СРЕДНЕЕ ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАНИЕ

М. С. СПИРИНА, П. А. СПИРИН

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Допущено
Министерством образования Российской Федерации
в качестве учебника для студентов образовательных учреждений
среднего профессионального образования

2-е издание, стереотипное

Москва
Издательский центр «Академия»
2011

УДК 519.21(075.32)
ББК 22.171: 22.172я723
С722

Рецензенты:
канд. техн. наук *Н.А. Сосина* (зав. кафедрой естественно-математических дисциплин филиала Самарского государственного экономического университета в г. Тольятти);
преподаватель Московского государственного колледжа информационных технологий *В. В. Котельников*

Спирина М.С.
С722 Теория вероятностей и математическая статистика : учебник для студ. учреждений сред. проф. образования / М.С.Спирина, П.А.Спирин. — 2-е изд., стер. — М. : Издательский центр «Академия», 2011. — 352 с.
ISBN 978-5-7695-8210-3

В учебнике приведены основные элементы комбинаторики, понятия и теоремы теории вероятностей, рассмотрены случайные величины и методы математической статистики — выборки, статистических испытаний и др.

Для студентов учреждений среднего профессионального образования.

УДК 519.21(075.32)
ББК 22.171: 22.172я723

Учебное издание

Спирина Марина Савельевна, Спирин Павел Алексеевич

Теория вероятностей и математическая статистика

Учебник

Редактор *Л. В. Честная*. Технический редактор *О. Н. Крайнова*
Компьютерная верстка: *Г. Ю. Никитина*. Корректоры: *В. А. Жилкина*, *Г. Н. Петрова*

Изд. № 102112114. Подписано в печать 29.04.2011. Формат 60×90/16. Гарнитура «Таймс». Бумага офс. № 1. Печать офсетная. Усл. печ. л. 22,0. Тираж 1 000 экз. Заказ № 4809

ООО «Издательский центр «Академия». www.academia-moscow.ru
125252, Москва, ул. Зорге, д. 15, корп. 1, пом. 26б.
Адрес для корреспонденции: 129085, Москва, пр-т Мира, 101В, стр. 1, а/я 48.
Тел./факс: (495) 648-0507, 616-00-29.
Санитарно-эпидемиологическое заключение № РОСС RU. AE51. Н 14964 от 21.12.2010.

Отпечатано в ОАО «Тверской полиграфический комбинат».
170024, г. Тверь, пр-т Ленина, 5.
Телефон: (4822) 44-52-03, 44-50-34. Телефон/факс (4822) 44-42-15
Home page — www.tverpk.ru Электронная почта (E-mail) — sales@tverpk.ru

ISBN 978-5-7695-8210-3

ПЕРЕЧЕНЬ МАТЕМАТИЧЕСКИХ СИМВОЛОВ И СОКРАЩЕНИЙ

Символы

$\subset$ — включение
$\in$ — принадлежность элемента множеству
$\notin$ — непринадлежность элемента множеству
$\bigcup$ — объединение множеств
$\bigcap$ — пересечение множеств
$\sum$ — суммирование
$\forall$ — квантор общности
$\exists$ — квантор существования
$|M|$, $n(M)$ — мощность множества M
U — универсальное множество
$\bigcup\limits_i A_i$ — объединение множеств A_i
$\bigcap\limits_i A_i$ — пересечение множеств A_i
mes — мера множества
$n!$ — факториал числа n, т.е. произведение натуральных чисел от 1 до n
$\exists!$ — существует единственный элемент
A_n^m — число размещений из n по m
C_n^m — число сочетаний из n по m
P_n — число перестановок из n элементов
B^n — n-я декартова степень множества B

Характеристики случайной величины:

$M(X)$, Mx, m_x — математическое ожидание случайной величины X
$D(X)$, Dx — дисперсия случайной величины X
$\sigma(X)$, σ — среднеквадратическое отклонение
$F(x)$ — функция распределения
$f(x)$ — плотность вероятности
$\mathbb{N}$ — множество натуральных чисел
$\mathbb{Z}$ — множество целых чисел
$\mathbb{Q}$ — множество рациональных чисел
$\mathbb{R}$ — множество действительных чисел
$\mathbb{R}_+$ — множество неотрицательных действительных чисел $[0, \infty)$
$1...n$, $\overline{1, n}$ — множество натуральных чисел от 1 до n включительно

Сокращения

СВ — случайная величина
ДСВ — дискретная случайная величина
НСВ — непрерывная случайная величина
ФР — функция распределения
ПВ — плотность вероятности
ТСЧ — таблица случайных чисел
$N(m; \sigma)$ — нормальное распределение с математическим ожиданием m и среднеквадратическим отклонением σ
$\chi^2(n)$ — распределение Пирсона с n степенями свободы
$\mathrm{St}(n)$ — распределение Стьюдента с n степенями свободы

ПРЕДИСЛОВИЕ

Теория вероятностей как наука сформировалась свыше 300 лет назад. Это одна из самых молодых математических дисциплин. В настоящее время знание основ теории вероятностей и математической статистики необходимо специалистам самых различных профессий: от социологов до экономистов. Только к середине XX в. вероятностные методы нашли реальное применение в экономике и психологии, медицине и фармакологии, социологии и юриспруденции, в различных прикладных науках, таких, как кибернетика, теория игр, системы массового обслуживания и т.д. Планирование производства, контроль качества, принятие решений в различных сферах, анализ экспериментальных данных не обходятся без математической статистики, которая является методологией наук.

Наибольшие трудности при изучении элементов комбинаторики и теории вероятностей традиционно вызывают приемы решения задач, так как сложно алгоритмизировать процесс вычислений. Определенные попытки все же были предприняты. Например, для решения ряда комбинаторных задач для соединений без повторений удобно пользоваться алгоритмом (см. рис. 1.3).

Рекомендуется обратить внимание на логические связки «и» и «или», которые нужно применять при объяснении решения, так как с их помощью можно выбрать соответственно действия умножения и сложения, применяемые при решении комбинаторных и вероятностных задач.

При разработке как теоретических, так и практических вопросов предпочтение отдавалось наглядным методам (изображение условия на отрезке, в виде кругов Эйлера и т.д.).

Для лучшего усвоения определений аналогичных и противоположных понятий отличительные особенности этих определений представлены в виде двойной записи — например, вправо (влево) (метод «Укрупнение дидактических единиц» П. М. и Б. П. Эрдниевых). Так, запись предложения «Величина математического ожидания m влияет на расположение кривой нормального распределения $f(x)$ относительно оси ординат: при возрастании (убывании) m кривая смещается вправо (влево)» заменяет два предложения «Величина математического ожидания m влияет на расположение кривой нормального распределения $f(x)$ относительно оси ординат: при возрастании m кривая смещается вправо» и «Величина математического ожидания m влияет на расположение кривой нормального распределения $f(x)$ относительно оси ординат: при убывании m кривая смещается влево».

Подбор задач осуществляется таким образом, чтобы охватить как можно большее количество их при доступности изложения.

ВВЕДЕНИЕ

> Случай играет в мире столь большую роль, что обыкновенно я стараюсь отвести ему как можно меньше места в уверенности, что и без моей помощи он позаботится о себе.
>
> *А. Дюма «Три мушкетера»*

Математическая статистика — математическая наука, изучающая методы обработки результатов наблюдений.

Задача математической статистики заключается в том, чтобы по выборке — подмножеству ограниченных данных — выявить с определенной степенью точности характеристики всего множества — генеральной совокупности.

Методы математической статистики основываются на тесно связанной с ней теорией вероятностей.

Теория вероятностей — математическая наука, изучающая закономерности случайных явлений.

Под *случайными* понимаются такие явления, которые при многократном повторении опыта протекают каждый раз по-иному. Исход таких опытов непредсказуем.

Вероятность — количественная мера неопределенности — число, которое выражает степень уверенности в наступлении того или иного события.

Предметом теории вероятностей является изучение вероятностных закономерностей массовых однородных случайных событий.

Теория вероятностей в настоящее время — обязательный *инструмент анализа ситуаций, включающих неопределенность.* Основной *задачей* теории вероятностей является установление математических законов для исследования случайных явлений массового характера и предвидения их на основании отдельных фактов.

? Почему изучаются именно *случайные явления*, и может ли случайность носить *массовый* характер?

В окружающем мире мы постоянно имеем дело с различными случайными явлениями, но сравнительно мало о них знаем. История развития знаний есть история борьбы человека с «Его Величеством — Случаем». С давних времен люди несерьезно относились к случайностям: их природа была неизведана и потому непредсказуема.

? О каких закономерностях может идти речь, когда дело касается всего лишь некоторой случайности?

Однако еще в XVI—XVII вв. случайности стали изучать в связи с установлением закономерностей в азартных играх, а также потребностями страхования. Так, один из крупнейших ученых того времени Б. Паскаль (1623—1662) по просьбе азартного игрока — знатного француза де Мере — решал задачи, связанные с вероятностью выпадения определенных чисел при бросании игральной кости (кубика).

? Если история свидетельствует о том, что теория вероятностей зародилась в Средние века в связи с анализом азартных игр, да и в настоящее время очень часто предметом ее рассмотрения являются задачи о картах, монетах, игральной кости, то насколько ее можно считать наукой и нужна ли в настоящее время такая «наука»?

Сохранилась переписка крупнейших ученых XVII в. Б. Паскаля и П. Ферма (1601—1665), в которой высказано предвидение того, что за отдельными частными задачами по поиску закономерностей случайных событий стоит фундаментальная наука, изучающая массовые случайные явления.

В процессе решения различных достаточно сложных задач, связанных с практическими проблемами страхования и демографии (народонаселения), были получены первые практические и теоретические результаты в области теории вероятностей такими учеными, как Д. Кардано (1501—1576), Б. Паскаль, А. Муавр (1667—1754), X. Гюйгенс (1629—1695), П. Ферма (1601—1665) и др.

Первая попытка систематизации знаний и превращения их в науку «Теория вероятностей» была осуществлена в начале XVIII в. Я. Бернулли (1654—1705) в книге «Искусство предположений», изданной в 1713 г. Дальнейшее развитие теории вероятностей было связано с проблемами развития естествознания, с военным делом. Такие науки, как кинетическая теория газов, теория стрельбы и другие, поставили проблемы, решаемые вероятностными методами.

Так, выдающимся французским математиком П. Лапласом (1749—1827) были опубликованы работы «Мемуар о вероятностях» (1778) и «Аналитическая теория вероятностей» (1812).

Выявление различных закономерностей, установление порядка в знаниях о природе, об обществе, о человеке ограничивали сферу влияния случайности на жизнь людей. У ученых даже появилась надежда настолько познать мир, чтобы оградить человека от воли Случая. В представлении П. Лапласа это могло быть некоторое воображаемое «сверхсущество», которое в полном объеме знает все прошлое и будущее о мире. В таком мире уже не останется места для случайности, все будет заранее предопределено и детерминировано.

? Такой строго упорядоченный мир, конечно, был бы очень удобен: никакие случайности не могли бы повлиять на жизнь человека; все было бы предсказуемо,... но это была бы, наверное, неинтересная, предсказуемая и скучная жизнь?

Однако знания о вероятностной природе окружающего мира, почерпнутые из естествознания, дают уверенность в том, что «сверхсущество Лапласа» останется лишь вымыслом, фантазией. Случайность никогда не будет «укрощена», так же как и никогда не будет до конца познан реальный мир.

Немецкий физик В. Гейзенберг открыл (1927) так называемый принцип неопределенности, согласно которому, зная положение электрона в пространстве в данный момент времени, невозможно точно определить направление его последующего движения.

? Может это и есть объяснение неисчерпаемости случайности, ее вездесущности и необузданности?

Свободолюбивый нрав Случая, его неукротимая фантазия основаны на неисчерпаемости мира и неполноте наших знаний о нем. Поэтому Случайность надо не избегать, не бояться ее, а познавать. Оказывается «поведение» Случайности тоже подчиняется законам. И задача таких математических дисциплин, как теория вероятностей и математическая статистика, выявить эти закономерности и использовать для обоснования различных теоретических выводов практических наблюдений.

Большой вклад в современную науку «Теория ошибок» внес немецкий математик К. Гаусс (1777—1855), одним из первых заметивший закономерности в случайных явлениях и возможность с их помощью давать прогнозы, т. е. делать *прогностические* выводы.

В XIX в. теория вероятностей сформировалась как стройная математическая дисциплина благодаря работам выдающегося русского математика П. Л. Чебышева (1821—1894), а также его учеников А. А. Маркова (1856—1922) и А. М. Ляпунова (1857—1918).

Свою первую работу по теории вероятностей «Опыт элементарного анализа теории вероятностей» (1845) П. Л. Чебышев написал в 25 лет! Благодаря сформулированному и доказанному этими учеными «Закону больших чисел» (в различных формах) теория вероятностей нашла применение во многих областях человеческой деятельности, в различных науках. Появилась возможность делать обобщенные выводы по сравнительно небольшим результатам, полученным в ходе экспериментов. Так, с помощью закона больших чисел экспериментальные данные о вероятности бракованных изделий в выборке десяти изделий дают возможность предсказать количество бракованных изделий во всей большой партии из 500 изделий. А знание вероятности появления букв в слове дает возможность расположить определенным (хорошо известным) способом буквы алфавита на клавиатуре ЭВМ.

Полученные научные результаты теории вероятностей и математической статистики были использованы для развития других наук, особенно статистической физики, генетики, теории информации и др.

Расцвет теории вероятностей и математической статистики во второй половине XX в. связан с именами отечественных ученых А. Н. Колмогорова, Б. В. Гнеденко, А. Я. Хинчина и др. Широкую известность приобрели фундаментальные работы зарубежных ученых: Г. Крамера, Дж. фон Неймана, Р. Фишера, М. Кендалла, А. Стьюарта и др. В настоящее время это одна из перспективных развивающихся областей знаний, которая служит инструментом различных исследований.

? Если так важна роль теории вероятностей при изучении массовых явлений, то что же изучает математическая статистика?

Целью любой науки является описание, объяснение и прогнозирование явлений окружающего мира на основе тех законов, которые выявлены, сформулированы и доказаны. Благодаря наблюдениям за окружающей действительностью (природой, обществом, человеком) выявляются отдельные закономерности. Ученые формулируют их в виде гипотез и для подтверждения своих выводов должны проводить многократные наблюдения за исследуемым явлением с сохранением первоначально заданных основных условий. Например, для выведения высокоурожайного сорта сельскохозяйственной культуры агроном проводит экспериментальную работу в течение ряда лет на нескольких участках, имеющих одинаковые агрохимические показатели: состав почвы, климатические условия, количество и качество вносимых удобрений и др. Сравнивая результаты на экспериментальных участках с контрольным, который отличается по некоторому параметру (например, новый способ посадки и ухода за картофелем), агроном сможет сделать вывод о том, подтвердилась ли гипотеза об эффективности этого способа посадки и ухода за картофелем: высокая урожайность достигается за счет нового способа посадки.

Для проведения селекционной работы агроному важно знать, какое минимальное количество картофеля с контрольного и экспериментальных участков необходимо сравнить, чтобы получить вывод, которому можно *доверять*. С другой стороны, его интересует, насколько *надежны* полученные результаты, какими критериями пользоваться при их сравнении.

Аналогичные проблемы возникают в различных исследованиях самых разнообразных областей знаний. Ответы на эти вопросы дает математическая статистика. «*Статистика — это наука о распределениях*», так характеризовал эту науку «отец кибернетики», выдающийся американский ученый Н. Винер (1894—1964), который внес большой вклад в развитие теории вероятностей.

Методы теории вероятностей и математической статистики нашли широкое практическое применение в медицине и фармакологии, социологии и юриспруденции, психологии и экономике. Современные научные направления, такие, как квантовая механика и статистическая теория решений, теория информации и теория игр, теория очередей и теория катастроф, теория массового обслуживания и теория надежности, теория автоматического управления и теория стрельбы, теория ошибок и общая теория связи, астрономия, геодезия, метеорология, — это далеко не полный список теоретических и прикладных наук, использующих методы теории вероятностей и математической статистики.

Красноречивы сами термины этих наук: «*математическое ожидание*», «*надежность*», «*доверительные интервалы*», «*доверительная вероятность*»!

Потребности естествознания способствовали развитию математической статистики и теории вероятностей, так как аппарат этих наук в состоянии описать многочисленные явления природы. Благодаря этому к концу XX в. теория вероятностей и математическая статистика оказались интенсивно развивающимися областями математики и превратились в строго выстроенную фундаментальную математическую дисциплину, способную решать задачи современных естественных и общественных наук, экономики и техники.

В основе большинства научных знаний лежит наблюдение. Для того чтобы установить общие закономерности, которым подчинены эти явления, необходимо за ними многократно наблюдать в одинаковых условиях. Несмотря на то что нельзя предвидеть результаты единичного наблюдения, благодаря систематизации и изучению результатов многократных наблюдений, можно предсказать результат серии измерений в этих наблюдениях.

Напомним, что теория вероятностей занималась построением математической модели теоретического закона некоторой случайной величины, установлением вероятности «сложного» события по вероятностям логически связанных с ним «простых» событий.

В отличие от теории вероятностей математическая статистика по результатам наблюдений над случайными явлениями оценивает их вероятность, а также на основании многократных наблюдений проверяет некоторые предположения относительно их вероятностных закономерностей.

При исследовании реальных экономических процессов необходимо обрабатывать большое количество статистической информации по различным показателям, которые по сути есть случайные величины. Конкретные исследуемые в практической деятельности реализации этих случайных величин ограничены, что не дает возможность использовать традиционные методы математического анализа.

Для исследования статистической информации разработаны методы математической статистики, которые позволяют получить необходимые знания об исследуемом объекте, произведя направленный анализ, сделать обоснованные выводы.

Под *статистическими данными*, которые подлежат исследованию методами математической статистики, понимают данные наблюдений за значениями некоторой случайной величины или совокупности случайных величин, характеризующих исследуемый процесс. Знание методов математической статистики, владение ими служат необходимым условием успешного анализа статистической информации.

В настоящее время основами теории вероятностей и математической статистики необходимо владеть специалистам самых различных профессий: от социологов до экономистов.

Математическая статистика берет начало из английской «школы политических арифметиков» (XVII в.) и немецкой описательной школы. Существенный вклад в развитие математической статистики внесли отечественные математики: П.Л.Чебышев, А.А.Марков, А.М.Ляпунов, А.Н.Колмогоров и другие, а также зарубежные математики, такие, как К.Гаусс, С.Пуассон, К.Пирсон, Р.Фишер, В.Госсет (Стъюдент) и др. С различного рода случайностями мы сталкиваемся на каждом шагу. Поможет ли изучение науки о закономерностях случайностей избежать ошибок в реальной жизни при встречах с ними? Научно доказано, что благодаря знаниям основ теории вероятности и математической статистики у нас формируется вероятностное мышление, которое особенно необходимо развивать в себе любому современному человеку, стремящемуся к достижению успеха. Такое вероятностное мышление может проявляться, в частности, в принятии к руководству некоторых прагматических установок, вытекающих из двух принципов — *«практической невозможности маловероятных событий»* и *«практической уверенности»* (см. подразд. 2.10), среди которых следующие:

- будем считать практически достоверным событие, вероятность которого близка к единице;
- будем считать практически невозможным событие, вероятность которого близка к нулю.

Благодаря применению этих установок в качестве жизненных правил, в критических жизненных ситуациях срабатывает *«вероятностная интуиция»* и людям удается избегать многих ошибок.

? Неужели изучение «сухой» математики может изменить характер, а в конечном итоге и судьбу человека?

Вы, уважаемые читатели, находитесь в том замечательном возрасте, когда человек (если он умеет мыслить критично) видит свои недостатки и не только стремится от них избавиться, но приобрести какие-то на его взгляд необходимые качества. А что значит «мыслить критично»? Какие личностные качества необходимо иметь молодому человеку нового поколения — поколения XXI в.?

Оказывается, независимо от выбранной специальности востребованы в настоящее время как раз критичность мышления, его вариативность, умение самостоятельно принимать решения, критично работать с любой литературой, в том числе с информацией, представленной в виде графиков, схем, таблиц, уметь самостоятельно систематизировать информацию, владеть приемами научного мышления. Все эти и другие качества характерны для творческой личности, а именно такие специалисты — компетентные, конкурентно способные — востребованы современным обществом. Проанализировав наличие в себе перечисленных личностных качеств, у вас может возникнуть вопрос: «Как же их приобрести, как их в себе воспитать?». Надеемся, что этот учебник окажет реальную помощь для формирования качеств, характерных для творческой личности. А теория вероятностей и математическая статистика (говорят: «стохастическая математика»), как никакой другой учебный предмет, способствуют позитивным изменениям личности.

В учебнике предусмотрены:

• сгруппированные по видам задачи, для того чтобы увидеть аналогичные, решаемые одним методом, по одним и тем же алгоритмам;

• разные методы решения одной задачи, приучающие к поиску оптимального варианта решения проблемы (в том числе и жизненной);

• сюжеты задач, связанные не только с будущей профессиональной деятельностью, но и с реальными «жизненными» ситуациями;

• задания на составление собственных (авторских) задач, с помощью которых вы не только учитесь творчески мыслить, но и лучше усваиваете новый материал, понимая «механизмы» работы формул, их применение на практике;

• представление информации в разных видах, в том числе в систематизированных, а также задания на самостоятельную систематизацию учебной информации;

• «вопросы для размышлений», а также различные парадоксы, способствующие приобретению критичного мышления и т.д.

? Зачем в каждом пункте указаны его содержание или цели его изучения?

Умение формулировать цель своей деятельности и предпринимать шаги для ее достижения также является характерной особенностью компетентного, конкурентно способного специалиста.

Итак, мы начинаем изучать теорию вероятностей — науку, позволяющую по вероятностям одних случайных событий находить вероятности других случайных событий, связанных какимлибо образом с первыми... Иначе, *теория вероятностей* — наука, выясняющая закономерности, которые возникают при взаимодействии большого числа случайных факторов.

ГЛАВА 1
ОСНОВНЫЕ ПОНЯТИЯ И ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

1.1. Элементы комбинаторики

В этом подразделе приведены виды комбинаторных соединений, определение вида соединений и вычисление их количества. При рассмотрении элементов комбинаторики использован язык теории множеств и математической логики, а также изображение заданных множеств графически с помощью диаграмм Эйлера—Венна либо представление множества на отрезке. Как и в математической логике, для нас будет очень важно правильно употреблять связки «и» и «или», что соответствует в математике действиям умножения и сложения.

Комбинаторикой называется раздел математики, в котором решаются задачи на составление различных комбинаций из конечного числа элементов и подсчет всех возможных таких комбинаций.

Комбинаторика имеет большое практическое применение при сортировке изделий и определении их пригодности. Комбинаторные методы используются в теории случайных процессов, статистике, математическом программировании, вычислительной математике, планировании экспериментов, а также при решении многих вопросов естествознания и техники.

? Как отличить комбинаторные задачи от других видов задач и каковы их особенности?

Существует три основных вида задач в зависимости от постановки вопроса и того элемента комбинаторики, который применяется при их решении. Рассмотрим такие задачи.

Задача 1.1. Сколько существует способов, чтобы рассадить по четырем креслам квартет из одноименной басни И.А.Крылова?

Задача 1.2. Сколькими способами можно выбрать в группе из 30 человек одного старосту и одного физорга?

Задача 1.3. Сколькими способами можно в группе из 30 человек направить пять студентов для участия в пробеге?

Во всех трех задачах надо дать ответ на вопрос: «Сколькими способами?». Но в первой задаче есть существенное отличие от других: необходимо установить порядок во всем конечном множестве элементов, т.е. получить упорядоченное множество.

Пусть M — конечное множество, состоящее из m элементов, $f: M \to \{1, 2, ..., m\}$ — функция, задающая порядок на M, т.е. правило, по которому каждому элементу множества M ставится в соответствие натуральное число от 1 до m, причем одному числу из $\{1, 2, ..., m\}$ соответствует один элемент из M.

Тогда пару $\langle M, f\rangle$ назовем ***упорядоченным множеством***, или ***перестановкой*** из m элементов.

Пример 1.1. Три отличника из 11 «А» класса — Иванов, Петров и Сидорова — должны выйти к президиуму и получить грамоты. Их можно вызвать в алфавитном порядке фамилий, т.е. Иванова — первым, Петрова — вторым, Сидорову — третьей. Тем самым на множестве $M=$ {Иванов, Петров, Сидорова} задана функция порядка f:

$$f : \begin{cases} f(\text{Иванов}) = 1; \\ f(\text{Петров}) = 2; \\ f(\text{Сидорова}) = 3. \end{cases}$$

Можно было бы ввести порядок другим способом (например, по возрасту, росту, массе и др.). На этом же множестве рассмотрим другую функцию g, джентльмены пропустили даму вперед (обычная очередь — тоже упорядоченное множество):

$$g : \begin{cases} g(\text{Иванов}) = 2; \\ g(\text{Петров}) = 3; \\ g(\text{Сидорова}) = 1. \end{cases}$$

Для того чтобы отличить разные упорядоченные множества, будем (для краткости) записывать их в виде *кортежа*: сначала первый (в соответствии с функцией f) элемент (т.е. прообраз 1), затем второй и т.д.:

$$\langle M, f\rangle = \left\langle f^{-1}(1),\ f^{-1}(2),\ \ldots,\ f^{-1}(m-1),\ f^{-1}(m)\right\rangle.$$

Тогда в примере $\langle M, f\rangle$ = ⟨Иванов, Петров, Сидорова⟩; $\langle M, g\rangle$ = = ⟨Сидорова, Иванов, Петров⟩.

? Сколько на одном множестве из n элементов может быть различных функций порядка?

Число таких функций на множестве из n элементов называется ***числом перестановок*** из n элементов и обозначается P_n. Понятно, что оно не зависит от рода самого множества, а зависит лишь от количества элементов n. Очевидно, что $P_1 = 1$.

Пример 1.2. Перестановки из двух элементов (шаров) в виде кортежа (рис. 1.1).

Имеем $P_2 = 2 = 1 \cdot 2$.

Чтобы найти перестановки из трех элементов, надо к каждому из P_2 вариантов перестановок добавить третий элемент (скажем, зеленый шар), который может располагаться для каждой перестановки на одном из трех мест (до, между, после). Поэтому $P_3 = P_2 \cdot 3 = 6$. Аналогично, $P_4 = P_3 \cdot 4 = 24$ и т.д., т.е. для получения числа перестановок из n элементов нужно в каждую из P_{n-1} перестановок из $n - 1$ элементов вставить n-й элемент на n свободных мест. Число таких вариантов равно

Рис. 1.1

$$P_n = nP_{n-1} = n(n-1)P_{n-2} = \ldots = n(n-1)\ \ldots\ 2P_1 = n(n-1)\ \ldots\ 2 \cdot 1.$$

Для обозначения произведения n натуральных чисел от 1 до n используется знак факториала (!). Итак,

$$P_n = n! \qquad (1.1)$$

Нужно запомнить, что $1! = 1$ и $0! = 1!/1 = 1$ (перестановки из одного элемента — один вариант, а вариант выбора пустого множества из себя также единственен! Последнее можно считать соглашением, или определением, факториала нуля).

Теперь понятно, что для решения задачи 1.1, когда важно лишь, кто на каком кресле будет сидеть, можно пронумеровать кресла; тогда любой вариант рассаживания будет соответствовать одной перестановке. Всего таких вариантов $4! = 24$.

Для решения задачи 1.2 вспомним, что поскольку функции старосты и физорга различны (A — староста, B — физорг и B — староста, A — физорг), существуют различные варианты: если выбирать из 30 человек одного старосту, то будет 30 вариантов. Для выбора физорга остается 29 вариантов.

Поскольку каждому варианту уже выбранного старосты соответствует 29 вариантов выбора физорга, то всего возможно $30 \times 29 =$ $= 870$ вариантов. Такое очевидное правило подсчета числа вариантов *не зависимых* между собой явлений, соответствующее числу элементов декартова произведения множеств, будем называть ***правилом произведения***.

Фактически в подмножестве $m = 2$ элементов установлен порядок (функциональная значимость). Мы говорим «порядок важен», понимая под этим порядок в записи элементов этого множества, например при получении различной стипендии в кассе колледжа: {староста, физорг} или {физорг, староста}.

Размещением из n элементов по m называется упорядоченное подмножество, содержащее m элементов множества, состоящего из n элементов. Число размещений из n элементов по m обозначается A_n^m (от фр. *arrangement* — размещение).

Таким образом, для решения задачи 1.2 нужно из множества учащихся, состоящего из 30 элементов выбрать упорядоченное подмножество из двух элементов. Тогда решением задачи 1.2 будет являться число A_{30}^2.

А в задаче 1.3 не важно, как побегут пять спортсменов, их роли в забеге одинаковы (предполагается, что фактор тактики отсутствует, например в спринте), т.е. из множества $n = 30$ элементов выбрали подмножество из $m = 5$ элементов, но без установления порядка в нем: вариант, когда уже отобранные спортсмены побежали, — единственный.

Сочетанием из n элементов по m называется (неупорядоченное) подмножество, содержащее m элементов множества, состоящего из n элементов. Число сочетаний из n элементов по m обозначается C_n^m (от фр. *combinaison* — комбинация, сочетание).

Пример 1.3. Пусть даны n элементов, из которых m — бракованные. Сколькими способами их можно выбрать? Пусть число сочетаний, т.е. число неупорядоченных подмножеств, содержащих по m бракованных изделий, равно C_n^m. Если из n элементов m — бракованных, то в остальных $n - m$ элементах — нет брака. Тогда каждому разбиению на два подмножества брак/небрак соответствуют $P_m P_{n-m}$ вариантов упорядочивания отдельно подмножества бракованных и подмножества небракованных изделий.

Чтобы получить все $n!$ способов в этом множестве, надо воспользоваться правилом произведения, учитывая ***и*** перестановки во множестве m элементов ($m!$), ***и*** перестановки во множестве $n - m$ элементов $(n - m)!$, ***и*** сочетания между ними C_n^m. Тогда по правилу произведения $P_n = P_m P_{n-m} C_n^m$, откуда формула числа сочетаний примет вид:

$$C_n^m = \frac{n!}{m!(n-m)!}. \tag{1.2}$$

Решение задачи 1.3 примет вид $C_{30}^5 = \dfrac{30!}{5!25!} = 142\,506.$

В отличие от этой задачи, где не важно, как бегут все пять спортсменов в общем забеге, в задаче про выборы физорга и старосты важен порядок во вновь образованном подмножестве из m элементов, т.е. имеем

$$A_n^m = C_n^m P_m = \frac{n!}{(n-m)!}.$$

Воспользовались правилом произведения, потому что ***и*** количество подмножеств из n по m надо учесть, ***и*** порядок во вновь созданных множествах из m элементов. Итак,

$$A_n^m = \frac{n!}{(n-m)!}. \tag{1.3}$$

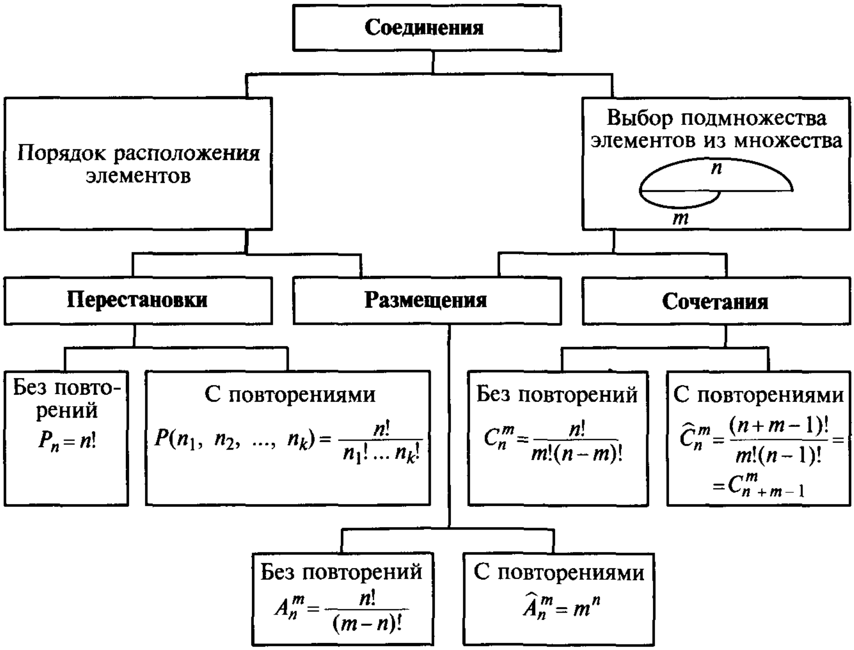


Рис. 1.2

В формулах размещений и сочетаний удобно выполнить сокращение на $(n - m)!$ элементов. Тогда $A_n^m = (n - m + 1) \ldots (n - 1)n$.

Заметим, что при $m = n$, $A_n^n = P_n = n!$, т.е. перестановки есть частный случай размещений, когда подмножество из m элементов «доросло» до самого множества из n элементов.

Элементы комбинаторики можно классифицировать по их отношению к установленному порядку в подмножествах (рис. 1.2).

1.2. Задачи на непосредственное применение формул комбинаторики

В этом подразделе вы научитесь применять разные типы соединений для решения конкретных задач.

Из всего разнообразия математических задач задачи по комбинаторике легко узнаются по формулировке поставленного вопроса: «Сколькими способами?», «Сколько вариантов?» и т.д.

При решении комбинаторных задач (без повторений) удобно использовать алгоритм (рис. 1.3), в котором требуется найти ответы на два «ключевых» вопроса:

- во всем ли множестве необходимо найти число возможных вариантов?

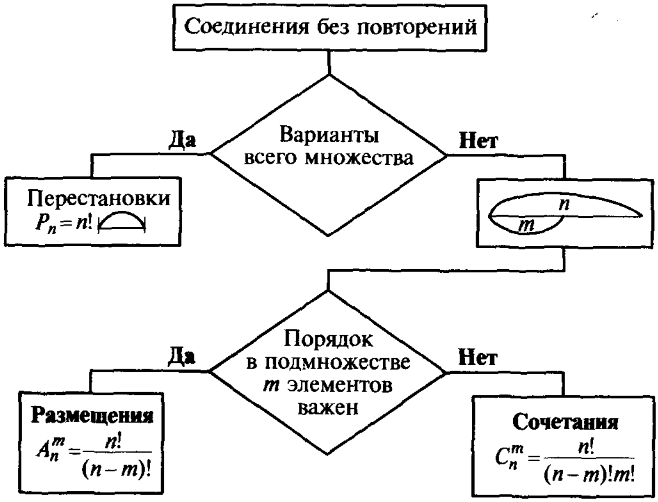


Рис. 1.3

- важен ли порядок расположения отдельных элементов в исследуемом множестве (подмножестве)?

Задача 1.4. Восемь студентов обменялись рукопожатиями. Сколько было рукопожатий?

Решение. В рукопожатии участвует «подмножество», состоящее из двух студентов ($m = 2$), тогда как все «множество» студентов составляет 8 человек ($n = 8$). Наглядное представление «на отрезке» соотношения подмножества и множества удобно для выбора соединения (рис. 1.4).

Так как в процессе рукопожатия порядок не важен, выбираем формулу сочетаний

$$C_8^2 = \frac{8!}{2!6!} = \frac{7 \cdot 8}{1 \cdot 2} = 28.$$

Задача 1.5. Сколькими способами можно составить трехцветный полосатый флаг из пяти различных по цвету отрезов материи?

Решение. Порядок важен, так как перестановка материи внутри трехцветного флага обозначает разные страны. Поэтому выбираем формулу размещений без повторений, где множество отрезов материи содержит $n = 5$, а подмножество — $m = 3$ цветов:

$$A_5^3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = 3 \cdot 4 \cdot 5 = 60.$$

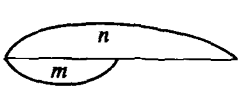


Рис. 1.4

Задача 1.6. Сколько словарей надо издать, чтобы можно было выполнять переводы с любого из шести языков на любой из них?

Решение. Множество включает $n = 6$ языков. Поскольку перевод есть отношение между двумя языками, то $m = 2$, причем порядок важен, так как, например, словари англо-русский и русско-английский имеют различное применение. Поэтому выбираем размещения без повторений:

$$A_6^2 = \frac{6!}{(6-2)!} = \frac{6!}{4!} = 5 \cdot 6 = 30.$$

Задача 1.7. Сколько имеется вариантов составления расписания на понедельник, если предметов у студентов 9, а в понедельник четыре пары занятий и предметы не повторяются?

Решение. а) Для студентов порядок не важен, поэтому выбираем сочетания без повторений: $C_9^4 = \frac{9!}{4!5!} = \frac{6 \cdot 7 \cdot 8 \cdot 9}{1 \cdot 2 \cdot 3 \cdot 4} = 126.$

б) Для преподавателей порядок важен, поэтому выбираем формулу размещений без повторений: $A_9^4 = \frac{9!}{5!} = 6 \cdot 7 \cdot 8 \cdot 9 = 3024.$

Задача 1.8. Сколькими способами можно расставить на книжной полке девять книг, среди которых есть трехтомник А.С.Пушкина?

Решение. Так как три тома, входящие в трехтомник, должны стоять рядом, причем по возрастанию номера тома слева направо, рассматриваем их как один элемент данного множества, в котором имеется еще шесть элементов. Поэтому выбираем перестановки без повторений во множестве, содержащем семь элементов:

$$P_7 = 7! = 5040.$$

Задача 1.9. Сколькими способами можно назначить в группе из 30 человек трех дежурных?

Решение. а) Если их роль в процессе дежурства одинакова, то порядок не важен, поэтому выбираем сочетания без повторений: $C_{30}^3 = \frac{30!}{3!27!} = 4060.$

б) Если порядок важен, т.е. во время дежурства их функциональные обязанности различны, то по формуле размещений без повторений имеем: $A_{30}^3 = \frac{30!}{27!} = 24360.$

Задача 1.10. Сколько существует шестизначных телефонных номеров, у которых: а) возможны любые цифры; б) все цифры различные?

Решение.

а) 1. Так как в шестизначном наборе телефонного номера возможны любые цифры, то на каждом из шести мест может встретиться любая из десяти цифр от 0 до 9. Необходимо из всех возможных десяти цифр выбрать лишь те шесть, которые будут использованы для шестизначных телефонных номеров. Поскольку в записи телефонных номеров порядок расположения цифр важен, по формуле размещений с повторениями имеем: $\bar{A}_{10}^6 = 10^6 = 1\,000\,000$.

2. Как известно, не бывает шестизначных номеров, начинающихся с нуля, поэтому надо подсчитать их количество и вычесть его из общего числа комбинаций. Число номеров, первая цифра у которых 0, найдем по формуле размещений с повторениями, «зафиксировав ноль», т.е. на каждом из пяти остальных возможных мест может встретиться любая из десяти цифр от 0 до 9. Тогда число таких комбинаций $\bar{A}_{10}^5 = 10^5 = 100\,000$.

3. Общее число шестизначных телефонных номеров, у которых могут быть любые, в том числе и повторяющиеся, цифры, равно разности

$$\bar{A}_{10}^6 - \bar{A}_{10}^5 = 10^6 - 10^5 = 1\,000\,000 - 100\,000 = 900\,000.$$

б) 1. Пусть теперь в шестизначном наборе все цифры различные. Необходимо из всех возможных десяти цифр выбрать лишь те шесть, которые используются для шестизначных телефонных номеров, причем никакая цифра не повторяется. Тогда по формуле размещений без повторений имеем

$$A_{10}^6 = \frac{10!}{(10-6)!} = 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 = 151\,200.$$

2. Поскольку шестизначных номеров, начинающихся с нуля, не бывает, надо подсчитать их количество и вычесть его из общего числа комбинаций. Число номеров, первая цифра у которых 0, найдем по формуле размещений без повторений, «зафиксировав ноль», т.е. на каждом из пяти оставшихся возможных мест могут встретиться цифры от 0 до 9. Тогда число таких комбинаций найдем по формуле размещений без повторений. Имеем

$$A_{10}^5 = \frac{10!}{(10-5)!} = 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 = 30\,240.$$

3. Общее число шестизначных телефонных номеров, у которых не может быть повторяющихся цифр, равно разности:

$$A_{10}^6 - A_{10}^5 = 151\,200 - 30\,240 = 120\,960.$$

Задача 1.11. Сколькими способами можно выделить делегацию в составе трех человек, выбирая их среди четырех супружеских пар, если:

а) в состав делегации входят любые трое из данных восьми человек;

б) делегация должна состоять из двух женщин и одного мужчины;

в) в делегацию не входят члены одной семьи?

Решение. а) Порядок не важен: $C_8^3 = \dfrac{8!}{3!5!} = 56.$

б) Выберем двух женщин из имеющихся четырех C_4^2 способами и одного мужчину из четырех C_4^1 способами. По правилу произведения (***и*** мужчина, ***и*** две женщины) имеем $C_4^2 \cdot C_4^1 = 24.$

в) Из четырех семей выбираем трех членов делегации четырьмя ($C_4^3 = \dfrac{4!}{3!1!} = 4$) способами. Но в каждой семье имеется по два способа выбора члена делегации. По правилу произведения $C_4^3 \cdot 2 \cdot 2 \cdot 2 = C_4^3 \cdot 2^3 = 4 \cdot 8 = 32.$

Задача 1.12. В колледже учится 2000 студентов. Можно ли утверждать, что хотя бы двое из них имеют одинаковые инициалы и имени, и фамилии?

Решение. В русском алфавите 33 буквы, из них ъ, ь, ы, й не могут быть использованы, поэтому $n = 33 - 4 = 29$. Каждая из 29 букв может быть инициалом ***и*** имени, ***и*** фамилии. По правилу произведения $29 \cdot 29 = 841 < 2000$. Значит, может быть лишь 841 различных вариантов, и среди 2000 студентов обязательно будут совпадения.

1.3. Треугольник Паскаля. Бином Ньютона

В этом подразделе вы узнаете о необыкновенной числовой конструкции — треугольнике Паскаля и его необычных свойствах, а также научитесь возводить двучлен (бином) в любую натуральную степень. С частными случаями этой формулы вы давно знакомы (при $n = 2$ и $n = 3$). Но в общем виде эту формулу нельзя записать без знаний комбинаторных операций.

Рассмотрим бесконечный числовой треугольник (рис. 1.5), построенный таким образом, что по краям каждой строки стоят единицы, а остальные числа получены при сложении двух стоящих над ними чисел предыдущей строки.

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & & & & & \dots\dots & & & & &
\end{array}
$$

Рис. 1.5

Этот треугольник назван по имени Б. Паскаля — математика и философа, который привел его в такой форме в работе «Трактат об арифметическом треугольнике».

Треугольник Паскаля обладает рядом интересных особенностей, рассмотрим некоторые из них.

1. Числа, составляющие треугольник Паскаля, представляют собой сочетания из n по m: C_n^m, где n — номер строки, начиная с нулевой: $n \in \mathbb{N} \cup \{0\}$, а m — номер числа в этой строке, начиная с нуля. При этом $C_0^0 = 1$, $C_n^1 = C_n^{n-1} = n$, $C_n^0 = C_n^n = 1$.

Элементы любой строки треугольника Паскаля обладают симметрией: т.е. равностоящие от концов каждой строки числа равны между собой, что становится очевидным, исходя из формулы сочетаний:

$$C_n^m = \frac{n!}{m!(n-m)!} = C_n^{n-m}. \tag{1.4}$$

2. Рассмотрим *доску Гальтона.* Она «похожа» на треугольник Паскаля, только на месте чисел стоят штыри. Скатываясь произвольно по доске, шарик диаметра, равного расстоянию между штырями, упруго ударяется о штыри и, отталкиваясь от них, двигается или влево, или вправо. Если сосчитать число вариантов прохождения шарика по доске к данному штырю, то оказывается, что оно полностью совпадает с соответствующим числом треугольника Паскаля, поскольку оно как раз является суммой вариантов прохождения к двум вышестоящим штырям. Например, ко второму штырю третьей строки шарик попадает с левого и правого штыря второй строки; число вариантов — $1+1=2$. А в треугольнике Паскаля на этом месте C_2^1. Или на втором месте пятой строки треугольника Паскаля стоит C_4^1 (рис. 1.6).

На соответствующий штырь доски Гальтона шарик попадает, отталкиваясь от левого и правого штырей четвертой строки, при этом число ударов в сумме $1+3=4$.

Элементы доски Гальтона обладают такой же симметрией, что и числа в треугольнике Паскаля.

3. Свойство симметрии чисел, равностоящих от концов каждой строки треугольника Паскаля, можно записать в виде формулы

$$C_n^m = C_n^{n-m}.$$

Действительно:

$$C_n^{n-m} = \frac{n!}{(n-m)!(n-(n-m))!} = \frac{n!}{(n-m)!m!} = C_n^m.$$

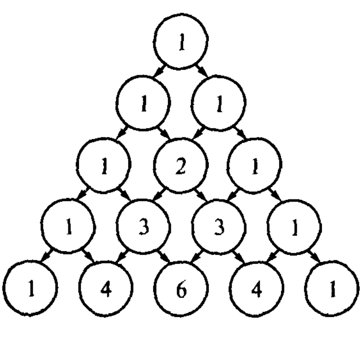


Рис. 1.6

Кроме математического доказательства полезно заметить, что число сочетаний из n по m есть число сочетаний пустот из n по $n-m$. Этот теперь уже философский факт находит отражение, в частности, в современной физике, где помимо движения электронов рассматривается движение «дырок».

4. Тот факт, что любое число треугольника Паскаля можно получить сложением двух чисел, стоящих над ним на предыдущей строке, можно записать в виде формулы:

$$C_{n-1}^{m} + C_{n-1}^{m-1} = C_n^m. \qquad (1.5)$$

Действительно:

$$C_{n-1}^{m} + C_{n-1}^{m-1} = \frac{(n-1)!}{(m-1)!\,(n-1-(m-1))!} + \frac{(n-1)!}{m!\,(n-1-m)!} =$$

$$= \frac{(n-1)!}{(m-1)!\,(n-m)!} + \frac{(n-1)!}{m!\,(n-m-1)!} =$$

$$= \frac{(n-1)!}{(m-1)!\,(n-m-1)!}\left(\frac{1}{n-m} + \frac{1}{m}\right) = \frac{n!}{m!\,(n-m)!} = C_n^m.$$

5. Для любой строки треугольника Паскаля справедлива формула: сумма элементов n-й строки равна 2^n:

$$\sum_{m=0}^{n} C_n^m = 2^n. \qquad (1.6)$$

6. Рассмотрим коэффициенты в разложении натуральных степеней двучлена (бинома) по степеням:

$$(a+b)^0 = \mathbf{1}$$
$$(a+b)^1 = \mathbf{1}a + \mathbf{1}b$$
$$(a+b)^2 = \mathbf{1}a^2 + \mathbf{2}ab + \mathbf{1}b^2$$
$$(a+b)^3 = \mathbf{1}a^3 + \mathbf{3}a^2b + \mathbf{3}ab^2 + \mathbf{1}b^3 \text{ и т.д.}$$

Оказывается, коэффициенты бинома в разложении по степеням также совпадают с соответствующими числами строки треугольника Паскаля, причем n изменяется от 0 до ∞, а m — от 0 до n.

Непосредственно можно проверить, что

$$(a+b)^3 = \sum_{m=0}^{3} C_3^m a^{3-m} b^m; \quad (a+b)^4 = \sum_{m=0}^{4} C_4^m a^{4-m} b^m \ldots,$$

а с помощью метода неполной математической индукции **можно** обобщить:

$$(a+b)^n = \sum_{m=0}^{n} C_n^m a^{n-m} b^m. \qquad (1.7)$$

Эта формула и получила название бинома Ньютона в честь ученого, обобщившего в 1676 г. известные до него формулы на случай, когда показатель n — произвольное натуральное число. Формула может быть доказана методом (полной) математической индукции. Числа C_n^m поэтому также называют *биномиальными коэффициентами.*

Бином Ньютона обладает следующими свойствами:

1) в разложении по степеням коэффициенты n-й степени бинома совпадают с $(n+1)$-й строкой треугольника Паскаля;

2) сумма коэффициентов в разложении n степени равна 2^n (см. формулу (1.4)). Действительно, при $a=b=1$ бином $(a+b)^n$ примет вид:

$$(1+1)^n = \sum_{m=0}^{n} C_n^m 1^{n-m} 1^m = \sum_{m=0}^{n} C_n^m = 2^n;$$

3) в разложении n степеней бинома содержится $n+1$ слагаемых, причем каждое слагаемое (одночлен) представляет собой произведение трех множителей: биномиального коэффициента C_n^m и степеней первого и второго члена бинома. Одночлен, стоящий на $(k+1)$-м месте (начиная с первого) в разложении бинома n-й степени, можно найти по формуле $T_{k+1} = C_n^k a^{n-k} b^k$;

4) в разложении в многочлене показатели степени первого члена бинома уменьшаются от одного слагаемого к другому на один (от n до 0), а показатели второго члена соответственно увеличиваются на один (от нуля до n). Поэтому сумма показателей в каждом слагаемом (одночлене) в разложении бинома равна n. Это должно быть очевидно из физических соображений размерности, если a и b — размерные величины, например метры.

Задача 1.13. Найти разложение $(1-\sqrt{2})^6$.
Решение.

$$(1-\sqrt{2})^6 = C_6^0 1^6(-\sqrt{2})^0 + C_6^1 1^5(-\sqrt{2})^1 + C_6^2 1^4(-\sqrt{2})^2 + C_6^3 1^3(-\sqrt{2})^3 +$$
$$+C_6^4 1^2(-\sqrt{2})^4 + C_6^5 1^1(-\sqrt{2})^5 + C_6^6 1^0(-\sqrt{2})^6 = 1+6(-\sqrt{2})+15\cdot 2+$$
$$+20(-2\sqrt{2})+15\cdot 4+6(-4\sqrt{2})+1\cdot 8 = 1-6\sqrt{2}+30-40\sqrt{2}+60-$$
$$-24\sqrt{2}+8 = 99-70\sqrt{2}.$$

Задача 1.14. Найти девятый член разложения $(2+\sqrt{a})^{12}$.
Решение. Девятый член $T_{k+1} = T_9$, значит, $k=8$, $n=12$.
По формуле $T_{k+1} = C_n^k a^{n-k} b^k$ найдем $T_9 = C_{12}^8 2^{12-8}(\sqrt{a})^8 = \dfrac{12!}{8!4!} 2^4 a^4$.

$$T_9 = \frac{12\cdot 11\cdot 10\cdot 9}{1\cdot 2\cdot 3\cdot 4}\cdot 2^4\cdot a^4 = 7920a^4.$$

Задача 1.15. Найти средний член разложения бинома Ньютона $\left(2x^2y - \frac{1}{x^2y}\right)^6$.

Решение. Средний член стоит на четвертом месте, так как всего слагаемых $6 + 1 = 7$. Тогда $k = 3$, $n = 6$:

$$T_4 = T_{1+3} = C_6^3(2x^2y)^3 \frac{1}{(x^2y)^3} = \frac{6!}{3!3!}\frac{2^3x^6y^3}{x^6y^3} = 160.$$

Задача 1.16. Разложить по степеням $(2 + i)^6$.

Решение. Используя свойства $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, получаем

$$(2+i)^6 = 64 + 6(2^5 i + 2i^5) + 15(2^4 i^2 + 2^2 i^4) + 20 \cdot 2^3 i^3 + i^6 = -117 + 44i.$$

1.4. Виды случайных событий. Операции над событиями

В этом подразделе вы познакомитесь с «алгеброй событий», научитесь описывать пространство элементарных событий. При изучении операций над событиями полезно провести аналогию между ними, операциями над множествами и операциями в математической логике.

Под *событием* принято понимать всякий факт, который может произойти в данных условиях. Совокупность условий, при которых событие может произойти, а может не произойти, назовем заданным *комплексом условий* S. Под *испытанием*, *опытом* или *экспериментом* понимают реализацию (воспроизведение) определенного комплекса условий S. ***Случайным*** называется такое ***событие***, которое может произойти, а может не произойти при заданном комплексе условий S. Каждое событие является совокупностью элементарных событий. ***Элементарным событием*** (***исходом***) назовем каждый из возможных результатов случайного испытания.

Множество всех возможных в результате испытаний элементарных событий называется *пространством* элементарных событий и обозначается $\Omega = \{\omega_1, \omega_2, \omega_3, ...\}$, а сами элементарные исходы — $\omega_1, \omega_2, \omega_3$ и т.д.

Можно заметить, что между различными событиями и явлениями окружающего мира существуют связи. Среди этого многообразия взаимосвязей наблюдаются такие, у которых исход однозначно предопределен, и такие, в которых исход неоднозначен.

? Подчиняются ли строгим математическим законам явления, носящие случайный характер?

В зависимости от результата исходов многократного воспроизведения заданного комплекса условий S события можно условно классифицировать следующим образом (рис. 1.7).

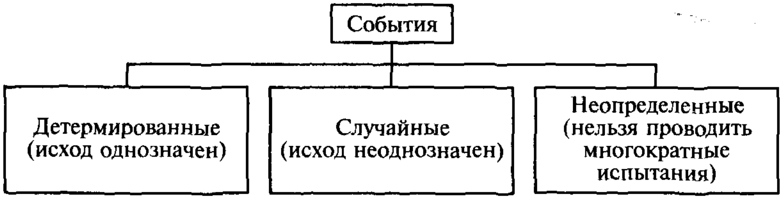


Рис. 1.7

Заранее известен исход во многих явлениях классической физики, поэтому классическую физику называют *детерминированной*. В то же время нельзя определить результат исхода войны или матча, так как их невозможно повторить, сохранив все начальные условия без изменений.

Если при повторении некоторого начального комплекса условий будущее состояние системы определено не однозначно, а лишь с некоторой вероятностью, то рассматривают случайные события: выявление бракованной детали, результат стрельбы по мишени и т.д.

Теория вероятностей изучает свойства массовых случайных событий, способных многократно повторяться при воспроизведении определенного комплекса условий, а также при взаимодействии большого числа случайных факторов.

Обозначать события принято прописными буквами латинского алфавита, например, A, B, C, указывая в фигурных скобках множество элементарных событий ω_i, из которых оно состоит: $A = \{\omega_1, \omega_4, \omega_9\}$.

