения и есть область определения композиции. Может получиться, что пересечение пусто. В этом случае построение сложной функции невозможно. Приведем математический пример. Пусть $f(x) = -(x-1)^2$, $g(y) = \sqrt{y}$. Видно, что значения f строго отрицательны при любом аргументе x, а функция g определена только при неотрицательных аргументах. Пересечение соответствующих областей пусто, и композиция невозможна.

Приведем примеры, в которых можно построить композицию функций.

1. $f(x) = \sin x$, $g(y) = y^2$. Тогда $(g \circ f)(x) = g(f(x)) = (\sin x)^2$. В данном случае можно считать, что все три множества X, Y и Z совпадают с числовой осью R. Тогда можно построить и композицию $f \circ g$: сначала действует функция g, а затем функция f; привычно обозначая аргумент через x, получаем, что $(f \circ g)(x) = f(g(x)) = \sin(x^2)$.

2. $f(x) = \cos x$, $g(y) = 1/y$. Композицию $(g \circ f)(x) = g(f(x)) = 1/\cos x$ можно построить, если сузить исходную область определения функции f до множества $X = R \setminus \left\{\frac{\pi}{2} + \pi k \mid k \in Z\right\}$, удалив те аргументы, при которых косинус принимает нулевое значение. В этом случае область значений f и область определения g можно выбрать как $R \setminus \{0\}$. Композиция $(f \circ g)(x) = f(g(x)) = \cos(1/x)$ имеет естественную область определения $x \neq 0$.

Операцию, которая двум функциям ставит в соответствие их композицию, также называют композицией. Из примеров можно сделать очевидное наблюдение.

∇1. *Композиция функций некоммутативна:*

$$g \circ f \neq f \circ g.$$

Внимание! Впервые нам встретилась операция, не удовлетворяющая закону коммутативности. В дальнейшем мы всегда будем обращать внимание на алгебраические свойства операций и отмечать подобные «отклонения от нормы».

Можно составить композицию и более чем двух функций. В самом деле, пусть заданы функции:

$$f\colon X \to Y;$$
$$g\colon Y \to Z;$$
$$h\colon Z \to W.$$

Воздействуем на элемент $x \in X$ функцией f, затем на $y = f(x) \in Y$ функцией g и, наконец, на $z = g(y) = g(f(x)) \in Z$ функцией h. Получим некоторый элемент $w = h(z) = h(g(y)) = h(g(f(x))) \in W$. Эта процедура определяет функцию из X в W, т.е. композицию трех функций. Совершенно ясно, что эту композицию можно составить двумя разными способами (рис. 2.28).

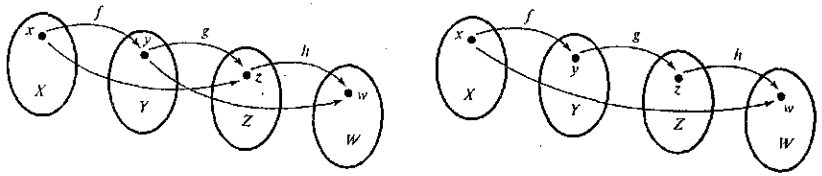


Рис. 2.28

Мы могли бы сначала построить композицию $g \circ f$, а затем – композицию этой композиции с h. С другой стороны, можно сначала построить композицию $h \circ g$, а затем – композицию этой композиции с f. В итоге получится одна и та же функция из X в W, которая и есть композиция трех функций.

∇2. *Композиция функций ассоциативна:*

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

Любой ассоциативный закон позволяет расставлять скобки в произвольном порядке и обозначать композицию трех функций как $h \circ g \circ f$. Ассоциативные законы будут встречаться на протяжении всего курса математики, в самых различных ее разделах. Вспомните ассоциативные законы операций над множествами: объединения и пересечения.

2.7. Обратная функция

Продолжим пример с функцией f, устанавливающей соответствие между студентами и партами (множества X и Y соответственно). Пусть студенты сидят по одному и все парты заняты, т.е. студентов столько же, сколько парт. В этом случае легко построить новую функцию g, которая ставит в соответствие каждому элементу $y \in Y$ (парте) единственный элемент $x \in X$ (студента, сидящего за этой партой). Проделанная нами процедура есть не что иное, как построение обратной функции.

Тот факт, что значений у функции f ровно столько же, сколько аргументов, математически означает следующее: уравнение $f(x) = y$ однозначно разрешимо относительно x для каждого $y \in Y$. Другими словами, каждый образ имеет единственный прообраз. Такая функция называется *взаимно однозначной*. Если провести прямую $y = y_0$, параллельную оси абсцисс, где y_0 – некоторое значение функции $f(x)$, эта прямая пересечет график взаимно однозначной функции в единственной точке. Как видно из приведенных ниже примеров, функцию можно сделать однозначной, если некоторым образом ограничить, «урезать» ее область значений и область определения.

Пусть задана взаимно однозначная функция f. Функция g, которая ставит в соответствие каждому элементу $y \in Y$ элемент $x \in X$, такой, что $y = f(x)$, называется ***обратной*** к f:

$$y = f(x) \Leftrightarrow x = g(y)$$

(рис. 2.29). Функцию f естественно называть исходной по отношению к обратной функции g. Другими словами, обратная функция ставит в соответствие элементам из области значений прямой функции их прообраз в области определения. У каждого элемента существует прообраз, и притом единственный, так как прямая функция по условию взаимно однозначна. Значит, построенное соответствие действительно задает функцию. Естественное обозначение для обратной функции: f^{-1}.

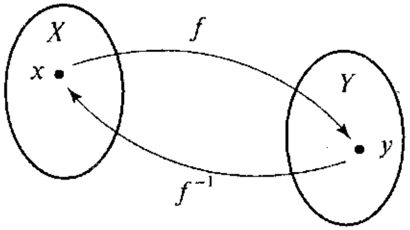


Рис. 2.29

Наш пример со студентами и партами позволяет сделать следующие наблюдения.

∇1. *Область определения обратной функции совпадает с областью значений исходной функции и наоборот:*

$$D_{f^{-1}} = E_f, \qquad E_{f^{-1}} = D_f.$$

∇2. *Функция, обратная к f^{-1}, есть функция f:*

$$(f^{-1})^{-1} = f.$$

Это позволяет называть функции f и f^{-1} взаимно обратными.

Что значит построить обратную функцию к функции $y = f(x)$? По существу, это разрешить данное уравнение относительно x, выразить x из этого уравнения (рис. 2.30).

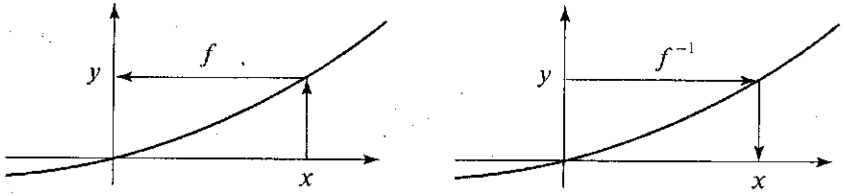


Рис. 2.30

Необходимо помнить, что обратная функция строится только для взаимно однозначных функций. Поэтому построение следует начинать с выбора подходящих областей определения и значений прямой функции.

Примеры.

1. $y = f(x) = x^2 - 3$.

Чтобы функция была взаимно однозначной, достаточно рассмотреть ее на следующих областях:

$$f\colon [0, \infty) \to [-3, \infty).$$

Выразим x из данного уравнения: $x = f^{-1}(y) = \sqrt{y+3}$. Построенная обратная функция, очевидно, есть взаимно однозначная функция, определенная на промежутке $[-3, \infty)$, с областью значений $[0, \infty)$:

$$f^{-1}\colon [-3, \infty) \to [0, \infty).$$

Область ее определения совпадает с областью значений прямой функции f, а область значений – с соответствующей областью определения (рис. 2.31).

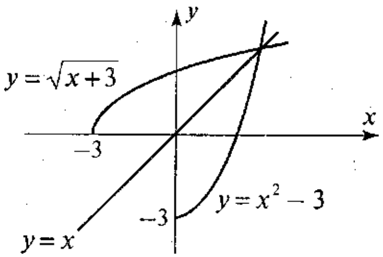


Рис. 2.31

2. $y = f(x) = \sin x$.

Функция становится взаимно однозначной, если рассмотреть ее на отрезке возрастания, например на $[-\pi/2, \pi/2]$, а область значений задать в виде отрезка $[-1, 1]$:

$$f: [-\pi/2, \pi/2] \to [-1, 1].$$

Функция $x = f^{-1}(y)$, где x определяется из уравнения $y = \sin x$, называется, как известно, *арксинусом*: $x = \arcsin y$ (рис. 2.32). Это такой угол от $-\pi/2$ до $\pi/2$, синус которого равен y. Естественная область определения арксинуса – отрезок $[-1, 1]$:

$$f^{-1}: [-1, 1] \to [-\pi/2, \pi/2].$$

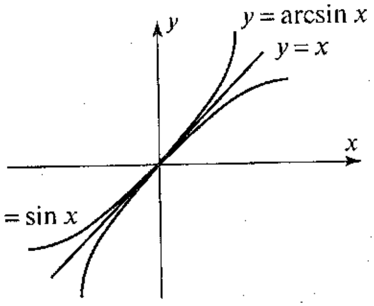


Рис. 2.32

Аргумент функции привычнее обозначать через x, а значение – через y. Поэтому часто пишут: $y = f^{-1}(x)$. Нетрудно убедиться в справедливости следующего утверждения.

∇3. *Графики функций $y = f(x)$ и $y = f^{-1}(x)$ симметричны относительно биссектрисы I и III квадрантов* (*прямой $y = x$*).

Задачи

1. Исследовать отношения ρ на множестве $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ (проверить, является ли данное отношение рефлексивным, симметричным, транзитивным, антисимметричным; если является – доказать, если нет – привести опровергающий пример):

а) $x\rho y$, если x делится на y;

б) $x\rho y$, если $x - y$ делится на 4;

в) $x\rho y$, если $x + y \in X$;

г) $x\rho y$, если $xy = y^2$;

д) $x\rho y$, если $x + y = 8$;

е) $x\rho y$, если $x < 2y$.

2. Исследовать отношения ρ на множестве $X = \{-4, -3, -2, -1, 0, 1, 2, 3\}$:

а) $x\rho y$, если $x < y^2$;

б) $x\rho y$, если $xy > 0$;

в) $x\rho y$, если $|x - y| < 3$.

3. Построить графики отношений из задач 1 и 2 (нанести на координатную плоскость точки, соответствующие упорядоченным парам отношения).

4. Исследовать отношения на множестве людей:

а) «быть братом» ($x\rho y$, если x есть брат y);

б) «любить» ($x\rho y$, если x любит y);

в) «быть отцом»;

г) «быть подчиненным»;

д) «быть в одной партии».

5. Исследовать отношения на множестве русских слов:

а) «иметь равное число букв»;

б) «быть синонимом»;

в) «рифмоваться».

6. Какому свойству удовлетворяет график отношения:

а) рефлексивного;

б) симметричного.

7. Дано множество $X = \{-4, -3, -2, 1, 2, 3\}$. Доказать, что следующие отношения есть отношения эквивалентности, и построить соответствующее разбиение множества X:

а) $x\rho y$, если $xy \geqslant 0$;

б) $x\rho y$, если $x - y$ делится на 2.

8. Доказать, что следующие отношения есть отношения частичного порядка, и построить диаграммы частично упорядоченных множеств:

а) отношение $\subseteq$ на множестве-степени $\mathcal{P}(A)$ множества $A = \{1, 2, 3\}$;

б) отношение на множестве $X = \{1, 2, 3, 5, 6, 10, 15, 30\}$: $x\rho y$, если y делится на x.

9. Исследовать функции (указать область определения, проверить свойства взаимной однозначности, четности, монотонности, ограниченности, периодичности):

а) $y = x^2$;

б) $y = x^3$;

в) $y = 1/x$;

г) $y = \sin x$;

д) $y = \operatorname{tg} x$;

е) $y = x$;

ж) $y = |x|$.

10. Привести пример функции:

а) определенной на множестве действительных чисел $\mathbf{R}$ со значениями во множестве целых чисел $\mathbf{Z}$;

б) определенной на $\mathbf{Z}$ со значениями в $\mathbf{R}$;

в) область определения и область значений – множество натуральных чисел, но функция не взаимно однозначная.

11. Найти композиции $g \circ f$ и $f \circ g$ (указать область определения и область значений):

а) $f(x) = 1/x$, $g(x) = \sin x$;

б) $f(x) = \operatorname{tg} x$, $g(x) = \sqrt{x}$;

в) $f(x) = 3$, $g(x) = x^2$.

Глава 3 ОСНОВЫ МАТЕМАТИЧЕСКОЙ ЛОГИКИ

Математика развивается по своим внутренним законам, а именно по законам логики. Математическая логика – это теория верных рассуждений. Слово «верных», как и везде в математике, не следует понимать в абсолютном смысле. Большинство математических дисциплин строится на основе некоторых исходных договоренностей, или аксиом, которые формулируются исходя из соображений здравого смысла. Далее из аксиом выводятся следствия, причем доказательство строится по законам логического вывода. Эти законы обеспечивает как раз математическая логика. Из другой системы аксиом и по другим законам вывода может получиться совершенно другая наука. Чудесная загадка математики, которую М. Клайн называет «необъяснимой эффективностью», состоит в том, что, несмотря на кажущийся произвол своего построения, математика в состоянии описывать реальный мир с удивительной точностью. Игра ума – вывод математических теорий по логическим законам – оказалась гораздо ближе к действительности, чем можно было ожидать.

Далее приводятся некоторые основные факты математической логики, которую еще называют формальной логикой. Формальной потому, что она позволяет проверить правильность рассуждений независимо от их содержания. Цепочки рассуждений в совершенно

разных областях математики и других наук можно одинаково описать на языке логики и убедиться в их справедливости или ошибочности.

3.1. Высказывания и логические связки

Основной объект математической логики – высказывание. ***Высказыванием*** называется повествовательное предложение, которое может быть классифицировано либо как истинное, либо как ложное, но не как и то и другое вместе. Можно привести сколько угодно высказываний, например:

«Сегодня идет снег»;

«П.И. Чайковский написал 10 опер»;

«Все каналы на Марсе пересохли»;

«Если я подпрыгну до потолка, то у меня заболит голова»;

«25 делится на 13».

Содержание высказывания несущественно: лишь бы это предложение могло быть либо истинным, либо ложным. При этом вовсе не обязательно указывать способ проверки истинности. Главное, что высказывание не может быть истинным и ложным одновременно. Если высказывание истинно, будем говорить, что его значение истинности – истина (или t от английского true); если ложно, то значение истинности – ложь (f от false).

Высказывания в математической логике обычно обозначаются прописными латинскими буквами: A, B, C и т.д. Для того чтобы из высказываний получать новые высказывания, применяются специальные операции – ***логические связки***. Рассмотрим пять основных логических связок. Сначала дадим неформальное объяснение. Однако оно чревато неточностями, поэтому дадим логическим операциям также строгое определение. Определить высказывание – значит указать, в каких случаях оно истинно, а в каких ложно.

Отрицание – это высказывание, которое получается из данного высказывания A с помощью слова «не». Отрицание можно обозначать по-разному: $\neg A$, $\sim A$, $\overline{A}$. Остановимся на обозначении $\neg A$.

Простое добавление слова «не» к высказыванию чаще всего будет противоречить языковым нормам. Поэтому в конкретных случаях требуется «перевод» полученного высказывания на русский язык. Пусть, например, A = «Завтра пойдет дождь». Что значит «Не (Завтра пойдет дождь)»: «Дождь пойдет не завтра», «Завтра пойдет не дождь» или «Завтра не пойдет дождь»? Здравый смысл подсказывает, что отрицанием высказывания A является третье предложение. Чтобы определить точно, дадим формальное определение отрицания.

Отрицанием $\neg A$ высказывания A называется такое высказывание, которое принимает значение f (ложно), если высказывание A истинно, и значение t (истинно), если высказывание A ложно.

В нашем примере этому условию удовлетворяет только третье предложение. Итак, $\neg A$ = «Завтра **не** пойдет дождь».

Дизъюнкция – это высказывание, которое получается из двух данных высказываний A и B с помощью союза «или». Дизъюнкция обозначается $A \vee B$. Допускается дизъюнкция любых, самых далеких по смыслу высказываний:

«Дважды два – четыре **или** Париж – столица Англии».

Мы уже обсуждали трудности толкования союза «или» в русском языке, когда давали определение объединения множеств. Этот союз можно понимать в исключающем и неисключающем смысле. Дизъюнкция строится с помощью неисключающего «или». Таким образом, ***дизъюнкция*** $A \vee B$ истинна, когда истинно по крайней мере одно из высказываний A и B или оба вместе. Другими словами, дизъюнкция ложна в том и только в том случае, когда оба высказывания ложны.

Конъюнкция – это высказывание, которое получается из двух данных высказываний A и B с помощью союза «и»:

«Сегодня четверг **и** Шекспир – это группа авторов».

Конъюнкция обозначается $A \wedge B$ (иногда $A \& B$). Смысл союза «и» подсказывает строгое определение конъюнкции: *конъюнкция* $A \wedge B$ высказываний A и B истинна в том и только в том случае, когда оба высказывания истинны. Мы уже пользовались союзом «и» при определении пересечения множеств.

Операции объединения и пересечения множеств не случайно определяются с помощью тех же союзов, что и дизъюнкция и конъюнкция высказываний. Не случайны также и похожие обозначения ($\vee$ и $\cup$; $\wedge$ и $\cap$). Далее мы убедимся, что алгебраические свойства операций над множествами и логических связок аналогичны.

Импликация образуется из высказываний A и B с помощью слов «если... то...». Получается высказывание вида «если A, то B». Напомним, что математическая логика носит формальный характер, содержанием высказываний она не занимается. Так, импликация может связывать далекие по смыслу высказывания:

«**Если** функция f взаимно однозначна, **то** красный конь купается»;

«**Если** $2 + 2 = 5$, **то** $3 + 3 = 10$»;

«**Если** Аннушка пролила масло, **то** будет дорожное происшествие».

На примере импликации хорошо видна разница между обычным языком и языком логики. В обычном языке сложное предложение «если A, то B» предполагает между A и B отношение посылки и следствия или же причины и обусловленного ею действия, как, например, в таком предложении: «Если заточить карандаш, то можно нарисовать дом». В логике импликация связывает любые два высказывания.

Импликация обозначается $A \Rightarrow B$, при этом говорят: «A влечет B» или «B при условии, что A», «B, если A», «A есть достаточное условие для B», «B есть необходимое условие для A».

Договорились, что *импликация* $A \Rightarrow B$ ложна в том и только в том случае, когда высказывание A истинно, а высказывание B ложно. Такое определение подсказано здравым смыслом: разумно считать импликацию истинной, если B истинно, независимо от значения A; если оба участника импликации ложны, импликация, естественно, также истинна. В единственном случае, когда «предпосылка» импликации истинна, а «вывод» ложен, импликация считается ложной. Следующие три импликации истинны:

«Если 7 – простое число, то $2 \times 2 = 4$» (истина $\Rightarrow$ истина);

«Если 8 – простое число, то $2 \times 2 = 4$» (ложь $\Rightarrow$ истина);

«Если 8 – простое число, то $2 \times 2 = 5$» (ложь $\Rightarrow$ ложь).

Импликация

«Если 7 – простое число, то $2 \times 2 = 5$» (истина $\Rightarrow$ ложь)

по определению ложна.

Эквиваленция образуется из высказываний A и B с помощью слов «...тогда и только тогда, когда...»:

«Треугольники подобны **тогда и только тогда, когда** их углы совпадают»;

«Луна – живое существо **тогда и только тогда, когда** Маккартни поет колыбельную».

Подчеркнем, что утверждение «A тогда и только тогда, когда B» не означает в логике, что составляющие высказывания A и B имеют одно и то же значение или один и тот же смысл.

Эквиваленция обозначается $A \Leftrightarrow B$. Синонимы для эквиваленции: «если A, то B, и если B, то A», «A в том и только в том случае, когда B», «A есть необходимое и достаточное условие для B», «B есть необходимое и достаточное условие для A». Разумное определение эквиваленции: *эквиваленция* истинна в том и только в том случае, когда высказывания A и B имеют одинаковое значение истинности (либо оба истинны, либо оба ложны). Поэтому следующие эквиваленции истинны:

«7 – простое число тогда и только тогда, когда $2 \times 2 = 4$» (истина $\Leftrightarrow$ истина);

«8 – простое число тогда и только тогда, когда $2 \times 2 = 5$» (ложь $\Leftrightarrow$ ложь).

Новые высказывания (отрицание, дизъюнкция, конъюнкция, импликация и эквиваленция) образуются из существующих высказываний с помощью операций, или *логических связок*, имеющих те же названия. Логические связки удобно определять с помощью т а б л и ц   и с т и н н о с т и, где перебираются все возможные комбинации значений истинности составляющих высказываний и указывается соответствующее значение истинности высказывания, полученного в результате действия логической операции:

Отрицание

A	$\neg A$
t	f
f	t

Дизъюнкция

A	B	$A \vee B$
t	t	t
t	f	t
f	t	t
f	f	f

Конъюнкция

A	B	$A \wedge B$
t	t	t
t	f	f
f	t	f
f	f	f

Импликация

A	B	$A \Rightarrow B$
t	t	t
t	f	f
f	t	t
f	f	t

Эквиваленция

A	B	$A \Leftrightarrow B$
t	t	t
t	f	f
f	t	f
f	f	t

Логические связки позволяют из простых высказываний получить новые, сколь угодно сложные высказывания. Рассмотрим пример:

«Если завтра будет дождь или снег, то я возьму зонт и надену пальто или свитер».

Введем обозначения:

A = «Завтра будет дождь»;

B = «Завтра будет снег»;

C = «Я возьму зонт»;

D = «Я надену пальто»;

E = «Я надену свитер».

Тогда наше предложение можно записать в виде

$$(A \vee B) \Rightarrow (C \wedge (D \vee E)).$$

В логике, как и в арифметике, операции делятся по старшинству. Это позволяет при записи сложных высказываний избегать большого количества скобок. **Порядок выполнения операций** таков: приоритет имеет отрицание, затем на одном уровне – дизъюнкция и конъюнкция, следующая связка – импликация и, наконец, самая последняя – эквиваленция. Поэтому составленное выше высказывание можно записать короче: $A \vee B \Rightarrow C \wedge (D \vee E)$. Одной пары скобок не избежать, так как высказывание $C \wedge D \vee E$ непонятно (операции $\wedge$ и $\vee$ имеют одинаковую силу). Приведем еще примеры записи высказываний без скобок:

$$A \wedge B \Rightarrow C \text{ вместо } (A \wedge B) \Rightarrow C;$$
$$A \Leftrightarrow B \Rightarrow C \text{ вместо } A \Leftrightarrow (B \Rightarrow C);$$
$$A \Leftrightarrow B \wedge C \text{ вместо } A \Leftrightarrow (B \wedge C);$$
$$\neg A \vee B \text{ вместо } (\neg A) \vee B.$$

Для сложных высказываний (их еще называют составными формулами) применяют также таблицы истинности, где перебирают все возможные значения истинности составляющих высказываний (элементарных формул, простых компонент, т.е. таких высказываний, которые нельзя представить как результат действия логических операций). Такая таблица содержит 2^n строк, где n – количество простых компонент. Таблицу истинности удобно составлять, разбивая сложное высказывание на более простые и добавляя вспомогательные столбцы со значениями истинности этих составляющих. Ниже приводится таблица истинности для высказывания $(A \vee B) \wedge (B \Rightarrow \neg C)$:

A	B	C	$A \vee B$	$\neg C$	$B \Rightarrow \neg C$	$(A \vee B) \wedge (B \Rightarrow \neg C)$
t	t	t	t	f	f	f
t	t	f	t	t	t	t
t	f	t	t	f	t	t
t	f	f	t	t	t	t
f	t	t	t	f	f	f
f	t	f	t	t	t	t
f	f	t	f	f	t	f
f	f	f	f	t	t	f

Высказывание содержит три простые компоненты A, B и C, поэтому в таблице $2^3 = 8$ строк.

3 Зак. № 2435 Турецкий

3.2. Логическая эквивалентность. Свойства логических операций

Введем на множестве всех высказываний отношение логической эквивалентности. Формула A ***логически эквивалентна*** формуле B, если их таблицы истинности совпадают. Это записывается как A eq B. Проверим, например, что $A \vee B$ eq $(A \Rightarrow B) \Rightarrow B$. Для этого составим таблицу истинности формулы $(A \Rightarrow B) \Rightarrow B$:

A	B	$A \Rightarrow B$	$(A \Rightarrow B) \Rightarrow B$
t	t	t	t
t	f	f	t
f	t	t	t
f	f	t	f

Видно, что построенная таблица совпадает с таблицей истинности дизъюнкции.

∇**1.** *Логическая эквивалентность есть отношение эквивалентности.*

В самом деле, из определения логической эквивалентности легко проверяются свойства рефлексивности, симметричности и транзитивности данного отношения.

Логически эквивалентные высказывания неразличимы с точки зрения формальной логики, поскольку они при каждом наборе значений компонент принимают одинаковые значения истинности. В любом рассуждении высказывание можно заменить на логически эквивалентное.

В следующей теореме отражены алгебраические свойства дизъюнкции, конъюнкции и отрицания, полностью аналогичные соответствующим свойствам объединения, пересечения и дополнения множеств (см. теоремы 1.1 и 1.2).

Теорема 3.1. *Следующие пары формул логически эквивалентны*:

1. $A \vee (B \vee C)$ eq $(A \vee B) \vee C$. 1′. $A \wedge (B \wedge C)$ eq $(A \wedge B) \wedge C$.
2. $A \vee B$ eq $B \vee A$. 2′. $A \wedge B$ eq $B \wedge A$.
3. $A \vee (B \wedge C)$ eq $(A \vee B) \wedge (A \vee C)$. 3′. $A \wedge (B \vee C)$ eq $(A \wedge B) \vee (A \wedge C)$.
4 и 4′. $\neg(\neg A)$ eq A.
5. $A \vee A$ eq A. 5′. $A \wedge A$ eq A.
6. $A \vee (A \wedge B)$ eq A. 6′. $A \wedge (A \vee B)$ eq A.
7. $\neg(A \vee B)$ eq $\neg A \wedge \neg B$. 7′. $\neg(A \wedge B)$ eq $\neg A \vee \neg B$.

Внимательно рассматривая эти логические эквивалентности, легко заметить аналогию со свойствами операций над множествами, если отношению логической эквивалентности поставить в соответствие отношение равенства множеств, а операциям дизъюнкции, конъюнкции и отрицания – соответственно операции объединения, пересечения и дополнения. Таким образом, для логических операций справедливы ассоциативные законы 1, 1′, коммутативные законы 2, 2′ и дистрибутивные законы 3, 3′. Алгебраические свойства логических операций впервые исследовал английский математик Джордж Буль.

Джордж Буль (1815–1864). Английский математик. Родился в Линкольне. Самоучка; самостоятельно изучил греческий, латинский, немецкий, французский и итальянский языки, затем математику. С 16 лет работал помощником учителя в школе. В 1849–1864 гг. – профессор математики в Куинс-колледже в Корке (Ирландия).

Научные интересы Буля были достаточно широки: философия, логика, математический анализ, теория вероятностей. В 1847 г. он опубликовал «Математический анализ логики, или Опыт исчисления дедуктивных умозаключений», в 1848 г. – «Логическое исчисление». Развил свою систему и изложил ее в трактате «Исследование законов мышления, на коих основаны математические теории логики и вероятностей» (1854), в котором свел логику к алгебраической форме. Логическое исчисление Буля получило название булевой алгебры.

Идеи Буля развивали Огастес де Морган и Чарлз Пирс, в XX в. – Бертран Рассел и Джон фон Нейман. В 40–50-х годах XX в. булева алгебра приобрела особенное значение в связи с развитием вычислительной техники, так как элементы электронных схем хорошо описываются языком формальной логики.

∇**2.** *Свойства ассоциативности и коммутативности позволяют определить дизъюнкцию $A_1 \vee A_2 \vee \ldots \vee A_n$ и конъюнкцию $A_1 \wedge A_2 \wedge \ldots \wedge A_n$ произвольного конечного набора высказываний. Независимо от порядка и от способа расстановки скобок получаются логически эквивалентные высказывания.*

Законы 6, 6′, как и для множеств, называются законами поглощения, а законы 7, 7′ – законами де Моргана.

Огастес де Морган (1806–1871). Шотландский математик и логик. Родился в Мадуре (Индия). С 1823 г. учился в Кембриджском университете. В 1828–1831 и 1836–1866 гг. – профессор Университетского колледжа в Лондоне.

Работы посвящены основаниям алгебры, арифметике, математическому анализу, теории вероятностей и логике. Один из основателей формальной алгебры. Ему принадлежат важные работы по математической логике. В 1847 г. опубликовал работу «Формальная логика, или Исчисление выводов, необходимых и возможных». Был инициатором применения логических исчислений к обоснованию теорем теории вероятностей. Изучал историю математики.

Основатель Лондонского королевского общества и его первый президент (с 1866), член Королевского астрономического общества.

Теорему 3.1 легко доказать с помощью таблиц истинности. Докажем, например, закон де Моргана 7: отрицание дизъюнкции логически эквивалентно конъюнкции отрицаний. Итак:

A	B	$A \vee B$	$\neg(A \vee B)$	$\neg A$	$\neg B$	$\neg A \wedge \neg B$
t	t	t	f	f	f	f
t	f	t	f	f	t	f
f	t	t	f	t	f	f
f	f	f	t	t	t	t

Как и для множеств, справедлив **принцип двойственности**: *если два высказывания логически эквивалентны и в них заменить дизъюнкцию на конъюнкцию и наоборот, полученные высказывания также логически эквивалентны.* Эквивалентности в теореме 3.1 расположены двойственными парами. Закон 4, 4′ является самодвойственным.

Логические эквивалентности теоремы 3.1 не зависят от содержания высказываний. Читателям предоставляется возможность самостоятельно привести примеры, придав высказываниям A, B, C конкретный смысл.

3.3. Тавтологии, или законы логики

Рассмотрим для примера высказывание $A \wedge (A \Rightarrow B) \Rightarrow B$: «Если A и из A следует B, то B». Создается впечатление, что независимо от содержания и значений истинности компонент A и B данное высказывание всегда истинно. Составим таблицу истинности для высказывания «Если на небе 13 звезд и из этого следует, что Пушкин застрелил Дантеса, то Пушкин застрелил Дантеса»:

A	B	$A \Rightarrow B$	$A \wedge (A \Rightarrow B)$	$A \wedge (A \Rightarrow B) \Rightarrow B$
t	t	t	t	t
t	f	f	f	t
f	t	t	f	t
f	f	t	f	t

В самом деле, высказывание принимает только значение t, какое бы значение ни принимали простые компоненты.

Высказывание, истинное при любых значениях его компонент, называется ***тождественно истинным*** или ***тавтологией***. Тот факт, что A — тавтология, обозначается как $\models A$. Тавтологии играют большую роль в формальной логике, поэтому их называют еще ***законами логики***. Приведем несколько законов логики, в справедливости которых легко убедиться, составив таблицы истинности:

1) $A \vee \neg A$ (закон исключенного третьего);
2) $\neg(A \wedge \neg A)$ (закон противоречия);
3) $[(A \Rightarrow B) \wedge (B \Rightarrow C)] \Rightarrow (A \Rightarrow C)$ (закон силлогизма);
4) $A \Leftrightarrow \neg(\neg A)$ (закон двойного отрицания);
5) $(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$ (закон контрапозиции).

Некоторым законам логики можно придать разумные словесные формулировки. Например, закон исключенного третьего гласит: «или A истинно, или A ложно», «или A, или не A истинно» (третьего не дано!). Закон противоречия: «не могут быть одновременно истинны A и не A».

Отрицание тавтологии, т.е. тождественно ложное высказывание, называется ***противоречием***. Следующая теорема продолжает аналогию между высказываниями и множествами (продолжается и нумерация законов из теоремы 3.1). Заметим, что тавтология соответствует универсальному множеству U, а противоречие — пустому множеству $\varnothing$.

Теорема 3.2.

8, 8′. *Если* $A \vee B$ *— тавтология,* $A \wedge B$ *— противоречие, то* B eq $\neg A$.
9. *Если* $A \vee B$ eq A *для всех* A, *то* B *— противоречие.*
9′. *Если* $A \wedge B$ eq A *для всех* A, *то* B *— тавтология.*
10. *Если* B *— противоречие, то* $A \vee B$ eq A.
10′. *Если* B *— тавтология, то* $A \wedge B$ eq A.
11. *Если* B *— тавтология, то* $A \vee B$ *— тавтология.*
11′. *Если* B *— противоречие, то* $A \wedge B$ *— противоречие.*

Теперь принцип двойственности можно сформулировать полнее.

Принцип двойственности. *Если два высказывания логически эквивалентны и в них заменить дизъюнкцию на конъюнкцию, тавтологию на противоречие и наоборот, полученные высказывания будут также логически эквивалентны.*

В заключение составим таблицу соответствия между понятиями теории множеств и математической логики:

Теория множеств	**Логика**
Множество	Высказывание
Объединение	Дизъюнкция
Пересечение	Конъюнкция
Дополнение	Отрицание
Универсальное множество	Тавтология
Пустое множество	Противоречие

3.4. Правила логического вывода

Задача математической логики – дать принцип рассуждения, т.е. теорию вывода, критерии для решения механическим путем вопроса о том, можно ли некоторую цепочку рассуждений считать правильной. При этом важна только форма высказываний, составляющих цепочку, а не их содержание и смысл. Основной инструмент построения правильного доказательства – тавтологические импликации. Поэтому в предыдущем параграфе мы уделили много внимания понятию и свойствам тавтологии.

Высказывание B называется ***логическим следствием*** высказываний $A_1, A_2, \ldots, A_m$, если B истинно всякий раз, когда каждое A_i, $i = 1, 2, \ldots, m$, истинно. Это записывается так:

$$A_1, A_2, \ldots, A_m \models B.$$

Высказывания $A_1, A_2, \ldots, A_m$ называются ***посылками*** логического следствия, а высказывание B – ***заключением***.

По-русски это звучит так: «Пусть $A_1, A_2, \ldots, A_m$. Тогда B». Большинство теорем в математике имеет именно такую структуру. Доказать теорему – значит доказать, что заключение действительно является логическим следствием посылок. В этом параграфе будет дан формальный инструмент правильного доказательства теорем и выявления логических ошибок в доказательстве. Самый простой способ – составить совместную таблицу истинности посылок и заключения, найти в ней строки, в которых посылки истинны, и убедиться, что заключение в этих строках также имеет значение t.

Пример.

$A, C, A \wedge B \Rightarrow \neg C \models \neg B$. Для доказательства составим таблицу истинности:

A	B	C	$A \wedge B \Rightarrow \neg C$	$\neg B$
t	t	t	f	f
t	t	f	t	f
t	f	t	t	t
t	f	f	t	t
f	t	t	t	f
f	t	f	t	f
f	f	t	t	t
f	f	f	t	t

В таблице нашлась только одна строка (заключена в рамку), в которой все посылки принимают значение истины (выделены полужирным). Видно, что заключение при этом также истинно. Таким образом, логическое следствие доказано. Этим доказывается любая теорема, которая по форме совпадает с данным логическим следствием. Чтобы подчеркнуть независимость доказательства от содержания высказываний, приведем несколько формулировок теорем данной логической структуры.

1. A = «x делится на 5»; B = «x – нечетное»; C = «x делится на 4». «Пусть x делится на 5 и на 4. Известно, что нечетные числа, кратные 5, не делятся на 4. Значит, x четно».
2. A = «ΔABC – равнобедренный»; B = «ΔABC – прямоугольный»; C = «ΔABC – равносторонний».

   «Пусть ΔABC равнобедренный и равносторонний. Если равнобедренный треугольник является прямоугольным, то он – не равносторонний. Значит, ΔABC – не прямоугольный».
3. A = «Он – светловолосый»; B = «Он – левша»; C = «Он – моряк».

   «Он – светловолосый моряк. Светловолосые левши не бывают моряками. Значит, он не левша».
4. A = «Он – образованный»; B = «У него болит голова»; C = «Он – дятел».

   «Пусть он – образованный дятел. Если у образованного болит голова, то он не дятел. Значит, у него не болит голова».

Приведем также пример нелогичного рассуждения:

$$A \Rightarrow B,\ C \Rightarrow D,\ A \vee C \models B \wedge D.$$

Пусть компоненты принимают следующие значения: $A = t$, $B = t$, $C = f$, $D = f$. В этом случае все посылки истинны, а заключение ложно. Содержательный пример: «Если идет дождь (A), то возьмем зонт (B). Если светит солнце (C), то наденем футболку (D). Пусть идет дождь или светит солнце ($A \vee C$). Тогда возьмем зонт и наденем футболку ($B \wedge D$)».

Было бы странно при доказательстве теорем (логических следствий) каждый раз строить таблицы истинности. Обычно доказательство теоремы $A_1, A_2, \ldots, A_m \models B$ представляет собой цепочку (конечную последовательность) высказываний $E_1, E_2, \ldots, E_r$, последнее из которых совпадает с доказываемым заключением: $E_r = B$. Какую же цепочку доказательства считать правильной? Представляется разумным следующее правило.

Любая формула в цепочке $E_1, E_2, \ldots, E_r$ доказательства теоремы $A_1, A_2, \ldots, A_m \models B$ есть:

- *либо посылка (одно из высказываний $A_1, A_2, \ldots, A_m$);*
- *либо логическое следствие из некоторых предыдущих высказываний.*

Нетрудно убедиться, что в цепочку доказательств можно ввести любую тавтологию, а также любое высказывание, логически эквивалентное какому-либо высказыванию цепочки.

Логическое следствие можно построить с помощью набора правил (правил вывода). Математически эти правила основаны на применении тождественно истинных импликаций. В самом деле, если высказывание вида «если A, то B» является тавтологией, ясно, что высказывание B логически следует из высказывания A.

Приведем примеры таких импликаций (с некоторыми из них мы уже знакомы), порождающих правила вывода:

1. $A \wedge (A \Rightarrow B) \Rightarrow B$: из A и $A \Rightarrow B$ выводится B. Это самое распространенное правило вывода имеет специальное название – правило modus ponens.
2. $\neg B \wedge (A \Rightarrow B) \Rightarrow \neg A$: из «не B» и $A \Rightarrow B$ выводится «не A».
3. $\neg A \wedge (A \vee B) \Rightarrow B$: из «не A» и «A или B» выводится B.
4. $A \wedge B \Rightarrow A$.
5. $A \Rightarrow A \vee B$.
6. $[(A \Rightarrow B) \wedge (B \Rightarrow C)] \Rightarrow (A \Rightarrow C)$ – знакомый нам закон силлогизма.
7. $(A \Rightarrow \neg A) \Rightarrow \neg A$ – принцип приведения к абсурду: если из A следует «не A», то из этого выводится «не A».
8. $(A \Rightarrow B) \Rightarrow (A \vee C \Rightarrow B \vee C)$.
9. $(A \Rightarrow B) \Rightarrow (A \wedge C \Rightarrow B \wedge C)$.

Построим цепочку доказательства теоремы «о светловолосых моряках». Для удобства обозначим высказывания другими буквами (P, Q, R вместо A, B, C):

$$P,\ R,\ P \wedge Q \Rightarrow \neg R \models \neg Q.$$

Итак:

1) R – посылка;

2) $P \wedge Q \Rightarrow \neg R$ – посылка;

3) $\neg(P \wedge Q)$ – вывели из высказываний 1 и 2 в силу тавтологии 2 $\neg B \wedge (A \Rightarrow B) \Rightarrow \neg A$ при $A = P \wedge Q$, $B = \neg R$;

4) $\neg P \vee \neg Q$ – заменили высказывание 3 на логически эквивалентное по закону де Моргана;

5) P – посылка;

6) $\neg Q$ – вывели из высказываний 4 и 5 по тавтологии 3 $\neg A \wedge (A \vee B) \Rightarrow B$, обозначив $A = \neg P$, $B = \neg Q$.

Таким образом, цепочка доказательства составлена по правилам логического вывода и последнее высказывание в ней совпадает с доказываемым заключением $\neg Q$. Запишем текст доказательства в соответствии с построенной цепочкой:

> Он моряк (1), а светловолосый левша не бывает моряком (2). Значит, он не светловолосый левша (3), т.е. не светловолосый или не левша (4). Но он светловолосый (5). Следовательно, он не левша (6), что и требовалось доказать.

Итак, мы познакомились с некоторыми правилами логичного доказательства теорем, основанными на понятии логического следствия.

На этом заканчивается знакомство с основами математической логики. Основное содержание главы – культура логичного рассуждения. Эта культура необходима не только в математике. Логичными, доказательными должны быть рассуждения исследователя в любой области знаний (а в гуманитарных науках – особенно). Убедительное, добротное исследование строится, на самом деле, по законам логики.

Задачи

1. Записать высказывания на языке логики с помощью логических связок:

а) Идет дождь или кто-то не выключил душ.

б) Если вечером будет туман, то Сергей или останется дома, или должен будет взять зонт.

в) Петр сядет, и он или Сергей будет ждать.

г) Ни Север, ни Юг не победил в гражданской войне.

д) Если я устал или голоден, я не могу заниматься.

е) Если Маша встанет и пойдет в школу, она будет довольна, а если она не встанет, она не будет довольна.

ж) Если мыши живут на Марсе, а мыши, читающие Фолкнера, говорят по-английски, то мыши не живут на Марсе.

2. Пусть C = «Сегодня ясно», R = «Сегодня идет дождь», S = «Сегодня идет снег», Y = «Вчера было пасмурно». Перевести на русский язык следующие высказывания:

а) $C \Rightarrow \neg(R \vee S)$;

б) $Y \Leftrightarrow C$;

в) $Y \wedge (C \vee R)$;

г) $Y \Rightarrow R \vee C$.

3. Составить таблицы истинности для следующих высказываний:

а) $P \Rightarrow (P \Rightarrow Q)$;

б) $(P \Rightarrow Q) \Leftrightarrow \neg P \vee Q$;

в) $P \Rightarrow \neg(Q \vee R)$;

г) $(P \Rightarrow Q \wedge R) \vee (\neg P \wedge Q)$.

4. Доказать с помощью таблиц истинности законы дистрибутивности теоремы 3.1.

5. Доказать закон силлогизма и закон контрапозиции.

6. Доказать тавтологию $\neg B \wedge (A \Rightarrow B) \Rightarrow \neg A$.

7. Пусть импликация $P \Rightarrow Q$ истинна. Что можно сказать о значении истинности высказывания $\neg P \wedge Q \Leftrightarrow P \vee Q$?

8. Пусть эквиваленция $P \Leftrightarrow Q$ истинна. Что можно сказать о значении истинности эквиваленции $P \Leftrightarrow \neg Q$?

9. Построить цепочку доказательства логического следствия $A \Rightarrow B, \neg(B \vee C) \models \neg A$. Придать высказываниям конкретные значения и привести текст доказательства.

Часть 2
ОСНОВЫ АЛГЕБРЫ И АНАЛИТИЧЕСКОЙ ГЕОМЕТРИИ

Нам задали алфавит, целый длиннющий алфавит.
И я его знаю: сперва идет «А», а потом все остальные буквы.

А. Линдгрен. Малыш и Карлсон

В данной части мы переходим к рассмотрению новых математических объектов. В первую очередь, речь идет о *векторах*. Развитие этого понятия представляет собой увлекательный путь от знакомых по школе векторов на плоскости и в пространстве (глава 4) к векторам в многомерном пространстве (глава 5) и далее – к элементам абстрактных векторных пространств (параграф 5.4 и глава 6). Подобное развитие идей характерно для математики, когда все повышающийся уровень абстракции позволяет многие конкретные математические объекты рассматривать впоследствии как частные случаи некоторых общих структур. В данном случае важнейшую роль играет понятие *операции* над элементами различных множеств, в частности операции над векторами. Нас будут особо интересовать алгебраические свойства операций, с которыми мы уже ознакомились в первой части: ассоциативность, коммутативность, дистрибутивность и т.д. Оказывается, одинаковыми алгебраическими свойствами, аналогичными свойствам векторов, обладают самые разнообразные объекты. Это и позволяет ввести понятие абстрактного векторного пространства (параграф 5.4), в котором данные свойства уже не выводятся, а постулируются.

С помощью векторов мы изучим свойства таких геометрических множеств, как *прямая* и *плоскость*, ознакомимся с системой координат. Главная идея аналитической геометрии состоит в том, что исследование геометрических объектов (линий, поверхностей и т.д.) можно заменить исследованием их уравнений. Через уравнения кривых в системе координат геометрия, по сути, смыкается с алгеброй. Важным инструментом исследования в геометрии становятся *матрицы* и *определители*. Эти алгебраические понятия представляют интерес и сами по себе.

Последняя глава данной части (глава 6) посвящена абстрактным алгебраическим структурам. Мы уже на более высоком уровне введем понятие операции и изучим некоторые множества, в которых операции обладают специфическими свойствами: *полугруппу, группу, кольцо, поле* и некоторые другие. Накопленный к этому моменту конкретный материал позволит нам привести много содержательных примеров.

Глава 4
АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ

4.1. Векторы на плоскости

4.1.1. Понятие вектора. Операции над векторами

Из школьного курса физики мы хорошо помним, что бывают физические величины, обладающие не только значением, но и направлением. Традиционный пример – сила: важно, в каком направлении она действует. Такие величины называют векторными. В подпараграфе 4.1.1 вводится математическое понятие вектора и изучаются основные свойства векторов.

В геометрии нет строгого определения точки (как нет строгого определения множества). Будем считать, что понятие точки интуитивно ясно, а такие множества точек, как отрезок, луч, прямая и плоскость, известны из школьной геометрии. Пусть даны две точки A и B на плоскости. ***Вектором*** называется направленный отрезок, идущий из точки A в точку B (рис. 4.1). Другими словами, вектор – это упорядоченная пара точек. Точка A называется *началом* вектора или *точкой приложения*, точка B – *концом* вектора.

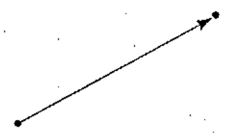


Рис. 4.1

Будем обозначать векторы или строчными латинскими буквами со стрелкой, или двумя прописными буквами по началу и концу: $\vec{a} = \overrightarrow{AB}$. Не исключен случай, когда начало и конец вектора совпадают. Такой вектор называется *нулевым* и обозначается $\vec{0}$. Расстояние между началом и концом вектора называется его *длиной*: $|\vec{a}| = |\overrightarrow{AB}| = AB$. Таким образом, длина вектора $\vec{0}$ равна нулю.

Два вектора называются *коллинеарными*, если они лежат на параллельных прямых или на одной и той же прямой. Коллинеарность обозначается как параллельность: $\vec{a} \parallel \vec{b}$. Как ясно из определения, коллинеарные векторы можно расположить четырьмя способами (рис. 4.2).

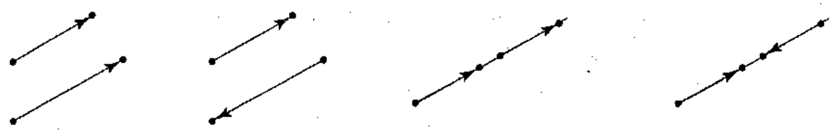


Рис. 4.2

С каждым вектором однозначно связан луч на плоскости: он начинается в начале вектора и продолжается до бесконечности в направлении, задаваемом вектором (рис. 4.3).

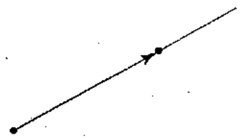


Рис. 4.3

Понятие *сонаправленных* векторов, т.е. коллинеарных векторов, имеющих одинаковое направление, вводится следующим образом (рис. 4.4).

- Два вектора, не лежащие на одной прямой, сонаправленны, если концы их лежат в одной полуплоскости.
- Два вектора, лежащие на одной прямой, сонаправленны, если луч одного из них содержится в луче другого.

Рис. 4.4

В противном случае коллинеарные векторы называются *антинаправленными* (рис. 4.5).

Рис. 4.5

Два вектора называются *равными* ($\vec{a}=\vec{b}$), если они сонаправленны и их длины равны (рис. 4.6). Условие сонаправленности очень важно. Векторы могут иметь одинаковую длину, но быть неколлинеарными или антинаправленными. В этом случае они не считаются равными (рис. 4.7).

Рис. 4.6

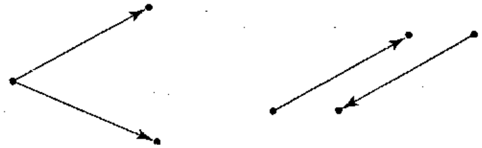


Рис. 4.7

Определение равенства векторов позволяет говорить о *свободных* векторах: данный вектор можно перемещать по плоскости параллельно самому себе, при этом получится вектор, равный данному.

∇1. *Равенство есть отношение эквивалентности на множестве векторов.*

В самом деле, по определению вектор равен сам себе (рефлексивность); если $\vec{a}=\vec{b}$, то и $\vec{b}=\vec{a}$ (симметричность); если $\vec{a}=\vec{b}$ и $\vec{b}=\vec{c}$, то $\vec{a}=\vec{c}$ (транзитивность). Действительно, это отношение эквивалентности.

Значит, фактически мы имеем дело не с отдельными векторами, а с классами эквивалентности по отношению равенства. Все векторы одного класса равны и, следовательно, неразличимы. Другими словами, вектор можно приложить к л ю б о й точке.

Чтобы работать с какими-либо математическими объектами, нужно ввести операции над ними. Определим операции над векторами, которые, как мы увидим, во многом напоминают операции над числами.

Сложение векторов. *Суммой* $\vec{a}+\vec{b}$ векторов $\vec{a}$ и $\vec{b}$ называется вектор, который идет из начала вектора $\vec{a}$ к концу вектора $\vec{b}$, если $\vec{b}$ приложен к концу $\vec{a}$. На рис. 4.8 $\overrightarrow{AB}=\vec{a}$, $\overrightarrow{BC}=\vec{b}$, $\overrightarrow{AC}=\vec{a}+\vec{b}$.

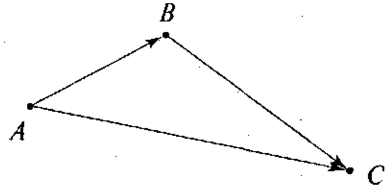


Рис. 4.8

∇2. *Сложение векторов коммутативно:*

$$\vec{a}+\vec{b}=\vec{b}+\vec{a}.$$

Чтобы в этом убедиться, построим параллелограмм *ABCD*, в котором $\overrightarrow{AB}=\overrightarrow{DC}=\vec{a}$, $\overrightarrow{BC}=\overrightarrow{AD}=\vec{b}$ (рис. 4.9).

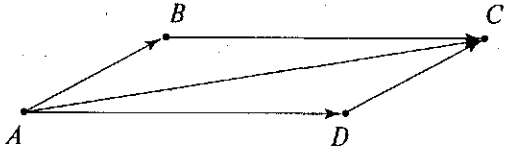


Рис. 4.9

Тогда по определению, с одной стороны, $\overrightarrow{AC}=\overrightarrow{AB}+\overrightarrow{BC}=\vec{a}+\vec{b}$, а с другой стороны, $\overrightarrow{AC}=\overrightarrow{AD}+\overrightarrow{DC}=\vec{b}+\vec{a}$. Отсюда и следует равенство $\vec{a}+\vec{b}=\vec{b}+\vec{a}$. Можно построить сумму двух векторов по-другому: для этого нужно приложить их к одной точке и достроить до параллелограмма (рис. 4.10). Диагональ параллелограмма, идущая из точки приложения векторов, и есть их сумма. В этом состоит известное из школьного курса физики «правило параллелограмма».

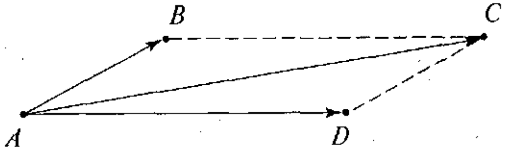


Рис. 4.10

∇3. *Сложение векторов ассоциативно:*

$$\vec{a}+(\vec{b}+\vec{c})=(\vec{a}+\vec{b})+\vec{c}.$$

Это видно из рис. 4.11.

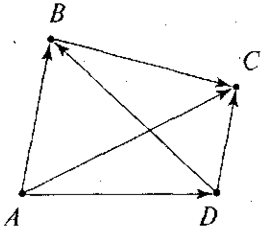


Рис. 4.11

Здесь $\overrightarrow{AD}=\vec{a}$, $\overrightarrow{DB}=\vec{b}$, $\overrightarrow{BC}=\vec{c}$. Тогда $\overrightarrow{DC}=\vec{b}+\vec{c}$, $\overrightarrow{AB}=\vec{a}+\vec{b}$. Значит, с одной стороны, $\overrightarrow{AC}=\overrightarrow{AD}+\overrightarrow{DC}=\vec{a}+(\vec{b}+\vec{c})$, а с другой – $\overrightarrow{AC}=\overrightarrow{AB}+\overrightarrow{BC}=(\vec{a}+\vec{b})+\vec{c}$. Этим и доказывается ассоциативность.

Ассоциативность сложения векторов позволяет определить сумму произвольного конечного набора векторов $\vec{a}_1+\vec{a}_2+...+\vec{a}_n$. Чтобы получить такую сумму, нужно второй вектор приложить к концу первого, третий – к концу второго и т.д., а сумму отложить от начала первого к концу последнего (рис. 4.12):

$$\overrightarrow{AB}+\overrightarrow{BC}+\overrightarrow{CD}+\overrightarrow{DE}+\overrightarrow{EF}=\overrightarrow{AF}.$$

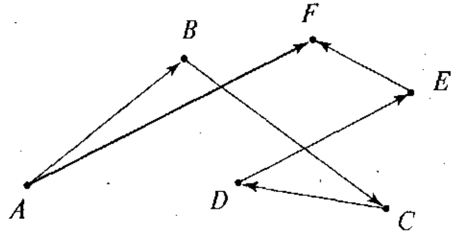


Рис. 4.12

∇4. *Для любого вектора* $\vec{a}$

$$\vec{a}+\vec{0}=\vec{0}+\vec{a}=\vec{a}.$$

∇5. *Для любого вектора* $\vec{a}=\overrightarrow{AB}$ *справедливо*

$$\overrightarrow{AB}+\overrightarrow{BA}=\overrightarrow{BA}+\overrightarrow{AB}=\vec{0}.$$

Это следует непосредственно из определения суммы векторов. Вектор $\overrightarrow{BA}$ называют *противоположным* (симметричным) вектору $\vec{a}$ и обозначают $-\vec{a}$. Последнее наблюдение можно переформулировать: каждый вектор имеет противоположный.

∇6. *Уравнение*

$$\vec{x}+\vec{b}=\vec{a}$$

имеет единственное решение при любых векторах $\vec{a}$, $\vec{b}$.

Действительно, достаточно прибавить к обеим частям уравнения вектор $-\vec{b}$: $\vec{x}=\vec{a}+(-\vec{b})$. Такой вектор называется *разностью* векторов $\vec{a}$ и $\vec{b}$ и обозначается $\vec{a}-\vec{b}$. Итак, разностью векторов $\vec{a}$ и $\vec{b}$ называется вектор $\vec{a}-\vec{b}$, такой, что в сумме с вектором $\vec{b}$ он дает вектор $\vec{a}$. На рис. 4.13 $\overrightarrow{AB}=\vec{a}$, $\overrightarrow{AC}=\vec{b}$, $\overrightarrow{CB}=\vec{a}-\vec{b}$.

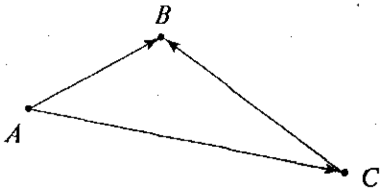


Рис. 4.13

Умножение вектора на число. *Произведение* $\lambda\vec{a}$ вектора $\vec{a}$ на число λ есть вектор:

- коллинеарный $\vec{a}$;
- имеющий длину $|\lambda|\cdot|\vec{a}|$;
- сонаправленный $\vec{a}$ при $\lambda>0$ и антинаправленный при $\lambda<0$.

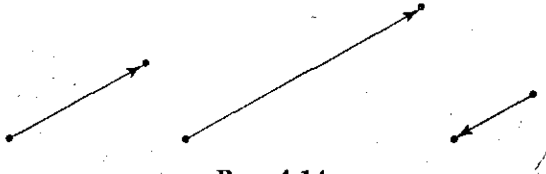


Рис. 4.14

На рис. 4.14 изображены векторы $\vec{a}$, $2\vec{a}$, $-0{,}5\vec{a}$. При $\lambda = 0$ получается вектор нулевой длины, т.е. нулевой вектор: $0 \cdot \vec{a} = \vec{0}$.

Справедливы следующие свойства операции умножения вектора на число.

∇**7.** *Ненулевые векторы $\vec{a}$ и $\vec{b}$ коллинеарны тогда и только тогда, когда существует число λ, такое, что $\vec{b} = \lambda\vec{a}$.*

Легко проверить, что $\lambda = \pm(|\vec{b}|/|\vec{a}|)$. Знак «+» берется, если векторы $\vec{a}$ и $\vec{b}$ сонаправленны, а «–» – если антинаправленны.

∇**8.** *Умножение вектора на число ассоциативно относительно умножения чисел:*

$$\lambda(\mu\vec{a}) = (\lambda\mu)\vec{a}.$$

Свойство ∇8 прямо следует из определения.

∇**9.** *Умножение вектора на число дистрибутивно относительно сложения чисел:*

$$(\lambda+\mu)\vec{a} = \lambda\vec{a} + \mu\vec{a}.$$

Доказательство. Свойство очевидно, если хотя бы одно из чисел λ, μ или вектор $\vec{a}$ равны нулю. Пусть оба числа и вектор ненулевые. Пусть λ и μ одного знака. Тогда векторы $\lambda\vec{a}$ и $\mu\vec{a}$ по определению сонаправленны. Поэтому $|\lambda\vec{a} + \mu\vec{a}| = |\lambda\vec{a}| + |\mu\vec{a}| = |\lambda|\cdot|\vec{a}| + |\mu|\cdot|\vec{a}| = (|\lambda| + |\mu|)\cdot|\vec{a}|$. Последнее равенство справедливо в силу дистрибутивности умножения относительно сложения для чисел.

С другой стороны, $|(\lambda + \mu)\vec{a}| = |\lambda + \mu|\cdot|\vec{a}| = (|\lambda| + |\mu|)\cdot|\vec{a}|$. Итак, длины сонаправленных векторов совпадают. Значит, они равны. Случай разных знаков λ и μ рассматривается аналогично.

∇**10.** *Умножение вектора на число дистрибутивно относительно сложения векторов:*

$$\lambda(\vec{a}+\vec{b}) = \lambda\vec{a} + \lambda\vec{b}.$$

Доказательство. Свойство очевидно, если хотя бы одно из чисел λ, μ или вектор $\vec{a}$ равны нулю. Пусть это не так. Рассмотрим случай, когда векторы $\vec{a}$ и $\vec{b}$ коллинеарны. Тогда $\vec{b} = \mu\vec{a}$ при некотором μ. Отсюда, в силу дистрибутивности ∇9 и ассоциативности ∇8, $\lambda(\vec{a} + \vec{b}) = \lambda(\vec{a} + \mu\vec{a}) = \lambda(1 + \mu)\vec{a} = (\lambda + \lambda\mu)\vec{a} = \lambda\vec{a} + (\lambda\mu)\vec{a} = \lambda\vec{a} + \lambda(\mu\vec{a}) = \lambda\vec{a} + \lambda\vec{b}$.

Осталось рассмотреть случай неколлинеарных векторов $\vec{a}$ и $\vec{b}$. Для доказательства рассмотрим рис. 4.15 в случае $\lambda > 1$.

Здесь $\overrightarrow{AC} = \vec{a}$, $\overrightarrow{AE} = \lambda\vec{a}$, $\overrightarrow{CB} = \vec{b}$, $\overrightarrow{ED} = \lambda\vec{b}$. Стало быть, $\overrightarrow{AB} = \vec{a} + \vec{b}$, $\overrightarrow{AD} = \lambda\vec{a} + \lambda\vec{b}$. Но из подобия треугольников ABC и ADE следует, что $\overrightarrow{AD} = \lambda\overrightarrow{AB} = \lambda(\vec{a} + \vec{b})$. Значит, вектор $\overrightarrow{AD}$ равен одновременно $\lambda\vec{a} + \lambda\vec{b}$ и $\lambda(\vec{a} + \vec{b})$. Свойство доказано. Случаи других значений λ рассматриваются аналогично.

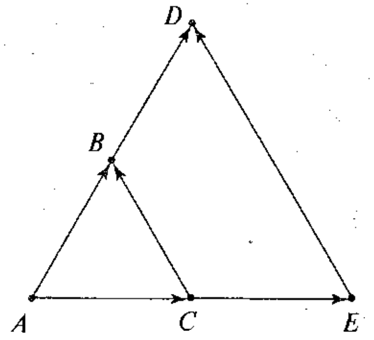


Рис. 4.15

∇**11.** $1\cdot\vec{a} = \vec{a}$, $(-\lambda)\vec{a} = -(\lambda\vec{a})$.

Свойства операций позволяют обращаться с векторами так же, как с числами: переносить из одной части равенства в другую с противоположным знаком, делить обе части на ненулевое число, приводить подобные члены и т.д. Например, из равенства $2\vec{a} + 3\vec{b} = 4\vec{a} - 2\vec{c}$ следует $\vec{a} = 1{,}5\vec{b} + \vec{c}$: мы перенесли $2\vec{a}$ вправо с минусом, $2\vec{c}$ – влево с плюсом, привели подобные члены $4\vec{a} - 2\vec{a} = 2\vec{a}$ и, наконец, разделили обе части равенства на 2.

4.1.2. Разложение по базису. Система координат

Теорема 4.1. *Пусть даны два ненулевых и неколлинеарных вектора $\vec{e}_1$ и $\vec{e}_2$. Тогда любой вектор $\vec{x}$ можно представить в виде*

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2,$$

и притом единственным образом.

Доказательство. Если вектор $\vec{x}$ коллинеарен одному из базисных векторов (например, вектору $\vec{e}_1$), то, в силу свойства ∇7, $\vec{x} = \pm(|\vec{x}|/|\vec{e}_1|)\vec{e}_1$ и можно принять $x_1 = \pm(|\vec{x}|/|\vec{e}_1|)$, $x_2 = 0$. В случае когда $\vec{x}$ неколлинеарен базисным векторам, существование чисел x_1 и x_2 доказывается с помощью следующих построений. Приложим векторы $\vec{e}_1$, $\vec{e}_2$ и $\vec{x}$ к одной точке: $\overrightarrow{AB}=\vec{e}_1$, $\overrightarrow{AC}=\vec{e}_2$, $\overrightarrow{AD}=\vec{x}$. Построим параллелограмм $AEDF$, в котором сторона AE продолжает AB, а сторона AF – сторону AC (рис. 4.16). Обозначим $\overrightarrow{AE}=\vec{r}_1$, $\overrightarrow{AF}=\vec{r}_2$. Очевидно, что $\vec{x}=\vec{r}_1+\vec{r}_2$. Однако по построению $\vec{r}_1 \parallel \vec{e}_1$, $\vec{r}_2 \parallel \vec{e}_2$ и по свойству ∇7 существуют числа x_1 и x_2, такие, что $\vec{r}_1=x_1\vec{e}_1$, $\vec{r}_2=x_2\vec{e}_2$. Итак, $\vec{x}=x_1\vec{e}_1+x_2\vec{e}_2$, что и требовалось доказать. Другое взаимное расположение векторов $\vec{e}_1$, $\vec{e}_2$ и $\vec{x}$ рассматривается аналогично.

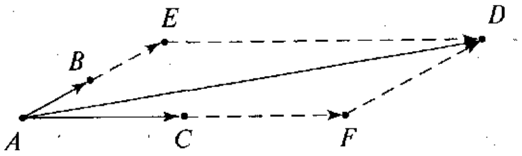


Рис. 4.16

Предположим, что разложение не единственно, т.е. найдутся другие числа x_1', x_2', при которых $\vec{x} = x_1'\vec{e}_1 + x_2'\vec{e}_2$. Вычитая одно равенство из другого и пользуясь дистрибутивностью, получаем $\lambda_1\vec{e}_1 + \lambda_2\vec{e}_2 = \vec{0}$, где $\lambda_1 = x_1' - x_1$, $\lambda_2 = x_2' - x_2$. Пусть, например, $x_1' \neq x_1$, т.е. $\lambda_1 \neq 0$. Тогда можно разделить полученное равенство на это число: $\vec{e}_1 + (\lambda_2/\lambda_1)\vec{e}_2 = \vec{0}$, откуда $\vec{e}_1 = -(\lambda_2/\lambda_1)\vec{e}_2$. Последнее равенство означает, что векторы $\vec{e}_1$ и $\vec{e}_2$ коллинеарны, а это противоречит условиям теоремы. Значит, предположение неверно и разложение единственно.

Говорят, что пара ненулевых неколлинеарных векторов $\vec{e}_1$, $\vec{e}_2$ образует *базис* на плоскости. Равенство $\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2$ называется *разложением* вектора $\vec{x}$ по данному базису, а числа x_1, x_2 – координатами вектора $\vec{x}$ при разложении по базису. Это записывается как $\vec{x} = (x_1, x_2)$.

С базисом на плоскости можно связать *систему координат*. Для этого зафиксируем точку O – начало координат. Каждой точке A на плоскости соответствует вектор $\overrightarrow{OA}$ – *радиус-вектор* этой точки. Координаты радиуса-вектора при разложении по базису $\vec{e}_1$, $\vec{e}_2$ называются ***координатами точки*** в построенной системе координат: $A(x_1, x_2)$ (рис. 4.17).

Самая распространенная система координат образуется двумя взаимно перпендикулярными векторами $\vec{e}_1$, $\vec{e}_2$, длина которых равна 1: $\vec{e}_1 \perp \vec{e}_2$, $|\vec{e}_1| = |\vec{e}_2| = 1$. Такая система координат называется *декартовой прямоугольной системой координат* (рис. 4.18). Вы, конечно, заметили, что именно эта система координат использовалась для построения графиков функций.

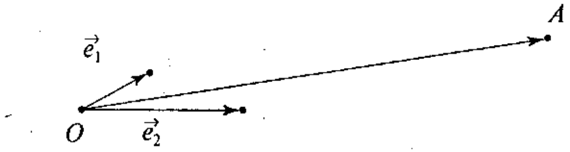


Рис. 4.17

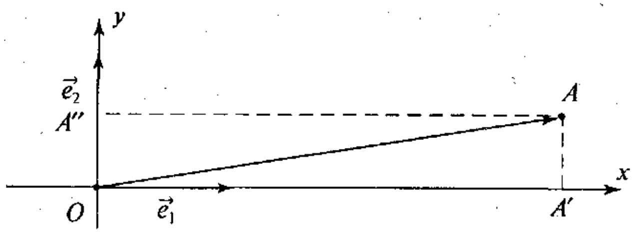


Рис. 4.18

Рене Декарт (1596–1650). Великий французский философ, математик, физик, физиолог. Учился в иезуитской школе, затем изучал медицину и право. Окончил университет в Пуатье (1616). Был на военной службе, много путешествовал. В 1618 г. заинтересовался математикой и физикой.

В главном математическом труде Декарта «Геометрия» (1637) заложены основы аналитической геометрии. Он утверждал, что единственным общим методом математики является алгебраический. Он ввел метод координат и понятие уравнения кривой. В связи с этим отрицательные числа получили пояснение как точки отрицательного направления на координатной оси. Получил важные результаты и в других разделах математики, в частности в теории алгебраических уравнений.

Декарт был основоположником философии картезианства (названного по его латинизированной фамилии Картезий). Он создал общую картину мира, исходя из предположения, что пространство сплошь заполнено материей, находящейся в непрерывном движении. Все природные процессы он сводил к пространственному перемещению. Разработал метод познания, согласно которому главная роль в научном исследовании отводилась разуму как решающему критерию оценки результатов исследования. Считал, что животный организм можно представить в виде машины, так как, по его мнению, все физиологические процессы сводятся к механическому движению.

В декартовой системе координат удобно построить оси: ось абсцисс x идет в направлении вектора $\vec{e}_1$, а ось ординат y — в направлении вектора $\vec{e}_2$. Оси делят плоскость на четыре квадранта. Опустив перпендикуляры из точки A на оси координат, мы получим проекции точки A' и A''. Из определений ясно, что первая координата точки A (ее абсцисса) равна длине отрезка OA', взятой со знаком «+» для точек I и II квадрантов и со знаком «–» для точек III и IV квадрантов. Аналогично вторая координата (ордината) точки A есть длина отрезка OA'' со знаком «+» для точек I и IV квадрантов и со знаком «–» для остальных точек.

Следующие свойства справедливы в любой системе координат (не только в декартовой).

∇12. *Если* $\vec{a} = (x_1, y_1)$, $\vec{b} = (x_2, y_2)$, *то*

$$\vec{a} + \vec{b} = (x_1 + x_2, y_1 + y_2).$$

Равенство следует из свойств операций над векторами. Если $\vec{e}_1$, $\vec{e}_2$ – базис, то $\vec{a} = x_1\vec{e}_1 + y_1\vec{e}_2$, $\vec{b} = x_2\vec{e}_1 + y_2\vec{e}_2$. Тогда $\vec{a} + \vec{b} = x_1\vec{e}_1 + y_1\vec{e}_2 + x_2\vec{e}_1 + y_2\vec{e}_2 = (x_1 + x_2)\vec{e}_1 + (y_1 + y_2)\vec{e}_2 = (x_1 + x_2, y_1 + y_2)$.

∇13. *Если* $\vec{a} = (x, y)$, *то*

$$\lambda\vec{a} = (\lambda x, \lambda y).$$

Свойство ∇13 проверяется аналогично.

∇14. *Вектор* $\overrightarrow{AB}$, *соединяющий точки* $A(x_1, y_1)$ *и* $B(x_2, y_2)$, *имеет координаты* $(x_2 - x_1, y_2 - y_1)$.

Пусть O — начало координат. Тогда по определению координат точки $\overrightarrow{OA} = (x_1, y_1)$, $\overrightarrow{OB} = (x_2, y_2)$ (рис. 4.19).

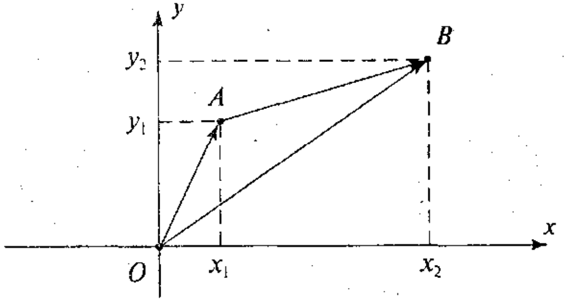


Рис. 4.19

Ясно, что $\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}$ и по свойству ∇12 $\overrightarrow{AB} = (x_2 - x_1, y_2 - y_1)$.

∇15. *Векторы* $\vec{a} = (x_1, y_1)$, $\vec{b} = (x_2, y_2)$ *коллинеарны тогда и только тогда, когда их координаты пропорциональны:*

$$x_1 y_2 = x_2 y_1.$$

Свойство следует из ∇7 и ∇13. Мы не написали $x_1/x_2 = y_1/y_2$, чтобы избежать возможного деления на нуль.

Обратите внимание, что следующее свойство справедливо только в декартовой системе координат.

∇16. *В декартовой системе координат длина вектора* $\vec{a} = (x, y)$ *равна*

$$|\vec{a}| = \sqrt{x^2 + y^2}.$$

Равенство справедливо по теореме Пифагора.

4.1.3. Скалярное произведение

Введем еще одну операцию над векторами. *Углом* между ненулевыми векторами $\vec{a}$ и $\vec{b}$ называется угол между соответствующими лучами, если оба вектора приложить к одной точке (рис. 4.20). Угол между векторами обозначают $\widehat{\vec{a}\,\vec{b}}$.

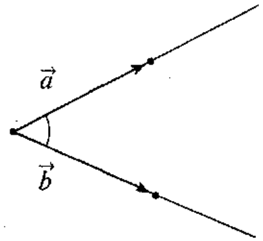


Рис. 4.20

Скалярным произведением векторов $\vec{a}$ и $\vec{b}$, обозначаемым $\vec{a} \cdot \vec{b}$, называется число, равное произведению их длин и косинуса угла между ними:

$$\vec{a} \cdot \vec{b} = |\vec{a}|\,|\vec{b}| \cos\varphi,$$

где $\varphi = \widehat{\vec{a}\,\vec{b}}$.

Следующие свойства скалярного произведения векторов вытекают прямо из определения.

∇17. *Скалярное произведение коммутативно:*

$$\vec{a}\cdot\vec{b}=\vec{b}\cdot\vec{a}.$$

∇18. *Для любого числа* λ *и любых векторов* $\vec{a}$ *и* $\vec{b}$ *справедливо равенство*

$$(\lambda\vec{a})\cdot\vec{b}=\lambda(\vec{b}\cdot\vec{a}).$$

∇19. *Скалярный квадрат равен квадрату длины вектора:*

$$\vec{a}\cdot\vec{a}=\vec{a}^2=|\vec{a}|^2.$$

∇20. *Если* $|\vec{e}|=1$, *то* $(\lambda\vec{e})\cdot(\mu\vec{e})=\lambda\mu$.

∇21. $\vec{a}\cdot\vec{b}=0$ *тогда и только тогда, когда векторы* $\vec{a}$ *и* $\vec{b}$ *перпендикулярны или один из них равен нулю.*

Как известно, чтобы получить проекцию точки на прямую, требуется опустить из точки перпендикуляр на эту прямую. *Проекцией* $\mathrm{pr}_{\vec{b}}\vec{a}$ вектора $\vec{a}$ на вектор $\vec{b}$ (или на направление $\vec{b}$) называется вектор, началом которого служит проекция начала $\vec{a}$, а концом — проекция конца $\vec{a}$ на прямую, содержащую $\vec{b}$. На рис. 4.21 $\overrightarrow{AB}=\vec{a}$, $\overrightarrow{A'B'}=\mathrm{pr}_{\vec{b}}\vec{a}$.

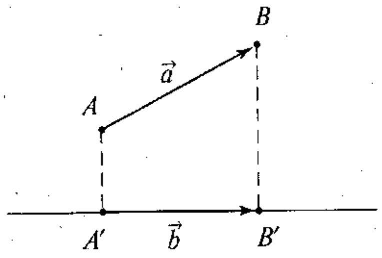


Рис. 4.21

∇22. *Равные векторы имеют равные проекции.*

∇23. $|\mathrm{pr}_{\vec{b}}\vec{a}|=|\vec{a}|\cos\varphi$, *где* $\varphi=\widehat{\vec{a}\,\vec{b}}$.

Это видно из рис. 4.22. Проведем AC параллельно прямой, содержащей $\vec{b}$. Тогда $|\mathrm{pr}_{\vec{b}}\vec{a}|=AC$ и $\angle BAC=\varphi$. Из треугольника ABC имеем $AC=AB\cos\varphi=|\vec{a}|\cos\varphi$.

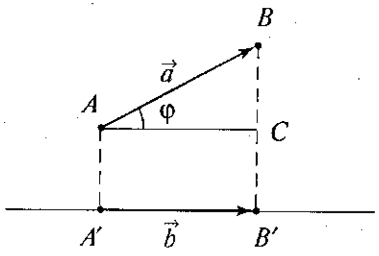


Рис. 4.22

∇24. $\vec{a}\cdot\vec{b}=\mathrm{pr}_{\vec{b}}\vec{a}\cdot\vec{b}$.

Свойство следует из определения скалярного произведения и ∇23.

∇25. *Проекция суммы векторов равна сумме их проекций:*

$$\mathrm{pr}_{\vec{c}}(\vec{a}+\vec{b})=\mathrm{pr}_{\vec{c}}\vec{a}+\mathrm{pr}_{\vec{c}}\vec{b}.$$

Это равенство иллюстрирует рис. 4.23, где $\overrightarrow{AB}=\vec{a}$, $\overrightarrow{BC}=\vec{b}$, $\overrightarrow{AC}=\vec{a}+\vec{b}$ и, стало быть, $\overrightarrow{A'B'}=\mathrm{pr}_{\vec{c}}\vec{a}$, $\overrightarrow{B'C'}=\mathrm{pr}_{\vec{c}}\vec{b}$, $\overrightarrow{A'C'}=\mathrm{pr}_{\vec{c}}(\vec{a}+\vec{b})$.

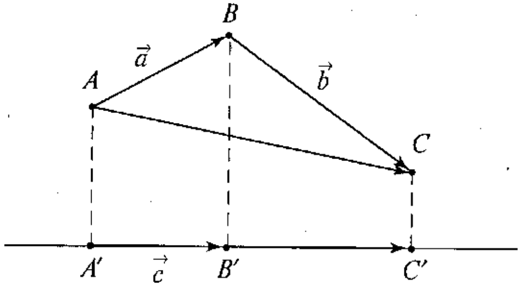


Рис. 4.23

Свойства проекции позволяют доказать следующее важное свойство скалярного произведения векторов.

∇26. *Скалярное произведение дистрибутивно относительно сложения векторов:*

$$(\vec{a}+\vec{b})\cdot\vec{c}=\vec{a}\cdot\vec{c}+\vec{b}\cdot\vec{c}.$$

Доказательство. Если один из векторов равен нулю, равенство очевидно. Пусть все три вектора ненулевые. По свойствам

$\nabla 24$ и $\nabla 25$ имеем $(\vec{a} + \vec{b}) \cdot \vec{c} = \mathrm{pr}_{\vec{c}}(\vec{a} + \vec{b}) \cdot \vec{c} = (\mathrm{pr}_{\vec{c}}\vec{a} + \mathrm{pr}_{\vec{c}}\vec{b}) \cdot \vec{c}$ и $\vec{a} \cdot \vec{c} + \vec{b} \cdot \vec{c} = \mathrm{pr}_{\vec{c}}\vec{a} \cdot \vec{c} + \mathrm{pr}_{\vec{c}}\vec{b} \cdot \vec{c}$. Обозначим через $\vec{e}$ вектор единичной длины, коллинеарный $\vec{c}$. Тогда найдутся числа λ, μ, ν, такие, что $\mathrm{pr}_{\vec{c}}\vec{a} = \lambda\vec{e}$, $\mathrm{pr}_{\vec{c}}\vec{b} = \mu\vec{e}$, $\vec{c} = \nu\vec{e}$. Значит, $(\mathrm{pr}_{\vec{c}}\vec{a} + \mathrm{pr}_{\vec{c}}\vec{b}) \cdot \vec{c} = (\lambda\vec{e} + \mu\vec{e}) \cdot (\nu\vec{e}) = \underline{(\lambda + \mu)\nu}$ и $\mathrm{pr}_{\vec{c}}\vec{a} \cdot \vec{c} + \mathrm{pr}_{\vec{c}}\vec{b} \cdot \vec{c} = (\lambda\vec{e}) \cdot (\nu\vec{e}) + (\mu\vec{e}) \cdot (\nu\vec{e}) = \underline{\lambda\nu + \mu\nu}$ (по свойству $\nabla 20$). Очевидно, что подчеркнутые выражения равны.

Итак, при вычислении скалярного произведения можно раскрывать скобки. Используем указанное свойство для вычисления скалярного произведения в декартовой системе координат. Пусть $\vec{e}_1$, $\vec{e}_2$ – базис декартовой системы координат, т.е. эти векторы перпендикулярны и имеют единичную длину. Тогда по определению $\vec{e}_1 \cdot \vec{e}_2 = 0$, $\vec{e}_1 \cdot \vec{e}_1 = 1$, $\vec{e}_2 \cdot \vec{e}_2 = 1$. Воспользуемся этой «таблицей умножения» для вывода следующего равенства.

$\nabla 27$. *Пусть в декартовой системе координат* $\vec{a} = (x_1, y_1)$, $\vec{b} = (x_2, y_2)$. *Тогда*

$$\vec{a} \cdot \vec{b} = x_1x_2 + y_1y_2.$$

Доказательство. В самом деле, $\vec{a} = x_1\vec{e}_1 + y_1\vec{e}_2$, $\vec{b} = x_2\vec{e}_1 + y_2\vec{e}_2$ и $\vec{a} \cdot \vec{b} = (x_1\vec{e}_1 + y_1\vec{e}_2) \cdot (x_2\vec{e}_1 + y_2\vec{e}_2) = (x_1x_2)(\vec{e}_1 \cdot \vec{e}_1) + (x_1y_2)(\vec{e}_1 \cdot \vec{e}_2) + (x_2y_1)(\vec{e}_2 \cdot \vec{e}_1) + (y_1y_2)(\vec{e}_2 \cdot \vec{e}_2) = (x_1x_2) \cdot 1 + (x_1y_2) \cdot 0 + (x_2y_1) \cdot 0 + (y_1y_2) \cdot 1 = x_1x_2 + y_1y_2$. Равенство доказано.

Итак, мы ознакомились с понятием и основными свойствами свободных векторов на плоскости. В параграфе 4.2 эти свойства переносятся на векторы в пространстве. В дальнейшем мы с удивлением обнаружим, что аналогичные свойства присущи гораздо более абстрактным и менее наглядным математическим объектам. От векторов в реальном трехмерном мире мы перейдем к векторам в n-мерном пространстве и далее – к линейному пространству с элементами произвольной природы.

4.2. Векторы в пространстве

Вектор в пространстве не отличается принципиально от вектора на плоскости. Мы рассмотрели сначала векторы на плоскости в целях простоты и большей наглядности изложения. Повторим кратко основное содержание параграфа 4.1 применительно к векторам в пространстве, отметив некоторые, не слишком существенные отличия.

Вектор в пространстве определяется так же, как на плоскости: это направленный отрезок, имеющий начало (точку приложения) и конец, или упорядоченная пара точек. Не меняется и определение коллинеарности: два вектора коллинеарны, если лежат на параллельных прямых или один из них нулевой. Чтобы повторить определение сонаправленности и антинаправленности, достаточно провести через два коллинеарных вектора плоскость: из школьного курса геометрии известно, что через две параллельные прямые всегда можно провести плоскость, и притом единственную. Соответственно, повторяется и определение равенства векторов. Итак, в пространстве мы тоже имеем дело не с отдельными векторами, а, по сути, с классами эквивалентности по отношению равенства. Это приводит к понятию свободного вектора: данный вектор можно переносить параллельно самому себе, т.е. приложить к любой точке.

Без изменений остаются определения операций над векторами: сложение и умножение на число. Соответственно справедливы и все свойства этих операций, отмеченные в параграфе 4.1. Вспомним их.

Сложение векторов	Умножение на число
$\vec{a} + \vec{b} = \vec{b} + \vec{a}$	$\lambda(\mu\vec{a}) = (\lambda\mu)\vec{a}$
$\vec{a} + (\vec{b} + \vec{c}) = (\vec{a} + \vec{b}) + c$	$(\lambda + \mu)\vec{a} = \lambda\vec{a} + \mu\vec{a}$
$\vec{a} + \vec{0} = \vec{0} + \vec{a} = \vec{a}$	$\lambda(\vec{a} + \vec{b}) = \lambda\vec{a} + \lambda\vec{b}$
$\vec{a} + (-\vec{a}) = \vec{0}$	$1 \cdot \vec{a} = \vec{a}$, $(-\lambda)\vec{a} = -(\lambda\vec{a})$

Три вектора $\vec{a}$, $\vec{b}$, $\vec{c}$ называются *компланарными*, если они лежат в одной плоскости. В следующей теореме описывается разложение вектора по базису в пространстве.

Теорема 4.2. *Пусть векторы* $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ *некомпланарны. Тогда любой вектор* $\vec{x}$ *можно представить в виде*

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3,$$

и притом единственным образом.

Доказательство.

Случай 1. Пусть вектор $\vec{x}$ коллинеарен одному из векторов $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ (например, $\vec{x} \parallel \vec{e}_1$). Тогда можно принять $x_1 = \pm(|\vec{x}| / |\vec{e}_1|)$, $x_2 = x_3 = 0$.

Случай 2. Пусть вектор $\vec{x}$ не коллинеарен ни одному из векторов $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ и лежит в одной плоскости с двумя из них (например, векторы $\vec{x}$, $\vec{e}_1$, $\vec{e}_2$ компланарны). Тогда можно воспользоваться теоремой 4.1 о разложении вектора по базису на плоскости: $\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2$, $x_3 = 0$.

Случай 3. Вектор $\vec{x}$ не лежит в одной плоскости ни с одной из пар векторов $\{\vec{e}_1, \vec{e}_2\}$, $\{\vec{e}_1, \vec{e}_3\}$, $\{\vec{e}_2, \vec{e}_3\}$. Тогда для доказательства проводятся построения, аналогичные построениям в теореме 4.1, только строится не параллелограмм, а параллелепипед (рис. 4.24).

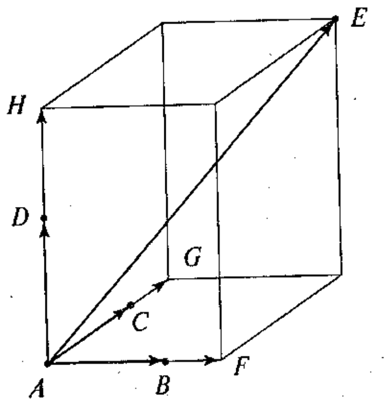


Рис. 4.24

Здесь $\overrightarrow{AB} = \vec{e}_1$, $\overrightarrow{AC} = \vec{e}_2$, $\overrightarrow{AD} = \vec{e}_3$, $\overrightarrow{AE} = \vec{x}$. Продолжим AB до AF, AC до AG, AD до AH так, чтобы на этих трех сторонах построить параллелепипед с диагональю AE. Обозначим $\overrightarrow{AF} = \vec{r}_1$, $\overrightarrow{AG} = \vec{r}_2$, $\overrightarrow{AH} = \vec{r}_3$. Тогда по построению $\vec{x} = \vec{r}_1 + \vec{r}_2 + \vec{r}_3$. Но, поскольку векторы $\vec{r}_i$ коллинеарны $\vec{e}_i$, найдутся числа $x_i = \pm(|\vec{r}_i|/|\vec{e}_i|)$, $i = 1, 2, 3$, такие, что $\vec{r}_i = x_i\vec{e}_i$. Это доказывает разложение. Единственность проверяется так же, как при доказательстве теоремы 4.1.

Тройка некомпланарных векторов в пространстве образует *базис*, а числа x_1, x_2, x_3 в разложении вектора $\vec{x}$ по базису называются *координатами* этого вектора.

Базис в пространстве порождает систему координат. Если отложить базисные векторы от начала координат, то за координаты точки принимаются координаты ее радиуса-вектора (вектора, идущего в данную точку из начала координат). Повторим свойства координат для пространства:

$$\vec{a} = (x_1, y_1, z_1),\ \vec{b} = (x_2, y_2, z_2) \Rightarrow \vec{a} + \vec{b} = (x_1 + x_2, y_1 + y_2, z_1 + z_2)$$

$$\vec{a} = (x, y, z)\lambda \Rightarrow \vec{a} = (\lambda x, \lambda y, \lambda z)$$

$$A(x_1, y_1, z_1),\ B(x_2, y_2, z_2) \Rightarrow \overrightarrow{AB} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$$

$$\vec{a} \parallel \vec{b} \Leftrightarrow x_1y_2 = x_2y_1,\ x_1z_2 = x_2z_1,\ y_1z_2 = y_2z_1$$

В декартовой системе координат $|\vec{a}| = \sqrt{x^2 + y^2 + z^2}$

Наибольший интерес представляет система координат, порожденная тремя взаимно перпендикулярными векторами $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ единичной длины. Такие векторы в пространстве можно расположить по-разному. Введем о р и е н т а ц и ю тройки векторов. Говорят, что векторы $\vec{a}$, $\vec{b}$, $\vec{c}$ образуют *правую* тройку, если наблюдатель, находящийся на конце третьего вектора $\vec{c}$, видит поворот от первого $\vec{a}$ ко второму $\vec{b}$ против часовой стрелки. Чтобы лучше воспринять это определение, представим себе пальцы правой руки: большой палец – вектор $\vec{a}$, указательный – $\vec{b}$ и средний – $\vec{c}$. Если мы «усядемся» на конец среднего пальца, то заметим, что поворот от большого к указательному происходит против часовой стрелки. При повороте по часовой стрелке векторы образуют *левую* тройку: это расположение первых пальцев левой руки.

Три вектора можно упорядочить шестью разными способами. При этом тройки $\{\vec{a}, \vec{b}, \vec{c}\}$, $\{\vec{b}, \vec{c}, \vec{a}\}$, $\{\vec{c}, \vec{a}, \vec{b}\}$ имеют одинаковую ориентацию, а тройки $\{\vec{b}, \vec{a}, \vec{c}\}$, $\{\vec{a}, \vec{c}, \vec{b}\}$, $\{\vec{c}, \vec{b}, \vec{a}\}$ – противоположную (на рис. 4.25, *а* и *б*, где $\overrightarrow{AB} = \vec{a}$, $\overrightarrow{AC} = \vec{b}$, $\overrightarrow{AD} = \vec{c}$, соответственно первая тройка – правая, а вторая – левая).

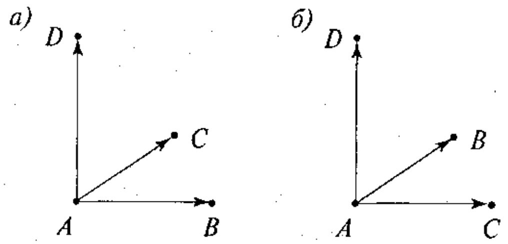


Рис. 4.25

Система координат называется *декартовой прямоугольной*, если образована тремя взаимно перпендикулярными векторами единичной длины, составляющими правую тройку. Базисные векторы иногда обозначают $\vec{i}$, $\vec{j}$, $\vec{k}$. Соответствующие оси есть ось абсцисс x, ось ординат y и ось аппликат z. Чтобы найти координаты точки A, опустим из нее перпендикуляр на плоскость xOy (спроецируем на эту плоскость) (рис. 4.26). Затем найдем проекцию полученной точки на оси x и y. Получим соответственно точки A' и A''. Абсцисса равна длине отрезка OA', ордината – длине отрезка OA'', взятых с соответствующими знаками. Аналогично строится точка A''' на оси z и определяется аппликата точки A.

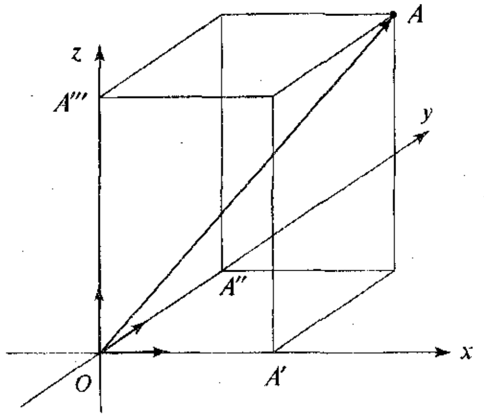


Рис. 4.26

Угол между двумя векторами определяется так же, как на плоскости: достаточно отложить эти векторы от одной точки и поместить на плоскость. Соответственно остаются без изменений определение скалярного произведения и его свойства:

$$\vec{a}\cdot\vec{b} = \vec{b}\cdot\vec{a}$$
$$(\lambda\vec{a})\cdot\vec{b} = \lambda(\vec{a}\cdot\vec{b})$$
$$\vec{a}^2 = \vec{a}\cdot\vec{a} = |\vec{a}|^2$$
$$(\vec{a}+\vec{b})\cdot\vec{c} = \vec{a}\cdot\vec{c} + \vec{b}\cdot\vec{c}$$
$$\vec{a}\cdot\vec{b} = 0 \Leftrightarrow \vec{a}\perp\vec{b}$$

В декартовой системе координат $\vec{a}\cdot\vec{b} = x_1x_2 + y_1y_2 + z_1z_2$

В пространстве вводятся еще две операции над векторами.

Векторное произведение. *Векторным произведением* двух векторов $\vec{a}$ и $\vec{b}$, обозначаемым $\vec{a}\times\vec{b}$, называется вектор, определяемый следующим образом:

- длина $\vec{a}\times\vec{b}$ равна $|\vec{a}|\,|\vec{b}|\sin\varphi$, где $\varphi = \widehat{\vec{a}\,\vec{b}}$;
- вектор $\vec{a}\times\vec{b}$ перпендикулярен векторам $\vec{a}$ и $\vec{b}$;
- векторы $\vec{a}$, $\vec{b}$, $\vec{a}\times\vec{b}$ образуют правую тройку.

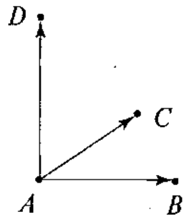


Рис. 4.27

На рис. 4.27 $\overrightarrow{AB} = \vec{a}$, $\overrightarrow{AC} = \vec{b}$, $\overrightarrow{AD} = \vec{a}\times\vec{b}$. Из определения ясно, что $\vec{a}\times\vec{b} = \vec{0}$, если один из них нулевой или они коллинеарны (в последнем случае $\sin\varphi = 0$).

Справедливы следующие свойства векторного произведения:

∇1. $\vec{a}\times\vec{a} = \vec{0}$.

∇2. $(\lambda\vec{a})\times\vec{b} = \lambda(\vec{a}\times\vec{b})$.

∇3. *Векторное произведение антикоммутативно:*

$$\vec{a}\times\vec{b} = -\vec{b}\times\vec{a}.$$

Еще одна некоммутативная операция!!

∇4. *Векторное произведение дистрибутивно относительно сложения векторов:*

$$(\vec{a}+\vec{b})\times\vec{c} = \vec{a}\times\vec{c} + \vec{b}\times\vec{c}.$$

∇5. *Длина* $\vec{a}\times\vec{b}$ *равна площади параллелограмма, построенного на векторах* $\vec{a}$ *и* $\vec{b}$.

Смешанное произведение. Пусть даны три вектора $\vec{a}$, $\vec{b}$, $\vec{c}$. *Смешанным произведением* называется число, равное скалярному произведению векторного произведения $\vec{a}\times\vec{b}$ на вектор $\vec{c}$:

$$\vec{a}\,\vec{b}\,\vec{c} = (\vec{a}\times\vec{b})\cdot\vec{c}.$$

∇**6.** *Смешанное произведение равно объему параллелепипеда, построенного на векторах $\vec{a}$, $\vec{b}$, $\vec{c}$ и взятого со знаком «+», если тройка правая, и со знаком «–», если тройка левая:*

$$\vec{a}\,\vec{b}\,\vec{c} = \pm V.$$

∇**7.** *В смешанном произведении не важно, в каком порядке брать векторное и скалярное произведение:*

$$\vec{a}\,\vec{b}\,\vec{c} = (\vec{a}\times\vec{b})\cdot\vec{c} = \vec{a}\cdot(\vec{b}\times\vec{c}).$$

Это свойство и позволяет обозначать смешанное произведение просто $\vec{a}\,\vec{b}\,\vec{c}$.

∇**8.** *Смешанное произведение равно нулю тогда и только тогда, когда векторы $\vec{a}$, $\vec{b}$, $\vec{c}$ компланарны.*

В самом деле, равенство $\vec{a}\,\vec{b}\,\vec{c} = 0$ означает, что векторы $\vec{a}\times\vec{b}$ и $\vec{c}$ перпендикулярны. Но $\vec{a}\times\vec{b}$ есть вектор, перпендикулярный $\vec{a}$ и $\vec{b}$. Стало быть, $\vec{a}\times\vec{b}$ перпендикулярен всем трем векторам одновременно. Это возможно лишь в случае, когда они лежат в одной плоскости.

В параграфе 4.3 вводится понятие определителя, с помощью которого мы выразим векторное и смешанное произведение через координаты векторов.

4.3. Определители 2-го порядка

Пусть четыре числа образуют таблицу из двух рядов и двух столбцов

$$A = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}.$$

Говорят, что задана квадратная матрица 2×2. ***Определителем*** этой матрицы (определителем 2-го порядка) называется число

$$\det A = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - b_1 a_2.$$

Например,

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 1\cdot 4 - 2\cdot 3 = -2.$$

Воспользуемся языком определителей, чтобы переформулировать условия коллинеарности векторов на плоскости и в пространстве. В самом деле, векторы $\vec{a} = (x_1, y_1)$ и $\vec{b} = (x_2, y_2)$ коллинеарны тогда и только тогда, когда их координаты пропорциональны, т.е. $x_1y_2 - x_2y_1 = 0$. Левая часть этого условия есть не что иное, как определитель, составленный из координат векторов.

∇**1.** *Два вектора $\vec{a} = (x_1, y_1)$ и $\vec{b} = (x_2, y_2)$ коллинеарны тогда и только тогда, когда*

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = 0.$$

Аналогично записывается условие пропорциональности координат векторов в пространстве.

∇**2.** *Два вектора $\vec{a} = (x_1, y_1, z_1)$ и $\vec{b} = (x_2, y_2, z_2)$ коллинеарны тогда и только тогда, когда*

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} = \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} = 0.$$

Определители используются для исследования систем линейных уравнений. Рассмотрим систему уравнений с двумя неизвестными

$$\begin{cases} a_1x + b_1y = c_1, \\ a_2x + b_2y = c_2, \end{cases}$$

где a_1, b_1, a_2, b_2 – коэффициенты при неизвестных;
c_1, c_2 – свободные члены.

Чтобы решить систему, домножим первое уравнение на b_2, второе на b_1 и вычтем второе из первого. Затем домножим первое на a_2, второе – на a_1 и вычтем первое из второго. Получим два равенства:

$$(a_1b_2 - a_2b_1)x = b_2c_1 - b_1c_2;$$
$$(a_1b_2 - a_2b_1)y = a_1c_2 - a_2c_1.$$

Введем три определителя:

$$\Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}, \quad \Delta_x = \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}, \quad \Delta_y = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}.$$

Определитель Δ составлен из коэффициентов при неизвестных, Δ_x получается из Δ заменой первого столбца на столбец свободных членов, Δ_y – аналогичной заменой второго столбца. По определению нетрудно заметить, что полученные равенства можно записать в виде

$$\Delta\cdot x = \Delta_x, \quad \Delta\cdot y = \Delta_y.$$

Рассмотрим возможные случаи.

1. $\Delta \neq 0$. В этом случае система имеет единственное решение

$$x = \Delta_x/\Delta, \quad y = \Delta_y/\Delta.$$

Это известное *правило Крамера*. Решим, например, систему

$$\begin{cases} 2x + y = 2, \\ x + 2y = 3, \end{cases}$$

в которой

$$\Delta = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3, \quad \Delta_x = \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} = 1, \quad \Delta_y = \begin{vmatrix} 2 & 2 \\ 1 & 3 \end{vmatrix} = 4$$

и, стало быть, $x = {}^1/_3$, $y = {}^4/_3$. Легко убедиться, что это действительно решение, подставив его в систему.

В частности, однородная система

$$\begin{cases} a_1x + b_1y = 0, \\ a_2x + b_2y = 0 \end{cases}$$

имеет при $\Delta \neq 0$ единственное решение $x = y = 0$.

2. $\Delta = 0$. Случай распадается на две возможности:

а) $\Delta_x = \Delta_y = 0$. Такая система имеет бесконечно много решений. На самом деле, коэффициенты и свободные члены уравнений системы в данном случае пропорциональны и одно уравнение следует из другого. Если $b_1 \neq 0$, любая пара вида $(x, (-a_1x + c_1)/b_1)$ образует решение системы;

б) $\Delta_x \neq 0$ или $\Delta_y \neq 0$. Такая система не имеет решений, так как одно из равенств противоречиво. Пусть, например, $\Delta_x \neq 0$. Тогда равенство $\Delta \cdot x = \Delta_x$ не может выполняться. В этом случае говорят, что система несовместна.

Итак, проведено полное исследование системы линейных уравнений с помощью определителей.-

Применим определители для вычисления векторного и смешанного произведения векторов в декартовой системе координат. Прежде всего составим таблицу векторного умножения для базисных векторов. Базисные векторы $\vec{i}$, $\vec{j}$, $\vec{k}$ взаимно перпендикулярны, имеют единичную длину и образуют правую тройку. Поэтому в силу определения векторного произведения

$$\vec{i} \times \vec{j} = \vec{k}, \quad \vec{j} \times \vec{k} = \vec{i}, \quad \vec{k} \times \vec{i} = \vec{j}.$$

Кроме того, векторный квадрат равен нулю:

$$\vec{i} \times \vec{i} = \vec{0}, \quad \vec{j} \times \vec{j} = \vec{0}, \quad \vec{k} \times \vec{k} = \vec{0}.$$

Пусть $\vec{a} = (x_1, y_1, z_1) = x_1\vec{i} + y_1\vec{j} + z_1\vec{k}$, $\vec{b} = (x_2, y_2, z_2) = x_2\vec{i} + y_2\vec{j} + z_2\vec{k}$. Тогда по свойствам векторного произведения справедливы равенства

$$\vec{a} \times \vec{b} = (x_1\vec{i} + y_1\vec{j} + z_1\vec{k}) \times (x_2\vec{i} + y_2\vec{j} + z_2\vec{k}) =$$

$$= \underbrace{x_1x_2(\vec{i} \times \vec{i}) + y_1y_2(\vec{j} \times \vec{j}) + z_1z_2(\vec{k} \times \vec{k})}_{\vec{0}} +$$

$$+ x_1y_2\underbrace{(\vec{i} \times \vec{j})}_{\vec{k}} + y_1x_2\underbrace{(\vec{j} \times \vec{i})}_{-\vec{k}} + y_1z_2\underbrace{(\vec{j} \times \vec{k})}_{\vec{i}} + z_1y_2\underbrace{(\vec{k} \times \vec{j})}_{-\vec{i}} +$$

$$+ x_1z_2\underbrace{(\vec{i} \times \vec{k})}_{-\vec{j}} + z_1x_2\underbrace{(\vec{k} \times \vec{i})}_{\vec{j}} =$$

$$= (y_1z_2 - z_1y_2)\vec{i} + (z_1x_2 - x_1z_2)\vec{j} + (x_1y_2 - y_1x_2)\vec{k}.$$

Приглядевшись повнимательнее к полученному выражению, мы заметим знакомые разности произведений. В самом деле,

$$\begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} = y_1z_2 - z_1y_2, \quad \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} = x_1z_2 - z_1x_2,$$

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = x_1y_2 - y_1x_2.$$

Стало быть, выведена формула

$$\boxed{\vec{a} \times \vec{b} = \left(\begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix}, \ -\begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix}, \ \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \right).}$$

Легко понять, как строятся определители, задающие координаты векторного произведения. Выпишем координаты векторов $\vec{a}$ и $\vec{b}$ друг под другом, т.е. составим таблицу шести чисел с двумя строками и тремя столбцами:

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{pmatrix}.$$

Чтобы получить первую координату векторного произведения, вычеркнем первый столбец. Оставшиеся четыре числа образуют нужный определитель. Далее вычеркиваем второй столбец и берем получившийся определитель с обратным знаком. Наконец, третья координата равна определителю, который остается после вычеркивания третьего столбца.

Зная выражение для координат векторного произведения, получим формулу смешанного произведения. Пусть $\vec{c} = (x_3, y_3, z_3)$. Тогда

$$\vec{a}\vec{b}\vec{c} = (\vec{a}\times\vec{b})\cdot\vec{c} = x_3\begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} - y_3\begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} + z_3\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}.$$

Вспоминая, что равенство нулю смешанного произведения есть необходимое и достаточное условие компланарности трех векторов (см. свойство ∇8 из параграфа 4.2), переформулируем этот факт через координаты в декартовой системе.

∇3. *Векторы* $\vec{a} = (x_1, y_1, z_1)$, $\vec{b} = (x_2, y_2, z_2)$, $\vec{c} = (x_3, y_3, z_3)$ *компланарны тогда и только тогда, когда*

$$x_3\begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} - y_3\begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} + z_3\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = 0.$$

4.4. Прямая на плоскости

4.4.1. Уравнение линии

Пусть выбрана система координат на плоскости. ***Уравнением линии*** называется такое уравнение $f(x, y) = 0$, которому удовлетворяют координаты каждой точки, лежащей на этой линии, и не удовлетворяют координаты ни одной из точек, не лежащих на ней. Уравнение $f(x, y) = 0$ называют еще *общим* уравнением линии или уравнением ***в неявной форме***. Таким образом, *линия* Γ есть геометрическое место точек плоскости, координаты которых удовлетворяют уравнению $f(x, y) = 0$, т.е.

$$\Gamma = \{(x, y) \mid f(x, y) = 0\}.$$

Линию называют также *кривой*. В дальнейшем мы будем рассматривать кривые в декартовой прямоугольной системе координат.

Примеры линий.

1. $x - y = 0$. Этому уравнению удовлетворяют точки, у которых абсцисса равна ординате, т.е. данное уравнение описывает биссектрису I и III квадрантов (рис. 4.28, *а*).

2. $x + y = 0$. Это биссектриса II и IV квадрантов (рис. 4.28, *б*).

3. $x^2 - y^2 = 0$. Этому уравнению удовлетворяют координаты точек обеих биссектрис (и только они) (рис. 4.28, *в*).

4. $x^2 + y^2 = 0$. Этому уравнению удовлетворяет единственная точка — начало координат (0, 0) (рис. 4.28, *г*). Такие линии называются вырожденными.

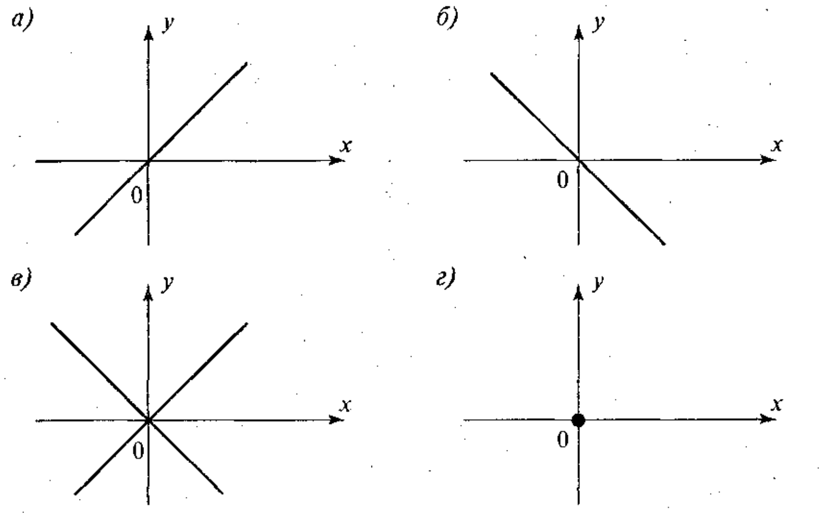


Рис. 4.28

∇1. *Кривая не обязательно является графиком некоторой функции.*

В аналитической геометрии обычно ставятся две задачи:

1) по заданным геометрическим свойствам кривой составить ее уравнение;

2) по заданному уравнению выяснить геометрические свойства кривой.

Приведем пример с окружностью.

1. Пусть $A_0(x_0, y_0)$ — произвольная точка на плоскости, R — положительное число. Составим уравнение окружности с центром в точке $A_0(x_0, y_0)$ и радиусом R. Пусть $A(x, y)$ — точка на окружности (рис. 4.29). По определению окружность есть геометрическое место точек, равноудаленных от центра на расстояние R. Вектор $\overrightarrow{A_0A}$ имеет координаты $(x - x_0, y - y_0)$, и, значит, квадрат его длины равен $(x - x_0)^2 + (y - y_0)^2 = R^2$.

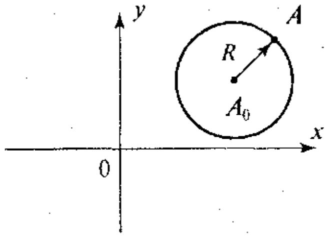


Рис. 4.29

Итак, если точка лежит на окружности, то ее координаты удовлетворяют уравнению

$$(x - x_0)^2 + (y - y_0)^2 - R^2 = 0.$$

И наоборот, если координаты точки удовлетворяют данному уравнению, то она лежит на окружности. Таким образом, найдено общее уравнение окружности с центром в заданной точке и заданным радиусом.

2. Пусть задано уравнение

$$x^2 + y^2 + 2ax + 2by + c = 0,$$

в котором $a^2 + b^2 > c$. Тогда уравнение можно переписать в виде

$$(x + a)^2 + (y + b)^2 - (a^2 + b^2 - c) = 0$$

или

$$(x+a)^2 + (y+b)^2 - \left(\sqrt{a^2+b^2-c}\right)^2 = 0.$$

Из предыдущего примера ясно, что данное уравнение есть уравнение окружности с центром в точке $A_0(-a, -b)$ и радиусом $R = \sqrt{a^2+b^2-c}$.

Представим себе, что мы едем на поезде и железная дорога имеет форму некоторой кривой. Пусть у нас есть подробное расписание движения, т.е. в каждый момент времени t мы можем определить, где находимся. Это значит, что в каждый момент времени нам известны наши координаты. Иначе говоря, заданы две функции $f_1(t)$ и $f_2(t)$, такие, что

$$\begin{cases} x = f_1(t), \\ y = f_2(t). \end{cases}$$

Такие уравнения называются *параметрическими* уравнениями кривой или уравнением *в параметрической форме*. Величина $t \in R$ есть *параметр*, который не обязательно, как в нашем примере, играет

роль времени. При задании параметрических уравнений кривой требуется указать, в каких пределах изменяется параметр, т.е. указать некоторое числовое множество T, из которого берутся значения параметра. Обычно это отрезок, конечный или бесконечный.

Построим параметрические уравнения окружности с центром в начале координат и радиусом R. В качестве параметра t удобно взять угол между радиусом-вектором произвольной точки на окружности и осью абсцисс. Из рис. 4.30 ясно, что справедливы равенства

$$x = R\cos t, \quad y = R\sin t, \quad t \in [0, 2\pi].$$

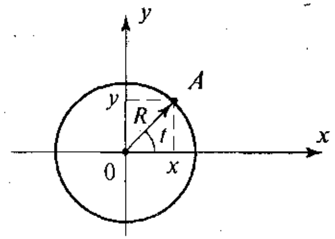


Рис. 4.30

Это и есть параметрические уравнения окружности. Окружность с центром в точке (x_0, y_0) имеет уравнения

$$x = R\cos t + x_0, \quad y = R\sin t + y_0, \quad t \in [0, 2\pi].$$

Чтобы перейти от параметрических уравнений кривой к общему уравнению, нужно исключить параметр t из параметрических уравнений. Для этого можно, например, выразить параметр из одного уравнения и подставить в другое. В случае окружности поступим иначе: возведем в квадрат равенства $x - x_0 = R\cos t$, $y - y_0 = R\sin t$ и сложим полученные равенства: $(x - x_0)^2 + (y - y_0)^2 = R^2(\cos^2 t + \sin^2 t) = R^2$. Это и есть общее уравнение окружности.

Обозначим радиус-вектор точки, лежащей на кривой, через $\vec{r}$. Тогда параметрические уравнения легко записать в виде *векторного* уравнения кривой

$$\vec{r} = (f_1(t), f_2(t)).$$

4.4.2. Уравнения прямой

Из школьного курса геометрии нам хорошо известно, что такое прямая. Попробуем определить прямую на языке векторов. Зафиксируем некоторую точку (x_0, y_0) на плоскости и зададим вектор $\vec{s}$ (так называемый *направляющий вектор*). Если откладывать от точки

(x_0, y_0) всевозможные векторы, коллинеарные данному вектору $\vec{s}$, то концы этих векторов составят множество точек, которое и является прямой. Построенная прямая проходит через точку (x_0, y_0) в направлении $\vec{s}$.

Обозначим через $\vec{r}_0 = \overrightarrow{OA}$ радиус-вектор точки (x_0, y_0), а через $\vec{r} = \overrightarrow{OC}$ – радиус-вектор произвольной точки (x, y) на прямой (рис. 4.31). Тогда по построению вектор $\vec{r} - \vec{r}_0 = \overrightarrow{AC}$ коллинеарен направляющему вектору $\vec{s} = \overrightarrow{AB}$. Это значит, что найдется действительное число t, такое, что $\vec{r} - \vec{r}_0 = t\vec{s}$ или $\vec{r} = \vec{r}_0 + t\vec{s}$. Перебирая все возможные значения параметра $t \in \boldsymbol{R}$, мы получим радиусы-векторы всех точек данной прямой. Итак, построено векторное уравнение прямой

$$\vec{r} = \vec{r}_0 + t\vec{s}.$$

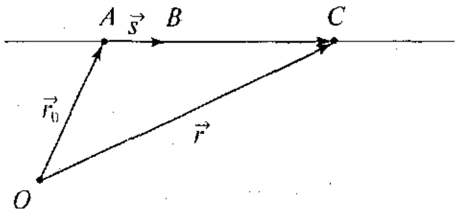


Рис. 4.31

Если $\vec{s} = (s_1, s_2)$, легко записать параметрические уравнения прямой:

$$\begin{cases} x = x_0 + ts_1, \\ y = y_0 + ts_2, \end{cases}$$

где $-\infty \leqslant t \leqslant \infty$.

Исключим параметр t. Для этого можно, например, первое равенство умножить на s_2, а второе – на s_1: $(x - x_0)s_2 = ts_1s_2$ и $(y - y_0)s_1 = ts_1s_2$. Тогда $(x - x_0)s_2 = (y - y_0)s_1$, откуда, раскрыв скобки, получим $s_2x - s_1y - s_2x_0 + s_1y_0 = 0$. Обозначим $A = s_2$, $B = -s_1$, $C = -s_2x_0 + s_1y_0$. Тогда получается общее уравнение прямой

$$Ax + By + C = 0.$$

На самом деле, мы проверили только то, что координаты точки, которая лежит на прямой, проходящей через точку (x_0, y_0) в направлении вектора $\vec{s} = (s_1, s_2)$, удовлетворяют этому уравнению. Нетрудно проверить и обратное утверждение. Итак, данное уравнение действительно является общим уравнением прямой. Из проделанных выкладок ясно, что прямая, заданная своим общим уравнением, имеет направляющий вектор $\vec{s} = (-B, A)$. Заметим, что в общее уравнение входят только первые степени x и y (в отличие, скажем,

от уравнения окружности). Поэтому прямая на плоскости есть линия *первого порядка*.

Пусть задан вектор $\vec{n}$, перпендикулярный направляющему вектору $\vec{s}$ (рис. 4.32). Вектор $\vec{n}$ – это так называемая *нормаль*. Тогда $\vec{n} \perp (\vec{r} - \vec{r}_0)$ и по свойству скалярного произведения справедливо равенство

$$\vec{n} \cdot (\vec{r} - \vec{r}_0) = 0,$$

которое называется *нормальным* уравнением прямой.

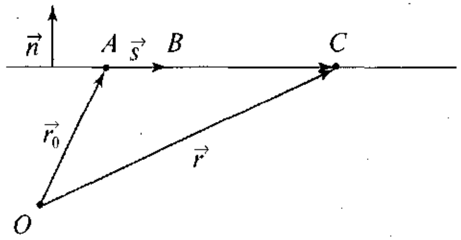


Рис. 4.32

Поскольку точка (x_0, y_0) лежит на прямой, ее координаты удовлетворяют общему уравнению $Ax_0 + By_0 + C = 0$. Вычитая это равенство из общего уравнения, получаем равенство $A(x - x_0) + B(y - y_0) = 0$, которое совпадает с нормальным уравнением при $\vec{n} = (A, B)$. Итак, если задано общее уравнение прямой, то нормаль к ней имеет координаты $\vec{n} = (A, B)$. При этом очевидно, что $C = -\vec{n} \cdot \vec{r}_0$.

Пусть задано общее уравнение прямой и $B \neq 0$. Тогда уравнение можно разделить на B и, обозначив $k = -A/B$, $b = -C/B$, получить хорошо знакомое «школьное» уравнение прямой

$$y = kx + b.$$

Другими словами, при $B \neq 0$ прямая представляет собой график линейной функции. Коэффициент k равен тангенсу угла наклона прямой к оси x и называется *угловым коэффициентом* прямой. Число b – это ордината точки пересечения прямой с осью y: говорят, что прямая отсекает отрезок b на оси y (рис. 4.33).

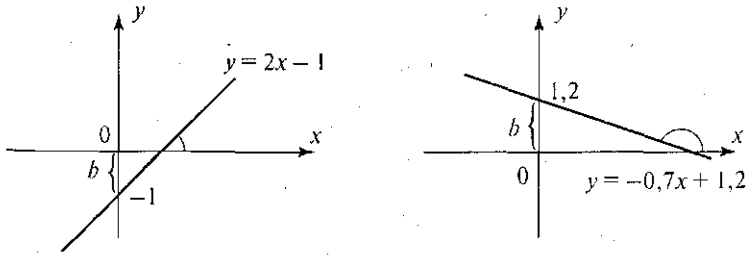


Рис. 4.33

Предположим теперь, что все коэффициенты общего уравнения не равны нулю. Если разделить общее уравнение на C, получится равенство $(A/C)x + (B/C)y + 1 = 0$, которое можно записать в виде $\frac{x}{-C/A} + \frac{y}{-C/B} = 1$ или, обозначив $a = -C/A$, $b = -C/B$, в виде

$$\frac{x}{a} + \frac{y}{b} = 1.$$

Такое уравнение называется «*уравнением в отрезках*». В самом деле, при $x = 0$ имеем $y = b$, а при $y = 0$ соответственно $x = a$. Таким образом, числа a и b есть отрезки, отсекаемые прямой на осях координат (рис. 4.34).

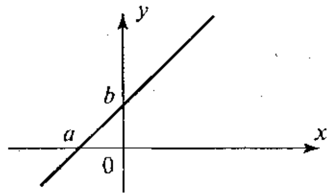


Рис. 4.34

В итоге получено шесть различных уравнений прямой:

Векторное: $\vec{r} = \vec{r}_0 + t\vec{s}$
Параметрические: $\begin{cases} x = x_0 + ts_1 \\ y = y_0 + ts_2 \end{cases}$
Общее: $Ax + By + C = 0$
Нормальное: $\vec{n} \cdot (\vec{r} - \vec{r}_0) = 0$
«Школьное»: $y = kx + b$
В отрезках: $\frac{x}{a} + \frac{y}{b} = 1$
$\vec{s} = (-B, A)$, $\vec{n} = (A, B)$, $C = -\vec{n} \cdot \vec{r}_0$

Пример.

Рассмотрим прямую, заданную общим уравнением $2x + 3y + 1 = 0$ с коэффициентами $A = 2$, $B = 3$, $C = 1$ (рис. 4.35).

Ее направляющий вектор $\vec{s} = (-B, A) = (-3, 2)$, вектор нормали $\vec{n} = (A, B) = (2, 3)$. В качестве начального вектора $\vec{r}_0$ можно взять радиус-вектор любой точки, лежащей на прямой, например $\vec{r}_0 = (x_0, y_0) = (1, -1)$. В самом деле, $2 \cdot 1 + 3 \cdot (-1) + 1 = 0$, значит,

данная точка принадлежит прямой. Проверим, например, что $C = -\vec{n} \cdot \vec{r}_0$: $-(2, 3) \cdot (1, -1) = -(2 \cdot 1 + 3 \cdot (-1)) = -(2 - 3) = 1 = C$.

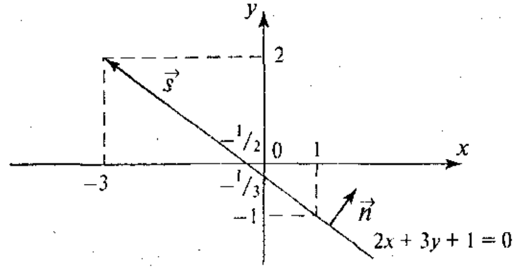


Рис. 4.35

Убедимся, что тот же результат получится, если в качестве $\vec{r}_0$ взять любую другую точку на прямой. Для этого выразим y через x, т.е. выпишем «школьное» уравнение прямой $y = (-^2/_3)x - {}^1/_3$. Из него, кстати, видно, что прямая наклонена к оси x под тупым углом ($k = -^2/_3 < 0$) и пересекает ось ординат в точке $(0, -^1/_3)$. Пусть произвольная точка прямой имеет абсциссу x_0. Тогда ее ордината равна $(-^2/_3)x_0 - {}^1/_3$. Посчитаем в этом случае величину $-\vec{n} \cdot \vec{r}_0$: $-(2, 3) \cdot (x_0, (-^2/_3)x_0 - {}^1/_3) = -(2x_0 + 3 \cdot (-^2/_3)x_0 + 3 \cdot (-^1/_3)) = -(2x_0 - 2x_0 - 1) = 1 = C$. Таким образом, неважно, какую именно точку взять за начальную.

Найденные векторы $\vec{s}$ и $\vec{r}_0$ дают следующие параметрические уравнения прямой:

$$\begin{cases} x = 1 - 3t, \\ y = -1 + 2t. \end{cases}$$

Выведем еще уравнение в отрезках. Для этого коэффициенты общего уравнения разделим на $-C = -1$: $-2x - 3y = 1$, или $\frac{x}{-1/2} + \frac{y}{-1/3} = 1$. Из уравнения видно, что прямая пересекает ось абсцисс в точке $(-^1/_2, 0)$. Точку пересечения с осью ординат $(0, -^1/_3)$ мы уже нашли.

4.4.3. Расположение прямой на плоскости. Взаимное расположение прямых

Рассмотрим некоторые частные случаи расположения прямой на плоскости. Напомним, что первый базисный вектор декартовой системы координат, соответствующий оси x, обозначается через $\vec{i}$, а второй (порождающий ось y) — через $\vec{j}$.

1. $\vec{s} \parallel \vec{i}$: вторая координата направляющего вектора $s_2 = 0$. Поскольку $\vec{s} = (-B, A)$, в этом случае $A = 0$. Такая прямая параллельна оси абсцисс (рис. 4.36, *а*) и ее уравнение можно записать в виде $y = -C/B$ ($B \neq 0$, так как коэффициенты A и B не могут обращаться в нуль одновременно).

2. $\vec{s} \parallel \vec{j}$. Тогда $s_1 = 0$, $B = 0$. Прямая параллельна оси ординат: $x = -C/A$. Как раз в этом случае прямая не является графиком линейной функции (рис. 4.36, *б*).

3. $\vec{r}_0 \parallel \vec{s}$: радиус-вектор точки на прямой коллинеарен направляющему вектору. Это может быть только в случае, когда прямая проходит через начало координат (рис. 4.36, *в*). Из коллинеарности векторов $\vec{r}_0$ и $\vec{s}$ следует пропорциональность их координат: $s_1y_0 - s_2x_0 = 0$, т.е. $C = 0$.

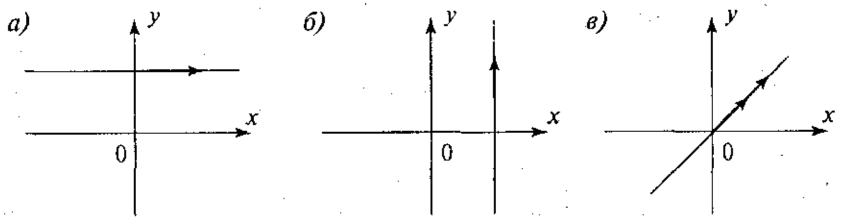


Рис. 4.36

Вопрос о взаимном расположении двух прямых можно сформулировать так: есть ли у них общие точки и если есть, то сколько? Пусть две прямые заданы своими векторными уравнениями:

$$\vec{r} = \vec{r}_1 + t\vec{s}_1;$$
$$\vec{r} = \vec{r}_2 + t\vec{s}_2.$$

Теперь вопрос сводится к тому, есть ли такие числа t_1, t_2, что выполняется равенство $\vec{r}_1 + t\vec{s}_1 = \vec{r}_2 + t\vec{s}_2$ (т.е. для каждой прямой нашлось значение параметра t, порождающее один и тот же радиус-вектор $\vec{r}$). Запишем данное уравнение по-другому:

$$t_1\vec{s}_1 - t_2\vec{s}_2 = \vec{r}_2 - \vec{r}_1$$

и рассмотрим различные случаи.

1. $\vec{s}_1$ и $\vec{s}_2$ неколлинеарны. Тогда они образуют базис на плоскости и любой вектор можно разложить по этому базису единственным образом. В частности, можно разложить и вектор $\vec{r}_2 - \vec{r}_1$. Это

значит, что числа t_1, t_2 в данном случае определяются однозначно. Итак, если направляющие векторы двух прямых $\vec{s}_1$ и $\vec{s}_2$ неколлинеарны, прямые пересекаются в единственной точке (рис. 4.37, *а*). Условие неколлинеарности направляющих векторов $\vec{s}_1$ и $\vec{s}_2$ или, что то же самое, неколлинеарности нормалей $\vec{n}_1$ и $\vec{n}_2$ можно записать через коэффициенты общих уравнений прямых. Если

$$A_1x + B_1y + C_1 = 0,$$
$$A_2x + B_2y + C_2 = 0,$$

то $\vec{n}_1 = (A_1, B_1)$, $\vec{n}_2 = (A_2, B_2)$. Эти векторы неколлинеарны, если их координаты непропорциональны, т.е. если

$$A_1B_2 - A_2B_1 \neq 0.$$

В частности, прямые могут быть перпендикулярны. Для этого достаточно, чтобы перпендикулярны были их направляющие векторы или нормали: $\vec{n}_1 \perp \vec{n}_2$. Условие перпендикулярности векторов — равенство нулю их скалярного произведения: $\vec{n}_1 \cdot \vec{n}_2 = A_1A_2 + B_1B_2 = 0$. Таким образом, получено условие перпендикулярности прямых:

$$A_1A_2 + B_1B_2 = 0.$$

2. $\vec{s}_1 \parallel \vec{s}_2$: направляющие векторы прямых коллинеарны. В этом случае имеется две возможности:

а) $(\vec{r}_2 - \vec{r}_1) \parallel \vec{s}_1$. Тогда прямые совпадают, т.е. все их точки общие (рис. 4.37, *б*).

б) векторы $\vec{r}_2 - \vec{r}_1$ и $\vec{s}_1$ неколлинеарны. В этом случае прямые параллельны, т.е. не имеют общих точек (рис. 4.37, *в*).

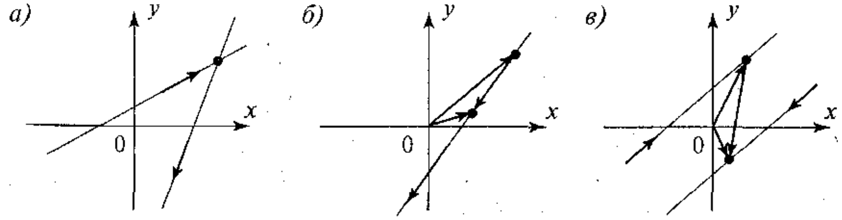


Рис. 4.37

Условие параллельности или совпадения прямых: $\vec{s}_1 \parallel \vec{s}_2$ или $\vec{n}_1 \parallel \vec{n}_2$. Пропорциональность координат через коэффициенты общих уравнений записывается в виде условия

$$A_1B_2 - A_2B_1 = 0.$$

Проанализируем взаимное расположение прямых с помощью определителей. Эта задача сводится к вопросу о существовании и числе решений системы линейных уравнений

$$\begin{cases} A_1x + B_1y + C_1 = 0, \\ A_2x + B_2y + C_2 = 0 \end{cases}$$

или, что то же самое,

$$\begin{cases} A_1x + B_1y = -C_1, \\ A_2x + B_2y = -C_2. \end{cases}$$

Из параграфа 4.3 нам известно, что в данном случае все зависит от величины соответствующих определителей. Система имеет единственное решение тогда и только тогда, когда

$$\Delta = \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} = A_1B_2 - A_2B_1 \neq 0$$

или для ненулевых коэффициентов

$$A_1/A_2 \neq B_1/B_2.$$

В этом случае прямые п е р е с е к а ю т с я.

Если $\Delta = 0$, возможны два случая: а) когда

$$\Delta_x = \begin{vmatrix} -C_1 & B_1 \\ -C_2 & B_2 \end{vmatrix} = \Delta_y = \begin{vmatrix} A_1 & -C_1 \\ A_2 & -C_2 \end{vmatrix} = 0,$$

система имеет бесконечно много решений (все коэффициенты пропорциональны). В этом случае прямые с о в п а д а ю т. При ненулевых коэффициентах это условие можно записать в виде

$$A_1/A_2 = B_1/B_2 = C_1/C_2;$$

б) если же хотя бы один из определителей Δ_x и Δ_y не равен нулю, система несовместна и прямые п а р а л л е л ь н ы.

Итак, мы пришли к одинаковым выводам, как при рассмотрении коллинеарности–неколлинеарности векторов, так и при исследовании системы линейных уравнений с помощью определителей.

Например, прямые

$$\begin{cases} 2x - 3y + 1 = 0, \\ x + 2y - 2 = 0 \end{cases}$$

пересекаются ($\Delta = 7 \neq 0$);

прямые

$$\begin{cases} x + 2y - 1 = 0, \\ 2x + 4y + 3 = 0 \end{cases}$$

параллельны (пропорциональны только коэффициенты при x и y);

прямые

$$\begin{cases} x + 2y - 1 = 0, \\ 2x + 4y - 2 = 0 \end{cases}$$

совпадают (все коэффициенты пропорциональны: $1/2 = 2/4 = -1/(-2)$).

4.4.4. Основные задачи о прямых

Предположим, что координаты заданы в декартовой прямоугольной системе координат.

Задача 1. *Провести прямую, проходящую через две заданные точки.*

Даны точки $A_1(x_1, y_1)$, $A_2(x_2, y_2)$. Требуется составить общее уравнение прямой, проходящей через них. В качестве направляющего вектора естественно взять вектор $\vec{s} = \overrightarrow{A_1A_2} = (x_2 - x_1, y_2 - y_1) = (-B, A)$. Значит, $A = y_2 - y_1$, $B = -(x_2 - x_1)$ и $\vec{n} = (A, B) = (y_2 - y_1, -(x_2 - x_1))$. В качестве начального вектора возьмем радиус-вектор одной из данных точек, например $\vec{r}_0 = (x_1, y_1)$. Если $\vec{r} = (x, y)$ – радиус-вектор произвольной точки на искомой прямой, то нормальное уравнение $\vec{n}\cdot(\vec{r} - \vec{r}_0) = 0$ запишется в виде $(y_2 - y_1)(x - x_1) - (x_2 - x_1)(y - y_1) = 0$. Отсюда уравнение искомой прямой можно представить в виде

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}$$

или с помощью определителя

$$\begin{vmatrix} x - x_1 & y - y_1 \\ x_2 - x_1 & y_2 - y_1 \end{vmatrix} = 0.$$

Пример.

Проведем прямую через точки (1, 1) и (–2, 3). Получим уравнение

$$(x - 1)/(-2 - 1) = (y - 1)/(3 - 1),$$

или

$$(x - 1)/(-3) = (y - 1)/2.$$

Значит, $2(x - 1) = -3(y - 1)$ и, следовательно, общее уравнение прямой $2x + 3y - 5 = 0$ (рис. 4.38). Легко проверить, что координаты данных точек действительно удовлетворяют этому уравнению. Выражая y через x, т.е. переходя к «школьному» уравнению, получим $y = (-{}^2/_3)x + {}^5/_3$: это прямая с отрицательным угловым коэффициентом $k = -{}^2/_3$, пересекающая ось y в точке $(0, {}^5/_3)$.

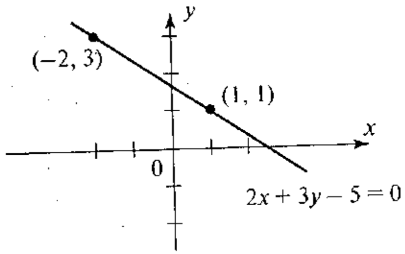


Рис. 4.38

Задача 2. *Провести прямую через заданную точку параллельно заданной прямой.*

Дана прямая $Ax + By + C = 0$ и точка $M(x_0, y_0)$. Поскольку мы строим прямую, параллельную данной, у них совпадают направляющие векторы и векторы нормали. Это означает, что первые два коэффициента искомой и заданной прямой совпадают. Осталось найти третий коэффициент. В качестве начального вектора естественно взять радиус-вектор заданной точки $\vec{r}_0 = (x_0; y_0)$. Нормаль к искомой прямой $\vec{n} = (A, B)$. Значит, третий коэффициент равен $-\vec{n} \cdot \vec{r}_0 = -Ax_0 - By_0$. Итак, общее уравнение прямой $Ax + By - Ax_0 - By_0 = 0$ или

$$A(x - x_0) + B(y - y_0) = 0.$$

Пример.

Проведем прямую параллельно прямой $x + 2y + 3 = 0$ через точку $M(1, 2)$. Общее уравнение имеет вид $(x - 1) + 2(y - 2) = 0$ или $x + 2y - 5 = 0$. Отсюда $y = -0{,}5x + 2{,}5$ (рис. 4.39): прямая имеет отрицательный наклон $k = -0{,}5$ и пересекает ось y в точке $(0, 2{,}5)$.

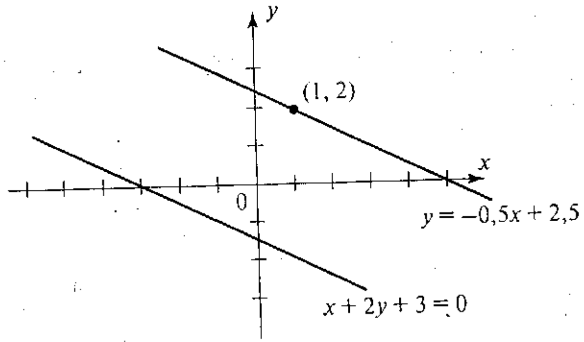


Рис. 4.39

Задача 3. *Провести прямую через заданную точку перпендикулярно заданной прямой.*

Искомая прямая перпендикулярна заданной прямой $Ax + By + C = 0$, следовательно, ее нормаль $\vec{n}$ совпадает с направляющим вектором $(-B, A)$. Значит, коэффициент при x равен $-B$, а коэффициент при y соответственно A. В качестве $\vec{r}_0$ снова возьмем радиус-вектор заданной точки $M(x_0, y_0)$. Тогда третий коэффициент равен $-\vec{n} \cdot \vec{r}_0 = Bx_0 - Ay_0$. Общее уравнение прямой имеет вид $-Bx + Ay + Bx_0 - Ay_0 = 0$ или

$$B(x - x_0) - A(y - y_0) = 0.$$

Пример.

Проведем прямую через точку $(1, 1)$ перпендикулярно прямой $x - y - 1 = 0$. Поскольку $A = 1$, $B = -1$, ее уравнение запишется в виде $-(x - 1) - (y - 1) = 0$, или $x + y - 2 = 0$. Отсюда $y = -x + 2$ (рис. 4.40): прямая параллельна биссектрисе II и IV квадрантов и проходит через точку $(0, 2)$.

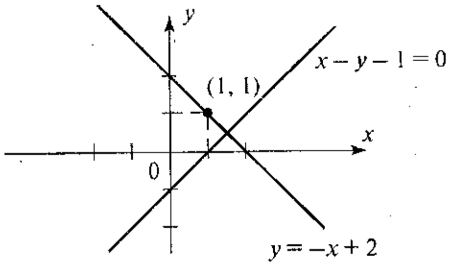


Рис. 4.40

4.5. Плоскость и прямая в пространстве

4.5.1. Поверхности и линии в пространстве

Пусть в пространстве задана система координат. Ограничимся случаем прямоугольной декартовой системы. Поверхность в пространстве задается уравнением с тремя неизвестными

$$F(x, y, z) = 0.$$

Такое уравнение называется *общим* уравнением поверхности, если ему удовлетворяют координаты всякой точки, лежащей на данной поверхности, и не удовлетворяют координаты никаких других точек.