Таким образом, случайное событие называют ***элементарным***, если оно удовлетворяет условиям:

• два элементарных случайных события никогда одновременно не реализуются в стохастическом эксперименте: $\omega_i \cap \omega_j = \varnothing$, где $i \neq j$;

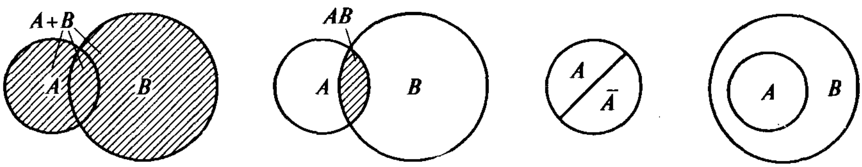
• конкретный исход эксперимента всегда есть элементарное событие, одно из множества элементарных событий Ω.

Элементарные события служат «кирпичиками», из которых строится описание любого события стохастического эксперимента.

Рассмотрим отдельные виды событий и отношения между ними. При этом будем использовать аналогию между языком теории множеств (с операциями объединения и пересечения множеств) и языком математической логики (с операциями дизъюнкции и конъюнкции высказываний соответственно).

Достоверным называется событие Ω (или U), если оно обязательно произойдет в данном испытании в результате выполнения комплекса условий S.

Невозможным называется событие $\varnothing$ (или V), если оно никогда не произойдет в данных испытаниях в результате выполнения



а *б* *в* *г*

Рис. 1.8

совокупности условий S. Например, «правильная» монета не может одновременно выпасть орлом и решкой.

Суммой (или *объединением*) событий A и B называется событие $A \cup B$ $(A+B)$, состоящее в наступлении *хотя бы* одного из событий — A или B (рис. 1.8, *а*). Для применения правила суммы используют «ключевое» слово «***или***». Например, гол в ворота команды «Динамо» засчитывается в том случае, если команда-противник забьет гол (A) или если динамовцы забьют гол в свои ворота.

Произведением событий A и B называется событие AB, $A \cap B$, состоящее в совместном выполнении *одновременно и* события A, ***и*** события B (рис. 1.8, *б*). Для применения правила произведения используют «ключевое» слово «***и***». Например: рекорд в беге на 100 м фиксируется, если преодолен рубеж прежнего рекорда и скорость попутного ветра была в пределах допустимого значения.

Несовместными называют события, если наступление одного из них в том же испытании исключает наступление другого. Например, несовместными будут события «деталь — стандартная» и «деталь является бракованной». Таким образом, два события A и B несовместны, если их произведение есть невозможное событие: $AB = \varnothing$.

Несколько событий образуют *полную группу событий*, если в результате испытаний произойдет хотя бы одно из них. В частности, совокупность несовместных событий A_i образует *полную группу событий*, если в результате единичных испытаний произойдет обязательно одно из этих событий:

$$\sum_{i=1}^{n} A_i = \Omega, \text{ где } n \in \mathbb{N},\ A_i A_j = \varnothing \text{ для } i \neq j.$$

Например, выпадение чисел от 1 до 6 при подбрасывании игральной кости.

Противоположными называются два несовместных события A и $\bar{A}$, образующие полную группу событий. Для противоположных событий справедливо тождество $\bar{\bar{A}} = A$ (двойное отрицание снимается). Произведение противоположных событий $A\bar{A} = \varnothing$ есть невозможное событие. Сумма противоположных событий $A + \bar{A} = \Omega$ — достоверное событие.

Так, противоположными являются события «поражение мишени» и «промах» («третьего не дано»), рис. 1.8, *в*.

Если из осуществления события A следует осуществление события B (говорят: A влечет за собой B): $A \subset B$, т.е. все элементарные события (могущие привести к реализации A, содержатся также и в B), то говорят, что из A следует B. Для высказывательных форм (утверждений), например при формулировке теоремы, принято обозначение $A \Rightarrow B$.

Например, из правильных ответов на экзамене следует получение пятерки (рис 1.8, *г*).

Равными ($A = B$), или *равносильными*, называют такие два события A и B, для которых *одновременно* выполняются условия $A \to B$, $B \to A$ ($A \subset B$, $B \subset A$). Например, победа в финале и получение спортсменами кубка.

Благоприятствующими событию A называют те элементарные исходы, при которых наступает изучаемое событие A.

Равновозможными называют такие элементарные события, которые при создании комплекса условий S имеют одинаковые шансы для их наступления. Так, равновозможны выпадения цифр 5 и 3 при одноразовом бросании симметричного кубика (игральной кости).

Пример 1.4. Пусть дано трехэлементное множество элементарных случайных событий $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Для него существует восемь различных подмножеств: $A_1 = \{\varnothing\}$, $A_2 = \{\omega_1\}$, $A_3 = \{\omega_2\}$, $A_4 = \{\omega_3\}$, $A_5 = \{\omega_1, \omega_2\}$, $A_6 = \{\omega_1, \omega_3\}$, $A_7 = \{\omega_2, \omega_3\}$, $A_8 = \{\omega_1, \omega_2, \omega_3\}$.

Классическим стохастическим экспериментом (т.е. моделью, на которой он изучается) является «игра в кости». Под «костью» понимают правильный шестигранный кубик с пронумерованными гранями. Элементарное событие ω_i — выпадение грани с номером i (число очков), где $1 \le i \le 6$. Найдем элементарные события, *благоприятствующие* событию B_k, если:

1) выпадение четной грани — событие B_1;
2) выпадение нечетной грани — событие B_2;
3) выпадение грани с номером, больше трех, — событие B_3;
4) выпадение грани с номером, не больше трех, — событие B_4;
5) выпадение грани с номером, меньше трех, — событие B_5;
6) выпадение грани с номером, не меньше трех, — событие B_6;
7) выпадение нечетной грани с номером, больше трех, — событие B_7;
8) выпадение нечетной грани с номером, не больше трех, — событие B_8;
9) выпадение нечетной грани с номером, больше пяти, — событие B_9.

Тогда события можно представить в виде множеств, состоящих из введенных элементарных событий: $B_1 = \{\omega_2, \omega_4, \omega_6\}$, $B_2 = \{\omega_1, \omega_3,$

$\omega_5\}$, $B_3 = \{\omega_4, \omega_5, \omega_6\}$, $B_4 = \{\omega_1, \omega_2, \omega_3\}$, $B_5 = \{\omega_1, \omega_2\}$, $B_6 = \{\omega_3, \omega_4, \omega_5, \omega_6\}$, $B_7 = \{\omega_5\}$, $B_8 = \{\omega_1, \omega_3\}$, $B_9 = \{\varnothing\}$.

Другими общепринятыми моделями для разбора решения комбинаторных или вероятностных задач могут быть шарики разного цвета, вынимаемые из непрозрачной урны, монеты или карты. Несмотря на то что в реальной жизни (на практике) такие задачи никто не решает, тем не менее, они удобны при разборах решений аналогичных задач своей наглядностью и абстрактной простотой, так как не «привязывают» внимание к конкретному содержанию.

1.5. Определения вероятности

Целью этого подраздела является знакомство с четырьмя различными способами определения вероятности событий.

Количественной мерой степени возможности осуществления события в заданном комплексе условий S служит *вероятность*. Вероятность события A обозначается $P(A)$ или $P\{A\}$. Дадим *классическое* определение вероятности.

Вероятностью события A называется отношение числа исходов m, благоприятствующих наступлению события A, к числу n всех несовместных равновозможных элементарных исходов, образующих полную группу событий:

$$P(A) = \frac{m}{n}. \tag{1.8}$$

Из определения следует, что вероятность удовлетворяет условиям

$$0 \le P(A) \le 1,\ P(\Omega) = 1,\ P(\varnothing) = 0.$$

Классическое определение вероятности применяется только в следующих случаях:

- число элементарных исходов конечно;
- результаты всех испытаний или наблюдений *равновозможны*;
- все равновозможные события образуют *полную группу попарно несовместных* событий.

Классическое определение вероятности впервые встречается у Я. Бернулли в книге «Искусство предположения» (1713). Более четкую формулировку можно найти в трудах А. Муавра, П. Лапласа, Ж. Лагранжа.

Следует уточнить, что такое определение содержит уже в себе самом понятие *равновероятности*, и в рамках классических (но наиболее понятных интуитивно) представлений от этого порочного круга избавиться не удается. Также недостатком такого подхода

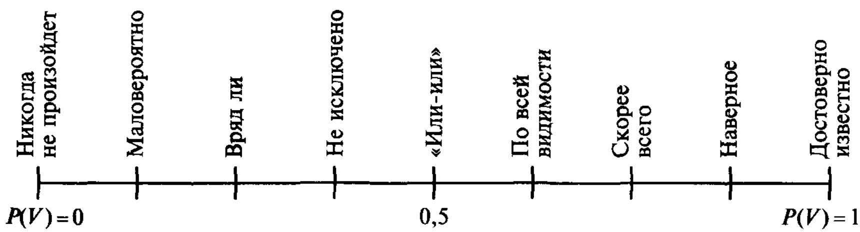


Рис. 1.9

служит тот очевидный факт, что классическая вероятность может быть лишь рациональным числом как отношение двух целых, хотя уже из определения геометрической вероятности следует, что она может являться иррациональным числом.

Вероятность условно можно изобразить в виде делений шкалы от 0 до 1 (рис. 1.9).

Если не удается посчитать m и n для вычисления вероятности из-за нарушения перечисленных условий, то можно воспользоваться или *геометрическим*, или *статистическим определением* вероятности.

Геометрическое определение вероятности связано с вероятностью попадания точки в некоторую область, являющуюся геометрическим объектом (множество точек отрезка, части плоскости или пространства). Пусть событие A представляет собой попадание точки в область $g \subset G$, а попадание в измеримую область G достоверно. Вычислим вероятность попадания произвольно взятой точки в некоторую область (событие A), которая пропорциональна *мере* этой части и не зависит от ее формы и расположения. Это может быть мера или длины, или площади, или объема и т.д., в зависимости от пространства элементарных событий. Обозначим это понятие меры через mes. Например, для подмножества плоскости $g \subset G \subset \mathbb{R}^2$ мера (mes) есть площадь (рис. 1.10).

Тогда ***геометрической вероятностью*** назовем отношение меры подмножества g к мере множества G:

$$P(A) = \frac{\operatorname{mes} g}{\operatorname{mes} G}. \tag{1.9}$$

Геометрическое определение вероятности используется при выполнении условий:

- на пространстве элементарных событий можно ввести меру;
- любые два элементарных события несовместны;
- вероятность события, заключающегося в попадании точки в бесконечно малую окрестность, зависит только от размеров этой окрестности.

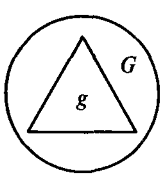


Рис. 1.10

Геометрическая вероятность — та же классическая вероятность, только для несчетного множества элементарных исходов. Соответственно априорная равновероятность записана в дифференциальной форме: $dP(g)/d\,\text{mes}\,g = \text{const}$.

Статистическое, или *частотное*, определение вероятности связано с понятием частоты событий. В отличие от классического определения вероятности для подсчета *относительной* частоты необходимо выполнить серии реальных испытаний.

Относительной частотой события A называется отношение числа m испытаний, при которых событие A произошло (или $m(A)$), к общему числу всех фактически произведенных испытаний n, при проведении которых событие A могло произойти или не произойти.

Обозначается относительная частота $W(A) = \frac{m}{n}$ или $P^*(A) = \frac{m(A)}{n}$.

При достаточно большом числе испытаний вероятность появления события A почти совпадает с частотой появления A, т.е. $P(A) \approx W(A)$, что отражается в законе больших чисел Я. Бернулли (см. подразд. 2.12). Массовые случайные события обладают свойством *статистической устойчивости частоты*: значения частоты появления события A, полученные в результате многократных испытаний, колеблются около некоторого постоянного числа, которое и является вероятностью этого события. Поэтому относительную частоту и называют *статистической вероятностью*. Отличие статистического определения вероятности от классического заключается в том, что по классическому определению вероятность вычисляется до проведения эксперимента, а согласно статистическому определению — по результату опытного исследования.

? Зачем так много определений дается для одного понятия «вероятность»? Обычно в точных науках, а в математике особенно, стараются ограничиться одним определением!

Дело в том, что каждое из этих определений «работает» только в своей области и имеет довольно узкий спектр применения (например, только для геометрических объектов или при статистических расчетах). Однако все эти определения, достаточно удобные для практического применения, можно заменить одним — аксиоматическим.

Для построения теории вероятностей как строгой (дедуктивной) науки необходимо познакомиться с аксиоматическим определением вероятности, которое было предложено А. Н. Колмогоровым в 30-е годы XX в. Неопределяемым понятием является произвольное множество Ω, называемое *пространством элементарных исходов*. На Ω построим множество всех его подмножеств F, которое назовем *пространством событий*, а его элементы — *случайными событиями*. На F определена алгебра событий, так как выполняются требования:

- $\varnothing \in F$, $\Omega \in F$;
- если $A \in F$ и $B \in F$, то $A + B \in F$, $AB \in F$, $\bar{A} \in F$.

Два события $A \in F$ и $B \in F$ называются попарно несовместными, если $AB = \varnothing$.

Определение вероятности сформулируем в виде *аксиом*:

1) каждому случайному событию A пространства событий F ставится в соответствие неотрицательное число $P(A) \geq 0$, называемое ***вероятностью*** (т. е. всюду на F определена числовая функция P: $F \to \mathbb{R}_+$);

2) вероятность достоверного события равна единице: $P(\Omega) = 1$;

3) если события A_1, A_2, ..., A_i попарно несовместны, то вероятность суммы попарно несовместных событий равна сумме вероятностей этих событий:

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i). \qquad (1.10)$$

Благодаря введению аксиоматического определения удалось представить теорию вероятностей в виде стройной математической науки.

Однако методы вычисления вероятностей на основе лишь одних определений не всегда эффективны, а порой и просто невозможны. На практике неизвестную вероятность часто находят косвенными методами: по известным вероятностям одних событий получают с помощью математических операций вероятности других с ними связанных событий. В основе таких методов поиска вероятностей лежат основные теоремы теории вероятностей, которые будут рассматриваться в следующем подразделе.

1.6. Некоторые теоремы теории вероятностей

В этом подразделе сформулируем и докажем некоторые свойства вероятности в виде теорем. Частным случаем аксиомы сложения является вероятность суммы несовместных событий.

Теорема 1.1. *Вероятность суммы несовместных событий* (рис. 1.11) *равна сумме вероятностей этих событий* (*ключевое слово* «***или***»):

$$P(A + B) = P(A) + P(B).$$

Доказательство. Пусть m — число равновозможных исходов, благоприятствующих событию A, несовместному с событием B; k —

число равновозможных исходов, благоприятствующих событию B. Обозначим через n общее число равновозможных элементарных исходов, образующих полную группу несовместных испытаний C. Пусть $P(A)=\dfrac{m}{n}$; $P(B)=\dfrac{k}{n}$. Так как по определению суммы $(A+B)$ имеют место случаи, когда осуществляется или событие A, или событие B, то $P(A+B)=\dfrac{m+k}{n}$.

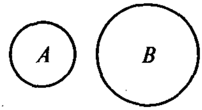


Рис. 1.11

Поэтому $P(A+B)=\dfrac{m+k}{n}=\dfrac{m}{n}+\dfrac{k}{n}=P(A)+P(B)$.

С геометрической точки зрения $\operatorname{mes} A=m$, $\operatorname{mes} B=k$, $\operatorname{mes} C=n$ (рис. 1.12), т.е. $P(A+B)=P(A)+P(B)$.

Теорема 1.2. *Сумма вероятностей полной группы событий равна единице*: $\sum\limits_{i=1}^{n} P(A_i)=1$.

Теорема 1.3. *Сумма вероятностей противоположных событий равна единице*:

$$P(A+\bar{A})=P(A)+P(\bar{A})=1.$$

Так как сумма противоположных событий есть событие достоверное, т.е. $A\bar{A}=\varnothing$, $A+\bar{A}=\Omega$ и $P(\Omega)=1$, вероятность события $\bar{A}$, *противоположного* событию A, равна

$$P(\bar{A})=1-P(A). \tag{1.11}$$

Очевидно, что вероятность произведения *противоположных* событий равна 0, т.е. $P(A\bar{A})=0$, так как это событие невозможное: $A\bar{A}=\varnothing$.

Теорема 1.4. *Если событие A влечет за собой событие B*, т.е. $A\subset B$, то $P(A)\le P(B)$.

Действительно, если $A\subset B$, то $P(A)+P(B\backslash A)=P(B)$ как вероятность суммы несовместных событий. Но разность $B\backslash A$ — тоже случайное событие, вероятность которого неотрицательна, $P(B\backslash A)\ge 0$; отсюда $P(A)\le P(B)$.

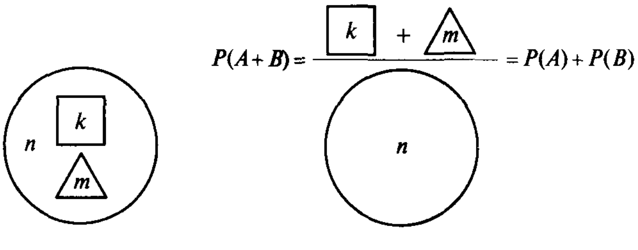


Рис. 1.12

Теорема 1.5. *Вероятность суммы двух совместных событий* (см. рис. 1.8, *а*, *б*) *вычисляется по формуле*:

$$P(A+B)=P(A)+P(B)-P(AB). \tag{1.12}$$

Доказательство. Известно, что так как события A и B совместные, то по определению сумма $A+B$ означает «или A, или B, или оба вместе», т.е. событие $A+B$ произойдет при наступлении одного из трех несовместных событий: либо $A\bar{B}$, либо $\bar{A}B$, либо AB. Тогда по теореме сложения вероятности несовместных событий имеем

$$P(A+B)=P(A\bar{B})+P(\bar{A}B)+P(AB).$$

Событие A произойдет при наступлении одного из двух несовместных событий $A\bar{B}$ или AB. Тогда по теореме сложения вероятностей несовместных событий имеем $P(A)=P(A\bar{B})+P(AB)$. Отсюда выразим вероятность $P(A\bar{B})=P(A)-P(AB)$.

Аналогично, событие B произойдет при наступлении одного из двух несовместных событий $\bar{A}B$ или AB. Тогда по теореме сложения вероятностей несовместных событий имеем $P(B)=P(\bar{A}B)+$ $+P(AB)$, откуда $P(\bar{A}B)=P(B)-P(AB)$.

Подставив полученные выражения в формулу вероятности суммы несовместных событий, имеем

$$P(A+B)=[P(A)-P(AB)]+[P(B)-P(AB)]+P(AB)=$$
$$=P(A)+P(B)-P(AB).$$

Проиллюстрируем это на примере. Известно, что $P(A)=m/n$; $P(B)=k/n$. Так как события A и B совместные, то среди $m+k$ благоприятных исходов содержатся l исходов, благоприятствующих и событию A, и событию B. Значит, $P(AB)=l/n$ (по определению вероятности произведения).

Так как по определению «$A+B$» означает или «A», или «B», или оба вместе, то благоприятствующих исходов при этом $(m+k-l)$. Поэтому

$$P(A+B)=\frac{m+k-l}{n}=\frac{m}{n}+\frac{k}{n}-\frac{l}{n}=P(A)+P(B)-P(AB).$$

С геометрической точки зрения ($\mathrm{mes}\,A=m$, $\mathrm{mes}\,B=k$, $\mathrm{mes}\,(AB)=$ $=l$, $\mathrm{mes}\,C=n$, рис. 1.13): $P(A+B)=P(A)+P(B)-P(AB)$.

Замечания. 1. Формула вероятности суммы несовместных событий является частным случаем полученной формулы, так как для несовместных событий вероятность их одновременного осуществления равна нулю, т.е. при $P(AB)=0$ формула (1.12) принимает вид: $P(A+B)=P(A)+P(B)$.

2. Для трех событий A, B и C справедлива формула:

$$P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC).$$

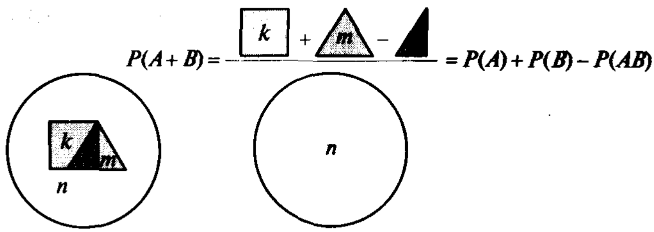


Рис. 1.13

3. Эти формулы полностью аналогичны формулам для суммы двух множеств, поэтому удобно рассматривать элементарный исход как точку в некоем пространстве событий Ω, а совокупность всех элементарных исходов, влекущих событие A, — подмножество $A \subset \Omega$. Тогда можно вообще не делать разницы между событием и соответствующим ему множеством.

Условная вероятность. Для событий A и B введем условную вероятность $P(A|B)$ (альтернативное обозначение $P_B(A)$) — вероятность того, что событие A произойдет при условии, что произошло событие B.

Условная вероятность того, что произойдет событие A, при условии, что событие B произошло, вычисляется по формуле

$$P(A|B) = \frac{P(AB)}{P(B)}, \text{ где } P(B) \neq 0. \tag{1.13}$$

Условная вероятность обладает следующими свойствами:

1) $P(\varnothing|B) = 0$; $P(\Omega|B) = 1$;

2) $P(A|A) = 1$;

3) если $A \subset B$, то $P(B|A) = 1$. Действительно, так как $A \subset B$, то $AB = A$ и $P(AB) = P(A)$. Поэтому $P(B|A) = \dfrac{P(A)}{P(A)} = 1$;

4) для несовместных событий $A_1A_2 = \varnothing$ справедливо $P_B(A_1 + A_2) = P_B(A_1) + P_B(A_2)$.

Теорема 1.6. *Вероятность совместного появления событий A и B равна произведению вероятности одного события на условную вероятность другого при условии, что первое событие произошло (ключевое слово «**и**»), и находится по формуле*

$$P(AB) = P(A|B)P(B) = P(B|A)P(A). \tag{1.14}$$

Событие A называется *независимым* от события B, если появление события B не изменит вероятность появления события A, т.е. $P(A|B) = P(A)$. Если A не зависит от B, то, согласно формуле (1.14), событие B не зависит от A, т.е. они *попарно* (*взаимно*) *независимы*. Тогда $P(A|B) = P(A|\overline{B})$.

Теорема 1.7. *Вероятность произведения независимых событий A и B равна произведению их вероятностей*:

$$P(AB) = P(A)P(B). \tag{1.15}$$

Полученное равенство может служить определением независимых событий. Если два события не являются независимыми, то они называются *зависимыми*. Заметим, что если событие A зависит от исхода события B, то $P(A|B) \neq P(A)$.

Пусть дано некоторое множество событий $A = \{A_1, ..., A_k\}$. Элементу $A_p \in A$ поставим в соответствие множество B_p, составленное из произвольных комбинаций (произведений) из числа оставшихся событий, где каждое представлено не более одного раза. Тогда события $A_1, ..., A_k$ называются *независимыми в совокупности*, если для любого $p \leq k$ и любого события $C \in B_p$ события A_p и C попарно независимы.

Следствие 1. Если события $A_1, ..., A_k$ независимы в совокупности, то любое их подмножество независимо в совокупности. Соответственно любая пара независима попарно.

Следствие 2. Если события $A_1, ..., A_k$ независимы в совокупности, то вероятность их совместного осуществления равна произведению их вероятностей, т.е.

$$P(A_1A_2...A_k) = P(A_1)P(A_2)...P(A_k).$$

Следствие 3. Вероятность совместного осуществления нескольких событий равна произведению вероятности одного из них на условную вероятность всех остальных, причем вероятность каждого последующего события вычисляется в предположении, что все предыдущие события уже произошли:

$$P(A_1A_2...A_n) = P(A_1)P_{A_1}(A_2)P_{A_1A_2}(A_3)\dots P_{A_1\dots A_{n-1}}(A_n).$$

Заметим, что порядок осуществления событий роли не играет.

Задача 1.17. В денежно-вещевой лотерее из 1000 билетов на 24 выпадают денежные выигрыши и на 10 — вещевые. Вы приобрели два билета. Какова вероятность выиграть *хотя бы* по одному билету?

Решение. Пусть A — выигрыш по первому билету, B — по второму билету. Так как выигрышных 24 + 10 = 34 билета, то $P(A) = 0{,}034$, а вероятность проигрыша по первому билету $P(\bar{A}) = 1 - P(A) = 0{,}966$. Чтобы найти вероятность выигрыша хотя бы по одному из двух билетов, от единицы вычитаем вероятность проигрыша сразу по двум билетам $P(\bar{A}\bar{B})$ (*и* по первому, *и* по второму). Так как приобретены два билета, то вероятность проигрыша по второму билету при условии, что первый проиграл, есть условная вероятность (так как теперь могут выиграть 34 билета из 999). Поэтому $P(B|\bar{A}) = \dfrac{34}{999}$, а $P(\bar{B}|\bar{A}) = \dfrac{965}{999}$.

Тогда $P = 1 - P(\bar{B}\bar{A}) = 1 - P(\bar{A})P(\bar{B}|\bar{A}) = 1 - \frac{966}{1000} \cdot \frac{965}{999} \approx 0,067.$

Задача 1.18. Для сигнализации об аварии установлены два независимо работающих сигнализатора. Вероятность того, что при аварии сработает первый сигнализатор, равна 0,95; второй — 0,9. Какова вероятность того, что при аварии сработает *только* один сигнализатор?

Решение. Обозначим буквой A событие «сработает только один сигнализатор», B — «сработает первый», C — «сработает второй сигнализатор». Тогда $P(B) = 0,95$; $P(\bar{B}) = 1 - P(B) = 1 - 0,95 = 0,05$; $P(C) = 0,9$; $P(\bar{C}) = 1 - P(C) = 1 - 0,9 = 0,1$.

Получить желаемый результат можно двумя способами: ***или*** сработает первый, а второй при этом не сработает, ***или*** сработает только второй, ***и*** при этом первый не сработает. Операции между событиями при этом имеют вид: $A = B\bar{C} + \bar{B}C$. Тогда вероятность того, что сработает только один сигнализатор:

$$P(A) = P(B)P(\bar{C}) + P(\bar{B})P(C) = 0,95 \cdot 0,1 + 0,9 \cdot 0,05 = 0,14.$$

Задача 1.19. Вероятность правильного оформления счета составляет 0,8. Во время аудиторской проверки были взяты для анализа два счета. Какова вероятность того, что среди взятых счетов будет правильно оформленный?

Решение. Пусть событие A — правильно оформлен хотя бы один счет, события B_1 и B_2 — правильно оформлены соответственно первый и второй счета. Известно, что $P(B_1) = P(B_2) = 0,8$; тогда $P(\bar{B}_1) = P(\bar{B}_2) = 0,2$. Так как могут быть правильно оформлены ***или*** первый, ***или*** второй, ***или*** оба счета вместе, то вероятность правильно оформить оба счета складывается из трех произведений:

$$P(A) = P(B_1)P(\bar{B}_2) + P(\bar{B}_1)P(B_2) + P(B_1)P(B_2) =$$
$$= 0,8 \cdot 0,2 + 0,2 \cdot 0,8 + 0,8 \cdot 0,8 = 0,96.$$

Итак, $P(A) = 0,96$.

Второй способ решения задачи: так как вероятность допустить ошибки в оформлении двух счетов равна вероятности произведения этих событий $P(\bar{B}_1\bar{B}_2) = P(\bar{B}_1)P(\bar{B}_2) = 0,04$, то вероятность противоположного события — правильно оформить оба счета равна $P(A) = 1 - 0,04 = 0,96$.

Эту задачу можно было решить и третьим способом — по правилу суммы совместных событий имеем:

$$P(A) = P(B_1) + P(B_2) - P(B_1)P(B_2) = 0,8 + 0,8 - 0,8 \cdot 0,8 = 0,96.$$

Задача 1.20. Прибор, работающий в течение суток, состоит из трех узлов, каждый из которых независимо от других может за это время выйти из строя. Неисправность любого из узлов приводит к прекращению работы прибора. Вероятность исправной работы в течение суток первого узла равна 0,9, второго — 0,85, третьего — 0,95.

а) С какой вероятностью прибор будет работать в течение суток безотказно?

б) Какова вероятность выхода из строя в течение суток хотя бы одного узла?

в) Какова вероятность выхода из строя в течение суток одного узла?

г) Какова вероятность того, что в течение суток не менее одного узла будут работать безотказно?

д) Какова вероятность того, что в течение суток не менее двух узлов выйдут из строя?

Решение. Обозначим через A, B и C события, соответствующие исправной работе в течение суток узлов I, II, и III соответственно, через D — событие, соответствующее безотказной работе прибора в течение суток.

Известно, что $P(A)=0,9$; $P(B)=0,85$; $P(C)=0,95$.

а) Так как события A, B и C независимые, то вероятность того, что ***и*** первый, ***и*** второй, ***и*** третий узелы в течение суток будут работать безотказно, находится с помощью теоремы о произведении независимых событий (1.15):

$$P(D)=P(ABC)=P(A)P(B)P(C)=0,9\cdot 0,85\cdot 0,95=0,72675\approx 0,727.$$

б) События $\bar{D}$ — «выход из строя в течение суток хотя бы одного узла» и D — «безотказная работа в течение суток» — противоположные.

Тогда $P(\bar{D})=1-0,727=0,273$.

в) Вероятности выхода из строя I, II, и III узлов соответственно равны $P(\bar{A})=1-0,9=0,1$; $P(\bar{B})=1-0,85=0,15$; $P(\bar{C})=1-0,85=$ $=0,05$. Тогда выход из строя в течение суток одного узла — событие $E=AB\bar{C}+\bar{A}BC+A\bar{B}C$, состоящее из суммы трех несовместных событий (один из узлов отказал, а остальные два работают), каждое из которых есть произведение независимых событий (работа одного из узлов).

Тогда $P(E)=0,9\cdot 0,85\cdot 0,05+0,1\cdot 0,85\cdot 0,95+0,9\cdot 0,15\cdot 0,95=$ $=0,246$.

г) Событие K — «не менее одного узла будут работать безотказно» противоположно событию $\bar{K}$ — «менее одного узла будут работать безотказно», т.е. выходу из строя всех трех узлов одновременно. Поэтому находим его вероятность по формуле

$$P(K)=1-P(\bar{K})=1-P(\bar{A})P(\bar{B})P(\bar{C})=1-0,1\cdot 0,15\cdot 0,05=0,9992.$$

На основании полученных данных можно сделать вывод о том, что событие «не менее одного узла будут работать безотказно» «практически достоверное».

д) Событие M — «не менее двух узлов выйдут из строя» противоположно событию $\bar{M}$ — «менее двух узлов выйдут из строя», т.е. или один окажется неисправным (E), или менее одного узла будут неисправны, т.е. все работают безотказно (D). Поэтому надо либо непосредственно найти вероятность события M — выход из строя двух (H) или трех ($\bar{K}$) узлов, либо находить вероятность противоположного события — «менее двух узлов выйдут из строя».

I вариант: $P(M) = P(H) + P(\bar{K})$, где событие

$$H = \bar{A}\bar{B}C + \bar{A}B\bar{C} + A\bar{B}\bar{C}.$$

II вариант: $P(M) = 1 - P(\bar{M}) = 1 - (P(E) + P(D)) = 1 - (0{,}246 + 0{,}727) = 0{,}027$.

Для решения задачи выбран второй вариант как более рациональный.

Задача 1.21. Найти вероятность того, что наудачу взятое двузначное число будет кратно или трем, или пяти.

Решение. Обозначим через A событие «число делится на пять ($a \vdots 5$)», через B — «число делится на три ($a \vdots 3$)». Тогда, чтобы число делилось на пятнадцать, надо чтобы оно делилось и на три, и на пять, т.е. по правилу произведения это $P(AB)$. Так как события A и B совместные, воспользуемся теоремой сложения для совместных событий $P(A + B) = P(A) + P(B) - P(AB)$. Рассмотрим множество двузначных чисел от 10 до 99. Их общее количество $n = 90$ $(99 - 9 = 90)$. Двузначных чисел, кратных пяти, всего $90 : 5 = 18$ $(m_A = 18)$. По признаку делимости на пять это числа, оканчивающиеся на 0 (их 9 от 10 до 90) и на 5 (их 9 от 15 до 95). Так как $m_A = 18$, то $P(A) = \dfrac{m_A}{n} = \dfrac{18}{90} = \dfrac{1}{5}$.

Количество чисел, кратных трем, можно сосчитать, учитывая, что на 3 делится каждое третье число от 10 до 90 , т.е. $m = n : 3 = 90 : 3 = 30$.

Тогда $m_B = 30$, а $P(B) = \dfrac{m_B}{n} = \dfrac{30}{90} = \dfrac{1}{3}$. Так как события A и B независимые, то $P(AB) = P(A)P(B) = \dfrac{1}{5} \cdot \dfrac{1}{3} = \dfrac{1}{15}$; поэтому $P(A + B) = P(A) + P(B) - P(AB) = \dfrac{1}{5} + \dfrac{1}{3} - \dfrac{1}{15} = \dfrac{7}{15}$.

Читателя не должен смущать факт того, что пришлось выполнять лишние на первый взгляд действия, сначала разделив 90 на 3, а потом на 90, и получать все те же 1/3. Эта кажущаяся нерациональность — следствие совпадения, что общее количество чисел ($N = 90$), идущих подряд, делится и на 3, и на 5 (и, как следствие, на 15). Предлагаем читателю самому убедиться в том, что если N не делится хотя бы на одно из чисел 3 или 5, то ответ будет не равен $1/3 + 1/5 - 1/15$. Тогда предложенный алгоритм прямого подсчета чисел будет единственно правильным.

Задача 1.22. Вероятность выхода прибора из строя в течение суток равна α, где α — малая положительная величина. Какова вероятность того, что в течение рабочей недели прибор будет работать безотказно?

Решение. Обозначим через A — выход прибора из строя в течение суток. Значит, $P(A) = \alpha$. Пусть B — выход прибора из строя в течение недели. Тогда вероятность работы прибора в течение суток $P(\bar{A}) = 1 - \alpha$ (выход из строя и работа прибора — события противоположные). Вероятность того, что всю рабочую неделю (6 дней) прибор работает безотказно, найдем, учитывая, что и в первый, и во второй, и ..., и в шестой день он работал с вероятностью $(1 - \alpha)$, тогда $P(\bar{B}) = (1 - \alpha)^6$. Так как α малая величина и ее степенями можно пренебречь, по формуле приближенных вычислений бинома найдем $P(\bar{B}) = (1 - \alpha)^6 \approx 1 - 6\alpha$. Например, при $\alpha = 0{,}03$, $P(\bar{A}) \approx 1 - 6 \cdot 0{,}03 = 1 - 0{,}18 = 0{,}82$.

Задача 1.23. В первом ряду сидят шесть юношей и четыре девушки. С этого ряда вызывают к доске подряд двух студентов. Какова вероятность вызвать подряд двух юношей?

Решение. Пусть A — вызов к доске юноши в первом случае, B — вызов к доске юноши во втором случае. Тогда $P(A) = \frac{6}{10} = \frac{3}{5}$; $P(B|A) = \frac{5}{9}$ (условная вероятность: после вызова к доске первого юноши их осталось пять из девяти студентов).

Поскольку к доске и в первом, и во втором случаях были вызваны юноши, применим теорему вероятности произведения зависимых событий: $P(AB) = P(A)P(B|A) = \frac{3}{5} \cdot \frac{5}{9} = \frac{1}{3}$.

Задача 1.24. Брошены две игральные кости (кубики). Найти вероятность того, что:

1) сумма выпавших очков равна 6;
2) сумма выпавших очков не более 10;
3) сумма выпавших очков равна 11, а разность — 2;
4) сумма равна 7, а разность — 5.

Решение. 1) Обозначим событие «сумма выпавших очков равна 6» через A. По правилу произведения число равновозможных исходов при бросании двух костей (и в первом, и во втором случаях 6 вариантов) равно $n = 6 \cdot 6 = 36$. Из них число благоприятных исходов — при $m = 5$. Это варианты $5 + 1$, $1 + 5$, $2 + 4$, $4 + 2$, $3 + 3$, которые можно записать с помощью элементарных событий: (ω_5, ω_1), (ω_1, ω_5), (ω_2, ω_4), (ω_4, ω_2), (ω_3, ω_3).

Тогда $P(A) = \frac{m}{n} = \frac{5}{36}$.

2) Обозначим событие «сумма выпавших очков не более 10» через B. Заметим, что число благоприятных вариантов очень большое, а неблагоприятных вариантов (событие $\bar{B}$) только три: $6 + 5$; $5 + 6$; $6 + 6$ или (ω_6, ω_5), (ω_5, ω_6), (ω_6, ω_6).

Поэтому $P(\bar{B}) = \frac{3}{36} = \frac{1}{12}$. Тогда $P(B) = 1 - P(\bar{B}) = 1 - \frac{1}{12} = \frac{11}{12}$.

3) Обозначим событие «сумма выпавших очков равна 11» через C, а «разность этих очков равна 2» — D. Но, если сумма двух целых чисел равна 11, то они разной четности, а если разность равна 2, то одинаковой. Поэтому эти события являются несовместными и вероятность произведения $P(CD) = 0$. (Это эквивалентно тому, что система $\begin{cases} a + b = 11; \\ |a - b| = 2 \end{cases}$ не имеет натуральных корней.)

4) Обозначим событие «сумма выпавших очков равна 7» через N, а «разность равна 5» — M. Число 7 может быть получено в шести благоприятных исходах: $1 + 6$; $6 + 1$; $2 + 5$; $5 + 2$; $3 + 4$; $4 + 3$, или (ω_1, ω_6), (ω_6, ω_1), (ω_2, ω_5), (ω_5, ω_2), (ω_3, ω_4), (ω_4, ω_3). Значит, $m = 6$, а $P(N) = \frac{6}{36} = \frac{1}{6}$; из них только два варианта 1 и 6, 6 и 1 имеют разность, равную 5. Поэтому условная вероятность $P(M|N) = \frac{2}{6} = \frac{1}{3}$. А вероятность произведения событий (и сумма равна 7, и разность равна 5) находится с помощью теоремы умножения зависимых событий: $P(AB) = P(N)P(M|N) = \frac{1}{6} \cdot \frac{1}{3} = \frac{1}{18}$.

Эту задачу легче решить в обратном порядке: сначала найти решение системы ($a + b = 7$, $|a - b| = 5$), а потом подсчитать соответствующую вероятность. Очевидно, система имеет два решения: $a = 6$, $b = 1$ и $a = 1$, $b = 6$, т.е. из 36 возможных вариантов два благоприятствуют этому событию. Тогда $P(AB) = 2/36 = 1/18$.

Задача 1.25. В лотерее разыгрываются 10000 билетов; из них на 10 билетов падает выигрыш по 200 тыс. р.; на 100 билетов — по 100 тыс. р.; на 500 билетов — по 25 тыс. р.; на 1000 билетов — по 5 тыс. р.

Какова вероятность выиграть не менее 25 тыс. р. по одному билету?

Решение. Введем обозначения событий:

A — выиграть не менее 25 тыс. р.;

B — выиграть 200 тыс. р.: $P(B) = \frac{10}{10000} = 0,001$;

C — выиграть 100 тыс. р.: $P(C) = \frac{100}{10000} = 0,01$;

D — выиграть 25 тыс. р.: $P(D) = \dfrac{500}{10000} = 0,05.$

Так как не менее 25 тыс. р. можно получить или в первом, или во втором, или в третьем случаях, то по теореме сложения вероятностей: $P(A) = P(B + C + D) = P(B) + P(C) + P(D) = 0,001 + 0,01 + 0,05 = 0,061$.

Задача 1.26. На военных учениях летчик получил задание «уничтожить три рядом расположенных склада». На борту самолета три бомбы, по одной для каждой цели. Вероятность попадания в первый склад 0,01, во второй — 0,08, в третий — 0,025. Любое попадание в результате детонации уничтожает все склады, самолет сверхзвуковой, и летчик не знает, были ли удачными его первые попытки. Какова вероятность того, что склады будут уничтожены?

Решение. Введем обозначения событий:

B — попадание в первый склад;

C — попадание во второй склад;

D — попадание в третий склад.

Тогда:

$$P(B) = 0,01;\ P(C) = 0,08;\ P(D) = 0,025 \text{ (по условию).}$$

Событие A — уничтожены или первый, или второй, или третий склад, значит, так как события B, C, D несовместные,

$$P(A) = P(B) + P(C) + P(D) = 0,01 + 0,08 + 0,025 = 0,115.$$

Задача 1.27. Какова вероятность того, что при бросании трех монет выпадает:

а) два орла;

б) не более двух орлов;

в) ни одного орла;

г) более трех орлов;

д) не более трех орлов?

Решение. Введем обозначения событий:

• A, B, C — выпадание орла в первом, втором и третьем испытаниях соответственно.

Тогда вероятности этих событий $P(A) = P(B) = P(C) = 1/2$ и $P(\bar{A}) = P(\bar{B}) = P(\bar{C}) = 1/2$;

• D — выпадание двух орлов по схеме $AB\bar{C}$, т. е. орел не выпал только в третьем испытании. Тогда $P(D) = P(A)P(B)P(\bar{C}) = 1/8$, так как события «выпадание орла» независимы друг от друга. Но два орла при трех подбрасываниях монет могут также выпасть в случаях: ***или*** если не выпадет орел только при втором подбрасывании (схема $A\bar{B}C$ — событие E), ***или*** только при первом подбрасывании (схема $\bar{A}BC$ — событие F). Так как события D, E, F несовместны, полная вероятность получается при сложении всех вариантов. Поэтому $P = 1/8 + 1/8 + 1/8 = 3/8$. Более подробно это объяснено в формуле Бернулли (см. подразд. 1.10);

• M — выпало не более двух орлов, значит, ***или*** 0, ***или*** 1, ***или*** 2 орлов,

противоположное событие $\bar{M}$ — более двух орлов, значит, три орла:

$$P(\bar{M}) = P(A)P(B)P(\bar{C}) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}, \text{ тогда } P(M) = 1 - \frac{1}{8} = \frac{7}{8};$$

• N — не выпало ни одного орла, значит, выпали все решки: схема $N = \bar{A}\bar{B}\bar{C}$; тогда $P(N) = 1/8$;

• O — выпало более трех орлов — невозможное событие, тогда $P(O) = 0$;

• Q — выпало не более трех орлов: $Q = (\bar{O})$. Тогда $P(Q) = 1 - P(O) = 1$, где Q — достоверное событие.

Задача 1.28. Из колоды в 32 карты наугад одну за другой берут две карты. Найти вероятность того, что:

а) взяты два короля;

б) взяты две карты пиковой масти;

в) взяты король и дама.

Решение. Введем обозначения событий:

A — взят первый король;

B — взят второй король;

C — взята одна пиковая карта;

D — взята вторая пиковая карта;

E — взята первая дама;

F — взята вторая дама.

Тогда

1) $P(A) = \dfrac{4}{32}$; $P(B|A) = \dfrac{3}{31}$; $P(M) = P(A)P(B|A) = \dfrac{4}{32}\cdot\dfrac{3}{31} = \dfrac{12}{992}$;

2) $P(C) = \dfrac{8}{32}$; $P(D|C) = \dfrac{7}{31}$; отсюда $P(C)P(D|C) = \dfrac{8}{32}\cdot\dfrac{7}{31} = \dfrac{56}{992} = \dfrac{7}{124}$;

3) $P(E|B) = \dfrac{47}{31}$; $P(B)P(E|B) = \dfrac{4}{32}\cdot\dfrac{4}{31} = \dfrac{16}{992} = \dfrac{1}{62}$.

Аналогично $P(A)P(F|A) = 1/62$. Суммируя, получаем $P = 1/31$.

Задача 1.29. В городе N число ясных дней в июле равно 25. Найти вероятность того, что:

а) первые два дня июля будут ясными;

б) первые три дня июля будут ясными.

Решение. Введем обозначения событий:

A_1 — первые два дня июля ясные;

A_2 — первые три дня июля ясные;

B — первый день июля ясный, тогда $P(B) = \dfrac{25}{31}$;

C — второй день июля ясный, тогда $P(C|B)=\dfrac{24}{30}$;

D — третий день июля ясный, тогда $P(D|BC)=\dfrac{23}{29}$.

Имеем: $P(A_1)=P(B)P(C|B)=\dfrac{25}{31}\cdot\dfrac{24}{30}=\dfrac{600}{930}=\dfrac{20}{31}$, т.е. $P(A_1)=\dfrac{20}{31}$;

$P(A_2)=P(BCD)P(B)P(C|B)P(D)=\dfrac{25}{31}\cdot\dfrac{24}{30}\cdot\dfrac{23}{29}=\dfrac{460}{899}$.

Задача 1.30. Вероятность выхода из строя станка в течение одного рабочего дня равна α (где α — малое число, квадратом которого можно пренебречь). Какова вероятность того, что за пять дней станок ни разу не выйдет из строя?

Решение. Введем обозначение событий:

A — выход из строя станка в течение одного рабочего дня;

B — выход из строя станка в течение пяти рабочих дней.

Тогда $P(A)=\alpha$, $P(\bar{A})=1-\alpha$ — вероятность того, что станок не выйдет из строя в течение одного дня. По правилу умножения вероятностей найдем вероятность того, что он не выйдет из строя в течение пяти дней (***и*** в первый, ***и*** во второй, ..., ***и*** в пятый). Возведя $P(\bar{A})=1-\alpha$ в пятую степень, по известной приближенной формуле при $\alpha\approx 0$ имеем $P(B)=(1-\alpha)^5\approx 1-5\alpha$.

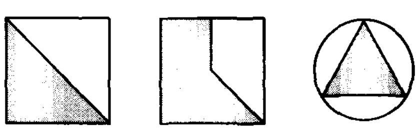
В задачах 1.31—1.35 воспользуемся геометрическим определением вероятности.

Задача 1.31. Найти вероятность попадания в паутину бабочки, оказавшейся в колодце различной формы (рис. 1.14).

Решение. Пусть на рисунках в каждом из колодцев паутина будет представлена в виде затемненного многоугольника. Обозначим события «попадание в паутину бабочки»: A — в первом случае (рис. 1.14, *а*); B — во втором случае (рис. 1.14, *б*); C — в третьем случае (рис. 1.14, *в*). Тогда

$$P(A)=\frac{1}{2};\ P(B)=\frac{5}{8};\ P(C)\approx\frac{S_\Delta}{S_O}=\frac{3\sqrt{3}r^2/4}{\pi r^2}=\frac{3\sqrt{3}}{4\pi}.$$

Задача 1.32. Внутри эллипса $\dfrac{x^2}{25}+\dfrac{y^2}{16}=1$ расположен круг с уравнением $x^2+y^2=9$ (рис. 1.15). Найти вероятность попадания



а *б* *в*

Рис. 1.14

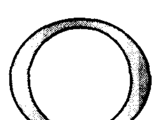


Рис. 1.15

точки в кольцо, ограниченное эллипсом и кругом, если за пределы эллипса точка не попадает.

Решение. Обозначим событие «попадание точки в кольцо» через A. Тогда

$$P(A)=\frac{S_{\text{кольца}}}{S_{\text{эл}}}=\frac{S_{\text{эл}}-S_{\text{кр}}}{S_{\text{эл}}}=\frac{5\cdot 4\pi-3^2\pi}{5\cdot 4\pi}=\frac{11}{20}.$$

Задача 1.33. «*Задача о встрече*». Молодые люди договорились встретиться с 12.00 до 13.00 в кафе, но не смогли назвать точное время своего прихода. Однако каждый пообещал ждать другого в течение 15 мин. Какова вероятность их встречи?

Решение. Обозначим через x — момент прихода первого участника рандеву, через y — время прихода второго, а их встречу — как событие A. Условие встречи запишем в виде неравенства: $|x-y|\le\frac{1}{4}$ (ч.).

В этой задаче пространство элементарных событий не может быть представлено в виде какого-то дискретного множества с помощью перечисления элементарных исходов, его составляющих. Удобнее изобразить это неравенство геометрически в прямоугольной системе координат, где за единичный отрезок выбран 1 ч. На рис. 1.16 в этом единичном квадрате заштриховано множество точек, соответствующих их возможной встрече, т.е. точки, расположенные между параллельными прямыми $x-y=\frac{1}{4}$ и $x-y=-\frac{1}{4}$. Интуитивно понятно, что момент прихода каждого участника равномерно распределен на единичном отрезке $[0, 1]$, а точка их прихода равномерно распределена в единичном квадрате. Тогда их возможные встречи соответствуют площади заштрихованной части квадрата ($S_{\text{встр}}$).

Поэтому вероятность их встречи найдем с помощью геометрического определения вероятности как отношение этих площадей $S_{\text{кв}}=1$ и $S_{\text{встр}}=1-\left(\frac{3}{4}\right)^2=\frac{7}{16}$. Отсюда

$$P(A)=\frac{S_{\text{встр}}}{S_{\text{кв}}}=\frac{7}{16}.$$

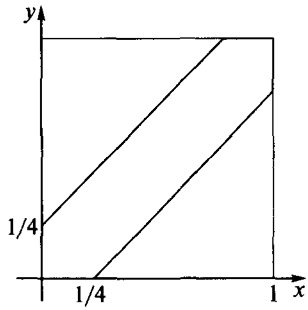


Рис. 1.16

? Первый пришедший ждет партнера уже более 10 мин. Какова вероятность того, что за оставшееся время встреча произойдет?

Задача 1.34. Вероятности появления независимых событий $A_1, A_2, \ldots, A_n$ равны соответственно $p_1, p_2, ..., p_n$. Найти вероятность появления *хотя бы одного* из событий $A_1, A_2, ..., A_n$.

Решение. 1) Пусть событие A происходит, если наступит хотя бы одно из событий A_i, где $i = \overline{1,n}$. Тогда вероятности непоявления событий $A_1, A_2, ..., A_n$, которые обозначим $\bar{A}_1, \bar{A}_2, ..., \bar{A}_n$ соответственно, равны $q_1 = 1 - p_1, ..., q_n = 1 - p_n$. Очевидно, $\bar{A} = \bar{A}_1 \bar{A}_2 ... \bar{A}_n$. Так как события $\bar{A}_1, \bar{A}_2, ..., \bar{A}_n$ независимы, имеем

$$P(A) = 1 - P(\bar{A}) = 1 - P(\bar{A}_1\ \bar{A}_2 ... \bar{A}_n) = 1 - q_1\ q_2\ ...\ q_n.$$

2) В случае, если события $A_1, A_2, ..., A_n$ имеют одинаковую вероятность $P(A_1) = P(A_2) = ... = P(A_n) = p$, то вероятность появления хотя бы одного события вычисляется по формуле $P(A) = 1 - q^n$.

Задача 1.35. На «проблемном» перекрестке в праздничные дни дорожно-транспортные происшествия (ДТП) совершаются с вероятностью 0,2. Какова вероятность того, что на этом перекрестке за три праздничных новогодних дня произойдет хотя бы одно ДТП?

Решение. Обозначим через A — событие «за три праздничных новогодних дня произойдет хотя бы одно ДТП». Пусть событие «произойдет одно ДТП» имеет вероятность $p = 0{,}2$, тогда противоположное событие «в праздничный новогодний день не произойдет ни одного ДТП» имеет вероятность $1 - p = q = 1 - 0{,}2 = 0{,}8$.

Тогда вероятность того, что за три праздничных новогодних дня произойдет хотя бы одно ДТП, равна

$$P(A) = 1 - P(\bar{A}) = 1 - (0{,}8)^3 = 1 - 0{,}512 = 0{,}488.$$

? В теории вероятностей вероятности элементарных событий всегда заранее известны?

Действительно, во всех рассмотренных задачах численные значения вероятности элементарных событий или были заранее известны, или вычислялись благодаря применению классического и геометрического определений вероятности. Но во многих задачах число благоприятных и равновозможных случаев приходится специально подсчитывать. В таких расчетах на помощь приходит комбинаторика.

1.7. Применение комбинаторики для подсчета вероятностей

В этом подразделе будем использовать комбинаторику для подсчета различных вариантов группировки множеств.

Задачи по определению вероятности легко узнать по постановке вопроса: «С какой вероятностью?», «Какова вероятность?»

и т.д. Иногда при подсчете числа благоприятных или равновозможных исходов применяют формулы комбинаторики.

Задача 1.36. В ящике 10 одинаковых деталей, помеченных номерами от 1 до 10. Наудачу берут 6 деталей. Найти вероятность того, что среди извлеченных деталей: а) будет деталь № 5; б) будут детали № 5 и 9.

Решение. Обозначим события:

A — «среди извлеченных деталей будет деталь № 5»; B — «среди извлеченных деталей будут детали № 5 и 9».

а) $P(A) = m/n$, где $n = C_{10}^6 = \dfrac{10!}{6!4!}$, а $m = C_9^5 = \dfrac{9!}{5!4!}$.

Чтобы найти число благоприятных исходов, откладываем нужную деталь (№ 5), так как от нее теперь не зависит общее число благоприятных исходов. Тогда имеем

$$P(A) = \frac{C_9^5}{C_{10}^6} = \frac{9!}{5!4!} : \frac{10!}{6!4!} = \frac{6}{10} = 0{,}6.$$

б) $P(B) = \dfrac{m}{n}$, где $m = C_8^4 = \dfrac{8!}{4!4!}$. Отложив две детали (№ 5 и 9), для подсчета числа благоприятных испытаний исключаем их из общего количества:

$$P(A) = \frac{C_8^4}{C_{10}^6} = \frac{8!}{4!4!} : \frac{10!}{6!4!} = \frac{5 \cdot 6}{9 \cdot 10} = \frac{1}{3}.$$

Задача 1.37. Какова вероятность того, что при десятикратном подбрасывании монеты орел выпадет три раза?

Решение. Пусть событие A — орел выпадет три раза. Найдем неизвестную вероятность с помощью классического определения $P(A) = m/n$. Определим неизвестные значения m и n.