Примеры поверхностей.

1. *Сфера.* Как можно ожидать по аналогии с окружностью, сфера с центром в точке (x_0, y_0, z_0) и радиусом R задается общим уравнением

$$(x-x_0)^2+(y-y_0)^2+(z-z_0)^2-R^2=0.$$

2. *Круговой цилиндр.* Пусть в плоскости xOy расположена окружность $(x-x_0)^2+(y-y_0)^2=R^2$. Если через каждую ее точку провести прямую, параллельную оси z, получится поверхность, которая называется «прямой круговой цилиндр». Рассекая эту поверхность любой плоскостью, параллельной плоскости xOy, мы получим окружность. Если некоторая точка (x_1, y_1, z_1) лежит на цилиндре, то любая точка вида (x_1, y_1, z) также будет лежать на этой поверхности. Таким образом, можно догадаться, что уравнение цилиндра не должно содержать переменной z и выглядит так же, как уравнение окружности на плоскости:

$$(x-x_0)^2+(y-y_0)^2-R^2=0.$$

Вообще, уравнение вида $F(x, y)=0$, не содержащее z, задает так называемую цилиндрическую поверхность. Она получается, если через каждую точку линии $F(x, y)=0$ на плоскости xOy провести прямую, параллельную оси z. Исходная линия называется в данном случае *направляющей линией* поверхности. Вся такая цилиндрическая поверхность состоит из прямых, параллельных оси z (так называемых *образующих*). Точно так же уравнения $F(x, z)=0$ и $F(y, z)=0$ задают цилиндрические поверхности с образующими, параллельными осям y и x соответственно.

Чтобы задать поверхность в параметрическом виде, уже недостаточно одного параметра. На поверхности можно построить сколько угодно железных дорог (кривых). (Помните, в подпараграфе 4.4.1 мы представили, что едем на поезде.) Поездка по каждой из них описывается с помощью своего расписания, т.е. своего параметра. Трудно предположить, что найдется один универсальный параметр, определяющий точки на всех этих дорогах. Однако оказывается, что для задания всех точек поверхности достаточно всего двух параметров. Уравнения

$$\begin{cases} x=f_1(u, v),\\ y=f_2(u, v),\\ z=f_3(u, v) \end{cases}$$

называются уравнениями поверхности *в параметрической форме.* Величины u и v – числовые параметры, изменяющиеся в заданных

пределах. Исключив параметры из параметрических уравнений, мы получим общее уравнение поверхности. Например, параметрические уравнения прямого кругового цилиндра, направляющая которого есть окружность радиусом R и с центром в начале координат на плоскости xOy, выглядят следующим образом:

$$\begin{cases} x=R\cos u,\\ y=R\sin u, \quad u\in[0, 2\pi], \quad -\infty\leqslant v\leqslant\infty.\\ z=v, \end{cases}$$

Кривая (линия) в пространстве – это пересечение двух поверхностей. Поэтому кривая задается не одним уравнением, а системой уравнений

$$\begin{cases} F_1(x, y, z)=0,\\ F_2(x, y, z)=0. \end{cases}$$

Например, любая окружность в пространстве может быть задана как пересечение двух сфер. Заранее скажем, что прямая в пространстве определяется как пересечение двух плоскостей.

4.5.2. Уравнения плоскости

Убедимся, что теория, описывающая плоскость в пространстве, во многом сходна с теорией прямых на плоскости. Чтобы вывести векторное уравнение плоскости, зададим на ней начальную точку с радиусом-вектором $\vec{r}_0=(x_0, y_0, z_0)$ и пару неколлинеарных векторов $\vec{p}$ и $\vec{q}$ (рис. 4.41). Пусть $\vec{r}=(x, y, z)$ – радиус-вектор произвольной точки на плоскости. Тогда очевидно, что вектор $\vec{r}-\vec{r}_0$ лежит на плоскости. Значит, его можно разложить по базису $\{\vec{p}, \vec{q}\}$, т.е. найдутся два числа u и v, такие, что $\vec{r}-\vec{r}_0=u\vec{p}+v\vec{q}$. Отсюда

$$\vec{r}=\vec{r}_0+u\vec{p}+v\vec{q}.$$

Перебирая все возможные действительные значения параметров u и v, мы получим радиусы-векторы всех точек плоскости. Таким образом, построено в е к т о р н о е уравнение плоскости. Если $\vec{p}=(p_1, p_2, p_3)$, $\vec{q}=(q_1, q_2, q_3)$, из векторного мы легко выведем п а р а м е т р и ч е с к и е уравнения плоскости

$$\begin{cases} x=x_0+p_1u+q_1v,\\ y=y_0+p_2u+q_2v, \quad u, v\in \mathbf{R}.\\ z=z_0+p_3u+q_3v, \end{cases}$$

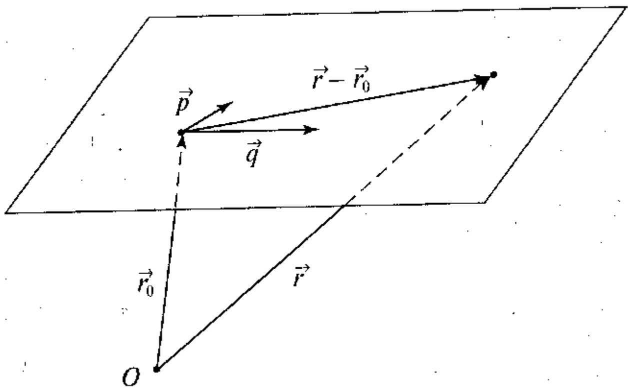


Рис. 4.41

Поскольку, как было отмечено, векторы $\vec{r}-\vec{r}_0$, $\vec{p}$, $\vec{q}$ лежат на одной плоскости, их смешанное произведение равно нулю:

$$(\vec{r}-\vec{r}_0)\vec{p}\vec{q}=0.$$

Вспомним, что смешанное произведение легко записывается через определитель 3-го порядка. Имея в виду, что $\vec{r}-\vec{r}_0 = (x-x_0,\ y-y_0,\ z-z_0)$, получаем еще одно уравнение плоскости:

$$(x-x_0)\begin{vmatrix} p_2 & p_3 \\ q_2 & q_3 \end{vmatrix}-(y-y_0)\begin{vmatrix} p_1 & p_3 \\ q_1 & q_3 \end{vmatrix}+(z-z_0)\begin{vmatrix} p_1 & p_2 \\ q_1 & q_2 \end{vmatrix}=0.$$

От этого уравнения перейдем к общему. Введем обозначения:

$$A=\begin{vmatrix} p_2 & p_3 \\ q_2 & q_3 \end{vmatrix},\quad B=-\begin{vmatrix} p_1 & p_3 \\ q_1 & q_3 \end{vmatrix},\quad C=\begin{vmatrix} p_1 & p_2 \\ q_1 & q_2 \end{vmatrix}.$$

Тогда

$$A(x-x_0)+B(y-y_0)+C(z-z_0)=0$$

или, если обозначить $D=-(Ax_0+By_0+Cz_0)$,

$$Ax+By+Cz+D=0.$$

Это и есть о б щ е е уравнение плоскости. Мы проверили, что точки плоскости действительно удовлетворяют такому уравнению. Верно и обратное утверждение: всякое уравнение такого вида задает некоторую плоскость. Плоскость в пространстве, как и прямая на плоскости, задается уравнением первой степени относительно координат. Поэтому плоскость – это поверхность *первого порядка*.

Пусть вектор $\vec{n}$ перпендикулярен плоскости (является нормалью). Тогда скалярное произведение нормали и вектора $\vec{r}-\vec{r}_0$ равно

нулю. Это равенство есть не что иное, как н о р м а л ь н о е уравнение плоскости

$$\vec{n}\cdot(\vec{r}-\vec{r}_0)=0.$$

Сравнивая нормальное и общее уравнения, убеждаемся, что $\vec{n}=(A,\ B,\ C)$, $D=-\vec{n}\cdot\vec{r}_0$. Как и у прямой, коэффициенты при координатах в общем уравнении составляют вектор нормали, а свободный член общего уравнения есть $-\vec{n}\cdot\vec{r}_0$.

При $C\neq 0$ из общего уравнения плоскости получаем более привычное уравнение, похожее на «школьное» уравнение прямой,

$$z=k_1x+k_2y+b,$$

где $k_1=-A/C$, $k_2=-B/C$, $b=-D/C$. В этом виде плоскость представляет собой график линейной функции двух переменных. Величина c – длина отрезка, отсекаемого плоскостью на оси z.

Если все коэффициенты общего уравнения не равны нулю, можно получить уравнение плоскости в о т р е з к а х

$$\frac{x}{a}+\frac{y}{b}+\frac{z}{c}=1$$

при $a=-D/A$, $b=-D/B$, $c=-D/C$. Плоскость отсекает отрезки a, b, c на осях координат.

Пример.

Рассмотрим плоскость с общим уравнением $3x+2y-2z+1=0$. Нормаль к ней $\vec{n}=(3,\ 2,\ -2)$. На этой плоскости лежит, к примеру, точка $(x_0,\ y_0,\ z_0)=(1,\ 2,\ 4)$. Тогда $-\vec{n}\cdot\vec{r}_0=-(3\cdot 1+2\cdot 2-2\cdot 4)=1=C$. Тот же результат получится, если взять любую другую точку плоскости в качестве начальной. Выражая z, получаем «школьное» уравнение $z=1{,}5x+y+0{,}5$. Уравнение в отрезках имеет вид $\frac{x}{-1/3}+\frac{y}{-1/2}+\frac{z}{1/2}=1$. Данная плоскость отсекает отрезки, равные $-1/3$, $-1/2$ и $1/2$ соответственно на осях x, y и z.

4.5.3. Расположение плоскости в пространстве. Взаимное расположение плоскостей

Проанализируем частные случаи расположения плоскости относительно декартовой системы координат.

1. $\vec{n}\parallel\vec{i}$: нормаль к плоскости параллельна оси x. Поскольку $\vec{i}=(1,\ 0,\ 0)$ для нормали $\vec{n}=(A,\ B,\ C)$ имеем $B=C=0$. Уравнение

принимает вид $Ax + D = 0$, или $x = -D/A$. В этом случае плоскость параллельна координатной плоскости yOz.

2. $\vec{n} \parallel \vec{j}$: $A = C = 0$ и плоскость параллельна xOz. Уравнение имеет вид $y = -D/B$.

3. $\vec{n} \parallel \vec{k}$: $A = B = 0$ и плоскость $z = -D/C$ параллельна плоскости xOy.

4. $\vec{n} \perp \vec{i}$: вектор нормали лежит в плоскости yOz и, стало быть, плоскость параллельна оси x. В этом случае $A = 0$, так как $\vec{n} \cdot \vec{i} = A \cdot 1 + B \cdot 0 + C \cdot 0 = A$.

5. $\vec{n} \perp \vec{j}$: $B = 0$ и плоскость параллельна оси y.

6. $\vec{n} \perp \vec{k}$: $C = 0$, плоскость параллельна оси z.

7. $\vec{n} \perp \vec{r}_0$: это возможно лишь в случае, когда плоскость проходит через начало координат. При этом $D = -\vec{n} \cdot \vec{r}_0 = 0$ и плоскость задается уравнением $Ax + By + Cz = 0$, которому, разумеется, удовлетворяет точка $(0, 0, 0)$.

Взаимное расположение двух плоскостей удобно анализировать с помощью их нормали. Пусть плоскости заданы общими уравнениями

$$A_1x + B_1y + C_1z + D_1 = 0,$$
$$A_2x + B_2y + C_2z + D_2 = 0$$

и, стало быть, имеют нормали $\vec{n}_1 = (A_1, B_1, C_1)$, $\vec{n}_2 = (A_2, B_2, C_2)$.

Рассмотрим различные случаи.

1. $\vec{n}_1$ и $\vec{n}_2$ неколлинеарны. В этом случае, разумеется, плоскости пересекаются по прямой. Условие неколлинеарности векторов нормалей – непропорциональность координат. Значит, хотя бы один из определителей

$$\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}, \quad \begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix}, \quad \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}$$

отличен от нуля. В частности, плоскости перпендикулярны, если перпендикулярны их нормали, т.е. их скалярное произведение равно нулю:

$$A_1A_2 + B_1B_2 + C_1C_2 = 0.$$

2. $\vec{n}_1 \parallel \vec{n}_2$: нормали коллинеарны. В этом случае плоскости, как и прямые, параллельны либо совпадают. Условие коллинеарности

$$\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} = \begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix} = \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix} = 0$$

или для ненулевых коэффициентов

$$A_1/A_2 = B_1/B_2 = C_1/C_2.$$

Если, кроме того, выполняется условие

$$A_1/A_2 = B_1/B_2 = C_1/C_2 = D_1/D_2,$$

то плоскости совпадают, а если не выполняется, то параллельны.

4.5.4. Основные задачи о плоскостях

Задача 1. *Провести плоскость, проходящую через три заданные точки.*

Заданы точки $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, $M_3(x_3, y_3, z_3)$. В качестве начального вектора можно выбрать радиус-вектор любой из них, например $\vec{r}_0 = (x_1, y_1, z_1)$. За базис на плоскости возьмем векторы $\vec{p} = \overline{M_1M_2} = (x_2 - x_1, \ y_2 - y_1, \ z_2 - z_1)$, $\vec{q} = \overline{M_1M_3} = (x_3 - x_1, y_3 - y_1, z_3 - z_1)$. Тогда, если радиус-вектор произвольной точки на плоскости обозначить через $\vec{r} = (x, y, z)$, уравнение примет вид

$$(x - x_1)\begin{vmatrix} y_2 - y_1 & z_2 - z_1 \\ y_3 - y_1 & z_3 - z_1 \end{vmatrix} - (y - y_1)\begin{vmatrix} x_2 - x_1 & z_2 - z_1 \\ x_3 - x_1 & z_3 - z_1 \end{vmatrix} +$$
$$+ (z - z_1)\begin{vmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{vmatrix} = 0.$$

Задача 2. *Провести плоскость, проходящую через заданную точку параллельно заданной плоскости.*

Заданы точка $M(x_0, y_0, z_0)$ и прямая $Ax + By + Cz + D = 0$. Ясно, что нормали данной и искомой плоскостей коллинеарны. Поэтому первые три коэффициента у них совпадают. Четвертый коэффициент равен $-\vec{n} \cdot \vec{r}_0 = -(A, B, C) \cdot (x_0, y_0, z_0) = -(Ax_0 + By_0 + Cz_0)$. Значит,

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$$

и есть уравнение искомой плоскости.

4.5.5. Прямая в пространстве

Прямая в пространстве есть не что иное, как пересечение двух плоскостей. Таким образом, прямая задается системой уравнений

$$\begin{cases} A_1x + B_1y + C_1z + D_1 = 0, \\ A_2x + B_2y + C_2z + D_2 = 0, \end{cases}$$

если данные плоскости пересекаются. Вспомним операции над множествами. Если прямая l лежит на пересечении плоскостей π_1 и π_2, то множество точек прямой есть пересечение множеств точек каждой из плоскостей: $l = \pi_1 \cap \pi_2$. Через прямую в пространстве проходит бесконечно много различных плоскостей. Поэтому ясно, что прямую можно определить системой уравнений бесконечно многими способами.

Если заданы точка на прямой с радиусом-вектором $\vec{r}_0 = (x_0, y_0, z_0)$ и направляющий вектор $\vec{s} = (s_1, s_2, s_3)$, то из рассуждений в параграфе 4.4 ясно, что данная прямая описывается векторным уравнением точно так же, как и прямая на плоскости:

$$\vec{r} = \vec{r}_0 + t\vec{s}, \quad -\infty \leqslant t \leqslant \infty$$

или в параметрической форме

$$\begin{cases} x = x_0 + ts_1, \\ y = y_0 + ts_2, \\ z = z_0 + ts_3. \end{cases}$$

Попытаемся найти направляющий вектор $\vec{s}$, если прямая задана системой уравнений. Пусть $\vec{n}_1$ и $\vec{n}_2$ – нормали к плоскостям, пересечение которых составляет данная прямая (рис. 4.42). Ясно, что прямая перпендикулярна каждой из нормалей. Поэтому в качестве ее направляющего вектора нужно взять вектор, перпендикулярный обеим нормалям. Мы точно знаем по крайней мере один такой вектор: это векторное произведение нормалей. Значит, векторное уравнение прямой в пространстве можно записать в виде

$$\vec{r} = \vec{r}_0 + t(\vec{n}_1 \times \vec{n}_2).$$

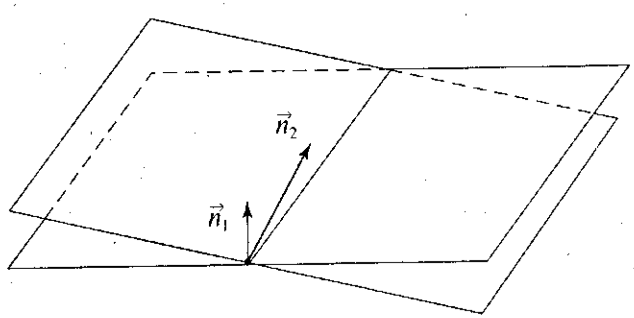


Рис. 4.42

Задачи

1. Вычислить длину вектора $\vec{a} = (6, 3, -2)$.

2. Даны две координаты вектора: $x = 4$, $y = -12$. Найти третью координату z, если $|\vec{a}| = 13$.

3. Даны точки $A(3, -1, 2)$ и $B(-1, 2, 1)$. Найти координаты векторов $\overrightarrow{AB}$ и $\overrightarrow{BA}$.

4. По данным векторам $\vec{a}$ и $\vec{b}$ построить векторы $\vec{a} + \vec{b}$, $\vec{a} - \vec{b}$, $\vec{b} - \vec{a}$, $-\vec{a} - \vec{b}$.

5. В треугольнике ABC вектор $\overrightarrow{AB} = \vec{m}$ и $\overrightarrow{AC} = \vec{n}$. Построить векторы: а) $(\vec{m} + \vec{n})/2$; б) $(\vec{m} - \vec{n})/2$; в) $(\vec{n} - \vec{m})/2$; г) $-(\vec{m} + \vec{n})/2$.

6. Проверить, что четыре точки $A(3, -1, 2)$, $B(1, 2, -1)$, $C(-1, 1, -3)$ и $D(3, -5, 3)$ служат вершинами трапеции.

7. Определить длину суммы и разности векторов $\vec{a} = (3, -5, 8)$ и $\vec{b} = (-1, 1, -4)$.

8. Векторы $\vec{a}$ и $\vec{b}$ взаимно перпендикулярны, вектор $\vec{c}$ образует с ними углы, равные $\pi/3$. Зная, что $|\vec{a}| = 3$, $|\vec{b}| = 5$, $|\vec{c}| = 8$, вычислить $(3\vec{a} - 2\vec{b}) \cdot (\vec{b} + 3\vec{c})$.

9. Даны векторы $\vec{a} = (4, -2, -4)$, $\vec{b} = (6, -3, 2)$. Вычислить а) $\vec{a} \cdot \vec{b}$; б) $(2\vec{a} - 3\vec{b}) \cdot (\vec{a} + 2\vec{b})$; в) $(\vec{a} + \vec{b})^2$.

10. Дано: $|\vec{a}| = 10$, $|\vec{b}| = 2$ и скалярное произведение $\vec{a} \cdot \vec{b} = 12$. Вычислить длину векторного произведения $|\vec{a} \times \vec{b}|$.

11. Векторы $\vec{a}$ и $\vec{b}$ взаимно перпендикулярны. Зная, что $|\vec{a}| = 3$, $|\vec{b}| = 4$, вычислить $|(\vec{a} + \vec{b}) \times (\vec{a} - \vec{b})|$.

12. Даны векторы $\vec{a} = (3, -1, 2)$, $\vec{b} = (1, 2, -1)$. Найти координаты векторных произведений: а) $\vec{a} \times \vec{b}$; б) $(2\vec{a} + \vec{b}) \times \vec{b}$; в) $(2\vec{a} - \vec{b}) \times (2\vec{a} + \vec{b})$.

13. Векторы $\vec{a}$, $\vec{b}$, $\vec{c}$, образующие правую тройку, взаимно перпендикулярны. Зная, что $|\vec{a}| = 4$, $|\vec{b}| = 2$, $|\vec{c}| = 3$, вычислить смешанное произведение $\vec{a}\,\vec{b}\,\vec{c}$.

14. Даны три вектора: $\vec{a} = (1, -1, 3)$, $\vec{b} = (-2, 2, 1)$ и $\vec{c} = (3, -2, 5)$. Вычислить $\vec{a}\,\vec{b}\,\vec{c}$.

15. Установить, компланарны ли векторы $\vec{a}$, $\vec{b}$, $\vec{c}$:

а) $\vec{a} = (2, 3, -1)$, $\vec{b} = (1, -1, 3)$, $\vec{c} = (1, 9, -11)$;

б) $\vec{a} = (3, -2, 1)$, $\vec{b} = (2, 1, 2)$, $\vec{c} = (3, -1, -2)$.

16. Определить, какие из точек $M_1(3, 1)$, $M_2(2, 3)$, $M_3(6, 3)$, $M_4(-3, -3)$, $M_5(3, -1)$, $M_6(-2, 1)$ лежат на прямой $2x - 3y - 3 = 0$ и какие не лежат на ней.

17. Определить точки пересечения прямой $2x - 3y - 12 = 0$ с координатными осями и построить прямую на чертеже.

18. Стороны AB, BC и AC треугольника ABC заданы соответственно уравнениями прямых $4x + 3y - 5 = 0$, $x - 3y + 10 = 0$, $x - 2 = 0$. Найти координаты его вершин.

19. Дана прямая $5x + 3y - 3 = 0$. Определить угловой коэффициент k прямой:

а) параллельной данной прямой;

б) перпендикулярной данной прямой.

20. Вычислить угловой коэффициент прямой, проходящей через две данные точки $M_1(2, -5)$, $M_2(3, 2)$.

21. Даны последовательные вершины выпуклого четырехугольника: $A(-3, 1)$, $B(3, 9)$, $C(7, 6)$ и $D(-2, -6)$. Определить точку пересечения его диагоналей.

22. Установить, какие из следующих пар прямых перпендикулярны:

а) $3x - y + 5 = 0$, $x + 3y - 1 = 0$;

б) $3x - 4y + 1 = 0$, $4x - 3y + 7 = 0$;

в) $6x - 15y + 7 = 0$, $10x + 4y - 3 = 0$.

23. Составить уравнение плоскости, проходящей через точку $M(2, 1, -1)$ и имеющей нормальный вектор $\vec{n} = (1, -2, 3)$.

24. Даны две точки $M_1(3, -1, 2)$ и $M_2(4, -2, -1)$. Составить уравнение плоскости, проходящей через точку M_1 перпендикулярно вектору $\overrightarrow{M_1M_2}$.

25. Установить, какие из следующих пар уравнений определяют параллельные плоскости:

а) $2x - 3y + 5z - 7 = 0$, $2x - 3y + 5z + 3 = 0$;

б) $4x + 2y - 4z + 5 = 0$, $2x + y + 2z - 1 = 0$.

26. Установить, какие из следующих пар уравнений определяют перпендикулярные плоскости:

а) $3x - y - 2z - 5 = 0$, $x + 9y - 3z + 3 = 0$;

б) $2x - 5y + z = 0$, $x + 2z - 3 = 0$.

27. Найти отрезки, отсекаемые плоскостью $3x - 4y - 24z - 12 = 0$ на осях координат.

28. Найти точки пересечения прямой

$$\begin{cases} 2x + y - z - 3 = 0, \\ x + y + z - 1 = 0 \end{cases}$$

с координатными плоскостями.

Глава 5
ЭЛЕМЕНТЫ ЛИНЕЙНОЙ АЛГЕБРЫ*

5.1. Пространство R^n

5.1.1. n-мерные векторы и операции над ними

С помощью системы координат нам удается отождествить вектор (или точку) на плоскости с упорядоченной парой действительных чисел. Если система координат задана, можно не делать различия между вектором (направленным отрезком) и парой чисел (его координатами). Точно так же вектор (точка) в пространстве задается упорядоченной тройкой чисел. Можно пойти дальше и выписать четверку действительных чисел (x_1, x_2, x_3, x_4). Современная физика имеет дело с пространством, включающим четвертую координату — время. Такая четверка чисел представляет собой координаты точки четырехмерного пространства. Отвлекаясь от реальных геометрических представлений, мы можем рассмотреть упорядоченный набор из n действительных чисел $\mathbf{x} = (x_1, x_2, ..., x_n)$ и по аналогии назвать его *n-мерным вектором*. Числа $x_1, x_2, ..., x_n$ называются *координатами* вектора $\mathbf{x}$.

К определению множества n-мерных векторов можно подойти и с другой стороны. Вспомним, что декартово произведение множества действительных чисел $\boldsymbol{R}$ само на себя состоит из всевозможных упорядоченных числовых пар. Это множество обозначают $\boldsymbol{R}^2$, и его можно отождествить с плоскостью. Множество $\boldsymbol{R} \times \boldsymbol{R} \times \boldsymbol{R} = \boldsymbol{R}^3$ состоит из упорядоченных троек и представляет собой трехмерное пространство. Точно так же, если осуществить декартово произведение $\boldsymbol{R}$ на себя n раз, получим совокупность всех n-мерных векторов. Это множество принято называть пространством $\boldsymbol{R}^n$.

Векторы $\mathbf{x} = (x_1, x_2, ..., x_n)$, $\mathbf{y} = (y_1, y_2, ..., y_n)$ *равны*, если они совпадают покоординатно:

$$\mathbf{x} = \mathbf{y} \iff x_1 = y_1, x_2 = y_2, ..., x_n = y_n.$$

Чтобы работать с математическим объектом, необходимо, как нам уже известно, определить операции над ним. Операции сложения и умножения n-мерных векторов (в дальнейшем – просто векторов) вводятся по аналогии с обычными векторами, т.е. покоординатно. Свойства обычных векторов, которые мы доказывали, становятся, таким образом, определениями для n-мерных векторов. Итак, пусть $\mathbf{x} = (x_1, x_2, ..., x_n)$, $\mathbf{y} = (y_1, y_2, ..., y_n) \in \boldsymbol{R}^n$, λ – число. Тогда по определению

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n),$$
$$\lambda\mathbf{x} = (\lambda x_1, \lambda x_2, ..., \lambda x_n).$$

Операции над n-мерными векторами обладают теми же алгебраическими свойствами, что и операции над векторами на плоскости и в пространстве:

1. *Сложение векторов коммутативно*: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
2. *Сложение векторов ассоциативно*: $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$.
3. $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$, *где* $\mathbf{0} = (0, 0, ..., 0)$.
4. $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$, *где* $-\mathbf{x} = (-x_1, -x_2, ..., -x_n)$.
5. *Умножение на число ассоциативно*: $\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}$.
6. *Умножение на число дистрибутивно относительно сложения чисел*: $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$.
7. *Умножение на число дистрибутивно относительно сложения векторов*: $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$.
8. $0 \cdot \mathbf{x} = \mathbf{0}$; $\lambda\mathbf{0} = \mathbf{0}$.
9. $1 \cdot \mathbf{x} = \mathbf{x}$; $(-\lambda)\mathbf{x} = -\lambda\mathbf{x}$.

Все эти свойства следуют непосредственно из определений операций. Вектор $\mathbf{0}$ называется *нулевым*, а вектор $-\mathbf{x}$ – *противоположным* вектору $\mathbf{x}$. Существование противоположного вектора $-\mathbf{x}$ позволяет ввести вычитание векторов: $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-\mathbf{y}) = (x_1 - y_1, x_2 - y_2, ..., x_n - y_n)$.

5.1.2. Линейная независимость. Базис

Понятие базиса на плоскости и в пространстве обобщим на пространство $\boldsymbol{R}^n$.

Система векторов $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ называется ***линейно зависимой***, если найдется набор чисел $\lambda_1, \lambda_2, ..., \lambda_k$, не все из которых равны нулю, такой, что выполняется равенство

$$\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + ... + \lambda_k\mathbf{x}_k = 0.$$

В противном случае система $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ называется ***линейно независимой***. В данном контексте «в противном случае» означает, что из

равенства $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + ... + \lambda_k\mathbf{x}_k = 0$ следует, что $\lambda_1 = \lambda_2 = ... = \lambda_k = 0$. Если система векторов линейно зависима или линейно независима, будем иногда говорить, что векторы системы линейно зависимы или независимы. Из определения, в частности, следует, что один вектор всегда линейно независим, если только он ненулевой. Кроме того, если система векторов $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ линейно зависима, то, добавляя к ней какие угодно новые векторы, мы снова получим линейно зависимую систему.

Спустимся из «заоблачных высот» пространства $\boldsymbol{R}^n$ в привычные $\boldsymbol{R}^2$ и $\boldsymbol{R}^3$, т.е. на плоскость и в трехмерное пространство.

∇**1.** *Два ненулевых вектора на плоскости линейно независимы тогда и только тогда, когда они неколлинеарны.*

В самом деле, пусть векторы $\vec{a}$ и $\vec{b}$ линейно независимы. Предположим, что они коллинеарны. Тогда $\vec{b} = \mu\vec{a}$, $\mu \neq 0$ и при $\lambda_1 = -\mu$, $\lambda_2 = -1$ имеем $\lambda_1\vec{a} + \lambda_2\vec{b} = \vec{0}$, что противоречит линейной независимости. С другой стороны, пусть векторы $\vec{a}$ и $\vec{b}$ неколлинеарны. Это означает (см. наблюдение ∇1 в параграфе 4.3), что

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = x_1y_2 - x_2y_1 \neq 0.$$

Предположим, что $\lambda_1\vec{a} + \lambda_2\vec{b} = \vec{0}$. Это означает, что

$$\begin{cases} \lambda_1x_1 + \lambda_2x_2 = 0; \\ \lambda_1y_1 + \lambda_2y_2 = 0. \end{cases}$$

Это система двух линейных уравнений относительно неизвестных λ_1 и λ_2. Определитель этой системы равен $x_1y_2 - x_2y_1$ и по предположению не равен нулю. Значит, данная система имеет единственное нулевое решение, т.е. $\lambda_1 = \lambda_2 = 0$ и векторы $\vec{a}$ и $\vec{b}$ линейно независимы.

∇**2.** *Любые три вектора на плоскости линейно зависимы.*

Действительно, если среди них есть линейно зависимая пара векторов, то и тройка также линейно зависима. Если же, скажем, векторы $\vec{a}$ и $\vec{b}$ линейно независимы, то они образуют базис на плоскости и третий вектор $\vec{c}$ представим в виде $\vec{c} = \mu_1\vec{a} + \mu_2\vec{b}$. Возьмем $\lambda_1 = -\mu_1$, $\lambda_2 = -\mu_2$, $\lambda_3 = 1$. Тогда $\lambda_1\vec{a} + \lambda_2\vec{b} + \lambda_3\vec{c} = -\mu_1\vec{a} - \mu_2\vec{b} + \vec{c} = \vec{0}$ и векторы линейно зависимы.

Аналогичные свойства справедливы и в трехмерном пространстве.

∇3. *Три ненулевых вектора в трехмерном пространстве линейно независимы тогда и только тогда, когда они некомпланарны.*

∇4. *Любые четыре вектора в трехмерном пространстве линейно зависимы.*

Вернемся к пространству $\boldsymbol{R}^n$. По аналогии справедливо следующее свойство.

∇5. *Любые $n + 1$ векторов пространства $\boldsymbol{R}^n$ линейно зависимы.*

Пример линейно независимой системы из n векторов дают векторы $\mathbf{e}_1 = (1, 0, 0, ..., 0, 0)$, $\mathbf{e}_2 = (0, 1, 0, ..., 0, 0)$, ..., $\mathbf{e}_{n-1} = (0, 0, 0, ..., 1, 0)$, $\mathbf{e}_n = (0, 0, 0, ..., 0, 1)$. В каждом векторе $\mathbf{e}_i$ единица стоит на i-м месте, а остальные координаты – нули. В самом деле, по определению операций легко убедиться, что $\lambda_1\mathbf{e}_1 + \lambda_2\mathbf{e}_2 + ... + \lambda_n\mathbf{e}_n = (\lambda_1, \lambda_2, ..., \lambda_n)$. Этот вектор может быть равен нулевому только тогда, когда все числа $\lambda_1, \lambda_2, ..., \lambda_n$ равны нулю. Это и означает, что данная система векторов линейно независима.

Пусть $\mathbf{x} = (x_1, x_2, ..., x_n) \in \boldsymbol{R}^n$ – произвольный вектор. Тогда, очевидно, справедливо равенство

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_n\mathbf{e}_n.$$

Говорят, что вектор $\mathbf{x}$ *разложим* по векторам $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$. Например, $(3, 2, 5, 4, 2) = (3, 0, 0, 0, 0) + (0, 2, 0, 0, 0) + (0, 0, 5, 0, 0) + (0, 0, 0, 4, 0) + (0, 0, 0, 0, 2) = 3 \cdot (1, 0, 0, 0, 0) + 2 \cdot (0, 1, 0, 0, 0) + 5 \cdot (0, 0, 1, 0, 0) + 4 \cdot (0, 0, 0, 1, 0) + 2 \cdot (0, 0, 0, 0, 1)$. Линейно независимая система векторов, по которой можно разложить произвольный вектор пространства $\boldsymbol{R}^n$, называется ***базисом*** этого пространства. Система $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ называется *основным базисом* $\boldsymbol{R}^n$.

Заметим, что любая пара неколлинеарных (линейно независимых) векторов образует базис на плоскости, а любая тройка некомпланарных (линейно независимых) векторов есть базис трехмерного пространства. Аналогичное свойство выполняется и в $\boldsymbol{R}^n$.

∇6. *Любые n линейно независимых векторов $\boldsymbol{R}^n$ образуют базис, причем любые $m < n$ линейно независимых векторов базиса $\boldsymbol{R}^n$ не образуют.*

Итак, минимальное количество векторов, которые могут составить базис $\boldsymbol{R}^n$, равно n. Число n называют *размерностью* пространства $\boldsymbol{R}^n$.

5.1.3. Скалярное произведение в $\boldsymbol{R}^n$. Норма вектора

Пусть $\mathbf{x} = (x_1, x_2, ..., x_n)$, $\mathbf{y} = (y_1, y_2, ..., y_n) \in \boldsymbol{R}^n$. ***Скалярным произведением*** двух векторов называется число

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + ... + x_ny_n.$$

Снова свойство векторов на плоскости и в трехмерном пространстве стало определением в $\boldsymbol{R}^n$. Остаются неизменными и свойства скалярного произведения:

1. *Скалярное произведение коммутативно*: $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.
2. $(\lambda\mathbf{x}) \cdot \mathbf{y} = \lambda(\mathbf{x} \cdot \mathbf{y})$.
3. *Скалярное произведение дистрибутивно относительно сложения векторов*: $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$.

Вспомним, что длину вектора $\vec{a} = (x, y, z)$ в трехмерном пространстве легко выразить через его (вектора) координаты:

$$|\vec{a}| = \sqrt{x^2 + y^2 + z^2}.$$

Видно, что $|\vec{a}| = \sqrt{\vec{a} \cdot \vec{a}}$.

Аналогично определяется ***норма*** вектора $\mathbf{x} = (x_1, x_2, ..., x_n)$ в $\boldsymbol{R}^n$, только обозначают ее так:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}.$$

Из определения выводятся следующие свойства нормы вектора:

1. $\|\mathbf{x}\| \geqslant 0$.
2. $\|\lambda\mathbf{x}\| = |\lambda| \cdot \|\mathbf{x}\|$.
3. $\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$.

Последнее свойство называется «неравенством треугольника». В случае плоскости оно отражает тот факт, что сторона треугольника всегда меньше суммы длин остальных сторон: в ΔABC вектор $\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC}$ и $|\overrightarrow{AC}| \leqslant |\overrightarrow{AB}| + |\overrightarrow{BC}|$. Равенство возможно лишь в вырожденном случае, когда вершина B лежит на стороне AC.

Итак, основные свойства геометрических векторов обобщены на случай n-мерного пространства $\boldsymbol{R}^n$.

5.2. Матрицы

5.2.1. Понятие матрицы. Основные операции над матрицами

При рассмотрении определителей 2-го порядка (см. параграф 4.3) мы уже вводили квадратные матрицы 2×2 как таблицы чисел, содержащие соответствующее число строк и столбцов. Точно так же можно построить таблицу с произвольным числом строк и столбцов (например, m строк и n столбцов). Такая таблица называется *прямоугольной матрицей* размерности $m\times n$. Числа, из которых состоит таблица, называют *элементами* матрицы и обычно нумеруют двумя индексами, обозначающими соответственно номер строки и номер столбца, в которых расположен данный элемент:

$$A=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Часто употребляют сокращенную запись: $A=(a_{ij})$ или $A=(a_{ij})^n_m$.

Две матрицы называются *равными*, если они имеют одинаковую размерность (т.е. одинаковое число строк и столбцов) и если числа, стоящие на соответствующих местах этих матриц, равны.

Если число строк матрицы равно числу ее столбцов, матрицу называют *квадратной*. В квадратной матрице число строк (равное числу столбцов) называют *порядком* матрицы. В частности, квадратная матрица порядка 1 — это просто действительное число.

Теперь вектор пространства $\boldsymbol{R}^n$ можно представить как матрицу, состоящую из одной строки (если координаты расположить горизонтально) или одного столбца (при вертикальном расположении). Соответственно говорят о векторе-строке или векторе-столбце:

$$\mathbf{x}=(x_1,\ x_2,\dots,\ x_n),\qquad \mathbf{x}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Вектор-строка есть матрица $1\times n$, вектор-столбец имеет размерность $n\times 1$.

Основные операции над матрицами – сложение матриц и умножение матрицы на число. Операции должны быть определены таким образом, чтобы не возникло противоречия с введенными ранее операциями над векторами из $\boldsymbol{R}^n$.

Итак, ***суммой матриц*** A и B одинаковой размерности называется матрица C такой же размерности, элементы которой равны суммам соответствующих элементов матриц A и B. Под соответствующими понимаются элементы, стоящие в одной и той же строке и одном и том же столбце:

$$A+B=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}+\begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix}=$$

$$=\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{pmatrix}.$$

Для ***умножения матрицы*** A ***на число*** нужно умножить на это число все элементы матрицы A:

$$\lambda A=\lambda\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}=\begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \dots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \dots & \lambda a_{2n} \\ \dots & \dots & \dots & \dots \\ \lambda a_{m1} & \lambda a_{m2} & \dots & \lambda a_{mn} \end{pmatrix}.$$

Например,

$$\begin{pmatrix} 2 & 1 & 3 \\ 1 & -3 & 5 \end{pmatrix}+\begin{pmatrix} 0 & 2 & -2 \\ 1 & 2 & -5 \end{pmatrix}=\begin{pmatrix} 2 & 3 & 1 \\ 2 & -1 & 0 \end{pmatrix},$$

$$3\begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \end{pmatrix}=\begin{pmatrix} 3 & 6 & 9 \\ -3 & 0 & 3 \end{pmatrix}.$$

Матрица, все элементы которой равны нулю, называется *нулевой* и обозначается O или O_{mn} (указано число строк и столбцов).

Из определения операций над матрицами вытекают их главные алгебраические свойства, совпадающие со свойствами операций над векторами:

5 Зак. № 2435 Турецкий

1. *Сложение матриц коммутативно*: $A + B = B + A$.
2. *Сложение матриц ассоциативно*: $A + (B + C) = (A + B) + C$.
3. $A + O = O + A = A$.
4. *Для матрицы* $-A = (-1)A$: $A + (-A) = (-A) + A = O$.
5. *Умножение на число ассоциативно*: $\lambda(\mu A) = (\lambda\mu)A$.
6. *Умножение на число дистрибутивно относительно сложения чисел*: $(\lambda + \mu)A = \lambda A + \mu A$.
7. *Умножение на число дистрибутивно относительно сложения матриц*: $\lambda(A + B) = \lambda A + \lambda B$.
8. $0 \cdot A = O$; $\lambda O = O$.
9. $1 \cdot A = A$; $(-\lambda)A = -\lambda A$.

Вместо $A + (-B)$ обычно пишут $A - B$, определяя таким образом вычитание. Из перечисленных свойств легко вывести некоторые следствия:

$$-(A + B) = -A - B;$$
$$-(-A) = A;$$
$$A + A = 2A, \quad A + A + A = 3A \quad \text{и т.д.}$$

5.2.2. Умножение матриц

В отличие от операций сложения и умножения на число операция умножения матрицы на матрицу определяется более сложно. Пусть заданы две матрицы A и B, причем число столбцов первой из них равно числу строк второй: например, матрица A имеет размерность $m \times n$, а матрица B — размерность $n \times k$. Если

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nk} \end{pmatrix},$$

то матрица размерности $m \times k$

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1k} \\ c_{21} & c_{22} & \dots & c_{2k} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mk} \end{pmatrix},$$

где $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}$ $(i = 1, 2, \dots, m;\ j = 1, 2, \dots, k)$, называется ***произведением матрицы A на матрицу B*** и обозначается AB или $A \cdot B$. Например,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} -1 & 1 & 1 \\ 2 & 1 & -1 \end{pmatrix} =$$

$$= \begin{pmatrix} 1\cdot(-1)+2\cdot 2 & 1\cdot 1+2\cdot 1 & 1\cdot 1+2\cdot(-1) \\ 3\cdot(-1)+4\cdot 2 & 3\cdot 1+4\cdot 1 & 3\cdot 1+4\cdot(-1) \\ 5\cdot(-1)+6\cdot 2 & 5\cdot 1+6\cdot 1 & 5\cdot 1+6\cdot(-1) \end{pmatrix} =$$

$$= \begin{pmatrix} 3 & 3 & -1 \\ 5 & 7 & -1 \\ 7 & 11 & -1 \end{pmatrix}.$$

Правило умножения матриц иногда формулируют следующим образом:

> ***Чтобы получить элемент, стоящий в i-й строке и j-м столбце произведения двух матриц, нужно элементы i-й строки первой матрицы умножить на соответствующие элементы j-го столбца второй матрицы и полученные произведения сложить.***

На математическом жаргоне иногда говорят: «нужно i-ю строку матрицы A умножить на j-й столбец матрицы B». Ясно, что для этого строка первой и столбец второй матрицы должны содержать одинаковое число элементов. В частности, при перемножении вектора-строки и вектора-столбца одной размерности (т.е. матриц размерности $1 \times n$ и $n \times 1$) получится число

$$(x_1, x_2, \dots, x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \dots + x_ny_n,$$

равное скалярному произведению $\mathbf{x} \cdot \mathbf{y}$ векторов $\mathbf{x}, \mathbf{y} \in R^n$. Наоборот, если перемножить вектор-столбец и вектор-строку, получится матрица размерности $n \times n$, в которой на месте (i, j) стоит произведение x_iy_j $(i, j = 1, 2, \dots, n)$.

Если квадратную матрицу $n \times n$ умножить на n-мерный вектор-столбец, в результате получится новый n-мерный вектор

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

где $y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n$, $i = 1, 2, \dots, n$. В частности, при умножении квадратной матрицы 2-го порядка A на двумерный вектор $\vec{r}$ получится двумерный вектор $\vec{p} = A\vec{r}$. Вспомним формулы преобразования координат на плоскости при повороте системы координат на угол α:

$$\begin{cases} x = x'\cos\alpha - y'\sin\alpha, \\ y = x'\sin\alpha + y'\cos\alpha, \end{cases} \qquad \begin{cases} x' = x\cos\alpha + y\sin\alpha, \\ y' = -x\sin\alpha + y\cos\alpha. \end{cases}$$

Если обозначить

$$\vec{r}_1 = \begin{pmatrix} x \\ y \end{pmatrix}; \quad \vec{r}_2 = \begin{pmatrix} x' \\ y' \end{pmatrix},$$

$$A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \quad B = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix},$$

связь между старыми и новыми координатами можно записать с помощью матричного умножения:

$$\vec{r}_1 = A\vec{r}_2, \quad \vec{r}_2 = B\vec{r}_1.$$

Выясним некоторые свойства умножения матриц.

∇**1.** *Умножение матриц некоммутативно: вообще говоря, $AB \neq BA$.* Например,

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 5 & 8 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 7 & 6 \\ 4 & 4 \end{pmatrix}.$$

Снова некоммутативная операция!!

∇**2.** $\lambda(AB) = (\lambda A)B = A(\lambda B)$.

Это свойство можно проверить непосредственным вычислением по определению. Так же проверяются и следующие свойства.

∇**3.** *Умножение матриц ассоциативно: $A(BC) = (AB)C$.*

Из ассоциативности умножения матриц следует, что произведение нескольких матриц A, B, C, D, ..., записанных в определенном порядке, не зависит от способа расстановки скобок. Поэтому можно говорить о произведении не только двух, но и большего числа матриц. Например, определено произведение четырех матриц $ABCD$, так как все пять возможных способов вычисления этого произведения

$$((AB)C)D,\ (A(BC))D,\ A((BC)D),\ A(B(CD)),\ (AB)(CD)$$

дают один и тот же результат.

∇**4.** *Умножение матриц дистрибутивно относительно сложения матриц:*

$$(A + B)C = AC + BC, \qquad C(A + B) = CA + CB.$$

Таким образом, в действиях с матрицами можно раскрывать скобки и выносить за скобки общий множитель.

Поскольку складывать и перемножать можно не любые матрицы, свойства $\nabla 3$ и $\nabla 4$ следует понимать так: если действия, указанные по одну сторону равенства, возможны, то возможны и действия, указанные по другую сторону равенства, и результаты в обеих частях одинаковы.

5.2.3. Транспонирование матриц

Рассмотрим произвольную матрицу размерности $m \times n$

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Если в ней заменить строки на столбцы, получится матрица размерности $n \times m$

$$A' = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix},$$

которая называется *транспонированной* по отношению к A. Можно проверить, что справедливы следующие правила транспонирования.

∇5. $(\alpha A + \beta B)' = \alpha A' + \beta B'$, *где* α, β — *какие-либо числа.*

∇6. $(AB)' = B'A'$.

Не доказывая свойство ∇6, уясним хотя бы, что слева и справа стоят матрицы одинаковой размерности. Действительно, если матрица A имеет размерность $m \times n$, а матрица B — размерность $n \times k$, то размерность AB есть $m \times k$ и, соответственно, $(AB)'$ имеет размерность $k \times m$. С другой стороны, B' имеет размерность $k \times n$, A' — размерность $n \times m$ и, стало быть, $B'A'$ тоже имеет размерность $k \times m$. Пусть, например,

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Тогда

$$AB = \begin{pmatrix} 1 & 1 \\ 3 & 1 \end{pmatrix}$$

и

$$(AB)' = \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix}.$$

При этом

$$B' = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad A' = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

и

$$B'A' = \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix}.$$

Заметим, что если считать n-мерный вектор $\mathbf{x}$ вектором-столбцом, то транспонированный вектор $\mathbf{x}'$ есть вектор-строка. Поэтому скалярное произведение двух векторов $\mathbf{x}, \mathbf{y} \in R^n$ можно представить в виде $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}'\mathbf{y}$.

5.2.4. Квадратные матрицы

Ранее уже отмечалось, что не всякие две матрицы можно сложить или перемножить, так как для осуществления таких операций необходимы известные соотношения между числами строк и столбцов. Это неудобство исчезает, если рассматривать только квадратные матрицы некоторого фиксированного порядка n. Любые две такие матрицы можно сложить или перемножить, а также умножить

на любое число. В результате всех этих операций снова получаются квадратные матрицы того же порядка n.

Элементы, стоящие на главной диагонали квадратной матрицы (идущей из левого верхнего в правый нижний угол), называются *диагональными*. Квадратная матрица, все диагональные элементы которой равны единице, а остальные — нулю, называется *единичной* и обозначается E или E_n, где n — ее порядок. Таким образом,

$$E = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Непосредственным вычислением проверяется следующее свойство.

∇7. *Для любой квадратной матрицы*

$$AE = EA = A.$$

Квадратная матрица A называется *обратимой*, если существует квадратная матрица X, удовлетворяющая соотношениям

$$AX = XA = E.$$

Матрица X называется *обратной* к матрице A.

∇8. *Если обратная матрица существует, то она единственна.*

В самом деле, пусть нашлась еще одна матрица Y, такая, что $AY = YA = E$. Умножая равенство $AY = E$ слева на матрицу X, получим $X(AY) = (XA)Y = EY = Y = XE = X$, т.е. $Y = X$.

Матрица, обратная к матрице A, обозначается A^{-1}. Если обратная матрица существует,

$$AA^{-1} = A^{-1}A = E.$$

Пусть, например,

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Тогда нетрудно убедиться, что

$$A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}.$$

Вообще, если

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

то

$$A^{-1} = \frac{1}{|A|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

где $|A| = ad - bc$ — определитель матрицы A (разумеется, формула справедлива при $|A| \neq 0$). Предлагаем читателю самостоятельно убедиться в этом равенстве. В предыдущем примере $|A| = -2$ и $4/(-2) = -2$, $(-2)/(-2) = 1$, $(-3)/(-2) = {}^3/_2$, $1/(-2) = -{}^1/_2$.

В определение обратной матрицы матрицы A и A^{-1} входят симметрично. Поэтому матрица A является, в свою очередь, обратной к матрице A^{-1}.

∇**9.** $(A^{-1})^{-1} = A$.

∇**10.** *Если квадратные матрицы A и B обратимы, то их произведение AB также обратимо и справедлива формула*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

В самом деле, $(AB)(B^{-1}A^{-1}) = (A(BB^{-1}))A^{-1} = (AE)A^{-1} = AA^{-1} = E$. Равенство $(B^{-1}A^{-1})(AB) = E$ проверяется аналогично.

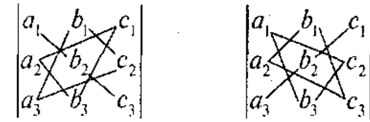
Итак, операция обращения действует на произведение матриц так же, как операция транспонирования: матрица, обратная произведению, есть произведение обратных матриц в противоположном порядке.

5.3. Определители и системы линейных уравнений

Обобщим понятие определителя 2-го порядка, рассмотренное в главе 4, на случай квадратной матрицы произвольного порядка n. Удобнее всего сделать это по индукции. Рассмотрим таблицу из девяти чисел, расположенных в три ряда и три столбца (квадратную матрицу 3×3). Определителем 3-го порядка называется число

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 + b_1c_2a_3 + c_1a_2b_3 - c_1b_2a_3 - b_1a_2c_3 - a_1c_2b_3.$$

Для запоминания этой громоздкой на первый взгляд формулы существует удобное мнемоническое «правило треугольников»:



Диагональ, идущая в таблице из левого верхнего угла (a_1) в правый нижний (c_3), называется главной, вторая диагональ (от c_1 к a_3) — побочной. Посмотрим, как образуются три слагаемых в определении, которые берутся со знаком «+». Сначала перемножаются элементы главной диагонали ($a_1b_2c_3$). Затем строятся два треугольника: один с основанием выше главной диагонали, а второй — ниже. Перемножаются элементы, стоящие в вершинах полученных треугольников: $b_1c_2a_3$ и $c_1a_2b_3$. Слагаемые со знаком «–» образуются точно так же, но относительно побочной диагонали.

Вычисление определителя 3-го порядка по определению, даже с использованием правила треугольников, — занятие утомительное и неблагодарное. Оказывается, определитель можно посчитать проще. Шесть слагаемых можно, например, сгруппировать следующим образом:

$$\Delta = a_1b_2c_3 + b_1c_2a_3 + c_1a_2b_3 - c_1b_2a_3 - b_1a_2c_3 - a_1c_2b_3 =$$
$$= a_1(b_2c_3 - b_3c_2) + a_2(b_3c_1 - b_1c_3) + a_3(b_1c_2 - b_2c_1).$$

Мы вынесли за скобки элементы a_1, a_2, a_3 первого столбца. Посмотрим внимательнее на величины в скобках. Возьмем, например, элемент a_1 и вычеркнем из определителя строку и столбец, на пересечении которых расположен этот элемент (в данном случае — первую строку и первый столбец). В результате останется определитель 2-го порядка

$$\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} = b_2c_3 - b_3c_2.$$

Такой определитель называется минором элемента a_1. *Минором* произвольного элемента называется определитель, который получается вычеркиванием строки и столбца, на пересечении которых расположен данный элемент. В частности, миноры элементов a_2 и a_3 есть соответственно определители

$$\begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} = b_1c_3 - b_3c_1, \quad \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} = b_1c_2 - b_2c_1.$$

Теперь видно, что в полученном разложении a_1 умножается на свой минор, a_2 — на минор с обратным знаком, а a_3 — снова на минор. *Алгебраическим дополнением* элемента определителя называется его минор, взятый со своим знаком, если сумма номеров строки и столбца, в которых расположен элемент, четна, и с обратным знаком, если сумма номеров нечетна. Элемент a_1 расположен в первой строке и первом столбце, сумма номеров $1 + 1 = 2$ четна, и алгебраическое дополнение равно минору. Для элемента a_2, расположенного во второй строке и первом столбце, сумма $2 + 1 = 3$ нечетна и алгебраическое дополнение равно минору с обратным знаком. У элемента a_3 сумма снова четна. Теперь ясно, что каждый из элементов a_1, a_2, a_3 умножается на свое алгебраическое дополнение. Если обозначить алгебраические дополнения элементов соответствующими прописными буквами (алгебраическое дополнение элемента a_1 – через A_1, элемента b_1 – через B_1 и т.д.), полученное нами равенство запишется в простом виде:

$$\Delta = a_1A_1 + a_2A_2 + a_3A_3.$$

Говорят, что определитель *разложили* по первому столбцу. Оказывается, определитель можно разложить по любому столбцу и любой строке. Справедлива следующая теорема.

Теорема 5.1. *Определитель равен сумме произведений элементов какого-нибудь столбца (или строки) на их алгебраические дополнения.*

Теорема доказывается непосредственным вычислением, как мы это сделали при разложении по первому столбцу. Итак, справедливы шесть равенств:

По столбцам	По строкам
$\Delta = a_1A_1 + a_2A_2 + a_3A_3$	$\Delta = a_1A_1 + b_1B_1 + c_1C_1$
$\Delta = b_1B_1 + b_2B_2 + b_3B_3$	$\Delta = a_2A_2 + b_2B_2 + c_2C_2$
$\Delta = c_1C_1 + c_2C_2 + c_3C_3$	$\Delta = a_3A_3 + b_3B_3 + c_3C_3$

Это правило сводит вычисление определителя 3-го порядка к вычислению трех определителей 2-го порядка, что, несомненно, проще. Чтобы узнать, с каким знаком брать минор в алгебраическом дополнении, полезно иметь в виду следующую схему:

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix},$$

где знаком «+» помечены места тех элементов, для которых алгебраические дополнения равны минорам, взятым с их собственными знаками.

Пример.

Вычислим определитель

$$\Delta = \begin{vmatrix} 2 & 4 & 6 \\ 5 & 12 & 19 \\ 3 & 9 & 17 \end{vmatrix},$$

разложив его по элементам первой строки. Итак,

$$\Delta = 2 \cdot \begin{vmatrix} 12 & 19 \\ 9 & 17 \end{vmatrix} - 4 \cdot \begin{vmatrix} 5 & 19 \\ 3 & 17 \end{vmatrix} + 6 \cdot \begin{vmatrix} 5 & 12 \\ 3 & 9 \end{vmatrix} =$$

$$= 2 \cdot (12 \cdot 17 - 9 \cdot 19) - 4 \cdot (5 \cdot 17 - 3 \cdot 19) + 6 \cdot (5 \cdot 9 - 3 \cdot 12) = 8.$$

Таким образом, вычисление определителя 3-го порядка сводится к вычислению трех определителей 2-го порядка. Точно так же разложением по строке или столбцу можно задать определитель 4-го порядка. Пусть задана квадратная матрица 4-го порядка

$$A = \begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{pmatrix}.$$

Назовем определителем этой матрицы (определителем 4-го порядка) число

$$|A| = \det A = \begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{vmatrix} =$$

$$= a_1 \begin{vmatrix} b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \\ b_4 & c_4 & d_4 \end{vmatrix} - b_1 \begin{vmatrix} a_2 & c_2 & d_2 \\ a_3 & c_3 & d_3 \\ a_4 & c_4 & d_4 \end{vmatrix} +$$

$$+ c_1 \begin{vmatrix} a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \\ a_4 & b_4 & d_4 \end{vmatrix} - d_1 \begin{vmatrix} a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \\ a_4 & b_4 & c_4 \end{vmatrix}.$$

Мы задали определитель 4-го порядка разложением по первой строке. Поскольку мы умеем вычислять определители 3-го порядка, вычисление определителя 4-го порядка по этой формуле не составит принципиальных трудностей (хотя будет достаточно громоздким). Принцип построения формулы очень прост: элемент строки умножается на определитель (минор), который получается вычеркиванием соответствующих строки и столбца и берется с соответствующим знаком. Разложение определителя 4-го порядка по любой строке и любому столбцу дает одинаковый результат. Проверим это.

Пример.

Посчитаем двумя способами определитель

$$\begin{vmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{vmatrix}.$$

1. Разложение по первой строке:

$$\begin{vmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{vmatrix} - \begin{vmatrix} 0 & 2 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{vmatrix} +$$

$$+(-1)\cdot\begin{vmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} - 2\cdot\begin{vmatrix} 0 & 1 & 2 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{vmatrix} =$$

$$= -1\cdot(-2) - 2\cdot(-1)\cdot(-2) = 2 - 4 = -2$$

(первый и третий определители равны нулю, так как содержат одинаковые столбцы (см. свойство ∇1 на с. 141); второй и четвертый определители разложили по первому столбцу).

2. Разложение по четвертому столбцу:

$$\begin{vmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{vmatrix} = (-2)\cdot\begin{vmatrix} 0 & 1 & 2 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{vmatrix} + \begin{vmatrix} 1 & 1 & -1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{vmatrix} -$$

$$-\begin{vmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 0 & 1 & 0 \end{vmatrix} + \begin{vmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 1 & 1 & 0 \end{vmatrix} =$$

$$= (-2)\cdot 2 - 1\cdot 1 - (-2) + \begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix} + \begin{vmatrix} 1 & -1 \\ 1 & 2 \end{vmatrix} = -4 - 1 + 2 - 2 + 3 = -2$$

(первый определитель разложили по первому столбцу; второй – по третьему; третий и четвертый – по первому столбцу).

Пусть мы умеем вычислять определители $(n-1)$-го порядка и задана квадратная матрица n-го порядка

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

Назовем *минором* элемента a_{ij} определитель $(n-1)$-го порядка, получаемый из матрицы A вычеркиванием i-й строки и j-го столбца. Назовем *алгебраическим дополнением* A_{ij} элемента a_{ij} его минор, если число $i+j$ четно, и минор с противоположным знаком, если $i+j$ нечетно. Тогда *определителем* матрицы A (определителем n-го порядка) называется число

$$|A| = \det A = a_{i1}A_{i1} + a_{i2}A_{i2} + \dots + a_{in}A_{in}.$$

Можно проверить, что определенное таким образом число не зависит от номера i-й строки, по которой разлагается определитель. Аналогично определитель n-го порядка можно разложить и по j-му столбцу:

$$|A| = \det A = a_{1j}A_{1j} + a_{2j}A_{2j} + \dots + a_{nj}A_{nj}.$$

При этом получится то же самое число.

Сформулируем кратко некоторые свойства определителей n-го порядка (их легко проверить непосредственным вычислением для определителей 2-го и 3-го порядка).

∇1. *Определитель равен нулю, если содержит две одинаковые строки (два одинаковых столбца) или нулевую строку (нулевой столбец).*

∇2. *Определитель меняет знак при перестановке двух строк (столбцов).*

∇3. *При умножении строки (столбца) на число определитель умножается на это число.*

∇4. *Определитель не меняется, если к строке (столбцу) прибавить любую другую строку (другой столбец), умноженный на произвольное число.*

Указанные свойства позволяют упростить вычисление определителя. С их помощью определитель последовательно преобразуют к такому виду, чтобы в какой-нибудь строке (столбце) все элементы, кроме одного, стали нулевыми.

Рассмотрим определитель из предыдущего примера:

$$\begin{vmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & -1 & -1 & 2 \\ 0 & 0 & 2 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{vmatrix} = -(-1)\begin{vmatrix} 0 & 2 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{vmatrix} =$$

$$= -(-1)\cdot 1\cdot(-2) = -2.$$

Мы вычли четвертый столбец из второго. При этом во втором столбце остался только один ненулевой элемент (-1). Минор этого элемента содержит только один ненулевой элемент в третьей строке (1), по которой его и разложили. Вычисление определителя существенно упростилось.

Добавим некоторые новые свойства определителей n-го порядка, связанные со свойствами квадратных матриц и n-мерных векторов.

∇5. *Определитель не меняется при транспонировании матрицы:*

$$|A| = |A'|.$$

∇6. *Определитель единичной матрицы равен единице:* $\det E = 1$.

∇7. *Если матрицу умножить на число* λ, *определитель умножится на* λ^n:

$$|\lambda A| = \lambda^n |A|.$$

Умножение одной строки на λ приводит к умножению определителя на это число. При умножении матрицы на число на него умножаются все n строк, что соответствует n-кратному умножению определителя на λ, т.е. умножению на λ^n.

∇8. *Определитель произведения матриц равен произведению определителей:*

$$|AB| = |A|\cdot|B|.$$

∇9. *Пусть* $|A| \neq 0$. *Тогда определитель обратной матрицы связан с определителем данной матрицы следующим соотношением:*

$$|A^{-1}| = \frac{1}{|A|}.$$

В самом деле, $|AA^{-1}| = |A|\cdot|A^{-1}| = |E| = 1$, откуда и следует данное равенство.

Мы применяли определители 2-го порядка для исследования системы двух линейных уравнений с двумя неизвестными. Точно так же с помощью определителей n-го порядка можно исследовать системы линейных уравнений с n неизвестными.

Рассмотрим систему линейных уравнений n-го порядка

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n. \end{cases}$$

Можно обозначить вектор неизвестных через $\mathbf{x}$, вектор правых частей уравнений через $\mathbf{b}$, а матрицу коэффициентов через A:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & & \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}.$$

Тогда по определению матричного произведения систему линейных уравнений можно записать короче:

$$A\mathbf{x} = \mathbf{b}.$$

При исследовании системы линейных уравнений главную роль, как и в двумерном случае, играет определитель $\Delta = \det A$ матрицы, составленный из коэффициентов системы. Его называют главным определителем системы. Наряду с главным, введем еще n определителей по следующему правилу: определитель Δ_j получается из определителя Δ заменой j-го столбца на столбец $\mathbf{b}$ правых частей уравнений ($j = 1, 2, \ldots, n$). Для системы n-го порядка оказывается справедливым тот же закон, что и для систем двух уравнений.

Правило Крамера.

1. Если $\Delta \neq 0$, система $A\mathbf{x} = \mathbf{b}$ имеет единственное решение
$$x_1 = \Delta_1/\Delta,\ x_2 = \Delta_2/\Delta,\ \ldots,\ x_n = \Delta_n/\Delta.$$
2. Если $\Delta = 0$, а хотя бы один из определителей $\Delta_1, \Delta_2, \ldots, \Delta_n$ не равен нулю, то система несовместна.
3. Если $\Delta = \Delta_1 = \Delta_2 = \ldots = \Delta_n = 0$, система имеет бесконечно много решений.

∇**10.** *Система линейных уравнений, в которой число уравнений равно числу неизвестных, имеет единственное решение тогда и только тогда, когда ее главный определитель отличен от нуля.*

Пример.

Рассмотрим систему 3-го порядка

$$\begin{cases} x_1 + 2x_2 + \ x_3 = 1, \\ \qquad\ \ x_2 + \ x_3 = 1, \\ 2x_1 + \ x_2 + 3x_3 = 7. \end{cases}$$

Решим ее сначала «школьным» способом, т.е. исключением переменных. Если вычесть второе уравнение из первого, получим $x_1 + x_2 = 0$, откуда $x_1 = -x_2$. Подставив это выражение в третье уравнение, получим уравнение $-x_2 + 3x_3 = 7$, которое сложим со вторым: $4x_3 = 8$, откуда $x_3 = 2$. Значит, из второго уравнения получим $x_2 = 1 - 2 = -1$, и, стало быть, $x_1 = 1$. Итак, система имеет решение $x_1 = 1$, $x_2 = -1$, $x_3 = 2$.

Теперь проверим найденное решение с помощью правила Крамера:

$$\Delta = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 2 & 1 & 3 \end{vmatrix} = 1 \cdot \begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} + 2 \cdot \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = 2 + 2 = 4$$

(мы разложили этот определитель по первому столбцу);

$$\Delta_1 = \begin{vmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 7 & 1 & 3 \end{vmatrix} = 1 \cdot \begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} - 2 \cdot \begin{vmatrix} 1 & 1 \\ 7 & 3 \end{vmatrix} + 1 \cdot \begin{vmatrix} 1 & 1 \\ 7 & 1 \end{vmatrix} =$$

$$= 2 - 2 \cdot (-4) + (-6) = 4;$$

(разложили по первой строке);

$$\Delta_2 = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 2 & 7 & 3 \end{vmatrix} = 1 \cdot \begin{vmatrix} 1 & 1 \\ 7 & 3 \end{vmatrix} + 2 \cdot \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = -4 + 0 = -4;$$

$$\Delta_3 = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 2 & 1 & 7 \end{vmatrix} = 1 \cdot \begin{vmatrix} 1 & 1 \\ 1 & 7 \end{vmatrix} + 2 \cdot \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = 6 + 2 = 8.$$

Значит, по правилу Крамера $x_1 = 4/4 = 1$, $x_2 = (-4)/4 = -1$, $x_3 = 8/4 = 2$. Как и следовало ожидать, получено то же самое решение.

С помощью правила Крамера можно сделать выводы о решении однородной системы линейных уравнений, т.е. системы, в которой правые части всех уравнений равны нулю:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = 0. \end{cases}$$

Очевидно, однородная система имеет нулевое решение $x_1 = x_2 = \ldots = x_n = 0$. Все определители $\Delta_1, \Delta_2, \ldots, \Delta_n$ равны нулю, так как содержат нулевой столбец. Поэтому, если главный определитель однородной системы равен нулю, система имеет бесконечно много решений.

∇**11.** *Однородная система линейных уравнений, в которой число уравнений равно числу неизвестных, имеет ненулевое решение тогда и только тогда, когда ее главный определитель равен нулю.*

Предположим, что матрица A коэффициентов системы обратима. Тогда равенство $A\mathbf{x} = \mathbf{b}$ можно умножить слева на обратную матрицу A^{-1} и сразу получить вектор решения:

$$A^{-1}(A\mathbf{x}) = (A^{-1}A)\mathbf{x} = E\mathbf{x} = \mathbf{x} = A^{-1}\mathbf{b}.$$

Итак, если матрица обратима, система имеет единственное решение, которое можно записать в виде $\mathbf{x} = A^{-1}\mathbf{b}$. Обратное утверждение также верно.

∇**12.** *Система линейных уравнений, в которой число уравнений равно числу неизвестных, имеет единственное решение тогда и только тогда, когда матрица ее коэффициентов обратима.*

Из свойств $\nabla 11$ и $\nabla 12$ следует важный вывод, который можно сформулировать в виде теоремы.

Теорема 5.2. *Матрица A обратима тогда и только тогда, когда* $\det A \neq 0$.

Можно проверить, что матрица, обратная матрице коэффициентов из предыдущего примера, имеет вид

$$A^{-1} = \begin{pmatrix} 1/2 & -5/4 & 1/4 \\ 1/2 & 1/4 & -1/4 \\ -1/2 & 3/4 & 1/4 \end{pmatrix}.$$

Значит, решение системы получится, если умножить эту матрицу на вектор правых частей уравнений:

$$\mathbf{x} = \begin{pmatrix} 1/2 & -5/4 & 1/4 \\ 1/2 & 1/4 & -1/4 \\ -1/2 & 3/4 & 1/4 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 7 \end{pmatrix} =$$

$$= \begin{pmatrix} 1/2 - 5/4 + 7/4 \\ 1/2 + 1/4 - 7/4 \\ -1/2 + 3/4 + 7/4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

Как мы и предполагали, получено то же решение.

5.4. Линейное пространство

5.4.1. Понятие и примеры линейного пространства

Покажем, что пространство $\boldsymbol{R}^n$ является частным случаем более общей структуры, называемой линейным пространством. При этом свойства пространства $\boldsymbol{R}^n$ превращаются в аксиомы линейного пространства. Итак, пусть имеется множество E, над элементами которого заданы две операции:

1) **сложение**. Для любых двух элементов x, y множества E однозначно определен некоторый элемент множества E, называемый *суммой* этих элементов и обозначаемый $x + y$;

2) **умножение на число**. Для любого элемента $x \in E$ и любого действительного числа λ определен некоторый элемент множества E, называемый *произведением* λ *на* x и обозначаемый λx или $\lambda \cdot x$.

Забегая вперед, можно сказать, что на множестве E заданы таким образом внутренняя и внешняя бинарные операции. Свойства бинарных операций подробно исследуются в главе 6.

Множество E называется ***линейным*** (или ***векторным***) ***пространством***, если операции сложения и умножения на число удовлетворяют следующим требованиям (аксиомам линейного пространства):

1°. Сложение ассоциативно:

$$x + (y + z) = (x + y) + z.$$

2°. Существует элемент $o \in E$, такой, что для всех $x \in E$

$$x + o = o + x$$

(элемент o называется *нейтральным* относительно сложения или *нулевым* элементом E).

3°. Для каждого элемента $x \in E$ существует элемент $-x \in E$, такой, что

$$x + (-x) = (-x) + x = o$$

(элемент $-x$ называется *противоположным* элементу x).

4°. Сложение коммутативно:

$$x + y = y + x$$

5°. Умножение на число ассоциативно:

$$\lambda(\mu x) = (\lambda\mu)x.$$

6°. Умножение на число дистрибутивно относительно сложения в E:

$$\lambda(x + y) = \lambda x + \lambda y.$$

7°. Умножение на число дистрибутивно относительно сложения чисел:

$$(\lambda + \mu)x = \lambda x + \mu x.$$

8°. Единица нейтральна при умножении:

$$1 \cdot x = x.$$

Следует еще раз подчеркнуть, что множество является линейным пространством только в том случае, когда над его элементами заданы две операции – сложения и умножения на число. Если в одном и том же множестве нам удастся задать операции двумя разными способами, мы получим р а з н ы е линейные пространства!

Примеры линейных пространств.

1. Совокупность свободных векторов на плоскости или в пространстве с операциями сложения векторов и умножения вектора на число образует линейное пространство. Это видно из основных свойств векторов. Нейтральный элемент есть вектор $\vec{0}$.

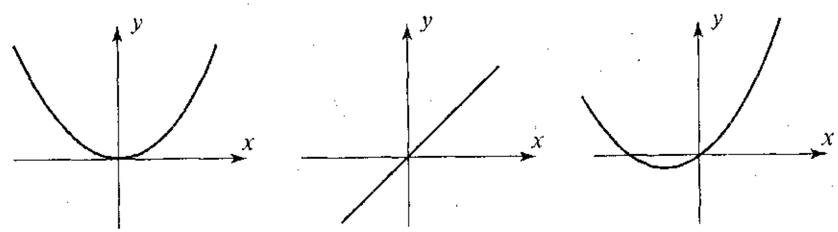
2. Пространство $\boldsymbol{R}^n$ есть линейное пространство с операциями сложения n-мерных векторов и умножения на число. Нейтральным элементом является вектор $\mathbf{0}$ с нулевыми координатами.

Пространство векторов — типичное линейное пространство. Поэтому линейное пространство называют еще векторным, а его элементы, независимо от их природы, векторами.

3. Совокупность прямоугольных матриц фиксированной размерности $m \times n$ образует линейное пространство относительно операций сложения матриц и умножения матрицы на число. Нейтральный элемент — нулевая матрица O. Кстати, пространство $\boldsymbol{R}^n$ — частный случай пространства матриц при $m = 1$.

4. Множество действительных чисел $\boldsymbol{R}$ есть линейное пространство с обычными арифметическими операциями сложения и умножения чисел. Нейтральный элемент — нуль.

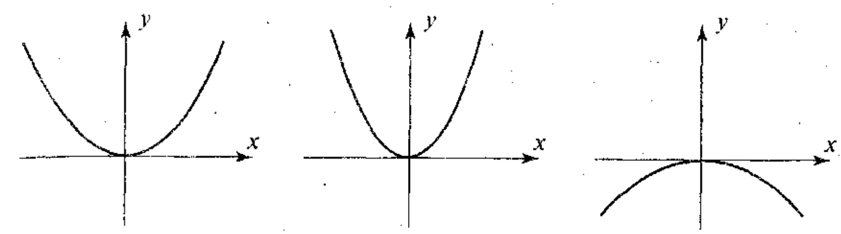
5. Множество числовых функций образует линейное пространство. Рассмотрим это множество. Естественным образом определяется операция сложения функций: функция $f + g$ каждому аргументу x ставит в соответствие значение $f(x) + g(x)$. Поэтому по графикам функций $y = f(x)$ и $y = g(x)$ легко построить график их суммы $y = f(x) + g(x)$ (рис. 5.1).



$y = x^2$ $\qquad$ $y = x$ $\qquad$ $y = x^2 + x$

Рис. 5.1

Операция умножения функции на число определяется аналогично: λf есть новая функция, которая каждому аргументу x ставит в соответствие значение $\lambda f(x)$. Мы хорошо знаем, что график новой функции $y = \lambda f(x)$ получается из графика функции $y = f(x)$ растяжением в λ раз по оси ординат. Растяжение в данном случае может быть и сжатием (при $|\lambda| < 1$), и даже отражением относительно оси абсцисс (при $\lambda < 0$) (рис. 5.2).



$y = x^2$ $\qquad$ $y = 2x^2$ $\qquad$ $y = -0{,}5x^2$

Рис. 5.2

Эти операции полностью соответствуют школьным представлениям о действиях над функциями.

Нетрудно убедиться, что множество числовых функций образует линейное пространство. В самом деле, сложение функций ассоциативно и коммутативно: $f + (g + h) = (f + g) + h$, $f + g = g + f$. Нейтральным элементом является тождественный нуль: $f(x) \equiv 0$. Противоположный элемент для каждой функции f есть функция $-f$. Кроме того, операция умножения функции на число удовлетворяет условиям 5°–8°:

$(\lambda\mu)f = \lambda(\mu f)$ (ассоциативность относительно умножения чисел);

$\lambda(f + g) = \lambda f + \lambda g$ (дистрибутивность относительно сложения функций);

$(\lambda + \mu)f = \lambda f + \mu f$ (дистрибутивность относительно сложения чисел);

$1 \cdot f = f$ (нейтральность действия единицы).

Итак, совокупность числовых функций действительно представляет собой линейное пространство.

6. Частный случай числовой функции — многочлен. Например, $x + 5$, $3x^2 + 2x + 1$, $1{,}5x^4 - 0{,}5x^2$ — это многочлены. В общем виде многочлен степени n можно представить как $a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0$, где $a_0, a_1, a_2, \ldots, a_{n-1}, a_n$ — действительные числа, называемые коэффициентами многочлена.

Правила сложения и умножения функций на число становятся исключительно простыми для многочленов. Чтобы сложить два многочлена, нужно сложить коэффициенты при равных степенях. Например, $(x^3 + 3x^2 + 2x + 2) + (2x^2 + 1) = (x^3 + 3x^2 + 2x + 2) + (0 \cdot x^3 + 2x^2 + 0 \cdot x + 1) = (1 + 0) \cdot x^3 + (3 + 2) \cdot x^2 + (2 + 0) \cdot x + (2 + 1) = x^3 + 5x^2 + 2x + 3$. Для умножения многочлена на число нужно каждый коэффициент умножить на это число. Например, $3(x^3 - 1{,}5x) = 3x^3 - 4{,}5x$. Нетрудно убедиться, что совокупность многочленов с такими операциями представляет собой линейное пространство.

Ассоциативность и коммутативность сложения позволяет складывать произвольное конечное число элементов, не задумываясь о порядке слагаемых и способе расстановки скобок. Например,

$$(a + b) + (c + d) = c + (a + (d + b)).$$

Операция сложения с противоположным элементом называется *вычитанием*. При этом сумму $x + (-y)$ обозначают $x - y$ и называют *разностью* элементов x и y.

Из аксиом выводятся простые следствия.

∇1. $-(-x) = x; \quad -(x + y) = -x - y.$

∇2. $0 \cdot x = o.$

∇3. $\lambda \cdot o = o.$

∇4. $mx = \underbrace{x + x + \ldots + x}_{m \text{ раз}}$ (*m — натуральное число*).

∇5. $(-\lambda)x = -(\lambda x).$

∇6. *Если $x \neq o$, то из $\lambda x = \mu x$ следует $\lambda = \mu$.*

∇7. Если $\lambda \neq 0$, *то из $\lambda x = \lambda y$ следует $x = y$.*

В качестве упражнения читатель может доказать эти свойства, опираясь на аксиомы.

5.4.2. Линейное подпространство

Подмножество F линейного пространства E называется ***линейным подпространством***, если выполняются следующие условия:

1) если $x, y \in F$, то $x + y \in F$;

2) если $x \in F$, $\lambda \in R$, то $\lambda x \in F$.

Другими словами, множество $F \subseteq E$ является линейным подпространством, если оно з а м к н у т о относительно операций сложения и умножения на число. Легко понять, что все восемь аксиом линейного пространства выполняются и в подпространстве. В частности, $o \in F$, поскольку $o = x + (-1)x$.

∇8. *Линейное подпространство само является линейным пространством.*

Примеры подпространств.

1. Самое простое подпространство – множество $F = \{o\}$, состоящее из одного элемента: нуля пространства E. В самом деле, $o + o = o \in F$ и $\lambda o = o \in F$.

2. Любое пространство E является своим собственным подпространством.

3. В качестве исходного пространства E возьмем совокупность свободных векторов на плоскости. Проведем на плоскости прямую и рассмотрим множество всех векторов, параллельных этой прямой. Очевидно, это множество есть подпространство пространства E. При сложении двух коллинеарных векторов снова получится вектор, им коллинеарный, то же самое при умножении вектора на число. Таким образом, определение подпространства выполняется. Итак, говоря коротко, прямая есть подпространство плоскости. Точно так же плоскость – подпространство трехмерного пространства.

4. Ранее рассматривались пространства функций и многочленов (см. подпараграф 5.4.1). Ясно, что второе есть подпространство первого.

5. Пусть определитель матрицы A равен нулю. Тогда множество решений однородной системы $A\mathbf{x} = 0$ образует подпространство пространства $\boldsymbol{R}^n$. В самом деле, пусть $\mathbf{x}$ и $\mathbf{y}$ – два решения однородной системы. Тогда $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = 0 + 0 = 0$, т.е. сумма $\mathbf{x} + \mathbf{y}$ также является решением однородной системы. Точно так же $A(\lambda\mathbf{x}) = \lambda(A\mathbf{x}) = \lambda 0 = 0$ и, значит, $\lambda\mathbf{x}$ – решение системы. Итак, множество решений однородной системы замкнуто относительно операций сложения и умножения на число и, следовательно, является подпространством.

6. Пусть $X = \{a_1, a_2, \ldots, a_m\}$ – конечное множество элементов из линейного пространства E. Определим множество F элементов $x \in E$ вида $x = \lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_m a_m$, где $\lambda_1, \lambda_2, \ldots, \lambda_m$ – произвольные числа. Такие элементы x называются *линейными комбинациями* элементов $a_1, a_2, \ldots, a_m$. Легко убедиться, что множество F линейных комбинаций является подпространством пространства E. Его называют *подпространством, порожденным элементами* $a_1, a_2, \ldots, a_m$, или *линейной оболочкой* векторов $a_1, a_2, \ldots, a_m$. Векторы $a_1, a_2, \ldots, a_m$ называются *системой образующих* подпространства F.

5.4.3. Норма и расстояние

Пусть E – векторное пространство. ***Нормой*** вектора $x \in E$ называется число $\|x\|$, обладающее следующими свойствами:

1°. $\|x\| \geqslant 0$; $\|x\| = 0$, тогда и только тогда, когда $x = o$.

2°. $\|\lambda x\| = |\lambda| \cdot \|x\|$ для любого числа λ и любого вектора $x \in E$.

3°. $\|x + y\| \leqslant \|x\| + \|y\|$ для любых векторов $x, y \in E$.

Длина обычного геометрического вектора есть норма в пространстве свободных векторов. Условие 3°, называемое неравенством треугольника, отражает в данном случае общеизвестный факт: длина стороны треугольника не превосходит суммы длин двух других сторон. Мы обращали на это внимание, когда рассматривали векторы. В пространстве $\boldsymbol{R}^n$ также вводилась норма для вектора $\mathbf{x} = (x_1, x_2, \ldots, x_n)$:

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}\,.$$

Свойства нормы (см. подпараграф 5.1.3) проверялись в соответствующих разделах. Оказывается, это не единственный способ задать норму в $\boldsymbol{R}^n$. В частности, каждая из формул

$$\|\mathbf{x}\| = |x_1| + |x_2| + \ldots + |x_n|,$$

$$\|\mathbf{x}\| = \max_{i=1,\ldots,n} |x_i|$$

определяет норму в $\boldsymbol{R}^n$. Читатель может убедиться, что аксиомы нормы в этих случаях выполняются. Итак, задать норму (нормировать пространство) можно по-разному.

Множество векторов $x \in E$, удовлетворяющих условию

$$\|x\| \leqslant 1,$$

называется *единичным шаром* в пространстве E. Вид единичного шара, разумеется, зависит от того, как задана норма. Посмотрим, какие «шары» получатся в пространстве $\boldsymbol{R}^2$, т.е. на плоскости, для различных норм. В данном случае мы не делаем различий между векторами и точками на плоскости (как концами соответствующих радиусов-векторов). Для обычной нормы $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$ получаем множество $\{(x_1, x_2) | \, x_1^2 + x_2^2 \leqslant 1\}$, т.е. обычный круг с радиусом, равным 1, и центром в начале координат (рис. 5.3). Норма $\|\mathbf{x}\| = |x_1| + |x_2|$ порождает множество $\{(x_1, x_2) | \, |x_1| + |x_2| \leqslant 1\}$. Нетрудно убедиться, что это квадрат со стороной, равной 2, и с вершинами в точках $(1, 1)$, $(1, -1)$, $(-1, -1)$, $(-1, 1)$ (рис. 5.4). Наконец, единичный шар для третьей нормы $\|\mathbf{x}\| = \max\{|x_1|, |x_2|\}$ тоже представляет собой квадрат, но со стороной, равной $\sqrt{2}$, и вершинами $(0, 1)$, $(1, 0)$, $(0, -1)$, $(-1, 0)$ (рис. 5.5).

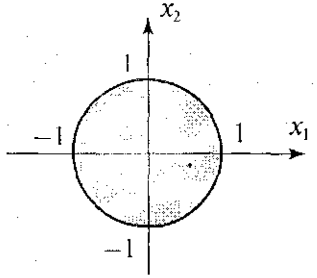


Рис. 5.3

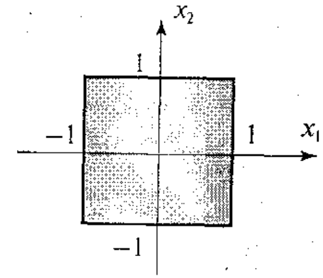


Рис. 5.4

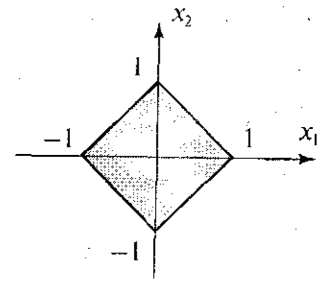


Рис. 5.5

Величину $\|x - y\|$ называют ***расстоянием между векторами*** x и y и обозначают $\rho(x, y)$. Это обобщение обычного геометрического расстояния между точками. Справедливы следующие простые свойства расстояния, вытекающие из аксиом нормы:

∇**1.** $\rho(x, x) = 0$; $\rho(x, y) > 0$ *при* $x \neq y$.

∇**2.** $\rho(x, y) = \rho(y, x)$.

∇**3.** $\rho(x, y) + \rho(y, z) \geqslant \rho(x, z)$.

Последнее свойство следует из неравенства треугольника:

$$\rho(x, z) = \|x - z\| = \|(x - y) + (y - z)\| \leqslant \|x - y\| + \|y - z\| = \\ = \rho(x, y) + \rho(y, z).$$

5.4.4. Скалярное произведение. Евклидово пространство

Обобщим понятие скалярного произведения геометрических свободных векторов и векторов из $\boldsymbol{R}^n$. Линейное пространство E называется ***евклидовым***, если каждой упорядоченной паре (x, y) векторов из E поставлено в соответствие некоторое действительное число $x \cdot y$, называемое ***скалярным произведением*** вектора x на вектор y и удовлетворяющее следующим свойствам:

1°. $x \cdot y = y \cdot x$ (коммутативность);

2°. $(\lambda x) \cdot y = \lambda(x \cdot y)$ (ассоциативность относительно умножения на число);

3°. $(x + y) \cdot z = x \cdot z + y \cdot z$ (дистрибутивность относительно сложения векторов);

4°. Если $x \neq o$, то $x \cdot x > 0$.

Свойства скалярного произведения позволяют обращаться с ним по обычным арифметическим правилам: раскрывать скобки и выносить за скобки общий множитель, а также выносить число за знак скалярного произведения. Это означает, в частности, что справедливы стандартные формулы приведения. Например,

$$(a + b) \cdot (a + b) = a \cdot a + 2a \cdot b + b \cdot b,$$
$$a \cdot a - b \cdot b = (a - b) \cdot (a + b).$$

Кроме того, нетрудно убедиться, что $o \cdot x = 0$ для любого вектора x. Обычное скалярное произведение удовлетворяет данным аксиомам (см. подпараграф 4.1.3).

Скалярное произведение порождает норму в пространстве E:

$$\|x\| = \sqrt{x \cdot x}.$$

Нетрудно проверить, что определенная таким образом норма удовлетворяет всем трем аксиомам нормы. Длина обычного геометрического вектора $\bar{x}$ как раз и равна квадратному корню из скалярного квадрата $\bar{x} \cdot \bar{x}$. Можно сделать вывод, что норма, определенная через скалярное произведение, есть обобщение геометрической длины.

Итак, мы обобщили, по существу, все геометрические понятия. Операциям над свободными векторами мы поставили в соответствие абстрактные операции умножения на число и сложения элементов (векторов) линейного пространства. Коллинеарности и компланарности векторов соответствует линейная зависимость конечного набора векторов. Кроме того, мы обобщили геометрическую длину вектора и расстояние между точками, а также скалярное произведение геометрических векторов. В заключение еще раз подчеркнем: путь от конкретных математических объектов к их абстрактным обобщениям весьма характерен в математике. В данном случае от обычного вектора как направленного отрезка мы пришли к вектору как элементу линейного пространства произвольной природы. В качестве промежуточного пункта мы рассмотрели пространство n-мерных векторов $\boldsymbol{R}^n$.

Задачи

1. Даны матрицы

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Найти произведения AB и BA.

2. Перемножить матрицы

$$A = \begin{pmatrix} 3 & 1 \\ -1 & 2 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

3. Даны матрицы

$$A = \begin{pmatrix} 2 & 3 \\ -2 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}.$$

Найти матрицы ABC, $(AB - BC)'$.

4. Найти определитель 3-го порядка

$$\begin{vmatrix} 1 & 0 & 2 \\ 3 & 2 & 6 \\ 1 & 2 & 1 \end{vmatrix}:$$

а) по правилу треугольников;
б) разложением по первому столбцу;
в) с помощью элементарных преобразований.

5. Найти определитель 4-го порядка

$$\begin{vmatrix} 1 & 1 & 0 & 2 \\ 2 & 3 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 2 & 3 & 1 \end{vmatrix}.$$

6. С помощью правила Крамера решить линейные системы:

а) $\begin{cases} 2x + 3y = 1, \\ x - y = 0; \end{cases}$

б) $\begin{cases} x + y + z = 1, \\ 2x - y - z = 2, \\ y + 2z = 3. \end{cases}$

7. Какие из следующих однородных систем имеют ненулевое решение?

а) $\begin{cases} x+2y=0, \\ -x+\ \ y=0; \end{cases}$

б) $\begin{cases} 2x-3y=0, \\ -4x+6y=0; \end{cases}$

в) $\begin{cases} 2x-\ \ 3y+\ \ \ \ z=0, \\ x+\ \ \ \ y-\ \ \ \ z=0, \\ -x+1{,}5y-0{,}5z=0. \end{cases}$

8. Убедиться, что следующие векторы линейно независимы:

$$\mathbf{b}_1=\begin{pmatrix}1\\2\\1\\0\end{pmatrix},\quad \mathbf{b}_2=\begin{pmatrix}0\\1\\0\\1\end{pmatrix},\quad \mathbf{b}_3=\begin{pmatrix}1\\0\\1\\0\end{pmatrix},\quad \mathbf{b}_4=\begin{pmatrix}0\\2\\3\\0\end{pmatrix}.$$

9. Найти три известные нормы вектора

$$\mathbf{x}=\begin{pmatrix}1\\-1\\-3\\2\\-7\end{pmatrix}.$$

Глава 6
АЛГЕБРАИЧЕСКИЕ СТРУКТУРЫ*

В предыдущих главах мы ознакомились с различными математическими объектами и действиями (операциями) над ними. Достаточно вспомнить операции сложения и умножения чисел, объединения и пересечения множеств, дизъюнкции и конъюнкции высказываний, сложения векторов, матриц и функций. В ходе изложения отмечались сходные свойства этих операций над объектами различной природы (например, ассоциативные, коммутативные и дистрибутивные законы). В данной главе вводится понятие операции над элементами множества произвольной природы и изучаются свойства таких абстрактных операций. Приводятся содержательные примеры, основанные на знакомых математических понятиях. Вводятся абстрактные алгебраические структуры на множествах с одной операцией (группы и полугруппы) и с двумя операциями (кольца, тела и поля). Основная идея состоит в том, чтобы изучать не свойства конкретных элементов конкретных множеств, а свойства операций над этими элементами. Это позволяет получать общие свойства множеств различной природы.

6.1. Бинарные операции

Вспомним некоторые определения из параграфа 2.4. Функцией (*отображением*) называется правило, которое каждому элементу x множества X (области определения) ставит в соответствие единственный элемент y множества Y (области значений). Декартовым произведением $X\times Y$ множеств X и Y называется множество всевозможных упорядоченных пар (x, y), $x\in X$, $y\in Y$.

Назовем ***бинарной операцией*** на множестве E отображение $E\times E$ в E. Таким образом, каждой паре (a, b) элементов множества E ставится в соответствие единственный элемент c того же множества. Бинарная операция по-другому называется *внутренним законом композиции* на множестве E. Будем обозначать бинарную операцию символом $\top$, а результат действия операции на элементы a, b назовем *композицией* этих элементов:

$$a\top b=c.$$

Множество E в данном случае называется *группоидом* относительно операции $\top$.

Примеры.

1. Операции над числами:
а) сложение $a+b=c$;
б) умножение $a\cdot b=c$;
в) возведение в степень $a^b=c$ на множестве натуральных чисел.
2. Операции над множествами:
а) объединение $A\cup B=C$;
б) пересечение $A\cap B=C$.
3. Логические операции:
а) дизъюнкция $A\vee B=C$;
б) конъюнкция $A\wedge B=C$.

4. Операции над векторами:

а) сложение $\vec{a}+\vec{b}=\vec{c}$;

б) векторное произведение $\vec{a}\times\vec{b}=\vec{c}$.

5. Операции над матрицами:

а) сложение $A + B = C$ (A и B – одинаковой размерности);

б) умножение $AB = C$ (число столбцов матрицы A равно числу строк матрицы B).

6. Операции над функциями:

а) сложение $f + g = h$;

б) композиция функций, заданных на некотором множестве $f \circ g = h$.

Рассмотрим основные свойства, которым могут удовлетворять бинарные операции.

Ассоциативность. Бинарная операция называется *ассоциативной*, если $\forall a, b, c \in E$ выполняется равенство

$$a\top(b\top c) = (a\top b)\top c.$$

Примеры.

1. Ассоциативны, как известно, операции сложения и умножения чисел, объединения и пересечения множеств, дизъюнкции и конъюнкции высказываний, сложения векторов и матриц, умножения матриц, сложения функций.

2. Операция возведения в целую положительную степень – не ассоциативный закон, так как в общем случае

$$a^{(b^c)} \neq (a^b)^c.$$

Например, $2^{(3^4)} = 2^{81}$, а $(2^3)^4 = 8^4 = 2^{12}$.

∇**1.** *Если закон ассоциативен, можно определить композицию конечного набора элементов $a_1\top a_2\top \ldots \top a_n$, результат которой не зависит от расстановки скобок.*

Коммутативность. Бинарная операция называется *коммутативной*, если $\forall a, b \in E$ выполняется равенство

$$a\top b = b\top a.$$

Примеры.

1. Коммутативны операции сложения и умножения чисел, объединения и пересечения множеств, дизъюнкции и конъюнкции высказываний.

2. Операции векторного произведения, композиции функций, произведения матриц не коммутативны.

3. Операция возведения в степень не коммутативна:

$$a^b \neq b^a.$$

∇**2.** *Если операция коммутативна и ассоциативна, то она полностью коммутативна: результат композиции $a_1\top a_2\top \ldots \top a_n$ не зависит от порядка элементов.*

Регулярный элемент. Элемент a называется *регулярным* относительно операции $\top$, если $\forall x, y \in E$ равенства

$$a\top x = a\top y \quad \text{и} \quad x\top a = y\top a$$

влекут $x = y$. В этом случае говорят, что можно *сократить* на a.

Если операция не коммутативна, следует различать левый регулярный элемент $a\top x = a\top y \Rightarrow x = y$ и правый регулярный элемент $x\top a = y\top a \Rightarrow x = y$. Таким образом, a есть регулярный элемент относительно $\top$, если он регулярен как слева, так и справа.

Примеры.

1. Для сложения любое число регулярно: $a + x = a + y \Rightarrow x = y$.

2. Для умножения регулярным является всякое число, кроме нуля, так как $0 \cdot x = 0 \cdot y$ не влечет $x = y$: так, $0 \cdot 2 = 0 \cdot 3 = 0$, но $2 \neq 3$. Поэтому, сокращая равенство $a \cdot x = a \cdot y$ на a, мы должны быть уверены, что $a \neq 0$. В $\boldsymbol{R}^+$ уже всякое число регулярно относительно умножения.

3. Пусть $a^x = a^y$. Тогда $x = y$, если $a \neq 0$ и $a \neq 1$. Напротив, $1^x = 1^y$ не влечет $x = y$. Значит, 0 и 1 не являются регулярными слева относительно возведения в натуральную степень. Точно так же $x^a = y^a$ влечет $x = y$ при всех $a \neq 0$. Следовательно, 0 не регулярен справа относительно возведения в степень.

4. Для множеств $A \cup X = A \cup Y$ не влечет $X = Y$, если $A \neq \varnothing$. Так, $\{1, 2, 3\} \cup \{4\} = \{1, 2, 3\} \cup \{3, 4\} = \{1, 2, 3, 4\}$, но $\{4\} \neq \{3, 4\}$. Значит, только пустое множество регулярно относительно объединения. Точно так же $A \cap X = A \cap Y$ не влечет $X = Y$, если A – не универсальное множество: $\{1, 2\} \cap \{2, 3\} = \{1, 2\} \cap \{2, 4\} = \{2\}$, но $\{2, 3\} \neq \{2, 4\}$. Значит, только универсальное множество регулярно относительно пересечения.

Нейтральный элемент. *Нейтральным* относительно закона $\top$ называют такой элемент $e \in E$, при котором $\forall x \in X$ справедливо

$$e\top x = x\top e = x.$$

Для некоммутативной операции различают левый и правый нейтральные элементы e' и e'': соответственно $e'\top x = x$ и $x\top e'' = x \quad \forall x \in X$.

Нейтральный элемент является одновременно левым и правым нейтральным.

∇**3.** *Если нейтральный элемент существует, то он единственный.*

В самом деле, пусть нашлись два нейтральных элемента e_1 и e_2, т.е. $e_1 \top x = x \top e_1 = x$ и $e_2 \top x = x \top e_2 = x \quad \forall x \in X$. В равенстве $x \top e_1 = x$ возьмем $x = e_2$: получим $e_2 \top e_1 = e_2$. Теперь в равенстве $e_2 \top x = x$ возьмем $x = e_1$: $e_2 \top e_1 = e_1$. Композиция $e_2 \top e_1$ одновременно равна e_1 и e_2. Значит, $e_1 = e_2$.

Из определения нейтрального элемента вытекает следующее наблюдение.

∇**4.** *Если нейтральный элемент существует, то он является регулярным.*

Примеры.

1. Нуль – нейтральный элемент относительно сложения: $0 + x = x + 0 = x$; единица – относительно умножения: $1 \cdot x = x \cdot 1 = x$. Нейтральный элемент относительно произвольной бинарной операции можно называть *единицей*.

2. Пустое множество $\varnothing$ – единица относительно объединения: $\varnothing \cup X = X \cup \varnothing = X$. Универсальное множество нейтрально относительно пересечения: $U \cap X = X \cap U = X$.

3. Нулевой вектор $\vec{0}$ (или $\mathbf{0} \in \boldsymbol{R}^n$) нейтрален относительно сложения векторов.

4. Единичная матрица E нейтральна относительно умножения квадратных матриц.

5. Функция «тождественный нуль» нейтральна относительно сложения числовых функций.

6. Операция возведения в степень не имеет левого нейтрального элемента, поскольку $e^x = x$ невозможно ни при каком x. Напротив, $x^e = x$ при любом x дает $e = 1$ и, значит, 1 есть правый нейтральный элемент. Так как нет левого нейтрального элемента, то операция возведения в степень не имеет нейтрального элемента.

Симметричные элементы. Пусть $\top$ есть бинарная операция на E, обладающая нейтральным элементом e. Говорят, что элемент $\bar{a} \in E$ *симметричен* элементу $a \in E$, если $\bar{a} \top a = e$. В этом случае элемент a называется *симметризуемым*. Симметричный элемент $\bar{a}$ называют также *обратным* или *противоположным*.

Примеры.

1. Если a – действительное число, то $-a$ симметрично ему относительно сложения: $a + (-a) = 0$. Если же $a \neq 0$, то $1/a$ симметрично ему относительно умножения: $a(1/a) = 1$.

2. Вектор $-\vec{a}$ противоположен вектору $\vec{a}$ относительно сложения векторов.

3. Если $\det A \neq 0$, то обратная матрица A^{-1} симметрична матрице A относительно умножения квадратных матриц.

4. Пусть f – взаимно однозначная функция из X на X. Тогда элементом, симметричным f относительно композиции функций, является обратная функция: $f^{-1} \circ f = i_X$.

∇**5.** *Элемент, симметричный нейтральному элементу e, – это он сам, так как $e \top e = e$.*

∇**6.** *Если закон $\top$ коммутативен и a имеет симметричный элемент $\bar{a}$, то обратным к $\bar{a}$ является a:*

$$\bar{\bar{a}} = a.$$

В самом деле, в этом случае $\bar{a} \top a = a \top \bar{a}$. Так как $\bar{a} \top a = e$, то $a \top \bar{a} = e$, т.е. a симметричен $\bar{a}$. При сложении $-(-a) = a$, при умножении $1/(1/a) = a$.

∇**7.** *Пусть закон $\top$ ассоциативен (но не обязательно коммутативен) и регулярный элемент a имеет симметричный элемент $\bar{a}$. Тогда a, в свою очередь, является симметричным для $\bar{a}$: $a \top \bar{a} = e$.*

Действительно, в этом случае $(a \top \bar{a}) \top a = a \top (\bar{a} \top a) = a \top e = a$. Однако $e \top a = a$. Таким образом, $(a \top \bar{a}) \top a = e \top a$. Поскольку a – регулярный, на него можно сократить: $a \top \bar{a} = e$. Это и означает, что a симметричен $\bar{a}$. В этом случае a и $\bar{a}$ называются *взаимно симметричными* (*взаимно обратными, взаимно противоположными*).

∇**8.** *Пусть закон $\top$ ассоциативен и регулярный элемент a имеет симметричный элемент $\bar{a}$. Тогда этот элемент единственный.*

Если бы нашелся еще один симметричный элемент a', мы имели бы $a' \top a = e$. Поскольку a регулярен, в силу наблюдения $\nabla 7$ $a \top a' = e$. Поэтому $(\bar{a} \top a) \top a' = \bar{a} \top (a \top a') = \bar{a} \top e = \bar{a}$. С другой стороны, $(\bar{a} \top a) \top a' = e \top a' = a'$. Значит, $a' = \bar{a}$.

∇**9.** *Пусть закон $\top$ ассоциативен и элементы a, b симметризуемы. Тогда композиция $a \top b$ также симметризуема и справедливо равенство*

$$\overline{a \top b} = \bar{b} \top \bar{a}.$$

Действительно, $(\bar{b} \top \bar{a}) \top (a \top b) = \bar{b} \top (\bar{a} \top a) \top b = \bar{b} \top e \top b = \bar{b} \top b = e$.

∇**10.** *Пусть закон* $\top$ *ассоциативен и элемент* b *симметризуем. Тогда уравнение*

$$x \top b = a$$

имеет единственное решение $x = a \top \bar{b}$.

Для сложения, таким образом, определяется разность: $x + b = a \Rightarrow x = a + (-b) = a - b$. Для умножения — частное от деления a на b: $x \cdot b = a \Rightarrow x = a \cdot (1/b) = a/b,\ b \neq 0$.

Дистрибутивность. Пусть на множестве E заданы две бинарные операции: $\top$ и $\perp$. Говорят, что закон $\top$ *дистрибутивен* (слева и справа) относительно закона $\perp$, если $\forall a, b, c \in E$ имеет место

$$a \top (b \perp c) = (a \top b) \perp (a \top c)$$

и

$$(b \perp c) \top a = (b \top a) \perp (c \top a).$$

Примеры.

1. Умножение дистрибутивно относительно сложения: $a \cdot (b + c) = a \cdot b + a \cdot c$. При изучении множеств мы уже обращали внимание, что сложение не дистрибутивно относительно умножения, так как $a + b \cdot c \neq (a + b) \cdot (a + c)$.
2. Операции объединения и пересечения множеств дистрибутивны одна относительно другой (см. параграф 1.6, теорема 1.1).
3. Возведение в степень дистрибутивно справа относительно умножения: $(b \cdot c)^a = (b^a) \cdot (c^a)$.

Изоморфизм. Пусть имеются два различных или совпадающих множества E и F. Пусть на множестве E задана бинарная операция $\top$, а на F — операция $\perp$. Множества E и F называются *изоморфными* относительно законов $\top$ и $\perp$, если существует взаимно однозначная функция f из E на F, такая, что $\forall a, b \in E$ выполняется равенство

$$f(a \top b) = f(a) \perp f(b).$$

Функция f при этом называется *изоморфизмом* E на F.

Пример.

Пусть E — множество $\boldsymbol{R}^+$ положительных действительных чисел, закон $\top$ есть умножение; F — множество $\boldsymbol{R}$ всех действительных

чисел, в котором $\perp$ — сложение. Функция $x \to \ln x$, т.е. $f(x) = \ln x$, есть изоморфизм, так как $\ln(a \cdot b) = \ln a + \ln b$, и, кроме того, эта функция взаимно однозначна: $\ln u = \ln v$ влечет $u = v$.

Алгебраические свойства двух изоморфных множеств можно считать одинаковыми: если закон $\top$ коммутативен на E, то закон $\perp$ также коммутативен на F; если для каждого элемента $x \in E$ существует симметричный элемент относительно закона $\top$, то и для каждого $y \in F$, соответствующего элементу x, существует элемент, симметричный относительно закона $\perp$, и т.д.

Изоморфные множества часто отождествляются. По существу, два изоморфных множества представляют собой два конкретных примера одной и той же алгебраической структуры. Понятие изоморфизма наглядно иллюстрирует основную идею алгебры: достаточно изучать не свойства элементов множеств, а лишь свойства операций, заданных на этих множествах.

∇**11.** *Пусть* f *— изоморфизм* E *на* F *и в* E *существует нейтральный элемент* e *относительно* $\top$. *Тогда* $f(e)$ *есть нейтральный элемент в* F *относительно операции* $\perp$.

∇**12.** *Если* f *— изоморфизм* E *на* F, *то* f^{-1} *— изоморфизм* F *на* E.

Изоморфизм позволяет заменить операцию $a \top b$ во множестве E следующими операциями: образуем элементы $a' = f(a)$ и $b' = f(b)$ множества F, а в F применим к ним операцию $\perp$, т.е. образуем элемент $c' = a' \perp b'$. Наконец, получим $a \top b = f^{-1}(c')$. Это целесообразно в том случае, когда операция $\perp$ в F более проста, чем операция $\top$ в E. Так поступают, заменяя с помощью логарифмов умножение сложением.

Таблицы. Конечный группоид удобно задавать таблицей, похожей на обычную таблицу умножения. Каждому столбцу и каждой строке таблицы соответствует элемент множества E. Первый элемент пары берется из соответствующего столбца, второй — из соответствующей строки, на пересечении получаем композицию этих элементов. Пусть, например, $E = \{a, b, c\}$. Зададим операцию $\top$ таблицей

$\top$	a	b	c
a	c	a	b
b	c	c	b
c	a	b	c

Пользоваться такой таблицей просто, как и таблицей умножения: например, $a \top a = c$, $a \top b = a$, $a \top c = b$, $b \top a = c$, $b \top b = c$ и т.д. Легко убедиться, что данная операция не является ни ассоциативной, ни коммутативной. Так, $a \top (b \top c) = a \top b = a$, но $(a \top b) \top c = a \top c = b$; $a \top b = a$, но $b \top a = c$.

Аддитивная и мультипликативная запись. Самые распространенные примеры бинарных операций – сложение и умножение чисел. Поэтому знаки этих операций часто используются для записи произвольной бинарной операции. При *аддитивной* форме записи бинарная операция обозначается знаком «+» ($a + b$); нейтральный элемент называется нулем и обозначается 0; симметричный (противоположный) элемент обозначается $-a$. При *мультипликативной* форме операция обозначается знаком «·» ($a \cdot b$); нейтральный элемент называется единицей и обозначается 1; симметричный (обратный) элемент обозначается a^{-1} или $1/a$.

6.2. Полугруппы, группы и подгруппы

Полугруппой относительно операции $\top$ называется группоид относительно $\top$, если операция $\top$ ассоциативна.

Следующая таблица определяет трехэлементную полугруппу:

$\top$	a	b	c
a	b	c	a
b	c	c	b
c	b	b	c

Читатель может сам убедиться, что такая операция $\top$ ассоциативна. Для примера: $a \top (b \top c) = a \top b = c$ и $(a \top b) \top c = c \top c = c$. Точно так же проверяются и все остальные требуемые равенства.

Если, кроме того, закон $\top$ коммутативен, полугруппа называется *абелевой* по имени великого норвежского математика Нильса Абеля. Следующая таблица задает трехэлементную абелеву полугруппу:

$\top$	a	b	c
a	a	c	b
b	c	b	a
c	b	a	c

Видно, например, что $a \top b = b \top a = c$ и т.д. Свойство коммутативности легко определить по таблице: таблица, задающая коммутативную операцию, симметрична относительно главной диагонали (диагонали, идущей из левого верхнего в правый нижний угол).

Примеры полугрупп.

1. Множество положительных чисел – абелева полугруппа относительно сложения.

2. Множество свободных векторов – абелева полугруппа относительно сложения.

3. Совокупность квадратных матриц одного порядка – неабелева полугруппа относительно умножения матриц.

4. Пусть словом называется произвольный конечный упорядоченный набор букв русского алфавита. Определим операцию «сложения» двух слов: результатом этой операции является слово, составленное п о д р я д из букв первого и второго слов. Например, если a = «Вася», b = «студент», то $a + b$ = «Васястудент», а $b + a$ = = «студентВася». Такая операция (ее называют *конкатенацией*), очевидно, ассоциативна и не коммутативна. Таким образом, множество слов образует неабелеву полугруппу относительно конкатенации.

Нильс Хенрих Абель (1802–1829). Норвежский математик. В Королевском парке в Осло стоит памятник Нильсу Абелю – скульптура сказочного юноши, попирающего двух поверженных чудовищ. Первое из чудовищ символизирует алгебраические уравнения пятой степени. Еще в последних классах школы Абелю показалось, что он нашел формулу для их решения. Никто в провинциальной Норвегии не смог тогда проверить доказательство. Абель сам нашел у себя ошибку, он уже знал, что не существует выражения для корней в радикалах. Абель доказал это, пользуясь абстрактными алгебраическими понятиями, которые впоследствии получили его имя.

К тому времени Абель был уже студентом университета в Осло (тогда Кристиании). Он был очень беден, и вначале стипендию ему выплачивали профессора из собственных средств. Затем он получил государственную стипендию, которая позволила ему провести два года за границей. В Норвегии были люди, которые понимали, сколь одарен Абель, но не было тех, кто бы смог понять его работы.

Во Франции Абель с интересом следил за математическими новостями, пользовался каждой возможностью увидеть П. Лапласа или А. Лежандра, С. Пуассона или О. Коши, но серьезных научных контактов с великими математиками ему установить не удалось. Представленная в академию рукопись об эллиптических функциях (второе поверженное чудовище) не была рассмотрена. Ее обнаружили только через сто лет.

В 1827 г. Абель возвратился на родину и получил лишь временную работу, заменив профессора, уехавшего в длительную экспедицию в Сибирь. Долги стали его уделом, несмотря на то, что трудился Абель с полной отдачей сил. К его работам пришло признание. Французские академики-математики об-

ратились к шведскому королю, правившему Норвегией, с просьбой помочь талантливому ученому. Тем временем у Абеля быстро прогрессировал туберкулез. 6 апреля 1829 г. Абеля не стало.

Группой называется группоид с бинарной операцией $\top$, в котором:

1°) закон $\top$ ассоциативен;

2°) существует нейтральный элемент относительно закона $\top$;

3°) для каждого элемента существует симметричный (обратный).

Если, кроме того, закон $\top$ коммутативен, группа называется *абелевой.*

Множества, обладающие групповыми свойствами, рассматриваются в самых различных разделах математики, а также в ядерной физике, теории относительности, кристаллографии и т.д. Один раз доказанная теорема о свойствах группы может быть применена в самых отдаленных областях для самых разнообразных объектов.

Зададим операцию $\top$ на трехэлементном множестве с помощью следующей таблицы:

$\top$	a	b	c
a	b	c	a
b	c	a	b
c	a	b	c

Можно проверить, что мы определили абелеву группу. Итак:

- Ассоциативность: например, $a\top(b\top c) = a\top b = c$ и $(a\top b)\top c = c\top c = c$ и т.д.
- Элемент c является нейтральным элементом относительно $\top$. Действительно, $c\top a = a\top c = a$, $c\top b = b\top c = b$ и $c\top c = c$. Наличие нейтрального элемента можно определить по таблице: в ней должны присутствовать строка и столбец, в которых все элементы множества расположены в том же порядке, в каком обозначены строки и столбцы. В данном примере это третья строка и третий столбец, соответствующие третьему элементу c.
- Для каждого элемента существует симметричный: $a\top b = b\top a = c$, $c\top c = c$, т.е. $\bar{a} = b$, $\bar{b} = a$, $\bar{c} = c$. В таблице это видно из того, что в каждой строке и каждом столбце единственный раз встречается нейтральный элемент c.
- Закон $\top$ коммутативен, так как таблица симметрична относительно главной диагонали.

Большой вклад в теорию групп внес французский математик Э. Галуа.

Эварист Галуа (1811–1832). Он прожил двадцать лет, всего пять из них занимался математикой. Математические работы, обессмертившие его имя, занимают чуть более 60 страниц.

В 15 лет Галуа открыл для себя математику и с тех пор, по словам одного из преподавателей, «был одержим демоном математики». Юноша отличался страстностью, неукротимым темпераментом, что постоянно приводило его к конфликтам с окружающими, да и с самим собой.

Галуа быстро оказался на уровне современной науки. Ему было 17 лет, когда его учитель Ришар констатировал: «Галуа работает только в высших областях математики». В неполных 18 лет он опубликовал первую работу. В те же годы Галуа два раза подряд не удается сдать экзамены в Политехническую школу, самое престижное учебное заведение того времени. В 1830 г. он был принят в Высшую нормальную школу, готовившую преподавателей. За год учебы в этой школе Галуа написал несколько работ.

Бурные июльские дни 1830 г. застали Галуа в стенах Нормальной школы. Его все больше захватывала новая страсть – политика. Он присоединился к набиравшей силы республиканской партии – Обществу друзей народа. Возник конфликт с директором школы, и в январе 1831 г. Галуа исключили из школы. Тогда же он передал в Парижскую академию наук рукопись своего исследования о решении уравнений в радикалах. Однако академия отвергла работу Галуа – слишком новы были изложенные в ней идеи. В это время Галуа находился в тюрьме. После освобождения уже в июле он вновь оказался в тюрьме Сент-Пелажи после попытки организовать манифестацию 14 июля (в годовщину взятия Бастилии). На этот раз он приговорен к 9 месяцам тюрьмы. За месяц до окончания срока заключения заболевшего Галуа перевели в больницу. В тюрьме он встретил свое двадцатилетие.

Галуа вышел на свободу в апреле, но ему было суждено прожить только один месяц. В конце мая он был смертельно ранен на дуэли. Работы Галуа посвящены проблеме разрешимости алгебраических уравнений. Так называемая теория Галуа составляет сегодня одну из самых глубоких глав алгебры. Работы Галуа были опубликованы лишь в 1846 г., а признание к ним пришло еще позже, в 70-х гг. XIX в., когда понятие группы постепенно становилось одним из основных математических объектов.

Примеры групп.

1. Множество целых чисел $\boldsymbol{Z}$ образует абелеву группу относительно сложения. Действительно, сложение ассоциативно; имеется нейтральный элемент 0; каждый элемент a имеет симметричный (противоположный) элемент $-a$; кроме того, сложение коммутативно.

Напротив, относительно умножения множество Z не является группой. Хотя умножение ассоциативно и нейтральный элемент существует (1), но для произвольного целого числа обратное число не является целым, а 0 вообще не имеет обратного элемента.

2. Множество Q^+ положительных рациональных и множество R^+ положительных действительных чисел – примеры абелевых групп относительно умножения (с нейтральным элементом 1).

3. Свободные геометрические векторы и векторы из R^n образуют абелеву группу относительно сложения векторов.

4. Совокупность квадратных матриц одного порядка с ненулевыми определителями образует неабелеву группу относительно умножения матриц.

5. Легко заметить, что первые четыре аксиомы линейного пространства (см. подпараграф 5.4.1) соответствуют определению абелевой группы относительно сложения (ассоциативность и коммутативность, существование нейтрального элемента o и противоположного $-x$). Итак, произвольное линейное пространство есть абелева группа относительно сложения векторов.

6. Пусть m – натуральное число, отличное от нуля. Зададим на множестве целых чисел Z отношение сравнимости по модулю m: $x\rho y$, если $x - y$ делится на m. Это отношение обозначается $x \equiv y$ (mod m) и является отношением эквивалентности. Все множество целых чисел разбивается на m классов эквивалентности таким образом, что числа из одного класса при делении на m дают одинаковый остаток: 0, 1, ..., $m - 1$. Эти классы принято обозначать ((0)), ((1)), ..., ((m – 1)). Множество классов эквивалентности состоит из m множеств и обозначается Z_m. Такое множество называется фактор-множеством множества Z относительно ρ. Для отношения сравнимости по модулю m фактор-множество называется множеством вычетов по модулю m.

Рассмотрим множество Z_2 вычетов по модулю 2, состоящее из двух классов: ((0)) – целые числа, делящиеся на 2 без остатка (т.е. четные), и ((1)) – дающие при делении на 2 остаток 1 (нечетные). Легко определить на этом двухэлементном множестве бинарную операцию «сложение по модулю 2», обозначается ⊕. Поскольку сумма двух четных и двух нечетных чисел дает четное число, а сумма четного и нечетного нечетна, данная операция задается следующей таблицей:

⊕	((0))	((1))
((0))	((0))	((1))
((1))	((1))	((0))

Легко убедиться, что множество вычетов Z_2 является абелевой группой. Ассоциативность проверяется непосредственно. Коммутативность операции ⊕ видна из того, что таблица симметрична относительно главной диагонали. Нейтральным элементом является

класс четных чисел ((0)), так как ((0)) ⊕ ((0)) = ((0)) и ((0)) ⊕ ((1)) = = ((1)) ⊕ ((0)) = ((1)). Каждый элемент является обратным самому себе: ((0)) ⊕ ((0)) = ((0)) и ((1)) ⊕ ((1)) = ((0)).

Точно так же можно определить операцию сложения по модулю 3 на множестве Z_3 вычетов по модулю 3:

⊕	((0))	((1))	((2))
((0))	((0))	((1))	((2))
((1))	((1))	((2))	((0))
((2))	((2))	((0))	((1))

Множество Z_3 – абелева группа относительно ⊕. Ассоциативность и коммутативность очевидны. Нейтральным элементом снова является класс ((0)) чисел, кратных 3. Из таблицы видно, что каждый элемент имеет симметричный: ((0)) симметричен сам себе, а ((1)) и ((2)) взаимно симметричны.

7. Представим себе такую игру. Трое участников сидят на стульях A, B и C. По команде ведущего они встают со своих мест, танцуют под музыку и, как только музыка прекращается, должны снова сесть на стулья (не обязательно на те же самые). Через несколько конов ведущий незаметно помечает какой-нибудь стул. Тот, кто занял отмеченный стул, проигрывает.

Итак, если вначале 1, 2 и 3-й участники занимали соответственно стулья A, B, C, то после первого кона они могут занять, например, стулья B, A, C, после второго – C, B, A и т.д. Каждый кон, таким образом, можно описать как п е р е с т а н о в к у трехэлементного множества $\{A, B, C\}$. Под *перестановкой* конечного множества X понимается взаимно однозначная функция, отображающая X на X. Нетрудно убедиться, что всего существует шесть различных перестановок трехэлементного множества. Зададим их в виде таблиц, из которых ясно, в какие элементы переходят «стулья» A, B и C:

A	B	C
A	B	C

e

A	B	C
B	A	C

p

A	B	C
A	C	B

q

A	B	C
C	B	A

r

A	B	C
B	C	A

s

A	B	C
C	A	B

t

Итак, последовательность конов нашей игры – это последовательные перестановки трехэлементного множества. Если вначале участники сидели на стульях A, B, C, а затем на стульях B, A, C, значит, произошла перестановка p. Если теперь применить перестановку r (B перейдет в B, C в A и A в C), то участники окажутся соответственно на стульях B, C, A. Заметим, что последовательное применение перестановок p и r дает тот же результат, что и при одной перестановке s. Если 2 раза подряд применить перестановку p, участники окажутся на тех же стульях, что и в начале игры: таким образом, дважды примененная перестановка p равносильна тождественной перестановке e. Легко определить композицию двух перестановок (*) как результат их последовательного выполнения и составить таблицу умножения для этой бинарной операции:

*	e	p	q	r	s	t
e	e	p	q	r	s	t
p	p	e	t	s	r	t
q	q	s	e	t	p	r
r	r	t	s	e	q	p
s	s	q	r	p	t	e
t	t	r	p	q	e	s

Можно проверить, что множество перестановок трехэлементного множества с таким законом композиции является группой. Ассоциативность очевидна из определения операции и легко проверяется по таблице умножения. Тождественная перестановка e является нейтральным элементом. Перестановки e, p, q, r симметричны самим себе, а перестановки s и t – взаимно симметричны. Итак, все свойства группы выполняются. Заметим, что построенная группа перестановок неабелева: например, $p * q = t$, а $q * p = s$.

8. С той же группой перестановок связана группа геометрических преобразований равностороннего треугольника, совмещающих его с самим собой. Таких преобразований шесть. Во-первых, это тождественное преобразование, которое оставляет треугольник на месте: вершины A, B, C переходят в A, B, C (рис. 6.1).

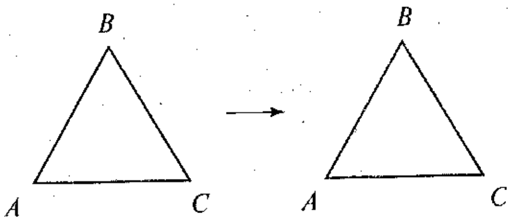


Рис. 6.1

Во-вторых, это симметричное отражение треугольника относительно трех высот. При отражении относительно высоты, опущенной из вершины C, вершины A, B меняются местами, т.е. A, B, C переходят в B, A, C (рис. 6.2, *а*); симметрия относительно высоты, опущенной из A, переводит A, B, C в A, C, B (рис. 6.2, *б*); наконец, последняя симметрия переводит A, B, C в C, B, A (рис. 6.2, *в*).

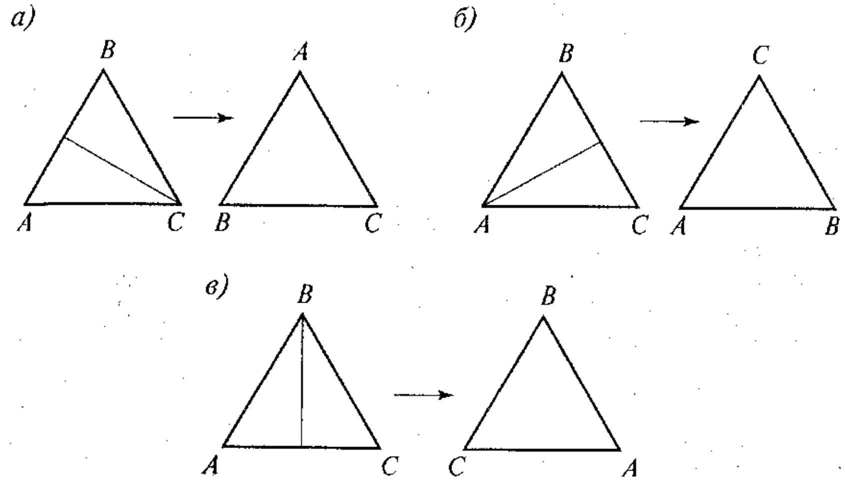


Рис. 6.2

Остаются еще два совмещающих преобразования: поворот против часовой стрелки на угол $2\pi/3$ относительно центра треугольника, т.е. A, B, C переходят в B, C, A (рис. 6.3, *а*), и поворот на угол $4\pi/3$, т.е. A, B, C переходят в C, A, B (рис. 6.3, *б*).

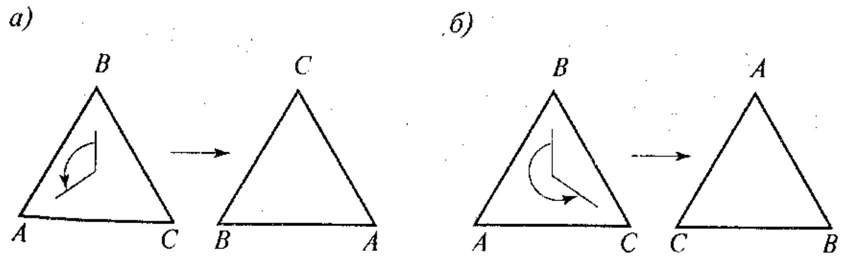


Рис. 6.3

Приглядевшись внимательно, мы замечаем, что шесть преобразований, совмещающих равносторонний треугольник с самим собой, описываются теми же перестановками e, p, q, r, s, t трехэлементого множества, что и игра из примера 7. Композиция $*$ двух преобразований есть результат их последовательного применения. Например, если сначала симметрично отразить треугольник относительно высоты, опущенной из вершины C (что соответствует перестановке p), а затем повернуть треугольник на угол $2\pi/3$ (перестановка s), вершины будут расположены в порядке C, B, A. Тот же результат получится при симметрии относительно высоты, опущенной из вершины B (перестановка r). Таким образом, $p * s = r$, что полностью соответствует таблице для композиции перестановок. Итак, мы приходим к выводу, что множество совмещающих преобразований равностороннего треугольника является неабелевой группой относительно композиции преобразований.

Группа перестановок n-элементного множества называется *симметрической* и обозначается S_n. Можно проверить, что данная группа содержит $1 \cdot 2 \cdots n = n!$ перестановок (читается: n-факториал). Группе S_4 соответствует группа из 24 геометрических преобразований правильного тетраэдра, совмещающих его с самим собой. Подобные геометрические аналогии обусловили активное применение теории групп для исследования формы правильных кристаллов в кристаллографии.

Подмножество группы, которое само является группой относительно того же закона, называется ***подгруппой***.

Примеры подгрупп.

1. Множество $\boldsymbol{Q}$ рациональных чисел — подгруппа группы $\boldsymbol{R}$ действительных чисел относительно сложения; множество целых чисел $\boldsymbol{Z}$ — подгруппа $\boldsymbol{Q}$; множество четных чисел — подгруппа группы $\boldsymbol{Z}$.

2. Подмножества $\{e, s, t\}$, $\{e, p\}$, $\{e, q\}$, $\{e, r\}$ группы S_3 перестановок трехэлементного множества являются подгруппами относительно композиции перестановок $*$ (см. примеры групп 7 и 8):

*	e	s	t
e	e	s	t
s	s	t	e
t	t	e	s

*	e	p
e	e	p
p	p	e

*	e	q
e	e	q
q	q	e

*	e	r
e	e	r
r	r	e

На языке группы преобразований, совмещающих равносторонний треугольник с самим собой, первая подгруппа состоит из тождественного преобразования и двух поворотов; остальные три — из тождественного преобразования и соответствующей симметрии.

∇**1.** *Подгруппа всегда содержит нейтральный элемент группы.*

6.3. Кольца. Тела. Поля

Рассмотрим некоторые алгебраические структуры, представляющие собой множества с двумя бинарными операциями. Допустим, имеется множество A, на котором заданы две операции. Обозначим их через «+» и «·» и будем называть первую операцию сложением, а вторую — умножением. Пусть A есть абелева группа относительно первого закона, а второй закон ассоциативен и дистрибутивен относительно первого. В этом случае множество A называется ***кольцом***. Итак, перечислим свойства кольца.

Сложение (A — абелева группа, т.е. $\forall x, y, z \in A$):

1°. $(x + y) + z = x + (y + z)$ (ассоциативность).

2°. $x + 0 = 0 + x = x$ (существование нейтрального элемента).

3°. $x + (-x) = 0$ (существование симметричного элемента).

4°. $x + y = y + x$ (коммутативность).

Умножение:

1°. $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ (ассоциативность).

2°. $(x + y) \cdot z = x \cdot z + y \cdot z$ (дистрибутивность справа).

3°. $z \cdot (x + y) = z \cdot x + z \cdot y$ (дистрибутивность слева).

Другими словами, множество A есть кольцо, если:

1) A есть абелева группа относительно сложения;

2) A есть полугруппа относительно умножения;

3) умножение дистрибутивно относительно сложения.

Если, кроме того, умножение коммутативно, т.е. $x \cdot y = y \cdot x$, кольцо называют *абелевым* (или *коммутативным*).

Кольцо A не обязано содержать нейтральный элемент (единицу) второй операции. Если кольцо содержит единицу относительно умножения, его называют *кольцом с единицей*.

Примеры колец.

1. Множество целых чисел с законами сложения и умножения образует абелево кольцо с единицей.

2. Множество четных чисел с законами сложения и умножения образует абелево кольцо без единицы.

3. В главе 5 (см. подпараграф 5.4.1) определялись многочлены степени n как функции вида $a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$. Наряду со сложением многочленов, естественным образом определяется операция умножения. Умножение многочленов осуществляется путем раскрытия скобок и приведения подобных членов. Например, $(x^2 + 2x)(x + 1) = x^3 + 3x^2 + 2x$. Нетрудно убедиться, что совокупность многочленов образует абелево кольцо с единицей (от-

носительно сложения и умножения). В самом деле, ассоциативность и коммутативность обеих операций очевидна, также очевидна дистрибутивность умножения относительно сложения. Нейтральным элементом относительно сложения (нулем) является многочлен $0 \cdot x^0$. Элемент, противоположный многочлену $a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$, есть многочлен $(-a_n)x^n + (-a_{n-1})x^{n-1} + \ldots + (-a_1)x + (-a_0)$. Нейтральный элемент относительно умножения (единица) есть многочлен $1 \cdot x^0$.

В кольце, разумеется, элемент, симметричный к симметричному относительно сложения, есть сам этот элемент: $-(-a) = a$.

В кольце можно определить вычитание:

$$a - b = a + (-b).$$

∇**1.** *Умножение дистрибутивно относительно вычитания:*

$$a \cdot (b - c) = a \cdot b - a \cdot c,$$
$$(b - c) \cdot a = b \cdot a - c \cdot a.$$

∇**2.** *Справедливы равенства*

$$(-a) \cdot (-b) = a \cdot b; \quad (-a) \cdot b = a \cdot (-b) = -(a \cdot b).$$

Телом называют кольцо с единицей, в котором каждый отличный от нуля элемент a обладает симметричным (обратным) относительно умножения a^{-1}. По-другому тело определяется так: кольцо называется телом, если множество отличных от нуля элементов этого кольца образует группу относительно умножения. Таким образом, в теле заданы две группы: относительно сложения (аддитивная) и относительно умножения (мультипликативная). Если мультипликативная группа в теле является абелевой (умножение коммутативно), тело называется ***полем***.

Примеры полей.

1. Множество рациональных чисел $\boldsymbol{Q}$ и множество действительных чисел $\boldsymbol{R}$.

2. Трехэлементное множество $\{a, b, c\}$, в котором операции сложения и умножения задаются следующими таблицами:

+	a	b	c
a	b	c	a
b	c	a	b
c	a	b	c

$\cdot$	a	b	c
a	a	b	c
b	b	a	c
c	c	c	c

Это пример конечного поля. Сложение здесь определено так же, как операция $\top$ в примере абелевой группы (см. параграф 6.2). Нейтральный элемент этой группы есть c, он противоположен сам

себе, а элементы a и b взаимно противоположны. Из второй таблицы видно, что a есть единица относительно умножения, элементы a и b обратны сами себе. Ассоциативность, коммутативность и дистрибутивность умножения относительно сложения читатели могут проверить самостоятельно непосредственно по таблицам. Итак, множество $\{a, b, c\}$ с заданными таким образом операциями сложения и умножения является конечным полем.

В поле определено деление на любой элемент $b \neq 0$: $\dfrac{a}{b} = a/b = a \cdot b^{-1} = b^{-1} \cdot a$.

∇**3.** *Пусть K – поле, $a, b \in K$, $a \neq 0$, $b \neq 0$. Тогда элементы*

$$a + b, \quad a - b, \quad a \cdot b, \quad a/b, \quad b/a$$

также принадлежат полю K.

Итак, вместе с любыми двумя элементами a и b поле содержит их сумму, разность, произведение и частное (если, конечно, делитель отличен от нуля). Арифметические операции в поле выполняются без ограничений, в нем можно «перемещаться» без препон, как по ровной местности. Отсюда, вероятно, поле и получило свое название.

∇**4.** *Если в поле $a \cdot b = 0$ и $a \neq 0$, то $b = 0$.*

В самом деле, $a^{-1} \cdot (a \cdot b) = (a^{-1} \cdot a) \cdot b = 1 \cdot b = b = 0$.

∇**5.** *В поле справедливы равенства*

$$\frac{a}{b} \pm \frac{c}{d} = \frac{a \cdot d \pm b \cdot c}{b \cdot d},$$
$$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d},$$
$$\frac{a}{b} \Big/ \frac{c}{d} = \frac{a \cdot d}{b \cdot c}.$$

Итак, в поле выполняются обычные арифметические правила сложения, вычитания, умножения и деления дробей. В произвольном поле справедливы также известные из школьного курса математики формулы приведения. Если обозначить $a + a$ как $2a$, сумму из n одинаковых слагаемых a как na, $a \cdot a$ как a^2, а произведение n одинаковых сомножителей a как a^n, окажутся справедливыми равенства

$$(a + b)^2 = a^2 + 2ab + b^2,$$
$$(a + b)(a - b) = a^2 - b^2$$

и т.д.

Понятие поля возникает в алгебре при исследовании корней многочлена (или, что то же самое, при разложении многочлена на множители). Решение зависит от того, какими числами разрешено пользоваться или, другими словами, над каким полем задан многочлен. Если учитель попросит разложить на множители многочлен $x^2 - 1$, то ученик 6-го класса легко справится с этой задачей: корни многочлена $x = \pm 1$ и $x^2 - 1 = (x - 1)(x + 1)$. О многочлене $x^2 - 3$ этот ученик скажет, что он на множители неразложим. Ученик же 7-го класса найдет корни $x = \pm\sqrt{3}$ и разложение $x^2 - 3 = (x - \sqrt{3})(x + \sqrt{3})$. Все дело в том, что в первом случае пользовались только рациональными числами (многочлен был задан над полем $\boldsymbol{Q}$), а во втором – всеми действительными числами, включая иррациональные (многочлен задан над полем $\boldsymbol{R}$). Для разложения многочлена $x^2 + 4$ уже недостаточно действительных чисел. Чтобы найти корни этого многочлена, нужно познакомиться с полем комплексных чисел.

Поле комплексных чисел. Чтобы определить комплексное число, введем специальное число i, квадрат которого равен -1. Это число называется *мнимой единицей*. Определим теперь множество $\boldsymbol{C}$ комплексных чисел вида $a + bi$, где $a, b \in \boldsymbol{R}$. При $b = 0$ получается обычное действительное число. Поэтому $\boldsymbol{R} \subset \boldsymbol{C}$. Два комплексных числа $a + bi$ и $c + di$ равны тогда и только тогда, когда $a = c$ и $b = d$. Сложение и умножение комплексных чисел выполняются по обычным правилам раскрытия скобок и приведения подобных членов. Пусть $\alpha = a + bi$, $\beta = c + di$. Тогда сложение задается следующим равенством:

$$\alpha + \beta = (a + c) + (b + d)i.$$

Легко убедиться, что такой закон сложения ассоциативен и коммутативен. Нейтральным элементом является число $0 + 0 \cdot i = 0$. Каждое число $\alpha = a + bi$ имеет симметричное (противоположное) число $-\alpha = (-a) + (-b)i$. Таким образом, множество $\boldsymbol{C}$ есть абелева группа относительно сложения. Легко определить и вычитание комплексных чисел:

$$\alpha - \beta = (a - c) + (b - d)i.$$

Умножение комплексных чисел определяется так:

$$\alpha \cdot \beta = (ac - bd) + (ad + bc)i.$$

Мы придем к этому определению, если аккуратно раскроем скобки и приведем подобные члены: $(a + bi)(c + di) = ac + adi + bci + bdi^2 = ac + (ad + bc)i + bd \cdot (-1) = (ac - bd) + (ad + bc)i$.

Проверим коммутативность умножения: по определению $\beta \cdot \alpha = (ca - db) + (cb + da)i = (ac - bd) + (ad + bc)i = \alpha \cdot \beta$. Мы воспользовались коммутативностью сложения и умножения действительных чисел.

Проверим дистрибутивность умножения относительно сложения. Пусть $\alpha_1 = a_1 + b_1 i$, $\alpha_2 = a_2 + b_2 i$, $\alpha_3 = a_3 + b_3 i$. Тогда $\alpha_1(\alpha_2 + \alpha_3) = \alpha_1((a_2 + a_3) + (b_2 + b_3)i) = (a_1(a_2 + a_3) - b_1(b_2 + b_3)) + (a_1(b_2 + b_3) + b_1(a_2 + a_3))i = ((a_1a_2 - b_1b_2) + (a_1b_2 + b_1a_2)i) + ((a_1a_3 - b_1b_3) + (a_1b_3 + b_1a_3)i) = \alpha_1\alpha_2 + \alpha_1\alpha_3$. Дистрибутивность доказана.

Ассоциативность проверяется непосредственно по определению путем аккуратных рутинных преобразований. Читатель может выполнить их самостоятельно.

Нейтральным элементом относительно умножения является число $1 + 0 \cdot i = 1$, совпадающее с обычной единицей. В самом деле, $(a + bi)(1 + 0 \cdot i) = (a \cdot 1 - b \cdot 0) + (a \cdot 0 + b \cdot 1)i = a + bi$ для любого комплексного числа $a + bi$. Попытаемся найти комплексное число, симметричное (обратное) данному числу $a + bi \neq 0$ относительно умножения. Обозначим его $x + yi$. Должно выполняться равенство $(a + bi)(x + yi) = 1 + 0 \cdot i$, что приводит к системе уравнений

$$\begin{cases} ax - by = 1, \\ bx + ay = 0. \end{cases}$$

Легко проверить, что система имеет единственное решение

$$x = \frac{a}{a^2 + b^2}, \quad y = \frac{-b}{a^2 + b^2},$$

так как $a^2 + b^2 \neq 0$ в силу того, что $a + bi \neq 0$. Итак, каждый ненулевой элемент имеет обратный относительно умножения:

$$(a + bi)^{-1} = \frac{a}{a^2 + b^2} + \frac{-b}{a^2 + b^2}i.$$

Таким образом, выполняются все условия поля. Множество $\boldsymbol{C}$ есть абелева группа относительно сложения, и все ненулевые элементы в нем образуют абелеву группу относительно умножения. Значит, множество $\boldsymbol{C}$ комплексных чисел является полем.

Теперь мы легко найдем корни многочлена $x^2 + 4$. Зная, что квадратный корень из -1 равен мнимой единице i, легко определить квадратный корень из любого отрицательного числа $a < 0$: $\sqrt{a} = \sqrt{|a|}\,i$. Значит, многочлен имеет корни $x = \pm 2i$ и $x^2 + 4 = (x - 2i)(x + 2i)$.

В общем случае многочлен степени n, заданный над полем комплексных чисел, имеет ровно n корней, если каждый корень считать столько раз, какова его кратность.

Итак, мы закончили рассмотрение абстрактных алгебраических структур. Благодаря понятию бинарной операции мы смогли исследовать свойства множеств самой разнообразной природы с единой точки зрения, а именно с точки зрения алгебраических свойств операций. Мы убедились, что совокупности различных математических объектов обладают похожими или одинаковыми алгебраическими свойствами. С самого начала курса математики мы следили за этими свойствами, отмечая, где только возможно, ассоциативные, коммутативные и дистрибутивные законы, а также их нарушение. На наш взгляд, идея алгебры об исследовании операций над элементами множеств вместо самих элементов проиллюстрирована достаточно убедительно.

Задачи

1. Будут ли следующие множества группой относительно обычного сложения чисел?

а) множество действительных чисел;

б) множество рациональных чисел;

в) множество неотрицательных рациональных чисел;

г) множество натуральных чисел;

д) множество четных целых чисел;

е) множество нечетных целых чисел;

ж) множество простых чисел.

2. Будут ли следующие множества группой относительно обычного умножения чисел?

а) множество действительных чисел;

б) множество положительных действительных чисел;

в) множество действительных чисел без нуля;

г) множество рациональных чисел, больших единицы;

д) множество неотрицательных рациональных чисел;

е) множество натуральных чисел;

ж) множество целых чисел.

3. Дано трехэлементное множество $M = \{a, b, c\}$. Операция на M задается следующей таблицей:

а)

$*$	a	b	c
a	a	a	a
b	b	b	b
c	c	c	c

б)

$*$	a	b	c
a	c	a	a
b	a	a	a
c	a	a	a

в)

$*$	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

Является ли M полугруппой?

4. Пусть $M = \{a, b, c, d\}$. Известно, что операции, определяемые следующими таблицами:

а)

$*$	a	b	c	d
a	b	c	d	b
b	c	d	b	c
c	d	b	c	d
d	b	c	d	b

б)

$*$	a	b	c	d
a	a	b	c	d
b	b	c	d	a
c	c	d	a	b
d	d	a	b	c

ассоциативны. Является ли M группой относительно каждой из этих операций? Является ли операция коммутативной?

5. Написать таблицы для следующих операций:

а) $x * y = y$ на множестве $M = \{0, 1, 2, 3, 4\}$;

б) $x * y = |x - y|$ на множестве $M = \{0, 1, 2, 3, 4\}$.

6. Пусть $A = \{1, 2, 3, 4\}$ и M – множество всех подмножеств множества A. Является ли M группой относительно операции:

а) пересечения множеств;

б) объединения?

7. Исследовать свойства указанных операций на множестве натуральных чисел:

а) $a * b = a - b$;

б) $a * b = |a - b|$;

в) $a * b = a/b$;

г) $a * b = (a - b)^2$;

д) $a * b = a^b$.

8. Являются ли следующие множества кольцом относительно обычного сложения и умножения чисел?

а) множество действительных чисел;

б) множество рациональных чисел;

в) множество целых чисел;

г) множество натуральных чисел;

д) множество чисел вида $a + b\sqrt{5}$, где a и b – любые целые числа;

е) множество четных целых чисел;

ж) множество чисел, кратных 6.

Являются ли данные множества полем?

Часть 3
ОСНОВЫ МАТЕМАТИЧЕСКОГО АНАЛИЗА

Я думаю, вполне логично считать,
что мир бесконечен.

Х.-Л. Борхес

Функция принадлежит к числу базовых математических понятий. В части I мы уже знакомились с функциями, в частности с числовыми. Математический анализ изучает свойства числовых функций более подробно. Основное внимание уделяется *непрерывным функциям*, а также *производной* и *интегралу*. Они определяются на основе понятия *предела*, с которого и начинается изучение математического анализа.

Глава 7
ТЕОРИЯ ПРЕДЕЛОВ

7.1. Предел последовательности

7.1.1. Последовательности

Зададимся детским вопросом (который, кстати, волновал многих мыслителей прошлого и до сих пор не утратил актуальности): «Сколько звезд на небе?» Помните Делового человека из «Маленького принца» Сент-Экзюпери? Он считал делом своей жизни занесги в каталог все эти «маленькие, блестящие, золотые штучки, которые иногда видны в воздухе». Маленький принц прервал его, когда он обнаружил звезду под номером 501 622 731. Наблюдая за его работой, мы можем отметить следующее. Во-первых, процесс пересчета звезд не имеет конца. Вот, кажется, пересчитаны все звезды на небе. Но появляется более мощный телескоп и работа продолжается. Во-вторых, каждая звезда приобретает свой уникальный номер, представляющий собой целое положительное, т.е. натуральное,

число. Итак, с одной стороны, совокупность звезд бесконечна, а с другой стороны, все звезды в ней пронумерованы. Такие совокупности в математике называются последовательностями.

Посмотрим на этот процесс по-другому. При подсчете каждому натуральному числу ставится в соответствие вполне определенная звезда. Это соответствие, как известно, называется функцией. Значит, работа звездочета состоит в построении функции, областью определения которой являются натуральные числа, а областью значений — совокупность звезд. Теперь можно дать строгое определение последовательности: ***последовательностью*** называется функция натурального аргумента. Значения такой функции называются *элементами* последовательности и обозначаются обычно строчными латинскими буквами с индексами: x_1, x_2 и т.д. Индекс показывает, какой номер имеет данный элемент, т.е. какому натуральному числу он соответствует. Произвольному натуральному числу n, таким образом, соответствует элемент x_n с номером n. Обычно при рассмотрении последовательностей не делается различий между функцией и множеством ее значений, поскольку аргументы у всех последовательностей одни и те же: числа натурального ряда 1, 2, 3, ... Это позволяет обозначать последовательности с помощью фигурных скобок: $\{x_1, x_2, ...\} = \{x_n\}$.

Аккуратный звездочет, занося звезду в каталог, как бы отмечает ее крестиком: ведь одна и та же звезда не должна упоминаться дважды. Однако любой астроном может по рассеянности занести в каталог того же самого белого карлика, что и в позапрошлом году, присвоив ему новый номер. Звездный каталог получится неправильным, однако с математической точки зрения он по-прежнему представляет собой последовательность. Ведь определение последовательности не требует, чтобы функция была взаимно однозначной. Итак, элементы последовательности могут повторяться, лишь бы они имели разные номера. Например, если мы изо дня в день будем пересчитывать птиц, пролетающих мимо нашего окна, то неизбежно посчитаем одного и того же воробья много раз. Ничего страшного: ведь каждый раз мы присваиваем ему новый номер.

Давайте вместо звезд подсчитаем, например, четные положительные числа. Мы получим последовательность, элементами которой являются числа: $\{2, 4, 6, ...\}$. Такая последовательность, естественно, называется *числовой*. Другими словами, числовая последовательность есть числовая функция натурального аргумента. В дальнейшем мы будем интересоваться именно числовыми последовательностями.

Последовательность, как и любое бесконечное множество, невозможно задать перечислением всех ее элементов. Естественно встает вопрос: как задать последовательность? Это можно сделать по-разному.

Первый, самый простой способ – в явном виде указать зависимость элемента последовательности от его номера, т.е. задать функцию аналитически. В применении к последовательностям говорят, что задается «формула общего члена» в виде $x_n = f(n)$. Задавая вместо n произвольное натуральное число, мы можем узнать, какой элемент (какое число) находится на данном месте в последовательности. Например, сотый номер имеет число $f(100)$, а миллионный – число $f(1\ 000\ 000)$. Так, с помощью одной формулы функции задается бесконечная совокупность элементов последовательности.

Примеры.

1. $x_n = n$. Данная последовательность представляет собой натуральный ряд: $x_n = \{1, 2, 3, 4, ...\}$. Для краткости будем иногда совмещать обе формы и писать $x_n = n = \{1, 2, 3, 4, ...\}$.

2. $x_n = 2n$. Это уже упоминавшаяся последовательность четных положительных чисел: $x_n = \{2, 4, 6, 8, ...\}$.

3. $x_n = 1/n$. Последовательность состоит из дробей, числитель которых равен 1, а знаменатель – номеру элемента: $x_n = \{1, {}^1/_2, {}^1/_3, {}^1/_4, ...\}$.

4. $x_n = 1$. Последовательность состоит из одних единиц, все ее элементы равны 1: $x_n = \{1, 1, 1, 1, ...\}$. Последовательность, все члены которой равны между собой, называют *стационарной*.

5. Рассмотрим последовательность, состоящую из единиц с чередующимися знаками: $x_n = \{1, -1, 1, -1, ...\}$. На нечетном месте в этой последовательности стоит 1, а на четном –1. Легко убедиться в том, что общий член такой последовательности задается формулой $x_n = (-1)^{n+1}$: при нечетном n число –1 возводится в четную степень и получается 1, а при четном – в нечетную.

Второй способ задания последовательности состоит в том, что ее последующие элементы вычисляются по предыдущим (или предыдущему). Например, мы говорим: «Чтобы получить следующий элемент последовательности, нужно извлечь квадратный корень из предыдущего элемента и к полученному числу прибавить единицу». Математически это можно записать так: $x_{n+1} = \sqrt{x_n} + 1$. Полностью ли определена последовательность? Нет, поскольку не задан ее первый элемент. Достаточно указать, что, например, $x_1 = 1$, и мы сможем найти элемент последовательности с любым номером: $x_2 = \sqrt{1} + 1 = 2$, $x_3 = \sqrt{2} + 1$, $x_4 = \sqrt{\sqrt{2}+1}+1$ и т.д. Такой способ задания последовательности называется ***рекуррентным***. Нам было бы весьма затруднительно задать нашу последовательность с помощью формулы общего члена.

Иногда одну и ту же последовательность можно задать и рекуррентно, и формулой общего члена. Из школьного курса математики вам знакомы две такие последовательности: арифметическая и геометрическая прогрессии. В арифметической прогрессии каждый последующий член равен сумме предыдущего с некоторым заданным числом: $x_{n+1} = x_n + a$, $x_1 = b$. Например, $\{1, 4, 7, 10, ...\}$ при $a = 3$, $b = 1$. Нетрудно убедиться, что общий член арифметической прогрессии задается формулой $x_n = b + (n - 1)a$. В геометрической прогрессии последующий элемент получается из предыдущего умножением на заданное число: $x_{n+1} = qx_n$, $x_1 = b$ или $x_n = bq^{n-1}$. Например, $\{2, 1, {}^1/_2, {}^1/_4, {}^1/_8, ...\}$ при $q = {}^1/_2$, $b = 2$.

7.1.2. Предел

Начнем с геометрического примера. Построим равнобедренный треугольник и впишем в него окружность (рис. 7.1).

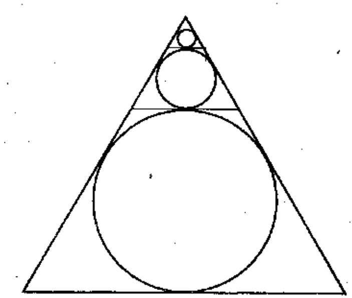


Рис. 7.1

Пусть ее диаметр d_1. Проведем касательную к вписанной окружности параллельно основанию треугольника. В полученный равнобедренный треугольник снова впишем окружность с диаметром d_2. Будем продолжать этот процесс, получая все более маленькие треугольники и все более маленькие вписанные окружности. Ясно, что теоретически процесс можно продолжать до бесконечности. Таким образом, можно построить последовательность $\{d_n\} = \{d_1, d_2, d_3, ...\}$ диаметров окружностей. Однако каким бы тонким карандашом мы ни пользовались, уже через несколько шагов окружности будут неотличимы от точки: диаметр становится все меньше и меньше, приближаясь к нулю. При этом на каждом шаге диаметр вписанной окружности имеет вполне определенное положительное значение. Говорят, что диаметр (последовательность диаметров) стремится к нулю, но никогда не становится равным нулю. Другими словами, элементы последовательности становятся «все ближе и ближе» к нулю. Неформально говоря, число 0 и есть в данном случае предел последовательности d_n.

Чтобы дать строгое определение предела, необходимо формализовать выражение «ближе и ближе». Назовем *окрестностью* точки x_0 любой интервал (a, b), содержащий эту точку: $x_0 \in (a, b)$. Такое определение окрестности вполне согласуется с обычным значением этого слова: в окрестности – значит вблизи. В дальнейшем мы будем использовать окрестности вида $(x_0 - \varepsilon, x_0 + \varepsilon)$, симметричные относительно x_0. Здесь ε – произвольное положительное число. Такую окрестность будем называть *окрестностью с радиусом* ε или просто ε-*окрестностью*. Тот факт, что число x принадлежит окрестности $(x_0 - \varepsilon, x_0 + \varepsilon)$, означает, что выполняются неравенства $x_0 - \varepsilon < x < x_0 + \varepsilon$, которые можно объединить в одно неравенство с помощью модуля:

$$|x - x_0| < \varepsilon.$$

Когда говорят, что элементы последовательности становятся все ближе и ближе к числу x_0, имеют в виду, что, начиная с некоторого номера, все элементы x_n попадают в ε-окрестность точки x_0, сколь бы мало ни было число ε. Это не означает, что каждый следующий элемент обязательно ближе к x_0, чем предыдущий. Просто для каждой окрестности существует такой номер N, что для всех $n > N$ числа x_n не покидают данную окрестность. По сути, мы уже дали строгое определение предела: число x_0 называется ***пределом*** последовательности $\{x_n\}$, если, начиная с некоторого номера, все ее элементы попадают в сколь угодно малую ε-окрестность точки x_0. Другими словами, каково бы ни было число $\varepsilon > 0$, найдется номер N, такой, что при всех $n > N$ выполняется неравенство $|x - x_0| < \varepsilon$ (рис. 7.2). Запишем это определение с помощью математических знаков $\forall$ (для любого) и $\exists$ (существует). Итак, число x_0 называется пределом последовательности $\{x_n\}$, если

$$\forall \varepsilon > 0 \ \exists N = N(\varepsilon) \quad \forall n: n > N \Rightarrow |x_n - x_0| < \varepsilon.$$

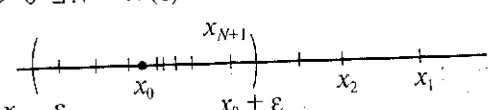


Рис. 7.2

Зависимость номера N от радиуса окрестности отражается в записи $N = N(\varepsilon)$. То, что x_0 – предел $\{x_n\}$, обозначается специальным знаком lim (от латинского limit – предел):

$$x_0 = \lim x_n = \lim_{n\to\infty} x_n.$$

Говорят также, что последовательность $\{x_n\}$ *стремится* или *сходится* к числу x_0 (когда n стремится к бесконечности): $x_n \to x_0$. Сама последовательность в этом случае называется *сходящейся*.

Таким образом, «доказать по определению», что x_0 есть предел последовательности $\{x_n\}$, означает указать способ определения по любому числу $\varepsilon > 0$ номера N, начиная с которого все члены последовательности не выходят за пределы ε-окрестности x_0. Это, в свою очередь, сводится к разрешению неравенства $|x_n - x_0| < \varepsilon$ относительно n.

Примеры.

1. $x_n = 1$: последовательность состоит из одних единиц. Очевидно, что все элементы этой последовательности лежат в окрестности числа 1 с произвольным радиусом. Таким образом, $\lim x_n = 1$. Это справедливо для любой последовательности, состоящей из одинаковых чисел.

∇**1.** *Если $x_n = c$ для всех n, то* $\lim x_n = c$.

2. $x_n = 1/n$. Интуитивно ясно, что члены этой последовательности по мере увеличения n становятся все меньше и меньше, приближаясь к нулю. Докажем по определению, что $\lim x_n = 0$. Зададимся произвольным числом $\varepsilon > 0$. Поскольку $x_n > 0$ при всех n, неравенство $|x_n - x_0| < \varepsilon$ в данном случае имеет очень простой вид: $1/n < \varepsilon$. Отсюда $n > 1/\varepsilon$. Например, при $\varepsilon = 0{,}1$ все члены последовательности, начиная с одиннадцатого, удовлетворяют неравенству $x_n < \varepsilon$, т.е. лежат в ε-окрестности нуля (числа от 1 до $^1/_{10}$ еще не попадают в эту окрестность, а $^1/_{11}$ и все последующие уже меньше 0,1). При $\varepsilon = 0{,}01$ этому условию удовлетворяют все x_n, начиная со сто первого. Итак, для любого $\varepsilon > 0$ нашелся номер $N = 1/\varepsilon$, такой, что при всех $n > N$ выполняется неравенство $|x_n - 0| < \varepsilon$. Этим доказано, что предел последовательности $\{x_n\}$ равен нулю.

3. $x_n = \dfrac{n+1}{n}$. Общий член этой последовательности можно записать по-другому: $x_n = 1 + 1/n$. С увеличением номера члены последовательности приближаются к единице. Неравенство $|x_n - x_0| < \varepsilon$ при $x_0 = 1$ совпадает с неравенством из предыдущего примера: $1/n < \varepsilon$. Значит, определению удовлетворяет тот же номер $N = 1/\varepsilon$. Итак, $\lim x_n = 1$.

Если x_0 есть предел последовательности $\{x_n\}$, все элементы x_{N+1}, x_{N+2} и т.д. попадают в произвольную окрестность точки x_0. Это значит, что за пределами этой окрестности могут оказаться только элементы $x_1, x_2, \ldots, x_N$. При достаточно большом значении ε все элементы последовательности могут лежать в ε-окрестности. Например, все $x_n = 1/n$ удовлетворяют неравенству $x_n < 1{,}5$, т.е. лежат в окрестности нуля, радиус которой равен 1,5. Таким образом, можно сделать следующее наблюдение.

∇2. *Пусть* $\lim x_n = x_0$. *Тогда вне любой окрестности точки* x_0 *находится разве что конечное число элементов последовательности* $\{x_n\}$. (Слова «разве что конечное» означают, что число элементов, лежащих вне окрестности, либо конечно, либо равно нулю.)

Последовательность называется *ограниченной*, если существует такое число $M > 0$, что для всех n выполняется неравенство $|x_n| < M$, т.е. все элементы последовательности лежат в интервале $(-M, M)$. Из свойства ∇2 вытекает еще одно свойство сходящейся последовательности.

∇3. *Если последовательность имеет предел, то она ограничена.*

В самом деле, зададим любую окрестность предела x_0. Ясно, что вне этой окрестности может лежать лишь конечное число элементов последовательности. Тогда найдется такой достаточно широкий интервал $(-M, M)$, который содержит внутри себя и окрестность (с бесконечным числом элементов последовательности), и оставшиеся элементы.

Очень важно понять, в каком случае число x_0 **не является** пределом последовательности $\{x_n\}$. Что значит не является пределом? Это значит, что найдется такая окрестность точки x_0, куда не попадают все элементы последовательности, начиная с некоторого номера. Другими словами, какое бы число N мы ни задали, обязательно найдется элемент x_{n_0} с номером $n_0 > N$, который лежит вне этой окрестности. Математически это можно записать так:

$$\exists \varepsilon_0 > 0 \quad \forall N > 0 \quad \exists n_0 > N\colon |x_{n_0} - x_0| \geqslant \varepsilon_0.$$

Последнее неравенство как раз и означает, что элемент последовательности с номером n_0 оказался вне окрестности с радиусом ε_0. Поскольку в определении утверждается существование конкретной окрестности и конкретного элемента, в обозначениях используется индекс 0. Итак, доказательство того, что x_0 не является пределом $\{x_n\}$, сводится к отысканию окрестности, которую покидает хотя бы один элемент с номером, большим произвольного N: как бы далеко мы ни продвинулись по последовательности, всегда найдется еще более далекий элемент, не лежащий в данной окрестности.

Докажем, например, что последовательность $x_n = (-1)^{n+1} = \{1, -1, 1, -1, \ldots\}$ не имеет предела. Сначала докажем, что единица не является пределом этой последовательности. В самом деле, выберем окрестность точки 1, радиус которой равен 0,5, т.е. интервал (0,5; 1,5). Какой бы номер N мы ни задали, найдется элемент с номером, большим N (а именно -1), не попадающий в эту окрестность. Таким образом, нашлось число $\varepsilon_0 = 0{,}5$, такое, что для любого N найдется элемент $x_{n_0} = -1$, лежащий вне ε_0-окрестности единицы. Значит, по определению единица не является пределом данной последовательности. Точно такие же рассуждения справедливы для любого числа x_0. Итак, если никакое x_0 не является пределом, то последовательность не имеет предела.

Справедливо следующее важное свойство предела последовательности.

∇4. *Если последовательность имеет предел, то он единственный.*

Д о к а з а т е л ь с т в о. Единственность предела вытекает непосредственно из определения. В самом деле, начиная с некоторого номера, все элементы последовательности находятся «вблизи» своего предела. Ясно, что они не могут находиться вблизи двух различных чисел. Докажем это строго. Будем рассуждать от противного. Предположим, что последовательность x_n имеет два предела $a \neq b$. Зададимся числом $\varepsilon = (b - a)/2$ (это половина расстояния между точками a и b). По определению предела существует число N_1, такое, что при $n > N_1$ все элементы последовательности лежат в ε-окрестности числа a, т.е. выполняется неравенство $|x_n - a| < \varepsilon$. Точно так же найдется номер N_2, такой, что при всех $n > N_2$ выполняется неравенство $|x_n - b| < \varepsilon$. Выберем из этих двух номеров максимальный: $N = \max\{N_1, N_2\}$. Тогда при всех $n > N$ должны выполняться оба неравенства. Однако этого не может быть, так как соответствующие окрестности точек a и b не пересекаются (рис. 7.3). Итак, предположив существование двух различных пределов, мы пришли к противоречию. Следовательно, предположение неверно и последовательность имеет единственный предел.

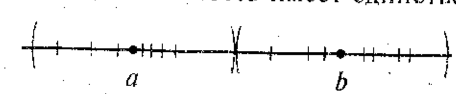


Рис. 7.3

7.1.3. Арифметические свойства предела

Над последовательностями можно осуществлять обычные арифметические операции, если договориться выполнять их почленно. Итак, суммой двух последовательностей называется последовательность, состоящая из сумм соответствующих элементов. Например, если $x_n = n$, $y_n = 1/n$, то $x_n + y_n = n + 1/n = \dfrac{n^2+1}{n} = \left\{2,\ 2\dfrac{1}{2},\ 3\dfrac{1}{3},\ \ldots\right\}$. Точно так же определяются разность $x_n - y_n$, произведение $x_n \cdot y_n$ и отношение x_n/y_n двух последовательностей. В последнем случае требуется оговорка, что $y_n \neq 0 \quad \forall n$, чтобы почленное деление было возможно. Аналогично задается умножение последовательности на число: последовательность αx_n состоит из элементов x_n, умноженных на α. Например, если $x_n = n$, то $2x_n = 2n = \{2, 4, 6, \ldots\}$.

Справедливы следующие арифметические свойства предела.

∇5. *Пусть существует предел последовательностей x_n и y_n:* $\lim x_n = x_0$, $\lim y_n = y_0$. *Тогда*

$$\lim (\alpha x_n) = \alpha x_0,$$
$$\lim (x_n \pm y_n) = x_0 \pm y_0,$$
$$\lim (x_n \cdot y_n) = x_0 \cdot y_0,$$
$$\lim (x_n/y_n) = x_0/y_0 \quad (y_n \neq 0,\ y_0 \neq 0).$$

Приведем доказательство того, что предел суммы равен сумме пределов, как пример типичного доказательства в теории пределов. При этом используется известное свойство модуля, что модуль суммы не превосходит суммы модулей: $|a + b| \leqslant |a| + |b|$. Поскольку множество действительных чисел есть линейное пространство относительно обычных операций сложения и умножения, а модуль числа удовлетворяет определению нормы, данное неравенство представляет собой частный случай неравенства треугольника.

Зададимся произвольным числом ε. Мы хотим доказать, что можно найти номер N, такой, что при $n > N$ выполняется неравенство $|(x_n + y_n) - (x_0 + y_0)| < \varepsilon$. Нам известно, что $x_n \to x_0$. Это означает, что, начиная с некоторого номера, величина $|x_n - x_0|$ становится меньше любого наперед заданного числа. В частности, в качестве такого числа можно взять $\varepsilon/2$. Итак, $\exists N_1$: $n > N_1 \Rightarrow |x_n - x_0| < \varepsilon/2$. Точно так же $\exists N_2$: $n > N_2 \Rightarrow |y_n - y_0| < \varepsilon/2$. Заметим, что по свойству модуля

$$|(x_n + y_n) - (x_0 + y_0)| =$$
$$= |(x_n - x_0) + (y_n - y_0)| \leqslant |x_n - x_0| + |y_n - y_0|.$$

Первое слагаемое меньше $\varepsilon/2$ при $n > N_1$, второе – при $n > N_2$. Выберем максимальный из номеров N_1, N_2 и обозначим его через N: $N = \max\{N_1, N_2\}$. Тогда при всех $n > N$ будут одновременно выполняться оба неравенства. В этом случае

$$|(x_n + y_n) - (x_0 + y_0)| \leqslant |x_n - x_0| + |y_n - y_0| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Итак, для произвольного $\varepsilon > 0$ мы нашли номер N, такой, что при всех $n > N$ выполняется неравенство $|(x_n + y_n) - (x_0 + y_0)| < \varepsilon$. А это и означает, что число $x_0 + y_0$ есть предел последовательности $x_n + y_n$. Свойство доказано. Остальные арифметические свойства предела доказываются по аналогичной схеме.

Обратим особое внимание на два арифметических свойства: константу можно выносить за знак предела и предел суммы равен сумме пределов. Из этих свойств вытекает более общее: предел линейной комбинации последовательностей равен линейной комбинации пределов (если, конечно, все пределы существуют). Такие свойства называются *линейными*. Итак, отыскание предела – линейная операция.

Благодаря арифметическим свойствам предела наши возможности для вычисления различных пределов значительно расширяются. Рассмотрим, например, последовательность $x_n = \dfrac{n+1}{n^2}$. Числитель и знаменатель не имеют пределов, поэтому нельзя применить свойство о пределе отношения. Однако данную последовательность легко преобразовать в сумму двух последовательностей: $x_n = \dfrac{1}{n} + \dfrac{1}{n^2}$. Каждое слагаемое стремится к нулю, следовательно, и сумма стремится к нулю: $x_n \to 0$.

Разберем более сложный пример. Пусть $x_n = \dfrac{2n^2 - 1}{3n + n^2}$. Снова числитель и знаменатель не имеют предела. Однако, если разделить их на n^2, можно устранить это неудобство:

$$x_n = \frac{2 - 1/n^2}{3/n + 1}.$$

Числитель есть разность двух последовательностей: первая состоит из одних двоек и ее предел равен 2, вторая, т.е. $1/n^2$, как легко убедиться, стремится к нулю. Значит, числитель сходится к $2 - 0 = 2$. Похожие рассуждения приводят к выводу, что знаменатель сходится к $0 + 1 = 1$. Стало быть, по арифметическим свойствам отношение имеет предел, равный $2/1 = 2$: $\quad x_n \to 2$.

Вы можете заметить, что два рассмотренных примера похожи: в обоих случаях общий член последовательности представляет собой отношение двух многочленов от n. Понятие многочлена (другое его название – полином) уже упоминалось в главе об алгебраических структурах (см. главу 6). Дробь, числитель и знаменатель которой есть многочлены, называется *рациональной*. Вот еще примеры рациональных дробей:

$$y_n = \frac{n^3+2n}{3n^4+2n^3+1},$$

$$z_n = \frac{2n^2+1}{4n^2+n}$$

и т.д. Чтобы применить к такой последовательности арифметические свойства предела, надо числитель и знаменатель разделить на старшую степень n, т.е. на максимальную из степеней числителя и знаменателя. В приведенных примерах это соответственно 4 и 2. Нетрудно убедиться, что в результате такого деления числитель и знаменатель уже имеют пределы и выполняется следующее правило.

∇**6.** *Если степень числителя меньше степени знаменателя, рациональная дробь стремится к нулю. Если степени числителя и знаменателя совпадают, предел рациональной дроби равен отношению их старших коэффициентов (т.е. коэффициентов при старшей степени n).*

Например, для указанных последовательностей $\lim y_n = 0$, $\lim z_n = 2/4 = 0{,}5$.

Относительно арифметических действий с последовательностями можно сделать следующие замечания.

∇**7.** *Если последовательность x_n имеет предел (сходится), а последовательность y_n расходится, их сумма $x_n + y_n$ расходится. Если обе последовательности x_n и y_n расходятся, то их сумма может быть как сходящейся, так и расходящейся последовательностью.*

Например, $x_n = (-1)^n = \{-1, 1, -1, 1, \ldots\}$, $y_n = (-1)^{n+1} = \{1, -1, 1, -1, \ldots\}$ – расходящиеся последовательности (не имеющие предела). При этом их сумма $x_n + y_n = \{0, 0, 0, \ldots\}$ сходится к нулю.

Пусть дана некоторая последовательность $x_n = \{x_1, x_2, x_3, \ldots\}$. Последовательно выбрав из нее элементы, например, только с четными номерами, мы получим новую последовательность, состоящую из чисел x_2, x_4, x_6 и т.д. Фактически сначала мы организовали последовательность номеров $n_k = 2k$, $k = 1, 2, 3, \ldots$, а затем выбрали из x_n элементы с номерами n_k. Важно, что при этом мы не возвра-

щались назад, т.е. каждый следующий номер был больше предыдущего: $n_1 < n_2 < n_3 < \ldots$. Последовательность элементов x_{n_k}, полученная таким образом, называется *подпоследовательностью* последовательности x_n. Ясно, что существует бесконечно много способов выделения подпоследовательности из данной последовательности (т.е. бесконечно много последовательностей номеров n_k). Например, для $x_n = n$

$$x_{2k} = \{2, 4, 6, \ldots\},$$
$$x_{2k-1} = \{1, 3, 5, \ldots\},$$
$$x_{k^2} = \{1, 4, 9, 16, \ldots\} \quad \text{и т.д.}$$

∇**8.** *Последовательность x_n имеет предел тогда и только тогда, когда все ее подпоследовательности имеют предел, и притом один и тот же.*

Это свойство обычно применяют для доказательства отсутствия предела. Например, последовательность $x_n = (-1)^n = \{-1, 1, -1, 1, \ldots\}$ имеет две стационарные подпоследовательности $x_{2k-1} = -1$ и $x_{2k} = 1$, сходящиеся соответственно к -1 и 1. Следовательно, последовательность x_n не имеет предела.

7.1.4. Бесконечно большие и бесконечно малые последовательности

Будем говорить, что последовательность *стремится к бесконечности*, если, начиная с некоторого номера, ее элементы становятся по модулю больше любого наперед заданного положительного числа. Итак, $x_n \to \infty$, если

$$\forall M > 0\ \exists N: n > N \Rightarrow |x_n| > M.$$

Если при этом все элементы последовательности, начиная с некоторого номера, остаются положительными, говорят, что последовательность сходится к «плюс бесконечности»: $\lim x_n = +\infty$, а если отрицательными – то к «минус бесконечности»: $\lim x_n = -\infty$. Например, последовательность $x_n = n = \{1, 2, 3, \ldots\}$ сходится к $+\infty$, а $x_n = -n = \{-1, -2, -3, \ldots\}$ – к $-\infty$. Последовательность $x_n = (-1)^n n = \{-1, 2, -3, 4, -5, 6, \ldots\}$ сходится к бесконечности, но не к $+\infty$ и не к $-\infty$. В частности, у рациональной дроби и числитель, и знаменатель стремятся к бесконечности (плюс или минус – в зависимости от знака их старшего коэффициента). Говорят, что имеет место неопределенность вида $\frac{\infty}{\infty}$. В подпараграфе 7.1.3 мы научились избавляться от этой неопределенности делением на старшую степень n, когда степень числителя не больше степени знаменателя. Легко убедиться, что если у рациональной дроби степень числителя больше степени знаменателя, то дробь стремится к $+\infty$, если старший коэффициент числителя положителен, и к $-\infty$, если отрицателен. Последовательность, сходящаяся к бесконечности, к $+\infty$ или к $-\infty$, называется ***бесконечно большой***. В свою очередь, последовательность, сходящаяся к нулю, называется ***бесконечно малой***.

∇**9.** *Если последовательность x_n ограничена, а последовательность y_n бесконечно большая, то их отношение x_n/y_n есть бесконечно малая последовательность.*

Например, $x_n = \dfrac{2}{n^2+1} \to 0$: числитель 2 ограничен, а знаменатель $n^2 + 1$ стремится к $+\infty$.

∇**10.** *Если найдется такое $a > 0$, что $x_n > a$ при всех n, а y_n — бесконечно малая последовательность, то их отношение x_n/y_n есть бесконечно большая последовательность.*

Условие отделенности от нуля ($x_n > a$) крайне важно. Пусть, например, числитель x_n также стремится к нулю и не существует такого числа $a > 0$, что $x_n > a$ при всех n. Тогда об отношении x_n/y_n нельзя сказать заранее ничего определенного: оно может иметь предел, стремиться к бесконечности или не иметь никакого предела. Говорят, что в этом случае имеет место неопределенность вида $\dfrac{0}{0}$.

Примеры.

1. $x_n = y_n = 1/n$. Тогда $x_n/y_n = 1 \to 1$.
2. $x_n = 1/n$, $y_n = 1/n^2$. Тогда $x_n/y_n = n \to +\infty$.
3. $x_n = (-1)^n/n = \{-1, {}^1/_2, -{}^1/_3, {}^1/_4, \ldots\}$, $y_n = 1/n = \{1, {}^1/_2, {}^1/_3, {}^1/_4, \ldots\}$.

Обе эти последовательности стремятся к нулю, а их отношение $x_n/y_n = \{-1, 1, -1, 1, \ldots\}$ не имеет предела.

Неопределенность другого вида ($\infty - \infty$) возникает при вычитании бесконечно большой последовательности из бесконечно большой.

Примеры.

1. $x_n = n + 1$, $y_n = n$: $x_n - y_n = 1 \to 1$.
2. $x_n = n^2$, $y_n = n$: $x_n - y_n \to +\infty$.

Ознакомимся с одним практическим приемом раскрытия неопределенности вида $\infty - \infty$. Рассмотрим последовательность $x_n = \sqrt{n+1} - \sqrt{n}$, представляющую собой разность бесконечно больших последовательностей. Домножим и разделим эту разность на сумму тех же величин (сопряженное выражение):

$$x_n = \frac{\left(\sqrt{n+1} - \sqrt{n}\right)\left(\sqrt{n+1} + \sqrt{n}\right)}{\sqrt{n+1} + \sqrt{n}}.$$

Воспользуемся известной «школьной» формулой для разности квадратов $(a - b)(a + b) = a^2 - b^2$ при $a = \sqrt{n+1}$, $b = \sqrt{n}$. Тогда $\left(\sqrt{n+1} - \sqrt{n}\right)\left(\sqrt{n+1} + \sqrt{n}\right) = (n + 1) - n = 1$ и

$$x_n = \frac{1}{\sqrt{n+1} + \sqrt{n}}.$$

Числитель этой дроби ограничен, а знаменатель стремится к бесконечности. Значит, дробь стремится к нулю: $\lim x_n = 0$. Домножение на сопряженное выражение — распространенный прием вычисления пределов с иррациональными числами (корнями).

7.1.5. Монотонные ограниченные последовательности

Ранее мы договорились, что последовательность есть функция, определенная на множестве натуральных чисел (см. подпараграф 7.1.1). Значит, для последовательности можно ввести известные свойства числовых функций, в частности монотонность. Итак, последовательность x_n называется:

возрастающей, если $x_{n+1} > x_n$ при всех n;
неубывающей, если $x_{n+1} \geqslant x_n$ при всех n;
убывающей, если $x_{n+1} < x_n$ при всех n;
невозрастающей, если $x_{n+1} \leqslant x_n$ при всех n.

В этих случаях последовательность называется *монотонной*.

Определение ограниченной последовательности дано в подпараграфе 7.1.2: для некоторого $M > 0$ имеет место неравенство $|x_n| < M$ при всех n. Последовательность является *ограниченной снизу* (числом M), если существует число M (необязательно положительное), такое, что $x_n > M$ при всех n, и *ограниченной сверху*, если $x_n < M$. Следующее утверждение принимается в математическом анализе за аксиому.

Аксиома Больцано — Вейерштрасса. *Всякая монотонная ограниченная последовательность имеет предел.*

7 Зак. № 2435 Турецкий

В частности, неубывающая и ограниченная сверху числом A последовательность имеет предел $x_0 \leqslant A$; невозрастающая и ограниченная снизу числом B – предел $x_0 \geqslant B$. Возрастающая последовательность, ограниченная сверху, напоминает лестницу, ведущую под потолок. Ее ступеньки расположены все ближе друг к другу, но как бы мы ни пригибались, в какой-то момент мы стукнемся головой о потолок. Вся разница в том, что последовательность упирается в свой потолок за бесконечное число шагов.

Не будем углубляться в проблему оснований анализа. Скажем лишь, что аксиома Больцано – Вейерштрасса эквивалентна утверждению о том, что между любыми двумя действительными числами лежит действительное число, не равное им.

Карл Теодор Вильгельм Вейерштрасс (1815–1897). Великий немецкий математик. Изучал право в Боннском университете (1834–1838), затем математику в Мюнстере. В течение многих лет Вейерштрасс был учителем одной из прусских гимназий. С 1856 г. – профессор Берлинского университета, где преподавал тридцать лет. Слава его лекций все возрастала; главным образом благодаря этим лекциям идеи Вейерштрасса стали достоянием математиков.

Основные работы Вейерштрасса посвящены математическому анализу, линейной алгебре и многим другим областям математики. Он построил логическое обоснование анализа, исходящее из предложенной им же теории действительных чисел. Своей славой Вейерштрасс обязан исключительной тщательности рассуждений, «вейерштрассовской строгости». Он разъяснил понятия минимума, функции, производной и таким образом устранил те неясности, которые оставались в формулировках основных понятий анализа. Он был воплощением математической скрупулезности. Вейерштрасс положил начало сведению принципов математического анализа к простейшим арифметическим понятиям, которое известно как арифметизация математики.

Давид Гильберт писал: «В основном это заслуга научной деятельности Вейерштрасса, что теперь в анализе существуют полное согласие и уверенность относительно таких способов рассуждения, которые основаны на понятии иррационального числа и предела вообще».

Бернард Больцано (1781–1848). Чешский математик, философ, богослов. Родился в Праге. Окончил философский и теологический факультеты Пражского университета. В 1805–1820 гг. – профессор истории религии этого университета. Из-за своих либеральных воззрений был отстранен от преподавания и лишен права публичных выступлений, как устных, так и в печати.

Основные работы относятся к теории множеств, математическому анализу, механике и физике. Внес важный вклад в математическую логику и теорию множеств, последней посвящены «Парадоксы бесконечного», изданные после смерти автора (1851). Выдвинул и обосновал идею арифметической теории действительных чисел. Ему принадлежит ряд теорем математического анализа, он уточнил понятия непрерывности и предела.

Занимался исследованием основных понятий физики. Считал необходимым аксиоматическое построение механики, исследовал понятие силы, построил теорию распространения волн. Занимался вопросами преобразования математики как науки и как предмета преподавания. Считал математический метод единственным методом научного исследования. В своей работе «Наукоучение», опубликованной в 1837 г., он пытался развить методологию всех наук по образцу математики. Большинство его работ опубликовано посмертно.

Приведем пример монотонной ограниченной последовательности. Пусть $x_n = \sqrt[n]{a}$ и $a > 1$. Тогда, очевидно, $\sqrt[n+1]{a} < \sqrt[n]{a}$, т.е. $x_{n+1} < x_n$, и последовательность убывает. С другой стороны, при всех n выполняется $x_n > 1$, так как число, меньшее единицы, при возведении в любую степень не становится больше единицы. Итак, данная последовательность убывает и ограничена снизу. Значит, она имеет предел. При $0 < a < 1$ последовательность возрастает и ограничена сверху (тоже единицей). Можно доказать, что при любом положительном a $\lim \sqrt[n]{a} = 1$.

С аксиомой Больцано – Вейерштрасса связаны два особых в математике числа: π и e.

Рассмотрим круг с радиусом, равным единице. Будем вписывать в него правильные многоугольники: сначала треугольник, потом квадрат, пятиугольник и т.д. (рис. 7.4). Этот процесс можно продолжать бесконечно. Если каждый раз измерять площадь многоугольника, получится числовая последовательность: S_1 – площадь треугольника, S_2 – площадь квадрата и т.д. Можно проверить, что площадь каждого следующего многоугольника больше площади предыдущего, т.е. $S_{n+1} > S_n$, и последовательность $\{S_n\}$ является возрастающей. Кроме того, ясно, что площади всех вписанных фигур не превышают площади круга. Таким образом, построенная последовательность ограничена сверху. Значит, в силу аксиомы Больцано – Вейерштрасса последовательность сходится. Доказано, что предел равен площади единичного круга, которая и обозначается символом π. Если вписывать правильные многоугольники в круг с радиусом R, легко доказать с помощью простых геометрических рассуждений, что их площади равны $R^2 S_n$. Следовательно, предел последовательности их площадей равен $\lim (R^2 S_n) = R^2 \lim S_n = \pi R^2$ – это известная формула площади круга.

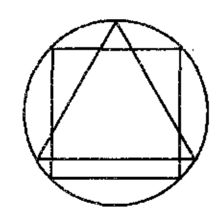


Рис. 7.4

7*

Число π является иррациональным. Это значит, что его нельзя представить в виде отношения k/m двух натуральных чисел. Десятичная запись рационального числа содержит либо конечное число знаков (например, $^1/_2 = 0{,}5$), либо бесконечно повторяющуюся группу цифр (период, его обычно заключают в скобки). Примеры периодических дробей: $^1/_3 = 0{,}33333... = 0{,}(3)$, $^1/_{11} = 0{,}090909... = 0{,}(09)$. Период может начинаться не сразу после десятичной точки: $^1/_7 = 0{,}1428562856... = 0{,}14(2856)$.

Иррациональное число в своей бесконечной десятичной записи не содержит периода, это бесконечная непериодическая десятичная дробь. На самом деле, π не просто иррациональное число. Некоторые иррациональные числа являются корнями алгебраических уравнений, т.е. нулями многочленов. Например, $\sqrt{2}$ есть корень уравнения $x^2 - 2 = 0$. Существуют, однако, иррациональные числа (их называют трансцендентными), которые не могут быть корнями никаких алгебраических уравнений. Именно к этой категории и принадлежит число π.

Мы помним, что $\pi = 3{,}141592...$. Многие математики затратили огромные усилия, чтобы вычислить последующие десятичные знаки числа π. Например, английский математик XIX в. Уильям Шенкс более чем за 20 лет вычислил 707 знаков числа π, которые в память об этом колоссальном труде выбиты на его надгробии. Однако в 1945 г. с помощью электронно-вычислительной машины было обнаружено, что Шенкс допустил ошибку в 520-м знаке и соответственно все следующие знаки, полученные им, неверны. В 1949 г. американский компьютер ENIAC за 70 часов посчитал 2000 знаков π, а в 1961 г. на ЭВМ IBM-7090 за 9 часов было вычислено 100 625 знаков. По иронии судьбы программу написал однофамилец У. Шенкса Дэниел Шенкс.

Другое важное число – это число e. Рассмотрим последовательность $x_n = (1 + 1/n)^n$. Основание $1 + 1/n$ стремится к единице, а степень n – к бесконечности. Получаем новую разновидность неопределенности: 1^∞. Воспользуемся аксиомой Больцано – Вейерштрасса. Можно проверить, что эта последовательность возрастает, т.е. $x_{n+1} > x_n$. Например, $x_1 = (1 + 1)^1 = 2$, $x_2 = (1 + {}^1/_2)^2 = 1{,}5^2 = 2{,}25$, $x_3 = (1 + {}^1/_3)^3 \approx 2{,}37$ и т.д. Опустим строгое доказательство этого факта. Так же доказывается, что $x_n < 3$ при всех n. Итак, данная последовательность возрастает и ограничена сверху. Значит, по аксиоме Больцано – Вейерштрасса она имеет предел. Этот предел и обозначают через e. Это, как и π, трансцендентное число. Первые знаки после запятой легко запомнить: после 2,7 идет 2 раза подряд год рождения Льва Толстого – 1828:

$$\lim (1 + 1/n)^n = e = 2{,}718281828... .$$

Вглядимся повнимательнее, как устроена последовательность, сходящаяся к e. В основании к единице прибавляется бесконечно малая последовательность $\alpha_n = 1/n$, а степень n есть бесконечно большая последовательность, обратная к α_n. Оказывается, любая последовательность, устроенная таким образом, имеет своим пределом число e.

∇11. *Пусть* $\alpha_n \to 0$. *Тогда*

$$\lim (1+\alpha_n)^{1/\alpha_n} = e.$$

Например,

$$\lim \left(1+\frac{1}{2n+1}\right)^{2n+1} = e,$$

$$\lim \left(1+\frac{1}{n^2}\right)^{n^2} = e$$

и т.д. Это свойство позволяет вычислять некоторые пределы. Рассмотрим данный прием на примере последовательности

$$x_n = \left(\frac{n+1}{n-1}\right)^n.$$

Прежде всего приведем основание степени к виду $1 + \alpha_n$:

$$\frac{n+1}{n-1} = \frac{(n-1)+2}{n-1} = 1 + \frac{2}{n-1}.$$

Затем выделим в показателе степени величину, обратную к добавке: $1/\alpha_n = (n - 1)/2$. Для этого домножим и разделим показатель степени на эту величину:

$$n = \left(\frac{n-1}{2}\right)\left(\frac{2}{n-1}\right)n.$$

Теперь исходную последовательность можно записать в виде

$$x_n = \left[\left(1+\frac{2}{n-1}\right)^{\frac{n-1}{2}}\right]^{\left(\frac{2}{n-1}\right)n} = \left[\left(1+\frac{2}{n-1}\right)^{\frac{n-1}{2}}\right]^{\frac{2n}{n-1}}.$$

Последовательность в квадратных скобках стремится к e, поскольку устроена именно так: основание есть сумма единицы и бесконечно малой последовательности, а показатель степени – бесконечно большая последовательность, обратная к этой бесконечно малой добавке. Величина в квадратных скобках возводится в степень $\frac{2n}{n-1}$, а эта последовательность, как мы знаем, стремится к 2: это рациональная дробь, степень числителя и знаменателя равна 1, предел равен отношению старших коэффициентов (см. ∇6 в подпараграфе 7.1.3). Итак, в конечном счете, $\lim x_n = e^2$.

Мы воспользовались еще одним свойством предела, а именно возможностью переходить к пределу отдельно в основании и показателе степени, если они оба существуют и конечны. Общая процедура понятна: выделить в основании степени группу «единица плюс бесконечно малая», выделить в показателе степени величину, обратную к этой бесконечно малой добавке, найти предел выражения в показателе степени.

Функция $y = e^x$ имеет специальное название – *экспонента*. Мы не раз встретимся с этой функцией в курсе математического анализа. Логарифм по основанию e называется *натуральным* и обозначается ln. Функция $y = \ln x$ является обратной к экспоненте.

7.2. Предел функции

7.2.1. Определение предела

Пусть заданы функция $y = f(x)$ и некоторая точка x_0. Можно организовать последовательность x_n, сходящуюся к x_0. Такой последовательности аргументов соответствует последовательность значений функции $y_n = f(x_n)$. Например, для функции $y = x^2$ и $x_0 = 0$ можно выбрать $x_n = 1/n = \{1, {}^1/_2, {}^1/_3, \ldots\} \to 0$. Соответствующая последовательность значений будет $y_n = \dfrac{1}{n^2} = \{1, {}^1/_4, {}^1/_9, \ldots\}$. Понятно, что последовательность $x_n \to x_0$ можно выбрать не единственным способом. Более того, таких последовательностей бесконечно много, и каждой из них соответствует своя последовательность значений y_n. Среди этих последовательностей могут оказаться сходящиеся, в том числе и к разным пределам, а также расходящиеся последовательности. Если же все такие последовательности значений y_n сходятся к одному числу A, его и называют пределом функции в точке x_0. Заметим, что в наших построениях не требовалось, чтобы функция была определена в точке x_0. Достаточно, чтобы были определены значения $f(x_n)$ вблизи x_0. Итак, дадим строгое определение.

Число A называется ***пределом*** функции $f(x)$ в точке x_0, если:

1) функция определена в некоторой окрестности точки x_0, кроме, быть может, самой этой точки;

2) для любой последовательности $x_n \to x_0$, такой, что $x_n \neq x_0$ при всех n, последовательность значений $y_n = f(x_n)$ сходится к A.

Записывается это следующим образом:

$$\lim_{x\to x_0} f(x) = A.$$

Вернемся к примеру с функцией $y = x^2$. Нетрудно проверить, что при $x_n = 1/n$ соответствующая последовательность значений $y_n = \dfrac{1}{n^2}$ сходится к нулю. Это верно и для произвольной $x_n \to 0$: $y_n = x_n^2 = x_n \cdot x_n$ и по арифметическим свойствам $\lim y_n = \lim x_n \cdot \lim x_n = 0$. Итак, получено

$$\lim_{x\to 0} x^2 = 0.$$

Легко вывести, что и в произвольной точке

$$\lim_{x\to x_0} x^2 = x_0^2.$$

Докажем по определению, что функция $y = \sin\dfrac{1}{x}$ не имеет предела в нуле. Организуем две последовательности, сходящиеся к нулю: $x_n^{(1)} = \dfrac{1}{\pi n}$ и $x_n^{(2)} = \dfrac{2}{\pi(4n+1)}$. Первой из них соответствует последовательность значений $y_n^{(1)} = \sin \pi n = 0$. Эта последовательность стационарная и сходится к нулю. Вторая последовательность значений $y_n^{(2)} = \sin(\pi(4n+1)/2) = \sin(\pi/2 + 2\pi n) = 1$ также стационарна и сходится к единице. Таким образом, нашлись две последовательности аргументов, сходящиеся к нулю, для которых соответствующие последовательности значений сходятся к разным пределам. Значит, функция не имеет предела в нуле.

Итак, *A есть предел $f(x)$ в x_0, если $f(x_n) \to A$*, какова бы ни была последовательность $x_n \to x_0$, $x_n \neq x_0$. Это определение предела функции связывают с именем немецкого математика Гейне.

Генрих Эдуард Гейне (1821–1881). Немецкий математик. Родился в Берлине. Учился в Геттингене и Берлине. Окончил Берлинский университет (1842). С 1844 г. работал в Боннском университете, с 1856 г. – профессор университета в Галле (1864–1865 – ректор).

Основные направления исследований – основания математики, математическая физика и теория функций. Работал вместе с Г. Кантором над обоснованием математического анализа, предложенным Вейерштрассом. Ему принадлежат исследования по теории пределов и по геометрии множеств.

Член-корреспондент Берлинской АН, член Геттингенского научного общества.

Нам известно определение бесконечного предела последовательности. Определение предела функции позволяет без труда ввести такие понятия, как предел $f(x)$ при $x \to \infty$ (через последовательности $x_n \to \infty$), а также бесконечный предел в точке x_0 (через последовательности значений $f(x_n) \to \infty$). Не будем останавливаться подробно на этих определениях. Таким образом, возможны следующие варианты:

1) конечный предел в конечной точке: например,

$$\lim_{x\to 2}(x+2)=4;$$

2) бесконечный предел в конечной точке: например,

$$\lim_{x\to 0}(1/x)=\infty;$$

3) конечный предел в бесконечности: например,

$$\lim_{x\to \infty}(1/x)=0;$$

4) бесконечный предел в бесконечности: например,

$$\lim_{x\to \infty}(x^2)=\infty.$$

В соответствующих случаях вместо бесконечности может фигурировать $+\infty$ или $-\infty$.

Поскольку определение предела функции дано на языке последовательностей, арифметические свойства предела последовательности переходят по наследству к пределу функции.

∇1. *Пусть функции $f(x)$ и $g(x)$ имеют конечные пределы в точке x_0 (конечной или бесконечной):*

$$\lim_{x\to x_0} f(x)=A;\quad \lim_{x\to x_0} g(x)=B.$$

Тогда

$$\lim_{x\to x_0}(\alpha f(x))=\alpha A,$$

$$\lim_{x\to x_0}(f(x)\pm g(x))=A\pm B,$$

$$\lim_{x\to x_0}(f(x)\cdot g(x))=AB,$$

$$\lim_{x\to x_0}(f(x)/g(x))=A/B$$

(если $B\neq 0$ и $g(x)\neq 0$ в некоторой окрестности x_0).

В частности, наследуются линейные свойства предела. Так, предел линейной комбинации функций равен линейной комбинации пределов:

$$\lim_{x\to x_0}(\alpha_1 f_1(x)+\ldots+\alpha_n f_n(x))=\alpha_1\lim_{x\to x_0} f_1(x)+\ldots+\alpha_n\lim_{x\to x_0} f_n(x).$$

Как и в случае последовательностей, благодаря арифметическим свойствам расширяются наши возможности при вычислении

пределов функций. Все свойства предела последовательности наследуются пределом функции при $x\to\infty$. Например, наследуются свойства предела рациональной дроби. Такой предел равен:

- нулю, если степень числителя меньше степени знаменателя, например,

$$\lim_{x\to\infty}\frac{x+1}{x^2+2x}=0;$$

- отношению старших коэффициентов, если степень числителя равна степени знаменателя, например,

$$\lim_{x\to\infty}\frac{2x+1}{x+2}=2;$$

- бесконечности, если степень числителя больше степени знаменателя, например,

$$\lim_{x\to\infty}\frac{x^2+1}{x-1}=\infty.$$

Наследуются также свойства пределов последовательностей, сходящихся к e:

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x=\text{e}.$$

Более того, если функция устроена как $f(x)=(1+\varphi(x))^{\frac{1}{\varphi(x)}}$ и при этом $\lim_{x\to x_0}\varphi(x)=0$, то

$$\lim_{x\to x_0} f(x)=\text{e}$$

(лишь бы к единице прибавлялось нечто, стремящееся к нулю, а в показателе степени была величина, обратная этому «нечто»). Например,

$$\lim_{x\to 0}(1+x)^{1/x}=\text{e}.$$

Неопределенность вида $\frac{0}{0}$, $\frac{\infty}{\infty}$ (когда числитель и знаменатель дроби стремятся к нулю или бесконечности) можно раскрыть различными способами. В следующем примере это сделать совсем просто:

$$\lim_{x\to 2}\frac{x^2-4}{x-2}.$$

Предел числителя и знаменателя при $x \to 2$ равен 0. Числитель можно разложить на множители по формуле разности квадратов: $x^2 - 4 = (x - 2)(x + 2)$, а затем сократить дробь на $x - 2$. Это допустимо, так как при вычислении предела предполагается, что $x \neq x_0$ (функция может не быть определенной в точке x_0). Итак, после преобразований предел превращается в $\lim\limits_{x\to 2}(x+2)$, неопределенность отсутствует и предел равен 4.

При раскрытии неопределенности можно воспользоваться так называемым замечательным пределом

$$\lim_{x\to 0}\frac{\sin x}{x}=1.$$

Позже, познакомившись с правилом раскрытия неопределенностей, основанным на производной, мы убедимся в справедливости данного предельного соотношения. На самом деле, если $\lim\limits_{x\to x_0}\varphi(x)=0$, то

$$\lim_{x\to x_0}\frac{\sin\varphi(x)}{\varphi(x)}=1.$$

Например,

$$\lim_{x\to 0}\frac{\sin 5x}{5x}=1.$$

Воспользуемся этим свойством для вычисления, например, предела

$$\lim_{x\to 0}\frac{\operatorname{tg} 2x}{x}$$

(неопределенность $\frac{0}{0}$). Преобразуем функцию:

$$\frac{\operatorname{tg} 2x}{x}=\frac{1}{\cos 2x}\frac{\sin 2x}{x}=2\frac{1}{\cos 2x}\frac{\sin 2x}{2x}$$

(мы вспомнили, что тангенс есть отношение синуса и косинуса, а затем домножили числитель и знаменатель на 2). В силу арифметических свойств

$$\lim_{x\to 0}\frac{\operatorname{tg} 2x}{x}=2\lim_{x\to 0}\frac{1}{\cos 2x}\lim_{x\to 0}\frac{\sin 2x}{2x}=2\cdot 1\cdot 1=2,$$

так как двойку можно вынести за знак предела, предел с косинусом не содержит неопределенности и равен 1, а последний предел — замечательный и тоже равен 1.

7.2.2. Пределы справа и слева

При построении последовательности $x_n \to x_0$ можно по-разному приближаться к точке x_0. В частности, все элементы последовательности x_n могут находиться правее x_0 ($x_n > x_0$). В этом случае говорят, что x_n стремится к x_0 справа. Обозначим это через $x_n \to x_0 + 0$. *Правой окрестностью* точки x_0 называется любой полуинтервал вида $[x_0, b)$. Если в определении предела функции оставить только последовательности, стремящиеся к x_0 справа, мы получим определение предела справа. Итак, число A называется ***пределом справа*** (или ***правым пределом***) функции $f(x)$ в точке x_0, если:

1) функция определена в некоторой правой окрестности точки x_0, кроме, быть может, самой этой точки;

2) для любой последовательности $x_n \to x_0 + 0$ последовательность значений $y_n = f(x_n)$ сходится к A.

Правый предел обозначают как

$$\lim_{\substack{x\to x_0\\ x>x_0}} f(x) \text{ или } f(x_0+0).$$

Аналогично определяется ***предел слева*** (или ***левый предел***). Для этого вводится понятие последовательности, сходящейся к x_0 слева. Левый предел обозначают как $\lim\limits_{\substack{x\to x_0\\ x<x_0}} f(x)$ или $f(x_0 - 0)$. На рис. 7.5 число A есть предел слева в точке x_0, а число B — предел справа. Видно, что эти числа могут не совпадать.

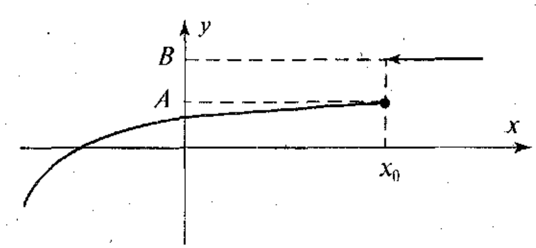


Рис. 7.5

∇2. *Для того чтобы функция $f(x)$ имела предел в конечной точке x_0, необходимо и достаточно, чтобы существовали правый и левый пределы $f(x)$ в этой точке и были равны между собой:*

$$f(x_0+0)=f(x_0-0)=\lim_{x\to x_0} f(x).$$

Задачи

1. Найти пределы последовательностей:

а) $\lim\limits_{n\to\infty}\dfrac{(n+1)^2}{n^3+2n-1}$; б) $\lim\limits_{n\to\infty}\dfrac{(n+1)^3-(n-1)^3}{(n+1)^2+(n-1)^2}$;

в) $\lim\limits_{n\to\infty}\left(\sqrt{n^2+1}-\sqrt{n^2-1}\right)$; г) $\lim\limits_{n\to\infty}\left(\sqrt{n+5}-\sqrt{n-1}\right)$;

д) $\lim\limits_{n\to\infty}\left(\dfrac{n+3}{n}\right)^{3n-1}$; е) $\lim\limits_{n\to\infty}\left(\dfrac{3n+5}{3n+2}\right)^{\frac{n+1}{3}}$.

2. Найти пределы функций:

а) $\lim\limits_{x\to 4}\dfrac{x^2+\sqrt{x}}{x-2}$; б) $\lim\limits_{x\to 3}\left(\dfrac{x}{x-2}\right)^2$;

в) $\lim\limits_{x\to\infty}\dfrac{4x^2-5}{3x^3+2}$; г) $\lim\limits_{x\to\infty}\dfrac{x^5-2x}{3x^5+2}$;

д) $\lim\limits_{x\to 4}\dfrac{x^2-10x+24}{x^2-9x+20}$; е) $\lim\limits_{x\to 3}\dfrac{x^2-2x-3}{x^2-3x}$;

ж) $\lim\limits_{x\to\infty}\left(1+\dfrac{1}{x}\right)^{\frac{x+1}{x}}$; з) $\lim\limits_{x\to\infty}\left(1-\dfrac{1}{x}\right)^{x}$;

и) $\lim\limits_{x\to 0}\dfrac{\operatorname{tg} x}{x}$; $\lim\limits_{x\to 0}\dfrac{\sin x}{\sin 5x}$.

Глава 8
НЕПРЕРЫВНЫЕ ФУНКЦИИ

8.1. Непрерывность в точке

8.1.1. Основные определения

Графики многих функций можно нарисовать не отрывая пера от бумаги. Такие «хорошие» функции называют непрерывными. В чем их особенность? Зафиксируем некоторую точку x_0 и будем двигаться по графику, приближаясь к x_0 сначала справа, а потом слева (рис. 8.1). При этом мы попадем в точку $(x_0, f(x_0))$, лежащую на графике.

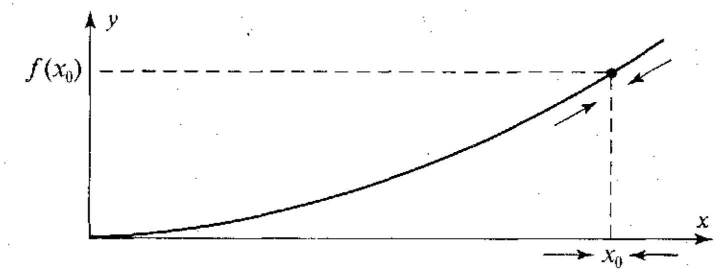


Рис. 8.1

Фактически мы убедились в том, что правый и левый пределы в точке x_0 совпадают и равны значению функции в этой точке. Итак, функция называется ***непрерывной в точке*** x_0, если:

1) функция определена в некоторой окрестности точки x_0, включая саму эту точку;

2) в точке x_0 существует предел функции и он равен значению функции в этой точке:

$$\lim_{x\to x_0} f(x)=f(x_0).$$

В противном случае будем говорить, что функция *разрывна* или имеет *разрыв* в точке x_0. Исходя из определения непрерывности, можно указать три возможные причины разрыва:

- функция не определена в точке x_0, например функция $y = 1/x$ в точке $x_0 = 0$ (рис. 8.2);
- функция не имеет предела в точке x_0. Это может быть в случае, когда правый и левый пределы существуют, но не совпадают. Например, функция

$$y=\begin{cases} x, & x \leqslant 1,\\ x+1, & x > 1.\end{cases}$$

В точке $x = 1$ левый предел равен 1, а правый 2 (рис. 8.3);

- функция имеет предел, но он не совпадает со значением функции в этой точке. Например, функция

$$y=\begin{cases} x, & x \neq 1,\\ 2, & x=1.\end{cases}$$

Мы как бы взяли одну точку (1, 1) на графике функции $y = x$ и подняли ее на одну единицу. В точке $x = 1$ предел существует и равен 1, но не совпадает со значением функции $f(1) = 2$ (рис. 8.4).

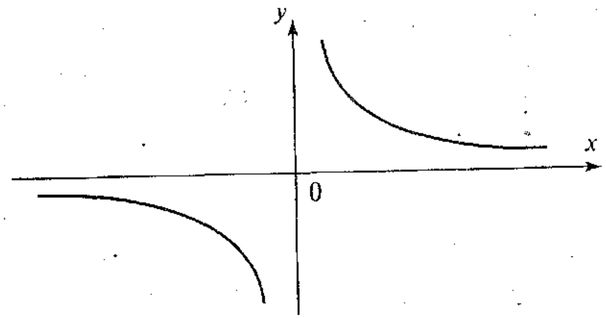


Рис. 8.2

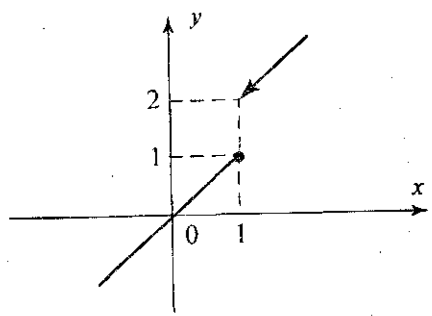


Рис. 8.3

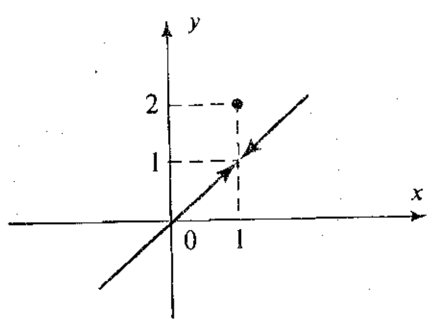


Рис. 8.4

Тот факт, что $\lim\limits_{x \to x_0} f(x) = f(x_0)$, по сути означает следующее: если аргументы функции находятся вблизи точки x_0, то ее значения попадают в малую окрестность $f(x_0)$. Другими словами, функция непрерывна в точке x_0, если **малому приращению аргумента соответствует малое приращение функции** δ. При разрыве, наоборот, сколь

угодно малому приращению аргумента может соответствовать приращение функции, большее некоторого фиксированного числа.

Арифметические свойства предела функции наследуются для непрерывных функций.

∇**1.** *Пусть функции $f(x)$ и $g(x)$ непрерывны в точке x_0. Тогда непрерывными в этой точке являются функции $\alpha f(x)$, $f(x) \pm g(x)$, $f(x)\,g(x)$ и $f(x)/g(x)$ (последняя – если $g(x) \neq 0$ в окрестности x_0).*

В частности, линейная комбинация непрерывных функций непрерывна.

Примеры непрерывных функций.

1. Самый простой пример – константа: $y \equiv c$. Эта функция непрерывна в любой точке. В самом деле, при любом приращении аргумента приращение функции равно нулю.

2. Функция $y = x$ также непрерывна во всех точках: приращение функции равно приращению аргумента и в определении непрерывности можно взять $\delta = \varepsilon$.

3. Любая целая степень x (функция $y = x^n$) непрерывна как произведение непрерывных функций.

4. Многочлен $y = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$ непрерывен во всех точках в силу арифметических свойств.

5. Можно доказать, что функции $y = \sin x$, $y = \cos x$, $y = e^x$ и многие другие непрерывны во всех точках x.

6. Функция $y = 1/x$ непрерывна во всех точках, кроме $x = 0$.

7. Функция $y = \operatorname{tg} x$ непрерывна во всех точках, кроме $x = \pi/2 + \pi k$ при любом целом k.

8.1.2. Непрерывность слева и справа. Классификация разрывов

В подпараграфе 8.1.2 вводится свойство, более слабое, чем непрерывность. Функция называется ***непрерывной слева*** в точке x_0, если левый предел в этой точке существует и равен значению функции:

$$f(x_0 - 0) = f(x_0).$$

Аналогично определяется ***непрерывность справа***:

$$f(x_0 + 0) = f(x_0).$$

На рис. 8.5, *а* функция непрерывна слева в точке x_0, так как левый предел совпадает со значением функции ($A = f(x_0)$), на рис. 8.5, *б* – непрерывна справа ($B = f(x_0)$).

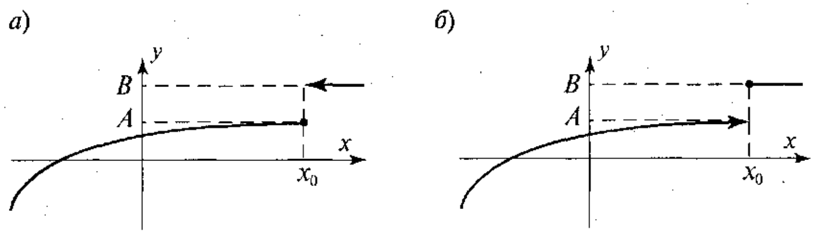


Рис. 8.5

Функция, непрерывная в точке, разумеется, непрерывна в ней слева и справа. Из определений вытекает также следующее наблюдение.

∇2. *Если функция непрерывна в точке x_0 как слева, так и справа, то она непрерывна в этой точке.*

Другими словами, для того чтобы функция f была непрерывной в точке x_0, необходимо и достаточно, чтобы все три числа $f(x_0 - 0)$, $f(x_0)$ и $f(x_0 + 0)$ имели смысл и были равны между собой. Когда говорят, что левый и правый пределы имеют смысл, это значит, что они конечны.

Если функция имеет предел в точке x_0 и при этом разрывна в этой точке, говорят, что в точке x_0 имеется *устранимый* разрыв. Как можно понять из этого названия, в данном случае легко восстановить непрерывность функции, т.е. устранить разрыв. В самом деле, возможны только два случая, когда предел есть, а непрерывности нет:

1) функция не определена в этой точке (точка на графике «выколота»). Например,

$$y = \frac{x^2 - x}{x - 1}.$$

Очевидно, что функция не определена при $x = 1$, однако предел найти очень легко, так как $f(x) = x$ при $x \neq 1$ и $\lim\limits_{x\to 1} f(x) = 1$. В этом случае восстановить непрерывность можно доопределив функцию $f(x)$ в точке $x = 1$:

$$y = \begin{cases} \dfrac{x^2 - x}{x - 1}, & x \neq 1, \\ 1, & x = 1, \end{cases}$$

т.е. вернув выколотую точку на график;

2) функция определена, но значение не совпадает с пределом. Такая функция нам уже встречалась (см. рис. 8.3):

$$y = \begin{cases} x, & x \neq 1, \\ 2, & x = 1. \end{cases}$$

Разрыв устраняется, если переопределить функцию в точке $x = 1$, а именно: вместо $f(1) = 2$ положить $f(1) = \lim\limits_{x\to 1} f(x) = 1$ (рис. 8.6).

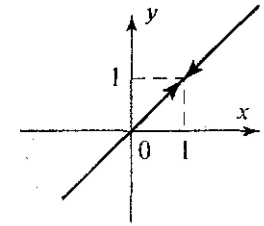


Рис. 8.6

Итак, непрерывность в точке устранимого разрыва восстанавливается путем доопределения или переопределения функции в данной точке. Соответственно, *неустранимый* разрыв возникает, если предела функции в данной точке не существует. Например, функция $y = 1/x$ не имеет предела при $x = 0$, и никаким доопределением или переопределением непрерывность не восстановить.

Если в точке x_0 имеют смысл оба числа $f(x_0 - 0)$ и $f(x_0 + 0)$ и при этом функция разрывна, говорят, что в точке x_0 имеет место *разрыв I рода*. Как ясно из предыдущих определений, разрыв I рода может быть как устранимым (если $f(x_0 - 0) = f(x_0 + 0)$), так и неустранимым. Все устранимые разрывы являются разрывами I рода.

На рис. 8.7 показаны шесть различных ситуаций разрыва I рода.

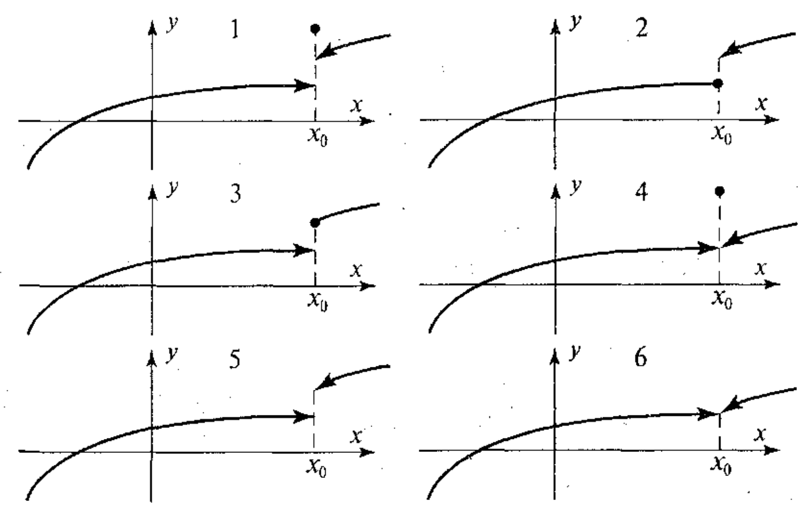


Рис. 8.7

В ситуациях 1–4 все три числа $f(x_0 - 0)$, $f(x_0)$ и $f(x_0 + 0)$ имеют смысл: в первом случае все они различны, во втором функция непрерывна слева, в третьем – справа. Это неустранимые разрывы I рода. Четвертый случай – существует предел, не совпадающий со значением функции: это устранимый разрыв I рода. В последних двух случаях (5 и 6) функция не определена в точке x_0: пятая ситуация – неустранимый разрыв, шестая – устранимый.

Если не существует конечного левого или правого предела (или обоих), говорят о *разрыве II рода* (такие разрывы еще называют *бесконечными*). Разрыв II рода имеет, например, функция $y = 1/x$ в точке $x = 0$ (левый предел равен $-\infty$, а правый $+\infty$) или функция $y = \operatorname{tg} x$ в точках $x = \pi/2 + \pi k$.

8.2. Непрерывность на отрезке

В параграфе 8.1 непрерывность рассматривалась как л о к а л ь н о е свойство функции, т.е. как непрерывность в точке. Функция называется ***непрерывной на отрезке*** $[a, b]$, если она непрерывна во всех точках интервала (a, b), непрерывна справа в точке a и непрерывна слева в точке b. Функции, непрерывные на отрезке, обладают рядом интересных свойств, которые формулируются в виде двух теорем Вейерштрасса и теоремы Коши.

Теорема 8.1 (первая теорема Вейерштрасса). *Если функция f непрерывна на отрезке $[a, b]$, то она ограничена на нем.*

В самом деле, если соединить две точки линией не отрывая пера от бумаги, нам не удастся уйти в бесконечность. В формулировке теоремы важна именно непрерывность на отрезке, т.е. на промежутке, включающем концы. Если хотя бы на одном из концов отрезка непрерывность нарушается, функция может и не быть ограниченной. Например, функция $y = 1/x$ непрерывна на полуинтервале $(0, 1]$, но имеет разрыв на левом конце в точке $x = 0$ (рис. 8.8). При этом функция не ограничена на $(0, 1]$: какое бы большое число $M > 0$ мы ни задали, найдется достаточно близкая к нулю точка, в которой $f(x) > M$. С другой стороны, если чуть-чуть отступить от нуля и рассмотреть отрезок $[\varepsilon, 1]$ при любом малом $\varepsilon > 0$, ситуация изменится: функция непрерывна на этом отрезке и ограничена на нем.

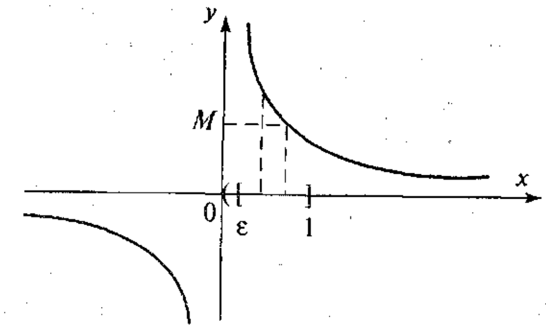


Рис. 8.8

Теорема 8.2 (вторая теорема Вейерштрасса). *Непрерывная на отрезке $[a, b]$ функция f достигает в некоторых точках отрезка $[a, b]$ своего максимума и минимума.*

Это значит, что существуют точки α, $\beta \in [a, b]$, такие, что

$$f(\alpha) \geqslant f(x) \quad \forall x \in [a, b], \qquad f(\beta) \leqslant f(x) \quad \forall x \in [a, b].$$

По-другому это записывается так:

$$\max_{x \in [a,b]} f(x) = f(\alpha), \qquad \min_{x \in [a,b]} f(x) = f(\beta).$$

Точки максимума α и минимума β могут находиться как внутри отрезка $[a, b]$, так и на его границах (совпадать с точками a или b). На рис. 8.9 точка минимума лежит внутри отрезка, а точка максимума – на его правой границе. Разумеется, может быть и несколько точек максимума и минимума: теорема утверждает только, что такие точки обязательно существуют.

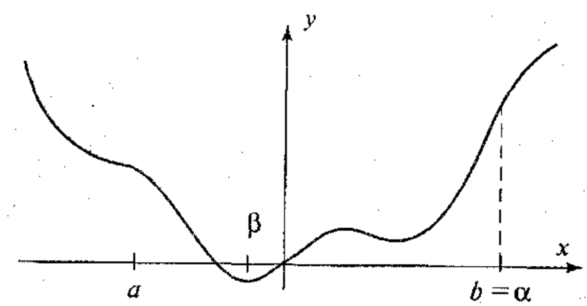


Рис. 8.9

Остается справедливым замечание о том, что важна непрерывность именно на отрезке. Та же функция $y = 1/x$ не достигает максимума на полуинтервале $(0, 1]$, но достигает и максимума, и минимума на любом отрезке $[\varepsilon, 1]$ (поскольку функция монотонно убывает, максимум достигается на левой границе, а минимум — на правой).

Теорема 8.3 (теорема Коши). *Пусть функция f непрерывна на отрезке $[a, b]$, а значения на концах $f(a)$, $f(b)$ не равны нулю и имеют разные знаки...*

Прервем формулировку теоремы и попытаемся предсказать ее заключение. Итак, представим себе: значения функции на концах имеют разные знаки, например $f(a) < 0$, а $f(b) > 0$ (рис. 8.10).

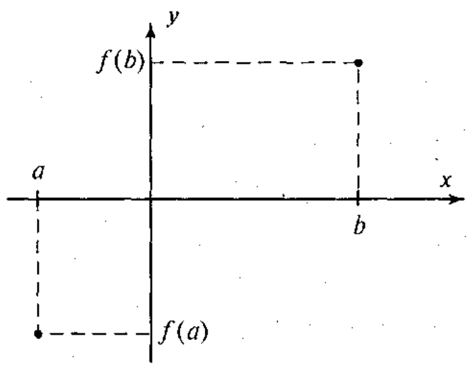


Рис. 8.10

Функция непрерывна на отрезке, значит, нарисовать ее график можно не отрывая пера от бумаги. Что произойдет при движении от точки $(a, f(a))$ к точке $(b, f(b))$ на графике? Разумеется, вы уже догадались: мы неминуемо пересечем ось x, ведь точки расположены по разные стороны от этой прямой. (Как Герда из «Снежной королевы», переходя с половины принцессы на королевскую половину, она неминуемо пересекала черту и оказывалась во власти беззакония.) Именно это и утверждает теорема Коши.

Теорема 8.3 (теорема Коши). *Пусть функция f непрерывна на отрезке $[a, b]$, а значения на концах $f(a)$, $f(b)$ не равны нулю и имеют разные знаки. Тогда на интервале (a, b) имеется по крайней мере одна точка c, такая, что*

$$f(c) = 0.$$

На рис. 8.11, *а* одна такая точка c. Конечно, если функция не монотонная, таких точек может быть несколько (рис. 8.11, *б*): в теореме говорится о том, что существует хотя бы одна такая точка.

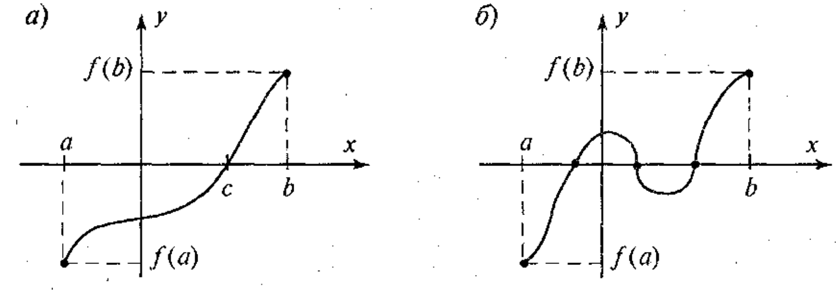


Рис. 8.11

Огюстен Луи Коши (1789–1857). Великий французский математик. Родился в Париже. Окончил Политехническую школу (1807) и Школу дорог и мостов (1810) в Париже. В 1810–1813 гг. работал инженером на сооружении военного порта в Шербуре. С 1816 г. – профессор Политехнической школы, в 1816–1830 гг. – Сорбонны, а в 1848–1857 – Коллеж де Франс.

Коши написал более 700 математических работ, в которых заложил основы современной математики – теории функций, математической физики, математического анализа. Развивал теорию определителей, интегральное исчисление, теорию дифференциальных уравнений. Сформулировал понятие непрерывности функции. Дал определение интеграла. Написал ряд работ в области геометрии, алгебры, теории чисел. Ввел понятие конечной группы. Занимался теорией упругости и оптикой. Научному творчеству Коши свойствен «глобальный» подход к решению поставленных проблем: зная результаты для бесконечного числа значений исследуемого объекта (что графически изображается в виде кривой), он выводил общие свойства функции.

Почетный член Петербургской АН (с 1831).

В теореме Коши, по сути, утверждается, что если функция на отрезке удовлетворяет ее условиям (непрерывна и имеет ненулевые значения разных знаков на концах), то на интервале (a, b) существует по крайней мере одно решение уравнения

$$f(x) = 0.$$

Например, уравнение $\cos x - x = 0$ имеет корень на интервале $(0, \pi)$, так как функция $y = \cos x - x$ непрерывна на отрезке $[0, \pi]$, $f(0) = 1 > 0$, а $f(\pi) = -1 - \pi < 0$.

Задачи

1. Функция определена следующим образом:

$$y=\begin{cases}0, & \text{при } x<0;\\ x, & \text{при } 0\leqslant x<1;\\ -x^2+4x-2, & \text{при } 1\leqslant x<3;\\ 4-x, & \text{при } x\geqslant 3.\end{cases}$$

Будет ли эта функция непрерывной?

2. Пусть

$$f(x)=\begin{cases}x+1, & \text{если } x\leqslant 1;\\ 3-ax^2, & \text{если } x>1.\end{cases}$$

При каком выборе числа a функция $f(x)$ будет непрерывной? Построить ее график.

3. Пусть

$$f(x)=\begin{cases}-2\sin x, & \text{если } x\leqslant -\pi/2;\\ A\sin x+B, & \text{если } -\pi/2<x<\pi/2;\\ \cos x, & \text{если } x\geqslant \pi/2.\end{cases}$$

Подобрать числа A и B так, чтобы функция $f(x)$ была непрерывной; построить ее график.

4. Функция $f(x)=\dfrac{x^2-1}{x^3-1}$ не определена при $x=1$. Каким значением $f(1)$ доопределить функцию, чтобы она стала непрерывной при $x=1$?

5. Какого рода разрывы имеют функции $y=\dfrac{\sin x}{x}$ и $y=\dfrac{\cos x}{x}$ при $x=0$?

6. Используя теорему Коши, убедиться в том, что уравнение $x^5-3x=1$ имеет по крайней мере один корень, заключенный между 1 и 2.

7. Показать, что уравнение $x\cdot 2^x=1$ имеет по крайней мере один положительный корень, не превосходящий 1.

8*. Пусть функция $y=[x]$ есть целая часть числа x, т.е. наибольшее целое число, не превосходящее x (например, $[3{,}5]=3$, $[-3{,}5]=-4$ и т.д.). Исследовать непрерывность и построить графики функций:

а) $y=[x]$;

б) $y=x-[x]$;

в) $y=(-1)^{[x]}$.

Глава 9
ОСНОВЫ ДИФФЕРЕНЦИАЛЬНОГО ИСЧИСЛЕНИЯ

9.1. Производная

9.1.1. Мгновенная скорость. Определение производной

Любая функция отражает зависимость одной величины от другой (аргумента). Часто мы говорим: «Как быстро растут цены!» При этом мы фиксируем не только функциональную зависимость цен от времени, но и скорость их изменения. Другие процессы происходят медленно: например, изменение средней температуры на Земле на один градус требует веков. Можно представить себе графики упомянутых функций: первый поднимается круто, второй – плавно, почти незаметно для глаза. Итак, нас интересует, насколько быстро меняется та или иная величина при изменении аргумента. Фактически речь идет о с к о р о с т и изменения функции. Ясно, что скорость не является постоянной. В частности, процесс потепления на Земле происходит все быстрее. Таким образом, возникает понятие «мгновенной» скорости, скорости в данный момент времени или при данном значении аргумента. Как же определяется мгновенная скорость, скажем, механического движения?

Чтобы разобраться в этом, поднимемся вместе с Галилео Галилеем на вершину Пизанской башни, бросим вниз камушек и будем за ним наблюдать. Галилей обнаружил факт, известный сейчас любому школьнику: расстояние, которое преодолел камень при падении с башни, зависит от времени следующим образом:

$$s(t)=\frac{gt^2}{2},$$

где через g обозначено ускорение свободного падения. Эта формула, разумеется, является приближенной, так как не учитывает сопротивления воздуха. Однако она достаточно точна для практических вычислений. Понятие скорости не вызывает затруднений, если скорость не меняется: скоростью называется путь, пройденный за единицу времени. Наблюдая за камушком и пользуясь формулой, мы убедимся, что путь за первую секунду меньше, чем за вторую и т.д.; значит, скорость движения камня непостоянна. Поступим следующим образом. Рассмотрим движение на малом промежутке времени от момента t_0 до момента $t_0 + \Delta t$ (Δt – малое число). Промежуток времени настолько мал, что скорость падения камня на этом промежутке можно считать постоянной. Разумеется, в этом случае скорость можно посчитать, разделив пройденный путь на время Δt. Путь легко вычислить, зная, что к моменту t_0 камень преодолел расстояние $s(t_0)$, а к моменту $t_0 + \Delta t$ – расстояние $s(t_0 + \Delta t)$. Значит, путь, пройденный за время Δt, составит $s(t_0 + \Delta t) - s(t_0)$. Поскольку в действительности скорость непостоянна, мы напишем приближенное равенство

$$v \approx \frac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}.$$

На самом деле, величина, посчитанная таким образом, есть с р е д н я я скорость на промежутке времени от t_0 до $t_0 + \Delta t$. Наше приближенное равенство будет тем точнее, чем меньше выбран промежуток времени Δt. Вот теперь мы и введем мгновенную скорость: устремим Δt к нулю и договоримся, что скорость в момент t_0 есть п р е д е л отношения пути и времени:

$$v(t_0) = \lim_{\Delta t \to 0} \frac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}.$$

Такой предел и называют ***мгновенной скоростью*** тела в данный момент времени. Используя формулу зависимости пути от времени, найдем мгновенную скорость:

$$s(t_0 + \Delta t) - s(t_0) = \frac{g}{2}((t_0 + \Delta t)^2 - t_0^2) = \frac{g}{2}(t_0^2 + 2t_0\Delta t + (\Delta t)^2 - t_0^2) =$$

$$= \frac{g}{2}\Delta t(2t_0 + \Delta t)$$

и, стало быть,

$$\lim_{\Delta t \to 0} \frac{s(t_0 + \Delta t) - s(t_0)}{\Delta t} = \lim_{\Delta t \to 0} \frac{g}{2}(2t_0 + \Delta t) = gt_0$$

(мы применили арифметические свойства предела: вынесли константу и посчитали предел суммы как сумму пределов). Итак, $v(t_0) = gt_0$, а в произвольный момент времени $v(t) = gt$. Скорость зависит от времени линейно и каждую следующую секунду возрастает в g раз.

Мгновенная скорость движения дает механическую интерпретацию понятия производной, которая, по существу, уже определена. Пусть задана функция $f(x)$. ***Приращением*** функции в точке x_0 при приращении аргумента Δx называется величина $\Delta f = f(x_0 + \Delta x) - f(x_0)$. ***Производной*** $f'(x_0)$ функции в точке x_0 называется предел отношения приращения функции к приращению аргумента при стремлении приращения аргумента к нулю:

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x},$$

если данный предел существует (важное уточнение, поскольку производная есть не у всех функций и не во всех точках).

Сравнивая определения мгновенной скорости и производной, мы немедленно замечаем, что мгновенная скорость есть не что иное, как производная функции пути:

$$v(t_0) = s'(t_0).$$

Ясно, что предел можно посчитать в произвольной точке x. Поэтому производная $f'(x)$ сама по себе является функцией (как и мгновенная скорость, которая изменяется со временем). Производную $f'(x)$ иногда обозначают $\frac{\mathrm{d}f}{\mathrm{d}x}$. Происхождение такого обозначения станет понятно позже, когда мы введем понятие дифференциала. Производная функции $y = f(x)$ также обозначается через y'.

Функция, имеющая производную в точке x_0, называется *дифференцируемой* в данной точке, а процесс отыскания производной – *дифференцированием*; найти производную функции – значит ее продифференцировать. Зарождение дифференциального исчисления связано с именами двух величайших мыслителей – Ньютона и Лейбница.

Исаак Ньютон (1643–1727). В 1665 г. Исаак Ньютон окончил Кембриджский университет и собирался начать работу там же, в своем родном Тринити-колледже. Однако чума, бушевавшая в Англии, заставила Ньютона уединиться на своей ферме, в Вулсторпе. «Чумные каникулы» затянулись почти на два года. Тогда и сделал молодой ученый почти все свои открытия в физике и математике. Он открыл закон всемирного тяготения и приступил с его помощью к исследованию планет.

Для того чтобы исследовать законы физики, Ньютону приходилось заниматься и математикой. В Вулсторпе Ньютон, решая задачи на проведение касательных к кривым, создал общий метод решения таких задач – метод флюксий (которые мы называем производными) и флюэнт (которые у Лейбница назывались дифференциалами). О дифференциальном и интегральном исчислении Ньютон подробно написал в своей самой значительной работе по математике «Метод флюксий» (1670–1671), опубликованной уже после его смерти. В этой работе были заложены основы математического анализа.

Когда Ньютон вернулся в Кембридж в 1666 г., он привез бесчисленные и бесценные результаты своих математических занятий в Вулсторпе. У него пока не было времени привести их в форму, пригодную для публикации, и он не торопился с этим. В 1669 г. он возглавил физико-математическую кафедру. В 1672 г. его выбрали членом Лондонского королевского общества (Английской АН).

В 1680 г. Ньютон начал работу над основным своим сочинением «Математические начала натуральной философии», в котором он задумал изложить свою систему мира. Выход книги был крупным событием в истории естествознания. В ней все величественное здание механики строится на основе аксиом движения, известных теперь как законы Ньютона.

Многие математические труды Ньютона так и не были своевременно опубликованы. Первые его сравнительно подробные публикации относятся к 1704 г. Среди них – «Рассуждения о квадратуре круга», посвященные дифференциальному и интегральному исчислению.

В 1688 г. Ньютона избрали в парламент, а в 1699 г. он переехал в Лондон, где получил пожизненное место директора Монетного двора.

Работы Исаака Ньютона надолго определили пути развития физики и математики. Значительная часть классической механики сохранилась в виде, созданном Ньютоном. Созданный им математический анализ открыл новую эпоху в математике.

Готфрид Вильгельм Лейбниц (1646–1716). Математика не была единственной страстью этого великого мыслителя. С юных лет ему хотелось познать природу в целом, и математика должна была стать главным средством в этом познании. Он был философом и лингвистом, историком и биологом, дипломатом и политическим деятелем, математиком и изобретателем. Научные и общественные планы Лейбница были грандиозны. Он мечтал о создании всемирной академии наук, о построении «универсальной науки». Он хотел выделить простейшие понятия, из которых по определенным правилам можно сформировать все сколь угодно сложные понятия. Лейбниц мечтал об универсальном языке, позволяющем записывать любые мысли в виде математических формул, причем логические ошибки должны проявляться в виде математических ошибок. Он думал о машине, которая выводит теоремы из аксиом, о превращении логических утверждений в арифметические.

Однако грандиозность замыслов уживалась у Лейбница с пониманием того, что можно реально осуществить. Он не мог организовать всемирную академию, но в 1700 г. организовал академию в Берлине, рекомендовал Петру I организовать академию в России. При создании Петербургской АН в 1725 г. пользовались планами Лейбница. Он прекрасно умел решать конкретные задачи и в математике. Создал новый тип арифмометра, который не только складывает и вычитает числа, но и умножает, делит, возводит в степень и извлекает квадратные и кубические корни, решил трудные геометрические задачи. Ввел понятие определителя и заложил основы теории определителей.

И все же Лейбниц всегда стремился рассмотреть любой вопрос под самым общим углом зрения. Особенно ярко проявились эти его качества, когда он открыл метод, позволяющий решать самые разнообразные математические задачи: он создал дифференциальное и интегральное исчисления, которые в другом варианте были построены, но не опубликованы Ньютоном. Он разработал также символику математического анализа, которая сохранилась до наших дней. В отличие от Ньютона Лейбниц потратил много сил, чтобы передать свой метод другим математикам, среди которых выделялись братья Якоб и Иоганн Бернулли. По его инициативе был создан журнал, в котором группа математиков оттачивала методы нового математического анализа. Смысл своей жизни Лейбниц видел в познании природы, в создании теорий, помогающих раскрыть ее законы.

Продифференцируем некоторые «школьные» функции и составим таблицу основных производных.

1. Константа $y \equiv c$. Очевидно, что приращение этой функции в любой точке равно нулю, так как ее значения не меняются. Итак, во всех точках $\Delta f/\Delta x = 0$ и, соответственно, $y' = 0$; константа дифференцируема на всей числовой оси и имеет нулевую производную. Это согласуется с нашими представлениями о производной как о скорости изменения функции: значения функции не меняются и скорость изменения, естественно, равна нулю:

$$c' = 0.$$

2. $y = x$. Для этой функции приращение функции совпадает с приращением аргумента: $\Delta f = (x + \Delta x) - x = \Delta x$. Таким образом,

$$\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta x}{\Delta x} = 1$$

(мы можем сократить на Δx, так как при вычислении предела предполагается, что $\Delta x \neq 0$). Итак,

$$x' = 1.$$

во всех точках x. Действительно, скорость изменения этой функции постоянна.

3. $y = x^2$. Фактически нам уже пришлось вычислить производную квадратичной функции при определении мгновенной скорости падения камня. Повторим вычисления:

$$\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{(x+\Delta x)^2 - x^2}{\Delta x} =$$

$$= \lim_{\Delta x \to 0} \frac{2x\Delta x + (\Delta x)^2}{\Delta x} = \lim_{\Delta x \to 0}(2x + \Delta x) = 2x.$$

Итак,

$$(x^2)' = 2x.$$

Производная этой функции уже не постоянна и линейно зависит от аргумента, как и мгновенная скорость движения.

4. $y = \sin x$. Для этой функции

$$\frac{\Delta f}{\Delta x} = \frac{\sin(x+\Delta x) - \sin x}{\Delta x} = \frac{2\cos\left(\frac{2x+\Delta x}{2}\right)\sin\left(\frac{\Delta x}{2}\right)}{\Delta x} =$$

$$= \cos\left(\frac{2x+\Delta x}{2}\right)\left[\frac{\sin(\Delta x/2)}{\Delta x/2}\right].$$

Величина в квадратных скобках стремится к единице при $\Delta x \to 0$, так как это замечательный предел. Значит,

$$\lim_{\Delta x\to 0}\frac{\Delta f}{\Delta x} = \lim_{\Delta x\to 0}\cos\left(\frac{2x+\Delta x}{2}\right) = \cos x.$$

Итак,

$$(\sin x)' = \cos x.$$

Аналогично получаем, что

$$(\cos x)' = -\sin x.$$

5. $y = \ln x$, $x > 0$. Вычисление производной этой функции – хорошая тренировка. Вспомним свойства логарифма:

$$\ln(a/b) = \ln a - \ln b, \qquad \ln(a^b) = b\ln a.$$

Значит,

$$\frac{\Delta f}{\Delta x} = \frac{\ln(x+\Delta x) - \ln x}{\Delta x} = \ln\left(\frac{x+\Delta x}{x}\right)^{1/\Delta x} = \ln\left(1+\frac{\Delta x}{x}\right)^{1/\Delta x} =$$

$$= \ln\left[\left(1+\frac{\Delta x}{x}\right)^{x/\Delta x}\right]^{1/x} = \frac{1}{x}\ln\left[\left(1+\frac{\Delta x}{x}\right)^{x/\Delta x}\right].$$

Величина в квадратных скобках стремится к e при $\Delta x \to 0$, поскольку основание есть сумма единицы с функцией, стремящейся к нулю, а показатель степени – величина, обратная этой малой добавке. Итак,

$$\lim_{\Delta x\to 0}\frac{\Delta f}{\Delta x} = \frac{1}{x}\ln e = \frac{1}{x},$$

так как по определению $\ln e = 1$. Итак,

$$(\ln x)' = \frac{1}{x}.$$

Такой результат трудно было предсказать, опираясь на интуицию.

Составим начальную таблицу производных, которые нам удалось посчитать по определению:

$f(x)$	$f'(x)$
c	0
x	1
x^2	$2x$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\ln x$	$\frac{1}{x}$

Приведем пример недифференцируемой функции. Пусть $y = |x|$ (рис. 9.1).

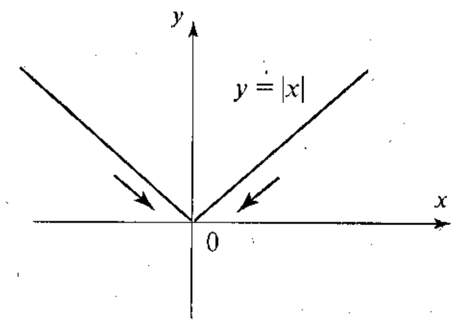


Рис. 9.1

Попытаемся найти производную этой функции в точке $x = 0$:

$$\frac{\Delta f}{\Delta x} = \frac{|0+\Delta x| - |0|}{\Delta x} = \frac{|\Delta x|}{\Delta x} = \begin{cases} 1 & \text{при } \Delta x > 0, \\ -1 & \text{при } \Delta x < 0. \end{cases}$$

Значит, предел $\frac{\Delta f}{\Delta x}$ справа равен 1, а предел слева равен -1, т.е. предела отношения приращения функции к приращению аргумента в этой точке не существует. Стало быть, не существует и производной: функция $y = |x|$ не дифференцируема в точке $x = 0$. Из этого примера хорошо видно, что непрерывная функция может быть недифференцируемой. Тем не менее справедливо обратное утверждение:

∇**1.** *Если функция $f(x)$ имеет производную в точке x_0, то она непрерывна в этой точке.*

Как вы уже поняли, производная также является функцией. Поэтому от нее в свою очередь можно взять производную. Так определяется *вторая производная* функции $f(x)$. Например, легко убедиться, что для функции $f(x) = x^3 + 3x^2 + 2x + 1$ вторая производная равна $f''(x) = 6x + 6$. Аналогично определяются третья и последующие производные.

9.1.2. Геометрический смысл производной

Пусть заданы функция $y = f(x)$, точка x_0. Зададим некоторое приращение Δx и проведем прямую через точки $A(x_0, f(x_0))$ и $B(x_0 + \Delta x, f(x_0 + \Delta x))$, лежащие на графике функции (рис. 9.2). Такая прямая называется *секущей*. Легко определить угловой коэффициент k секущей, т.е. тангенс угла наклона этой прямой к оси абсцисс. В прямоугольном треугольнике ABC, где катет AC проведен параллельно оси x, а катет BC – параллельно оси y, угол BAC совпадает с углом наклона секущей относительно оси абсцисс. Значит,

$$k = \operatorname{tg}\alpha = \operatorname{tg}\angle BAC = \frac{BC}{AC}.$$

Заметим, что $AC = \Delta x$, а $BC = f(x_0 + \Delta x) - f(x_0) = \Delta f$. Таким образом,

$$k = \frac{\Delta f}{\Delta x}.$$

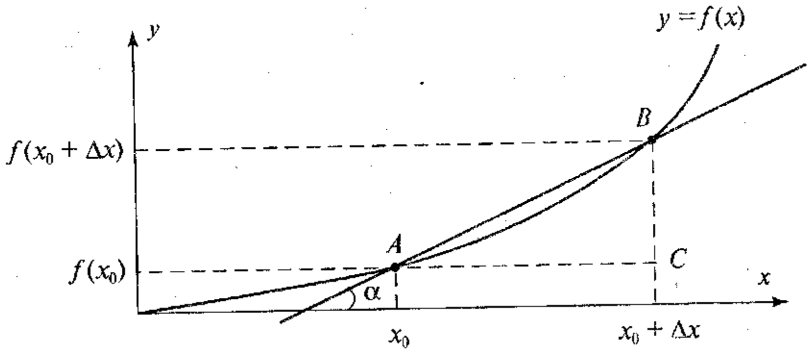


Рис. 9.2

Будем уменьшать приращение аргумента Δx, т.е. приближать точку B к точке A на графике. В конце концов, при $\Delta x \to 0$ точки сольются, а секущая превратится в *касательную* (рис. 9.3). Эта прямая наиболее плотно прилегает к графику функции в точке $A(x_0, f(x_0))$ и не имеет других общих точек с графиком функции.

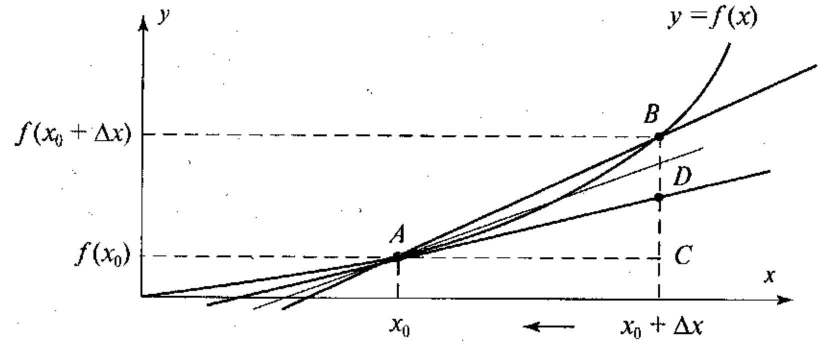


Рис. 9.3

Итак, в пределе секущая перешла в касательную. Чтобы найти угловой коэффициент касательной, перейдем к пределу в угловом коэффициенте секущей:

$$k_{\text{кас}} = \lim_{\Delta x \to 0} k = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}.$$

Это выражение нам уже знакомо: данный предел равен п р о и з в о д н о й функции $f(x)$ в точке x_0. Таким образом, мы приходим к следующему выводу:

> угловой коэффициент касательной к графику функции $f(x)$ равен производной функции в данной точке.

Итак,

$$k_{\text{кас}} = f'(x_0),$$

если, разумеется, производная существует.

Рассмотрим для примера функцию $y = x^2$. Касательная к графику в точке $x_0 = 1$ имеет угловой коэффициент, равный $2x_0 = 2 \cdot 1 = 2$.

Геометрическая интерпретация производной позволяет определять точки, в которых производная отсутствует. Это точки графика, в которых невозможно провести касательную, т.е. описанная процедура перехода от секущей к касательной не приводит к результату. Ранее рассматривался пример функции $y = |x|$ в точке 0 (см. рис. 9.1). Любая секущая правее нуля совпадает с прямой $y = x$, а любая секущая левее нуля – с прямой $y = -x$. Поэтому, приближаясь к нулю с разных сторон, мы получим разные прямые, т.е. правый и левый пределы отношения приращений не совпадают. Любые изломы, острые углы на графике – это точки, в которых нет производной. Их называют точками нарушения гладкости функции, а функция, дифференцируемая во всех точках отрезка, называется *гладкой* на этом отрезке.

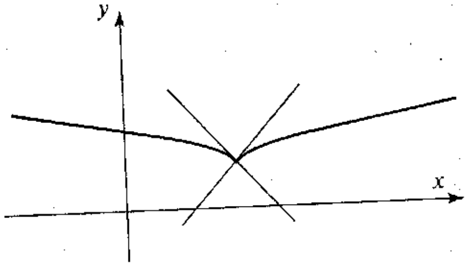


Рис. 9.4

На рис. 9.4 изображены «касательные», которые получаются при приближении к точке излома справа и слева: эти прямые не совпадают.

Геометрическая интерпретация производной позволяет ввести важное в математическом анализе понятие дифференциала функции. Вернемся к рис. 9.3. Истинное приращение функции равно отрезку BC. Однако, если приращение аргумента Δx мало, приращение функции приблизительно равно отрезку CD (точка D получается при движении не по секущей, а по касательной): $BC \approx CD$. Поскольку, как мы выяснили, $\operatorname{tg} \angle BAC = f'(x_0)$, из прямоугольного треугольника ADC легко найти $CD = AC \operatorname{tg} \angle BAC = f'(x_0)\Delta x$. Таким образом, $\Delta f = f(x_0 + \Delta x) - f(x_0) \approx f'(x_0)\Delta x$. Правую часть этого приближенного равенства называют *дифференциалом* функции $f(x)$ в точке x_0 и обозначают $df(x_0)$. Таким образом, дифференциал представляет собой приближенное значение приращения функции, если «двигаться» по касательной, а не по секущей. Приближенное равенство будет тем точнее, чем меньше приращение аргумента Δx. Пусть $f(x) = x$. Тогда $df(x) = dx = x' \cdot \Delta x = 1 \cdot \Delta x = \Delta x$, т.е. приращение аргумента есть не что иное, как дифференциал функции $f(x) = x$. Поэтому обычно пишут $df(x) = f'(x)dx$. Например, $d(\sin x) = \cos x\, dx$.

9.1.3. Правила дифференцирования

Производная определяется через предел. Поэтому линейные свойства предела наследуются при дифференцировании.

1°. Линейные свойства дифференцирования. Константу можно выносить за знак производной, производная суммы (разности) равна сумме (разности) производных:

$$(cf(x))' = cf'(x),$$
$$(f(x) \pm g(x))' = f'(x) \pm g'(x).$$

(Подразумевается, что производные $f'(x)$ и $g'(x)$ существуют.) Например, $(2\sin x)' = 2\cos x$, $(3x^2 + 4\cos x)' = 6x - 4\sin x$ и т.д. Пользуясь линейностью, можно найти производную логарифма по произвольному основанию. Учитывая, что $\log_a x = \dfrac{\ln x}{\ln a}$ $(x > 0,\ a > 0)$, имеем

$$(\log_a x)' = \frac{1}{\ln a}(\ln x)' = \frac{1}{x \ln a}.$$

2°. Производная произведения. Пусть существуют производные $f'(x)$ и $g'(x)$. Тогда

$$(f(x) \cdot g(x))' = f'(x)g(x) + f(x)g'(x)$$

(производная произведения есть производная первой функции, умноженная на вторую, плюс производная второй, умноженная на первую).

Доказательство. По определению

$$(f(x) \cdot g(x))' = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x)g(x)}{\Delta x}.$$

Прибавим и вычтем в числителе величину $f(x)g(x + \Delta x)$:

$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x)g(x+\Delta x) + f(x)g(x+\Delta x) - f(x)g(x)}{\Delta x}.$$

Представим дробь в виде суммы двух дробей:

$$\lim_{\Delta x \to 0} g(x+\Delta x)\frac{f(x+\Delta x) - f(x)}{\Delta x} + \lim_{\Delta x \to 0} f(x)\frac{g(x+\Delta x) - g(x)}{\Delta x}.$$

В первом пределе $g(x + \Delta x)$ стремится к $g(x)$, а дробь – к $f'(x)$, во втором пределе дробь стремится к $g'(x)$. Итак,

$$(f(x) \cdot g(x))' = f'(x)g(x) + f(x)g'(x),$$

что и требовалось доказать.

Примеры.

1. $f(x) = x \sin x$: $f'(x) = (x)' \cdot \sin x + x \cdot (\sin x)' = 1 \cdot \sin x + x \cdot \cos x = \sin x + x \cos x$.

8 Зак. № 2435 Турецкий

2. Вспомним известные нам формулы: $(x)' = 1 = 1 \cdot x^0$, $(x^2)' = 2x = 2 \cdot x^1$. Они наталкивают на мысль, что производную степени x, т.е. функции x^n, можно получить так: показатель степени n умножить на x в степени на единицу меньше:

$$(x^n)' = nx^{n-1}.$$

Докажем эту формулу по индукции, пользуясь формулой для производной произведения. Итак, формула справедлива при $n = 1$ и $n = 2$. Пусть она справедлива при некотором n. Проверим, верна ли она для $n + 1$:

$$(x^{n+1})' = (x \cdot x^n)' = 1 \cdot x^n + x \cdot (nx^{n-1}) = x^n + nx^n = (n+1)x^n,$$

формула верна. Например, $(x^3)' = 3x^2$, $(x^4)' = 4x^3$ и т.д. Позже мы убедимся, что данная формула верна не только для целых степеней x.

3°. Производная частного. Пусть существуют производные $f'(x)$ и $g'(x)$, причем $g(x) \neq 0$ в окрестности точки x. Тогда

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}.$$

Найдем, например, производную тангенса:

$$(\operatorname{tg} x)' = \left(\frac{\sin x}{\cos x}\right)' = \frac{(\sin x)'\cos x - \sin x(\cos x)'}{\cos^2 x} =$$

$$= \frac{\cos x \cdot \cos x - \sin x \cdot (-\sin x)}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}.$$

Мы воспользовались известной формулой $\sin^2 x + \cos^2 x = 1$. Итак,

$$(\operatorname{tg} x)' = \frac{1}{\cos^2 x}.$$

Аналогично вычисляется производная котангенса:

$$(\operatorname{ctg} x)' = -\frac{1}{\sin^2 x}.$$

4°. Производная сложной функции. Пусть существуют производная $g'(x)$ в точке x и производная $f'(y)$ в точке $y = g(x)$. Тогда

$$(f(g(x)))' = f'(g(x))g'(x),$$

т.е. производная внешней функции в точке $g(x)$ умножается на производную внутренней функции в точке x.

Примеры.

1. $y = (2x+5)^2$: внешняя функция – квадрат $f(y) = y^2$, аргументом которого является внутренняя функция $g(x) = 2x + 5$. Производная квадрата – удвоенный аргумент, т.е. $f'(g(x)) = 2(2x+5)$. Производная аргумента $(2x+5)' = (2x)' + (5)' = 2 + 0 = 2$. Итак, $y' = 2(2x+5) \cdot 2 = 4(2x+5)$. Рассуждая аналогично при вычислении производной функции $y = f(ax+b)$, получаем $y' = af'(ax+b)$. Например, $(\sin(3x-1))' = 3\cos(3x-1)$.

2. $y = \sin x^2$: внешняя функция – синус, внутренняя – квадрат. Значит,

$$y' = \cos x^2 \cdot 2x = 2x\cos x^2.$$

3. $y = (\sin x)^2$: внешняя функция – квадрат, внутренняя – синус. Поэтому

$$y' = 2\sin x \cos x = \sin 2x.$$

5°. Производная обратной функции. Пусть существует не равная нулю производная функции $f(y)$ в точке $y = f^{-1}(x)$. Тогда

$$(f^{-1}(x))' = \frac{1}{f'(f^{-1}(x))};$$

чтобы найти производную обратной функции, нужно найти производную прямой функции, подставить в нее в качестве аргумента значение $f^{-1}(x)$ и разделить единицу на полученную величину.

Примеры.

1. $y = \arcsin x$. Функция арксинуса обратна функции синуса. Производная прямой функции есть косинус. Значит, по формуле

$$(\arcsin x)' = \frac{1}{\cos(\arcsin x)}.$$

Формулу можно преобразовать. Обозначим $\arcsin x = \alpha$. Тогда по определению арксинуса $\sin\alpha = x$ и, стало быть, $\cos\alpha = \sqrt{1-x^2}$. Итак,

$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}.$$

Аналогично находится и производная арккосинуса:

$$(\arccos x)' = -\frac{1}{\sqrt{1-x^2}}.$$

2. $y = a^x$, $a > 0$. Эта показательная функция обратна логарифму по основанию a. Воспользуемся формулой для производной обрат-

ной функции. Производная прямой функции есть $(\log_a x)' = \dfrac{1}{x \ln a}$. Подставим в нее в качестве аргумента величину $f^{-1}(x) = a^x$, получим $\dfrac{1}{a^x \ln a}$. Затем разделим единицу на полученное выражение:

$$(a^x)' = a^x \ln a.$$

Например, $(3^x)' = 3^x \ln 3$. Для экспоненты получим

$$(e^x)' = e^x,$$

так как $\ln e = 1$. Итак, производная экспоненты есть сама экспонента! Это единственная функция, обладающая таким свойством.

Вернем один долг: найдем производную степенной функции $y = x^a$ с произвольным, а не обязательно натуральным показателем степени. Такая функция не всегда определена при отрицательных x (например, при $a = {}^1/_2$ имеем $y = \sqrt{x}$). Поэтому будем полагать $x > 0$. Заметим, что $e^{\ln x} = x$. Поэтому степенную функцию можно представить в виде

$$x^a = e^{a \ln x}.$$

Получилась сложная функция: наружная функция – экспонента, аргумент – функция $a \ln x$. Производная экспоненты – она сама (не забудем, что аргументом ее является $a \ln x$), производная внутренней функции равна $(a \ln x)' = a/x$. Тогда по формуле производной сложной функции имеем

$$(x^a)' = e^{a \ln x}\left(\frac{a}{x}\right) = x^a\left(\frac{a}{x}\right) = ax^{a-1}.$$

В самом деле,

$$(x^a)' = ax^{a-1}$$

при любой степени a.

Например, производная для $y = \sqrt{x} = x^{1/2}$

$$y' = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}},$$

для $y = 1/x = x^{-1}$

$$y' = (-1)x^{-2} = -\frac{1}{x^2}$$

и т.д.

Дополним нашу таблицу производных:

$f(x)$	$f'(x)$
x^a, $x > 0$	ax^{a-1}
$\log_a x$, $x > 0$, $a > 0$	$\dfrac{1}{x \ln a}$
a^x, $a > 0$	$a^x \ln a$
$\operatorname{tg} x$	$\dfrac{1}{\cos^2 x}$
$\operatorname{ctg} x$	$-\dfrac{1}{\sin^2 x}$
$\arcsin x$	$\dfrac{1}{\sqrt{1-x^2}}$
$\arccos x$	$-\dfrac{1}{\sqrt{1-x^2}}$

Правила вычисления производных позволяют вычислять дифференциалы:

$$\boxed{\begin{gathered} d(cf) = c\,df \\ d(f \pm g) = df \pm dg \\ d(fg) = f\,dg + g\,df \\ d\left(\frac{f}{g}\right) = \frac{g\,df - f\,dg}{g^2} \end{gathered}}$$

Например,

$$d(2e^x + \cos x) = 2e^x dx - \sin x\, dx = 2d(e^x) + d(\cos x);$$

$$d(x \sin x) = (x \sin x)'dx = (1 \cdot \sin x + x \cos x)dx = \\ = \sin x\, dx + x\, d(\sin x).$$

9.2. Исследование функций с помощью производных

9.2.1. Монотонность

Производная – мощный инструмент для исследования числовых функций. В частности, по знаку производной можно определить, возрастает функция или убывает. Интуитивно ясно, что к графику возрастающей функции нельзя провести касательную, имеющую отрицательный наклон (рис. 9.5). Стало быть, производная возрастающей функции (равная угловому коэффициенту касательной) должна быть положительной. То же и для убывающей функции: все касательные к графику таких функций имеют отрицательный наклон.

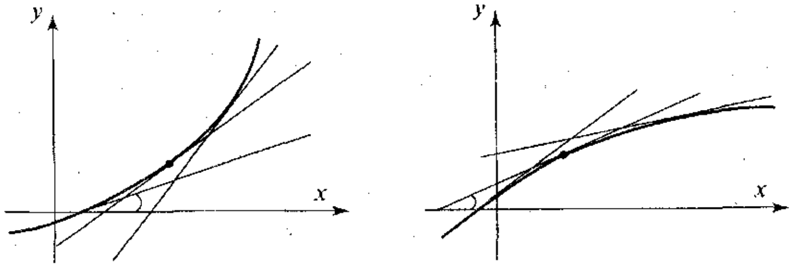


Рис. 9.5

Эти факты формулируются в виде следующей теоремы.

Теорема 9.1. *Пусть функция $f(x)$ непрерывна на отрезке $[a, b]$ и имеет положительную (отрицательную) производную на (a, b). Тогда функция возрастает (убывает) на отрезке $[a, b]$.*

Итак, получены критерии возрастания и убывания функции. С помощью этих критериев можно выделять участки монотонности. Например, $x' = 1 > 0$ при всех x и функция $y = x$ строго возрастает на всей числовой оси (рис. 9.6). Для функции $y = x^2$ производная $y' = 2x$ отрицательна при $x < 0$ и положительна при $x > 0$. Значит, функция убывает на участке $(-\infty, 0)$ и возрастает на участке $(0, +\infty)$ (рис. 9.7).

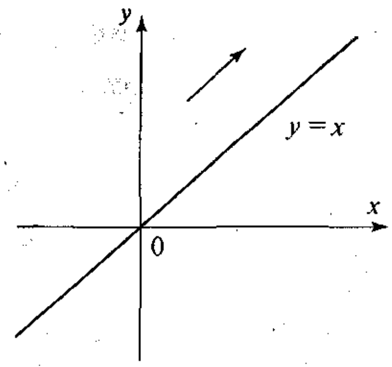


Рис. 9.6

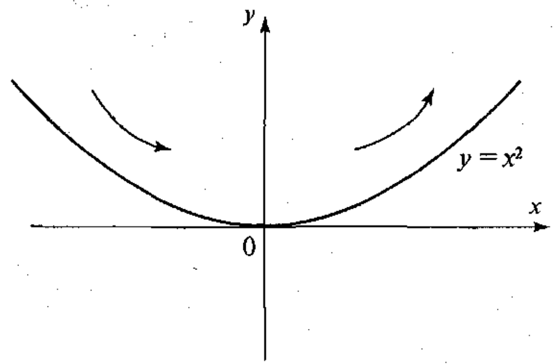


Рис. 9.7

9.2.2. Максимумы и минимумы

Точка x_0 называется точкой ***локального максимума*** функции $f(x)$, если найдется такая окрестность точки x_0, в которой $f(x)$ определена, что для любой точки x из этой окрестности выполняется

$$f(x) \leqslant f(x_0),$$

и ***локального минимума***, если

$$f(x) \geqslant f(x_0).$$

Локальность означает, что в данной точке достигается наибольшее (наименьшее) значение среди всех точек вблизи точки x_0. На рис. 9.8 в точках x_1 и x_3 – локальные минимумы, а в точке x_2 – максимум. Локальный максимум или минимум называется *локальным экстремумом*.

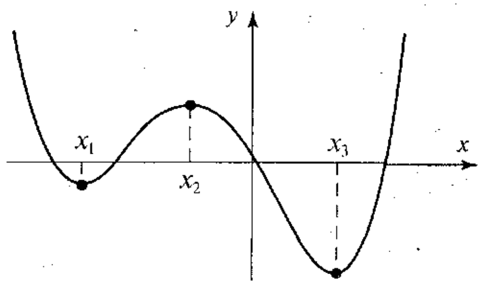


Рис. 9.8

В следующей теореме формулируется необходимое условие экстремума.

Теорема 9.2 (теорема Ферма). *Если функция $f(x)$ достигает в точке x_0 локального экстремума и в этой точке существует производная f, то $f'(x_0) = 0$.*

Д о к а з а т е л ь с т в о. Предположим, что в точке x_0 – локальный максимум. Пусть $x = x_0 + \Delta x$ – произвольная точка из окрестности x_0. Тогда по определению максимума $\Delta f = f(x) - f(x_0) \leqslant 0$. Если x лежит справа от x_0, то $\Delta x > 0$. Тогда $\Delta f/\Delta x \leqslant 0$. По условию производная в точке x_0, т.е. предел этого отношения, существует. Поэтому

$$f'(x_0) \leqslant 0.$$

Для точек, лежащих слева, $\Delta x < 0$ и поэтому

$$f'(x_0) \geqslant 0.$$

Из этих двух неравенств и следует, что $f'(x_0) = 0$. Случай локального минимума разбирается аналогично.

Геометрически это означает, что касательная в точках максимума и минимума (если она существует) параллельна оси x (рис. 9.9). В самом деле, функция возрастает слева от точки максимума (касательные имеют положительный наклон) и убывает справа (наклон касательных отрицательный) (см. рис. 9.9, *а*). Поскольку производная, а значит и касательная, в точке x_0 существует, она неизбежно будет параллельна оси абсцисс. В случае минимума ситуация аналогична (см. рис. 9.9, *б*).

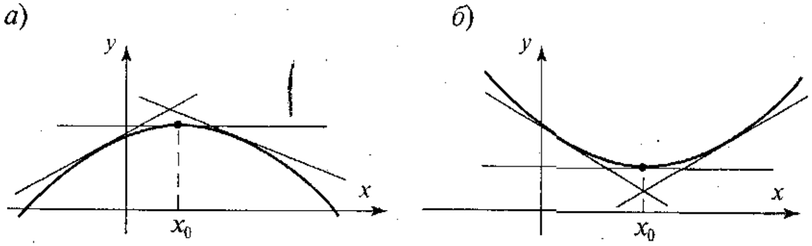


Рис. 9.9

Пьер Ферма (1601–1665). Французский математик. Родился в Бомон-де-Ломань. Учился в школе ордена кордельеров в Бомоне, затем слушал право в Тулузском университете. С 1631 г. – советник кассационной палаты Тулузского парламента. Работа советника в парламенте не мешала Ферма заниматься математикой. Постепенно он приобрел славу одного из первых математиков Франции, хотя и не писал книг (научных журналов еще не было). Его работы стали известны благодаря переписке с коллегами. Он соперничал с Рене Декартом в создании аналитической геометрии, общих методов решения задач на максимум и минимум. Его приемы построения касательных к кривым, вычисления площадей криволинейных фигур проложили дорогу к созданию дифференциального и интегрального исчисления. Фактически не определив производной, он вывел алгоритм вычисления скорости, представляющий собой предельный переход в отношении приращения пути к приращению времени. С переписки Пьера Ферма и Блеза Паскаля отсчитывает свою историю теория вероятностей. Имя Ферма носит основной принцип геометрической оптики, в силу которого свет в неоднородной среде выбирает путь, занимающий наименьшее время (впрочем, Ферма считал, что скорость света бесконечна, и формулировал принцип более туманно). Однако больше всего прославили Ферма работы по теории чисел.

Одна из самых таинственных и фантастических историй в математике связана с так называемой великой теоремой Ферма. На полях книги древнегреческого математика Диофанта Ферма оставил заметку: «Невозможно разложить куб на два куба... и вообще никакую степень, большую квадрата, на две степени с тем же показателем. Я открыл этому поистине чудесное доказательство, но эти поля для него слишком узки». Другими словами, уравнение $x^n + y^n = z^n$ при $n > 2$ не имеет решений в натуральных числах x, y, z. То, что Ферма не оставил доказательства, никого не удивило – он почти не оставил доказательств своих арифметических теорем. Однако с тех пор на протяжении трехсот лет многие выдающиеся умы пытались доказать эту теорему. Она была доказана в некоторых частных случаях (Эйлером – для $n = 4$ и впоследствии для $n = 3$, Лежандром и Дирихле – для $n = 5$). Затем математики поняли, что теорему Ферма достаточно доказать лишь для простых показателей n. Серьезные исследования великой теоремы связаны с именем немецкого математика Э. Куммера (1810–1893), посвятившего доказательству всю жизнь и рассмотревшего много частных случаев. В XX в. вся мощь современной математики в сочетании с использованием компьютера была брошена на доказательство теоремы Ферма, и в последнее время появились обнадеживающие сообщения, что доказательство, возможно, найдено.

Ферма сформулировал также много других утверждений о числах. С работ Ферма началась новая математическая наука – теория чисел.

Вернемся к теореме Ферма. Условие равенства нулю производной есть лишь н е о б х о д и м о е (но не достаточное) условие экстремума по двум причинам:

- во-первых, в точке экстремума может не быть производной. Например, для функции $y = |x|$ точка $x = 0$ есть точка локального минимума (см. рис. 9.1). Однако в этой точке нет производной и нельзя провести касательную;

- во-вторых, точка, производная в которой равна нулю, тем не менее может не быть точкой экстремума. К примеру, функция $y = x^3$ имеет в точке $x = 0$ нулевую производную, но эта точка не является ни точкой максимума, ни точкой минимума: в любой окрестности точки 0 есть точки с меньшими значениями (слева) и бо́льшими значениями (справа) (рис. 9.10).

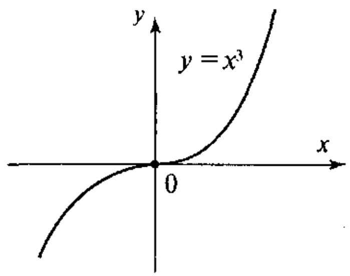


Рис. 9.10

Точки, в которых производная равна нулю, называются *стационарными* или точками, «подозрительными» на экстремум. В них может быть экстремум, а может и не быть – требуется дополнительное исследование.

На самом деле, вовсе не обязательно, чтобы производная существовала в самой точке x_0. Как видно из примера с модулем, достаточно, чтобы производная была справа и слева от точки и имела разные знаки (разный наклон касательных).

Теорема 9.3. *Если функция $f(x)$ непрерывна в окрестности точки x_0 и имеет производную $f'(x) \leqslant 0$ справа от x_0 и $f'(x) \geqslant 0$ слева от x_0, то x_0 есть точка максимума (если знаки чередуются в обратном порядке, то x_0 есть точка минимума).*

В теореме 9.3 сформулированы уже д о с т а т о ч н ы е условия экстремума: если они выполняются, то в данной точке – локальный экстремум.

Примеры.

1. Исследуем функцию $f(x) = 2x^3 - 9x^2 + 12x - 1$. Стационарные точки найдем из условия $f'(x) = 6x^2 - 18x + 12 = 0$, откуда $x_1 = 1$, $x_2 = 2$. Производную можно записать в виде $f'(x) = 6(x - 1)(x - 2)$. Легко заметить, что слева от точки $x_1 = 1$ производная положительна, а справа – отрицательна. Значит, это точка локального максимума. Аналогично проверяется, что точка $x_2 = 2$ – точка локального минимума. Полученные результаты позволяют выделить участки монотонности: при $x < 1$ функция возрастает, на интервале $(1, 2)$ – убывает, а при $x > 2$ – снова возрастает. Обычно результаты исследования функции оформляются в виде таблицы. В нашем примере получим следующую таблицу:

Интервал	$(-\infty, 1)$	1	$(1, 2)$	2	$(2, \infty)$
$f'(x)$	+	0	–	0	+
$f(x)$	↗	max	↘	min	↗

2. Исследуем аналогично функцию $f(x) = x^3 - 3x^2 + 3x + 2$. Производная $f'(x) = 3x^2 - 6x + 3 = 3(x - 1)^2$, значит, единственная стационарная точка $x_0 = 1$. Однако при переходе через эту точку производная не меняет знака (остается положительной). Следовательно, локального экстремума в этой точке нет.

Более сильные условия содержатся в следующей теореме.

Теорема 9.4. *Пусть функция $f(x)$ дважды дифференцируема в точке x_0 и $f'(x_0) = 0$. Если при этом $f''(x_0) < 0$, то x_0 – точка максимума, а если $f''(x_0) > 0$, то точка минимума.*

Например, для $f(x) = 2x^3 - 9x^2 + 12x - 1$ из предыдущего примера $f''(x) = 12x - 18$ и $f''(1) = -6 < 0$, $f''(2) = 6 > 0$. Стало быть, $x = 1$ – точка максимума, а $x = 2$ – минимума.

Большой практический интерес представляют задачи поиска максимума и минимума функций на заданном промежутке (так называемого *условного экстремума*). Вторая теорема Вейерштрасса (см. параграф 8.2) гарантирует, что непрерывная функция достигает своего максимального и минимального значения на отрезке $[a, b]$. Если функция еще и дифференцируема на интервале (a, b), поиск экстремумов сильно упрощается. Ясно, что максимум (минимум) может в этом случае достигаться либо в одной из стационарных точек, либо на одном из концов отрезка. Поэтому если функция непрерывна на $[a, b]$ и дифференцируема на (a, b), максимум и минимум на отрезке $[a, b]$ можно найти следующим образом:

- определить все стационарные точки $x_1, x_2, \ldots, x_k$, лежащие на $[a, b]$, и значения функции $f(x_1), f(x_2), \ldots, f(x_k)$ в этих точках;
- вычислить значения $f(a), f(b)$ функции на концах отрезка;
- найти максимальное и минимальное из чисел $f(a), f(b), f(x_1), f(x_2), \ldots, f(x_k)$.

Ясно, что максимум (минимум) функции на отрезке может достигаться и в нескольких точках. Найдем, например, максимум и минимум функции $f(x) = \frac{1}{2}\cos 2x + \sin x$ на отрезке $[0, \pi/2]$. Стационарные точки находим из уравнения $f'(x) = -\sin 2x + \cos x = 0$. Пользуясь формулой синуса двойного угла, получаем $\cos x(1 - 2\sin x) = 0$. Легко обнаружить, что на отрезке $[0, \pi/2]$ имеется две стационарные точки $x_1 = \pi/6$ и $x_2 = \pi/2$ (в данном случае стационарная точка совпала с правым концом отрезка). Найдем значения функции в стационарных и граничных точках отрезка: $f(0) = 0{,}5$; $f(\pi/6) = {}^1/_4 + {}^1/_2 = 0{,}75$; $f(\pi/2) = -{}^1/_2 + 1 = 0{,}5$. Видно, что максимальное значение функции равно 0,75 и достигается в точке $x = \pi/6$, а минимальное равно 0,5 и достигается в двух точках, а именно на обоих концах отрезка.

Задачи

1. Найти производные функций:

а) $y = 3x^2 - 5x + 1$; б) $y = x^4 - \frac{1}{3} + 2{,}5x^2 - 0{,}3x + 0{,}1$;

в) $y = \sqrt[3]{x} + \sqrt[3]{2}$; г) $y = 2\sqrt{x} - \frac{1}{x} + \sqrt[4]{3}$;

д) $y = (x^2 - 3x + 3)(x^2 + 2x - 1)$; е) $y = \left(\sqrt{x} + 1\right)\left(\frac{1}{\sqrt{x}} - 1\right)$;

ж) $y = \frac{x+1}{x-1}$; з) $y = \frac{2}{x^3 - 1}$;

и) $y = \sin x + \cos x$; к) $y = \cos^2 x$;

л) $y = \sin 3x$; м) $y = 3\sin(3x + 5)$;

н) $y = \cos x - \frac{1}{3}\cos^3 x$; о) $y = \frac{x}{1 - \cos x}$;

п) $y = \ln(x^2)$; р) $y = \operatorname{tg} e^x$;

с) $y = \sin x \cdot e^{\cos x}$; т) $y = e^{-x^2} \ln x$.

2. Найти угловой коэффициент касательной, проведенной к параболе $y = x^2$:

а) в начале координат;

б) в точке (3, 9);

в) в точке (–2, 4).

3. В каких точках угловой коэффициент касательной к кубической параболе $y = x^3$ равен 3?

4. На параболе $y = x^2$ взяты две точки с абсциссами $x_1 = 1$ и $x_2 = 2$. Через эти точки проведена секущая. В какой точке параболы касательная к ней будет параллельна проведенной секущей?

5. Вычислить приближенно, используя формулу конечных приращений: а) $1{,}02^3$; б) $\sqrt{0{,}97}$; в) $\frac{1}{1{,}04}$.

6. Найти вторую производную функций:

а) $y = x^2 - 3x + 2$;

б) $y = \cos^2 x$.

7. Показать, что точка $x = 0$ есть точка минимума функции

$$y = 3x^4 - 4x^3 + 12x^2 + 1.$$

8. Показать, что функция

$$y = 2x^3 + 3x^2 - 12x + 1$$

убывает на интервале (–2, 1).

9. Найти интервалы монотонности функции

$$y = x^3 - 3x^2 - 9x + 14$$

и построить по точкам ее график в интервале (–2, 4).

10. Найти максимумы и минимумы функции

$$y = 2x^3 - 3x^2.$$

Глава 10
ОСНОВЫ ИНТЕГРАЛЬНОГО ИСЧИСЛЕНИЯ

Существует операция, обратная операции дифференцирования, – интегрирование функции. Любые обратные операции в математике устроены обычно сложнее, чем прямые. Хорошим примером такого различия служат операции умножения и деления. Мы можем перемножать любые целые числа: в результате снова получится целое число. Однако при делении целого числа на целое получается чаще всего не целое число. Значит, операция деления выводит за пределы множества целых чисел. Более того, мы не можем делить на нуль. Интегрирование функции (операция, обратная нахождению производной) значительно сложнее дифференцирования. Если вычисление производных подчиняется четким правилам, то вычисление интегралов во многом можно считать искусством. Мы ознакомимся лишь с простейшими приемами интегрирования.

10.1. Первообразная. Неопределенный интеграл

10.1.1. Основные определения

Пусть на интервале (a, b) задана непрерывная функция $f(x)$. Функция $F(x)$ называется ***первообразной*** функцией для f на (a, b), если для всех $x \in (a, b)$ выполняется

$$F'(x) = f(x).$$

Итак, первообразная – это такая функция, производная которой равна данной функции $f(x)$. Вспоминая пример с камушком, брошенным с Пизанской башни, можно сказать, что функция пути есть первообразная функции скорости, так как скорость – это производная пути. Зная формулы дифференцирования, мы можем подобрать первообразные для некоторых простых функций. Например, первообразная функции $f(x) \equiv 1$ есть функция $F(x) = x$: действительно, $(x)' = 1$. Первообразная косинуса есть синус: $(\sin x)' = \cos x$ и т.д. Вообще говоря, найти первообразную какой-то функции достаточно трудно.

Одна из главных проблем в математике – проблема существования: определив какой-либо объект, мы задаемся вопросом о том, существует ли он в достаточно большом числе случаев? Что касается первообразной, ответ дает следующая теорема, которую мы примем к сведению без доказательства.

Теорема 10.1. *Всякая непрерывная на (a, b) функция имеет на этом интервале непрерывную первообразную.*

Обратим внимание на некоторые свойства первообразной. Заметим, например, что не только x есть первообразная единицы. Эта функция имеет и другие первообразные: например, $F(x) = x + 1$, $F(x) = x - 2$ и т.д. Вообще, любая функция вида $F(x) = x + C$, где C – постоянное число, будет первообразной функции $f(x) \equiv 1$. Это связано с тем, что производная константы равна нулю. Справедливо следующее общее свойство.