Так как число испытаний 10 и в каждом из них было два варианта исхода (орел и решка), по правилу произведения, или по формуле размещений с повторениями n, число равновозможных исходов равно $n = 2^{10}$.

Тогда m — число благоприятных исходов, когда из 10 опытов в трех будет ожидаемый орел. Так как порядок появления орлов и решек не важен, используем формулу сочетаний $m = C_{10}^3$.

Итак, $P(A) = \dfrac{C_{10}^3}{2^{10}} = \dfrac{15}{128}$.

Задача 1.38. Какова вероятность сесть за праздничный (круглый) стол рядом с другом, если за столом будет 10 человек?

Решение. $P(A) = m/n$, где событие A — осуществление желания сесть рядом с другом.

Тогда n — число равновозможных событий, не зависящее от порядка рассаживания гостей за столом, находим по формуле перестановок из 10 элементов: $n = P_{10} = 10!$; m — число благоприятных исходов. Так как рядом с другом можно сесть за стол $2 \cdot 10 = 20$ способами (10 вариантов, когда друг слева, и 10 вариантов, когда друг справа), а остальные 8 человек могут сесть $P_8 = 8!$ способами, то по правилу произведения $m = 20 \cdot 8!$, тогда $P(A) = \dfrac{20 \cdot 8!}{10!} = \dfrac{2}{9}$.

Более простое решение: друг сядет или слева, или справа (благоприятствующие события), или на одно из оставшихся семи мест: $P(A) = 2/(2+7) = 2/9$.

Задача 1.39. Среди 15 инвестиционных фондов пять — «пирамиды». Какова вероятность того, что, приобретая наудачу по одной акции трех фондов, инвестор вложит все деньги в «пирамиды»?

Решение. $P(A) = m/n$, где событие A — приобретение трех акций инвестиционных фондов.

Тогда m — число «благоприятных» исходов, когда три акции приобретены у пяти фондов-«пирамид», причем в подмножестве m порядок приобретения акций не важен. Поэтому выбираем формулу сочетаний без повторений: $m = C_5^3$; n — равные возможности приобрести три акции у любого из 15 инвестиционных фондов, поэтому $n = C_{15}^3$. Тогда

$$P(A) = \frac{C_5^3}{C_{15}^3} = \frac{5!}{3!2!} : \frac{15!}{3!12!} = \frac{5!3!12!}{3!2!15!} = \frac{3 \cdot 4 \cdot 5}{13 \cdot 14 \cdot 15} = \frac{2}{91}.$$

Задача 1.40. Забыты три последние цифры телефонного номера друга. Какова вероятность набрать нужный номер, если сохранилась в памяти следующая информация:

а) цифры различные;

б) цифры повторяются;

в) все три цифры одинаковые?

Решение. $P(A) = m/n$, где событие A_i — набрать нужный номер в i-м задании.

Имеем m — благоприятный исход единственный — когда у друга зазвонит телефон, т.е. $m = 1$:

а) n — количество способов набрать забытые три цифры из 10 имеющихся на телефонном аппарате, зависит от порядка (так как переставленные цифры дадут разные номера). Тогда выбираем формулу размещений без повторений: $n = A_{10}^3 = \dfrac{10!}{(10-3)!} = \dfrac{10!}{7!} = 720$. Поэтому $P(A_1) = \dfrac{1}{720}$;

б) если среди трех забытых цифр могут встретиться одинаковые, то используем формулу размещений с повторениями $n = \bar{A}_{10}^3 = 10^3 = 1\,000$. Тогда $P(A_2) = \dfrac{1}{1000}$;

в) если в телефонном номере три последние цифры одинаковые, то таких вариантов всего 10 (тройки одинаковых цифр от 0 до 9). Отсюда $P(A_3) = \frac{1}{10}$.

Задача 1.41. Шахматную секцию посещают семь студентов первого курса, пять студентов второго курса и шесть студентов третьего курса. Какова вероятность того, что в финальной игре на первенство колледжа по шахматам противниками будут однокурсники?

Решение. Обозначим события:

A — играют однокурсники;

B_i — играет студент i-го курса.

Так как события B_1, B_2 и B_3 несовместны (играют только однокурсники), по теореме о сумме несовместных событий имеем:

$$P(A) = P(B_1) + P(B_2) + P(B_3).$$

Согласно классическому определению вероятности, подсчитаем число m благоприятных исходов и число n равновозможных, единственно возможных и несовместных исходов, образующих полную группу событий. Так как два шахматиста в финальной игре — однокурсники, причем порядок их выбора не имеет значения (т.е. не важно, кто из них играет белыми и кто черными фигурами), то используем формулу сочетаний. Имеем число благоприятных исходов $m_1 = C_7^2$, $m_2 = C_5^2$ и $m_3 = C_6^2$ и число равновозможных исходов для общего числа шахматистов $(7 + 5 + 6 = 18)$ $n = C_{18}^2$. Тогда вероятность каждого из событий B_i равна:

$$P(B_1) = \frac{C_7^2}{C_{18}^2};\ P(B_2) = \frac{C_5^2}{C_{18}^2}\ \text{и}\ P(B_3) = \frac{C_6^2}{C_{18}^2},$$

откуда

$$P(A) = \frac{C_7^2}{C_{18}^2} + \frac{C_5^2}{C_{18}^2} + \frac{C_6^2}{C_{18}^2} = \frac{C_7^2 + C_5^2 + C_6^2}{C_{18}^2} = \frac{21 + 10 + 15}{153} = \frac{46}{153} \approx 0,3.$$

К этому же ответу приводят и другие рассуждения.

Число благоприятных исходов m можно получить, если два шахматиста ***или*** с первого курса, ***или*** со второго курса, ***или*** с третьего курса. Так как события B_1, B_2 и B_3 несовместны, а порядок в подмножестве игроков не важен, то $m = C_7^2 + C_5^2 + C_6^2$, откуда

$$P(A) = \frac{C_7^2 + C_5^2 + C_6^2}{C_{18}^2} = \frac{21 + 10 + 15}{153} = \frac{46}{153} \approx 0,3.$$

Задача 1.42. На книжной полке произвольно расставлены девять книг, среди которых четырехтомник М.Ю.Лермонтова. С какой вероятностью все тома поэта будут стоять рядом?

Решение. Пусть событие A — вероятность, что тома Лермонтова будут стоять рядом. Число равновозможных исходов равно числу перестановок во всем множестве, состоящем из девять книг, т.е. $P_9 = 9!$ Число благоприятных исходов равно числу перестановок в подмножестве, состоящем из четырех книг Лермонтова, т.е. $P_4 = 4!$, а также числу перестановок во множестве, состоящем из оставшихся пяти книг и «связки» томов Лермонтова, т.е. $P_6 = 6!$ Тогда число благоприятных исходов по правилу произведения равно $P_4 \cdot P_6$, а вероятность события A равна $P(A) = \frac{P_4 \cdot P_6}{P_9} = \frac{4!6!}{9!} = \frac{246!}{6!7 \cdot 8 \cdot 9} = \frac{1}{21}$.

К числу более сложных задач на подсчет вероятности с помощью формул комбинаторики относятся задачи на *гипергеометрические распределения* (см. подразд. 2.2).

В общем виде формулировка задачи может выглядеть следующим образом.

Задача 1.43. В ящике N деталей, среди которых M бракованных (рис. 1.17). Какова вероятность того, что среди n наудачу вынутых деталей окажется m бракованных?

Решение. Пусть событие A — среди n наудачу вынутых деталей окажется m бракованных. В общем количестве N деталей, если M — число бракованных, то $N - M$ — число небракованных деталей. Если из взятых n деталей m — брак, то остальные $n - m$ деталей — не брак. Каждая стрелка на рис. 1.6 вида $b \to a$ символизирует собой число C_a^b способов выбора. Нужно лишь определить, какие события благоприятствуют нужному набору бракованных деталей, а также общее число равновозможных вариантов.

Число равновозможных испытаний не зависит от порядка выбора любых n из имеющихся N деталей и вычисляется по формуле сочетаний C_N^n, так как порядок в подмножестве не важен. Для подсчета числа благоприятствующих исходов нужно выбрать m бракованных деталей из M — общего числа бракованных, что можно сделать C_M^m способами, а также остальные $n - m$ деталей выбирать из небракованных, которых всего было $N - M$. Это можно сделать C_{N-M}^{n-m} способами. А так как надо взять ***и*** те ***и*** другие детали, то по правилу произведения всего будет $C_M^m C_{N-M}^{n-m}$ благоприятствующих исходов. Тогда вероятность того, что среди n наудачу вынутых деталей окажется m бракованных, вычисляется по формуле: $P(A) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$.

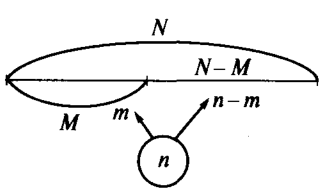


Рис. 1.17

Задача 1.44. На улицах одного из районов города разместили 100 телефонов-автоматов. Найти вероятность того, что среди наудачу проверенных четырех телефонов-автоматов из 100 развешанных три будут с дефектами, если стало известно, что в партии телефонов-автоматов было 10 дефектных.

Решение. Обозначим через A — событие «среди четырех телефонов-автоматов из 100 развешанных три — с дефектами». Применив формулу из предыдущей задачи, имеем: $P(A) = \dfrac{C_{10}^3 C_{90}^1}{C_{100}^4}$.

Задача 1.45. Какова вероятность выиграть по двум билетам лотереи, если приобретено 10 билетов и известно, что среди них 50 % выигрышных?

Решение. Обозначим через A — событие «выиграть по двум билетам», тогда $P(A) = m/n$. Найдем m и n, зная, что 50 % от 10 составляют 5 выигрышных и 5 проигрышных билетов. Имеем:

n — число равновозможных испытаний, когда из 10 имеющихся выиграли оба билета (порядок не важен). Поэтому $n = C_{10}^2$;

m — благоприятные случаи, когда ***или*** эти 2 билета взяли сразу из выигрышных ($m = C_5^2$), ***или*** один из билетов выиграл C_5^1, а другой проиграл — C_5^1. Так как купили оба билета (***и*** один ***и*** другой), то, применив правило произведения, имеем: $C_5^1 C_5^1 = 5 \cdot 5$. Тогда вероятность события A равна

$$P(A) = \frac{C_5^2}{C_{10}^2} + \frac{C_5^1 C_5^1}{C_{10}^2}.$$

Чтобы найти вероятность ***или*** первого, ***или*** второго случая, надо сложить эти вероятности.

$$P(A) = \frac{5!}{2!3!} : \frac{10!}{2!8!} + 5 \cdot 5 : \frac{10!}{2!8!} = \frac{5!}{2!3!} \cdot \frac{2!8!}{10!} + \frac{25 \cdot 2!8!}{10!} =$$
$$= \frac{4 \cdot 5}{9 \cdot 10} + \frac{25 \cdot 2}{9 \cdot 10} = \frac{7}{9}.$$

Задача 1.46. В вазе 10 белых и 8 алых роз. Наудачу берут два цветка. Какова вероятность того, что они разного цвета?

Решение. Обозначим через A — событие «розы разного цвета». Всего $10 + 8 = 18$ цветков. Найдем в формуле $P(A) = m/n$ число n равновозможных и число m благоприятных исходов: $n = C_{18}^2$ (так как порядок не важен); $m = C_{10}^1 C_8^1$ (чтобы цветки были разного цвета, достаточно взять из 10 одну белую розу C_{10}^1 способами ***и*** из 8 одну алую розу C_8^1 способами). Найдем вероятность события A:

$$P(A) = \frac{C_{10}^1 C_8^1}{C_{18}^2} = \frac{10 \cdot 8 \cdot 2!16!}{18!} = \frac{10 \cdot 8 \cdot 2}{17 \cdot 18} = \frac{80}{153}.$$

Задача 1.47. Два равносильных шахматиста играют в шахматы. Что вероятнее: выиграть одному из них две партии из четырех или три из шести? (ничьи не принимаются).

Решение. Введем обозначения событий:

A — выиграл две партии из четырех;

B — выиграл три партии из шести.

Для равносильных шахматистов вероятность выигрыша в одной партии $p=\frac{1}{2}$ и вероятность проигрыша в одной партии $q=\frac{1}{2}$. Варианты выигрышей находят по формуле сочетаний, так как порядок не важен. Тогда по правилу произведения надо умножить эти сочетания на соответствующее число выигрышей и проигрышей:

$$P(A)=C_4^2p^2q^2=\frac{4!}{2!2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2=\frac{6}{2^4}=\frac{6}{16}=\frac{3}{8};$$

$$P(B)=C_6^3p^3q^3=\frac{6!}{3!3!}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^3=\frac{20}{2^4}=\frac{20}{64}=\frac{5}{16}.$$

Так как $\frac{6}{16}>\frac{5}{16}$, то $P(A)>P(B)$, т.е. вероятнее выиграть две партии из четырех.

Задача 1.48. Двенадцать рабочих получили путевки в четыре дома отдыха: «Жигули» — 3 места, «Прилесье» — 3 места, «Сосновый бор» — 2 места, «Усолье» — 4 места. С какой вероятностью три друга — Иванов, Петров и Сидоров — попадут в один дом отдыха?

Решение. Введем обозначения событий:

A — друзья отправятся в «Жигули»;

B — друзья отправятся в «Прилесье»;

C — друзья отправятся в «Усолье»;

D — друзья отправятся в «Сосновый бор»;

M — друзья попадут в один дом отдыха.

Поскольку в «Сосновый бор» было только две путевки, $P(D)=0$.

Тогда вероятность попасть в каждый дом отдыха будет определяться частным от деления числа m благоприятных исходов на число n равновозможных. Так как исходы не зависят от порядка получения трех путевок, то m и n определяются через сочетания:

$$P(A)=\frac{C_3^3}{C_{12}^3}=\frac{1}{220};\ P(B)=\frac{C_3^3}{C_{12}^3}=\frac{1}{220};\ P(C)=\frac{C_4^3}{C_{12}^3}=\frac{4}{220}.$$

Поскольку втроем можно поехать только в один дом отдыха: ***или*** в «Жигули», ***или*** в «Прилесье», ***или*** в «Усолье», по теореме сложения вероятностей несовместных событий имеем:

$$P(M)=P(A)+P(B)+P(C)+P(D)=$$
$$=\frac{1}{220}+\frac{1}{220}+\frac{4}{220}+0=\frac{6}{220}\approx 0{,}027.$$

1.8. Формула полной вероятности

В этом подразделе вы узнаете, как находить вероятность события, которое произойдет одновременно с одним из некоторых попарно несовместных событий. Однако задачи на еще не известную формулу можно решить с помощью логических связок «*и*» и «*или*». В качестве примера рекомендуем выполнить задание 1.1 (см. контрольные задания к гл. 1).

Задача 1.49. Туристу предстоит выбрать маршрут от развилки O до конечного пункта M. Перед ним схема дорог (рис. 1.18). Какова вероятность того, что турист попадет в пункт M?

Решение. Находясь в пункте O, туристу предстоит выбрать одну из дорог: или H_1, или H_2, или H_3 (см. рис. 1.18). Поскольку у туриста нет предпочтений при выборе дорог, вероятность выбора каждой из дорог равна:

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3},$$

где события выбора дорог H_1, H_2 и H_3 — несовместны и равновероятны.

Пусть событие A — приход туриста в пункт M. Так как событие A — «приход туриста в пункт M» и выбор одной из дорог независимы друг от друга, то учтем ***и*** то, какую дорогу он выбрал, ***и*** то, с какой вероятностью он попадет по этой дороге в пункт M. Тогда по теореме о произведении независимых событий AH_1 — приход в M по дороге H_1, $P(A|H_1) = \frac{1}{3}$ (из H_1 — три дороги, одна из которых нужная) аналогично $P(A|H_2) = \frac{1}{2}$ (из H_2 — две дороги). Приняв гипотезу, что он пойдет через H_3, имеем $P(A|H_3) = 0$ (из H_3 — четыре дороги, но ни одна из них не ведет в M). Тогда вероятность того, что турист придет в пункт M ***или*** по первой, ***или*** по второй, ***или*** по третьей дороге, вычислим по правилу суммы несовместных событий: $P(A) = \frac{1}{3}\cdot\frac{1}{3} + \frac{1}{2}\cdot\frac{1}{3} + 0\cdot\frac{1}{3} = \frac{5}{18}$.

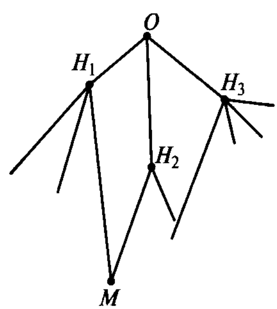


Рис. 1.18

Отсюда имеем формулу для вычисления вероятности события, которое обязательно произойдет с одним из несовместных событий H_i:

$$P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + P(A|H_3)P(H_3).$$

Рисунок схемы дорог, который был использован для того, чтобы лучше понять условие задачи, применяется и в тех случаях, когда содержание задачи не связано с представлением на карте. Иллюстрация условия в виде графа-дерева помогает увидеть возможные гипотезы (H_i), которые изображают первым ярусом графа-дерева, а также варианты осуществления события в каждой из них (события $A|H_i$), которые изображают вторым ярусом графа-дерева. Само событие A представляют вершиной графа, в которую ведут лишь те ветви графа-дерева, которые устанавливают связи, соответствующие содержанию задачи.

Выведем теперь формулу полной вероятности.

Пусть некоторые попарно несовместные события H_1, H_2, ..., H_n, образуют полную группу событий, т.е. $P(H_1) + P(H_2) + ... + P(H_n) = 1$, а событие A наступает только одновременно с одним из этих попарно несовместных событий, которые назовем *гипотезами* по отношению к самому событию A. Это означает, что произошло одно из несовместных событий A и H_1, A и H_2, ..., A и H_n, или по правилу произведения несовместных событий AH_1, AH_2, ..., AH_n. В этом случае для любого события A справедлива *формула полной вероятности*:

$$P(A) = \sum_{i=1}^{n} P(A|H_i)P(H_i). \tag{1.16}$$

Таким образом, ***вероятность события А равна сумме произведений условных вероятностей этого события по каждой из гипотез на вероятность самих гипотез.***

Задача 1.50. Для комплектования некоторого изделия поступают одноименные детали с двух заводов B_1 и B_2, причем с завода B_1 — 60 %, а с завода B_2 — 40 % всей продукции. Из каждых 100 деталей стандартными оказываются 85 деталей с первого завода и 78 деталей со второго завода. Найти вероятность того, что взятая наудачу деталь окажется стандартной (событие A), т.е. требуется найти безусловную вероятность события A.

Решение. Вероятность взять стандартную деталь с первого завода $P_1(A) = 0,85$ — это вероятность того, что деталь стандартна (событие A) при условии, что ее привезли с первого завода, т.е. условная вероятность $P(A|B_1)$, со второго завода — $P(A|B_2) = 0,78$. По условию вероятность поступления детали с первого завода равна $P(B_1) = 0,6$, со второго завода — $P(B_2) = 0,4$; тогда по формуле условной вероятности $P(AB_1) = P(A|B_1)P(B_1)$; $P(AB_2) = P(A|B_2)P(B_2)$.

Так как событие A равносильно тому, что деталь стандартна и получена с первого или второго завода, то $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$, т.е.

$$P(A) = 0,85 \cdot 0,6 + 0,78 \cdot 0,4 = 0,822.$$

1.9 Формула Байеса. Вероятность оценки гипотез

В этом подразделе вы научитесь узнавать характерные ситуации, в которых интересующее нас событие A наступает одновременно с одной из попарно несовместных гипотез, образующих полную группу событий, и осуществлять переоценку этих гипотез на основе найденной вероятности самого события A.

Пусть в ходе испытания событие A осуществилось вместе с одним из n попарно несовместных событий (гипотез) H_1, H_2, ..., H_n, образующих полную группу событий, т.е. $\sum_{i=1}^{n} P(H_i) = 1$. Найти вероятность каждой отдельной гипотезы H_1, H_2, ..., H_n при условии выполнения события A можно по формуле Байеса.

Вероятность гипотезы H_i при условии, что событие A произошло, обозначим $P(H_i|A)$.

По теореме произведения зависимых событий вероятность одновременного осуществления и события A, и гипотезы H_i равна:

$$P(AH_i) = P(H_i|A)P(A) = P(A|H_i)P(H_i).$$

Тогда вероятность гипотезы H_i при условии, что событие A произошло, найдем, выразив из последних двух произведений неизвестный множитель $P(H_i|A)$ по формуле

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A)}.$$

Заменив $P(A)$ суммой произведений по формуле полной вероятности, получим

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_{k=1}^{n} P(A|H_k)P(H_k)}. \qquad (1.17)$$

Формула Байеса дает возможность *переоценивать вероятности гипотез*, принятых до испытаний, по результатам вновь произведенных испытаний (уточнение гипотез).

Задача 1.51. Имеются три урны с шарами: в первой — 4 белых и 6 красных, во второй — 7 белых и 3 красных, в третьей — 8 белых и 2 красных. Бросают игральную кость. При выпадании одного, двух, трех очков вынимают шар из первой урны, четырех очков — из второй, другого количества очков — из третьей урны.

а) Найти вероятность того, что вынутый шар — белый.

б) Вынутый шар оказался белым. Из какой урны наиболее вероятно он мог быть извлечен?

Решение. Примем в качестве гипотез событие H_i — извлечь шар из i-й урны, где $i = \{1, 2, 3\}$.

а) Найдем вероятности гипотез.

Вероятность выпадания одной из трех граней с одним, двумя и тремя очками $P(H_1) = 3/6 = 1/2$ — вероятность первой гипотезы (H_1).

Вероятность второй гипотезы (H_2) — выпадение четырех очков — равна $P(H_2) = 1/6$.

Вероятность третьей гипотезы (H_3) — выпадение других очков, т.е. пяти или шести, равна $P(H_3) = 2/6 = 1/3$. Поскольку гипотезы составляют полную группу событий, сумма их вероятностей равна единице:

$$P(H_1) + P(H_2) + P(H_3) = 1/2 + 1/6 + 1/3 = 1.$$

Заметим, что условие выбора гипотезы могло быть записано отношением 3 : 2 : 1.

Вероятность извлечения белого шара (событие A) при условии, что он вынут из i-й урны, равна $P(A|H_i)$. Тогда полная вероятность события A — «извлечение белого шара из одной из трех урн» — равна:

$$P(A) = \sum P(A|H_i)P(H_i) = 2/5 \cdot 1/2 + 7/10 \cdot 1/6 + 4/5 \cdot 1/3 = 7/12.$$

б) Для того чтобы установить, из какой урны наиболее вероятно извлечь белый шар, надо сравнить вероятности извлечения белого шара из каждой урны. Вычислим вероятность гипотез H_i, где $i = \{1, 2, 3\}$, при условии, что событие A произошло, по формуле (1.17), занесем все данные в таблицу (табл. 1.1) и вычислим

Таблица 1.1

События, гипотезы	Вероятности		
	$P(H_i)$	$P(A\|H_i)$	$P(H_i\|A)$
H_1	$P(H_1) = 1/2$	$P(A\|H_1) = 2/5$	$P(H_1\|A) = \dfrac{1/2 \cdot 2/5}{7/12} = 12/35$
H_2	$P(H_2) = 1/6$	$P(A\|H_2) = 7/10$	$P(H_2\|A) = \dfrac{1/6 \cdot 7/10}{7/12} = 1/5 = 7/35$
H_3	$P(H_3) = 1/3$	$P(A\|H_3) = 4/5$	$P(H_3\|A) = \dfrac{1/3 \cdot 4/5}{7/12} = 16/35$
$P(A)$	$P(A) = \sum_{i=1}^{3} P(H_i)P(A\|H_i) = \frac{1}{2}\cdot\frac{2}{5}+\frac{1}{6}\cdot\frac{7}{10}+\frac{1}{3}\cdot\frac{4}{5}=\frac{35}{60}=\frac{7}{12}$		

вероятности гипотез H_i при условии, что событие A произошло, по формуле (1.17).

Из трех вариантов наибольшая вероятность извлечь белый шар из третьей урны, так как вероятность $P(H_3|A) = 16/35$ максимальная. Поэтому можно сделать вероятностный вывод о том, что белый шар скорее всего извлечен из третьей урны.

? Почему в учебниках по теории вероятностей так часто встречаются задачи «без сюжета», например о шарах, игральных костях, картах и др.?

Эти задачи, называемые ключевыми (см. задачу 1.51), позволяют увидеть характерные особенности задач отдельного вида, а также общие приемы их решения. Анализ условия и решения таких задач помогает решать аналогичные задачи с иным содержанием.

Задача 1.52. Статистика запросов кредитов в банке такова: 10 % — государственные органы, 30 % — другие банки, остальное — физические лица. Вероятности невозврата кредита соответственно равны: 0,01; 0,05 и 0,2. Найти вероятность невозврата очередного кредита. Начальнику кредитного отдела доложили, что получено сообщение о невозврате кредита, но в факсе имя клиента плохо пропечатано. Какова вероятность того, что данный кредит не возвращает другой банк?

Решение. 1) Вероятность невозврата найдем по формуле полной вероятности. Пусть H_1 означает, что запрос поступил от государственного органа, H_2 — от банка, H_3 — от физического лица и событие A — невозврат рассматриваемого кредита. Тогда

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + P(H_3)P(A|H_3) =$$
$$= 0{,}1 \cdot 0{,}01 + 0{,}3 \cdot 0{,}05 + 0{,}6 \cdot 0{,}2 = 0{,}136.$$

2) Вероятность того, что данный кредит не возвращает другой банк, найдем по формуле Байеса:

$$P(H_2|A) = \frac{P(H_2)P(A|H_2)}{P(A)} = \frac{0{,}015}{0{,}136} = \frac{15}{136} \approx \frac{1}{9}.$$

Задача 1.53. Из 10 студентов, пришедших на экзамен по математике, трое подготовились отлично, четверо — хорошо, двое — удовлетворительно и один студент вообще не готовился к экзамену. Из 20 экзаменационных вопросов первые три студента могут ответить на все 20 вопросов, хорошо подготовившиеся — на 16, удовлетворительно подготовившиеся — на 10, неподготовившийся студент помнит ответы лишь на пять вопросов с лекций, на которых присутствовал.

Экзаменующийся студент ответил на все три вопроса. Какова вероятность того, что он действительно выучил весь материал?

Решение. Введем обозначения событий:

A — студент ответил на N_i вопроса;

H_1 — студент подготовился отлично;

H_2 — студент подготовился хорошо;

H_3 — студент подготовился удовлетворительно;

H_4 — студент к экзамену не готов.

Тогда $P(H_1)=0{,}3$; $P(H_2)=0{,}4$; $P(H_3)=0{,}2$; $P(H_4)=0{,}1$.

Вероятность ответить на все три вопроса для тех, кто подготовился:

отлично — $P(A|H_1)=1$;

хорошо — $P(A|H_2)=\dfrac{C_{16}^3}{C_{20}^3}=\dfrac{16}{20}\cdot\dfrac{15}{19}\cdot\dfrac{14}{18}\approx 0{,}491$ (ответил ***и*** на 1-й, ***и*** на 2-й, ***и*** на 3-й вопросы, а после каждого ответа количество предлагаемых вопросов на один уменьшалось);

удовлетворительно — $P(A|H_3)=\dfrac{C_{10}^3}{C_{20}^3}=\dfrac{10}{20}\cdot\dfrac{9}{19}\cdot\dfrac{8}{18}\approx 0{,}105.$

Для неподготовленного студента $P(A|H_4)=\dfrac{C_{16}^5}{C_{20}^3}=\dfrac{5}{20}\cdot\dfrac{4}{19}\cdot\dfrac{3}{18}\approx$ $\approx 0{,}009.$

По формуле Байеса найдем вероятность того, что ответивший студент выучил весь материал:

$$P(H_i|A)=\frac{0{,}3\cdot 1}{0{,}3\cdot 1+0{,}4\cdot 0{,}491+0{,}2\cdot 0{,}105+0{,}1\cdot 0{,}009}\approx 0{,}58.$$

Так как вероятность невелика, преподаватель не сразу ставит оценку «отлично», а задает дополнительные вопросы.

Задача 1.54. Комплекты деталей изготовляют опытный мастер и ученик. Опытный мастер допускает брак в пяти случаях из 100, а ученик портит каждую пятую деталь. В комплект входят две детали. Для контроля наугад взяли один комплект. Какова вероятность того, что

а) выбранный комплект оказался бракованным;

б) выбранный комплект сделан мастером при условии, что он бракованный?

Решение. Введем обозначения событий: A — комплект бракованный; B_1 — деталь сделана мастером; B_2 — деталь сделана учеником. Вероятность того, что мастер допустит брак, равна $P(\bar{B}_1)=1/20$, ученик — $P(\bar{B}_2)=1/5$. Тогда, поскольку в комплект входят две детали, множество возможных событий включает гипотезы, образующие полную группу событий:

H_1 — обе детали без брака, причем $H_1=B_1B_2$;

H_2 — брак допустил только мастер, причем $H_2=\bar{B}_1B_2$;

H_3 — брак допустил только ученик, причем $H_3=B_1\bar{B}_2$;

H_4 — обе детали бракованные, причем $H_4 = \bar{B}_1\bar{B}_2$.

Найдем вероятности гипотез, учитывая, что каждая из них есть произведение независимых событий:

$$P(H_1) = P(B_1)P(B_2) = [1 - P(\bar{B}_1)][1 - P(\bar{B}_2)] = \frac{19}{20}\cdot\frac{4}{5} = \frac{76}{100} = 0{,}76;$$

$$P(H_2) = P(\bar{B}_1)[1 - P(\bar{B}_2)] = \frac{1}{20}\cdot\left(1 - \frac{1}{5}\right) = \frac{1}{20}\cdot\frac{4}{5} = \frac{4}{100} = 0{,}04;$$

$$P(H_3) = [1 - P(\bar{B}_1)]P(\bar{B}_2) = \left(1 - \frac{1}{20}\right)\cdot\frac{1}{5} = \frac{19}{20}\cdot\frac{1}{5} = \frac{19}{100} = 0{,}19;$$

$$P(H_4) = P(\bar{B}_1)P(\bar{B}_2) = \frac{1}{20}\cdot\frac{1}{5} = \frac{1}{100} = 0{,}01.$$

Сложив вероятности всех гипотез, убеждаемся, что их сумма равна $\sum\limits_{i=1}^{4} P(H_i) = 0{,}76 + 0{,}04 + 0{,}19 + 0{,}01 = 1$, и они действительно образуют полную группу событий.

Найдем условные вероятности события A — «комплект бракованный» для каждой гипотезы: $P(A|H_1) = 0$; $P(A|H_2) = 1$; $P(A|H_3) = 1$; $P(A|H_4) = 1$.

а) Тогда, согласно формуле полной вероятности, вероятность того, что выбранный комплект оказался бракованным, равна:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + P(H_3)P(A|H_3) + P(H_4)P(A|H_4), \text{ или}$$

$$P(A) = 0{,}76\cdot 0 + 0{,}04\cdot 1 + 0{,}19\cdot 1 + 0{,}01\cdot 1 = 0{,}24.$$

б) Применив формулу Байеса, получим вероятность того, что выбранный комплект сделан мастером, при условии, что он бракованный:

$$P(H_2|A) = \frac{P(H_2)P(A|H_2)}{P(A)} = \frac{0{,}04\cdot 1}{0{,}24} = \frac{1}{6}.$$

Задача 1.55. При обследовании пациента у врача возникли подозрения на одно из двух близких по симпотомам заболеваний H_1 и H_2. Их вероятность в данной ситуации $P(H_1) = 0{,}6$; $P(H_2) = 0{,}4$, т.е. у пациента обнаружена болезнь или H_1, или H_2. Для уточнения диагноза пациенту проводят дополнительное обследование. В случае положительной реакции вероятность первого заболевания 0,9, отрицательной — 0,1. Для второго заболевания положительная и отрицательная реакции равновероятны.

В результате двукратного проведения дополнительного обследования реакция дважды оказалась отрицательной. Необходимо найти вероятность каждого заболевания H_1 и H_2 после дообследования.

Решение. Обозначим через A событие, заключающееся в отрицательной реакции.

Для заболевания H_1 вероятность $P(A|H_1) = 0{,}1 \cdot 0{,}1 = 0{,}01$ (***и*** в первом, ***и*** во втором случаях реакция отрицательная). Аналогично для H_2 вероятность $P(A|H_2) = 0{,}5 \cdot 0{,}5 = 0{,}25$. Найдем вероятность первого заболевания в случае, если произошло событие A, по формуле Байеса:

$$P(H_1|A) = \frac{P(A|H_1)P(H_1)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2)} = \frac{0{,}01 \cdot 0{,}6}{0{,}01 \cdot 0{,}6 + 0{,}25 \cdot 0{,}4} = \frac{6}{106}.$$

Аналогично для второго заболевания имеем:

$$P(H_2|A) = \frac{P(A|H_2)P(H_2)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2)} = \frac{0{,}25 \cdot 0{,}4}{0{,}01 \cdot 0{,}6 + 0{,}25 \cdot 0{,}4} = \frac{100}{106}.$$

Поэтому врач остановился на гипотезе, что у больного второе заболевание.

1.10. Независимые повторные испытания. Формула Бернулли

В этом подразделе рассмотрим вычисления вероятности при независимых повторных *испытаниях.*

Рассмотрим часто встречающиеся ситуации:

• контроль качества продукции показал наличие m бракованных изделий из общего числа n изделий, причем для каждого из изделий вероятность брака p есть величина постоянная;

• в результате приобретения n лотерейных билетов выявляется число m выигрышей, причем для каждого из лотерейных билетов вероятность выигрыша p есть величина постоянная.

Эти примеры обладают общими особенностями:

1) каждому испытанию соответствует один из двух возможных исходов, которые условно назовем «успех» и «неуспех». Эти исходы являются между собой противоположными и несовместными событиями;

2) вероятность «успеха» p есть величина постоянная;

3) во всех типичных ситуациях проводились независимые испытания, т.е. вероятность наступления события при каждом испытании не зависела от результата других аналогичных испытаний.

Перечисленные ситуации являются *независимыми повторными испытаниями*, т.е. вероятность появления события A в любом из них не зависит от исходов в других.

Пусть производятся независимые повторные испытания, в каждом из которых возможно появление событий $A_1, A_2, ..., A_k$ с постоянными вероятностями $p_1, p_2, ..., p_k$. События $A_1, A_2, ..., A_k$ образуют полную группу, т.е. в каждом испытании должно происходить одно из них, поэтому $\sum_{i=1}^{k} p_i = 1$. Эксперимент, при котором $k=2$, т.е. возможно появление лишь двух событий (например, выпадение орла или решки), называется *испытаниями Бернулли*, или *схемой Бернулли*.

Пусть в испытании Бернулли вероятность события $A_1 = A$, условно называемого *успехом*, равна p. Назовем противоположное событие $A_2 = \bar{A}$ — неуспех или неудача. Тогда вероятность того, что событие A не появится, равна $q = 1 - p$, так как $P(A) + P(\bar{A}) = 1$. Например, для события, которое осуществится в трех из четырех повторных испытаний, возможны варианты: $\bar{A}AAA$, $A\bar{A}AA$, $AA\bar{A}A$, $AAA\bar{A}$. Их число равно $C_n^m = C_4^3 = 4$.

Пусть производится n испытаний Бернулли. Найдем вероятность $P_n(m)$ того, что в этих n испытаниях событие A осуществится ровно m раз. Так как вероятность появления события A в каждом из m (***и*** в 1-м, ***и*** во 2-м, ... ***и*** в m-м случаях) случаях равна p, то по *правилу умножения вероятностей* вероятность появления этого события равна p^m. Но так как известно, что событие A не должно появиться в остальных $n - m$ испытаниях, то по правилу произведения вероятность его непоявления равна $q^{n-m} = (1 - p)^{n-m}$.

Последовательность выполнения и невыполнения события A не зависит от порядка их осуществления в этих испытаниях. Поэтому количество таких вариантов находится с помощью сочетаний $C_n^m = \dfrac{n!}{m!(n-m)!}$. Так как приходится учитывать ***и*** число вариантов осуществления события A, ***и*** вероятность его появления, ***и*** вероятность его непоявления, то по правилу произведения находим формулу вероятности числа успехов m в n испытаниях Бернулли:

$$P_n(m) = C_n^m p^m q^{n-m}. \tag{1.18}$$

Формула Бернулли, названная по имени швейцарского математика Я. Бернулли, имеет очень широкое применение в прикладных науках, так как она описывает реальные жизненные ситуации: из проверяемых n изделий вероятность обнаружить m бракованных; из посаженных n семян известной всхожести p взойдут m; из зарегистрированных n новорожденных вероятность рождения m мальчиков.

Задача 1.56. Вероятность того, что лампа перегорит после 1000 ч работы, равна 0,8.

а) Какова вероятность того, что, проработав 1000 ч, три из пяти ламп будут продолжать гореть?

б) Найти вероятность того, что из пяти ламп не менее трех будут гореть.

Решение. а) По условию задачи работа каждой лампы в течение 1000 ч — испытание, поэтому $n = 5$. Событие A — работа трех ламп происходит в трех испытаниях, т.е. $m = 3$. В одиночном испытании (одной лампы) вероятность того, что лампа перегорит, равна $p = 0{,}8$, а вероятность того, что лампа будут гореть, равна $q = 1 - p = 1 - 0{,}8 = 0{,}2$. По формуле Бернулли имеем:

$$P(A) = P_5(3) = \frac{5!}{3!2!}0{,}2^3 \cdot 0{,}8^2 = 10 \cdot 0{,}008 \cdot 0{,}64 = 0{,}0512.$$

б) Обозначим через B событие «из пяти ламп не менее трех будут гореть».

Условие «не менее трех ламп будут гореть» означает, что могут гореть ***или*** три лампы (событие A_3), ***или*** четыре лампы (событие A_4), ***или*** пять ламп (событие A_5). Тогда по формуле сложения вероятностей несовместных событий имеем: $P(B) = P(A_3) + P(A_4) + P(A_5)$. Так как вероятность каждого из этих событий находится по формуле Бернулли $P_5(m) = C_5^m p^m q^{5-m}$, где $m = 3,\ 4,\ 5$, имеем:

$$P(B) = P_5(3) + P_5(4) + P_5(5) =$$
$$= 10 \cdot 0{,}2^3 \cdot 0{,}8^2 + 5 \cdot 0{,}2^4 \cdot 0{,}8 + 1 \cdot 0{,}2^5 \cdot 0{,}8^0 =$$
$$= 0{,}0579 \approx 0{,}06.$$

Итак, ***применение формулы Бернулли предполагает выполнение ряда условий***:

- число опытов фиксировано и ограничено;
- опыты независимы (результат одного испытания не влияет на результаты остальных);
- результатом испытания может быть лишь один из двух взаимно исключающих исходов (успех или неуспех);
- вероятность успеха в каждом опыте одна и та же.

Решим задачу 1.47, используя формулу Бернулли.

В первом случае $n_1 = 4$, $m_1 = 2$. Так как выигрыши и проигрыши равновероятны, то $p = q = 1/2$. Тогда вероятность выиграть две партии из четырех равна $P_4(2) = C_4^2 p^2 q^2 = \frac{4!}{2!2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{6}{16} = \frac{3}{8}$.

Во втором случае $n_2 = 6$, $m_2 = 3$, $p = q = 1/2$. Тогда аналогично $P_6(3) = C_6^3 p^3 q^3 = \frac{6!}{3!3!}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^3 = \frac{5}{16}$. Так как $P_4(2) > P_6(3)$, то вероятнее выиграть две партии из четырех (получили тот же результат).

Задача 1.57. Банк имеет шесть филиалов. С вероятностью 0,2 независимо от других каждый филиал может заказать крупную сумму денег. Какова вероятность того, что есть заявка от первого отделения, если поступило две заявки?

Решение. Задачу надо решать с помощью формулы Бернулли. Это условная вероятность $P(B|A)$, где событие A «поступило две заявки», а событие B «поступила заявка от первого отделения». Тогда вероятность события A найдем по формуле Бернулли, имея подмножество 2 из всего множества отделений, равного шести.

Вероятность произведения событий найдем, учитывая *и* то, что 0,2 — вероятность поступления заявки от первого отделения, *и* то, что поступила еще одна (вторая) заявка с вероятностью, которую найдем по формуле Бернулли из пяти оставшихся отделений банка. Отсюда по формуле для условной вероятности имеем:

$$P(B/A) = \frac{P(A \cdot B)}{P(A)} = \frac{0,2C_5^1 \cdot 0,2^1 \cdot 0,8^4}{C_6^2 \cdot 0,2^2 \cdot 0,8^4} = \frac{1}{3}.$$

1.11. Наивероятнейшее число наступления события в схеме Бернулли

На практике иногда требуется знать наивероятнейшее число наступления события в схеме Бернулли, т.е. при каком m и фиксированном n вероятность $P_n(m)$ принимает наибольшее значение.

Обозначим это наивероятнейшее число через m_0 и найдем его, применив формулу Бернулли.

Рассмотрим отношение:

$$\frac{P_n(m)}{P_n(m-1)} = \frac{C_n^m p^m q^{n-m}}{C_n^{m-1} p^{m-1} q^{n-m+1}} = \frac{n!(m-1)!(n-m+1)!\,p}{m!(n-m)!\,n!\,q} = \frac{(n-m+1)p}{mq} =$$

$$= \frac{np - mp + p}{mq} = \frac{np - m(1-q) + p}{mq} = \frac{mq}{mq} + \frac{(np+p) - m}{mq} =$$

$$= 1 + \frac{p(n+1) - m}{mq}.$$

В зависимости от значений параметров n, m, p, q возможны три варианта:

$$P_n(m) > P_n(m-1) \text{ при } m < p(n+1);$$

$$P_n(m) = P_n(m-1) \text{ при } m = (n+1)p;$$
$$P_n(m) < P_n(m-1) \text{ при } m > (n+1)p.$$

Наивероятнейшее число m_0 должно удовлетворять условиям:

$$P_n(m_0 - 1) \le P_n(m_0); \quad P_n(m_0 + 1) \le P_n(m_0),$$

откуда $(n+1)p - 1 \le m_0 \le (n+1)p$. Учитывая, что $p = 1 - q$, после преобразований $np + p - 1 = np - q$ имеем

$$np - q \le m_0 \le np + p, \tag{1.19}$$

т.е. наивероятнейшая частота m_0 находится в интервале $m_0 \in [np - q;\ np + p]$, длина которого равна единице. Так как наивероятнейшая частота может выражаться только целым числом, она может принимать или одно значение, если границы выражены дробными числами, или два значения, если границы сами являются целыми числами. Тогда нужно сравнить вероятности на концах.

Задача 1.58. В результате многолетних наблюдений вероятность дождя 21 июля в городе N составляет 0,3. Найти наивероятнейшее число дождливых дней 21 июля на ближайшие 30 лет.

Решение. По условию задачи $p = 0{,}3$; $q = 0{,}7$; $n = 30$.

Так как $np - q \le m_0 \le np + p$, то наивероятнейшее число дождливых дней m_0 найдем из двойного неравенства:

$$30 \cdot 0{,}3 - 0{,}7 \le m_0 \le 30 \cdot 0{,}3 + 0{,}3, \text{ или } 8{,}3 \le m_0 \le 9{,}3.$$

В этом интервале находится лишь одно целое решение $m_0 = 9$, т.е. с наибольшей долей вероятности можно утверждать, что в ближайшие 30 лет в этот день будет дождь лишь в девяти случаях.

Задача 1.59. В питомнике 40 вакцинированных кроликов и 10 контрольных. Осуществили проверку 14 кроликов и результат зарегистрировали. Определить наивероятнейшее число появления контрольного кролика.

Решение. По условию $n = 14$; $p = 10/50 = 0{,}2$; $q = 40/50 = 0{,}8$, тогда

$$14 \cdot 0{,}2 - 0{,}8 \le m_0 \le 14 \cdot 0{,}2 + 0{,}2, \text{ или } 2 \le m_0 \le 3.$$

Задача имеет два решения: контрольных кроликов будет ***или*** 2, ***или*** 3. Тогда можно подставить эти числа в формулу Бернулли и убедиться, что вероятности равны.

Задача 1.60. В ящике 100 стандартных деталей и 20 бракованных. Из ящика берут деталь, регистрируют ее качество и возвращают на место. Наивероятнейшее число достать стандартную деталь равно 15. Сколько деталей успели проверить?

Решение. По условию задачи $m_0 = 15$, вероятность достать стандартную деталь составляет $p = \frac{100}{120} = \frac{5}{6}$; $q = \frac{1}{6}$. Подставив значения в двойное неравенство (1.19), найдем n:

$$\frac{5}{6}n-\frac{1}{6}\le 15\le \frac{5}{6}n+\frac{5}{6}.$$

Решением неравенства относительно n будет $\frac{85}{5}\le n\le\frac{91}{5}$, или $17\le n\le 18{,}2$. Итак, проверили ***или*** 17, ***или*** 18 деталей.

Анализируя двойное неравенство для нахождения наивероятнейшего числа успеха в n опытах, можно заметить особую роль произведения np, которое можно рассматривать как среднее число успехов в n испытаниях, т.е. $m_0\approx np$.

Задача 1.61. Клиентов сбербанка обслуживают два филиала. Первый филиал за рабочий день обслужил 120 клиентов, второй — 140 клиентов. Вероятность того, что эти клиенты взяли деньги со счетов, составляет соответственно 0,94 и 0,8. Найти наивероятнейшее число клиентов, взявших деньги со своих счетов. Какой из филиалов обслуживает больше клиентов?

Решение. $n_1=120$; $n_2=140$; $p_1=0{,}94$; $p_2=0{,}8$. Найдем m_{01} и m_{02}: $m_{01}\approx n_1p_1=120\cdot 0{,}94=112{,}8\approx 113$; $m_{02}\approx n_2p_2=140\cdot 0{,}8=112$. Больше клиентов обслуживает первый филиал, так как $113>112$.

Однажды в США при проверке программы космических исследований было установлено, что ракета стоимостью несколько миллионов долларов могла взорваться сразу после запуска. Как выяснилось, в записанной на ЭВМ программе управления полетом пропущен знак — точка с запятой. Последствия такой «ничтожно» малой ошибки могли оказаться катастрофическими.

? Какова же вероятность допустить подобную ошибку и существуют ли математические законы, способные определить степень риска даже для таких *редких* явлений?

Ответы на эти вопросы дает формула Пуассона.

1.12. Формула Пуассона

В этом подразделе вы научитесь решать задачи по определению вероятностей редких явлений по формуле Пуассона.

Если в каждом отдельном независимом испытании вероятность одного из событий p или q близка к нулю, то события называют ***редкими***. Редкими можно считать события: появление ошибки на некоторой странице в книге, телефонный звонок в квартиру за сутки, количество осадков, выпавших за июнь в городе N, и др.

Для определения вероятности таких явлений применяется асимптотическая формула Пуассона, названная по имени французского математика С. Пуассона, $P_n(m) \approx \frac{\lambda^m}{m!}e^{-\lambda}$, где $\lambda = np$ ($\lambda = \text{const}$, $n \to \infty$).

Теорема 1.8. *Если вероятность p события A в каждом повторном испытании связана с числом независимых испытаний n, которое достаточно велико, то вероятность того, что в n независимых испытаниях событие A произойдет m раз, приближенно находится по формуле*

$$P_n(m) \approx \frac{\lambda^m}{m!}e^{-\lambda} \quad (\lambda = np = \text{const}). \tag{1.20}$$

Доказательство. По формуле Бернулли имеем:

$$P_n(m) = \frac{n!}{m!(n-m)!}p^m q^{n-m} = \frac{(n-m+1)\ldots(n-1)n}{m!}p^m q^{n-m}.$$

Выразим p и q через λ и n. Так как $\lambda = np$, то $p = \frac{\lambda}{n}$, а $q = 1 - \frac{\lambda}{n}$. Тогда формула примет вид

$$P_n(m) = \frac{(n-m+1)\ldots(n-1)n}{m!}\left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m}.$$

Найдем приближенное значение вероятности при $n \to \infty$ с помощью предела:

$$P_n(m) \approx \lim_{n\to\infty} \frac{(n-m+1)\ldots(n-1)n}{m!}\left(\frac{\lambda}{n}\right)^m \cdot \left(1 - \frac{\lambda}{n}\right)^{n-m} =$$

$$= \frac{\lambda^m}{m!}\lim_{n\to\infty}\left(\frac{n-m+1}{n}\ldots\frac{n-1}{n}\frac{n}{n}\right)\left(1-\frac{\lambda}{n}\right)^{-m}\left(1-\frac{\lambda}{n}\right)^{n} =$$

$$= \frac{\lambda^m}{m!}\lim_{n\to\infty}\left(1-\frac{m-1}{n}\right)\ldots\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-m}e^{-\lambda} = \frac{\lambda^m}{m!}e^{-\lambda}.$$

При выводе формулы использован второй замечательный предел:

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \lim_{n\to\infty}\left[\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda} =$$

$$= \left[\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda} = \left[\lim_{x\to 0}(1+x)^{1/x}\right]^{-\lambda} = e^{-\lambda}.$$

Пределы остальных двучленов равны единице при $n \to \infty$:

$$\lim_{n\to\infty}\left(1-\frac{m}{n}\right) = 1, \text{ где } 0 < m < n, \quad \lim_{n\to\infty}\left(1-\frac{1}{n}\right) = 1 \text{ и т.д.}$$

Итак, закон Пуассона применяется для определения вероятности появления m событий, происходящих независимо друг от друга с постоянной вероятностью (средней интенсивностью), причем число испытаний n достаточно велико ($n \to \infty$), а вероятность появления события в каждом испытании p мала, т.е. $p \to 0$ (или $q \to 0$).

Приближенные значения вероятности по формуле Пуассона приведены в табл. П. 1.

Задача 1.62. Прядильщица обслуживает 1 000 веретен. Вероятность обрыва нити на одном веретене в течение 1 мин равна 0,002. Найти вероятность того, что в течение 1 мин обрыв произойдет более чем на трех веретенах.

Решение. По условию задачи $n = 1\,000$; $p = 0{,}002$; $m > 3$.

Так как обрыв нити на каждом веретене может либо произойти, либо не произойти, то речь идет о независимых повторных испытаниях. Тот факт, что вероятность обрыва нити мала, дает возможность использовать для решения формулу Пуассона для редких явлений.

Имеем $\lambda = np = 1\,000 \cdot 0{,}002 = 2$.

Тогда $P_n\{k > 3\} = 1 - P_n\{k \le 3\} = 1 - (P_n(0) + P_n(1) + P_n(2) + P_n(3))$.

Используя формулу Пуассона, имеем:

$$P(m > 3) = 1 - \sum_{k=0}^{3} \frac{2^k}{k!} e^{-2} = 1 - e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!}\right) =$$

$$= 1 - \frac{1}{e^2}\left(1 + 2 + 2 + \frac{4}{3}\right) \approx 0{,}143.$$

Задача 1.63. Радиоаппаратура состоит из 1 000 электроэлементов. Вероятность отказа одного из них в течение года работы равна 0,001 и не зависит от состояния других элементов. Какова вероятность отказа:

а) двух элементов;

б) не менее двух и не более четырех элементов;

в) не менее двух элементов в год?

Решение. Независимые повторные испытания при $p = 0{,}001$ вычисляют по формуле Пуассона для редких явлений. Тогда $\lambda = np = 1\,000 \cdot 0{,}001 = 1$. Найдем вероятность по табл. П. 1.

а) $P_{1000}(2) = P(2,\ 1) = 0{,}184$;

б) $P_{1000}\{2 \le m \le 4\} = P_{1000}(2) + P_{1000}(3) + P_{1000}(4) = 0{,}184 + 0{,}061 + 0{,}015 = 0{,}26$;

в) $P_{1000}\{m \ge 2\} = 1 - P_{1000}\{m < 2\} = 1 - (P_{1000}(0) + P_{1000}(1)) = 1 - (0{,}368 + 0{,}368) = 0{,}264$.

Задача 1.64. АТС получает в среднем 300 вызовов в час. Какова вероятность того, что в указанную минуту будут два вызова?

Решение. Количество вызовов в среднем можно найти, если вычислить число вызовов в минуту, т.е. $\lambda = 300 : 60 = 5$. Найдем $P(2) = \frac{5^2}{2!}e^{-5} = \frac{25}{2e^5} \approx 0,09.$

Задача 1.65. Вероятность выхода из строя кодового замка в течение месяца равна 2 %. Какова вероятность того, что в партии из 600 кодовых замков, установленных фирмой на входных дверях домов, 20 замков выйдут из строя в течение месяца?

Решение. По условию $n = 600$; $m = 20$, $p = 0,02$, $q = 0,98$. Если применять формулу Бернулли, $P_{600}(20) = C_{600}^{20} \cdot 0,2^{20} \cdot 0,98^{580}$, то подсчет вероятности будет весьма сложен.

1.13. Локальная и интегральная теоремы Муавра—Лапласа

В этом подразделе вы научитесь определять вероятность событий в независимых повторных испытаниях, каждое из которых имеет постоянную вероятность p при больших значениях n. В таких случаях, когда формулой Бернулли пользоваться неудобно из-за громоздких вычислений, для поиска вероятности применяют формулу Муавра—Лапласа.

На практике часто возникает потребность вычислять вероятность $P_n(m)$ при достаточно больших значениях параметров. В таких случаях применяют локальную приближенную формулу Муавра—Лапласа, выведенную для частного случая английским математиком А. Муавром и впоследствии обобщенную французским математиком П. Лапласом.

Локальная формула Муавра—Лапласа. Если в n независимых испытаниях событие A происходит с постоянной вероятностью p, которая не очень близка к 0 и 1, то при достаточном большом количестве испытаний вероятность того, что событие A произойдет m раз, приближенно равна

$$P_n(m) \approx \frac{\varphi(x)}{\sqrt{npq}}, \tag{1.21}$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $x = \frac{m - np}{\sqrt{npq}}$.

Для функции $\varphi(x)$ составлена таблица ее значений, которая публикуется в справочной литературе (см. табл. П. 2).

Тогда задача 1.65 (см. подразд. 1.12) с помощью локальной формулы Муавра—Лапласа может быть решена так. При $n = 600$; $m = 20$; $p = 0,02$; $q = 0,98$ найдем:

1) $\sqrt{npq} = \sqrt{600 \cdot 0,02 \cdot 0,98} \approx 3,43;$

2) $x = \frac{m - np}{\sqrt{npq}} = \frac{20 - 600 \cdot 0,02}{3,43} \approx 2,33.$

Значение функции $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ для $x=2,33$ определяем по табл. П. 2:

$$\varphi(2,33)\approx\frac{1}{\sqrt{2\pi}}e^{-(2,33)^2/2}\approx 0,026.$$

Тогда искомая вероятность $P_{600}(20)$ равна

$$P_{600}(20)=\frac{\varphi(x)}{\sqrt{npq}}=\frac{0,026}{3,43}=0,00758.$$

Свойства функции $\varphi(x)$:

1) функция $\varphi(x)$ — четная;

2) $x=\pm1$ — точки перегиба графика функции;

3) таблица значений функции $\varphi(x)$ составлена для $0\le x\le 5$, так как при $x\ge 5$ $\varphi(x)\to 0$;

4) формула применяется при $npq\ge 10$.

Интегральная формула Лапласа. Пусть требуется найти вероятность суммы $\sum\limits_{m=m_1}^{m_2}P_n(m)$ на некотором интервале $[m_1, m_2]$. Для подсчета такой вероятности применяют интегральную формулу Лапласа.

Если в n независимых испытаниях событие A происходит с постоянной вероятностью p, которая отличается от 0 и 1, то при достаточно большом значении n вероятность того, что частота m события A находится в интервале $[m_1, m_2]$, приближенно равна

$$P_n(m_1\le m\le m_2)\approx\Phi(x_2)-\Phi(x_1), \tag{1.22}$$

где $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-t^2/2}dt$,

причем $x_1=\frac{m_1-np}{\sqrt{npq}}$, $x_2=\frac{m_2-np}{\sqrt{npq}}$.

Функция $\Phi(x)$ является первообразной функции $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ из локальной формулы Муавра—Лапласа и называется *функцией* (*интегралом*) *Лапласа*. Она принимает значения в интервале (–0,5; 0,5), причем $\Phi(-x)=-\Phi(x)$, т.е. функция Лапласа — нечетная, $\Phi(+\infty)=0,5$; $\Phi(0)=0$ (подробнее см. подразд. 2.6). Таблица значений функции Лапласа $\Phi(x)$ приведена в приложении (см. табл. П. 3).

Задача 1.66. В каждом из 700 независимых испытаний на брак появление стандартной лампочки происходит с постоянной вероятностью 0,65. Найти вероятность того, что при таких условиях появление бракованной лампочки произойдет чаще, чем в 230 испытаниях, но реже, чем в 270 случаях.

Решение. Введем обозначение:

событие A — появление бракованной лампочки.

По условию задачи $n = 700$; $p = 1 - 0{,}65 = 0{,}35$; $q = 0{,}65$. Найдем $P_{700}(230 < m < 270)$, если $m_1 = 230$, $m_2 = 270$. Применим интегральную формулу Лапласа:

1) $\sqrt{npq} = \sqrt{700 \cdot 0{,}35 \cdot 0{,}65} = 12{,}6$;

2) $x_1 = \dfrac{230 - 700 \cdot 0{,}35}{12{,}6} = -1{,}19$; $x_2 = \dfrac{270 - 700 \cdot 0{,}35}{12{,}6} = 1{,}98$.

Значения функции $\Phi(x)$ при $x = -1{,}19$ и $x_2 = 1{,}98$ возьмем из табл. П. 3:

$$\Phi(x_1) = \Phi(-1{,}19) = -\Phi(1{,}19) = -0{,}383;$$
$$\Phi(x_2) = \Phi(1{,}98) = 0{,}476.$$

Искомая вероятность

$$P_{700}(230 < m < 270) = \Phi(x_2) - \Phi(x_1) = 0{,}476 + 0{,}383 = 0{,}859.$$

Задача 1.67. В каждом из 500 независимых испытаний на всхожесть событие A — прорастание исследуемого семени — происходит с постоянной вероятностью 0,4. Найти вероятность того, что среди 500 посаженных семян взойдет менее 235.

Решение. По условию задачи $n = 500$; $p = 0{,}4$, $m_1 = 0$; $m_2 = 235$; $q = 1 - 0{,}4 = 0{,}6$. Найдем: $P_{500}(m < 235)$:

1) $\sqrt{npq} = \sqrt{500 \cdot 0{,}4 \cdot 0{,}6} \approx 11$;

2) $x_1 = \dfrac{m_1 - np}{\sqrt{npq}} = \dfrac{0 - 500 \cdot 0{,}4}{11} \approx -18$; $x_2 = \dfrac{m_2 - np}{\sqrt{npq}} = \dfrac{235 - 500 \cdot 0{,}4}{11} = 3{,}18$;

3) $\Phi(x_1) = \Phi(-18) = -\Phi(18) \approx -0{,}5$; $\Phi(x_2) = \Phi(3{,}18) = -0{,}499$;

4) $P_{500}(m < 235) = P_{500}(0 < m < 235) = \Phi(x_2) - \Phi(x_1) = 0{,}499 - (-0{,}5) = 0{,}999$.

В табл. П. 3 даны значения функции $\Phi(x)$ лишь до $x = 5$, так как для всех $x \geq 5$ отклонение функции Лапласа $\Phi(x)$ от 0,5 пренебрежимо мало, можно воспользоваться асимптотическим разложением.

Задача 1.68. Прибор выходит из строя при отказе (неисправности) его микросхемы. Вероятность отказа в течение 1 ч работы прибора 0,02. С какой вероятностью за 100 ч эксплуатации прибора микросхему придется менять три раза?

Решение. По условию задачи: $p = 0{,}02$; $n = 100$; $m = 3$; $q = 0{,}98$. Тогда $\lambda = np = 100 \cdot 0{,}02 = 2 < 10$, поэтому применим формулу

Пуассона для редких явлений: $P_{100}(3) \approx \frac{2^3}{3!}e^{-2}$, и, воспользовавшись табл. П. 1, сразу получим приближенный ответ:

$$P_{100}(3) \approx P(3;\ 2) \approx 0{,}18045 \approx 0{,}18.$$

Сравним логический ответ с ответом при использовании формулы Лапласа:

$$x = \frac{m - np}{\sqrt{npq}} = \frac{3-2}{\sqrt{100 \cdot 0{,}02 \cdot 0{,}98}} \approx \frac{1}{\sqrt{1{,}96}} \approx \frac{1}{1{,}4} \approx 0{,}714.$$

Тогда значение $\varphi(x) \approx \varphi(0{,}714) \approx 0{,}31$ определим по табл. П. 2, откуда

$$P(0{,}714) \approx \frac{0{,}308}{\sqrt{1{,}96}} \approx \frac{0{,}31}{1{,}4} \approx 0{,}22.$$

Разница в значениях вероятности связана с различиями в условиях использования формул Пуассона и Лапласа. Формула Лапласа обычно применяется при $\lambda > 10$, и поэтому она дала менее точный результат.

Контрольные вопросы

1.1. Дайте определение понятию «перестановки».

1.2. Что называется сочетаниями (без повторений)?

1.3. Дайте определение понятию «размещение без повторений».

1.4. Как находится размещение с повторениями?

1.5. Какое событие называется достоверным? Чему равна его вероятность?

1.6. Какое событие называются недостоверным? Чему равна его вероятность?

1.7. Какие события называются совместными?

1.8. Какие события называются зависимыми?

1.9. Какие события называются равновозможными?

1.10. Какие события образуют полную группу событий?

1.11. Какие события называются противоположными?

1.12. Что называется вероятностью события? Дайте классическое определение вероятности.

1.13. Дайте геометрическое определение вероятности.

1.14. Что называется относительной частотой события?

1.15. Какие аксиомы составляют аксиоматическое определение вероятности?

1.16. Что называют пространством элементарных исходов?

1.17. Как найти вероятность объединения двух совместных событий?

1.18. Как определить вероятность объединения двух несовместных событий?

1.19. Как найти вероятность противоположного события?

1.20. Чему равна вероятность полной группы событий?

1.21. Как найти вероятность двух независимых событий?

1.22. Чему равна вероятность произведения зависимых событий?

1.23. Как определить условную вероятность событий?

1.24. Приведите формулу для определения полной вероятности событий.

1.25. Как найти вероятность гипотезы?

1.26. Как найти вероятность независимых повторных испытаний?

1.27. В каких случаях применяется формула Пуассона?

1.28. В каких случаях применяется формула Лапласа?

1.29. В каких случаях применяется интегральная формула Муавра—Лапласа?

Контрольные задания

Задание 1. В ущелье пропала горнолыжная экспедиция. Карта ущелья прилагается (см. рис. 1.18). Место последней связи с пропавшими альпинистами отмечено буквой M. Необходимо срочно оказать помощь.

Для оказания помощи необходимо выполнить следующее:

1) назначить составы групп спасения и распределить роли, выбрав командира, инструктора-проводника, спасателей, корреспондента газеты, оформителя;

2) проложить маршрут спасения. Поиск маршрута осуществляет инструктор с помощью кубика согласно накладываемым требованиям:

- маршрут H_1 — четные грани;
- маршрут H_2 — грани, кратные трем;
- маршрут H_3 — грани, кратные пяти;

3) выбрать стратегию спасения, если известна метеосводка на ближайшее время в зоне поиска. На проложенных трех тропинках с запада на восток соответственно ожидается:

- селевой поток с вероятностью 0,8 на маршруте H_1;
- возможность оползней 25 % на маршруте H_2;
- сход снежных лавин с вероятностью 0,6 на маршруте H_3;

4) подсчитать вероятность спасения по проложенному маршруту;

5) известно, что пропавших альпинистов удалось спасти одной из групп спасателей. Какова вероятность того, что именно ваша группа спасения нашла пропавших альпинистов;

6) подготовить репортаж об участии вашей группы спасения в поиске горнолыжной экспедиции.

Задание 1.2. Придумайте социально значимую аналогичную игру по предложенному или собственному графу.

Задание 1.3. Обозначьте ветви первого яруса графа через H_i, где $i \in \mathbb{N}$.

а) Найдите формулу для вычисления вероятности события A при любом числе ветвей.

б) Известно, что событие A произошло. Найдите формулу для вычисления вероятности события H_i.

Задачи для самостоятельного решения

Элементы комбинаторики

1.1. *Вычислите устно*:

1) $\dfrac{8!-7!}{7!}$; 2) $\dfrac{5!+6!}{4!}$; 3) $\dfrac{P_7-P_5}{6!}$; 4) $\dfrac{P_9-8!}{P_7}$; 5) $\dfrac{6!-P_5}{5!}$; 6) $\dfrac{8!}{P_7+P_6}$;
7) $\dfrac{9!}{7!}\cdot\dfrac{7!}{8!}$.

1.2. *Сократите дроби*:

1) $\dfrac{(n-1)!}{n!}$; 2) $\dfrac{(n-1)!}{(n-3)!}$; 3) $\dfrac{A_9^3+A_9^2}{P_8}$; 4) $\dfrac{A_x^3}{A_x^2}$; 5) $\dfrac{2k(2k-1)!}{(2k)!}$; 6) $\dfrac{A_{n+1}^3}{A_n^2}$;
7) $\dfrac{(n-1)!\,n}{(n+1)n!}$.

1.3. *Решите уравнения*:

1) $\dfrac{C_{x+1}^{x-1}}{C_x^{x-3}}=\dfrac{4}{5}$; 2) $\dfrac{A_{k+1}^4P_{k-4}}{P_{k-1}}=15$; 3) $C_{x+1}^{x-4}=\dfrac{7}{15}A_{x+1}^3$; 4) $C_x^{x-3}+C_x^{x-4}=$
$=11C_{x+1}^2$; 5) $A_{2k}^3=100A_k^2$; 6) $C_k^3:C_k^5=2:3$.

1.4. *Решите задачи.*

1) В группе из 26 человек выбирают актив: старосту, физорга, профорга и культорга. Сколькими способами могут избрать актив группы?

2) Сколько различных спортивных прогнозов могут дать болельщики перед началом первенства по футболу, если в высшей лиге участвуют 15 команд и разыгрываются три медали: золотая, серебряная, бронзовая?

3) Сколькими способами в бригаде из шести операторов можно распределить три путевки в профилакторий, на турбазу и в дом отдыха?

4) Сколько существует вариантов, чтобы из букв слова «ученик» составить различные кортежи длиной 4?

5) Сколькими способами можно устроить на работу 8 выпускников факультета программирования программистами в пять различных вычислительных центров?

6) На железнодорожной станции имеется шесть запасных путей. Сколькими способами можно расставить на них четыре поезда?

7) Придумайте и решите аналогичную задачу.

1.5. *Решите задачи.*

1) Из 25 студентов группы на беседу с деканом пригласят пятерых. Сколькими способами это можно сделать?

2) Из восьми преподавателей математики нужно послать на курсы повышения квалификации трех человек. Сколькими способами это можно сделать?

3) Сколькими способами можно устроить на летнюю практику 10 студентов на три предприятия города?

4) Из 20 рабочих нужно выделить восемь человек для работы на новом участке. Сколькими способами это можно сделать?

5) В читальном зале 12 студентов дожидаются своих заказов на книги. Сколько способов обслужить читателей имеет библиотекарь, если до обеденного перерыва могут быть выполнены только пять заказов?

6) По сведениям геологоразведки, один из 12 участков земли может содержать нефть. Однако компания имеет средства для бурения только семи скважин. Сколько способов отбора для бурения семи различных скважин имеется у компании?

7) Придумайте и решите аналогичную задачу.

1.6. *Решите задачи.*

1) Компания имеет четыре отдела: производственный, снабжения, менеджмента и маркетинга. Количество людей в отделах 25, 36, 24 и 15 соответственно. Каждый отдел собирается послать одного представителя на ежегодную встречу с директором компании. Сколько различных групп для встречи можно составить из числа работников компании?

2) Ученый-химик желает исследовать эффект влияния на скорость химического процесса трех переменных: давления, температуры и катализаторов. Он намерен использовать три температуры, два значения давления и три типа катализаторов. Сколькими способами ученый может управлять реакцией, если пожелает использовать все возможные комбинации давления, температуры и типов катализаторов?

3) Авиакомпания совершает четыре рейса между Самарой и Москвой, а также два рейса между Москвой и Нью-Йорком. Сколькими способами можно заказать билет из Самары до Нью-Йорка, если рейсы осуществляются в разные дни?

4) От Самары до Москвы можно добраться на пяти различных поездах, а от Москвы до Одессы на четырех. Сколькими способами можно заказать билет из Самары до Одессы через Москву?

5) В колледже три спортивные секции — футбольная, баскетбольная и волейбольная. Спортсмены каждой секции были на соревнованиях в четырех различных городах России, в которых оставили о себе хорошие воспоминания и диск с фильмом о своем городе. Сколько дисков для этих поездок они имели?

6) В городе три района, в каждом из которых по пять филиалов магазинов сети «Магнит». Сколько существует маршрутов, по которым в конце рабочего дня денежные потоки попадают на единый счет в банке?

7) Придумайте и решите аналогичную задачу.

1.7. *Решите задачи.*

1) Сколько существует различных телефонных номеров, состоящих из шести цифр? Сколько существует таких телефонных номеров, в которых цифры не повторяются?

2) К космическим полетам готовятся 8 инженеров-исследователей, 10 врачей и 5 командиров экипажей. Сколько различных возможных экипажей можно составить из них для будущих полетов в космос, если в каждый экипаж входит один командир, три взаимозаменяемых инженера-исследователя и четыре врача разного профиля?

3) Слово «теория» составлено из букв разрезной азбуки. Наудачу извлекаются четыре карточки и складываются в ряд друг за другом в порядке появления. Сколько различных комбинаций можно составить из этих четырех букв? Сколько наборов из четырех букв можно составить из этого слова?

4) Директор предприятия рассматривает заявления о приеме на работу пяти выпускников колледжа. На предприятии имеются три вакансии. Сколькими способами директор может заполнить эти вакансии? Сколько существует способов приема на работу молодых специалистов, если эти вакансии на различные должности?

5) Из 25 студентов группы отбирают троих для участия в спортивных соревнованиях. Сколько имеется вариантов составить команду, если соревнования проходят по одному виду спорта? Сколько имеется вариантов составить команду, если соревнования проводят по различным видам спорта?

6) Среди 30 студентов группы разыгрываются четыре пригласительных билета в театр. Сколько вариантов имеется составить команду «счастливчиков», выигравших билет? Как изменится это число, если окажется, что билеты предлагают на различные спектакли?

7) Придумайте и решите аналогичную задачу.

1.8. *Решите задачи.*

1) В ящике 1 болт и 15 винтиков разных размеров. Нужно подобрать два винтика. Сколькими вариантами можно это сделать?

2) В ящике 7 болтов и 15 винтиков разных размеров. Нужно подобрать два болта и три винтика. Сколькими вариантами можно это сделать?

3) В школе олимпийского резерва обучаются 12 лыжников и 15 конькобежцев. Сколько имеется способов сформировать из них команду на соревнования по зимним видам спора, в которую должны войти три лыжника и четыре конькобежца?

4) Сколькими способами можно привезти в колледж 12 новых компьютеров на двух машинах, если на каждой машине можно разместить не более пяти ЭВМ?

5) Группу из 16 студентов должны разбить на подгруппы для работы в разных компьютерных классах. Сколько существует возможных вариантов формирования подгрупп, если в трех компьютерных классах соответственно 5, 4 и 7 работающих ЭВМ?

6) Из 15 красных и 7 белых гладиолусов формируют букеты. Сколькими способами можно составить букеты из четырех красных и трех белых гладиолусов?

7) Придумайте и решите аналогичную задачу.

1.9. *Решите задачи.*

1) Сколько существует способов поставить на книжную полку в беспорядке собрание сочинений, состоящее из семи томов?

2) Сколькими способами можно рассадить восемь гостей за круглым столом?

3) Сколько всевозможных кортежей длиной семь можно составить из слова «кислота»?

4) Сколько всевозможных кортежей длиной десять можно составить из букв слова «математика»?

5) На пяти одинаковых карточках написаны буквы. На трех из них — буква «и», на остальных — буква «л». Эти карточки выкладывают в ряд наудачу. Сколько вариантов получить при этом слово «лилии»?

6) Сколько всевозможных кортежей длиной десять можно составить из букв слова «программа»?

7) Придумайте и решите аналогичную задачу.

1.10. *Решите задачи.*

1) Сколько различных пятизначных чисел можно составить так, чтобы любые две соседние цифры были различными?

2) Четыре рассорившиеся подружки садятся в один поезд из восьми вагонов на одной и той же станции метро. Сколькими способами они могут проехать в этом поезде, не сев в один вагон?

3) Из цифр 3, 4, 5, 6 составлены четырехзначные числа. Сколько вариантов таких чисел можно найти, если среди найденных нет чисел, заканчивающихся на 36?

4) Сколькими способами можно построить кортежи из букв слова «грамматика»?

5) Сколькими способами можно поставить на полку четырехтомник В.Яна, двухтомник А.Ахматовой и трехтомник А.Блока так, чтобы книги каждого автора стояли рядом?

6) На полке стоят 10 книг, пять из них собрание сочинений Л.Н.Толстого. Сколько вариантов расстановки книг на полке существует, при условии что все пять томов Л.Н.Толстого должны стоять рядом?

1.11. *Решите задачи.*

1) Сколько различных трехзначных номеров для автомобилей одной серии можно составить из нечетных цифр?

2) Сколько различных трехзначных номеров для автомобилей одной серии можно составить из четных цифр?

3) В первенстве факультета по баскетболу участвуют восемь команд. Победители первенства занимают по одному первому, второму и третьему месту, а две команды, занявшие последние места, выбывают из дальнейших соревнований этого года. Сколько возможно различных вариантов результата первенства при таком условии?

4) Сколько различных неудачных попыток может сделать посторонний человек, подбирающий код к замку в автоматической камере хранения, если число цифр и букв на диске равно восьми, а всего шесть дисков?

5) Хватит ли кортежей длины не более четырех из знаков азбуки Морзе (• и –), чтобы передать все различные фразы (и числа) русского алфавита?

6) Сколько различных четырехзначных номеров одной серии для автомобилей можно составить из четных цифр?

1.12. *Решите задачи.*

1) Сколько двузначных чисел можно составить из нечетных цифр так, чтобы:

а) использовались любые из них;

б) цифры не повторялись;

в) использовались одинаковые цифры?

2) Сколько трехзначных номеров можно составить из нечетных цифр так, чтобы:

а) использовались любые из них;

б) цифры не повторялись;

в) использовались одинаковые цифры?

3) Сколько двузначных чисел можно составить из цифр, кратных трем так, чтобы:

а) использовались любые из них;

б) цифры не повторялись;

в) использовались одинаковые цифры?

4) Сколько трехзначных номеров можно составить из цифр 1, 2, 3, 4, 5, 6, 7 так, чтобы:

а) использовались любые из них;

б) цифры не повторялись;

в) использовались одинаковые цифры?

5) Сколько трехзначных номеров можно составить из цифр 3, 4, 5, 6, 7, 8, 9 так, чтобы:

а) использовались любые из них;

б) цифры не повторялись;

в) использовались одинаковые цифры?

6) Сколько двузначных чисел можно составить из цифр 3, 4, 5, 6, 7, 8 так, чтобы:

а) использовались любые из них;

б) цифры не повторялись;

в) использовались одинаковые цифры?

7) Придумайте и решите аналогичную задачу.

Виды случайных событий. Операции над событиями

1.13. *Решите задачи.* Подбрасывают два игральных кубика. Укажите количество событий, входящих в пространство элементарных событий этого испытания. Какие элементарные события *благоприятствуют* событию:

1) на обоих кубиках выпало одинаковое число очков;
2) сумма очков на обоих кубиках равна четырем;
3) сумма очков на обоих кубиках не превышает четырех;
4) сумма очков на обоих кубиках не менее четырех;
5) сумма очков на обоих кубиках менее четырех;
6) сумма очков на обоих кубиках равна 14?
7) Придумайте и решите аналогичную задачу.

1.14. *Решите задачи.* Подбрасывают два игральных кубика.

а) Укажите количество событий, входящих в пространство элементарных событий этого испытания.

б) Сколько элементарных событий *благоприятствуют* событию:

1) на обоих кубиках выпало максимальное число очков;
2) разность очков на обоих кубиках равна 4;
3) разность очков на обоих кубиках не превышает 4;
4) разность очков на обоих кубиках не менее 4;
5) разность очков на обоих кубиках менее 4;
6) разность очков на обоих кубиках равна 14?
7) Придумайте и решите аналогичную задачу.

1.15. *Решите задачи.* Подбрасывают три игральных кубика:

1) Укажите количество событий, входящих в пространство элементарных событий этого испытания.

2) Сформулируйте шесть различных вопросов об элементарных событиях и ответьте на них.

1.16. *Решите задачи.* Являются ли совместными следующие события:

1) при трехкратной стрельбе по мишени попадание не более одного раза и промах два раза;

2) двузначное число — четное и двузначное число кратно трем;

3) при трехкратном подбрасывании правильной монеты выпадение орла и выпадение цифры не менее трех раз;

4) при трехкратной стрельбе по мишени не менее одного попадания и промах при втором выстреле;

5) при подбрасывании двух правильных монет выпадение орла и выпадение решки не более одного раза;

6) при трехкратном подбрасывании правильной монеты однократное выпадение орла и выпадение решки не более двух раз?

1.17. *Приведите примеры*:

1) двух совместных событий;

2) двух несовместных событий;

3) двух равновозможных событий;

4) двух противоположных событий;

5) двух независимых событий;

6) двух зависимых событий.

1.18. *Решите задачи.* Перечислите события, образующие полную группу событий в следующих испытаниях:

1) при подбрасывании правильной игральной кости;

2) при подбрасывании двух правильных монет;

3) при подбрасывании правильной игральной кости и правильной монеты;

4) при подбрасывании двух правильных игральных костей;

5) при подбрасывании трех правильных монет;

6) при подбрасывании трех правильных игральных костей.

1.19. *Решите задачи.* Пусть A, B и C — три произвольных события. Придумайте задачу с сюжетом, содержащую вопрос, который соответствует некоторому событию M, описываемому заданной формулой:

1) $M = \bar{A}B\bar{C}$;

2) $M = \bar{A}\bar{B}\bar{C}$;

3) $M = \bar{A}B\bar{C} + \bar{A}\bar{B}C + A\bar{B}\bar{C}$;

4) $M = A + B + C$;

5) $M = \bar{A}BC + A\bar{B}C + AB\bar{C}$;

6) $M = AB + AC + BC$.

Определения вероятности. Некоторые теоремы теории вероятностей

1.20. *Решите задачи.*

1) Слово «спаниель» составлено из букв разрезной азбуки. Наудачу извлекают карточки и складывают в ряд друг за другом в порядке появления.

а) Сколько возможных соединений можно составить из букв этого слова?

б) Какова вероятность получить при этом слово «апельсин»?

в) Какова вероятность получить при этом слово «лес»?

2) Слово «ромашка» составлено из букв разрезной азбуки. Наудачу извлекают карточки и складывают в ряд друг за другом в порядке появления.

а) Сколько возможных соединений можно составить из букв этого слова?

б) Какова вероятность получить при этом слово «мошкара»?

в) Какова вероятность получить при этом слово «кара»?

3) Слово «соратница» составлено из букв разрезной азбуки. Наудачу извлекают карточки и складывают в ряд друг за другом в порядке появления.

а) Сколько возможных соединений можно составить из букв этого слова?

б) Какова вероятность получить при этом слово «стационар»?

в) Какова вероятность получить при этом слово «сорт»?

4) Слово «множества» составлено из букв разрезной азбуки. Наудачу извлекают карточки и складывают в ряд друг за другом в порядке появления.

а) Сколько возможных соединений можно составить из букв этого слова?

б) Какова вероятность получить при этом слово «жеманство»?

в) Какова вероятность получить при этом слово «межа»?

5) Слово «минотавр» составлено из букв разрезной азбуки. Наудачу извлекают карточки и складывают в ряд друг за другом в порядке появления.

а) Сколько возможных соединений можно составить из букв этого слова?

б) Какова вероятность получить при этом слово «норматив»?

в) Какова вероятность получить при этом слово «мотив»?

6) Слово «тарификация» составлено из букв разрезной азбуки. Наудачу извлекают карточки и складывают в ряд друг за другом в порядке появления.

а) Сколько возможных соединений можно составить из букв этого слова?

б) Какова вероятность получить при этом слово «ратификация»?

в) Какова вероятность получить при этом слово «акация»?

7) Составьте и решите аналогичную задачу со словом «коршун».

1.21. *Решите задачи*:

1) Какова вероятность того, что наудачу выбранное целое число от 40 до 70 является кратным 6?

2) Какова вероятность того, что наудачу выбранное целое число от 10 до 60 (включительно) кратно 15?

3) Какова вероятность того, что наудачу выбранное целое число от 1 до 30 (включительно) является делителем числа 30?

4) Какова вероятность того, что наудачу выбранное целое число от 20 до 80 (включительно) является делителем числа 80?

5) Какова вероятность того, что наудачу выбранное целое число от 30 до 90 (включительно) является делителем числа 30?

6) Какова вероятность того, что наудачу выбранное целое число от 10 до 70 (включительно) является делителем числа 35?

7) Придумайте и решите аналогичную задачу.

1.22. *Решите задачи.*

1) На участке теплосети длиной 800 м произошла авария. Какова вероятность того, что повреждение находится не далее 100 м от середины участка?

2) На плоскости проведены параллельные прямые, образующие полосы, расстояние между которыми чередуется попеременно и составляет 1,5 и 8 см. Какова вероятность того, что наудачу брошенный круг радиусом 2,5 см не пересечет ни одной из этих прямых?

3) В течение часа к причалу должны подойти два теплохода. Стоянка каждого теплохода у причала по расписанию составляет 20 мин. Определите вероятность встречи судов у причала, если моменты подхода их к причалу в течение часа независимы и равновозможны.

4) Две одинаковые монеты радиусом R' расположены внутри круга радиусом $R_{кр}$, в который наудачу бросают точку. Найти вероятность того, что эта точка попадет на одну из этих монет, если монеты не перекрываются.

5) Два друга условились встретиться в промежутке времени от 19 ч до 19 ч 30 мин, причем каждый обещал ждать другого не более 10 мин. Из двух событий «встреча произойдет» и «встреча не произойдет», какое наиболее вероятно?

6) Талон на прием к врачу рассчитан на 0,5 ч. Обычно на прием врач тратит не менее 10 мин. Пациент опаздывает, но других пациентов пока нет. Какова вероятность того, что пациент успеет на прием в отведенное для него время?

7) Придумайте и решите аналогичную задачу.

1.23. *Решите задачи.*

1) На карточке лотереи «Спортлото» необходимо зачеркнуть шесть из 49 возможных чисел от 1 до 49. В случае угадывания всех шести чисел есть шанс выиграть максимальную для этого тиража сумму денег. С какой вероятностью можно угадать все шесть номеров?

2) По сведениям геологоразведки, три из 15 участков земли, по всей вероятности, содержат газ. Однако компания имеет средства для бурения только девять скважин. Какова вероятность обнаружения всех трех месторождений газа?

3) Группа туристов, состоящая из 12 юношей и 8 девушек, выбирает по жребию хозяйственную команду в составе четырех человек. Какова вероятность того, что в числе избранных окажется двое юношей и две девушки?

4) На тепловой электростанции 15 сменных инженеров, из них три женщины. В смену заняты четыре человека. Найти вероятность того, что в случайно выбранную смену будут работать два мужчины и две женщины.

5) Фирма предлагает в продажу со склада партию из 15 компьютеров, 6 из которых — с дефектами. Колледж приобретает 7 из них, не зная о возможных дефектах. Найти вероятность того, что среди приобретенных компьютеров окажутся два с дефектами.

6) Для участия в судебном процессе из 20 потенциальных кандидатов, среди которых 6 женщин и 15 мужчин, выбирают 6 присяжных заседателей. Какова вероятность того, что после отбора в группе окажутся только две женщины?

7) Придумайте и решите аналогичную задачу.

1.24. *Решите задачи.*

1) На кафедре работает 12 преподавателей. С какой вероятностью дни рождения каждого из преподавателей придутся на разные месяцы года?

2) Замок имеет четырехзначный цифровой код. Наугад выбираются четыре цифры. Какова вероятность открыть замок, если известно, что все цифры различны?

3) Замок имеет четырехзначный цифровой код. Наугад выбираются четыре цифры. Какова вероятность открыть замок, если известно, что цифры в шифре могут повторяться?

4) Наудачу набранный номер состоит из пяти цифр. Найдите вероятность того, что все цифры в нем различные.

5) В награду за хорошую учебу 10 студентов группы получили билеты на концерт, при этом места расположены на одном ряду. Какова вероятность того, что у вас и вашего друга места окажутся рядом?

6) На стеллаже в библиотеке случайным образом расставлены 10 учебников, из которых шесть по теории вероятностей. Библиотекарь снимает с полки наудачу три учебника. Какова вероятность того, что хотя бы один из снятых учебников по теории вероятностей?

7) Придумайте и решите аналогичную задачу.

1.25. *Решите задачи (разными способами).*

1) В ходе исследования работы городского транспорта проводился социологический опрос пассажиров. Оказалось, что 64 % населения используют только муниципальный транспорт, а 38 % только частные микроавтобусы. Найдите вероятность того, что случайно выбранный пассажир пользуется любым из возможных видов транспорта.

2) В магазин компьютерной техники поставляются компьютеры из городов *A*, *B* и *C*. Вероятность получения компьютера из города *A* равна $p_1 = 0{,}5$, из *B* — $p_2 = 0{,}2$. Найдите вероятность того, что очередной компьютер будет получен из города *C*.

3) В научно-исследовательском институте работает 120 человек, из них 70 знают английский язык, 60 — немецкий язык, а 50 — английский и немецкий языки. Какова вероятность того, что выбранный наудачу сотрудник не знает ни одного языка?

4) Два станка работают независимо друг от друга, причем вероятность бесперебойной работы первого станка в течение определенного времени равна 0,8, а другого — 0,85. Какова вероятность выхода из строя обоих станков в течение данного промежутка времени?

5) Три стрелка независимо друг от друга стреляют по мишени. Вероятности их попадания соответственно равны 0,85; 0,7 и 0,9. Какова вероятность того, что промахнулись все три стрелка?

6) Известно, что в городе N 25 % населения смотрят рекламу по телевизору, 37 % населения слушают ее по радиоприемнику, причем 12 % населения как слушают радиорекламу, так и смотрят телерекламу. Чему равна вероятность того, что случайно отобранный житель города N знаком по крайней мере хотя бы с одной из этих рекламных передач?

7) Придумайте и решите аналогичную задачу.

Сложение совместных событий

1.26. *Решите задачи.*

1) Правовой центр получил приглашение для оказания юридических услуг от двух фирм. Вероятность получения заказа от фирмы *A* предположительно 0,54, от фирмы *B* — 0,62. С какой вероятностью адвокаты правового центра получат не менее одного заказа?

2) Найдите вероятность того, что взятое наудачу двузначное число окажется кратным либо 3, либо 5, либо 3 и 5 одновременно.

3) Два станка работают независимо друг от друга. Вероятность бесперебойной работы первого станка в течение времени *t* равна 0,9, второго — 0,8. Определите вероятность бесперебойной работы хотя бы одного из двух станков в течение времени *t*.

4) Электронный прибор состоит из двух последовательно включенных блоков. Вероятность выхода из строя за 1 месяц работы первого блока равна 1/3, второго — 1/4, обоих — 1/6. Найдите вероятность безаварийной работы прибора в течение месяца.

5) Во время учебных маневров два танка пытаются прорваться в расположение противотанковой обороны противника. Какова вероятность того, что подбит хотя бы один танк, если вероятность подбить один танк равна 2/3, два танка — 2/5?

6) В целях изучения спроса в большом универмаге установили камеру для подсчета числа входящих покупателей. Если в магазин входят одновременно два покупателя, причем один идет перед другим, то первый из них будет учтен электронным устройством с вероятностью 0,96, второй — с вероятностью 0,93, оба — с вероятностью 0,91. Найдите вероятность того, что устройство ска-

нирует по крайней мере одного из двух входящих вместе покупателей.

7) Придумайте и решите аналогичную задачу.

1.27. *Решите задачи.*

1) В компьютере одновременно работают две независимые программы. Вероятность того, что первая программа даст сбой, составляет 0,3, а вторая — 0,4. Найти вероятность того, что:

а) обе программы дадут сбой;

б) сбой произошел;

в) обе программы не дадут сбоя;

г) хотя бы одна программа даст сбой;

д) хотя бы одна будет работать без сбоя;

е) только одна программа даст сбой;

ж) будет не менее одного сбоя.

2) Вероятность того, что потребитель увидит рекламу ультразвуковой стиральной машинки по телевидению, равна 0,7. Вероятность того, что потребитель увидит эту рекламу на рекламном стенде, равна 0,4. Найдите вероятность того, что потребитель увидит:

а) обе рекламы, если учесть, что оба события независимы;

б) хотя бы одну рекламу (эффективность рекламы);

в) только одну рекламу;

г) не более одной рекламы;

д) не менее одной рекламы;

е) менее одной рекламы;

ж) более одной рекламы.

3) В автопробеге участвуют три автомобиля, причем первый может сойти с маршрута с вероятностью 0,16, второй — с вероятностью 0,12, третий — с вероятностью 0,2. Определите вероятность того, что к финишу:

а) прибудут все автомобили;

б) прибудут два автомобиля;

в) прибудут по крайней мере два автомобиля;

г) прибудет не более двух автомобилей;

д) прибудет только один автомобиль;

е) прибудут не менее трех автомобилей;

ж) прибудет более трех автомобилей.

4) Для фирмы, занимающейся строительством терминалов для аэропортов, вероятность получить контракт в одной из европейских стран равна 0,5, а в одной из азиатских стран — 0,7. Найдите вероятность того, что фирма:

а) получит контракт хотя бы в одной стране;

б) получит контракт только в одной из стран;

в) получит контракт не более чем в одной стране;

г) получит контракт не менее чем в одной стране;

д) получит контракт или в одной из европейских стран, или в одной из азиатских стран;

е) не получит контракт;

ж) получит контракт и в одной из европейских стран, и в одной из азиатских стран.

5) Вероятность того, что покупатель, собирающийся приобрести компьютер и пакет прикладных программ, приобретет только компьютер, равна 0,35. Вероятность, что покупатель купит только пакет программ, равна 0,2. Найдите вероятность того, что:

а) будут куплены или компьютер, или пакет программ, или компьютер и пакет программ вместе;

б) будут куплены либо компьютер, либо пакет прикладных программ;

в) покупатель откажется от приобретений;

г) покупатель совершит только одну покупку;

д) покупатель совершит не менее одной покупки.

6) По предварительной оценке некоторой брокерской фирмы, 35 % потенциальных вкладчиков покупают акции, а 26 % покупают облигации, с которыми имеет дело эта фирма. Некий бизнесмен интересуется деятельностью этой брокерской фирмы. С какой вероятностью он:

а) купит и акции, и облигации;

б) купит только акции или только облигации;

в) не станет иметь деловых контактов с этой брокерской фирмой;

г) обязательно будет иметь деловые контакты с этой брокерской фирмой?

7) Придумайте и решите аналогичную задачу.

1.28. *Решите задачи.*

1) Два программиста пишут компьютерные программы. Вероятность того, что программа, составленная первым программистом, будет работать, равна 0,5; вторым — 0,7. Сформулируйте четыре вопроса о вычислении вероятности по заданному условию и решите задачу.

2) В библиотеке стоят три компьютера. Вероятность исправной работы первого компьютера равна 0,7; второго — 0,8; третьего — 0,5. Сформулируйте четыре вопроса о вычислении вероятности по заданному условию и решите задачу.

1.29. *Решите задачи.*

1) За компьютерами в компьютерном классе сидят 10 юношей и 5 девушек. Сломались три компьютера. Найдите вероятность того, что все, у кого сломались компьютеры, юноши.

2) Универсам размещает свою рекламу на местном телеканале, считая, что 70 % телезрителей этого канала — потенциальные клиенты магазина. Социологический опрос показал, что 73 % телезрителей помнят о рекламе универсама. Оцените, чему равен процент людей, которые являются потенциальными покупателями универсама и могут вспомнить его рекламу.

3) Социологический опрос абитуриентов показал, что 45 % из них выбирают специальность программиста из-за того, что она востребована на рынке труда, 65 % — из-за родителей, а 36 % — из-за того и другого вместе. Основываясь на результатах социологического опроса, ответьте, являются ли два вида предпочтений абитуриентов независимыми друг от друга? Объясните свой ответ.

4) Вновь воспользуемся результатами социологического опроса, описанного в предыдущей задаче. Пусть из группы абитуриентов случайно выбраны трое. Чему равна вероятность того, что все трое выбирают специальность программиста потому, что прислушались к мнению родителей?

5) Во время маркетинговых исследований наибольшие проблемы возникают в двух ситуациях: при отказе граждан заполнять анкеты и при их отсутствии дома. Предыдущие исследования показали, что вероятность застать граждан дома ровна 0,65. В то же время исследователи надеются, что 92 % опрошенных граждан согласятся заполнять анкеты. Основываясь на этой информации, определите процент заполненных анкет.

6) Вероятность роста цен на потребительские товары в следующем месяце составляет 0,4, вероятность роста цен на нефть — 0,3, а вероятность одновременного роста цен на нефть и на потребительские товары — 0,12. Основываясь на этой информации, определите, являются ли цены на нефть и потребительские товары независимыми друг от друга. Объясните свой ответ.

7) Придумайте и решите аналогичную задачу.

1.30. *Решите задачи.*

1) Консультационная фирма получила приглашение на выполнение двух заказов от немецкой и финской корпораций. Правление фирмы оценивает в 80 % вероятность реального получения заказа от немецкой корпорации и в 40 % вероятность получения заказа от обеих корпораций. Оцените вероятность реального получения заказа от финской корпорации.

2) Правление кредитного банка оценивает вероятность невозвращения кредита в течение пяти лет обанкротившимися фирмами в 15 %. В то же время известно, что 30 % кредитовавшихся в банке фирм обанкротились. Оцените вероятность, с которой недавно обанкротившийся клиент этого банка не сможет в течение пяти лет вернуть банку долг.

3) Финансовый аналитик прогнозирует, что если за следующий квартал процентная ставка упадет не менее чем на 0,5 %, то с вероятностью 75 % выпуск продукции некоторой отрасли будет расти. В то же время он считает, что за этот период с вероятностью 25 % процентная ставка может упасть не менее чем на 0,5 %. Оцените вероятность падения процентной ставки за этот период при одновременном росте выпуска продукции этой отраслью.

4) Строительная компания размещает рекламу в местной газете. По мнению менеджеров этой компании, 10 % читателей газеты являются потенциальными клиентами этой строительной компании. Социологическое исследование, проведенное строительной компанией, показало, что 70 % ее клиентов узнали о компании через местную газету. С какой вероятностью читатели обращают внимание на рекламу этой строительной компании на страницах газеты?

5) Вероятность того, что выпускник колледжа продолжит образование в вузе, равна 0,8. Вероятность того, что он продолжит образование в вузе и одновременно начнет работать по специальности, равна 0,6. С какой вероятностью выпускник колледжа после его окончания начнет работать по специальности?

6) Отдел маркетинга ВАЗа изучает спрос потенциальных покупателей легковых автомобилей. По результатам социологического исследования, 40 % потенциальных покупателей отдают предпочтение техническим характеристикам, а 35 % кроме технических характеристик считают важным возможность сервисного обслуживания. Оцените в результатах этого социологического исследования долю потенциальных покупателей, которые отдают предпочтение возможности сервисного обслуживания.

7) Придумайте и решите аналогичную задачу.

Формула полной вероятности. Формула Байеса

1.31. *Решите задачи.* Пусть в трех урнах находятся шары черного и белого цвета. Введем обозначения:

H_i — выбрали i-ю урну (гипотезы);

A — вынули из урны белый шар (событие);

$\bar{A}$ — вынули из урны черный шар.

Запишите с помощью символов события:

1) выбрали либо первую, либо вторую урну;

2) выбрали какую-либо урну, но не третью;

3) вынули белый шар из третьей урны;

4) вынули черный шар из второй урны;

5) выбрали третью урну и вынули из нее черный шар;

6) выбрали одну из трех урн и вынули из нее черный шар.

7) Придумайте и решите аналогичную задачу.

1.32. Дайте словесную трактовку (расшифруйте) события, записанного с помощью символов (см. рис. 1.1):

1) AH_2;

2) $\bar{A}H_3$;

3) $\bar{A}H_1 + \bar{A}H_2 + \bar{A}H_3$;

4) $\bar{A}H_3$;

5) $\bar{A}/H_2$;

6) $\bar{A}/H_3$.

7) Придумайте и решите аналогичную задачу.

1.32. *Решите задачи.* Известно, что в каждой из трех урн по одному черному и одному белому шару. Запишите события символически и найдите вероятности событий:

1) извлечь белый шар из третьей урны;

2) извлечь черный шар из второй урны;

3) выбрать первую урну и затем вынуть из нее белый шар;

4) выбрать первую или третью урну и затем достать из нее черный шар;

5) достать черный шар, который оказался из второй урны;

6) достать белый шар, который оказался не из первой урны.

7) Придумайте и решите аналогичную задачу.

1.33. *Решите задачи.*

1) Имеется три урны. В первой находится пять белых и три черных шаров, во второй — шесть белых и два черных, в третьей — десять белых. Вынимают наугад один шар. Урна выбирается тоже наугад. Найдите вероятность того, что этот шар белый.

2) В ящике находятся детали, из которых 12 изготовлены на первом станке, 20 — на втором и 16 — на третьем. Вероятности того, что детали, изготовленные на первом, втором и третьем станках, стандартные, соответственно равны 0,9; 0,8 и 0,6. Найдите вероятность того, что взятая наугад деталь окажется стандартной.

3) Результаты статистических исследований в медицине показывают, что если пациент болен некоторым инфекционным заболеванием, то тест даст положительный результат для 90 %, а если не болен, то тест может дать положительный результат для 7 % проверяемых. Этому виду инфекции, согласно статистическим исследованиям, подвержено только 0,3 % населения. Пусть некоторому случайно выбранному пациенту сделан анализ и получен положительный результат. Найдите вероятность того, что он действительно заражен этим видом инфекции.

4) Судоходная компания в течение лета организует круизы по Волге. Очевидно, что наибольшую прибыль компания получит, если каюты корабля будут полностью заняты туристами. Представитель туристического агентства, сотрудничающий с компанией, предсказывает, что если курс доллара не повысится, то вероятность того, что корабль будет заполнен в течение сезона, равна 0,93, в противном случае — с вероятностью 0,85. По оценкам экономистов, вероятность того, что в течение сезона доллар подорожает по отношению к рублю, равна 0,26. Найти вероятность того, что билеты на все круизы будут проданы.

5) В группе учащихся из 30 человек 12 юношей, а остальные — девушки. Половина юношей и треть девушек живут в общежитии.

Найдите вероятность того, что случайно выбранный учащийся группы живет в общежитии.

6) В вычислительной лаборатории имеется шесть клавишных автоматов и четыре полуавтомата. Вероятность того, что за время выполнения некоторого расчета автомат не выйдет из строя, равна 0,95, для полуавтомата вероятность равна 0,8. Студент производит расчет на машине, выбранной наудачу. Найдите вероятность того, что до окончания расчета машина выйдет из строя.

7) Придумайте и решите аналогичную задачу.

1.34. *Решите задачи.*

1) Из пункта A в пункт B можно добраться тремя маршрутами. Водитель выбирает дорогу наугад. Если он поедет по первому маршруту, то вероятность того, что он попадет в пункт B за сутки, равна 0,6; по второму — 0,3; по третьему — 0,1.

а) Найдите вероятность того, что водитель приедет в пункт B в течение суток.

б) Водитель приехал в пункт B в течение суток. Какова вероятность того, что он ехал по второму маршруту?

2) Железнодорожный билет до Москвы можно купить в одной из трех касс. Вероятность купить билет в первой кассе равна 1/2, во второй — 1/3, в третьей — 1/6. Вероятность того, что билетов в кассе уже нет, составляет для первой кассы 1/8, для второй — 1/6, для третьей — 1/4.

а) Какова вероятность того, что билет до Москвы удалось приобрести?

б) Билет до Москвы купить удалось. В какой кассе вероятнее всего был приобретен билет?

3) Известно, что 5 % мужчин и 0,25 % женщин — дальтоники. Будем считать, что количество мужчин и женщин одинаково. В кабинет к врачу-окулисту должен зайти очередной пациент.

а) Какова вероятность того, что он страдает дальтонизмом?

б) Очередной пациент страдает дальтонизмом. Какова вероятность того, что это мужчина?

4) Телеграфное сообщение состоит из сигналов «точка» и «тире». Статистические свойства помех таковы, что искажаются в среднем 2/5 сообщений «точка» и 1/3 сообщений «тире». Известно, что среди передаваемых сигналов «точка» и «тире» встречаются в отношении 5 : 3.

а) С какой вероятностью можно принять сигнал «точка»?

б) Определите вероятность того, что принят передаваемый сигнал, если принят сигнал «точка».

5) В поисках учебника по дискретной математике студент решил обойти три библиотеки. В каждой из библиотек равновероятно, что книга есть в ее фонде или ее нет. Если книга в библиотечном фонде имеется, то равновероятно, что она либо на руках у читателей, либо на библиотечной полке.

а) Что более вероятно: достанет студент учебник или нет, если библиотеки комплектуются самостоятельно (независимо друг от друга)?

б) Учебник по дискретной математике получен. Какова вероятность того, что он из третьей библиотеки?

6) Рабочий обслуживает три станка, на которых обрабатывается равное количество однотипных деталей. На первом станке было изготовлено 70 % деталей без брака, на втором — 80 % и на третьем — 90 %. На склад поступили детали только с этих трех станков.

а) Определите вероятность того, что деталь, взятая наугад со склада, окажется без брака.

б) Взятая наугад со склада деталь оказалась без брака. Определите вероятность того, что эта деталь изготовлена на втором станке.

7) Придумайте и решите аналогичную задачу.

1.35. *Решите задачи.*

1) По линии связи передается кодированный с помощью букв A, B, C текст. Вероятности передачи отдельных букв таковы: $p(A) = 0{,}5$; $p(B) = 0{,}3$; $p(C) = 0{,}2$. Вероятности искажения при передаче отдельных букв равны соответственно: 0,01; 0,03; 0,02. Установлено, что сигнал из двух букв принят без искажений. Чему равна вероятность того, что передавался сигнал AB?

Указание. Найдите полную вероятность события «сигнал из двух букв принят без искажений».

2) Из 18 стрелков пять попадают в мишень с вероятностью 0,8; семь — с вероятностью 0,7; четыре — с вероятностью 0,6 и два — с вероятностью 0,5. Наудачу выбранный стрелок произвел выстрел, но в мишень не попал. К какой из групп вероятнее всего принадлежал этот стрелок?

Указание. Найдите полную вероятность события «выбранный наугад стрелок промахнулся» при гипотезах, что он из первой, второй, третьей или четвертой групп. Найдите, какая из гипотез наиболее вероятна.

3) У рыбака есть три излюбленных места для рыбалки. Эти места он посещает с одинаковой вероятностью. Вероятность того, что при одном забрасывании удочки рыба клюнет в первом месте, близка к 1/3; во втором — 1/2; в третьем — 1/4. Известно, что рыбак забрасывал удочку три раза, а вытащил только одну рыбу. Какова вероятность того, что он рыбачил в первом из излюбленных мест?

Указание. Найдите полную вероятность события «рыбак поймал одну рыбу, три раза забросив удочку» при гипотезах, что он рыбачил в первом, втором или третьем местах.

4) Из партии в пять изделий наугад взято одно изделие, оказавшееся бракованным. Количество бракованных изделий равновозможно любое. Какое предположение о количестве бракованных изделий наиболее вероятно?

Указание. Найдите полную вероятность события «взятое наугад изделие бракованное» при гипотезах, что в партии одно, два, три, четыре, пять бракованных изделий. Определите, какая из гипотез имеет наибольшую вероятность.

5) Определите вероятность того, что среди 1 000 лампочек нет ни одной неисправной, если из взятых наугад 100 лампочек все оказались исправными. Предполагается, что число неисправных лампочек из 1 000 штук равновозможно от 0 до 5.

Указание. Найдите полную вероятность события «выбрано 100 исправных лампочек» при гипотезах, что в партии имеется 0, 1, 2, 3, 4, 5 неисправных лампочек. Найдите вероятность гипотезы «в партии 0 неисправных лампочек».

6) На вход радиолокационного устройства с вероятностью p поступает полезный сигнал с помехой, а с вероятностью $(1-p)$ — только одна помеха. Если поступает полезный сигнал с помехой, то устройство регистрирует наличие какого-то сигнала с вероятностью p_1; если только помеха — с вероятностью p_2. Известно, что устройство зарегистрировало наличие какого-то сигнала. Найдите вероятность того, что в его составе имеется полезный сигнал.

Указание. Найти полную вероятность события «зарегистрирован какой-то сигнал».

7) Придумайте и решите аналогичную задачу.

1.36. *Решите задачи.*

1) Стрелки A, B, C попадают в мишень с вероятностью соответственно $p(A)=0{,}6$; $p(B)=0{,}5$ и $p(C)=0{,}4$. После первого выстрела в мишень попали две пули из трех.

а) Определите, что вероятнее: стрелок C попал по мишени или нет.

б) Мишень оказалась пораженной. С какой вероятностью в нее попал первый стрелок A?

2) Вероятность того, что клиент банка не вернет заем в период экономического роста, равна 0,06, в период экономического кризиса — 0,18. По прогнозам аналитиков, вероятность того, что начнется период экономического роста, равна 0,56. Определите, что вероятнее: случайно выбранный клиент банка вернет или не вернет полученный кредит.

3) На заводе изготовляют болты. На станках A, B, C производят соответственно 25, 35 и 40 % всех болтов, брак составляет соответственно 5, 4 и 2 %. Случайно выбранный болт оказался дефек-

тным. Какова вероятность того, что он был изготовлен на станке A, на станке B, на станке C?

4) Экономист-аналитик условно подразделяет экономическую ситуацию в стране на «хорошую», «посредственную» и «плохую» и оценивает их вероятности для данного момента времени в 0,15; 0,60 и 0,25 соответственно. Платежеспособность населения возрастает с вероятностью 0,7, когда ситуация «хорошая», с вероятностью 0,25, когда ситуация «посредственная», и с вероятностью 0,2, когда ситуация «плохая». Известно, что в настоящий момент платежеспособность населения изменилась. Какое из состояний экономики страны наиболее вероятно в этих условиях?

5) Фирма собирается заключить контракт на поставку оборудования в одну из стран СНГ. Если конкуренты фирмы не станут одновременно претендовать на заключение аналогичного контракта, то вероятность его получения оценивается в 0,4, в противном случае — в 0,3. По оценкам экспертов, вероятность того, что конкуренты выдвинут свои предложения по заключению контракта, равна 0,35. Определите, что вероятнее: фирма заключит этот контракт или нет.

6) Агент по недвижимости пытается продать участок земли под застройку. Он полагает, что если экономическая ситуация в регионе не будет ухудшаться, то участок будет продан в ближайшие шесть месяцев с вероятностью 0,8; в противном случае вероятность продажи участка уменьшится до 0,6. Экономист, консультирующий агента, предсказывает с вероятностью, равной 0,3, что экономическая ситуация в регионе в течение следующих шести месяцев будет ухудшаться. Определите, что вероятнее: в течение ближайших шести месяцев участок будет продан или его не удастся продать.

7) Придумайте и решите аналогичную задачу.

1.37. ***Решите задачи (самостоятельно сформулируйте вопросы по вычислению вероятности гипотез).***

1) Исходя из результатов предыдущих исследований специалисты по нефтеразведке считают, что вероятность наличия нефти в месте предполагаемого бурения скважины равна 0,4. На завершающем этапе разведки проводится сейсмический тест, который имеет определенную степень надежности: если на проверяемом участке есть нефть, то результат теста будет положительным в 85 % случаев; если нефти нет, то в 12 % случаев результат теста может быть ошибочно положительным. Сейсмический тест указал наличие нефти.