∇**1.** *Если $F(x)$ – первообразная для функции $f(x)$ на (a, b) и C – постоянная, то функция $F(x) + C$ также является первообразной для f.*

Действительно,

$$(F(x) + C)' = F'(x) + C' = f(x) + 0 = f(x).$$

Таким образом, если мы нашли одну первообразную, то фактически определили бесконечное множество первообразных: прибавив любую константу, мы снова получим первообразную. Однако может ли функция иметь и какие-то другие первообразные? Оказывается, нет. Это следует из второго свойства.

∇**2.** *Пусть $F_1(x)$ и $F_2(x)$ – первообразные для $f(x)$. Тогда они отличаются на (a, b) на некоторую постоянную C.*

В самом деле, зададим функцию $F(x) = F_2(x) - F_1(x)$ и найдем ее производную в произвольной точке $x \in (a, b)$:

$$F'(x) = (F_2(x) - F_1(x))' = F_2'(x) - F_1'(x) = f(x) - f(x) = 0,$$

так как по условию обе функции $F_1(x)$ и $F_2(x)$ есть первообразные для f. Итак, функция $F(x)$ имеет нулевую производную во всех точках интервала (a, b). Значит, эта функция есть константа: $F(x) \equiv C$ и $F_2(x) = F_1(x) + C$.

Из этих двух свойств можно сделать следующий вывод:

> если $F(x)$ есть какая-либо первообразная для $f(x)$ на (a, b), то любая первообразная для f выражается формулой $F(x) + C$, где $C = \text{const}$.

Никаких других первообразных у функции f нет!

Неопределенным интегралом от непрерывной на интервале (a, b) функции $f(x)$ называется произвольная ее первообразная. Неопределенный интеграл обозначают так:

$$\int f(x)dx.$$

Из определения следует, что если $F(x)$ есть некоторая первообразная для f на интервале (a, b), то неопределенный интеграл от f на этом интервале равен

$$\int f(x)dx = F(x) + C.$$

Следующее свойство вытекает из определения.

∇**3.** *Производная от интеграла равна подынтегральной функции и, наоборот, интеграл от производной равен самой функции:*

$$\left(\int f(x)dx\right)' = f(x),$$

$$\int f'(x)dx = f(x) + C.$$

Процесс отыскания неопределенного интеграла называется *интегрированием*. Независимая переменная не обязательно должна обозначаться через x. Мы должны понимать, что $\int f(x)dx$, $\int f(t)dt$ или $\int f(z)dz$ – всего лишь разная запись для одной и той же функции и неважно, как обозначить ее аргумент. Например,

$$\int \sin x dx = \int \sin t dt.$$

Это равенство следует понимать так: если в функцию, стоящую справа, подставить аргумент функции, стоящей слева, получится одна и та же функция.

10.1.2. Таблица простейших интегралов

Функция называется *элементарной*, если ее значение можно найти с помощью конечного числа элементарных арифметических действий: сложения, вычитания, умножения, деления и возведения в степень. Иногда к этому набору добавляют взятие синуса и некоторые другие «школьные» операции, а также взятие функции от функции. По правилам дифференцирования можно найти производную любой элементарной функции. В результате снова непременно получится элементарная функция. Однако первообразная и неопределенный интеграл от элементарной функции могут не быть элементарными. В этом принципиальное отличие интегрирования от дифференцирования. К счастью, многие интересные в математической практике элементарные функции интегрируются в элементарных функциях (т.е. их первообразные есть элементарные функции). Воспользовавшись таблицей основных производных, найдем некоторые простые интегралы.

1. Степенная функция $y = x^a$. При дифференцировании степенной функции получается степенная же функция с показателем степени, на единицу меньшим, и с коэффициентом, равным показателю степени. Ясно, что степень a можно получить, если продифференцировать степенную функцию со степенью $a + 1$: $(x^{a+1})' = (a + 1)x^a$. Теперь, чтобы справа осталась в чистом виде функция x^a, разделим обе части этого равенства на $a + 1$ и воспользуемся линейными свойствами производной:

$$\left(\frac{x^{a+1}}{a+1}\right)' = x^a.$$

Заметим, что это можно проделать лишь при $a + 1 \neq 0$, т.е. при $a \neq -1$. Итак, найдена первообразная степенной функции. Значит, по определению

$$\int x^a dx = \frac{x^{a+1}}{a+1} + C, \quad a \neq -1.$$

В частности,

$$\int x dx = \frac{x^2}{2} + C,$$

$$\int x^2 dx = \frac{x^3}{3} + C,$$

$$\int \frac{1}{x^2} dx = \frac{x^{-1}}{-1} + C = -\frac{1}{x} + C,$$

$$\int \sqrt{x} dx = \frac{x^{3/2}}{3/2} + C = \frac{2}{3} x\sqrt{x} + C,$$

$$\int \frac{1}{\sqrt{x}} dx = \frac{x^{1/2}}{1/2} + C = 2\sqrt{x} + C$$

и т.д. Для интегралов от дробей пользуются более короткой записью:

$$\int \frac{1}{x^2} dx = \int \frac{dx}{x^2}.$$

Осталось рассмотреть случай $a = -1$, т.е. интеграл от функции $y = x^{-1} = \frac{1}{x}$. Для положительных x нам известно, что $(\ln x)' = \frac{1}{x}$. При $x < 0$ определена функция $\ln(-x)$, и по формуле производной сложной функции $(\ln(-x))' = \frac{1}{-x} \cdot (-1) = \frac{1}{x}$. Эти случаи можно объединить в одну формулу:

$$(\ln |x|)' = \frac{1}{x}, \quad x \neq 0.$$

Значит, на любом интервале, не содержащем нуля, выполняется

$$\int \frac{dx}{x} = \ln |x| + C.$$

2. Показательная функция $y = a^x$. Производная показательной функции равна самой этой функции, умноженной на $\ln a$: $(a^x)' = a^x \ln a$. Рассуждая как в предыдущем случае, приходим к равенству

$$\left(\frac{a^x}{\ln a}\right)' = a^x.$$

Поэтому

$$\int a^x dx = \frac{a^x}{\ln a} + C.$$

В частности,

$$\int e^x dx = e^x + C;$$

можно сказать, что с точностью до постоянной интеграл от экспоненты равен ей самой. Итак, еще раз подтверждается уникальность этой функции: не только производная, но и неопределенный интеграл от экспоненты равен самой экспоненте!

3. Тригонометрические функции. С помощью известных производных получаем следующие равенства:

$$\int \sin x\, dx = -\cos x + C,$$

$$\int \cos x\, dx = \sin x + C,$$

$$\int \frac{dx}{\cos^2 x} = \operatorname{tg} x + C.$$

Кроме того, зная производную арксинуса, находим

$$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C.$$

Вот, собственно, и весь (довольно скудный) набор интегралов, которые можно найти, непосредственно используя таблицу производных. Составим таблицу интегралов вместе с соответствующими производными.

Производная	Интеграл
$(x^a)' = ax^{a-1}$	$\int x^a dx = \frac{x^{a+1}}{a+1} + C, \quad a \neq -1$
$(\ln x)' = \frac{1}{x}$	$\int \frac{dx}{x} = \ln\lvert x\rvert + C$
$(a^x)' = a^x \ln a$	$\int a^x dx = \frac{a^x}{\ln a} + C$
$(\sin x)' = \cos x$	$\int \cos x\, dx = \sin x + C$
$(\cos x)' = -\sin x$	$\int \sin x\, dx = -\cos x + C$
$(\operatorname{tg} x)' = \frac{1}{\cos^2 x}$	$\int \frac{dx}{\cos^2 x} = \operatorname{tg} x + C$
$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$	$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C$

10.1.3. Линейные свойства интеграла

Неопределенный интеграл определяется через производную. Поэтому линейные свойства производной наследуются интегралом.

∇4.
$$\int cf(x)dx = c\int f(x)dx + C,$$

$$\int (f(x) \pm g(x))dx = \int f(x)dx \pm \int g(x)dx + C,$$

где C = const. Таким образом, константу можно выносить за знак интеграла, интеграл от суммы (разности) функций равен сумме (разности) интегралов. Разумеется, равенство интегралов следует понимать с точностью до произвольной постоянной. Два линейных свойства, как обычно, можно объединить в одно: интеграл от линейной комбинации функций равен линейной комбинации интегралов. В частности, для двух функций данное свойство выглядит так:

$$\int (c_1 f_1(x) + c_2 f_2(x))dx = c_1\int f_1(x)dx + c_2\int f_2(x)dx + C.$$

Выведем это равенство из линейных свойств производной. По определению неопределенного интеграла слева в этом равенстве стоит какая-то первообразная функции $c_1f_1(x) + c_2f_2(x)$. С другой стороны, имеет место равенство

$$\left(c_1\int f_1(x)dx + c_2\int f_2(x)dx\right)' = c_1\left(\int f_1(x)dx\right)' + c_2\left(\int f_2(x)dx\right)' =$$
$$= c_1f_1(x) + c_2f_2(x),$$

т.е. правая часть равенства без постоянной C также есть первообразная функции $c_1f_1(x) + c_2f_2(x)$. Прибавляя произвольную постоянную, получаем произвольную первообразную, т.е. неопределенный интеграл.

В общем виде линейность интеграла записывается для любого конечного набора функций:

$$\int\left(\sum_{i=1}^{n} c_i f_i(x)\right)dx = \sum_{i=1}^{n} c_i\int f_i(x)dx + C.$$

Например,

$$\int (2x^2 - 3x + 1)dx = \frac{2}{3}x^3 - \frac{3}{2}x^2 + x + C,$$

$$\int\left(5\sin x + \frac{2}{x}\right)dx = -5\cos x + 2\ln\lvert x\rvert + C,$$

$$\int (0{,}5\cos x - 4e^x)dx = 0{,}5\sin x - 4e^x + C$$

и т.д.

10.2. Приемы интегрирования

Линейные свойства интеграла увеличивают наши возможности для вычисления интегралов. Однако мы по-прежнему можем интегрировать весьма ограниченный круг функций. Методы, изложенные в параграфе 10.2, существенно расширяют этот круг.

10.2.1. Метод замены переменной

Начнем с примера. Попытаемся найти интеграл от функции $\cos 2x$. Как же выглядит функция, производная которой равна $\cos 2x$? Ясно, что это должен быть синус, причем аргумент этой функции (синуса) равен $2x$. По правилу дифференцирования сложной функции $(\sin 2x)' = \cos 2x \cdot (2x)' = 2\cos 2x$. Чтобы справа остался $\cos 2x$, разделим обе части этого равенства на 2:

$$\left(\frac{\sin 2x}{2}\right)' = \cos 2x.$$

Из этого равенства ясно, что

$$\int \cos 2x\,\mathrm{d}x = \frac{\sin 2x}{2} + C.$$

Преобразуем исходный интеграл, разделив и домножив его на 2:

$$\int \cos 2x\,\mathrm{d}x = \frac{1}{2}\int \cos 2x \cdot 2\mathrm{d}x.$$

Вглядевшись внимательнее в подынтегральную функцию, можно заметить, что она устроена особым образом: функция представляет собой произведение с л о ж н о й функции ($\cos 2x$) на п р о и з в о д н у ю е е а р г у м е н т а ($(2x)' = 2$). Теперь нам точно известна функция, производная которой совпадает с подынтегральной, это $\sin 2x$. Фактически мы взяли интеграл от косинуса

$$\int \cos t\,\mathrm{d}t = \sin t + C$$

и подставили в него аргумент $2x$. Не забудем, что перед интегралом есть коэффициент $\frac{1}{2}$. Так что в итоге получается уже известный нам результат: $\frac{\sin 2x}{2}$.

Разберемся, как мы действовали. Сначала исходный интеграл мы преобразовали таким образом, чтобы под интегралом оказалось произведение сложной функции на производную аргумента,

т.е. привели к виду $f(\varphi(x))\varphi'(x)$. Затем мы нашли интеграл от внешней функции $f(t)$ и в конце концов подставили в него исходный аргумент $\varphi(x)$. Говорят, что в исходном интеграле осуществлена *замена переменной* $\varphi(x) = t$. Наши действия описываются следующим равенством:

$$\boxed{\int f(\varphi(x))\varphi'(x)\,\mathrm{d}x = \int f(t)\,\mathrm{d}t + C}$$

Данное равенство на первый взгляд парадоксально: ведь слева стоит функция от x, а справа – от t. Это нужно понимать следующим образом: справа сначала вычисляют интеграл от f, а затем в полученную функцию от t подставляют вместо t аргумент $\varphi(x)$. Тогда в обеих частях равенства будет функция от x. Равенство легко доказывается с помощью формулы производной сложной функции. В самом деле, пусть $F(t)$ – первообразная функции $f(t)$, т.е. $F'(t) = f(t)$. Тогда справа стоит сложная функция $F(\varphi(x)) + C$. Ее производная равна

$$(F(\varphi(x)) + C)' = F'(\varphi(x))\,\varphi'(x) = f(\varphi(x))\,\varphi'(x),$$

т.е. как раз подынтегральная функция в левой части равенства. Равенство доказано.

У читателя должен появиться естественный вопрос: зачем делать замену переменной? Это может потребоваться, если интеграл в правой части $\int f(t)\mathrm{d}t$ проще, чем исходный (например, он табличный). В разобранном примере именно так и произошло: мы не знали интеграла от $\cos 2x$, но хорошо знали интеграл от $\cos t$. Итак, если встречается подынтегральная функция специального вида $f(\varphi(x))\,\varphi'(x)$, можно попытаться сделать замену переменной $\varphi(x) = t$ – вдруг полученный интеграл окажется проще?

Примеры.

1. **Линейная замена переменной.** Пусть задана функция f, а требуется найти интеграл от сложной функции $f(ax + b)$, т.е.

$$\int f(ax+b)\,\mathrm{d}x.$$

Поступим как в примере с косинусом: разделим и домножим этот интеграл на a:

$$\int f(ax+b)\,\mathrm{d}x = \frac{1}{a}\int f(ax+b)\,a\mathrm{d}x.$$

Теперь подынтегральная функция стала именно такой, что можно применить формулу замены переменной: это произведение сложной функции $f(ax + b)$ на производную ее аргумента $(ax + b)' = a$. Делая замену переменной $ax + b = t$, приходим к равенству

$$\int f(ax+b)\,\mathrm{d}x = \frac{1}{a}\int f(t)\,\mathrm{d}t + C.$$

Разобранный ранее пример – это частный случай полученной формулы.

Приведем примеры, когда такая замена переменной приводит к табличному интегралу. Это уже известный нам интеграл от косинуса

$$\int \cos(ax+b)\,dx = \frac{1}{a}\int \cos t\,dt = \frac{\sin t}{a} = \frac{\sin(ax+b)}{a} + C$$

(на практике произвольную постоянную обычно пишут в конце преобразований). Аналогично

$$\int \sin(ax+b)\,dx = -\frac{\cos(ax+b)}{a} + C,$$

$$\int (ax+b)^n\,dx = \frac{(ax+b)^{n+1}}{(n+1)a} + C,$$

$$\int e^{ax+b}\,dx = \frac{e^{ax+b}}{a} + C$$

и т.д.

В частности,

$$\int \sin 3x\,dx = -\frac{\cos 3x}{3} + C,$$

$$\int (2x+1)^3\,dx = \frac{(2x+1)^4}{8} + C,$$

$$\int \frac{1}{2x-1}\,dx = \frac{\ln|2x-1|}{2} + C.$$

2. $\int 2e^{x^2} x\,dx = \int e^{x^2}(x^2)'\,dx = \int e^t\,dt = e^t = e^{x^2} + C.$

Подынтегральная функция имеет требуемый вид, поэтому мы сделали замену переменной $x^2 = t$. Всегда можно убедиться в правильности вычисления интеграла, продифференцировав полученную функцию. При этом должна получиться подынтегральная функция. Опуская произвольную постоянную (ее п[redacted]изводная равна нулю), имеем

$$(e^{x^2})' = e^{x^2}(x^2)' = e^{x^2}\cdot 2x = 2e^{x^2}x,$$

т.е. интеграл найден верно.

3. $\int \frac{x}{\sqrt{1-x^2}}\,dx = -\frac{1}{2}\int \frac{-2x}{\sqrt{1-x^2}}\,dx = -\frac{1}{2}\int \frac{1}{\sqrt{1-x^2}}(1-x^2)'\,dx =$

$$= -\frac{1}{2}\int \frac{dt}{\sqrt{t}} = -\frac{1}{2}(2\sqrt{t}) = -\sqrt{t} = -\sqrt{1-x^2} + C.$$

Делением и умножением на -2 привели интеграл к нужному виду, воспользовавшись тем, что $(1-x^2)' = -2x$, а затем сделали замену $1 - x^2 = t$. Полученный интеграл от $1/\sqrt{t}$ – это табличный интеграл от t^a при $a = -1/2$. Он равен $\frac{t^{1/2}}{1/2} = 2\sqrt{t}$. После этого, как и следует, подставили вместо t исходный аргумент $1 - x^2$.

Проверка подтверждает правильность вычислений:

$$\left(-\sqrt{1-x^2}\right)' = -\frac{1}{2\sqrt{1-x^2}}(1-x^2)' = -\frac{-2x}{2\sqrt{1-x^2}} = \frac{x}{\sqrt{1-x^2}}.$$

4. $\int \operatorname{tg} x\,dx = \int \frac{\sin x}{\cos x}\,dx = -\int \frac{-\sin x}{\cos x}\,dx = -\int \frac{1}{\cos x}(\cos x)'\,dx =$

$$= -\int \frac{dt}{t} = -\ln|t| = -\ln|\cos x| + C.$$

Делением и умножением на -1 привели интеграл к нужному виду, затем сделали замену $\cos x = t$, получили табличный интеграл и заменили в нем t на $\cos x$.

Создается впечатление, что почти всегда требуется разделить и домножить подынтегральную функцию на какое-нибудь число, чтобы привести ее к удобному для применения формулы виду. Действительно, этот прием часто приводит к успеху. Однако как догадаться, на какое число делить и умножать? Нельзя ли предложить процедуру, позволяющую определять этот коэффициент автоматически? Для этого мы применим понятие дифференциала функции. Рассмотрим еще раз формулу замены переменной. Под знаком интеграла находится выражение $\varphi'(x)dx$, представляющее собой не что иное, как дифференциал функции $\varphi(x)$. Поэтому

$$\int f(\varphi(x))\varphi'(x)\,dx = \int f(\varphi(x))\,d\varphi(x) + C.$$

Теперь, просто введя обозначение $\varphi(x) = t$, приходим к той же самой формуле замены [redacted]ременной. Фактически нам удалось исходное подынтегральное[redacted]ражение записать через t и dt.

Предположим теперь, что подынтегральная функция не соответствует нужному виду. Попытаемся тем не менее сделать замену переменной $\varphi(x) = t$. Замена будет удачной, если мы сможем выразить подынтегральное выражение только через новую независимую переменную t и ее дифференциал dt (как в случае, когда под знаком интеграла стоит $f(\varphi(x))\varphi'(x)$). Рассмотрим, например, интеграл

$$\int e^{x^2} x\,dx.$$

В нем есть группа $x\mathrm{d}x$, почти совпадающая с дифференциалом $\mathrm{d}(x^2)$. Это наталкивает на мысль, что замена $x^2 = t$ может привести к успеху. При такой замене $e^{x^2} = e^t$, а $\mathrm{d}t = 2x\mathrm{d}x$, откуда $x\mathrm{d}x = \mathrm{d}t/2$. Итак, мы выразили подынтегральное выражение через t и $\mathrm{d}t$: $e^{x^2}x\mathrm{d}x = e^t\mathrm{d}t/2$. Коэффициент $^1/_2$ выносим за знак интеграла и получаем

$$\int e^{x^2} x\,\mathrm{d}x = \frac{1}{2}\int e^t\,\mathrm{d}t = \frac{1}{2}e^t = \frac{e^{x^2}}{2} + C.$$

Заметьте, что мы не приводили подынтегральную функцию к специальному виду, а множитель $^1/_2$ нашли автоматически из выражения для дифференциала новой переменной. Таким образом, наш подход более или менее формализовался: после замены переменной требуется выразить подынтегральное выражение только через t и $\mathrm{d}t$. Тогда, возможно, новый интеграл окажется знакомым или более легким для вычисления.

10.2.2. Метод интегрирования по частям

Прежде всего вспомним формулу производной произведения:

$$(u(x)v(x))' = u'(x)v(x) + u(x)v'(x)$$

(мы переписали формулу в несколько других, общепринятых обозначениях). Ясно, что первообразные функций, стоящих в левой и правой части этого равенства, совпадают с точностью до произвольной постоянной. Поэтому

$$\int (u(x)v(x))'\,\mathrm{d}x = \int (u'(x)v(x) + u(x)v'(x))\,\mathrm{d}x + C.$$

Интеграл от производной равен самой функции, а интеграл от суммы равен сумме интегралов. Значит,

$$u(x)v(x) = \int u'(x)v(x)\,\mathrm{d}x + \int u(x)v'(x)\,\mathrm{d}x + C.$$

Равенство можно переписать в виде

$$\boxed{\int u(x)v'(x)\,\mathrm{d}x = u(x)v(x) - \int u'(x)v(x)\,\mathrm{d}x + C}$$

Эта формула называется формулой *интегрирования по частям.* Итак, если под интегралом стоит произведение функции на производную некоторой другой функции, такой интеграл можно пытаться найти, пользуясь формулой интегрирования по частям. Это разумно делать, если исходный интеграл от $u(x)v'(x)$ найти трудно, а интеграл от $u'(x)v(x)$ – табличный или более легкий для вычисления. Учитывая, что $v'(x)\,\mathrm{d}x = \mathrm{d}v$, а $u'(x)\,\mathrm{d}x = \mathrm{d}u$, формулу интегрирования по частям можно записать короче:

$$\int u\,\mathrm{d}v = uv - \int v\,\mathrm{d}u + C.$$

Примеры.

1. $\int x\sin x\,\mathrm{d}x$.

Если под интегралом стоит произведение функций, важно определить, какую из них считать функцией $u(x)$, а какую – производной $v'(x)$. От этого многое зависит: ситуация может либо упроститься (получится более простой интеграл), либо нет. Примем, например, $\sin x$ за $u(x)$, а x за $v'(x)$. Раз $v'(x) = x$, легко определить, что $v(x) = x^2/2$. Применив формулу интегрирования по частям, получим

$$\int x\sin x\,\mathrm{d}x = \sin x\frac{x^2}{2} - \int \frac{x^2}{2}\cos x\,\mathrm{d}x + C.$$

Мы учли, что $u'(x) = (\sin x)' = \cos x$. Интеграл в правой части нисколько не проще, чем исходный. В данном случае интегрирование по частям не дает результата. Попробуем обозначить наоборот: $u(x) = x$, $v'(x) = \sin x$. Тогда $v(x) = -\cos x$, $u'(x) = 1$, и, интегрируя по частям, получаем

$$\int x\sin x\,\mathrm{d}x = x(-\cos x) - \int 1\cdot(-\cos x)\,\mathrm{d}x = -x\cos x + \int \cos x\,\mathrm{d}x.$$

Теперь в правой части стоит табличный интеграл от косинуса, что действительно упрощает вычисление исходного интеграла:

$$\int x\sin x\,\mathrm{d}x = -x\cos x + \sin x + C.$$

Проверим, что интеграл найден правильно:

$$(-x\cos x + \sin x)' = -(\cos x - x\sin x) + \cos x = x\sin x.$$

2. Аналогичный пример:

$$\int x e^x\mathrm{d}x \overset{(u=x,\ v'=e^x)}{=} x e^x - \int e^x\mathrm{d}x = x e^x - e^x = e^x(x-1) + C.$$

3. Иногда функция под интегралом не выглядит явно как произведение. Тем не менее интегрирование по частям можно применять и в таком случае, считая вторым сомножителем единицу. Рассмотрим, например, интеграл

$$\int \arcsin x\,\mathrm{d}x = \int \arcsin x\cdot 1\,\mathrm{d}x.$$

Обозначим $u(x) = \arcsin x$, $v'(x) = 1$. Тогда

$$u'(x) = \frac{1}{\sqrt{1-x^2}},\quad v(x) = x$$

и по формуле интегрирования по частям

$$\int \arcsin x\,dx = x\arcsin x - \int \frac{x}{\sqrt{1-x^2}}dx + C.$$

Интеграл в правой части равенства мы уже находили с помощью замены переменной $1 - x^2 = t$:

$$\int \frac{x}{\sqrt{1-x^2}}dx = -\frac{1}{2}\int \frac{d(1-x^2)}{\sqrt{1-x^2}} = -\frac{1}{2}\int \frac{dt}{\sqrt{t}} =$$

$$= -\sqrt{t} = -\sqrt{1-x^2} + C.$$

Таким образом,

$$\int \arcsin x\,dx = x\arcsin x + \sqrt{1-x^2} + C.$$

Проверку читатель может сделать самостоятельно.

4. Аналогично

$$\int \ln x\,dx \overset{(u=\ln x,\ v'=1)}{=} x\ln x - \int x\frac{1}{x}dx =$$

$$= x\ln x - \int 1\,dx = x\ln x - x + C.$$

Подводя итог, заметим, что с помощью формул замены переменной и интегрирования по частям можно интегрировать достаточно сложные функции. Встретив сложный на первый взгляд интеграл, можно попытаться применить один из этих приемов – а вдруг получится?

10.3. Определенный интеграл Римана

10.3.1. Основные определения

В подпараграфе 10.3.1 определяется новый математический объект, который тоже называется интегралом. Начнем с примера. Эту механическую задачу решал Исаак Ньютон: пусть известна скорость движения тела $v(t)$ на промежутке времени $a \leqslant t \leqslant b$. Требуется определить путь s, проделанный телом за это время. Задача, как видим, в каком-то смысле обратна той, в которой мы находили по известной зависимости пути от времени $s(t)$ мгновенную скорость и которая привела нас к определению производной. Будем рассуждать следующим образом. Разобьем отрезок времени на n равных

промежутков точками $a = t_0 < t_1 < \ldots < t_n = b$. Можно считать промежутки разбиения настолько малыми, что скорость движения на них примерно постоянна и равна скорости в произвольной точке промежутка τ_i (например, в середине):

$$v(t) \approx v(\tau_i),\ t_i \leqslant t < t_{i+1},\ i = 0, 1, \ldots, n-1.$$

Тогда путь на каждом малом промежутке легко посчитать: он равен скорости, умноженной на длину промежутка $\Delta = t_{i+1} - t_i = (b-a)/n$, а полный путь равен сумме этих расстояний:

$$s \approx v(\tau_0)\Delta + v(\tau_1)\Delta + \ldots + v(\tau_{n-1})\Delta = \frac{b-a}{n}\sum_{i=0}^{n-1} v(\tau_i).$$

Приближенное равенство будет тем точнее, чем меньше промежутки, на которые разбит отрезок $[a, b]$, т.е. чем больше их количество n. Равенство становится т о ч н ы м, если устремить n к ∞:

$$s = \lim_{n\to\infty}\frac{b-a}{n}\sum_{i=0}^{n-1} v(\tau_i).$$

Если предел в правой части этого равенства существует, его называют определенным интегралом от функции $v(t)$ на отрезке $[a, b]$ и обозначают почти так же, как неопределенный интеграл:

$$s = \int_a^b v(t)dt.$$

Дадим определение еще раз для произвольной функции $f(x)$, заданной на отрезке $[a, b]$, $a < b$. Зададим разбиение отрезка на n равных промежутков длины $(b-a)/n$ точками $a = x_0 < x_1 < \ldots < x_n = b$. Пусть ξ_i – произвольная точка из промежутка $[x_i, x_{i+1})$. Составим *интегральную сумму*

$$\sum_{i=0}^{n-1} f(\xi_i)\frac{b-a}{n}$$

из произведений значения функции в промежуточной точке на длину маленького промежутка. Предел интегральной суммы при $n \to \infty$, **если он существует и не зависит от выбора промежуточных точек**, называется ***определенным интегралом*** или ***интегралом Римана*** от функции $f(x)$ на отрезке $[a, b]$:

$$\int_a^b f(x)dx = \lim_{n\to\infty}\frac{b-a}{n}\sum_{i=0}^{n-1} f(\xi_i).$$

Итак, определенный интеграл, в отличие от неопределенного, – это **число**, а не функция. Заметим, что неважно, какой буквой обозначается независимая переменная:

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b f(t)\,\mathrm{d}t.$$

Границы отрезка $[a, b]$ называются *пределами интегрирования*. Функция $f(x)$, для которой существует определенный интеграл на $[a, b]$, называется *интегрируемой на* $[a, b]$.

Замечание. На самом деле, строгое определение несколько сложнее. Отрезок $[a, b]$ разбивается на не обязательно равные промежутки, в которых выбирается произвольная точка ξ_i. Составляется интегральная сумма

$$\sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i).$$

Затем устремляют к нулю максимальный промежуток разбиения $\Delta = \max\{x_{i+1} - x_i,\ i = 0, 1, \ldots, n-1\}$ и определенным интегралом называют предел интегральных сумм

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\Delta\to 0} \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i),$$

если он существует и не зависит от разбиения и выбора промежуточных точек. Можно доказать, что определения с равномерным и произвольным разбиением отрезка $[a, b]$ эквивалентны.

Обычно определенный интеграл не считают по определению. Попытаемся проделать вычисления в достаточно простом случае. Рассмотрим функцию $f(x) = x$ на отрезке $[0, 1]$. Разобьем отрезок на n равных частей, длина которых равна $1/n$. В качестве промежуточных точек ξ_i возьмем середины промежутков. Середина первого промежутка – точка $1/(2n)$. Каждая следующая середина получается прибавлением длины промежутка разбиения $1/n$. Таким образом, точки

$$\xi_i = \frac{1}{2n} + \frac{i}{n} = \frac{2i+1}{2n}, \quad i = 0, 1, \ldots, n-1.$$

Соответственно,

$$f(\xi_i) = \xi_i = \frac{2i+1}{2n}.$$

Составим интегральную сумму

$$\sum_{i=0}^{n-1} \frac{2i+1}{2n}\frac{1}{n} = \frac{1}{2n^2}\sum_{i=0}^{n-1}(2i+1).$$

Требуется сложить n нечетных чисел от 1 до $2n-1$: например, $1+3=4$, $1+3+5=9$, $1+3+5+7=16$ и т.д. Заметим, что полученные суммы равны квадрату количества слагаемых. Докажем по индукции, что

$$\sum_{i=0}^{n-1}(2i+1) = n^2.$$

Мы проверили, что это верно при $n = 2, 3, 4$. Пусть это верно при некотором n. Покажем, что тогда формула верна и при $n+1$:

$$\sum_{i=0}^{n}(2i+1) = \sum_{i=0}^{n-1}(2i+1) + (2n+1).$$

По предположению сумма в правой части равна n^2. Поэтому

$$\sum_{i=0}^{n}(2i+1) = n^2 + 2n + 1 = (n+1)^2,$$

т.е. формула верна. Итак, интегральная сумма равна

$$\frac{1}{2n^2}n^2 = \frac{1}{2},$$

независимо от n. Ее предел при $n \to \infty$, разумеется, равен $^1/_2$. Можно показать, что тот же результат получится при произвольном выборе промежуточных точек ξ_i. Таким образом,

$$\int_0^1 x\,\mathrm{d}x = \frac{1}{2}.$$

Георг Фридрих Бернхард Риман (1826–1866). Великий немецкий математик. Риман, пожалуй, больше, чем кто-либо другой, повлиял на развитие современной математики. Он был сыном деревенского священника. Уже в школе начал читать труды Эйлера, Лежандра и других математиков. В 1846 г. поступил в Геттингенский университет, где слушал лекции Гаусса. Его учителем и другом стал другой видный математик Дирихле, оказавший влияние на научное развитие Римана. В 1854 г. Риман становится приват-доцентом Геттингенского университета, а в 1857 г. – профессором. Болезненный, как и Абель, он провел последние месяцы жизни в Италии, где умер в сорокалетнем возрасте. За свою короткую жизнь Риман опубликовал сравнительно небольшое число работ, тем не менее каждая из них была и остается важной, а некоторые посвящены совершенно новым и плодотворным областям математики.

Исследования Римана относятся к теории функций, геометрии, математической физике и другим разделам математики. С его именем связано развитие неевклидовой геометрии. Его подход позволяет рассматривать пространства Евклида и Лобачевского как частные случаи более общего, риманова, пространства. Более того, Риман предположил существование других геометрий, многие из которых впоследствии с пользой были введены в математику и математическую физику. В математическом анализе Риман дал строгое определение определенного интеграла и получил условия его существование. Он плодотворно исследовал функции, аргументом которых являются комплексные числа (теория функций комплексного переменного). Его именем названы различные математические объекты (интеграл Римана, римановы поверхности и др.). Известны леммы и теоремы Римана.

10.3.2. Геометрический смысл интеграла Римана

Чтобы проиллюстрировать определенный интеграл геометрически, рассмотрим следующую задачу. Пусть на отрезке $[a, b]$ задана непрерывная функция $y = f(x) \geqslant 0$. Требуется найти площадь под графиком функции на отрезке $[a, b]$ (эту фигуру обычно называют криволинейной трапецией). Разобьем отрезок $[a, b]$ на n малых промежутков точками $a = x_0 < x_1 < \ldots < x_n = b$. Для простоты промежутки можно считать равными. Площадь криволинейной трапеции S равна сумме площадей S_i маленьких трапеций, построенных на промежутках разбиения (рис. 10.1). Выберем в каждом промежутке произвольную точку ξ_i и построим прямоугольник с основанием на промежутке $[x_i, x_{i+1})$ и высотой, равной $f(\xi_i)$. Если промежуток разбиения мал, площадь построенного прямоугольника примерно равна площади криволинейной трапеции на промежутке $[x_i, x_{i+1})$, т.е. $S_i \approx f(\xi_i)(x_{i+1} - x_i)$.

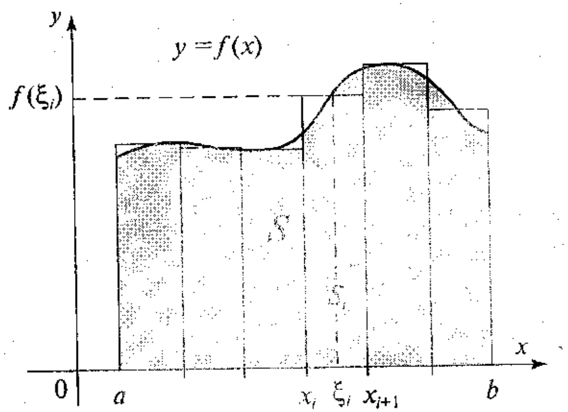


Рис. 10.1

Таким образом, площадь всей криволинейной трапеции приблизительно равна сумме площадей построенных прямоугольников:

$$S \approx \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i).$$

Приближенное равенство будет тем точнее, чем мельче разбиение отрезка $[a, b]$. В пределе, устремив число промежутков разбиения n к бесконечности, мы получим точное равенство. Однако в правой части стоит интегральная сумма функции $f(x)$ на отрезке $[a, b]$. Значит, предел правой части равен определенному интегралу

$$S = \int_a^b f(x)\mathrm{d}x.$$

Итак, интеграл Римана равен площади криволинейной трапеции под графиком функции $f(x) \geqslant 0$ на отрезке $[a, b]$. Найденный интеграл

$$\int_0^1 x\,\mathrm{d}x = \frac{1}{2}$$

легко вычислить как площадь прямоугольного треугольника ABC (рис. 10.2):

$$S = \frac{1}{2} \cdot AB \cdot BC = \frac{1}{2} \cdot 1 \cdot 1 = \frac{1}{2}.$$

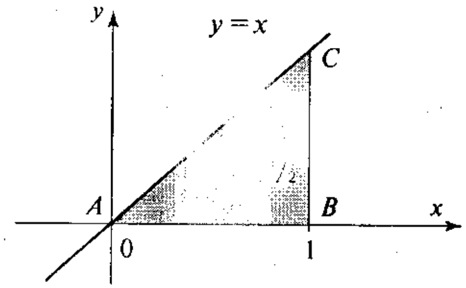


Рис. 10.2

Чуть позже мы убедимся, что если $f(x) < 0$ (график лежит ниже оси x), интеграл Римана равен соответствующей площади, взятой со знаком «–» (рис. 10.3).

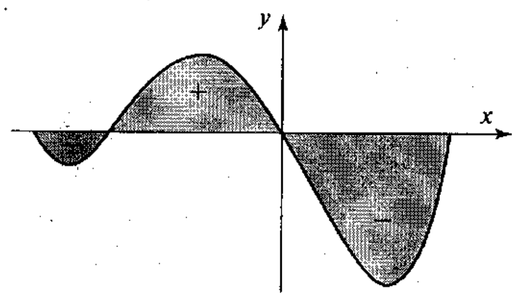


Рис. 10.3

10.3.3. Формула Ньютона – Лейбница

Интеграл Римана определяется через предел интегральных сумм достаточно сложным образом. Ясно, что вычислять интеграл по определению крайне затруднительно. Мы убедились в этом на простом примере. Как же считать определенные интегралы? Обратимся еще раз к примеру с механическим движением тела. Мы выяснили, что путь, пройденный на промежутке времени $[a, b]$, равен интегралу от скорости движения:

$$s = \int_a^b v(t)\mathrm{d}t.$$

С другой стороны, если зависимость пути от времени задается функцией $s(t)$, этот же путь можно посчитать как разность пути, пройденного к моменту b, и пути, пройденного к моменту a:

$$s = s(b) - s(a).$$

Таким образом, справедливо равенство

$$\int_a^b v(t)\mathrm{d}t = s(b) - s(a).$$

Однако функция пути $s(t)$ – это первообразная функции скорости $v(t)$. Более того, любая первообразная $\tilde{s}(t)$ функции $v(t)$ имеет вид $\tilde{s}(t) = s(t) + C$. Поэтому

$$\tilde{s}(b) - \tilde{s}(a) = s(b) - s(a) = \int_a^b v(t)\mathrm{d}t.$$

Значит, чтобы посчитать интеграл от скорости, достаточно найти любую первообразную этой функции и вычислить разность ее значений на концах отрезка интегрирования. Это справедливо для любой непрерывной на $[a, b]$ функции $f(x)$.

Теорема 10.2 (теорема Ньютона – Лейбница). *Пусть функция $f(x)$ непрерывна на отрезке $[a, b]$ и $F(x)$ – любая ее первообразная. Тогда справедлива формула (Ньютона – Лейбница)*

$$\boxed{\int_a^b f(x)\mathrm{d}x = F(b) - F(a)}$$

Разность значений функции на концах обозначают как $F($

Поэтому формулу можно записать короче:

$$\int_a^b f(x)\mathrm{d}x = F(x)\Big|_a^b.$$

Например,

$$\int_a^b 1 \cdot \mathrm{d}x = b - a,$$

т.е. интеграл от единицы равен длине промежутка интегрирования (рис. 10.4).

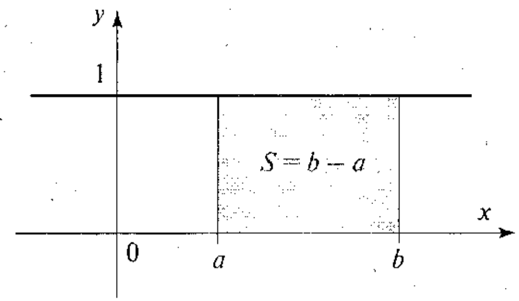


Рис. 10.4

Формула Ньютона – Лейбница, во-первых, устанавливает связь между определенным и неопределенным интегралами. Ведь произвольная первообразная функции – это и есть неопределенный интеграл. Итак, определенный интеграл от непрерывной функции равен разности значений на концах неопределенного интеграла от этой функции. Во-вторых, формула дает мощный инструмент для вычисления определенных интегралов, поскольку мы умеем искать неопределенные интегралы различными методами.

Примеры.

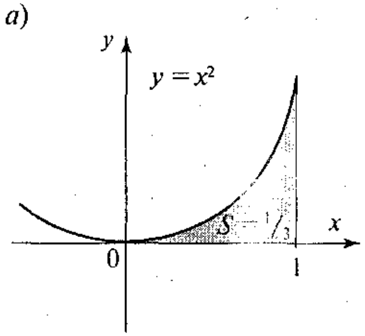
1. $\int_0^1 x\,\mathrm{d}x = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}.$

Этот интеграл мы уже нашли по определению (см. рис. 10.2).

2. $\int_0^1 x^2\,\mathrm{d}x = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3} - 0 = \frac{1}{3}$ (рис. 10.5, a).

3. $\int_0^{\pi/2} \sin t\,\mathrm{d}t = -\cos t\Big|_0^{\pi/2} = (-\cos \pi/2) - (-\cos 0) = 0 - (-1) = 1$

ис. 10.5, $б$).



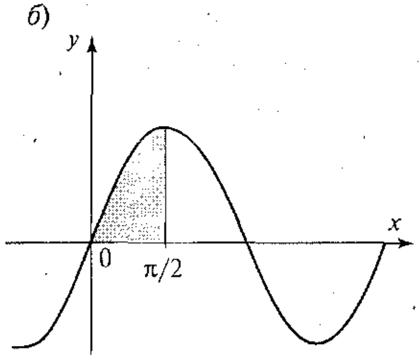


Рис. 10.5

Мы еще раз подчеркнули, что неважно, какой буквой обозначать независимую переменную.

4. $\int\limits_0^1 e^x\,dx = e^x\Big|_0^1 = e^1 - e^0 = e - 1.$

10.3.4. Свойства интеграла Римана

Определенный интеграл есть предел интегральных сумм. Поэтому линейные свойства предела наследуются интегралом Римана: константа выносится за знак интеграла, интеграл от суммы (разности) функций равен сумме (разности) интегралов.

∇**1.** *Пусть функции $f(x)$ и $g(x)$ интегрируемы на $[a, b]$. Тогда*

$$\int\limits_a^b cf(x)\,dx = c\int\limits_a^b f(x)\,dx,$$

$$\int\limits_a^b (f(x) \pm g(x))\,dx = \int\limits_a^b f(x)\,dx \pm \int\limits_a^b g(x)\,dx.$$

Эти свойства, как всегда, можно обобщить: определенный интеграл от линейной комбинации функций равен линейной комбинации интегралов. Например,

$$\int\limits_1^2 (2x - 0{,}5x^2)\,dx = 2\int\limits_1^2 x\,dx - 0{,}5\int\limits_1^2 x^2\,dx = 2\cdot\frac{x^2}{2}\Bigg|_1^2 - 0{,}5\cdot\frac{x^3}{3}\Bigg|_1^2 =$$

$$= 2\left(\frac{4}{2} - \frac{1}{2}\right) - 0{,}5\left(\frac{8}{3} - \frac{1}{3}\right) = 2\cdot\frac{3}{2} - 0{,}5\cdot\frac{7}{3} = 3 - \frac{7}{6} = 1\frac{5}{6}.$$

Пусть задана непрерывная функция $y = f(x)$ на отрезке $[a, b]$ и точка c лежит внутри отрезка (рис. 10.6).

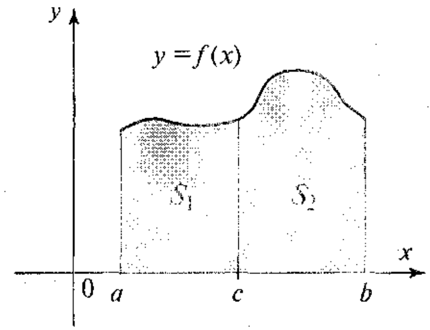


Рис. 10.6

Тогда ясно, что площадь криволинейной трапеции S на отрезке $[a, b]$ распадается на сумму двух площадей: площадь S_1 под графиком на отрезке $[a, c]$ и S_2 на отрезке $[c, b]$. Каждая площадь равна соответствующему определенному интегралу. Значит, выполняется следующее свойство.

∇**2.** *Если функция $f(x)$ интегрируема на $[a, b]$, то*

$$\int\limits_a^b f(x)\,dx = \int\limits_a^c f(x)\,dx + \int\limits_c^b f(x)\,dx.$$

Поскольку мы договорились, что площадь входит в интеграл с учетом знака, данная формула справедлива, независимо от знака функции $f(x)$. Свойство $\nabla 2$ называют *аддитивным* свойством определенного интеграла. Например, с одной стороны,

$$\int\limits_0^{2\pi} \sin t\,dt = -\cos t\Big|_0^{2\pi} = (-\cos 2\pi) - (-\cos 0) = (-1) - (-1) = 0.$$

С другой стороны,

$$\int\limits_0^{2\pi} \sin t\,dt = \int\limits_0^{\pi} \sin t\,dt + \int\limits_{\pi}^{2\pi} \sin t\,dt = -\cos t\Big|_0^{\pi} + (-\cos t)\Big|_{\pi}^{2\pi} =$$

$$= (1 - (-1)) + (-1 - 1) = 2 - 2 = 0.$$

Это хорошо видно на графике (рис. 10.7): площади под графиком синуса на участке от 0 до π и от π до 2π совпадают, но в интеграл входят с разными знаками. Поэтому в итоге интеграл на $[0, 2\pi]$ равен нулю.

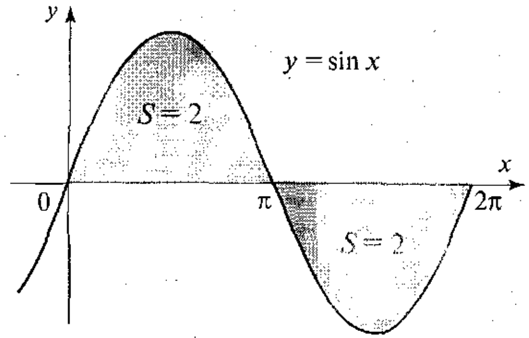


Рис. 10.7

Мы определили интеграл Римана от $f(x)$ на $[a, b]$ при $a < b$. Полезно расширить это определение, полагая

$$\int_a^a f(x)\mathrm{d}x = 0,$$

$$\int_a^b f(x)\mathrm{d}x = -\int_b^a f(x)\mathrm{d}x.$$

При таком расширенном понимании символа $\int\limits_a^b$ аддитивное свойство выполняется независимо от расположения точек a, b, c (лишь бы существовал интеграл на наибольшем из отрезков). Например,

$$\int_0^1 x\,\mathrm{d}x = \int_0^2 x\,\mathrm{d}x + \int_2^1 x\,\mathrm{d}x = \int_0^2 x\,\mathrm{d}x - \int_1^2 x\,\mathrm{d}x = \left(\frac{4}{2} - 0\right) - \left(\frac{4}{2} - \frac{1}{2}\right) = \frac{1}{2}.$$

Из геометрического смысла интеграла Римана ясно, что если функция не ограничена на отрезке $[a, b]$, она не может быть интегрируемой на этом отрезке. В самом деле, нам никак не найти площадь под графиком неограниченной функции. Например, нетрудно убедиться, что не существует интеграла

$$\int_0^1 \frac{\mathrm{d}x}{x}.$$

Наоборот, если функция интегрируема, она является ограниченной на отрезке интегрирования. Сформулируем это в виде свойства.

∇3. *Если функция $f(x)$ интегрируема на отрезке $[a, b]$, то она является ограниченной на этом отрезке.*

10.3.5. Замена переменной и интегрирование по частям

Коль скоро установлена простая связь между определенным и неопределенным интегралом, при вычислении интеграла Римана можно пользоваться формулами замены переменной и интегрирования по частям. Естественно, эти формулы будут выглядеть несколько иначе.

Замена переменной в определенном интеграле. При замене переменной $\varphi(x) = t$ в определенном интеграле должны измениться пределы интегрирования: если исходная переменная изменялась от a до b, то новая переменная t изменяется в пределах от $\varphi(a)$ до $\varphi(b)$. Значит, формула для определенного интеграла записывается следующим образом:

$$\boxed{\int_a^b f(\varphi(x))\varphi'(x)\mathrm{d}x = \int_{\varphi(a)}^{\varphi(b)} f(t)\mathrm{d}t}$$

Примеры.

1. $$\int_0^2 e^{x^2} x\,\mathrm{d}x \overset{x^2=t,\ x\mathrm{d}x=\mathrm{d}t/2}{=} \frac{1}{2}\int_0^4 e^t\,\mathrm{d}t = \frac{1}{2}e^t\Big|_0^4 = \frac{e^4-1}{2}.$$

Исходная переменная x изменялась от 0 до 2, значит, новая переменная $t = x^2$ меняется от 0 до 4.

2. $$\int_1^e \frac{\ln x}{x}\mathrm{d}x \overset{\ln x=t,\ \mathrm{d}x/x=\mathrm{d}t}{=} \int_{\ln 1}^{\ln e} t\,\mathrm{d}t = \int_0^1 t\,\mathrm{d}t = \frac{t^2}{2}\Big|_0^1 = \frac{1}{2}.$$

3. $$\int_1^2 \frac{e^{1/x}\,\mathrm{d}x}{x^2} \overset{1/x=t,\ \mathrm{d}x/x^2=-\mathrm{d}t}{=} -\int_1^{1/2} e^t\,\mathrm{d}t = \int_{1/2}^1 e^t\,\mathrm{d}t = e^t\Big|_{1/2}^1 = e - \sqrt{e}.$$

Интегрирование по частям. Формула интегрирования по частям для определенного интеграла отличается от соответствующей формулы для неопределенного интеграла тем, что содержит в правой части пределы интегрирования:

$$\boxed{\int_a^b u(x)v'(x)\mathrm{d}x = u(x)v(x)\Big|_a^b - \int_a^b u'(x)v(x)\mathrm{d}x}$$

Примеры.

1. $\int\limits_0^1 x e^{-x} dx \overset{u=x,\ v'=e^{-x},\ u'=1,\ v=-e^{-x}}{=} (-xe^{-x})\Big|_0^1 + \int\limits_0^1 e^{-x} dx =$

$= -1\cdot e^{-1} - 0 + \left(-e^{-x}\right)\Big|_0^1 = -\dfrac{1}{e} - \dfrac{1}{e} + 1 = 1 - \dfrac{2}{e} = \dfrac{e-2}{e}.$

2. $\int\limits_0^{\pi/2} x\cos x\, dx \overset{u=x,\ v'=\cos x,\ u'=1,\ v=\sin x}{=} x\sin x\Big|_0^{\pi/2} - \int\limits_0^{\pi/2} \sin x\, dx =$

$= (\pi/2)\sin \pi/2 - 0 + \cos x\Big|_0^{\pi/2} = \pi/2 + \cos \pi/2 - \cos 0 = \pi/2 - 1.$

10.4. Интегралы на бесконечных промежутках

В данном параграфе определяются интегралы Римана от функций, заданных на бесконечных промежутках. С геометрической точки зрения такой интеграл представляет собой площадь «бесконечно длинной» фигуры.

Пусть функция определена при $x \geqslant a$ и интегрируема на любом конечном отрезке $[a, b]$. Может оказаться так, что существует предел

$$\lim_{b\to\infty} \int\limits_a^b f(x)dx.$$

Тогда этот предел тоже называют интегралом от функции $f(x)$ на промежутке $[a, \infty)$ и обозначают

$$\int\limits_a^\infty f(x)dx.$$

Рассмотрим, например, интеграл

$$\int\limits_0^\infty e^{-x} dx = \lim_{b\to\infty} -e^{-x}\Big|_0^b = \lim_{b\to\infty}(-e^{-b}+1) = 1 \quad (\text{рис. } 10.8).$$

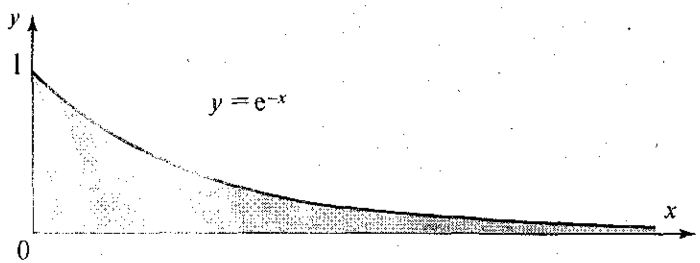


Рис. 10.8

Аналогично определяется интеграл от $-\infty$ до конечной точки b:

$$\int\limits_{-\infty}^b f(x)dx = \lim_{a\to-\infty} \int\limits_a^b f(x)dx.$$

Например,

$$\int\limits_{-\infty}^0 \frac{dx}{1+x^2} = \lim_{a\to-\infty} \int\limits_a^0 \frac{dx}{1+x^2} = \lim_{a\to-\infty} \operatorname{arctg} x\Big|_a^0 = \lim_{a\to-\infty} (\operatorname{arctg} 0 - \operatorname{arctg} a) =$$

$$= 0 - (-\pi/2) = \pi/2 \quad (\text{рис. } 10.9).$$

Мы учли, что

$$(\operatorname{arctg} x)' = \frac{1}{1+x^2}$$

(убедитесь в этом самостоятельно) и что арктангенс стремится к $-\pi/2$ при $x \to -\infty$.

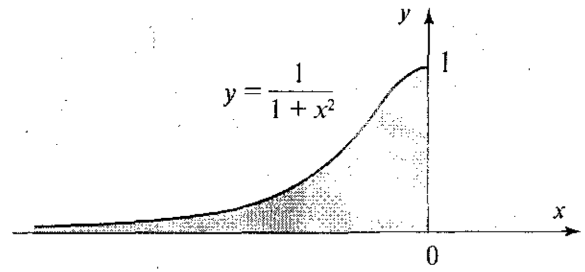


Рис. 10.9

С интегралами на бесконечных промежутках мы встретимся очень скоро, в следующей части, при изучении теории вероятностей.

Задачи

1. Найти неопределенные интегралы методом замены переменной:

а) $\int (3x+1)^2 dx$; б) $\int \dfrac{dx}{(2x-3)^3}$; в) $\int \dfrac{x\,dx}{\sqrt{1+x^2}}$;

г) $\int \dfrac{x^2 dx}{x^3+1}$; д) $\int x \sin x^2 dx$; е) $\int e^{\sin x} \cos x\, dx$.

2. Найти неопределенные интегралы методом интегрирования по частям:

а) $\int x\sin 2x\,dx$; б) $\int x e^{-x}\,dx$; в) $\int \arccos x\,dx$;

г) $\int x^2 \ln x\,dx$; д) $\int x\cos^2 x\,dx$.

3. Вычислить определенные интегралы:

а) $\int\limits_0^{\pi} \cos x\,dx$; б) $\int\limits_1^2 x^3\,dx$; в) $\int\limits_{-1}^{1} e^t\,dt$;

г) $\int\limits_0^1 \sqrt{1+x}\,dx$; д) $\int\limits_{-2}^{-1} \frac{dx}{(11+5x)^3}$; е) $\int\limits_0^1 (e^x - x + 2)\,dx$;

ж) $\int\limits_0^1 \frac{x\,dx}{(x^2+1)^2}$; з) $\int\limits_0^{\pi/2} \frac{\cos x}{\sin^2 x}\,dx$; и) $\int\limits_1^2 \frac{e^{1/x}}{x^2}\,dx$;

к) $\int\limits_0^1 x e^{-x}\,dx$; л) $\int\limits_0^{\pi/2} x\cos x\,dx$; м) $\int\limits_0^{e-1} \ln(x+1)\,dx$.

4. Вычислить интегралы на бесконечных промежутках:

а) $\int\limits_1^{\infty} \frac{dx}{x^4}$; б) $\int\limits_0^{\infty} e^{-2x}\,dx$; в) $\int\limits_{-\infty}^{0} \frac{2x\,dx}{(x^2+1)^2}$.

Часть 4
ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

...И случай, бог – изобретатель.

А.С. Пушкин

Средь шумного бала, случайно...

А.К. Толстой

Мы переходим к изучению совершенно новых математических объектов, связанных с понятием *случайности.* Случайность как явление реальной жизни хорошо нам знакома. В языке это закреплено множеством поговорок и выражений: случай может быть «счастливым» и «несчастным», «случайными» бывают попутчики, встречи, звонки и т.д. Мы говорим «дело случая», подразумевая, например, исход хоккейного матча. Многое в нашей жизни случайно, т.е. не обусловлено четкими причинно-следственными связями. Случайность чаще всего и выступает противоположностью обусловленности или, как говорят, детерминированности.

Итак, даже на бытовом уровне мы различаем два вида событий – детерминированные и случайные. В детерминированном событии мы точно знаем, что данная причина приведет к единственному и вполне определенному следствию. Мы нажимаем на кнопку, и лампа загорается: между действием и результатом существует четкая однозначная связь. Исход случайного события, напротив, непредсказуем, он зависит от случайных факторов. Классический пример такого события – подбрасывание монеты или кубика. Нельзя узнать заранее, на какую сторону упадет монета или какая грань выпадет на кубике. Далее мы будем говорить о случайных событиях на математическом языке и сформулируем основные условия, позволяющие считать событие случайным. Оказывается, одной только непредсказуемости результата для этого недостаточно: требуется еще возможность неограниченного повторения случайного опыта и дополнительное свойство устойчивости.

Существуют также бытовые представления о *вероятности.* Например, оценивая шансы на выигрыш, мы неосознанно подсчитываем его вероятность. Математическое понятие вероятности как раз и возникло из анализа карточной игры. Мы называем событие «невероятным», нисколько не задумываясь о математическом содержании этого понятия. Отметим, что на самом деле математическая и бытовая трактовки вероятностных понятий достаточно далеки друг от друга.

Остановимся еще раз на роли случайности в математике и реальности. Многие процессы сопровождаются неизвестными, неопределенными воздействиями: помехами, шумами, измерительными ошибками и т.д. Быть может, эти явления и не случайны по сути, но причины, порождающие их, нам неизвестны. Для математического описания таких явлений удобно считать, что неопределенные факторы имеют случайную природу. Итак, случайность с этой точки зрения – это инструмент, с помощью которого легче исследовать неопределенные явления.

С другой стороны, случайность – не выдуманная математическая абстракция, она, безусловно, существует в реальности. Это значит, что мы не в состоянии получить полную информацию о явлении не только из-за нашего неумения, неспособности, неточных приборов, но и объективно: так устроена природа. В физике микромира хорошо известен принцип неопределенности Гейзенберга, согласно которому невозможно одновременно точно определить положение и скорость частицы. Великий американский математик Норберт Винер в своей книге «Кибернетика» иллюстрировал необходимость введения случайности в описание природы на примере астрономии и метеорологии. Движение небесных тел жестко обусловлено законами механики. Можно составить, например, сколь угодно точное расписание солнечных и лунных затмений на сотни лет вперед или назад. Метеорологические явления, напротив, подвержены влиянию многих неопределенных факторов. Будь мы хоть семи пядей во лбу, мы не сможем учесть все силы, действующие в настоящий момент на облака. Движение облаков подчиняется случайным законам. Точно так же невозможно получить полную информацию обо всех силах, действующих в обществе. Общественное развитие – результат совместного действия многих случайных факторов. Это же можно сказать и о поведении отдельного человека, о формировании его психологического состояния. Большинство математических методов, применяемых в социологии, психологии и других гуманитарных науках, опирается на представление о случайном характере явлений в этих областях.

Любое развитие, эволюция происходят за счет постепенного накапливания случайных изменений. Теория эволюции Чарлза Дарвина и теория наследственности Грегора Менделя дают этому прекрасное подтверждение. Изменение биологического вида – это результат случайных мутаций, происходящих на протяжении многих поколений. Интересно, что сын Чарлза Дарвина, Джордж Дарвин, создал так называемую теорию приливной эволюции, объясняющую возникновение приливной волны через случайные столкновения частиц воды. Таким образом, и механические явления, казалось бы, полностью подчиненные законам Ньютона, несут в себе элемент случайности. Очень интересна философская связь между случайностью и направленностью времени. Лауреат Нобелевской премии Илья Пригожин в своих исследованиях по неравновесной термодинамике по существу доказал необратимость времени, подтвердив философские воззрения Анри Бергсона.

Дадим краткий исторический очерк теории вероятностей как раздела математики. Проникновение случайности в математику, которое в конечном счете привело к созданию и развитию этой науки, имеет длинную историю, начало которой лежит в Древней Греции. Детерминист и материалист Демокрит (родился около 470 до н.э.) утверждал: «Люди сотворили себе кумира из случая как прикрытие для присущего им недомыслия». Великий философ Аристотель (384–322 до н.э.), напротив, писал: «Уничтожение случая ведет за собой нелепые последствия. Есть многое, что совершается не по необходимости, а случайно».

Возникновение вероятностных понятий в математике связывают с перепиской двух великих французов Блеза Паскаля и Пьера Ферма, посвященной задаче о справедливом разделении ставок в игре. Интерес к задачам, связанным с вероятностями, формировался прежде всего под влиянием развития страхового дела, но частные вопросы, побудившие известных математиков поразмыслить над этим предметом, были поставлены в связи с играми в кости и карты. Кавалер де Мере (который вовсе не был заядлым игроком, а серьезно интересовался наукой) обратился к Паскалю по поводу так называемой задачи об очках. Паскаль завязал переписку с Ферма, и они вдвоем установили некоторые основные положения теории вероятностей (1654). Когда Христиан Гюйгенс узнал об этой переписке, он попытался дать собственное решение, в результате чего появилась его книга «О расчетах при азартных играх» (1657), первый трактат по теории вероятностей. Гюйгенс писал: «...при внимательном изучении предмета читатель заметит, что он занимается не только игрой, а что здесь даются основы теории глубокой и весьма интересной». Итак, теория вероятностей зародилась как наука в XVII в.

под влиянием интереса к такому несерьезному занятию, как азартные игры. Дальнейшее развитие теории вероятностей в XVIII в. связано с именами таких математиков, как Якоб Бернулли (1654–1705), Даниил Бернулли (1700–1782), Абрахам де Муавр (1667–1754) и Жорж Луи Бюффон (1707–1788). Можно сказать, что к середине XVIII в. вероятность завоевала прочное место в математике. Ее применяют также в демографии, при оценках ошибок наблюдений, постепенно возникают вероятностные подходы в различных разделах естествознания.

Значительный вклад в теорию вероятностей внесли русские математики. В первую очередь это Пафнутий Львович Чебышёв (1821–1894) и его ученики А.М. Ляпунов (1857–1918) и А.А. Марков (1856–1922). Их результаты – большой шаг в науке и ее практическом применении. Традицию продолжают советские математики. Первая попытка построить теорию вероятностей на основе системы аксиом была предпринята математиком С.Н. Берштейном в 1917 г. Окончательно аксиоматический подход к вероятности сложился в работах выдающегося математика А.Н. Колмогорова (его статья «Об основных понятиях теории вероятностей» была опубликована в 1933 г.). Аксиоматика Колмогорова признана математическим миром и составляет фундаментальную основу теории вероятностей.

Переход к изучению случайных явлений, их математическое описание на языке вероятности потребуют некоторого умственного напряжения. Не пожалеем усилий, ведь это расширит наши представления не только о математике, но и об окружающем мире.

Глава 11
СЛУЧАЙНЫЕ СОБЫТИЯ

11.1. Понятие случайного события

11.1.1. Классификация событий

Мы уже дали некоторую начальную классификацию, разделив события на *детерминированные* и *случайные*. Рассмотрим основные черты (их как раз и «покрывает тайна») случайного события как математического объекта. Какие наблюдения можно сделать, к примеру, при бросании монеты? Во-первых, исход этого эксперимента

непредсказуем. Однако возможных исходов всего два: может выпасть либо «орел», либо «решка». Каждый раз, бросая монету, мы получаем один из двух возможных результатов. Говоря «каждый раз», мы подразумеваем, что эксперимент можно повторять. Более того, его можно повторять неограниченное число раз. Реальное бросание монеты всегда происходит по-разному: один раз мы подбрасываем монету сильнее, другой раз – слабее и т.д. Короче говоря, условия реального эксперимента при повторении неизбежно меняются. При математическом определении случайного события (эксперимента) мы абстрагируемся от возможной неодинаковости условий и считаем, что эксперимент можно повторить неограниченное число раз при одинаковых условиях. Это, наряду с непредсказуемостью единичного опыта, составляет важную характеристику случайного события.

Случайный характер явления может проявиться лишь при многократном повторении эксперимента. Поэтому уникальное событие, даже если его исход и нельзя предсказать, не может считаться случайным с математической точки зрения. Например, исход выборов президента непредсказуем, но данный эксперимент невозможно повторять много раз в одинаковых условиях. Такие события (с неясным исходом, но и не случайные) называют *неопределенными*.

Давайте при бросании монеты будем наблюдать, например, за количеством выпадения «орлов». Пусть монета подбрасывалась n раз. Если при этом выпало μ «орлов», доля выпавших «орлов» составляет $\nu = \mu/n$. Число μ называют обычно *частотой* данного случайного события, а число ν – *относительной частотой*. Будем откладывать по оси x количество испытаний n, а по оси y – относительную частоту ν. Если мы бросим монету один раз, выпадет либо «решка», либо «орел»; доля орлов составит либо 0, либо 1. При двух бросаниях «орел» может не появиться ни разу, а может выпасть один или два раза, при этом относительная частота ν будет равна соответственно 0, $^1/_2$ или 1. Однако при увеличении n мы заметим, что наш график ведет себя очень интересным образом: доля «орлов» ν начинает колебаться вблизи величины $^1/_2$; говорят, что относительная частота выпадения «орлов» стабилизируется около этого значения (рис. 11.1).

Если повторить серию бросаний, мы получим, разумеется, новую кривую. Однако она будет обладать тем же замечательным свойством: при увеличении числа испытаний относительная частота события имеет тенденцию к стабилизации. Можно сказать, что большие отклонения от уровня $^1/_2$ хотя и не исключены, но будут происходить все реже и реже. То же самое можно сказать и о частоте выпадения «решек». Такое поведение характерно для любого случайного события и носит название *закона статистической устойчивости*, который, к сожалению, мы пока не можем сформулировать строго (его точная формулировка будет получена в конце курса теории вероятностей). Пока можно утверждать лишь следующее.

> При неограниченном возрастании числа случайных экспериментов относительная частота каждого исхода имеет тенденцию к стабилизации.

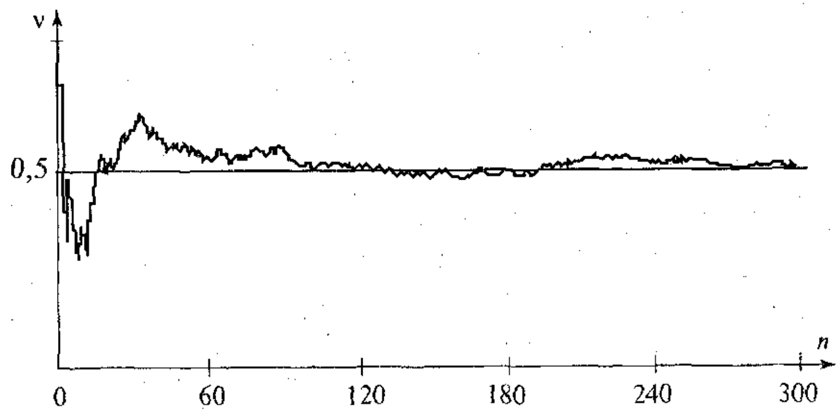


Рис. 11.1

Итак, случайное событие характеризуется тремя особенностями:

- неопределенностью исхода единичного эксперимента;
- возможностью неограниченного повторения в одинаковых условиях;
- стабилизацией относительной частоты.

Ниже мы с удовольствием убедимся, что величина, около которой стабилизируется относительная частота, как раз и представляет собой вероятность случайного события. Случайные события в теории вероятностей обычно обозначают прописными латинскими буквами A, B, C и т.д. Примеры случайных событий традиционны. Уже упоминалось бросание монеты и игрального кубика. Многие задачи теории вероятностей связаны со случайными экспериментами с «урной» (под урной понимается ящик, коробка и т.д.). Пусть, например, в урне находятся 4 белых и 6 черных шаров. Тщательно перемешаем шары и наугад вытащим один шар из урны. Скажем,

что произошло событие A, если мы достали белый шар. Это событие является случайным, если соблюдать чистоту эксперимента – каждый раз тщательно перемешивать шары и доставать шар не глядя. Похожие события происходят, например, при розыгрыше тиража «Спортлото».

Будем придерживаться следующей терминологии. Случайное событие наступает или не наступает в результате *случайного эксперимента* (*случайного опыта*). Например, случайное событие «выпал орел» может наступить в результате бросания монеты.

11.1.2. Действия над случайными событиями

Как всегда, определив какой-нибудь математический объект, мы должны установить правила работы с ним. Для этого чаще всего мы задаем действия или операции над объектами.

Говорят, что событие B *следует* из события A, если событие B происходит всегда, когда произошло событие A. Например, событие A состоит в том, что при бросании кубика выпала единица, а событие B – что выпало нечетное число. Ясно, что если выпадает единица, то выпадает и нечетное число, т.е. B наступает всегда, если происходит A. Это обозначается тем же символом, что и подмножество: $A \subset B$. Далее мы убедимся в том, что существует глубокая связь между случайными событиями и множествами. Поэтому и определения действий над событиями очень похожи на определения операций над множествами. Будем говорить о *равенстве* двух событий A и B, если из A следует B и из B следует A.

Суммой двух событий A и B называется событие $A + B$, состоящее в том, что произошло событие A **или** событие B. Здесь мы в третий раз сталкиваемся с союзом «или» (первый раз мы пользовались этим словом при определении объединения множеств, второй раз – когда определяли логическую конъюнкцию). В данном случае «или» снова употребляется в неисключающем значении: A или B *означает, что произошло либо событие A, либо событие B, либо оба* эти события одновременно. Из определения ясно, что сложение событий удовлетворяет коммутативному и ассоциативному законам.

∇1.
$$A + B = B + A,$$
$$A + (B + C) = (A + B) + C.$$

Коммутативность и ассоциативность позволяют складывать любое число событий в любом порядке. Из определений вытекает также следующее свойство.

∇**2.** *Из события A следует сумма этого события с любым событием B*:

$$A \subset A + B.$$

Примеры.

1. **Бросание кубика.** Пусть случайное событие C состоит в том, что выпавшее на грани число кратно трем. Такое событие есть сумма $A + B$ двух событий: A = «выпадет 3» и B = «выпадет 6». Событие D, состоящее в том, что выпадет четное число, представляет собой сумму трех событий («выпадет 2», «выпадет 4» и «выпадет 6»). Это же событие в соответствии с ассоциативным законом можно представить в виде суммы двух событий: «двойка» и «четверка или шестерка»; «четверка» и «двойка или шестерка» и т.д.

2. **Стрельба по мишени.** Пусть стрелок сделал 5 выстрелов по мишени. Событие C = «стрелок попал не меньше трех раз» есть сумма трех событий:

$$C = A_3 + A_4 + A_5,$$

где A_k = «стрелок попал ровно k раз», $k = 3, 4, 5$.

Событие D = «стрелок попал хотя бы один раз» произойдет, если стрелок попал один раз, или два раза, или три раза и т.д. Значит, в тех же обозначениях

$$D = A_1 + A_2 + A_3 + A_4 + A_5.$$

3. **Одновременное бросание двух кубиков.** Пусть C = «произведение выпавших чисел четно». Произведение будет четным, если хотя бы на одном кубике выпадет четное число (при этом не исключено, что четное число выпадет на обоих кубиках сразу):

$$C = A + B,$$

где события A и B состоят в том, что четное число выпадет на первом или втором кубике соответственно.

События A и B называются *несовместными*, если они не могут произойти одновременно. В примерах 1 и 2 слагаемые представляют собой несовместные события. В примере 3 события A и B не являются несовместными.

Произведением двух событий A и B называется событие AB, состоящее в том, что событие A и событие B произошли одновременно. Умножение событий так же, как и сложение, коммутативно и ассоциативно.

∇**3.**
$$AB = BA,$$
$$A(BC) = (AB)C.$$

По определению справедливо следующее свойство.

∇**4.** *Из события AB следуют событие A и событие B*:

$$AB \subset A, \quad AB \subset B.$$

Сложение и умножение событий удовлетворяют двум дистрибутивным законам.

∇**5.**
$$A(B + C) = AB + AC,$$
$$A + BC = (A + B)(A + C).$$

Последнее равенство выглядит непривычно, однако очень напоминает дистрибутивный закон для операций над множествами, если сложение заменить на объединение, а умножение – на пересечение. Аналогия не случайна, вскоре мы убедимся в этом. Доказательство дистрибутивных законов опирается на определения и на свойства ∇2, ∇3. Проведите доказательство самостоятельно.

Примеры.

1. **Эксперимент с урной.** Пусть в урне находятся шары трех видов: белые, черные и пестрые – наполовину белые, наполовину черные. Событие A означает, что вынули шар с белым цветом, а событие B – шар с черным цветом. Тогда произведение этих событий AB состоит в том, что вынули пестрый шар (события A и B произошли одновременно). Очевидно, что если вынули пестрый шар, то вынули шар с белым цветом (событие A) и с черным цветом (событие B). Значит, в самом деле из произведения событий следуют оба сомножителя.

2. **Одновременное бросание двух кубиков.** Событие C состоит в том, что сумма чисел на обоих кубиках четна. Это может произойти, когда оба числа четные или оба нечетные. Пусть A_1 = «на первом кубике четное число», B_1 = «на первом кубике нечетное число», A_2, B_2 – те же события для второго кубика. Тогда

$$C = A_1A_2 + B_1B_2.$$

Событие $\overline{A}$ называется *противоположным* событию A, если оно состоит в том, что не произошло событие A. Например, выпадение «орла» при бросании монеты противоположно выпадению «решки». Если A – выпадение пятерки при бросании кубика, то $\overline{A}$ состоит в том, что не выпала пятерка, т.е. выпала или единица, или двойка, или тройка, или четверка, или шестерка. Таким образом,

$$\overline{A_5} = A_1 + A_2 + A_3 + A_4 + A_6,$$

где A_k = «выпало число k».

Событие называется *достоверным*, если оно происходит всегда. Например, событие, состоящее в том, что при бросании монеты выпадет «орел» или «решка», является достоверным. Соответственно, событие, не наступающее никогда, называется *невозможным*. Из определения ясно, что произведение несовместных событий – невозможное событие. К примеру, одновременное выпадение «орла» и «решки» невозможно.

События образуют *полный набор*, если они несовместны, а их сумма есть достоверное событие. Например, при бросании монеты полный набор состоит из двух событий: выпадение «орла» или «решки». Действительно, событие «выпал орел или решка» является достоверным. При бросании кубика можно найти несколько полных наборов. К примеру, шесть событий A_k, $k = 1 \div 6$, состоящих в том, что выпала грань с числом k, образуют полный набор. Два события «выпало четное число» и «выпало нечетное число» также образуют полный набор.

11.1.3. Эмпирическая вероятность

Легко заметить, что одно случайное событие происходит чаще другого. Например, при бросании кубика четное число выпадает чаще, чем кратное трем. Обычно при этом мы делаем вывод, что первое событие (то, которое бывает чаще) имеет больший шанс произойти, чем второе (более редкое). Таким образом, шанс случайного события мы интуитивно связываем с его частотой. Тем не менее между этими понятиями существует принципиальное различие: шансы оцениваются заранее, до опыта, а частота может быть определена только экспериментально, после проведения серии случайных опытов в одинаковых условиях. По результатам экспериментов (по прошлому опыту) мы оцениваем шанс случайного события произойти в будущем.

Далее будет введено понятие вероятности как числовой характеристики шанса, возможности появления случайного события. Вероятность присуща случайному событию, ее можно посчитать заранее, до опыта. Однако связь шанса с частотой события несомненна. Поэтому интересно предварительно выяснить некоторые свойства частоты, которые должны перейти по наследству к математической вероятности. Итак, пусть случайный эксперимент проводится n раз в одинаковых условиях и событие A, за которым мы наблюдаем, случилось $\mu(A)$ раз (это количество успешных экспериментов). Относительная частота события A есть

$$\nu_n(A) = \frac{\mu(A)}{n}.$$

Относительную частоту называют еще ***эмпирической вероятностью*** именно потому, что по частоте события мы оцениваем возможность его появления в будущем. Обозначим эмпирическую вероятность через $\tilde{P}_n(A)$. С буквы P начинается слово Probability («вероятность»), тильда указывает на то, что пока это не настоящая вероятность, а эмпирическая, индекс n напоминает о количестве случайных экспериментов. Итак,

$$\tilde{P}_n(A) = \frac{\mu(A)}{n}.$$

Справедливы следующие свойства эмпирической вероятности.

∇1. *Для любого случайного события A*

$$0 \leqslant \tilde{P}_n(A) \leqslant 1.$$

Действительно, эмпирическая вероятность по определению неотрицательна. С другой стороны, число успешных опытов $\mu(A)$ не превышает общего числа n проведенных экспериментов. Поэтому выполняется второе неравенство.

∇2. *Пусть A и B – несовместные события. Тогда*

$$\tilde{P}_n(A+B) = \tilde{P}_n(A) + \tilde{P}_n(B),$$

т.е. эмпирическая вероятность суммы несовместных событий равна сумме их эмпирических вероятностей.

В самом деле, пусть в результате n случайных экспериментов событие A наступило $\mu(A)$ раз, а событие B – соответственно $\mu(B)$ раз. Поскольку события A и B несовместны (не наступают одновременно), событие $A + B$ (A или B) произошло при этом $\mu(A) + \mu(B)$ раз. Значит,

$$\tilde{P}_n(A+B) = \frac{\mu(A)+\mu(B)}{n} = \frac{\mu(A)}{n} + \frac{\mu(B)}{n} = \tilde{P}_n(A) + \tilde{P}_n(B).$$

Свойство ∇2 выполняется не только для двух несовместных событий, но и для любого конечного набора попарно несовместных событий $A_1, A_2, \ldots, A_k$:

$$\tilde{P}_n\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \tilde{P}_n(A_i).$$

∇3. *Пусть события $A_1, A_2, \ldots, A_k$ образуют полный набор. Тогда*

$$\sum_{i=1}^{k} \tilde{P}_n(A_i) = 1.$$

Набор событий $A_1, A_2, ..., A_k$ полон. Значит, их сумма представляет собой достоверное событие, т.е. в результате данного случайного эксперимента может наступить только одно из событий A_i и никакое другое. Поэтому событие $A_1 + A_2 + ... + A_k$ происходит ровно столько раз, сколько было экспериментов, т.е. n раз:

$$\mu(A_1) + \mu(A_2) + ... + \mu(A_k) = n.$$

Таким образом,

$$\tilde{P}_n\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \tilde{P}_n(A_i) = \frac{\mu(A_1)+\mu(A_2)+...+\mu(A_k)}{n} = \frac{n}{n} = 1.$$

Как мы уже указывали, существует тесная связь между эмпирической вероятностью (экспериментально вычисляемой относительной частотой) и вероятностью (характеристикой возможности появления события, определяемой до опыта). В силу этого вероятность должна унаследовать основные свойства эмпирической вероятности. Перечислим их, не давая пока формального определения вероятности:

- значение вероятности находится между нулем и единицей;
- вероятность суммы несовместных событий равна сумме их вероятностей;
- сумма вероятностей событий, составляющих полный набор, равна единице.

Определяя случайное событие, мы обратили внимание на то, что относительная частота (и, стало быть, эмпирическая вероятность) при неограниченном увеличении числа испытаний n стабилизируется около некоторого значения. Немецкий математик Р. Мизес предлагал определить вероятность $P(A)$ случайного события A через предел его эмпирической вероятности:

$$P(A) = \lim_{n\to\infty} \tilde{P}_n(A).$$

Однако это невозможно, так как указанного предела, строго говоря, просто не существует. Далее мы выясним, в каком именно смысле следует понимать стабилизацию относительной частоты. Другое возражение против данного определения состоит в том, что такую «вероятность» невозможно посчитать заранее, не проводя бесконечного количества экспериментов. Итак, можно сформулировать требования, которым должно удовлетворять любое математическое определение вероятности. Во-первых, это должна быть величина, которую можно вычислить для любого случайного события, не проводя опытов. Во-вторых, должны выполняться три перечисленных выше свойства, унаследованные от эмпирической вероятности.

11.2. Классическое определение вероятности

11.2.1. Классическая схема

В подпараграфе 11.2.1 будет наконец дано первое определение вероятности, понятия, которое мы анонсировали в начале главы. Повторимся: вероятность есть числовая характеристика возможности появления случайного события. При этом предполагается, что условия эксперимента могут быть воспроизведены неограниченное число раз. Это нематематическое определение носит скорее интуитивный характер. Придадим ему более точный смысл.

Рассмотрим некоторый случайный эксперимент. Пусть в результате данного эксперимента может произойти несколько исходов (случайных событий). К примеру, при бросании монеты возможны два исхода: «орел» или «решка». При бросании кубика может произойти шесть различных исходов (может выпасть число от 1 до 6). Назовем исход *благоприятным* для случайного события A, если событие A следует из такого исхода. Пусть, например, событие A состоит в том, что выпавшее на грани кубика число четно. Благоприятными для этого события будут три исхода эксперимента: «двойка», «четверка» и «шестерка».

Будем называть *равновозможными* исходы, имеющие одинаковые шансы. Равновозможность определяется нестрого, однако нам, безусловно, ясно, какие именно исходы считаются равновозможными. Например, оба исхода при бросании монеты и все шесть исходов при бросании кубика равновозможны. Равновозможность исходов есть синоним некой симметрии случайного эксперимента. Например, чтобы все исходы бросания кубика были равновозможными, его грани должны быть равноправными, т.е. кубик должен быть абсолютно симметричным по форме и однородным по внутреннему строению. Если в кубике вблизи одной из граней находится свинцовый шарик, то кубик в основном будет падать именно на эту грань: симметрия эксперимента нарушится, и исходы перестанут быть равновозможными.

Будем предполагать, что данный эксперимент имеет N возможных исходов, все они равновозможны и несовместны, т.е. никакие два из них не могут наступить одновременно. ***Вероятностью*** $P(A)$ события A называется отношение числа благоприятных исходов $m(A)$ к общему числу N несовместных равновозможных исходов:

$$\boxed{P(A)=\frac{m(A)}{N}}$$

Данное равенство называется *классическим определением вероятности.*

Вероятность можно вычислять в процентах. Например, выражения $P(A) = 90\%$ и $P(A) = 0{,}9$ эквивалентны.

Формула классической вероятности удивительно похожа на выражение для эмпирической вероятности. Однако они отличаются по существу. И числитель, и знаменатель в определении вероятности можно посчитать заранее, до опыта, а priori. Эмпирическая же вероятность, как уже не раз подчеркивалось, вычисляется только по результатам экспериментов. Итак, классическое определение удовлетворяет первому требованию, предъявляемому к определению вероятности: ее можно посчитать, не проводя экспериментов. Убедимся, что выполняются также три свойства, полученные в наследство от эмпирической вероятности.

∇1. *Для любого случайного события A*

$$0 \leqslant P(A) \leqslant 1.$$

Во-первых, по определению вероятность неотрицательна. Во-вторых, число благоприятных исходов $m(A)$ не больше общего числа исходов N. Поэтому $P(A) \leqslant 1$.

∇2. *Пусть события A и B несовместны. Тогда*

$$P(A + B) = P(A) + P(B).$$

В самом деле, пусть для события A благоприятными являются $m(A)$ исходов эксперимента, а для события B – соответственно $m(B)$ исходов. Тогда для события $A + B$ благоприятными являются все исходы, благоприятные как для A, так и для B. Поскольку события A и B несовместны, их ровно $m(A) + m(B)$ (нет исходов, благоприятных для A и B одновременно). Значит, по определению

$$P(A+B)=\frac{m(A)+m(B)}{N}=\frac{m(A)}{N}+\frac{m(B)}{N}=P(A)+P(B).$$

∇3. *Пусть события A_1, A_2, ..., A_k образуют полный набор. Тогда*

$$P(A_1) + P(A_2) + ... + P(A_k) = 1.$$

Действительно, события попарно несовместны, значит, событие $A_1 + A_2 + ... + A_k$ имеет ровно $m(A_1) + m(A_2) + ... + m(A_k)$

благоприятных исходов. Однако данное событие достоверно. Стало быть, все исходы эксперимента для него благоприятны. Поэтому $m(A_1) + m(A_2) + ... + m(A_k) = N$ и равенство выполняется.

Итак, требования, предъявляемые к определению вероятности, выполняются. Классическое определение, конечно, нельзя считать полностью удовлетворительным, так как оно содержит не совсем четкое понятие равновозможности. По сути, определение вероятности использует условие равной вероятности исходов. Однако с помощью данного определения можно практически вычислять вероятности многих событий.

Примеры.

1. Рассмотрим хорошо знакомый нам эксперимент с бросанием монеты. У него может быть два несовместных и равновозможных исхода: «орел» и «решка». Значит, вероятность появления «орла» равна $^1/_2$, так как благоприятным для этого события является один исход. На графике зависимости эмпирической вероятности от количества экспериментов стабилизация происходит как раз около этого значения. Вероятность появления решки также равна $^1/_2$.

2. Еще один знакомый пример: бросание кубика. В этом эксперименте могут случиться шесть исходов, они несовместны и равновозможны (при соблюдении условий симметрии, как это обсуждалось ранее). Значит, вероятность выпадения каждой грани равна $^1/_6$. Пусть событие A состоит в том, что выпадет четное число. Для этого события благоприятны три исхода: «двойка», «четверка» и «шестерка». Значит, $P(A) = {}^3/_6 = {}^1/_2$.

3. В урне находятся 4 белых и 6 черных шаров. Какова вероятность, что вынутый наугад шар будет белым? Всего эксперимент имеет десять исходов (можно вынуть любой из десяти шаров). Благоприятными будут четыре исхода. Значит, вероятность этого события равна $^4/_{10} = 0{,}4$. Соответственно, вероятность вынуть черный шар равна 0,6. Слово «наугад» часто используется в задачах теории вероятностей. Говоря «наугад», мы подразумеваем, что эксперимент носит случайный характер и все его исходы равновозможны. В данном случае имеется в виду, что шары каждый раз тщательно перемешиваются, что шар достают не глядя и т.д.

4. Бросают два кубика одновременно. Какова вероятность того, что на обоих кубиках выпадет одинаковое число? Исход данного эксперимента описывается упорядоченной парой чисел: первое число – то, что выпало на первом кубике, второе – то, что на втором. Нетрудно убедиться, что таких пар $\{i, j\}$, $i, j = 1 \div 6$, ровно 36. Можно их перечислить: $\{1, 1\}$, $\{1, 2\}$, ..., $\{1, 6\}$, $\{2, 1\}$, ...,$\{2, 6\}$, ..., $\{6, 1\}$, ..., $\{6, 6\}$.

Благоприятными исходами являются шесть пар с совпадающими числами. Значит, вероятность данного события равна $^6/_{36} = {}^1/_6$.

Посчитаем, например, вероятность того, что произведение выпавших на кубике чисел равно 12. Благоприятны следующие исходы: {2, 6}, {3, 4}, {4, 3}, {6, 2}. Значит, вероятность равна $^4/_{36} = {}^1/_9$. Так можно вычислить вероятность любого события, связанного с данным случайным экспериментом.

11.2.2. Комбинаторика и схемы выбора

Чтобы пользоваться классическим определением вероятности, нужно уметь подсчитывать общее число исходов эксперимента и число благоприятных исходов. Такой подсчет сводится к перебору вариантов. Раздел математики, в котором исследуются различные задачи на перебор, называется *комбинаторикой.* Рассмотрим некоторые комбинаторные задачи и выясним, как их решение может помочь при вычислении вероятности по классической формуле.

Задача о рассаживании. *Из группы в n человек требуется рассадить за столом m человек ($m \leqslant n$). Сколькими способами это можно сделать?*

Пронумеруем m стульев. Тогда на первый стул можно усадить одного из n человек. Пусть первое место уже занято. Тогда на второе остается $n - 1$ претендент. Каждая из n возможностей занять первое место сочетается с $n - 1$ возможностью для второго. Таким образом, существует $n(n - 1)$ вариантов рассаживания на первые два стула. После того как заняты первые два места, кандидат на третье место выбирается из оставшихся $n - 2$ желающих, причем каждый из них может оказаться в компании с любой из $n(n - 1)$ возможных пар, занимающих первое и второе места. Процесс рассаживания продолжается до тех пор, пока не будут заняты все m стульев. Число возможностей усадить человека на каждый следующий стул уменьшается на единицу. Последний, m-й стул может занять соответственно $n - (m - 1) = n - m + 1$ человек. Итак, всего получается $n(n - 1)(n - 2) \cdots (n - m + 1)$ вариантов рассаживания. Искомое число способов рассаживания называется ***числом размещений из n по m*** и обозначается A_n^m. Таким образом,

$$A_n^m = n(n-1)(n-2)\cdots(n-m+1).$$

В случае когда $m = n$, задача о рассаживании превращается в задачу о количестве перестановок n-элементного множества. Мы уже сталкивались с этой задачей в курсе алгебры при изучении группы

перестановок. Формула для числа размещений при $m = n$ превращается в следующее равенство:

$$A_n^n = n(n-1)(n-2)\cdots 2\cdot 1$$

(на последний стул будет претендовать один человек). Произведение n первых натуральных чисел называется «*n-факториал*» и обозначается $n!$. Например, $1! = 1$, $2! = 1\cdot 2 = 2$, $3! = 1\cdot 2\cdot 3 = 6$, $4! = 1\cdot 2\cdot 3\cdot 4 = 24$ и т.д. Чтобы избежать проблем, договариваются, что $0! = 1$ (так же, как $a^0 = 1$). Итак, мы получили, что число размещений из n по n, или число перестановок из n, равно $n!$:

$$A_n^n = n!$$

Число A_n^m также можно выразить через факториалы. Для этого умножим и разделим правую часть равенства на произведение $(n - m)(n - m - 1)\cdots 2\cdot 1$:

$$A_n^m = \frac{n(n-1)\cdots(n-m+1)(n-m)(n-m-1)\cdots 2\cdot 1}{(n-m)(n-m-1)\cdots 2\cdot 1}.$$

Нетрудно заметить, что в числителе стоит произведение всех натуральных чисел от 1 до n, т.е. $n!$, в знаменателе – произведение всех натуральных чисел от 1 до $n - m$, т.е. $(n - m)!$. Итак,

$$\boxed{A_n^m = \frac{n!}{(n-m)!}}$$

К примеру, $A_4^2 = \frac{4!}{2!} = 12$, т.е. четверых человек можно рассадить по двое 12 способами.

Задача о выборе. *Сколькими способами можно выбрать m человек из группы в n человек ($m \leqslant n$)?*

Задача очень похожа на предыдущую. Отличие состоит в том, что при рассаживании был важен порядок, в котором располагались m выбранных человек, а в данной задаче речь идет о неупорядоченном выборе. В первой задаче каждому варианту выбора m человек соответствует $m!$ вариантов рассаживания (количество перестановок из m). Значит, способов рассаживания в $m!$ раз больше, чем способов выбора. Количество способов выбора называется ***числом сочетаний из n по m*** и обозначается C_n^m. Итак,

$$C_n^m = \frac{A_n^m}{m!},$$

откуда, вспоминая формулу для числа размещений, получаем

$$\boxed{C_n^m = \frac{n!}{m!(n-m)!}}$$

Учитывая, что $0! = 1$, получаем следующие естественные равенства.

∇1. *При любом n*

$$C_n^0 = 1, \quad C_n^n = 1.$$

В самом деле, выбрать 0 человек из n (т.е. не выбрать ни одного) можно, разумеется, только одним способом. То же можно сказать и о выборе n человек из n.

Из формулы для числа сочетаний следует еще один вывод.

∇2. $$C_n^m = C_n^{n-m}.$$

Действительно, способов выбрать m человек из n ровно столько же, сколько способов оставить $n - m$ человек на месте (т.е. тоже в каком-то смысле «выбрать»).

Не столь очевидно следующее свойство числа сочетаний.

∇3. $$C_{n+1}^m = C_n^{m-1} + C_n^m, \quad m = 0, 1, \ldots, n.$$

Равенство проверяется непосредственно по формуле для числа сочетаний:

$$C_n^{m-1} + C_n^m = \frac{n!}{(m-1)!(n-m+1)!} + \frac{n!}{m!(n-m)!}.$$

Общий знаменатель этих дробей есть $m!(n - m + 1)!$. Чтобы привести к общему знаменателю, числитель первой дроби нужно умножить на m, а числитель второй дроби – на $(n - m + 1)$. Значит,

$$C_n^{m-1} + C_n^m = \frac{m \cdot n! + (n-m+1) \cdot n!}{m!(n-m+1)!} = \frac{(n+1) \cdot n!}{m!(n-m+1)!} =$$

$$= \frac{(n+1)!}{m!(n-m+1)!} = C_{n+1}^m.$$

Равенство доказано.

Последнее свойство позволяет находить C_n^m с помощью замечательного «треугольника Паскаля»:

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{array}$$

Строки треугольника нумеруются, начиная с нуля. В строке с номером n расположены числа C_n^m, $m = 0, 1, \ldots, n$. При этом каждое число в следующей $(n + 1)$-й строке вычисляется как сумма двух своих верхних соседей в n-й строке, т.е. по формуле $C_{n+1}^m = C_n^{m-1} + C_n^m$.

Числа C_n^m применяются в знаменитой формуле бинома Ньютона, *предназначенной для вычисления степени от суммы двух чисел:*

$$\boxed{(a+b)^n = C_n^0 a^n + C_n^1 a^{n-1} b + \ldots + C_n^m a^m b^{n-m} + \ldots + C_n^{n-1} a b^{n-1} + C_n^n b^n}$$

Нетрудно увидеть, что этой формуле удовлетворяют известные «школьные» формулы разложения квадрата и куба суммы:

$$(a + b)^2 = a^2 + 2ab + b^2,$$
$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3.$$

Бином Ньютона позволяет вычислить любую степень суммы. В связи с тем что числа C_n^m используются в этой известной формуле, их называют *биномиальными коэффициентами.* Из бинома Ньютона следует важное свойство биномиальных коэффициентов.

∇4. $$C_n^0 + C_n^1 + \ldots + C_n^{n-1} + C_n^n = 2^n.$$

В самом деле, это равенство непосредственно выводится из формулы бинома при $a = b = 1$: слева будет 2^n, а справа – сумма биномиальных коэффициентов, так как все степени a и b равны единице. Свойство означает, что сумма чисел в n-й строке треугольника Паскаля равна соответствующей степени двойки. В этом легко убедиться: в нулевой строке стоит одна единица (это 2^0), сумма двух единиц в первой строке есть 2^1, во второй строке единица, двойка и единица дают в сумме 4, т.е. 2^2. Сумма чисел в следующих строках равна 8, 16 и т.д.

Рассмотрим теперь, как комбинаторные формулы для числа размещений и числа сочетаний применяются в задачах теории вероятностей. Как мы уже отмечали, многие случайные события моделируются экспериментами с урной и шарами. Шары из урны можно доставать по-разному: шар можно каждый раз возвращать в урну, а можно этого не делать; выбранные шары можно упорядочивать или не упорядочивать и т.д. Таким образом, существуют различные схемы выбора. В каждой из этих схем общее число исходов и число благоприятных исходов подсчитываются по-разному. Рассмотрим основные схемы выбора и соответствующие задачи.

I. Схемы без возвращения.

1. Схема без упорядочения. Предполагается, что в урне n шаров и эксперимент состоит в том, что наугад вынимают m шаров, порядок при этом не важен. Вынуть m шаров – это все равно, что m раз вынуть по одному шару, не возвращая их обратно в урну. Поэтому такая ситуация описывается схемой без возвращения и без упорядочения. Понятно, что общее число исходов этого случайного эксперимента равно числу способов выбрать m шаров из n, т.е. числу сочетаний:

$$N = C_n^m.$$

Задача. *В урне 10 шаров, из них 3 белых. Какова вероятность того, что из четырех наугад выбранных шаров ровно один будет белый? Какова вероятность, что белых шаров будет ровно два?*

Р е ш е н и е. Всего исходов $N = C_{10}^4$. Посчитаем это число, чтобы заодно освоить технику таких вычислений:

$$N = C_{10}^4 = \frac{10!}{4!6!} = \frac{1\cdot 2\cdots 6\cdot 7\cdot 8\cdot 9\cdot 10}{4!\cdot 1\cdot 2\cdots 6}.$$

Видно, что можно сократить числитель и знаменатель на $6! = 1\cdot 2\cdots 6$:

$$N = \frac{7\cdot 8\cdot 9\cdot 10}{1\cdot 2\cdot 3\cdot 4}.$$

Осталось сделать несколько сокращений: в числителе есть восьмерка, а в знаменателе – произведение $2\cdot 4$, кроме того, можно сократить 9 и 3. В итоге получаем

$$N = 7\cdot 3\cdot 10 = 210.$$

В первом случае при благоприятном исходе среди четырех шаров один белый, а остальные три – черные (событие A). Белый шар можно выбрать тремя способами (их всего три), три черных шара можно выбрать C_7^3 способами, так как черных шаров в урне семь. Каждый из трех белых шаров может сочетаться с любой из C_7^3 черных троек (рис. 11.2). Таким образом, благоприятных исходов

$$m = 3C_7^3 = 3\frac{7!}{3!4!} = 3\frac{5\cdot 6\cdot 7}{1\cdot 2\cdot 3} = 3\cdot 5\cdot 7 = 105.$$

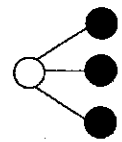


Рис. 11.2

Значит, искомая вероятность

$$P(A) = \frac{m}{N} = \frac{105}{210} = \frac{1}{2}.$$

Найдем число благоприятных исходов во втором случае (два белых, два черных шара – событие B). Пару белых шаров можно выбрать $C_3^2 = 3$ способами. Если это шары с номерами 1, 2 и 3, то соответствующие пары – {1, 2}, {1, 3} и {2, 3}. Для пары черных шаров число способов выбора

$$C_7^2 = \frac{7!}{2!5!} = \frac{6\cdot 7}{2} = 21.$$

Каждая пара белых может сочетаться с каждой парой черных (рис. 11.3). Поэтому всего благоприятных исходов $m = 3\cdot 21 = 63$. Значит, вероятность второго события (B)

$$P(B) = \frac{63}{210} = 0{,}3.$$

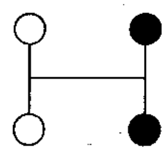


Рис. 11.3

2. Схема с упорядочением. Так же, как в предыдущем случае, наугад достают m шаров из урны с n шарами. Разница состоит в том, что теперь важен порядок, в котором вынимают шары. Можно представить, что шары пронумерованы и, вынимая по одному шару, мы записываем их номера. По-прежнему шары в урну не возвращают. Ясно, что в этом случае общее число исходов равно числу размещений из n по m:

$$N = A_n^m.$$

Задача. *В урне находятся карточки с цифрами от 0 до 5. Наугад достают две карточки и складывают подряд. Какова вероятность того, что полученное двузначное число кратно семи?*

Р е ш е н и е. Всего исходов $N = A_6^2 = 5 \cdot 6 = 30$. Благоприятные исходы — это числа 14, 21, 35, 42, т.е. $m = 4$. Значит, искомая вероятность

$$P(A) = \frac{4}{30} = \frac{2}{15}.$$

II. Схемы с возвращением.

1. Схема без упорядочения. В этом эксперименте из урны с n шарами m раз выбирают наугад по одному шару и возвращают шар обратно. Результатом опыта являются всевозможные наборы из m шаров, отличающиеся составом. Наборы из одних и тех же шаров, но расположенных в различном порядке, считаются одинаковыми (схема без упорядочения). При этом отдельные наборы могут содержать повторяющиеся элементы. Если шары пронумерованы, то, например, при $m = 3$ наборы $\{1, 1, 2\}$ и $\{1, 2, 1\}$ неразличимы, а набор $\{1, 1, 4\}$ отличается от любого из предыдущих.

Такую схему можно пояснить следующим образом. Пусть имеется n ящиков и мы хотим разложить в них m предметов. Можно класть по несколько предметов в один ящик. Пусть в первый ящик попало r_1 предметов, во второй — r_2 и т.д. Тогда результатом эксперимента будут всевозможные наборы натуральных чисел $r_1, r_2, ..., r_n$, удовлетворяющие условию

$$r_1 + r_2 + ... + r_n = m,$$

так как всего разложили m предметов. Ящики соответствуют шарам, а число r_k — это количество k-х шаров среди выбранных m шаров. Изобразим n ящиков в виде $n + 1$ черточки, а предметы — в виде m звездочек. Пусть, к примеру, $n = 4$, $m = 3$. Тогда изображение

|**| | |*|

соответствует ситуации, когда в первом ящике два предмета, во втором и в третьем — ни одного и в четвертом ящике — один предмет. Если вернуться к схеме с шарами, это означает, что среди трех выбранных шаров два имеют номер 1 и один шар с номером 4 (рис. 11.4). При таком обозначении в начале и конце обязательно стоят черточки, а остальные $n - 1$ черточка и m звездочек могут располагаться в произвольном порядке. Таким образом, задача сводится к вычислению количества способов поставить m звездочек на $n + m - 1$ возможных мест. Ясно, что число таких способов равно числу сочетаний из $n + m - 1$ по m. Это и есть общее число исходов нашего эксперимента:

$$N = C_{n+m-1}^m.$$

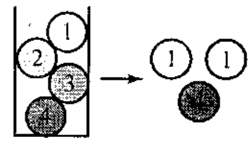


Рис. 11.4

Рассмотрим снова случай $n = 4$, $m = 3$. По формуле получается $C_{4+3-1}^3 = C_6^3 = 20$ возможных исходов. Посчитаем число исходов независимо. Во-первых, это все тройки из одинаковых элементов: $\{1, 1, 1\}$, $\{2, 2, 2\}$, $\{3, 3, 3\}$, $\{4, 4, 4\}$. Таких троек всего четыре. Во-вторых, это тройки с двумя видами шаров: $\{1, 1, 2\}$, $\{1, 1, 3\}$, $\{1, 1, 4\}$, $\{2, 2, 1\}$, $\{2, 2, 3\}$, $\{2, 2, 4\}$, $\{3, 3, 1\}$, $\{3, 3, 2\}$, $\{3, 3, 4\}$, $\{4, 4, 1\}$, $\{4, 4, 2\}$, $\{4, 4, 3\}$. Их 12. И наконец, это четыре тройки из разных шаров: $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 3, 4\}$ и $\{2, 3, 4\}$. Всего действительно получилось $4 + 12 + 4 = 20$ вариантов.

Задача. *В кондитерской продается семь видов пирожных. Очередной покупатель выбил чек на четыре пирожных. Найти вероятность того, что заказаны:*

а) пирожные одного вида;

б) пирожные разных видов;

в) по два пирожных разных видов.

Р е ш е н и е. Выбор пирожных укладывается в схему выбора с возвращением без упорядочения. Поэтому общее число исходов равно $N = C_{7+4-1}^4 = C_{10}^4 = 210$. В первом случае благоприятных исходов 7 (наборы из пирожных каждого из семи видов). Значит, вероятность

$$P(A) = \frac{7}{210} = \frac{1}{30}.$$

Во втором случае благоприятными являются всевозможные наборы из четырех различных пирожных, выбранных из семи (порядок не важен). Ясно, что это число сочетаний из 7 по 4:

$$m = C_7^4 = \frac{7!}{3!4!} = \frac{5 \cdot 6 \cdot 7}{1 \cdot 2 \cdot 3} = 35.$$

Поэтому вероятность второго события

$$P(B)=\frac{35}{210}=\frac{1}{6}.$$

Рассмотрим третий случай. Благоприятный исход представляет собой две пары одинаковых пирожных. Таких наборов ровно столько, сколько различных *пар* можно составить из семи пирожных, т.е. $m = C_7^2 = 21$. Значит, вероятность этого события

$$P(C)=\frac{21}{210}=0{,}1.$$

2. Схема с упорядочением. Снова выбирают m шаров из n, возвращая шары обратно. Однако в этом случае важно, в каком порядке доставали шары. На первом месте может быть любой из n шаров, на втором – тоже любой и т.д. При этом каждый шар может комбинироваться с каждым. Всего получается

$$N=\underbrace{n\cdot n\cdots n}_{m}=n^m$$

возможных исходов.

Задача. *Телефонная книга раскрывается наудачу и выбирается случайный номер телефона. Считая, что телефонные номера состоят из семи цифр, причем все комбинации равновероятны, найти вероятность того, что все цифры в номере различны.*

Р е ш е н и е. Заметим, что условие задачи разрешает любые номера (такие, например, как 0012413, 0123456 и даже 0000000). Поскольку всего цифр 10, а номера семизначные, общее число номеров равно $N = 10^7 = 10\ 000\ 000$. На самом деле, это все целые числа от 0 до 9 999 999. Разумеется, их десять миллионов. Благоприятные исходы составляют все различные наборы из семи цифр, отличающиеся также порядком. Значит, благоприятных исходов

$$m=A_{10}^7=\frac{10!}{3!}.$$

Итак,

$$P(A)=\frac{4\cdot5\cdot6\cdot7\cdot8\cdot9\cdot10}{10\cdot10\cdot10\cdot10\cdot10\cdot10\cdot10}=\frac{2\cdot3\cdot7\cdot9}{5\cdot5\cdot5\cdot5\cdot10}=0{,}06048.$$

Из рассмотренных примеров и задач ясно, что схема выбора шаров из урны – не более чем образ, удобный способ описания многих случайных экспериментов, возможно, и не связанных с урной и шарами.

11.2.3. Геометрическая вероятность

До сих пор мы рассматривали случайные эксперименты с конечным числом исходов (задачи с монетой, кубиком и урной). Приведем теперь пример эксперимента, число исходов в котором бесконечно. Классическим в теории вероятностей является эксперимент со случайным «бросанием точки». Итак, пусть на координатной плоскости задана некоторая область Ω, например квадрат с вершинами в начале координат и в точке $\{1, 1\}$ (рис. 11.5). Мы, зажмурившись, бросаем точку в эту область. Говорят, что точка брошена н а у г а д. Под этим подразумевают, что все исходы такого эксперимента равновозможны, все точки области Ω равноправны. Если мы специально не прицеливаемся, такой эксперимент вполне можно считать случайным. Его исход изображается координатами точки попадания, т.е. парой чисел $\{x, y\}$. В частности, для единичного квадрата $0 \leqslant x,\ y \leqslant 1$. Понятно, что эксперимент имеет бесконечно много исходов, при этом исходы несовместны и равновозможны. Пусть внутри области Ω задана область $A \subseteq \Omega$. Это может быть, например, маленький квадрат со стороной $^1/_2$ в левом нижнем углу квадрата Ω. Будем считать, что произошло событие A (обозначим его так же, как и область), если брошенная наугад точка попала в область A. Какова вероятность события A?

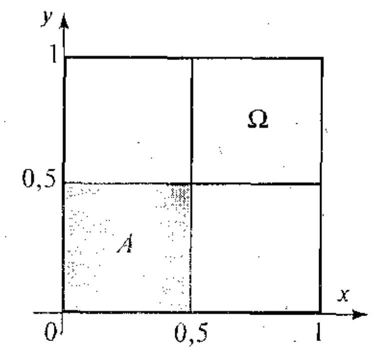


Рис. 11.5

Попытаемся распространить классическую схему на случай с бесконечным числом исходов. Общее число исходов естественно представить как площадь $S(\Omega)$ всей области Ω. Тогда благоприятным исходам соответствует площадь $S(A)$ области A. Вероятность по классическому определению равна отношению числа благоприятных исходов к общему числу несовместных и равновозможных

10 Зак. № 2435 Турецкий

исходов. В нашем случае получается отношение соответствующих площадей:

$$\boxed{P(A)=\frac{S(A)}{S(\Omega)}}$$

В частности, площадь единичного квадрата $S(\Omega) = 1$, площадь маленького квадратика $S(A) = {}^1/_4$ и, стало быть, $P(A) = {}^1/_4$.

Сформулируем определение геометрической вероятности в более общем виде.

> *Вероятность* случайного события равна отношению площади области, благоприятствующей данному событию, к площади, соответствующей всевозможным исходам эксперимента.

Понятно, что, если точку бросают не в область на плоскости, а на отрезок прямой, площадь следует заменить на длину; если в область пространства, то роль площади будет играть объем и т.д. Не вдаваясь в подробности, скажем, что мы столкнулись здесь с важным математическим понятием меры множества. Вместо слов «число исходов», «площадь области», «длина отрезка», «объем тела» можно сказать «мера множества». В каком-то смысле мера обобщает понятие количества элементов множества: мы измеряем множество (в частности, множество точек на плоскости) с помощью площади соответствующей области.

Заметим, что геометрическая вероятность удовлетворяет свойствам эмпирической вероятности, которые мы признали обязательными для любого математического определения вероятности. Отношение площадей неотрицательно и не превышает единицы, так как площадь части области не больше площади всей области. Несовместные события в геометрической схеме описываются непересекающимися областями. Площадь объединения таких областей равна сумме их площадей. Поэтому геометрическая вероятность суммы несовместных событий равна сумме их вероятностей. Наконец, полный набор событий изображается разбиением области Ω на непересекающиеся области, объединение которых совпадает со всей областью Ω. Поэтому сумма геометрических вероятностей событий, образующих полный набор, равна единице.

Приведем примеры задач, которые сводятся к вычислению геометрической вероятности.

1. **Задача о встрече.** *Вася и Оля условились о свидании у памятника Пушкину между девятью и десятью часами. Они договорились, что каждый ждет другого в течение 15 минут, а затем уходит. Какова вероятность, что они встретятся?*

Пусть x – момент прихода Васи (отсчитываем его от девяти часов в долях часа: например, если он пришел полдесятого, то $x = 0{,}5$), y – момент Олиного прихода. Исход эксперимента можно описать парой чисел $\{x, y\}$ или точкой на плоскости с такими координатами. При этом обе координаты могут принимать любые значения от 0 до 1. Таким образом, область Ω, соответствующая всем возможным исходам, – это единичный квадрат (рис. 11.6). Благоприятный исход (когда Вася и Оля встретятся) означает, что между моментами x и y прошло не больше 15 минут, или четверти часа. Это описывается неравенством $|x - y| \leqslant {}^1/_4$. Раскрывая модуль, получим $x - {}^1/_4 \leqslant y \leqslant x + {}^1/_4$. Этому неравенству удовлетворяют точки, лежащие между прямыми $y = x - {}^1/_4$ и $y = x + {}^1/_4$ (заштрихованная область A на рис. 11.6). Нетрудно посчитать площадь благоприятной области. Она равна площади квадрата минус площадь двух незаштрихованных прямоугольных треугольников. Катеты треугольников равны ${}^3/_4$, и их суммарная площадь равна $({}^3/_4)^2 = {}^9/_{16}$. Значит, искомая площадь $S(A) = 1 - {}^9/_{16} = {}^7/_{16}$. Вероятность встречи

$$P(A)=\frac{S(A)}{S(\Omega)}=\frac{7/16}{1}=7/16.$$

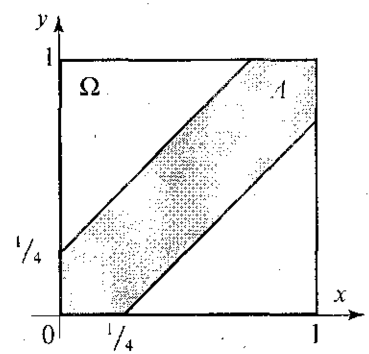


Рис. 11.6

2. **Задача Бюффона.** *На плоскость равномерно нанесены параллельные прямые, расстояние между которыми равно h. На эту плоскость наугад бросают иглу длиной l ($l \leqslant h$). Какова вероятность того, что игла пересечет одну из параллельных прямых?*

Исход эксперимента (т.е. положение брошенной наугад иглы) можно описать двумя числами: d – расстояние от центра иглы A до ближайшей прямой, и φ – угол между прямой и иглой (рис. 11.7). Величина d может принимать значения от 0 до $h/2$ (половина расстояния между прямыми), а угол φ может меняться от 0 до $\pi/2$. Итак, область Ω всевозможных исходов есть прямоугольник на плоскости $\{\varphi, d\}$, площадь которого равна $S(\Omega) = \pi h/4$. При каких же значениях d и φ игла пересечет прямую? Построим прямоугольный треугольник ABC с гипотенузой AB, совпадающей с половиной иглы. Игла пересекает прямую, если катет AC этого треугольника больше, чем расстояние d от точки A до прямой. Очевидно, что катет

$$AC = \frac{l}{2}\sin\varphi.$$

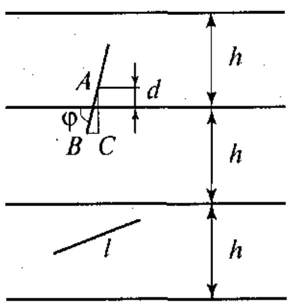


Рис. 11.7

Значит, игла пересекает прямую при условии

$$d \leqslant \frac{l}{2}\sin\varphi,$$

т.е. область, благоприятная для пересечения прямой, находится на плоскости ниже графика функции $d = \frac{l}{2}\sin\varphi$ (заштрихованная область на рис. 11.8). Осталось найти площадь благоприятной области. Вспомним, что площадь под графиком функции равна определенному интегралу от этой функции. Таким образом,

$$S(A) = \int_0^{\pi/2} \frac{l}{2}\sin\varphi\, d\varphi = \frac{l}{2}(-\cos\varphi)\Big|_0^{\pi/2} = \frac{l}{2}.$$

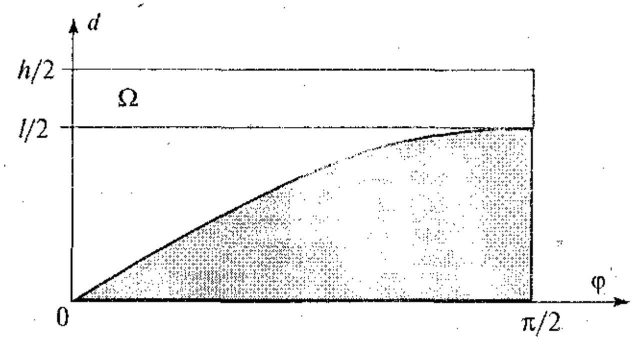


Рис. 11.8

Итак, вероятность того, что игла пересечет прямую, равна

$$P(A) = \frac{S(A)}{S(\Omega)} = \frac{l/2}{\pi h/4} = \frac{2l}{\pi h}.$$

Интересно, что решение задачи Бюффона подсказывает способ экспериментального определения числа π. В самом деле, если бросать иглу много раз, то относительная частота пересечений $\nu_n(A)$ стабилизируется около значения вероятности $P(A)$. Поэтому при большом числе экспериментов n можно считать, что

$$\pi \approx \frac{2l}{\nu_n(A)h}.$$

Такие эксперименты проводились. Например, в 1901 г. Лаццарини проделал 3408 бросаний иглы и получил приближенное значение $\pi = 3{,}1415929$, фантастически близкое к истинному. Кстати, автор данной задачи Жорж Луи Бюффон, впервые определивший геометрическую вероятность в 1777 г., известен как автор 36 увлекательно написанных томов «Естественной истории», а также знаменитого рассуждения о стиле («Стиль – это человек»).

Итак, мы рассмотрели классическое определение вероятности и его обобщение в виде геометрической вероятности. Ограниченность классической схемы равновозможных исходов очевидна. Далеко не все реальные случайные эксперименты обладают той симметрией, которая необходима для применения классической формулы.

В параграфе 11.3 вероятность вводится с помощью набора аксиом. Это еще один пример аксиоматического подхода, знакомого нам по курсу алгебры. Сначала мы рассматривали свойства векторов на плоскости и в пространстве, затем ввели n-мерные векторы, после чего перешли к абстрактному векторному пространству, свойства которого постулировались. Таким образом обычный вектор становился частным случаем более общего математического понятия. Ход рассуждений при определении вероятности вполне укладывается в такую схему. Сначала исследуются свойства эмпирической вероятности, затем вводятся классическая и геометрическая вероятности, и, наконец, вероятность определяется аксиоматически.

11.3. Аксиоматический подход к вероятности

11.3.1. Пространство элементарных исходов

Из вышесказанного ясно, что любое случайное событие можно рассматривать только в связи с некоторым случайным экспериментом. В результате этого эксперимента можно однозначно определить, произошло данное событие или не произошло. Каждый случайный эксперимент характеризуется некоторым набором возможных исходов (не обязательно конечным). Различают две разновидности исходов. *Сложный* (или *составной*) исход можно представить в виде суммы других исходов. Например, при бросании кубика исход «выпало четное число» есть, как мы знаем, сумма трех исходов, которые коротко можно назвать «двойка», «четверка» и «шестерка». В противоположность сложному *элементарный* исход случайного эксперимента неразложим в сумму других исходов. Предполагается также, что два различных элементарных исхода не могут произойти одновременно, т.е. они попарно несовместны. Заметим, что понятие элементарного исхода определяется на интуитивном уровне. Это базовое понятие, не имеющее строгого математического определения, как точка или множество. Тем не менее каждому случайному эксперименту мы безошибочно можем приписать совокупность всех элементарных исходов:

- они неразложимы в сумму других исходов;
- они попарно несовместны;
- никаких других элементарных исходов в результате данного эксперимента произойти не может.

Эту совокупность принято называть ***пространством элементарных исходов*** (или элементарных событий) и обозначать Ω. Отметим, что совокупность элементарных исходов очень напоминает совокупность всех исходов в классическом определении вероятности. Однако в данном случае не предполагается, что все элементарные исходы

равновозможны. Если это условие выполняется, то пространство элементарных исходов совпадает с совокупностью всех возможных исходов эксперимента (а вероятность, как мы убедимся далее, совпадает с классической вероятностью). Фактически многие примеры пространств элементарных исходов уже приведены ранее.

Примеры.

1. **Бросание монеты.** В данном эксперименте возможны два несовместных элементарных исхода, которые можно обозначить как «О» («орел») и «Р» («решка»). Значит, $\Omega = \{О, Р\}$.

2. **Бросание кубика.** Пусть исход ω_k, $k = 1, 2, \ldots, 6$ состоит в том, что выпало число k. Тогда $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$.

3. **Бросание двух монет.** Результат этого эксперимента – пара, состоящая из «орлов» и «решек». Всего возможно четыре элементарных исхода: $\Omega = \{ОО, ОР, РО, РР\}$. Заметим, что данный эксперимент соответствует схеме с возвращением и упорядочением: можно представить себе урну с двумя шарами, обозначенными «О» и «Р», из которой 2 раза достают шар и каждый раз возвращают его обратно, запоминая, в каком порядке вынимались шары. Количество исходов (т.е. число элементов пространства Ω) равно $n^m = 2^2 = 4$. То же самое пространство получится при двукратном бросании одной монеты.

Подобной схемой описывается одновременное бросание n монет или n-кратное подбрасывание одной монеты (2^n элементарных исходов), бросание n кубиков (6^n исходов) и т.д. Например, при бросании двух кубиков пространство элементарных исходов состоит из 36 исходов и описывается парами натуральных чисел от 1 до 6:

$$\Omega = \{(i, j), \quad i, j = 1, 2, \ldots, 6\}.$$

4. **Раскладывание предметов по ящикам.** Пусть m предметов случайным образом раскладываются по n ящикам ($m \leqslant n$, в ящик кладут по одному предмету). Данный эксперимент описывается схемой выбора без возвращения с упорядочением: из n ящиков наугад выбирают m штук – те, в которые положат предметы, а затем раскладывают предметы в произвольном порядке. Такой эксперимент имеет A_n^m элементарных исходов. Например, при $m=2$, $n=3$ будет $A_3^2 = 2 \cdot 3 = 6$ элементарных исходов. Если ящики обозначать черточками, а предметы буквами a и b, то

$$\Omega = \{|a|b|\ |, |b|a|\ |, |a|\ |b|, |b|\ |a|, |\ |a|b|, |\ |b|a|\}.$$

Если предметы неразличимы, получится схема без возвращения и без упорядочения: выбирают ящики, а порядок расположения предметов неважен. В этом случае пространство Ω содержит C_n^m элементарных исходов. В частности, при $m = 2$, $n = 3$

$$\Omega = \{|\ast|\ast|\quad|,\ |\ast|\quad|\ast|,\ |\quad|\ast|\ast|\}$$

(звездочки обозначают неразличимые предметы). Если разрешить класть по нескольку неразличимых предметов в один ящик, эксперимент описывается схемой с возвращением и без упорядочения. Число элементарных исходов в этом случае равно C_{n+m-1}^m. При $m = 2$, $n = 3$ получается $C_4^2 = 6$ элементарных исходов:

$$\Omega = \{|\ast|\ast|\quad|,\ |\ast|\quad|\ast|,\ |\quad|\ast|\ast|,\ |\ast\ast|\quad|\quad|,\ |\quad|\ast\ast|\quad|,\ |\quad|\quad|\ast\ast|\}.$$

5. **Стрельба по мишени.** В примерах 1–4 пространство элементарных исходов состояло из конечного числа исходов. В данном случае каждый элементарный исход – попадание в конкретную точку мишени, поэтому пространство Ω бесконечно. Если мишень представить как область на координатной плоскости, элементарный исход можно изобразить в виде координат точки попадания. Например, для круглой мишени с радиусом 1 м и центром в начале координат

$$\Omega = \{(x, y) \mid x^2 + y^2 \leqslant 1\}$$

(недаром мы обозначаем одной и той же буквой пространство элементарных исходов и область попадания в геометрическом определении вероятности).

11.3.2. Алгебра случайных событий

Установим теперь, как обещали, связь между случайными событиями и множествами. С каждым случайным экспериментом мы связали его пространство (множество) элементарных исходов. Теперь случайное событие удобно и естественно трактовать как п о д м н о ж е с т в о пространства Ω. Будем считать, что случайное событие A состоит из тех элементарных исходов, из которых это событие следует:

$$A = \{\omega \in \Omega \mid \omega \subset A\}.$$

Другими словами, говорят, что событие A *произошло* в том и только в том случае, когда в результате данного случайного эксперимента наступил один из элементарных исходов $\omega \in \Omega$, из которого следует событие A.

Рассмотрим, например, такие события, связанные с бросанием кубика (через x обозначим выпавшее число):

A = «x четно»;
B = «x кратно трем»;
C = «$x > 3$»;
D = «$2 \leqslant x < 5$».

Тогда, если ω_k снова обозначает элементарный исход, состоящий в том, что выпало число k, данные события можно изобразить следующими подмножествами Ω:

$A = \{\omega_2, \omega_4, \omega_6\}$;
$B = \{\omega_3, \omega_6\}$;
$C = \{\omega_4, \omega_5, \omega_6\}$;
$D = \{\omega_2, \omega_3, \omega_4\}$.

Интерпретируя случайное событие как некоторое множество элементарных исходов (подмножество пространства Ω), легко перевести на язык множеств операции и свойства случайных событий (см. подпараграф 11.1.2).

Событие B *следует* из события A, если соответствующее множество A есть п о д м н о ж е с т в о множества B.

Достоверному событию соответствует в с е п р о с т р а н с т в о Ω: это событие наступает в результате данного эксперимента всегда, поскольку всегда имеет место один из элементарных исходов $\omega \in \Omega$. Пространство элементарных исходов играет роль универсального множества по отношению к случайным событиям, связанным с данным экспериментом. *Невозможное* событие, разумеется, описывается п у с т ы м м н о ж е с т в о м ∅.

Сумма событий A и B есть не что иное, как о б ъ е д и н е н и е соответствующих множеств:

$$A + B = A \cup B.$$

В самом деле, по определению сумма событий – это событие, состоящее в том, что произошло событие A или событие B. Значит, имеет место один из элементарных исходов, влекущий событие A или событие B. Это и означает, что событию $A + B$ соответствует множество элементарных исходов, принадлежащих множеству A или множеству B, т.е. именно объединение этих множеств. Например, в предыдущих обозначениях

$$A + C = \{\omega_2, \omega_4, \omega_6\} \cup \{\omega_4, \omega_5, \omega_6\} = \{\omega_2, \omega_4, \omega_5, \omega_6\}.$$

Действительно, событие $A + C$ = «x четно или $x > 3$» наступает тогда и только тогда, когда выпадает двойка, четверка, пятерка или шестерка.

Произведение событий есть п е р е с е ч е н и е соответствующих множеств:

$$AB = A \cap B.$$

Это те элементарные исходы, из которых следуют оба события A и B. Например,

$$AC = \{\omega_2, \omega_4, \omega_6\} \cap \{\omega_4, \omega_5, \omega_6\} = \{\omega_4, \omega_6\},$$

т.е. событие AC = «x четно и $x > 3$» происходит лишь при двух элементарных исходах: при четверке и шестерке.

События A и B *несовместны*, если пересечение соответствующих множеств пусто: $A \cap B = \varnothing$.

Событие, *противоположное* событию A, есть дополнение $\overline{A}$ множества A. Смысл этого события в том, что не произошло событие A, поэтому ему соответствует множество всех элементарных исходов, из которых не следует A. Например,

$$\overline{C} = \overline{\{\omega_4, \omega_5, \omega_6\}} = \{\omega_1, \omega_2, \omega_3\}$$

есть событие «$x \leqslant 3$».

События образуют *полный набор*, если соответствующие множества не пересекаются, а их объединение совпадает со всем пространством Ω. Например, при бросании кубика события E = «$x \leqslant 2$», F = «$2 < x \leqslant 4$» и G = «$x \geqslant 5$» составляют полный набор:

$$E = \{\omega_1, \omega_2\};$$
$$F = \{\omega_3, \omega_4\};$$
$$G = \{\omega_5, \omega_6\}$$

и

$$E \cup F \cup G = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\} = \Omega.$$

Алгебраические свойства случайных событий (коммутативность, ассоциативность и дистрибутивность) теперь не нуждаются в специальном обосновании и вытекают из соответствующих свойств множеств. Более того, на случайные события распространяются и все остальные свойства множеств. Перечислим их еще раз.

1. $A + B = B + A$	1′. $AB = BA$
2. $A + (B + C) = (A + B) + C$	2′. $A(BC) = (AB)C$
3. $A(B + C) = AB + AC$	3′. $A + BC = (A + B)(A + C)$
4. $A + \varnothing = A$	4′. $A\Omega = A$
5. $A + \Omega = \Omega$	5′. $A\varnothing = \varnothing$
6. $A + A = A$	6′. $AA = A$
7. $A + \overline{A} = \Omega$	7′. $A\overline{A} = \varnothing$
8. $\overline{\Omega} = \varnothing$	8′. $\overline{\varnothing} = \Omega$
9 и 9′. $\overline{\overline{A}} = A$	
10. $A + AB = A$	10′. $A(A + B) = A$
11. $\overline{A+B} = \overline{A}\,\overline{B}$	11′. $\overline{AB} = \overline{A} + \overline{B}$

Свойствам можно придать естественные словесные формулировки. Например, свойство 7: в результате данного эксперимента любое событие либо происходит, либо нет. Свойство 7′: событие не может одновременно произойти и не произойти.

Свойства объединены попарно по признаку двойственности, так же как это делалось при формулировке свойств множеств и логических операций. Итак, главный вывод, к которому мы пришли: случайные события отождествляются с подмножествами пространства элементарных исходов. В дальнейшем можно не делать различий между событиями и соответствующими множествами. Более того, случайное событие можно определять как подмножество Ω. Ниже устанавливается соответствие между понятиями теории множеств и теории вероятностей (подобно тому, как это делалось в логике для высказываний):

Теория множеств	**Теория вероятностей**
Множество	Случайное событие
Объединение $A \cup B$	Сумма $A + B$
Пересечение $A \cap B$	Произведение AB
Непересекающиеся множества	Несовместные события
Разбиение	Полный набор событий
Дополнение	Противоположное событие
Универсальное множество	Достоверное событие Ω
Пустое множество	Невозможное событие

Далее будет введена вероятность случайного события как функция, заданная на подмножествах пространства Ω. Прежде чем определять эту функцию, следует задать ее область определения. Разумно поставить следующее условие: если мы умеем вычислять вероятность событий A и B, мы должны уметь вычислять вероятность событий $A + B$, AB, а также вероятность противоположных событий. Для этого область определения функции вероятности должна быть замкнута относительно этих операций.

Система $\mathcal{F}$ подмножеств некоторого множества Ω называется ***алгеброй*** при выполнении следующих условий:

1) $\Omega \in \mathcal{F}$;

2) если $A \in \mathcal{F}$, $B \in \mathcal{F}$, то:

- $A \cup B \in \mathcal{F}$;
- $A \cap B \in \mathcal{F}$;
- $A \setminus B \in \mathcal{F}$.

Таким образом, алгебра множеств – это система подмножеств некоторого множества Ω, замкнутая относительно операций объединения, пересечения и дополнения.

∇**1.** *Пустое множество принадлежит любой алгебре* $\mathcal{F}$:

$$\varnothing \in \mathcal{F}.$$

В самом деле, поскольку по определению $\Omega \in \mathcal{F}$ и алгебра замкнута относительно дополнения, то $\Omega \setminus \Omega = \varnothing \in \mathcal{F}$.

∇**2.** *Если* $A \in \mathcal{F}$, *то* $\overline{A} \in \mathcal{F}$.

Действительно, из того, что $A \in \mathcal{F}$ и $\Omega \in \mathcal{F}$, следует, что $\Omega \setminus A = \overline{A} \in \mathcal{F}$.

Из определения ясно, что одно и то же множество Ω порождает различные алгебры. Самая «бедная» алгебра состоит всего из двух множеств – пустого и самого Ω:

$$\mathcal{F}_0 = \{\varnothing, \Omega\}.$$

Любое подмножество $A \subset \Omega$ порождает четырехэлементную алгебру:

$$\mathcal{F}_A = \{\varnothing, \Omega, A, \overline{A}\}.$$

Множество-степень $\mathcal{P}(\Omega)$ множества Ω (совокупность всех подмножеств Ω) также образует алгебру. Это самая «богатая» алгебра, порождаемая множеством Ω. В дальнейшем удобно работать именно с этой алгеброй $\mathcal{F} = \mathcal{P}(\Omega)$. Итак, все случайные события, связанные с пространством элементарных исходов, образуют алгебру случайных событий (подмножеств Ω).

11.3.3. Аксиоматическое определение вероятности

Теперь, когда мы четко определили математический статус случайного события, настало время дать строгое определение вероятности (систему аксиоматического обоснования построил А.Н. Колмогоров в 1933 г.). Числовая функция $P(A)$, заданная на алгебре $\mathcal{F}$ подмножеств пространства элементарных исходов Ω, называется ***вероятностью*** случайного события A, если она удовлетворяет следующим свойствам (аксиомам):

А1. $P(A) \geqslant 0 \quad \forall A \in \mathcal{F}$.

А2. $P(\Omega) = 1$.

А3. *Если* $A \cap B = \varnothing$, *то*

$$P(A \cup B) = P(A) + P(B).$$

Последнее свойство называется *аксиомой конечной аддитивности.*

На языке случайных событий две последние аксиомы вероятности можно сформулировать следующим образом.

А2. *Вероятность достоверного события равна единице.*

А3. *Если события A и B несовместны, то*

$$P(A + B) = P(A) + P(B).$$

Тройка $\{\Omega, \mathcal{F}, P\}$ называется *вероятностным пространством*, связанным с данным случайным экспериментом.

Андрей Николаевич Колмогоров (1903–1987). Выдающийся советский математик. Родился в Тамбове. Окончил Московский университет (1925). С 1931 г. – профессор Московского университета. С 1939 г. – академик.

Построил аксиоматику теории вероятностей и внес большой вклад в теорию случайных процессов. Получил важные результаты в логике, топологии, теории дифференциальных уравнений, функциональном анализе, механике. Колмогорову принадлежат многочисленные работы по применению математических методов в военном деле (теория стрельбы), биологии и технике, а также работы по применению статистических методов к задачам контроля качества продукции, к математической лингвистике. Много работал над усовершенствованием школьных программ по математике. Автор и редактор школьных учебников. С 1968 г. – член Академии педагогических наук.

А.Н. Колмогоров был президентом Московского математического общества (1964–1966). Иностранный член Парижской АН, Лондонского королевского общества и ряда других академий наук. Лауреат Государственной премии СССР (1941), Ленинской премии (1965).

Пример.

Рассмотрим бросание кубика. Пространство элементарных исходов содержит шесть исходов ω_k, соответствующих выпадению числа k:

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}.$$

В качестве алгебры возьмем множество-степень $\mathcal{F} = \mathcal{P}(\Omega)$. Тогда любое случайное событие представляет собой конечное множество элементарных исходов ω_k. Обозначим через $m(A)$ количество элементарных исходов, принадлежащих данному событию A (для невозможного события положим $m(\varnothing) = 0$). Например, для события A = «выпало четное число» = $\{\omega_2, \omega_4, \omega_6\}$ это число равно $m(A) = 3$. Определим функцию вероятности следующим образом:

$$P(A) = \frac{m(A)}{6}.$$

Убедимся, что такая функция удовлетворяет трем аксиомам вероятности. В самом деле, эта функция, очевидно, неотрицательна (выполняется первая аксиома). Кроме того,

$$m(\Omega) = m(\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}) = 6 \text{ и } P(\Omega) = 6/6 = 1$$

(выполняется вторая аксиома). Далее, пусть события A и B несовместны, т.е. $A \cap B = \varnothing$. Тогда объединение множеств A и B содержит все элементарные исходы множества A и множества B. Значит, $m(A + B) = m(A) + m(B)$ и

$$P(A+B)=\frac{m(A)+m(B)}{6}=\frac{m(A)}{6}+\frac{m(B)}{6}=P(A)+P(B)$$

(выполняется третья аксиома).

Нетрудно заметить, что определенная таким образом вероятность совпадает с классической вероятностью, равной отношению числа благоприятных исходов к общему числу несовместных и равновозможных исходов. Равновозможность в данном случае имеет смысл равновероятности:

$$P(\omega_1) = P(\omega_2) = \ldots = P(\omega_6) = {}^1/_6.$$

В самом деле, общее число исходов $N = 6$, а исходы, составляющие множество A, – это и есть исходы, благоприятные для события A (те, из которых это событие следует).

Однако можно построить и другую функцию вероятности. Зададим, например, шесть чисел $p_k \geqslant 0$, $k = 1, 2, \ldots, 6$, таких, что

$$p_1 + p_2 + \ldots + p_6 = 1.$$

Эти числа можно трактовать как вероятности элементарных исходов ω_k. Вероятность любого события A можно определить как сумму вероятностей элементарных исходов, составляющих это событие. Если событие A состоит из элементарных исходов $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_s}$, полагаем

$$P(A) = p_{i_1} + p_{i_2} + \ldots + p_{i_s}.$$

Все аксиомы вероятности легко проверяются. В частности, при $p_1 = p_2 = \ldots = p_6 = {}^1/_6$ получается предыдущий случай классической вероятности. Ясно, что существует бесконечно много наборов неотрицательных чисел p_k, сумма которых равна единице. Различным наборам соответствуют разные функции вероятности $P(A)$. Значит, можно сделать следующий вывод.

∇**3.** *Вероятность не определяется однозначно системой аксиом.*

> Выбор конкретной модели (конкретного вероятностного пространства) осуществляется на основе не вошедших в систему аксиом дополнительных соображений с проверкой практикой и опытом.

Например, симметричный и однородный кубик порождает классическую схему с $p_1 = p_2 = \ldots = p_6 = {}^1/_6$. Если к грани, противоположной шестерке, прикреплена свинцовая пластинка, т.е. шестерка выпадает гораздо чаще, чем все остальные грани, можно положить, например, $p_1 = p_2 = \ldots = p_5 = 0{,}1$, $p_6 = 0{,}5$ или даже $p_1 = p_2 = \ldots = p_5 = 0$, $p_6 = 1$. Все зависит от конкретных условий случайного эксперимента.

Если множество $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ конечно и $\mathcal{F} = \mathcal{P}(\Omega)$, из аксиомы конечной аддитивности следует, что набор неотрицательных чисел $p_1, p_2, \ldots, p_n$, сумма которых равна единице, полностью определяет вероятность любого события $A \in \mathcal{F}$.

Из аксиом вероятности вытекают следующие важные следствия.

1°. *Если события $A_1, A_2, \ldots, A_k$ попарно несовместны, то*

$$P(A_1 + A_2 + \ldots + A_k) = P(A_1) + P(A_2) + \ldots + P(A_k).$$

Равенство следует из ассоциативности сложения событий и аксиомы конечной аддитивности А3.