2) На заводе установлена система аварийной сигнализации. Когда возникает аварийная ситуация, звуковой сигнал срабатывает с вероятностью 0,9. Звуковой сигнал может сработать случайно и без аварийной ситуации с вероятностью 0,05. Реальная веро-

ятность аварийной ситуации равна 0,006. Известно, что звуковой сигнал сработал.

3) Маркетинговые исследования показали, что для некоторого товара известно, что его успех на рынке возможен с вероятностью 0,65, если товар действительно удачный, и с вероятностью 0,25 в случае, если он неудачен. Из предыдущего опыта известно, что новый товар может иметь успех на рынке с вероятностью 0,65. Новый товар прошел выборочную проверку и имел успех.

4) Два автомата производят одинаковые детали, которые поступают на общий конвейер. Производительность первого автомата вдвое больше производительности второго автомата. Первый автомат производит в среднем 65 % деталей отличного качества, а второй — 82 % деталей отличного качества. Наугад взятая с конвейера деталь оказалась отличного качества.

5) Охотник произвел три выстрела по лисе. Вероятности убить лису первым, вторым, третьим выстрелом соответственно равны 40, 50 и 70 %. Одним попаданием лису можно убить с вероятностью 0,2, двумя — 0,6, тремя — наверняка. Лиса убита.

6) В трех группах программистов учится по 20 студентов в каждой: в первой — 18 юношей и две девушки, во второй — 12 юношей и восемь девушек, в третьей — 15 юношей и пять девушек. Вероятность выбора студентов на олимпиаду по программированию в первой группе в два раза больше, чем во второй, а во второй — в три раза больше, чем в третьей. Участвовал в олимпиаде юноша.

Схема Бернулли

1.38. *Решите задачи.*

1) Для данного баскетболиста вероятность попадания мяча в кольцо при каждом броске равна 0,4. Определите наиболее вероятную ситуацию — попадание трех мячей при четырех бросках мяча или попадание четырех мячей при пяти бросках мяча, если броски считаются независимыми.

2) Для данного баскетболиста вероятность попадания мяча в кольцо равна 0,6. Баскетболист выполнил серию из четырех бросков. Какова вероятность того, что при этом было ровно три попадания?

3) Среди коконов некоторой партии содержится 20 % цветных. Какова вероятность того, что среди шести случайно отобранных из партии коконов четыре окажутся цветными?

4) Вероятность того, что замаскированный «противник» находится на обстреливаемом участке, равна 0,3; вероятность попадания в этом случае при каждом отдельном выстреле равна 0,2. Для поражения достаточно одного попадания. Какова вероятность поражения при двух выстрелах?

5) Применяемый метод лечения приводит к выздоровлению в 90 % случаев. Какова вероятность того, что из пяти больных поправится не менее четырех человек?

6) Случайно встреченный человек с вероятностью 0,2 может оказаться брюнетом, с вероятностью 0,3 — шатеном, с вероятностью 0,4 — блондином, с вероятностью 0,1 — рыжим. Какова вероятность того, что среди случайно встреченных лиц окажется:

а) не менее четырех блондинов;

б) хотя бы один рыжий;

в) три блондина и три брюнета?

7) Придумайте и решите аналогичную задачу.

1.39. *Решите задачи.*

1) Среди семян пшеницы 0,6 % семян сорняков. Какова вероятность при случайном отборе 1000 семян обнаружить:

а) не менее трех семян сорняков;

б) не более 16 семян сорняков;

в) ровно шесть семян сорняков;

г) только семена пшеницы?

2) Книга в 500 страниц содержит 50 опечаток. Оцените вероятность того, что на случайно выбранной странице:

а) не менее трех опечаток;

б) не более восьми опечаток;

в) ровно четыре опечатки;

г) ни одной опечатки.

3) Вероятность быть зарегистрированной счетчиком у частицы, вылетевшей из радиоактивного источника, равна 1/10000. Предположим, что за время наблюдения из источника вылетело 30000 частиц. Какова вероятность того, что счетчик:

а) зарегистрировал более 10 частиц;

б) не зарегистрировал ни одной частицы;

в) зарегистрировал ровно три частицы;

г) зарегистрировал менее трех частиц?

4) В ходе аудиторской проверки фирмы случайным образом были отобраны пять счетов. Согласно статистике, известно, что 2 % счетов содержат ошибки. Найдите вероятность того, что аудитор найдет:

а) только один счет с ошибкой;

б) хотя бы один счет с ошибкой;

в) не более одного счета с ошибкой;

г) все счета без ошибок.

5) Продавец ювелирного магазина заметил, что вероятность продажи украшения при единичном контакте с покупателем равна приблизительно 0,02. В течение рабочего дня к продавцу обратилось 100 посетителей. Чему равна вероятность того, что:

а) он продал ровно одно изделие;

б) ни одно изделие не продано;

в) он продал по крайней мере одно изделие;

г) он продал не более одного изделия?

6) В столовой сварили компот из 300 абрикосов и разлили его в 450 стаканов. С какой вероятностью в вашем стакане:

а) окажутся два абрикоса;

б) не окажется ни одного абрикоса;

в) окажется не менее двух абрикосов;

г) окажется не более двух абрикосов?

7) Придумайте и решите аналогичную задачу.

1.40. *Решите задачи.*

1) Партия из 40 деталей содержит пять бракованных. Какова вероятность того, что среди 10 случайно взятых деталей:

а) бракованных нет;

б) две детали бракованные;

в) не менее двух бракованных деталей;

г) не более двух бракованных деталей;

д) менее двух бракованных деталей;

е) более двух деталей с браком?

2) В студенческой группе из 24 человек шестеро занимаются спортом. Какова вероятность того, что среди семи пришедших на спортивный праздник студента будет насчитываться спортсменов:

а) ни одного;

б) ровно четыре;

в) более четырех;

г) не более четырех;

д) менее четырех;

е) не менее четырех?

3) Устройство состоит из трех независимо работающих основных элементов. Устройство перестает работать, если отказывает хотя бы один элемент. Вероятность отказа каждого из элементов за некоторое время t равна 0,2. Найдите вероятность безотказной работы устройства за этот период, если:

а) работают все основные элементы;

б) отказали два основных элемента;

в) отказали не менее двух основных элементов;

г) отказали менее двух основных элементов;

д) отказали не более двух основных элементов;

е) отказали более двух основных элементов при условии, что каждый неисправный элемент замещается резервными с той же вероятностью выйти из строя.

4) Два шахматиста играют 10 результативных партий (без ничьих). Для первого игрока вероятность выиграть составляет 2/3, а для второго 1/3 в каждой партии. Чему равна вероятность выигрыша:

а) первым игроком во всех партиях;

б) первым игроком половины партий;

в) вторым игроком двух партий;

г) вторым игроком не менее двух партий;

д) вторым игроком более двух партий;

е) вторым игроком не более двух партий?

5) Для прядения смешаны поровну белый и окрашенный хлопок. Какова вероятность того, что среди пяти случайно отобранных волокон смеси обнаружатся:

а) ровно два окрашенных;

б) ровно три окрашенных;

в) не менее двух окрашенных?

6) Контролер проверяет изделия, каждое из которых независимо от других может с вероятностью p оказаться дефектным. Какова вероятность того, что из десяти проверенных изделий:

а) только одно оказалось дефектным;

б) только два оказались дефектными;

в) не более двух оказались дефектными;

г) не более восьми оказались дефектными?

7) Придумайте и решите аналогичную задачу.

1.41. *Решите задачи.*

1) Медиками установлено, что 94 % лиц, которым сделаны прививки против туберкулеза, приобретают иммунитет против этого заболевания. Какова вероятность того, что среди 100 000 граждан, сделавших прививки, 5 800 не защищены от заболевания туберкулезом?

2) Вероятность встретить на улице однокурсника равна 0,002. Какова вероятность того, что среди 1 200 случайных прохожих вы встретите:

а) трех однокурсников;

б) не менее трех однокурсников;

в) более трех однокурсников;

г) не более трех однокурсников;

д) менее трех однокурсников;

е) ни одного однокурсника?

3) Вероятность выздоровления больного в результате нового способа лечения гриппа возросла до 0,8. Какова вероятность того, что во время эпидемии из 1000 больных с одного участка вылечатся и не получат осложнений:

а) 15 больных;

б) не менее 15 больных;

в) менее 15 больных;

г) более 15 больных;

д) не более 15 больных;

е) все больные этого участка?

4) Баскетболист забрасывает штрафной мяч с вероятностью 0,75. Какова вероятность того, что из 30 бросков будет:

а) 25 попаданий;
б) не менее 25 попаданий;
в) менее 25 попаданий;
г) более 25 попаданий;
д) не более 25 попаданий;
е) 30 попаданий?

5) По данным ремонтной мастерской, в течение гарантийного срока выходят из строя в среднем 12 % кинескопов. Какова вероятность того, что из 50 наугад выбранных кинескопов проработают гарантийный срок:

а) 47 кинескопов;
б) не менее 47 кинескопов;
в) менее 47 кинескопов;
г) более чем 47 кинескопов;
д) не более 47 кинескопов;
е) 50 кинескопов?

6) С помощью зенитной установки обстреливают мишень. Вероятность попадания в цель составляет 0,7. Какова вероятность того, что из 80 произведенных на штабных учениях выстрелов достигнут цели:

а) 75 выстрелов;
б) не менее 75 выстрелов;
в) менее 75 выстрелов;
г) не более 75 выстрелов;
д) более 75 выстрелов;
е) все выстрелы?

7) Придумайте и решите аналогичную задачу.

Наивероятнейшее число наступления события

1.42. *Решите задачи.*

1) Сколько изюма в среднем должны содержать калорийные булочки для того, чтобы вероятность иметь в булочке хотя бы одну изюмину была не менее 0,99?

2) Известно, что вероятность выпуска сверла повышенной хрупкости (брак) равна 0,02. Сверла укладывают в коробки по 100 штук. Какое наименьшее количество сверл нужно класть в коробку для того, чтобы с вероятностью, не меньшей 0,9, в ней было не менее 1000 исправных?

3) Пусть вероятность быть зарегистрированной счетчиком у частицы, вылетевшей из радиоактивного источника, равна 1/10 000. Предположим, что за время наблюдения из источника вылетело 30 000 частиц. Какое наименьшее число частиц должно вылететь из источника для того, чтобы с вероятностью, большей 0,99, счетчик зарегистрировал более трех частиц?

4) В магазин зашли 12 покупателей. Вероятность того, что любой из них не уйдет без покупки, — 0,2. Какова вероятность того, что покупку сделают четыре покупателя?

5) Из всей продукции обувной фабрики 42 % составляют изделия высшего сорта. Сколько пар сапог высшего сорта можно надеяться найти среди 75 пар сапог, поступивших с этой обувной фабрики в магазин?

6) С помощью станка-автомата изготовлено 90 деталей. Вероятность изготовления деталей высшего сорта составляет 78 %. Какое среднее число деталей первого сорта ожидается получить на этом станке, если 95 % продукции составляют детали первого и высшего сортов?

7) Придумайте и решите аналогичную задачу.

1.43. *Решите задачи.*

1) Станок-автомат изготовляет 80 деталей в день. Какова вероятность того, что изготовленная деталь высшего сорта, если обычно в течение рабочей смены 68 деталей высшего сорта?

2) Обычно число ясных дней в ноябре в нашей местности равно 12. В этом году синоптики предсказали на ноябрь 10 ясных дней. Какова степень вероятности прогноза?

3) Обычно в партии из 20 принтеров три принтера бракованные. Поставщик гарантирует, что в новой партии наиболее вероятен брак у двух принтеров. Какова степень гарантии поставщика?

4) Вероятность попадания в цель при одном выстреле равна 0,8. Найдите наиболее вероятное число попаданий при пяти выстрелах и соответствующую этому вероятность.

5) Вероятность того, что покупателю нужна мужская обувь 42 размера, равна 0,2. Найдите наиболее вероятное число покупателей, интересующихся 42 размером обуви из шести человек, находящихся в данный момент в магазине, и соответствующую этому числу вероятность.

6) Вероятность рождения мальчиков равна 0,515. Найдите наивероятнейшее число девочек из 300 новорожденных и соответствующую этому числу вероятность.

7) Придумайте и решите аналогичную задачу.

1.44. *Решите задачи (сформулируйте пять различных вопросов к каждой задаче).*

1) Вероятность появления в продаже бракованного компьютера равна 0,004. Оказалось, что за весь период работы фирмы продано 1 000 компьютеров.

2) Прядильщица обслуживает 1 000 веретен. Вероятность обрыва нити в течение одной минуты на одном веретене составляет 0,002.

3) Предприятие отправило на реализацию 5 000 качественных микрокалькуляторов. Однако в пути они могут с вероятностью 0,0002 повредиться.

4) Вакцина против гриппа дает иммунитет в 99 % случаях. В осенний период в поликлинике были вакцинированы 2 000 человек.

5) В магазин поступило 3 000 бутылок минеральной воды, причем вероятность появления при перевозке разбитой бутылки составляет 0,3 %.

6) В нашем городе вероятность того, что человек увидит цветную рекламу, равна 0,2. Для статистических исследований случайно отобраны 10 человек.

ГЛАВА 2
СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

2.1. Случайные величины и их числовые характеристики

Случайная величина — одно из важнейших понятий теории вероятностей. Случайные величины обозначают прописными буквами латинского алфавита, а их возможные значения — строчными.

Большинство экспериментов завершаются появлением некоторого числа X:

- в опыте по подбрасыванию n раз монеты — X (числа выпаданий решки);
- при стрельбе по мишени из n опытов — X (числа точных попаданий) и т.д.

В приведенных примерах число X обозначает величину, характеризующую некоторое случайное событие (опыт, эксперимент). В зависимости от исхода конкретного испытания эта величина принимает различные числовые значения, поэтому называется случайной.

Поскольку большинство величин окружающего мира являются случайными, их исследование представляется весьма важным в теории вероятностей.

2.1.1. Функция распределения случайной величины

В этом подразделе дано определение функции распределения и приведены ее свойства.

Результат эксперимента ξ будем называть ***случайной величиной*** (СВ), если для любого $x \in \mathbb{R}$ неравенство $\xi < x$ является событием, т.е. определена вероятность $P(\xi < x)$. Эта вероятность как функция от x называется *функцией распределения* (ФР) *случайной величины* ξ и обозначается $F_\xi(x)$.

Итак, ***функцией распределения*** называют вероятность того, что случайная величина ξ в результате испытания примет значение, меньшее x:

$$F_\xi(x) = P(\xi < x). \tag{2.1}$$

Там, где очевидно, о какой случайной величине идет речь, ее функцию распределения будем обозначать просто $F(x)$.

Свойства функции распределения:

1) функция распределения монотонно не убывает на $\mathbb{R}$, т.е. $\forall\, x_1, x_2 \in \mathbb{R}$, если $x_1 < x_2$, то $F(x_1) \le F(x_2)$. Это очевидно, так как событие $(\xi < x_2)$ является суммой двух несовместных событий $(\xi < x_1)$ и $(x_1 \le \xi < x_2)$. Значит, $p(\xi < x_2) = p(\xi < x_1) + p(x_1 \le \xi < x_2) \ge p(\xi < x_1)$ в силу неотрицательности вероятности;

2) так как событие $\xi < +\infty$ — достоверное, а $\xi < -\infty$ — невозможное, то $\exists \lim\limits_{x\to-\infty} F(x) = 0$, $\exists \lim\limits_{x\to+\infty} F(x) = 1$;

3) поскольку функция распределения монотонна и ограничена на $\mathbb{R}$, она может иметь не более чем счетное множество точек разрыва первого рода;

4) функция распределения непрерывна слева при любом значении x: $\lim\limits_{y\to x-0} F(y) = F(x)$;

5) вероятность того, что случайная величина ξ примет значение в полуинтервале $[a, b)$, равна

$$P\{a \le \xi < b\} = F(b) - F(a). \tag{2.2}$$

Эти свойства непосредственно вытекают из определения функции распределения.

2.1.2. Дискретные случайные величины

В этом подразделе вы научитесь составлять аналитическое выражение функции распределения дискретной случайной величины (ДСВ) и строить ее график, вычислять вероятность попадания ДСВ в заданный интервал.

Случайная величина называется ***дискретной***, если она принимает отдельные, изолированные возможные значения с определенными (ненулевыми) вероятностями. Тогда каждому элементарному исходу X ставится в соответствие одно из не более чем счетного набора пар чисел (x_1, p_1), ..., (x_n, p_n), $n \le \infty$.

Правило, устанавливающее связь между значением случайной величины и ее вероятностью, называется ***законом распределения случайной величины***.

Случайные величины обозначают прописными буквами латинского алфавита X, Y, ..., а значения, которые они принимают, — соответствующими строчными: x, y, ...

Например, дискретная случайная величина X представляет собой конечный (или бесконечный) ряд чисел x_1, x_2, x_3, ..., x_n, ... Если заданы вероятности p_1, p_2, p_3, ..., p_n, ..., то его называют также ***рядом распределений***.

Обычно закон распределения случайной величины задается в виде таблицы, в первой строке которой расположены значения случайной величины, а во второй — соответствующие им вероятности:

x_i	x_1	x_2	x_3	...	x_n
p_i	p_1	p_2	p_3	...	p_n

При этом сумма вероятностей всех возможных значений случайной величины X равна 1:

$$\sum_{i=1}^{n} p_i = 1.$$

Задача 2.1. В результате подбрасывания двух игральных костей появляется некоторое число X — случайная величина, характеризующая сумму выпавших очков с определенной вероятностью. Найти закон распределения такой случайной величины X.

Решение. Число равновозможных исходов $n = 6 \cdot 6 = 36$, а число благоприятных исходов, например, для $x = 4$ может быть получено тремя способами: $4 = 1 + 3 = 2 + 2 = 3 + 1$.

Поэтому соответствующая $x = 4$ вероятность равна $p = \frac{3}{36}$.

Закон распределения такой случайной величины можно задать таблицей:

x_i	2	3	4	5	6	7	8	9	10	11	12
p_i	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Дискретная случайная величина считается *заданной*, если указан закон ее распределения, т.е. известны все значения ДСВ и вероятность каждого из них.

Очевидно, что можно было бы еще учесть выпадение сумм вида 0, 1, 13, 14, ..., а также нецелые числа и т.д., но поскольку ни одно испытание этому не благоприятствует, рассматриваются только реальные испытания. Поэтому для практического применения формул для ДСВ достаточно учитывать *только те x_i, для которых* $p_i \neq 0$.

Поскольку каждому значению x ДСВ ставится в соответствие ее вероятность, то закон распределения ДСВ можно задавать с помощью функции распределения ДСВ.

Функцией распределения $F(x)$ ДСВ ξ называется вероятность события $\xi < x$: $F(x) = P(\xi < x)$. Очевидно, она обладает всеми общими свойствами функции распределения.

Свойства функции распределения ДСВ:

пусть задана ДСВ X: (x_i, p_i), $p_i \neq 0$, $\sum p_i = 1$. Тогда:

1) функция распределения непрерывна при $x \neq x_i$ и имеет разрыв первого рода при $x = x_i$, равный p_i;
2) функция распределения постоянна на полуинтервале $(x_i, x_{i+1}]$;
3) $F(x_i + 0) - F(x_i) = p_i$;
4) свойство накопительной вероятности:

$$F(x) = \sum_{x_i < x} p_i. \qquad (2.3)$$

Использование этого свойства удобно при моделировании, например на ЭВМ, где сначала указывают пределы суммирования, не зависящие от суммируемых величин.

График функции распределения произвольной ДСВ представляет собой «возрастающую ступеньку» (рис. 2.1). График функции распределения ДСВ, заданной аналитически формулой

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le x_1; \\ p_1 & \text{при} \quad x_1 < x \le x_2; \\ 1 & \text{при} \quad x > x_2, \end{cases}$$

приведен на рис. 2.1.

Введем θ-функцию: $\theta(x) = \begin{cases} 0, & \text{если} \quad x \le 0; \\ 1, & \text{если} \quad x > 0. \end{cases}$

Очевидно, что это функция распределения «неслучайной» СВ, принимающей значение 0 с вероятностью 1. Тогда ФР ДСВ, заданной рядом распределений, будет иметь вид $F(x) = \sum_i p_i \theta(x - x_i)$. Проверьте, что все свойства функции распределения ДСВ выполнены. Таким образом, суммирование идет по всем i, и необязательно представлять ряд распределений в виде ранжированного ряда, т.е. для машинного задания требуется меньшее число сравнений, а значит, машинного времени.

Пусть φ — некоторая детерминированная функция, определенная на пространстве элементарных исходов Ω случайной величины X. Тогда каждому возможному значению x_i случайной величины X соответствует определенное значение $y_i = \varphi(x_i)$. В таком случае исходу y_i благоприятствует элементарный исход x_i с той же вероятностью p_i, т.е. функция φ задает новое пространство элементарных

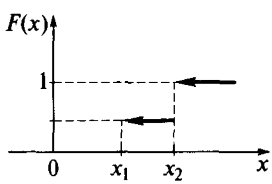


Рис. 2.1

исходов φ (Ω), на котором задана случайная величина Y, называемая ***функцией одного случайного аргумента*** $Y = \varphi(X)$.

Если одному значению y_i соответствуют различные значения $x_1, \ldots, x_k$, то полная вероятность осуществления y_i равна сумме вероятностей всех исходов, влекущих y_i, т.е.

$$p(X = x_1 \text{ или... или } X = x_k) = \sum_{i=1}^{k} P(X = x_i) = \sum_{i=1}^{k} p_i.$$

Задача 2.2. Дискретная случайная величина X задана рядом распределений:

x_i	–2	2	5
$P(x_i)$	0,35	0,42	0,23

Составить закон распределения ДСВ $Y = X^2$.

Решение. Составим закон распределения ДСВ $Y = X^2$:

x_i	–2	2	5
$(x_i)^2$	4	4	25
$P(x_i)$	0,35	0,42	0,23

Так как двум различным значениям СВ X ($x = -2$, $x = 2$) соответствуют равные значения СВ Y ($y = 4$), то составим новый закон распределения ДСВ $Y = X^2$, сложив вероятности, соответствующие этим значениям СВ X:

y_i^2	4	25
$P(y_i)$	0,77	0,23

? В повседневной жизни часто используется среднее значение случайных величин (например, «на выполнение этой работы потребуется в среднем два часа», «на концерте присутствовало около трехсот человек», «обычно очередь в кассу в соседнем магазине составляет пять-шесть человек»). Как охарактеризовать случайную величину не рядом чисел, а одним-двумя числовыми значениями?

2.1.3. Числовые характеристики дискретной случайной величины

Для описания дискретной случайной величины иногда удобнее пользоваться не законом распределения, а числовыми характеристиками: модой, медианой, математическим ожиданием, дисперсией и среднеквадратическим отклонением.

Модой ДСВ (будем обозначать Mo(X)) называется такое значение дискретной случайной величины, вероятность которого наибольшая.

Значение случайной величины, вероятность которого минимальная, называется *антимодой* (антимода выбирается из пространства реальных событий, ибо всегда можно ввести «фантастическое значение» СВ и приписать ему нулевую вероятность).

В задании 2.1 наибольшую вероятность $p = 1/6$ имеет $x = 7$. Значит, в этом случае мода равна 7, антимода — 2 и 12 ($p = 1/36$).

Ряд распределений может не иметь моды.

Задача 2.3. Ряд распределений выпадения грани при подбрасывании одного кубика имеет вид 1, 2, 3, 4, 5, 6, так как выпадение любой грани происходит с вероятностью $\frac{1}{6}$. Составить закон распределения ДСВ X и найти моду.

Решение. Закон распределения для ДСВ X задается таблицей, из которой видно, что у всех значений ДСВ вероятность одинакова. Значит, в этом случае мода не может быть указана.

x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

В некоторых случаях ДСВ имеет не одну, а несколько мод.

Задача 2.4. По наблюдениям метеорологов, среднесуточная температура в первой половине февраля имела следующий ряд распределения: –18, –15, –18, –18, –15, –12, –12, –5, –10, –7, –12, –18, –20, –15, –12. Составить закон распределения ДСВ — среднесуточной температуры и найти моду.

Решение. Составим закон распределения ДСВ — среднесуточной температуры, ранжируя ее значения в порядке возрастания:

x_i	–20	–18	–15	–12	–10	–7	–5
p_i	1/15	4/15	3/15	4/15	1/15	1/15	1/15

В данном случае наибольшую вероятность $p = 4/15$ имеют два значения ДСВ: $x = -18$ и $x = -12$. Значит, мода ДСВ X равна Mo(X) = $= -12$ и Mo(X) $= -18$.

Моду как характеристику ДСВ часто используют в социологических исследованиях, например при определении рейтинга популярности того или иного политического деятеля или певца. При этом в качестве ДСВ выступает число голосов, отданных, например, за любимого певца при социологическом опросе. Результат такого социологического исследования может быть использован, например, при составлении репертуара музыкальных радио- и телепередач, при поиске объекта для взятия интервью для прессы, при составлении концертных репертуаров и т.д., т.е. имеет конкретное значение в деловом бизнесе, а значит, влечет за собой экономический эффект.

Так, результат социологического опроса молодежи: «Какой музыкой вы увлекаетесь, какой отдаете предпочтение?» (4,2 % ответили: симфонической, 38,7 % — эстрадной, 28 % — джазовой и т.д.), служит источником информации для выпуска кассет или компакт-дисков.

Медианой ДСВ (будем обозначать Ме(X)) называется среднее по положению в пространстве событий значение дискретной случайной величины.

Если ранжировать (упорядочить в порядке возрастания или убывания) ряд распределений ДСВ, то в ряду с нечетным количеством членов медиана есть значение ДСВ на «среднем месте».

Номер места n вычисляется по формуле

$$n = \frac{N+1}{2}, \tag{2.4}$$

где N — количество (конечное) элементов в ряду распределений.

Если в ряду распределений четное число членов, то медианой являются два значения с номерами $n - 0{,}5 = [N/2]$ и $n + 0{,}5 = [N/2 + 1]$. Так, большие пальцы рук являются пятым и шестым, если считать слева направо ($N = 10$).

Задача 2.5. Учет производительности труда станочников цеха № 3 за смену задан рядом распределений:

Номер по списку	1	2	3	4	5	6	7	8
Производительность, дет./смену	52	52	53	54	56	57	57	57

Найти моду и медиану ДСВ X.

Решение. Мода (*самая «модная»* производительность труда) равна 57 (деталей за смену), значит, Мо(X) = 57.

Так как $N = 8$, то медиана вычисляется с помощью определения номера места n по формуле $n = \frac{N+1}{2} = \frac{8+1}{2} = 4{,}5$ тогда $n - 0{,}5 = 4$, $n + 0{,}5 = 5$.

Среднее арифметическое значение ДСВ на четвертом и пятом местах $\frac{54+56}{2} = 55$, значит, Ме(X) = 55.

Заметим, что в случае если работник с табельным номером 1 не вышел на работу, медиана равна Me(X) = 56 (из семи значений на четвертом месте), если не вышел работник с номером 8 — Me(X) = 54 (на четвертом месте).

Для нахождения медианы ранжирование ряда распределений является обязательным условием. Медиана характеризует не только количественную, но и качественную сторону некоторого события. Например, средний капитал фирмы может определять и стабильность, и надежность.

Одна из самых важных характеристик ДСВ — математическое ожидание.

Математическим ожиданием ДСВ называется сумма произведений значений случайной величины на их вероятности (обозначается $M(X)$ или Mx):

$$M(X) = \sum_i x_i p_i \tag{2.5}$$

в случае, если существует сумма $M|x| = \sum_i |x_i| p_i$.

Бытовой и практический смысл математического ожидания — это среднее значение ДСВ. Например, при некотором измерении (скажем, титровании) величины C было получено N результатов $C_1, C_2, \ldots, C_N$. Если они относительно близко лежат: $(C_{max} - C_{min})/C_{min} \ll 1$, т.е. нет грубых промахов, то все полученные результаты равновероятны: $p_k = 1/N = p$; тогда математическое ожидание ДСВ C равно $M(C) = \sum_{k=1}^{N} C_k p_k = p\sum_{k=1}^{N} C_k = \frac{1}{N}\sum_{k=1}^{N} C_k$, т.е. среднему арифметическому. Аналогично среднее значение некоторой функции f от ДСВ X вычисляется по формуле

$$Mf(X) = \sum_i f(x_i) p_i. \tag{2.6}$$

Если ДСВ задана законом распределения с конечным числом элементов:

X_i	x_1	x_2	x_3	...	x_{n-1}	x_n
p_i	p_1	p_2	p_3	...	p_{n-1}	p_n

то математическое ожидание $M(X)$ находится по формуле

$$M(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n = \sum_{i=1}^{n} x_i p_i.$$

? Любая ли ДСВ может иметь математическое ожидание?

Действительно, ДСВ может не обладать математическим ожиданием, так как в случае бесконечного ряда распределений ряд

$\sum_{i=1}^{\infty} p_i$ должен сходиться к единице, а ряд $\sum_{i=1}^{\infty} x_i p_i$ может расходиться, например, $p_i = 6/(\pi i)^2$, $x_i = i^2$, $i = 1...\infty$.

Пример 2.1 [*«петербургская игра»*, см. прил. 2(6)]. Игрок подбрасывает монету (если выпадает орел, то он выигрывает 1 р., при очередном бросании он выигрывает в два раза больше) до тех пор, пока не выпадает решка. На этом игра прекращается. Вычислим математическое ожидание *выигрыша*: $x_1 = 1$, $p_1 = \frac{1}{2}$; $x_2 = 2$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$, ..., $x_n = 2^{n-1}$, $p_n = \frac{1}{2^n}$.

$M(X) = \sum_{n=1}^{\infty} x_n p_n = \sum_{n=1}^{\infty} 2^{n-1} \frac{1}{2^n} = \frac{1}{2} \sum_{n=1}^{\infty} 1 = \infty$. Таким образом, невозможно определить даже разумные взносы участников игры.

ДСВ может не обладать математическим ожиданием, если ряд $\sum_{i=1}^{\infty} x_i p_i$ существует (сходится), а ряд $\sum_{i=1}^{\infty} |x_i| p_i$ расходится. Поясним смысл обязательного существования суммы ряда из модулей: согласно теоремам Римана, если существует $\sum_{i=1}^{\infty} x_i p_i$, но не существует $\sum_{i=1}^{\infty} |x_i| p_i$, то надлежащей перестановкой членов ряда можно подогнать сумму $\sum_{i=1}^{\infty} x_i p_i$ к любому действительному числу с заданной точностью, т.е. понятие математического ожидания превращалось бы в абсурд. Если существуют обе суммы, то любая перестановка членов обоих рядов не влияет на итоговые суммы. Очевидно, что вышеизложенное относится к случаям, когда число вариант ряда распределений счетно.

Если число вариант ряда распределения ДСВ конечно, то такая ДСВ будет обладать математическим ожиданием.

Задача 2.6. Из 100 лотерейных билетов в тридцати выигрыш составляет 100 тыс. р., в десяти — 200 тыс. р., в пяти — 300 тыс. р., в одном — 1 млн р. Найти числовые характеристики выигрыша.

Решение. Случайная величина X — выигрыш — принимает значения $x_1 = 0$, $x_2 = 100$ тыс. р., $x_3 = 200$ тыс. р., $x_4 = 300$ тыс. р., $x_5 =$ $= 1$ млн р.

Вероятность того, что СВ X принимает соответственно значения:

$$p_1 = \frac{54}{100} = 0,54;\ p_2 = \frac{30}{100} = 0,3;\ p_3 = \frac{10}{100} = 0,1;$$

$$p_4 = \frac{5}{100} = 0,05;\ p_5 = \frac{1}{100} = 0,01.$$

Тогда закон распределения этой ДСВ имеет вид:

x_i	0	100 тыс.	200 тыс.	300 тыс.	1 млн
p_i	0,54	0,30	0,10	0,05	0,01

Числовые характеристики выигрыша:

1) Мо = 0, так как наибольшая вероятность $p = 0{,}54$ (отсюда поговорка «в азартные игры с государством не играю», очень велика вероятность проигрыша). Антимода равна 1 млн;

2) Ме = 200 тыс. — «идеальное равновесие» находится на третьем месте. Но эти характеристики не отвечают на вопрос: «Каков ожидаемый выигрыш?»;

3) $M(X) = 0 \cdot 0{,}54 + 100$ тыс. р. $\cdot\, 0{,}3 + 200$ тыс. р. $\cdot\, 0{,}1 + 300$ тыс. р. $\times$ $\times\, 0{,}05 + 1\,000$ тыс. р. $\cdot\, 0{,}01 = 75$ тыс. р. — это и есть среднее значение выигрыша, поэтому лотерейный билет должен стоить никак не меньше 75 тыс. р.

Свойства математического ожидания:

1) математическое ожидание постоянной величины равно этой постоянной $M(C) = C = \text{const}$;

2) постоянный множитель можно выносить за знак математического ожидания:

$$M(kX) = kM(X), \quad k = \text{const};$$

3) математическое ожидание суммы случайных величин равно сумме их математических ожиданий: $M(X + Y) = M(X) + M(Y)$;

4) математическое ожидание произведения *независимых* случайных величин равно произведению их математических ожиданий: $M(XY) = M(X)M(Y)$.

Математическое ожидание можно найти всегда, если задан закон распределения ДСВ.

В жизни, в быту, в других науках часто используют средние значения различных случайных величин: средняя продолжительность жизни, среднемесячная температура, средний надой молока, среднее потребление электроэнергии, средняя потребительская корзина, центр масс системы и т.д.

Задача 2.7. Найти математическое ожидание случайной величины $Y = 5X + 9$, если известно, что $M(X) = 2{,}5$.

Решение. Зная свойства математического ожидания, имеем:

$$M(Y) = M(5X + 9) = M(5X) + M(9) = 5M(X) + 9 = 5 \cdot 2{,}5 + 9 =$$
$$= 7{,}5 + 9 = 16{,}5.$$

Задача 2.8. Найти математическое ожидание случайной величины:

а) X — суммы очков, выпавших при подбрасывании двух кубиков;

б) T — произведения числа очков на их гранях.

Решение. Пусть Y и Z — случайные величины, выпавшие одновременно на первом и втором кубиках соответственно. Эти случайные величины имеют одинаковый ряд распределений (одинаково распределены).

y_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

Тогда их математические ожидания можно найти по известной формуле:

$$M(Y)=\sum_{i=1}^{6} y_i p_i;\ M(Z)=\sum_{i=1}^{6} z_i p_i,\ \text{или}\ M(Y)=M(Z)=$$
$$=(1+2+3+4+5+6)1/6=3{,}5.$$

а) Математическое ожидание суммы очков

$$M(X)=M(Y+Z)=M(Y)+M(Z)=3{,}5+3{,}5=7.$$

б) Так как случайные величины Y и Z независимы, имеем:

$$M(T)=M(XY)=M(X)M(Y)=3{,}5\cdot 3{,}5=12{,}25.$$

Задача 2.9. Контролеры проводили проверку качества изготовления болтов двумя бригадами. Отклонение длины болта от заданных размеров (стандарта) в миллиметрах для каждой из бригад есть случайные величины X_1 и X_2 соответственно, заданные таблично:

X_{1i}	−10	−6	−2	1	3	5	8	10
p_i	1/16	1/8	1/4	1/16	1/4	1/16	1/8	1/16

X_{2i}	−2	−1	0	2	3	4	4
p_i	1/4	1/4	1/16	1/16	1/8	1/8	1/8

Какая бригада работает лучше?

Решение. Сравним математические ожидания двух случайных величин X_1 и X_2, заданных рядами распределений:

$$M(X_1)=(-10)\cdot\frac{1}{16}+(-6)\cdot\frac{1}{8}+(-2)\cdot\frac{1}{4}+1\cdot\frac{1}{16}+$$
$$+3\cdot\frac{1}{4}+5\cdot\frac{1}{16}+8\cdot\frac{1}{8}+10\cdot\frac{1}{16}=\frac{7}{8};$$

$$M(X_2) = (-2) \cdot \frac{1}{4} + (-1) \cdot \frac{1}{4} + 0 \cdot \frac{1}{16} +$$
$$+2 \cdot \frac{1}{16} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} + 5 \cdot \frac{1}{8} = \frac{7}{8}.$$

? Можно ли по полученным результатам сделать вывод о качестве работы этих бригад?

Надежда на то, что математические ожидания случайных величин X_1 и X_2 покажут, какая из бригад работает лучше, не оправдались. Равные математические ожидания «уравняли» явно различное качество работы этих бригад.

Таким образом, можно сделать вывод: недостаточно знаний одного математического ожидания — оно может совпадать для явно различных распределений этих случайных величин.

Если сравнить распределение этих случайных величин на числовой прямой (рис. 2.2), то видно, что случайная величина X_1 (рис. 2.2, *а*) имеет большой разброс на отрезке $[-10; 10]$, а случайная величина X_2 (рис. 2.2, *б*) сосредоточена вокруг $M(X_2) = 7/8$.

Очевидно, что «лучшим» считают такое распределение, которое менее отклоняется от среднего значения — стандарта, т. е. интуитивно ожидается результат: лучше работает вторая бригада. В частности, значительное отклонение от стандарта для первой бригады могло послужить поводом, что ее в отличие от второй бригады лишили премии.

Чтобы определить степень *«сосредоточенности»* ДСВ вокруг ее математического ожидания, надо найти среднее отклонение случайной величины X от ее математического ожидания $M(X)$, т.е. $M[X - M(X)]$. Так как такое отклонение всегда равно нулю и такая характеристика не может являться ответом на поставленную задачу, то находят математическое ожидание не самого отклонения, а его квадрата.

Дисперсией дискретной случайной величины X называется математическое ожидание квадрата отклонения этой величины X от

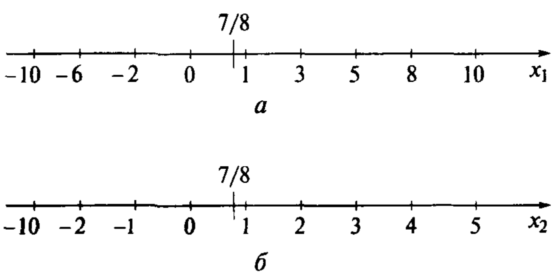


Рис. 2.2

ее математического ожидания. Дисперсия ДСВ X обозначается $D(X)$ или Dx. Тогда

$$D(X) = M[X - M(X)]^2. \qquad (2.7)$$

Найдем дисперсию в задаче 2.9:

x_{1i}	-10	-6	-2	1	3	5	8	10
$X_1 - M(X_1)$	$-10\frac{7}{8}$	$-6\frac{7}{8}$	$-2\frac{7}{8}$	$\frac{1}{8}$	$2\frac{1}{8}$	$4\frac{1}{8}$	$7\frac{1}{8}$	$9\frac{1}{8}$
$[X_1 - M(X_1)]^2$	$118\frac{17}{64}$	$47\frac{17}{64}$	$8\frac{17}{64}$	$\frac{1}{64}$	$4\frac{33}{64}$	$17\frac{1}{64}$	$50\frac{49}{64}$	$83\frac{17}{64}$
p_{1i}	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
$p_1[X_1 - M(X_1)]^2$	$7\frac{401}{1024}$	$5\frac{465}{512}$	$2\frac{17}{256}$	$\frac{1}{1024}$	$1\frac{33}{256}$	$1\frac{65}{1024}$	$6\frac{177}{512}$	$5\frac{209}{1024}$

Тогда $D(X_1) = M[X_1 - M(X_1)]^2 = 9$.

x_{2i}	-2	-1	0	2	3	4	5
$X_2 - M(X_2)$	$-2\frac{7}{8}$	$-1\frac{7}{8}$	$-\frac{7}{8}$	$1\frac{1}{8}$	$2\frac{1}{8}$	$3\frac{1}{8}$	$4\frac{1}{8}$
$[X_2 - M(X_2)]^2$	$8\frac{17}{64}$	$4\frac{17}{64}$	$\frac{49}{64}$	$1\frac{17}{64}$	$4\frac{33}{64}$	$9\frac{49}{64}$	$17\frac{1}{64}$
p_{2i}	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$p_2[X_2 - M(X_2)]^2$	$2\frac{17}{256}$	$1\frac{17}{256}$	$\frac{49}{1024}$	$\frac{81}{256}$	$\frac{189}{512}$	$1\frac{113}{512}$	$2\frac{65}{512}$

$$D(X_2) = M[X_2 - M(X_2)]^2 = 6\frac{934}{1024} \approx 7.$$

Итак, если отклонение от среднего значения длины (стандарта) — дисперсия — небольшое, то считается, что станки работают нормально. Аналогично весы с меньшим отклонением от стандарта (с меньшей дисперсией) считаются лучше.

Свойства дисперсии случайной величины:

1) дисперсия постоянной величины равна нулю

$$D(C) = 0, \text{ где } C = \text{const};$$

2) дисперсия — всегда неотрицательная величина

$$D(X) \geq 0, \text{ так как } p_i(x) > 0 \text{ и } [X - M(X)]^2 \geq 0;$$

3) при вынесении постоянного множителя за знак дисперсии необходимо возвести его в квадрат: $D(kX) = k^2 D(X)$, где $k = \text{const}$;

4) для независимых случайных величин X и Y дисперсия суммы равна сумме дисперсий слагаемых: $D(X + Y) = D(X) + D(Y)$;

5) если существуют $M(X)$ и $M(X^2)$, то дисперсия ДСВ равна математическому ожиданию квадрата ДСВ минус квадрат ее математического ожидания:

$$D(X) = M(X^2) - [M(X)]^2. \tag{2.8}$$

Действительно, так как математическое ожидание есть постоянная величина, то

$$D(X) = M[X - M(X)]^2 = M[X^2 - 2XM(X) + M^2(X)] =$$
$$= M(X^2) - [M(X)]^2.$$

Формула (2.8) более удобна для вычисления дисперсии.

Например, дисперсию $D(X)$ в задаче 2.9 можно определить таким способом:

x_{2i}	–10	–6	–2	1	3	5	8	10
x_i^2	100	36	4	1	9	25	64	100
p_i	1/16	1/8	1/4	1/16	1/4	1/16	1/8	1/16
$p_i x_i^2$	25/4	9/2	1	1/16	9/4	25/16	8	25/4

Тогда $M(X^2) = 9\frac{3}{4}$, поэтому

$$D(X) = M(X^2) - [M(X)]^2 = 9\frac{3}{4} - \left(\frac{7}{8}\right)^2 = 9\frac{3}{4} - \frac{49}{64} = 8\frac{63}{64} \approx 9,$$

x_i'	–2	–1	0	1	2	3	4	5
$(x_i')^2$	4	1	0	1	4	9	16	25
p_i	1/4	1/4	1/16	0	1/16	1/8	1/8	1/8
$p_i(x_i')^2$	1	1/4	0	0	1/4	9/8	2	25/8

$$M(X')^2 = 7\frac{7}{8};\; D(X') = M(X')^2 - [M(X')]^2 = 7\frac{7}{8} - \left(\frac{7}{8}\right)^2 = 7\frac{7}{64} \approx 7.$$

? Если случайная величина X в этой задаче измерялась в миллиметрах, то в каких единицах измеряется $D(X)$? Если бы случайной величиной X в этой задаче были ошибки при получении сдачи в магазине, то дисперсия в этом случае измерялась бы в «квадратных» рублях.

Размерность дисперсии ДСВ не совпадает с размерностью самой ДСВ, что вызывает определенные трудности при подсчетах. Поэтому на практике чаще используют не саму дисперсию, а среднеквадратическое отклонение σ.

Среднеквадратическим отклонением случайной величины X называется корень квадратный из дисперсии этой ДСВ:

$$\sigma(X) = \sqrt{D(X)}. \qquad (2.9)$$

В задаче 2.9 $\sigma(X) = \sqrt{9} = 3$; $\sigma(X') = \sqrt{7} \approx 2{,}7$.

Среднеквадратическое отклонение также является *мерой разброса* значений ДСВ около ее математического ожидания. В экономике среднеквадратическое отклонение называют *стандартным отклонением*.

Задача 2.10. Составить закон распределения ДСВ X, принимающей значения $x_1 = 1$, $x_2 = 3$, $x_3 = 4$, если известно, что $M(X) = 2{,}3$, а $D(X) = 1{,}21$.

Решение. 1) Обозначим неизвестные вероятности, соответствующие заданным значениях ДСВ X, через p_1, p_2, p_3. Тогда для них справедливо требование $p_1 + p_2 + p_3 = 1$. Согласно условию, можно составить систему уравнений, выразив $M(X)$ — в первом уравнении, а во втором уравнении $M(X^2)$ через $D(X)$ из формулы $D(X) = M(X^2) - [M(X)]^2$.

Отсюда $M(X^2) = D(X) + [M(X)]^2$. Тогда система примет вид:

$$\begin{cases} p_1 + 3p_2 + 4p_3 = 2{,}3; \\ p_1 + 9p_2 + 16p_3 = 1{,}21 + 2{,}3^2; \\ p_1 + p_2 + p_3 = 1. \end{cases}$$

2) Решим систему трех уравнений с тремя неизвестными (методом Гаусса):

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 2{,}3 \\ 1 & 9 & 16 & 6{,}5 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 1{,}3 \\ 0 & 8 & 15 & 5{,}5 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 1{,}3 \\ 0 & 0 & 3 & 0{,}3 \end{array}\right).$$

Найдем значения неизвестных: $p_1 = 0{,}4$, $p_2 = 0{,}5$, $p_3 = 0{,}1$.

Задача 2.11. ДСВ может принимать только два значения x_1 и x_2, причем $x_1 < x_2$. Найти закон распределения ДСВ X, если известно, что $p_1 = 0,5$; $M(X) = 3$; $D(X) = 1$.

Решение. 1) Найдем неизвестное значение p_2 из требования $p_1 + p_2 = 1$. Имеем $p_2 = 1 - 0,5 = 0,5$.

2) Выразим $M(X)$, используя определение математического ожидания. Согласно условию, можно составить уравнение: $0,5x_1 + 0,5x_2 = 3$, т.е. $x_1 + x_2 = 6$.

3) Выразим $D(X)$, используя определение дисперсии. Согласно условию, можно составить уравнение: $0,5x_1^2 + 0,5x_2^2 = 10$, т.е. $x_1^2 + x_2^2 = 20$.

4) Решим полученную систему уравнений: .

$$\begin{cases} x_1 + x_2 = 6; \\ x_1^2 + x_2^2 = 20 \end{cases} \Rightarrow \begin{cases} x_1 + x_2 = 6; \\ (x_1 + x_2)^2 - 2x_1x_2 = 2 \end{cases} \Rightarrow \begin{cases} x_1 + x_2 = 6; \\ x_1x_2 = 8. \end{cases}$$

5) Так как по условию известно, что $x_1 < x_2$, то решением системы будут значения ДСВ: $x_1 = 2$ и $x_2 = 4$.

Задача 2.12. Найти $M(Z_1)$, $M(Z_2)$, $M(Z_3)$, $D(Z_1)$, $D(Z_2)$, если известно, что случайные величины X и Y независимы, причем $M(X) = 3$; $M(Y) = 4$; $D(X) = 5$; $D(Y) = 2$, для ДСВ $Z_1 = 2X + 3Y$; $Z_2 = 5X - 4Y$; $Z_3 = (3X - 2Y)^2$.

Решение. 1) Так как по условию случайные величины X и Y независимы, то воспользуемся свойствами математического ожидания. Тогда имеем:

$M(Z_1) = M(2X + 3Y) = 2M(X) + 3M(Y) = 2 \cdot 3 + 3 \cdot 4 = 18$;
$M(Z_2) = M(5X - 4Y) = 5M(X) - 4M(Y) = 5 \cdot 3 - 4 \cdot 4 = -1$;
$M(Z_3) = M(3X - 2Y)^2 = M(9X^2 - 12XY + 4Y^2) = 9M(X^2) - 12M(X)M(Y) + 4M(Y^2)$.

2) Найдем $M(X^2)$ и $M(Y^2)$, воспользовавшись формулой для вычисления дисперсии:

$D(X) = M(X^2) - [M(X)]^2$. Отсюда имеем $M(X^2) = D(X) + [M(X)]^2 = 5 + 9 = 14$ и $M(Y^2) = D(Y) + [M(Y)]^2 = 2 + 16 = 18$.

3) Вычислим $M(Z_3) = M(3X - 2Y)^2 = 9 \cdot 14 - 12 \cdot 3 \cdot 4 + 4 \cdot 18 = 54$.

4) Найдем дисперсию, используя ее свойства:

$D(Z_1) = D(2X + 3Y) = 4D(X) + 9D(Y) = 4 \cdot 5 + 9 \cdot 2 = 38$;
$D(Z_2) = D(5X - 4Y) = 25D(X) + 16D(Y) = 25 \cdot 5 + 16 \cdot 2 = 157$.

Иногда ДСВ удобно задавать графически. При этом на оси абсцисс откладывают значения ДСВ, а на оси ординат — вероятность (или частоту) их появления. Полученные точки соединяют отрезками. Такой график называют *«многоугольник распределений»*.

Многоугольник распределений еще раз проявляет дискретный (прерывистый) характер ДСВ.

В следующих подразделах рассмотрим отдельные виды распределения ДСВ, научимся составлять закон их распределения, функцию распределения ДСВ, находить числовые характеристики отдельных видов ДСВ, осуществлять поиск неизвестных данных ДСВ по заданным, узнавать вид ДСВ в реальной ситуации по содержанию задания.

2.2. Биномиальное распределение

В этом подразделе рассмотрим определение и числовые характеристики биноминального распределения.

Биномиальным называется закон распределения ДСВ X — числа появления события в n независимых испытаниях, в каждом из которых вероятность появления события p — величина постоянная, причем вероятность возможного значения $X = m$ вычисляется по формуле Бернулли:

$$P\{X = m\} = C_n^m p^m q^{n-m}, \text{ где } q = 1 - p;\ m = 0,\ 1,\ \ldots,\ n. \qquad (2.10)$$

Распределение называется биномиальным, так как правую часть формулы (2.10) можно рассматривать как общий член разложения бинома Ньютона, в частности нормировка

$$\sum_{m=0}^{n} C_n^m p^m q^{n-m} = (p+q)^n = 1^n = 1.$$

Таким образом, вероятность события $X = m$ из n возможных полностью совпадает с соответствующим членом разложения бинома Ньютона, стоящем на $(m+1)$-м месте справа.

ДСВ X, распределенную по биномиальному закону с параметрами n и p, обозначим $X \in \mathrm{Bi}(n,\ p)$.

Тогда ряд распределений биномиального закона имеет вид:

x_i	0	1	2	...	m	...	$n-1$	n
p_i	q^n	npq^{n-1}	$C_n^2 p^2 q^{n-2}$	...	$C_n^m p^m q^{n-m}$	...	$np^{n-1}q$	p^n

Найдем числовые характеристики биномиального закона распределений: математическое ожидание $M(X)$, дисперсию $D(X)$, среднеквадратическое отклонение $\sigma(X)$.

Пусть в каждом из n независимых испытаний исследуемое событие A имеет вероятность появления p. Тогда вероятность того, что это событие не произойдет, $q = 1 - p$.

Закон распределения такой ДСВ имеет вид:

x_i	0	1
p_i	q	p

Тогда для одного испытания $M(X) = 0 \cdot q + 1 \cdot p = p$, а для n испытаний

$$M(X) = np,$$

$$M(X^2) = 0^2 \cdot q + 1^2 \cdot p = p.$$

По формуле $D(X) = M(X^2) - [M(X)]^2 = p - p^2 = p(1 - p) = pq$, так как $p + q = 1$.

Так как отдельные события x_i — независимые, а событие X — сумма x_i, то

$$D(x_1 + x_2 + \ldots + x_n) = pq + pq + \ldots + pq = npq. \qquad (2.11)$$

Итак, при биномиальном распределении $D(X) = npq$, а $\sigma(X) = \sqrt{D(X)} = \sqrt{npq}$.

Задача 2.13. Определить математическое ожидание, дисперсию и среднеквадратическое отклонение числа правильных заключений на проверяемые балансы, если аудитору на заключение представлены шесть балансов, а аудитор допускает ошибку при проверке бухгалтерского баланса с вероятностью $\frac{1}{3}$. Найти *функцию* распределения для ДСВ X — числа правильных заключений на проверяемые балансы.

Решение. 1) По условию имеем $n = 6$; $p = \frac{1}{3}$; $q = \frac{2}{3}$; тогда $M(X) = np = 6 \cdot \frac{1}{3} = 4$; $D(X) = npq = 6 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{3}$; $\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{4}{3}} \approx 1{,}15$.

2) Дискретная случайная величина X — число правильных заключений на проверяемые балансы — имеет *биномиальное* распределение, так как каждый из балансов может с постоянной вероятностью $\frac{1}{3}$ быть с ошибкой. Возможные значения ДСВ X есть натуральные числа, не превосходящие шести: $x_i = i$, $i = \overline{1,6}$, или ноль: $x_0 = 0$.

3) Используя формулу Бернулли, можно составить *функцию распределения* для *биномиальной случайной величины* X — числа правильных заключений на проверяемые балансы, при $n = 6$, $p = \frac{1}{4}$, $q = \frac{3}{4}$.

Тогда *функция распределения* примет вид:

$$F(x) = \begin{cases} 0 & \text{при } x \le 0; \\ \sum\limits_{0 \le k < x} C_6^k \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{6-k} & \text{при } 0 < x \le 6; \\ 1 & \text{при } x > 6. \end{cases}$$

Задача 2.14. Проверка качества микрокалькуляторов (МК) показала, что из каждых ста МК имеют дефекты в среднем 25 штук.

1) Составить ряд распределений вероятностей для X — числа исправных МК из взятых наудачу шести из них. Построить график этого распределения.

2) Найти числовые характеристики этой случайной величины.

3) Составить функцию распределения и построить ее график.