2°. *Если события $A_1, A_2, \ldots, A_k$ образуют полный набор, то*

$$P(A_1) + P(A_2) + \ldots + P(A_k) = 1.$$

Это следует из свойства 1° и аксиомы А2: поскольку события образуют полный набор, они попарно несовместны, а их сумма совпадает с Ω. Значит,

$$P(A_1) + P(A_2) + \ldots + P(A_k) = P(A_1 + A_2 + \ldots + A_k) = P(\Omega) = 1.$$

3°. $P(\overline{A}) = 1 - P(A)$.

Действительно, $A\overline{A} = \varnothing$ (эти события несовместны) и $A + \overline{A} = \Omega$ (они образуют полный набор). Значит, $P(A) + P(\overline{A}) = 1$, откуда и вытекает данное равенство. Например, если событие A = «выпадает число, кратное трем» = $\{\omega_3, \omega_6\}$, то в классической схеме $P(A) = 2/6 = {}^1/_3$. При этом $\overline{A}$ = «выпадает число, не кратное трем» = $\{\omega_1, \omega_2, \omega_4, \omega_5\}$ и $P(\overline{A}) = 4/6 = {}^2/_3 = 1 - P(A)$.

4°. *Вероятность невозможного события равна нулю:*

$$P(\varnothing) = 0.$$

Равенство следует из свойства 3° и аксиомы А2:

$$P(\varnothing) = P(\overline{\Omega}) = 1 - P(\Omega) = 1 - 1 = 0.$$

5°. *Вероятность любого события не превосходит единицы:*

$$P(A) \leqslant 1.$$

В самом деле, по свойству 3° вероятность $P(A) = 1 - P(\overline{A})$. Однако в силу аксиомы А1 имеем $P(\overline{A}) \geqslant 0$. Значит, $P(A) \leqslant 1$.

Итак, аксиоматически определенная вероятность удовлетворяет всем условиям, которые должны быть унаследованы от эмпирической вероятности, а именно:

- вероятность принимает значения между нулем и единицей;
- вероятность суммы несовместных событий равна сумме их вероятностей;
- сумма вероятностей событий, образующих полный набор, равна единице.

Кроме того, вероятность любого события $A \in \mathcal{F}$ определяется до опыта, не требует для своего вычисления проведения эксперимента. Это вполне согласуется с общей концепцией, согласно которой вероятность рассматривается как величина шанса, присущая случайному событию a priori.

6°. *Если событие B следует из события A, то $P(A) \leqslant P(B)$.*

Если $A \subset B$, то справедливо тождество

$$B = A \cup (\overline{A} \cap B).$$

Это нетрудно доказать непосредственно по определению или убедиться с помощью диаграммы Венна. На рис. 11.9 большой круг – это множество B, круг поменьше внутри его – множество A. Тогда заштрихованная область изображает множество $\overline{A} \cap B$ (точки множества B, не принадлежащие A). Очевидно, большой круг есть объединение маленького круга и заштрихованной области.

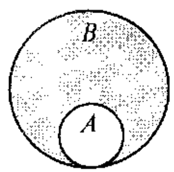


Рис. 11.9

Кроме того, ясно, что множества A и $\overline{A} \cap B$ не пересекаются: $A \cap (\overline{A} \cap B) = (A \cap \overline{A}) \cap B = \varnothing \cap B = \varnothing$. Значит, в силу аксиомы конечной аддитивности

$$P(B) = P(A \cup (\overline{A} \cap B)) = P(A) + P(\overline{A} \cap B) \geqslant P(A).$$

7° (теорема сложения). *Вероятность суммы событий равна сумме вероятностей минус вероятность их произведения:*

$$P(A + B) = P(A) + P(B) - P(AB).$$

Если множества A и B конечны, то их объединение содержит столько элементов, сколько имеется в каждом из них, за вычетом тех, что принадлежат им обоим одновременно (т.е. посчитанных дважды). Докажем равенство для произвольных множеств. По свойствам операций над множествами справедливы следующие тождества:

$$A \cup B = A \cup (\overline{A \cap B} \cap B),$$
$$B = (A \cap B) \cup (\overline{A \cap B} \cap B).$$

Убедимся, например, в первом:

$$A \cup (\overline{A \cap B} \cap B) = (A \cup \overline{A \cap B}) \cap (A \cup B) =$$
$$= (A \cup \overline{A} \cup \overline{B}) \cap (A \cup B) =$$
$$= (\Omega \cup \overline{B}) \cap (A \cup B) = \Omega \cap (A \cup B) = A \cup B.$$

В обоих равенствах объединяются непересекающиеся множества. Значит, можно применять аксиому конечной аддитивности. Введем обозначение: $\overline{A \cap B} \cap B = C$. Из второго равенства получаем

$$P(B) = P(A \cap B) + P(C),$$

откуда $P(C) = P(B) - P(A \cap B)$. Из первого равенства следует

$$P(A \cup B) = P(A) + P(C).$$

Если подставить найденное выражение для $P(C)$, сразу получим, что

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Это и есть доказываемое равенство, если переписать его на языке множеств.

Пусть, например, событие A состоит в том, что выпало четное число, а событие B – в том, что выпало число, кратное трем. Тогда событие $A + B$ = «выпало четное или кратное трем число» происходит, если выпадает двойка, тройка, четверка или шестерка:

$$A + B = \{\omega_2, \omega_3, \omega_4, \omega_6\}$$

и

$$P(A + B) = 4/6 = {}^2/_3.$$

С другой стороны, событие AB, состоящее в том, что выпало четное число, кратное трем, происходит только при выпадении шестерки: $AB = \{\omega_6\}$, т.е. $P(AB) = {}^1/_6$. Кроме того, $P(A) = {}^3/_6$, $P(B) = {}^2/_6$ и

$$P(A) + P(B) - P(AB) = \frac{3+2-1}{6} = 4/6 = {}^2/_3.$$

11.4. Условная вероятность

11.4.1. Определение условной вероятности

При вычислении вероятности случайного события (например, по классической формуле) мы вынуждены учитывать все возможные исходы эксперимента, поскольку все они равновозможны и нам неизвестно, какой именно произойдет. Другое дело, если из независимых источников (от Господа Бога, от доброжелателя и пр.) мы получили некоторую информацию об исходе эксперимента. Эта информация, возможно, не носит исчерпывающего характера, но тем не менее сужает круг возможных исходов. При этом шансы того или иного события могут возрасти или упасть. Картежники говорят: «Знать бы, что в прикупе лежит...», т.е. дополнительные знания повышают шанс выиграть.

Рассмотрим пример с кубиком. Если мы ничего не знаем об исходе опыта и находимся в условиях симметричной классической схемы, вероятность того, что выпадет двойка, равна $^1/_6$. Если же нам сообщили, что выпало четное число, то шансы выпадения двойки возрастают. Теперь общее число возможных исходов уменьшается до трех (двойка, четверка и шестерка) и вероятность того, что на самом деле выпала двойка, равна $^1/_3$. Если же нас интересует выпадение тройки, то, разумеется, информация о четности снижает вероятность данного события до нуля.

Пусть дополнительная информация формулируется так: «Известно, что в результате эксперимента произошло событие B». Зная это, мы хотим подсчитать вероятность некоторого события A. Такую вероятность (п р и у с л о в и и, что произошло событие B) называют *условной вероятностью* события A и обозначают $P(A|B)$. В предыдущем примере B = «выпало четное число». Если A = «выпала двойка» и C = «выпала тройка», то $P(A|B) = {}^1/_3$, $P(C|B) = 0$.

Приведем еще пример. Пусть студент из 28 билетов выучил 4 билета с четными номерами и 12 с нечетными. По классической схеме вероятность вытянуть «счастливый», выученный, билет (событие A) равна

$$P(A)=\frac{16}{28}=\frac{4}{7}$$

(всего 28 исходов, благоприятных – 16). Допустим, когда он пришел на экзамен, ему сообщили, что на столе экзаменатора осталось всего 14 билетов, притом все они с нечетными номерами (событие B). Теперь общее число исходов равно 14, а благоприятных – 12, так как он выучил 12 «нечетных» билетов. Значит,

$$P(A\mid B)=\frac{12}{14}=\frac{6}{7}.$$

Рассмотрим общую ситуацию в классическом случае. Пусть случайный эксперимент имеет N равновозможных исходов. Допустим, известно, что произошло событие B. Какова условная вероятность события A? Ясно, что общее число исходов уменьшается до тех,

которые благоприятствуют событию B (раз оно точно произошло, никаких других исходов произойти не могло). Это число обозначим, как обычно, через $m(B)$. Какие же исходы в данном случае будут благоприятными для события A? Конечно, те, которые благоприятствуют одновременно событиям A и B, т.е. благоприятные для их произведения. Их количество равно $m(AB)$. Итак, условная вероятность

$$P(A\mid B)=\frac{m(AB)}{m(B)}.$$

Разделив числитель и знаменатель на общее число исходов N, получим изящную формулу

$$P(A\mid B)=\frac{m(AB)/N}{m(B)/N}=\frac{P(AB)}{P(B)}.$$

Очевидно, что в примере с экзаменационными билетами эта формула выполняется: 14 исходов, благоприятных для события B, и 12 исходов, благоприятных как для A, так и для B.

Равенство

$$\boxed{P(A\mid B)=\frac{P(AB)}{P(B)}}$$

принимается за о п р е д е л е н и е ***условной вероятности*** события A при условии, что произошло событие B (разумеется, предполагается, что $P(B) > 0$). Из определения легко вывести некоторые простые свойства условной вероятности.

∇1. $P(A|A) = 1$.

∇2. *Если* $B \subset A$, *то* $P(A|B) = 1$.

∇3. *Для любого события B с ненулевой вероятностью*

$$P(\Omega\mid B) = 1,\quad P(\varnothing\mid B) = 0.$$

∇4. *Если события A_1 и A_2 несовместны, то*

$$P(A_1 + A_2\mid B) = P(A_1\mid B) + P(A_2\mid B).$$

Используя дистрибутивный закон, получаем $(A_1 + A_2)B = A_1B + A_2B$. Однако по условию A_1 и A_2 несовместны, т.е. $A_1A_2 = \varnothing$. Значит, $(A_1B)(A_2B) = \varnothing$ и события A_1B и A_2B также несовместны. Таким образом, вероятность их суммы равна сумме вероятностей: $P((A_1 + A_2)B) = P(A_1B + A_2B) = P(A_1B) + P(A_2B)$, откуда по определению

$$P(A_1+A_2|B)=\frac{P((A_1+A_2)B)}{P(B)}=$$

$$=\frac{P(A_1B)}{P(B)}+\frac{P(A_2B)}{P(B)}=P(A_1|B)+P(A_2|B).$$

∇5 (теорема умножения).

$$P(AB)=P(A\,|\,B)P(B).$$

Данное утверждение должно вызвать недоумение: ведь мы определили условную вероятность через вероятность произведения событий. Как же пользоваться этой формулой для вычисления $P(AB)$? На самом деле, формула бессмысленна в рамках заданного вероятностного пространства $\{\Omega, \mathcal{F}, P\}$. Однако при решении реальных задач часто можно найти условные вероятности исходя из некоторых дополнительных практических соображений, не прибегая к определению.

Не вдаваясь в математические тонкости, скажем лишь, что, по существу, это означает, что вводится новое вероятностное пространство, в котором условные вероятности, вычисленные по определению, совпадают с вычисленными из практических соображений.

Задача. *Пусть в урне находятся 4 белых и 6 черных шаров. Из урны наугад вынимают 2 шара, не возвращая их обратно. Какова вероятность того, что первый из выбранных шаров белый, а второй — черный?*

Р е ш е н и е (по классической схеме). Всего эксперимент имеет $N=A_{10}^2=9\cdot 10=90$ исходов: столько упорядоченных пар можно достать из урны с 10 шарами. Благоприятными являются пары с белым шаром на первом месте (4 возможности) и черным — на втором (6 возможностей). Значит, благоприятных исходов $4\cdot 6=24$ и искомая вероятность равна

$$P(A)=\frac{24}{90}=\frac{4}{15}.$$

Р е ш е н и е (по теореме умножения). Обозначим через $Б_1$ и $Ч_2$ события, состоящие в том, что первым вынули белый шар, а вторым — черный. Нас интересует вероятность события $Б_1Ч_2$. По теореме умножения

$$P(Б_1Ч_2)=P(Б_1)P(Ч_2\,|\,Б_1).$$

Вероятность события $Б_1$ равна $4/10={}^2/_5$ (всего шаров 10, белых 4). Если первым вынули белый шар (произошло событие $Б_1$), то в урне осталось 3 белых и 6 черных шаров. Теперь вероятность достать

черный шар равна $6/9={}^2/_3$, а это и есть условная вероятность события $Ч_2$ при условии, что произошло событие $Б_1$. Итак,

$$P(A)=P(Б_1Ч_2)=\frac{2}{5}\frac{2}{3}=\frac{4}{15}.$$

Естественно, получен тот же результат, что и в классической схеме. Условную вероятность мы посчитали, не прибегнув к определению, из разумных соображений.

Теорему умножения можно обобщить на случай произведения n событий.

∇5′. $P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$.

Равенство выводится из теоремы умножения по индукции, что предлагается проделать читателю самостоятельно.

Задача. *В кармане лежит n ключей, из которых только один подходит к замку. Ключи по одному достают из кармана. Какова вероятность того, что нужный ключ будет вынут при k-м извлечении?*

Р е ш е н и е. Обозначим через A_k событие, состоящее в том, что нужный ключ вынули на k-м шаге, соответственно $\overline{A}_k$ состоит в том, что на этом шаге достали не тот ключ. Событие, вероятность которого требуется найти, есть произведение $\overline{A}_1\overline{A}_2\cdots\overline{A}_{k-1}A_k$ (первые $k-1$ раз доставали не тот ключ, а на k-й раз вынули нужный). Очевидно, что

$$P(\overline{A}_1)=\frac{n-1}{n}$$

($n-1$ ненужных из n ключей);

$$P(\overline{A}_2|\overline{A}_1)=\frac{n-2}{n-1}$$

(если в первый раз достали неправильный ключ, в кармане осталось $n-1$ ключей, из них $n-2$ не подходят);

$$P(\overline{A}_3|\overline{A}_1\overline{A}_2)=\frac{n-3}{n-2}$$

и так далее:

$$P(\overline{A}_{k-1}|\overline{A}_1\overline{A}_2\cdots\overline{A}_{k-2})=\frac{n-k+1}{n-k+2}.$$

Если $k-1$ раз доставали неправильный ключ, то в кармане осталось $n-k+1$ ключей, из них один подходящий. Значит,

$$P(A_k|\overline{A}_1\overline{A}_2\cdots\overline{A}_{k-1})=\frac{1}{n-k+1}.$$

Итак,

$$P(\overline{A}_1\overline{A}_2\cdots\overline{A}_{k-1}A_k)=\frac{n-1}{n}\frac{n-2}{n-1}\frac{n-3}{n-2}\cdots\frac{n-k+1}{n-k+2}\frac{1}{n-k+1}=\frac{1}{n}$$

(после сокращений остается только знаменатель первой дроби и числитель последней).

11.4.2. Независимость событий

Часто на практике появление одного события никак не влияет на шансы появления другого. Например, при многократном подбрасывании монеты результаты каждого эксперимента не зависят от результатов предыдущих опытов (если, конечно, соблюдаются одинаковые условия). Такие события естественно назвать независимыми. Математически независимость означает, что условная вероятность некоторого события совпадает с его вероятностью (безусловной вероятностью). Итак, говорят, что событие A *не зависит* от события B, если

$$P(A|B)=P(A).$$

∇1. *Если событие A не зависит от события B, то и событие B не зависит от события A.*

В самом деле, справедлива следующая цепочка равенств:

$$P(B|A)=\frac{P(BA)}{P(A)}=\frac{P(AB)}{P(A)}=\frac{P(A|B)P(B)}{P(A)}=\frac{P(A)P(B)}{P(A)}=P(B).$$

Первое равенство – определение условной вероятности, затем используется коммутативность умножения ($BA=AB$). В следующем равенстве вероятность произведения записывается по теореме умножения, и, наконец, используется тот факт, что событие A не зависит от события B. Итак, если одно событие не зависит от другого, то последнее также не зависит от первого. Это дает полное основание называть такие события ***независимыми***.

∇2. *Если события A и B независимы, то*

$$P(AB)=P(A)P(B).$$

Равенство следует из теоремы умножения и определения независимости. Часто это равенство используется в качестве

определения независимых событий. Если математическая модель (вероятностное пространство) случайного эксперимента адекватно отражает его реальные свойства, то событиям, независимым в действительности, соответствуют независимые (по определению) события.

Пусть, например, бросают две монеты одновременно. Рассмотрим два события: A = «на первой монете выпал «орел» и B = «на второй монете выпал «орел». Ясно, что данные события независимы в житейском, практическом смысле. Убедимся в том, что они независимы и по определению. Пространство Ω состоит в данном случае из четырех исходов: $\Omega=\{ОО, РР, ОР, РО\}$, где, например, РО означает, что на первой монете выпала «решка», а на второй – «орел». Тогда ясно, что каждому из событий A и B благоприятствуют два исхода:

$$A=\{ОО, ОР\},\quad B=\{РО, ОО\},$$

т.е. $P(A)=P(B)={}^1/_2$. Событие AB («на обеих монетах выпал «орел») состоит из единственного исхода ОО, и $P(AB)={}^1/_4$. Однако $P(A)P(B)=({}^1/_2)({}^1/_2)={}^1/_4=P(AB)$ и события независимы по определению. Значит, выбрана хорошая математическая модель эксперимента (четыре несовместных равновозможных исхода, классическая схема вычисления вероятности).

Следует различать попарную независимость событий и независимость в совокупности.

События A_1, A_2, ..., A_n называются *независимыми в совокупности*, если они попарно независимы и каждое из них не зависит от произведения любого набора из остальных событий.

Следующий пример иллюстрирует различие понятий попарной независимости и независимости в совокупности. Пусть в урне находятся четыре карточки с числами 2, 3, 5 и 30. Из урны наугад вынимают одну карточку (число на ней обозначим через x). Рассмотрим следующие события: A_1 = «x четно», A_2 = «x кратно трем» и A_3 = «x кратно пяти». Эксперимент имеет четыре исхода, причем каждому перечисленному событию благоприятствуют два из них: карточка с соответствующим простым числом или карточка с числом 30. Значит,

$$P(A_1)=P(A_2)=P(A_3)={}^1/_2.$$

События A_1A_2, A_1A_3, A_2A_3 происходят только в том случае, когда вынимают карточку «30». Значит,

$$P(A_1A_2)=P(A_1A_3)=P(A_2A_3)=\frac{1}{4}=\frac{1}{2}\frac{1}{2}$$

и события п о п а р н о независимы. При этом

$$P(A_1|A_2A_3) = P\,(x \text{ четно}|x \text{ кратно трем и пяти}) = 1$$

(если x кратно трем и пяти одновременно, то достали карточку «30»: в этом случае x обязательно четно). Таким образом,

$$P(A_1|A_2A_3) \neq P(A_1)$$

и события не являются независимыми в с о в о к у п н о с т и.

∇**2′.** *Если события $A_1, A_2, ..., A_n$ независимы в совокупности, то*

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n).$$

Задача. *Стрелок стреляет по мишени 3 раза. Он попадает в мишень с вероятностью $p = 0,6$. Какова вероятность, что он не попадет по мишени хотя бы один раз?*

Р е ш е н и е. Пусть событие A_k ($k = 1, 2, 3$) означает, что стрелок попал в мишень при k-й попытке. Соответственно, $\overline{A_k}$ — событие, состоящее в том, что стрелок не попал в мишень. По условию $P(A_k) = 0,6$. Значит, $P(\overline{A_k}) = 1 - 0,6 = 0,4$. Рассматриваемое событие противоположно событию «стрелок не попал ни разу», которое представляет собой произведение $\overline{A_1}\,\overline{A_2}\,\overline{A_3}$. Поэтому

$$P(A) = 1 - P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}).$$

Можно проверить, что события $\overline{A_1}$, $\overline{A_2}$, $\overline{A_3}$ независимы в совокупности. Поэтому

$$P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}) = P(\overline{A_1})P(\overline{A_2})P(\overline{A_3}) = 0,4 \cdot 0,4 \cdot 0,4 = 0,064.$$

Окончательно получаем

$$P(A) = 1 - 0,064 = 0,936.$$

11.4.3. Формула Байеса

Начнем с задачи.

Задача. *В двух группах занимаются соответственно 20 и 30 студентов, причем в первой группе 5 юношей, а во второй — 3. Какова вероятность того, что выбранный наугад студент — юноша?*

Обозначим через A событие, состоящее в том, что выбрали юношу. Разумеется, эту простую задачу можно решить по классической схеме. Всего в двух группах 50 студентов, из них $5 + 3 = 8$ юношей. Поэтому $P(A) = 8/50 = {}^4/_{25}$. Однако данная задача принадлежит к

большой группе задач, не все из которых решаются столь просто по классической формуле.

Пусть события B_1 и B_2 означают соответственно, что выбранный студент учится в первой или второй группе. Эти события являются, по сути, п р е д п о с ы л к а м и события A: если ни одно из них не происходит, не происходит и событие A. Кроме того, эти события не могут произойти одновременно. Математически это означает, что события-предпосылки несовместны и событие A содержится в сумме этих событий. Задача, таким образом, состоит в том, чтобы найти вероятность некоторого события, обладающего набором событий-предпосылок. Выбор в этом случае можно описать двухступенчатой схемой: сначала наугад выбирается предпосылка (в данном случае группа), а затем проводится испытание (выбирается студент). В следующей теореме содержится так называемая *формула полной вероятности*, позволяющая решать подобные задачи.

Теорема 11.1. *Пусть события $B_1, B_2, ..., B_n$ попарно несовместны и событие A содержится в их сумме:*

$$A \subset B_1 + B_2 + ... + B_n.$$

Тогда вероятность события A можно вычислить по следующей формуле:

$$\boxed{P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + ... + P(B_n)P(A|B_n) = \sum_{k=1}^{n} P(B_k)P(A|B_k)}$$

Д о к а з а т е л ь с т в о. Поскольку A содержится в сумме событий B_k, имеет место равенство $A(B_1 + B_2 + ... + B_n) = A$ или

$$A = AB_1 + AB_2 + ... + AB_n.$$

Кроме того, события B_k попарно несовместны, т.е. $B_iB_j = \varnothing$, если $i \neq j$. Значит, и события $AB_1, AB_2, ..., AB_n$ также попарно несовместны: $(AB_i)(AB_j) = A(B_iB_j)A = A\varnothing A = \varnothing$. Как мы знаем, вероятность суммы попарно несовместных событий равна сумме их вероятностей, т.е.

$$P(A) = P(AB_1) + P(AB_2) + ... + P(AB_n) = \sum_{k=1}^{n} P(AB_k).$$

По теореме умножения

$$P(AB_k) = P(B_kA) = P(B_k)P(A|B_k).$$

Из двух последних равенств и выводится формула полной вероятности.

Решим теперь ту же задачу с применением выведенной формулы: $P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2)$. Очевидно, что вероятность события B_1 (студент из первой группы) равна $P(B_1) = 20/50 = {}^2/_5$. Аналогично $P(B_2) = 30/50 = {}^3/_5$. Вероятность того, что выбранный студент – юноша, при условии, что он из первой группы, равна $P(A|B_1) = 5/20 = {}^1/_4$ (так как в группе 20 студентов, из которых 5 – юноши). Точно так же $P(A|B_2) = 3/30 = {}^1/_{10}$. Итак,

$$P(A) = \frac{2}{5}\frac{1}{4} + \frac{3}{5}\frac{1}{10} = \frac{1}{10} + \frac{3}{50} = \frac{8}{50} = \frac{4}{25}.$$

Ответ, как и следовало ожидать, совпадает с полученным по классической формуле.

Чаще всего набор несовместных предпосылок выбирают таким образом, чтобы их сумма совпадала со всем пространством элементарных исходов:

$$B_1 + B_2 + \ldots + B_n = \Omega.$$

В этом случае условия теоремы 11.1 автоматически выполняются. События-предпосылки принято называть *гипотезами*.

Для события с предпосылками представляет интерес еще один вопрос: пусть известно, что произошло событие A; какова вероятность, что произошла данная предпосылка? К примеру, в рассмотренной задаче требуется найти вероятность того, что выбранный студент учится в первой или второй группе, если известно, что он юноша. Другими словами, требуется найти условную вероятность $P(B_k|A)$. Эту вероятность называют *апостериорной вероятностью* события-предпосылки, в отличие от *априорной вероятности* $P(B_k)$. Апостериорную вероятность находят при условии, что событие-следствие уже произошло. Она представляет собой уточненное, пересчитанное значение вероятности с учетом дополнительной информации. Формула для апостериорной вероятности (так называемая *формула Байеса*) выводится из формулы полной вероятности.

Теорема 11.2. *Пусть события $B_1, B_2, \ldots, B_n$ попарно несовместны и событие A содержится в их сумме:*

$$A \subset B_1 + B_2 + \ldots + B_n.$$

Тогда при $k = 1, 2, \ldots, n$ справедлива формула

$$\boxed{P(B_k | A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}}$$

Доказательство. По определению условной вероятности и теореме умножения справедливы равенства

$$P(B_k | A) = \frac{P(B_k A)}{P(A)} = \frac{P(B_k)P(A|B_k)}{P(A)}.$$

Чтобы получить формулу Байеса, осталось только заменить $P(A)$ согласно формуле полной вероятности. Обратите внимание, что индекс в сумме изменился с k на i: теперь буква k обозначает номер конкретного события-предпосылки и не может использоваться для переменного индекса.

Томас Байес (1702–1761). Родился в Лондоне. Получил домашнее образование. Существует предположение, что его учителем был известный английский математик, один из создателей теории вероятностей Абрахам де Муавр, который в это время занимался в Лондоне частным преподаванием. Отец Томаса был членом Королевского общества, имел духовный сан и служил пастором в одной из лондонских церквей. Т. Байес также получил духовный сан и начал помогать отцу в выполнении его пасторских обязанностей.

В 1720 г. в Танбридже, в 50 км от Лондона, была организована религиозная секта. Вскоре туда переехал и Т. Байес. В 1731 г. он написал трактат «Божественная доброта...» на религиозные темы. При этом Байес интересовался математикой, хотя его математические работы и не были опубликованы при жизни. С 1742 г. – член Королевского общества. В 1752 г. он отошел от дел. Байес прожил в Танбридже до своей кончины, наступившей 17 апреля 1761 г.

В 1763 г. в «Философских трудах Королевского общества» были опубликованы работы Байеса по теории вероятностей под названием «Опыт решения задачи по теории вероятностей покойного достопочтенного мистера Байеса, члена Королевского общества». В этой работе изложены основные принципы и понятия теории вероятностей, которые почти в неизменном виде сохранились до наших дней. Это, например, определения несовместных и независимых событий, понятие вероятности как числовой характеристики шанса и т.д. Байес доказал ряд важных теорем, в частности теорему о сложении вероятностей несовместных событий.

«Формулы Байеса» в работе Байеса нет. Это название появилось только в конце XIX в. Формула носит имя Байеса, так как полученные им результаты имеют отношение к оценке вероятности после того, как стали известны результаты испытания (a posteriori). Важная работа Байеса была фактически забыта, и его имя связывалось только с известной формулой, которую великий Лаплас причислил к основным принципам теории вероятностей.

Приведем еще один пример применения формулы полной вероятности и формулы Байеса.

Задача. *В первой урне 1 белый и 9 черных шаров, во второй – 1 черный и 5 белых. Из каждой урны выбрали наугад по одному шару, а оставшиеся ссыпали в третью урну, из которой наугад вынули 1 шар. Найти вероятность того, что:*

а) шар, вынутый из третьей урны, будет белым;

б) из обеих урн вынули белые шары, если шар, вынутый из третьей урны, оказался белым.

Р е ш е н и е. Обозначим через A событие, состоящее в том, что из третьей урны вынули белый шар. Введем события-предпосылки:

B_1 = «из первой урны вынули белый шар, из второй – черный»;

B_2 = «из второй урны вынули белый шар, из первой – черный»;

B_3 = «из обеих урн вынули белые шары»;

B_4 = «из обеих урн вынули черные шары».

События A и B_1, B_2, B_3, B_4 удовлетворяют условиям теорем 11.1 и 11.2. Во-первых, события B_k несовместны, во-вторых, очевидно, что $A \subset B_1 + B_2 + B_3 + B_4$ (если произошло A, то непременно произошло одно из B_k). Значит, можно применить формулу полной вероятности:

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) + P(B_4)P(A|B_4),$$

т.е. для вычисления искомой вероятности требуется найти восемь чисел – вероятности событий B_k и условные вероятности.

Событие B_1 есть произведение $Б_1Ч_2$ двух независимых событий (из первой урны вынули белый шар, из второй – черный). Поэтому

$$P(B_1) = P(Б_1)P(Ч_2).$$

Из условия задачи непосредственно вытекает, что $P(Б_1) = {}^1/_{10}$ (один белый из десяти) и $P(Ч_2) = {}^1/_6$ (один черный из шести). Значит,

$$P(B_1) = \frac{1}{10}\frac{1}{6}.$$

В случае если из первой урны достали белый шар, а из второй – черный, в третьей урне оказалось 9 черных и 5 белых шаров. Вероятность достать белый шар при этом равна

$$P(A|B_1) = \frac{5}{14}.$$

Точно так же определяем, что

$$P(B_2) = \frac{9}{10}\frac{5}{6}, \quad P(A|B_2) = \frac{5}{14},$$

$$P(B_3) = \frac{1}{10}\frac{5}{6}, \quad P(A|B_3) = \frac{4}{14},$$

$$P(B_4) = \frac{9}{10}\frac{1}{6}, \quad P(A|B_4) = \frac{6}{14}.$$

Окончательно получаем

$$P(A) = \frac{1}{10}\frac{1}{6}\frac{5}{14} + \frac{9}{10}\frac{5}{6}\frac{5}{14} + \frac{1}{10}\frac{5}{6}\frac{4}{14} + \frac{9}{10}\frac{1}{6}\frac{6}{14} =$$

$$= \frac{1\cdot1\cdot5 + 9\cdot5\cdot5 + 1\cdot5\cdot4 + 9\cdot1\cdot6}{10\cdot6\cdot14} = \frac{304}{840} = \frac{38}{105}.$$

Вторая задача решается с помощью формулы Байеса. Требуется найти апостериорную вероятность $P(B_3|A)$:

$$P(B_3|A) =$$

$$= \frac{P(B_3)P(A|B_3)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) + P(B_4)P(A|B_4)} =$$

$$= \frac{20/840}{304/840} = \frac{20}{304} = \frac{5}{76}.$$

Особое значение формула Байеса приобретает для таких экспериментов, в которых предпосылки B_1, B_2, ... B_n непосредственно наблюдать невозможно, хотя их априорные вероятности $P(B_k)$ и условные вероятности $P(A|B_k)$, $k = 1, 2, \ldots n$, можно определить из дополнительных опытов. Такая ситуация имеет место, например, если отсутствует прибор, регистрирующий данные события, или же применение такого прибора приводит к разрушению предмета наблюдения (разрушающий контроль). В этих случаях переоценка вероятностей предпосылок после опыта может быть проведена на основании наблюдаемого события A, тесно связанного с событиями-предпосылками. Такой подход часто используется при медицинской и технической диагностике, при определении причин аварий и во многих других областях.

11.5. Схема испытаний Бернулли

Пусть в результате некоторого случайного испытания может произойти или не произойти определенное событие A. Если событие произошло, будем называть испытание *успешным*, а само событие – *успехом*. Испытание повторяется n раз. При этом соблюдаются следующие условия:

- вероятность успеха $P(A) = p$ в каждом испытании одна и та же;
- результат любого испытания не зависит от исходов предыдущих испытаний.

Такая последовательность испытаний с двумя исходами (успех/неудача) называется *последовательностью независимых испы-*

таний Бернулли или короче – *схемой Бернулли.* Схема Бернулли представляется весьма важной с точки зрения философии теории вероятностей, так как в ней отражается концепция случайности, принятая в математике. По сути дела, любое случайное событие можно рассматривать как успех некоторого эксперимента в последовательности независимых испытаний, условия которых сохраняются неизменными.

Примеры.

1. Монету подбрасывают n раз, успехом считается выпадение «орла». Вероятность успеха при каждом испытании $p = P(\text{О}) = {}^1/_2$.

2. В урне 4 белых и 6 черных шаров. Наугад вынимают шар и возвращают обратно. Испытание успешно, если вынули белый шар. В этом случае $p = 0{,}4$ (4 белых шара из 10).

Задача. *Найти вероятность $P_n(k)$ того, что в схеме Бернулли из n независимых испытаний произошло ровно k успехов* ($k = 0, 1, \ldots, n$).

Прежде чем решить эту задачу в общем виде, рассмотрим подробнее частный случай схемы Бернулли. Пусть $n = 3$, т.е. проводится три испытания. Обозначим через A_k успех в k-м испытании, $k = 1, 2, 3$. В результате каждого испытания может случиться успех (событие A_k) или неудача (событие $\overline{A}_k$). Тогда исход данного эксперимента можно представить в виде произведения трех событий, каждое из которых есть либо A_k, либо $\overline{A}_k$. Легко перечислить все возможные исходы:

$\overline{A}_1\overline{A}_2\overline{A}_3$ – ни одного успеха ($k = 0$);

$A_1\overline{A}_2\overline{A}_3$, $\overline{A}_1A_2\overline{A}_3$, $\overline{A}_1\overline{A}_2A_3$ – один успех ($k = 1$);

$A_1A_2\overline{A}_3$, $A_1\overline{A}_2A_3$, $\overline{A}_1A_2A_3$ – два успеха ($k = 2$);

$A_1A_2A_3$ – все испытания успешны ($k = 3$).

Очевидно, что перечисленные восемь исходов несовместны. Теперь нетрудно подсчитать вероятность события, состоящего, например, в том, что ровно два испытания оказались успешными, т.е. величину $P_3(2)$. Данное событие содержит три несовместных исхода. Следовательно,

$$P_3(2) = P(A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3) =$$
$$= P(A_1A_2\overline{A}_3) + P(A_1\overline{A}_2A_3) + P(\overline{A}_1A_2A_3).$$

Каждый из указанных исходов представляет собой произведение трех независимых событий – двух успехов и одной неудачи. Значит, вероятность исхода равна произведению соответствующих вероятностей. Вероятность успеха равна p, вероятность неудачи равна $q = 1 - p$

(это принятое обозначение). Таким образом, вероятность каждого такого исхода равна $p \cdot p \cdot q = p^2q$:

$$P(A_1A_2\overline{A}_3) = P(A_1\overline{A}_2A_3) = P(\overline{A}_1A_2A_3) = p^2q,$$

а искомая вероятность равна сумме трех одинаковых слагаемых:

$$P_3(2) = 3p^2q.$$

Например, при бросании монеты $p = q = 0{,}5$ и

$$P_3(2) = 3 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{3}{8}.$$

При извлечении шара из урны с 4 белыми и 6 черными шарами (успех – белый шар) $p = 0{,}4$, $q = 0{,}6$ и

$$P_3(2) = 3 \cdot 0{,}4^2 \cdot 0{,}6 = 0{,}288.$$

В рассмотренном примере хорошо просматривается общий подход к решению поставленной задачи. Пусть проделано n независимых испытаний. Событие, состоящее в том, что было ровно k успешных испытаний, есть сумма некоторого числа несовместных исходов вида

$$A_1\overline{A}_2 \cdots \overline{A}_iA_{i+1} \cdots A_n$$

(состоящих из k успехов и $n - k$ неудач). Значит, искомая вероятность равна сумме вероятностей этих исходов. Каждый такой исход равен произведению n независимых событий, из которых ровно k – успех (его вероятность равна p), а $n - k$ – неудача (с вероятностью $q = 1 - p$). Таким образом, вероятность каждого такого исхода по свойству независимых событий равна

$$\underbrace{p \cdots p}_{k}\,\underbrace{q \cdots q}_{n-k} = p^kq^{n-k}.$$

Осталось подсчитать количество слагаемых. Если обозначить успех единицей, а неудачу – нулем, вопрос сводится к следующему: сколько существует наборов из n единиц и нулей, содержащих ровно k единиц? При $n = 3$, $k = 2$ это наборы 110, 101, 011. Очевидно, что таких наборов ровно столько, сколько существует способов выбрать k позиций из n. Ясно, что это есть число сочетаний из n по k. Таким образом, искомая вероятность равна сумме C_n^k одинаковых слагаемых p^kq^{n-k}. Получена ***формула Бернулли*** вероятности k успехов в n независимых испытаниях:

$$\boxed{P_n(k) = C_n^kp^kq^{n-k}}$$

Якоб Бернулли (1654–1705). Один из родоначальников великой династии швейцарских ученых. Родился в Базеле. В Швейцарии Базель, свободный имперский город с 1263 г., был средоточием науки. Еще во времена Эразма Роттердамского его университет был важным научным центром. Науки и искусства процветали в Базеле под управлением купеческого патрициата. К этому базельскому патрициату принадлежала семья Бернулли. С конца XVII в. эта семья в каждом поколении давала выдающихся ученых.

Якоб Бернулли изучал теологию, а его брат Иоганн – медицину, но, прочитав работы Лейбница, оба они решили стать математиками. В 1687 г. Якоб возглавил кафедру математики в Базельском университете, где преподавал до своей смерти. Одним из его учеников был отец великого Леонарда Эйлера. Якоб начал переписываться с Лейбницем в 1687 г. Затем, постоянно обмениваясь мыслями с Лейбницем, не раз вступая в ожесточенное соперничество друг с другом, оба брата начали делать открытия, вошедшие в сокровищницу математической мысли. Список их результатов длинен и содержит многое из того, что входит сейчас в учебники по дифференциальному и интегральному исчислению, теории вероятностей и другим разделам математики.

Якоб Бернулли был одним из первых исследователей в теории вероятностей, чему посвящена его неоконченная книга «Искусство предположений», опубликованная посмертно в 1713 г. Книга вышла с предисловием Николая Бернулли, его племянника, сына Иоганна. Николай Бернулли пишет: «Автор ставил своей целью показать исключительную пользу, которую может оказать в вопросах гражданской жизни до сего времени мало разработанная часть математики, имеющая своим предметом измерение вероятностей».

В своей работе Якоб Бернулли высказал общие соображения о природе случайных событий и вероятности. Он развил теорию перестановок и сочетаний и показал, как с помощью комбинаторики можно решать разнообразные вероятностные задачи об азартных играх. Многие результаты из книги Бернулли стали классическими, основополагающими в теории вероятностей. Достаточно назвать формулу Бернулли и теорему Бернулли, в которой устанавливается тесная связь между наблюдаемой частотой случайного события и его вероятностью. Благодаря работам Якоба Бернулли начался новый этап в развитии теории вероятностей, обусловленный новым подходом к самому понятию вероятности.

Из формулы Бернулли и равенств $C_n^n = C_n^0 = 1$ легко вывести простые следствия.

∇**1.** $$P_n(n) = p^n, \quad P_n(0) = q^n.$$

Проиллюстрируем применение формулы Бернулли на следующем примере.

Задача. *В урне 3 белых и 2 черных шара. Из урны 4 раза наугад вынимают по 2 шара и возвращают обратно. Найти вероятность того, что:*

а) ровно 3 раза вынули шары разного цвета;

б) хотя бы один раз вынули шары разного цвета.

Решение. Очевидно, что мы имеем дело со схемой Бернулли и в первом задании требуется найти вероятность $P_4(3)$. Под успехом понимается событие A, состоящее в том, что из урны вынули шары разного цвета. Единственное, что требуется для успешного применения формулы Бернулли, – найти вероятность p успеха в единичном испытании. При каждом испытании из урны вынимают 2 шара. В урне 5 шаров, следовательно, такое испытание имеет C_5^2 исходов:

$$N = C_5^2 = \frac{5!}{2!3!} = \frac{4 \cdot 5}{2} = 10.$$

Благоприятными являются исходы с шарами разного цвета: при этом каждый из 3 белых шаров может сочетаться с любым из 2 черных. Значит, число благоприятных исходов равно

$$m = 3 \cdot 2 = 6.$$

Итак,

$$p = P(A) = \frac{6}{10} = 0{,}6.$$

Теперь мы готовы применить формулу Бернулли:

$$P_4(3) = C_4^3 p^3 q^1 = 4 \cdot (0{,}6)^3 \cdot 0{,}4 = 0{,}3456.$$

Перейдем ко второму заданию. Выражение «хотя бы один раз» означает, что успех случился один, или два, или три, или четыре раза. Стало быть, искомая вероятность равна сумме $P_4(1) + P_4(2) + P_4(3) + P_4(4)$. Вычислять эту сумму по формуле было бы весьма утомительно. Возможен и другой подход: событие «хотя бы один раз произошел успех» противоположно событию «ни разу не было успеха». Поэтому искомая вероятность равна

$$1 - P_4(0) = 1 - q^4 = 1 - (0{,}4)^4 = 0{,}9744.$$

По существу, мы воспользовались следующим свойством бернуллиевых вероятностей.

∇**2.** $$P_n(0) + P_n(1) + \ldots + P_n(n-1) + P_n(n) = 1.$$

В самом деле, по формуле Бернулли

$$P_n(0) + P_n(1) + \ldots + P_n(n-1) + P_n(n) =$$
$$= C_n^0 p^0 q^n + C_n^1 p^1 q^{n-1} + \ldots + C_n^{n-1} p^{n-1} q^1 + C_n^n p^n q^0.$$

Вглядевшись повнимательнее, мы немедленно узнаем в выражении справа формулу бинома Ньютона для $(q + p)^n$. Итак,

$$P_n(0) + P_n(1) + \ldots + P_n(n-1) + P_n(n) = (q + p)^n.$$

11 Зак. № 2435 Турецкий

Однако $q + p = 1$, откуда и выводится доказываемая формула. Например, при $n = 3$ формула принимает вид

$$q^3 + 3q^2p + 3qp^2 + p^3 = (q + p)^3 = 1.$$

Задачи

Операции над событиями

1. Кубик бросают один раз. Выписать, из каких элементарных исходов состоят следующие события:

A = «выпало не более 2 очков»;

B = «выпало 3 очка»;

C = «выпало не менее 4 очков».

Образуют ли эти события полный набор? Являются ли они равновозможными?

2. Кубик подбрасывают 2 раза. Описать пространство элементарных исходов Ω и подмножества, соответствующие следующим событиям:

A = «оба раза выпало одинаковое число очков»;

B = «ни разу не выпало число 6»;

C = «оба раза выпало число, большее 3»;

3. Три стрелка сделали по одному выстрелу в мишень. Событие A_k состоит в том, что k-й стрелок попал в цель ($k = 1, 2, 3$). С помощью операций над событиями выразить следующие события через A_1, A_2, A_3:

а) первый и второй стрелки поразили мишень;

б) первый стрелок не поразил мишень;

в) в мишени по крайней мере одна пуля;

г) в мишени три пули;

д) в мишени ни одной пули.

Классическое определение вероятности

4. Из колоды карт (36 карт) наугад выбирается одна. Какова вероятность, что выбранная карта – туз?

5. В урне 3 белых и 3 черных шара. Какова вероятность того, что два наудачу выбранных шара имеют разный цвет?

6. В урне 2 синих, 2 красных и один желтый шар. Определить вероятность того, что среди взятых наугад двух шаров не будет желтого.

7. Числа 1, 2, 3, 4, 5 написаны на пяти карточках. Наугад последовательно выбираются три карточки и ставятся слева направо. Какова вероятность того, что полученное трехзначное число будет четным?

8. Телефонный номер состоит из семи цифр. Найти вероятность того, что первая цифра номера – двойка.

9. На перекрестке установлен автоматический светофор, в котором одну минуту горит зеленый свет и полминуты – красный, затем снова одну минуту – зеленый и полминуты – красный и т.д. В случайный момент времени к перекрестку подъезжает автомобиль. Какова вероятность того, что он проедет перекресток без остановки?

Вероятность суммы и произведения событий

10. Два стрелка, для которых вероятность попадания в мишень равна 0,8 и 0,7, производят по одному выстрелу в мишень. Найти вероятность:

а) двух попаданий в мишень;

б) хотя бы одного попадания в мишень.

11. В урне лежат 4 красных и 6 синих шаров. Последовательно вынимают 2 шара без возвращения их обратно. Какова вероятность того, что первый шар будет синим, а второй – красным?

12. В первой урне 5 белых и 10 черных шаров, а во второй – 6 белых и 9 черных. Из обеих урн наудачу извлекают по одному шару. Какова вероятность, что оба шара одного цвета?

13. Бросают 2 кубика. Найти вероятность того, что произведение выпавших очков будет четным.

Формулы полной вероятности и Байеса

14. В первой урне 3 белых и 7 черных шаров, во второй – 6 белых и 3 черных. Из первой урны во вторую переложили шар, цвет которого неизвестен. Затем из второй урны вынули шар. Какова вероятность того, что вынутый шар – черный?

15. В первой урне 2 белых и 4 черных шара, во второй – 1 белый и 3 черных. Из первой урны во вторую наудачу переложили 2 шара, затем из второй урны вынули шар. Найти вероятность того, что:

а) этот шар – белый;

б) переложили шары разных цветов, если вынутый шар оказался черным.

16. В урне шар неизвестного цвета – с равной вероятностью белый или черный. В урну опускают один белый шар, перемешивают

и извлекают один шар. Он оказался белым. Какова вероятность того, что в урне остался белый шар?

Схема Бернулли

17. Два кубика подбрасывают 10 раз. Найти вероятность того, что:

а) ровно 4 раза произведение выпавших очков равно шести;

б) хотя бы один раз произведение равно шести.

18. В урне 6 белых и 4 черных шара. Берут по 2 шара и возвращают их обратно. Опыт проводят 5 раз. Найти вероятность того, что:

а) ровно 3 раза взяли шары разного цвета;

б) хотя бы один раз взяли шары разного цвета.

19. Пусть вы играете с равносильным противником в игру, ничья в которой невозможна. Что вероятнее выиграть: три партии из четырех или пять из восьми?

Глава 12
СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

12.1. Понятие случайной величины

Часто исход случайного эксперимента выражается некоторым числом, как, например, при бросании кубика. Если исход опыта и не описывается числами, можно «принудительно» приписать каждому исходу числовое значение. Так, при бросании монеты можно договориться, что «орел» обозначается единицей, а «решка» — нулем. На самом деле, эта величина есть просто-напросто количество выпавших «орлов»: их может быть 0 штук (выпала «решка») или 1 штука (выпал «орел»). В этих случаях случайный эксперимент характеризуется набором чисел, связанных с каждым элементарным исходом. Этот набор, «облако» чисел можно рассматривать как одно «случайное» число, принимающее значения в зависимости от того, какой произошел исход. Набор чисел может быть конечным или бесконечным: это зависит от количества элементарных исходов эксперимента, т.е. от вероятностного пространства. Неформально говоря, такое число, принимающее случайные значения, и называется случайной величиной.

Примеры.

1. В урне 2 белых и 4 черных шара. Из урны наугад вынимают 4 шара. Будем характеризовать исход эксперимента количеством вынутых белых шаров. Это число может принимать три значения: 0 (ни одного белого), 1 и 2. Всего эксперимент имеет $C_6^4 = 15$ элементарных исходов. При этом наша случайная величина принимает нулевое значение в одном исходе (все шары черные), значение 1 в $2C_4^3 = 8$ исходах (один белый — двумя способами, и остальные 3 черных выбираются из четырех возможностей); наконец, значение 2 она принимает в $C_4^2 = 6$ исходах (оба белых шара с любой парой из четырех черных).

2. Проводятся n независимых испытаний Бернулли. Следим за количеством успехов k. Ясно, что такая случайная величина принимает $n + 1$ значение от 0 до n включительно.

3. Обратная ситуация: независимые испытания проводятся до тех пор, пока не случится ровно k успехов. Число, характеризующее данный эксперимент, — количество n проведенных испытаний. Ясно, что такая случайная величина может принимать любые натуральные значения, не меньшие k (если провели меньше, чем k, испытаний, в них не может быть k успехов). Итак, $n = k, k + 1, \ldots$, т.е. случайная величина принимает бесконечное количество значений. Такое бесконечное количество мы называли счетным: набор значений образует числовую последовательность.

4. Пусть точку наугад бросают на отрезок $[0, 1]$. Каждому исходу эксперимента естественным образом соответствует число — координата точки попадания. В данном случае случайная величина также имеет бесконечное количество значений. Но эта бесконечность «больше», чем в примере 3: значения невозможно пронумеровать, они не составляют последовательности. Такие множества имеют так называемую мощность континуум.

5. Измеряется некоторая физическая величина, например температура больного. Каждое измерение (случайный эксперимент) характеризуется случайным числом — измеренным значением. Данная случайная величина принимает значения из некоторого промежутка — например, от 35 до 42 °C.

Когда каждому элементарному исходу случайного эксперимента мы ставим в соответствие число, на самом деле мы определяем некоторую числовую функцию. Строгое определение случайной величины базируется именно на понятии функции. Пусть задано вероятностное пространство $\{\Omega, \mathcal{F}, P\}$, где, напомним, Ω — пространство элементарных исходов, $\mathcal{F}$ — алгебра событий и P — функция вероятности, удовлетворяющая аксиомам Колмогорова. *Случайной величиной* ξ называется числовая функция $\xi = \xi(\omega)$, $\omega \in \Omega$, заданная на пространстве элементарных событий Ω:

$$\xi: \Omega \to \boldsymbol{R}.$$

Строго говоря, случайной величиной называется не всякая такая функция. Существует дополнительное условие: для любого действительного числа x функция $\xi(\omega)$ определяет случайное событие

$$A_x = \{\omega | \xi(\omega) < x\},$$

состоящее в том, что случайная величина приняла значение, меньшее заданного x. Естественно потребовать, чтобы была возможность вычислить вероятность такого события. На математическом языке это означает, что событие A_x принадлежит алгебре $\mathcal{F}$. Итак, повторим.

> ***Случайной величиной*** называется функция $\xi = \xi(\omega)$, $\omega \in \Omega$, такая, что $A_x = \{\omega | \xi(\omega) < x\} \in \mathcal{F}$ при любом $x \in \boldsymbol{R}$.

Рассмотрим простой пример бросания монеты. Пространство элементарных событий состоит из двух исходов: $\Omega = \{\text{О}, \text{Р}\}$. В качестве алгебры $\mathcal{F}$ удобно взять совокупность всех подмножеств Ω, а вероятность P задается классической схемой $P(\text{О}) = P(\text{Р}) = {}^1/_2$. Мы договорились, что случайную величину (число выпавших «орлов») можно задать так: «орлу» соответствует единица, «решке» – нуль, т.е. $\xi(\text{О}) = 1$, $\xi(\text{Р}) = 0$.

Рассмотрим события A_x при различных значениях x. Если $x \leqslant 0$, событие A_x никогда не наступает, так как случайная величина принимает только два значения, нуль и единицу, и не бывает отрицательной. Значит, в этом случае $A_x = \varnothing$. Пусть, к примеру, $x = 0{,}1$. Тогда $A_{0,1} = \{\omega | \xi(\omega) < 0{,}1\}$ и очевидно, что такому событию благоприятствует лишь один исход: выпадение «решки». Значит, $A_{0,1} = \{\text{Р}\}$. То же можно сказать про все события A_x, если $0 < x \leqslant 1$. Если же $x > 1$, то событие A_x является достоверным, так как при любом исходе эксперимента значение случайной величины будет меньше x. В этом случае $A_x = \{\text{О}, \text{Р}\} = \Omega$. Итак,

$$A_x = \begin{cases} \varnothing, & x \leqslant 0, \\ \{\text{Р}\}, & 0 < x \leqslant 1, \\ \Omega, & x > 1. \end{cases}$$

Видно, что во всех случаях выполняется условие $A_x \in \mathcal{F}$.

12.2. Конечные случайные величины

12.2.1. Закон распределения конечной случайной величины

Случайная величина, принимающая конечное число значений, называется ***конечной случайной величиной***. Пусть пространство элементарных исходов конечно: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. Тогда, как уже указывалось, в качестве алгебры $\mathcal{F}$ удобно взять множество $\mathcal{P}(\Omega)$ всех подмножеств Ω. Вероятность P любого случайного события, связанного с данным экспериментом, полностью определяется набором неотрицательных чисел $p_i = P(\omega_i)$, $i = 1, 2, \ldots, n$, таких, что $p_1 + p_2 + \ldots + p_n = 1$. Такое вероятностное пространство удобно представить с помощью таблицы

$$\begin{pmatrix} \omega_1 & \omega_2 & \ldots & \omega_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}.$$

Зададим случайную величину на таком вероятностном пространстве. Поскольку число элементарных исходов конечно, получится конечная случайная величина. Функцию $\xi(\omega)$, заданную на конечном числе аргументов, также удобно задать т а б л и ч н ы м способом:

$$\begin{pmatrix} \omega_1 & \omega_2 & \ldots & \omega_n \\ x_1 & x_2 & \ldots & x_n \end{pmatrix}.$$

Будем пока предполагать, что все числа x_k различны. Случайная величина принимает значение x_k, если произошел исход ω_k, вероятность которого равна p_k. Другими словами, случайная величина ξ принимает значение x_k с вероятностью p_k. Еще точнее: вероятность события $\{\xi(\omega) = x_k\}$ равна p_k. Конечная случайная величина полностью определяется своими значениями и их вероятностями. Поэтому таблица

$$\xi = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}, \quad p_1 + p_2 + \ldots + p_n = 1,$$

часто отождествляется с самой случайной величиной и называется ***законом распределения*** конечной случайной величины. Закон распределения можно записать короче:

$$\xi = \begin{pmatrix} x_i \\ p_i \end{pmatrix}, \quad i = 1, 2, \ldots, n.$$

Примеры.

1. При бросании монеты случайная величина ξ, равная количеству выпавших «орлов», принимает значения 0 и 1 с вероятностями $^1/_2$, так как эксперимент имеет два равновероятных исхода:

$$\xi = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}.$$

2. При бросании кубика случайная величина, равная выпавшему числу, принимает шесть различных значений от 1 до 6 с равными вероятностями:

$$\xi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

3. Пусть в урне 4 белых и 6 черных шаров. Из урны наугад вынимают один шар. Случайная величина ξ – число оставшихся в урне белых шаров. Эта случайная величина может равняться трем (если вынули белый шар) и четырем (если вынули черный). Соответствующие вероятности равны 0,4 и 0,6:

$$\xi = \begin{pmatrix} 3 & 4 \\ 0,4 & 0,6 \end{pmatrix}.$$

4. Согласно концепции случайной величины одно число можно также считать случайной величиной, принимающей единственное значение с вероятностью 1:

$$c = \begin{pmatrix} c \\ 1 \end{pmatrix}.$$

Такую «случайную» величину, естественно, называют *постоянной случайной величиной.*

Разным элементарным исходам может быть поставлено в соответствие одно и то же значение случайной величины: среди чисел $x_1, x_2, \ldots, x_n$ могут встретиться повторы. В этом случае также легко посчитать вероятность события $\{\omega | \xi(\omega) = x_k\}$. Это событие состоит из всех элементарных исходов ω, для которых $\xi(\omega) = x_k$. Поскольку элементарные исходы несовместны, вероятность такого события равна сумме вероятностей соответствующих исходов. В этом случае для построения закона распределения нужно «склеить» столбцы, отвечающие тем элементарным исходам, на которых значения случайной

величины совпадают. Таким образом, любую конечную случайную величину после «склейки» можно изобразить в виде таблицы

$$\xi = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}, \quad p_1 + p_2 + \ldots + p_n = 1.$$

Другой наглядный способ изображения конечной случайной величины – графический. Отложим по оси абсцисс значения x_i случайной величины, а по оси ординат – соответствующие вероятности p_i, и точки соединим отрезками прямых (рис. 12.1). Получится график в виде гребенки, который называют *графиком функции вероятностей* конечной случайной величины.

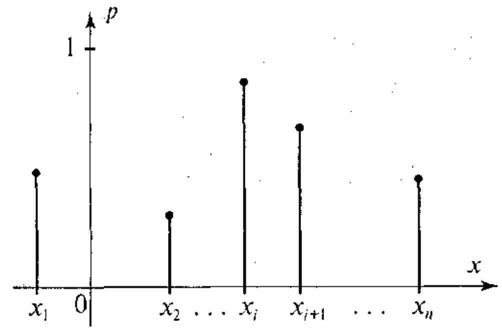


Рис. 12.1

Примеры.

1. Монету бросают 2 раза. Случайная величина – количество выпавших «орлов». Эксперимент имеет четыре равновероятных исхода: «РР», «ОР», «РО» и «ОО». Первому исходу соответствует нулевое значение случайной величины (оба раза «решка»), второму и третьему исходам соответствует значение 1, а четвертому – значение 2. Вероятность каждого исхода равна $^1/_4$. Значит,

$P(\xi(\omega) = 0) = P(\text{PP}) = {}^1/_4$;

$P(\xi(\omega) = 1) = P(\{\text{OP, PO}\}) = P(\text{OP}) + P(\text{PO}) = {}^1/_4 + {}^1/_4 = {}^1/_2$;

$P(\xi(\omega) = 2) = P(\text{OO}) = {}^1/_4$

и случайная величина имеет распределение

$$\xi = \begin{pmatrix} 0 & 1 & 2 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

(при построении закона «склеили» столбцы, соответствующие второму и третьему исходам). На рис. 12.2 изображен график функции вероятности указанной случайной величины.

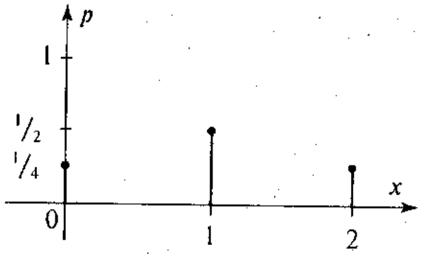


Рис. 12.2

2 (**биномиальное распределение**). Двукратное бросание монеты можно рассматривать как реализацию схемы Бернулли при $n = 2$, $p = q = {}^1/_2$. Если считать успехом выпадение «орла» при одном бросании, случайная величина равна числу успехов при двух независимых испытаниях. Эта случайная величина есть частный случай случайной величины, равной числу успехов в n испытаниях Бернулли с вероятностью успеха p. Мы уже выясняли, что такая случайная величина принимает $n + 1$ значение от 0 до n. Событие $\xi(\omega) = k$ состоит в том, что при n испытаниях произошло ровно k успешных. Значит, вероятности каждого значения можно определить по формуле Бернулли

$$P(\{\xi(\omega) = k\}) = P_n(k) = C_n^k p^k q^{n-k},$$

где $q = 1 - p$. В связи с тем что вероятности определяются с помощью биномиальных коэффициентов C_n^k, говорят, что данная случайная величина имеет *биномиальное распределение*, или саму случайную величину называют *биномиальной*. Распределение определяется параметрами n и p, поэтому биномиальная случайная величина обозначается как $B_{n,p}$. Итак, биномиальное распределение описывается следующей таблицей:

$$B_{n,p} = \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix} =$$

$$= \begin{pmatrix} 0 & 1 & \dots & k & \dots & n-1 & n \\ q^n & npq^{n-1} & \dots & C_n^k p^k q^{n-k} & \dots & np^{n-1}q & p^n \end{pmatrix}.$$

По свойству биномиальных коэффициентов выполняется основное свойство закона распределения конечной случайной величины: сумма вероятностей во второй строке таблицы равна единице.

В частности, при $n = 2$ и $n = 3$ биномиальное распределение принимает вид

$$B_{2,p} = \begin{pmatrix} 0 & 1 & 2 \\ q^2 & 2pq & p^2 \end{pmatrix},$$

$$B_{3,p} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ q^3 & 3pq^2 & 3p^2q & p^3 \end{pmatrix}.$$

Пусть, например, в урне 2 белых и 3 черных шара. Из урны вынимают 2 шара и возвращают обратно. Опыт повторяют 3 раза. Успешным считается эксперимент, в котором вынимают шары разного цвета. Найдем вероятность p успеха в единичном эксперименте. Всего возможно $C_5^2 = 10$ исходов. Успешными будут $2 \cdot 3 = 6$ исходов (любой белый с любым черным). Значит, $p = 0{,}6$, $q = 1 - p = 0{,}4$ и число успехов имеет следующее распределение:

$$B_{3,\,0{,}6} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0{,}064 & 0{,}288 & 0{,}432 & 0{,}216 \end{pmatrix}.$$

Легко проверить, что сумма вероятностей действительно равна единице.

3. Рассмотрим эксперимент из примера 1 параграфа 12.1: в урне 2 белых и 4 черных шара, вынимают наугад 4 шара и случайная величина равна количеству белых шаров среди вынутых. Мы выяснили, что всего возможно 15 равновероятных исходов. Ни одного белого шара не вынимают в одном случае, один белый будет при 8 исходах, а два белых – при 6 исходах. Таким образом, данная случайная величина задается следующим законом распределения:

$$\xi = \begin{pmatrix} 0 & 1 & 2 \\ {}^1/_{15} & {}^8/_{15} & {}^6/_{15} \end{pmatrix}.$$

Заметим, что указанная случайная величина не является биномиальной: реализована схема без возвращения, которую нельзя считать последовательностью независимых испытаний.

12.2.2. Совместное распределение случайных величин

Пусть заданы две конечные случайные величины:

$$\xi = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ p_1 & p_2 & \dots & p_m \end{pmatrix},$$

$$\eta = \begin{pmatrix} y_1 & y_2 & \dots & y_n \\ q_1 & q_2 & \dots & q_n \end{pmatrix}.$$

Событие $\{\xi = x_i\}\{\eta = y_j\}$ состоит в том, что одновременно случайная величина ξ принимает значение x_i, а случайная величина η — значение y_j. Назовем вероятности таких событий *совместными вероятностями* и обозначим их через p_{ij}:

$$p_{ij} = P(\xi = x_i, \eta = y_j), \quad i = 1, 2, \ldots, m; j = 1, 2, \ldots, n.$$

Набор точек $\{x_i, y_j\}$ вместе с совместными вероятностями p_{ij} образуют ***совместное распределение*** случайных величин ξ и η.

∇1. *Сумма всех совместных вероятностей равна единице*:

$$\sum_{i,j} p_{ij} = 1.$$

Доказательство. Зафиксируем индекс i и рассмотрим сумму $p_{i1} + p_{i2} + \ldots + p_{in}$. Вероятности произведения можно выразить через условные вероятности:

$$P(\xi = x_i, \eta = y_j) = P(\xi = x_i)P(\xi = x_i \mid \eta = y_j).$$

Поэтому

$$p_{i1} + p_{i2} + \ldots + p_{in} = \sum_{j=1}^{n} P(\xi = x_i)P(\xi = x_i \mid \eta = y_j).$$

Приглядевшись внимательнее, мы заметим, что сумма в правой части равенства есть не что иное, как вероятность $P(\xi = x_i) = p_i$, вычисленная по формуле полной вероятности. Значит,

$$p_{i1} + p_{i2} + \ldots + p_{in} = p_i.$$

Однако сумму всех чисел p_{ij} можно представить в виде суммы m слагаемых вида $p_{i1} + p_{i2} + \ldots + p_{in}$. Окончательно получаем

$$\sum_{i,j} p_{ij} = p_1 + p_2 + \ldots + p_m.$$

Сумма в правой части по определению случайной величины равна единице, что и доказывает свойство.

Отметим, что, если бы мы зафиксировали индекс j, аналогичные рассуждения привели бы нас к равенству

$$p_{1j} + p_{2j} + \ldots + p_{mj} = q_j,$$

из которого также непосредственно вытекает доказываемое свойство.

Совместное распределение двух конечных случайных величин удобно задать с помощью таблицы (матрицы) из m строк и n столбцов:

$\xi \backslash \eta$	y_1	y_2	…	y_j	…	y_n	
x_1	p_{11}	p_{12}	…	p_{1j}	…	p_{1n}	p_1
x_2	p_{21}	p_{22}	…	p_{2j}	…	p_{2n}	p_2
⋮	…	…	…	…	…	…	⋮
x_i	p_{i1}	p_{i2}	…	p_{ij}	…	p_{in}	p_i
⋮	…	…	…	…	…	…	⋮
x_m	p_{m1}	p_{m2}	…	p_{mj}	…	p_{mn}	p_m
	q_1	q_2	…	q_j	…	q_n	1

Как мы выяснили, в этой матрице сумма чисел по строкам равна вероятностям p_i, а сумма по столбцам равна вероятностям q_j, что и отражено в таблице. Таким образом, можно сделать следующий вывод.

> Зная совместное распределение конечных случайных величин ξ и η, можно восстановить законы распределения величин ξ и η.

Обратное утверждение неверно. Можно привести пример, когда распределение обеих случайных величин известно, но не все совместные вероятности поддаются определению без дополнительной информации.

В таблице вероятности p_i и q_j записываются как бы «на полях». Поэтому распределения случайных величин ξ и η называют *маргинальными* (margin — поле) по отношению к их совместному распределению.

Примеры.

1. Пусть в урне по одному белому и черному шару. Из урны 2 раза наугад вынимают один шар и возвращают обратно. Случайные величины ξ и η — количество белых шаров, вынутых соответственно в первый и второй раз. Данные случайные величины принимают значения 0 и 1 с равными вероятностями $^1/_2$:

$$\xi = \begin{pmatrix} 0 & 1 \\ ½ & ½ \end{pmatrix}, \quad \eta = \begin{pmatrix} 0 & 1 \\ ½ & ½ \end{pmatrix}.$$

Найдем их совместное распределение, т.е. вычислим вероятности p_{ij}, $i, j = 1, 2$. Всего возможны четыре равновероятных исхода, которые понятным образом можно обозначить {○, ●}, {●, ○}, {●, ●}, {○, ○}. Ситуация {○, ●}, означающая, что в первый раз вынули белый шар, а во второй – черный, соответствует событию $\{\xi = 1\}\{\eta = 0\}$. Аналогично разбираются и другие ситуации. Значит, $p_{ij} = {}^1/_4$ при всех $i, j = 1, 2$. Таблица совместного распределения выглядит следующим образом:

	0	1	
0	$^1/_4$	$^1/_4$	$^1/_2$
1	$^1/_4$	$^1/_4$	$^1/_2$
	$^1/_2$	$^1/_2$	1

2. Ситуация та же, что в примере 1: в урне по одному белому и черному шару. Теперь шар вынимают один раз, случайная величина ξ – количество вынутых белых шаров, η – количество оставшихся белых шаров. Нетрудно понять, что случайные величины принимают те же значения 0 и 1 с равными вероятностями $^1/_2$. Однако их совместное распределение в данном случае будет другим. В самом деле, нельзя взять один белый шар ($\xi = 1$) так, чтобы в урне остался тоже один белый шар ($\eta = 1$). Значит, событие $\{\xi = 1\}\{\eta = 1\}$ невозможно. Точно так же невозможно событие $\{\xi = 0\}\{\eta = 0\}$, т.е. $p_{11} = p_{22} = 0$. События $\{\xi = 1\}\{\eta = 0\}$ и $\{\xi = 0\}\{\eta = 1\}$ состоят в том, что вынули соответственно белый или черный шар. Это равновероятные события с вероятностью $^1/_2$, т.е. $p_{12} = p_{21} = {}^1/_2$. Совместное распределение выглядит так:

	0	1	
0	0	$^1/_2$	$^1/_2$
1	$^1/_2$	0	$^1/_2$
	$^1/_2$	$^1/_2$	1

Видно, что одинаковые по форме случайные величины имеют совершенно разные совместные распределения. В чем же отличие рассмотренных примеров? В первом из них значение случайной величины ξ никак не влияет на значение η. Во втором же примере их значения жестко связаны: если $\xi = 1$, то $\eta = 0$ (если взяли один белый шар, то белых шаров не осталось ни одного) и наоборот. В данных примерах мы столкнулись с важным понятием независимости случайных величин.

Две конечные случайные величины называются *независимыми*, если события $\{\xi = x_i\}$ и $\{\eta = y_j\}$ независимы при всех $i = 1, 2, ..., m$; $j = 1, 2, ..., n$. В противном случае случайные величины *зависимы*. Если случайные величины независимы, то по известным распределениям величин ξ и η очень просто построить совместное распределение:

$$p_{ij} = P(\xi = x_i)P(\eta = y_j) = p_i q_j, \quad i = 1, 2, ..., m; j = 1, 2, ..., n.$$

Именно такой случай мы рассмотрели в примере 1. Если же случайные величины зависимы (как в примере 2), для того чтобы построить совместное распределение, потребовались дополнительные соображения, учитывающие специфику задачи.

12.2.3. Действия над конечными случайными величинами

Мы ввели новый математический объект – случайные величины. Теперь требуется определить некоторые арифметические операции над ними. Итак, пусть снова заданы две конечные случайные величины:

$$\xi = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ p_1 & p_2 & \dots & p_m \end{pmatrix},$$

$$\eta = \begin{pmatrix} y_1 & y_2 & \dots & y_n \\ q_1 & q_2 & \dots & q_n \end{pmatrix}.$$

Их ***суммой*** называется случайная величина $\xi + \eta$, значениями которой являются всевозможные суммы $x_i + y_j$, $i = 1, 2, ..., m$; $j = 1, 2, ..., n$, с совместными вероятностями $p_{ij} = P(\xi = x_i, \eta = y_j)$.

Произведением этих случайных величин называется случайная величина $\xi\eta$, значениями которой являются всевозможные произведения $x_i y_j$ с теми же вероятностями p_{ij}.

Определения суммы и произведения корректны, так как сумма всех совместных вероятностей равна единице (см. подпараграф 12.2.2). Значит, в результате сложения и умножения действительно получается случайная величина.

Приведем алгебраические свойства операций над случайными величинами, вытекающие из определений.

Ассоциативность:

$$(\xi + \eta) + \zeta = \xi + (\eta + \zeta), \qquad (\xi\eta)\zeta = \xi(\eta\zeta).$$

Коммутативность:

$$\xi + \eta = \eta + \xi, \qquad \xi\eta = \eta\xi.$$

Дистрибутивность:

$$\xi(\eta + \zeta) = \xi\eta + \xi\zeta.$$

Под равенством случайных величин понимается одинаковый закон распределения. Как обычно, ассоциативные и коммутативные законы позволяют строить суммы $\xi_1 + \xi_2 + ... + \xi_n$ и произведения $\xi_1\xi_2 \cdots \xi_n$ произвольного количества случайных величин (неважно, в каком порядке).

Итак, закон распределения суммы и произведения случайных величин легко построить, зная их совместное распределение. Если случайные величины независимы, совместное распределение восстанавливается по законам распределения ξ и η: $p_{ij} = p_i q_j$.

Примеры.

1 (**сложение с постоянной и умножение на постоянную**). Пусть c – постоянная случайная величина. Все возможные суммы и произведения значений случайной величины ξ и c есть соответственно числа $x_i + c$ и cx_i. Вероятность того, что одновременно $\xi = x_i$, а постоянная принимает значение c, равна p_i. Поэтому сумма ξ с постоянной и произведение ξ на постоянную случайную величину c распределены следующим образом:

$$\xi + c = \begin{pmatrix} x_1 + c & x_2 + c & \dots & x_n + c \\ p_1 & p_2 & \dots & p_n \end{pmatrix},$$

$$c\xi = \begin{pmatrix} cx_1 & cx_2 & \dots & cx_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}.$$

Сложение с постоянной можно интерпретировать как сложение случайной величины с числом, а умножение на постоянную – как умножение на число.

2. Пусть ξ и η – случайные величины из примера 1 подпараграфа 12.2.2 (количество белых шаров, вынутых из урны с одним белым и одним черным шаром и возвращенных в урну, соответственно в первый и второй раз). Построим закон распределения их суммы $\xi + \eta$. Сумму можно интерпретировать как общее количество вынутых белых шаров. Случайные величины независимы, так как независимы события $\{\xi = x_i\}$ и $\{\eta = y_j\}$ ($x_1 = y_1 = 0$, $x_2 = y_2 = 1$). Если $\xi = 0$, $\eta = 0$, то $\xi + \eta = 0$ *и вероятность этого события равна* $^1/_4$. Аналогично разбираются остальные комбинации. В итоге

$$\xi + \eta = \begin{pmatrix} 0 & 1 & 1 & 2 \\ ¼ & ¼ & ¼ & ¼ \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ ¼ & ½ & ¼ \end{pmatrix}$$

(«склеили» второй и третий столбцы).

Данный пример представляет собой частный случай более общей ситуации. Пусть проводятся два независимых испытания с вероятностью успеха p, случайная величина ξ есть число успехов в первом испытании, η – число успехов во втором испытании. Очевидно, что данные случайные величины независимы и имеют одинаковые законы распределения: они принимают значение 0 с вероятностью $q = 1 - p$ (неудача) и значение 1 с вероятностью p (успех):

$$\xi = \eta = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}.$$

Говорят, что эти случайные величины имеют *распределение Бернулли* (или сами величины называют *бернуллиевыми*). Легко построить закон распределения суммы $\xi + \eta$. Всего возможно четыре комбинации значений ξ и η: (0, 0), (1, 0), (0, 1), (1, 1). Сумма принимает нулевое значение в первом случае (вероятность этого события равна q^2), *значение 1 – во втором и в третьем* (*с вероятностью* pq) и значение 2 – в последнем (с вероятностью p^2). После «склейки» второго и третьего случаев (дающих одинаковое значение суммы) получаем закон распределения

$$\xi + \eta = \begin{pmatrix} 0 & 1 & 2 \\ q^2 & 2pq & p^2 \end{pmatrix}.$$

Полученное распределение нам хорошо знакомо: это не что иное, как распределение биномиальной случайной величины $B_{2,p}$, т.е. распределение числа успехов в двух независимых испытаниях Бернулли. Такой результат не должен явиться неожиданностью: ведь если ξ – число успехов в первом испытании, а η – число успехов во втором, то их сумма, разумеется, есть число успехов в двух испытаниях. Справедлив и более общий результат.

Теорема 12.1. *Пусть* $\xi_1, \xi_2, ..., \xi_n$ *– независимые бернуллиевы случайные величины*:

$$\xi_1 = \xi_2 = ... = \xi_n = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}.$$

Тогда их сумма есть биномиальная случайная величина:

$$\boxed{B_{n,p} = \xi_1 + \xi_2 + ... + \xi_n}$$

Примем этот факт без доказательства как достаточно естественный: если ξ_k можно интерпретировать как число успехов в k-м испытании схемы Бернулли, то их сумма, конечно, есть число успехов в n испытаниях.

3. Сложим случайные величины из примера 2 подпараграфа 12.2.2 (ξ – количество вынутых белых шаров, η – количество

оставшихся). Они распределены так же, как рассмотренные выше, но, в отличие от них, зависимы. На самом деле, нетрудно убедиться, что эти случайные величины связаны равенством $\eta = 1 - \xi$. Вообще, если для некоторых случайных величин выполняется равенство

$$\eta = a\xi + b,$$

где a и b – постоянные, то такие величины являются зависимыми. В этом случае говорят, что они *линейно зависимы*. Ясно, что сумма $\xi + \eta$ может принимать только одно значение 1 (в случаях, когда $\xi = 1$, $\eta = 0$ или $\xi = 0$, $\eta = 1$). Остальные комбинации значений невозможны. Итак, сумма данных случайных величин есть постоянная:

$$\xi + \eta = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Этот результат можно было предвидеть: ведь сумма ξ и η, конечно же, равна общему количеству белых шаров.

4 (**сложение и умножение одинаковых случайных величин**). Пусть ξ – число, выпавшее на грани кубика при одном бросании:

$$\xi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

Найдем сумму $\xi + \xi$ этой случайной величины с самой собой. Складываются зависимые величины. Поэтому возможны далеко не все комбинации значений. Например, $P(\xi = 1, \xi = 2) = 0$, так как ξ не может одновременно принимать разные значения. При этом $P(\xi = i, \xi = i) = P(\xi = i) = 1/6$, $i = 1, 2, \ldots, 6$. Таким образом, сумма распределена по следующему закону:

$$\xi + \xi = \begin{pmatrix} 2 & 4 & 6 & 8 & 10 & 12 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

Аналогичными рассуждениями проверяется, что для любой конечной случайной величины справедливо

$$\xi + \xi = \begin{pmatrix} 2x_1 & 2x_2 & \ldots & 2x_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix} = 2\xi.$$

То же и при умножении случайной величины на самое себя: возможны только события $\{\xi = x_i\}\{\xi = x_i\}$ и их вероятность равна p_i. Поэтому

$$\xi\xi = \xi^2 = \begin{pmatrix} x_1^2 & x_2^2 & \ldots & x_n^2 \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}.$$

12.3. Числовые характеристики конечных случайных величин

12.3.1. Математическое ожидание

Пусть случайная величина связана с некоторым случайным экспериментом. Проводя серию экспериментов, мы каждый раз будем получать какое-то ее значение в зависимости от того, какой произошел исход. Очень интересен вопрос о с р е д н е м значении случайной величины. Такой вопрос возникает в азартных играх (а именно они стимулировали развитие теории вероятностей), когда случайная величина есть некий денежный выигрыш. Хотелось бы заранее оценить, на какой средний выигрыш можно рассчитывать, начиная игру. Пусть, например, игра состоит в бросании кубика (игральной кости): выигрыш в рублях при каждом бросании равен числу, выпавшему на грани. Пусть кубик бросили n раз. При этом единица выпала m_1 раз, двойка – m_2 раз и т.д. ($m_1 + m_2 + \ldots + m_6 = n$). Общий выигрыш составит

$$S = 1 \cdot m_1 + 2 \cdot m_2 + 3 \cdot m_3 + 4 \cdot m_4 + 5 \cdot m_5 + 6 \cdot m_6$$

рублей. Значит, на одно бросание приходится средний выигрыш, равный

$$s = \frac{S}{n} = \frac{1 \cdot m_1 + 2 \cdot m_2 + 3 \cdot m_3 + 4 \cdot m_4 + 5 \cdot m_5 + 6 \cdot m_6}{n} =$$
$$= 1\frac{m_1}{n} + 2\frac{m_2}{n} + 3\frac{m_3}{n} + 4\frac{m_4}{n} + 5\frac{m_5}{n} + 6\frac{m_6}{n}.$$

К сожалению, эту величину можно посчитать только после опытов.

Обратим внимание на то, что величины m_i/n, $i = 1, 2, \ldots, 6$, есть не что иное, как относительные частоты выпадения числа i. Из закона статистической устойчивости (фундаментального свойства случайного события) нам известно, что при большом числе испытаний относительная частота стабилизируется около значения вероятности. Поэтому, чтобы оценить средний выигрыш до игры, можно заменить в формуле относительные частоты на соответствующие вероятности – ведь вероятности мы можем посчитать, не проводя экспериментов. Таким образом,

$$s \approx m = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4 + 5 \cdot p_5 + 6 \cdot p_6.$$

В нашем случае $p_1 = p_2 = \ldots = p_6 = 1/6$ и

$$m = \frac{1+2+3+4+5+6}{6} = 3{,}5.$$

С бросанием кубика связана случайная величина, принимающая значения от 1 до 6 с равными вероятностями $^1/_6$:

$$\xi = \begin{pmatrix} i \\ 1/6 \end{pmatrix}, \quad i = 1, 2, \ldots, 6.$$

Значит, приближенное значение среднего выигрыша мы подсчитывали по формуле

$$m = x_1 p_1 + x_2 p_2 + \ldots + x_6 p_6.$$

Такая сумма для произвольной конечной случайной величины ξ называется ***математическим ожиданием*** ξ и обозначается $M\xi$. Итак, математическим ожиданием конечной случайной величины

$$\xi = \begin{pmatrix} x_i \\ p_i \end{pmatrix}, \quad i = 1, 2, \ldots, m,$$

называется число

$$\boxed{M\xi = x_1 p_1 + x_2 p_2 + \ldots + x_m p_m = \sum_{i=1}^{m} x_i p_i}.$$

Математическое ожидание случайной величины, как мы выяснили, характеризует ее среднее, «ожидаемое», значение. Конечно, величина $M\xi$ не обязательно присутствует среди значений x_i случайной величины (как это и произошло в разобранном примере).

Примеры.

1. Пусть случайная величина имеет равновероятные значения:

$$\xi = \begin{pmatrix} x_1 & x_2 & \ldots & x_m \\ 1/m & 1/m & \ldots & 1/m \end{pmatrix}.$$

Тогда по определению

$$M\xi = x_1 \frac{1}{m} + x_2 \frac{1}{m} + \ldots + x_m \frac{1}{m} = \frac{x_1 + x_2 + \ldots + x_m}{m},$$

т.е. математическое ожидание такой случайной величины совпадает со средним арифметическим ее значений.

2. Пусть в кошельке 5 монет по одному рублю, 3 пятирублевые монеты и по одной монете достоинством 50 и 100 рублей. Всего в кошельке 10 монет общим достоинством 170 рублей. «Среднее достоинство» монеты в кошельке составляет 17 рублей:

$$17 = \frac{1 \cdot 5 + 5 \cdot 3 + 50 \cdot 1 + 100 \cdot 1}{10} = 1 \cdot 0{,}5 + 5 \cdot 0{,}3 + 50 \cdot 0{,}1 + 100 \cdot 0{,}1.$$

Видно, что средний номинал монеты равен математическому ожиданию случайной величины

$$\xi = \begin{pmatrix} 1 & 5 & 50 & 100 \\ 0{,}5 & 0{,}3 & 0{,}1 & 0{,}1 \end{pmatrix},$$

равной номиналу монеты, случайно вынутой из кошелька. Этот пример еще раз убеждает в том, что математическое ожидание есть среднее значение случайной величины.

Математическому ожиданию можно придать весьма наглядную механическую интерпретацию. Представим себе стержень (отрезок оси абсцисс), на который в точках x_i нанизаны шары массой p_i (рис. 12.3). Используя школьные знания по физике, можно вывести, что центр тяжести такой системы находится в точке с координатой, равной $M\xi$.

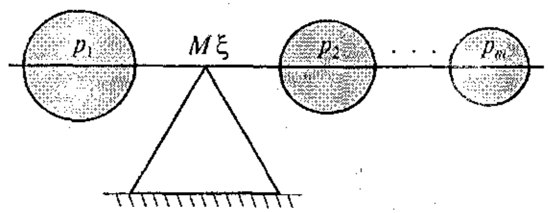


Рис. 12.3

Математическое ожидание обладает следующими свойствами.

1°. *Математическое ожидание постоянной равно ей самой:*

$$Mc = c.$$

В самом деле, по определению математическое ожидание случайной величины

$$c = \begin{pmatrix} c \\ 1 \end{pmatrix}$$

равно $Mc = c \cdot 1 = c$.

2°. *Если случайная величина ξ принимает только неотрицательные значения, то $M\xi \geqslant 0$.*

3°. *Константу можно выносить за знак математического ожидания:*

$$M(c\xi) = cM\xi.$$

Свойства 2° и 3° непосредственно следуют из определения математического ожидания.

4°. *Математическое ожидание суммы (разности) случайных величин равно сумме (разности) их математических ожиданий:*

$$M(\xi \pm \eta) = M\xi \pm M\eta.$$

По определению суммы случайных величин и математического ожидания

$$M(\xi+\eta)=\sum_{i,j}(x_i+y_j)p_{ij},$$

где p_{ij} — совместные вероятности ξ и η. Разобьем выражение в правой части на две суммы:

$$M(\xi+\eta)=\sum_{i,j}x_i p_{ij}+\sum_{i,j}y_j p_{ij}.$$

Первую сумму можно представить в виде

$$x_1(p_{11}+p_{12}+\ldots+p_{1n})+x_2(p_{21}+p_{22}+\ldots+p_{2n})+\ldots+$$
$$+x_m(p_{m1}+p_{m2}+\ldots+p_{mn}).$$

Однако по свойству совместных вероятностей сумма в каждой i-й скобке равна p_i (сумма по строкам в таблице совместного распределения). Значит, рассматриваемая сумма равна

$$x_1p_1+x_2p_2+\ldots+x_mp_m=M\xi.$$

Точно так же преобразовываем вторую сумму, используя тот факт, что $p_{1j}+p_{2j}+\ldots+p_{mj}=q_j$ (сумма по столбцам):

$$y_1q_1+y_2q_2+\ldots+y_nq_n=M\eta.$$

Эти два равенства и доказывают свойство для суммы. Для разности доказательство дословно повторяется.

Свойства 3° и 4° — л и н е й н ы е свойства математического ожидания. Их, как обычно, можно объединить в одно: *математическое ожидание линейной комбинации случайных величин равно линейной комбинации их математических ожиданий:*

$$M(c_1\xi_1+c_2\xi_2+\ldots+c_n\xi_n)=c_1M\xi_1+c_2M\xi_2+\ldots+c_nM\xi_n.$$

Воспользуемся этим свойством для вычисления математического ожидания биномиальной случайной величины. Как мы выяснили ранее,

$$B_{n,p}=\xi_1+\xi_2+\ldots+\xi_n,$$

где $\xi_1=\xi_2=\ldots=\xi_n=\begin{pmatrix}0 & 1\\ q & p\end{pmatrix}$ — бернуллиевы величины.

Значит,

$$MB_{n,p}=M\xi_1+M\xi_2+\ldots+M\xi_n.$$

Математическое ожидание бернуллиевой величины легко вычисляется по определению:

$$M\xi_i=0\cdot q+1\cdot p=p.$$

Значит, $MB_{n,p}$ есть сумма n одинаковых слагаемых, равных p, т.е.

$$\boxed{MB_{n,p}=np}$$

Пример.

В урне 4 белых и 6 черных шаров. Из урны наугад вынимают шар и возвращают обратно. Опыт повторяют 10 раз, причем успех достигается, если вынимают белый шар. Случайная величина ξ — число успешных испытаний. Это биномиальная случайная величина при $n=10$ и $p=0{,}4$ (вероятность успеха). Значит, $M\xi=MB_{10,\,0,4}=10\cdot0{,}4=4$: среднее число успехов в 10 испытаниях равно четырем.

5°. *Для любой случайной величины справедливо равенство*

$$M(\xi-M\xi)=0.$$

Равенство вытекает из свойств 4°, 1°:

$$M(\xi-M\xi)=M\xi-M(M\xi)=M\xi-M\xi=0$$

(математическое ожидание постоянной, в частности, равной $M\xi$, равно ей самой).

Случайная величина $\mathring{\xi}=\xi-M\xi$ называется *центрированной* (по отношению к ξ), ее математическое ожидание равно нулю. Операция вычитания математического ожидания из случайной величины называется *центрированием.*

6°. *Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий:*

$$M(\xi\eta)=M\xi M\eta$$

(ξ *и* η *независимы*).

По определению

$$M(\xi\eta)=\sum_{i,j}x_iy_jp_{ij}.$$

Однако для независимых случайных величин $p_{ij}=p_iq_j$, поэтому

$$M(\xi\eta)=\sum_{i,j}x_iy_jp_iq_j=\left(\sum_{i=1}^{m}x_ip_i\right)\left(\sum_{j=1}^{n}y_jq_j\right)=M\xi M\eta.$$

Запомним:

- математическое ожидание суммы равно сумме математических ожиданий для **любых** случайных величин;
- математическое ожидание произведения равно произведению математических ожиданий только для **независимых** случайных величин.

Итак, математическое ожидание характеризует среднее значение случайной величины. Говорят, что математическое ожидание – характеристика положения или центральных тенденций.

12.3.2. Дисперсия и среднеквадратичное отклонение

Зная характеристику среднего поведения случайной величины – математическое ожидание, хорошо было бы оценить, насколько случайная величина отклоняется от среднего, насколько велик ее р а з б р о с. Мы не можем выбрать в качестве характеристики разброса величину $M(\xi - M\xi)$ (среднее отклонение от математического ожидания), так как она всегда равна нулю. Поэтому разброс характеризуют средним квадратом отклонений, или дисперсией. ***Дисперсией*** конечной случайной величины ξ называется число

$$D\xi = M(\xi - M\xi)^2.$$

Случайная величина $(\xi - M\xi)^2$ распределена по закону

$$\begin{pmatrix} (x_i - M\xi)^2 \\ p_i \end{pmatrix},$$

и, по определению математического ожидания, дисперсия вычисляется по следующей формуле:

$$\boxed{D\xi = \sum_{i=1}^{n} (x_i - M\xi)^2 p_i}$$

Дисперсию иногда обозначают как $\sigma^2(\xi)$ или σ^2_ξ. Если случайная величина ξ измеряется, например, в метрах, то ее дисперсия $D\xi$ измеряется в $м^2$. Чтобы привести характеристику разброса к единицам измерения ξ, нужно извлечь из нее квадратный корень. Полученное число $\sigma_\xi = \sqrt{D\xi}$ называется ***среднеквадратичным отклонением*** или ***стандартным отклонением*** случайной величины ξ. В русской литературе для обозначения среднеквадратичного отклонения принята аббревиатура СКО.

1°. *Дисперсия любой случайной величины неотрицательна:*

$$D\xi \geqslant 0.$$

При этом $D\xi = 0$ *тогда и только тогда, когда случайная величина постоянна.*

Свойство 1° следует из определения.

2°. *Константа выносится из-под знака дисперсии с квадратом:*

$$D(c\xi) = c^2 D\xi.$$

В самом деле, по определению дисперсии и свойствам математического ожидания

$$D(c\xi) = M[c\xi - M(c\xi)]^2 = M(c\xi - cM\xi)^2 = M(c^2(\xi - M\xi)^2) =$$
$$= c^2 M(\xi - M\xi)^2 = c^2 D\xi.$$

В частности, $D(-\xi) = D\xi$.

3°. *Сдвиг на константу не меняет дисперсии:*

$$D(\xi + c) = D\xi.$$

В самом деле,

$$D(\xi + c) = M[\xi + c - M(\xi + c)]^2 = M(\xi + c - M\xi - c)^2 =$$
$$= M(\xi - M\xi)^2 = D\xi.$$

В частности, в качестве константы можно взять $-M\xi$. Тогда получаем, что *дисперсия центрированной случайной величины* $\mathring{\xi} = \xi - M\xi$ *равна дисперсии* ξ.

Свойства 2° и 3° объединяются в одно:

$$D(a\xi + b) = a^2 D\xi.$$

4°. *Дисперсия суммы независимых случайных величин равна сумме их дисперсий:*

$$D(\xi \pm \eta) = D\xi + D\eta$$

(ξ *и* η *независимы*).

Свойство 4° доказывается следующей цепочкой равенств:

$$D(\xi + \eta) = M[(\xi + \eta) - M(\xi + \eta)]^2 = M[(\xi - M\xi) + (\eta - M\eta)]^2 =$$
$$= M[(\xi - M\xi)^2 + 2(\xi - M\xi)(\eta - M\eta) + (\eta - M\eta)^2] =$$
$$= M(\xi - M\xi)^2 + M(\eta - M\eta)^2 + 2M[(\xi - M\xi)(\eta - M\eta)] =$$
$$= D\xi + D\eta + 2M[(\xi - M\xi)(\eta - M\eta)].$$

По условию случайные величины ξ и η независимы. Можно проверить, что в этом случае и центрированные величины $\xi - M\xi$, $\eta - M\eta$

также независимы. Значит, математическое ожидание их произведения равно произведению математических ожиданий:

$$M[(\xi - M\xi)(\eta - M\eta)] = M(\xi - M\xi)M(\eta - M\eta).$$

Оба сомножителя в последнем произведении равны нулю, что и доказывает свойство 4°. Последнее свойство обобщается для n попарно независимых случайных величин: *если* $\xi_1, \xi_2, \ldots, \xi_n$ *попарно независимы, то*

$$D(\xi_1 + \xi_2 + \ldots + \xi_n) = D\xi_1 + D\xi_2 + \ldots + D\xi_n.$$

5°. *Дисперсия равна «среднему квадрату минус квадрат среднего»:*

$$D\xi = M\xi^2 - (M\xi)^2.$$

Именно эта формула часто применяется при практическом вычислении дисперсии. Она выводится из определения и из свойств математического ожидания:

$$D\xi = M(\xi - M\xi)^2 = M[\xi^2 - 2\xi M\xi + (M\xi)^2] =$$

$$= M\xi^2 - 2M(\xi M\xi) + M(M\xi)^2 =$$

$$= M\xi^2 - 2M\xi M\xi + (M\xi)^2 = M\xi^2 - (M\xi)^2.$$

Примеры.

1. В урне 3 черных и 2 белых шара. Из урны наугад вынимают 2 шара. Случайная величина ξ — число белых шаров среди вынутых. Нетрудно построить закон распределения ξ (читатель может проделать выкладки самостоятельно):

$$\xi = \begin{pmatrix} 0 & 1 & 2 \\ 0{,}3 & 0{,}6 & 0{,}1 \end{pmatrix}.$$

Математическое ожидание $M\xi = 0 \cdot 0{,}3 + 1 \cdot 0{,}6 + 2 \cdot 0{,}1 = 0{,}8$. Посчитаем дисперсию по определению:

$$D\xi = (0 - 0{,}8)^2 \cdot 0{,}3 + (1 - 0{,}8)^2 \cdot 0{,}6 + (2 - 0{,}8)^2 \cdot 0{,}1 =$$

$$= 0{,}64 \cdot 0{,}3 + 0{,}04 \cdot 0{,}6 + 1{,}44 \cdot 0{,}1 = 0{,}36.$$

Теперь вычислим дисперсию по формуле. Случайная величина ξ^2 имеет распределение

$$\xi^2 = \begin{pmatrix} 0 & 1 & 4 \\ 0{,}3 & 0{,}6 & 0{,}1 \end{pmatrix}$$

и $M\xi^2 = 1$. Значит,

$$D\xi = 1 - 0{,}8^2 = 0{,}36.$$

Результаты, разумеется, совпали.

2. Воспользуемся формулой для вычисления дисперсии биномиальной случайной величины $B_{n,p}$, которая представляет собой сумму n независимых бернуллиевых величин

$$\xi_i = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \quad i = 1, 2, \ldots, n.$$

Значит, $DB_{n,p} = D\xi_1 + D\xi_2 + \ldots + D\xi_n$. Ясно, что $\xi^2 = \xi$ и $M\xi = M\xi^2 = p$. Тогда дисперсия каждой бернуллиевой величины $D\xi_i = p - p^2 = p(1 - p) = pq$, и таких слагаемых в сумме n штук. Таким образом,

$$\boxed{DB_{n,p} = npq}$$

Центрирование случайной величины не меняет ее дисперсии. Это следует из того, что математическое ожидание центрированной величины равно нулю:

$$D\overset{\circ}{\xi} = M(\xi - M\xi - 0)^2 = D\xi.$$

Случайная величина

$$\xi^* = \frac{\overset{\circ}{\xi}}{\sqrt{D\xi}} = \frac{\xi - M\xi}{\sqrt{D\xi}}$$

называется *стандартизованной* (по отношению к ξ) или просто *стандартизацией* ξ. Стандартизованная случайная величина имеет нулевое математическое ожидание и единичную дисперсию:

$$M\xi^* = \frac{1}{\sqrt{D\xi}} M(\xi - M\xi) = 0,$$

$$D\xi^* = \frac{1}{D\xi} D(\xi - M\xi) = \frac{1}{D\xi} D\xi = 1.$$

Особенность стандартизованной величины в том, что она не имеет размерности. Делением на СКО мы приводим случайную величину к безразмерному виду. Часто это бывает удобно.

12.3.3. Ковариация и коэффициент корреляции

Ранее мы определили характеристики положения и разброса конечной случайной величины, теперь введем величину, характеризующую связь, з а в и с и м о с т ь между двумя случайными величинами. ***Ковариацией*** двух случайных величин ξ и η (или ковариацией между ξ и η) называется число

$$\mathrm{cov}(\xi, \eta) = M(\overset{\circ}{\xi}\,\overset{\circ}{\eta}) = M[(\xi - M\xi)(\eta - M\eta)].$$

Из определения вытекают следующие простые свойства ковариации.

1°. $\operatorname{cov}(\xi, \eta) = M(\xi\eta) - M\xi M\eta$.

2°. *Ковариация коммутативна*:

$$\operatorname{cov}(\xi, \eta) = \operatorname{cov}(\eta, \xi).$$

3°. $\operatorname{cov}(\xi, \xi) = D\xi$.

4°. $D(\xi + \eta) = D\xi + D\eta + 2\operatorname{cov}(\xi, \eta)$.

Следующее свойство важно при оценке степени зависимости двух случайных величин.

5°. *Если случайные величины ξ и η независимы, то их ковариация равна нулю.*

Для независимых величин ξ и η их центрированные величины также независимы. Поэтому

$$\operatorname{cov}(\xi, \eta) = M\mathring{\xi}M\mathring{\eta} = 0,$$

так как математическое ожидание центрированной случайной величины равно нулю.

Этому свойству можно придать более практическую формулировку: *если* $\operatorname{cov}(\xi, \eta) \neq 0$, *то случайные величины* ξ и η *зависимы.*

Итак, ненулевая ковариация свидетельствует о зависимости случайных величин. Обратное утверждение неверно. Можно привести примеры как независимых, так и зависимых случайных величин, ковариация между которыми равна нулю.

Размерность ковариации равна произведению размерностей случайных величин ξ и η. Чтобы получить безразмерную характеристику зависимости, ковариацию делят на произведение СКО ξ и η. Полученное число называется ***коэффициентом корреляции*** между случайными величинами ξ и η:

$$r_{\xi,\eta} = \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{D\xi}\sqrt{D\eta}} = \frac{M[(\xi - M\xi)(\eta - M\eta)]}{\sqrt{D\xi}\sqrt{D\eta}}.$$

Разумеется, предполагается, что случайные величины ξ и η имеют ненулевые дисперсии. Справедливы следующие свойства коэффициента корреляции.

1°. $r_{\xi,\eta} = M(\xi^*\eta^*)$.

Это видно из определения.

2°. *Коэффициенты корреляции между ξ и η и между их стандартизациями совпадают*:

$$r_{\xi,\eta} = r_{\xi^*,\eta^*}.$$

В самом деле, поскольку $M\xi^* = M\eta^* = 0$, $D\xi^* = D\eta^* = 1$, коэффициент корреляции

$$r_{\xi^*,\eta^*} = \frac{M[(\xi^* - M\xi^*)(\eta^* - M\eta^*)]}{\sqrt{D\xi^*}\sqrt{D\eta^*}} = M(\xi^*\eta^*) = r_{\xi,\eta}.$$

3°. $|r_{\xi,\eta}| \leqslant 1$.

Для доказательства введем вспомогательную случайную величину $\zeta = ((\xi - M\xi) + \lambda(\eta - M\eta))^2$, где λ — некоторое действительное число. Случайная величина ζ принимает только неотрицательные значения. Поэтому $M\zeta \geqslant 0$ при любом λ. Распишем это неравенство подробнее:

$$M\zeta = M[(\xi - M\xi) + \lambda(\eta - M\eta)]^2 =$$
$$= M(\xi - M\xi)^2 + 2\lambda M[(\xi - M\xi)(\eta - M\eta)] + \lambda^2 M(\eta - M\eta)^2 =$$
$$= D\eta\lambda^2 + 2\operatorname{cov}(\xi, \eta)\lambda + D\xi \geqslant 0.$$

Полученное квадратное неравенство имеет вид $a\lambda^2 + 2b\lambda + c \geqslant 0$ и выполняется при всех значениях λ. При этом старший коэффициент $a = D\eta > 0$. Вспоминая школьный курс алгебры, делаем вывод, что квадратный трехчлен $a\lambda^2 + 2b\lambda + c \geqslant 0$ имеет неположительный дискриминант:

$$b^2 - ac \leqslant 0.$$

Подставим значения коэффициентов:

$$(\operatorname{cov}(\xi, \eta))^2 - D\xi D\eta \leqslant 0,$$

откуда

$$\frac{(\operatorname{cov}(\xi, \eta))^2}{D\xi D\eta} = \left(\frac{\operatorname{cov}(\xi, \eta)}{\sqrt{D\xi}\sqrt{D\eta}}\right)^2 = r_{\xi,\eta}^2 \leqslant 1.$$

Извлекая квадратный корень из последнего неравенства, получаем требуемое свойство: $|r_{\xi,\eta}| \leqslant 1$.

Следующие два свойства касаются независимости случайных величин. Первое из них прямо следует из соответствующего свойства ковариации.

4°. *Если ξ и η независимы, то $r_{\xi,\eta} = 0$ (если $r_{\xi,\eta} \neq 0$, то ξ и η зависимы).*

5°. *Коэффициент корреляции равен* ± 1 *тогда и только тогда, когда случайные величины линейно зависимы:*

$$|r_{\xi,\eta}| = 1 \Leftrightarrow \eta = a\xi + b.$$

Доказательство.

1. Пусть $r_{\xi,\eta} = 1$. Поскольку $M\xi^* = M\eta^* = 0$, справедливо

$$D(\xi^* - \eta^*) = M(\xi^* - \eta^*)^2.$$

Продолжим равенство, раскрывая скобки:

$$M(\xi^* - \eta^*)^2 = M\xi^{*2} + M\eta^{*2} - 2M(\xi^*\eta^*) = \\ = D\xi^* + D\eta^* - 2r_{\xi,\eta} = 1 + 1 - 2 = 0$$

(последнее равенство получено с учетом того, что $D\xi^* = D\eta^* = 1$ и $r_{\xi,\eta} = 1$). Итак, $D(\xi^* - \eta^*) = 0$ и по свойству дисперсии получаем, что $\xi^* - \eta^* = 0$. Запишем это равенство подробнее:

$$\frac{\xi - M\xi}{\sqrt{D\xi}} - \frac{\eta - M\eta}{\sqrt{D\eta}} = 0,$$

откуда $\eta = a\xi + b$, где

$$a = \frac{\sqrt{D\eta}}{\sqrt{D\xi}}, \quad b = M\eta - \frac{\sqrt{D\eta}}{\sqrt{D\xi}} M\xi.$$

2. Случай $r_{\xi,\eta} = -1$ разбирается аналогично, только вместо $D(\xi^* - \eta^*)$ вычисляется $D(\xi^* + \eta^*)$. При этом получается

$$a = -\frac{\sqrt{D\eta}}{\sqrt{D\xi}}, \quad b = M\eta + \frac{\sqrt{D\eta}}{\sqrt{D\xi}} M\xi.$$

Далее, если $\eta = a\xi + b$, то $M\eta = aM\xi + b$, $D\eta = a^2 D\xi$. Поэтому

$$r_{\xi,\eta} = \frac{M[(\xi - M\xi)(\eta - M\eta)]}{\sqrt{D\xi}\sqrt{D\eta}} = \frac{M[(\xi - M\xi)(a\xi - aM\xi)]}{|a| D\xi} =$$

$$= \frac{aM(\xi - M\xi)^2}{|a| D\xi} = \frac{aD\xi}{|a| D\xi} = \frac{a}{|a|} = \pm 1.$$

Утверждение доказано.

Посчитаем коэффициент корреляции между случайными величинами ξ и η из примера, в котором урна содержит по одному белому и черному шару (ξ – количество вынутых белых шаров,

η – количество оставшихся белых). Эти величины распределены по законам

$$\xi = \begin{pmatrix} 0 & 1 \\ ½ & ½ \end{pmatrix}, \quad \eta = \begin{pmatrix} 0 & 1 \\ ½ & ½ \end{pmatrix}.$$

Легко подсчитать, что $M\xi = M\eta = {}^1/_2$, $D\xi = D\eta = {}^1/_4$. Значит,

$$\xi^* = \begin{pmatrix} -1 & 1 \\ ½ & ½ \end{pmatrix}, \quad \eta^* = \begin{pmatrix} -1 & 1 \\ ½ & ½ \end{pmatrix}$$

и

$$\xi^*\eta^* = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

есть постоянная. Ее математическое ожидание, естественно, равно -1. Значит,

$$r_{\xi,\eta} = M(\xi^*\eta^*) = -1,$$

откуда следует линейная зависимость ξ и η. Однако по смыслу задачи $\eta = 1 - \xi$, т.е. $\eta = a\xi + b$ при $a = -1$, $b = 1$. Действительно, $\sqrt{D\xi} = \sqrt{D\eta} = {}^1/_2$, т.е. $a = -1$, а $b = {}^1/_2 + 1 \cdot {}^1/_2 = 1$.

Итак, ковариация и коэффициент корреляции позволяют довольно хорошо (хотя и не в полной мере) охарактеризовать степень зависимости между случайными величинами:

- если они независимы, то $\mathrm{cov}(\xi, \eta) = 0$ $(r_{\xi,\eta} = 0)$;
- если $\mathrm{cov}(\xi, \eta) \neq 0$ $(r_{\xi,\eta} \neq 0)$, то они зависимы;
- если $r_{\xi,\eta} = \pm 1$, то они линейно зависимы.

12.4. Функция распределения

Как вы помните, при определении случайной величины ξ особая важность придавалась тому, чтобы можно было посчитать вероятность события $\{\xi < x\}$ для любого числа x. Эта вероятность играет большую роль при описании случайных величин. Меняя x, мы будем получать различные значения вероятности $P(\xi < x)$, т.е. данная вероятность есть функция от x. Функция действительной переменной

$$F_\xi(x) = P(\xi < x)$$

называется ***функцией распределения*** случайной величины ξ. Построим функцию распределения нескольких случайных величин, для которых мы уже выясняли строение событий $\{\xi < x\}$.

1. Бросание монеты. Пусть случайная величина ξ есть число «орлов», выпавшее при одном бросании:

$$\xi = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

При $x \leqslant 0$ событие $\{\xi < x\}$ невозможно; при $0 < x \leqslant 1$ это событие состоит в том, что выпала «решка»; при $x > 1$ данное событие достоверно: что бы ни выпало, значение ξ равно 0 или 1 и заведомо меньше x. Итак,

$$F_\xi(x) = P(\xi < x) = \begin{cases} P(\varnothing) & = & 0, & x \leqslant 0, \\ P(\{\text{P}\}) & = & \frac{1}{2}, & 0 < x \leqslant 1, \\ P(\Omega) & = & 1, & x > 1 \end{cases}$$

(рис. 12.4).

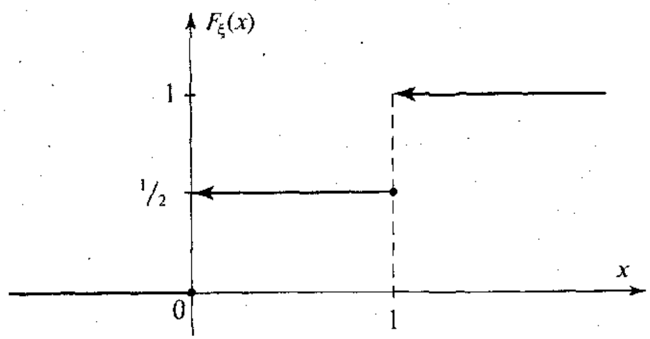


Рис. 12.4

Процедура построения функции распределения в этом примере подсказывает, как строится функция в общем случае для конечной случайной величины:

$$\xi = \begin{pmatrix} x_i \\ p_i \end{pmatrix}, \quad i = 1, 2, \dots, m.$$

Предположим, что значения случайной величины расположены по возрастанию: $x_1 < x_2 < \dots < x_m$. Тогда понятно, что при x событие $\{\xi < x\}$ невозможно: нет значений ξ, меньших x. На пр жутке $x_1 < x \leqslant x_2$ данное событие состоит в том, что ξ приним значение x_1 и соответственно вероятность такого события равна На следующем промежутке $x_2 < x \leqslant x_3$ этому событию благопри ствуют два значения ξ:

$$\{\xi < x\} = \{\xi = x_1\} + \{\xi = x_2\}$$

и вероятность равна $p_1 + p_2$. Вообще, при $x_i < x \leqslant x_{i+1}$

$$\{\xi < x\} = \{\xi = x_1\} + \{\xi = x_2\} + \dots + \{\xi = x_i\}$$

и $P(\xi < x) = p_1 + p_2 + \dots + p_i$. При $x > x_m$ событие становится достоверным: какое бы значение ни приняла случайная величина, оно заведомо меньше x. Итак,

$$F_\xi(x) = P(\xi < x) = \begin{cases} 0, & x \leqslant x_1, \\ p_1, & x_1 < x \leqslant x_2, \\ p_1 + p_2, & x_2 < x \leqslant x_3, \\ \dots\dots\dots\dots & \\ p_1 + p_2 + \dots + p_i, & x_i < x \leqslant x_{i+1}, \\ \dots\dots\dots\dots & \\ p_1 + p_2 + \dots + p_{m-1}, & x_{m-1} < x \leqslant x_m, \\ p_1 + p_2 + \dots + p_m = 1, & x > x_m. \end{cases}$$

Короче это можно записать так:

$$F_\xi(x) = \sum_{x_i < x} p_i.$$

Теперь ясно, как строить график функции распределения конечной случайной величины. Функция распределения является ступенчатой: от $-\infty$ до x_1 включительно функция равна нулю, в точке x_1 происходит скачок на величину p_1 и функция остается постоянной до x_2 включительно; затем – следующий скачок на p_2: функция приняла значение $p_1 + p_2$ и остается постоянной до x_3 и т.д. Последний скачок на p_m происходит в точке x_m, и функция равна единице от x_m до $+\infty$ (рис. 12.5).

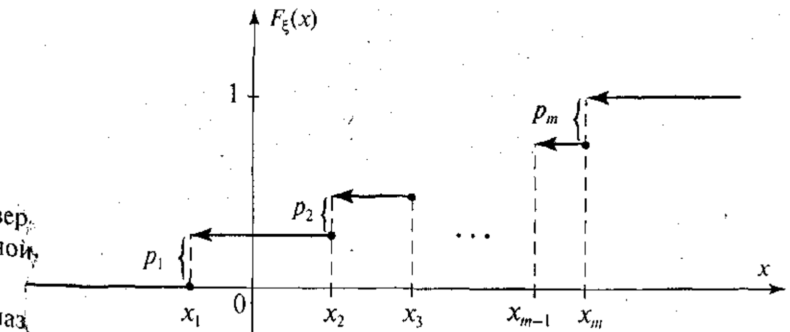


Рис. 12.5

2. Бросание точки. Точку наугад бросают на отрезок $[0, 1]$. Случайная величина ξ – координата точки попадания. При $x \leqslant 0$ событие $\{\xi < x\}$ невозможно; при $0 < x \leqslant 1$ данное событие состоит в том, что точка попала на промежуток $[0, x)$; при $x > 1$ – это достоверное событие: куда бы ни попала точка, ее координата непременно меньше x. Таким образом,

$$F_\xi(x) = P(\xi < x) = \begin{cases} P(\varnothing) & = 0, & x \leqslant 0, \\ P(\xi \in [0, x)) & = x, & 0 < x \leqslant 1, \\ P(\Omega) & = 1, & x > 1. \end{cases}$$

(Вероятность попадания на промежуток $[0, 1)$ легко вычисляется по формуле геометрической вероятности: длина всего отрезка равна 1, а длина $[0, 1)$ равна x; их отношение равно x.) График данной функции распределения представлен на рис. 12.6.

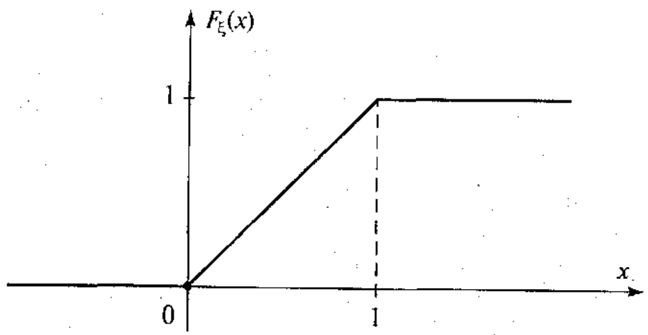


Рис. 12.6

Из определения функции распределения выводятся следующие равенства.

∇1.
$$P(\xi \geqslant x) = 1 - F_\xi(x),$$
$$P(a \leqslant x < b) = F_\xi(b) - F_\xi(a).$$

Первое равенство справедливо, так как событие $\{\xi \geqslant x\}$ противоположно событию $\{\xi < x\}$. Второе получается из того факта, что при $a < b$ событие $\{\xi < b\}$ есть сумма несовместных событий $\{\xi < a\}$ и $\{a \leqslant x < b\}$. Поэтому

$$P(\xi < b) = P(\xi < a) + P(a \leqslant x < b),$$

откуда

$$P(a \leqslant x < b) = P(\xi < b) - P(\xi < a) = F_\xi(b) - F_\xi(a).$$

Приведенные примеры хорошо иллюстрируют основные свойства функции распределения.

1°. *При любом x выполняется неравенство* $0 \leqslant F_\xi(x) \leqslant 1$.

Это справедливо, поскольку любое значение функции распределения есть вероятность.

2°. *Функция распределения есть неубывающая функция.*

В самом деле, пусть $x_1 \leqslant x_2$. Тогда из события $\{\xi < x_1\}$ следует событие $\{\xi < x_2\}$. Поэтому $P(\xi < x_1) \leqslant P(\xi < x_2)$ и, значит, $F_\xi(x_1) \leqslant F_\xi(x_2)$.

3°.
$$F_\xi(-\infty) = \lim_{x \to -\infty} F_\xi(x) = 0,$$
$$F_\xi(+\infty) = \lim_{x \to +\infty} F_\xi(x) = 1.$$

При $x \to -\infty$ событие $\{\xi < x\}$ стремится к невозможному и вероятность соответственно стремится к нулю. При стремлении x к $+\infty$ событие становится достоверным.

4°. *Функция распределения непрерывна слева, т.е.*

$$\lim_{x \to x_0 + 0} F_\xi(x) = F_\xi(x_0).$$

Примем это свойство без доказательства. Приведенные примеры подтверждают его: в точках разрыва функция непрерывна слева.

12.5. Непрерывные случайные величины

Начнем со знакомого примера: точку наугад бросают на отрезок $[0, 1]$, случайная величина ξ – координата точки попадания. Пусть $x_0 \in [0, 1]$ – произвольная точка из отрезка. Зададимся на первый взгляд странным вопросом: какова вероятность того, что при бросании мы попадем в точку x_0? Другими словами: какова вероятность того, что случайная величина ξ примет значение x_0? Будем рассуждать с точки зрения геометрической вероятности. Вероятность попадания точки в отрезок $[x_1, x_2] \subseteq [0, 1]$ пропорциональна длине этого отрезка. Поскольку длина отрезка $[0, 1]$ равна единице,

$$P(\xi \in [x_1, x_2]) = x_2 - x_1.$$

Значит, вероятность попасть в отрезок $[x_0 - \varepsilon, x_0 + \varepsilon]$ равна 2ε. При уменьшении ε событие $\{\xi \in [x_0 - \varepsilon, x_0 + \varepsilon]\}$ становится все ближе к событию $\{\xi = x_0\}$. В пределе они совпадают. Таким образом,

$$P(\xi = x_0) = \lim_{\varepsilon \to 0} P(\xi \in [x_0 - \varepsilon, x_0 + \varepsilon]) = \lim_{\varepsilon \to 0} 2\varepsilon = 0.$$

Мы пришли к парадоксальному выводу.

Вероятность попадания в конкретную точку x_0 отрезка равна нулю:

$$P(\xi = x_0) = 0.$$

Нам известно, что конечная случайная величина может принимать только конечное число значений с ненулевой вероятностью. В рассмотренном примере мы столкнулись с совершенно другим поведением случайной величины: она принимает бесконечное число значений (это все числа из отрезка $[0, 1]$) и при этом вероятность принять каждое конкретное значение равна нулю. Этим свойством обладает любая случайная величина, распределенная на целом промежутке, т.е. принимающая все значения из отрезка, интервала или на всей числовой оси. Ясно, что мы имеем дело со случайной величиной другого типа, а именно с непрерывной случайной величиной. Далее мы дадим ей строгое определение.

Мы описывали конечную случайную величину с помощью закона распределения в виде таблицы. Это было возможно именно потому, что число значений конечно и каждое принимается с ненулевой вероятностью. Как же описать случайную величину, для которой вероятность принять данное значение равна нулю? Таблицу в этом случае построить невозможно.

Ситуация имеет прозрачную физическую аналогию. Предположим, отрезок $[a, b]$ «изготовлен» из какого-то материала и мы хотим измерить его м а с с у. Масса отдельной точки, разумеется, равна нулю. Однако, если задана п л о т н о с т ь ρ вещества, из которого «сделан» отрезок, вычисление массы не составит труда, для этого нужно просто умножить плотность на длину отрезка: $m = \rho(b - a)$. В общем случае плотность является переменной величиной и задается функцией $\rho(x)$. Применяя стандартные рассуждения интегрального исчисления, получаем простую формулу для массы:

$$m = \int_a^b \rho(x)\,\mathrm{d}x.$$

Итак, масса отрезка равна интегралу Римана от плотности.

Не поступить ли нам аналогичным образом и в случае вероятности попадания в отрезок? Так как вероятность попасть в каждую точку равна нулю, введем функцию п л о т н о с т и в е р о я т н о с т и. Тогда вероятность попадания в отрезок можно задать через интеграл от плотности.

Случайная величина ξ называется ***непрерывной случайной величиной***, если существует функция $f_\xi(x)$, такая, что

$$\boxed{P(\xi \in [a, b]) = \int_a^b f_\xi(x)\,\mathrm{d}x}$$

Функция $f_\xi(x)$ называется *плотностью вероятности* или *плотностью распределения* случайной величины ξ. Предполагается, что плотность распределения непрерывна всюду, за исключением, быть может, конечного числа точек (в которых имеются разрывы I рода). Если такая функция нашлась, можно считать, что случайная величина полностью определена: ведь мы можем вычислить вероятность ее попадания в любой промежуток числовой оси. При этом автоматически получается, что

$$P(\xi = x_0) = \int_{x_0}^{x_0} f_\xi(x)\,\mathrm{d}x = 0.$$

Вспомнив геометрический смысл интеграла Римана (см. подпараграф 10.3.1), можно сделать вывод, которым мы будем активно пользоваться.

∇1. ***Вероятность того, что непрерывная случайная величина принимает значения из отрезка $[a, b]$, равна площади под графиком плотности вероятности на этом отрезке.***

Подход, основанный на плотности вероятности, позволяет считать, что любая непрерывная случайная величина распределена на всей числовой оси: если она не принимает значений в некотором промежутке (вероятность попадания в него равна нулю), это означает, что плотность вероятности на этом промежутке — нулевая. В примере с бросанием точки на отрезок $[0, 1]$ легко убедиться, что плотность вероятности имеет вид

$$f_\xi(x) = \begin{cases} 0, & x \notin [0, 1], \\ 1, & x \in [0, 1] \end{cases}$$

(рис. 12.7).

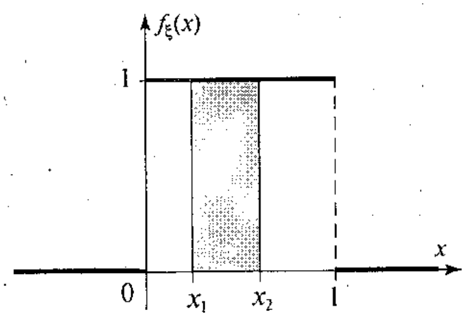


Рис. 12.7

В самом деле, если $[x_1, x_2] \subseteq [0, 1]$, то

$$P(x_1 \leqslant \xi \leqslant x_2) = \int_{x_1}^{x_2} f_\xi(x)\mathrm{d}x = \int_{x_1}^{x_2} \mathrm{d}x = x_2 - x_1,$$

а вероятность попасть в любой промежуток вне отрезка $[0, 1]$ равна нулю. Видно, что $f_\xi(x)$ имеет разрывы I рода в точках $x = 0$ и $x = b$. Данная случайная величина представляет собой частный случай равномерной случайной величины. Далее мы изучим такие случайные величины подробнее.

Из определения легко вывести следующие равенства.

∇2. *Для любой непрерывной случайной величины*

$$P(a \leqslant \xi \leqslant b) = P(a < \xi \leqslant b) = P(a \leqslant \xi < b) = P(a < \xi < b) =$$

$$= \int_a^b f_\xi(x)\mathrm{d}x.$$

Событие $\{a \leqslant \xi \leqslant b\}$ есть сумма несовместных событий $\{\xi = a\}$ и $\{a < \xi \leqslant b\}$. Поэтому $P(a \leqslant \xi \leqslant b) = P(\xi = a) + P(a < \xi \leqslant b) = 0 + P(a < \xi \leqslant b)$. Точно так же проверяются и остальные равенства.

∇3. *Функция распределения непрерывной случайной величины имеет вид*

$$\boxed{F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)\mathrm{d}t}$$

Вспомним, что функция распределения есть по определению вероятность события $\{\xi < x\}$. Значит,

$$F_\xi(x) = P(\xi < x) = P(-\infty < \xi < x) = \int_{-\infty}^{x} f_\xi(t)\mathrm{d}t.$$

Здесь переменная интегрирования обозначена через t, чтобы не путать ее с аргументом функции распределения. Иногда данное равенство принимают за определение непрерывной случайной величины: случайную величину называют *непрерывной*, если ее функция распределения выражается таким образом через функцию $f_\xi(x)$.

Обратим внимание на то, что для описания функции распределения непрерывной случайной величины используется несобственный интеграл, который мы изучили в курсе математического анализа.

Функция плотности распределения обладает следующими свойствами.

1°. *Функция $f_\xi(x)$ неотрицательна при всех x.*

2° (условие нормировки). *Справедливо равенство*

$$\boxed{\int_{-\infty}^{+\infty} f_\xi(x)\mathrm{d}x = 1}$$

Событие $\{-\infty < \xi < +\infty\} = \{\xi \in \boldsymbol{R}\}$ достоверно, т.е. его вероятность равна 1. Однако

$$P(-\infty < \xi < +\infty) = \int_{-\infty}^{+\infty} f_\xi(x)\mathrm{d}x,$$

откуда и следует условие нормировки. Другими словами: *площадь под графиком функции плотности распределения на всей числовой оси равна единице.*

Пример (равномерное распределение).

Говорят, что непрерывная случайная величина имеет *равномерное распределение* на отрезке $[a, b]$ (*равномерно распределена* на отрезке), если ее плотность распределения имеет вид

$$f_\xi(x) = \begin{cases} 0, & x \notin [a, b], \\ c = \text{const}, & x \in [a, b]. \end{cases}$$

Сама случайная величина в этом случае называется *равномерной* или *равномерно распределенной* на отрезке $[a, b]$. Плотность распределения равномерной случайной величины постоянна на отрезке $[a, b]$ и равна нулю вне этого отрезка.

Из условия нормировки легко найти значение, которое принимает плотность равномерного распределения при $x \in [a, b]$. Полная площадь под графиком равна площади прямоугольника с высотой c и основанием $b - a$: $S = c(b - a) = 1$, откуда

$$c = \frac{1}{b - a}$$

(рис. 12.8).

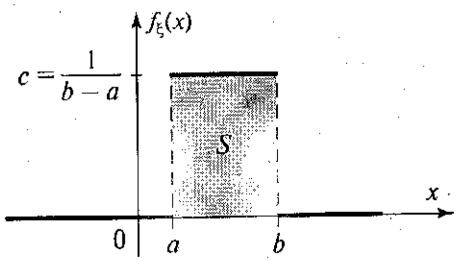


Рис. 12.8

Интегрируя плотность распределения, получаем функцию распределения:

если $x \leqslant a$, то $F_\xi(x) = 0$ (под интегралом нулевая функция);

если $a < x \leqslant b$, то

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)\,\mathrm{d}t = \int_{a}^{x} \frac{\mathrm{d}t}{b-a} = \frac{1}{b-a}t\Big|_a^x = \frac{x-a}{b-a}$$

(на графике это отрезок прямой, соединяющий точки $(a, 0)$ и $(b, 1)$);

если $x > b$, то $F_\xi(x) = 1$ (рис. 12.9).

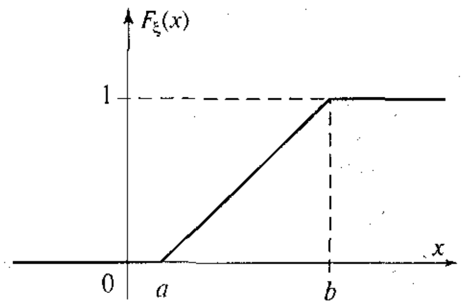


Рис. 12.9

Интегрируя плотность вероятности, получаем

$$P(x_1 \leqslant \xi \leqslant x_2) = \int_{x_1}^{x_2} \frac{\mathrm{d}x}{b-a} = \frac{x_2 - x_1}{b-a}.$$

Значит, равномерную случайную величину можно интерпретировать как координату точки, брошенной наугад на отрезок $[a, b]$ таким образом, что вероятность попадания в отрезок $[x_1, x_2] \subseteq [a, b]$ пропорциональна его длине. Мы уже рассматривали частный случай такой случайной величины при $a = 0$, $b = 1$.

3°. *В точках непрерывности плотность вероятности равна производной функции распределения:*

$$F'_\xi(x) = f_\xi(x).$$

Это равенство выводится из того, что функция распределения есть интеграл от плотности. Часто этот факт также принимают за определение непрерывной случайной величины: случайную величину называют *непрерывной*, если ее функция распределения имеет производную всюду, за исключением, быть может, конечного числа точек. Эту производную и называют плотностью распределения.

Для равномерного распределения $F'_\xi(x) = 0$ при $x < a$ и $x > b$; при $a < x < b$

$$F'_\xi(x) = \left(\frac{x-a}{b-a}\right)' = \frac{1}{b-a},$$

т.е. действительно $F'_\xi(x) = f_\xi(x)$, $x \neq a$, $x \neq b$. В точках $x = a$ и $x = b$ функция распределения не имеет производной.

Для непрерывных случайных величин, так же как для конечных, можно построить совместное распределение, задавая вероятности событий $\{\xi \in [a, b], \eta \in [c, d]\}$ для различных отрезков $[a, b]$ и $[c, d]$. Кроме того, можно определить действия над непрерывными случайными величинами. С математической точки зрения это более сложная процедура, чем для конечных величин. Поэтому не будем углубляться в математические тонкости. Скажем только, что две непрерывные случайные величины ξ и η называются *независимыми*, если события $\{a \leqslant \xi \leqslant b\}$ и $\{c \leqslant \eta \leqslant d\}$ независимы, какие бы отрезки $[a, b]$ и $[c, d]$ мы ни взяли.

12.6. Числовые характеристики непрерывных случайных величин

12.6.1. Математическое ожидание

Математическое ожидание непрерывной случайной величины определяется по аналогии с математическим ожиданием конечной случайной величины. Пусть ξ – непрерывная случайная величина с плотностью распределения $f_\xi(x)$. Выделим на числовой оси отрезок $[-N, N]$ и разобьем его на n равных частей (рис. 12.10).

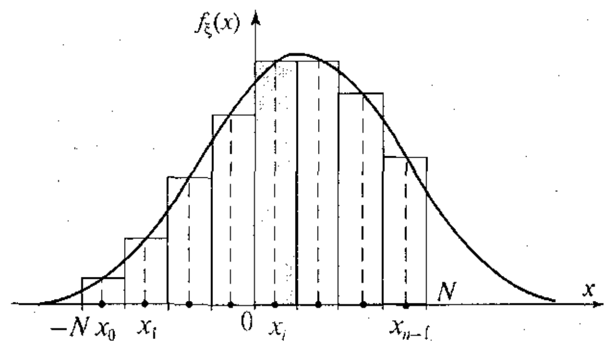


Рис. 12.10

Возьмем в каждом промежутке произвольную точку x_i. Можно считать, что промежутки разбиения настолько малы, что плотность вероятности на них примерно постоянна и равна $f_\xi(x_i)$. Тогда вероятность попасть в i-й промежуток примерно равна $p_i = f_\xi(x_i)\Delta x_i$, где Δx_i – длина этого промежутка. Среднее значение ξ на отрезке $[-N, N]$ примерно равно

$$\sum_{i=0}^{n-1} x_i p_i = \sum_{i=0}^{n-1} x_i f_\xi(x_i)\Delta x_i,$$

и равенство будет тем точнее, чем меньше промежутки разбиения. Поэтому необходимо перейти к пределу при $n \to \infty$. Кроме того, непрерывная случайная величина распределена на всей числовой оси. Значит, и число N нужно устремить к бесконечности. Итак, среднее значение непрерывной случайной величины равно

$$\lim_{N\to\infty}\lim_{n\to\infty}\sum_{i=0}^{n-1} x_i f_\xi(x_i)\Delta x_i.$$

Приглядевшись внимательнее, замечаем, что построенный предел есть не что иное, как несобственный интеграл Римана от функции $xf_\xi(x)$ в пределах от $-\infty$ до $+\infty$. Среднее значение случайной величины есть ее математическое ожидание. Таким образом, ***математическим ожиданием*** непрерывной случайной величины называется число

$$\boxed{M\xi = \int_{-\infty}^{+\infty} xf_\xi(x)\,dx}$$

(если соответствующий интеграл существует).

Найдем по определению математическое ожидание равномерной случайной величины. Здравый смысл подсказывает, что среднее значение координаты точки, случайно брошенной на отрезок $[a, b]$, равно середине отрезка. Убедимся в этом:

$$M\xi = \int_{-\infty}^{+\infty} xf_\xi(x)\,dx = \int_a^b \frac{x\,dx}{b-a} = \frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b =$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2},$$

это и есть середина отрезка $[a, b]$.

Можно проверить, что все свойства математического ожидания, выведенные для конечных случайных величин, справедливы и для непрерывных случайных величин.

> 1°. $M(c\xi) = cM\xi$.
> 2°. $M(\xi \pm \eta) = M\xi \pm M\eta$.
> 3°. $M(\xi - M\xi) = 0$.
> 4°. *Для независимых случайных величин*
> $$M(\xi\eta) = M\xi M\eta.$$

Говорят, что распределение случайной величины *симметрично* (случайная величина *симметрична*), если график плотности распределения симметричен относительно математического ожидания (т.е. относительно прямой $x = M\xi$):

$$f_\xi(M\xi + x) = f_\xi(M\xi - x).$$

12.6.2. Дисперсия и среднеквадратичное отклонение

Дисперсия для непрерывной случайной величины определяется так же, как для конечной случайной величины:

$$D\xi = M(\xi - M\xi)^2$$

и по-прежнему характеризует разброс случайной величины относительно среднего значения. Дисперсия вычисляется через интеграл:

$$\boxed{D\xi = \int_{-\infty}^{+\infty} (x - M\xi)^2 f_\xi(x)\,dx}$$

(если интеграл существует).

Найдем дисперсию равномерной случайной величины:

$$D\xi = \int_{-\infty}^{+\infty} (x - M\xi)^2 f_\xi(x)\,dx = \frac{1}{b-a}\int_a^b \left(x - \frac{a+b}{2}\right)^2 dx =$$

$$= \frac{1}{b-a}\left.\frac{\left(x-\frac{a+b}{2}\right)^3}{3}\right|_a^b = \frac{(b-a)^3 - (a-b)^3}{3\cdot 8\cdot(b-a)} = \frac{2(b-a)^3}{24(b-a)} = \frac{(b-a)^2}{12}.$$

Итак, чем больше отрезок, на котором распределена равномерная случайная величина, тем больше ее дисперсия, т.е. тем больше ее разброс относительно среднего значения. Ничего неожиданного в этом результате нет.

Свойства дисперсии остаются неизменными.

1°. $D(c\xi) = c^2 D\xi$.

2°. $D(\xi + c) = D\xi$.

3°. *Для независимых случайных величин*

$$D(\xi + \eta) = D\xi + D\eta.$$

4°. $D\xi = M\xi^2 - (M\xi)^2$.

Посчитаем дисперсию равномерной случайной величины второй раз по формуле «средний квадрат минус квадрат среднего»:

$$M\xi^2 = \int_{-\infty}^{+\infty} x^2 f_\xi(x)\,dx = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a}\left.\frac{x^3}{3}\right|_a^b =$$

$$= \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

и

$$D\xi = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 =$$

$$= \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} = \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12}.$$

Результат, конечно, совпал с вычисленным по определению.

Квадратный корень из дисперсии называют среднеквадратичным отклонением (СКО). Непрерывную случайную величину можно стандартизовать:

$$\xi^* = \frac{\xi - M\xi}{\sqrt{D\xi}}.$$

Математическое ожидание стандартизованной величины равно нулю, а дисперсия – единице.

12.6.3. Квантили

Квантилью случайной величины ξ порядка p называется число x_p, такое, что вероятность события $\{\xi < x_p\}$ равна p. Разумеется, квантиль можно определить и для конечной, и для непрерывной случайной величины. Для непрерывных случайных величин квантили удобно интерпретировать с помощью функции распределения и плотности распределения. Для непрерывной величины квантиль x_p есть корень уравнения

$$F_\xi(x) = p.$$

Найдем квантили равномерной случайной величины (рис. 12.11):

$$\frac{x-a}{b-a} = p,$$

откуда $x_p = a + p(b - a)$.

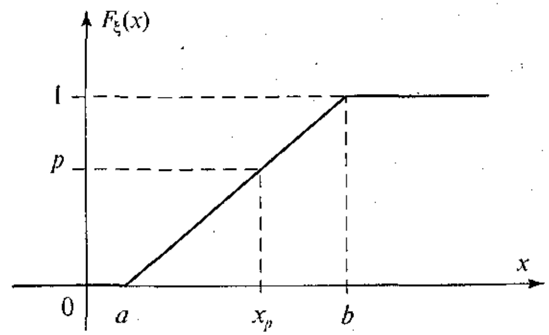


Рис. 12.11

Площадь под графиком плотности распределения левее некоторого x равна значению функции распределения $F_\xi(x)$. Поэтому площадь левее квантили x_p равна p (рис. 12.12). Эта наглядная интерпретация понятия квантили будет в дальнейшем активно использоваться.

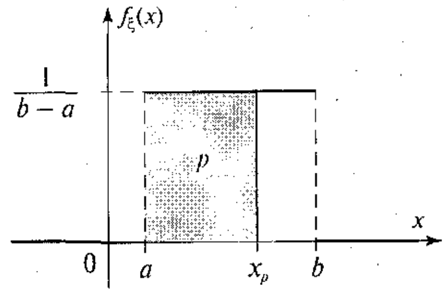


Рис. 12.12

Квантиль порядка 0,9 можно называть 90-процентной квантилью. Некоторые квантили имеют специальные названия. Квантиль $x_{1/2}$ порядка 0,5 (50-процентная квантиль) называется *медианой* распределения m_ξ. Площадь левее и правее медианы под графиком плотности распределения равна 0,5 (рис. 12.13). Для симметричных распределений медиана совпадает с математическим ожиданием: так как график симметричен относительно математического ожидания, то и площади левее и правее него равны. Для несимметричных распределений медиана, наряду с математическим ожиданием, характеризует среднее поведение, центральные тенденции случайной величины. В некоторых случаях медиана дает более точную оценку среднего поведения, чем математическое ожидание.

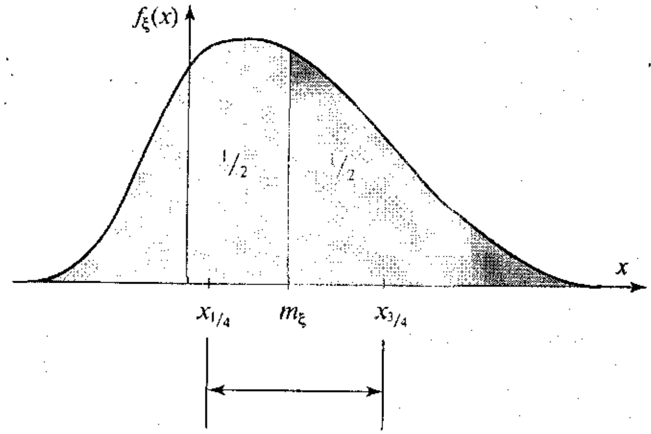


Рис. 12.13

Квантили порядка $^1/_4$ и $^3/_4$ называются соответственно *нижней* и *верхней квартилями*. Левее нижней квартили и правее верхней распределено по 25% значений случайной величины. Значит, между квартилями лежит половина всех значений (точнее, вероятность попасть в промежуток между квартилями равна 50%). Расстояние между квартилями $x_{3/4} - x_{1/4}$ (*межквартильный диапазон*) – дополнительная характеристика разброса случайной величины (см. рис. 12.13). Если этот диапазон невелик, то случайная величина не слишком сильно разбросана.

12.7. Нормальное распределение

12.7.1. Определение и свойства

В подпараграфе 12.7.1 вводится случайная величина, имеющая особое значение в теории вероятностей и математической статистике. Эта случайная величина – математическое чудо, наглядное под-

тверждение загадочного, почти мистического сближения математики с реальностью. Мы увидим, как довольно экзотический математический объект неожиданно становится мощным средством описания действительности.

Говорят, что случайная величина ξ распределена по ***нормальному закону*** (имеет ***нормальное распределение***) с параметрами m и σ ($\sigma > 0$), если она имеет плотность распределения

$$\boxed{f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}}$$

Нормальное распределение называют еще *гауссовым* по имени одного из его первооткрывателей Карла Фридриха Гаусса. Тот факт, что случайная величина имеет нормальное распределение с параметрами m, σ, обозначают так: $\xi \sim N(m, \sigma)$.

Карл Фридрих Гаусс (1777–1855). Великий немецкий математик, астроном, геодезист. Родился в Брауншвейге в семье поденщика. Брауншвейгский герцог обратил внимание на вундеркинда и позаботился о его обучении. В 1795–1798 гг. Гаусс учился в Геттингенском университете. С 1807 г. был профессором родного университета и одновременно директором университетской астрономической обсерватории. Его относительная обособленность, владение в равной мере прикладной и чистой математикой, занятия астрономией, использование латинского языка – на всем этом лежит отпечаток XVIII столетия. Однако в его трудах ощущался дух новой эпохи. Как и его современники Кант, Гете, Бетховен и Гегель, он стоял в стороне от больших политических битв, но в своей области самым энергичным образом выразил новые идеи своего века.

Творчество Гаусса было чрезвычайно разносторонним. Его исследования посвящены высшей алгебре, теории чисел, теории вероятностей, геодезии, небесной механике, теоретической астрономии, теории электричества и магнетизма. В 1801 г. он написал работу «Арифметические исследования», посвященную теории чисел и высшей алгебре. Некоторые результаты из этой работы стали классическими. Гаусс доказал основную теорему алгебры о том, что всякое алгебраическое уравнение с действительными коэффициентами имеет корень. Кроме того, он строго изложил теорию комплексных чисел и получил другие важные результаты в алгебре.

В 1820 г. ему было поручено произвести геодезическую съемку Ганновера. Для этого он разработал эффективные вычислительные методы (в том числе метод наименьших квадратов, широко применяющийся и поныне). Практически его исследования привели к созданию нового научного направления – высшей геодезии. С практикой геодезии связаны и его геометрические исследования. Его основная работа в этом направлении – «Общие изыскания о кривых поверхностях» (1827) – содержит много новых геометрических положений. Гаусс пришел к идее возможности неевклидовой геометрии, но не опубликовал этих результатов. Они были открыты заново русским математиком Н.И. Лобачевским и венгерским математиком Я. Больяи.

Гаусс занимался также астрономией. Ему удалось вычислить орбиту малой планеты Цереры. Решение этой сложной задачи принесло ученому известность, и он был приглашен заведовать кафедрой математики и астрономии. Результаты своих исследований по астрономии Гаусс объединил в фундаментальном труде «Теория движения небесных тел». Статуя в Геттингене изображает Гаусса и его младшего коллегу, физика Вильгельма Вебера, работающими над изобретением электрического телеграфа. Это относится к 1833–1834 гг., когда Гаусс начал интересоваться физикой. В этот период он выполнил большую экспериментальную работу по земному магнетизму.

Гаусс интересовался теорией вероятностей. В 1809 г. он заново открыл нормальное распределение (найденное впервые Муавром в 1730 г. и оставшееся незамеченным), которое называют его именем. Гаусс принадлежит к плеяде великих ученых-энциклопедистов, оставивших яркий след во многих областях знания.

На рис. 12.14 приведен график плотности нормального распределения (мы уже пользовались этим графиком для иллюстраций без упоминания названия распределения). График похож на рисунок, с помощью которого Маленький принц тестировал детей и взрослых. Как вы помните, взрослые считали, что на рисунке изображена шляпа, тогда как на самом деле это был удав, проглотивший слона. Итак, шляпа или удав – неважно, главное – хорошо запомнить, как выглядит этот график, потому что мы будем многократно им пользоваться.

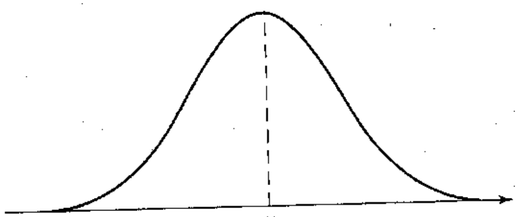


Рис. 12.14

Из определения вытекают основные особенности графика плотности нормального распределения:

- график симметричен относительно прямой $x = m$;
- функция достигает максимума в точке $x = m$;
- график приближается к нулю при возрастании $|x|$:

$$\lim_{x\to-\infty} f_\xi(x) = \lim_{x\to+\infty} f_\xi(x) = 0.$$

Вид графика подсказывает смысл параметров m и σ. Поскольку график симметричен относительно прямой $x = m$, можно предположить, что среднее значение (математическое ожидание) такой слу-

чайной величины равно m. Вычисление соответствующего интеграла подтверждает это предположение. Итак,

$$\boxed{M\xi = m}.$$

Кроме того, по определению точка m является также и модой, и медианой нормального распределения.

Смысл параметра σ станет понятным, если нарисовать графики плотности с различными значениями этого параметра (рис. 12.15).

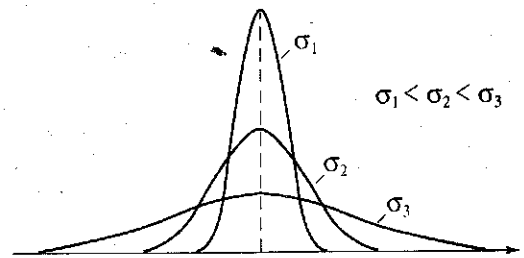


Рис. 12.15

Мы замечаем, что при меньших значениях σ график имеет более острую и узкую вершину и круче спадает к нулю (говорят, что распределение имеет короткие «хвосты»). В этом случае вероятность того, что случайная величина примет значения, далекие от m, мала (она задается площадью под «хвостами»). Наиболее вероятно попасть в малую окрестность математического ожидания. Значит, разброс случайной величины невелик. При увеличении σ, наоборот, график приобретает плоскую и широкую вершину, а «хвосты» растягиваются. Это говорит об увеличении разброса случайной величины. Таким образом, ясно, что параметр σ характеризует разброс нормальной случайной величины. Вычисления подтверждают, что σ есть ее СКО, т.е.

$$\boxed{D\xi = \sigma^2}$$

По плотности распределения строится функция распределения нормальной случайной величины

$$F_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt.$$

На рис. 12.16 изображены графики функции распределения при двух значениях параметра σ. Чем меньше σ, тем круче подъем графика.

Кроме того, из симметрии плотности распределения следует, что график функции нормального распределения симметричен относительно точки с координатами $(m, 0{,}5)$. Параметры m и σ полностью определяют нормально распределенную случайную величину. Нормальное распределение с параметрами $m = 0$, $\sigma = 1$ называется *стандартным* и задается плотностью

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

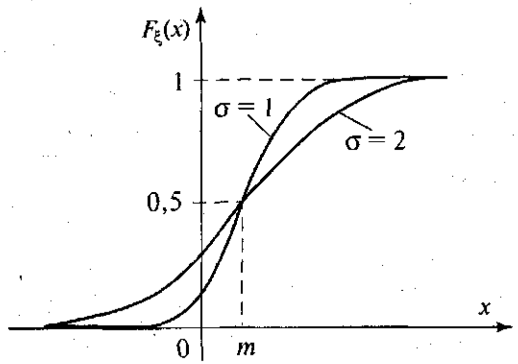


Рис. 12.16

Стандартную нормальную случайную величину иногда обозначают как Ξ^*. Функция распределения стандартной нормальной случайной величины обозначается через $\Phi(x)$ (рис. 12.17):

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

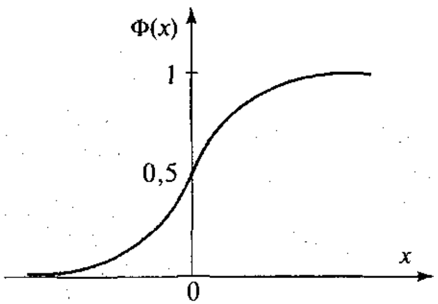


Рис. 12.17

Интеграл от плотности нормального распределения не берется в элементарных функциях. Поэтому функция $\Phi(x)$ задается с помощью таблиц. В любом учебнике по теории вероятностей или статистике приводятся таблицы этой функции. Подчеркнем еще раз геометрический смысл функции распределения: площадь под кривой плотности стандартного нормального распределения левее точки x равна $\Phi(x)$ (рис. 12.18).

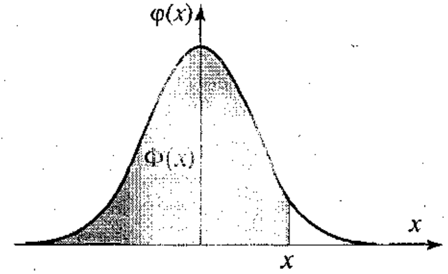


Рис. 12.18

∇**1.** *Если* $\xi \sim N(m, \sigma)$, *то* $\dfrac{\xi - m}{\sigma} \sim N(0, 1)$. *Другими словами, стандартная нормальная случайная величина есть стандартизация произвольной нормальной случайной величины:* $\xi^* = \Xi^*$.

Обозначим стандартизацию ξ через η:

$$\eta = \frac{\xi - m}{\sigma}.$$

Тогда

$$F_\eta(x) = P(\eta < x) = P\left(\frac{\xi - m}{\sigma} < x\right) = P(\xi < \sigma x + m) =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\sigma x + m} e^{-\frac{(t-m)^2}{2\sigma^2}}\,dt.$$

Сделаем замену переменной $z = \dfrac{t - m}{\sigma}$. Тогда $dz = dt/\sigma$ и полученный интеграл можно переписать через новую переменную:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\,dz = \Phi(x).$$

Итак, функция распределения случайной величины η (стандартизации нормальной случайной величины ξ) совпадает с $\Phi(x)$, т.е. с функцией распределения стандартной нормальной случайной величины. Тем самым свойство доказано.

Из доказательства вытекает связь между функциями распределения произвольной и стандартной нормальной величины:

$$F_\xi(x) = \Phi\left(\frac{x-m}{\sigma}\right).$$

∇2. $$\Phi(-x) = 1 - \Phi(x).$$

В самом деле, справедливо равенство

$$\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-t^2/2}\,dt = \frac{1}{\sqrt{2\pi}}\int\limits_{x}^{+\infty} e^{-t^2/2}\,dt,$$

которое легко доказывается заменой $z = -t$. Величина в левой части равенства есть по определению $\Phi(-x)$. Величина в правой части – площадь под кривой плотности стандартного нормального распределения от x до $+\infty$, т.е. вероятность события $\{\xi^* \geqslant x\}$ для стандартной нормальной случайной величины ξ^*. Это событие противоположно событию $\{\xi^* < x\}$. Поэтому $P(\xi^* \geqslant x) = 1 - P(\xi^* < x) = 1 - \Phi(x)$ (рис. 12.19). Равенство доказано.

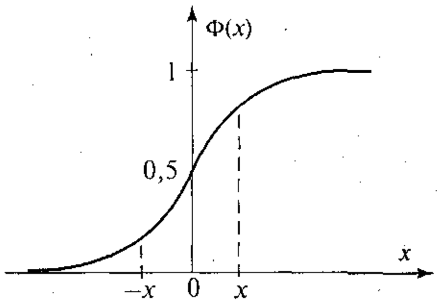


Рис. 12.19

Можно было рассуждать с геометрических позиций. Рассмотрим график плотности стандартного нормального распределения (рис. 12.20).

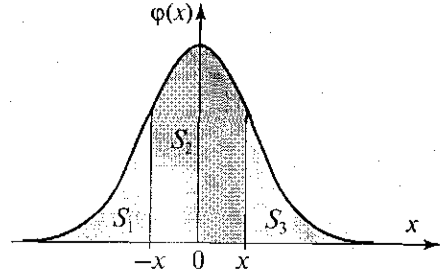


Рис. 12.20

Обозначим площадь под ним левее точки $-x$ через S_1; площадь между $-x$ и x – через S_2, оставшуюся площадь (правее x) – через S_3. Тогда, во-первых, из симметричности графика плотности следует, что $S_1 = S_3$. Во-вторых, из условия нормировки $S_1 + S_2 + S_3 = 1$ (вся площадь под графиком плотности равна единице). Значит, $S_1 + (S_1 + S_2) = 1$. Но по смыслу функции распределения $S_1 = \Phi(-x)$, $S_1 + S_2 = \Phi(x)$. Значит, $\Phi(-x) + \Phi(x) = 1$, откуда и следует доказываемое равенство.

Иногда вместо функции $\Phi(x)$ табулируется функция

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{0}^{x} e^{-t^2/2}\,dt,$$

равная площади под графиком плотности стандартного нормального распределения от 0 до x (рис. 12.21).

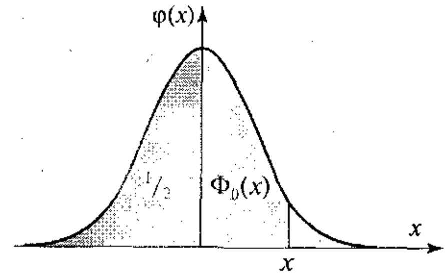


Рис. 12.21

В силу симметрии

$$\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{0} e^{-t^2/2}\,dt = \frac{1}{2}.$$

Поэтому между функциями $\Phi(x)$ и $\Phi_0(x)$ существует простая зависимость:

$$\Phi(x) = \frac{1}{2} + \Phi_0(x).$$

∇3. *Функция* $\Phi_0(x)$ *нечетна*:

$$\Phi_0(-x) = -\Phi_0(x)$$

(рис. 12.22).

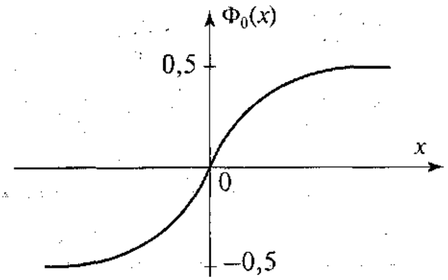


Рис. 12.22

В самом деле,

$$\Phi_0(-x) = \Phi(-x) - \frac{1}{2} = 1 - \Phi(x) - \frac{1}{2} =$$

$$= \frac{1}{2} - \left(\frac{1}{2} + \Phi_0(-x)\right) = -\Phi_0(x).$$

С помощью функций $\Phi(x)$ и $\Phi_0(x)$ легко вычислить вероятности попадания значений нормальной случайной величины в заданные промежутки. Пусть $\xi \sim N(m, \sigma)$. Тогда по свойству функции распределения $P(a \leqslant \xi \leqslant b) = F_\xi(b) - F_\xi(a)$ (рис. 12.23). Однако $F_\xi(x) = \Phi\left(\frac{x-m}{\sigma}\right) = \frac{1}{2} + \Phi_0\left(\frac{x-m}{\sigma}\right)$. Поэтому вероятность попадания в отрезок

$$\boxed{P(a \leqslant \xi \leqslant b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right) = \Phi_0\left(\frac{b-m}{\sigma}\right) - \Phi_0\left(\frac{a-m}{\sigma}\right)}$$

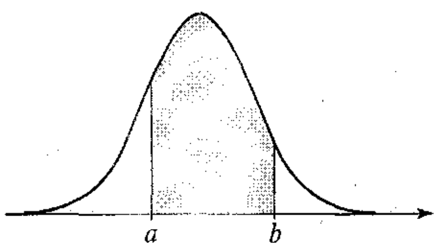


Рис. 12.23

Пусть отрезок симметричен относительно математического ожидания: $a = m - \varepsilon$, $b = m + \varepsilon$ (рис. 12.24). Тогда $\frac{a-m}{\sigma} = -\frac{\varepsilon}{\sigma}$, $\frac{b-m}{\sigma} = \frac{\varepsilon}{\sigma}$ и

$$\Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(-\frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1,$$

$$\Phi_0\left(\frac{\varepsilon}{\sigma}\right) - \Phi_0\left(-\frac{\varepsilon}{\sigma}\right) = 2\Phi_0\left(\frac{\varepsilon}{\sigma}\right).$$

Таким образом,

$$\boxed{P(|\xi - m| \leqslant \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1 = 2\Phi_0\left(\frac{\varepsilon}{\sigma}\right)}$$

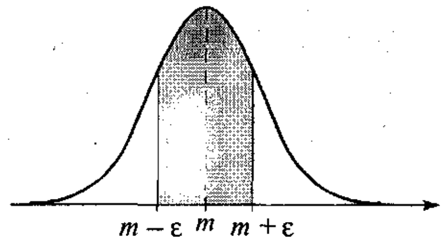


Рис. 12.24

Особенно простыми эти формулы становятся в случае, когда число ε кратно среднеквадратичному отклонению ξ, т.е. когда $\varepsilon = k\sigma$. Подставляя это значение в полученные равенства, имеем

$$\boxed{P(|\xi - m| \leqslant k\sigma) = 2\Phi(k) - 1 = 2\Phi_0(k)}$$

При $k = 3$ получаем $P(|\xi - m| \leqslant 3\sigma) = 2\Phi_0(3)$. По таблице находим $\Phi_0(3) = 0{,}4987$. Значит, вероятность равна $2 \cdot 0{,}4987 = 0{,}9974$. Итак, примерно 99,7% значений нормальной случайной величины распределено в промежутке от $m - 3\sigma$ до $m + 3\sigma$ (рис. 12.25). Вероятность попасть за пределы этого промежутка, на «хвосты», составляет менее 0,3%.

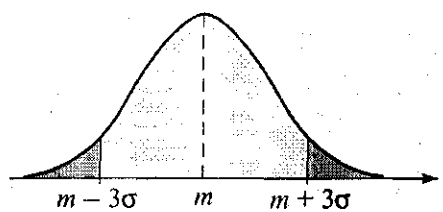


Рис. 12.25

∇4 (**«правило трех сигм»**). *Практически все значения нормальной случайной величины находятся в промежутке* $[m - 3\sigma,\ m + 3\sigma]$.

Квантили стандартного нормального распределения обычно обозначают как z_p. Для них справедливо следующее равенство.

∇5. $$z_{1-p} = -z_p.$$

Рассмотрим график плотности стандартного нормального распределения (рис. 12.26). Площадь под графиком левее квантили z_p по определению равна p. Значит, площадь правее этой точки равна $1 - p$. Такая же площадь расположена левее точки z_{1-p}. Итак, площади левее z_{1-p} и правее z_p равны. Поскольку график симметричен относительно оси ординат, из этого следует, что эти точки расположены на одинаковом расстоянии от нуля.

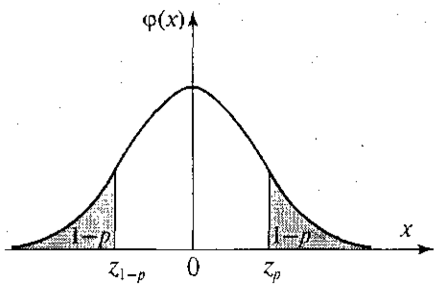


Рис. 12.26

∇6. *Пусть* $\xi \sim N(m, \sigma)$. *Тогда квантиль* x_p *случайной величины* ξ *связана с квантилью стандартного нормального распределения следующим соотношением*:

$$x_p = m + \sigma z_p.$$

Соотношение следует из определения квантилей и равенства

$$F_\xi(x) = \Phi\left(\frac{x-m}{\sigma}\right).$$

При решении задач, связанных с нормальным распределением, часто используют таблицы значений функции $\Phi(x)$ или $\Phi_0(x)$.

Задача. *Для случайной величины* $\xi \sim N(m, \sigma)$ *известно, что 14,7% ее значений меньше 12, а 40% – больше 17,2. Найти* m *и* σ.

Решение. Воспользуемся определением функции распределения:

$$0{,}147 = P(\xi < 12) = F_\xi(12) = \Phi\left(\frac{12-m}{\sigma}\right),$$

$$0{,}4 = P(\xi > 17{,}2) = 1 - F_\xi(17{,}2) = 1 - \Phi\left(\frac{17{,}2-m}{\sigma}\right).$$

В таблице (см. Приложение 1) приводятся значения $\Phi(x)$ только для неотрицательных аргументов (т.е. значения, не меньшие 0,5). Из выписанных равенств получим

$$\Phi\left(\frac{m-12}{\sigma}\right) = 1 - \Phi\left(\frac{12-m}{\sigma}\right) = 1 - 0{,}147 = 0{,}853,$$

$$\Phi\left(\frac{17{,}2-m}{\sigma}\right) = 1 - 0{,}4 = 0{,}6.$$

Найдем по таблице, каким аргументам соответствуют значения 0,853 и 0,6: ближайшие значения 0,25 и 1,05 ($\Phi(0{,}25) \approx 0{,}6$, $\Phi(1{,}05) \approx 0{,}853$). Значит,

$$\begin{cases} \dfrac{m-12}{\sigma} = 1{,}05, \\ \dfrac{17{,}2-m}{\sigma} = 0{,}25. \end{cases}$$

Складывая эти равенства, получаем $1{,}3\sigma = 5{,}2$, откуда $\sigma = 4$. Подставляя это значение в первое уравнение, находим, что $m = 16{,}2$ (рис. 12.27).

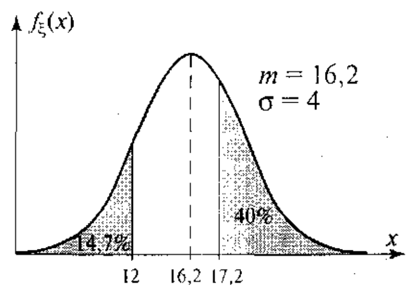


Рис. 12.27

12.7.2. Важная роль нормального распределения

Здесь мы увидим, почему такая, на первый взгляд, искусственная конструкция, как нормальная случайная величина, играет столь важную роль в теории вероятностей и при описании реальности. Сначала установим связь между биномиальным и нормальным распределением.

Теорема 12.2 (теорема Муавра – Лапласа). *Пусть $B_{n,p}$ — биномиальная случайная величина, $B^*_{n,p}$ — ее стандартизация, Ξ^* — стандартная нормальная случайная величина. Тогда для любых a, b выполняется предельное соотношение*

$$\lim_{n\to\infty} P(a \leqslant B^*_{n,p} < b) = P(a \leqslant \Xi^* < b) = \Phi(b) - \Phi(a).$$

При $a = -\infty$, $b = x$ это соотношение записывается в виде

$$\boxed{\lim_{n\to\infty} P(B^*_{n,p} < x) = \Phi(x)}$$

Говорят, что случайная величина $B^*_{n,p}$ *асимптотически нормальна*, т.е. при большом числе испытаний n она ведет себя приблизительно так же, как нормальная случайная величина.

Нам известно, что $\mathrm{M}B_{n,p} = np$, $\mathrm{D}B_{n,p} = npq$, т.е.

$$B^*_{n,p} = \frac{B_{n,p} - np}{\sqrt{npq}}.$$

Поэтому неравенство $a \leqslant B_{n,p} < b$ эквивалентно неравенству $\frac{a-np}{\sqrt{npq}} \leqslant B^*_{n,p} < \frac{b-np}{\sqrt{npq}}$. Значит, предельное соотношение можно переписать в виде

$$P(a \leqslant B_{n,p} < b) \approx \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right) =$$

$$= \Phi_0\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{a-np}{\sqrt{npq}}\right).$$

Этот результат позволяет решать задачи, связанные с биномиальным распределением (при больших n они требуют громоздких вычислений), с помощью таблиц функции стандартного нормального распределения.

Пусть, к примеру, монету подбрасывают 100 раз и нас интересует вероятность того, что число выпавших «орлов» находится в промежутке от 40 до 65. Точная формула выглядит устрашающе:

$$P(40 \leqslant B_{100,\,0,5} < 65) = P_{100}(40) + P_{100}(41) + \ldots + P_{100}(64),$$

дальнейший расчет проводится по формулам биномиальных вероятностей. Воспользовавшись теоремой Муавра – Лапласа, мы посчитаем эту вероятность приближенно. Для этого нам понадобится найти всего лишь два числа в таблице значений $\Phi_0(x)$. В данном случае $np = 50$, $\sqrt{npq} = \sqrt{25} = 5$. Значит,

$$P(40 \leqslant B_{100,\,0,5} < 65) \approx \Phi_0(3) - \Phi_0(-2) =$$
$$= \Phi_0(3) + \Phi_0(2) = 0{,}499 + 0{,}477 = 0{,}976.$$

Такое упрощение сложных вычислений составляет практическое значение теоремы Муавра – Лапласа. Однако реальная ее ценность этим не исчерпывается. Фундаментальность этого результата связана с тем, что схема испытаний Бернулли и, в частности, биномиальная случайная величина моделирует общую концепцию случайного события. Любое случайное событие можно интерпретировать как успех в серии независимых испытаний. Поэтому теорема Муавра – Лапласа утверждает, по сути, глубокую связь нормального распределения с проявлениями случайности в мире.

Абрахам де Муавр (1667–1754). Английский математик, член Лондонского королевского общества (с 1697). Родился в Витри-ле-Франсуа (Франция). В 1685–1688 гг. находился в заключении как протестант. После «варфоломеевской ночи» эмигрировал в Англию. Зарабатывал на жизнь частными уроками. С 1703 г. был в дружбе с Исааком Ньютоном.

Работал в области теории комплексных чисел, теории вероятностей. В 1733 г. вывел функцию нормального распределения как аппроксимацию биномиального распределения. Доказал важную теорему, которая потом была названа его именем.

Пьер Симон Лаплас (1749–1827). Великий французский математик, физик и астроном. Родился в Нормандии в крестьянской семье. Учился в школе бенедиктинцев. Отличался замечательной памятью и способностями, благодаря чему быстро овладел несколькими языками, а также изучил математику и астрономию. В 1766 г. приехал в Париж, где с помощью Даламбера получил место профессора в Парижской артиллерийской школе. После Великой французской революции Лаплас принимал деятельное участие в реорганизации системы образования во Франции и в создании Высшей нормальной и Политехнической школ. В 1790 г. был председателем Палаты мер и весов, в 1795 г. вошел в состав руководства Бюро долгот, в 1799 г. – министр внутренних дел. Наполеон удостоил его многих почестей, как и Людовик XVIII. Лаплас легко менял свои политические привязанности. Впрочем, такая неустойчивость позволила ему продолжать математическую деятельность при всех политических изменениях во Франции.

Научные интересы Лапласа были разносторонними. Важнейшие направления его исследований – математика, небесная механика и математическая физика. Один из его фундаментальных трудов – пятитомный «Трактат о небесной механике» (1798–1825). Лаплас завершил создание небесной механики на основе закона всемирного тяготения Ньютона. Он доказал, что этот закон полностью поясняет движение планет Солнечной системы. Исследовал закономерности движения Юпитера и Сатурна, особенности лунной орбиты. Разработал теорию приливов и отливов, развил и обосновал гипотезу Канта о происхождении Солнечной системы из туманности. Хорошо известен ответ Лапласа Наполеону, который попытался упрекнуть его за то, что в его книге нет упоминания о Боге: «Государь, я не нуждаюсь в этой гипотезе».

Другой большой труд Лапласа – «Аналитическая теория вероятностей» (1812). Многие позднейшие открытия теории вероятностей можно обнаружить в этом трактате, где подробно рассмотрены азартные игры, геометрические вероятности, схема Бернулли и ее связь с нормальным распределением, теоремы сложения и умножения вероятностей, понятие математического ожидания и многое другое. Лаплас также спас от забвения и заново сформулировал теорию, набросок которой дал Томас Байес, малоизвестный английский священник, чьи работы были опубликованы посмертно в 1763–1764 гг. Эта теория стала известна как теория вероятностей a posteriori. Еще при жизни Лапласа «Аналитическая теория вероятностей» издавалась трижды.

Продолжим обсуждение теоремы. Вспомним, что биномиальная случайная величина $B_{n,p}$ представляет собой сумму n независимых бернуллиевых случайных величин

$$\xi_i = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \quad i = 1,\ 2,\ \ldots,\ n,$$

с математическим ожиданием $M = \mathrm{M}\xi_i = p$ и дисперсией $\sigma^2 = \mathrm{D}\xi_i = pq$. Значит, предельное соотношение в теореме можно переписать следующим образом:

$$\lim_{n\to\infty} P\left(\frac{\xi_1+\xi_2+\ldots+\xi_n - nM}{\sigma\sqrt{n}} < x\right) = \Phi(x)$$

или – при больших n –

$$P\left(\frac{\xi_1+\xi_2+\ldots+\xi_n - nM}{\sigma\sqrt{n}} < x\right) \approx \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

Оказывается, что данное предельное соотношение при достаточно необременительных предположениях выполняется для л ю б ы х независимых случайных величин ξ_i.

Теорема 12.3 (центральная предельная теорема). *Пусть случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ независимы, одинаково распределены с математическим ожиданием M и конечной дисперсией σ^2. Тогда справедливо предельное соотношение*

$$\lim_{n\to\infty} P\left(\frac{\xi_1+\xi_2+\ldots+\xi_n - nM}{\sigma\sqrt{n}} < x\right) = \Phi(x).$$

Соотношение можно переписать по-другому: при больших n

$$P(\xi_1+\xi_2+\ldots+\xi_n < x) \approx \Phi\left(\frac{x-nM}{\sigma\sqrt{n}}\right).$$

В правой части равенства стоит не что иное, как функция распределения нормальной случайной величины $\xi \sim N(nM,\ \sigma\sqrt{n})$. Значит, сумма независимых случайных величин имеет асимптотически нормальное распределение. Это очень важный результат, поскольку реальные явления обычно подвержены действию м н о г и х независимых случайных факторов. В силу центральной предельной теоремы их суммарное воздействие можно приближенно описать с помощью нормального распределения. В прикладных задачах погрешность измерения различных величин рассматривается обычно как нормальная случайная величина, поскольку погрешность суммирует влияние многих случайных воздействий (неточность прибора, изменчивость условий измерения и т.д.). На самом деле, требование равных математических ожиданий и дисперсий в центральной предельной теореме можно ослабить.

Центральную предельную теорему хорошо иллюстрирует наглядное пособие, называемое по имени его изобретателя *доской Гальтона* (рис. 12.28). Прибор представляет собой ящик или наклонную доску. Вверху прибора – воронка, внизу – перегородки, образующие несколько отделений. Пространство между воронкой и отделениями занято рядами игл, расположенных в шахматном порядке. Если ссыпать в воронку мелкие шарики, диаметр которых меньше расстояния между иглами, то столкновения с иглами будут вызывать отклонение шариков от вертикального пути. В итоге шарики будут расположены в отделениях неравномерно: не только прямо под воронкой, но и сбоку. Если соединить верхние шарики в отделениях плавной кривой, то эта кривая будет удивительно напоминать график плотности нормального распределения. Это связано с тем, что шарик на своем пути много раз сталкивается с иглами, каждый раз изменяя направление движения. Столкновения случайны, и суммарное воздействие многих случайных столкновений приводит к тому, что вероятность попасть в каждое отделение примерно описывается нормальным законом: попасть в средние отделения более вероятно, а на края — менее вероятно.

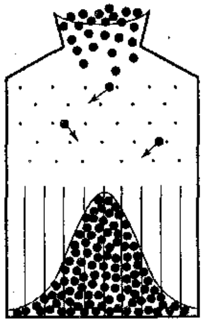


Рис. 12.28

Фрэнсис Гальтон (1822–1911). Английский психолог, антрополог, метеоролог, один из основателей биометрии и метода тестов.

Не будучи профессиональным математиком, много сделал для применения статистических методов в естественных науках и криминалистике. Его девизом было: «Где возможно — считай». Сидя на лекции, он мог подсчитывать покашливания и ерзанья в аудитории, используя полученные данные как мерило внимательности. Первым начал составлять карты погоды, открыв таким образом пути движения антициклонов. Изучал медицину, хотя никогда не работал врачом, объездил немало стран, увлекался географией, антропологией и антропометрией (изучал и обмерял туземцев Южной Африки). Произведение его кузена Чарлза Дарвина «О происхождении видов», в котором подробно рассмотрены проблемы наследственности, побудило Гальтона к изучению наследственности физических и умственных способностей. В работе ему понадобились статистические данные о мужчинах, женщинах и детях во многих поколениях. На основе этих данных Гальтон организовал развлекательный павильон на Международной лондонской выставке (1884), где каждый посетитель за три пенса мог измерить и оценить свои духовные и физические возможности. Он также применил прогрессивный метод полицейской идентификации — снятие отпечатков пальцев («Отпечатки пальцев», 1892). В книге «Естественная наследственность» (1889) описал сконструированный им прибор для демонстрации нормального закона распределения.

Нельзя, однако, абсолютизировать значение нормального распределения. Не все случайные величины в мире распределены по нормальному закону. Известный французский ученый Анри Пуанкаре (кстати сказать, один из авторов термина «нормальное распределение») шутил по этому поводу: «Все верят в нормальный закон: математики — потому, что они думают, что физики наблюдают его на опыте; физики же — потому, что думают, что математики могут доказать теоретически, что он должен выполняться». Тем не менее на практике, если явление подвержено действию многих случайных факторов, их суммарное воздействие вполне оправданно можно описать с помощью нормального закона.

Жюль Анри Пуанкаре (1854–1912). Выдающийся французский математик, физик, астроном и философ. Родился в Нанси. Окончил Политехническую школу (1875) и Горную школу (1879) в Париже. С 1879 г. преподавал на Каннском, с 1881 г. — на Парижском факультетах наук (с 1886 г. — профессор). С 1893 г. — член Бюро долгот.

Основные исследования посвящены теории чисел, алгебре, теории дифференциальных уравнений, математической физике, небесной механике, основаниям математики. Во многих рассуждениях прогнозировал дальнейшее развитие науки. Автор свыше 1000 работ, которые можно отнести к различным научным направлениям. Исследовал фигуры равновесия вращающейся жидкости. Две основные его работы по небесной механике — «Новые методы небесной механики» и «Лекции по небесной механике» — содержат важные результаты в различных областях математики. Большой цикл работ Пуанкаре относится к теории дифференциальных уравнений, в которой получил многочисленные и важные результаты (теоремы Пуанкаре, уравнения Пуанкаре и т.д.). *Использовал геометрию Лобачевского, известна его интерпретация* неевклидова пространства. В работе «О динамике электрона» независимо от А. Эйнштейна развил математические следствия «постулата относительности». Как философ является основателем конвенционализма.

На этом завершается изучение основ теории вероятностей. В заключение хотелось бы еще раз обратить внимание на красоту этой математической теории. Построенная на основе абстрактных аксиом, она с удивительной точностью описывает явления реальной жизни. На этом примере хорошо видно, как реальность и математика оказывают влияние друг на друга. Уникальная иллюстрация этого — нормальный закон распределения: математическая абстракция становится мощным инструментом исследования действительности.

Задачи

Конечные случайные величины

1. Случайная величина ξ есть число «орлов» при четырех бросаниях монеты. Найти закон распределения ξ и ее математическое ожидание.

2. В урне 4 красных, 3 синих, 1 зеленый и 2 желтых шара. Выбирают наугад 4 шара. Случайная величина ξ — число красных шаров среди вынутых. Найти закон распределения ξ и ее математическое ожидание.

3. Начертить график функции распределения случайной величины

$$\xi = \begin{pmatrix} -1 & 0 & 2 \\ 0{,}2 & 0{,}5 & 0{,}3 \end{pmatrix}.$$

4. Бросают 2 кубика. Найти закон распределения и математическое ожидание суммы выпавших очков.

5. Даны законы распределения независимых случайных величин:

$$\xi = \begin{pmatrix} 0 & 1 & 3 \\ 0{,}2 & 0{,}6 & 0{,}2 \end{pmatrix}, \quad \eta = \begin{pmatrix} 2 & 3 \\ 0{,}4 & 0{,}6 \end{pmatrix}.$$

Найти закон распределения суммы $\xi + \eta$. Проверить равенство $D(\xi + \eta) = D\xi + D\eta$.

6. При однократном бросании кубика ξ – число выпавших «орлов». Найти закон распределения случайной величины ξ^2, $M\xi^2$ и $D\xi^2$.

7. Даны две независимые случайные величины

$$\xi = \begin{pmatrix} -1 & 0 & 1 \\ 0{,}1 & 0{,}4 & 0{,}5 \end{pmatrix}, \quad \eta = \begin{pmatrix} -1 & 0 & 1 \\ 0{,}5 & 0{,}2 & 0{,}3 \end{pmatrix}.$$

Убедиться в том, что $\operatorname{cov}(\xi, \eta) = 0$.

Непрерывные случайные величины

8. Распределение случайной величины называется *треугольным* на отрезке $[a, b]$ ($\xi \sim R(a, b)$), если ее плотность распределения имеет вид, как на рис. 12.29. Найти:

а) высоту треугольника c из условия нормировки;

б) математическое ожидание и дисперсию треугольной случайной величины;

в) 95-процентную квантиль для распределения $R(1, 3)$.

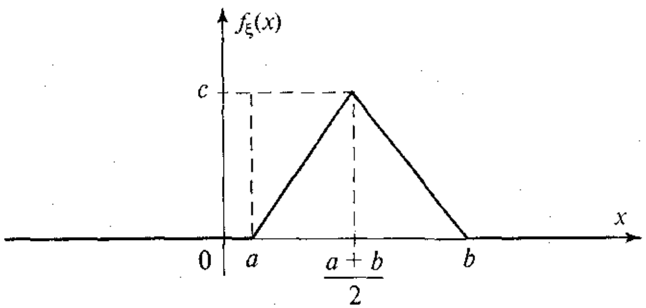


Рис. 12.29

9. Распределение случайной величины называется *экспоненциальным*, если ее плотность распределения задается формулой

$$f_\xi(x) = \begin{cases} 0, & x \leqslant 0, \\ \lambda e^{-\lambda x}, & x > 0 \end{cases}$$

(рис. 12.30). Найти $M\xi$ и $D\xi$.

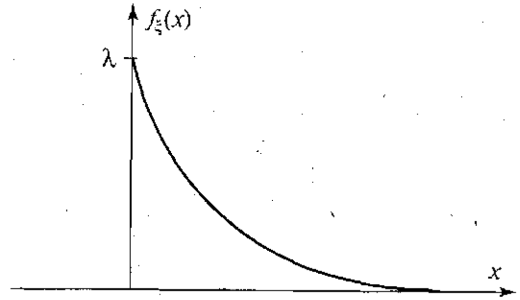


Рис. 12.30

10. Найти математическое ожидание и дисперсию случайной величины ξ, распределенной равномерно на отрезке $[2, 8]$.

Нормальное распределение

Указание. При решении пользоваться таблицами $\Phi(x)$ или $\Phi_0(x)$, пояснять решение на графике плотности нормального распределения.

11. Пусть $\xi \sim N(1, 0{,}5)$. Какое событие вероятнее: $\{\xi \leqslant 0{,}3\}$ или $\{\xi > 0{,}6\}$?

12. Пусть $\xi \sim N(m, 1)$. Известно, что 2,3% всех значений ξ отрицательны. Найти m и определить вероятность того, что $0{,}5 \leqslant \xi \leqslant 3{,}5$.

13. Пусть $\xi \sim N(0, \sigma)$. Известно, что 38,3% всех значений ξ лежат на отрезке $[0, 1]$. Найти σ и определить, какова вероятность того, что $\xi > 1{,}5$.

14. Пусть $\xi \sim N(m, \sigma)$. Известны две квантили распределения: $x_{0,9} = 1{,}141$, $x_{0,975} = 1{,}48$. Найти квантиль $x_{0,95}$.

15. Пусть $\xi \sim N(1, 1)$. Найти вероятность того, что при пяти испытаниях ровно три значения ξ попадут в отрезок $[0{,}5, 1{,}5]$.

13 Зак. № 2435 Турецкий

Часть 5
ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Но, может быть, когда-нибудь к среднему придем...

Б. Окуджава

Глава 13
ЗАДАЧИ ОЦЕНИВАНИЯ ПАРАМЕТРОВ

13.1. Основные понятия

13.1.1. Что такое математическая статистика?

Математическая статистика – раздел математики, очень близкий к теории вероятностей. Лучше сказать, математическая статистика в решении своих специфических задач пользуется результатами теории вероятностей. Однако следует иметь в виду, что это другая, отдельная наука, решающая в каком-то смысле обратные задачи по сравнению с задачами теории вероятностей. Я.И. Хургин в своей интересной, популярной книге «Да... Нет... Может быть...» образно поясняет различную природу этих задач следующим примером.

Типичная задача теории вероятностей. *Задана вероятность p выловить окуня. Какова вероятность того, что среди 200 выловленных рыб окажется ровно 3 окуня?*

Типичная задача математической статистики. *Выловили 200 рыб, среди них оказалось 3 окуня. Какова вероятность выловить окуня?*

Из этого примера хорошо видна принципиальная разница задач, которые решают эти математические науки. В задачах теории вероятностей мы исходим из того, что вероятностное пространство задано. Это означает, что заданы множество элементарных исходов Ω и вероятность P любого события из алгебры $\mathcal{F}$. Имея эти данные, мы можем п р е д с к а з а т ь поведение любой случайной величины, связанной с данным случайным экспериментом: если случайная величина конечна, закон распределения строится в виде таблицы; для непрерывных случайных величин задается плотность их распределения. В частности, в задаче с окунями реализуется схема испытаний Бернулли с заданной вероятностью успеха, по которой предсказывается вероятность появления заданного числа успехов (конкретного значения биномиальной случайной величины). По формуле Бернулли $P_{200}(3) = C_{200}^3 p^3 q^{197}$.

В математической статистике единственный объект, с которым мы работаем, – это данные эксперимента. Чаще всего результаты эксперимента выражаются значениями некоторой случайной величины. В примере с рыбами нам известно, что в результате 200 опытов случайная величина (число выловленных окуней) приняла значение 3. По этим данным требуется оценить вероятность успеха, т.е., по сути, в о с с т а н о в и т ь строение вероятностного пространства, которое в каком-то смысле лучше всего описывает результаты эксперимента. Исходя из закона статистической устойчивости, согласно которому при большом числе испытаний относительная частота стабилизируется около вероятности, можно предположить, что в данном случае вероятность успеха $p = 3/200$. Ясно, что этот результат не является точным, вместо точного значения вероятности получена ее оценка. Случись в этот день другой улов окуней, оценка была бы другой. Важно лишь, что при большом числе выловленных рыб вероятность ошибки была бы меньше. В дальнейшем мы придадим этим общим рассуждениям строгий математический смысл. Пока можно лишь заметить, что ответы в математической статистике имеют вероятностный характер.

Итак, в теории вероятностей вероятностное пространство задано и требуется предсказать возможное поведение случайной величины; в математической статистике, наоборот, известны лишь реализовавшиеся значения случайной величины, по которым реконструируется вероятностное пространство. Говорят, что по экспериментальным данным строится вероятностная модель явления, соответствующая этим данным (возможно, в каком-то смысле наилучшая). Задача математической статистики коротко формулируется как и н т е р п р е т а ц и я данных.

13.1.2. Генеральная совокупность и выборка

Пусть случайному эксперименту поставлена в соответствие некоторая случайная величина ξ с функцией распределения $F_\xi(x)$. В результате каждого опыта случайная величина принимает вполне определенное значение. Теоретически эксперименты можно проводить бесконечно. Можно представить себе большую коробку (урну), в которую мы будем складывать карточки с полученными значениями ξ. Образованная таким образом гипотетическая совокупность всех возможных значений случайной величины ξ называется ***генеральной совокупностью*** с функцией распределения $F_\xi(x)$. Функцию распределения можно заменить табличным законом распределения или плотностью распределения (соответственно для конечных или непрерывных случайных величин). Определение генеральной совокупности не носит строгого математического характера. Это понятие относится к тем математическим объектам, которые вводятся на интуитивном уровне (как множество, точка и др.).

Итак, пусть имеется генеральная совокупность, связанная с некоторым случайным экспериментом. Например, при однократном бросании монеты случайная величина ξ (число выпавших «орлов») принимает значение 0 и 1 с равной вероятностью $^1/_2$. Тогда генеральная совокупность (напомним, что она получается в результате б е с к о н е ч н о г о числа бросаний) будет состоять из равного числа нулей и единиц. Теперь эксперимент по бросанию монеты можно заменить экспериментом по случайному выбору карточки из «урны с генеральной совокупностью»: результатом будет с равной вероятностью нуль или единица. Если выбор осуществить n раз (с возвращением карточки обратно в урну), мы получим n конкретных чисел $x_1, x_2, \ldots, x_n$, другими словами, n значений случайной величины ξ. Это соответствует тому, что n раз проведен определенный случайный эксперимент. Такой набор значений случайной величины называется ***выборкой*** объема n из генеральной совокупности с функцией распределения $F_\xi(x)$, а числа $x_1, x_2, \ldots, x_n$ – ***выборочными значениями*** случайной величины ξ.

Говоря проще, выборка есть набор значений случайной величины, полученных в результате n единичных экспериментов (или одного составного эксперимента, предоставляющего собой n испытаний). Однако на выборку можно взглянуть и по-другому. В самом деле, наш составной эксперимент можно повторять неограниченное число раз. Каждый раз будет получаться новая выборка $x_1, x_2, \ldots, x_n$.

Поскольку все эти числа – реализовавшиеся значения некоторой случайной величины, каждое из них, в свою очередь, можно интерпретировать как случайную величину, распределенную так же, как генеральная совокупность. В таком случае выборка есть не что иное, как набор из n случайных величин, независимых и одинаково распределенных. Обозначение $\{x_1, x_2, \ldots, x_n\}$ имеет, таким образом, двоякий смысл: с одной стороны, это выборка, набор конкретных чисел, с другой – набор случайных величин. Набор случайных величин называют ***случайным вектором***. При такой трактовке конкретная выборка рассматривается как представитель, реализовавшееся значение случайного вектора.

Все, что находится в распоряжении математической статистики при решении своих задач, – это выборка. На основании выборки в статистике делаются выводы о параметрах и свойствах распределения генеральной совокупности. Однако в силу того, что выборка есть набор с л у ч а й н ы х величин, все статистические выводы (оценки, решения) также являются случайными. Поэтому в статистике широко применяется аппарат теории вероятностей.

Бесконечная генеральная совокупность – математическая абстракция. На практике, конечно, имеют дело с конечными совокупностями. Важно лишь, чтобы объем выборки был мал по сравнению с объемом реальной генеральной совокупности. Примером служат опросы общественного мнения. Пусть нас интересует оценка населением своего экономического положения. Каждый житель страны обладает такой оценкой и может быть отождествлен с конкретным значением этой случайной величины. Генеральная совокупность в данном случае – все население, выборка же строится на основании опроса конечного числа респондентов, небольшого по сравнению с объемом генеральной совокупности.

Классифицируя задачи теории вероятностей, мы выделяли различные схемы выбора из урны: с возвращением и без возвращения. Выбирать из «урны с генеральной совокупностью» тоже можно по-разному. Однако, если урна содержит достаточно много элементов, а выборка составляет лишь их малую часть, различие между двумя этими схемами выбора стирается. В пределе, при бесконечном объеме «урны» (а именно так определяется генеральная совокупность) различие пропадает. При определении выборки мы для простоты предположили, что реализуется выбор с возвращением, или так называемый «простой выбор». Однако на практике выборка часто образуется по схеме выбора без возвращения. Если, например, проводится опрос населения, то каждый респондент обычно опрашивается один раз: опрошенный как бы исключается из генеральной совокупности. В дальнейшем мы не будем уточнять, как именно выбираются значения из генеральной совокупности.

13.1.3. Эмпирическое распределение

Каждой выборке $\{x_1, x_2, \ldots, x_n\}$ можно поставить в соответствие конечную случайную величину, принимающую эти значения с равными вероятностями $1/n$:

$$\xi_n = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ 1/n & 1/n & \ldots & 1/n \end{pmatrix}.$$

Это распределение называется ***выборочным*** или ***эмпирическим***. Как и для любой конечной случайной величины, для эмпирической случайной величины можно построить ступенчатую функцию распределения: она называется *выборочной функцией распределения*. Кроме того, можно вычислить все числовые характеристики выборочной случайной величины ξ_n: математическое ожидание, дисперсию, СКО и др. Все эти величины снабжаются определением «выборочный»: *выборочное математическое ожидание* (его обычно называют *выборочным средним*), *выборочная дисперсия* и т.д. Например, выборочное среднее (его обозначают через $\bar{x}$) есть не что иное, как среднее арифметическое значений выборки:

$$M\xi_n = \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Соответственно, выборочная дисперсия s^2 равна

$$D\xi_n = s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Аналогично вычисляются остальные выборочные характеристики. Поскольку значениями эмпирического распределения являются выборочные значения из некоторой генеральной совокупности, неудивительно, что многие свойства генеральной совокупности в каком-то смысле передаются по наследству выборочной случайной величине. Поэтому выборочное распределение и его числовые характеристики играют большую роль при оценке параметров генеральной совокупности. Чуть позже мы придадим этим замечаниям строгий математический смысл.

13.1.4. Первичная обработка выборки

1. Упорядочение. Выборочные характеристики зависят лишь от значений $x_1, x_2, \ldots, x_n$ выборки и не зависят от их порядка. Кроме того, выборочные значения, записанные в порядке их регистрации, обычно труднообозримы и неудобны для дальнейшего анализа. Поэтому чаще всего имеют дело с выборкой, упорядоченной по возрастанию. Не вводя новых обозначений, будем считать, что выборка уже упорядочена: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$. Такая упорядоченная выборка называется *вариационным рядом*. Разность между максимальным и минимальным элементом выборки $x_n - x_1 = w$ называется *размахом выборки*.

2. Частотный анализ. Пусть выборка $\{x_1, x_2, \ldots, x_n\}$ содержит k различных чисел $z_1, z_2, \ldots, z_k$, причем z_i встречается n_i раз ($i = 1, 2, \ldots, k$). Число n_i называется *частотой* элемента выборки z_i. Ясно, что

$$\sum_{i=1}^{k} n_i = n.$$

Совокупность пар (z_i, n_i) называется *статистическим рядом* выборки. Статистический ряд удобнее всего представлять в виде таблицы: в первой строке значения z_i, во второй — их частоты n_i.

Величина $\nu_i = n_i/n$ называется *относительной частотой* значения z_i. Разумеется,

$$\sum_{i=1}^{k} \nu_i = 1.$$

Если в эмпирической случайной величине ξ_n «склеить» столбцы с одинаковыми значениями, то вероятность каждого отдельного значения z_i равна $n_i \dfrac{1}{n} = n_i/n = \nu_i$; относительная частота каждого значения есть вероятность этого значения при записи закона распределения выборочной случайной величины.

Пусть значения выборки $z_1, z_2, \ldots, z_k$ упорядочены по возрастанию. Тогда можно подсчитать *накопленные частоты* $n_1 + n_2 + \ldots + n_i$ и *накопленные относительные частоты* $\nu_1 + \nu_2 + \ldots + \nu_i$ значения z_i. По определению накопленная частота последнего значения z_k равна n, а накопленная относительная частота — единице. Эти условия служат для проверки правильности подсчета частот.

3. Группировка. При большом объеме выборки ее элементы объединяют в группы (классы), представляя результаты опытов в виде *группированного статистического ряда*. Для этого интервал, содержащий все значения выборки, разбивают на k непересекающихся интервалов. Удобнее всего разбивать на равные интервалы. При этом считается, что правая граница интервала принадлежит следующему интервалу (т.е. на самом деле это полуинтервалы, которые включают левую границу и не включают правую). Последний интервал, конечно, включает правую границу. После этого подсчитываются частоты – количество n_i элементов выборки, попавших в i-й интервал. Полученный статистический ряд в первой строке содержит середины z_i интервалов группировки, а во второй строке – частоты n_i попадания в соответствующий интервал. Наряду с частотами подсчитываются относительные частоты $\nu_i = n_i/n$, накопленные частоты и накопленные относительные частоты. Результаты обычно сводятся в *таблицу частот группированной выборки*; процесс формирования такой таблицы называется *частотной табуляцией* выборки.

Пример.

По просьбе автора его сыновья провели следующий эксперимент: книгу «Винни-Пух и все-все-все» открывали на случайной странице, где выбирали случайное слово. При этом фиксировали длину этого слова. В результате 20 опытов получена следующая выборка:

$$4, 1, 4, 5, 1, 13, 4, 10, 2, 4, 7, 2, 2, 4, 6, 4, 5, 6, 2, 4.$$

Соответствующий вариационный ряд:

$$1, 1, 2, 2, 2, 2, 4, 4, 4, 4, 4, 4, 4, 5, 5, 6, 6, 7, 10, 13.$$

Размах выборки $w = 13 - 1 = 12$. Выборка содержит 8 различных значений и описывается следующим статистическим рядом:

z_i	1	2	4	5	6	7	10	13
n_i	2	4	7	2	2	1	1	1

Эмпирический закон распределения после «склейки» имеет вид

$$\xi_n = \begin{pmatrix} 1 & 2 & 4 & 5 & 6 & 7 & 10 & 13 \\ 0{,}1 & 0{,}2 & 0{,}35 & 0{,}1 & 0{,}1 & 0{,}05 & 0{,}05 & 0{,}05 \end{pmatrix}$$

(соответствующие вероятности равны $\nu_i = n_i/n$).

Выборочная функция распределения (функция распределения выборочной случайной величины) имеет ступенчатый вид и строится как любая функция распределения конечной случайной величины: левее наименьшего значения она равна нулю; в каждой точке x_k

происходит скачок на величину вероятности p_k; правее наибольшего значения функция равна единице (рис. 13.1).

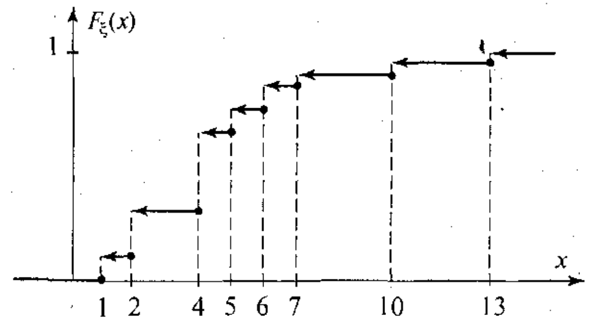


Рис. 13.1

Проведем частотную табуляцию данных. Для этого разобьем промежуток изменения значений выборки [1, 13] на 6 частей длиной 2. В результате получим следующую таблицу частот:

Номер интервала i	Границы интервала	Середина интервала z_i	Частота n_i	Накопленная частота	Относительная частота $\nu_i = n_i/n$	Накопленная относительная частота
1	1–3	2	6	6	0,3	0,3
2	3–5	4	7	13	0,35	0,65
3	5–7	6	4	17	0,2	0,85
4	7–9	8	1	18	0,05	0,9
5	9–11	10	1	19	0,05	0,95
6	11–13	12	1	20	0,05	1

Нетрудно проверить равенства $n_1 + n_2 + \ldots + n_6 = 20$, $\nu_1 + \nu_2 + \ldots + \nu_6 = 1$. Последние значения в столбцах накопленных частот, разумеется, равны $n = 20$ и 1 (для относительных частот).

13.2. Точечные оценки

13.2.1. Оценки и их классификация

Будем предполагать, что имеется выборка $\{x_1, x_2, \ldots, x_n\}$ из генеральной совокупности с функцией распределения $F_\xi(x)$. Для удобства опустим индекс ξ в обозначении функции распределения. Пусть функция распределения $F(x)$ на самом деле зависит от неизвестного параметра ϑ:

$$P(x_k < x) = F(x, \vartheta).$$

Одна из главных задач математической статистики – оценить значение параметра ϑ, имея в распоряжении только выборку. Например, нам известно, что генеральная совокупность распределена по биномиальному закону при 10 испытаниях. Неизвестным параметром в этом случае является вероятность p успеха в единичном испытании. Иногда требуется оценить несколько параметров. К примеру, у нормально распределенной генеральной совокупности требуется оценить математическое ожидание m и СКО σ; у равномерного распределения – границы отрезка $[a, b]$ и т.д. ***Оценкой*** (***точечной оценкой***) параметра ϑ называется произвольная функция от значений выборки

$$\hat{\vartheta}_n = \hat{\vartheta}_n(x_1, x_2, ..., x_n).$$

Индекс n в обозначении оценки напоминает, что она получена по выборке объема n, «крышка» показывает, что это не истинное значение параметра, а его оценка. Произвольную функцию от выборки называют еще *статистикой.*

∇1. *Оценка $\hat{\vartheta}_n$ является случайной величиной.*

Это следует из того, что выборка, по существу, есть набор случайных величин. Значит, оценка обладает всеми характеристиками случайной величины: имеет функцию распределения, математическое ожидание, дисперсию и т.д.

Разумеется, оценки в таком широком понимании (как произвольная функция выборки) не представляют практического интереса. Другими словами, не всякая статистика удовлетворительна в качестве оценки данного параметра. Хотелось бы получить в каком-то смысле х о р о ш у ю оценку. Качество оценки определяется поразному.

Оценка называется *несмещенной*, если при любом объеме выборки n ее математическое ожидание совпадает с истинным значением параметра:

$$M\hat{\vartheta}_n = \vartheta.$$

Разность $M\hat{\vartheta}_n - \vartheta$ называется *смещением* оценки $\hat{\vartheta}_n$. Несмещенная оценка имеет нулевое смещение.

Разумные соображения подсказывают, что оценка должна улучшаться, если учитываются результаты большего количества экспериментов: чем больше объем выборки, тем лучше оценка. Поскольку оценка является случайной величиной, это требование формализуют на вероятностном языке. Оценка называется *состоятельной*, если при увеличении объема выборки вероятность того, что оценка мало отличается от истинного значения, приближается к единице. Формально это записывается в виде предельного соотношения:

$$\forall \varepsilon > 0: \lim_{n\to\infty} P(\left|\hat{\vartheta}_n - \vartheta\right| < \varepsilon) = 1.$$

Итак, как бы мало ни было отклонение ε, событие $\{|\hat{\vartheta}_n - \vartheta| < \varepsilon\}$, состоящее в том, что оценка и истинное значение параметра близки по модулю, приближается к достоверному при $n \to \infty$. В этом случае говорят: оценка $\hat{\vartheta}_n$ при $n \to \infty$ *сходится по вероятности* к истинному значению параметра ϑ. Определение можно переписать по-другому:

$$\forall K > 0: \lim_{n\to\infty} P(\left|\hat{\vartheta}_n - \vartheta\right| \geqslant K) = 0,$$

т.е. вероятность больших отклонений оценки от истинного значения стремится к нулю при $n \to \infty$.

Теорема 13.1. *Если $\hat{\vartheta}_n$ – несмещенная оценка параметра ϑ и ее дисперсия стремится к нулю при $n \to \infty$ ($D\hat{\vartheta}_n \to 0$), то данная оценка является состоятельной.*

Разумеется, нам бы хотелось получать несмещенные и состоятельные оценки параметров. Предположим, мы нашли две различные оценки параметра и обе они несмещенные и состоятельные. Какая оценка лучше? Как выбрать между разными оценками? Качество оценки характеризуют *средним квадратом ошибки*:

$$\delta = M(\hat{\vartheta}_n - \vartheta)^2.$$

Для несмещенных оценок ($M\hat{\vartheta}_n = \vartheta$) этот показатель есть простонапросто дисперсия оценки $D\hat{\vartheta}_n$. Если $\hat{\vartheta}_{n1}$ и $\hat{\vartheta}_{n2}$ – две несмещенные оценки параметра ϑ и $D\hat{\vartheta}_{n1} < D\hat{\vartheta}_{n2}$, то говорят, что первая оценка *эффективнее* второй. Несмещенная оценка называется *наиболее эффективной* (или просто *эффективной*), если она имеет минимальную дисперсию среди всех несмещенных оценок данного параметра. Полученные в дальнейшем оценки параметров будут, как правило, обладать перечисленными хорошими свойствами: они будут несмещенными, состоятельными и эффективными. В математической статистике разработаны мощные методы получения хороших оценок (метод максимального правдоподобия, метод наименьших квадратов и др.). К сожалению, мы не имеем возможности останавливаться здесь на особенностях вывода оценок. Однако мы подробно прокомментируем результаты.

13.2.2. Частота и вероятность

Сформулируем важную теорему. Она принадлежит к так называемым *законам больших чисел*. Эта теорема как бы «перекидывает мостик» от теории вероятностей к математической статистике. В ней устанавливается связь между частотой и вероятностью, известная нам в форме эмпирического закона статистической устойчивости. Теорема придает этому закону строгий математический смысл, что было обещано нами в начале курса теории вероятностей.

Теорема 13.2 (теорема Бернулли). *Пусть μ_n – число успехов в n испытаниях Бернулли, p – вероятность успеха в единичном испытании. Тогда относительная частота успеха сходится по вероятности к вероятности p. Другими словами, для любого $\varepsilon > 0$ выполняется предельное соотношение*

$$\boxed{\lim_{n\to\infty} P\left(\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right) = 1}$$

Фундаментальность этой теоремы очевидна. Ведь схема испытаний Бернулли под видом успеха допускает произвольное случайное событие. Поэтому частота успеха – это на самом деле частота любого случайного события A, а вероятность успеха p – вероятность этого события $P(A)$. Из теоремы Бернулли наконец-то становится ясен смысл туманной фразы о том, что «относительная частота стабилизируется около вероятности». Мизес был не прав, когда утверждал, что относительная частота сходится к вероятности (см. конец подпараграфа 11.1.3). Сходимости в обычном смысле нет, но есть сходимость по вероятности: вероятность больших отклонений частоты от вероятности становится малой при большом числе испытаний. Из определения, данного в подпараграфе 13.2.1, понятно, что теорема Бернулли допускает другую, «статистическую», формулировку: *относительная частота есть состоятельная оценка вероятности.*

Убедимся в том, что относительная частота является несмещенной оценкой. Вспомним лишь, что число успехов μ_n есть биномиальная случайная величина $B_{n,p}$. Значит,

$$\mathrm{M}\left(\frac{\mu_n}{n}\right) = \frac{1}{n}\mathrm{M}\mu_n = \frac{1}{n}\mathrm{M}\,B_{n,p} = \frac{1}{n}np = p,$$

т.е. математическое ожидание относительной частоты совпадает с вероятностью успеха.

Состоятельность оценки можно вывести также с помощью теоремы 13.1 (не пользуясь теоремой Бернулли). Для этого вычислим дисперсию относительной частоты:

$$\mathrm{D}\left(\frac{\mu_n}{n}\right) = \frac{1}{n^2}\mathrm{D}\,B_{n,p} = \frac{npq}{n^2} = \frac{pq}{n}.$$

Видно, что дисперсия стремится к нулю при $n \to \infty$, а при этом условии несмещенная оценка является состоятельной. Известно также (примем этот факт без доказательства), что оценка вероятности через относительную частоту является эффективной. Итак, справедлив следующий вывод.

∇1. *Относительная частота есть несмещенная, состоятельная и эффективная оценка вероятности.*

Таким образом, вычисление относительной частоты события на основании многих экспериментов – не более чем способ и з м е р е н и я вероятности (так же, как взвешивание – способ измерения массы). Вероятность, a priori присущая случайному событию, измеряется с помощью относительной частоты по результатам многочисленных опытов. Можно сказать, что расставлены все точки над i в непростом вопросе о взаимосвязи частоты и вероятности, фундаментальной проблеме описания случайности. Еще раз обратим внимание на то, что для измерения, оценки параметра (в данном случае вероятности) используется только выборка. Никаких других данных для статистических оценок не требуется.

13.2.3. Оценка функции распределения

Пусть в нашем распоряжении имеется выборка $\{x_1, x_2, \ldots, x_n\}$ из генеральной совокупности с функцией распределения $F(x)$. Функция распределения $\hat{F}_n(x)$ эмпирической случайной величины

$$\xi_n = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ 1/n & 1/n & \ldots & 1/n \end{pmatrix}$$

есть вероятность события $\{\xi_n < x\}$:

$$\hat{F}_n(x) = P(\xi_n < x).$$

Пусть среди значений выборки имеется $\mu_n(x)$ чисел, которые меньше данного числа x. Тогда

$$\hat{F}_n(x) = \frac{\mu_n(x)}{n}.$$

Покажем, что выборочная функция распределения $\hat{F}_n(x)$ есть оценка функции распределения $F(x)$ генеральной совокупности. Зададимся числом x и применим схему Бернулли. Будем считать успехом событие, состоящее в том, что выборочное значение меньше x. Поскольку каждое значение из выборки есть случайная величина с функцией распределения $F(x)$, вероятность успеха $p = F(x)$. Число успехов равно $\mu_n(x)$, а относительная частота успеха равна $\mu_n(x)/n$ и совпадает с выборочной функцией распределения. Итак, выборочная функция распределения $\hat{F}_n(x)$ представляет собой относительную частоту успеха, а функция распределения генеральной совокупности $F(x)$ – вероятность успеха. Нам известно, что относительная частота есть несмещенная состоятельная оценка вероятности. Значит, $\hat{F}_n(x)$ действительно является несмещенной, состоятельной и эффективной оценкой функции распределения $F(x)$:

$$\mathrm{M}\,\hat{F}_n(x) = F(x),$$

$$\lim_{n\to\infty} P\left(\left|\hat{F}_n(x) - F(x)\right| < \varepsilon\right) = 1.$$

Стало быть, выборочная функция распределения дает нам представление об истинной функции распределения генеральной совокупности. Если n велико, то почти наверняка выборочная функция распределения приблизительно равна функции распределения генеральной совокупности. Выборочное распределение рассматривают как статистический аналог распределения генеральной совокупности. Геометрически это означает, что ступенчатый график $\hat{F}_n(x)$ проходит вблизи графика $F(x)$.

На рис. 13.2 приведен график выборочной функции распределения из примера о длине слов из «Винни-Пуха» (см. подпараграф 13.1.4). Вид графика подсказывает, что длина слов распределена по закону, близкому к нормальному. На том же рисунке приводится график функции нормального распределения, проходящий достаточно близко к графику $\hat{F}_n(x)$. Скорее всего, при большем количестве экспериментов приближение было бы еще лучше.

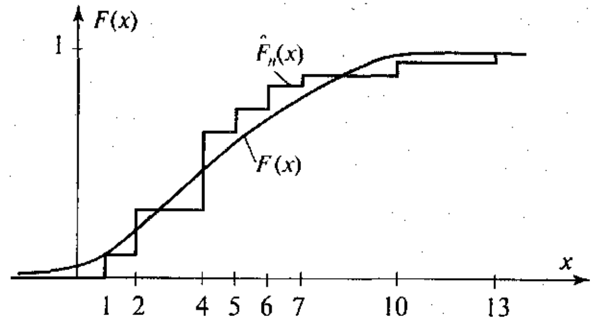


Рис. 13.2

13.2.4. Гистограмма и полигон

На практике выборки большого объема из непрерывных распределений обычно подвергаются группировке: интервал изменения значений выборки разбивается на малые промежутки, а затем подсчитываются частоты n_i попадания значений выборки в каждый i-й промежуток.

Для оценки плотности распределения генеральной совокупности используется специальный график – ***гистограмма***. Строится она следующим образом. Пусть длина каждого маленького промежутка равна h. Построим на i-м промежутке как на основании прямоугольник высотой $\frac{n_i}{nh}$. Тогда площадь прямоугольника будет равна $\nu_i = n_i/n$, т.е. относительной частоте попадания значений выборки в данный интервал. Однако нам известно, что относительная частота события при больших n оценивает его вероятность. Вероятность же попадания в i-й интервал равна площади под кривой плотности распределения генеральной совокупности на этом промежутке. Итак, площадь прямоугольника гистограммы примерно равна площади под кривой плотности распределения. Значит, верхняя часть контура гистограммы дает приближенное представление о графике плотности распределения совокупности. Таким образом, гистограмма есть статистический аналог плотности распределения так же, как выборочная функция распределения – статистический аналог функции распределения генеральной совокупности.

Пользуясь частотной табуляцией выборки из предыдущего примера, построим гистограмму. На рис. 13.3 гистограмма изображена на фоне «гипотетической» плотности нормального распределения.

Видно, что приближение не слишком удовлетворительно. Причиной может быть малый объем выборки или неверная догадка о нормальности распределения длины слов. Тем не менее гистограмма дает наглядное представление об истинной плотности распределения генеральной совокупности.

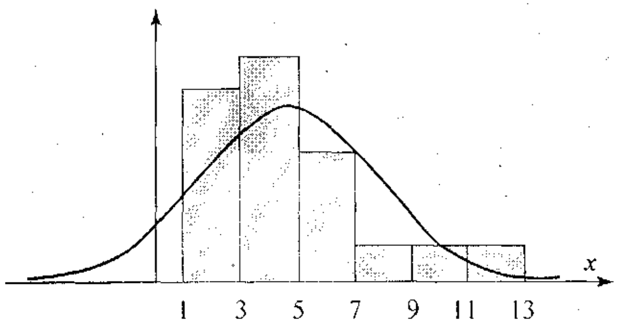


Рис. 13.3

Нетрудно посчитать, что суммарная площадь прямоугольников гистограммы равна единице. Это статистический аналог условия нормировки для плотности распределения. Построенную гистограмму было бы правильнее называть *гистограммой относительных частот*. По аналогии с этой гистограммой строится *гистограмма частот*: высота прямоугольников равна n_i/h и площадь гистограммы совпадает с объемом выборки n. Внешне гистограмма частот не отличается от гистограммы относительных частот, только каждый прямоугольник в n раз выше.

Если соединить отрезками середины верхних сторон прямоугольников гистограммы, получится еще одно графическое представление для плотности распределения – ***полигон***. Мы обозначали середины интервалов группировки через z_i. Поэтому, соединяя точки с координатами $\left(z_i, \dfrac{n_i}{nh}\right)$, получаем *полигон относительных частот*, а с координатами $(z_i, n_i/h)$ – *полигон частот*. На рис. 13.4 показано, как из гистограммы получается полигон.

Наряду с описанными в статистике используются также *гистограммы* и *полигоны накопленных частот* (*накопленных относительных частот*). Гистограмма накопленных относительных частот строится точно так же, как обычная гистограмма, но высота прямоугольника равна накопленной относительной частоте $\nu_1 + \nu_2 + \ldots + \nu_i$. Высота последнего прямоугольника равна единице. Сумма относительных частот приближенно равна сумме соответствующих вероятностей,

т.е. интегралу от плотности распределения от левой границы выборки до текущей точки. Можно считать, что такое суммирование приближенно заменяет интегрирование от $-\infty$. Однако такой интеграл от плотности распределения есть не что иное, как функция распределения $F(x)$. Поэтому гистограмма накопленных относительных частот (как и выборочная функция распределения) является аналогом функции распределения генеральной совокупности (рис. 13.5). Гистограмма накопленных частот и соответствующие полигоны строятся аналогично.

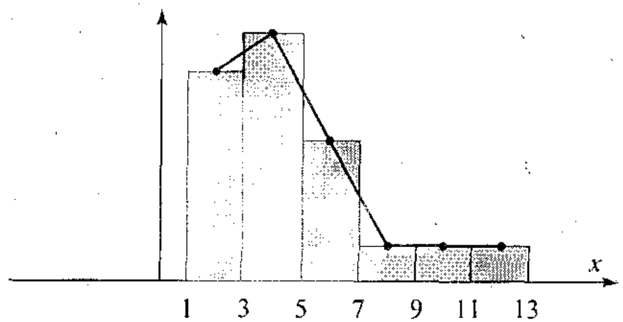


Рис. 13.4

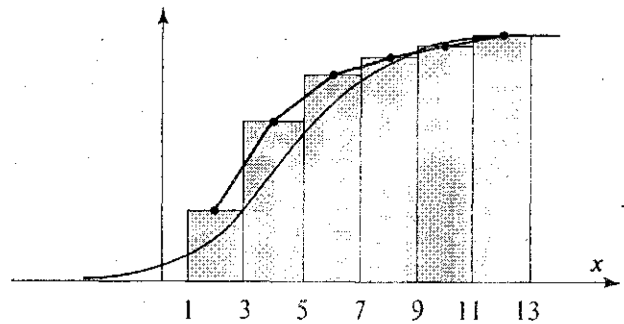


Рис. 13.5

13.2.5. Выборочные характеристики как оценки

Мы уже упоминали о важной роли, которую играет выборочное распределение при оценивании параметров генеральной совокупности. На самом деле, часто в качестве оценок просто берут соответствующие выборочные характеристики.

1. Выборочное среднее. Покажем, что выборочное среднее

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

дает несмещенную и состоятельную оценку математического ожидания

$$M = Mx_1 = Mx_2 = ... = Mx_n = M\xi$$

(каждый элемент выборки имеет такое же распределение, как случайная величина ξ, порождающая данную генеральную совокупность; их общее математическое ожидание обозначено через M). Найдем математическое ожидание оценки $\overline{x}$:

$$M\overline{x} = M\left(\frac{x_1 + x_2 + ... + x_n}{n}\right) = \frac{1}{n}(Mx_1 + Mx_2 + ... + Mx_n) = \frac{nM}{n} = M,$$

т.е. $\overline{x}$ есть несмещенная оценка математического ожидания. Для проверки состоятельности этой оценки найдем ее дисперсию, предварительно обозначив дисперсию генеральной совокупности через σ^2:

$$\sigma^2 = Dx_1 = Dx_2 = ... = Dx_n = D\xi.$$

Итак,

$$D\overline{x} = D\left(\frac{x_1 + x_2 + ... + x_n}{n}\right) = \frac{1}{n^2}(Dx_1 + Dx_2 + ... + Dx_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Видно, что дисперсия оценки стремится к нулю при $n \to \infty$. Поэтому данная оценка является состоятельной.

∇**1.** *Выборочное среднее есть несмещенная состоятельная оценка математического ожидания.*

Так, выборочное среднее для выборки из примера с длиной слов (см. подпараграф 13.1.4)

$$\overline{x} = \frac{4+1+4+5+1+13+4+10+2+4+7+2+2+4+6+4+5+6+2+4}{20} = 4{,}5.$$

Можно проверить, что в случае нормально распределенной генеральной совокупности выборочное среднее есть также и эффективная оценка математического ожидания.

2. Выборочная дисперсия. Естественной оценкой дисперсии является, как можно догадаться, выборочная дисперсия

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Найдем математическое ожидание выборочной дисперсии. Для этого преобразуем выражение для s^2 (через M, как и раньше, обозначено математическое ожидание генеральной совокупности):

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}((x_i - M) - (\overline{x} - M))^2 =$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - M)^2 - 2(\overline{x} - M)\sum_{i=1}^{n}(x_i - M) + n(\overline{x} - M)^2\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - M)^2 - 2\frac{1}{n}(\overline{x} - M)\sum_{i=1}^{n}(x_i - M) + (\overline{x} - M)^2$$

(под знаком суммы в каждом слагаемом прибавили и вычли M, а затем раскрыли скобки по формуле квадрата разности). Рассмотрим выражение

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - M) = \frac{x_1 + x_2 + ... + x_n - nM}{n} =$$

$$= \frac{x_1 + x_2 + ... + x_n}{n} - M = \overline{x} - M.$$

Значит, второе слагаемое равно $-2(\overline{x} - M)^2$. Поэтому в итоге получаем, что

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - M)^2 - (\overline{x} - M)^2.$$

Такое представление выборочной дисперсии более удобно для вычисления математического ожидания:

$$Ms^2 = \frac{1}{n}\sum_{i=1}^{n}M(x_i - M)^2 - M(\overline{x} - M)^2.$$

Поскольку M – это математическое ожидание генеральной совокупности, а значит, каждого выборочного значения, под знаком суммы стоит величина

$$M(x_i - Mx_i)^2 = Dx_i = \sigma^2$$

(дисперсия генеральной совокупности). Следовательно, первое слагаемое равно

$$\frac{1}{n}n\sigma^2 = \sigma^2.$$

Нам известно, что математическое ожидание выборочного среднего совпадает с математическим ожиданием генеральной совокупности.

Стало быть, второе слагаемое есть не что иное, как дисперсия выборочного среднего:

$$M(\bar{x}-M)^2 = M(\bar{x}-M\bar{x})^2 = D\bar{x} = \frac{\sigma^2}{n}$$

(эту величину мы уже вычисляли). Итак,

$$Ms^2 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2.$$

Видно, что выборочная дисперсия s^2 является смещенной оценкой дисперсии σ^2 генеральной совокупности со смещением $-\sigma^2/n$. Смещение стремится к нулю при $n \to \infty$. Однако хотелось бы получить несмещенную оценку дисперсии. Ее легко получить из смещенной, вспомнив линейные свойства математического ожидания, а именно достаточно умножить s^2 на дробь $\frac{n}{n-1}$:

$$M\left(\frac{n}{n-1}s^2\right) = \frac{n}{n-1}Ms^2 = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2.$$

Таким образом, несмещенная оценка дисперсии получается, если выборочную дисперсию умножить на n и разделить на $n-1$. В результате получится оценка

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2;$$

она отличается от выборочной дисперсии только коэффициентом: в выборочной дисперсии сумма делится на n, а в несмещенной оценке S^2 — на $n-1$.

Итак, получено две оценки дисперсии: первая — выборочная дисперсия s^2 является смещенной, вторая — S^2 — несмещенной. Можно проверить (не будем утяжелять изложение громоздкими вычислениями), что обе эти оценки состоятельны.

∇**2.** *Выборочная дисперсия*

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

и величина

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

дают состоятельные оценки дисперсии σ^2 генеральной совокупности (соответственно смещенную и несмещенную).

Для вычисления выборочной дисперсии можно вывести более удобную формулу, если вспомнить, что дисперсия равна «среднему квадрату минус квадрат среднего». Поэтому

$$\boxed{s^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2}$$

Воспользуемся этой формулой для вычисления оценки дисперсии в примере с длиной слов. Обычно процедура вычисления выборочной дисперсии оформляется в виде таблицы. В нашем случае получим следующие таблицы:

i	x_i	x_i^2	i	x_i	x_i^2
1	4	16	11	7	49
2	1	1	12	2	4
3	4	16	13	2	4
4	5	25	14	4	16
5	1	1	15	6	36
6	13	139	16	4	16
7	4	16	17	5	25
8	10	100	18	6	36
9	2	4	19	2	4
10	4	16	20	4	16

$\sum x_i$	$\bar{x}$	$\sum x_i^2$	$\frac{1}{n}\sum x_i^2$	s^2
90	4,5	540	27	$27-4{,}5^2=6{,}75$

Итак, $s^2 = 6{,}75$. Значит, $S^2 = \frac{20}{19}s^2 \approx 7{,}01$.

3. Квантили. Медиана. Мода. *Выборочная квантиль* легко строится по вариационному ряду выборки (т.е. по упорядоченной выборке). К примеру, 90-процентная выборочная квантиль — это значение, левее которого расположены 90% значений вариационного ряда. Соответственно, *выборочная медиана* — это середина вариационного ряда, значение, расположенное на одинаковом расстоянии от левой и правой границы выборки. Если объем выборки есть нечетное число $n = 2k-1$, то выборочная медиана есть k-е значение x_k (выборка считается упорядоченной). Например, если выборка содержит $15 = 2\cdot 8-1$ значений, то медиана — это восьмой элемент вариационного ряда: слева и справа от него расположено по 7 значений. Если же $n = 2k$ — четное число, за медиану принимается среднее арифметическое между k-м и $(k+1)$-м значениями. В примере с длиной слов (объем выборки равен 20) медиана есть $(4 + 4)/2 = 4$, так как и на 10-м, и на 11-м месте в вариационном ряду стоят четверки:

1, 1, 2, 2, 2, 2, 4, 4, 4, **4**, **4**, 4, 4, 5, 5, 6, 6, 7, 10, 13.

Нетрудно проверить, что выборочная медиана является несмещенной оценкой математического ожидания генеральной совокупности.

Выборочная мода – это наиболее вероятное, т.е. чаще всего встречающееся, значение в выборке. Так, в нашем примере чаще всего встречались слова из четырех букв, т.е. выборочная мода равна 4.

13.3. Некоторые статистические распределения

Основной метод математической статистики при решении различных задач состоит, как мы убедимся, в следующем. Иногда удается отыскать такую статистику (случайную величину, вычисляемую по выборке), распределение которой известно, несмотря на то, что неизвестны параметры распределения генеральной совокупности. С помощью такой статистики можно получить информацию о случайной величине из генеральной совокупности. Подобные результаты представляются нам почти фантастичными: только на основании выборочных значений можно построить случайную величину с известным законом распределения! Многие важные статистики распределены по специальным законам. К ним, в частности, относятся распределение χ^2 («хи-квадрат») и распределение Стьюдента. Мы не можем обойтись без рассмотрения этих распределений, хотя их определение выглядит громоздко и несколько экзотично.

13.3.1. χ^2-распределение

Пусть $\xi_1, \xi_2, \ldots, \xi_k$ – независимые случайные величины, распределенные по стандартному нормальному закону:

$$\xi_1, \xi_2, \ldots, \xi_k \sim N(0, 1).$$

Говорят, что сумма квадратов этих случайных величин распределена ***по закону*** χ^2 с k степенями свободы. Эту случайную величину обозначают $\chi^2(k)$:

$$\chi^2(k) = \xi_1^2 + \xi_2^2 + \ldots + \xi_k^2.$$

Запись $\xi \sim \chi^2(k)$ также означает, что случайная величина ξ распределена по закону χ^2 с k степенями свободы. Графики плотности распределения $\chi^2(k)$ при различных k изображены на рис. 13.6.

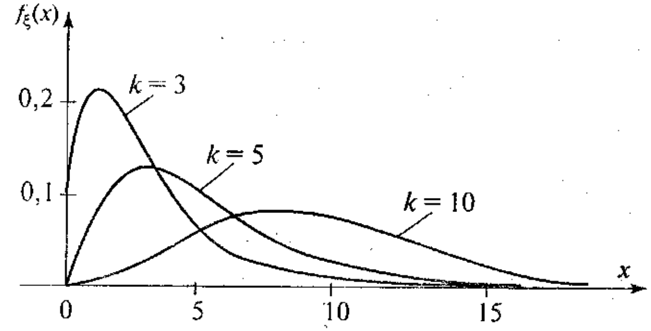


Рис. 13.6

∇**1.** *Случайная величина $\chi^2(k)$ имеет нулевую плотность распределения при $x \leqslant 0$.*

Это ясно из определения, так как данная случайная величина есть сумма квадратов и не может принимать отрицательные значения (вероятность попасть в любой интервал левее нуля равна нулю, поэтому и плотность распределения там равна нулю).

∇**2.** *При большом числе степеней свободы k распределение $\chi^2(k)$ близко к нормальному.*

На рис. 13.6 этот факт иллюстрируется графиком при $k = 10$.

Для квантилей χ^2-распределения составлены таблицы, которые обычно приводятся в учебниках по математической статистике (см. Приложение 1).

Проверяется, что математическое ожидание случайной величины, распределенной по закону χ^2 с k степенями свободы, равно k:

$$M\chi^2(k) = k.$$

13.3.2. Распределение Стьюдента

Пусть случайная величина ξ распределена по стандартному нормальному закону: $\xi \sim N(0, 1)$. Разделим ξ на корень из $\chi^2(k)/k$ (т.е. из случайной величины, распределенной по закону χ^2 с k степенями свободы, деленной на k). Говорят, что полученная случайная величина имеет ***распределение Стьюдента*** с k степенями свободы.

Данная случайная величина и соответствующий закон распределения обозначаются через $t(k)$:

$$t(k) = \frac{\xi}{\sqrt{\chi^2(k)/k}}.$$

Графики плотности распределения Стьюдента приведены на рис. 13.7 при различном числе степеней свободы. Из вида графиков и из определения можно сделать некоторые наблюдения о свойствах распределения Стьюдента.

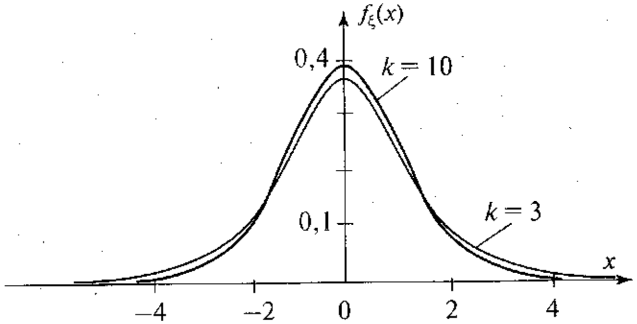


Рис. 13.7

∇**3.** *Распределение Стьюдента симметрично, причем*

$$\mathrm{M}t(k) = 0.$$

∇**4.** *При больших k распределение Стьюдента близко к стандартному нормальному распределению* $N(0, 1)$.

Квантили распределения Стьюдента с разным числом степеней свободы также можно найти в таблицах (см. Приложение 1).

В начале XX в. директор крупного пивоваренного завода, принадлежащего пивному магнату Гиннессу (помните «Книгу рекордов Гиннесса»?), случайно прочитал книгу по теории вероятностей. После этого он послал младшего служащего **Уильяма Госсета** в Центр статистических исследований в Лондоне. Центром руководил крупнейший специалист в области математической статистики Чарлз Пирсон, основатель журнала «Биометрика», в котором публиковались статистические работы.

Молодой человек проявил неординарные способности и вскоре получил важные результаты. Однако устав компании Гиннесса запрещал публиковать результаты исследований во избежание промышленного шпионажа. Тем не менее в 1908 г. в журнале «Биометрика» появилась серия блестящих статей, подписанных псевдонимом «Студент» (Student). Эти работы совершили переворот в математической статистике, а Уильям Госсет (вместе с открытым им распределением) так и вошел в историю математики под именем «Стьюдент».

13.3.3. Распределение Фишера

На основе распределения χ^2 вводится еще одна случайная величина. Пусть заданы случайные величины, распределенные по закону χ^2 с m и n степенями свободы: $\chi^2(m)$ и $\chi^2(n)$. Разделим каждую из них на соответствующее число степеней свободы и возьмем отношение полученных случайных величин. Говорят, что образованная таким образом случайная величина имеет ***распределение Фишера*** с m и n степенями свободы. Новую случайную величину (и ее закон распределения) обозначают $F(m, n)$. Таким образом,

$$F(m, n) = \frac{\chi^2(m)/m}{\chi^2(n)/n}.$$

Из определения видно, что данная случайная величина не может принимать отрицательные значения.

∇5. *Случайная величина $F(m, n)$ имеет нулевую плотность распределения при $x \leqslant 0$.*

Графики функций распределения Фишера при различном числе степеней свободы изображены на рис. 13.8. Квантили $F_p(m, n)$ распределения Фишера с m и n степенями свободы можно найти в соответствующих таблицах (см. Приложение 1).

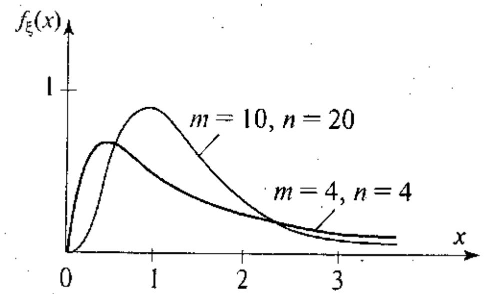


Рис. 13.8

Распределения χ^2, Стьюдента и Фишера играют важную роль при решении многих задач математической статистики. С некоторыми из них мы ознакомимся далее.

13.4. Интервальные оценки

13.4.1. Доверительные интервалы

Иногда важно получить информацию не о значении параметра, а о границах его изменения, найти интервал, в котором находится значение параметра. Поскольку для построения такого интервала у нас нет ничего, кроме выборки, ясно, что границы интервала будут случайными величинами. Значит, выводы о границах изменения параметра будут носить вероятностный характер.

Пусть задано малое число α, так называемый *уровень значимости* (на практике обычно берут α, равное 0,01, 0,05 или 0,1). Предположим, что по выборке удалось построить интервал

$$(\vartheta_1,\ \vartheta_2) = (\vartheta_1(x_1,\ x_2,\ \ldots,\ x_n),\ \vartheta_2(x_1,\ x_2,\ \ldots,\ x_n)),$$

удовлетворяющий равенству

$$P(\vartheta_1 < \vartheta < \vartheta_2) = 1 - \alpha.$$

Другими словами, построен интервал, в который истинное значение параметра ϑ попадает с достаточно большой вероятностью $1 - \alpha$. Данное выражение не совсем точно: ведь ϑ не является случайной величиной и нельзя говорить, что это значение куда-то попадает с какой-то вероятностью. Точнее будет сказать: случайный интервал $(\vartheta_1, \vartheta_2)$ накрывает истинное значение параметра с заданной, достаточно большой вероятностью. Важно, что найденный интервал не зависит от значения параметра ϑ. Тогда интервал $(\vartheta_1, \vartheta_2)$ называют ***доверительным интервалом*** для параметра ϑ с *доверительной вероятностью* $1 - \alpha$. Если, к примеру, $\alpha = 0{,}05$, то строится доверительный интервал с доверительной вероятностью 0,95 (или 95-процентный доверительный интервал).

Часто доверительный интервал находят как интервал, симметричный относительно точечной оценки параметра:

$$(\vartheta_1,\ \vartheta_2) = (\hat{\vartheta}_n - \Delta,\ \hat{\vartheta}_n + \Delta).$$

Именно с такими доверительными интервалами мы будем иметь дело в конкретных примерах. При построении доверительных интервалов будут использоваться замечательные факты, о которых упоминалось ранее: по выборке мы будем находить случайные величины с известным распределением.

13.4.2. Доверительные интервалы для параметров нормального распределения

Пусть дана выборка $x_1,\ x_2,\ \ldots,\ x_n$ из нормально распределенной генеральной совокупности:

$$x_i \sim N(m,\ \sigma), \quad i = 1,\ 2,\ \ldots,\ n,$$

и при различных условиях требуется найти доверительные интервалы для параметров m и σ.

Доверительные интервалы для математического ожидания

Построим доверительный интервал для параметра m нормальной генеральной совокупности. Рассмотрим два случая.

1. Дисперсия σ^2 известна. Рассматривая выборочное среднее как оценку математического ожидания, мы нашли его математическое ожидание и дисперсию. В данном случае

$$\mathrm{M}\bar{x} = m, \quad \mathrm{D}\bar{x} = \frac{\sigma^2}{n}.$$

Если все x_i распределены по нормальному закону, то и выборочное среднее тоже имеет нормальное распределение. Таким образом,

$$\bar{x} \sim N(m,\ \sigma/\sqrt{n}).$$

Стандартизуя эту случайную величину, мы получим статистику, распределенную по стандартному нормальному закону:

$$\boxed{U = \frac{\bar{x} - m}{\sigma/\sqrt{n}} \sim N(0,\ 1)}$$

Это **первый замечательный факт** математической статистики: независимо от истинного значения математического ожидания m, нам известно распределение статистики U.

∇**1.** *Если выборка $x_1,\ x_2,\ \ldots,\ x_n$ распределена по нормальному закону $N(m,\ \sigma)$, то статистика $U = \dfrac{\bar{x} - m}{\sigma/\sqrt{n}}$ имеет стандартное нормальное распределение.*

Пользуясь этим фактом, построим доверительный интервал для m, симметричный относительно выборочного среднего $\bar{x}$. Итак, требуется найти радиус интервала Δ, для которого выполняется неравенство

$$P(|m - \bar{x}| < \Delta) = 1 - \alpha.$$

Разделим обе части неравенства в скобках на величину $\sigma/\sqrt{n}$:

$$P\left(\frac{|\bar{x}-m|}{\sigma/\sqrt{n}} < \frac{\Delta}{\sigma/\sqrt{n}}\right)=1-\alpha.$$

(ведь $|m - \bar{x}| = |\bar{x} - m|$). По определению статистики U данное выражение переписывается как

$$P\left(|U| < \frac{\Delta}{\sigma/\sqrt{n}}\right)=1-\alpha.$$

Отсюда видно, что случайная величина U попадает в интервал $\left(-\frac{\Delta}{\sigma/\sqrt{n}}, \frac{\Delta}{\sigma/\sqrt{n}}\right)$, симметричный относительно нуля, с вероятностью $1 - \alpha$. Но статистика U распределена по стандартному нормальному закону.

Рассмотрим рис. 13.9.

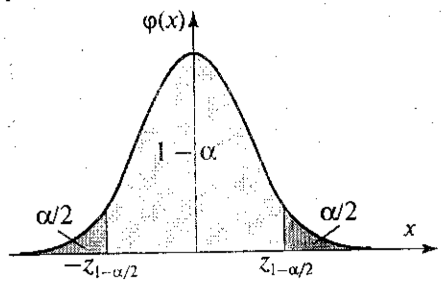


Рис. 13.9

Стандартная нормальная величина попадает в симметричный относительно нуля интервал с вероятностью $1 - \alpha$. Значит, суммарная площадь «хвостов» равна α. Поскольку интервал симметричен, каждый хвост имеет площадь $\alpha/2$. Таким образом, площадь левее правой границы интервала равна $\alpha/2 + 1 - \alpha = 1 - \alpha/2$. Это означает, что правая граница интервала совпадает с квантилью $z_{1-\alpha/2}$ стандартного нормального распределения порядка $1 - \alpha/2$. Соответственно левая симметричная граница равна $z_{\alpha/2} = -z_{1-\alpha/2}$. Учитывая, что радиус этого интервала равен $\frac{\Delta}{\sigma/\sqrt{n}}$, получаем, что $\frac{\Delta}{\sigma/\sqrt{n}} = z_{1-\alpha/2}$. Отсюда искомый радиус доверительного интервала

$$\Delta = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Если для краткости ввести обозначение $u_\alpha = z_{1-\alpha/2}$, получим следующий результат.

∇2. *Если дисперсия σ^2 нормальной генеральной совокупности известна, то доверительный интервал для математического ожидания m имеет вид*

$$\left(\bar{x}-u_\alpha\frac{\sigma}{\sqrt{n}},\quad \bar{x}+u_\alpha\frac{\sigma}{\sqrt{n}}\right).$$

Пример.

Продолжим пример с длиной слов в книге о Винни-Пухе (см. подпараграф 13.1.4). Рассматривалась выборка объемом $n = 20$ с выборочным средним $\bar{x} = 4{,}5$. Предположим, что выборка имеет нормальное распределение с дисперсией $\sigma^2 = 7$. Зададим уровень значимости $\alpha = 0{,}05$ и найдем по таблице (см. Приложение 1), что

$$u_{0,05} = z_{0,975} = 1{,}96.$$

Тогда радиус доверительного интервала равен

$$u_\alpha\frac{\sigma}{\sqrt{n}} = 1{,}96\frac{\sqrt{7}}{\sqrt{20}} \approx 1{,}16$$

и 95-процентный доверительный интервал для m имеет вид

$$(4{,}5 - 1{,}16,\ 4{,}5 + 1{,}16) = (3{,}34,\ \ 5{,}66).$$

2. Дисперсия σ^2 неизвестна. В этом случае построение доверительного интервала опирается на следующую важную теорему математической статистики.

Теорема 13.3. *Пусть выборка распределена по нормальному закону $N(m, \sigma)$ и s^2 — выборочная дисперсия. Тогда случайная величина $n\frac{s^2}{\sigma^2}$ распределена по закону χ^2 с $n - 1$ степенью свободы:*

$$\boxed{n\frac{s^2}{\sigma^2} \sim \chi^2(n-1)}$$

Это **второй замечательный факт**: какова бы ни была истинная дисперсия σ^2, распределение величины $n\frac{s^2}{\sigma^2}$ известно. Из теоремы следует, что

$$\sqrt{\frac{n}{n-1}}\frac{s}{\sigma} \sim \sqrt{\chi^2(n-1)/(n-1)}$$

(разделили на $n - 1$ и извлекли корень).

Однако

$$\sqrt{\frac{n}{n-1}}s = S,$$

где $S^2 = \dfrac{n}{n-1}s^2$ – несмещенная оценка дисперсии, поэтому

$$\frac{S}{\sigma} \sim \sqrt{\chi^2(n-1)/(n-1)}.$$

Вспомним, что такая случайная величина фигурирует в определении закона распределения Стьюдента; если на нее разделить стандартную нормальную величину, полученная величина будет иметь распределение Стьюдента с $n - 1$ степенью свободы. Поскольку у нас уже есть стандартная нормальная случайная величина, а именно статистика $U = \dfrac{\bar{x}-m}{\sigma/\sqrt{n}}$, разделим ее на величину S/σ:

$$\frac{U}{S/\sigma} = \frac{\bar{x}-m}{S\sigma/(\sigma\sqrt{n})} = \frac{\bar{x}-m}{S/\sqrt{n}}.$$

Неизвестный параметр σ чудесным образом исчез: удалось найти случайную величину с известным распределением, в которую σ не входит. Эту случайную величину называют *статистикой Стьюдента* или *t-статистикой*. Таким образом, мы пришли к **третьему замечательному факту** статистики.

∇3. *Если $\bar{x}$ – выборочное среднее, а S^2 – несмещенная оценка дисперсии для нормально распределенной выборки, то*

$$\boxed{t = \frac{\bar{x}-m}{S/\sqrt{n}} \sim t(n-1)}$$

Нетрудно заметить удивительное сходство первого и третьего замечательных фактов: истинное значение СКО σ заменяется на его несмещенную оценку, а стандартное нормальное распределение – на распределение Стьюдента. Поэтому построение доверительного интервала в данном случае осуществляется точно так же, как и при известной дисперсии. При этом используются соответствующие квантили распределения Стьюдента $t_{1-\alpha/2}(n-1)$, которые для краткости обозначим через τ_α. Приведем окончательный результат.

∇4. *Если дисперсия σ^2 нормальной генеральной совокупности неизвестна, то доверительный интервал для математического ожидания m имеет вид*

$$\left(\bar{x} - \tau_\alpha \frac{S}{\sqrt{n}},\quad \bar{x} + \tau_\alpha \frac{S}{\sqrt{n}}\right).$$

В нашем примере несмещенная оценка дисперсии $S^2 = 7{,}01$. Найдем по таблице (см. Приложение 1) квантиль распределения Стьюдента порядка 0,975 с 19 степенями свободы ($n = 20$):

$$\tau_{0,05} = t_{0,975}(19) = 2{,}093.$$

Значит, 95-процентный доверительный интервал для m в данном случае имеет вид (3,26, 5,74) (несколько шире, чем при известной дисперсии).

Для квантилей любого непрерывного распределения выполняется очевидное свойство: если $p_1 < p_2$, то $x_{p_1} < x_{p_2}$, т.е. площадь левее точки x_{p_2} больше, чем площадь левее точки x_{p_1} (рис. 13.10).

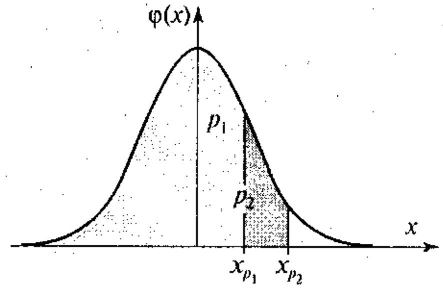


Рис. 13.10

Разумеется, это справедливо и для нормального распределения, и для распределения Стьюдента. Поскольку $u_\alpha = z_{1-\alpha/2}$, $\tau_\alpha = t_{1-\alpha/2}(n-1)$, то с уменьшением уровня значимости α эти коэффициенты увеличиваются, что при прочих равных условиях приводит к расширению доверительного интервала. Это ясно и по смыслу: уменьшение α означает увеличение доверительной вероятности $1 - \alpha$. Естественно, более широкий интервал накрывает истинное значение параметра с большей вероятностью.

∇5. *Чем больше доверительная вероятность, тем шире доверительный интервал.*

Если в рассмотренном примере с известной дисперсией задать $\alpha = 0{,}01$, то $u_{0,01} = z_{0,995} = 2{,}576$ (см. Приложение 1) и соответствующий 99-процентный доверительный интервал будет иметь вид (2,98, 6,02).

С другой стороны, видно, что радиус доверительного интервала обратно пропорционален корню из объема выборки. При увеличении объема выборки доверительный интервал сужается и точность интервального оценивания увеличивается.

∇**6.** *При увеличении объема выборки n точность интервального оценивания параметра m растет пропорционально* $\sqrt{n}$.

Значит, чтобы повысить точность в 2 раза, выборка должна быть больше в 4 раза.

Приведенные факты заслуживают того, чтобы повторить их еще раз:

Для нормально распределенной выборки
$U = \dfrac{\bar{x} - m}{\sigma/\sqrt{n}} \sim N(0, 1)$
$n\dfrac{s^2}{\sigma^2} \sim \chi^2(n-1)$
$t = \dfrac{\bar{x} - m}{S/\sqrt{n}} \sim t(n-1)$

Эти важные факты широко используются при решении различных задач математической статистики. Учитывая связь между смещенной и несмещенной оценками дисперсии, последнее выражение можно записать в виде

$$t = \frac{\bar{x} - m}{s/\sqrt{n-1}} \sim t(n-1).$$

Доверительный интервал для дисперсии

При построении доверительного интервала для дисперсии воспользуемся вторым замечательным фактом

$$n\frac{s^2}{\sigma^2} \sim \chi^2(n-1).$$

Зададим уровень значимости α. Пусть $\chi^2_{\alpha/2}(n-1)$ и $\chi^2_{1-\alpha/2}(n-1)$ квантили распределения $\chi^2(n-1)$ соответствующего поряд[к]а (рис. 13.11). Ясно, что площади левее первой и правее второй ква[н]тили равны $\alpha/2$. Поэтому вероятность попасть в промежуток д[ля] случайной величины $\chi^2(n-1)$ равна $1-\alpha$. Значит,

$$P\left(\chi^2_{\alpha/2}(n-1) < n\frac{s^2}{\sigma^2} < \chi^2_{1-\alpha/2}(n-1)\right) = 1-\alpha.$$

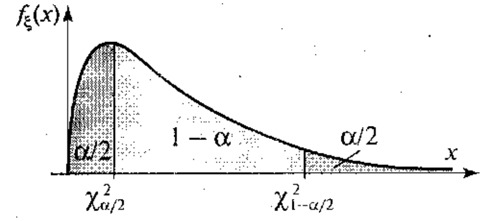


Рис. 13.11

Неравенство можно переписать для обратных величин:

$$P\left(\frac{1}{\chi^2_{1-\alpha/2}(n-1)} < \frac{\sigma^2}{ns^2} < \frac{1}{\chi^2_{\alpha/2}(n-1)}\right) = 1-\alpha,$$

откуда

$$P\left(\frac{ns^2}{\chi^2_{1-\alpha/2}(n-1)} < \sigma^2 < \frac{ns^2}{\chi^2_{\alpha/2}(n-1)}\right) = 1-\alpha.$$

Таким образом, построен доверительный интервал для σ^2.

∇**7.** ***Если выборка распределена по нормальному закону и s^2 – выборочная дисперсия, то доверительный интервал для дисперсии σ^2 имеет вид***

$$\left(\frac{ns^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{ns^2}{\chi^2_{\alpha/2}(n-1)}\right).$$

Продолжим пример с длиной слов ($n = 20$, $s^2 = 6{,}75$). Найдем по таблице требуемые квантили при $\alpha = 0{,}05$:

$$\chi^2_{0,975}(19) = 32{,}9, \quad \chi^2_{0,025}(19) = 8{,}91.$$

Легко посчитать, что доверительный интервал для дисперсии имеет вид (4,26, 15,73).

Задачи

1. Дана выборка

7, 3, 3, 6, 4, 5, 1, 2, 1, 3.

а) Построить вариационный и статистический ряды, определить [ра]змах выборки.

б) Найти выборочное среднее $\bar{x}$, выборочную дисперсию s^2, [не]смещенную оценку дисперсии S^2, выборочную медиану.

в) Построить выборочную функцию распределения.

2. Подбросить кубик 10 раз. Оценить вероятность выпадения числа, кратного трем. Сравнить оценку с точным значением вероятности. Повторить эксперимент при $n = 40$.

3. Провести статистическое исследование роста студентов в вашей группе. Сгруппировать выборку. Построить группированный статистический ряд, гистограмму частот и полигон накопленных частот.

4. Найти по таблицам квантили:

а) стандартного нормального распределения при $p = 0{,}9$, $0{,}95$, $0{,}975$, $0{,}99$;

б) распределения Стьюдента: $t_{0,75}(12)$, $t_{0,95}(22)$, $t_{0,975}(10)$, $t_{0,995}(27)$;

в) χ^2-распределения: $\chi^2_{0,005}(10)$, $\chi^2_{0,025}(17)$, $\chi^2_{0,2}(11)$, $\chi^2_{0,95}(28)$;

г) распределения Фишера: $F_{0,9}(2, 2)$, $F_{0,9}(5, 10)$, $F_{0,975}(10, 15)$.

5. Предполагая, что рост студентов в вашей группе распределен по нормальному закону, построить 95- и 99-процентные доверительные интервалы:

а) для математического ожидания (дисперсия неизвестна);

б) для дисперсии.

Глава 14
ЗАДАЧИ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

14.1. Статистические гипотезы

14.1.1. Понятие статистической гипотезы. Классификация гипотез

Наряду с задачами оценивания параметров большую группу задач математической статистики составляют так называемые задачи проверки статистических гипотез. Пусть дана выборка $\{x_1, x_2, ..., x_n\}$ из генеральной совокупности, порождаемой некоторой случайной величиной ξ. ***Статистической гипотезой*** называется предположение относительно параметров или вида распределения наблюдаемой случайной величины. Например, после анализа выборки, т.е. после того, как мы провели частотную табуляцию, определили

выборочные характеристики, построили гистограмму или полигон и т.д., мы делаем предположение о том, что данная случайная величина распределена по нормальному закону с такими-то параметрами. Теперь нужно принять решение: противоречат экспериментальные данные высказанной гипотезе или нет. Процесс принятия решения называется *проверкой статистической гипотезы*, а алгоритм проверки гипотезы, в соответствии с которым по экспериментальным данным принимается решение принять или отвергнуть гипотезу, — *решающим правилом*. Поскольку мы выдвигали гипотезу, опираясь только на случайные выборочные значения, наши выводы будут носить вероятностный характер. Мы не дадим точного ответа: да или нет. Можно будет лишь с некоторой долей уверенности (с некоторой вероятностью) утверждать, что данные не противоречат (или противоречат) предположению.

Гипотеза называется *простой*, если она однозначно определяет распределение генеральной совокупности. Так, гипотеза вида $\xi \sim N(0{,}5, 2)$ является простой. В противном случае гипотеза называется *сложной*. Например, предполагается, что $\xi \sim N(m, 2)$, где параметр m лежит в промежутке $[0, 2]$.

Статистические гипотезы можно разделить на следующие группы.

1. Гипотезы о параметрах распределения. Эти гипотезы представляют собой предположение о значении некоторых параметров распределения генеральной совокупности. Пусть, например, известно, что совокупность распределена по нормальному закону. Высказывается гипотеза о том, что параметр m имеет конкретное значение m_0 или лежит в заданном промежутке $[a, b]$. Другая разновидность гипотез о параметрах состоит в предположении о том, что параметры (например, математические ожидания) в двух выборках равны между собой. Гипотезы о параметрах распределения можно выдвигать, располагая достаточно большой информацией о генеральной совокупности или имея весомые основания считать известным ее закон распределения.

2. Гипотезы о виде распределения. Это более общие гипотезы, они выдвигаются в условиях недостаточной информации о генеральной совокупности. «Глядя» на выборку, мы выдвигаем гипотезу о том, по какому закону распределена совокупность (например, по равномерному или по нормальному). Проверка гипотезы о нормальности может помочь при дальнейшей обработке выборки: если случайную величину уверенно можно считать нормальной, то к ней применимы все теоремы о нормальных величинах, в частности, имеется возможность строить доверительные интервалы для параметров.

Проверяемая гипотеза называется *нулевой* и обычно обозначается H_0. Наряду с H_0 рассматривается *альтернативная* (конкурирующая) гипотеза H_1, т.е. та гипотеза, которая будет принята в случае, если нулевая гипотеза отвергается. Пусть, к примеру, рассматривается гипотеза о значении параметра m нормальной совокупности:

$$H_0\colon m = m_0.$$

Для этой гипотезы можно выдвинуть различные альтернативы:

$$H_1^{(1)}\colon m \neq m_0;$$
$$H_1^{(2)}\colon m > m_0;$$
$$H_1^{(3)}\colon m < m_0;$$
$$H_1^{(4)}\colon m = m_1,\ m_1 \neq m_0.$$

Выбор альтернативной гипотезы определяется конкретной формулировкой задачи.

14.1.2. Общая схема проверки гипотез

При формировании решающего правила используется та же идея, что и при построении доверительных интервалов, а именно требуется найти случайную величину (так называемую *статистику критерия*), удовлетворяющую двум основным требованиям:

1) ее значение можно посчитать только на основании выборки;

2) ее распределение известно в предположении, что нулевая гипотеза верна.

После того как такая статистика выбрана, на числовой оси выделяется область, попадание в которую для этой случайной величины маловероятно (*критическая область*). Малая вероятность задается, как и в доверительных интервалах, малым числом α (*уровнем значимости*). Основной принцип проверки гипотез состоит в следующем.

> Маловероятное событие считается **невозможным**.
> Событие с большой вероятностью считается **достоверным**.

По выборке можно найти значение статистики критерия. Если это конкретное число попало в критическую область (попадание в которую, напомним, маловероятно, если нулевая гипотеза верна), делается вывод о том, что экспериментальные данные противоречат выдвинутой гипотезе. В самом деле, если бы гипотеза была верна, то в соответствии с известным распределением статистики критерия вероятность попадания ее выборочного значения в критическую область была бы мала. Если это событие (маловероятное и, следовательно, невозможное) все же произошло, это неизбежно означает, что предположение о верности гипотезы ошибочно.

При таком подходе самое сложное – выбрать подходящую статистику критерия. В этом нам помогут замечательные статистические факты, на которые мы опирались при построении доверительных интервалов, а также некоторые дополнительные сведения. Говорят, что такой подход к проверке статистических гипотез основан на *статистическом критерии*, или *критерии значимости*.

Построение решающего правила на основе критерия значимости можно разбить на следующие основные шаги:

1. Сформулировать нулевую (H_0) и альтернативную (H_1) гипотезы.
2. Задать уровень значимости α.
3. Выбрать статистику Z критерия для проверки гипотезы H_0.
4. Найти плотность распределения статистики критерия $f_Z(x) = f_Z(x\,|\,H_0)$ **в предположении, что гипотеза H_0 верна.**
5. Определить на числовой оси критическую область $V_{кр}$ из условия
$$P(Z \in V_{кр}\,|\,H_0) = \alpha$$
(условная вероятность того, что Z попадает в область $V_{кр}$ при условии, что гипотеза H_0 верна). Область $\boldsymbol{R}\setminus V_{кр}$ в этом случае называется *областью принятия решения*. Условия, задающие критическую область, называются просто *критерием*.
6. По выборке вычислить выборочное значение Z_s статистики критерия.
7. Принять решение:
   - если $Z_s \in V_{кр}$, гипотеза H_0 отклоняется (т.е. принимается гипотеза H_1);
   - если $Z_s \in \boldsymbol{R}\setminus V_{кр}$, гипотеза H_0 принимается.

Принятое решение, разумеется, носит вероятностный, случайный характер. Поэтому обычно применяют более осторожные формулировки. Вместо того чтобы сказать «гипотеза H_0 отклоняется», говорят «данные эксперимента не подтверждают гипотезу H_0», «гипотеза не согласуется с экспериментом» и т.д.

Отметим следующую особенность.

∇1. *Значение уровня значимости не определяет критическую область однозначно.*

В самом деле, зная плотность распределения статистики Z, можно выделить сколько угодно областей на числовой оси, вероятность попадания в которые равна α. В частности, этому условию удовлетворяют области $\{x > z_{1-\alpha}\}$ (рис. 14.1, *а*), $\{x < z_\alpha\}$ (рис. 14.1, *б*) или $\{x > z_{1-\alpha/2},\ x < z_{\alpha/2}\}$ (рис. 14.1, *в*), где через z_p обозначены квантили распределения статистики Z.

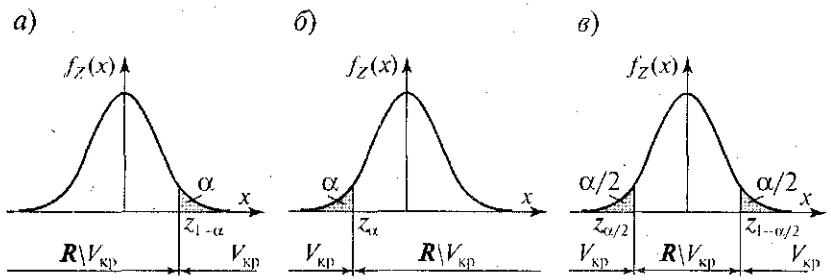


Рис. 14.1

Именно эти критические области обычно и применяются. Критерий в этих случаях называют соответственно *правосторонним*, *левосторонним* или *двусторонним*. На практике выбор критической области обычно определяется видом альтернативной гипотезы.

14.1.3. Ошибки при проверке гипотез

Принятие решения на основе статистического критерия носит случайный характер. Мы что-то решили, однако существует еще и реальное положение вещей, о котором мы и строим свои гипотезы. Возможны следующие ситуации.

1. Гипотеза H_0 верна, и она не отвергается. Это хорошо: наше решение отражает истинное положение.

2. Гипотеза H_0 верна, но она отвергается. В этом случае говорят, что допущена *ошибка I рода*. Рассмотрим ситуацию внимательнее. Поскольку нулевая гипотеза верна, статистика Z действительно имеет то распределение, на основании которого принималось решение. Тем не менее выборочное значение статистики попало в критическую область. Вероятность этого события по определению равна уровню значимости α.

∇2. *Вероятность ошибки I рода равна уровню значимости критерия.*

Так как уровень значимости задается произвольно, в нашей власти снизить вероятность ошибки I рода до сколь угодно низкого уровня.

3. Гипотеза H_0 неверна, и она отвергается. Снова все в порядке: отвергнута неверная гипотеза.

4. Гипотеза H_0 неверна, но она не отвергается. Тогда говорят, что допущена *ошибка II рода*. В этой ситуации выборочное значение попало в область принятия решения, тогда как гипотеза H_0 на самом деле неверна. Если распределение статистики Z известно и в предположении, что верна альтернативная гипотеза H_1, можно посчитать вероятность ошибки II рода: это условная вероятность того, что Z попадает в область $R\backslash V_{кр}$ при условии, что верна гипотеза H_1. Вероятность ошибки II рода обычно обозначают через β:

$$\beta = P(Z \in R\backslash V_{кр} \mid H_1).$$

Нам хотелось бы сделать вероятность ошибок поменьше. К сожалению, известно, что в большинстве случаев нельзя добиться минимального значения вероятностей α и β одновременно. Поступают обычно следующим образом: фиксируют вероятность α ошибки I рода, а затем добиваются минимума вероятности β ошибки II рода. За счет чего можно уменьшить β при фиксированном значении α? Только за счет выбора критической области: при заданной альтернативе H_1 критическую область выбирают таким образом, чтобы значение β (вероятность принять неверную гипотезу) было наименьшим из возможных. В этом случае вероятность отвергнуть неверную гипотезу $1 - \beta$ максимальна. Это число называют *мощностью критерия*. Таким образом, задача состоит в построении наиболее мощного критерия при заданном уровне значимости.

Пусть мы производим на продажу какой-то товар. Обычно, если объем выпуска большой, невозможно проверить качество каждого изделия. В этом случае применяется выборочный контроль: случайным образом отбираются экземпляры для проверки и на основании ее результатов делается вывод о том, хороша или нет вся партия. Фактически проверяется нулевая гипотеза о том, что партия товара удовлетворяет требованиям качества. Естественная альтернатива состоит в том, что вся партия некачественная. Ошибка I рода в данном случае состоит в том, что фактически хорошая продукция забракована. Это невыгодно производителю товара. Поэтому ошибку I рода называют иногда «ошибкой продавца». Наоборот, при ошибке II рода принят товар на самом деле некачественный. Это невыгодно потребителю, и ошибку II рода называют «ошибкой покупателя».

14.1.4. Проверка гипотез и доверительные интервалы

Пусть проверяется нулевая гипотеза о значении некоторого параметра ϑ:

$$H_0\colon \vartheta = \vartheta_0$$

при альтернативе $H_1\colon \vartheta \neq \vartheta_0$.

Представляется естественным в качестве статистики критерия выбрать о ц е н к у $\hat{\vartheta}_n$ параметра ϑ. В самом деле, чаще всего распределение оценки становится известным в предположении, что нулевая гипотеза верна. Например, выборочное среднее $\bar{x}$ для нормальной выборки с известной дисперсией σ^2 распределено по нормальному закону $N(m, \sigma/\sqrt{n})$. Если предположить, что $m = m_0$ (справедлива гипотеза H_0), то закон распределения оценки $\bar{x}$ полностью определен.

Пусть при этом найден доверительный интервал для параметра ϑ с доверительной вероятностью $1 - \alpha$, симметричный относительно оценки $\hat{\vartheta}_n$:

$$P(\hat{\vartheta}_n - \Delta < \vartheta < \hat{\vartheta}_n + \Delta) = 1 - \alpha.$$

Неравенство вида $a - \Delta < b < a + \Delta$ записывается через модуль: $|a - b| < \Delta$. Величины a и b входят в это неравенство симметрично. Поэтому его можно записать в виде $b - \Delta < a < b + \Delta$: если b лежит в Δ-окрестности a, то и a лежит в Δ-окрестности b (рис. 14.2).

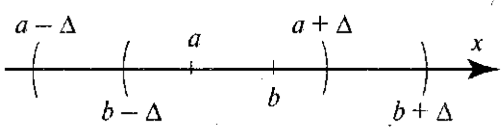


Рис. 14.2

Возвращаясь к $a = \hat{\vartheta}_n$ и $b = \vartheta$, получаем, что

$$P(\vartheta - \Delta < \hat{\vartheta}_n < \vartheta + \Delta) = 1 - \alpha,$$

т.е. оценка попадает в окрестность истинного значения с вероятностью $1 - \alpha$.

Предположим теперь, что верна гипотеза H_0 и истинное значение параметра ϑ равно ϑ_0. В этом случае событие

$$\vartheta_0 - \Delta < \hat{\vartheta}_n < \vartheta_0 + \Delta$$

происходит с большой вероятностью $1 - \alpha$ и в соответствии со схемой проверки статистических гипотез может считаться достоверным. По сути, область, в которой выполняется данное неравенство, является областью принятия решения. Если значение оценки, вычисленное по выборке, не удовлетворяет этому неравенству (а вероятность этого α мала), можно сделать вывод о том, что данные не подтверждают нулевую гипотезу. Критическая область имеет вид

$$\hat{\vartheta}_n < \vartheta_0 - \Delta, \qquad \hat{\vartheta}_n > \vartheta_0 + \Delta.$$

Уровень значимости α для доверительного интервала становится уровнем значимости при проверке гипотезы на основе статистики $\hat{\vartheta}_n$.

Возможны и более сложные случаи, но оценка и доверительный интервал часто помогают решить задачу проверки гипотезы о значении параметра.

14.2. Гипотезы о параметрах нормального распределения

14.2.1. Гипотеза о математическом ожидании

Пусть известно, что выборка $\{x_1, x_2, \ldots, x_n\}$ получена из генеральной совокупности, распределенной по нормальному закону. Выдвигается гипотеза о значении математического ожидания совокупности:

$$H_0\colon m = m_0.$$

Рассмотрим сначала простейшую альтернативу $H_1\colon m \neq m_0$.

Решающее правило для проверки этой гипотезы строится на основе доверительного интервала для математического ожидания. Ранее уже построен доверительный интервал с доверительной вероятностью $1 - \alpha$ вида $(\bar{x} - \Delta, \bar{x} + \Delta)$, радиус которого равен

$$\Delta = u_\alpha \frac{\sigma}{\sqrt{n}}$$

в случае известной дисперсии и

$$\Delta = \tau_\alpha \frac{S}{\sqrt{n}},$$

если дисперсия неизвестна. Напомним, что через u_α и τ_α обозначены квантили порядка $1-\alpha/2$ соответственно стандартного нормального распределения и распределения Стьюдента с $n-1$ степенью свободы, а S^2 — несмещенная оценка дисперсии σ^2. Итак,

$$P(\bar{x} - \Delta < m < \bar{x} + \Delta) = 1 - \alpha,$$

или

$$P(m - \Delta < \bar{x} < m + \Delta) = 1 - \alpha.$$

Значит, если $m = m_0$ (нулевая гипотеза верна), то

$$P(|\bar{x} - m_0| < \Delta) = 1 - \alpha$$

и в качестве критической области можно выбрать область, в которой это неравенство не выполняется:

$$\bar{x} < m_0 - \Delta, \qquad \bar{x} > m_0 + \Delta.$$

Проверяется, что такой критерий является наиболее мощным, т.е. обеспечивает при заданном уровне значимости α наименьшую вероятность β ошибки II рода. Это представляется весьма правдоподобным. Действительно, если конкретное значение $\bar{x}$ оказалось по модулю далеко от предполагаемого значения математического ожидания m_0, то с большой долей уверенности можно утверждать, что гипотеза $m = m_0$ не подтверждается экспериментом. Таким образом, получены двусторонние критерии для проверки гипотезы о математическом ожидании нормальной совокупности.

∇**1.** *Гипотеза H_0: $m = m_0$ при альтернативе H_1: $m \neq m_0$ для нормально распределенной совокупности не подтверждается экспериментом при уровне значимости α, если выборочное среднее $\bar{x}$ удовлетворяет условиям:*

- *при известной дисперсии*

$$\bar{x} < m_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{x} > m_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}};$$

- *при неизвестной дисперсии*

$$\bar{x} < m_0 - t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}}, \quad \bar{x} > m_0 + t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}}.$$

Заметим, что критические области можно записать в виде

$$\left|\frac{\bar{x} - m_0}{\sigma/\sqrt{n}}\right| > z_{1-\alpha/2},$$

$$\left|\frac{\bar{x} - m_0}{S/\sqrt{n}}\right| > t_{1-\alpha/2}(n-1),$$

т.е. в качестве статистики критерия можно взять величины

$$U = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}} \sim N(0, 1), \quad t = \frac{\bar{x} - m_0}{S/\sqrt{n}} \sim t(n-1).$$

В последнем случае критерий называется *критерием Стьюдента.*

Рассмотрим теперь альтернативу H_1: $m > m_0$.

В качестве статистики критерия снова выберем величины U или t. Ясно, что «неблагоприятной» областью для нулевой гипотезы теперь становятся те значения $\bar{x}$, которые лежат намного правее m_0. Поэтому критическая область в этом случае выбирается на правом «хвосте» соответствующих распределений: критерий является правосторонним. Чтобы площадь правого «хвоста» под графиком плотности была равна α, его граница должна совпадать с квантилью порядка $1-\alpha$:

$$U > z_{1-\alpha} \quad \text{(при известной дисперсии)};$$
$$t > t_{1-\alpha}(n-1) \quad \text{(при неизвестной дисперсии)}.$$

Если же выбрана альтернатива H_1: $m < m_0$, то критическая область лежит на левом «хвосте» распределения левее квантили порядка α:

$$U < z_\alpha \quad \text{(при известной дисперсии)};$$

$$t < t_\alpha(n-1) \quad \text{(при неизвестной дисперсии)}.$$

Сформулируем полученный результат.

∇**2.** *Гипотеза H_0: $m = m_0$ при альтернативе H_1: $m > m_0$ ($m < m_0$) для нормально распределенной совокупности не подтверждается экспериментом при уровне значимости α, если выборочное среднее $\bar{x}$ удовлетворяет условиям:*

- *при известной дисперсии*

$$\bar{x} > m_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$$

$$\left(\bar{x} < m_0 + z_\alpha\frac{\sigma}{\sqrt{n}}\right);$$

- *при неизвестной дисперсии*

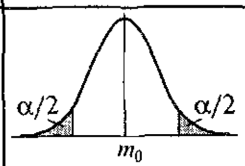
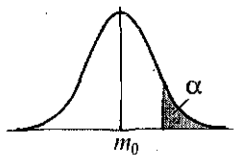
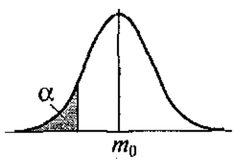
$$\bar{x} > m_0 + t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}$$

$$\left(\bar{x} < m_0 + t_\alpha(n-1)\frac{S}{\sqrt{n}}\right).$$

Обратим внимание на то, что в силу симметричности стандартного нормального распределения и распределения Стьюдента при малых α квантили порядка α отрицательны, т.е. граница критической области расположена в данном случае левее m_0.

Сведем полученные результаты в таблицу (табл. 14.1).

Таблица 14.1

H_1	Критерий		Критическая область
	σ известна	σ неизвестна	
$m \neq m_0$	$\bar{x} > m_0 + u_\alpha \dfrac{\sigma}{\sqrt{n}}$, $\bar{x} < m_0 - u_\alpha \dfrac{\sigma}{\sqrt{n}}$	$\bar{x} > m_0 + \tau_\alpha \dfrac{S}{\sqrt{n}}$, $\bar{x} < m_0 - \tau_\alpha \dfrac{S}{\sqrt{n}}$	
$m > m_0$	$\bar{x} > m_0 + z_{1-\alpha} \dfrac{\sigma}{\sqrt{n}}$	$\bar{x} > m_0 + t_{1-\alpha}(n-1) \dfrac{S}{\sqrt{n}}$	
$m < m_0$	$\bar{x} < m_0 + z_\alpha \dfrac{\sigma}{\sqrt{n}}$	$\bar{x} < m_0 + t_\alpha(n-1) \dfrac{S}{\sqrt{n}}$	

Пример.

Рассмотрим тот же самый пример с длиной слов, в котором получилось $\bar{x} = 4{,}5$ (см. подпараграф 13.1.4). Будем для простоты предполагать, что дисперсия $\sigma^2 = 7$ известна. Проверим гипотезу H_0: $m = 4$ против альтернативы H_1: $m \neq 4$ при уровне значимости $\alpha = 0{,}05$. По таблицам найдено $u_{0,05} = z_{0,975} = 1{,}96$ (см. подпараграф 13.4.2). Таким образом, критическая область имеет вид

$$\bar{x} < 3{,}84, \qquad \bar{x} > 5{,}16$$

и выборочное среднее не удовлетворяет этим условиям. Значит, нет оснований отвергать нулевую гипотезу. Делаем вывод, что гипотеза не противоречит данным эксперимента.

Проверим ту же гипотезу против альтернативы H_1: $m > 4$.

Найдем по таблице (см. Приложение 1) $z_{1-\alpha} = z_{0,95} = 1{,}645$. Критическая область: $\bar{x} > 4 + 1{,}645\sqrt{7/20} \approx 4{,}973$. Среднее снова не попало в критическую область. Поэтому гипотеза не отвергается.

Предположим тем не менее, что на самом деле $m = 5$. Найдем в этом случае вероятность β ошибки II рода. Это вероятность того, что выборочное среднее не попало в критическую область при том, что $\bar{x} \sim N(5, \sqrt{7/20})$:

$$\beta = P(\bar{x} < 4{,}973) = \Phi\left(\frac{4{,}973-5}{\sqrt{7/20}}\right) \approx$$

$$\approx \Phi(-0{,}05) = 1 - \Phi(0{,}05) = 1 - 0{,}5199 = 0{,}4801.$$

Таким образом, вероятность ошибки II рода весьма высокая — 48,01%. Это может быть вызвано несколькими причинами:

недостаточно нормальным распределением совокупности;

небольшим объемом выборки (при большем объеме, вполне возможно, гипотеза была бы отвергнута);

слишком произвольным «назначением» дисперсии.

Тем не менее мы вынуждены оперировать теми данными, которыми располагаем. Нужно только по возможности полнее и точнее их интерпретировать.

Если рассмотреть другую нулевую гипотезу, а именно H_0: $m = 3$, против той же альтернативы, то граница критической области смещается на единицу влево: $\bar{x} > 3{,}973$. В этом случае мы отвергаем нулевую гипотезу и принимаем альтернативную: данные противоречат предположению. Действительно, альтернативная гипотеза представляется более правдоподобной, чем нулевая.

14.2.2. Гипотеза о равенстве математических ожиданий

Данные часто можно сгруппировать в соответствии с каким-либо признаком. Предположим, выборка содержит данные о доходах. Такую выборку можно разбить на две группы: например, доходы студентов и всех остальных, доходы бизнесменов и небизнесменов и т.п. Возникает естественный вопрос: является ли фактор, по которому сгруппированы данные, важным с точки зрения выбранного показателя (уровня доходов)? Говорят: является ли фактор *статистически значимым*? О значимости фактора можно судить по средним значениям в каждой выборке. Если средние не сильно различаются, можно достаточно уверенно предположить, что фактор не является значимым. Данная гипотеза формулируется как гипотеза о равенстве математических ожиданий для двух независимых выборок.

Пусть даны две выборки $\{x_1, x_2, \ldots, x_n\}$ и $\{y_1, y_2, \ldots, y_m\}$ из нормальных генеральных совокупностей $N(m_x, \sigma_x)$, $N(m_y, \sigma_y)$. Проверяется гипотеза

$$H_0\colon m_x = m_y.$$

Решающее правило очень похоже на правило, которое применяют при проверке гипотезы о значении математического ожидания. Как обычно, рассматриваются разные случаи.

1. Дисперсии σ_x^2, σ_y^2 известны. Пусть $\bar{x}$ и $\bar{y}$ – выборочные средние для двух выборок. Рассмотрим разность $\bar{x} - \bar{y}$. Математическое ожидание разности в предположении равных математических ожиданий равно нулю:

$$\mathrm{M}(\bar{x} - \bar{y}) = m_x - m_y = 0.$$

Найдем дисперсию разности:

$$\mathrm{D}(\bar{x}-\bar{y}) = \mathrm{D}\bar{x} + \mathrm{D}\bar{y} = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}.$$

Значит, $\bar{x} - \bar{y} \sim N(0, \sqrt{\mathrm{D}\bar{x} + \mathrm{D}\bar{y}})$ и, стало быть,

$$\frac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} \sim N(0, 1).$$

Найдена статистика, распределенная по стандартному нормальному закону независимо от истинных значений математического ожидания (в предположении, что нулевая гипотеза выполняется). Дальше применяется стандартная схема. При альтернативе $H_1\colon m_x \neq m_y$ применяется двусторонний критерий

$$\frac{|\bar{x}-\bar{y}|}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} > u_\alpha,$$

где через u_α снова обозначена квантиль стандартного нормального распределения порядка $1 - \alpha/2$. Для альтернатив $H_1\colon m_x > m_y$ и $H_1\colon m_x < m_y$ применяются односторонние критерии соответственно

$$\frac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} > z_{1-\alpha} \quad \text{и} \quad \frac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} < z_\alpha.$$

2. Дисперсии σ_x^2, σ_y^2 неизвестны, но равны. Предположим, что $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Тогда, как ясно из предыдущего,

$$\mathrm{D}(\bar{x}-\bar{y}) = \sigma^2 \frac{m+n}{mn}$$

и

$$\frac{\bar{x}-\bar{y}}{\sigma}\sqrt{\frac{mn}{m+n}} \sim N(0, 1).$$

Оценим неизвестную дисперсию σ^2. Пусть S_x^2, S_y^2 – несмещенные оценки дисперсии по обеим выборкам:

$$\mathrm{M}S_x^2 = \mathrm{M}S_y^2 = \sigma^2.$$

Рассмотрим статистику

$$S^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2},$$

для которой

$$\mathrm{M}S^2 = \frac{n-1}{n+m-2}\mathrm{M}S_x^2 + \frac{m-1}{n+m-2}\mathrm{M}S_y^2 = \sigma^2.$$

Таким образом, величина S^2 также есть несмещенная оценка дисперсии рассматриваемых нормальных совокупностей.

По теореме 13.3

$$(n-1)\frac{S_x^2}{\sigma^2} \sim \chi^2(n-1), \quad (m-1)\frac{S_y^2}{\sigma^2} \sim \chi^2(m-1).$$

Можно проверить, что при этом

$$(n-1)\frac{S_x^2}{\sigma^2} + (m-1)\frac{S_y^2}{\sigma^2} \sim \chi^2(n+m-2).$$

Значит,

$$\frac{(n-1)S_x^2 + (m-1)S_y^2}{(n+m-2)\sigma^2} = \frac{S^2}{\sigma^2} \sim \frac{\chi^2(n+m-2)}{n+m-2},$$

или

$$\frac{S}{\sigma} \sim \sqrt{\frac{\chi^2(n+m-2)}{n+m-2}}.$$

Если разделить на эту величину статистику

$$\frac{\bar{x}-\bar{y}}{\sigma}\sqrt{\frac{mn}{m+n}} \sim N(0, 1),$$

получится по определению случайная величина, распределенная по закону Стьюдента с $n + m - 2$ степенями свободы:

$$\frac{\bar{x}-\bar{y}}{S}\sqrt{\frac{mn}{m+n}} \sim t(n+m-2).$$

Эту величину, не зависящую от неизвестной дисперсии, можно выбрать в качестве статистики критерия. Для альтернативы H_1: $m_x \neq m_y$ получаем двусторонний критерий

$$\frac{|\bar{x}-\bar{y}|}{S}\sqrt{\frac{mn}{m+n}} > \tau_\alpha,$$

где через τ_α обозначена квантиль распределения Стьюдента с $n + m - 2$ степенями свободы порядка $1 - \alpha/2$.

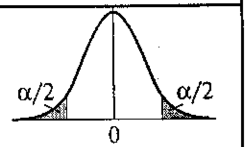
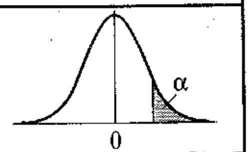
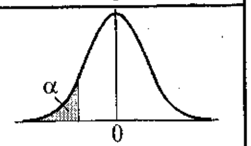
Для альтернатив H_1: $m_x > m_y$ и H_1: $m_x < m_y$ получаем односторонние критерии соответственно

$$\frac{\bar{x}-\bar{y}}{S}\sqrt{\frac{mn}{m+n}} > t_{1-\alpha}(n+m-2),$$

$$\frac{\bar{x}-\bar{y}}{S}\sqrt{\frac{mn}{m+n}} < t_{\alpha}(n+m-2).$$

Обозначив через $A = \dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n}+\dfrac{\sigma_y^2}{m}}}$ и $B = \dfrac{\bar{x}-\bar{y}}{S}\sqrt{\dfrac{mn}{m+n}}$, сведем полученные результаты в таблицу (табл. 14.2).

Таблица 14.2

H_1	Критерий		Критическая область
	σ_x^2, σ_y^2 известны	$\sigma_x^2 = \sigma_y^2$ неизвестна	
$m_x \neq m_y$	$\lvert A\rvert > u_\alpha$	$\lvert B\rvert > \tau_\alpha$	
$m_x > m_y$	$A > z_{1-\alpha}$	$B > t_{1-\alpha}(n + m - 2)$	
$m_x < m_y$	$A < z_\alpha$	$B < t_\alpha(n + m - 2)$	

Пример.