Решение. 1) По условию имеем $n = 6$; $p = 0{,}75$; $q = 0{,}25$. Так как любой калькулятор может быть либо исправным, либо не исправным с постоянной вероятностью, то число исправных МК распределяется по биномиальному закону. Для составления ряда распределений найдем по формуле Бернулли соответствующие вероятности для всех возможных значений $m = \{0, 1, 2, 3, 4, 5\}$, округлив вероятность до тысячных.

$P_6(0) = 0,25^6 \approx 0,0002 \approx 0,000$; $\quad P_6(1) = 6 \cdot 0,75^1 \cdot 0,25^5 \approx 0,004$;

$P_6(2) = C_6^2 \cdot 0,75^2 \cdot 0,25^4 \approx 0,033$; $\quad P_6(3) = C_6^3 \cdot 0,75^3 \cdot 0,25^3 \approx 0,132$;

$P_6(4) = C_6^4 \cdot 0,75^4 \cdot 0,25^2 \approx 0,297$; $\quad P_6(5) = C_6^5 \cdot 0,75^5 \cdot 0,25^1 \approx 0,356$;

$P_6(6) = 0,75^6 \approx 0,178$.

2) Тогда закон распределения этой ДСВ можно представить в виде ряда:

x	0	1	2	3	4	5	6
P	0,000	0,004	0,033	0,132	0,297	0,356	0,178

3) Графическое представление этого биномиального закона распределения — *многоугольник распределений* — имеет вид (рис. 2.3).

4) Числовые характеристики этой биномиальной ДСВ X найдем по формулам:

$$M(X) = np = 6 \cdot 0{,}75 = 4{,}5;$$
$$D(X) = npq = 6 \cdot 0{,}75 \cdot 0{,}25 \approx 1{,}125;$$
$$\sigma(X) = \sqrt{D(X)} = \sqrt{1{,}125} \approx 1{,}06.$$

5) Функцию распределения ДСВ X представим в виде системы

$$F(x) = \begin{cases} 0 & \text{при } x \le 0; \\ 0,000 & \text{при } 0 < x \le 1; \\ 0,004 & \text{при } 1 < x \le 2; \\ 0,037 & \text{при } 2 < x \le 3; \\ 0,169 & \text{при } 3 < x \le 4; \\ 0,466 & \text{при } 4 < x \le 5; \\ 0,822 & \text{при } 5 < x \le 6; \\ 1 & \text{при } x > 6. \end{cases}$$

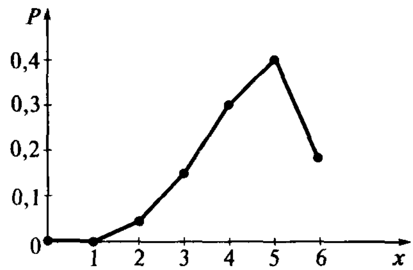


Рис. 2.3

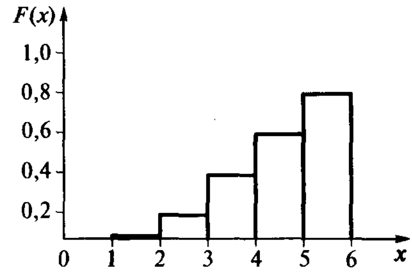


Рис. 2.4

Тогда график функции распределения имеет ступенчатый вид (рис. 2.4).

Задача 2.15. Стрелок производит по мишени три выстрела. Вероятность попадания в мишень при каждом выстреле $\frac{1}{3}$. Построить многоугольник распределения числа попаданий и найти все основные характеристики этой ДСВ.

Решение. Вероятность попадания $p = \frac{1}{3}$, а вероятность промаха $q = 1 - p = \frac{2}{3}$.

Рассмотрим все возможные варианты, учитывая, что данное распределение носит биномиальный характер, и вероятность появления ДСВ находят по формуле Бернулли $P_n^k = C_n^k p^k q^{n-k}$:

1) все промахи: $n = 3$, $m = 0$, тогда $P_3^0 = C_3^0 p^0 q^3 = \frac{8}{27}$;

2) одно попадание: $n = 3$, $m = 1$, отсюда

$$P_3^1 = C_3^1 p^1 q^2 = 3 \cdot \frac{1}{3} \cdot \left(\frac{2}{3}\right)^2 = \frac{4}{9};$$

3) два попадания: $n = 3$, $m = 2$, тогда $P_3^2 = C_3^2 p^2 q = 3 \cdot \left(\frac{1}{3}\right)^2 \cdot \frac{2}{3} = \frac{2}{9}$;

4) три попадания: $n = 3$, $m = 3$, поэтому

$$P_3^3 = C_3^3 p^3 q^0 = 1 \cdot \left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^0 = \frac{1}{27}.$$

При этом сумма вероятностей должна быть равна единице, так как рассмотрены все варианты, представляющие полную группу событий:

$$\frac{8}{27} + \frac{4}{9} + \frac{2}{9} + \frac{1}{27} = 1.$$

Поэтому на практике достаточно вычислять $(n-1)$-е значение p_i, а последнее, наиболее трудно вычисляемое, считать из условия *нормировки вероятности*.

Ряд распределений имеет вид 0; 1; 2; 3. Закон распределений оформим в виде таблицы

x_i	0	1	2	3
p_i	8/27	4/9	2/9	1/27

Математическое ожидание $M(X) = np = 3 \cdot 1/3 = 1$

$$\left(\text{или } 0 \cdot \frac{8}{27} + 1 \cdot \frac{4}{9} + 2 \cdot \frac{2}{9} + 3 \cdot \frac{1}{27} = 1\right);$$

дисперсия $D(X) = npq = 3 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{3}$; *среднеквадратическое отклонение* $\sigma = \sqrt{\frac{2}{3}} \approx 0{,}816$.

На рис. 2.5 представлены многоугольник распределений (*а*) и график функции распределения (*б*). «*Выколотые*» точки означают, что при этом аргументе значение функции распределения берется на нижней ветви графика.

Задача 2.16. ДСВ X распределена по биномиальному закону с $M(X) = 6$ и $D(X) = 2$. Найти вероятность попадания X в интервал [2; 5].

Решение. 1) Так как ДСВ X распределена по биномиальному закону и на интервале [2; 4] может принимать только значения 2, 3, 4, то

$$P(2 \le X \le 5) = P(X=2) + P(X=3) + P(X=4) + P(X=5) =$$
$$= C_n^2 p^2 q^{n-2} + C_n^3 p^3 q^{n-3} + C_n^4 p^4 q^{n-4} + C_n^5 p^5 q^{n-5}.$$

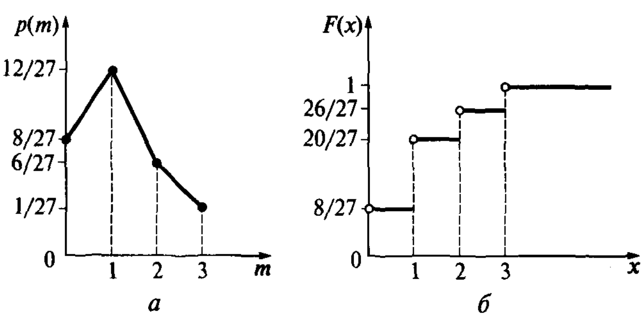


Рис. 2.5

2) Зная, что $M(X) = 6$ и $D(X) = 2$, найдем значение q для биномиального распределения из уравнений: $np = 6$, $npq = 2$, откуда $q = \frac{1}{3}$. Тогда $p = \frac{2}{3}$, $n = 9$.

3) Подставим найденные значения $n = 9$, $p = \frac{2}{3}$ и $q = \frac{1}{3}$ и определим вероятность попадания X в интервал [2;5]:

$$P(2 \le X \le 5) = C_9^2\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^7 + C_9^3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^6 +$$
$$+ C_9^4\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^5 + C_4^5\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^4.$$

Гипергеометрические распределения. Одним из видов распределения ДСВ являются гипергеометрические распределения, с вычислением вероятности которых мы познакомились в подразд. 1.7 (см. задачи 1.43—1.46).

Дискретная случайная величина X имеет ***гипергеометрическое распределение****, если она принимает значения* 1, 2, 3, ..., min (n, M) *с вероятностью, вычисляемой по формуле*

$$P(A) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n},$$

где $m = 1, 2, 3, ..., \min(n, M)$; $m \le N$, $n \le N$, n, M, $N \in \mathbb{N}$.

Пусть дискретная случайная величина X описывает m объектов, обладающих некоторым отличительным свойством по сравнению с n объектами случайной выборки без повторений (без возвратов), извлеченной из всего множества N объектов, обладающих этим свойством. Такое распределение дискретной случайной величины X называют *гипергеометрическим*.

? В чем главное отличие гипергеометрического распределения ДСВ от биномиального распределения?

Если в биномиальных распределениях проводились *независимые* повторные испытания, то в случаях гипергеометрических распределений повторные испытания являются *зависимыми*.

Числовые характеристики гипергеометрического распределения случайной величины X можно найти по формулам:

1) математическое ожидание

$$M(X) = n\frac{M}{N}; \tag{2.12}$$

2) дисперсия

$$D(X) = n\frac{M}{N-1}\left(1-\frac{M}{N}\right)\left(1-\frac{n}{N}\right). \qquad (2.13)$$

Функция распределения гипергеометрических распределений при $N \to \infty$ стремится к соответствующей функции распределения биномиальных распределений.

Гипергеометрические распределения широко применяются в работе отделов технического контроля (ОТК) при статистической оценке качества продукции с помощью выборочного обследования, когда производится *безвозвратная выборка.*

Задача 2.17. В партии, состоящей из 10 микрокалькуляторов, семь — стандартных. Контролер ОТК наудачу проверил два МК. Составить закон распределения числа обнаруженных стандартных МК и найти числовые характеристики этой случайной величины.

Решение. Число стандартных МК есть ДСВ. Обозначим ее X. Для подсчета вероятностей этих значений используем гипергеометрическое распределение (рис. 2.6).

Закон распределения этой ДСВ имеет вид:

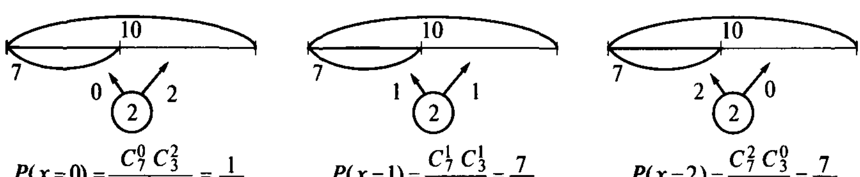
x_i	0	1	2
p_i	$\frac{1}{15}$	$\frac{7}{15}$	$\frac{7}{15}$

Найдем математическое ожидание и дисперсию по общей формуле:

$$M(X) = 0\cdot\frac{1}{15} + 1\cdot\frac{7}{15} + 2\cdot\frac{7}{15} = \frac{7}{15} \approx 1,4;$$

$$M(X^2) = 0\cdot\frac{1}{15} + 1\cdot\frac{7}{15} + 4\cdot\frac{7}{15} = \frac{7}{3} = 2\frac{1}{3}, \text{ тогда } D(X) \approx 2,33 - 1,96 = 0,37.$$

Среднеквадратическое отклонение: $\sigma(X) = \sqrt{0,37} \approx 0,6.$



$$P(x=0) = \frac{C_7^0\, C_3^2}{C_{10}^2} = \frac{1}{15} \qquad P(x=1) = \frac{C_7^1\, C_3^1}{C_{10}^2} = \frac{7}{15} \qquad P(x=2) = \frac{C_7^2\, C_3^0}{C_{10}^2} = \frac{7}{15}$$

Рис. 2.6

Сравним результат с соответствующими числовыми характеристиками, полученными по формулам, предназначенным для гипергеометрических распределений.

Имеем для $N = 10$, $M = 7$, $n = 2$, $M(X) = 2 \cdot \dfrac{7}{10} = 1{,}4$;

$$D(X) = 2 \cdot \frac{7}{9}\left(1 - \frac{7}{10}\right)\left(1 - \frac{2}{10}\right) = \frac{28}{75} \approx 0{,}37.$$

? В независимых повторных испытаниях находят вероятность ДСВ X — числа успехов, а можно ли найти вероятность ДСВ X — числа испытаний до первого успеха?

Например, ДСВ X — число испытаний до первого попадания в цель при стрельбе, ДСВ X — число испытаний до появления первого бракованного изделия при контроле качества и т.д.

Подобные распределения будут рассмотрены в следующем подразделе.

2.3. Геометрическое распределение

Вероятности p_k для значений 1, 2, ..., k образуют геометрическую прогрессию с первым членом p и знаменателем q, поэтому такое распределение называют геометрическим.

В последовательности независимых испытаний Бернулли (p — вероятность успеха в каждом испытании, q — вероятность неуспеха) рассмотрим случайную величину X — номер испытания, являющегося первым успехом. По смыслу X — ДСВ, так как множество реальных значений X является счетным множеством.

Дискретная случайная величина X имеет ***геометрическое распределение***, если она принимает значения 1, 2, 3, ... с вероятностями $P(X = k) = q^{k-1}p$. Геометрическое распределение будем обозначать $X \in G(p)$. Найдем ряд распределений ДСВ $X \in G(p)$, т.е. все пары вида $(k, p(k))$. Так как первый успех произошел в k-м испытании, то все предыдущие $k - 1$ испытаний являлись неуспехами. Тогда по свойству произведения независимых испытаний

$$P_k = \underbrace{qq\ldots q}_{k-1}\,p = q^{k-1}p, \text{ где } k = 1, 2, \ldots \tag{2.14}$$

$x_k = k$	1	2	3	...	k
p_k	p	qp	q^2p	...	$q^{k-1}p$

Проверим нормировку ряда:

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} q^{k-1}p = p\sum_{k=1}^{\infty} q^{k-1} = p\sum_{k=0}^{\infty} q^k = p\frac{1}{1-q} = \frac{p}{p} = 1,$$

так как ряд представляет собой убывающую ($p_k/p_{k-1} = q < 1$) *геометрическую прогрессию*. Поэтому это распределение называется ***геометрическим***. Вычислим основные характеристики ДСВ X:
математическое ожидание

$$\begin{aligned} M(X) &= \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1} = p\frac{d}{dq}\sum_{k=1}^{\infty} q^k = \\ &= p\frac{d}{dq}\left(\frac{q}{1-q}\right) = \frac{p}{(1-q)^2} = \frac{1}{p}; \end{aligned} \tag{2.15}$$

$$[M(X)]^2 = \sum_{k=1}^{\infty} k^2pq^{k-1} = p\frac{d}{dq}\sum_{k=1}^{\infty} kq^k = p\frac{d}{dq}\left(q\sum_{k=1}^{\infty} kq^{k-1}\right) =$$

$$= p\frac{d}{dq}\left(q\frac{1}{(1-q)^2}\right) = p\frac{1+q}{(1-q)^3} = \frac{1+q}{p^2};$$

дисперсию

$$D(X) = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}. \tag{2.16}$$

Задача 2.18. Студент подготовил из 40 экзаменационных билетов 32 и мечтает, что преподаватель разрешит ему выбрать выученный билет. Составить ряд распределений числа X возможных попыток взять билет до появления первого «знакомого» билета, если преподаватель остановил студента после четвертой попытки. Найти числовые характеристики этой случайной величины.

Решение. Вероятность того, что студент возьмет выученный билет, равна 0,8. Случайная величина X — число испытаний до появления первого выученного билета. Составим ряд распределений, найдем функцию распределения ДСВ X, построим ее график. Найдем все числовые характеристики (ограничиться тремя-пятью испытаниями).

Обозначим через n — число «испытаний», через p — вероятность взять выученный билет. Тогда $p = 0{,}8$, $q = 1 - p = 0{,}2$. Найдем вероятности $p_4(1)$, $p_4(2)$, $p_4(3)$, $p_4(4)$.

Так как случайная величина X — число возможных попыток до появления первого выученного билета, воспользуемся геометрической вероятностью:

$$P_4(1) = p = 0{,}8;$$

$P_4(2) = qp = 0{,}8 \cdot 0{,}2 = 0{,}16;$

$P_4(3) = q^2p = 0{,}2^2 \cdot 0{,}8 = 0{,}04 \cdot 0{,}8 = 0{,}032;$

$P_4(4) = q^3p = 0{,}2^3 \cdot 0{,}8 = 0{,}008 \cdot 0{,}8 = 0{,}0064.$

x_i	1	2	3	4
p_i	0,8	0,16	0,032	0,0064

Математическое ожидание $M(X) = \dfrac{1}{p} = 1{,}25;$

дисперсия $D(X) = \dfrac{q}{p^2} = \dfrac{0{,}2}{0{,}64} = 0{,}03125;$

среднеквадратическое отклонение $\sigma = \sqrt{0{,}03125} = 0{,}176.$

? Почему в рассмотренной задаче не выполняется условие нормировки: $\sum\limits_{k=1}^{4} p_k = 0{,}9984 \neq 1$?

Дело в том, что в случае геометрических распределений *условие нормировки* выполняется при $k \to \infty$, а в рассмотренной задаче были даны лишь четыре первые значения ДСВ X.

2.4. Закон распределения Пуассона

В этом подразделе рассмотрим распределения Пуассона, которые характеризуют количество событий, появившихся за некоторый промежуток времени или пространства (например, количество ДТП за неделю, количество телефонных звонков за час, количество прорывов газопровода на всем его протяжении и т.д.).

Формула Бернулли для биномиального распределения случайной величины при больших значениях m и n или при малых значениях p и q дает значительные вычислительные погрешности при округлениях.

Если в выражении $P_n(k)$ для биномиального распределения ДСВ $P_n(k) = C_n^k p^k (1-p)^{n-k}$ зафиксировать значение k и устремить n к бесконечности так, что произведение np является постоянным числом λ (т.е. $np = \lambda = \text{const}$), то $\lim\limits_{n\to\infty} P_n(k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$, где $k \in \mathbb{N}$.

Дискретная случайная величина X имеет ***распределение Пуассона***, *если она принимает значения* 0, 1, 2, ..., k, ...*с вероятностью*

$$P(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}. \tag{2.17}$$

k	0	1	2	...	r	...
P_k	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\lambda^2 e^{-\lambda}/2$	...	$\lambda^r e^{-\lambda}/r!$	...

Случайную величину X, имеющую распределение Пуассона, принято обозначать $X \in \Pi(\lambda)$. Легко проверить нормировку распределения Пуассона для постоянных λ с помощью рядов:

$$p_0 + p_1 + p_2 + \ldots = e^{-\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \ldots\right) = e^{-\lambda}e^{\lambda} = 1,$$

где выражение в скобках представляет собой разложение e^{λ} в ряд Тейлора в нуле. Этот ряд сходится к e^{λ} при любом λ.

График функции распределения Пуассона для отдельного значения параметра λ представляет собой дискретный ряд точек (рис. 2.7). Если их соединить ломаной, то получится *многоугольник распределений*. Разным значениям λ соответствуют разные распределения Пуассона. Значения $\dfrac{\lambda^k}{k!}e^{-\lambda}$ при различных значениях λ рассчитаны и обычно приводятся в справочной литературе в виде таблиц.

Так как биномиальные распределения при определенных условиях ($np = \lambda$) переходят в распределения Пуассона, то последние являются *предельными* для биномиальных распределений. Тот факт, что в распределениях Пуассона вероятность стремится к нулю, дает основание называть это распределение «*законом редких явлений*».

Итак, ***функция распределения Пуассона*** имеет вид

$$F(x) = \begin{cases} 0 & \text{при } x \le 0; \\ \sum\limits_{0 \le k < x} \dfrac{\lambda^k}{k!}e^{-\lambda} & \text{при } x > 0. \end{cases} \quad (2.18)$$

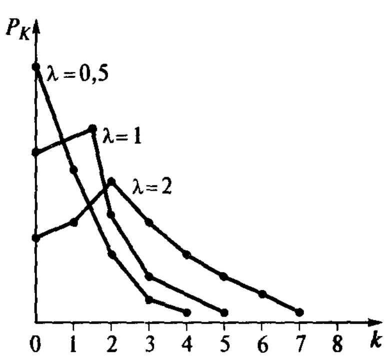


Рис. 2.7

? Если распределение Пуассона характерно для *редких явлений*, то по-видимому, важно уметь находить вероятность появления события, произошедшего *хотя бы один раз* при n независимых повторных испытаниях.

Вероятность появления события, произошедшего хотя бы один раз при n независимых повторных испытаниях, можно найти, учитывая, что события $k \ge 1$ и $k < 1$ (т.е. $k = 0$) — противоположные. Поэтому $P(n,k \geq 1) = 1 - P(n,k<1) = 1 - P(n,0) = 1 - e^{-\lambda}$, т.е. $P(n,k \geq 1) = 1 - e^{-\lambda}$.

Математическое ожидание и дисперсия СВ, имеющей распределение Пуассона, равны параметру λ:

$$M(X) = \lambda, \; D(X) = \lambda. \tag{2.19}$$

Действительно,

$$M(X) = \sum_{m=0}^{\infty} mP(m) = e^{-\lambda} \sum_{m=0}^{\infty} m \frac{\lambda^m}{m!} = e^{-\lambda} \sum_{m=1}^{\infty} m \frac{\lambda^m}{m!} =$$

$$= e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^{m-1}\lambda}{(m-1)!} = e^{-\lambda}\lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda.$$

Аналогично $D(X) = M(X^2) - [M(X)]^2 = M(X)^2 - \lambda^2$.

Вычислим $M(X^2)$:

$$M(X^2) = e^{-\lambda} \sum_{m=0}^{\infty} m^2 \frac{\lambda^m}{m!} = e^{-\lambda} \sum_{m=1}^{\infty} m^2 \frac{\lambda^m}{m!} = e^{-\lambda} \sum_{m=1}^{\infty} m \frac{\lambda^m}{(m-1)!} =$$

$$= e^{-\lambda}\lambda \sum_{m=1}^{\infty} m \frac{\lambda^{m-1}}{(m-1)!} = e^{-\lambda}\lambda \frac{\partial}{\partial\lambda} \sum_{m=1}^{\infty} \frac{\lambda^m}{(m-1)!} = e^{-\lambda}\lambda \frac{\partial}{\partial\lambda}\left(\lambda \sum_{m-1=0}^{\infty} \frac{\lambda^{m-1}}{(m-1)!}\right) =$$

$$= e^{-\lambda}\lambda \frac{\partial}{\partial\lambda}(\lambda e^{\lambda}) = \lambda^2 + \lambda,$$

откуда

$$D(X) = M(X^2) - [M(X)]^2 = \lambda. \tag{2.20}$$

По закону Пуассона распределяются, например, вызовы абонента в течение часа на телефонной станции, число дефектов вновь введенной автомобильной дороги и другие редкие явления.

Заметим, что распределения Пуассона применяют в *теории массового обслуживания* — разделе теории вероятностей, который изучает многие явления в экономических, социальных и технических областях — системы массового обслуживания (СМО): различные системы связи, системы снабжения, транспорт, медицинское обслуживание, торговлю и др. В теории массового обслуживания одним из главных является понятие ***потока событий*** — последовательности событий, наступающих в случайные моменты времени. В теории массового обслуживания рассматривается так называемый *простейший*, или *пуассоновский*, *поток* с интенсивностью λ, который обладает рядом характерных свойств, среди которых *число заявок* (событий), поступивших за единицу времени X, имеет *распределение Пуассона* с параметром λ, причем среднее число заявок, поступивших за время t, вычисляется по формуле $M(X_t) = \lambda t$.

На описываемые системы накладываются требования:

- вероятность появления события постоянна для любых двух интервалов одинаковой длины (свойство стационарности);
- вероятность, что событие произойдет в короткий временной интервал, пропорциональна величине интервала;
- вероятность, что в короткий временной интервал произойдут одновременно два события, равна нулю (свойство ординарности);
- вероятность появления событий во временно́м интервале не зависит от начала интервала;
- появление событий в различных временны́х интервалах не зависит одно от другого (отсутствие последействия).

? Потоки событий происходят во времени, а время непрерывно. Можно ли с помощью знаний теории вероятностей находить вероятности не дискретных, а непрерывных явлений?

Действительно, многообразный окружающий мир описывается законами, характерными не только для дискретных, но и для непрерывных функций. Одна из особенностей непрерывных случайных величин — невозможность указать два ее соседних значения, как и на числовой прямой невозможно указать координаты двух соседних точек. Начиная со следующего подраздела, будут рассматриваться непрерывные случайные величины.

2.5. Непрерывные случайные величины и их числовые характеристики

В этом подразделе дадим определения непрерывной случайной величины (НСВ), плотности вероятности и приведем формулы для вычисления числовых характеристик НСВ.

Функция распределения ДСВ изменяется скачкообразно: количество взошедших семян, бракованных изделий, попаданий в мишень и т.д. Но может оказаться, что функция распределения изменяется непрерывно и вообще не имеет скачков: рост или масса ребенка как функция времени, напряжение электрического тока в городской сети, выход из строя электроприборов и т.д. В таком случае СВ принимает не только определенные значения, но и все действительные значения в интервалах между определенными и называется непрерывной случайной величиной.

2.5.1. Плотность распределения вероятностей

В этом подразделе вы научитесь составлять функцию распределения НСВ (интегральную) и функцию плотности вероятности НСВ (дифференциальную), а также строить их графики, вычислять вероятность попадания НСВ в заданный интервал.

Случайная величина ξ называется ***непрерывной***, если функция ее распределения $F_\xi(x) = P(\xi < x)$, $\forall x \in \mathbb{R}$ представима в виде

$$F(x) = \int_{-\infty}^{x} f_\xi(t)dt, \tag{2.21}$$

т.е. существует функция $f_\xi(t)$, такая, что функция распределения представима в виде интеграла. В этом случае функция $f_\xi(t)$ называется ***плотностью вероятности*** НСВ ξ. Так же, как и для функции распределения, будем опускать индекс, указывающий на СВ там, где это не приведет к недоразумениям.

Каждому промежутку (a, b) из области значений НСВ соответствует определенная вероятность $P(a \le x < b)$ того, что значение этой случайной величины попало в указанный полуинтервал.

На промежутке (a, b) невозможно перечислить все значения случайной величины, так как множество его точек несчетно, но можно утверждать, что вероятность каждого отдельно взятого значения НСВ есть бесконечно малая величина (из геометрического определения вероятности). Поэтому для НСВ имеет смысл говорить о вероятности не конкретного значения, а некоего интервала значений.

? Означает ли это, что вероятность некоторого возможного события равна нулю? До сих пор равной нулю была вероятность только невозможного события! Нет ли здесь противоречия?

На самом деле в этом, парадоксальном на первый взгляд, факте, нет ничего удивительного. Понять причины исследуемого явления помогут представления о плотности вероятности и знание дифференциального исчисления.

Так как плотность вероятности НСВ ξ является подынтегральной функцией для функции распределения, а с помощью функции распределения находится вероятность попадания в некоторый интервал $P(a \le x < b) = F(b) - F(a)$, то можно найти вероятность попадания в бесконечно малый интервал Δx (как «интеграл в отдельной точке», т.е. с равными пределами интегрирования). Очевидно, $P(a \le x < a) = F(a) - F(a) = 0$. И с геометрической точки зрения зная, что вероятность попадания в заданный интервал представляется в виде площади прямоугольной трапеции, то вероятность попадания в Δx стремится к нулю.

Из курса дифференциального исчисления известно, что монотонная ограниченная функция имеет не более чем счетное множество точек разрыва, а также производные всюду, за исключением не более чем счетного множества. На этом множестве должны существовать соответствующие односторонние производные.

Свойства функции распределения НСВ:

1) функция распределения непрерывна на $\mathbb{R}$ как функция верхнего предела в несобственном интеграле с «закрепленным», хотя и бесконечным, нижним пределом: $F(x) \in C(\mathbb{R})$;

2) $P(a \le \xi < b) = F(b) - F(a)$, но так как $\forall \varepsilon > 0 \; P(x = b) < \varepsilon$ для НСВ, то

$$P(a \le \xi \le b) = F(b) - F(a). \tag{2.22}$$

Свойства плотности вероятности $f(x)$:

1) $$F'_x = \frac{d}{dx}\int_{-\infty}^{x} f(t)dt = f(x). \tag{2.23}$$

Эта формула, наряду с определением функции распределения, характеризует связь между функцией распределения и плотностью вероятности;

2) $f(x)$ определена почти всюду (множество K точек, где функция распределения не дифференцируема, не более чем счетно);

3) $f(x) \ge 0$, так как $F(x)$ — неубывающая функция;

4) $f(x)$ интегрируема на $\mathbb{R}$;

5) функция распределения есть та первообразная плотности вероятности, которая обращается в ноль на $-\infty$;

6) $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ как вероятность достоверного события;

7) вероятность нахождения величины ξ в интервале $-\infty \le a \le \xi \le$ $\le b \le +\infty$ равна

$$P\{a \le \xi \le b\} = F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \\ = \int_{a}^{b} f(x)dx \tag{2.24}$$

(по формуле Ньютона—Лейбница).

Закон распределения НСВ принято задавать с помощью функции плотности вероятности $f(x)$, так как часто $F(x)$ может не выражаться в элементарных функциях.

2.5.2. Числовые характеристики непрерывной случайной величины

В этом подразделе вы научитесь находить числовые характеристики НСВ, осуществлять поиск неизвестных данных НСВ по заданным.

Известно, что к основным числовым характеристикам случайной величины относятся мода, медиана, математическое ожидание и дисперсия.

Модой НСВ X называется такое ее значение, при котором плотность вероятности максимальная. Случайная величина может иметь несколько мод.

С геометрической точки зрения мода — значение аргумента x, при котором график функции плотности распределения принимает максимальное значение.

Нахождение моды — известная задача дифференциального исчисления поиска экстремума на множестве. Если функция $f(x)$ дифференцируема на интервале, то ищут «подозрительные» на локальный экстремум точки $\{x_{0i}\}$, а из них выбирают $\max f(x_{0i})$, который нужно сравнить со значениями $f(x)$ на границах интервала.

Медианой НСВ X называется такое ее значение μ, для которого равновероятно, что *случайная величина x больше или меньше* μ:

$$P(X<\mu)=P(X>\mu)=\frac{1}{2}.$$

В случае, когда ось симметрии кривой распределения $y=f(x)$ совпадает с прямой $x=a$, выполняется соотношение равной вероятности для $x=\mu$ в точке a, и тогда $\mu=a$. В общем случае медиана есть корень алгебраического уравнения $F(x)=\frac{1}{2}$ или интегрального уравнения $\int\limits_{-\infty}^{\mu} f(t)dt=\int\limits_{\mu}^{+\infty} f(t)dt=\frac{1}{2}$.

С геометрической точки зрения медиана делит площадь под графиком функции плотности вероятности на две равные части.

Математическим ожиданием НСВ называется интеграл

$$M(X)=\int\limits_{-\infty}^{\infty} xf(x)dx, \tag{2.25}$$

в том случае, если он существует и существует также интеграл $M|X|=\int\limits_{-\infty}^{\infty}|x|f(x)dx.$

Для функции от НСВ справедливо утверждение, что она распределена так же, как и сама НСВ. Поэтому *математическим ожиданием* функции G от НСВ X называется интеграл

$$M(G)=\int\limits_{-\infty}^{+\infty} G(t)f(t)dt. \tag{2.26}$$

С *геометрической точки зрения* математическое ожидание случайной величины равно абсциссе центра тяжести площади, ограниченной кривой распределения и осью абсцисс. В случае, когда кривая распределения *симметрична* относительно прямой $x = m$, математическое ожидание также совпадает с этой абсциссой. При этом математическое ожидание, мода и медиана равны между собой:

$$M(X) = \text{Mo} = \text{Me} = \mu.$$

Дисперсию НСВ X находят по формуле

$$D(X) = \int_{-\infty}^{\infty} [x - M(X)]^2 f(x)dx. \tag{2.27}$$

Дисперсия характеризует степень рассеяния значений случайной величины от среднего значения. Все свойства математического ожидания и дисперсии, сформулированные для ДСВ, сохраняются и для НСВ. Как и для ДСВ, формула дисперсии имеет и другой вид:

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - [M(X)]^2. \tag{2.28}$$

Задача 2.19. Плотность вероятности случайной величины X задана функцией распределения:

$$p(x) = \begin{cases} 0 & \text{при} \quad x \le 0; \\ Ax^2 & \text{при} \quad 0 < x \le 1; \qquad A = \text{const}; \\ 0 & \text{при} \quad x > 1. \end{cases}$$

1) Найти функцию распределения и построить ее график. 2) Построить график плотности вероятности. 3) Найти числовые характеристики НСВ X. 4) Найти вероятность попадания на интервал $\left(\frac{1}{3}; \frac{5}{3}\right)$.

Решение. 1) Найдем константу A из условия нормировки вероятности:

$$1 = \int_{-\infty}^{+\infty} p(t)\,dt = \int_0^1 At^2 dt = \frac{A}{3}, \text{ откуда } A = 3.$$

Очевидно, при $x \le F(x) = \int_{-\infty}^{x} p(t)\,dt = \int_0^x 3t^2 dt = x^3$, при $0 \le x \le 1$, т.е.

$$F(x) = \begin{cases} 0 & \text{при} \quad x < 0; \\ x^3 & \text{при} \quad x \in [0;1]; \\ 1 & \text{при} \quad x > 1. \end{cases}$$

2) Графики функции распределения (рис. 2.8, *а*) и плотности вероятности (рис. 2.8, *б*) представлены на рис. 2.8.

3) Найдем числовые характеристики:

$$M(X)=\int_a^b xf(x)=\int_0^1 3x^3dx=\frac{3}{4};$$

$$D(X)=\int_a^b x^2f(x)\,dx-M(X)^2=\int_0^1 3x^4dx-\left(\frac{3}{4}\right)^2=\frac{3}{5}-\frac{9}{16}\approx$$

$$\approx 0{,}6-0{,}56=0{,}04;$$

$$\sigma(X)=\sqrt{D}\approx\sqrt{0{,}04}=0{,}2.$$

Моду находят в процессе исследования на экстремум плотности вероятности $f(x)=p(x)=3x^2$ по известному алгоритму, причем при $x<0$ и $x>1$ моды заведомо нет.

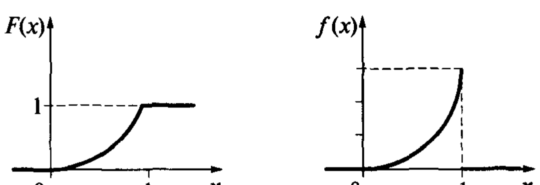
Согласно алгоритму исследования на экстремум плотности вероятности (ПВ) выполним действия:

- найдем производную $f'(x)=(3x^2)'=6x$ при $0<x<1$;
- приравняем производную к нулю и найдем критические точки: $6x=0$, т.е. $x=0$. Следовательно, внутри интервала (0; 1), где ПВ дифференцируема, моды нет;
- сравним значения ПВ в точках, где она недифференцируема, т.е. при $x=0$ и $x=1$: $p(0)=0$, $p(1)=3$. Следовательно, $\text{Mo}(X)=3$;
- для вычисления медианы решим уравнение $F(x>\mu)=F(x<\mu)=0{,}5$: очевидно, его корень должен находиться в интервале (0;1), т.е. $x^3=0{,}5$. Откуда $x\approx\sqrt[3]{0{,}5}\approx 0{,}79$. Поэтому медиана равна $\text{Me}(X)=0{,}79$.

4) Вероятность попадания случайной величины X в интервал $\left(\frac{1}{3};\frac{5}{3}\right)$ находится с учетом того, что $F\left(\frac{5}{3}\right)=1$.

Тогда $P\left\{\frac{1}{3}<x<\frac{5}{3}\right\}=F\left(\frac{5}{3}\right)-F\left(\frac{1}{3}\right)=1-\left(\frac{1}{3}\right)^3=1-\frac{1}{27}=\frac{26}{27}$.

С геометрической точки зрения вероятность попадания случайной величины X в интервал $\left(\frac{1}{3};\frac{5}{3}\right)$ равна площади криволиней-



а *б*

Рис. 2.8

ной трапеции, которая ограничена сверху кривой распределения, снизу — осью абсцисс, слева — прямой $x=\frac{1}{3}$, справа — прямой $x=\frac{5}{3}$.

К важнейшим непрерывным распределениям относятся *нормальное, равномерное и показательное распределения*, с которыми вы познакомитесь в следующих подразделах.

2.6. Нормальное распределение и его числовые характеристики

В этом подразделе вы научитесь составлять функцию распределения отдельных видов НСВ и строить их графики, находить числовые характеристики данного вида НСВ, узнавать по содержанию задания вид НСВ в реальной ситуации.

Самым распространенным в природе, экономике, социологии и других науках является нормальное распределение непрерывной случайной величины. С помощью нормального распределения можно описать плотность вероятности НСВ в тех случаях, когда отклонения от средней случайной величины появляются за счет различных явлений, воздействующих независимо одно от другого, но примерно в одинаковой степени, причем, чем больше суммируется таких случайных величин, тем результат точнее. Все эти явления не зависят друг от друга, но, воздействуя на процесс изготовления примерно с одинаковой силой, обусловливают то, что закон, по которому изменяется НСВ (например, размер конкретной детали), описывается *нормальным* распределением.

Самое точное изготовление детали с заданными размерами — «*эталон*» — будет соответствовать *математическому ожиданию* m, разброс фактических значений случайной величины размера детали — понятию *дисперсии* (точнее — *среднеквадратическому отклонению* σ).

Случайная величина с нормальным распределением существует в интервале $(-\infty;\ \infty)$ и описывается законами:

- плотности вероятности $f(x)$, называемой «кривой Гаусса» (рис. 2.9, *а*)

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}, \qquad (2.29)$$

где σ и m — параметры нормального распределения, причем $\sigma>0$;

- функции распределения $F(x)$ (рис. 2.9, *б*):

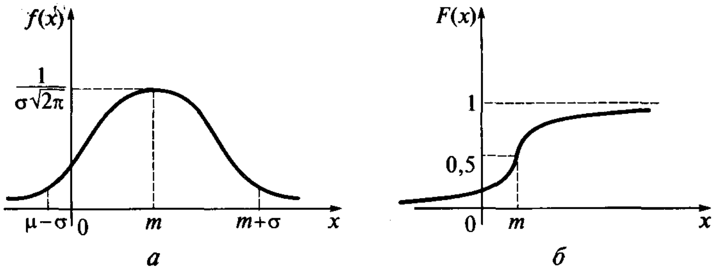


Рис. 2.9

$$F(x)=\int_{-\infty}^{x} f(y)dy=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}}dy. \qquad (2.30)$$

Подстановкой $t=\dfrac{y-m}{\sigma}$ интеграл приводится к виду

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-m}{\sigma}} e^{-t^2/2}dt=0{,}5+\frac{1}{\sqrt{2\pi}}\int_{0}^{\frac{x-m}{\sigma}} e^{-t^2/2}dt.$$

Поэтому для удобства вводится нечетная функция $\Phi(x)=$ $=\dfrac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt$, называемая ***функцией Лапласа***. Функцию Лапласа называют также «***интегралом вероятности***», или «***функцией ошибок***». Очевидно, что $\Phi(0)=0$, $\Phi(+\infty)=1/2$, $\Phi(-x)=-\Phi(x)$.

Математическое ожидание $M(X)$ случайной величины X, распределенной нормально, равно

$$M(X)=\int_{-\infty}^{+\infty} xf(x)dx=m, \qquad (2.31)$$

дисперсия равна

$$D(X)=\int_{-\infty}^{+\infty}(x-m)^2 f(x)dx=\sigma^2, \qquad (2.32)$$

поэтому параметр σ — среднеквадратическое отклонение.

Случайную величину X, распределенную нормально с параметрами σ и m, обозначают $X\in N(m,\sigma)$.

На практике для вычисления значений функции Лапласа используются таблицы, которые приводятся в справочной литературе (табл. П. 3).

Вероятность попадания в интервал (α,β) НСВ, распределенной по нормальному закону, можно найти с помощью функции Лапласа $\Phi(x)$ по формуле

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - m}{\sigma}\right) - \Phi\left(\frac{\alpha - m}{\sigma}\right).$$

Величины параметров нормального распределения СВ X непосредственно влияют на форму кривой $f(x)$: при $x = m$ она принимает максимальное значение, равное $\frac{1}{\sigma\sqrt{2\pi}}$. Поэтому с увеличением (уменьшением) σ максимальная ордината убывает (возрастает) и кривая становится более пологой, приближаясь к оси Ox.

Величина математического ожидания m влияет на расположение кривой $f(x)$ относительно оси ординат: при возрастании (убывании) m кривая смещается вправо (влево).

Поэтому с помощью подстановки $t = \frac{x - m}{\sigma}$ можно получить функцию $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ плотности вероятности, график которой симметричен относительно оси Oy. Такая кривая соответствует нормированному закону нормального распределения с параметрами $m = 0$ и $\sigma = 1$. Величину $X \in N(0,1)$ называют стандартно нормальной. Ее функция распределения имеет вид $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ (интеграл Лапласа, см. табл. П. 3).

? С функцией Лапласа и табл. П. 3 мы уже встречались в подразд. 1.13, но там значения аргумента и вероятности определялись по другим формулам.

Действительно, функцию Лапласа применяют при подсчете вероятностей для биномиальных распределений ДСВ, при больших значениях n. Возможность использования функции Лапласа для биномиальных распределений ДСВ может быть обоснована законом больших чисел (см. подразд. 2.10).

2.7. Равномерные распределения

В этом подразделе вы познакомитесь с равномерным распределением НСВ, которое встречается в тех ситуациях, когда мы имеем дело с различными циферблатами (часы, весы, физические приборы и т.д.). Также равномерное распределение возникает как распределение ошибок при округлении чисел.

Равномерным называется распределение таких случайных величин, все значения которых лежат на некотором отрезке [a; b] и имеют постоянную плотность вероятности на этом отрезке.

Плотность вероятности задается формулой

$$f(x)=\begin{cases}0 & \text{при}\, x<a;\\ h=\text{const} & \text{при}\, a\le x\le b;\\ 0 & \text{при}\, x>b.\end{cases} \qquad (2.33)$$

При равномерном распределении график плотности вероятности представлен на рис. 2.10.

Поскольку площадь прямоугольника равна $h(b-a)=1$, то $h=\dfrac{1}{b-a}$. Тогда функцию плотности вероятности можно записать в другом виде:

$$f(x)=\begin{cases}0 & \text{при}\;\; x<a;\\ \dfrac{1}{b-a} & \text{при}\;\; a\le x\le b;\\ 0 & \text{при}\;\; x>b.\end{cases} \qquad (2.34)$$

Видно, что равномерное распределение задается указанием соответствующего отрезка, поэтому СВ X, распределенная равномерно на [a, b], обозначается $X\in U[a,\,b]$ (от англ. *uniform* — равномерный).

Для равномерного распределения нельзя указать моду, так как все вероятности принимают одинаковые значения.

Математическое ожидание НСВ при равномерном распределении равно:

$$M(X)=\int_a^b \frac{x}{b-a}dx=\frac{1}{b-a}\frac{x^2}{2}\bigg|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{b+a}{2}, \qquad (2.35)$$

т.е. среднему арифметическому концов отрезка: $M(X)=\dfrac{a+b}{2}$.

И действительно, используя начальные сведения о симметрии прямоугольника, можно было предположить, что математическое ожидание при равномерном распределении есть среднее арифметическое концов отрезка.

Дисперсию равномерного распределения найдем по формуле

$$D(X)=M(X^2)-[M(X)]^2.$$

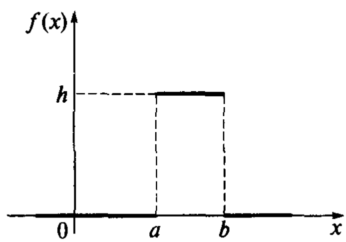


Рис. 2.10

Действительно, $M(X^2)=\int\limits_a^b \frac{x^2dx}{b-a}=\frac{1}{b-a}\frac{x^3}{3}\Bigg|_a^b=\frac{b^2+ab+a^2}{3}$. Тогда

$$D(X)=\frac{b^2+ab+a^2}{3}-\frac{b^2-2ab+a^2}{4}=\frac{(b-a)^2}{12}. \qquad (2.36)$$

Среднеквадратическое отклонение

$$\sigma(X)=\sqrt{\frac{(b-a)^2}{12}}=\frac{b-a}{2\sqrt{3}}, \qquad (2.37)$$

так как $b>a$ по условию.

Случайные величины с равномерным распределением встречаются в тех случаях, когда по условиям эксперимента случайная величина X принимает значения в конечном промежутке $[a;\ b]$, причем все значения *равновероятны*:

- X — время ожидания автобуса на остановке (случайная величина X равномерно распределена на отрезке $[0;\ b]$, где b — интервал движения между автобусами);
- X — ошибка при взвешивании некоторого предмета, полученная при округлении результата до целого значения (в этом случае $x\in[-0{,}5;\ 0{,}5)$, если цена деления шкалы весов равна единице) и др.

Функция распределения НСВ, равномерно распределенной на отрезке $[a;\ b]$, имеет вид

$$F(x)=\begin{cases}0 & \text{при}\quad x<a;\\ \dfrac{x-a}{b-a} & \text{при}\quad a\le x\le b;\\ 1 & \text{при}\quad x>b.\end{cases} \qquad (2.38)$$

Ее график изображен на рис. 2.11.

Если некоторый отрезок $[\alpha;\ \beta]$ длиной l целиком содержится в отрезке $[a;\ b]$, то ***вероятность попадания в него случайной величины*** $X\in U[a,\ b]$, распределенной равномерно, находят по формуле

$$P\{\alpha\le x\le\beta\}=\int\limits_\alpha^\beta f(x)dx=\int\limits_\alpha^\beta\frac{dx}{b-a}=\frac{1}{b-a}(\beta-\alpha)=\frac{l}{b-a}, \qquad (2.39)$$

где l — длина отрезка $[\alpha;\ \beta]$.

Это значит, что вероятность попадания в любую фиксированную область отрезка $[a;\ b]$ пропорциональна длине этой области. Вероятность попадания в область $[\alpha;\ \beta]$ можно записать и в виде

$$P(\alpha\le x\le\beta)=\frac{\mathrm{mes}[\alpha;\ \beta]}{\mathrm{mes}[a;\ b]},$$

что соответствует геометрическому определению вероятности. Отсюда можно сделать вывод о том, что непрерывное равномерное распределение влечет за собой геометрические вероятности.

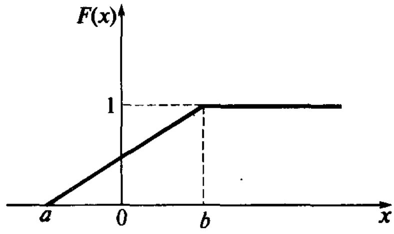


Рис. 2.11

Задача 2.20. НСВ X задана функцией распределения

$$F(x)=\begin{cases}0 & \text{при } x\le -2;\\ A(x+2) & \text{при } -2<x\le 3,\ A=\text{const};\\ 1 & \text{при } x>3.\end{cases}$$

Найти: 1) вероятность того, что в результате испытания X примет значение в интервале (0; 2); 2) плотность вероятности $f(x)$; 3) числовые характеристики НСВ X.

Решение. 1) Найдем постоянную A. Дифференцируя ФР, найдем следующее выражение для плотности вероятности:

$$F'(x)=f(x)=\begin{cases}0 & \text{при } x\le -2;\\ A & \text{при } -2<x\le 3;\\ 0 & \text{при } x>3,\end{cases}$$

откуда следует, что X равномерно распределена на [–2; 3], тогда $A=\dfrac{1}{5}$.

2) Найдем вероятность попадания значения X в интервал (0; 2):

$$P\{0<x<2\}=F(b)-F(a)=\left(\frac{x}{5}+\frac{2}{5}\right)\Big|_0^2=\frac{2}{5}.$$

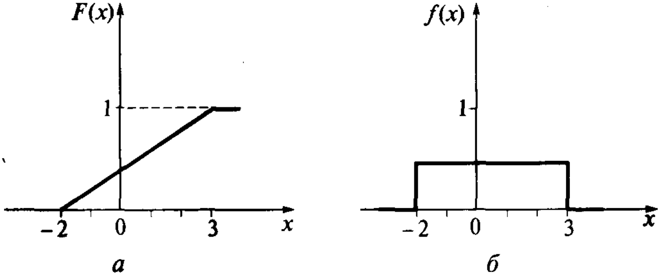


Рис. 2.12

3) Математическое ожидание

$$M(X) = (-2+3)/2 = 1/2;$$

среднеквадратическое отклонение $\sigma(X) = 5/(3\sqrt{2})$.

График функции распределения $F(x)$ приведен на рис. 2.12, *а*, а плотности вероятности — на рис. 2.12, *б*.

2.8. Показательное распределение

В этом подразделе будет рассмотрено показательное распределение НСВ, которое встречается, когда имеют дело с распределением времени совершенно случайных событий.

Задача 2.21. Найти закон распределения времени перегорания лампочки, если вероятность Λ ее перегорания не зависит от предыдущей работы лампочки, а зависит только от будущей работы.

Решение. Введем обозначение событий:

A — лампочка не перегорела до момента времени t [$P(A)$ — вероятность того, что лампочка не перегорела до момента времени t];

B — лампочка перегорит в следующие Δt единиц времени;

C — лампочка перегорит в промежуток времени $(t, t+\Delta t)$.

Очевидно, что $P(C) = P(B|A) = P(AB)/P(A) = \Lambda(\Delta t)$.

Пусть СВ X — время замены лампочки. Очевидно, что момент перегорания лампочки может быть абсолютно произвольным, начиная с некоторого момента начала отсчета времени: сделаем его равным нулю.

Таким образом, $0 < t < t+\Delta t$. Тогда X — НСВ. Обозначим ее функцию распределения $F(x)$, а плотность вероятности — $f(x)$. С учетом возможности произвольного выбора начала отсчета $f(x) = 0$ при $x < 0$. Поэтому далее будем рассматривать плотность вероятности и функцию распределения только при $x > 0$.

Тогда $P(A) = 1 - F(t)$.

$$P(AB) = F(t+\Delta t) - F(t) = \int_t^{t+\Delta t} f(x)\,dx.$$

При малых Δt вероятность $\Lambda(\Delta t)$ должна быть пропорциональна самому интервалу, т.е. $\Lambda(\Delta t) \approx \lambda\Delta t$.

$$P(C) = \frac{P(AB)}{P(A)} = \frac{\int_t^{t+dt} f(t)\,dt}{1-F(t)} = \Lambda.$$ Разделим части уравнения на Δt и вычислим предел при $\Delta t \to 0$:

$$\lim_{\Delta t \to 0} \frac{\frac{1}{\Delta t}\int\limits_{t}^{t+\Delta t} f(x)\,dx}{1-F(t)} = \frac{f(t)}{1-F(t)} = \lambda.$$

Вспомнив, что $f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}(F(t)-1)$, получим дифференциальное уравнение с разделяющимися переменными $\frac{d(F(t)-1)}{F(t)-1} =$ $= -\lambda dt$, которое легко интегрируется: $F(t) - 1 = Ce^{-\lambda t}$. Постоянную интегрирования найдем из условия $F(0) = 0$. Тогда $C = -1$ и получим

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le 0; \\ 1 - e^{-\lambda x} & \text{при} \quad x \ge 0. \end{cases}$$

Итак, ***показательным*** (*экспоненциальным*) называют распределение непрерывной случайной величины X, которое описывается *функцией распределения*:

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le 0; \\ 1 - e^{-\lambda x} & \text{при} \quad x \ge 0. \end{cases} \tag{2.40}$$

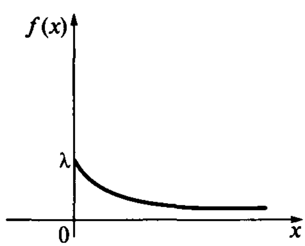
Тогда *плотность вероятности* показательного распределения равна

$$f(x) = \begin{cases} 0 & \text{при} \quad x < 0; \\ \lambda e^{-\lambda x} & \text{при} \quad x \ge 0, \end{cases} \tag{2.41}$$

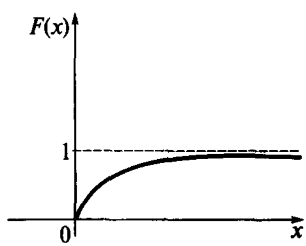
где λ — постоянная положительная величина.

Графики плотности вероятности (рис. 2.13, *а*) и функции распределения (рис. 2.13, *б*) приведены на рис. 2.13.

Показательное распределение, обозначаемое $X \sim E(\lambda)$, широко применяется в приложениях, в частности, в теории надежности,



а



б

Рис. 2.13

одним из основных понятий которой является *функция надежности*. Также закон показательного распределения используется в системах массового обслуживания. В частности, по показательному закону изменяется промежуток времени между двумя последовательными событиями или соседними заявками для простейших потоков в теории массового обслуживания (см. подразд. 2.4).

Пусть непрерывная случайная величина X распределена по показательному закону. Найдем ее математическое ожидание:

$$M(X)=\int_0^{+\infty}\lambda x e^{-\lambda x}dx=\frac{1}{\lambda}\int_0^{+\infty}ye^{-y}dy=\frac{1}{\lambda}. \tag{2.42}$$

Таким образом, *математическое ожидание* показательного распределения равно величине, обратной параметру λ. Поэтому, если переменная x — время некоего процесса, то $M(X)=1/\lambda$ имеет смысл времени релаксации этого процесса, когда плотность вероятности уменьшается в e раз. Тогда вероятность, что событие произойдет (например, лампочка перегорит) за время $1/\lambda$, равна $F(1/\lambda)-F(0)=1-1/e=0{,}632$.

Найдем *дисперсию*:

$$D(X)=\int_0^{+\infty}\lambda x^2e^{-\lambda x}dx-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}\left(\int_0^{+\infty}y^2e^{-y}dy-1\right)=\frac{1}{\lambda^2}, \tag{2.43}$$

откуда $M(X)=\sigma(X)=1/\lambda$.

? Получили, что показательное распределение определяется всего одним параметром λ.

Действительно, замеченная нами особенность показательного распределения указывает на его преимущество по сравнению с распределениями, зависящими от большего числа параметров. Обычно параметры неизвестны и в математической статистике приходится находить их оценки (приближенные значения). Очевидно, проще оценить один параметр, чем два, три и т.д.