Продолжим исследование примера с длиной слов. При проведении эксперимента оказалось, что выборка

4, 1, 4, 5, 1, 13, 4, 10, 2, 4, 7, 2, 2, 4, 6, 4, 5, 6, 2, 4

естественным образом разбивается на две группы: 11 слов с длиной

4, 4, 5, 13, 4, 2, 2, 4, 6, 5, 2

взяты из первой половины книги, а остальные 9 с длиной

1, 1, 10, 4, 7, 2, 4, 6, 4

– из второй. Будем рассматривать эти наборы как две выборки ($n = 11$, $m = 9$). Выборочные средние таковы:

$$\bar{x} = 4{,}64, \quad \bar{y} = 4{,}33.$$

Несмещенные оценки дисперсии:

$$S_x^2 = 9{,}45, \quad S_y^2 = 8{,}75,$$

откуда

$$S^2 = (10S_x^2 + 8S_y^2)/18 = 6{,}57, \quad S = 2{,}56.$$

(Обратите внимание, что данная оценка дисперсии не совпадает с полученной ранее (см. с. 405). Это неудивительно, так как применялись разные методы подсчета.)

Таким образом, выборочное значение статистики критерия

$$\frac{\bar{x}-\bar{y}}{S}\sqrt{\frac{mn}{m+n}} = \frac{4{,}64-4{,}33}{2{,}56}\sqrt{99/20} \approx 0{,}269.$$

При альтернативе H_1: $m_x \neq m_y$ по таблице (см. Приложение 1) найдем

$$\tau_{0,05} = t_{0,975}(18) = 2{,}101.$$

Видно, что значение статистики не попало в критическую область и гипотеза о равенстве средних не отвергается: среднюю длину слов в обеих половинах книги можно считать одинаковой.

Проверим гипотезу при альтернативе H_1: $m_x > m_y$.

Находим, что

$$t_{0,95}(18) = 1{,}734.$$

Статистика снова оказывается в области принятия решения: нулевая гипотеза не отвергается.

14.2.3. Гипотеза о дисперсии

Пусть выборка $\{x_1, x_2, \ldots, x_n\}$ взята из нормальной генеральной совокупности. Высказывается гипотеза о значении дисперсии

$$H_0\colon \sigma^2 = \sigma_0^2.$$

Пусть s^2 – выборочная дисперсия, S^2 – несмещенная оценка дисперсии. Тогда по теореме 13.3 в предположении, что гипотеза верна, случайная величина

$$\chi^2 = n\frac{s^2}{\sigma_0^2} = (n-1)\frac{S^2}{\sigma_0^2}$$

распределена по закону χ^2 с $n - 1$ степенью свободы. Выберем эту величину в качестве статистики критерия. При заданном уровне значимости α справедливо

$$P(\chi^2_{\alpha/2}(n-1) < \chi^2 < \chi^2_{1-\alpha/2}(n-1)) = 1 - \alpha$$

(в неравенстве фигурируют соответствующие квантили распределения χ^2). Поэтому при альтернативе H_1: $\sigma^2 \neq \sigma_0^2$ критическая область, естественно, имеет вид

$$\chi^2 < \chi^2_{\alpha/2}(n-1), \quad \chi^2 > \chi^2_{1-\alpha/2}(n-1)$$

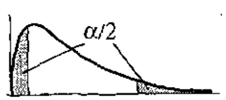
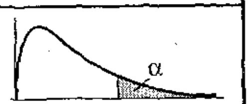
(двусторонний критерий).

При альтернативах H_1: $\sigma^2 > \sigma_0^2$ и H_1: $\sigma^2 < \sigma_0^2$ получаем односторонние критические области соответственно

$$\chi^2 > \chi^2_{1-\alpha}(n-1) \text{ и } \chi^2 < \chi^2_{\alpha}(n-1).$$

Составим таблицу результатов (табл. 14.3).

Таблица 14.3

H_1	Критерий	Критическая область
$\sigma^2 \neq \sigma_0^2$	$n\frac{s^2}{\sigma_0^2} < \chi^2_{\alpha/2}(n-1)$, $n\frac{s^2}{\sigma_0^2} > \chi^2_{1-\alpha/2}(n-1)$	
$\sigma^2 > \sigma_0^2$	$n\frac{s^2}{\sigma_0^2} > \chi^2_{1-\alpha}(n-1)$	
$\sigma^2 < \sigma_0^2$	$n\frac{s^2}{\sigma_0^2} < \chi^2_{\alpha}(n-1)$	

Пример.

Продолжим рассмотрение нашего примера с длиной слов ($n = 20$, $s^2 = 6{,}75$).

Проверим гипотезу H_0: $\sigma^2 = 7$.

Статистика критерия принимает значение

$$\chi^2 = 20\frac{6{,}75}{7} = 19{,}286.$$

Найдем в таблице (см. Приложение 1) нужные квантили при $\alpha = 0{,}1$:

$$\chi^2_{0,05}(19) = 10{,}1, \quad \chi^2_{0,95}(19) = 30{,}1.$$

Статистика не попадает в двустороннюю критическую область. Поэтому при альтернативе H_1: $\sigma^2 \neq 7$ нулевая гипотеза не отвергается.

14.3. Гипотеза о функции распределения. Критерий χ^2

Классифицируя статистические гипотезы, мы разделили их на две группы: гипотезы о параметрах и гипотезы о функции распределения. Поскольку методы проверки гипотез о параметрах хорошо развиты в основном для нормальной совокупности, вопрос о виде распределения приобретает особую важность. Прежде чем применять, скажем, критерий Стьюдента при проверке гипотезы о математическом ожидании, хорошо было бы убедиться, что выборка на самом деле распределена по нормальному закону. Закон распределения полезно знать и во многих других случаях.

Пусть $\{x_1, x_2, \ldots, x_n\}$ – выборка наблюдений некоторой случайной величины ξ. Наше предположение состоит в том, что генеральная совокупность имеет функцию распределения $F(x)$. Подобное предположение можно сделать на основе предварительного анализа выборки: построения гистограммы или полигона, частотной табуляции и т.д. Если, к примеру, частота попадания в центральные интервалы существенно выше, чем в крайние (гистограмма имеет выраженный максимум), если частота примерно монотонно убывает к краям, – тогда с достаточным основанием можно высказать предположение о нормальности распределения. Если же максимум частоты приходится на край или имеется, скажем, два выраженных «пика», то неразумно предполагать нормальность такого распределения. Итак, выдвигается гипотеза H_0: генеральная совокупность имеет функцию распределения $F(x)$ против альтернативы, что функция распределения не такова.

Проверка гипотезы H_0 осуществляется на основе частотной табуляции. Разобьем промежуток изменения значений выборки на r интервалов. Для простоты будем считать их равными. Пусть n_i, $i = 1, 2, \ldots, r$ – частота попадания значений выборки в i-й интервал

$[x^{(i)}, x^{(i+1)})$. Эту частоту называют *наблюдаемой*. Предполагая, что гипотеза H_0 верна, т.е. нам известна функция распределения генеральной совокупности, мы можем найти вероятности p_i попадания случайной величины в i-й интервал:

$$p_i = P(\xi \in [x^{(i)}, x^{(i+1)})) = F(x^{(i+1)}) - F(x^{(i)}).$$

Если, к примеру, найдено $p_i = 0{,}3$, это означает, что теоретически при большом числе испытаний примерно 30% всех значений случайной величины должны попадать в i-й интервал. Скажем, при $n = 100$ это количество составит примерно $0{,}3 \cdot 100 = 30$. Вообще, произведение np_i есть *теоретическая* (или *ожидаемая*) *частота* попадания значений случайной величины в данный интервал. Напомним, что ожидаемая частота может быть вычислена в предположении о том, что выполняется нулевая гипотеза. Мерой расхождения между истинным и теоретическим распределением генеральной совокупности может служить величина

$$\delta = \sum_{i=1}^{r} c_i (n_i - np_i)^2,$$

где коэффициенты c_i можно выбирать достаточно произвольно. Чем δ меньше, тем ближе наблюдаемые частоты к ожидаемым, тем точнее описывается истинное распределение теоретической функцией распределения $F(x)$. Данная величина вычисляется по выборке, если предположить, что гипотеза H_0 верна. Чтобы применить стандартную схему проверки гипотез, нам хотелось бы знать распределение этой случайной величины. Тогда ее можно выбрать в качестве статистики критерия. Оказывается, коэффициенты c_i можно подобрать таким образом, что распределение этой статистики при больших n примерно подчиняется закону χ^2.

Теорема 14.1 (теорема Пирсона). *Пусть* $c_i = \dfrac{1}{np_i}$. *Тогда при* $n \to \infty$ *распределение меры расхождения* δ *стремится к распределению* χ^2 *с* $r - 1$ *степенью свободы. Другими словами, при больших* n *можно считать, что*

$$\boxed{\sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(r-1)}$$

Меру расхождения в этом случае естественно обозначить через χ^2:

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}.$$

При $r = 2$ теорема, по сути, следует из теоремы Муавра – Лапласа. В этом случае каждый опыт, в результате которого получается выборочное значение, имеет два исхода: значение может попасть либо в первый, либо во второй интервал. Попадание в первый интервал можно считать успехом, попадание во второй – неудачей. Выборка, таким образом, становится результатом n испытаний Бернулли. В соответствии с этим можно обозначить частоту успехов n_1 через μ_n, вероятность успеха p_1 – через p. Тогда $n_2 = n - \mu_n$, $p_2 = 1 - p = q$.

Известно, что число успехов в n испытаниях Бернулли имеет биномиальное распределение $B_{n,p}$. Вычислим меру расхождения:

$$\chi^2 = \sum_{i=1}^{2} \frac{(n_i - np_i)^2}{np_i} = \frac{(\mu_n - np)^2}{np} + \frac{(n - \mu_n - nq)^2}{nq}.$$

После преобразований

$$n - \mu_n - nq = n - \mu_n - n(1 - p) = np - \mu_n$$

получаем

$$\chi^2 = \frac{(\mu_n - np)^2}{np} + \frac{(\mu_n - np)^2}{nq} = \frac{(\mu_n - np)^2 (q + p)}{npq} = \frac{(\mu_n - np)^2}{npq},$$

так как $q + p = 1$. Вспоминая, что

$$\mathrm{M}\mu_n = \mathrm{M}B_{n,p} = np, \qquad \mathrm{D}\mu_n = \mathrm{D}B_{n,p} = npq,$$

замечаем, что в данном случае

$$\chi^2 = (B^*_{n,p})^2,$$

где $B^*_{n,p} = \dfrac{\mu_n - np}{\sqrt{npq}}$ есть стандартизация биномиальной случайной величины. По теореме Муавра – Лапласа при $n \to \infty$ эта стандартизация стремится к стандартному нормальному распределению $N(0, 1)$. Значит, по определению ее квадрат стремится к распределению $\chi^2(1)$ с одной степенью свободы. Это и есть утверждение теоремы Пирсона при $r = 2$.

Распределение статистики χ^2 можно считать известным, если выполняется гипотеза H_0 и объем выборки n достаточно велик. «Плохими» с точки зрения соответствия нулевой гипотезе являются большие значения χ^2. Поэтому критическую область разумно выбрать на правом «хвосте» распределения $\chi^2(r - 1)$ (рис. 14.3). При заданном уровне значимости α определим квантиль $\chi^2_{1-\alpha}(r - 1)$ этого распределения порядка $1 - \alpha$. Тогда

$$P(\chi^2 > \chi^2_{1-\alpha}(r - 1)) = \alpha.$$

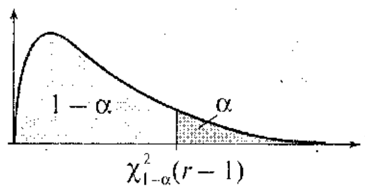


Рис. 14.3

Неравенство $\chi^2 > \chi^2_{1-\alpha}(r-1)$ является критерием при проверке гипотезы H_0 и называется ***критерием*** χ^2.

∇1. *Гипотеза H_0 о том, что генеральная совокупность имеет функцию распределения $F(x)$, не согласуется с данными эксперимента, если выборочное значение статистики χ^2 удовлетворяет неравенству*

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} > \chi^2_{1-\alpha}(r-1).$$

Чарлз Пирсон (1857–1936). Английский математик, биолог, статистик, философ. Член Лондонского королевского общества с 1896 г. Родился в Лондоне. Окончил Кембриджский университет (1875), учился в Гейдельбергском и Берлинском университетах. В 1884–1911 гг. – профессор Лондонского университетского колледжа, с 1911 г. – директор Лаборатории евгеники Лондонского университета, заслуженный профессор.

Основные работы Пирсона относятся к математической статистике. Разработал статистические методы в задачах наследственности и эволюции, критерии согласованности результатов эксперимента с теоретическими представлениями. Предложил ввести системы кривых распределения (кривые Пирсона) в качестве способа математического описания явлений природы, развивал методы оценивания параметров для решения практических задач. Написал ряд работ по истории науки. В работе В.И. Ленина «Материализм и эмпириокритицизм» упомянут как махист, отрицающий объективный характер законов природы. В 1900 г. основал журнал «Биометрика», в котором публиковались выдающиеся работы по математической статистике, в том числе работы Стьюдента.

Критерий χ^2, в отличие от рассмотренных ранее, является п р и б л и ж е н н ы м. Он применим лишь при больших объемах выборки. На практике обычно полагают $n \geqslant 50$. Кроме того, теоретические частоты не должны быть слишком малы. Считают, что выборка хорошо сгруппирована для применения критерия χ^2, если для всех интервалов $np_i \geqslant 5$. Если это условие не выполняется, соседние интервалы объединяют в один.

Запишем χ^2 в более удобном для вычислений виде:

$$\chi^2 = \sum_{i=1}^{r} \frac{n_i^2 - 2n_i np_i - n^2 p_i^2}{np_i} = \sum_{i=1}^{r} \frac{n_i^2}{np_i} - 2\sum_{i=1}^{r} n_i + n\sum_{i=1}^{r} p_i$$

(в числителе возвели разность в квадрат, а потом разделили каждое слагаемое на знаменатель). Теперь учтем очевидные равенства:

$$n_1 + n_2 + \ldots + n_r = n, \qquad p_1 + p_2 + \ldots + p_r = 1.$$

Тогда получается, что

$$\chi^2 = \sum_{i=1}^{r} \frac{n_i^2}{np_i} - n.$$

Задача. *Пусть в n испытаниях Бернулли произошло μ_n успехов. Проверить гипотезу о том, что вероятность успеха равна p (а вероятность неудачи, стало быть, $q = 1 - p$).*

Из предыдущих рассуждений ясно, что критическая область описывается в этом случае неравенством

$$\frac{(\mu_n - np)^2}{npq} > \chi^2_{1-\alpha}(1).$$

Естествоиспытатель Бюффон, о котором мы уже упоминали, бросал монету $n = 4040$ раз. При этом «орел» выпал $\mu_n = 2048$ раз, а «решка» – 1992 раза. Проверим гипотезу о том, что вероятность выпадения «орла» равна $p = {}^1/_2$. В данном случае статистика принимает значение

$$\chi^2 = \frac{(2048 - 4040 \cdot 0{,}5)^2}{4040 \cdot 0{,}5 \cdot 0{,}5} = \frac{28^2}{1010} = 0{,}776.$$

При 5-процентном уровне значимости

$$\chi^2_{0{,}95}(1) = 3{,}84.$$

Значение статистики не попало в критическую область. Поэтому делаем вывод, что гипотеза согласуется с данными эксперимента.

Проверим из интереса гипотезу о том, что вероятность выпадения «орла» равна $p = 0{,}3$ ($q = 0{,}7$). Тогда

$$\chi^2 = \frac{(2048 - 4040 \cdot 0{,}3)^2}{4040 \cdot 0{,}3 \cdot 0{,}7} = \frac{848^2}{800} = 1010{,}53,$$

что находится далеко за границей критической области. Гипотеза отвергается: она противоречит экспериментальным данным.

Выдвигая гипотезу H_0 о функции распределения генеральной совокупности, мы предполагали, что параметры распределения известны. На практике обычно бывает не так и приходится оценивать некоторые параметры по выборке. В этом случае критерий χ^2 несколько видоизменяется: число степеней свободы соответствующего распределения уменьшается на m.

Теорема 14.2 (теорема Пирсона). *Пусть* $c_i = \frac{1}{np_i}$ *и m параметров функции распределения* $F(x)$ *оцениваются по выборке. Тогда при* $n \to \infty$ *распределение меры расхождения* δ *стремится к распределению* χ^2 *с* $r - m - 1$ *степенью свободы. Другими словами, при больших n можно считать, что*

$$\boxed{\sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(r - m - 1)}$$

Значит, проверка гипотезы осуществляется в этом случае точно так же, но границей критической области является квантиль порядка $1 - \alpha$ распределения χ^2 с $r - m - 1$ степенью свободы (рис. 14.4).

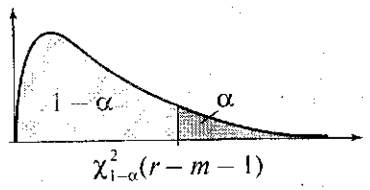


Рис. 14.4

∇**2.** *Пусть по выборке оцениваются m параметров функции распределения* $F(x)$. *Тогда гипотеза* H_0 *о том, что генеральная совокупность имеет функцию распределения* $F(x)$, *не согласуется с данными эксперимента, если выборочное значение статистики* χ^2 *удовлетворяет неравенству*

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} > \chi^2_{1-\alpha}(r - m - 1).$$

Пример.

Проверим гипотезу о том, что длина слов из нашего постоянного примера распределена по нормальному закону. Параметры распределения оцениваются по выборке:

$$\bar{x} = 4{,}5, \qquad S = 2{,}65.$$

Проверяется гипотеза H_0: $\xi \sim N(4{,}5,\ 2{,}65)$.

Оцениваются два параметра, т.е. $m = 2$. Воспользуемся результатами частотной табуляции. Промежуток изменения значений выборки разбивался на $r = 6$ интервалов. Поскольку проверяется гипотеза о нормальности, первый интервал следует расширить влево

до $-\infty$, а последний – вправо до ∞. Теоретические вероятности посчитаем по таблице значений функции $\Phi(x)$. Например,

$$p_1 = P(\xi < 3) = F_\xi(3) = \Phi\left(\frac{3 - 4{,}5}{2{,}65}\right) =$$

$$= \Phi(-0{,}566) = 1 - \Phi(0{,}566) \approx 1 - 0{,}7157 = 0{,}2843,$$

$$p_2 = P(3 \leqslant \xi < 5) = F_\xi(5) - F_\xi(3) = \Phi\left(\frac{5 - 4{,}5}{2{,}65}\right) - \Phi\left(\frac{3 - 4{,}5}{2{,}65}\right) =$$

$$= \Phi(0{,}188) - \Phi(-0{,}566) \approx 0{,}5753 - 0{,}2843 = 0{,}291$$

и т.д. Результаты принято оформлять в виде таблицы, в которой указываются интервалы разбиения, наблюдаемые и ожидаемые частоты, а также слагаемые статистики χ^2. В нашем случае получим следующую таблицу:

Номер интервала i	Границы интервала	Наблюдаемая частота n_i	Ожидаемая частота np_i
1	Менее 3	6	5,686
2	3–5	7	5,82
3	5–7	4	5,022
4	7–9	1	2,58
5	9–11	1	0,75
6	Не менее 11	1	0,142

Правильность расчета ожидаемых частот легко проверить: их сумма должна быть равна объему выборки:

$$np_1 + np_2 + \ldots + np_r = n(p_1 + p_2 + \ldots + p_r) = n \cdot 1 = n.$$

В данном случае действительно

$$5{,}686 + 5{,}82 + 5{,}022 + 2{,}58 + 0{,}75 + 0{,}142 = 20.$$

Заметим, что ожидаемые частоты в трех последних интервалах слишком малы. Поэтому нам придется объединить эти интервалы в один (при этом число интервалов разбиения становится $r = 4$):

Номер интервала i	Границы интервала	Наблюдаемая частота n_i	Ожидаемая частота np_i	Слагаемые $\frac{(n_i - np_i)^2}{np_i}$
1	Менее 3	6	5,686	0,017
2	3–5	7	5,82	0,239
3	5–7	4	5,022	0,208
4	Не менее 7	3	3,472	0,064

Суммарное значение статистики $\chi^2 = 0{,}529$. Число степеней свободы равно $r - m - 1 = 4 - 2 - 1 = 1$. Зададимся уровнем значимости $\alpha = 0{,}05$ и найдем по таблице (см. Приложение 1)

$$\chi^2_{0,95}(1) = 3{,}84.$$

Видно, что статистика не попала в критическую область. Поэтому гипотеза о нормальности не отвергается (нулевая гипотеза не противоречит экспериментальным данным). Следует, конечно, иметь в виду, что исследовалась небольшая выборка ($n = 20$) и в последнем интервале ожидаемая частота оказалась меньше 5. Если бы мы располагали выборкой бóльшего объема, то, возможно, получили бы другие результаты.

14.4. Однофакторный дисперсионный анализ

Мы уже рассматривали проблему влияния некоторого фактора на результаты измерений (вспомним оценку статистической значимости фактора). Задача интерпретировалась как проверка гипотезы о равенстве средних двух независимых выборок. Часто, однако, в соответствии с одним фактором данные можно разбить более чем на две группы. Скажем, исследуется влияние возраста на политические пристрастия населения. Ясно, что разумно разбить опрашиваемых как минимум на 4-5 возрастных групп. В этом случае (когда число выборок больше двух) применяется метод, известный под названием *«однофакторный дисперсионный анализ Р. Фишера»*.

Пусть результаты наблюдений составляют r независимых выборок (групп), полученных из r нормально распределенных генеральных совокупностей, которые имеют, вообще говоря, различные средние $m_1, m_2, \ldots, m_r$ и равные дисперсии σ^2 (дисперсия неизвестна). Каждая группа содержит n_j значений, $j = 1, 2, \ldots, r$. Общее число наблюдений равно n:

$$n_1 + n_2 + \ldots + n_r = n.$$

Проверяется гипотеза о равенстве средних во всех r выборках, т.е. H_0: $m_1 = m_2 = \ldots = m_r$.

Нулевая гипотеза является сложной: предполагается лишь, что математические ожидания совпадают (например, равны некоторому числу m), но о значении этого математического ожидания никаких предположений не делается. Альтернативная гипотеза состоит в том, что хотя бы две выборки имеют различные средние.

Обозначим i-й элемент j-й выборки через x_{ij}, $i = 1, 2, \ldots, n_j$; $j = 1, 2, \ldots, r$. В каждой выборке можно найти свое выборочное среднее $\bar{x}_j$ (его называют *групповым средним*):

$$\bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}.$$

Если объединить все r выборок в одну, определяется *общее среднее*

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{n_j} x_{ij}$$

(среднее арифметическое всех наблюдений).

Если нулевая гипотеза справедлива (математические ожидания совпадают), несмещенную оценку неизвестной дисперсии σ^2, как известно, легко получить, разделив сумму квадратов отклонений от общего среднего на $n - 1$. Основная идея дисперсионного анализа заключается в разбиении суммы квадратов отклонений на две компоненты. Одна из них характеризует рассеяние данных между группами (отклонение групповых средних от общего, обусловленное не только случайными вариациями измерений, но и влиянием группирующего фактора). Другая компонента описывает рассеяние внутри групп (отклонение значений от групповых средних, обусловленное только случайными вариациями). Если межгрупповое рассеяние существенно выше внутригруппового, достаточно уверенно можно сделать вывод о статистической значимости фактора, в соответствии с которым данные были сгруппированы.

∇1 (основное тождество дисперсионного анализа).

$$\boxed{\sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij} - \bar{x})^2 = \sum_{j=1}^{r} n_j(\bar{x}_j - \bar{x})^2 + \sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2}$$

Для проверки тождества нужно воспользоваться очевидным равенством

$$x_{ij} - \bar{x} = (\bar{x}_j - \bar{x}) + (x_{ij} - \bar{x}_j)$$

и учесть, что

$$\sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)(\bar{x}_j - \bar{x}) = 0$$

в силу определения средних $\bar{x}_j$ и $\bar{x}$. Читатель может самостоятельно провести аккуратные выкладки, подтверждающие данное тождество.

Тождество обычно записывают в виде $Q = Q_1 + Q_2$, где Q – общая сумма квадратов отклонений наблюдений от общего среднего, Q_1 – сумма квадратов отклонений групповых средних от общего среднего («сумма квадратов между группами»); Q_2 – сумма квадратов отклонений наблюдений от групповых средних («сумма квадратов внутри групп»). Итак,

> общая сумма квадратов отклонений от среднего есть сумма квадратов между группами плюс сумма квадратов внутри групп.

Дальнейшие рассуждения основаны на следующей теореме, которая приводится без доказательства.

Теорема 14.3. *Если нулевая гипотеза верна, то случайные величины Q_1/σ^2, Q_2/σ^2 имеют распределение χ^2 соответственно с $r - 1$ и $n - r$ степенями свободы.*

Теорема 14.3 позволяет получить две несмещенные оценки неизвестной дисперсии σ^2 (в дополнение к известной оценке $S^2 = \dfrac{Q}{n-1}$):

$$S_1^2 = \frac{1}{r-1}Q_1 = \frac{1}{r-1}\sum_{j=1}^{r} n_j(\bar{x}_j - \bar{x})^2,$$

$$S_2^2 = \frac{1}{n-r}Q_2 = \frac{1}{n-r}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2.$$

В самом деле, например,

$$\mathrm{M}S_1^2 = \mathrm{M}\left(\frac{Q_1}{r-1}\right) = \frac{\sigma^2}{r-1}\mathrm{M}\left(\frac{Q_1}{\sigma^2}\right) = \frac{\sigma^2}{r-1}\mathrm{M}[\chi^2(r-1)] =$$

$$= \frac{\sigma^2}{r-1}(r-1) = \sigma^2$$

(вспомнили, что математическое ожидание случайной величины χ^2 равно числу степеней свободы). Точно так же проверяется, что

$$\mathrm{M}S_2^2 = \sigma^2.$$

Оценка S_1^2 характеризует рассеяние групповых средних, обусловленное как обычными случайными вариациями, так и влиянием исследуемого фактора; оценка S_2^2 характеризует рассеяние данных внутри групп. Значительное превышение величины S_1^2 над S_2^2 можно объяснить различием средних в группах, т.е. влиянием группирующего фактора. Замечательно то, что нам известно распределение случайной величины S_1^2/S_2^2 в предположении, что нулевая гипотеза верна.

∇**2.** *Если нулевая гипотеза верна, то случайная величина $F = S_1^2/S_2^2$ распределена по закону Фишера с $r - 1$, $n - r$ степенями свободы.*

В самом деле,

$$\frac{S_1^2}{S_2^2} = \frac{Q_1/(r-1)}{Q_2/(n-r)}.$$

Разделим числитель и знаменатель на неизвестную дисперсию σ^2:

$$\frac{S_1^2}{S_2^2} = \frac{\dfrac{Q_1}{\sigma^2}\Big/(r-1)}{\dfrac{Q_2}{\sigma^2}\Big/(n-r)} = \frac{\chi^2(r-1)/(r-1)}{\chi^2(n-r)/(n-r)} \sim F(r-1,\ n-r).$$

Итак, мы оказались в условиях общей схемы принятия гипотез; найдена статистика, которую можно вычислить по выборочным данным и распределение которой известно при условии, что нулевая гипотеза справедлива. Здравый смысл подсказывает, что «плохими» с точки зрения нулевой гипотезы являются большие значения статистики F (когда оценка S_1^2 много больше оценки S_2^2). Соответственно, при заданном уровне значимости α критическая область находится на правом «хвосте» распределения Фишера, т.е. правее квантили порядка $1 - \alpha$ (рис. 14.5).

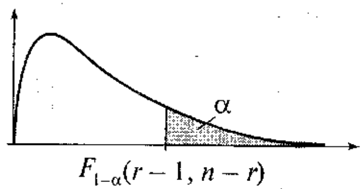


Рис. 14.5

∇**3.** *Фактор, в соответствии с которым сгруппированы данные, можно признать статистически значимым, если выборочное значение статистики F удовлетворяет неравенству*

$$F > F_{1-\alpha}(r-1,\ n-r).$$

В этом случае гипотеза о равенстве математических ожиданий не подтверждается экспериментальными данными.

Чтобы вычислить статистику F, необходимо найти числа Q_1 и Q_2. Непосредственное их вычисление по формулам достаточно громоздко. Поэтому на практике пользуются более удобными формулами, аналогичными формуле для выборочной дисперсии («средний квадрат минус квадрат среднего»). Формулы получаются, если раскрыть квадраты разности под знаком суммы. Обычно сначала вычисляют Q и Q_1, а затем $Q_2 = Q - Q_1$. Введем обозначения для сумм значений в каждой выборке

$$A_j = \sum_{i=1}^{n_j} x_{ij}$$

и для суммы всех значений

$$A = \sum_{j=1}^{r} A_j = \sum_{j=1}^{r}\sum_{i=1}^{n_j} x_{ij}.$$

Тогда формулы для вычисления Q и Q_1 имеют вид

$$Q = \sum_{j=1}^{r}\sum_{i=1}^{n_j} x_{ij}^2 - \frac{1}{n}A^2,$$

$$Q_1 = \sum_{j=1}^{r}\frac{1}{n_j}A_j^2 - \frac{1}{n}A^2.$$

Пример.

Рассмотрим пример из сборника задач по математике под редакцией А.В. Ефимова. Три группы учащихся изучали математику по различным методикам. В конце года провели тестовый контроль знаний у случайно отобранных учеников из каждой группы (соответственно 7, 5 и 3 человека). Учащиеся допустили следующее число ошибок:

1-я группа: 1, 3, 2, 1, 0, 2, 1;
2-я группа: 2, 3, 2, 1, 4;
3-я группа: 4, 5, 3.

При уровне значимости $\alpha = 0{,}05$ требуется проверить гипотезу об отсутствии влияния различных методик обучения на результаты, т.е., по существу, о равенстве математических ожиданий числа ошибок во всех трех группах:

$$H_0\colon m_1 = m_2 = m_3.$$

В нашем случае число групп $r = 3$; $n_1 = 7$, $n_2 = 5$, $n_3 = 3$; общее число испытуемых $n = n_1 + n_2 + n_3 = 15$. Вычислим вспомогательные величины:

$$A_1 = 10, \quad A_2 = 12, \quad A_3 = 12, \quad A = 34;$$

$$\sum_{j=1}^{r}\sum_{i=1}^{n_j} x_{ij}^2 = 104.$$

Значит,

$$Q = 104 - \frac{1}{15}\cdot 34^2 \approx 26{,}93;$$

$$Q_1 = \frac{1}{7}\cdot 10^2 + \frac{1}{5}\cdot 12^2 + \frac{1}{3}\cdot 12^2 - \frac{1}{15}\cdot 34^2 \approx 14{,}02;$$

$$Q_2 = Q - Q_1 = 26{,}93 - 14{,}02 = 12{,}91.$$

Выборочное значение статистики

$$F = \frac{Q_1/(r-1)}{Q_2/(n-r)} = \frac{14{,}02/2}{12{,}91/12} \approx 6{,}52.$$

Найдем нужную квантиль распределения Фишера:

$$F_{0,95}(2,\ 12) = 3{,}89.$$

Убедитесь в этом самостоятельно по более подробным таблицам, чем в Приложении 1 (см., например, задачник под ред. А.В. Ефимова).

Видно, что статистика попала правее квантили, т.е. в критическую область. Значит, гипотеза H_0 отклоняется: методика обучения значимо влияет на результат.

Задачи

1. Исследовалась длительность песен группы «Beatles». Для $n = 36$ случайно выбранных песен средняя длительность $\bar{x} = 154{,}5$ с. Предполагая, что длительность песен составляет нормальную совокупность с дисперсией $\sigma^2 = 16$ с^2, проверить гипотезу о том, что средняя длительность равна 150 с. Принять $\alpha = 0{,}01$. Проверить гипотезу при альтернативах:
   а) $m \neq 150$;
   б) $m > 150$;
   в) $m < 150$.

2. В условиях задачи 1 найти вероятность ошибки II рода при альтернативе $m > 150$, если средняя длительность фактически равна 160 с.

3. Решить задачу 1 в случае, если дисперсия σ^2 неизвестна, а выборочная дисперсия $s^2 = 25$ с^2.

4. С помощью критерия χ^2 проверить гипотезу о том, что длина слов в книге о Винни-Пухе распределена по равномерному закону на отрезке от 1 до 20 (воспользоваться экспериментальными данными из главы 13).

5. Собрать данные о росте студентов в вашей группе. Проверить гипотезу о том, что рост распределен по нормальному закону (принять $\alpha = 0{,}05$).

6. Предполагая, что рост девушек и юношей распределен по нормальному закону с одинаковыми дисперсиями, проверить гипотезу о том, что средний рост девушек и юношей одинаков (при уровне значимости $\alpha = 0{,}05$).

7. Самостоятельно собрать данные о длине слов в книгах трех писателей: К. Чуковского, А.С. Пушкина и А. Кристи. Применяя метод дисперсионного анализа, проверить гипотезу о том, что средняя длина слов у этих трех авторов одинакова ($\alpha = 0{,}01$).

Часть 6
ИНФОРМАТИКА

Вот опять окно...

М. Цветаева

Переходим к той части учебника, в которой вы познакомитесь с компьютером. Слово «информатика» (в буквальном его понимании) не совсем точно отражает ее содержание, но по сложившейся традиции так называют все, связанное с работой на компьютере. Поэтому не будем спорить о терминах, а попытаемся разобраться в существе дела.

Студенты гуманитарных факультетов, которым я читаю лекции по высшей математике, часто задают вопрос: «Зачем нам это нужно?» И я стараюсь по мере сил честно и полно обосновать необходимость изучения математики. Мои взгляды на эту проблему изложены в предисловии к данному учебнику. Говоря кратко, я вижу здесь две стороны: первая – математика есть часть культуры, одно из высших достижений цивилизации; вторая – прикладное значение математических методов в различных областях знания, не связанных непосредственно с математикой.

Если важность математики требуется доказывать, то необходимость применения компьютера, кажется, не нуждается в обосновании. За последние годы компьютер занял прочное место в нашей жизни и не собирается сдавать позиции. Наоборот, он быстро завоёвывает все новые и новые области в быту, бизнесе, науке, производстве и т.д. Еще десять лет назад мы бы не поверили, что будем встречать компьютеры буквально на каждом шагу. Поэтому, надеюсь, у вас не возникает вопроса «Зачем нужен компьютер?» Любой человек, и в частности представитель гуманитарной специальности, не может в наше время оставаться «компьютерно безграмотным». Неумение обращаться с компьютером может пагубно сказаться на профессиональной карьере специалиста в любой области.

Цели «компьютерной» части учебника можно сформулировать следующим образом. Первая задача – дать начальные сведения о компьютере, вылечить «компьютеробоязнь». Многие, особенно гуманитарии (не в обиду будет сказано), страшатся подойти к компьютеру, включить его, запустить нужную программу. «Компьютерная

15 Зак. № 2435 Турецкий

грамотность» необходима современному человеку так же, как и умение читать, писать и считать. Вторая задача — помочь специалисту-гуманитарию сориентироваться в разнообразии существующего программного обеспечения. Данный учебник не следует рассматривать как руководство для пользователя по конкретным программам (более подробно будет освещен разве что вездесущий Norton Commander). Однако вы найдете здесь классификацию наиболее часто используемых программ, их основные характеристики и возможности. Учебник, на мой взгляд, должен создать необходимую базу для дальнейшего освоения конкретных программных продуктов, дать возможность лучше разобраться в своих потребностях, выбрать нужную для работы программу.

Следует сделать одно важное замечание. Так сложилось, что, практически все компьютерные термины имеют английское происхождение. Кроме того, многие хорошие полезные программы работают «по-английски» и «не понимают» русского языка. Поэтому, чтобы освоить компьютер, нужно хотя бы на начальном уровне знать английский. Начиная преподавать информатику, я всегда обращаюсь к студентам с призывом: «Изучайте английский язык!» Знакомство с компьютером станет для вас намного проще, если английские термины или сообщения на экране не будут приводить вас в замешательство.

Глава 15 ОСНОВЫ КОМПЬЮТЕРНОЙ ГРАМОТНОСТИ

15.1. Из истории компьютера

Большинство персональных компьютеров, заполонивших российский рынок, создано могучей фирмой IBM (Ай-би-эм). Сейчас трудно себе представить, что история персонального компьютера IBM насчитывает не более двадцати лет (первый такой компьютер был создан в 1981 г.). Вычислительные же устройства (а слово «компьютер» произошло от английского compute — вычислять), появились значительно раньше. История компьютера полна драматизма, а главное — связана с именами многих великих ученых разных стран,

увлеченных идеей автоматизации вычислений. Их замыслы нашли свое воплощение лишь в XX в. при соответствующем уровне развития техники.

Очень давно возникла потребность упрощения счета. В древности для этого применяли счетные палочки, камешки и другие подобные предметы. Примерно в III в. в странах Средиземноморья появилось первое счетное устройство — *абак*, или обыкновенные счеты. Долгое время счеты оставались единственным средством быстрых расчетов.

Следующим шагом на долгом пути к компьютеру стали *костяшки Непера*, усовершенствованные счетные палочки, которые в конце жизни изобрел шотландский математик Джон Непер (1550–1617). Кстати говоря, Непер известен как изобретатель логарифмов. Его сочинение «Описание удивительной таблицы логарифмов» (1614) содержит определение логарифмов, объяснение их свойств и таблицы логарифмов. Математическое определение логарифма, данное Непером, оказало значительное влияние на развитие математического анализа.

В XVII в. идея создания вычислительного устройства овладела двумя величайшими умами. В 1644 г. Блез Паскаль создал свою суммирующую машину — «*паскалину*», представлявшую собой довольно громоздкий ящик с шестеренками. Суммирующая машина позволяла достаточно быстро складывать большие числа. Паскаль многократно пытался усовершенствовать машину и создал более пятидесяти ее вариантов. Однако ему не удалось приспособить ее для удобного умножения и деления. С этой задачей справился *калькулятор Лейбница*, изобретенный в 1673 г. С его помощью можно было выполнять все четыре арифметических действия. Кроме того, Лейбниц изобрел механическое устройство для интегрирования. Впоследствии механические калькуляторы, или *арифмометры*, нашли широкое распространение при сложных и громоздких расчетах.

Следующий важный этап истории компьютера относится уже к первой половине XIX в. и связан с именем английского математика и экономиста Чарлза Бэббиджа (1792–1871). Жизнь и труды Бэббиджа можно назвать историей великой мечты. Ему не удалось самому осуществить свои проекты — техника была еще не готова к восприятию таких смелых идей. Однако замыслы и планы Бэббиджа были почти буквально реализованы в нашем веке в виде первых электронно-вычислительных машин (ЭВМ).

В 1822 г. Бэббидж сконструировал машину для табулирования функций. С 1823 г. он работал над постройкой *разностной машины*.

Бэббидж подготовил полный комплект чертежей машины, но построил лишь ее часть. В 1834 г. Бэббидж приходит к идее универсальной *аналитической машины* – прообразу современных компьютеров. По его замыслу аналитическая машина должна была выполнять вычисления без участия человека (именно из-за ограниченной скорости работы человека (ввод данных и запись результатов) расчеты на арифмометре производились крайне медленно). В основу проекта аналитической машины Бэббиджа легли несколько идей, сыгравших основополагающую роль в развитии компьютеров. Прежде всего, эта машина должна была работать по программе, т.е. выполнять предусмотренный заранее набор инструкций. Программы и исходные данные должны были вводиться в машину с помощью перфокарт – специальных карт из плотной бумаги, информация на которых кодируется в виде отверстий (перфокарта – перфорированная карта, или карта с отверстиями). Перфокарты к этому времени уже использовались для автоматического управления ткацким станком Жаккарда (изобретенным французским инженером Ж.М. Жаккаром в 1804 г.). По мысли Бэббиджа, данные должны были обрабатываться по программе в специальном устройстве аналитической машины – «мельнице». Идея независимого устройства для вычислений – процессора – полностью осуществлена в компьютере XX в. Еще одна идея Бэббиджа, реализованная в современных ЭВМ, – наличие «склада» для запоминания данных и промежуточных результатов. Неотъемлемая часть современных компьютеров – память, прообразом которой и был «склад» аналитической машины Бэббиджа. Итак, проект аналитической машины предусматривал:

- работу по программе;
- автоматический ввод программ и данных;
- наличие процессора;
- наличие памяти.

Это характерно и для современных компьютеров.

Вся дальнейшая жизнь Бэббиджа была, по существу, посвящена усовершенствованию проекта аналитической машины. Проект оказался слишком сложным для технического воплощения средствами того времени. В работе Бэббиджу помогала дочь Байрона, графиня Ада Лавлейс (1815–1852). В 1842 г. она написала работу, в которой развила основы программирования для аналитической машины, и вполне заслуженно считается первым программистом. В ее честь назван один из языков программирования – Ада.

В жизни Чарлз Бэббидж считался чудаком с неуживчивым характером. Он боролся за математическую строгость, выискивал неточности в работах современных математиков. Много времени и сил он посвятил безуспешной борьбе с шарманщиками на английских улицах. Тем не менее он завоевал известность не только своим неосуществленным проектом аналитической машины, но и важными математическими результатами (в частности, в области теории вероятностей и математической статистики). Многие годы он заведовал кафедрой математики в Кембриджском университете. Невозможно переоценить важность идей Бэббиджа в создании и развитии компьютеров.

Некоторые части проекта Бэббиджа были реализованы еще в XIX в. Его сын, Г.П. Бэббидж, построил малую аналитическую машину, которая хранится в Лондонском музее наук. Шведский изобретатель П.Г. Шютц построил по идеям Бэббиджа разностную машину (1853), которая завоевала золотую медаль на Всемирной выставке в Париже, а впоследствии была приобретена Дудлеевской обсерваторией (Олбани, США), где успешно работала до 1924 г. Однако в полной мере идеи Бэббиджа были воплощены в жизнь лишь в нашем веке.

Итак, начался отсчет времени до создания настоящего компьютера. Параллельное развитие техники и науки готовило для этого базу. В 1890 г. появился *табулятор Холлерита*, устройство для автоматической обработки информации на перфокартах. Табулятор успешно применяли для обработки результатов переписи населения США. По сути дела, это было готовое устройство для ввода информации с перфокарт. Табуляторы и арифмометры были распространены очень широко, и в 1924 г. была образована специальная фирма по их разработке и продаже, получившая название IBM (International Business Machines). Именно в этой фирме осуществились мечты Чарлза Бэббиджа. Постепенно развиваясь, фирма IBM превратилась в наши дни в настоящего компьютерного гиганта.

Математическую основу будущего компьютера составила логика Джорджа Буля. Оказалось, что булева двоичная логика (т.е. логика с двумя возможными состояниями – истиной и ложью) очень хорошо применима для описания электротехнических переключательных систем. Ведь переключатель (триггер) также может находиться только в двух состояниях – выключенном и включенном. Впервые связь логики и электротехники была отмечена в докторской диссертации выдающегося американского математика Клода Шеннона, создателя теории информации. Эта работа была написана в 1936 г. Тогда же математические основы работы автоматических устройств были заложены в трудах английского математика Алана Тьюринга (1912–1954). «Машина Тьюринга» – это мысленное устройство для доказательства теорем, автомат, работающий по законам булевой логики. Тьюринг принимал самое активное участие в английских компьютерных разработках в годы войны. Сходными вопросами (передача, обработка и хранение информации) интересовался и великий Норберт Винер (1894–1964), основоположник кибернетики. Математические работы Шеннона, Тьюринга, Винера и других ученых стали научной основой автоматических ЭВМ, а электрический переключатель, триггер, – основным их элементом.

Непосредственным же толчком к созданию компьютера стала, как это ни печально, вторая мировая война. Потребности фронта (баллистические расчеты, расшифровка радиосообщений противника и т.д.) сделали насущной необходимостью создание быстродействующего автоматического устройства для счета. Современные технические возможности позволили создать на основе идей Бэббиджа первые ЭВМ по обе стороны линии фронта. В 1941 г. немецкий инженер Конрад Цузе, по сути, «переоткрыл» идеи Бэббиджа и создал первую ЭВМ. Аналогичная работа велась в Англии и США. В 1943 г. создана первая английская ЭВМ «Колосс», которая успешно использовалась для расшифровки вражеских кодированных сообщений. Тогда же американец Говард Эйкен на одном из предприятий фирмы IBM построил машину «Марк-1». Интересно, что в своей работе Эйкен пользовался чертежами аналитической машины Чарлза Бэббиджа. Так круг замкнулся и произошла связь времен. Упомянутые машины в качестве основного элемента (переключателя) использовали электромеханическое реле. Поэтому они были громоздки и обладали небольшим быстродействием.

Над созданием машин, подобных машинам Цузе и Эйкена, работали одновременно несколько групп исследователей. Начиная с 1943 г. группа специалистов под руководством Джона Мочли и Преспера Экерта начала конструировать ЭВМ уже на основе электронных ламп, а не реле (электронная лампа, как известно, также может находиться лишь в двух состояниях: в одну сторону ток проходит, лампа открыта, а в другую – не проходит, лампа закрыта). Работа была успешно завершена к 1946 г. созданием машины ENIAC (Electronic Numerical Integrator And Computer, т.е. электронный цифровой интегратор и вычислитель). Эта машина была по размерам вдвое больше, чем «Марк-1» (6 м высотой, 16 м длиной), но зато работала в тысячи раз быстрее за счет применения электронных

ламп. Правда, для того, чтобы задать ей программу, приходилось в течение нескольких часов или даже нескольких дней подсоединять нужным образом провода. Следующая модель Мочли и Экерта – машина EDVAC – уже хранит программу в своей памяти.

В 1945 г. к компьютерным разработкам был привлечен известный математик Джон фон Нейман, который подготовил доклад о машине EDVAC. Доклад получил широкую известность, так как в нем фон Нейман четко сформулировал основные принципы работы компьютера как универсального устройства для обработки информации. В дальнейшем принципы фон Неймана были воплощены в конкретных машинах. Можно сказать, что конструкция современного компьютера реализует основные принципы, отмеченные в докладе 1945 г. Так называемая а р х и т е к т у р а фон Неймана предполагает обязательное наличие в компьютере следующих устройств:

- арифметическо-логического устройства (для выполнения соответствующих операций);
- устройства управления (для обеспечения выполнения программ);
- памяти (для хранения программ и данных);
- внешних устройств (для ввода и вывода информации).

Видно, что принципы фон Неймана – это во многом переосмысленные идеи Бэббиджа. В современном компьютере арифметическо-логическое и управляющее устройства часто совмещены в виде центрального процессора. Тем не менее в общих чертах компьютер в наши дни имеет архитектуру фон Неймана.

Дальнейшее развитие компьютера происходило лавинообразно. В 1949 г. английский ученый Морис Уилкс построил первый компьютер в соответствии с принципами фон Неймана – машину EDSAC. В 1951 г. создан первый коммерческий американский компьютер LEO. В 1952 г. с помощью мощного компьютера UNIVAC известный политический обозреватель Уолтер Кронкайт предсказал итоги президентских выборов в США.

Работа по созданию компьютеров началась и в СССР. Под руководством академика С.А. Лебедева в 1951 г. в Киеве разработана первая в Советском Союзе ЭВМ под названием МЭСМ (малая электронно-счетная машина), а в Москве – БЭСМ (1952). Машина БЭСМ-6 этой серии не уступала лучшим зарубежным ЭВМ. Ее быстродействие достигало миллиона операций в секунду. К сожалению, впоследствии советские разработки копировали американские образцы (машины серий СМ ЭВМ и ЕС ЭВМ), что привело к отставанию в этой области.

Компьютерная революция была подготовлена стремительным развитием техники. В 1947 г. был изобретен транзистор, который заменил электронную лампу. Это позволило создавать компьютеры той же и большей мощности, но гораздо меньших размеров. Постепенное удешевление транзисторов и изобретение магнитной памяти позволяют наладить коммерческий выпуск компьютеров. Первый коммерческий мини-компьютер PDP-8 стоил 20 000 долларов и был размером с холодильник. В 1956–1959 гг. американские ученые Джек Килби и Роберт Нойс изобрели *интегральную схему*, или чип (chip): тысячи транзисторов удалось разместить на одном кремниевом кристалле. Компьютеры на интегральных схемах становятся все более миниатюрными и быстродействующими. В воздухе 70-х годов носится идея персонального компьютера, т.е. удобного настольного компьютера, предназначенного для одного пользователя. В 1971 г. Эдвард Хофф создает первый микропроцессор Intel-4004, т.е. интегральную схему, заменяющую собой центральный процессор компьютера. На базе его улучшенной модификации Intel-8080 в 1974 г. Эд Робертс строит первый микрокомпьютер «Альтаир». Несмотря на то что «Альтаир» работал без клавиатуры, без экрана и имел небольшую память, это был огромный шаг вперед. Компьютер, стоивший всего 397 долларов, стал доступен широкому кругу любителей. В первые же месяцы было продано несколько тысяч комплектов машины.

Сотни молодых людей увлеклись идеей усовершенствования микрокомпьютера. Покупатели снабжали «Альтаир» клавиатурой, дисплеем, увеличивали память. Начали издавать журналы для любителей компьютера, разрабатывались полезные программы. В магазинах стали продавать детали, из которых легко можно было собрать компьютер самостоятельно. Начался компьютерный бум. В 1975 г. Пол Аллен и Билл Гейтс разработали язык программирования для «Альтаира» – знаменитый Бейсик, значительно упростивший общение с компьютером. Впоследствии они основали фирму Microsoft, ставшую еще одним компьютерным гигантом. Программы этой фирмы (в частности, MS DOS и Windows, с которыми мы ознакомимся чуть позже) занимают сейчас ведущее место на рынке программных продуктов. В 1977 г. два американских студента Стефен Возняк и Стивен Джобс собрали в гараже первый настоящий персональный компьютер «Эппл». Компьютер, оснащенный клавиатурой и экраном дисплея, обладал гораздо большими возможностями, чем его предшественник «Альтаир». Молодые люди быстро стали миллионерами. В корпорации IBM решили начать коммерческий выпуск

персональных компьютеров. Группе, занимавшейся разработкой компьютера, была предоставлена полная свобода в использовании любых деталей, производимых другими фирмами. В августе 1981 г. фирма IBM выпустила свой первый персональный компьютер IBM PC на базе самого мощного в то время процессора Intel-8088. Программное обеспечение было поручено молодой фирме Microsoft. Новый компьютер быстро завоевал популярность и рынок. Эра персональных компьютеров началась!

15.2. Зачем нужен компьютер?

И все-таки – зачем? Компьютер создавался для ускорения и упрощения вычислений. Безусловно, и в наши дни эта его функция остается одной из основных. Однако постепенно компьютер начал играть гораздо более важную роль, чем просто устройство для быстрого счета. Возможности современного компьютера кажутся фантастичными. Тем не менее это реальность нашего времени и нашей повседневной жизни.

Говоря общими словами, современный компьютер представляет собой **автоматическое устройство для обработки информации**. На вход компьютера через внешние устройства ввода поступает самая разнообразная информация: это могут быть числа, тексты, графическое изображение, звук и т.д. Затем информация кодируется, преобразуется в числовой вид – ведь компьютер может работать только с числами. С помощью различных программ процессор компьютера обрабатывает введенную информацию. Последний этап – вывод информации в удобном и понятном для человека виде (распечатанный текст или рисунок, график на экране и т.д.), осуществляемый посредством разнообразных внешних устройств.

Охарактеризуем кратко основные сферы применения персонального компьютера.

1. Подготовка текстов. Это, пожалуй, самый распространенный вид работы на компьютере, с которым сталкиваются так называемые непрофессиональные пользователи (те, чья профессия не связана с компьютером, т.е. большинство работающих на компьютере). С помощью специальных программ (*текстовых редакторов*) документ проходит полный цикл обработки: от написания до редактирования и красивой печати с использованием различных шрифтов и других полиграфических достижений. Для высококачественной печати применяются специальные программы – *издательские системы*.

2. Обработка больших массивов данных. В современной науке, бизнесе, производстве приходится иметь дело с огромным объемом информации. Это могут быть результаты социологического опроса, сведения о спросе, предложении и ценах на товары, данные о состоянии больного, учетная информация о сотрудниках фирмы и многое другое. Чтобы извлечь из этого потока информации рациональное зерно, не обойтись без компьютера. Для хранения и обработки больших массивов информации используются специальные программы — *системы управления базами данных* и *электронные таблицы.*

3. Образование. Компьютер нашел широкое применение в школе и вузе. С каждым годом растет количество *обучающих программ* по самым разным предметам. Хорошая обучающая программа отличается от учебника тем, что дает возможность ученику не только пассивно получать сведения, но и в какой-то мере участвовать в процессе обучения, проводить самостоятельные эксперименты и расчеты, проверять свои знания с помощью тестов. Обучающая программа использует все возможности компьютера — цвет, графику, мультипликацию, звук, видеоматериалы и т.д. В последнее время в виде компьютерных программ издаются различные справочники, словари и даже энциклопедии: если вы решили прочитать статью, скажем, о президенте Кеннеди, вы не только увидите его портрет, но и сможете услышать его речи, увидеть небольшой фильм о нем и быстро получить справки о связанных с ним людях и событиях. Такие программы, сочетающие в себе различные средства общения, называются программами *мультимедиа.*

4. Связь. Благодаря компьютеру современный мир стал поистине миром без границ. С помощью компьютеров, объединенных в локальные и глобальные сети, осуществляется быстрая и качественная связь практически на любое расстояние. При этом используются спутники связи и другие технические достижения. Самый распространенный вид компьютерной связи — *электронная почта.* Со своего компьютера вы можете послать письмо в любую точку планеты и получить ответ в тот же день. Работа ученых, бизнесменов и людей других специальностей становится гораздо эффективнее благодаря быстрому обмену информацией по электронной почте. Никакой тоталитарный политический режим уже не в состоянии перекрыть этот канал общения. Знаком нашего времени стал *Internet* (Интернет) — глобальная информационная сеть, с помощью которой можно получать любую нужную информацию из любого удаленного источника. Через Интернет вы можете получить каталоги или статьи многих мировых научных журналов, ознакомиться с фондами крупнейших библиотек, получить какие угодно интересующие вас сведения. Теперь в Интернете можно прочитать новый роман, который еще не успели издать, узнать последние новости в области кино, моды, медицины. Перечисление возможных тем можно продолжать до бесконечности. Каждый пользователь, имеющий доступ в Интернет, может найти в этой громадной сети что-нибудь интересное для себя. Россия активно включается в мировой информационный процесс. Успешно реализуется программа подключения к Интернету российских университетов. Еще одна область применения компьютерной связи — *дистанционное обучение.* Не выезжая из страны, вы можете учиться, например, в одном из университетов Калифорнии, сдавать экзамены и зачеты, получить соответствующий диплом.

5. Игры и развлечения. С этой областью применения компьютера вы, безусловно, знакомы. Любой современный школьник хоть раз да играл в какую-нибудь увлекательную компьютерную игру. На самом деле, в индустрии компьютерных игр заняты крупнейшие фирмы, игры разрабатываются и программируются большими коллективами специалистов. Задача взрослых — соблюдать меру, не дать развиться увлечению компьютерными играми в страсть, поглощающую все время ребенка. Чрезмерное увлечение играми на компьютере (как, впрочем, и вообще компьютером) имеет негативные последствия для здоровья и психики. Однако в том, что современные дети (в отличие от их родителей) не боятся компьютера, большую роль сыграли профессионально сделанные компьютерные игры. Через игру компьютер становится частью жизни ребенка. Это помогает ему в дальнейшем решать с помощью компьютера серьезные взрослые задачи. Компьютерные игры получили новое развитие в так называемых системах *виртуальной реальности,* с помощью которых вы можете ощутить себя средневековым рыцарем, космическим путешественником или детективом, расследующим запутанное преступление.

6. Медицина. Все более активно компьютеры применяются в медицинских обследованиях. Примером служит компьютерная томография — исследование органов, позволяющее выявить патологию на ранних стадиях заболевания. Развивается компьютерная расшифровка кардиограмм, компьютер применяется также для других исследований сердца и мозга. В медицине находят применение *экспертные системы,* с помощью которых диагностируют заболевания по набору признаков и данных, выявляют нездоровые тенденции и т.д. Следует отметить, что диагностические системы работают эффективно только с участием опытных врачей и специалистов.

Невозможно переоценить роль компьютера и во многих других областях человеческой деятельности – **науке, бизнесе, производстве** и т.д. Существует много различных программ. Это и специальные программы, помогающие ученым различных, в том числе и гуманитарных, специальностей. Это и бухгалтерские программы, применяемые сейчас практически повсеместно. Это и программы автоматического управления станками, приборами и целым производством. Более знакомый пример – системы управления климатом, компьютерные кондиционеры, поддерживающие нужный режим влажности и температуры в помещениях. Можно еще долго перечислять сферы применения компьютера – от торговли до управления государством. Ясно только, что компьютер стал неотъемлемой, незаменимой частью современной жизни. Разумеется, это приводит к необходимости освоения компьютера самыми «широкими слоями населения», включая и студентов-гуманитариев.

15.3. Устройство персонального компьютера

Среди непрофессиональных пользователей зачастую бытует некое мистическое отношение к компьютеру. Это отражается в таких выражениях, как «компьютер решил», «компьютер нашел», «компьютер предсказал» и т.д. Следует иметь в виду, что компьютер не работает сам по себе, он лишь строго выполняет инструкции программ, написанных программистами. После того как специальным образом закодированная программа вводится в память компьютера, процессор последовательно считывает и выполняет команды программы. При этом не обязательно выполняется команда из следующей ячейки памяти: возможен переход к командам из других ячеек или к командам, уже выполнявшимся ранее (цикл). Все эти разветвления предусматривает при написании программы программист. Таким образом, компьютер представляет собой всего лишь послушный инструмент, полностью подчиненный введенной программе. Словосочетание «искусственный интеллект» характеризует скорее сложность программы, высокий уровень естественного человеческого интеллекта, заставляющего компьютер решать непростые задачи.

Итак, не будем бояться компьютера и подойдем к нему поближе. Мы непременно увидим три устройства. Первое – это «ящик», расположенный горизонтально или вертикально, с тумблером для включения компьютера. «Ящик» называется *системным блоком* и содержит основные элементы компьютера, его «сердце» и «мозг» – процессор, память и многое другое. Мы обязательно увидим также еще два устройства: клавиатуру – устройство для ввода информации и дисплей – экран для вывода информации в виде текста или рисунков. Клавиатура и дисплей относятся к большой группе *периферийных устройств* компьютера, которые предназначены для ввода и вывода. Итак, устройства, составляющие компьютер, делятся на две группы:

- системный блок;
- периферийные устройства.

Далее мы ознакомимся также с основными компьютерными понятиями – «единица информации», «файловая система» и т.д.

15.3.1. Системный блок

Откроем кожух системного блока и заглянем внутрь. Мы увидим множество разных деталей, расположенных на специальных пластинах, или платах. Самая большая плата называется материнской. На ней-то и находится «сердце» компьютера – ***процессор***. Процессор – это большая интегральная схема (БИС), состоящая из сотен тысяч (или даже миллионов) транзисторов и соединений. На самом деле процессор вовсе не большой, и это громадное количество деталей умещается на кремниевой пластинке размером с ноготь.

Двоичная система счисления и единицы информации. Вспомним, что Клод Шеннон и Алан Тьюринг обнаружили связь между булевой двоичной логикой и электрическими цепями, составленными из переключателей. В частности, схемы из переключателей хорошо моделируют основные логические операции – конъюнкцию, дизъюнкцию, отрицание и т.д. Именно поэтому переключатель, триггер, и стал основным элементом компьютера. Сначала это были электромеханические переключатели, затем – электронные лампы, транзисторы и, наконец, интегральные схемы, чипы. По мере развития компьютеров стало ясно, что с помощью триггеров удобно осуществлять не только логические, но и арифметические операции. Одно из состояний триггера можно обозначить единицей (логическая истина), другое – нулем (ложь). Оказывается, любое число можно изобразить с помощью этих двух цифр. Для этого идеально подходит **двоичная система счисления**. Двоичная система имеет всего две цифры – единицу и нуль, основание этой системы – двойка. Каждый разряд двоичного числа соответствует степени двойки: последняя цифра целого числа – это количество единиц (2^0), предпоследняя – количество двоек (2^1), затем идут четверки, восьмерки и т.д. Например, двоичное число 1101 обозначает десятичное число 13:

$$1101 = 1 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3 = 1 + 4 + 8 = 13.$$

Создатели компьютера отказались от привычной нам десятичной системы, так как работа с десятичными числами была медленной и неэффективной. Все числа в компьютере закодированы в двоичном виде.

Как известно, переключатель может находиться только в двух возможных состояниях. Эти два состояния – истина и ложь – соответствуют минимальному количеству информации, которое можно получить о каком-либо событии. Действительно, самое меньшее, что можно узнать о событии, – это сведения о том, произошло оно или нет. Говорят, что при этом мы получаем один *бит* информации. Английское слово «bit» означает «кусочек», небольшое количество. Кроме того, это сокращение от binary digit, что значит «двоичная единица». Процессор считывает из памяти биты информации, кодирующие программу и числовые данные. Однако процессор перерабатывает информацию не по одному биту, а порциями. Поэтому вводятся более крупные единицы информации. Информация, содержащаяся в восьми битах, называется *байтом*. Таким образом, байт – это произвольный набор из восьми единиц и нулей, от 00000000 до 11111111. На каждом месте может быть либо нуль, либо единица. Поэтому всего существует $2^8 = 256$ различных байтов.

Для работы с текстами каждый символ должен быть представлен числом. Каждый символ в компьютере кодируется с помощью одного байта. Буквам латинского и русского алфавитов (большим и малым), цифрам, знакам препинания и арифметических действий, а также некоторым вспомогательным символам соответствует вполне определенный набор из восьми битов, т.е. вполне определенное число от 0 до 256. Например, цифра «1» кодируется байтом 00110001 (это число 49), большая латинская буква «А» – байтом 01000001 (65), а малая «а» – байтом 01100001 (97). Чтобы определить, какой символ закодирован данным байтом, применяются различные таблицы кодировки. Байты от 0 до 127 задают символы в соответствии с международным стандартом, это так называемая таблица ASCII-кодов. Символы, закодированные байтами от 128 до 256, могут варьироваться. В частности, этими байтами кодируются буквы национальных алфавитов в разных странах. В России существует свой стандарт кодировки символов. Первая половина этой таблицы

совпадает с таблицей ASCII-кодов, вторая содержит русские буквы и некоторые другие символы.

Наряду с байтом определяются следующие единицы информации: 1024 байта составляют один *килобайт* (1 Кбайт). На стандартной странице текста содержится примерно 2 килобайта информации. Еще более крупные единицы: 1 *мегабайт* (Мбайт) – это 1024 килобайта, 1 *гигабайт* (Гбайт) – 1024 мегабайта. Мегабайт можно представить себе как книгу из 500 страниц, а гигабайт – это целая библиотека. Итак,

1 байт	=	8 бит
1 Кбайт	=	1024 байт
1 Мбайт	=	1024 Кбайт
1 Гбайт	=	1024 Мбайт

Процессор обрабатывает информацию порциями, как бы откусывая по кусочкам. Недаром английское слово «byte» (байт) созвучно слову «bite» (кусать). Наименьшая порция битов, которую может обработать процессор за один такт своей работы, называется *разрядностью* процессора. Разрядность – важная характеристика мощности процессора. Первые процессоры были 8-разрядными, т.е. «откусывали» за раз по одному байту, затем появились 16- и 32-разрядные процессоры, обрабатывающие порции по 2 и 4 байта соответственно. Важно не только, какую порцию битов может переработать процессор за один такт, но и сколько таких тактов он совершает за одну секунду. Чем больше эта величина, тем более быстродействующим является процессор, тем больше операций он выполнит за данное время. Количество тактов в секунду называется *тактовой частотой* процессора. Тактовая частота обычно измеряется в мегагерцах (количество единиц в секунду).

Этапы развития персональных компьютеров совпадают с созданием новых поколений процессоров, обладающих все большей разрядностью и тактовой частотой. Первые персональные компьютеры IBM PC и их модификация IBM PC/XT работали на базе 8-разрядного процессора Intel-8088. Затем он был заменен более совершенным процессором Intel-8086 с тактовой частотой 4,77 МГц: эти процессоры за одну секунду обрабатывали 4,77 миллиона порций информации в один байт. Несколько позже появились компьютеры, в которых тот же процессор работал с частотой 8 и 10 МГц. В 1985 г. был создан новый компьютер IBM PC/AT на базе процессора следующего поколения – Intel-80286. Это был уже 16-разрядный процессор с тактовой частотой от 10 до 25 МГц. Его производительность в десятки раз превышала мощность первых персональных компьютеров. Однако в настоящее время этот процессор устарел и снят с производства. Следует отметить, что процесс морального и физического устаревания компьютерной техники происходит стремительно – то, что еще недавно казалось передовым и высокоэффективным, перестает удовлетворять новым требованиям, не может работать с новыми программами. Затем появился 32-разрядный процессор 80386 (так называемая «тройка») с тактовой частотой от 25 до 40 МГц. Но и эта колоссальная скорость оказалась недостаточной для программы Windows фирмы Microsoft, завоевавшей рынок в 1990–1991 гг. Поэтому был разработан процессор 80486 («четверка»), работавший в 2-3 раза быстрее. Однако и это еще не предел: в 1993 г. фирма Intel выпускает процессор Pentium (собственное имя для 80586) с тактовой частотой от 60 до 100 МГц. Современный Pentium работает с частотой 166–200 МГц и выше.

Правильный выбор компьютера определяется теми задачами, которые вы собираетесь решать. Для простой обработки текстов вполне достаточно «четверки» и даже «тройки». Если же вы работаете с новыми программами, в частности в среде Windows 95, вам потребуется компьютер с процессором не хуже чем Pentium.

Кроме процессора на материнской плате находятся интегральные схемы ***памяти*** компьютера, предназначенной для хранения программ и данных. Эта память недолговечна: вся информация стирается из памяти после выключения компьютера. Поэтому память называется *оперативной*. Объем оперативной памяти также является важной характеристикой качества компьютера. Первый микрокомпьютер «Альтаир» обладал оперативной памятью объемом всего в 256 байт. В современных персональных компьютерах объем оперативной памяти достигает десятков мегабайтов. Для выполнения простых функций – редактирование текстов, простейшие банковские программы – достаточно иметь один мегабайт оперативной памяти. При работе в системе Windows 95 требуется не меньше 8 Мбайт, а лучше – 16 или 32 Мбайт.

Итак, оперативная память хранит данные только на время сеанса – от включения до выключения компьютера. Разумеется, хотелось бы иметь возможность хранить информацию (программы и данные) длительное время. Для этого предназначены ***магнитные диски***. Устройства для работы с дисками – *дисководы* – также находятся в системном блоке. Магнитные диски бывают двух разновидностей – *гибкие* и *жесткие*.

Гибкие магнитные диски (*дискеты*) различаются по диаметру: побольше, 5,25-дюймовые, и поменьше, диаметром 3,5 дюйма. Стандартная дискета диаметром 5,25 дюйма может хранить 1,2 Мбайт

информации, а 3,5-дюймовая – 1,44 Мбайт. Таким образом, на одной дискете может поместиться текст достаточно толстой книги. В системном блоке обычно есть два дисковода для дискет разного диаметра. Дискеты позволяют переносить программы и документы с одного компьютера на другой и хранить информацию, не используемую постоянно на компьютере. Набор дискет – это архив постоянного пользователя компьютера.

Жесткий магнитный диск (*винчестер*) предназначен для постоянного хранения информации, используемой при работе с компьютером: системных программ, часто применяемого программного обеспечения. С винчестером значительно удобнее работать на персональном компьютере. Важная характеристика компьютера – емкость его винчестера. Компьютеры с процессором 80386 имеют винчестеры емкостью от 40 до 120 Мбайт, «четверки» – от 120 до 540 Мбайт. Современные стандартные винчестеры хранят колоссальный объем информации до 4 Гбайт. Еще более мощные диски применяются при работе в сетях.

В каком же виде хранится информация на магнитных дисках?

Файловая система. Итак, нам уже известно, что процессор обрабатывает информацию порциями байтов. Вспомним, однако, какое громадное количество байтов может уместиться, к примеру, на жестком магнитном диске, – это многие тысячи и миллионы 8-битовых «кусочков». Как не запутаться и не заблудиться в таком обилии информации? Это напоминает стол, заваленный бумагами и бумажками, – здесь все вперемешку: конспекты лекций, телефонные счета, чьи-то адреса и телефоны, рисунки, любовные записки и т.д. Нужно срочно навести порядок! Простейший способ – сгруппировать бумаги по темам и разложить в отдельные папки. Именно этот принцип и применяется для организации информации на магнитных дисках и в памяти компьютера. Байты группируются по назначению (вот эти байты составляют программу редактирования текстов, а эти байты – сам текст), сгруппированные байты образуют так называемые *файлы*. Происхождение этого слова вполне понятно: по-английски file означает «папка для бумаг», «досье». Как и в папках, в файлах размещаются подборки однотипной информации. Это позволяет избежать хаоса и довольно легко обращаться с имеющейся информацией.

Таким образом, информационной единицей хранения информации на диске и в памяти является файл. В отдельных файлах находятся исполняемые программы (набор закодированных инструкций), тексты документов и тексты программ, написанные на каком-либо языке программирования, числовые данные, таблицы, описания рисунков и т.д. Компьютер обращается с этими упорядоченными наборами данных как с единым целым. Самое важное, что имеется возможность копировать, переносить файлы с одной дискеты на другую, с дискеты на винчестер, с магнитного диска в память и т.д. При этом сохраняется целостность информации, не теряется ее смысловая организация.

Работать с папками для бумаг не очень удобно, если они не подписаны. Аккуратный работник офиса обязательно надпишет на папке, какие сведения в ней содержатся, когда она заведена, сколько в ней листов. Точно так же обстоит дело и с файлами. Удобство работы с ними обусловлено тем, что каждый файл имеет собственное **имя**. Имя файла позволяет без труда отыскать нужную программу, таблицу или письмо. Имя файла состоит из двух частей – это собственно *имя* и *расширение*. Между именем и расширением файла ставится точка, например, «myfile.txt» или «abcdefgh.xyz».

Имя файла может содержать от одного до восьми символов (но не меньше одного – безымянных файлов не бывает!). В имени могут использоваться латинские буквы, цифры, а также символы «$» (знак доллара), «#» («решетка»), «&» (логическое «и», другое название этого символа – «амперсенд»), «@» (этот знак называют поразному: «черепашка», «собачка», строгое название – коммерческое at, знак принадлежности фирме), «!», «%», «-» (дефис), «_» (знак подчеркивания), скобки «(» и «)». Компьютер не различает в имени файла строчные и прописные буквы. Поэтому можно пользоваться наравне и теми и другими. Нельзя пользоваться символами «.», «,», «:», «;», «?», «\», «/», «*», «<», «>», «|», «+», «=» и знаком пробела (это значит, что не может быть имени файла, состоящего из двух и более слов). Не рекомендуется пользоваться в именах файлов русскими буквами – многие программы не воспринимают русских имен. Скажем сразу, что в системе Windows 95 нет ограничений на длину имени файла и допустимые символы: имя файла может быть длинным, содержать пробелы и русские буквы.

Вторая часть имени файла – расширение – обычно указывает, к какому типу принадлежит данный файл. Расширение может содержать до трех символов. В отличие от имени расширение может и вовсе отсутствовать. Однако работать с файлами, имеющими расширения, значительно удобнее. В расширении можно пользоваться теми же символами, что и в имени файла. Культура работы на компьютере требует, чтобы расширение файла несло информацию о типе

этого файла. Например, сразу ясно, что файл с расширением «.txt» содержит текст какого-то документа (это так называемый *текстовый* файл). То же можно сказать и про файл с расширением «.doc». Файл с расширением «.sys» – системный, предназначен для общего обслуживания работы компьютера. Среди стандартных расширений необходимо запомнить три: это расширения **«.exe»**, **«.com»**, **«.bat»**. Файлы с этими расширениями – это *исполняемые* файлы. Как только в оперативную память попадает файл с таким расширением, процессор немедленно начинает выполнять содержащиеся в нем инструкции.

На диске может находиться большое количество файлов, различных по содержанию и назначению. Чтобы избежать путаницы, файлы объединяются в **каталоги** (как книги в библиотеке). Каталог обычно содержит файлы, близкие по назначению, скажем, исполняемый файл какой-либо программы и вспомогательные файлы с данными для этой программы. По-английски каталог называется directory, поэтому слово «*директория*» также часто употребляется в компьютерной литературе. Каталог имеет имя длиной до восьми разрешенных символов. У каталога также может быть расширение, однако имена каталогов с расширениями используются довольно редко. Каталог может содержать подкаталоги, те, в свою очередь, – свои подкаталоги и т.д. Каталог, к которому принадлежит данный каталог, называется по отношению к нему *родительским* и может обозначаться двумя точками: «..». На диске всегда имеется каталог, не имеющий родительского. Такой каталог называется *корневым*. Корневой каталог обозначается буквой латинского алфавита с двоеточием. Стандартные обозначения корневых каталогов таковы:

на 5,25-дюймовой дискете – А:

на 3,5-дюймовой дискете – В:

на винчестере – С:

Следует отметить, что жесткий диск может быть разделен на несколько *логических дисков*. Все они расположены на одном винчестере, но для компьютера представляют собой как бы независимые диски. Они обозначаются буквами С:, D:, Е: и т.д.

Таким образом, на каждом диске есть корневой каталог, содержащий файлы и каталоги, которые, в свою очередь, могут содержать файлы и каталоги, и т.д. Говорят, что на диске имеется *дерево каталогов* (растущее из корневого каталога диска). Структура дерева каталогов может быть сколь угодно сложной и разветвленной. С помощью каталогов и подкаталогов файлы на диске строго упорядочены, каждый файл, как и карточка в библиотеке, находится в отведенном для него «ящичке». Дерево каталогов позволяет точно указать местонахождение любого файла. Предположим, вы занимаетесь исследованием биографий поэтов. На диске D: в каталоге USERS заведен каталог IVANOV (с вашей фамилией), в котором есть подкаталог POETS. В этом каталоге вы завели отдельные каталоги для каждого поэта: PUSHKIN, LERMONT (фамилия усечена, потому что в имени каталога не может быть больше 8 символов), FET и т.д. (рис. 15.1). В каталоге PUSHKIN имеется файл duel. dat, содержащий сведения о дуэли. Итак, указан путь, по которому можно дойти до этого файла по дереву каталогов. Такой путь определяет *полное имя* файла, которое образуется из всех нужных каталогов с помощью значка \ (обратная косая черта, так называемый обратный слэш, backslash). В нашем примере полное имя файла выглядит так:

D:\USERS\IVANOV\POETS\PUSHKIN \duel.dat

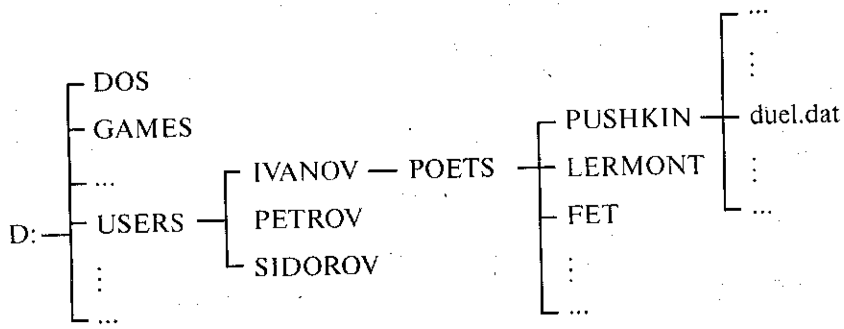


Рис. 15.1

На диске не может быть двух файлов с одинаковыми полными именами: полное имя файла уникально для данного диска. Это не значит, что в другом каталоге не может быть файла с таким же именем (например, файл duel.dat в каталоге LERMONT). Но полные имена этих файлов будут отличаться! Поэтому перенос файла из одного каталога в другой есть не что иное, как переименование этого файла, изменение его полного имени.

Вернемся к устройству системного блока компьютера. На материнской плате имеются специальные разъемы, так называемые ***слоты расширения***. В них могут устанавливаться дополнительные платы, которые применяются для расширения возможностей стандартного компьютера. Слоты позволяют подключить дополнительную память, различные внешние устройства – устройство воспроизведения звука, модем или факс для компьютерной связи по телефонным каналам и многое-многое другое. Таким образом, конфигурация компьютера изменяется достаточно просто: для этого нужно вставить в слот новую плату расширения. Такая простота наращивания мощности и возможностей позволяет говорить о так называемой открытой архитектуре компьютера.

Кроме перечисленного, в системном блоке находятся также блок питания, громкоговоритель для воспроизведения простых звуков и мелодий, а также некоторые вспомогательные детали и устройства.

Итак, основные устройства системного блока:

- процессор;
- оперативная память;
- дисководы гибких дисков и винчестер;
- слоты расширения;
- блок питания.

15.3.2. Периферийные устройства

Мы уже говорили о том, что два периферийных устройства всегда можно увидеть рядом с компьютером. Конечно, это ***клавиатура*** и ***монитор*** (дисплей).

Клавиатура

Клавиатура предназначена для ввода информации в компьютер. Легче всего работать с клавиатурой будет тем, у кого есть опыт печатания на пишущей машинке, расположение букв (латинских и русских) на клавиатуре и пишущей машинке, как правило, совпадает. Работа на клавиатуре требует соблюдения простых правил:

не следует загрязнять клавиатуру (ведь это достаточно сложное электронное устройство, содержащее микросхемы и другие детали);

не надо сильно стучать по клавишам, движения должны быть легкие, короткие и отрывистые;

клавиатура должна быть расположена на столе устойчиво.

Клавиши на клавиатуре обычно разбивают на несколько групп.

Символьные клавиши (символьная клавиатура). Это большая группа клавиш, обычно белого цвета, расположенных в центральной части клавиатуры. Самая длинная клавиша в нижнем ряду – клавиша *пробела* (Space). Черным цветом на буквенных клавишах традиционно обозначены латинские буквы, а красным – русские. Клавиатуру принято обозначать по первым буквам в верхнем буквенном ряду: английский стандарт называется QWERTY, русский – ЙЦУКЕНГ. Как и у пишущей машинки, у клавиатуры есть два регистра. Если это буквенная клавиша, то нижний регистр соответствует строчным буквам, а верхний – прописным. Для клавиш с другими символами соответствующий символ пишется внизу или вверху. К примеру, на клавише со знаками «:» и «;» точка с запятой принадлежит нижнему регистру, а двоеточие – верхнему. При включении компьютера клавиатура находится на нижнем регистре. Это *режим по умолчанию* (по-английски – default, так обозначают режим, в котором компьютер находится в исходном состоянии). Чтобы набрать символ верхнего регистра, необходимо вместе с нужной клавишей нажать клавишу Shift (таких клавиш две – слева и справа от символьной клавиатуры). Например, чтобы набрать прописную букву L, левой рукой мы нажмем на левый Shift, а правой – на клавишу с этой буквой. Можно зафиксировать верхний регистр. Для этого нажимают клавишу CapsLock. В этом случае регистры меняются местами: основной регистр – верхний, символы нижнего регистра набираются при одновременном нажатии клавиши Shift. При включении режима фиксированного верхнего регистра (CapsLock) загорается соответствующая индикаторная лампочка в правом верхнем углу клавиатуры. Повторное нажатие клавиши CapsLock отменяет этот режим (лампочка гаснет).

Переход от латинских букв к русским осуществляется благодаря специальным программам, так называемым *драйверам* клавиатуры. Таких драйверов довольно много. В зависимости от того, какой драйвер используется на вашем компьютере, вы нажимаете ту или иную комбинацию клавиш. Традиционные комбинации: две клавиши Shift (левая и правая одновременно), клавиши Ctrl и Alt (назначение этих вспомогательных клавиш мы еще обсудим) или какие-то другие. Обычно эти инструкции можно прочитать на экране при загрузке компьютера.

Клавиши управления курсором. Обычный *курсор* – это мигающая черточка, обозначающая текущее положение на экране. Особенно важно знать, где вы «находитесь», при редактировании текста. Курсор может иметь и другую форму. Разнообразные курсоры (в виде стрелок, часов, ладоней и др.) можно встретить при работе в системе Windows. Часто под курсором понимают светящийся прямоугольник, показывающий текущий пункт при выборе из меню. Такой

курсор вы увидите, работая с программой Norton Commander. Независимо от вида и назначения курсор можно перемещать по экрану с помощью клавиш управления курсором. Во-первых, это четыре серые клавиши со стрелками ←, ↓, → и ↑. Кроме того, это клавиши Home и End (обычно они переводят курсор в начало и конец строки), а также PageUp (PgUp) и PageDown (PgDn) для «листания» экранов (страниц редактируемого текста). Обратите внимание, что серые клавиши управления курсором дублируются белыми клавишами в правой части клавиатуры. Назначение этих клавиш изменится, если нажать клавишу NumLock. В этом случае они предназначены для ввода цифр (их еще называют *цифровой клавиатурой*). При этом загорается соответствующий индикатор. Отключение режима блокировки цифр (NumLock) происходит при повторном нажатии этой клавиши. К клавишам управления курсором можно отнести и клавишу *табуляции* Tab, передвигающую курсор вправо на размер красной строки. В некоторых программах эта клавиша может играть и другую роль.

Клавиши редактирования. Существует два режима редактирования текста: вставка и замена. В режиме *вставки* набранный символ **вставляется над курсором**, а строка раздвигается на один символ вправо. В режиме *замены* символ **пишется поверх имеющегося символа** и строка не раздвигается. Чтобы перейти с режима вставки (обычно именно этот режим включен по умолчанию) на режим замены, нажимают клавишу Insert (Ins). Как правило, при этом изменяется внешний вид курсора. Обратный переход – повторное нажатие этой клавиши. В правом верхнем углу символьной клавиатуры расположена серая клавиша, которая называется Backspace (обратный пробел или «забой»). Обычно она обозначается длинной левой стрелкой ⟵. С помощью этой клавиши стирается символ, расположенный **слева от курсора**. Символ непосредственно **над курсором** можно стереть с помощью клавиши Delete (Del). Результат стирания будет различным в зависимости от режима редактирования. В режиме вставки строка при этом сдвигается влево на один символ, а в режиме замены на месте стертого символа остается пробел. Клавиши Ins, Del дублируются на цифровой клавиатуре (в режиме NumLock они соответствуют символам нуля и десятичной точки).

Функциональные клавиши. В верхнем ряду клавиатуры находятся клавиши от F1 до F12. Это так называемые функциональные клавиши. Они имеют различное значение в разных программах. Мы познакомимся с употреблением этих клавиш в системе Norton Commander. Чаще всего с помощью клавиши F1 вызывают помощь, подсказку.

Специальные клавиши. Эти клавиши обычно не соответствуют никаким символам. Они изменяют значение других клавиш или выполняют другие специальные функции. С некоторыми из них мы уже ознакомились.

Shift – переход с одного регистра на другой;

CapsLock – включает/выключает режим фиксированного верхнего регистра;

NumLock – включает/выключает режим блокировки цифр на цифровой клавиатуре;

Enter – очень важная клавиша (серая, в виде угла), применяется для запуска исполняемых программ, для подтверждения выбора пункта меню и т.д.; дублируется в правой части клавиатуры;

Esc – обычно применяется для выхода из программы (от английского слова Escape – избежать);

Ctrl (Control, управление), Alt (Alternative – второе значение, альтернатива) – эти клавиши употребляются не сами по себе, а только в сочетании с другими клавишами, изменяя при этом их значение; в различных программах разные комбинации действуют по-разному. Например, комбинация трех клавиш Ctrl – Alt – Del перезагружает компьютер.

С краю цифровой клавиатуры есть серые (по английски grey – серый) клавиши *, +, – (их так и называют – Grey *, Grey + и Grey –). Они предназначены не только для ввода соответствующих символов, эти клавиши могут иметь и самостоятельное значение.

Существуют разнообразные программы-тренажеры, помогающие пользователю освоить клавиатуру. С них разумно начинать знакомство с компьютером.

Монитор

Монитор (дисплей) компьютера очень похож на обычный телевизор. Как и телевизоры, мониторы бывают цветными и монохромными (одноцветными). Монитор предназначен для вывода на экран информации в виде текста и рисунков (а также мультипликации,

видеофильмов и т.д.). Соответственно он может работать в одном из двух режимов: текстовом или графическом. В *текстовом* режиме экран монитора разбивается на 25 строк, в каждой из которых может быть выведено 80 символов. Как известно, один символ кодируется в компьютере с помощью одного байта. Поэтому доступный набор состоит из 256 символов. С большинством символов мы уже ознакомились при изучении клавиатуры. В *графическом* режиме экран монитора состоит из отдельных точек (*пикселей*), каждая из которых может быть темной или светлой на монохромных дисплеях или одного из нескольких цветов на цветных. Количество точек по горизонтали и вертикали называется *разрешающей способностью* монитора. Чем выше разрешающая способность, тем лучше выводимые на экран графики и рисунки.

Главные характеристики монитора – его разрешающая способность и количество доступных цветов. Современные мониторы имеют разрешающую способность 640×480, т.е. на них выводится 640 пикселей по горизонтали и 480 – по вертикали. Некоторые мониторы обладают разрешающей способностью 1024×768 и выше. Обычно доступны 16 или 256 цветов, но возможны режимы с многими тысячами и даже миллионами цветов.

Принтер

Третье по важности периферийное устройство, без которого не обходится сейчас ни один разумный пользователь компьютера, – это печатающее устройство, ***принтер***. Принтер предназначен для вывода результатов на бумагу в виде текста, таблиц, графиков и т.д. Существуют тысячи моделей принтеров, с которыми может работать компьютер IBM PC. Они различаются по конструкции, скорости и качеству печати, а также, разумеется, по цене. В зависимости от устройства принтеры бывают:

- матричные;
- лазерные;
- струйные.

Матричные принтеры – наиболее распространенный тип принтеров для персональных компьютеров. В основном это разнообразные модели фирмы «Epson». Печатающая головка матричного принтера имеет вертикальный ряд иголок, которые при движении головки вдоль бумаги в нужный момент ударяют по красящей ленте. Таким образом на бумаге формируются символы и графические изображения. Дешевые принтеры имеют 9 иголок. Более качественную печать обеспечивают 24-игольчатые принтеры.

Лазерный принтер устроен почти так же, как ксерокс: изображение переносится на бумагу со специального барабана, который электризуется в нужных местах лазером и притягивает частички краски. Лазерные принтеры обеспечивают очень высокое качество печати, близкое к типографскому. Самые распространенные модели – принтеры LaserJet известной компьютерной фирмы Hewlett Packard. Лазерный принтер работает гораздо быстрее матричного, поэтому он незаменим при больших объемах печати.

В *струйном* принтере изображение формируется микрокаплями специальных чернил, выдуваемыми на бумагу через капилляры печатающей головки. Струйные принтеры могут печатать цветной текст. Для этого используется комбинация чернил трех основных цветов. Качество печати на струйном принтере значительно выше, чем на матричном, тем не менее стоят они ненамного дороже.

Правильный выбор принтера, разумеется, зависит от потребностей пользователя: нужна ли цветная печать, каковы объемы распечатки, какое качество печати необходимо, требуется ли печать разными шрифтами и т.д. Следует обращать внимание на возможность печатания русских букв. Русификация принтера может достигаться аппаратными или программными средствами. В первом случае русский шрифт встроен в запоминающее устройство принтера. При программной русификации для печати русского текста требуется предварительно запустить соответствующую программу-драйвер.

Другие периферийные устройства

Кроме упомянутых ранее, к персональному компьютеру можно подключать и другие периферийные устройства. Перечислим некоторые из них.

Мышь – устройство, которое представляет собой небольшую коробочку, обычно с тремя клавишами и шариком, подключенную к компьютеру длинным проводом («мышка с хвостом»). Мышь все чаще применяется вместо клавиатуры для ввода информации в компьютер. Современные программы работают таким образом, что для выполнения любого сложного действия (выбора пункта меню, запуска программы и т.д.) достаточно переместить мышь по столу (при этом перемещается и курсор на экране), а затем нажать на клавишу. Для удобства работы с мышью используются специальные коврики, курсор при этом лучше «слушается», повторяя движения шарика.

Устройство для чтения компакт-дисков. Наряду с магнитными дисками в последнее время для хранения компьютерной информации все шире применяются лазерные компактные диски (так называемые CD ROM: CD – это компакт-диск, ROM – это Read-Only Mode, т.е. предназначены они только для чтения, а не для записи). Компьютерные лазерные диски выглядят точно так же, как хорошо вам знакомые диски с музыкальными записями. Лазерные диски могут хранить гораздо больше информации, чем магнитные (до 650 Мбайт). На лазерных дисках распространяются некоторые большие по объему программные пакеты, а также мультимедиа-программы, сочетающие в себе звук, изображение и т.д. (энциклопедии, мощные обучающие программы и др.).

Графопостроитель (плоттер) – устройство для вывода на бумагу чертежей и графиков.

Сканер – устройство для считывания графической информации в компьютер. С помощью сканера любую картинку можно превратить в файл данных, удобный для обработки. Специальные программы позволяют редактировать полученный таким образом рисунок, изменять цвета, дорисовывать детали, комбинировать с другими рисунками. Для считывания цветных рисунков существуют цветные сканеры. Сканер незаменим в издательском деле при компьютерном наборе текстов.

Модем – устройство для подключения компьютера к телефонной сети. С помощью модема компьютер может «общаться» с другими компьютерами через локальные и глобальные сети. Модем – это первый шаг в Интернет, к неограниченным источникам информации.

Звуковая плата (саундбластер) – улучшает музыкальные возможности компьютера, позволяет прослушивать на компьютере лазерные аудиодиски. К звуковой плате обычно прилагаются колонки для воспроизведения звука и микрофон. С помощью этих устройств компьютер, подключенный в сеть, может устанавливать аудиосвязь. Многие программы (особенно мультимедиа и игровые) используют звуковую плату для музыкального сопровождения и звуковых эффектов.

Трекбол – представляет собой перевернутую мышь: для управления курсором вращается шарик на подставке. Применяется обычно в портативных компьютерах.

Возможности подключения к компьютеру внешних устройств практически не ограничены. К компьютеру можно подключать фотоаппарат, видеокамеру, диктофон и многое другое. Открытая архитектура позволяет делать это достаточно легко и безболезненно. Требуются лишь переходное устройство, плата расширения и соответствующее программное обеспечение.

15.4. Работа в системе Norton Commander

Итак, мы ознакомились с устройством компьютера в общих чертах. Конечно, мы не вдавались в технические детали, да это и не требуется для обыкновенного пользователя. Теперь нам осталось, не страшась, включить компьютер. Запомним порядок включения: сначала включаются все нужные периферийные устройства (принтер – если требуется, обязательно – монитор, на нем есть кнопка или тумблер). После этого можно включать и сам компьютер: обычно это делается с помощью большой кнопки или переключателя на корпусе системного блока. Чаще всего эта кнопка обозначена словом «Power». Сразу после включения начинается загрузка в оперативную память самой главной компьютерной программы – **операционной системы**, под управлением которой и играет «оркестр» компьютера. Мы поговорим об операционной системе персонального компьютера чуть позже, когда приобретем некоторый опыт работы. Вместе с операционной системой загружаются другие нужные и полезные программы – например, русификаторы клавиатуры и монитора; программа-драйвер для пользования мышью и др. Загрузка сопровождается выводом на экран сообщений. Чаще всего это информация об успешном завершении той или иной операции или инструкция по пользованию какой-то программой – к примеру, на какие клавиши следует нажать, чтобы перейти на русские буквы. Иногда, если что-то не в порядке, на экране появляются сообщения об ошибках. В этом случае разумнее всего обратиться за помощью к специалисту. Будем, однако, надеяться, что загрузка прошла успешно и перед вами на экране возникли... Да-да, я уверен почти на сто процентов, что перед вами на экране появились две голубые панели со списками каталогов и файлов. Это значит, что начала работать самая распространенная в мире программа под названием **Norton Commander** (разработанная под руководством известного программиста Питера Нортона). Без преувеличения можно сказать, что та или иная версия этой программы установлена сейчас практически на каждом персональном компьютере.

Программа Norton Commander широко использует две чрезвычайно важные компьютерные идеи, благодаря которым общение с компьютером стало удобным и наглядным, а сам компьютер стал доступным непрофессиональному пользователю. Это идея **меню** и идея **окна**. Вместо того чтобы выдавать компьютеру команду (а это значит, как бы там ни было, набирать ее текст), вы всего лишь выбираете нужное действие из предлагаемых возможностей (точно так же, как блюдо в ресторане). Выбор пункта меню обычно осуществляется нажатием одной клавиши (чаще всего это клавиша Enter) или кнопки мыши. Работа на компьютере превращается в последовательный выбор: выбранный пункт может порождать другое меню и т.д. Такие меню называются рассыпающимися; вы столкнетесь с ними практически в любой современной программе. Особый тип меню – так называемые *кнопки*. Пункты меню расположены в маленьких окошках, которые выглядят как кнопки. Нужный пункт выбирается «нажатием» кнопки с помощью клавиши Enter или мыши. Иногда кнопки выглядят объемными и при нажатии утапливаются, как настоящие.

Концепция окна, оконная структура, также сыграла революционную роль в упрощении работы на компьютере. При работе программы экран разбивается на несколько прямоугольных участков – *окон*, в каждом из которых отображается своя информация. Удобная организация окон существенно облегчает восприятие, помогает сконцентрировать внимание на важнейших моментах работы. Развитие этой идеи привело к созданию программы Windows (что и означает «окна»). Окна в системе Norton Commander более скромные, но их применение помогает лучше организовать, упорядочить информацию на экране. Программа Norton Commander запускается с помощью исполняемого файла NC.EXE, который, как правило, находится в каталоге NC, или NORTON.

Как нам уже известно, любая работа на компьютере есть, по сути, работа с файлами. Чтобы облегчить эту работу, и предназначен Norton Commander. Что же мы видим на двух голубых панелях (которые на самом деле являются окнами на экране)? На них выведено содержимое, оглавление одного из каталогов на диске (рис. 15.2). В верхней строке панели указано имя этого каталога. Прописными буквами в оглавлении обозначены каталоги, строчными – файлы. Один из каталогов (на левой или правой панели) в каждый данный момент является *текущим*: его имя в ы д е л е н о, т.е. написано буквами другого цвета и на другом фоне (если стандартное сочетание, скажем, – белое на голубом, то выделенное – черным по серому). Точно так же выделено имя одного из каталогов или файлов в оглавлении: этот каталог или файл является текущим (к нему относятся все ваши действия в данный момент). Это выделение осуществляется с помощью движущегося светового окошка, курсора, который можно перемещать по оглавлению клавишами управления курсором, меняя текущий файл.

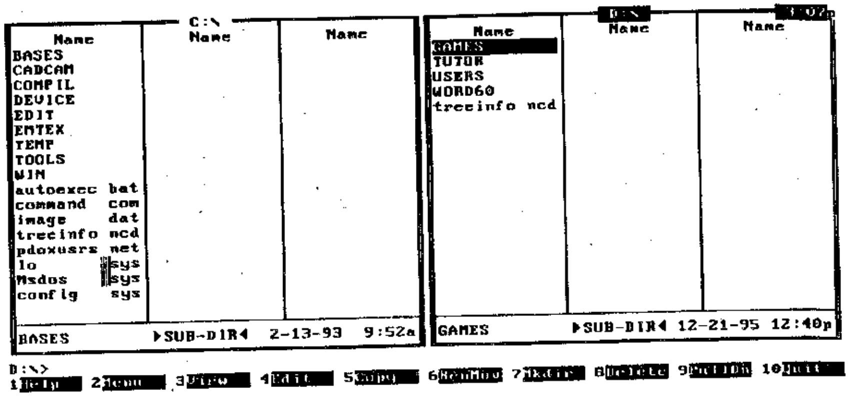


Рис. 15.2

15.4.1. Выбор диска и каталога. Просмотр оглавления

Нужный каталог выбирается в два этапа: сначала – диск, а затем – каталог на этом диске. Для выбора диска на левой панели нажимают одновременно клавиши [Alt] – [F1]. Перед вами появляется небольшое меню, состоящее из латинских букв: например, от A до D (рис. 15.3). Заголовок меню – Drive letter (буква диска) с надписью «Choose left drive» («Выбирайте диск на левой панели»). Это значит, что доступны четыре диска: два гибких (A и B) и два логических диска на винчестере (C и D). Диск A, как правило, выделен курсором, который перемещается клавишами «вправо» или «влево» (перемещение циклическое, как в большинстве нортоновских меню: дойдя вправо до последнего пункта, вы переходите к первому, и наоборот). Чтобы выбрать диск, подведите окошко к нужной букве и нажмите на [Enter]. Вместо этого можно просто нажать клавишу с нужной буквой. То же самое можно сделать с помощью мыши: на экране есть специальный курсор мыши, обычно это красный прямоугольник. Курсор перемещается по экрану при перемещении мыши. Чтобы выбрать диск, достаточно подвести курсор к нужной букве и нажать на правую клавишу мыши. Итак, диск выбран и на панели перед вами появляется оглавление корне-

Рис. 15.3

вого каталога данного диска. Чтобы перейти в нужный вам каталог (говорят – «войти» в каталог), нужно просто-напросто подвести к нему курсор (световое окошко) и нажать на [Enter]. Все ваши перемещения по каталогам отражаются в верхней строке панели: меняется имя текущего каталога. Войти в каталог можно и с помощью мыши. Во всех каталогах, за исключением корневого, имеется родительский каталог, обозначенный двумя точками. «Ходить» по оглавлению очень легко с помощью клавиш «вверх», «вниз» и клавиш листания страниц [PgUp], [PgDn]. Клавиша [Home] переводит курсор в начало каталога, клавиша [End] – в конец. Чтобы быстро перейти в корневой каталог, нажимают клавиши [Ctrl] – [\]. Выбор диска на правой панели осуществляется клавишами [Alt] – [F2], а дальнейший выбор каталога – точно так же, как на левой панели. Переход от одной панели к другой осуществляется клавишей табуляции [Tab].

15.4.2. Запуск программ

Как вы помните, файлы с расширениями **.com**, **.exe** и **.bat** являются исполняемыми программами. Norton Commander позволяет запустить такую программу на выполнение исключительно просто. Предположим, вы решили отдохнуть и поиграть в тетрис. Найдем файл tetris.com (очевидно, искать его нужно в каталоге GAMES), поставим на него курсор и нажмем клавишу [Enter]. Альтернативный способ, как всегда, – с помощью мыши. Для запуска текущего исполняемого файла нужно 2 раза быстро нажать левую клавишу мыши, или, говоря на профессиональном жаргоне, «дважды кликнуть мышкой». После этого программа начинает выполняться.

Можно запустить программу и «вручную». Под панелями Norton Commander находится строка с так называемым *приглашением* операционной системы. Приглашение представляет собой полное имя текущего каталога со значком «>»: например,

C:\USERS\IVANOV>

Для запуска программы после приглашения можно написать ее имя и нажать на [Enter]. Если файл находится в текущем каталоге, достаточно набрать только имя, без расширения. Система сама разберется, что данный файл является исполняемой программой. Не возбраняется, впрочем, набрать и имя с расширением. Если вы хотите запустить файл из другого каталога (расположенного, возможно, на другом диске), нужно набрать его **полное имя**:

C:\USERS\IVANOV>d:\games\tetris

Неважно, какими буквами набирать имя файла – строчными или прописными.

Как видите, запустить программу с помощью Norton Commander гораздо проще.

15.4.3. Работа с файлами и каталогами

В начале разговора о системе Norton Commander мы заявили, что эта программа облегчает пользователю работу с файлами. Что же можно делать с файлами или каталогами при работе на компьютере? Любые файлы или каталоги можно:

- копировать;
- переносить;
- удалять.

Часто возникает потребность сохранить файлы в другом каталоге или на другом диске так, чтобы при этом они остались на старом месте. Такая операция называется ***копированием***. Культура работы на компьютере предполагает регулярное копирование нужных файлов с винчестера на дискеты. Если требуется просто переместить файлы в другой каталог, не оставляя их на своем месте, пользуются операцией ***переноса***. По сути дела, перенести файл означает дать файлу новое полное имя. Поэтому перенос часто называют ***переименованием***. Операция ***удаления*** не нуждается в особых комментариях: при удалении файлы исчезают из каталога. Следует иметь в виду, что удаление из каталога не означает немедленного физического уничтожения, стирания файла с диска. Поэтому существуют специальные программы восстановления стертых файлов. Тем не менее операцией удаления следует пользоваться крайне осторожно.

Копировать, переносить и удалять можно как о т д е л ь н ы й файл, так и г р у п п у файлов (каталогов). Группа состоит из *помеченных* файлов или каталогов, их имена выделяются в оглавлении желтым цветом. Пометить, или выбрать, файл (каталог) можно тремя способами.

1. Выбор отдельного файла (каталога). Для этого нужно подвести к этому файлу курсор и нажать клавишу [Ins]. Отменить выбор отдельного файла можно повторным нажатием [Ins].

2. Выбор файлов по маске. Прежде чем обсудить этот способ выбора, рассмотрим употребление специальных символов ? и * в именах файлов. Эти символы позволяют с помощью одного имени обозначать сразу несколько файлов. Имя, в котором используются специальные символы, называется *маской* или *шаблоном*.

Вопросительный знак (?) заменяет в имени файла любой отдельный символ в данной позиции. Например, запись chapter?.txt обозначает все файлы с расширением .txt, имя которых начинается буквами chapter, а восьмая буква имени – любая. Например, файлы CHAPTER1.TXT, CHAPTER2.TXT, CHAPTERX.TXT, CHAPTER_.TXT и т.д. описываются данным шаблоном, а файлы CHAPTER.TXT и CHAPTER1.DOC – нет.

Звездочка (*) в имени файла означает, что в этой или в оставшихся позициях могут быть любые символы. Например, за шаблоном chap*.txt скрываются файлы CHAPTER.TXT, CHAP1.TXT, CHAPEL.TXT и т.д. Шаблон *.doc описывает все файлы текущего каталога с расширением .DOC. Шаблон myfile.* – это все файлы с именем MYFILE и с любым расширением (MYFILE.TXT, MYFILE.DOC, MYFILE.C и т.д.). Маска *.* задает в с е файлы текущего каталога. Ясно, что по определению обозначение *.* означает то же, что ????????.??? .

Символы ? и * называются символами *замены* (иногда – подстановочными, глобальными или родовыми символами). По-английски они называются wildcards, что означает карту-джокер.

Чтобы выбрать группу файлов по маске (Select the files), нажимают клавишу «серый плюс» (серая клавиша справа от цифровой клавиатуры). Появляется окошко, в котором по умолчанию задан шаблон *.*. Вы можете задать свой шаблон, пользуясь символами замены. По окошку можно «перемещаться» клавишами «вправо» и «влево», стирать символы клавишами [Backspace] и [Del]. Окончательный выбор шаблона – клавишей [Enter]. Отменить выбор по маске (Deselect) можно с помощью клавиш «серый минус», задавая шаблон файлов, отметку которых вы хотите отменить.

3. Инверсный выбор. Чтобы все выделенные файлы сделать невыделенными, а все невыделенные – выделить, нажимают серую клавишу [*]. Этот способ удобен, если нужно отметить много файлов в каталоге и их нельзя удобно описать с помощью шаблона. Тогда отметим немногие оставшиеся и клавишей [*] инвертируем выбор: нужные файлы станут отмеченными, а остальные – нет.

Если выбрана группа файлов, их имена на панели изображаются желтым цветом. Внизу панели содержатся сведения о числе выбранных файлов и их общем размере.

16 Зак. № 2435 Турецкий

Копирование, перенос или удаление одного файла (каталога) или группы файлов (каталогов) осуществляется с помощью функциональных клавиш. В последней строке экрана содержится так называемая *подсказка* с перечислением клавиш и их функций. Вы увидите:

... 5 Copy 6 RenMov ... 8 Delete

Значит, для копирования (Copy) применяется клавиша F5; для переименования/переноса (Rename/Move) – клавиша F6; для удаления (Delete) – клавиша F8. Эти три операции выполняются почти одинаково. Рассмотрим подробнее процедуру копирования.

Копирование. Если в текущем каталоге есть отмеченные файлы, то они и будут копироваться; если выбранных файлов нет, копируется текущий файл. После нажатия F5 в центре экрана появится запрос о том, куда надо копировать файл или файлы (рис. 15.4).

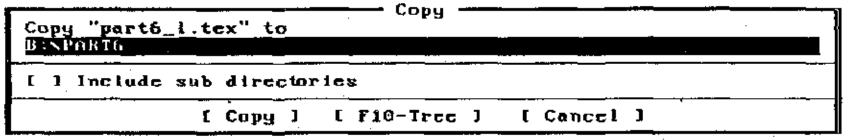


Рис. 15.4

По умолчанию предлагается имя каталога с другой панели. Поэтому перед копированием разумно открыть на неактивной панели тот каталог, куда вы собираетесь копировать файлы с активной панели. В этом случае достаточно нажать Enter. Если же вы хотите скопировать файлы в другой каталог и/или под другими именами, можно внести эти новые имена вместо предлагаемого по умолчанию. Для группового изменения имен файлов можно пользоваться символами замены ? и *. Предположим, вы отметили все файлы с расширением .bak и хотите скопировать их на пятидюймовую дискету с теми же именами, но с расширением .txt. Тогда вам следует написать a:.txt. Для быстрого поиска нужного каталога для копирования можно использовать кнопку F10 – Tree: при выборе этой операции появится дерево каталогов, по которому можно быстро «дойти» до требуемого каталога. После выбора места копирования нажимают Enter. При копировании файлов на экран выводится полоска-диаграмма, графически изображающая процесс копирования. Отказаться от копирования можно клавишей Esc.

Если вы хотите скопировать целиком каталог (каталог оказался среди выбранных имен или курсор указывает на каталог), необходимо включить режим «Include sub directories» («Включая подкаталоги»). Для этого следует переместить курсор в рамку слева от этой надписи и нажать клавишу «Пробел». При этом на месте курсора появляется крестик (рис.15.5). В этом режиме копируются и все подкаталоги данного каталога вместе с файлами, т.е. все «дерево», произрастающее из данного каталога.

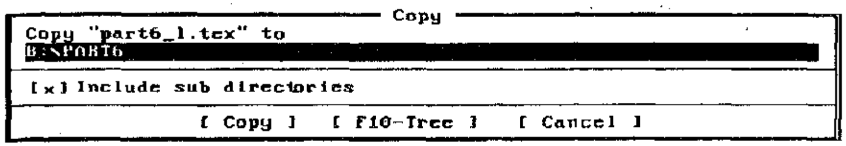


Рис. 15.5

Типичная ошибка при копировании возникает в ситуации, когда файл с данным именем уже присутствует в том каталоге, куда осуществляется копирование. Так бывает, если вы регулярно копируете одни и те же файлы на дискету.

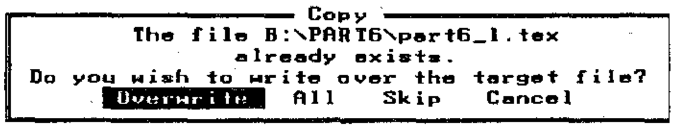


Рис. 15.6

В этом случае на экран выдается сообщение «The file... already exists. Do you wish to write over the target file?» («Файл... уже существует. Вы хотите записать поверх старого файла?») (рис. 15.6). Клавишами «вправо» и «влево» можно выбрать нужный ответ из следующих вариантов в кнопочном меню:

- Overwrite – разрешить копирование; при этом существовавший файл с тем же именем будет уничтожен;
- All – разрешить копирование этого и всех последующих файлов без запросов;
- Skip – не копировать данный файл;
- Cancel – вообще отказаться от копирования (то же самое можно сделать с помощью клавиши Esc).

Перенос. Переименование (перенос) файлов осуществляется с помощью клавиши F6. Для переименования вы должны задать новое имя файла, если работаете с одним текущим файлом, или имена – при работе с группой выделенных файлов (для задания новых имен можно пользоваться символами замены). Пересылка в Norton Commander осуществляется так же, как копирование: по умолчанию файлы пересылаются в каталог, открытый на неактивной панели, выдается сообщение, если файл с таким именем уже существует, и т.д.

Удаление. Для удаления файлов используется клавиша [F8]. При этом удаляется либо текущий файл, либо группа отмеченных файлов. Перед удалением Norton Commander спрашивает, действительно ли вы хотите удалить выбранные файлы. Можно ответить либо «Delete» («Удалить»), либо «Cancel» («Отменить»). Разумеется, отменить удаление можно и клавишей [Esc]. При удалении одного файла положительный ответ приводит к немедленному удалению без дополнительных запросов. Если же удаляется группа файлов, запрос на удаление выводится по каждому файлу индивидуально. Можно выбрать один из следующих вариантов ответа («нажать» соответствующую кнопку):

- Delete – удалить данный файл;
- All – удалить данный файл и все остальные без запросов;
- Skip – не удалять данный файл;
- Cancel – отказаться от удаления (или клавишей [Esc]).

При удалении каталогов нужно, как и при копировании, включить режим Include sub directories. Если данный режим выключен, то возможно удаление только пустых каталогов.

Для **создания нового каталога** на диске нажимают клавишу [F7] (в строке-подсказке вы найдете 7 MkDir, т.е. Make Directory – создать каталог). При нажатии этой клавиши появляется окошко, в которое можно внести имя вновь создаваемого каталога. После ввода имени нажмите на клавишу [Enter]. Новый каталог появится в оглавлении на активной панели. Если «зайти» в новый каталог, мы увидим, что он не содержит файлов. Единственное его содержимое – родительский каталог, обозначаемый двумя точками.

15.4.4. Работа с текстовыми файлами

Значительную долю времени работы на компьютере занимают операции с текстами. Текстовый файл можно:

- просматривать;
- редактировать;
- создавать.

Посмотрим на строку-подсказку:

... 3 View 4 Edit ...

Таким образом, просмотр (View) текстового файла осуществляется с помощью клавиши [F3], а редактирование – с помощью клавиши [F4].

Просмотр. Нажав [F3], вы сможете просмотреть текстовый файл, выделенный курсором в данный момент. Для перемещения по тексту используются клавиши управления курсором [←], [→], [↓], [↑], [PgUp], [PgDn]. Клавиши [Home] и [End] перемещают соответственно в начало и конец файла. Конец просмотра – клавиша [Esc].

Редактирование. Чтобы отредактировать текущий текстовый файл, нажимают клавишу [F4]. Экран редактирования отличается от экрана просмотра наличием курсора (мигающей черточки), показывающего положение текущего символа в тексте. Встроенный редактор Norton Commander достаточно прост, тем не менее он позволяет выполнять основные операции с текстом:

- редактирование в режиме вставки или замены (клавиша [Ins]);
- работа с блоками:
  копирование;
  перенос;
  удаление;
- поиск строки в тексте;
- поиск строки с заменой на другую строку;
- печать текстового файла на принтере.

Не будем останавливаться на этих процедурах редактирования. Мы ознакомимся с ними достаточно подробно при изучении текстовых редакторов. Клавиши, необходимые для каждой операции, указаны в строке-подсказке редактора.

Чтобы сохранить отредактированный файл, нажимают клавишу [F2]. Выход из режима редактирования – клавиша [Esc]. Если вы забыли сохранить изменения, Norton Commander спросит вас при выходе, записывать или нет данный файл.

Можно редактировать любой текстовый файл, не обязательно текущий. Для этого вместо [F4] нажимают [Shift] [F4] и по запросу вводят имя файла, требующего редактирования. Если при этом ввести имя файла, которого на диске нет, будет создан новый файл с этим именем. Поэтому комбинацию клавиш [Shift] [F4] обычно применяют для **создания** нового текстового файла.

15.4.5. Работа с панелями

Как указывалось ранее, переходить с панели на панель (т.е. менять активный каталог) можно с помощью клавиши табуляции Tab. Предусмотрены также следующие удобные действия с панелями:

убрать/вернуть левую панель	Ctrl F1
убрать/вернуть правую панель	Ctrl F2
убрать/вернуть обе панели	Ctrl O
поменять панели местами	Ctrl U

Убрать панель (панели) с экрана бывает необходимо, если предыдущая программа оставила на экране какие-то сообщения. При возвращении в Norton Commander эти сообщения будут скрыты за панелями. Для просмотра сообщений убирают панели. Панели возвращаются на место повторным нажатием соответствующих клавиш.

15.4.6. Дополнительные возможности

Быстрый переход к нужному файлу (каталогу). Если оглавление на панели содержит большое количество каталогов и файлов, поиск нужного файла может занять много времени. Чтобы быстро перейти к файлу, нажмите клавишу Alt и, не отпуская ее, наберите первые буквы имени этого файла или каталога. Предположим, вы хотите перейти в каталог USER. Нажмите Alt и клавишу с буквой U. Введенная буква появится в окошке в правом нижнем углу экрана (рис. 15.7). Если в оглавлении есть несколько каталогов (файлов), имя которых начинается с этой буквы, можно уточнить поиск, набрав вторую, третью и последующие буквы имени.

Рис. 15.7

Быстрый поиск файла на диске. Для быстрого поиска файла во всех каталогах текущего диска следует нажать клавиши Alt F7 и набрать справа от надписи «File Name» имя требуемого файла. В имени можно использовать символы замены ? и *. Можно также задать строку, содержащуюся в этом файле, справа от слова Containing («содержащий»). После этого нужно переместить светящийся курсор на окошко с надписью «OK» и нажать на Enter. В процессе поиска Norton Commander будет выводить на экран имена найденных файлов. К любому из них можно перейти, если подвести к нему курсор и нажать на Enter. В случае неудачного поиска можно поменять диск (Change Drive) и повторить поиск на другом диске.

Быстрый просмотр файлов. Norton Commander позволяет быстро просмотреть содержимое нескольких файлов. Для этого нажимают клавиши Ctrl Q. Неактивная панель при этом становится окном, в котором можно просмотреть содержимое текущего текстового файла на активной панели. Если текущим является каталог или файл с расширением .EXE или .COM, то в окне быстрого просмотра выводится сообщение об этом.

Изменение вида вывода информации на панели. Оглавление на панели может выводиться в двух режимах: кратком (Brief – режим по умолчанию) и полном (Full). В *кратком* режиме выводятся только имена и расширения файлов и каталогов. Вывод осуществляется в три столбца. В *полном* режиме на информацию о каждом каталоге и файле отводится одна строка на панели. При этом наряду с именем файла выводятся его размер (Size), дата и время создания (Date, Time). Для каталогов вместо размера выводится слово SUB-DIR (Subdirectory – подкаталог). Родительский каталог обозначается UP-DIR (каталог верхнего уровня).

Чтобы изменить режим вывода, нужно войти в нортоновское меню. Это делается с помощью клавиши F9. (В строке-подсказке указано: ... 9 PullDn (от Pulldown – рассыпающееся меню).) Первый и последний пункты меню, расположенного в верхней строке экрана, называются Left и Right и относятся соответственно к левой и правой панели. Выбрать пункт меню в системе Norton Commander можно двумя способами: либо подвести к нужному пункту курсор и нажать на Enter, либо нажать букву, выделенную в названии пункта. Например, переход к пункту Right можно сделать так: F9 – R. При этом вы попадете в вертикальное подменю, первые два пункта которого и отражают режим вывода информации: **B**rief и **F**ull. Для перехода к полной форме вывода нажмите клавишу с буквой F (при этом выбранный режим будет отмечен специальной галочкой). Итак, для правой панели быстрый выбор п о л н о й формы вывода:

F9 – R – F;

выбор краткого режима:

F9 – R – B.

Для левой панели R заменяется на L.

Оглавление на панели может выводиться не только в разных режимах, но и в различном порядке. Это зависит от того, по какому принципу упорядочены каталоги и файлы, какой выбран режим сортировки. Имеются следующие возможности:

- в алфавитном порядке имен;
- в алфавитном порядке расширений;
- в порядке убывания даты и времени создания;
- в порядке убывания размера файлов;
- в порядке записи каталогов и файлов на диск.

Наиболее удобным представляется упорядочение по расширению (этот порядок обычно и принят по умолчанию). В этом случае файлы сгруппированы по назначению: все текстовые файлы с расширением .TXT собраны вместе, все программы с расширением .EXE – тоже и т.д.

Чтобы задать способ упорядочения, также пользуются пунктами Left и Right нортоновского меню. В этих пунктах есть подпункты Name (имя), Extension (расширение), Time (время), Size (размер) и Unsorted (без сортировки) с соответствующими выделенными буквами. Выбор режима сортировки осуществляется либо с помощью клавиши Enter, либо нажатием клавиши с выделенной буквой.

Можно обойтись и без меню. В этом случае для выбора режима сортировки применяются следующие клавиши:

Ctrl F3 – сортировка по имени;
Ctrl F4 – сортировка по расширению;
Ctrl F5 – сортировка по времени;
Ctrl F6 – сортировка по размеру;
Ctrl F7 – естественный порядок.

Разумеется, мы ознакомились далеко не со всеми возможностями системы Norton Commander. Наша задача – дать представление об общих, наиболее часто используемых в повседневной работе функциях этой системы (см. Приложение 2). Некоторые дополнительные возможности (в частности, создание пользовательского меню и обработка расширений) будут разобраны позднее. Остальные вы освоите самостоятельно, изучая руководства и прибегая к подсказкам самой системы. Помощь по данному пункту меню или процедуре вы получите, нажав клавишу F1 (Help).

И, наконец, последняя функция: **выход** (Quit) из системы Norton Commander осуществляется клавишей F10. Если вы нажали эту клавишу, Norton спросит вас: «Do you want to quit Norton Commander?» («Вы хотите выйти из Norton Commander?»). Два возможных ответа – Yes и No (да и нет).

15.5. Основные команды операционной системы MS DOS

Мы уже упоминали, что всей работой компьютера управляет одна специальная программа, которая называется *операционной системой.* Операционная система помогает загружать программы, распределяет память, манипулирует данными, согласовывает действия периферийных устройств – короче говоря, делает всю невидимую, «черную», работу в компьютере. Подавляющее большинство компьютеров IBM и им подобных оснащено операционной системой **MS DOS**, т.е. дисковой операционной системой фирмы Microsoft. Помните молодого программиста Билла Гейтса, сочинившего язык Бейсик для компьютера «Альтаир»? Именно он возглавляет компьютерную империю, программное обеспечение которой используется на миллионах компьютеров в мире.

Итак, вы нажали клавишу F10, ответили Yes и покинули «уютные стены» Norton Commander. Перед вами – черный экран с приглашением DOS. Никаких подсказок, никаких окон, никаких меню... К счастью, ситуация не безвыходна – почти все, что умеет Norton Commander, умеет и операционная система.

Конечно, развитие персональных компьютеров шло в обратном порядке: сначала была создана операционная система и компьютер научился выполнять ее команды, а затем Питер Нортон разработал свою «палочку-выручалочку». Однако на практике непрофессиональный пользователь начинает работать сразу в системе Norton Commander и лишь в крайних случаях вынужден прибегать к помощи операционной системы. Естественный ход освоения компьютера подсказывает именно такое расположение материала – сначала Norton Commander, затем MS DOS. Тем не менее не надо теряться перед черным экраном в отсутствие спасительных голубых панелей.

Команды операционной системы набираются с клавиатуры после приглашения DOS. Неважно, какими буквами вы набираете команду – строчными или прописными. Введенную команду можно отредактировать с помощью клавиш управления курсором и клавиш стирания символов ($\boxed{\text{Del}}$ и $\boxed{\text{Backspace}}$). Ввод новых символов может осуществляться в режиме вставки и замены (режим изменяется, как всегда, с помощью клавиши $\boxed{\text{Ins}}$). Чтобы команда начала работать, после ее набора следует нажать клавишу $\boxed{\text{Enter}}$. Все команды операционной системы устроены одинаково. В их состав входят:

- ключевое слово (собственно команда);
- параметры (т.е. объекты действия команды, например имя создаваемого каталога или копируемого файла). Параметры отделяются от команды и друг от друга пробелами;
- ключи (ключ уточняет действие команды, задает режим ее выполнения). Ключ представляет собой символ после косой черты, например /w.

Параметры и ключи – необязательные компоненты команд. Бывают команды, состоящие из одного ключевого слова. Однако большинство команд не имеют смысла без указания нужных параметров. Разберем основные команды операционной системы MS DOS.

15.5.1. Команды работы с дисками и каталогами

Смена диска. Чтобы перейти на другой диск, достаточно набрать его имя и нажать на $\boxed{\text{Enter}}$. Например, если вы находились на диске С: и хотите перейти на диск А:, вы наберете «А:» после приглашения системы:

C:\>A:

Создание каталога. Используется команда **md** (Make Directory). Команда имеет обязательный параметр – имя создаваемого каталога. Ее синтаксис:

md *имя каталога*

Например, чтобы создать подкаталог PETROV каталога USERS, нужно набрать команду md petrov.

Переход в другой каталог. Используется команда **cd** (Change Directory). Обязательный параметр – имя каталога, в который вы желаете перейти. Синтаксис – такой же, как у предыдущей команды:

cd *имя каталога*

В качестве имени каталога следует указывать путь до него от текущего каталога или от корневого каталога диска. Пусть, например, на диске С: имеются каталоги GAMES и USERS. В последнем имеется подкаталог IVANOV, который, в свою очередь, содержит подкаталог PROGRAMS. Вы находитесь в каталоге IVANOV. Чтобы перейти в каталог PROGRAMS, достаточно набрать

cd programs

Если вы хотите перейти в каталог GAMES и набрали

cd games,

будет выдано сообщение об ошибке

Invalid directory

(неправильный каталог), так как не указан путь от корневого каталога. Правильная команда:

cd c:\games

Уничтожение каталога. Используется команда **rd** (Remove Directory) с обязательным параметром – именем уничтожаемого каталога:

rd *имя каталога*

Обратите внимание: уничтожать таким образом можно только пустые каталоги!

Просмотр каталогов. Используется команда **dir**. Если команда не содержит параметров, на экран выводится оглавление текущего каталога. После оглавления указывается количество содержащихся в каталоге файлов и их общий объем, а также количество байтов, свободных на данном диске. Каталог выводится в режиме «прокрутки», и, если он достаточно большой, мы увидим на экране только последнюю его часть. Для вывода оглавления порциями применяется ключ /p:

dir/p

После каждой порции, помещающейся на один экран, выводится надпись

Press any key to continue...

(«Для продолжения нажмите любую клавишу...»). Нажатие любой клавиши возобновляет вывод оглавления.

Оглавление каталога выводится в полной форме (это соответствует режиму Full в Norton Commander). Чтобы сэкономить экранное место, предусмотрена краткая форма вывода: оглавление выводится в пять колонок и содержит только имена файлов и каталогов без информации об их объеме, дате и времени создания.

Для вывода в краткой форме применяется ключ /w:

dir/w

Допускается сочетание различных ключей. Например, команда

dir/p/w

позволяет выводить оглавление в краткой форме экранными порциями.

Команда dir может содержать в качестве параметра имя каталога, если вы хотите просмотреть содержимое какого-то другого, не текущего каталога. В этом случае она имеет вид

dir *имя каталога*

К примеру, не выходя из своего каталога, вы просматриваете оглавление каталога GAMES:

dir c:\games

Имя файла можно указать также с использованием символов замены. Использование ключей при этом вполне допустимо. Например, команда

dir c:\games*.com/w

выводит на экран только файлы из каталога GAMES с расширением .COM, причем в краткой форме.

Задание списка каталогов для поиска файлов. Используется команда **path** (путь). Параметрами этой команды служит список имен каталогов, разделенных точкой с запятой:

path *имя каталога*; *имя каталога...*

Дело в том, что, если файл назван только по имени (без полного пути), операционная система ищет его только в текущем каталоге. Чтобы не утруждаться набором полных имен, можно заранее указать список каталогов для поиска файлов с помощью команды path. Пусть, например, игра в «шарики» (файл lines.exe) расположена на диске D: в каталоге GAMES в подкаталоге LINES. Если была введена команда

path d:\games\lines

то для запуска игры достаточно набрать просто lines.

15.5.2. Команды работы с файлами

Копирование. Используется команда **copy**. Обязательный параметр – имя копируемого файла (файлов). Его можно назвать параметром «откуда». Необязательный параметр – «куда» – имя каталога или файла (файлов), в который осуществляется копирование. Для группового копирования можно использовать в имени символы замены. Синтаксис команды в общем виде:

copy *«откуда» «куда»*

По команде copy копируются файлы, указанные в параметре «откуда». Если полное имя файлов не указано, то копируются файлы из т е к у щ е г о каталога.

- Если в параметре «куда» указан только каталог, то файлы копируются в этот каталог с теми же именами. Например, по команде

copy myfile.txt a:\works

файл MYFILE.TXT из текущего каталога скопируется в каталог WORKS на диске A:.

- Если в параметре «куда» указано имя файла, то файл копируется с этим именем. По команде

copy *.txt a:*.doc

все файлы с расширением .TXT текущего каталога будут скопированы в корневой каталог диска A: с расширением .DOC.

- Если второй параметр «куда» опущен, копирование осуществляется в т е к у щ и й каталог. Например, по команде

copy a:*.*

все файлы корневого каталога диска A: скопируются в текущий каталог.

Следует упомянуть одну важную особенность операционной системы MS DOS. Все основные периферийные устройства операционная система интерпретирует как файлы с заданными стандартными именами. Например, принтеру соответствует файл PRN, клавиатуре – файл CON. Благодаря этой особенности команда copy приобретает некоторые дополнительные функции.

1. *Создание текстового файла.* Эта операция интерпретируется как копирование из файла CON (с клавиатуры) в некоторый текстовый файл. К примеру, по команде

copy con bestfile.txt

в текущем каталоге будет создан файл BESTFILE.TXT. Сразу после этой команды можно ввести текст – содержимое этого файла. Ввод заканчивается одновременным нажатием клавиш $\boxed{\text{Ctrl}}$ $\boxed{\text{Z}}$, а затем клавиши $\boxed{\text{Enter}}$.

2. *Вывод файла на печать.* Распечатать файл – это значит скопировать его в файл PRN. Например, по команде

copy report.txt prn

файл REPORT.TXT из текущего каталога будет распечатан на принтере.

Переименование. Используется команда **ren** (Rename). Она имеет два обязательных параметра – старое и новое имя файла:

ren *старое имя новое имя*

В имени, разумеется, можно использовать символы ? и *. Если дисковод и путь в старом имени опущены, переименовываются файлы из текущего каталога. Например, по команде

ren *.bak *.txt

все файлы текущего каталога с расширением .BAK получат расширение .TXT.

Удаление. Используется команда **del** (Delete). Обязательный параметр – имя (имена) удаляемых файлов:

del *имя файла*

Например,

del badfile.txt

– удаление файла BADFILE.TXT из текущего каталога;

del d:\users*.bak

– удаление всех файлов с расширением .BAK из каталога USERS на диске D:.

Если вы хотите удалить все файлы текущего каталога, т.е. наберете команду del *.*, то операционная система спросит вас:

Are you sure (Y/N)?

(Вы уверены?). Для подтверждения нажмите Y и Enter, для отказа от удаления N и Enter.

Просмотр текстового файла. Используется команда **type**. Обязательный параметр – имя файла:

type *имя файла*

По этой команде содержимое текстового файла выводится на экран. Приостановить вывод можно одновременным нажатием клавиш Ctrl и S. При повторном нажатии этих клавиш вывод продолжается. Вывод файла прерывается нажатием Ctrl C.

15.5.3. Служебные команды

Ver. Операционная система MS DOS постоянно развивается. Разрабатываются все новые ее версии. Версию MS DOS принято обозначать двумя цифрами, разделенными точкой: первая цифра – номер версии, вторая – номер модификации, например MS DOS 5.2. Иногда бывает полезно узнать, какая версия установлена на вашей машине. Некоторые программы, например, не работают со старыми версиями. По команде ver на экран выводится номер установленной версии MS DOS.

Date. С помощью этой команды компьютеру можно задать текущую дату. Современные компьютеры снабжены собственными кварцевыми часами и не «забывают» дату – она хранится в специальном участке памяти, который питается от батарейки и не стирается при выключении машины. Важно, чтобы на компьютере была установлена правильная дата – тогда вы сможете проследить реальную историю создания ваших файлов. По команде date на экран выводится текущая дата и предоставляется возможность задать новую дату. По умолчанию дата выводится и вводится в «американском» формате мм/дд/гг: сначала месяц, затем число (день) и год. Например, 09/15/96 – это 15 сентября 1996 г.

Time. Эта команда позволяет задать текущее время. Как и дата, время не исчезает из памяти компьютера при его выключении.

Cls. С помощью этой команды очищается экран.

Замечание. Рассмотренные команды относятся к так называемым *внутренним* командам MS DOS, т.е. программы для их выполнения «запрятаны» в самой системе и не требуют отдельных исполняемых файлов. Наряду с внутренними имеются *внешние* команды. Для каждой из них существует отдельный файл с расширением .COM или .EXE. Обычно все эти файлы хранятся в каталоге DOS на жестком диске. Среди внешних команд можно отметить команды **tree** (для вывода на экран дерева каталогов), **format** (для форматирования дисков) и др. Следует иметь в виду, что все исполняемые файлы (файлы с расширениями .EXE, .COM и .BAT) операционная система также рассматривает как внешние команды.

15.5.4. Командный файл

Среди исполняемых файлов особое место занимают файлы с расширением .BAT. Их называют ***командными*** (или ***пакетными***, от слова «batch» – пакет). В отличие от exe- и com-файлов, bat-файлы представляют собой текстовые файлы. Соответственно, с ними можно работать как с обычными текстовыми файлами, т.е. создавать, просматривать и редактировать. В каждой строке командного файла содержится внутренняя или внешняя команда MS DOS. При запуске bat-файла эти команды последовательно выполняются. Пусть, например, файл DELBAK.BAT состоит из двух строк:

```
cls
del *.bak
```

При работе этого файла сначала очищается экран, а затем удаляются все файлы с расширением .BAK текущего каталога.

Чаще всего в корневом каталоге диска С: имеется командный файл **AUTOEXEC.BAT**. Это единственный файл, который выполняется автоматически при загрузке системы. В нем обычно содержатся команды, задающие основные параметры работы компьютера, команды, русифицирующие клавиатуру и экран, и т.д. Файл AUTOEXEC.BAT всегда содержит команду path со списком каталогов для поиска файлов, чтобы при работе не набирать полные имена наиболее часто используемых программ.

Итак, мы убедились, что можно обходиться без Norton Commander. Тем не менее ясно, что работать в Norton Commander значительно удобнее. Вспомним, что для запуска исполняемой программы в операционной системе достаточно набрать имя файла после приглашения DOS и нажать клавишу Enter. Чтобы вернуться в Norton Commander, наберите nc (имя файла NC.EXE), если, конечно, файл autoexec.bat содержит соответствующую команду path. В противном случае придется набрать полное имя (например, c:\nc\nc).

15.6. Основы работы в Windows 95

15.6.1. От Windows к Windows 95

Система **Windows** (впоследствии – Windows 3.1), разработанная в фирме Microsoft в конце 80-х годов, первоначально задумывалась как еще одна оболочка операционной системы MS DOS, надстройка над ней (нечто вроде улучшенного варианта Norton Commander). Однако довольно быстро эта система завоевала всемирную п… лярность. Все чаще на экранах мониторов можно было … вместо привычных голубых панелей эмблему Windows. Де… что Windows – это нечто большее, чем вспомогательная пр… раз- облегчающая работу на компьютере. В системе Windows реа… а- на совершенно новая концепция компьютерной среды, что … р лило предоставить много возможностей и удобств как для раз… чиков программ, так и для пользователей.

На базе этого популярного продукта фирмы Microsoft и создана новая операционная система – **Windows 95**, которая уже не нуждается в старой доброй MS DOS. Компьютеры с процессорами фирмы Intel и системой Windows (этот симбиоз называют Wintel) фактически стали международным стандартом. Сейчас, пожалуй, уже не найти компьютера, на котором не установлена система Windows 95. Продажа европейской версии началась 24 августа 1995 г., русская версия распространяется с 10 ноября 1995 г.

В чем преимущество системы Windows 95 по сравнению с другими? Коротко это преимущество можно выразить тремя словами – **единый пользовательский интерфейс**. Под *интерфейсом* обычно понимаются средства общения человека с компьютером. Тысячи программ, работавших до Windows, были снабжены своим собственным интерфейсом. Это касается вида экрана, системы меню, функций клавиш и т.д. Например, клавиша F3 в Norton Commander, как мы знаем, предназначена для просмотра текстовых файлов. В редакторе ЛЕКСИКОН эта же клавиша применяется для выделения фрагментов текста. И наоборот, одинаковые функции в разных программах выполняются по-разному: в одной программе выбор пункта меню осуществляется клавишей Enter, в другой – пробелом и т.д. Таких примеров можно привести десятки и сотни. Непрофессиональный пользователь, обычный человек, не знакомый с тонкостями программирования, разводил руками в недоумении: почему, переходя от одной программы к другой, он должен полностью переучиваться? Это приводило к путанице, ошибкам и недоразумениям. Возникла естественная задача унификации пользовательской среды, необходимо было создать такие условия, в которых самые разнообразные программы управлялись бы одинаковыми средствами. Эта задача и была решена с помощью Windows. Любая программа, работающая в среде Windows, обеспечивает сходные средства общения между пользователем и компьютером. Если вы научились работать с одной Windows-программой, вам не составит труда освоить и другую, и третью. Вы увидите на экране такие же окошки, такие же меню, одни и те же клавиши будут выполнять одинаковые функ… ии. Даже цветовая гамма на экране не будет существенно меняться …рограммы к программе.

…к, единый интерфейс безусловно удобен для пользователя. …е удобства система Windows дает и разработчикам программ, просма…ммистам. Казалось бы, удобства для работы программистов содер…лжны интересовать простого пользователя, однако на самом ке …это важно и для него: если для создания программ существуют напри…

простые и эффективные средства, в результате будет разработана качественная программа, отвечающая всем пользовательским запросам. Унифицированные элементы интерфейса – меню, окна, кнопки и т.д. – как кубики из детского конструктора, с их помощью опытный программист складывает нужную программу. Такой подход упрощает технологию программирования, позволяет писать программы быстро, качественно и в едином стиле.

Перечислим также другие особенности Windows, которые делают эту систему весьма популярной.

- **Многозадачность.** Экран монитора рассматривается в Windows как рабочий стол, на котором удобно размещены все необходимые бумаги и инструменты, все находится под рукой. За каждой программой закреплено свое окно. При этом одновременно можно запускать несколько программ (приложений). Физически на обычном компьютере программы не могут выполняться параллельно. Однако время и память делятся между задачами таким образом, что для пользователя создается полная иллюзия одновременной работы программ в различных окнах.
- ***Аппаратная независимость.*** *Старые DOS-программы достаточно* «капризны» в смысле использования периферийных устройств. Для конкретной программы требуется обычно вполне определенный тип монитора, принтера и других устройств или же необходима специальная перенастройка. При работе Windows-приложений настройку на внешние устройства выполняет сама система Windows: об этом можно не тревожиться ни пользователю, ни программисту.
- **Совместимость с DOS-программами.** Большинство DOS-программ можно запускать, не выходя из Windows. Поэтому переход к Windows не вынуждает отказываться от старых, но хорошо зарекомендовавших себя программ. Кроме того, в среде Windows 95 предусмотрен специальный режим работы – сеанс MS DOS, упрощающий работу старых DOS-программ.
- **Длинные имена файлов.** В Windows 95 разрешается давать файлам и каталогам имена длиной до 255 символов. Более того, имена могут состоять из нескольких слов (т.е. содержать пробелы) и задаваться русскими буквами.

И, наконец, еще одно преимущество Windows 95, о котором, возможно, следовало упомянуть в первую очередь: в этой среде просто приятно работать. На экране перед вами – красивые изображения с хорошо подобранными цветами. Вам не нужно помнить и вводить с клавиатуры команды и имена файлов – вместо этого вы работаете с их графическими изображениями. Удобство и простота работы – основная привлекательная черта Windows 95.

15.6.2. Основные понятия Windows 95

Итак, добро пожаловать в Windows 95! Все, с чем оперирует система, принято называть *объектами.* Объект – это программа, группа программ, файл, документ и многое другое. Любой объект изображается маленькой картинкой – *значком* (по-английски – icon). С объектами различной природы можно производить сходные действия (выделение, копирование, удаление, изменение свойств и т.д.). Это намного упрощает освоение Windows 95 и работу в этой системе.

Все управление в Windows 95 можно осуществлять с помощью мыши. На экране всегда виден курсор мыши, имеющий различную форму. Вид курсора зависит от того, в какой точке экрана он расположен. Курсор принимает вид стрелки при его переводе на меню, границы окна и т.д. Если курсор имеет вид песочных часов, это значит, что работает какая-то программа, идут вычисления или загрузка нового окна. Дальнейшие действия станут доступны тогда, когда курсор примет первоначальный вид.

Большинство операций в Windows 95 выполняется с помощью левой клавиши мыши (левая – по умолчанию, однако для левшей ее можно заменить на правую). Не будет преувеличением, если мы скажем, что почти всю работу можно сделать тремя движениями:

- *щелкнуть* мышью (click) – курсор мыши подвести к нужному объекту и кратковременно нажать на рабочую клавишу мыши; с помощью этой операции обычно выбирается объект на экране (меню, окно и т.д.);
- *дважды щелкнуть* мышью (double click) – быстро нажать на клавишу 2 раза подряд;
- *перетащить* мышью (drag and drop) – курсор установить на объект, нажать на клавишу мыши и, не отпуская ее, «передвинуть» объект на новое место.

В отличие от Windows 3.1, в Windows 95 многие функции отданы правой клавише мыши (левой для левшей). С некоторыми из таких операций мы ознакомимся чуть позже.

Мы уже говорили о том, что экран Windows 95 организован как *рабочий стол* (по-английски это называется Desktop), на котором можно удобно разместить все необходимое (рис. 15.8).

Что же мы видим на рабочем столе? Если использовать старую терминологию MS DOS, это значки, соответствующие различным каталогам и файлам. В Windows 95 применяются понятия папки и ярлыка. По сути, *папка* (folder) – синоним каталога MS DOS. Некоторые папки являются стандартными для Windows 95, остальные создаются пользователем. На рабочем столе, как правило, размещаются стандартные папки и те из пользовательских, которые наиболее часто применяются при работе. Кроме папок, на рабочем столе часто помещают специальные объекты, называемые *ярлыками*. Ярлык (Shortcut) – это значок, связанный с конкретным файлом. Дважды щелкнув на ярлыке, мы получаем мгновенный доступ к соответствующему файлу: исполняемый файл начинает работать, документ (например, текстовый файл) загружается в свою среду (в текстовый редактор). Английский термин дословно как раз и означает быстрый доступ.

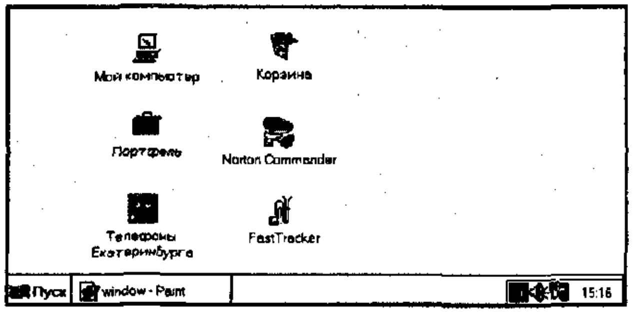


Рис. 15.8

На рабочем столе автора, изображенном на рис. 15.8, находятся следующие стандартные папки:

• **Мой компьютер** – самая большая папка, содержащая папки всех доступных дисков, а также папку для работы с принтерами и папку «Панель управления», необходимую для задания режимов работы всех устройств компьютера;

• **Корзина** – папка для удаления объектов. Удаленный файл (папка, ярлык) находится в корзине до тех пор пока корзина не очищена. Их можно «достать» из корзины и таким образом отменить удаление.

• **Портфель** – папка, необходимая в случае, когда пользователь работает то на стационарном, то на переносном компьютере и требуется периодическое обновление файлов на стационарном компьютере. Это осуществляется через портфель переносного компьютера.

Если компьютер подключен к сети, появляются некоторые дополнительные стандартные папки: «**Сетевое окружение**», «**Входящие**», «**Internet**» и др. Кроме стандартных папок, на рабочем столе находятся ярлыки нашего старого знакомого Norton Commander, а также музыкальной программы FastTracker (увлечение старшего сына автора) и справочной системы «Телефоны Екатеринбурга».

В нижней части экрана расположена так называемая *панель задач* (Task Bar). Это очень удобное средство управления работой компьютера в системе Windows 95. Панель задач содержит кнопку «Пуск», которая открывает Главное меню, а также кнопки всех работающих в данный момент приложений. На рис. 15.8 видна кнопка графического редактора Paint с загруженным в него рисунком. Чтобы переключиться в окно работающей программы, даже если его не видно на экране, достаточно щелкнуть по соответствующей кнопке на панели задач.

15.6.3. Разные окна, разные меню

Чтобы раскрыть папку, нужно дважды щелкнуть на ее значке. После этого папка раскрывается в ***окно***.

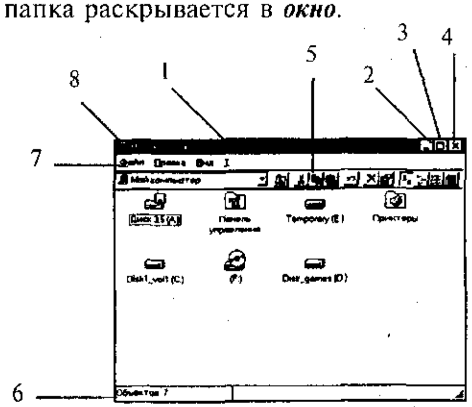


Рис. 15.9

Окно – одно из ключевых понятий Windows. Рассмотрим подробнее, что можно делать со стандартным окном Windows 95 и с помощью каких его элементов. На рис. 15.9 – раскрытое окно папки «Мой компьютер», где *1* – заголовок окна; *2* – кнопка минимизации; *3* – кнопка максимизации; *4* – кнопка закрытия окна; *5* – панель инструментов; *6* – строка состояния; *7* – меню окна; *8* – кнопка системного меню.

Окно можно:

• **переместить** – для этого нужно поместить курсор на заголовок окна (обычно на синем фоне) и перетащить мышью;

• **изменить размеры** – переместить курсор на границу окна, при этом он примет форму двусторонней стрелки; затем перетащить мышью границу, добиваясь нужного размера; если вы «тащите» за угол окна, то одновременно меняются размеры по вертикали и горизонтали;

• **максимизировать** (развернуть на весь экран) – для этого нужно щелкнуть кнопку максимизации в правом верхнем углу; после максимизации появляется кнопка, с помощью которой окно возвращается к исходным размерам;

• **минимизировать** (оставить только кнопку на панели задач) – щелкнуть кнопкой минимизации;

• **просмотреть** – если размеры окна таковы, что вся информация в него не помещается, появляются *полосы прокрутки* (вертикальная и/или горизонтальная); чтобы просмотреть оставшуюся часть окна, нужно нажать на соответствующую кнопку прокрутки (т.е. щелкнуть на ней мышью), указатель в полосе прокрутки показывает, какая часть окна просмотрена; если щелкнуть мышью в полосе прокрутки, окно будет прокручиваться быстрее;

• **закрыть** – щелкнуть по кнопке закрытия окна.

Большинство окон имеют свое ***меню.*** Строка меню расположена под заголовком окна. Стандартные пункты меню – «Файл» (работа с файлами, конкретное содержание зависит от программы), «Правка» (редактирование), «Вид» (задание внешнего вида окна), «?» (помощь). Выбрать пункт меню можно по-разному:

нажать $\boxed{\text{Alt}}$ или $\boxed{\text{F10}}$, клавишами управления курсором подвести его к нужному пункту и нажать на $\boxed{\text{Enter}}$;

вызвать по клавише быстрого выбора (hotkey): нажать вместе с клавишей $\boxed{\text{Alt}}$ клавишу подчеркнутой буквы;

дважды щелкнуть мышью на нужном пункте.

Отказаться от выбора можно с помощью клавиши $\boxed{\text{Esc}}$. Чаще всего после выбора пункта в горизонтальном меню появится вертикальное подменю: в Windows реализована та же идея рассыпающихся меню (Pulldown menu), что и в Norton Commander. Однако благодаря возможностям графического изображения меню могут выглядеть гораздо привлекательнее. Пункты меню в различных Windows-программах заданы не текстом, а картинками, значками.

Не все пункты меню могут быть доступны в данный момент. Например, если файл не открыт, мы не можем его переименовать. Такие недоступные пункты выводятся бледными буквами.

Рядом с пунктом меню часто указываются клавиши быстрого вызова, с помощью которых можно выполнить это действие, не прибегая к меню. Выбор пункта с треугольником приводит к вызову подменю. Если после пункта меню стоит многоточие, это означает, что после выбора этого пункта откроется *диалоговое окно*, предназначенное для запроса и ввода информации.

Галочками (флажками) отмечаются пункты, задающие различные режимы работы: режим можно включить или отключить. На рис. 15.10 – меню «Вид». Флажками отмечены пункты «Панель инструментов» и «Строка состояния» (их можно убрать из окна). *Панель инструментов* (Tool Bar) – набор кнопок, с помощью которых можно дублировать основные операции с окном и объектами, расположенными в окне. Кстати, это характерная черта Windows 95; большинство операций осуществляются р а з н ы м и способами. *Строка состояния* (Status Bar) – последняя строка окна, в которой находится информация об объектах окна и текущих действиях.

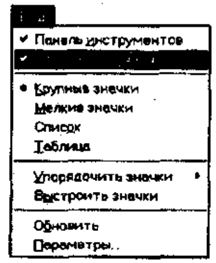


Рис. 15.10

Кружки (●) – выбор одной возможности из нескольких. Например, объекты в окне могут быть представлены четырьмя способами: крупными или мелкими значками, списком или таблицей, в зависимости от того, насколько подробную информацию о них мы хотим видеть. Выбрать режим можно и с помощью одной из четырех кнопок в правой части панели инструментов.

Важным элементом Windows 95 является так называемое *контекстное меню*, возникающее на экране при щелчке правой кнопкой мыши (рис. 15.11). Меню называется контекстным потому, что его вид зависит от того, на каком именно объекте или в каком месте экрана вы щелкнули мышью. Почти все контекстные меню содержат пункт **«Свойства»** – удобное средство для просмотра и изменения свойств объекта. Набор доступных свойств зависит от вида выбранного объекта: у папки – одни свойства, у ярлыка – другие и т.д.

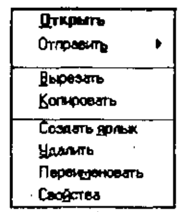


Рис. 15.11

Окно «Свойства» представляет собой разновидность *диалогового окна*, предназначенного для ввода информации пользователем. Перечислим некоторые стандартные элементы диалогового окна:

- **флажок** (Check Box) ✓ для включения/выключения некоторого режима;
- **переключатель** (Option Button) для выбора одного режима из нескольких (эти элементы нам уже знакомы по меню);
- **список** (List Box) для выбора объекта (как правило, файла или папки);
- **скрытый список** – список, который раскрывается при нажатии кнопки с треугольником (рис. 15.12);

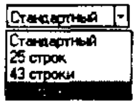


Рис. 15.12

- **строка ввода** (Text Box) для ввода текста (чаще всего имени файла) (рис.15.13);

Рис. 15.13

- **ползунок** (Slider) – средство для выбора значения параметра в пределах некоторого интервала (например, с помощью ползунка можно задать скорость перемещения указателя мыши) (рис. 15.14);

Рис. 15.14

- **счетчик** (Spin Box) – позволяет установить нужное числовое значение: щелчок по верхней кнопке увеличивает значение на единицу, щелчок по нижней – уменьшает (рис. 15.15). К примеру, с помощью счетчика устанавливается текущее время.

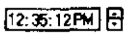


Рис. 15.15

За кнопкой «Пуск» скрывается *Главное меню*, с помощью которого можно осуществить основные операции в Windows 95. Как правило, Главное меню содержит следующие пункты:

- **Программы** – доступ к папкам с наиболее часто используемыми программами. Меню «Программы» обычно содержит пункты

«**Стандартные**» (набор удобных программ, таких, как Калькулятор, текстовый редактор Блокнот, графический редактор Paint, текстовый редактор WordPad и др.), «**Сеанс MS DOS**» (для перехода в режим DOS), «**Автозагрузка**» (программы, запускаемые автоматически при загрузке Windows 95), а также другие группы программ, которые пользователь может внести в это меню.

- **Документы** – доступ к документам (текстовым, графическим, музыкальным и другим файлам), с которыми пользователь работал в последнее время. Список может содержать до 15 документов.
- **Настройка** – изменение параметров устройств компьютера и панели задач.
- **Поиск** – поиск нужной папки или файла.
- **Справка** – доступ к обширной справочной системе Windows 95.
- **Выполнить** – запуск нужной программы. Если вы не помните точное имя, можно воспользоваться кнопкой «Обзор».
- **Завершение работы** – вызывается перед выключением компьютера.

15.6.4. Операции над объектами. Работа с файлами и папками

С любыми объектами в Windows 95 (значками, ярлыками, папками, файлами) осуществляются сходные действия. С помощью контекстного меню объекты можно **создать**, а также **переименовать**. Прежде чем работать с объектами, их можно **выделить**. Чтобы выделить один объект, по нему достаточно щелкнуть мышью. Если вы хотите выделить группу объектов, расположенных подряд, на последнем объекте нужно щелкнуть мышью одновременно с нажатием клавиши Shift. Следующие группы выделяются щелчком вместе с клавишами Ctrl (на первом объекте) и Ctrl – Shift (на последнем).

Чаще всего выделяются группы файлов или папок. Это можно делать либо в окне «Мой компьютер», либо с помощью специального средства «**Проводник**» (оно доступно, например, через контекстное меню кнопки «Пуск»). Окно «Проводника» состоит из двух панелей (рис. 15.16). На левой панели представлены все папки, доступные на компьютере (этакое «супердерево» со всеми дисками и устройствами); на правой – содержимое текущей раскрытой папки. Плюсом отмечены папки, имеющие подпапки: щелкнув по плюсу, их можно раскрыть дальше (тогда плюс становится минусом, щелчок по которому сворачивает данную ветку). Если на правой панели отмечены некоторые файлы, их, как и в Norton Commander, можно **скопировать** (просто потащить мышью в нужную папку на правой панели); **переместить** (потащить с нажатой клавишей [Alt]); **удалить** (перетащить в корзину или нажать клавишу [Del]).

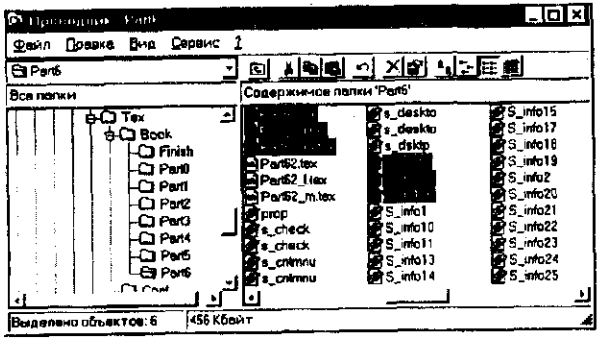


Рис. 15.16

Кроме того, любые выделенные объекты можно **скопировать** или **вырезать** в *буфер обмена*. В дальнейшем объекты можно **вставить** из буфера. Буфер обмена (Clipboard) – универсальное средство Windows 95 для обмена информацией между различными окнами и программами, область памяти, в которой хранится информация о файлах, участках текста, рисунках и т.д.

Мы ознакомились с основными приемами работы в Windows 95 (см. Приложение 2). Навыки можно приобрести лишь в ходе общения с компьютером на практике. В заключение посоветуем почаще обращаться к справочной системе Windows 95, чтобы получить квалифицированную и своевременную помощь. Включайте компьютер – и за работу!

Упражнения и вопросы

1. Выбрать правильный ответ. Операционная система MS DOS – это: а) устройство; б) программа; в) файл.

2. Какие из следующих имен файлов допустимы в MS DOS, а какие нет: MYFILE.TXT, MY FILE.DOC, ME&YOU.(2), 2.FILE.EXE, AAAAAAAAA.BBBB, ONE_WAY.STR, МОЙФАЙЛ.XYZ, @1234567.#89, AbCd.EfG.

3. Какую команду MS DOS следует выполнить, чтобы:

а) скопировать все файлы с расширением .txt из каталога STUDENT на диске C: в текущий каталог;

б) просмотреть каталог USERS диска D: в краткой форме;

в) переименовать все файлы текущего каталога с расширением .old в файлы с расширением .new;

г) создать каталог MYCAT в текущем каталоге;

д) выяснить версию MS DOS?

4. Что означает команда

copy con d:\users\catl\bestfile.txt ?

5. Файл BALDA.COM находится в каталоге GAMES на диске C:. Какая команда должна содержаться в файле AUTOEXEC.BAT, чтобы загрузить игру командой balda из текущего каталога?

6. Какая комбинация клавиш перезагружает компьютер?

7. Какие операции можно производить с любым файлом?

8. Объяснить назначение клавиш [F3], [F4], [F5], [F6], [F7], [F8] в Norton Commander.

9. Какие действия можно осуществлять с окном Windows 95 и каким образом?

10. Что происходит, если информация не помещается в окно Windows 95?

Глава 16
КЛАССИФИКАЦИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

В предыдущей главе рассматривались различные сферы применения компьютера. Каждая специфическая задача требует своего адекватного программного обеспечения. В настоящий момент в мире разработаны тысячи программ на любой случай, поэтому разумно провести классификацию программ в соответствии с теми задачами, которые эти программы решают. Следует иметь в виду, что невозможно перечислить все существующие программы той или иной группы. В этом, впрочем, нет и необходимости – большинство пользователей применяют весьма ограниченный круг программ, хорошо зарекомендовавших себя в своей области. Это добротные программы, созданные в солидных фирмах, которые производят программное обеспечение. Над их разработкой трудился большой коллектив специалистов. В каждой группе мы назовем лишь несколько программ, проверенных временем и опытом пользователей. Мы будем указывать, в какой среде работает данная программа, т.е. специально называть Windows-программы как наиболее перспективные и развивающиеся.

16.1. Текстовые редакторы

Уже упоминалось, что работа с текстами – самая распространенная область применения компьютера. Практически любой пользователь – как профессиональный, так и непрофессионал, встречается с необходимостью подготовки тех или иных документов. Это могут быть письма, статьи, отчеты, справки, рекламные проспекты и т.д. Все эти документы создавались и раньше, без применения компьютера. И до сих пор не все перешагнули через психологический барьер и отказались от пишущей машинки в пользу компьютера. Однако можно с уверенностью сказать, что подготовка текстов без использования компьютера в наши дни выглядит анахронизмом. Те, кто знаком с рынком секретарских рабочих мест, знают, что умение пользоваться компьютером – непременное условие при приеме на работу. Большинство ученых, посылая статью в редакцию, больше не вписывает формулы черными чернилами, а пользуется соответствующими программами подготовки научных текстов.

Компьютер предоставляет простые и удобные средства для обработки текстов, делает эту рутинную работу более приятной и эффективной. Потребности компьютерной подготовки документов привели к созданию огромного количества специальных программ. Их называют *текстовыми редакторами* (Word Processors). Всего существует несколько сотен текстовых редакторов. По назначению различают следующие их разновидности:

- редакторы программ;
- редакторы документов;
- издательские системы;
- редакторы научных текстов.

Возможности этих программ различны – от подготовки небольших документов простой структуры до набора, оформления и издания книг и журналов. Тем не менее многие из этих программ обладают общими свойствами, позволяют выполнять одинаковые операции с текстом. Хороший текстовый редактор обязательно имеет меню (пункты которого заданы текстом или пиктограммами), систему «горячих клавиш» для основных операций и помощь.

Редакторы программ предназначены не для непрофессиональных пользователей, а для программистов. Эти редакторы помогают в написании программ на одном из языков программирования (Си, Паскаль и др.). Редакторы научных текстов в основном удовлетворяют потребностям представителей точных наук – математиков, физиков, химиков, инженеров и т.д., чьи тексты изобилуют формулами. Поэтому мы не будем подробно останавливаться на особенностях этих программ. Наше внимание будет сконцентрировано на редакторах документов как наиболее подходящих для непрофессиональных пользователей и специалистов-гуманитариев. Мы отметим также некоторые черты издательских систем.

16.1.1. Редакторы документов

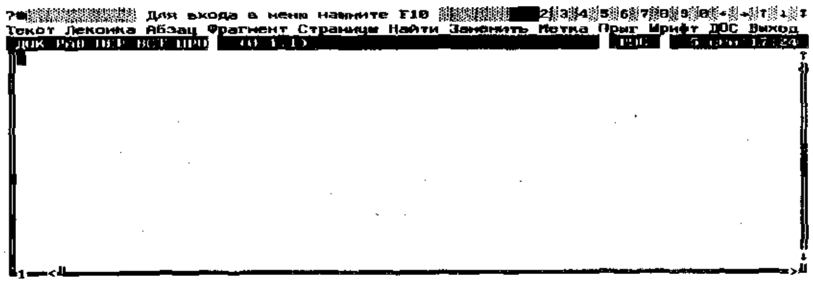
В качестве простейшего средства редактирования можно применять встроенный редактор Norton Commander (помните клавишу F4?). Среди наиболее распространенных в мире назовем Microsoft Word, WordPerfect, WordStar, Multi-Edit, встроенные редакторы Windows 95 – Блокнот и WordPad. В России большое распространение получил редактор ЛЕКСИКОН. Это программный продукт российского производства (созданный коллективом под руководством Е.Н. Веселова), а не русифицированная программа (как, например, русская версия Word). Интерфейс на русском языке, а также возможности обработки текста привлекают к этой программе многих непрофессиональных пользователей. ЛЕКСИКОН вполне подойдет тем, кому нужен простой инструмент для подготовки небольших и несложных документов, не требующих высокого полиграфического качества.

Рассмотрим некоторые стандартные действия с текстовыми файлами.

Перемещение по тексту. В большинстве текстовых редакторов экран представляет собой поле для редактируемого текста. В это поле загружается содержимое существующего файла или вносится новый текст. Текст, разумеется, состоит из символов. Под *словом* понимается набор символов между двумя пробелами. Текст в текстовом файле разбит на *строки*, каждая строка заканчивается невидимым символом конца строки. В поле редактирования всегда имеется курсор, отмечающий текущую позицию, текущий символ текста. Чаще всего позиция курсора (номер строки и столбца) отображается на экране. По тексту можно перемещаться во все стороны, листать страницы, возвращаться к началу и переходить к концу файла.

Во многих редакторах возможно быстрое перемещение к заданной строке текста. На рис. 16.1, *а* и *б* представлены соответственно экраны редактора ЛЕКСИКОН и редактора Microsoft Word 7.0.

а)



б)

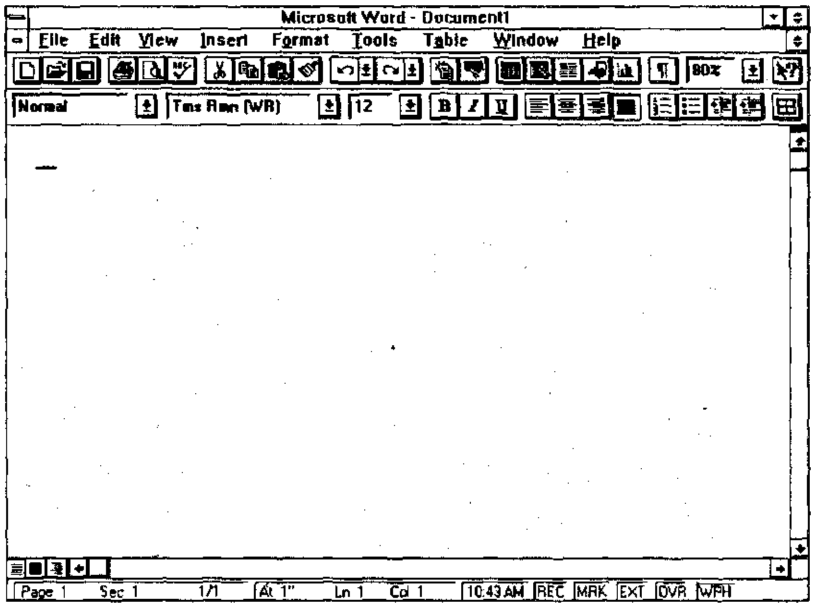


Рис. 16.1

Режимы редактирования.

• **Режим вставки/замены** (Insert/Overwrite). Об этих режимах мы упоминали при описании клавиатуры. В режиме вставки при наборе символа строка раздвигается вправо, а в режиме замены вновь набранный символ вписывается поверх имеющегося символа в текущей позиции. В большинстве редакторов по умолчанию принят

режим вставки, переключение режимов осуществляется клавишей $\boxed{\text{Ins}}$. Обычно можно определить, какой выбран режим, по виду курсора или по сообщению на экране.

• **Автоматический перевод строки** (AutoIndent). Если этот режим выключен, перевод строки осуществляется вручную, чаще всего клавишей $\boxed{\text{Enter}}$. Режим автоматического перевода позволяет вводить текст, не заботясь о новой строке. При этом строка может переводиться с переносом слов (если редактор имеет эту функцию) либо без переноса, т.е. целыми словами.

• **Выравнивание краев.** Чаще всего применяется автоматическое выравнивание правого края. В этом режиме редактор вставляет в строку нужное количество пробелов, чтобы все строки были одинаковой длины.

Операции корректировки текста. Большинство редакторов выполняют следующие операции:

- удаление символа;
- удаление слова;
- удаление строки;
- удаление от текущей позиции до конца строки;
- удаление от текущей позиции до конца файла;
- отмена последнего удаления (unkill, undelete, undo);
- расщепление строки;
- «склеивание» строк.

Работа с блоками. Под *блоком* (или *фрагментом*) понимается участок текста редактируемого файла. Различают два основных вида блоков: строчные и прямоугольные. *Строчный* блок состоит из нескольких идущих подряд целых строк, а *прямоугольный* – это просто прямоугольная область на экране. Прежде всего блок **отмечают**: на экране отмеченный блок выделяется цветом фона и символов. Дальнейшие действия с блоком очень напоминают действия с файлами. Блок можно:

- копировать;
- переносить;
- удалять.

Место для копирования или переноса блока определяется текущей позицией курсора. При копировании блок остается на старом месте и появляется на новом, при переносе – исчезает со старого места. Результат операций с блоками зависит от того, в каком режиме находится редактор – вставки или замены.

В режиме вставки при копировании/переносе текст раздвигается (для строчного блока – вниз, для прямоугольного – вправо), при удалении текст сжимается (для строчного – вверх, для прямоугольного – влево).

В режиме замены при копировании/переносе блок оказывается поверх существующего текста, а при удалении на месте блока остается пустое место.

На рис. 16.2 виден результат удаления прямоугольного блока в режиме вставки.

```
В лесу родилась елочка,   родилась елочка,
В лесу она росла.         она росла.
Зимой и летом стройная,    летом стройная,
Зеленая была.              была.
```

Рис. 16.2

Контекстный поиск (поиск с заменой). С помощью этой функции можно найти в тексте заданное слово или фразу («контекст»). Существуют различные режимы поиска. В частности, поиск может осуществляться целыми словами. Если вы задали контекст «файл», то в этом режиме редактор будет искать только слово «файл». Если же этот режим поиска отключен, то будут найдены все слова с этим сочетанием букв («файловый», «файлы» и т.д.).

При поиске с заменой надо указать фразу для поиска и фразу для замены. При этом редактор может каждый раз спрашивать – заменять или нет, либо заменить все без спроса (глобально). Например, вы хотите во всем тексте заменить слово «коммунистический» на слово «демократический». Тогда можно смело выбирать режим глобальной замены.

Форматирование текста. Любой текст имеет несколько параметров, основными из которых являются левая и правая граница, а также, возможно, величина красной строки. При форматировании текста его параметры приводятся в соответствие с заданными. При этом может быть отформатирован либо весь текст, либо текущий абзац, либо выделенный блок.

Использование различных шрифтов. Эта функция очень удобна при необходимости красивой печати документа. В простейшем варианте это модификация одного и того же шрифта с различным выделением символов – подчеркивание, курсив или жирный шрифт. В более сложных редакторах поддерживаются различные шрифты – как латинские, так и русские. Поскольку система Windows по своей

природе является графической оболочкой, наибольшее количество шрифтов поддерживается именно Windows-редакторами. На рис. 16.3 приведен фрагмент текста, набранного в редакторе Microsoft Word с использованием разных шрифтов.

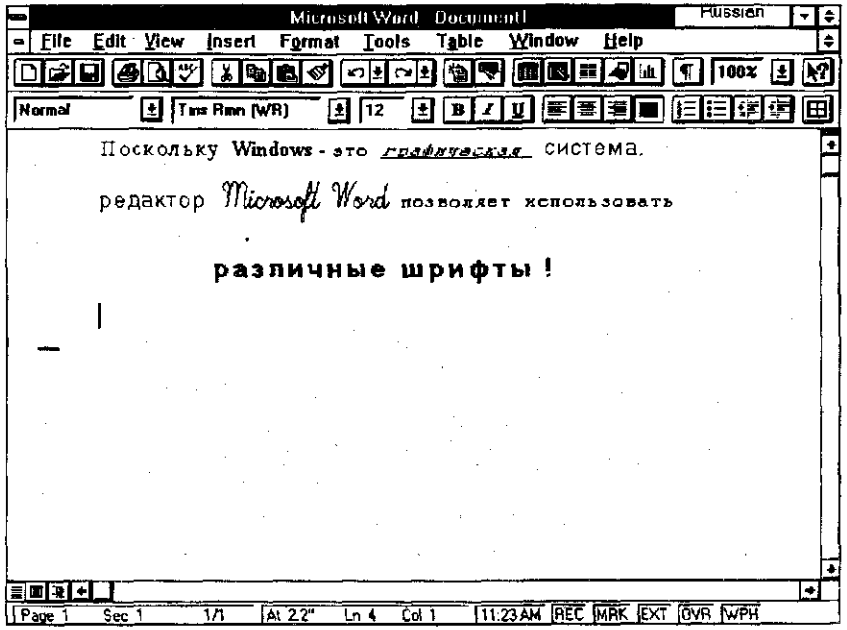


Рис. 16.3

Работа со страницами. При выводе документа на принтер важно, каким образом пронумерованы страницы. Большинство редакторов автоматически разбивают текст на страницы. Однако часто это бывает неудобно, поэтому предусматривается возможность разбить на страницы вручную. Редакторы автоматически нумеруют страницы. При этом можно указать, с какого номера начинается нумерация, сколько оставить ненумерованных страниц и т.д. Часто можно ввести автоматическую печать верхних и нижних заголовков страниц (*колонтитулов*). Таким образом, текстовые редакторы позволяют определить, в каком виде документ следует вывести на печать.

Одновременная работа с несколькими документами. Иногда удобно работать с несколькими текстами сразу. В большинстве редакторов это можно осуществить за счет оконной системы: разные файлы загружаются в разные окна редактора. Важно, что при этом допускается копирование и перенос блоков между различными файлами. Возможны дополнительные операции с текстовыми файлами:

- обработка и нумерация сносок;
- набор текста в несколько столбцов;
- создание таблиц и построение диаграмм;
- вставка рисунков в текст;
- проверка правописания;
- построение оглавлений, индексов и т.д.

16.1.2. Издательские системы

Для подготовки рекламных буклетов, оформления журналов и книг используются специальные издательские системы. Они позволяют готовить для типографской печати сложные документы высокого качества. Издательские системы обычно могут работать с текстовыми файлами, набранными с помощью редакторов документов. Благодаря специальным средствам можно превратить текст в макет будущего издания. Основная операция, для которой используются издательские системы, – это *верстка*, т.е. размещение текста по страницам документа, вставка рисунков, оформление текста разными шрифтами и т.д.

В основном издательские системы построены по принципу WYSIWYG (от английской фразы «What You See Is What You Get» – «Что вы видите, то и получите»). В этих системах документ на экране выглядит точно так же, как и при выводе на печать.

В качестве наиболее популярных издательских систем можно назвать **Aldus PageMaker** и **Ventura Publisher** (сейчас – Corel Ventura). Следует отметить, что современные редакторы документов приближаются по своим свойствам к издательским системам. В частности, Microsoft Word для Windows также построен по принципу WYSIWYG и позволяет готовить к печати весьма сложные и высококачественные документы. Для набора книг, журналов, газет со сложной структурой текста разумно использовать издательские системы.

16.2. Графические редакторы

Графическим редактором называется программа, позволяющая создавать и редактировать картинки на экране компьютера. Типичный представитель таких программ – **Paint**, включенный в набор стандартных программ Windows 95. На рис. 16.4 показан результат работы младшего сына автора в этом графическом редакторе.

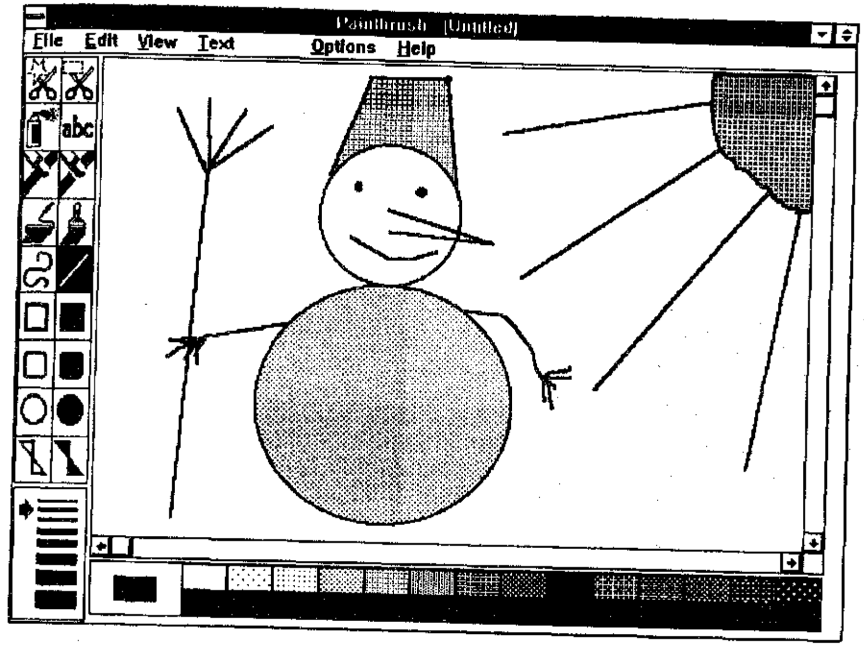


Рис. 16.4

Графический редактор обычно имеет несколько меню, с помощью которых пользователь может выбрать:

- цвет;
- тип и толщину линии;
- тип закраски;
- шрифт для надписей;
- стандартные элементы рисунка (круги, прямоугольники и т.д.).

Кроме того, можно выполнить стандартные действия с рисунком:

- выделить область (для копирования, переноса или удаления);
- закрасить («залить») замкнутую область рисунка;
- увеличить/уменьшить рисунок или его часть;
- «склеить» несколько рисунков;
- стереть часть рисунка и т.д.

Пункты меню в графическом редакторе обычно изображаются в виде понятных пиктограмм (карандаш, резинка, клей, ножницы, распылитель краски и т.д.). Вся работа в графическом редакторе чаще всего осуществляется с помощью мыши.

Следует отметить, что хороший графический редактор позволяет обрабатывать изображения, полученные путем сканирования. Кроме того, рисунки, созданные в графических редакторах, записываются в файлы стандартного формата. Эти файлы можно в дальнейшем включать в тексты с помощью текстовых редакторов и издательских систем. К стандартным графическим файлам относятся файлы с расширениями .PCX, .BMP, .GIF и др.

Профессионалов из издательств или рекламных агентств не удовлетворит редактор Paintbrush. Им потребуется более мощная программа, например **Corel Draw** или **Adobe PhotoShop**, позволяющая проделывать с рисунками все мыслимые преобразования (ретуширование, сглаживание линий и многое другое), работать с трехмерными изображениями и т.д.

16.3. Системы управления базами данных

Как уже говорилось, часто приходится иметь дело с огромными массивами информации, с данными, между которыми существуют разнообразные связи и зависимости. Для хранения и обработки таких массивов, безусловно, потребовалась автоматизация, применение компьютера. В 70-х годах была найдена чрезвычайно удобная форма организации данных для компьютерной обработки, получившая название *базы данных* (Database). Идея базы данных основана на таком привычном и простом способе записи информации, как т а б л и ц а. В самом деле, четкая и несложная структура (строки и столбцы) делает таблицу практически универсальным средством отображения информации любой природы. В одну таблицу можно свести сведения самого разнообразного характера – имена, числа, даты и многое другое. Можно, к примеру, составить таблицу, содержащую данные о студентах факультета:

Фамилия	Номер группы	Дата рождения	Семейное положение
Иванов И.И.	ПС-101	13.03.81	Холост
Петрова А.П.	ПС-101	25.10.82	Замужем
Сидоров П.С.	ПС-102	10.01.81	Холост
...	...	...	...

Разумеется, в такую таблицу можно включить и любые другие сведения о студентах. Если структура таблицы задана и известно, какие

данные следует поместить в каком столбце, дальнейшее заполнение таблицы не составит труда. С такой таблицей легко работать: добавлять данные о новых студентах; вычеркивать соответствующие строки, если студент отчислен; располагать строки в определенном порядке (например, по алфавиту) и т.д. Итак, таблица – удобная и достаточно универсальная структура для хранения и обработки информации.

В главе, посвященной основаниям математики, мы знакомились с таким важным понятием, как отношение. Рассматривались только отношения между парами объектов, т.е. бинарные отношения. Однако можно ввести отношение между любым количеством предметов, *n*-арное отношение. Не вдаваясь в математические подробности, скажем, что таблица как раз и задает такое отношение, а каждая ее строка – это совокупность предметов, находящихся между собой в данном отношении. В нашем примере отношение «студент» задает связь между четырьмя элементами – фамилией, номером группы, датой рождения и семейным положением. Каждая строка, соответствующая конкретному студенту, есть представитель, отдельный «экземпляр» данного отношения. Именно простая структура таблицы и ее математическая сущность как отношения (по-английски – relation), позволили Е.Ф. Кодду в 1970 г. сформулировать концепцию **реляционной базы данных**.

Фактически реляционная база данных – это таблица, специальным образом организованная в виде файла. При переходе от таблицы к базе данных меняется терминология. Строка в базе данных называется *записью*. Каждая запись состоит из *полей* (элементов столбцов). По-английски запись называется record, а поле – field. Например, в приведенной выше таблице каждому студенту соответствует одна запись, состоящая из четырех полей.

Обычная таблица имеет «шапку», которая, по существу, определяет структуру таблицы: каждый столбец имеет свое название, по которому ясно, во-первых, какого характера данные заносятся в этот столбец, и, во-вторых, какой он должен быть ширины. Например, понятно, что для даты достаточно отвести восемь позиций, для номера группы – скажем, шесть, ну а для фамилии – с запасом, позиций пятнадцать-двадцать. Шапке таблицы соответствует *структура* базы данных, которая представляет собой описание полей каждой записи. В структуре базы для каждого поля задается

- имя (name);
- тип (type);
- ширина (width).

Данные в таблицах могут иметь совершенно различную природу. Бывают данные описательного характера – это слова или фразы, короче говоря, текст. Этим данным соответствуют поля *символьного* типа. Другая очевидная разновидность данных – числа, порождающие поля *числового* типа. Кроме того, как отдельный тип данных можно выделить *даты*. Наконец, специально выделяются данные типа «да/нет». К примеру, семейное положение можно было бы задавать именно таким образом. Если человек состоит в браке – писать «да», не состоит – «нет». Обозначать такие данные можно по-разному: плюс – минус, единица – ноль или еще как-нибудь. Важно лишь, что в таких столбцах может быть одно из двух возможных значений. Таким данным отвечают поля *логического* типа. Итак, поля базы данных могут иметь один из следующих типов:

- символьный (Character);
- числовой (Numeric);
- дата (Date);
- логический (Logical).

Если переписать нашу таблицу в виде

Фамилия	Номер группы	Дата рождения	Семейное положение
Иванов И.И.	101	13.03.81	–
Петрова А.П.	101	25.10.82	+
Сидоров П.С	102	10.01.81	–
...	...	...	...

мы получим базу данных с полями всех четырех типов.

Имена полей похожи на имена файлов, только в различных системах накладываются разные ограничения на длину имени и допустимые символы. Скажем, создавая базу данных в соответствии с нашей таблицей, можно задать следующую структуру (описание полей):

Номер	Имя	Тип	Ширина
1	NAME	C	20
2	GROUP	N	3
3	BIRTHDAY	D	8
4	FAMILY	L	1

Заметим, что для полей числового типа кроме ширины обычно указывается число знаков после запятой (dec). Если это число равно нулю, значит, в данное поле заносятся только целые числа. Перечисленные четыре типа полей составляют основу любой системы для работы с базами данных.

В некоторых системах вводятся, кроме названных, некоторые дополнительные типы данных. Это может быть, к примеру, некий достаточно длинный текст (биография или трудовая книжка). Такие длинные тексты выделяются в особый тип данных – *комментарии*, или *memo-поля*. Некоторые системы позволяют в качестве данных хранить и обрабатывать графические образы, порождающие поля типа *Picture*.

Для работы с базами данных предназначены специальные программы, которые получили название ***систем управления базами данных*** (СУБД). Большинство современных СУБД обладают двоякой природой. С одной стороны, это средства обработки информации, организованной в виде баз данных. С этой точки зрения такие программы представляют несомненный интерес для пользователя любой специальности. С другой стороны, СУБД – это инструмент для написания специальных программ обработки баз данных на особом языке программирования. Эта сторона СУБД интересна в основном программистам, а непрофессионалы имеют дело уже с результатом их труда. Часто пользователи, запуская какую-либо программу, даже не догадываются, что в ней, по существу, происходит обработка нескольких сложных баз данных. Тем не менее полезно иметь представление об основных операциях с базами данных:

- создание;
- просмотр;
- корректировка;
- добавление/удаление записей;
- поиск данных в базе;
- сортировка/упорядочение.

Наряду с этими стандартными операциями могут быть реализованы некоторые дополнительные возможности. К примеру, на основе информации из базы данных могут быть сформированы и выданы на печать различные документы (отчеты, справки и т.д.). Возможна арифметическая обработка числовых полей базы: вычисление средних, суммарных значений и другие бухгалтерские функции. Кроме того, развитые СУБД поддерживают работу с несколькими базами данных одновременно, позволяют установить связь или зависимость между ними, что на самом деле отражает реальную структуру информации.

Среди СУБД, доступных обычному пользователю, для простейших операций с базами данных можно упомянуть программы PC-File, Reflex, Q&A. К достаточно сложным СУБД относятся известные программы серии DBase фирмы Ashton-Tate (позднее приобретенные компьютерным гигантом – фирмой Borland). Эти программы были одним из первых инструментальных средств для реализации сложных алгоритмов обработки информации. Широко распространена совместимая с DBase система FoxPro (в настоящее время принадлежащая фирме Microsoft). Сейчас разработана версия этой системы для работы в среде Windows. Можно также назвать системы Clipper, Paradox (с версией под Windows), RBase, Clarion.

Отметим, что все эти программы недостаточно просты для применения непрофессиональными пользователями, хотя их развитие идет именно в этом направлении: совершенствуются их интерфейс, система меню, помощи, «горячих клавиш» и т.д. Короче говоря, они становятся доступнее для обычных пользователей.

На рис. 16.5 показана операция создания базы данных о студентах на этапе определения структуры, описания полей, в системе FoxPro 2.5. Стандартный файл базы данных имеет расширение .DBF (Database File). В данном случае создается файл STUDENT.DBF.

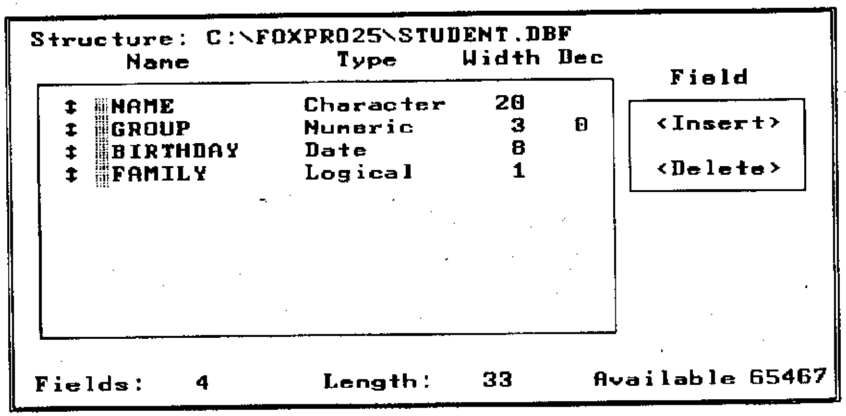


Рис. 16.5

16.4. Электронные таблицы

Под ***электронной таблицей*** понимают программу, предназначенную для работы с большими таблицами чисел. Такие программы иногда называют *табличными процессорами.* Они оперируют с данными более простой и однородной природы, чем в СУБД, однако привлекают пользователей наглядностью ввода и вывода информации, возможностями быстрого пересчета результатов при изменении данных и т.д.

При работе с электронной таблицей на экран выводится прямоугольная таблица, состоящая из *клеток* (cells). Клетки, как правило, нумеруются буквами по столбцам и цифрами по строкам. К примеру, клетка A1 – левая верхняя клетка таблицы. Чаще всего таблица бывает значительно больше размеров одного экрана. Поэтому по ней можно перемещаться во все стороны с помощью клавиш управления курсором. На экране мы видим текущий фрагмент таблицы. В клетках могут находиться:

- числа;
- пояснительный текст;
- формулы.

Последняя возможность означает, что содержимым клетки может быть некоторая формула для вычисления по данным (числам), расположенным в других клетках. При этом на экран отображается не формула, а результат расчета по ней. Как только меняются исходные данные в клетках-аргументах, сразу же пересчитывается значение в клетке с формулой. Именно эта особенность делает электронные таблицы столь привлекательными.

Предположим, вам необходимо регулярно вычислять какую-то величину по одной и той же формуле, но с меняющимися исходными данными. Скажем, вы считаете подоходный налог со своего заработка за последние три месяца. Вы вносите новые данные в электронную таблицу, а нужное вам число пересчитывается автоматически.

Среди наиболее популярных табличных процессоров можно назвать программы Lotus 1-2-3, Quattro Pro, различные версии SuperCalc и Microsoft Excel (под Windows).

На рис. 16.6 приведен фрагмент экрана русифицированной версии Excel. В таблице имеются клетки всех трех типов: числа в клетках A1, B1 и C1; текст в клетках E1, F1 и E3, F3; формулы для суммы и произведения соответственно в клетках G1 и G3. Как только одно из чисел будет изменено, значения суммы и произведения немедленно будут пересчитаны.

Заметим, что набор доступных функций в электронной таблице Excel достаточно широк. Он включает основные арифметические, тригонометрические, а также некоторые бухгалтерские и статистические функции. В частности, можно посчитать статистические характеристики выборки (среднее, дисперсию, СКО и др.), значения типичных функций распределения, квантили и многое другое.

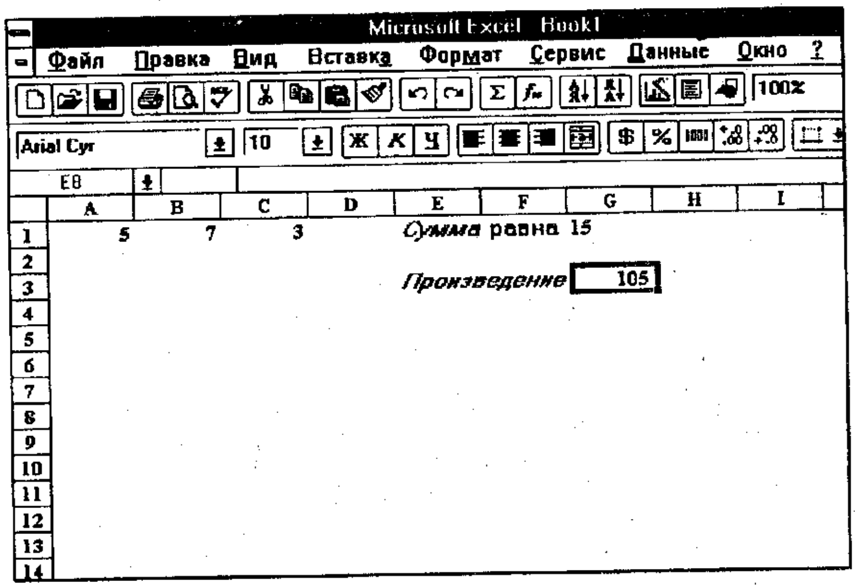


Рис. 16.6

Большинство современных электронных таблиц предоставляют пользователям дополнительные возможности. В частности, по числовым данным в клетках программа строит различные графики. На рис. 16.7 изображена парабола, построенная по значениям в клетках A1–A10.

Табличный процессор Excel, к примеру, дает широкий выбор графиков: обычный график функции, гистограмма, столбиковая и круговая диаграммы и т.д. Графики размещаются в нужной части таблицы, хорошо иллюстрируя числовые данные. Удобство состоит в том, что графики, так же как и формулы, пересчитываются и перерисовываются при изменении исходных данных.

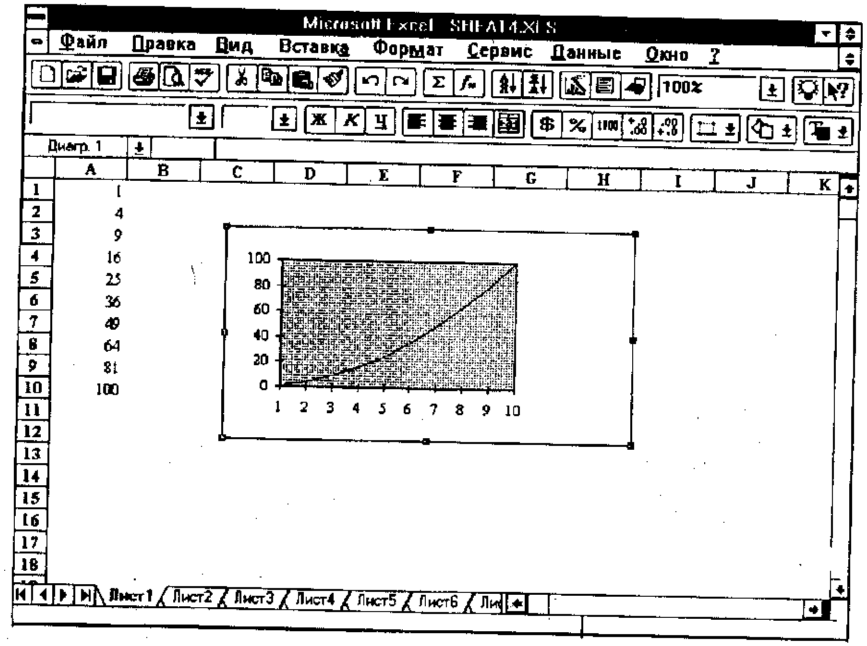


Рис. 16.7

16.5. Архиваторы

Как уже отмечалось, необходимо регулярно копировать наиболее часто используемые файлы с винчестера на дискеты. Ведь потеря информации может произойти по самым разным причинам – магнитный диск может физически испортиться, вы можете стереть нужный файл по ошибке, информация может быть уничтожена компьютерным вирусом и т.д. Однако для копирования большого количества объемных файлов требуется очень много дискет. В таком архиве трудно разобраться, его обновление представляет собой весьма трудоемкую операцию. Было бы превосходно иметь файлы, хранящие тот же объем информации, но занимающие на диске меньше места. Оказывается, эта задача имеет решение. Существуют специальные способы упаковать файл, т.е., используя различные алгоритмы сжатия информации, создать копию файла значительно меньшего размера. Более того, упаковка позволяет объединить сжатые копии нескольких файлов в один файл – *архив*. Таким образом, разумно хранить файлы в упакованном, *архивированном* виде. Естественно, должна быть предусмотрена возможность «разархивации», преобразования архивного файла в исходный вид.

Для архивации/разархивации файлов разработаны специальные ***программы-архиваторы*** (упаковщики). В них применяются разные принципы сжатия информации. На самом деле, задача архивации файла может быть сформулирована на математическом языке как задача кодирования/декодирования. Теория кодирования – достаточно сложная и активно развивающаяся отрасль прикладной математики. Ее развитие обусловлено потребностями и колоссальными возможностями современной связи. Не будем углубляться в технику сжатия файлов. Отметим лишь, что в ее основе лежат различные алгоритмы кодирования информации.

Как правило, программы-архиваторы позволяют:

- помещать копии файлов в сжатом виде в архивный файл;
- извлекать файлы из архива;
- просматривать оглавление архива;
- добавлять в архив новые файлы;
- удалять из архива отдельные файлы;
- архивировать целиком каталоги;
- защищать архивы с помощью пароля;
- выводить файлы из архива на экран и на печать;
- создавать многотомные архивы.

Многотомные архивы необходимы в случае, когда архивный файл не помещается на одну дискету. Тогда архив разбивается на несколько дискет.

Наиболее распространенные программы-упаковщики имеют примерно одинаковые возможности, и ни одна из них не превосходит другие по всем параметрам: одни программы архивируют быстрее, другие обеспечивают лучшую степень сжатия файлов. Среди наиболее популярных архиваторов можно упомянуть программы ARJ, PKZIP (PKUNZIP – для разархивации), LHA, PAK, ICE и др. Программа PKZIP/PKUNZIP выделяется непревзойденной скоростью работы. Программа ARJ отличается разнообразным сервисом и умеет создавать архивы на нескольких дискетах.

Последние версии Norton Commander (начиная с 4.0) предоставляют удобные средства работы с архивными файлами, созданными различными архиваторами. Клавиши [Alt] – [F5] (Compress) позволяют создать архивный файл из выделенной группы файлов или одного текущего файла (каталога). По умолчанию используется архиватор ARJ, но можно выбрать и другой упаковщик с помощью кнопки Select method (рис. 16.8).

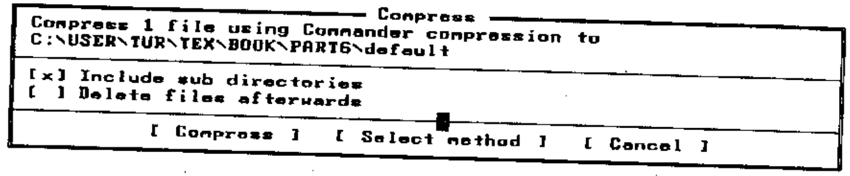


Рис. 16.8

С помощью клавиш [Alt] – [F6] (Decompress) файлы извлекаются из архивов в исходном виде.

16.6. Антивирусные программы

Тем, кто уже работал на компьютере, хорошо знакома такая ситуация: вы приходите к приятелю, чтобы переписать у него нужную вам программу, вставляете дискету в дисковод и в этот момент... приятель хватает вас за руку, не дает вам открыть дисковод на панели Norton Commander, а сам запускает какую-то программу. Вы спрашиваете: «Что случилось?», и он объясняет, что необходимо проверить вашу дискету – а вдруг она **заражена вирусом**? Он боится, что ваша дискета заразит его компьютер и это приведет к ужасным последствиям.

Что же это такое – компьютерный вирус? Неужели и в самом деле существуют микробы компьютерных болезней, которые передаются с диска на диск? И что это за болезни? Конечно же, имеется в виду не вирус в микробиологическом смысле. Под *компьютерным вирусом* понимается не что иное, как специальная небольшая п р о г р а м м а, которая может «приписываться», присоединяться к другим программам, становиться их частью (говорят – «заражать» эти программы). При выполнении зараженной программы управление может быть передано вирусу и тогда будут выполняться те команды, которые в нем запрограммированы, т.е. задуманные программистом вируса.

Первоначально компьютерные вирусы предназначались для защиты программ от несанкционированного использования и копирования, т.е. как средство обеспечения авторского права на программный продукт. Однако со временем среди программистов появилось много любителей (а на самом деле высококвалифицированных профессионалов), которые программировали вирусы для собственного удовольствия. Поскольку вирус – это программа, она может осуществлять любые действия на компьютере, доступные программированию. Вирус может:

- мешать правильной работе программ;
- портить файлы;
- засорять оперативную память ненужной информацией;
- замедлять работу компьютера;
- выдавать на экран посторонние сообщения,

а также производить любые другие вредные действия, запрограммированные «мастером» (от безобидного исполнения гимна США в День независимости до необратимой порчи всей информации на жестком диске).

Разные вирусы ведут себя по-разному. Некоторые могут не проявлять своей вредной сущности до выполнения определенного условия – заданный день недели или обращение к какому-то конкретному файлу. Другие начинают свою разрушительную работу сразу после заражения. Чаще всего пользователь обнаруживает появление вируса не сразу, а когда на его компьютере уже заражены многочисленные файлы. В какой-то момент появляются внешние признаки вируса: компьютер начинает работать медленно, программы выполняются с ошибками, которых прежде не было, изменяется размер файлов (ведь к ним добавляется «кусочек» программы-вируса) и т.д.

В связи с распространением компьютерных вирусов следует предостеречь как от излишней беспечности (использование непроверенных дискет и т.д.), так и от чрезмерной боязни вирусов, «вирусофобии». Ведь вирус является обычной программой и не может совершать никаких сверхъестественных действий. В частности, нужно иметь в виду, что заразиться вирусом могут, как правило, только исполняемые файлы, т.е. файлы с расширениями .COM и .EXE. Правда, недавно появились вирусы, заражающие файлы Microsoft Word с расширением .DOC. Большинство же файлов с данными могут быть только испорчены вирусом. Перезапись таких файлов не может привести к заражению компьютера. Для того чтобы компьютер заразился вирусом, необходимо, чтобы на нем хотя бы один раз была выполнена зараженная программа.

Среди разумных мер борьбы с вирусом можно назвать регулярное копирование информации (заражение винчестера не так трагично, если есть архивные копии нужных файлов на дискетах), а также разграничение доступа к каталогам (каждый пользователь работает только в своем каталоге и не трогает «чужие» файлы).

Однако этих мер недостаточно, и поэтому разработаны специальные ***антивирусные программы.*** Эти программы различаются по цели применения, по степени защиты, по количеству выявляемых вирусов и т.д. По назначению они делятся на следующие группы:

- программы-детекторы (обнаруживают зараженные файлы);
- программы-доктора, фаги («лечат» зараженные программы и диски);
- программы-ревизоры (следят за изменениями в файлах и системных областях дисков);
- программы-фильтры (перехватывают действия вирусов и выдают сообщения о них).

Большинство современных антивирусных программ сочетают в себе различные качества и могут одновременно быть и детектором, и ревизором, и доктором. Следует подчеркнуть, что программы-детекторы обнаруживают только те вирусы, которые им известны. Они могут не заметить вирус, недавно написанный или просто не включенный в список. Таким образом, если программа-детектор не обнаруживает вирусов на машине, это еще не означает, что все в полном порядке.

К наиболее популярным антивирусным программам относятся Aidstest Д.Н. Лозинского, Norton Antivirus, Dr. Web и др. Aidstest сейчас установлен практически на каждом компьютере, эта программа постоянно усовершенствуется, количество выявляемых ею вирусов все время растет. Norton Antivirus и Dr. Web могут обнаруживать вирусы даже в заархивированных файлах.

Чтобы предотвратить заражение компьютера, необходимо регулярно проводить профилактическое антивирусное обследование (это похоже на обязательную флюорографию человека). Не работайте с дискетами, не проверив их антивирусной программой. Если же все-таки заражение произошло, не отчаивайтесь. Хорошая антивирусная программа поможет вам вылечить компьютер. Разумнее же всего в этом случае обратиться к специалисту, который лучше разберется в ситуации и выберет правильную стратегию лечения.

16.7. Программы для связи

Локальные сети. Часто пользователям, которые работают или учатся вместе, необходимы одни и те же программы. Исследователи, занимающиеся одной проблемой, могут обрабатывать на компьютере одни и те же данные. Результаты, полученные одним ученым, могут быть использованы другим как исходные данные. Наконец, может возникнуть необходимость совместного использования периферийных устройств. Например, нескольким пользователям нужен принтер, и выгоднее иметь один хороший принтер на всех, чем плохой на каждого. В этих случаях разумно объединить группу компьютеров в *сеть* (network). Когда компьютеры, объединенные в сеть, расположены в одном здании или близко друг от друга, сеть называют *локальной.*

Простейшая сеть состоит из двух компьютеров, соединенных кабелем через специальные разъемы. Связь между двумя компьютерами может быть осуществлена в системе Norton Commander версии 4.00 и старше с помощью специальной команды Link (в пунктах Left и Right нортоновского меню). Один из компьютеров определяется как ведущий (Master, хозяин), а второй – как подчиненный (Slave, раб). После соединения вы можете выводить на свои панели каталоги дисков на втором компьютере и работать с ними как со своими собственными: копировать, переносить и удалять файлы; запускать программы и т.д.

Следующим шагом стало обеспечение совместной работы уже не двух, а нескольких компьютеров. Физически объединение осуществляется с помощью специальных кабелей (зачастую это дорогостоящие оптико-волоконные кабели). Один из компьютеров сети управляет ее работой. Это, как правило, самый мощный компьютер в сети, обладающий объемным жестким диском, большой оперативной памятью и высокой производительностью. Управляющий компьютер (администратор сети) называется *сервером.* Если ваш компьютер подключен к локальной сети, вам нет необходимости хранить все данные и программы на своем винчестере – вы можете пользоваться тем, что хранится на сервере. Обычно в локальной сети доступны следующие возможности:

- совместное использование компьютерных ресурсов (места на диске, оперативной памяти и т.д.);
- обмен файлами между компьютерами;
- использование общих периферийных устройств;
- управление доступом к ресурсам сети (система защиты информации, паролей, приоритетов и т.д.).

Для работы в сети необходимо специальное программное обеспечение. Самой популярной сетевой программой можно назвать Novell NetWare различных версий, а также Windows for Workgroups, работающая в системе Windows.

Интернет. Мы уже не раз упоминали о глобальной информационной сети Интернет (Internet). Глобальные сети объединяют компьютеры, находящиеся на больших расстояниях друг от друга, возможно, в разных городах и разных странах. Для межкомпьютерной связи используются надземные и подземные кабели, телефонные линии, системы радиорелейной и спутниковой связи. В настоящее время тысячи компьютеров в мире подключены к сети Интернет. Связь российских компьютеров с Интернетом осуществляют такие сети, как RelCom, RosNet, университетская сеть RuNet и др.

Главное назначение Интернета – обеспечить доступ к разнообразной информации, хранящейся на серверах этой сети в разных точках земного шара. Существует несколько основных разновидностей сервиса, которые предоставляются пользователям Интернета.

1. ***Электронная почта (E-mail).*** Самой распространенной операцией при работе в сети можно считать обмен файлами между компьютерами. Значительная доля в файловом обмене принадлежит, конечно, текстовым файлам. Пересылка текстового файла напоминает посылку обычного письма, только происходит во много раз быстрее, практически мгновенно. Поэтому эту разновидность компьютерной связи назвали *электронной почтой* (Electronic Mail, сокращенно E-mail).

Каждый пользователь электронной почты получает свой уникальный а д р е с. Сейчас принято указывать свой электронный адрес наряду с почтовым адресом и номером телефона. Например, адрес автора данного учебника выглядит так:

Vladimir.Turetsky@usu.ru

Значок @ применяется во всех интернетовских адресах. Он разделяет обозначение конкретного пользователя (в данном случае имя и фамилию) и обозначение узла связи (основного сервера, управляющего почтой многих адресатов, в данном случае это usu, т.е. Ural State University, Уральский государственный университет). И то и другое имя может состоять из нескольких уточняющих слов, разделенных точками. Уточнение, как правило, определяется цепочкой серверов, через которые осуществляется выход к данному серверу. В написанном выше адресе имя сервера – короткое, это говорит о том, что университет имеет непосредственный выход в Интернет через спутниковую антенну. Последнее слово в имени сервера – это обычно обозначение страны (ru – от Russia, Россия; по старой памяти в российских адресах до сих пор используется обозначение su – от Soviet Union). В некоторых интернетовских адресах вместо имени страны употребляются обозначения организации (com – коммерческие, gov – правительственные, ac – академические, edu – образовательные и т.д.). Разумно составленный электронный адрес содержит информацию о пользователе и его месте работы.

Для обеспечения электронной почты применяются специальные программы (мэйлеры, mailer). Если компьютеры одного узла объединены в локальную сеть под управлением Novell NetWare, чаще всего для электронной почты применяется программа Pegasus Mail. Имеются и другие программы – DMail, BMail и др. В основном программы-мэйлеры обеспечивают пользователю следующие функции:

- пересылка текстовых файлов по адресу или списку адресов;
- кодирование/декодирование информации;
- пересылка нетекстовых файлов в закодированном виде;
- подтверждение получения письма (Return Receipt);
- редактирование писем с помощью встроенного редактора;
- ведение записной книжки с часто используемыми адресами;
- возможность ответа на письмо (Reply);
- возможность пересылки полученного письма другому адресату (Forward);
- посылка копии письма по указанному адресу;
- преобразование писем в текстовые файлы (Extract)

и другие услуги.

Следует быть осторожным при пересылке русского текста. Это возможно без потери информации лишь при внутрироссийском обмене. При пересылке русского письма за рубеж информация, скорее всего, будет испорчена.

2. ***FTP.*** Этот сервис используется для пересылки файлов. Используя команды ftp (от File Transfer Protocol – протокол пересылки файлов), можно соединиться с удаленным компьютером, просмотреть его каталоги, скопировать с него нужные файлы. Для этого сервер должен быть известен сети как ftp-сервер. Для соединения с ним, разумеется, необходимо знать его адрес. Существуют специальные программы, помогающие осуществлять ftp-обмен. Например, это CuteFTP, работающая в системе Windows.

3. ***WWW.*** Специальные WWW-серверы (от WorldWide Web – всемирная паутина) предоставляют в Интернет текстовую и графическую информацию в виде так называемых «страниц информации» (home page). Если вы связались с WWW-сервером, на экране появится изображение этой страницы. Некоторые слова в тексте будут отмечены другим цветом. Щелкнув на них мышью, вы получите дополнительную информацию. При этом можно попасть на другой сервер, расположенный, возможно, в другой стране.

Все мировые информационные агентства, многие журналы, газеты, научные и образовательные организации имеют свои страницы в Интернете. Попадая на эти страницы, вы можете узнать самые последние новости, ознакомиться со свежими номерами журналов, просмотреть каталоги ведущих мировых библиотек, а затем и получить рефераты или тексты интересующих вас статей. Нужную информацию вы можете сохранить для себя в виде файла соответствующего формата (текстового, графического и т.д.). Некоторые страницы дополнительно оснащены видео- и звуковым сопровождением. Путешествие по Интернету – не только весьма полезное, но и очень увлекательное занятие.

Для работы с WWW-серверами предусмотрено специальное программное обеспечение. Наиболее популярные программы – Netscape Navigator и Microsoft Internet Explorer. Они, как и все Windows-программы, содержат удобное меню, упрощающее поиск нужной информации. На рис. 16.9 и 16.10 – страницы Уральского государственного университета и Американского Математического Общества, вызванные программой Netscape Navigator. Каждая страница содержит меню для дальнейшего получения информации.

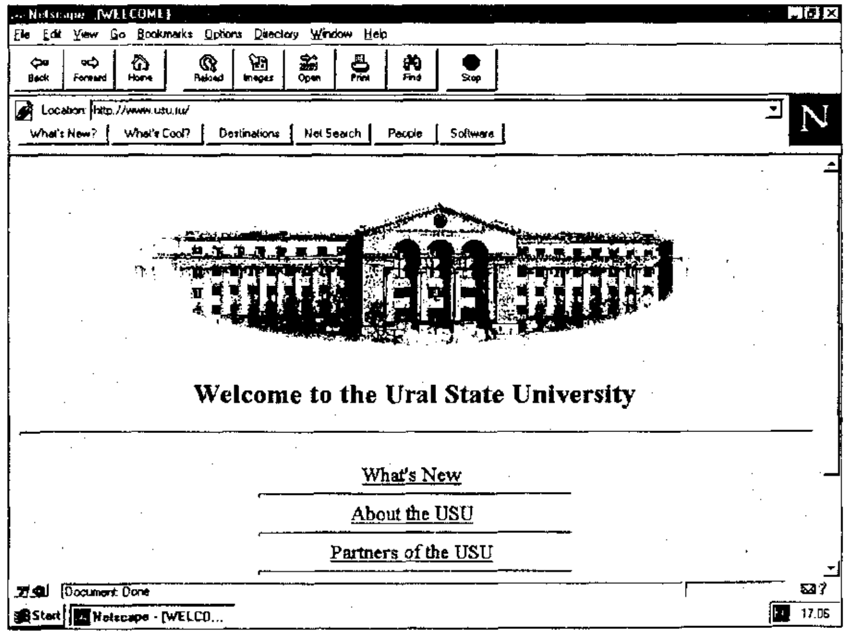


Рис. 16.9

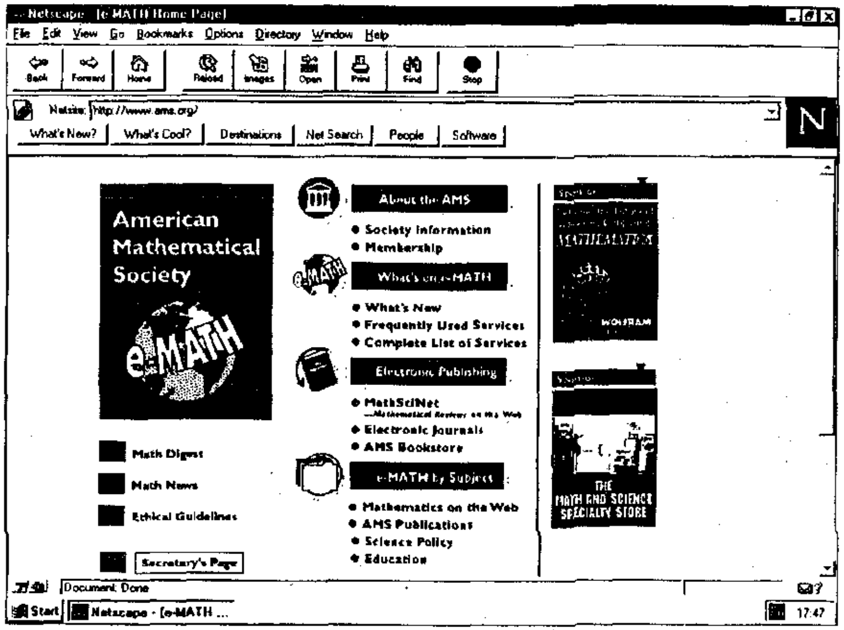


Рис. 16.10

В Интернете можно найти практически все – от кулинарных рецептов до последних постановлений правительства, от хит-парада до новинок моды. Эта информация «спрятана» в миллионах компьютеров по всему свету. Поэтому поиск нужного вам сервера превращается в непростую задачу. Программа Netscape Navigator оборудована мощной справочной системой: нужную информацию можно искать, указав тему запроса, название организации или даже ключевое слово. Чем точнее вы сформулировали запрос, тем быстрее будет найден нужный сервер. Следует иметь в виду, что поиск проводится по всему миру, так что он может занять сравнительно большое время. Поиск можно начать, если выбрать пункт Net Search. Специальный пункт (People) предназначен для поиска в Интернете конкретного человека (многие известные в мире люди имеют свои интернетовские странички).

В последнее время развивается новая концепция Интернета, в соответствии с которой вы можете получать по сети не только нужную информацию, но и специальные программы для обработки этой информации. Таким образом, нет необходимости иметь у себя все программы, вы можете взять нужную программу там, где она разработана, – быть может, в самом отдаленном уголке планеты.

Заключение

Вот и конец учебника. Не знаю, как у вас, а у меня ощущение, что мы переплыли море. Мы плыли вместе, и надеюсь, что пустились в путь не зря. Прошло два года, вы повзрослели, набрались ума-разума. Я не рассчитываю на то, что вы полюбили математику и компьютер больше своих основных предметов. Такая цель и не ставилась. Однако мне хотелось бы, чтобы вы:

- почувствовали красоту и полезность математики;
- познакомились с основными математическими объектами;
- уяснили их главные свойства и связи между ними;
- приобрели навык в решении математических задач;
- запомнили имена великих математиков;
- перестали бояться компьютера.

Следующий шаг теперь – и это самое главное – применение полученных знаний и умений в дальнейшей учебе и, как это ни странно, в жизни. На следующих курсах вы начнете изучать математические методы, используемые в близких вам областях. Вы убедитесь, что абстрактные значки, формулы, теоремы, которыми вам «забивали головы» в течение двух лет, на самом деле нужны и полезны. Тем более это верно в отношении компьютера.

Математические методы, применяемые в гуманитарной сфере, по преимуществу основаны на результатах математической статистики. Этот раздел математики вы изучали в этом курсе. Однако существует множество других статистических методов, не вошедших в данный учебник. Это касается, к примеру, непараметрических критериев, широко применяемых в психологии и социологии. Надеюсь, вы освоите их значительно проще благодаря полученным базовым знаниям. В вашей работе, возможно, окажутся необходимыми и другие разделы прикладной математики – методы оптимизации, дифференциальные уравнения и т.д.

В заключение я повторю те же слова, которые были сказаны в начале книги: «Плывите. Теперь дело за вами...» И спасибо – за внимание и терпение.

Статистические таблицы

Функция распределения $\Phi(x)$ стандартной нормальной случайной величины $N(0, 1)$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7234	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	09957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997

Квантили z_p стандартного нормального распределения

p	0,90	0,95	0,975	0,99	0,995	0,999
z_p	1,282	1,645	1,960	2,326	2,576	3,090

Квантили $\chi^2_p(k)$ распределения «хи-квадрат»

k \ p	0,05	0,1	0,2	0,3	0,7	0,8	0,9	0,95	0,975	0,99	0,995
1	0,00393	0,0158	0,0642	0,148	1,07	1,64	2,71	3,84	5,02	6,63	7,88
2	0,103	0,211	0,446	0,713	2,41	3,22	4,61	5,99	7,38	9,21	10,6
3	0,352	0,584	1,00	1,42	3,67	4,64	6,25	7,81	9,35	11,3	12,8
4	0,711	1,06	1,65	2,19	4,88	5,99	7,78	9,49	11,1	13,3	14,9
5	1,15	1,61	2,34	3,00	6,06	7,29	9,24	11,1	12,8	15,1	16,7
6	1,64	2,20	3,07	3,83	7,23	8,56	10,6	12,6	14,4	16,8	18,5
7	2,17	2,83	3,82	4,67	8,38	9,80	12,0	14,1	16,0	18,5	20,3
8	2,73	3,49	4,59	5,53	9,52	11,0	13,4	15,5	17,5	20,1	22,0
9	3,33	4,17	5,38	6,39	10,7	12,2	14,7	16,9	19,0	21,7	23,6
10	3,94	4,87	6,18	7,27	11,8	13,4	16,0	18,3	20,5	23,2	25,2
11	4,57	5,58	6,99	8,15	12,9	14,6	17,3	19,7	21,9	24,7	26,8
12	5,23	6,30	7,81	9,03	14,0	15,8	18,5	21,0	23,3	26,2	28,3
13	5,89	7,04	8,63	9,93	15,1	17,0	19,8	22,4	24,7	27,7	29,8
14	6,57	7,79	9,47	10,8	16,2	18,2	21,1	23,7	26,1	29,1	31,3
15	7,26	8,55	10,3	11,7	17,3	19,3	22,3	25,0	27,5	30,6	32,8
16	7,96	9,31	11,2	12,6	18,4	20,5	23,5	26,3	28,8	32,0	34,3
17	8,67	10,1	12,0	13,5	19,5	21,6	24,8	27,6	30,2	33,4	35,7
18	9,39	10,9	12,9	14,4	20,6	22,8	26,0	28,9	31,5	34,8	37,2
19	10,1	11,7	13,7	15,4	21,7	23,9	27,2	30,1	32,9	36,2	38,6
20	10,9	12,4	14,6	16,3	22,8	25,0	28,4	31,4	34,2	37,6	40,0
21	11,6	13,2	15,4	17,2	23,9	26,9	29,6	32,7	35,5	38,9	41,4
22	12,3	14,0	16,3	18,1	24,9	27,3	30,8	33,9	36,8	40,3	42,8
23	13,1	14,8	17,2	19,0	26,0	28,4	32,0	35,2	38,1	41,6	44,2
24	13,8	15,7	18,1	19,9	27,1	29,6	33,2	36,4	39,4	43,0	45,6
25	14,6	16,5	18,9	20,9	28,2	30,7	34,4	37,7	40,6	44,3	46,9
26	15,4	17,3	19,8	21,8	29,2	31,8	35,6	38,9	41,9	45,6	48,3
27	16,2	18,1	20,7	22,7	30,3	32,9	36,7	40,1	43,2	47,0	49,6
28	16,9	18,9	21,6	23,6	31,4	34,0	37,9	41,3	44,5	48,3	51,0
29	17,7	19,8	22,5	24,6	32,5	35,1	39,1	42,6	45,7	49,6	52,3
30	18,5	20,6	23,4	25,5	33,5	36,3	40,3	43,8	47,0	50,9	53,7
35	22,5	24,8	27,8	30,2	38,9	41,8	46,1	49,8	53,2	57,3	60,3
40	26,5	29,1	32,3	34,9	44,2	47,3	51,8	55,8	59,3	63,7	66,8
45	30,6	33,4	36,9	39,6	49,5	52,7	57,5	61,7	65,4	70,0	73,2
50	34,8	37,7	41,4	44,3	54,7	58,2	63,2	67,5	71,4	76,2	79,5
75	56,1	59,8	64,5	68,1	80,9	85,1	91,1	96,2	100,8	106,4	110,3
100	77,9	82,4	87,9	92,1	106,9	111,7	118,5	124,3	129,6	135,0	140,2

Квантили $t_p(k)$ распределения Стьюдента

k \ p	0,75	0,9	0,95	0,975	0,99	0,995	0,999
1	1,000	3,078	6,314	12,706	31,821	63,657	318
2	0,816	1,886	2,920	4,303	6,965	9,925	22,3
3	0,765	1,638	2,353	3,182	4,541	5,841	10,2
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307
60	0,679	1,296	1,671	2,000	2,390	2,704	3,307
120	0,677	1,282	1,645	1,960	2,326	2,576	3,090

Квантили $F_p(m, n)$ распределения Фишера, $p = 0,9$

n \ m	1	2	5	10	20	30	40	60	120
1	39,86	49,50	57,24	60,19	61,74	62,26	62,53	62,79	63,06
2	8,53	9,00	9,29	9,39	9,44	9,46	9,47	9,47	9,48
3	5,54	5,46	5,31	5,23	5,18	5,17	5,16	5,15	5,14
4	4,54	4,32	4,05	3,92	3,84	3,82	3,80	3,79	3,78
5	4,06	3,78	3,45	3,30	3,21	3,17	3,16	3,14	3,12
6	3,78	3,46	3,11	2,94	2,84	2,80	2,78	2,76	2,74
7	3,59	3,26	2,88	2,70	2,59	2,56	2,54	2,51	2,49
8	3,46	3,11	2,73	2,54	2,42	2,38	2,36	2,34	2,32
9	3,36	3,01	2,61	2,42	2,30	2,25	2,23	2,21	2,18
10	3,29	2,92	2,52	2,32	2,20	2,16	2,13	2,11	2,08
15	3,07	2,70	2,27	2,06	1,92	1,87	1,85	1,82	1,79
20	2,97	2,59	2,16	1,94	1,79	1,74	1,71	1,68	1,64
25	2,92	2,53	2,09	1,87	1,72	1,66	1,63	1,59	1,56
30	2,88	2,49	2,05	1,82	1,67	1,61	1,57	1,54	1,50
40	2,84	2,44	2,00	1,76	1,61	1,54	1,51	1,47	1,42
60	2,79	2,39	1,95	1,71	1,54	1,48	1,44	1,40	1,35
120	2,75	2,35	1,90	1,65	1,48	1,41	1,37	1,32	1,26

Справки по программному обеспечению

1. Norton Commander

Переход между панелями – [Tab]
Переход в каталог – Сделать текущим и нажать [Enter]
Переход в корневой каталог – [Ctrl] [\]
Переход в родительский каталог – [Ctrl] [PgDn]
Создать каталог (MkDir) – [F7]
Открыть диск на левой панели – [Alt] [F1]
Открыть диск на правой панели – [Alt] [F2]

Выбор группы файлов

Включить/исключить файл – [Ins]
Включить/исключить по шаблону – Серый [+] / [–]
Инверсный выбор – Серая [*]

Работа с файлом или с группой файлов

Копирование (Copy) – [F5]
Перенос/переименование (RenMov) – [F6]
Удаление (Delete) – [F8]
Архивирование (Comp) – [Alt] [F5]
Извлечение из архива (DeComp) – [Alt] [F6]

Работа с текстовым файлом

Просмотр (View) – [F3]
Редактирование (Edit) – [F4]
Создание – [Shift] [F4]

Работа с панелями

Убрать/вернуть обе панели – [Ctrl] [O]
Убрать/вернуть нетекущую панель – [Ctrl] [P]
Убрать/вернуть левую панель – [Ctrl] [F1]
Убрать/вернуть правую панель – [Ctrl] [F2]
Поменять панели местами – [Ctrl] [U]

Сортировка файлов текущей панели

По имени – [Ctrl] [F3]
По расширению – [Ctrl] [F4]
По времени создания – [Ctrl] [F5]
По размеру – [Ctrl] [F6]
В порядке создания – [Ctrl] [F7]

Дополнительные возможности

Быстрый переход к файлу в текущем каталоге – [Alt] буква
Поиск файла на диске – [Alt] [F7]
Быстрый переход по дереву (Tree) – [Alt] [F10]
Просмотр и выполнение предыдущих команд (History) – [Alt] [F8]
Меню Norton Commander (PullDn) – [F9]
Пользовательское меню (Menu) – [F2]
Помощь (Help) – [F1]

2. Windows 95

Системные операции и работа с окнами

Помощь – [F1]
Войти в меню – [Alt] или [F10]
Закрыть меню – [Esc]
Закрыть окно, завершить программу – [Alt] [F4]
Перейти к следующему открытому окну – [Alt] [Tab]

Перейти к следующему окну документа	– Ctrl F6
Открыть Главное меню	– Ctrl Esc
Открыть контекстное меню	– Shift F10
Копировать экран в буфер	– PrtScr
Копировать активное окно в буфер	– Alt PrtScr

Операции с выделенными объектами

Удаление в корзину	– Del
Безвозвратное удаление	– Shift Del
Переименование	– F2
Поиск папки или файла	– F3
Вызов окна свойств	– Alt Enter

Операции с диалоговыми окнами

Движение вперед	– Tab
Движение назад	– Shift Tab
Закрытие диалогового окна	– Esc
Выбор в списке / отказ от выбора	– Space

Операции в прикладных программах

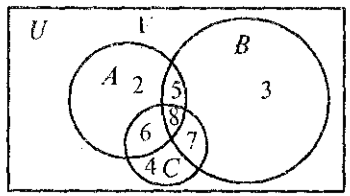
Перейти на конец/начало строки	– End / Home
Перейти на страницу вверх/вниз	– PgUp / PgDn
Перейти в начало/конец документа	– Ctrl End / Home
Выделение фрагмента текста	– Shift стрелка
Отменить последнее преобразование	– Ctrl Z
Перенести в буфер	– Ctrl X
Скопировать в буфер	– Ctrl C
Вставить из буфера	– Ctrl V

Ответы, указания, решения

Глава 1

1. Неверно. **2. а)** Целые числа, кратные 6; **б)** окружность с единичным радиусом и центром в начале координат; **в)** точка (0, 0). **3.** $A = \{1\}$, $B = \{\{1\}, 2\}$, $C = \{\{\{1\}, 2\}, 3\}$. **4. а)** Обозначим левое множество через A, правое – через B. Пусть $x \in A$. Значит, $x = 6y$ для некоторого целого числа y. Обозначим $3y = u$, $2y = v$. Это целые числа, и, очевидно, $x = 2u$, $x = 3v$. Итак, $A \subseteq B$. И наоборот: пусть $x \in B$, т.е. $x = 2u$, $x = 3v$ при некоторых целых u и v. Значит, $2u = 3v$ и число u кратно трем, т.е. делится нацело на 3. Обозначим $u/3 = y$, тогда ясно, что $x = 2u = 2 \cdot (3y) = 6y$. Это означает, что $x \in A$; **б), в)** – аналогично. **5.** $A \cup B$ состоит из четных чисел и чисел, кратных трем; $A \cap B$ – числа, кратные шести; $\overline{A} \cup \overline{B} = \overline{A \cap B}$ – числа, не кратные шести; $\overline{A} \cap \overline{B}$ – нечетные числа, не кратные трем. **6.** $A \cup B$ – прямоугольные и равносторонние треугольники; $A \cap B = \varnothing$, $\overline{A} \cap B = B$ – все равносторонние; $A \cap \overline{B} = A$ – все прямоугольные; $\overline{A} \cup B = \overline{A}$ – все непрямоугольные; $A \cup \overline{B} = \overline{B}$ – все неравносторонние. **7. а)** $\mathcal{P}(A) = \{\varnothing, \{\{1,2\}\}, \{\{3\}\}, \{1\}, \{\{1, 2\}, \{3\}\}, \{\{1, 2\}, 1\}, \{\{3\}, 1\}, A\}$; **б)** $\mathcal{P}(A) = \{\varnothing, \{1\}, \{\{1\}\}, A\}$; **в)** $\mathcal{P}(A) = \{\varnothing\}$; **г)** $\mathcal{P}(A) = \{\varnothing, \{\varnothing\}\}$. **8. а)** $A \cup (\overline{A} \cap B) = (A \cup \overline{A}) \cap (A \cup B) = U \cap (A \cup B) = A \cup B$; **б)** воспользоваться законом де Моргана; **в)** вынести A за скобки; **г)** дважды применить закон де Моргана; **д)** применить закон де Моргана и раскрыть скобки по дистрибутивному закону. **9. а)** $\overline{A} \cup B$; **б)** U; **в)** U; **г)** C.

10.

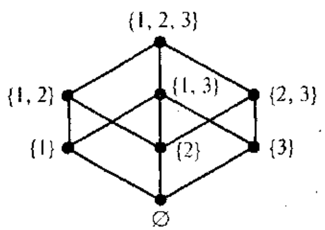


1) $\overline{A \cup B \cup C}$; 2) $A \cap \overline{B \cup C}$; 3) $B \cap \overline{A \cup C}$; 4) $C \cap \overline{A \cup B}$; 5) $A \cap B \cap \overline{C}$; 6) $A \cap C \cap \overline{B}$; 7) $B \cap C \cap \overline{A}$; 8) $A \cap B \cap C$.

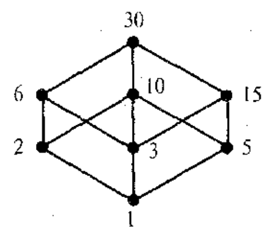
Глава 2

1. а) Рефлексивно: каждое число x делится само на себя. Не симметрично: если x делится на y, не обязательно y делится на x. Транзитивно: пусть x делится на y, y делится на z, т.е. $x = ky$, $y = mz$ при некоторых целых k и m. Значит, $x = k(mz) = (km)z = nz$, где $n = km$ – целое. Значит, действительно, x делится на z. Антисимметрично: если x делится на y и y делится на x, то ясно, что $x = y$;

б) рефлексивно: $x - x = 0$ делится на 4. Симметрично: если $x - y$ делится на 4, то и $y - x = -(x - y)$ тоже делится на 4. Транзитивно. Не антисимметрично; **в)** не рефлексивно: например, при $x = 8$ имеем $x + x = 16 \notin X$. Симметрично: если $x + y \in X$, то и $y + x \in X$. Не транзитивно: возьмем $x = 7$, $y = 2$, $z = 3$: $x + y = 9 \in X$, $y + z = 5 \in X$, но $x + z = 10 \notin X$. Не антисимметрично; **г)** рефлексивно, симметрично, транзитивно, антисимметрично (это отношение равенства на X); **д)** не рефлексивно, симметрично, не транзитивно, не антисимметрично; **е)** рефлексивно: для любого числа из X выполняется $x < 2x$. Не симметрично: например, $x = 1$, $y = 3$. Не транзитивно: при $x = 4$, $y = 3$, $z = 2$ получается $4 < 2\cdot 3$, $3 < 2\cdot 2$, но $4 = 2\cdot 2$. **2. а)** Не рефлексивно: не подходят $x = -1, 0, 1$. Не симметрично: $x = -1$, $y = 2$. Не транзитивно: при $x = 2$, $y = -2$, $z = 1$ выполняется $2 < (-2)^2$, $-2 < 1^2$, но не выполняется $2 < 1^2$; **б)** рефлексивно, симметрично, транзитивно, не антисимметрично (это отношение «быть одного знака»); **в)** рефлексивно, симметрично, транзитивно, не антисимметрично. **4. а)** Не рефлексивно, не симметрично (x может быть братом y, а y – сестрой x), транзитивно, не антисимметрично; **б)** не рефлексивно, не симметрично, не транзитивно, не антисимметрично; **в)** не рефлексивно, не симметрично, не транзитивно, не антисимметрично; **г)** рефлексивно (если каждый является своим собственным подчиненным), не симметрично, транзитивно, антисимметрично; **д)** рефлексивно, симметрично, транзитивно, не антисимметрично. **5.** Все отношения рефлексивные, симметричные, транзитивные, не антисимметричные (отношения эквивалентности). **6. а)** График содержит точки биссектрисы I и III квадрантов; **б)** график симметричен относительно этой биссектрисы. **7. а)** Это отношение «быть одного знака». Множество разбивается на два класса – отрицательные и положительные числа; **б)** это отношение «иметь одинаковую четность». Два класса эквивалентности – четные и нечетные числа. **8. а)** Для любых множеств $X \subseteq X$ – рефлексивность; $X \subseteq Y$, $Y \subseteq Z \Rightarrow X \subseteq Z$ – транзитивность; $X \subseteq Y$, $Y \subseteq X \Rightarrow X = Y$ – антисимметричность. $\mathcal{P}(A) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}$.



8. а)



8. б)

Пустое множество покрывается одноэлементными: к примеру, $\varnothing \subset \{1\}$, и никакое другое множество не является подмножеством множества $\{1\}$. Двухэлементные множества не покрывают пустое, так как, например, $\varnothing \subset \{1\} \subset \{1, 2\}$. Каждое одноэлементное множество покрывается двумя двухэлементными, все двухэлементные – множеством A; **б)** рассуждения аналогичные. Двойка, тройка и пятерка покрывают единицу: например, 2 делится на 1 и нет такого числа, не равного 1 и 2, чтобы 2 на него делилось. Точно так же шестерка не покрывает единицу (6 делится на 2, а 2 делится на 1), зато покрывает двойку и тройку, и т.д. **9.** Везде предполагается, что область значений совпадает со всей числовой осью. **а)** $x \in \boldsymbol{R}$, $y = x^2$ не взаимно однозначная, четная, убывает при $x \leqslant 0$ и возрастает при $x > 0$, ограниченная снизу, не является периодической; **б)** $x \in \boldsymbol{R}$, $y = x^3$ взаимно однозначная, нечетная, возрастающая, не ограниченная, не является периодической; **в)** $x \neq 0$, $y = 1/x$ не взаимно однозначная, нечетная, убывающая при $x < 0$ и $x > 0$, не ограниченная, не является периодической; **г)** $x \in \boldsymbol{R}$, $y = \sin x$ не взаимно однозначная, нечетная, убывает при $-\pi/2 + 2\pi k \leqslant x \leqslant \pi/2 + 2\pi k$ и возрастает при $\pi/2 + 2\pi k \leqslant x \leqslant 3\pi/2 + 2\pi k$, ограниченная, периодическая с периодом 2π; **д)** $x \neq \pi/2 + \pi k$, $y = \operatorname{tg} x$ не взаимно однозначная, нечетная, возрастает при $-\pi/2 + \pi k < x < \pi/2 + \pi k$, не ограниченная, периодическая с периодом π; **е)** $x \in \boldsymbol{R}$, $y = x$ взаимно однозначная, нечетная, возрастающая, не ограниченная, не является периодической; **ж)** $x \in \boldsymbol{R}$, $y = |x|$ не взаимно однозначная, четная, убывает при $x \leqslant 0$ и возрастает при $x > 0$, ограниченная снизу, не является периодической. **10. а)** $y = \lfloor x \rfloor$ (целая часть x); **б)** $y = 1/x$; **в)** $f(x) \equiv 1$.

11. а) $g \circ f = \sin \dfrac{1}{x}$, $x \neq 0$, $f \circ g = \dfrac{1}{\sin x}$, $x \neq \pi k$; **б)** $g \circ f = \sqrt{\operatorname{tg} x}$, $\pi k \leqslant x \leqslant \pi/2 + \pi k$, $f \circ g = \operatorname{tg} \sqrt{x}$, $x \geqslant 0$; **в)** $g \circ f \equiv 9$, $f \circ g \equiv 3$, $x \in \boldsymbol{R}$.

Глава 3

3. а)

P	Q	$P \Rightarrow (P \Rightarrow Q)$
t	t	t
t	f	f
f	t	t
f	f	t

б)

P	Q	$(P \Rightarrow Q) \Leftrightarrow \neg P \vee Q$
t	t	t
t	f	t
f	t	t
f	f	t

в)

P	Q	R	$P \Rightarrow \neg(Q \vee R)$
t	t	t	f
t	t	f	f
t	f	t	f
t	f	f	t
f	t	t	t
f	t	f	t
f	f	t	t
f	f	f	t

г)

P	Q	R	S
t	t	t	t
t	t	f	t
t	f	t	t
t	f	f	f
f	t	t	t
f	t	f	t
f	f	t	t
f	f	f	t

$$S = (P \Rightarrow Q \wedge R) \vee (\neg P \wedge Q)$$

4. Доказательство дистрибутивного закона 3:

A	B	C	$B \wedge C$	$A \vee (B \wedge C)$	$A \vee B$	$A \vee C$	$(A \vee B) \wedge (A \vee C)$
t	t	t	t	t	t	t	t
t	t	f	f	t	t	t	t
t	f	t	f	t	t	t	t
t	f	f	f	t	t	t	t
f	t	t	t	t	t	t	t
f	t	f	f	f	t	f	f
f	f	t	f	f	f	t	f
f	f	f	f	f	f	f	f

5. Доказательство закона контрапозиции:

A	B	$A \Rightarrow B$	$\neg B \Rightarrow \neg A$	$(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$
t	t	t	t	t
t	f	f	f	t
f	t	t	t	t
f	f	t	t	t

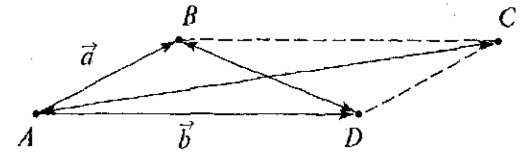
6. Составить таблицу истинности. **7.** Импликация $P \Rightarrow Q$ истинна, значит, не может быть ситуации, когда P истинно, Q ложно. Таким образом, таблица истинности должна содержать только 3 строки. **8.** Ложна. **9.** 1) $\neg B \wedge \neg C$ (по закону де Моргана получили высказывание, логически эквивалентное второй посылке). 2) $\neg B$ (из конъюнкции логически следует одно из ее высказываний). 3) $A \Rightarrow B$ (1-я посылка). 4) $\neg B \wedge (A \Rightarrow B)$ (конъюнкция 2-го и 3-го шагов). 5) $\neg A$ (из «не B» и $A \Rightarrow B$ выводится «не A»).

Глава 4

1. 7. **2.** ±3. **3.** $\overrightarrow{AB} = (-4, 3, -1)$, $\overrightarrow{BA} = (4, -3, 1)$.

4. Пусть $\vec{a} = \overrightarrow{AB}$, $\vec{b} = \overrightarrow{AD}$. Тогда:

$\vec{a} + \vec{b} = \overrightarrow{AC}$;
$\vec{a} - \vec{b} = \overrightarrow{BD}$;
$\vec{b} - \vec{a} = \overrightarrow{DB}$;
$-\vec{a} - \vec{b} = \overrightarrow{CA}$.



6. Составить векторы $\overrightarrow{AB}$, $\overrightarrow{BC}$, $\overrightarrow{CD}$, $\overrightarrow{DA}$; по признаку найти пару коллинеарных и пару неколлинеарных векторов. **7.** $|\vec{a} + \vec{b}| = 6$, $|\vec{a} - \vec{b}| = 14$. **8.** Раскрыть скобки и привести подобные члены. Ответ: -62. **9.** Найти координаты перемножаемых векторов. **а)** 22; **б)** -200; **в)** 129. **10.** 16. **11.** Раскрыть скобки. Ответ: 24. **12.** Воспользоваться формулой для координат векторного произведения через определители. **а)** $(-3, 5, 7)$; **б)** $(-6, 13, 14)$; **в)** $(-12, 20, 28)$. **13.** 24. **14.** Посчитать определитель 3-го порядка. Ответ: -7. **15.** Проверить с помощью определителей. **а)** Компланарны; **б)** не компланарны. **16.** Подставить координаты точек в уравнение прямой. Ответ: M_1, M_3, M_4 лежат на прямой, M_2, M_5, M_6 не лежат. **17.** $a = -C/A = 6$, $b = -C/B = -4$. **18.** Найдем координаты вершины B, т.е. точку пересечения прямых AB и BC. Для этого решим систему

$$\begin{cases} 4x + 3y - 5 = 0, \\ x - 3y + 10 = 0. \end{cases}$$

Складывая уравнения, получаем $5x = -5$, откуда $x = -1$. Подставим во второе уравнение: $y = 3$. Итак, $B(-1, 3)$. Аналогично $A(2, -1)$, $C(2, 4)$. **19.** Угловой коэффициент данной прямой равен $k = -A/B = -{}^5/_3$. **а)** $k = -{}^5/_3$; **б)** $k = {}^3/_5$. **20.** $k = \dfrac{y_1 - y_2}{x_1 - x_2} = 7$. **21.** Составить уравнения диагоналей AC и BD и найти точку их пересечения. Ответ: $(1, 3)$. **22.** Условие перпендикулярности: $A_1A_2 + B_1B_2 = 0$. Ответ: **а)** и **в)** перпендикулярны; **б)** не перпендикулярны. **23.** $A = 1$, $B = -2$, $C = 3$, $D = -(Ax_0 + By_0 + Cz_0) = 3$. Ответ: $x - 2y + 3z + 3 = 0$. **24.** Вектор нормали $\vec{n} = \overline{M_1M_2} = (1, -1, -3)$. Далее решаем по схеме задачи 23.

Ответ: $x - y - 3z + 2 = 0$. **25.** У параллельных плоскостей коллинеарные нормали, т.е. векторы (A, B, C). **а)** Параллельны; **б)** не параллельны. **26.** Условие перпендикулярности плоскостей: $A_1A_2 + B_1B_2 + C_1C_2 = 0$. **а)** Перпендикулярны; **б)** не перпендикулярны. **27.** $a = -D/A = 4$, $b = -D/B = -3$, $c = -D/C = -0{,}5$. **28.** Решить три системы уравнений, дополнив уравнение прямой уравнениями координатных плоскостей ($z = 0$, $y = 0$ и $x = 0$). Ответ: $(2, -1, 0)$; $({}^4/_3, 0, -{}^1/_3)$; $(0, 2, -1)$.

Глава 5

1. $AB = \begin{pmatrix} 3 & 4 \\ 5 & 10 \end{pmatrix}$, $BA = \begin{pmatrix} 5 & 6 \\ 5 & 8 \end{pmatrix}$. **2.** $AB = \begin{pmatrix} -2 & 2 & 3 \\ 3 & -3 & -1 \\ -1 & 1 & 1 \end{pmatrix}$.

3. $ABC = \begin{pmatrix} 7 & 6 \\ -3 & -6 \end{pmatrix}$, $(AB - BC)' = \begin{pmatrix} 6 & -1 \\ 4 & -3 \end{pmatrix}$.

4. а) $\begin{vmatrix} 1 & 0 & 2 \\ 3 & 2 & 6 \\ 1 & 2 & 1 \end{vmatrix} = 1\cdot 2\cdot 1 + 3\cdot 2\cdot 2 + 0\cdot 6\cdot 1 - 2\cdot 2\cdot 1 - 2\cdot 6\cdot 1 - 0\cdot 3\cdot 1 =$
$= -2$;

б) $\begin{vmatrix} 1 & 0 & 2 \\ 3 & 2 & 6 \\ 1 & 2 & 1 \end{vmatrix} = 1\cdot\begin{vmatrix} 2 & 6 \\ 2 & 1 \end{vmatrix} - 3\cdot\begin{vmatrix} 0 & 2 \\ 2 & 1 \end{vmatrix} + 1\cdot\begin{vmatrix} 0 & 2 \\ 2 & 6 \end{vmatrix} = -10 + 12 - 4 = -2$;

в) $\begin{vmatrix} 1 & 0 & 2 \\ 3 & 2 & 6 \\ 1 & 2 & 1 \end{vmatrix} =$ (вычтем из 3-го столбца удвоенный 1-й) $=$

$= \begin{vmatrix} 1 & 0 & 0 \\ 3 & 2 & 0 \\ 1 & 2 & -1 \end{vmatrix} =$ (разложим по 3-му столбцу) $= (-1)\cdot\begin{vmatrix} 1 & 0 \\ 3 & 2 \end{vmatrix} = -2$.

5. 31. **6. а)** $\Delta = \begin{vmatrix} 2 & 3 \\ 1 & -1 \end{vmatrix} = -5$, $\Delta_x = \begin{vmatrix} 1 & 3 \\ 0 & -1 \end{vmatrix} = -1$, $\Delta_y = \begin{vmatrix} 2 & 1 \\ 1 & 0 \end{vmatrix} = -1$; $x = \Delta_x/\Delta = {}^1/_5$, $y = \Delta_y/\Delta = {}^1/_5$; **б)** $x = 1$, $y = -3$, $z = 3$. **7.** Однородная система имеет ненулевое решение тогда и только тогда, когда ее главный определитель равен нулю. Ответ: **а)** не имеет; **б)** и **в)** имеют.

8. Составить определитель из координат векторов:

$\begin{vmatrix} 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 2 \\ 1 & 0 & 1 & 3 \\ 0 & 1 & 0 & 0 \end{vmatrix} =$ (разложим по 4-й строке) $= \begin{vmatrix} 1 & 1 & 0 \\ 2 & 0 & 2 \\ 1 & 1 & 3 \end{vmatrix} =$

$=$ (разложим по 1-й строке) $= \begin{vmatrix} 0 & 2 \\ 1 & 3 \end{vmatrix} - \begin{vmatrix} 2 & 2 \\ 1 & 3 \end{vmatrix} = -2 - 4 = -6 \neq 0$. Значит, векторы линейно независимы. **9.** $\|\mathbf{x}\|_1 = \sqrt{1^2 + (-1)^2 + (-3)^2 + 2^2 + (-7)^2} = 8$; $\|\mathbf{x}\|_2 = |1| + |-1| + |-3| + |2| + |-7| = 14$; $\|\mathbf{x}\|_3 = \max\{1, 1, 3, 2, 7\} = 7$.

Глава 6

1. а), б) Да; **в)** нет (нет противоположных элементов); **г)** нет (нет нейтрального элемента); **д)** да; **е)** нет (нет нейтрального элемента); **ж)** нет (это не группоид, так как сумма простых чисел может не быть простым числом). **2. а)** Нет (для нуля нет обратного элемента); **б)** да; **в)** да; **г), д), е), ж)** нет (нет обратных элементов). **3. а)** Является (проверяется непосредственно); **б)** нет: $a * (b * b) = a * a = c$, $(a * b) * b = a * b = a$; **в)** является. **4. а)** Не является группой, так как нет нейтрального элемента (видно по первой строке таблицы: при умножении a ни в одном случае не получается a). Операция коммутативна; **б)** является группой: нейтральный элемент a, противоположные элементы: $a^{-1} = a$, $b^{-1} = d$, $c^{-1} = c$. Операция коммутативна.

5. а)

*	0	1	2	3	4
0	0	1	2	3	4
1	0	1	2	3	4
2	0	1	2	3	4
3	0	1	2	3	4
4	0	1	2	3	4

б)

*	0	1	2	3	4
0	0	1	2	3	4
1	1	0	1	2	3
2	2	1	0	1	2
3	3	2	1	0	1
4	4	3	2	1	0

6. Нейтральный элемент относительно пересечения — множество A, относительно объединения — пустое множество. **а), б)** В обоих случаях нет противоположных элементов. Значит, это не группы. **7.** Проверить свойства ассоциативности, коммутативности, наличие нейтрального и противоположных элементов. **8. а), б)** Кольцо и поле; **в)** кольцо, но не поле (нет обратных элементов по умножению);

г) не кольцо (нет обратных элементов по сложению); д) кольцо и поле; **е), ж)** кольцо, но не поле (нет обратных элементов по умножению).

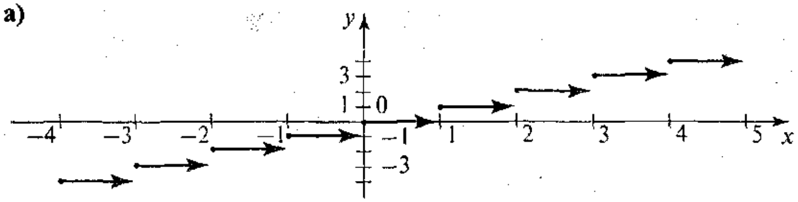
Глава 7

1. а) Степень числителя меньше степени знаменателя. Ответ: 0; **б)** раскрыть скобки в числителе и знаменателе. Ответ: 3; **в), г)** домножить числитель и знаменатель на сопряженное выражение. Ответ: 0; **д)** $\lim\limits_{n\to\infty}\left(\frac{n+3}{n}\right)^{3n-1}=\lim\limits_{n\to\infty}\left[\left(1+\frac{3}{n}\right)^{\frac{n}{3}}\right]^{\frac{3(3n-1)}{n}}=e^9$; **е)** $e^{1/3}$. **2. а)** 9; **б)** 9; **в)** степень числителя меньше степени знаменателя. Ответ: 0; **г)** степени числителя и знаменателя равны. Предел равен отношению старших коэффициентов. Ответ: $^1/_3$; **д)** $\lim\limits_{x\to4}\frac{x^2-10x+24}{x^2-9x+20}=\lim\limits_{x\to4}\frac{(x-4)(x-6)}{(x-4)(x-5)}=$ $=\lim\limits_{x\to4}\frac{x-6}{x-5}=2$; **е)** аналогично. Ответ: $^4/_3$; **ж)** $e^0=1$; **з)** $e^{-1}=1/e$; **и)** воспользуемся замечательным пределом: $\lim\limits_{x\to0}\frac{\operatorname{tg}x}{x}=\lim\limits_{x\to0}\frac{\sin x}{x\cos x}=$ $=\lim\limits_{x\to0}\frac{\sin x}{x}\lim\limits_{x\to0}\frac{1}{\cos x}=1\cdot1=1$; **к)** $^1/_5$.

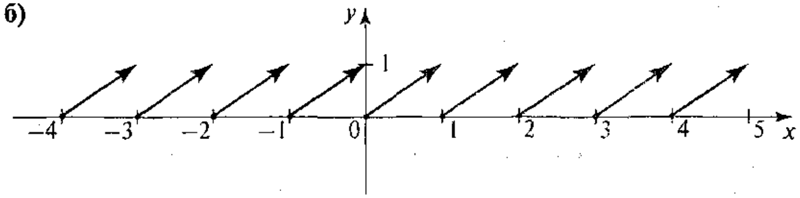
Глава 8

1. Достаточно найти пределы слева и справа в точках $x=0$, $x=1$, $x=3$. Для этого посчитаем значения функции по левой и правой формулам: $y(0-0)=0$, $y(0+0)=0$; $y(1-0)=1$, $y(1+0)=$ $=-1^2+4\cdot1-2=1$; $y(3-0)=-3^2+4\cdot3-2=1$, $y(3+0)=4-3=1$. Функция непрерывна. **2.** Функция будет непрерывной при $f(1-0)=$ $=f(1+0)$, т.е. при $3-a=2$, откуда $a=1$. **3.** Найдем A и B из условий $f(-\pi/2-0)=f(-\pi/2+0)$, $f(\pi/2-0)=f(\pi/2+0)$. Получим систему уравнений $\begin{cases}-A+B=2\\ A+B=0\end{cases}$, откуда $A=-1$, $B=1$. **4.** Найдем предел $\lim\limits_{x\to1}\frac{x^2-1}{x^3-1}=$ $=\lim\limits_{x\to1}\frac{x+1}{x^2+x+1}=$ $^2/_3$. Итак, $f(1)=$ $^2/_3$. **5.** Функция $\frac{\sin x}{x}$ имеет устранимый разрыв I рода (замечательный предел); функция $\frac{\cos x}{x}$ — бесконечный разрыв II рода. **6.** Найти значения функции на границах промежутка и убедиться, что они разного знака. **7.** Применить теорему Коши на промежутке $[-1, 1]$. **8.** Все эти функции имеют неустранимые разрывы I рода в целых точках ($x=0, 1, 2, \ldots$ и $x=-1, -2, -3, \ldots$).

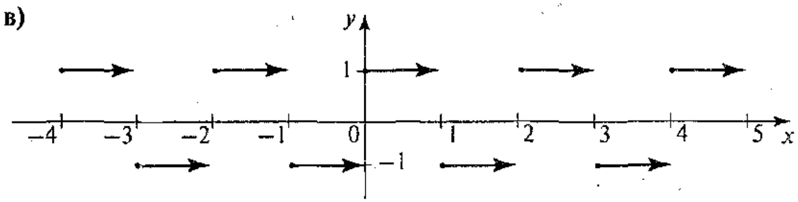
а)



б)



в)



Глава 9

1. а) Применяем формулу $(x^n)'=nx^{n-1}$ и линейные свойства производной: $y'=6x-5$; **б)** $y'=4x^3+5x-0{,}3$; **в)** $y'=\frac{1}{3\sqrt[3]{x^2}}$; **г)** $y'=\frac{1}{\sqrt{x}}+$ $+\frac{1}{x^2}$; **д)** производная произведения: $y'=(x^2-3x+3)'(x^2+2x-1)+$ $+(x^2-3x+3)(x^2+2x-1)'=(2x-3)(x^2+2x-1)+(x^2-3x+3)\times$ $\times(2x+2)=4x^3-3x^2-8x+9$; **е)** $y'=-\frac{1}{2\sqrt{x}}-\frac{1}{2x\sqrt{x}}$; **ж)** производная частного: $y'=\frac{(x+1)'(x-1)-(x+1)(x-1)'}{(x-1)^2}=-\frac{2}{(x-1)^2}$; **з)** $y'=-\frac{6x^2}{(x^3-1)^2}$;

и) $y' = \cos x - \sin x$; **к)** производная сложной функции: сначала дифференцируем внешнюю функцию (в данном случае – квадрат), затем – аргумент (косинус). Ответ: $y' = 2\cos x(-\sin x) = -\sin 2x$; **л)** $y' = 3\cos 3x$; **м)** $y' = 9\cos(3x+5)$; **н)** $y' = -\sin x - \cos^2 x(-\sin x) = -\sin x(1-\cos^2 x) = -\sin^3 x$; **о)** $y' = \dfrac{1-\cos x + x\sin x}{(1-\cos x)^2}$; **п)** $y' = \dfrac{1}{x^2}\cdot 2x = \dfrac{2}{x}$;

р) $y' = \dfrac{1}{\cos^2 e^x}e^x = \dfrac{e^x}{\cos^2 e^x}$; **с)** $y' = e^{\cos x}(\cos x - \sin^2 x)$;

т) $y' = \dfrac{e^{-x^2}(1-2x^2\ln x)}{x}$. **2.** Угловой коэффициент касательной равен значению производной в данной точке, $y' = 2x$. **а)** $k = 0$; **б)** $k = 2\cdot 3 = 6$; **в)** $k = 2\cdot(-2) = -4$. **3.** В точках $(1, 1)$ и $(-1, -1)$. **4.** Точки лежат на параболе, поэтому $y_1 = 1$, $y_2 = 4$ и угловой коэффициент секущей $k = \dfrac{y_2 - y_1}{x_2 - x_1} = 3$. Приравняем производную к значению k: $2x = 3$, откуда $x = 1{,}5$. Ответ: в точке $(1{,}5,\ 2{,}25)$. **5.** Формула конечных приращений: $f(x+\Delta) \approx f(x) + f'(x)\Delta$. **а)** $f(x) = x^3$, $f'(x) = 3x^2$, $x = 1$, $\Delta = 0{,}02$: $1{,}02^3 \approx 1 + 3\cdot 0{,}02 = 1{,}06$ (точное значение 1,061208); **б)** 0,985; **в)** 0,96. **6.** **а)** $y'' = 2$; **б)** $y'' = -2\cos 2x$. **7.** $y' = 12x^3 - 12x^2 + 24x$, $y'(0) = 0$; $y'' = 36x^2 - 24x + 24$, $y''(0) = 24 > 0$. Значит, это точка минимума. **8.** Найти производную и убедиться, что $y' < 0$ на данном промежутке. **9.** Возрастает при $x < -1$, $x > 3$, убывает при $-1 < x < 3$. **10.** $x = 0$ – максимум, $x = 1$ – минимум.

Глава 10

1. а) Линейная замена $3x + 1 = t$, $3dx = dt$, $\int (3x+1)^2\,dx = \dfrac{1}{3}\int t^2\,dt = \dfrac{t^3}{9} = \dfrac{(3x+1)^3}{9} + C$. Проверка: $\left(\dfrac{(3x+1)^3}{9}\right)' = \dfrac{3(3x+1)^2\cdot 3}{9} = (3x+1)^2$ – подынтегральная функция; **б)** замена $2x - 3 = t$. Ответ: $-\dfrac{1}{4(2x-3)^2} + C$;

в) замена $1 + x^2 = t$, $2xdx = dt$, $\int \dfrac{x\,dx}{\sqrt{1+x^2}} = \dfrac{1}{2}\int\dfrac{dt}{\sqrt{t}} = \dfrac{1}{2}\int t^{-1/2}dt = \dfrac{1}{2}\dfrac{t^{1/2}}{½} = \sqrt{t} = \sqrt{1+x^2} + C$; **г)** $\frac{1}{3}\ln\left|x^3+1\right| + C$ (замена $x^3 + 1 = t$); **д)** $-\dfrac{1}{2}\cos x^2 + C$ (замена $x^2 = t$); **е)** $e^{\sin x} + C$ (замена $\sin x = t$). **2. а)** Обозначим: $u = x$, $v' = \sin 2x$. Тогда $u' = 1$, $v = -\dfrac{\cos 2x}{2}$. Применим формулу интегрирования по частям: $\int x\sin 2x\,dx = -\dfrac{x\cos 2x}{2} + \dfrac{1}{2}\int\cos 2x\,dx = -\dfrac{x\cos 2x}{2} + \dfrac{\sin 2x}{4} + C$; **б)** $-e^{-x}(1 + x) + C$ ($u = x$, $v' = e^{-x}$, $u' = 1$, $v = -e^{-x}$);

в) $x\arccos x - \sqrt{1-x^2} + C$ ($u = \arccos x$, $v' = 1$); **г)** $\dfrac{x^3\ln x}{3} - \dfrac{x^3}{9} + C$;

д) $\int x\cos^2 x\,dx = \int x\dfrac{\cos 2x + 1}{2}dx = \dfrac{1}{2}\int x\,dx + \dfrac{1}{2}\int x\cos 2x\,dx$. Первое слагаемое равно $\dfrac{x^2}{4}$. Во втором интеграле применим формулу интегрирования по частям при $u = x$, $v' = \cos 2x$. Ответ: $\dfrac{x^2}{4} + \dfrac{x\sin 2x}{4} + \dfrac{\cos 2x}{8} + C$.

3. а) $\int\limits_0^{\pi}\cos x\,dx = \sin x\Big|_0^{\pi} = \sin\pi - \sin 0 = 0 - 0 = 0$; **б)** $3\dfrac{3}{4}$; **в)** $e - \dfrac{1}{e}$;

г) замена переменной: $1 + x = t$, $dx = dt$, $\int\limits_0^1\sqrt{1+x}\,dx = \int\limits_1^2\sqrt{t}\,dt = \dfrac{2}{3}t^{3/2}\Big|_1^2 = \dfrac{2}{3}(2\sqrt{2} - 1)$; **д)** $\dfrac{7}{72}$; **е)** интеграл от суммы равен сумме интегралов.

$\int\limits_0^1(e^x - x + 2)dx = \int\limits_0^1 e^x\,dx - \int\limits_0^1 x\,dx + \int\limits_0^1 2\,dx = e^x\Big|_0^1 - \dfrac{x^2}{2}\Big|_0^1 + 2x\Big|_0^1 = e + \dfrac{1}{2}$; **ж)** замена $x^2 + 1 = t$, $2xdx = dt$. Тогда $\int\limits_0^1\dfrac{x\,dx}{(x^2+1)^2} = \dfrac{1}{2}\int\limits_1^2\dfrac{dt}{t^2} = \dfrac{1}{2}\left(-\dfrac{1}{t}\right)\Big|_1^2 = \dfrac{1}{4}$;

з) замена $\sin x = t$. Ответ: -1; **и)** замена $\dfrac{1}{x} = t$, $-\dfrac{dx}{x^2} = dt$. $\int\limits_1^2\dfrac{e^{1/x}}{x^2}dx = -\int\limits_1^{1/2}e^t dt = \int\limits_{1/2}^1 e^t dt = e - \sqrt{e}$; **к)** интегрирование по частям: $u = x$, $v' = e^{-x}$. Тогда $\int\limits_0^1 xe^{-x}dx = (-xe^{-x})\Big|_0^1 + \int\limits_0^1 e^{-x}dx = -\dfrac{1}{e} + (-e^{-x})\Big|_0^1 = 1 - \dfrac{2}{e}$; **л)** $\pi/2 - 1$. **м)** 1. **4. а)** $^1/_3$; **б)** $^1/_2$; **в)** -1.

Глава 11

1. Да, эти события образуют полный набор. Если исход ω_i состоит в том, что выпало число i, то $A=\{\omega_1, \omega_2\}$, $B=\{\omega_3\}$, $C=\{\omega_4, \omega_5, \omega_6\}$. Видно, что $A+B+C=\Omega$. Эти события не являются равновозможными. **2.** Пространство элементарных событий состоит из пар целых чисел (i, j), $i, j=1\div 6$, где i и j – числа, выпавшие на первом и втором кубике. $A=\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$; $B=\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5)\}$; $C=\{(4, 4), (4, 5), (4, 6), (5, 4), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\}$. **3. а)** A_1A_2; **б)** $\overline{A_1}$; **в)** $A_1+A_2+A_3$; **г)** $A_1A_2A_3$; **д)** $\overline{A_1}\,\overline{A_2}\,\overline{A_3}$. **4.** Всего исходов 36, благоприятных 4, $P(A)=4/36={}^1/_9$. **5.** Всего исходов $N=C_6^2=15$, благоприятных $m=3\cdot 3=9$ (каждый из трех белых с каждым из трех черных). $P(A)=9/15={}^3/_5$. **6.** Всего исходов $N=C_5^2=10$, благоприятных $m=C_4^2=6$ (2 шара из четырех синих и красных). $P(A)=0,6$. **7.** Всего исходов $N=A_5^3=3!C_5^3=60$. Чисел с двойкой на конце можно составить $A_4^2=12$ штук (столько двузначных чисел можно составить из оставшихся цифр 1, 3, 4, 5). Столько же чисел с четверкой на конце. Итак, благоприятных исходов 24. $P(A)=24/60={}^2/_5$. **8.** 0,1. **9.** ${}^2/_3$. **10. а)** Вероятность произведения независимых событий равна произведению вероятностей: $P(A)=P(A_1A_2)=P(A_1)P(A_2)=0,8\cdot 0,7=0,56$, где A_1, A_2 – попадание первого и второго стрелка; **б)** событие «хотя бы одно попадание» противоположно событию «ни одного попадания»: $P(A)=1-P(\overline{A_1}\,\overline{A_2})=1-P(\overline{A_1})P(\overline{A_2})=1-(1-0,8)(1-0,7)=0,94$.

11. По теореме умножения $\frac{6}{10}\cdot\frac{4}{9}=\frac{4}{15}$. **12.** $P(A)=P(Б_1Б_2+Ч_1Ч_2)=$ (сумма несовместных событий) $=P(Б_1Б_2)+P(Ч_1Ч_2)=$ (произведения независимых событий) $=P(Б_1)P(Б_2)+P(Ч_1)P(Ч_2)=\frac{5}{15}\cdot\frac{6}{15}+\frac{10}{15}\cdot\frac{9}{15}=\frac{120}{225}=\frac{8}{15}$. **13.** ${}^3/_4$. **14.** Две гипотезы: B_1 – из первой урны переложили белый шар, B_2 – черный. По формуле полной вероятности $P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)=0,3\cdot 0,3+0,7\cdot 0,4=0,37$. **15.** B_1 – переложили два белых, B_2 – два черных, B_3 – черный и белый. Всего исходов $N=C_6^2=15$. Два белых можно переложить одним способом: $P(B_1)={}^1/_{15}$; два черных – $C_4^2=6$ способами: $P(B_2)={}^6/_{15}$; $P(B_3)={}^8/_{15}$. Если переложили 2 белых, то во второй урне стало 3 белых и 3 черных: $P(A|B_1)={}^3/_6$. Аналогично $P(A|B_2)={}^1/_6$, $P(A|B_3)={}^2/_6$.

а) $P(A)=\frac{1}{15}\cdot\frac{3}{6}+\frac{6}{15}\cdot\frac{1}{6}+\frac{8}{15}\cdot\frac{2}{6}=\frac{25}{90}=\frac{5}{18}$; **б)** по формуле Байеса

$$P(B_3|C)=\frac{\frac{8}{15}\cdot\frac{4}{6}}{\frac{1}{15}\cdot\frac{3}{6}+\frac{6}{15}\cdot\frac{5}{6}+\frac{8}{15}\cdot\frac{4}{6}}=\frac{32}{65}.$$

16. ${}^2/_3$. **17.** Благоприятное событие (произведение равно шести) происходит в случаях (1, 6), (2, 3), (3, 2), (6, 1), т.е. вероятность успеха равна $4/36={}^1/_9$. **а)** $P_{10}(4)=C_{10}^4\left(\frac{1}{9}\right)^4\left(\frac{8}{9}\right)^6$; **б)** «хотя бы один раз» – значит, не ноль раз. Ответ: $1-\left(\frac{8}{9}\right)^{10}$. **18.** Вероятность успеха $p={}^{24}/_{45}$. Применить формулу Бернулли. **19.** Три из четырех.

Глава 12

1. Это биномиальная случайная величина:

$$\xi=B_{4,\,1/2}=\begin{pmatrix}0 & 1 & 2 & 3 & 4\\ \frac{1}{16} & \frac{4}{16} & \frac{6}{16} & \frac{4}{16} & \frac{1}{16}\end{pmatrix},\quad M\xi=np=4\cdot\frac{1}{2}=2.$$

2. Случайная величина ξ может принимать значения от 0 до 4. Всего исходов $N=C_{10}^4=210$. Ни одного красного ($\xi=0$): благоприятных исходов $C_6^4=15$ (4 шара выбираются из 6 синих, зеленых и желтых). Один красный: благоприятных исходов $4C_6^3=80$ (каждый красный с любыми тремя другими). Два красных – соответственно $C_4^2\cdot C_6^2=90$ (любые два красных с любыми двумя другими). Три красных: $C_4^3\cdot 6=24$. Четыре красных можно вынуть в одном случае. Итак,

$$\xi=\begin{pmatrix}0 & 1 & 2 & 3 & 4\\ \frac{15}{210} & \frac{80}{210} & \frac{90}{210} & \frac{24}{210} & \frac{1}{210}\end{pmatrix},\quad M\xi=1,6.$$

4. $$\xi=\begin{pmatrix}2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12\\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36}\end{pmatrix},\quad M\xi=7.$$

5. Вычислим предварительно $D\xi$ и $D\eta$. Поскольку $\xi=\begin{pmatrix}0 & 1 & 3\\ 0,2 & 0,6 & 0,2\end{pmatrix}$, $\xi^2=\begin{pmatrix}0 & 1 & 9\\ 0,2 & 0,6 & 0,2\end{pmatrix}$, получаем $M\xi=1,2$, $M\xi^2=2,4$, $D\xi=2,4-1,2^2=$

$= 0{,}96$. Аналогично $D\eta = 0{,}24$. Сумма распределена по следующему закону: $\xi + \eta = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ 0{,}08 & 0{,}36 & 0{,}36 & 0{,}08 & 0{,}12 \end{pmatrix}$, и $M(\xi + \eta) = 3{,}8$, $M(\xi + \eta)^2 = 15{,}64$, $D(\xi + \eta) = 15{,}64 - 3{,}8^2 = 1{,}2 = D\xi + D\eta$.

6. $\xi^2 = \xi = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$, $M\xi^2 = 1/2$, $D\xi^2 = 1/4$. **7.** $M\xi = 0{,}4$, $M\eta = -0{,}2$, $\xi\eta = \begin{pmatrix} -1 & 0 & 1 \\ 0{,}28 & 0{,}52 & 0{,}2 \end{pmatrix}$ (произведение -1 получается в двух случаях: $\xi = -1$, $\eta = 1$ и $\xi = 1$, $\eta = -1$, вероятность равна $0{,}1 \cdot 0{,}3 + 0{,}5 \cdot 0{,}5 = 0{,}28$; остальные вероятности вычисляются аналогично). $M(\xi\eta) = -0{,}08$, $\mathrm{cov}(\xi, \eta) = M(\xi\eta) - M\xi\, M\eta = -0{,}08 - 0{,}4 \cdot (-0{,}2) = 0$.

8. а) $c = \dfrac{2}{b-a}$; **б)** $M\xi = \dfrac{a+b}{2}$, $D\xi = \dfrac{(b-a)^2}{24}$; **в)** 2,6838. **9.** $M\xi = \dfrac{1}{\lambda}$, $D\xi = \dfrac{1}{\lambda^2}$ (вычислить соответствующие несобственные интегралы).

10. $M\xi = 5$, $D\xi = 3$. **11.** $P(\xi \leqslant 0{,}3) = F_\xi(0{,}3) = \Phi\left(\dfrac{0{,}3-1}{0{,}5}\right) = \Phi(-1{,}4) = 1 - \Phi(1{,}4) = 1 - 0{,}9192 = 0{,}0808$, $P(\xi > 0{,}6) = 1 - P(\xi \leqslant 0{,}6) = 1 - F_\xi(0{,}6) = 1 - \Phi\left(\dfrac{0{,}6-1}{0{,}5}\right) = 1 - \Phi(-0{,}8) = \Phi(0{,}8) = 0{,}7881$. Второе событие вероятнее: это ясно и без вычислений: и 0,3, и 0,6 расположены левее среднего. Значит, площадь под кривой плотности распределения правее 0,6 больше $1/2$, а площадь левее 0,3 – меньше.

12. $P(\xi < 0) = F_\xi(0) = \Phi\left(\dfrac{0-m}{1}\right) = \Phi(-m) = 1 - \Phi(m) = 0{,}023$. Отсюда $\Phi(m) = 0{,}977$ и по таблице находим, что $m = 2$. Тогда $P(0{,}5 \leqslant \xi \leqslant 3{,}5) = P(|\xi - 2| \leqslant 1{,}5) =$ (для симметричного интервала) $= 2\Phi\left(\dfrac{1{,}5}{1}\right) - 1 = 2\Phi(1{,}5) - 1 = 0{,}8664$. **13.** $P(\xi \in [0, 1]) = \Phi\left(\dfrac{1-0}{\sigma}\right) - \Phi\left(\dfrac{0-0}{\sigma}\right) = \Phi(1/\sigma) - \Phi(0) = \Phi(1/\sigma) - 0{,}5 = 0{,}383$, откуда $\Phi(1/\sigma) = 0{,}883$. По таблице находим, что $1/\sigma = 1{,}19$, $\sigma \approx 0{,}84$. Значит, $P(\xi > 1{,}5) = 1 - \Phi\left(\dfrac{1{,}5-1}{0{,}84}\right) \approx 1 - \Phi(0{,}6) = 0{,}2743$. **14.** По условию $P(\xi < 1{,}141) = 0{,}9$,

$P(\xi < 1{,}48) = 0{,}975$. Значит, $\Phi\left(\dfrac{1{,}141-m}{\sigma}\right) = 0{,}9$, $\Phi\left(\dfrac{1{,}48-m}{\sigma}\right) = 0{,}975$. Отсюда

$$\begin{cases} \dfrac{1{,}141-m}{\sigma} = 1{,}282 \\ \dfrac{1{,}48-m}{\sigma} = 1{,}96 \end{cases}$$

и $\sigma = 0{,}5$, $m = 0{,}5$. Поэтому $x_{0{,}95} = m + \sigma z_{0{,}95} = 0{,}5 + 0{,}5 \cdot 1{,}645 = 1{,}3225$. **15.** По формуле Бернулли $P_5(3) = C_5^3 p^3 q^2$, где вероятность успеха $p = P(\xi \in [0{,}5,\ 1{,}5]) = 2\Phi(0{,}5) - 1 = 2 \cdot 0{,}6915 - 1 = 0{,}383$.

Глава 13

1. б) $\bar{x} = 3{,}5$, $s^2 = 3{,}65$, $S^2 \approx 4{,}06$, медиана равна 3.

Глава 14

1. а) Применяем двусторонний критерий: $m_0 \pm z_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 150 \pm z_{0{,}995} \cdot \dfrac{4}{6} = 150 \pm 2{,}576 \cdot \dfrac{4}{6} = 150 \pm 1{,}717$. Критическая область $\bar{x} > 151{,}717$, $\bar{x} < 148{,}283$. Выборочное среднее $\bar{x} = 154{,}5$ попало в критическую область: данные не подтверждают гипотезу; **б)** правосторонний критерий $\bar{x} > m_0 + z_{1-\alpha}\dfrac{\sigma}{\sqrt{n}} = 150 + 2{,}326 \cdot \dfrac{4}{6} = 151{,}55$ – гипотеза не подтверждается; **в)** гипотеза не отвергается по левостороннему критерию. **2.** Ошибка II рода произойдет, если выборочное среднее попадет в область принятия решения ($\bar{x} < 151{,}55$), тогда как истинное среднее равно 160, т.е. $\bar{x} \sim N(160, 4/6)$. Итак, $\beta = P(\bar{x} < 151{,}55) = \Phi\left(\dfrac{151{,}55-160}{4/6}\right) = \Phi(-12{,}675) \approx 0$. **3.** Применить критерий Стьюдента: квантили стандартного нормального распределения заменить квантилями распределения Стьюдента с 35 степенями свободы; дисперсию заменить на несмещенную оценку. **4.** Гипотеза отвергается при $\alpha = 0{,}1$.

Глава 15

1. Программа. **2.** Неправильные имена: MY FILE.DOC (имя из двух слов), 2.FILE.EXE (точка в имени), AAAAAAAAA.BBBB (имя из девяти символов, расширение из четырех), МОЙФАЙЛ.XYZ (имя по-русски). **3.** **а)** copy c:\student*.txt; **б)** dir d:\users/w; **в)** ren *.old *.new; **г)** md mycat; **д)** ver. **4.** Создать с клавиатуры текстовый файл BESTFILE.TXT в подкаталоге CAT1 каталога USERS на диске D:. **5.** path c:\ games. **6.** Ctrl Alt Del. **7.** Копировать, переносить, удалять.

Список литературы

Книги о математике и математиках

1. *Штейнгауз Г.* Задачи и размышления. М.: Мир, 1974.
2. *Клайн М.* Математика. Утрата определенности. М.: Мир, 1984.
3. *Клайн М.* Математика. Поиск истины. М.: Мир, 1988.
4. *Петер Р.* Игра с бесконечностью. М.: Просвещение, 1968.
5. *Боголюбов А.Н.* Математики, механики: Биограф. справ. Киев: Наук. думка, 1983.
6. Энциклопедический словарь юного математика / Сост. А.П. Савин. М.: Педагогика, 1983.
7. *Стройк Д.Я.* Краткий очерк истории математики. М.: Наука, 1984.

К части 1

8. *Столл Р.Р.* Множества. Логика. Аксиоматические теории. М.: Просвещение, 1968.
9. *Фор Р., Кофман А., Дени-Папен М.* Современная математика. М.: Мир, 1966.
10. *Важенин Ю.М.* Множества, логика, алгоритмы: Учеб. пособие. Екатеринбург: УрГУ, 1997.

К части 2

11. *Ефимов Н.В.* Краткий курс аналитической геометрии. М.: Физматгиз, 1962.
12. *Погорелов А.В.* Аналитическая геометрия. М.: Наука, 1968.
13. *Клетеник Д.В.* Сборник задач по аналитической геометрии. М.: Наука, 1975.
14. *Мальцев А.И.* Основы линейной алгебры. М.: Наука, 1970.

К части 3

15. *Никольский С.М.* Курс математического анализа. М.: Наука, 1983. Т. I.
16. *Берман Г.Н.* Сборник задач по курсу математического анализа. М.: Наука, 1977.
17. *Пизо Ш., Заманский М.* Курс математики: Алгебра и анализ. М.: Наука, 1971.

К частям 4 и 5

18. *Салий В.Н.* Введение в вероятность и теория информации для психологов: Учеб. пособие для психол. специальностей ун-тов. Саратов: Изд-во Сарат. ун-та, 1977.

19. *Чистяков В.П.* Курс теории вероятностей. М.: Наука, 1978.
20. *Гмурман В.Е.* Введение в теорию вероятностей и математическую статистику. М.: Высш. шк., 1966.
21. *Кремлев А.Г., Шелементьев Г.С.* Основные понятия теории вероятностей: Учеб. пособие. Свердловск: Изд-во Урал. ун-та, 1991.
22. Сборник задач по математике для втузов / Под ред. А.В. Ефимова. М.: Наука, 1990. Ч. 3: Теория вероятностей и математическая статистика.

К части 6

23. *Фигурнов В.Э.* IBM PC для пользователя. М.: Финансы и статистика, 1994.
24. *Кенин А.М., Печенкина Н.С.* Окно в мир компьютеров. Екатеринбург: Антарес-94, 1994.
25. *Петроченков А.В.* MS-DOS – не вопрос! Смоленск: РИЦ «ТОК», 1993.
26. *Денисов В.* Windows 95 с самого начала. СПб.: Питер, 1996.

Учебное издание

Владимир Яковлевич Турецкий

МАТЕМАТИКА И ИНФОРМАТИКА

Редактор *И.В. Мартынова*
Корректор *М.В. Литвинова*
Компьютерная верстка *О.В. Савостиной*
Оформление серии *Е.А. Доний*

ЛР № 070824 от 21.01.93

Подписано в печать 25.08.2000.
Формат 60×90/16. Усл.-печ. л. 35,0.
Тираж 6000 экз.
Заказ № 2435.

Издательский Дом «ИНФРА-М»,
127214, Москва, Дмитровское ш., 107.
Тел. (095) 485-70-63, 485-74-00.
Факс (095) 485-53-18. Робофакс 485-54-44
E-mail: books@infra-m.ru
http://www.infra-m.ru

Отпечатано в Тульской типографии.
300600, г. Тула, пр. Ленина, 109.

ISBN 5-16-000171-9

9 785160 001715