Найдем вероятность попадания НСВ X, распределенной по показательному закону, на интервал $(a;\ b)$. Воспользуемся общей формулой вероятности попадания НСВ X на заданный интервал:

$P(a<X<b)=F(b)-F(a)$, где $F(b)=1-e^{-\lambda b}$; $F(a)=1-e^{-\lambda a}$.
Тогда имеем

$$P(a<X<b)=F(b)-F(a)=1-e^{-\lambda b}-(1-e^{-\lambda a})=e^{-\lambda a}-e^{-\lambda b}.$$

Функция надежности. Назовем ***элементом*** некоторое устройство вне зависимости от степени его сложности. Рассмотрим время работы такого элемента в *системах массового обслуживания*. Обозначим через $t=0$ начальный момент времени работы элемента. Пусть через некоторое время T произошел отказ в его работе. Тогда T — непрерывная случайная величина — время безотказной работы элемента. Вероятность отказа работы элемента за некоторое время продолжительностью t определяется ***функцией распределения*** $F(t) = P(T < t)$.

Поэтому *вероятность безотказной работы* за время t как событие, противоположное вероятности отказа, можно найти по формуле $R(t) = P(T > t) = 1 - F(t)$, где $R(t)$ — ***функция надежности***. Продолжительность безотказной работы элемента — случайная величина, соответствующая показательному закону распределения НСВ: $F(t) = 1 - e^{-\lambda t}$, где λ — интенсивность отказов, т.е. среднее число отказов в единицу времени. Поэтому функция надежности принимает вид

$$R(t) = 1 - F(t) = e^{-\lambda t}. \tag{2.44}$$

Функцию надежности, заданную зависимостью $R(t) = e^{-\lambda t}$, называют *показательным законом надежности* и используют для вычисления *вероятности безотказной работы* элемента за время t.

Основное свойство показательного закона надежности заключается в том, что вероятность безотказной работы элемента за время t не зависит от времени его работы до рассматриваемого временно́го интервала, а зависит лишь от длительности интервала t при известной интенсивности отказов λ. Таким образом, условная вероятность безотказной работы элемента на интервале времени t (при условии его безотказной работы до этого временно́го интервала) равна безусловной вероятности, т.е. прошлый опыт работы не учитывается.

Ввиду своей простоты функция надежности нашла широкое применение в СМО.

Задача 2.22. Длительность безотказной работы элемента распределена по показательному закону при $t > 0$: $F(t) = 1 - e^{-0,02t}$. Найти вероятность того, что за время $t = 40$ ч:

а) элемент не откажет;

б) элемент выйдет из строя.

Решение. Условие задачи указывает на показательный закон надежности для времени $t = 40$ ч.

Вероятность безотказной работы за 40 ч равна

$$R(40) = e^{-0,02 \cdot 40} = e^{-0,8} = 0,45.$$

Поэтому вероятность противоположного события — отказа или выхода из строя элемента за 40 ч — $F(40) = 1 - 0,45 = 0,55$.

Задача 2.23. Время t — случайная величина, за которую в железнодорожном депо расформировывают состав через горку. За 1 ч на горке может быть расформировано шесть поездов. Определить вероятность того, что время, за которое состав будет расформирован:

а) меньше 0,5 ч;

б) меньше 24 мин, но больше 6 мин.

Решение. Так как вероятность расформировать состав подчиняется показательному закону, обозначим среднее число поездов в час как интенсивность отказов: $\lambda = 6$ ч$^{-1}$.

За время $t = 0{,}5$ ч вероятность расформировать состав равна

$$F(0{,}5) = P(T < 0{,}5) = 1 - e^{-6\cdot 0{,}5} = 1 - e^{-3} = 1 - 0{,}05 = 0{,}95.$$

Вероятность того, что время, за которое состав будет расформирован, при $t_1 = 6$ мин $= 0{,}1$ ч, $t_2 = 24$ мин $= 0{,}4$ ч составляет

$$P\{0{,}1 < T < 0{,}4\} = e^{-6\cdot 0{,}1} - e^{-6\cdot 0{,}4} = 0{,}55 - 0{,}09 = 0{,}46.$$

2.9. Распределения, связанные с нормальными

Кроме рассмотренных в предыдущих подразделах законов распределения НСВ, на практике чаще всего применяют два «специальных» закона распределения, получаемые из нормального — распределения Пирсона (хи-квадрат) и Стъюдента.

Нормальное распределение является исключительным в теории вероятностей, поскольку при достаточно общих условиях многие распределения стремятся по вероятности к нормальному (см. подразд. 2.10).

2.9.1. Распределение χ^2 (распределение Пирсона)

Распределение хи-квадрат, или распределение Пирсона, названо по имени английского математика К. Пирсона (1857—1936). С увеличением числа степеней свободы распределение Пирсона медленно приближается к нормальному.

Пусть независимые случайные величины $X_1, X_2, \ldots, X_k$ являются стандартно нормально распределенными величинами, т.е.

$$X_i \in N(0, 1), \text{ где } i = 1, 2, \ldots, k.$$

Распределение случайной величины

$$X(k) = \chi^2(k) = \sum_{i=1}^{k} X_i^2 = X_1^2 + X_2^2 + \ldots + X_k^2$$

называется ***распределением хи-квадрат с k степенями свободы***, а сама величина $\chi^2(k) \geq 0$ — величиной *хи-квадрат с k степенями свободы*. Если величины $X_1, X_2, \ldots, X_k$ не являются независимыми (т.е. между ними существует $l \geq 1$ функционально независимых уравнений связи), то *число независимых случайных величин* будет равно разности между числом суммируемых случайных величин и числом связей, ограничивающих свободу изменения этих величин, т.е. $k-l$. Если все уравнения связи линейны, то величина $\chi^2(k) = X_1^2 + X_2^2 + \ldots + X_k^2$ будет также распределена по χ^2, но с $k-l$ степенями свободы. Число степеней свободы является единственной числовой характеристикой таких случайных величин, следовательно, все числовые характеристики СВ $\chi^2(k)$ должны зависеть от единственного параметра k.

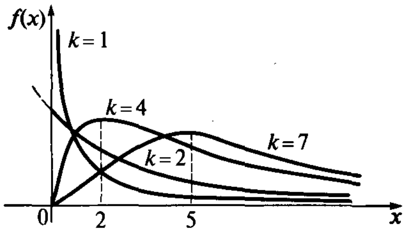


Рис. 2.14

Формула для определения функции распределения СВ $X(k) = \chi^2(k) = \sum_{i=1}^{k} X_i^2 = X_1^2 + X_2^2 + \ldots + X_k^2$ имеет довольно сложный вид:

$$F_{\chi^2}(x) = S_k(1)\frac{1}{2(2\pi)^{k/2}}\int_0^x e^{-\frac{t}{2}} t^{\frac{k-2}{2}}\,dt. \qquad (2.45)$$

Функция распределения позволяет написать выражение для плотности вероятности, которая оказывается элементарной функцией:

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ \dfrac{S_k(1)}{2(2\pi)^{k/2}} e^{-\frac{x}{2}} x^{\frac{k}{2}-1} & \text{при } x > 0, \end{cases}$$

где $S_k(1)$ — нормированный коэффициент, равный площади единичной $(k-1)$-мерной сферы.

Графики плотности вероятности для различных значений k представлены на рис. 2.14.

Из рис. 2.14 и уравнения (2.45) видно, что распределения с $k = 1$ и $k = 2$ являются особенными, так как не обращаются в ноль при $x = 0$. При $k = 1$ распределение не имеет моды, при $k = 2$ мода расположена в нуле: $\text{Mo}[\chi^2(2)] = 0$.

Свойства распределения хи-квадрат:

1) $\chi^2(k) \geq 0$;
2) $M[\chi^2(k)] = k$;
3) $D[\chi^2(k)] = 2k$;
4) $\text{Mo}[\chi^2(k)] = k - 2$;
5) если СВ $\chi^2(k_1)$ и $\chi^2(k_2)$ независимые, то их сумма имеет распределение хи-квадрат с числом степеней свободы $k_1 + k_2$, т.е.

$$\chi^2(k_1) + \chi^2(k_2) = \chi^2(k_1 + k_2).$$

Таким образом, $\chi^2(k)$-распределение зависит лишь от числа степеней свободы k, причем с увеличением k это распределение, согласно центральной предельной теореме (см. подразд. 2.12), медленно стремится к нормальному. Так, при $k > 30$ распределение случайной величины $Z = 2\chi^2 - \sqrt{2k}$ приближается к стандартному нормальному распределению $N(0,1)$.

2.9.2. Распределение Стьюдента

Распределение Стьюдента (псевдоним английского статистика В. Госсета) с увеличением числа степеней свободы быстро приближается к нормальному.

Пусть Z и V — независимые случайные величины, причем $Z \in N(0,1)$, а $V \in [\chi^2(k)]$.

Тогда распределение случайной величины $T_k = \dfrac{Z}{\sqrt{V/k}}$ называют St-распределением Стьюдента с k степенями свободы. С возрастанием числа степеней свободы это распределение быстро приближается к нормальному. График плотности вероятности распределения Стьюдента аналогичен графику нормального распределения (рис. 2.15).

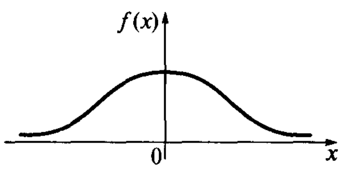


Рис. 2.15

Свойства распределения Стьюдента:

1) $M(T_k) = \mathrm{Mo}(T_k) = \mathrm{Me}(T_k) = 0$;

2) $D(T_k) = \dfrac{k}{k-2}$ и существует только при $k > 2$.

Значения χ^2-распределения и St-распределения (Стьюдента), зависящие лишь от степени свободы, табулированы.

2.10. Понятие о законе больших чисел

Закон больших чисел — это ряд теорем, в каждой из которых для различных условий устанавливают факт приближения средних характеристик большого числа опытов к постоянным неслучайным величинам.

В практической деятельности особое значение имеют такие события, осуществление которых можно предсказать, поскольку их вероятность близка к нулю или единице. Поэтому одной из главных задач теории вероятности является установление закономерностей, при которых вероятность их осуществления близка к нулю или единице. Так, если событие имеет очень малую вероятность (например, $p < 0{,}02$), то на практике в единичном испытании можно считать, что это событие не произойдет, несмотря на существование, пусть и малого, шанса того, что событие все же может произойти. Такую уверенность дает *принцип «практической невозможности маловероятных событий»* (I). Аналогично, если событие имеет вероятность, близкую к единице (например, $p > 0{,}98$), то при единичном испытании применяют *принцип «практической уверенности»* (II), который заключается в том, что можно считать: при такой большой вероятности событие обязательно произойдет.

? Как определить границу, при которой то или иное событие считают маловероятным (или достоверным)?

Необходимо учитывать последствия наступления этого события. Вероятность, с которой могло осуществиться некоторое событие, называют *уровнем значимости.*

Математические законы теории вероятностей получены в результате обобщения закономерностей массовых явлений природы и общества. Под массовостью в данном случае понимается значительное число повторяющихся испытаний в одинаковых или сходных условиях. При изучении массовых явлений особую роль играет группа теорем, известная в математике под названием «закон больших чисел». Благодаря этим теоремам, устанавливаются закономерности, возникшие в результате наложения большого числа случайных факторов. Так, одной из установленных закономерностей СВ является предсказуемость результатов: при определенных условиях СВ начинает себя вести не как случайная. При суммировании большого числа случайных величин закон распределения их суммы при соблюдении ряда условий близок к нормальному. Выявление условий, при которых можно применять перечисленные принципы (I и II), сформулированы в теоремах, которые называются *центральными предельными теоремами.*

С другой стороны, под законом больших чисел понимают давно установленное (наблюдаемое) свойство устойчивости массовых случайных явлений, смысл которого в том, что средний результат действия большого числа случайных явлений становится практически неслучайным, т. е. может быть достаточно точно предсказан. Благодаря закону больших чисел появляется возможность делать научные прогнозы случайных явлений с достаточно высокой точностью, а также оценивать точность этих прогнозов. В теории вероятностей под законом больших чисел принято понимать совокупность теорем, устанавливающих связь между средним арифметическим большого числа случайных величин и средним арифметическим их математических ожиданий. Различные формы закона больших чисел имеют большое практическое применение, так как составляют теоретическую базу математической статистики. Однако особую роль играют различные формы *центральной* *предельной теоремы*, так как устанавливают условия возникновения нормального закона распределения.

Этот закон распределения применяют в случае, когда исследуемую случайную величину можно представить в виде суммы достаточно большого числа независимых слагаемых. Так, некоторые экономические показатели представляют суммой большого числа слагаемых, каждое из которых дает незначительный вклад в общую сумму — некоторый показатель, распределение которого близко к нормальному.

В основе закона больших чисел лежит понятие сходимости случайных величин по вероятности.

Последовательность случайных величин X_1, X_2, ..., X_n, ... ***сходится по вероятности*** к случайной величине X, если для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1 \text{ или } \lim_{n\to\infty} P\{|X_n - X| \geq \varepsilon\} = 0.$$

Сходимость по вероятности можно записать символически: $X_n \xrightarrow{P} X$.

Рассмотрим некоторые формы закона больших чисел.

2.10.1. Неравенство Маркова

Неравенство Маркова названо по имени русского математика, ученика П.Л.Чебышева, А.А.Маркова (1856—1922).

Неравенство Маркова. *Для любой неотрицательной случайной величины X, имеющей конечное математическое ожидание, и $\forall\varepsilon > 0$, вероятность того, что она примет значение, превосходящее ε, не больше, чем отношение математического ожидания этой случайной величины $M(X)$ к значению ε:*

$$P\{X > \varepsilon\} \leq \frac{M(X)}{\varepsilon}, \text{ или } P\{X \leq \varepsilon\} > 1 - \frac{M(X)}{\varepsilon}. \qquad (2.46)$$

Задача 2.24. Средний срок службы холодильника пять лет. Оценить вероятность того, что данный холодильник не прослужит более 20 лет.

Решение. Пусть X — срок службы холодильника. Используя неравенство Маркова, имеем

$$P(X \leq 20) \geq 1 - 5/20 = 0{,}75.$$

Задача 2.25. Сумма вкладов в сбербанке составляет $4 \cdot 10^6$ р. Вероятность того, что случайно выбранный вклад не превышает 1000 р., равна 0,8. Оценить число вкладчиков в этом банке.

Решение. Пусть в банке n вкладчиков. Обозначим через X величину случайно снятого вклада. Найдем математическое ожидание вкладов отдельных вкладчиков $M(X) = 4 \cdot 10^6/n$. Пусть $\varepsilon = 1\,000 = 10^3$. Согласно условию задачи имеем $P|x < 1\,000| = 0{,}8$. Используя неравенство Маркова, эту вероятность можно записать в виде

$$0{,}8 = P(X \le \varepsilon) > 1 - \frac{4 \cdot 10^6}{10^3 n}.$$

Отсюда $4\,000/n > 0{,}2$, поэтому $n < 20\,000$. Таким образом в этом банке могут содержать вклады не более 20 тыс. вкладчиков.

Неравенство Маркова применяют для оценки вероятности неотрицательных случайных величин, для которых неизвестен закон распределения. Если $M(X) > \varepsilon$, то для оценки вероятности используют другие неравенства, так как $0 \le P\{X > \varepsilon\} \le 1$.

2.10.2. Неравенство Чебышева

В отличие от неравенства Маркова, неравенство П. Л. Чебышева (1821—1894) используется для любых, а не только неотрицательных случайных величин. Неравенство Чебышева устанавливает нижнюю границу вероятности и показывает, что эта граница зависит лишь от дисперсии.

Неравенство Чебышева. *Если СВ X имеет конечное математическое ожидание и дисперсию, то для любого* $\varepsilon > 0$ *справедливы неравенства:*

$$P\{|X - M(X)| \le \varepsilon\} > 1 - \frac{D(X)}{\varepsilon^2} \text{ и } P\{|X - M(X)| > \varepsilon\} \le \frac{D(X)}{\varepsilon^2}, \quad (2.47)$$

или вероятность того, что отклонение СВ X от ее математического ожидания по модулю не превосходит положительного числа $\varepsilon > 0$ ***больше чем на*** $1 - \frac{D(X)}{\varepsilon^2}$.

Задача 2.26. Стандартная длина детали $(50 \pm 0{,}32)$ см. Оценить вероятность того, что случайно взятая деталь окажется по длине не меньше, чем 49,5 см и не более 50,5 см.

Решение. По условию задачи имеем $M(X) = 50$ см, среднеквадратическое отклонение $\sigma = 0{,}32$. Тогда $D(X) = 0{,}32^2 = 0{,}1024$. Требуемое условие запишем, используя неравенство Чебышева для оценки вероятности

$P\{49{,}5 \le x \le 50{,}5\} = P\{|x - 50| \le 0{,}5\} \Rightarrow \varepsilon = 0{,}5$. Тогда имеем

$$P\{|x - 50| < 0{,}5\} > 1 - 0{,}1/0{,}5^2 = 1 - 0{,}4 = 0{,}6.$$

Задача 2.27. Электростанция обслуживает сеть из 15000 ламп, вероятность включения каждой из которых ранним осенним вечером равна 0,8. Какова вероятность того, что число ламп, включенных в сеть осенним вечером, отличается от своего математического ожидания не менее чем на 200?

Решение. По условию задачи имеем $p = 0{,}8$; $n = 15000$; $\varepsilon = 200$. Так как включения каждой из ламп независимы, то X — число включенных ламп — подчиняется биномиальному закону. Найдем числовые характеристики СВ X:

$$M(X) = np = 15000 \cdot 0{,}8 = 12000;\ D(X) = np(1-p) = \\ = 15000 \cdot 0{,}2 \cdot 0{,}8 = 2400.$$

Тогда, согласно неравенству Чебышева, имеем

$$P(|X - M(X)| > 200) \le 2400/200^2 = 0{,}06.$$

2.10.3. Теорема Чебышева

Теорема Чебышева кроме теоретического, имеет важное практическое значение. В математической статистике теорема Чебышева обосновывает выборочный метод, который дает возможность по характеристикам выборки изучать всю генеральную совокупность (см. подразд. 3.1).

Теорема Чебышева. *Если $X_1, X_2, \ldots, X_n$ — последовательность попарно независимых СВ, имеющих конечные математическое ожидание и дисперсию, ограниченную некоторой постоянной величиной C: $D(X_1) \le C$, $D(X_2) \le C$, ..., $D(X_n) \le C$, ..., то среднее арифметическое этих случайных величин сходится по вероятности к среднему арифметическому их математических ожиданий*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} M(X_i),\ \text{или для любого}\ \varepsilon > 0,$$

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i M(X_i)\right| < \varepsilon\right\} = 1.$$

Доказательство. Пусть $X_1, X_2, \ldots, X_n$ — последовательность независимых СВ, имеющих конечные математическое ожидание и дисперсию, ограниченную некоторой постоянной величиной C. Тогда используем неравенство Чебышева для оценки вероятности отклонения среднего значения СВ от среднего значения их математического ожидания. Это значит, что для любого $\varepsilon > 0$ справедливо утверждение:

$$P\left\{\left|\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i M(X_i)\right| < \varepsilon\right\} \ge 1 - \frac{D\left(\frac{1}{n}\sum_i X_i\right)}{\varepsilon^2} \ge 1 - \frac{C}{n\varepsilon^2}. \quad (2.48)$$

Действительно, если в неравенстве Чебышева учесть, что $\forall i$ имеет место неравенство $D(X_i) \le C$ и, значит $\sum D(X_i) \le nC$, то

$$\frac{D\left(\frac{1}{n}\sum_i X_i\right)}{\varepsilon^2} = \frac{\frac{1}{n^2} D\left(\sum_i X_i\right)}{\varepsilon^2} = \frac{D\sum_i X_i}{n^2\varepsilon^2} = \frac{\sum_i D(X_i)}{n^2\varepsilon^2} \le \frac{nC}{n^2\varepsilon^2} = \frac{C}{n\varepsilon^2}.$$

Переходя к пределу при $n \to \infty$, имеем

$$P\left\{\left|\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i M(X_i)\right| < \varepsilon\right\} \ge 1.$$

Но так как вероятность не превышает единицы, то справедливо

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i M(X_i)\right| < \varepsilon\right\} = 1.$$

Это означает, что среднее арифметическое $\bar{X}$ случайных величин $X_1, X_2, ..., X_n$ при $n \to \infty$ мало отличается от среднего арифметического их математических ожиданий $M(X_1), M(X_2), ..., M(X_n)$, т.е. обладает *свойством устойчивости*.

Для одинаково распределенных случайных величин с конечной дисперсией D обозначим математическое ожидание $M(X_i) = m$. Тогда теорема Чебышева принимает более простой вид:

$$P\left\{\left|\frac{1}{n}\sum_i x_i - m\right| < \varepsilon\right\} \ge 1 - \frac{D}{n\varepsilon^2}. \qquad (2.49)$$

Рассмотрим применение теоремы Чебышева на практике. В теории измерений, проведя n измерений некоторой величины x, можно получить различные значения $x_1, x_2, ..., x_n$. Тогда в качестве приближенного значения измеряемой величины x можно принять среднее арифметическое наблюдаемых значений $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i =$ $= \frac{1}{n}(x_1 + x_2 + ... + x_n)$. Причем, чем больше будет произведено опытов, тем точнее будет результат, т.е. при неточности измерительных приборов (большая дисперсия), увеличивая число измерений, теоретически можно получить результат со сколь угодно высокой точностью (но не превышающей точность прибора).

Задача 2.28. Для определения средней продолжительности горения лампы в партии из 200 одинаковых ящиков взяли на выбор по одной лампочке из каждого ящика. Оценить вероятность того, что продолжительность горения выборки (200 шт.) отличается от средней продолжительности горения всех ламп (партии) по модулю меньше, чем на 5 ч, если по ГОСТу стандартное отклонение не превышает 7 ч.

Решение. Пусть X_i — продолжительность горения электролампы, взятой из i-го ящика, так как $\sigma \le 7$ ч, $D_i = \sigma_i^2 \le (7 \text{ ч})^2 = 49 \text{ ч}^2 = C$.

Общая средняя продолжительность горения в выборке — $(x_1 + x_2 + \ldots + x_{200})/200$. Средняя продолжительность горения одной лампы в партии $m = (Mx_1 + Mx_2 + \ldots + Mx_{200})/200$. Оценим снизу вероятность

$$P\left(\left|\frac{1}{200}\sum_{i=1}^{200} x_i - m\right| < 5\right).$$

Поскольку лампы стандартные, то по теореме Чебышева со значениями параметров $C = 49 \text{ ч}^2$; $\varepsilon = 5$ ч; $n = 200$ получаем

$$P \ge 1 - 49 \text{ ч}^2/200 \cdot 5^2 = 1 - 0{,}0098 = 0{,}9902.$$

Задача 2.29. Сколько раз надо измерить данную величину, истинное значение которой a, чтобы с вероятностью $P \ge 0{,}95$ можно было утверждать, что среднее арифметическое значение этих измерений отличается от a по модулю меньше чем на 2, если прибор позволяет измерять со среднеквадратическим отклонением меньше 10?

Решение. Пусть X_i — результат i-го измерения. По условию $D(x) < 10^2 = C$.

Тогда надо найти n, при котором $P(|(x_1 + \ldots + x_n)/n - a| < 2) \ge 0{,}95$.

Так как $P\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - a\right| < \varepsilon\right) \ge 1 - \frac{C}{n\varepsilon^2}$, где $C = 100$, $\varepsilon = 2$, то

$$1 - 100/4n \le 0{,}95; \quad n \ge \frac{100}{4 \cdot 0{,}05} = 500,$$

т.е. надо произвести минимум 500 измерений.

2.10.4. Теорема Бернулли

Исторически первой формулировкой закона больших чисел была теорема Бернулли (1713), которая является частным случаем теоремы Чебышева.

Теорема Бернулли. *Пусть вероятность успеха некоторого события A из серии n независимых повторных испытаний равна $P(A) = p$, среди которых m — число успешных испытаний, тогда для любого $\varepsilon > 0$ имеет место предел*

$$\lim_{n\to\infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1.$$

Доказательство. Оценим вероятность того, что отклонение числа m появления события в n испытаниях от математического ожидания не превосходит некоторого $\varepsilon > 0$. Так как испытания независимы, то ДСВ X подчиняется закону распределения Бернулли и имеет числовые характеристики: $M(X) = np$, $D(X) = np(1-p) = npq$.

Рассмотрим другую величину $Y = X/n$, имеющую соответственно характеристики:

$$M(Y) = M(X/n) = 1/nM(X) = p;$$
$$D(Y) = D(X/n) = 1/n^2D(X) = pq/n.$$

Применим неравенство Чебышева для ДСВ Y. Тогда

$$P\left\{\left|\frac{m}{n} - p\right| \geq \varepsilon\right\} \leq \frac{pq}{n\varepsilon^2}, \text{ или } P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} > 1 - \frac{pq}{n\varepsilon^2}.$$

При увеличении числа экспериментов ($n \to \infty$) вероятность отклонения частоты от математического ожидания стремится к нулю: $P\left\{\left|\frac{m}{n} - p\right| \geq \varepsilon\right\} \leq \frac{pq}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$, а для противоположного события

$$P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} \xrightarrow[n\to\infty]{} 1. \qquad (2.50)$$

Эти записи являются разными формами записи теоремы Бернулли. Переходя к пределу, имеем $\lim\limits_{n\to\infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1.$

Задача 2.30. Из 1000 микрокалькуляторов, поступивших на базу, 200 было реализовано, и среди них 25 оказались бракованными. По этим данным как по случайной выборке определить вероятность того, что во всей партии бракованных микрокалькуляторов окажется не более 15 и не менее 10 %.

Решение. 1) Найдем статистическую вероятность (частоту) бракованных МК: $p^* = \frac{m}{n} = \frac{25}{200} = 0{,}125.$

2) Найдем наибольшее отклонение частоты появления бракованных МК от вероятности по модулю, если число $n = 1000$:

$$\left|\frac{m}{n} - p\right| = 0{,}025.$$

3) Найдем искомую вероятность по формуле

$$P\left(\left|\frac{x}{n} - p\right| < \varepsilon\right) > 1 - \frac{pq}{n\varepsilon^2}.$$

Тогда $P\left(\left|\frac{m}{n}-p\right|<0,025\right)>1-\frac{0,125(1-0,125)}{1000\cdot 0,025^2}=0,825,$

т.е. с вероятностью, большей, чем 0,825, можно утверждать, что бракованных микрокалькуляторов будет от 10 до 15 %.

Теорема Бернулли означает, что относительная частота числа m успехов в n испытаниях Бернулли обладает свойством *статистической устойчивости*, т.е. p^* мало отличается от вероятности, полученной с помощью расчетов. На этом свойстве основана надежность доверительных оценок вероятности в математической статистике.

Эта теорема дает возможность использовать известные формулы для расчета числовых характеристик СВ и применять соответствующие таблицы (функции Лапласа). А так как действие множества различных других законов распределения при их неограниченном увеличении также подчиняется закону нормального распределения, то ***центральная предельная теорема*** приобретает широкое практическое значение. Благодаря выводу из этой теоремы для взаимно независимых СВ при $n\to\infty$ можно найти числовые характеристики по формулам, соответствующим данному виду распределений, а оценивать вероятность их суммарного действия по формулам, соответствующим закону нормального распределения (через функцию Лапласа).

Пусть повторные испытания происходят с постоянной вероятностью p. Найти вероятность выполнения неравенства $\left|\frac{m}{n}-p\right|\le\varepsilon$, т.е. в серии из n испытаний частота m/n отличается по модулю от вероятности не более чем на $\varepsilon>0$.

Итак, надо найти значение вероятности $P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)$. Из неравенства $\left|\frac{m}{n}-p\right|\le\varepsilon$ имеем $-\varepsilon\le\frac{m-np}{n}\le\varepsilon$. Умножив на $\sqrt{\frac{n}{pq}}>0$, получим

$$-\varepsilon\sqrt{\frac{n}{pq}}\le\frac{(m-np)}{\sqrt{npq}}\le\varepsilon\sqrt{\frac{n}{pq}}.$$

Используя интегральную теорему Лапласа и нечетность функций Лапласа $\Phi(-x)=-\Phi(x)$, имеем

$$P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)\approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right). \tag{2.51}$$

Задача 2.31. Вероятность появления нестандартного элемента питания равна 0,05. Найти вероятность того, что среди случайно отобранных 500 элементов питания относительная частота появления нестандартного элемента питания отклонится от вероятности не более чем на 0,03.

Решение. По условию задачи имеем $p = 0{,}05$; $n = 500$; $\varepsilon = 0{,}03$. Тогда $q = 0{,}95$.

Оценим неизвестную вероятность по формуле

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

Тогда $P\left(\left|\frac{m}{500} - 0{,}05\right| \le 0{,}03\right) \approx 2\Phi\left(0{,}03\sqrt{\frac{500}{0{,}05 \cdot 0{,}95}}\right) \approx 2\Phi(3).$

По табл. П. 3 найдем $\Phi(3) = 0{,}49865$, тогда $2\Phi(3) \approx 2 \cdot 0{,}49865 \approx$ $\approx 0{,}9972$.

Задача 2.32 (*опыт Бюффона*). Один из первых математиков, изучавших вероятности событий, француз Ж.Бюффон (1707—1788), для выявления действия закона больших чисел подбрасывал монету 4040 раз, причем орел появлялся в 2048 случаях.

Найти вероятность того, что при повторении опыта Бюффона относительная частота появления орла отклонится от вероятности по модулю не более чем в опыте Бюффона.

Решение. По условию задачи относительная частота появления орла в опыте Бюффона $\left(\frac{m}{n}\right) = \frac{2048}{4040} \approx 0{,}507$, где $n = 4040$, $m = 2048$, а теоретическая вероятность появления орла $p = \frac{1}{2}$. Тогда оценим

$$P\left(\left|\frac{m}{n} - \frac{1}{2}\right| \le \left|\frac{m}{n} \cdot \frac{1}{2}\right| = 0{,}007\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right),$$

или

$$P\left(\left|\frac{m}{n} - \frac{1}{2}\right| \le 0{,}507\right) = 2\Phi(0{,}007 \cdot 2\sqrt{1000}) =$$

$$= 2\Phi(0{,}443) = 2 \cdot 0{,}171 = 0{,}342.$$

Результат, полученный в задаче 2.32, еще раз показывает, что для большей уверенности в результатах нужно проводить большее число независимых испытаний.

Таблица 2.1

Опыты	Число испытаний	Относительная частота	Отклонение относительной частоты от вероятности
Бюффона	4040	0,5069	0,0069
Первый Пирсона	12000	0,5016	0,0016
Второй Пирсона	24000	0,5005	0,0005

Действительно, в истории науки известны результаты английского статистика К. Пирсона по подсчету числа выпадения орла в опытах с монетой. Сравнение их с результатами опыта Бюффона приведено в табл. 2.1.

Эти опыты служат подтверждением теоретического вывода о существовании статистической устойчивости частот как уникального свойства случайных событий.

Задача 2.33. На опытном участке посажено 1000 семян пшеницы с вероятностью прорастания каждого семени $p = 0,9$. Оценить отклонение частоты всхожести семян от их вероятности при вероятности этого события $P = 0,995$.

Решение. Пусть $n = 1\,000$ — число независимых испытаний; A — событие, состоящее в том, что модуль отклонений частоты всхожести семян от их вероятности не превышает некоторого числа ε. Найдем это число.

Итак, $P\left(\left|\frac{m}{n} - 0,9\right| < \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{1000}{0,1 \cdot 0,9}}\right)$, где $q = 0,1$.

Тогда

$$P\left(\left|\frac{m}{n} - 0,9\right| < \varepsilon\right) \approx 2\Phi\left(\varepsilon\frac{\sqrt{1000}}{0,3}\right) = 0,995 \text{ или } \Phi\left(\varepsilon\frac{\sqrt{1000}}{0,3}\right) \approx 0,4975.$$

По таблице функции Лапласа $\Phi(x)$ найдем соответствующее значение x. Получаем соотношение $x = \varepsilon\frac{\sqrt{1000}}{0,3} \approx 2,81$, откуда

$$\varepsilon \approx \frac{2,81 \cdot 0,3}{\sqrt{1000}} = 0,027.$$

Этот ответ означает, что с вероятностью 0,995 частота прорастания семян отличается от вероятности по модулю на 0,027.

Задача 2.34. Вероятность появления события в каждом из независимых испытаний равна 0,5. Найти число испытаний n, при котором с вероятностью 0,7698 можно ожидать, что относительная частота появления события отклонится от его вероятности по модулю меньше, чем на 0,02.

Решение: $p = 0{,}5$; $q = 0{,}5$; $\varepsilon = 0{,}02$,

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) = 2\Phi\left(0{,}02\sqrt{\frac{n}{0{,}5 \cdot 0{,}5}}\right) = 0{,}7698.$$

Откуда $\Phi(0{,}04\sqrt{n}) = 0{,}3849$; по таблицам функции Лапласа находим аргумент, где значение функции Лапласа наиболее близко к 0,3949. Тогда $0{,}04\sqrt{n} = 1{,}2$. Следовательно, $n = 900$.

2.10.5. Центральная предельная теорема

В теории вероятностей известна большая группа утверждений, объединенных названием «центральные предельные теоремы». Смысл этих теорем: при достаточно общих условиях, налагаемых на независимые случайные величины, функция распределения суммы n независимых случайных величин (или нормированной суммы) сходится по вероятности к нормальной функции распределения при $n \to \infty$.

Теорема 2.1. *Если $X_1, X_2, \ldots, X_n$ — независимые случайные величины, распределенные одинаково, то в пределе при $n \to \infty$ функция распределения случайной величины $X = \sum_{i=1}^{n} X_i$ неограниченно приближается к нормальной функции распределения.*

Теорема 2.2. *Если $X_1, X_2, \ldots, X_n$ — независимые случайные величины, распределенные одинаково и имеющие математическое ожидание m и дисперсию σ^2, то в пределе при $n \to \infty$ функция распределения случайной величины $X = \sum_{i=1}^{n} (X_i - m)/\sigma\sqrt{n}$ неограниченно приближается к нормальной функции распределения:*

$$P\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - m) < x\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt. \qquad (2.52)$$

Закон больших чисел позволяет теоретически обосновывать устойчивость, т.е. практически — неслучайность основных эмпирических характеристик случайных величин: выборочных математического ожидания, дисперсии, выборочной функций распределения и т.д.

Пусть для изучения некоторого явления проводится серия экспериментов, порождающих случайную величину X. В первом эксперименте случайная величина X приняла значение x_1, во втором — x_2, ..., в n-м — x_n.

Такие наблюдения (x_1, x_2, ..., x_n) образуют выборку, представляющую собой независимые и одинаково распределенные случайные величины. Тогда к ним применимы теоремы Чебышева и Бернулли, с помощью которых можно обосновать статистическую устойчивость основных выборочных характеристик.

Теория вероятностей имеет дело с величинами, событиями и процессами, вероятность которых заранее известна. Но на практике далеко не всегда можно указать точное значение вероятности того или иного события или явления.

? Например, какова вероятность поймать форель в реке, карпа в озере? Какова вероятность того, что лампочка не перегорит в течение месяца?

Это далеко не праздные вопросы. Так, из-за того, что не были учтены наиболее часто встречающиеся размеры одежды или обуви, так трудно было найти подходящий размер, например 37, 38 размер женской обуви или 42—44 размер мужской обуви, зато полки магазинов были завалены маловстречающимися размерами. А это уже экономические проблемы, ведущие к вопросу рентабельности как магазинов, так и предприятий-изготовителей.

Не менее тревожна и противоположная ситуация: магазины перегружены товаром ходового размера, а тот, кто не вписывается в общие стандарты, не может найти подходящую одежду или обувь.

Для того чтобы учесть массовый спрос на изделия и распределение совокупности объектов по некоторому признаку, пользуются методами *математической статистики*.

Методы математической статистики используют в тех случаях, когда изучают распределение большой совокупности предметов или явлений по определенному признаку.

Так, для планирования экономики страны социологи интересуются распределением множества людей по возрасту, а работники сельского хозяйства — распределением пахотных земель по урожайности. Для предприятий-изготовителей и магазинов, реализующих товар, одинаково важно знать распределение людей по размерам и предпочтениям (с учетом моды), для фармацевтов и врачей — распределение больных по их реакции на определенное лекарство.

В русском журнале «Статистический вестник» за 1915 г. появилась статья, критикующая интендантскую службу за то, что в «качестве образца для армейской обуви был прислан один сапог, «надо полагать, среднеарифметический» для всех ног в армии».

Авторы статьи обращали внимание на то, что для правильного снабжения армии сапогами надо мыслить не только средними арифметическими величинами, но и учитывать совокупность размеров и частот, т.е. применять на практике знания математической статистики (см. гл. 3).

Контрольные вопросы

2.1. Что такое случайная величина?

2.2. Какие случайные величины называются дискретными?

2.3. Что задает случайную величину?

2.4. Что показывает математическое ожидание случайной величины и как найти математическое ожидание ДСВ?

2.5. Что показывает дисперсия ДСВ и по каким формулам она вычисляется?

2.6. Перечислите свойства математического ожидания.

2.7. Какими свойствами обладает дисперсия?

2.8. Как найти среднеквадратическое отклонение?

2.9. Что называется модой ДСВ и как ее найти?

2.10. Что называется медианой НСВ и как ее найти?

2.11. Что называется плотностью вероятности случайной величины и какими свойствами она обладает?

2.12. Что называется функцией распределения случайной величины и какими свойствами она обладает?

2.13. Как связаны между собой плотность вероятности и функция распределения СВ?

2.14. Какие случайные величины называются непрерывными?

2.15. Какие числовые характеристики имеет биномиальный закон распределения?

2.16. В каких случаях применяется закон распределения Пуассона для ДСВ?

2.17. Как вычисляются числовые характеристики закона редких явлений?

2.18. В каких случаях применяется геометрический закон распределения ДСВ?

2.19. Как вычисляются числовые характеристики в случае геометрического распределения ДСВ?

2.20. В каких случаях применяются и по какой теореме вычисляются геометрические распределения ДСВ?

2.21. Как найти математическое ожидание НСВ?

2.22. Приведите формулу для определения дисперсии НСВ.

2.23. Как найти среднеквадратическое отклонение НСВ?

2.24. В каких случаях применяется закон равномерного распределения НСВ?

2.25. Как вычисляются числовые характеристики равномерного распределения НСВ?

2.26. Как вычисляются числовые характеристики нормального распределения НСВ?

2.27. В каких случаях применяется нормальный закон распределения НСВ?

2.28. Когда применяется закон показательного распределения НСВ?

2.29. Как вычисляются числовые характеристики показательного распределения НСВ?

2.30. В чем заключается смысл закона больших чисел?

2.31. Как зависит форма кривой Гаусса от параметров нормального распределения НСВ?

2.32. Как вычисляется вероятность попадания в заданный интервал при нормальном распределении НСВ?

2.33. Как вычисляется вероятность попадания в заданный интервал в случае равномерного распределения НСВ?

2.34. Как вычисляется вероятность попадания в заданный интервал для показательного распределения НСВ?

2.35. Что называется модой НСВ и как ее найти?

2.36. Что называется медианой НСВ и как ее найти?

Задачи для самостоятельного решения

Дискретные случайные величины

2.1. *Решите задачи.*

Случайная величина X задана рядом распределения.

1)

x_i	3	5	7	11	12
p_i	0,14	0,20	0,39	0,17	?

2)

x_i	–2	–1	0	1	2
p_i	0,15	0,21	0,13	0,32	?

3)

x_i	1	3	4	6	7
p_i	0,12	0,16	0,15	0,28	?

4)

x_i	2	4	6	8	9
p_i	0,23	0,17	0,18	0,25	?

5)

x_i	–1	1	2	4	6
p_i	0,18	0,27	0,12	0,32	?

6)

x_i	2	3	5	8	11
p_i	0,21	0,15	0,18	0,25	?

а) Найдите недостающее значение вероятности;

б) найдите функцию распределения $F(x)$;

в) постройте ее график;

г) определите числовые характеристики ДСВ X: моду, медиану, математическое ожидание, дисперсию и среднеквадратическое отклонение.

2.2. *Решите задачи.*

Для следующих задач: а) составьте закон распределения ДСВ X, постройте ее график; б) составьте функцию распределения этой случайной величины, постройте ее график; в) найдите все числовые характеристики этой ДСВ:

1) Вероятность попадания стрелком в мишень равна $\frac{3}{4}$. Стрелок сделал четыре выстрела. Случайная величина X — число попаданий.

2) В магазин вошли четыре покупателя. Вероятность сделать покупку для каждого из вошедших в магазин равна 0,3. Случайная величина X — число покупок.

3) В мастерской ремонтируют пять машин. Вероятность того, что любая из машин отремонтирована, равна 0,2. Случайная величина X — число отремонтированных машин.

4) Для участия в олимпиаде по программированию в колледже были отобраны три юноши и три девушки. Три победителя будут участвовать в зональной олимпиаде. Пусть X — число девушек среди финалистов.

5) Кандидат на выборах в губернаторы считает, что 20 % избирателей этого региона поддерживают его избирательную платформу. Для участия в теледебатах были приглашены четыре избирателя из общего числа избирателей этой губернии. Случайная величина X — число избирателей, поддерживающих данного кандидата.

6) Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,1. Случайная величина X — число отказавших элементов в одном опыте.

7) Придумайте и решите аналогичную задачу.

Геометрические распределения вероятности

2.3. *Решите задачи.*

Для следующих задач: а) составьте законы распределения ДСВ X, постройте ее график; б) составьте функцию распределения ДСВ, постройте ее график; в) найдите все числовые характеристики этой ДСВ:

1) Из орудия производится стрельба по цели до первого попадания. Вероятность попадания в цель $p = 0,6$ при каждом выстреле. Случайная величина X — число выстрелов при попадании в цель, если в наличии три снаряда.

2) Вероятность дозвониться до друга с первой попытки равна $\frac{1}{3}$. Случайная величина X — дозвониться до друга, если в вашем распоряжении пять жетонов.

3) На конечной остановке «отдыхают» пять автобусов разного маршрута, среди которых только один — необходимый вам. Вероятность того, что любой из них уйдет в рейс, 0,2. Случайная величина X — число наблюдаемых вами автобусов, вышедших в рейс.

4) На книжной полке стоит четырехтомник А. С. Пушкина. Случайная величина X — взять том, в котором напечатан роман «Пиковая дама».

5) При прыжках в высоту вероятность взять планку (личный рекорд) равна некоторому значению $p < 1$. Спортсмен повторяет

попытки, пока не добьется успеха. Пусть X — число попыток, не превышающее четырех, причем обычно среднее число попыток равно пяти.

6) Вероятность принятия бизнес-плана равна 0,7. Случайная величина X — число попыток принять бизнес-план.

7) Придумайте и решите аналогичную задачу.

Гипергеометрические распределения

2.4. *Решите задачи.*

Для следующих задач: а) составьте законы распределения ДСВ X, постройте ее график; б) составьте функцию распределения ДСВ, постройте ее график; в) найдите все числовые характеристики этой ДСВ:

1) В партии из десяти деталей имеется восемь стандартных. Берут наугад две детали. Случайная величина X — число стандартных деталей среди отобранных.

2) В партии из десяти деталей три бракованных. Случайная величина X — число бракованных деталей среди трех отобранных.

3) Патруль, состоящий из семи солдат и трех офицеров, обходит участок. Случайная величина X — число офицеров в группе караула, состоящей из трех человек.

4) В продаже 12 красных и 8 белых гвоздик. Составляют букеты, содержащие по пять цветов. Случайная величина X — число белых гвоздик в букете.

5) Среди 16 победителей пяти олимпиад по различным предметам в этом году семь студентов факультета программирования. Случайная величина X — число будущих программистов, занявших призовые места в этих пяти олимпиадах.

6) В библиотеке среди 20 книг, стоящих на полке, восемь по математической статистике. Случайная величина X — число книг по математической статистике из четырех взятых с этой полки.

7) Придумайте и решите аналогичную задачу.

2.5. *Решите задачи.*

Составьте закон распределения ДСВ и постройте многоугольник распределений по имеющимся данным:

1) $x_1 = 1$, $x_2 = 2$, $x_3 = 4$, если известно, что $M(X) = 1{,}9$; $M(X^2) = 4{,}9$;
2) $x_1 = 2$, $x_2 = 3$, $x_3 = 4$, если известно, что $M(X) = 2{,}6$; $D(X) = 0{,}44$;
3) $x_1 = 2$, $x_2 = 4$, $x_3 = 5$, если известно, что $M(X) = 4$; $D(X) = 0{,}18$;
4) $x_1 = 1$, $x_2 = 4$, $x_3 = 5$, если известно, что $M(X) = 4$; $D(X) = 2{,}4$;
5) $x_1 = 1$, $x_2 = 5$, $x_3 = 7$, если известно, что $M(X) = 4$; $D(X) = 4{,}2$;
6) $x_1 = 2$, $x_2 = 5$, $x_3 = 6$, если известно, что $M(X) = 4$; $D(X) = 2{,}8$.

2.6. *Решите задачи.*

Найдите математическое ожидание для ДСВ Y, Z, W, если известно:

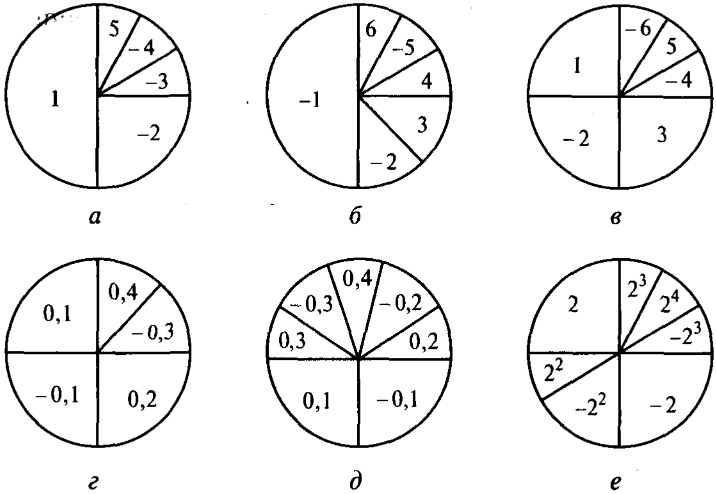


Рис. 2.16

1) $M(X) = 6$, где $Y = X + 3$, $Z = 3X - 5$, $W = X^2$;
2) $M(X) = 8$, где $Y = 2X - 3$, $Z = X + 5$, $W = X^2 + 3$;
3) $M(X) = 5$, где $Y = 3X + 4$, $Z = 2X - 6$, $W = X^2 + 2$;
4) $M(X) = 4$, где $Y = 6X + 2$, $Z = X + 5$, $W = X^2$;
5) $M(X) = 7$, где $Y = 2X + 7$, $Z = X - 5$, $W = X^2 - 2$;
6) $M(X) = 3$, где $Y = 5X - 3$, $Z = X + 5$, $W = X^2 - 3$.

2.7. *Решите задачи.*

Стрелок стреляет по мишени. При значительной угловой скорости вращения он не различает секторы мишени и стреляет наугад. Призовой фонд (тыс. р.) каждого сектора указан в соответствующем секторе мишени (рис. 2.16).

Составьте ряд распределения ДСВ X — попадания в мишень, и определите числовые характеристики X. Примите решение, стоит ли играть в такую игру.

Геометрические распределения

2.8. *Решите задачи.*

Самостоятельно сформулируйте задание.

1) Баскетболист бросает мяч в корзину до первого попадания. Вероятность попадания равна 0,6.

2) У охотника четыре патрона. Он стреляет по зайцу до тех пор, пока не попадет или пока не закончатся патроны. Известно, что вероятность попадания равна 0,25.

3) У дежурного гостиницы в кармане шесть ключей от разных комнат. Вынув наугад ключ, он пробует открыть дверь ближайшей комнаты.

4) Вероятность связаться с абонентом по телефону при каждой попытке равна 0,8.

5) Автомобиль встречает по дороге четыре светофора, каждый из которых пропустит его с вероятностью $\frac{1}{4}$.

6) Для испытания пряжи на крепость, соответствующую высшему сорту, испытывают образцы на прочность. Для таких испытаний берут образец стандартной длины из одного мотка (одной партии). Пусть вероятность прочности нити этой партии равна 0,9. Обычно берут не более четырех образцов нити.

2.9. *Решите задачи.*

Для следующих задач: а) составьте закон распределения ДСВ X, постройте ее график; б) составьте функцию распределения ДСВ; в) найдите все числовые характеристики ДСВ; г) ответьте на поставленный вопрос:

1) Универсам осуществляет контроль чеков. Случайная величина X — число покупателей, у которых был обнаружен обвес, распределена по закону Пуассона и составляет шесть покупателей из ста. Универсам обслужил 500 покупателей. С какой вероятностью обвесы покупателей не будут выявлены?

2) Согласно статистическим данным, вероятность того, что среднестатистический 30-летний молодой человек проживет еще один год, составляет 99,8 %. Страховой агент заключает договор с десятью 30-летними молодыми людьми, причем с каждым на сумму 1 млн р. со страховым взносом 2 тыс. р. Случайная величина X — размер прибыли этой страховой компании от страхования таких клиентов. Какова вероятность того, что страховая компания от этих договоров прибыли не получит?

3) Служащие корпорации при работе с финансовыми документами допускают в среднем 4 % ошибок. Приглашенный для проверки точности финансовых документов аудитор отобрал для проверки шесть документов. Случайная величина X — число ошибок в финансовых документах. Какова вероятность того, что ошибки в документах не обнаружатся?

4) Число опечаток, которые допускает секретарь при работе с документацией, распределено по закону Пуассона со средним значением пять опечаток на одной странице. Случайная величина X — число опечаток на одной странице, если секретарь напечатала восемь страниц. Какова вероятность того, что опечатки в документах не обнаружатся?

5) Маркетинговая служба ювелирного магазина установила, что вероятность покупки при консультации продавцами покупателей составляет около 0,02. Случайная величина X — число покупок в магазине, если к продавцам за консультацией обратились 100 покупателей. Какова вероятность того, что покупатели не приобретут товар даже после консультации продавцов?

6) Опытный врач при установлении диагноза пациента ошибается в среднем в двух случаях из ста. В течение месяца к этому врачу обратились за консультацией 200 пациентов. Случайная величина X — число диагностических ошибок этого врача. Какова вероятность того, что врач не допустит ошибок в диагнозе?

2.10. *Решите задачи.*

Найдите $M(Z_1)$, $M(Z_2)$, $M(Z_3)$, $D(Z_1)$, $D(Z_2)$, если известно, что случайные величины X и Y независимы:

1) $M(X) = 4$, $M(Y) = 5$, $D(X) = 3$, $D(Y) = 4$ для ДСВ $Z_1 = 3X + 5Y$, $Z_2 = 7X - 2Y$, $Z_3 = (X + 2Y)^2$;

2) $M(X) = 2$, $M(Y) = 2$, $D(X) = 3$, $D(Y) = 9$ для ДСВ $Z_1 = 4X - 2Y$, $Z_2 = 2X + 4Y$, $Z_3 = (4X + 3Y)^2$;

3) $M(X) = 5$, $M(Y) = 3$, $D(X) = 3$, $D(Y) = 8$ для ДСВ $Z_1 = 5X + 2Y$, $Z_2 = 3X - 4Y$, $Z_3 = (2X + 5Y)^2$;

4) $M(X) = 8$, $M(Y) = 7$, $D(X) = 3$, $D(Y) = 7$ для ДСВ $Z_1 = 6X - 2Y$, $Z_2 = 4X + 5Y$, $Z_3 = (6X - 2Y)^2$;

5) $M(X) = 9$, $M(Y) = 8$, $D(X) = 3$, $D(Y) = 3$ для ДСВ $Z_1 = 7X + 6Y$, $Z_2 = 6X - Y$, $Z_3 = (8X + Y)^2$;

6) $M(X) = 6$, $M(Y) = 9$, $D(X) = 3$, $D(Y) = 2$ для ДСВ $Z_1 = 8X - 5Y$, $Z_2 = X + 2Y$, $Z_3 = (X - 9Y)^2$.

Непрерывные распределения случайных величин

2.11. *Решите задачи.*

Случайная величина X задана функцией распределения $F(x)$: а) постройте график функции распределения $F(x)$; б) найдите плотность вероятности $f(x)$ по заданной функции распределения $F(x)$ и постройте ее график; в) найдите вероятность попадания в заданный интервал $(a;b)$, если:

1) $F(x) = \begin{cases} 0 & \text{при} \quad x \le 0; \\ \dfrac{x}{3} & \text{при} \quad 0 < x \le 3; \\ 1 & \text{при} \quad x > 3 \end{cases}$ **на интервале (1; 3);**

2) $F(x) = \begin{cases} 0 & \text{при} \quad x \le 0; \\ \dfrac{x^2}{9} & \text{при} \quad 0 < x \le 3; \\ 1 & \text{при} \quad x > 3 \end{cases}$ на интервале (1; 2);

3) $F(x) = \begin{cases} 0 & \text{при} \quad x \le 2; \\ (x-2)^2 & \text{при} \quad 2 < x \le 3; \\ 1 & \text{при} \quad x > 3 \end{cases}$ на интервале $2{,}5 < x < 3$;

4) $F(x)=\begin{cases}0 & \text{при} \quad x\le 0;\\ \dfrac{x^2}{2} & \text{при} \quad 0<x\le\sqrt{2};\\ 1 & \text{при} \quad x>\sqrt{2}\end{cases}$ на интервале $x\in(\frac{1}{6};\frac{1}{3})$;

5) $F(x)=\begin{cases}0 & \text{при} \quad x\le 0;\\ \dfrac{3}{4}x & \text{при} \quad 0<x\le 1\dfrac{1}{3};\\ 1 & \text{при} \quad x>1\dfrac{1}{3}\end{cases}$ на интервале $x\in(0;\frac{1}{3})$;

6) $F(x)=\begin{cases}0 & \text{при} \quad x\le 0;\\ 0{,}8x & \text{при} \quad 0<x\le 2;\\ 1 & \text{при} \quad x>2\end{cases}$ на интервале $x\in(1;1{,}5)$.

2.12. *Решите задачи.*

Случайная величина X задана функцией распределения $F(x)$: а) найдите вероятность того, что в результате испытаний НСВ X попадет в заданный интервал (0; 0,5); б) постройте график функции распределения НСВ в его области определения; в) найдите плотность вероятности НСВ X и постройте ее график; г) найдите числовые характеристики НСВ X, если:

1) $F(x)=\begin{cases}0 & \text{при} \quad x\le 0;\\ x^2 & \text{при} \quad 0<x\le 1;\\ 1 & \text{при} \quad x>1;\end{cases}$ 2) $F(x)=\begin{cases}0 & \text{при} \quad x\le -1;\\ \dfrac{x}{3}+\dfrac{1}{3} & \text{при} \quad -1<x\le 2;\\ 1 & \text{при} \quad x>2;\end{cases}$

3) $F(x)=\begin{cases}0 & \text{при} \quad x\le 0;\\ \dfrac{x}{6} & \text{при} \quad 0<x\le 6;\\ 1 & \text{при} \quad x>6;\end{cases}$ 4) $F(x)=\begin{cases}0 & \text{при} \quad x\le 1;\\ \dfrac{x-1}{2} & \text{при} \quad x\in(1;3];\\ 1 & \text{при} \quad x>3;\end{cases}$

5) $F(x)=\begin{cases}0 & \text{при} \ x\le -2;\\ \dfrac{x}{4}+\dfrac{1}{2} & \text{при} \ -2<x\le 2;\\ 1 & \text{при} \ x>2;\end{cases}$ 6) $F(x)=\begin{cases}0 & \text{при} \ x\le 1;\\ (x-1)^2 & \text{при} \ x\in(1;2];\\ 1 & \text{при} \ x>2.\end{cases}$

2.13. *Решите задачи.*

Непрерывная случайная величина X подчинена закону распределения с плотностью вероятности $f(x)$: а) постройте график плотности распределения $f(x)$; б) найдите функцию распределения $F(x)$ и постройте ее график; в) найдите числовые характеристики НСВ X, если:

1) $f(x)=\begin{cases}0 & \text{при} \quad x\notin(-\frac{\pi}{2};\frac{\pi}{2});\\ \frac{1}{2}\cos x & \text{при} \quad x\in(-\frac{\pi}{2};\frac{\pi}{2})\end{cases}$ на интервале $0\le x\le\frac{\pi}{4}$;

2) $f(x)=\begin{cases}0 & \text{при} \quad x\notin(2;2,5);\\ 2x-4 & \text{при} \quad x\in(2;2,5)\end{cases}$ на интервале $x\in(1;2,5)$;

3) $f(x)=\begin{cases}0 & \text{при} \quad x\in(0;\pi);\\ \frac{1}{2}\sin x & \text{при} \quad x\notin(0;\pi)\end{cases}$ на интервале $x\in(\frac{\pi}{6};\frac{2\pi}{3})$;

4) $f(x)=\begin{cases}0 & \text{при} \quad x\notin(0;\frac{\pi}{4});\\ 2\cos 2x & \text{при} \quad x\in(0;\frac{\pi}{4})\end{cases}$ на интервале $x\in(\frac{\pi}{6};\frac{\pi}{3})$;

5) $f(x)=\begin{cases}0 & \text{при} \quad x\notin(0;\frac{\pi}{3});\\ \frac{3}{2}\sin 3x & \text{при} \quad x\in(0;\frac{\pi}{3})\end{cases}$ на интервале $x\in(\frac{\pi}{6};\frac{\pi}{4})$;

6) $f(x)=\begin{cases}\frac{1}{3}\sin x & \text{при} \quad x\in(0;\pi);\\ 0 & \text{при} \quad x\notin(0;\pi)\end{cases}$ на интервале $x\in(0;\frac{\pi}{3})$.

2.14. *Решите задачи.*

Случайная величина X распределена по показательному закону с плотностью вероятности $f(x)$. Найдите: а) вероятность попадания НСВ X на интервал $(\alpha;\beta)$; б) функцию распределения этой НСВ X; в) числовые характеристики $M(X)$, $D(X)$, $\sigma(X)$:

1) $f(x)=\begin{cases}0 & \text{при} \quad x<0;\\ 6e^{-6x} & \text{при} \quad x\ge 0,\end{cases}$

$\alpha=0,23$; $\beta=0,32$;

2) $f(x)=\begin{cases}0 & \text{при} \quad x<0;\\ 7e^{-7x} & \text{при} \quad x\ge 0,\end{cases}$

$\alpha=0,45$; $\beta=0,62$;

3) $f(x)=\begin{cases}0 & \text{при} \quad x<0;\\ 4e^{-4x} & \text{при} \quad x\ge 0,\end{cases}$

$\alpha=0,38$, $\beta=0,49$;

4) $f(x)=\begin{cases}0 & \text{при} \quad x<0;\\ 5e^{-5x} & \text{при} \quad x\ge 0,\end{cases}$

$\alpha=0,81$, $\beta=1,23$;

5) $f(x) = \begin{cases} 0 & \text{при} \quad x < 0; \\ 8e^{-8x} & \text{при} \quad x \ge 0, \end{cases}$ $\alpha = 1{,}2,\ \beta = 2{,}3;$

6) $f(x) = \begin{cases} 0 & \text{при} \quad x < 0; \\ 9e^{-9x} & \text{при} \quad x \ge 0, \end{cases}$ $\alpha = 3{,}4,\ \beta = 5{,}3.$

2.15. *Решите задачи.*

Время безотказной работы компьютера распределено по показательному закону с надежностью $R(t)$. Найдите вероятность того, что компьютер проработает t часов, если:

1) $R(t) = 0{,}002e^{-0{,}002t}$, $t = 2\,000$;
2) $R(t) = 0{,}003e^{-0{,}003t}$, $t = 4\,000$;
3) $R(t) = 0{,}004e^{-0{,}004t}$, $t = 5\,000$;
4) $R(t) = 0{,}006e^{-0{,}006t}$, $t = 1\,000$;
5) $R(t) = 0{,}007e^{-0{,}007t}$, $t = 500$;
6) $R(t) = 0{,}008e^{-0{,}008t}$, $t = 600$.

2.16. *Решите задачи.*

Непрерывная случайная величина X распределена нормально. Найдите:

а) функцию распределения НСВ X;
б) ее плотность вероятности;
в) числовые характеристики $M(X)$, $\sigma(X)$;
г) вероятность попадания НСВ X на интервал $(\alpha;\beta)$, если:

1) $\text{Mo} = 2$, $D(X) = 4$, где $\alpha = 3$, $\beta = 5$;
2) $\text{Mo} = 3$, $D(X) = 16$, где $\alpha = 2$, $\beta = 5$;
3) $\text{Mo} = 4$, $D(X) = 9$, где $\alpha = 2$, $\beta = 6$;
4) $\text{Me} = 5$, $D(X) = 4$, где $\alpha = 3$, $\beta = 5$;
5) $\text{Me} = 2$, $D(X) = 9$, где $\alpha = 5$, $\beta = 7$;
6) $\text{Me} = 3$, $D(X) = 25$, где $\alpha = 4$, $\beta = 5$.

2.17. *Решите задачи.*

Непрерывная случайная величина X распределена нормально. Найдите: а) функцию распределения НСВ X; б) ее числовые характеристики; в) вероятность попадания НСВ X на интервал $(\alpha;\beta)$, если:

1) $f(x) = \dfrac{1}{4\sqrt{2\pi}} e^{-\frac{(x-2)^2}{32}}$, где $\alpha = 3$, $\beta = 5$;

2) $f(x) = \dfrac{1}{3\sqrt{2\pi}} e^{-\frac{(x-5)^2}{18}}$, где $\alpha = 2$, $\beta = 4$;

3) $f(x) = \dfrac{1}{2\sqrt{2\pi}} e^{-\frac{(x-8)^2}{8}}$, где $\alpha = 3$, $\beta = 6$;

4) $f(x) = \dfrac{1}{5\sqrt{2\pi}} e^{-\frac{(x+2)^2}{50}}$, где $\alpha = 2$, $\beta = 5$;

5) $f(x) = \dfrac{1}{4\sqrt{2\pi}} e^{-\frac{(x-6)^2}{32}}$, где $\alpha = 4$, $\beta = 5$;

6) $f(x) = \frac{1}{6\sqrt{2\pi}} e^{-\frac{(x-7)^2}{72}}$, где $\alpha = 2$, $\beta = 6$.

2.18. *Решите задачи.*

Непрерывная случайная величина X распределена равномерно на отрезке $[a, b]$. Найдите: а) функцию распределения НСВ X и постройте ее график; б) плотность вероятности НСВ X и постройте ее график; в) ее числовые характеристики; г) вероятность попадания НСВ X на интервал $(\alpha; \beta)$, если:

1) $a = 2$, $b = 5$; интервал $(3; 5)$;
2) $a = 1$, $b = 5$; интервал $(2; 4)$;
3) $a = 2$, $b = 7$; интервал $(4; 5)$;
4) $a = 3$, $b = 8$; интервал $(4; 7)$;
5) $a = 4$, $b = 10$; интервал $(3; 8)$;
6) $a = 5$, $b = 12$; интервал $(8; 10)$.

2.19. *Решите задачи.*

В следующих задачах найдите: а) плотность вероятности СВ и постройте ее график; б) интегральную функцию распределения СВ и постройте ее график; в) числовые характеристики СВ; г) вероятность попадания СВ в заданный интервал:

1) В универмаге осуществляется контроль чеков. Покупатели подходят к кассе в среднем 12 чел./ч. Случайная величина X — число покупателей в универмаге — распределена равномерно. Найдите вероятность того, что с 15 до 16 ч кассир обслужит не более восьми покупателей.

2) Интервал движения автобуса № 5, движущегося строго по расписанию, составляет 10 мин. Время ожидания автобуса на остановке распределено равномерно. Найдите вероятность того, что ожидать автобус придется не более 3 мин.

3) Поезда метрополитена идут с интервалом 2 мин. Время ожидания поезда — равномерно распределенная случайная величина. С какой вероятностью пассажир, пришедший на платформу, будет ожидать поезд более 1 мин?

4) Время работы холодильника до первого отказа подчинено показательному закону с интенсивностью отказов $5 \cdot 10^{-4}$ ($ч^{-1}$). Найдите вероятность того, что за 2000 ч эксплуатации холодильник не выйдет из строя.

5) Время работы телевизора подчинено показательному закону. Среднее время ремонта телевизоров составляет 12 дней. Определите вероятность того, что для ремонта телевизора понадобится не менее 15 дней.

6) Время между двумя сбоями компьютера распределено по показательному закону с интенсивностью 4 ч. Найдите вероятность того, что между двумя сбоями временной интервал будет от двух до трех часов.

2.20. *Решите задачи.*

1) Текущая цена акции может быть смоделирована по нормальному закону с математическим ожиданием 15 усл. ед. и среднеквадратическим отклонением 0,3 усл.ед. Найдите вероятность того, что цена акции:

а) не превышает 15,5 усл. ед.;

б) не ниже 15,5 усл. ед.;

в) ниже 15,4 усл. ед.;

г) заключена в пределах от 15,1 до 15,4 усл. ед.;

д) с помощью правила трех сигм установите границы, в которых будет находиться текущая цена акций.

2) Масса одного батона хлеба есть случайная величина, подчиненная нормальному закону распределения с математическим ожиданием 600 г и среднеквадратическим отклонением 6 г. Найдите вероятность того, что масса взятого для контроля батона из этой партии будет:

а) не меньше 605 г;

б) не больше 580 г;

в) заключено в пределах от 580 до 600 г;

г) больше 610 г;

д) с помощью правила трех сигм установите границы, в которых будет находиться масса одного батона хлеба.

3) Волжский автомобильный завод запускает в производство новый двигатель. Конструкторы двигателя предполагают, что благодаря ему средняя длина пробега автомобиля составит 160 тыс. км, а среднее стандартное отклонение — 30 тыс. км (средняя длина пробега автомобиля — случайная величина, подчиненная нормальному закону распределения). Найдите вероятность того, что длина пробега автомобиля с таким двигателем составит:

а) не менее 110 км;

б) не более 170 км;

в) заключена в пределах от 110 до 180 км;

г) больше 180 км;

д) с помощью правила трех сигм установите границы, в которых будет находиться длина пробега автомобиля.

4) Ежеквартальный выпуск продукции на заводе распределен по нормальному закону, где среднее значение $m = 152$ тыс.ед. продукции в квартал, а среднеквадратическое отклонение $\sigma = 40$ тыс.ед. Найдите вероятность того, что ежеквартальный выпуск продукции на заводе составит:

а) не менее 100 тыс.ед. продукции;

б) не более 190 тыс.ед. продукции;

в) более 200 тыс.ед. продукции;

г) заключен в пределах от 110 до 180 тыс.ед.;

д) с помощью правила трех сигм установите границы, в которых будет находиться ежеквартальный выпуск продукции на заводе.

5) Автомат штампует детали. Контролируемый размер детали есть случайная величина, распределенная по нормальному закону с параметрами $m = 52$ см, $\sigma = 0{,}4$ см. Найдите вероятность того, что контролируемый размер детали составит:

а) не менее 51,9 см;

б) не более 52,3 см;

в) более 53 см;

г) заключен в пределах от 51,97 до 52,03 см;

д) с помощью правила трех сигм установите границы, в которых будет находиться контролируемый размер детали.

6) Изменение индекса ценных бумаг на фондовой бирже может быть смоделировано как НСВ с параметрами $m = 1$, $\sigma = 0{,}1$. Найдите вероятность того, что на следующих торгах индекс ценных бумаг будет:

а) не более 1;

б) не менее 1;

в) менее 0,98;

г) заключен в пределах от 0,97 до 1,03;

д) с помощью правила трех сигм установите границы, в которых будет находиться индекс ценных бумаг.

Показательные распределения

2.21. *Решите задачи*:

1) Радиосистема, состоящая из 1000 элементов, с интенсивностью отказов $\lambda_i = 10^{-6}$ отк./ч, прошла испытание и принята заказчиком. Определите вероятность безотказной работы радиосистемы за время $t = 1000$ ч.

2) Холодильник имеет постоянную интенсивность отказа, равную $\lambda = 10^{-5}$ отк./ч. Какова вероятность того, что он откажет после гарантийного срока 20 000 ч?

3) Срок службы жесткого диска компьютера — случайная величина, распределенная по показательному закону со средним значением 12 тыс. ч. Определите вероятность безотказной службы жесткого диска свыше 18 тыс.ч.

4) Срок службы батареек микрокалькулятора подчиняется показательному закону распределения с $\lambda = \dfrac{1}{12}$. Какова доля батареек со сроком службы более 15 дней?

5) Менеджер по рекламе считает, что время, в течение которого телезрители удерживают в памяти телерекламу, есть случайная величина, распределенная по показательному закону с $\lambda = \dfrac{1}{6}$. Вычислите долю телезрителей, способных вспомнить телерекламу через две недели.

6) Компьютерный программист использует показательное распределение для оценки надежности своих программ. После того

как им были найдены 8 своих ошибок, он убедился, что время до нахождения следующей ошибки подчинено показательному закону распределения с $\lambda = 0{,}25$. Найдите среднее время, потраченное на поиск первой ошибки, и вычислите вероятность того, что десятую ошибку можно найти в интервале от 3 до 9 дней.

Закон больших чисел

2.22. *Решите задачи.*

В результате n независимых испытаний найдены некоторые значения случайной величины X: $x_1, x_2, ..., x_n$, имеющие математическое ожидание $M(X)$ и дисперсию $D(X)$. Оцените снизу вероятность того, что модуль разности между средним арифметическим наблюдаемых значений СВ $\left|\frac{1}{n}\sum_{i=0} x_i - M(X)\right| < \varepsilon$, если известно, что;

1) $n = 100,\ M(X) = 8,\ D(X) = 1,\ \varepsilon = \frac{1}{4}$;

2) $n = 150,\ M(X) = 6,\ D(X) = 2,\ \varepsilon = \frac{1}{2}$;

3) $n = 120,\ M(X) = 10,\ D(X) = 3,\ \varepsilon = \frac{1}{5}$;

4) $n = 110,\ M(X) = 7,\ D(X) = 1,\ \varepsilon = 0{,}3$;

5) $n = 130,\ M(X) = 10,\ D(X) = 4,\ \varepsilon = 0{,}2$;

6) $n = 200,\ M(X) = 12,\ D(X) = 2,\ \varepsilon = 0{,}4$.

2.23. *Решите задачи.*

Известны значения $M(X)$ и $D(X)$. Оцените с помощью неравенства Чебышева $P\{a \le x \le b\}$:

1) $M(X) = 12,\ D(X) = 5,\ a = 2,\ b = 16$;

2) $M(X) = 4,\ D(X) = 1,\ a = 3,\ b = 15$;

3) $M(X) = 7,\ D(X) = 2,\ a = 1,\ b = 9$;

4) $M(X) = 10,\ D(X) = 3,\ a = 4,\ b = 18$;

5) $M(X) = 5,\ D(X) = 4,\ a = 2,\ b = 17$;

6) $M(X) = 8,\ D(X) = 3,\ a = 5,\ b = 12$.

2.24. *Решите задачи* (с помощью неравенства Чебышева для биномиальных распределений).

1) Волжская ГЭС обслуживает сеть из 20 000 объектов, вероятность включения каждого из которых в вечернее время зимой составляет 0,85. С какой вероятностью число объектов, включенных в сеть зимним вечером, будет отличаться от математического ожидания более, чем на 600.

2) Устройство состоит из десяти независимо работающих элементов. Вероятность отказа в течение суток для каждого элемента равна 0,08. Оцените вероятность того, что модуль разности между

числом отказавших элементов и средним числом отказов за сутки будет меньше трех.

3) Осветительная сеть зала включает десять параллельно подключенных ламп. Вероятность того, что в течение суток лампа будет включена, составляет 0,75. Оцените вероятность того, что модуль разности между числом включенных ламп и средним числом включенных ламп за сутки окажется не больше двух.

4) Средняя длина детали 200 мм, дисперсия равна 0,2. Оцените вероятность того, что случайно взятая деталь окажется длиной не менее 195 мм и не более 205 мм.

5) Вероятность того, что число автобусов красного цвета в автобусном парке, состоящем из 110 автобусов, отличается от своего математического ожидания менее чем на 17.

6) Вероятность того, что число сотрудников с высшим образованием в коллективе, состоящем из 300 человек, отличается от своего математического ожидания менее чем на 10.

2.25. *Решите задачи.*

Используя заданный ряд распределения ДСВ, найдите и оцените вероятность того, что $|x - M(x)| < \varepsilon$

1) $\varepsilon = 5$;

x_i	1	2	3	4	5	6	7
p_i	0,09	0,15	0,24	0,15	0,23	0,1	0,04

2) $\varepsilon = 3$;

x_i	1	2	3	4	5	6
p_i	0,05	0,10	0,25	0,20	0,10	0,30

3) $\varepsilon = 4$;

x_i	1	2	3	5	6	8
p_i	0,25	0,05	0,10	0,20	0,30	0,10

4) $\varepsilon = 2$;

x_i	2	3	5	6	8	10
p_i	0,20	0,10	0,20	0,30	0,05	0,15

5) $\varepsilon = 2,5$;

x_i	1	2	3	4	5	8
p_i	0,10	0,05	0,10	0,30	0,20	0,25

6) $\varepsilon = 4,5$;

x_i	1	2	4	5	8	10
p_i	0,15	0,20	0,20	0,10	0,25	0,10

2.26. *Решите задачи.*

1) Средний срок службы майора 6 лет. Оцените снизу вероятность того, что этот майор не прослужит более 15 лет.

2) Вероятность выпуска нестандартных микрокалькуляторов равна 24 %. Оцените вероятность того, что в партии из 100 калькуляторов число нестандартных менее 64.

3) Вероятность приобретения бракованного компьютера составляет 2 %. Оцените снизу вероятность того, что из 80 ЭВМ, приобретенных колледжем, число бракованных не более двух.

4) Вероятность поражения цели при выстреле составляет 40 %. Оцените вероятность того, что при 150 выстрелах попаданий будет не более 100.

5) Средний расход воды в деревне Водохлебовке составляет 20 000 л в день. Оцените вероятность того, что в самый знойный день жители деревни Водохлебовки израсходуют не более 80 000 л воды.

6) Среднее число выпускников колледжа, получивших красный диплом, составляет ежегодно 25 чел. Оцените вероятность того, что в этом учебном году число отличников не превысит 30 чел.

7) Придумайте аналогичную задачу.

Дополнительные задачи

2.27. После ответа студента на вопросы экзаменационного билета экзаменатор задает ему дополнительные вопросы. Преподаватель прекращает задавать дополнительные вопросы, как только студент обнаружил незнание заданного вопроса. Вероятность того, что студент ответит на любой из дополнительных вопросов, равна 0,9. Найдите наивероятнейшее число заданных вопросов.

2.28. Вероятность того, что стрелок попадет в мишень, равна 0,8. Стрелку выдают патроны до тех пор, пока он не промахнется. Найдите наивероятнейшее число выданных стрелку патронов.

2.29. Два стрелка производят по одному выстрелу в мишень. Вероятность попадания в нее первым стрелком равна 0,5, вторым — 0,4. Составьте закон распределения числа попаданий в мишень.

2.30. Вероятность того, что в библиотеке необходимая студенту книга свободна, равна 0,3. Составьте закон распределения количества библиотек, которые посетит студент, если в городе четыре библиотеки.

2.31. Охотник стреляет по дичи до первого попадания, но успевает сделать не более четырех выстрелов. Составьте закон распределения числа промахов, если вероятность попадания в цель при одном выстреле равна 0,7. Вычислите математическое ожидание и дисперсию этой случайной величины.

2.32. В отделение скорой помощи поступает в среднем 300 вызовов в сутки. Какова вероятность того, что в указанный час будет пять вызовов?

2.33. Учебник издан тиражом 100 тыс. экз. Вероятность того, что учебник сброшюрован неправильно, равна 0,0001. Найдите вероятность того, что тираж содержит пять бракованных книг.

2.34. Устройство состоит из 1 000 элементов, работающих независимо друг от друга. Вероятность отказа любого элемента в течение суток равна 0,002. Найдите вероятность того, что за сутки из строя выйдут три элемента.

2.35. Завод отправил на базу 500 изделий. Вероятность повреждения изделия в пути равна 0,002. Какова вероятность того, что в пути будет повреждено:

а) три изделия;

б) менее трех изделий;

в) более трех изделий;

г) хотя бы одно изделие?

2.36 Магазин получил 1 000 бутылок минеральной воды. Вероятность того, что бутылка окажется разбитой, равна 0,003. Найдите вероятность того, что магазин получит:

а) две разбитые бутылки;

б) менее двух разбитых бутылок;

в) не менее двух разбитых бутылок;

г) хотя бы одну разбитую бутылку.

2.37. Станок-автомат штампует детали. Вероятность того, что изготовленная деталь окажется бракованной, равна 0,01. Найдите вероятность того, что среди 200 деталей окажется ровно четыре бракованные.

2.38. Вероятность выигрыша по одному лотерейному билету равна 0,01. Сколько нужно купить билетов, чтобы выиграть хотя бы по одному из них с вероятностью не менее 0,95?

2.39. НСВ распределена по показательному закону с $\lambda = 3$. Найдите вероятность того, что в результате испытаний X попадет в интервал $(0{,}13; 0{,}7)$.

2.40. НСВ распределена по экспоненциальному закону с $\lambda = 0{,}1$. Найдите вероятность попадания в интервал $(1{,}2)$.

2.41. Длительность времени выхода из строя элемента имеет показательное распределение с $\lambda = 0{,}00004$ с$^{-1}$. Найдите вероятность того, что за время $t = 100$ ч:

а) элемент откажет;

б) элемент не откажет.

2.42. Испытывают три независимых элемента. Длительность времени безотказной работы элементов распределена по показательному закону: для первого элемента $\lambda = 0{,}1$ ч$^{-1}$, для второго $\lambda = 0{,}2$ ч$^{-1}$, для третьего $\lambda = 0,3$ ч$^{-1}$. Найдите вероятность того, что в интервале времени 0,5 ч:

а) откажет только один элемент;

б) откажут только два элемента;

в) откажут все три элемента;

г) не будет отказов.

2.43. Средняя масса детали 300 г, дисперсия равна 0,4. Оцените вероятность того, что случайно взятая деталь окажется массой от 295 до 305 г.

2.44. НСВ $X \in N(10,2)$. Найдите вероятность попадания СВ в интервал (12,14).

2.45. НСВ $X \in N(20,5)$. Найдите вероятность попадания НСВ в интервал (15,25).

2.46. Автомат изготовляет стальные шарики. Шарик считается годным, если отклонение X диаметра шарика от проектного размера по абсолютному значению меньше 0,7 мм. Считая, что случайная величина X распределена нормально со среднеквадратическим отклонением $\sigma = 0,4$ мм, найдите, сколько в среднем будет годных шариков среди 100 изготовленных.

2.47. Случайная величина X распределена нормально с математическим ожиданием $M(X) = 10$. Вероятность попадания X в интервал (10,20) равна 0,3. Чему равна вероятность попадания X в интервал (0,10)?

2.48. Рост взрослого мужчины является случайной величиной, распределенной по нормальному закону. Пусть ее математическое ожидание равно 170 см, а дисперсия — 36 см^2. Вычислите вероятность того, что хотя бы один из наудачу выбранных четырех мужчин будет иметь рост от 168 до 172 см.

2.49. Диаметр изготовленной детали является случайной величиной, распределенной по нормальному закону. Дисперсия ее равна 0,0001 см^2, а математическое ожидание — 2,5 см. Найдите границы, в которых с вероятностью 0,9973 заключен диаметр наугад взятой детали.

2.50. Найдите дисперсию и среднеквадратическое отклонение случайной величины X, распределенной равномерно на интервале (2,8).

2.51. Цена деления шкалы амперметра равна 0,1 А. Показания округляют до ближайшего целого числа. Найдите вероятность того, что при отсчете будет сделана ошибка, превышающая 0,02 А.

2.52. Автобусы некоторого маршрута идут строго по расписанию. Интервал движения составляет 5 мин. Найдите вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус менее двух минут.

2.53. Цена деления шкалы измерительного прибора равна 0,2. Показания прибора округляют до ближайшего целого числа. Найдите вероятность того, что при отсчете будет сделана ошибка:

а) меньшая 0,04;
б) бо́льшая 0,05;
в) бо́льшая 0,18.

2.54. Сумма всех вкладов в сбербанке составляет $4 \cdot 10^7$ р. Вероятность изъятия вклада в 100 тыс. р. составляет 75 %. Сколько вкладчиков в этом банке?

2.55. Средняя урожайность пшеницы в Самарской области составляет 20 ц/га. В области существует 160 отобранных в случайном порядке опытных участков площадью по 1 га, засаженных пшеницей. Какова вероятность того, что средняя величина урожайности на этих опытных участках будет отличаться от урожайности во всей области не более чем на 1 %, т.е. заключена в пределах от 19,8 до 20,2 ц/га.

2.56. В результате n независимых испытаний найдены некоторые значения случайной величины X, имеющей математическое ожидание $M(X)$ и дисперсию $D(X)$. Оцените вероятность того, что модуль разности между средним арифметическим наблюдаемых значений СВ и $M(X)$ $\left|\frac{1}{n}\sum_{i=0}^{n} X_i - M(x)\right| < \varepsilon$:

а) $n = 300,\ M(X) = 15,\ D(X) = 2,\ \varepsilon = 1/4$;
б) $n = 250,\ M(X) = 16,\ D(X) = 5,\ \varepsilon = 1/2$.

2.57. 1) Известны значения $M(X)$ и $D(X)$. Оцените с помощью неравенства Чебышева $P\{a \le x \le b\}$:

а) $M(X) = 15,\ D(X) = 6,\ a = 12,\ b = 16$;
б) $M(X) = 14,\ D(X) = 7,\ a = 12,\ b = 15$.

2) При каких значениях ε выполнится неравенство $P\{|x - M(x)| \le \varepsilon\} \ge 0{,}995$, если значение $M(X)$ взято из предыдущей задачи.

2.58. Случайная величина X подчинена закону распределения с плотностью вероятности $f(x)$:

1) $f(x) = \begin{cases} \frac{2}{3}x - \frac{2}{9}x^2 & \text{при} \quad x \in [0;3] \\ 0 & \text{при} \quad x \notin [0;3] \end{cases}$ **на интервале (1;2);**

2) $f(x) = \begin{cases} 0 & \text{при} \quad x \notin (0;2) \\ x - \frac{1}{4}x^2 & \text{при} \quad x \in [0;2] \end{cases}$ **на интервале (1,5;2,5);**

3) $f(x) = \begin{cases} -\frac{3}{4}x^2 + \frac{9}{2}x - 6 & \text{при } x \in [2;4] \\ 0 & \text{при } x \notin [2;4) \end{cases}$ **на интервале** $x \in (1{,}5;2{,}5)$**;**

4) $f(x) = \begin{cases} 0 & \text{при } x \notin [0;2] \\ \frac{3}{4}x - \frac{3}{16}x^2 & \text{при } x \in [0;2] \end{cases}$ **на интервале** $x \in \left[\frac{1}{3};1\right]$;

5) $f(x)=\begin{cases}0 & \text{при } x\notin[0;2]\\ x-\dfrac{x^3}{4} & \text{при } x\in[0;2]\end{cases}$ на интервале $x\in\left[\dfrac{1}{4};\dfrac{1}{2}\right]$;

6) $f(x)=\begin{cases}0 & \text{при } x\notin[2;4]\\ -\dfrac{3}{4}x^2+\dfrac{9}{2}x-6 & \text{при } x\in[2;4]\end{cases}$ на интервале $x\in\left[\dfrac{1}{4};\dfrac{1}{2}\right]$.

Постройте графики плотности вероятности.

2.59. Найдите пределы: а) $\lim\limits_{\sigma\to 0}\dfrac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}dt$;

б) $\lim\limits_{b\to a+0} F(x,\, a,\, b)$, где $F(x,\, a,\, b)$ — функция распределения равномерно распределенной на $[a,\, b]$ случайной величины.

ГЛАВА 3
ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

3.1. Выборочный метод

Выборочный метод является одним из методов математической статистики.

Математическая статистика — раздел математики, посвященный методам сбора, анализа и обработки результатов статистических данных наблюдений для научных и практических целей. Математическая статистика тесно связана с теорией вероятности и опирается на ее выводы.

3.1.1. Задачи и методы математической статистики

В процессе рассмотрения любой научной проблемы исследователь ставит перед собой и решает следующие задачи: 1) описание явления; 2) анализ и прогноз; 3) выработка оптимального решения.

Методы математической статистики применяют в тех случаях, когда изучают распределение *массовых явлений*, т.е. большой совокупности предметов или явлений, распределенных *по определенному признаку.*

Пусть подлежит изучению совокупность однородных объектов, объединенных общим признаком или свойством качественного или количественного характера. Отдельные элементы такой совокупности называются ее членами. Число членов совокупности составляет ее *объем.*

Изучение всей совокупности чаще всего невозможно или нецелесообразно из-за значительных материальных затрат, порчи или уничтожения объекта исследования. Так, невозможно получить объективную и полную информацию о доходе населения всего региона, т.е. каждого его обитателя. В связи с порчей объекта исследования невозможно получить достоверную информацию о качестве, например партии лекарственных средств или продуктов питания.

Основная *задача* математической статистики заключается в исследовании всей совокупности по выборочным данным в зависимости от поставленной цели, т.е. изучение вероятностных свойств совокупности: закона распределения, числовых характеристик и т.д. для принятия управленческих решений в условиях неопределенности.

3.1.2. Виды выборки

В этом подразделе рассмотрим различные виды выборки; условия, которые необходимо выполнять при применении выборочного метода, а также такие понятия, как варианта, вариационный ряд, относительная частота, статистическое распределение выборки.

Пусть исследуется объект, состоящий из достаточного большого числа N элементов. Случайным образом выберем из этого множества N элементов подмножество n элементов и с его помощью осуществим исследование всего множества по некоторому признаку. Цель математической статистики — оценить характеристики генеральной совокупности по выборочным данным. Так как зачастую невозможно описать все N элементов исследуемого объекта, с помощью выборки опишем n элементов, на основании чего получим набор исследуемых значений x_i, где $i = \overline{1,\ n}$. Итогом такого исследования может быть вывод о распределении признака во всем объекте исследования, которое обычно неизвестно.

Сам процесс получения чисел x_i случаен и зависит от многих факторов. Поэтому целесообразно рассматривать совокупность полученных результатов x_i $(i = \overline{1,\ n})$ как набор одинаково распределенных случайных величин X_i, одной из реализаций которого является набор x_i.

Случайной выборкой (*выборкой*) объема n называется упорядоченный набор $X_1, X_2, ..., X_n$ независимых одинаково распределенных случайных величин.

Если рассмотреть весь исследуемый объект аналогичным образом, то результаты $x_1, ..., x_N$ будут соответствовать упорядоченному набору N одинаково распределенных случайных величин $X_1, ..., X_n$, который назовем *генеральной совокупностью.*

Понятие генеральной совокупности и выборки будем употреблять как для наборов чисел, так и для случайных величин в зависимости от контекста. Так как случайные величины $X_1, ..., X_N$ одинаково распределены и описывают один и тот же признак, то в дальнейшем будем рассматривать обобщенное понятие «случайная величина X», которое имеет определенную функцию распределения и числовые характеристики: математическое ожидание и др.

Таким образом, с помощью выборочного метода исследуется не вся генеральная совокупность, а выборка $(x_1, x_2, ..., x_n)$ как результат ограниченного ряда наблюдений. Затем по вероятностным свойствам данной выборки выносится суждение о генеральной совокупности. Для получения выборки применяют различные методы отбора. Объекты исследования после изучения можно возвратить (не возвращать) в генеральную совокупность, что соответствует повторной (бесповторной) выборке.

Выборка называется *репрезентативной*, или *представительной*, если она хорошо воспроизводит генеральную совокупность, т.е. вероятностные свойства выборки совпадают или близки к свойствам самой генеральной совокупности.

Итак, результативность применения выборочного метода повышается при соблюдении следующих условий:

• количество исследуемых элементов выборки должно быть *достаточно для выводов*, т.е. выборка должна быть представительна, или *репрезентативна*.

Так, достаточное количество деталей в партии, проверяемой на качество (брак), устанавливается с помощью законов теории вероятностей и математической статистики;

• элементы выборки должны быть *разнообразны*, *взяты случайно*, т.е. должен соблюдаться принцип *рандомизации*;

• изучаемый признак должен быть *характерен* (типичен) для всех элементов множества изучаемых объектов, т.е. для всей генеральной совокупности;

• изучаемый признак должен являться *существенным* для всех элементов данного класса.

Изменение признака статистической совокупности, изучаемого выборочным методом, называется *вариацией*, а наблюдаемые значения признака x_i — *вариантой*. *Абсолютной частотой* (*частотой*, или *частостью*) варианты x_i называется число членов совокупности (генеральной или выборки), имеющих значение x_i (т.е. это число частиц i-го сорта).

Ранжированная группировка вариант по отдельным значениям признака (или по интервалам изменения), т.е. последовательность вариант, расположенная в порядке возрастания, называется *вариационным рядом*. Любую числовую функцию $\gamma(x_1, x_2, ..., x_n)$ от результатов наблюдений $x_1, x_2, ..., x_n$ исследуемой случайной величины называют *статистикой*.

Принято объем генеральной совокупности обозначать N, ее абсолютные частоты — N_i, объем выборки — n, ее абсолютные частоты — n_i. Очевидно, что

$$\sum_i N_i = N; \quad \sum_i n_i = n.$$

Отношение частоты к объему совокупности называется *относительной частотой*, или ***статистической вероятностью***, и обозначается W_i или p_i^*:

$$W_i = p_i^* = \frac{n_i}{n}.$$

Если количество вариант велико или близко к объему выборки (при дискретном распределении), а также если выборка производится из непрерывной генеральной совокупности, то вариационный ряд составляют не по отдельным (*точечным*) значениям, а по *интервалам* значений генеральной совокупности. Вариационный ряд, представленный таблицей, построенный с помощью процедуры группировки, будем называть *интервальным*. При составлении интервального вариационного ряда первая строка таблицы заполняется равными по длине интервалами значений исследуемой совокупности, вторая — соответствующими абсолютными или относительными частотами.

Пусть из некоторой генеральной совокупности в результате n наблюдений извлечена выборка объемом n. *Статистическим распределением выборки* называется перечень вариант и соответствующих им абсолютных или относительных частот. Точечный вариационный ряд *абсолютных частот* может быть представлен таблицей:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

причем $\sum_{i=1}^{k} n_i = n$.

Точечный вариационный ряд *относительных частот* представляют таблицей:

x_i	x_1	x_2	...	x_k
p_i^*	p_1^*	p_2^*	...	p_k^*

причем $\sum_{i=1}^{k} p_i^* = 1$.

При построении интервального распределения существуют правила выбора числа интервалов или величины каждого интервала. Критерием здесь служит оптимальное соотношение: при увеличении числа интервалов улучшается репрезентативность, но увеличивается объем данных и время на их обработку. Разность $x_{\max} - x_{\min}$ между наибольшим и наименьшим значениями вариант называют *размахом* выборки.

Для подсчета числа интервалов k обычно применяют эмпирическую формулу Стерджесса:

$$k = 1 + 3{,}322 \lg n \tag{3.1}$$

(подразумевается округление до ближайшего целого).

Соответственно, **величину каждого интервала *h* можно вычислить** по формуле

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n}. \qquad (3.2)$$

За начало первого интервала рекомендуется **брать величину**, равную

$$x_{\min} = x_{\max} - 0{,}5h.$$

Каждый интервал должен содержать не менее пяти вариант. В том случае, когда число вариант в интервале меньше пяти, соседние интервалы принято объединять.

В американском журнале «Литературное обозрение» с помощью статистических методов было проведено исследование прогнозов относительно исхода предстоящих в 1936 г. выборов президента США. Основными претендентами на этот пост были Ф. Д. Рузвельт и А. М. Ландон. В качестве источника для генеральной совокупности исследуемых американцев были взяты справочники телефонных абонентов. Из этих справочников случайным образом были отобраны 4 млн адресов, по которым редакция журнала разослала открытки с просьбой высказать свое отношение к кандидатам на пост президента. Обработав результаты опроса, журнал опубликовал социологический прогноз о том, что на предстоящих выборах с большим перевесом победит А. М. Ландон и ...ошибся: победу на выборах одержал Ф. Д. Рузвельт. И это не единственный пример просчетов социологических исследований методами математической статистики.

? Так может статистические методы малоэффективны, не дают надежных (гарантированных) результатов?

Однако виноваты не методы, а редакция журнала, которая неумело ими пользовалась. В ходе исследования были допущены две серьезные ошибки:

- в первой половине XX в. телефоны имела лишь зажиточная часть населения США — главы семейств, которые придерживались тех же политических взглядов, что и А. М. Ландон;
- ответы поступили не от всех респондентов, а только от уверенных в себе деловых людей, которые благодаря профессиональным навыкам привыкли отвечать на различные письма и запросы. Этот слой общества также поддерживал Ландона.

Поэтому выборку, произведенную журналом «Литературное обозрение», нельзя признать репрезентативной. Таким образом, необходимо тщательно продумывать, как осуществлять выборку из генеральной совокупности.

Так, социологи Дж. Геллап и Э. Роунер, проведя анализ 4 тыс. анкет, правильно предсказали победу на тех же выборах Ф. Д. Рузвельта. При выборе респондентов они учли, что общество распадается на различные, достаточно однородные по своим политическим взглядам, группы. Поэтому даже относительно малочисленные выборки из таких социологических слоев дают достаточно точные результаты прогнозов. Этот простой пример подчеркивает фундаментальную роль математической статистики не только в социологии и экономике, но и в политике.

3.2. Графическое представление эмпирических данных

Графически эмпирические данные можно представить в виде кумуляты — графика эмпирической функции распределения, а также полигона частот — для точечного вариационного ряда и гистограммы частот — для интервального вариационного ряда.

При дискретном (точечном) распределении признака X получают график в виде ломаной линии, соединяющей точки с координатами (x_k, n_k), где x_k — варианты выборки; n_k — соответствующие им частоты.

При непрерывном (интервальном) распределении признака X интервал, в котором заключены все наблюдательные значения признака, разбивают на частичные интервалы и находят сумму частот вариант, попавших в этот интервал.

3.2.1. Эмпирическая функция распределения. Кумулята

В этом подразделе рассмотрим график эмпирической функции распределения — кумуляту.

Пусть $x_1, x_2, ..., x_n$ — выборка из генеральной совокупности, представленной случайной величиной X. Каждая генеральная совокупность как случайная величина имеет некоторую функцию распределения $F(x)$. Обычно она неизвестна. Задача заключается в том, что необходимо оценить функцию распределения $F(x)$ этой СВ X. По выборке можно найти эмпирическую функцию распределения $F^*(x)$.

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, которая определяет для каждого значения x относительную частоту события $X<x$.

Теоретической функцией распределения называют функцию распределения генеральной совокупности, которая определяет вероятность события $X<x$. На основании закона больших чисел (в форме Бернулли) эмпирическая функция распределения выборки $F^*(x)$ служит для приближенного представления теоретической функции распределения генеральной совокупности. Итак,

$$F^*(x)=\sum_{x_i<x} p_i^* = \frac{1}{n}\sum_{x_i<x} n_i = \frac{n_x}{n}, \qquad (3.3)$$

где n — объем выборки; $n_x = \sum\limits_{x_i<x} n_i$ — число вариант, значения которых меньше x.

Отличие эмпирической функции распределения $F^*(x)$ от реальной (теоретической) $F(x)$ заключается в том, что для ее составления берется не вся генеральная совокупность, а выборка, и вероятность p_i заменяется относительной частотой p_i^*. Таким образом, теоретическая функция $F(x)$ определяет *вероятность* события $X<x$, а эмпирическая функция $F^*(x)$ — *относительную частоту* этого же события.

Следует заметить, что вероятность p_i — тоже относительная частота, только примененная к самой генеральной совокупности: $p_i = N_i/N$.

Свойства эмпирической функции распределения (аналогичны свойствам теоретической функции распределения ДСВ, см. подразд. 2.1):

- значения $F^*(x)$ принадлежат отрезку $[0;1]$;
- $F^*(x)$ — неубывающая функция;
- если задана ранжированная последовательность вариант x_1, x_2, ..., x_n, то

$$F^*(x_1)=0,\ F^*(x_n+0)=1.$$

Итак, если выборка задана вариационным рядом (x_i, p_i^*), то эмпирическая функция распределения имеет вид

$$F^*(x)=\left\{\begin{array}{ll} 0 & \text{при} \quad x\le x_1; \\ n_1/n & \text{при } x_1<x\le x_2; \\ (n_1+n_2)/n & \text{при } x_2<x\le x_3; \\ (n_1+n_2+n_3)/n & \text{при } x_3<x\le x_4; \\ \ldots & \qquad \ldots \\ (n_1+n_2+\ldots+n_{n-1})/n & \text{при } x_{n-1}<x\le x_n; \\ 1 & \text{при} \quad x>x_n. \end{array}\right\} = \qquad (3.4)$$

$$=\sum_{i=1}^{n}\theta(x-x_i)p_i^*.$$

График эмпирической функции распределения называют ***кумулятой***. Для выборки, взятой из генеральной совокупности, имеющей вид ДСВ, кумулята носит характер ступенчатой ломаной (рис. 3.1). Для выборки, взятой из генеральной совокупности, имеющей вид НСВ, для построения кумуляты точки с координатами $(x_i;\ F^*(x_i))$ соединяют отрезками. В таком случае кумулята представляет собой непрерывную ломаную.

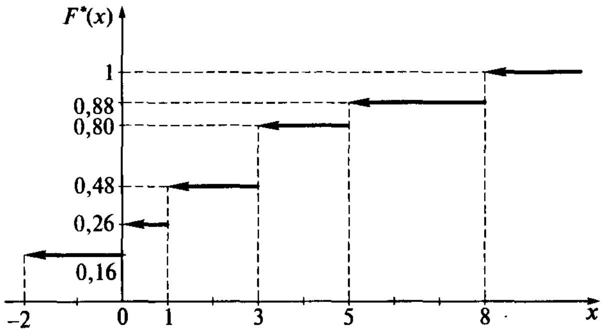


Рис. 3.1

Вариационный ряд накопительных частот (x_i, p_i^*) называют также кумулятивным рядом.

Задача 3.1. Контролер ОТК анализировал отклонение длины деталей в миллиметрах от стандарта на основе выборки, состоящей из 50 деталей. По результатам выборки построить эмпирическую функцию распределения:

x_i	–2	0	1	3	5	8
n_i	8	5	11	16	4	6

Решение. Объем выборки равен $n = 8+5+11+16+4+6 = 50$. Тогда вариационный ряд относительных частот имеет вид:

x_i	–2	0	1	3	5	8
n_i/n	0,16	0,10	0,22	0,32	0,08	0,12

В соответствии с формулой (3.4) построим эмпирическую функцию распределения:

$$F^*(x) = \begin{cases} 0 & \text{при} \quad x \le -2; \\ 0,16 & \text{при} \quad -2 < x \le 0; \\ 0,16+0,10=0,26 & \text{при} \quad 0 < x \le 1; \\ 0,26+0,22=0,48 & \text{при} \quad 1 < x \le 3; \\ 0,48+0,32=0,8 & \text{при} \quad 3 < x \le 5; \\ 0,8+0,08=0,88 & \text{при} \quad 5 < x \le 8; \\ 0,88+0,12=1 & \text{при} \quad x > 8. \end{cases}$$

Соответствующий кумулятивный ряд примет вид:

x_i	–2	0	1	3	5	8
n_i/n	0,16	0,10	0,22	0,32	0,08	0,12
p^*	0,16	0,26	0,48	0,80	0,88	1

Построим график соответствующей эмпирической функции распределения (см. рис. 3.1).

Кумулята дает возможность понимать графически представленную информацию, например, ответить на вопросы: «Определить число деталей, у которых отклонение от стандарта:

а) менее 3 мм;

б) не менее 3 мм».

Значение эмпирической функции распределения при $x = 3$ равно $F^*(3) = 0{,}80$. Тогда число деталей, у которых отклонение от стандарта менее 3 мм, найдем как произведение $n(x < 3) = 0{,}80 \cdot 50 = 40$ (дет.), а число деталей, у которых отклонение от стандарта не менее 3 мм, равно $n(x \geq 3) = (1 - 0{,}80) \cdot 50 = 10$ (дет.).

Если задан интервальный вариационный ряд, то для составления эмпирической функции находят середины интервалов $x_i + h/2$ и по ним получают эмпирическую функцию аналогично точечному вариационному ряду.

Таким образом, вся часть выборки, попавшая в интервал $(x_i, x_i + h]$, как бы «концентрируется» в середине этого интервала — в точке $x_i + h/2$.

3.2.2. Полигон и гистограмма

В этом подразделе вы научитесь строить полигоны частот и гистограммы — графики вариационных рядов.

Пусть результаты выборки количественного признака X из генеральной совокупности представлены вариационным рядом. Кумулята, будучи функцией распределения выборки, служит ее интегральной вероятностной характеристикой. Для изучения локальных свойств нужна функция, аналогичная ряду распределения или плотности вероятности.

Точечный вариационный ряд наглядно можно представить с помощью полигона частот, а интервальный — с помощью гистограммы.

Полигоном частот называют ломаную, отрезки которой соединяют точки с координатами (x_1, n_1), (x_2, n_2), ..., (x_k, n_k) для полигона абсолютных частот и точки с координатами (x_1, p_1^*), (x_2, p_2^*), ..., (x_k, p_k^*) для полигона относительных частот.

Для их построения, исходя из условия задачи, необходимо составить точечный (либо интервальный) вариационный ряд абсолютных или относительных частот. Если задан интервальный вариационный ряд, то для построения полигона частот необходимо найти середины интервалов $\Delta_1, \Delta_2, \dots, \Delta_k$ — точки $x_1, x_2, \dots, x_k$ (если крайние интервалы не ограничены, то для них не нужно, а точнее, невозможно, найти середину). Затем в декартовой системе координат надо отложить на оси абсцисс все возможные значения вариант $x_1, x_2, \dots, x_k$, а на оси ординат — соответствующие им абсолютные частоты $n_1, n_2, \dots, n_k$ (или относительные частоты $p_1^*, p_2^*, \dots, p_k^*$). Для построения полигона абсолютных (относительных) частот необходимо соединить отрезками полученные точки с координатами $(x_1, n_1), (x_2, n_2), \dots, (x_n, n_k)$ (или, соответственно, $(x_1, p_1^*), (x_2, p_2^*), \dots, (x_k, p_k^*)$).

Гистограммой абсолютных (относительных) частот называют ступенчатую фигуру, состоящую из прямоугольников, основания которых есть частичные интервалы длиной h (одинаковой для всех интервалов!), а высоты равны отношению n_i/h, т.е. пропорциональны частоте интервала n_i. Отношение n_i/h называют *плотностью частоты*, а отношение p_i^*/h — *плотностью относительной частоты*.

Для построения гистограммы необходимо найти *размах* выборки — ее границы, т.е. $x_{\max}$ и $x_{\min}$, длину интервалов $h = \dfrac{x_{\max} - x_{\min}}{k}$, а также k — число интервалов.

При удачном подборе интервалов гистограмма и полигон дают представление о графике функции плотности вероятности распределения генеральной совокупности, что используется для формулировки предположения о виде исследуемого теоретического закона распределения (см. подразд. 3.6.6).

Заметим, что если за единицу высоты принять $1/h$, площадь каждого i-го прямоугольника равна произведению основания на высоту: (hn_i/h), т.е. частоте i-го интервала (или сумме частот вариант, входящих в этот интервал). Тогда площадь всей гистограммы частот равна сумме всех частот $\sum_{i=1}^{k} n_i = n$, т.е. объему выборки, причем площадь и высота каждого столбца гистограммы пропорциональны частоте попадания наблюдений в данный интервал группировки. Если в качестве высот прямоугольников выбрать отношения n_i/hn, то площадь фигуры под гистограммой относительных частот равна единице, так как $S = h\sum_{i=1}^{k} \dfrac{n_i}{hn} = 1$.

Таким образом, площадь гистограммы абсолютных (относительных) частот равна сумме всех абсолютных (относительных) частот, т.е. объему выборки (единице).

Задача 3.2. Представить графическое распределение размеров заработную платы сотрудников фирмы за неделю (в усл. ед.), если они получили следующую заработную плату:

152,74; 176,66; 162,48; 167,72; 181,09; 155,00; 196,17; 169,60; 172,88; 182,47; 181,69; 186,91; 190,10; 176,14; 192,70; 178,59; 167,27; 175,14; 160,00; 177,46; 165,18; 167,77; 178,46; 165,00; 185,20; 157,02; 172,14; 192,22; 179,40; 191,03; 188,68; 169,51; 200,15; 178,47; 176,33; 179,05; 180,95; 174,28; 175,00; 178,45; 150,10; 176,86; 187,71; 168,33; 195,00; 172,37; 179,04; 182,05; 186,19; 190,05; 196,27; 209,28; 203,16; 168,52; 200,00; 196,30.

Решение. Найдем минимальное и максимальное значения вариант и объем выборки: $x_{max} = 209{,}28$; $x_{min} = 150{,}10$, $n = 56$.

Применим формулу Стерджесса и подсчитаем число интервалов:

$$h = \frac{x_{max} - x_{min}}{k} = \frac{x_{max} - x_{min}}{1 + 3{,}322 \lg n} = \frac{209{,}28 - 150{,}10}{1 + 3{,}322 \lg 56} \approx 10.$$

Построим интервальный вариационный ряд с интервалом $h = 10$:

Интервалы заработной платы	150—160	160—170	170—180	180—190	190—200	200—210
n_i	4	11	18	10	9	4

Для построения гистограммы вычислим высоты прямоугольников, т.е. отношения n_i/h (*плотности частоты*), где все значения $h_i = 10$:

Интервалы заработной платы	150—160	160—170	170—180	180—190	190—200	200—210
n_i	4	11	18	10	9	4
n_i/h_i	0,4	1,1	1,8	1,0	0,9	0,4

Для построения точечного вариационного ряда найдем середины интервалов:

x_i	155	165	175	185	195	205
n_i	4	11	18	10	9	4
n_i/h_i	0,4	1,1	1,8	1,0	0,9	0,4

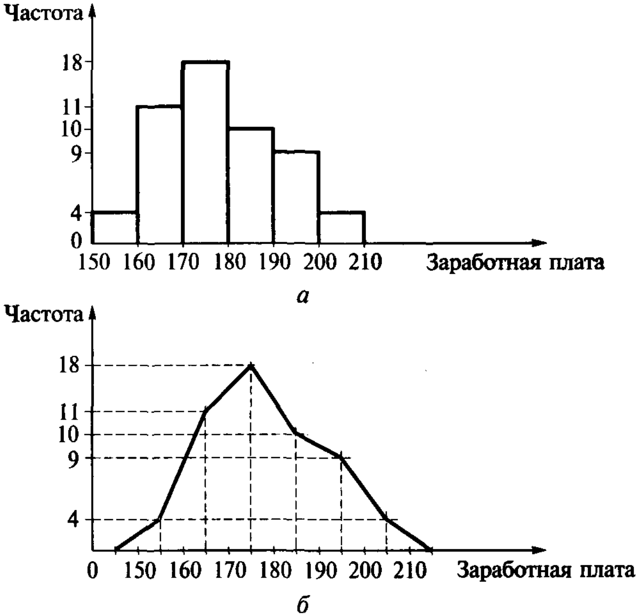


Рис. 3.2

Составим графики распределения заработной платы (рис. 3.2). Интервальному вариационному ряду соответствует гистограмма (рис. 3.2, *а*); точечному вариационному ряду — полигон (рис. 3.2, *б*).

Задача 3.3. Даны результаты изменения напряжения (в вольтах) в электросети. Составить вариационный ряд и начертить график распределения напряжения, если значения напряжения следующие:

227, 215, 230, 232, 223, 220, 228, 222, 221, 226, 226, 215, 218, 220, 216, 220, 225, 212, 217, 220.

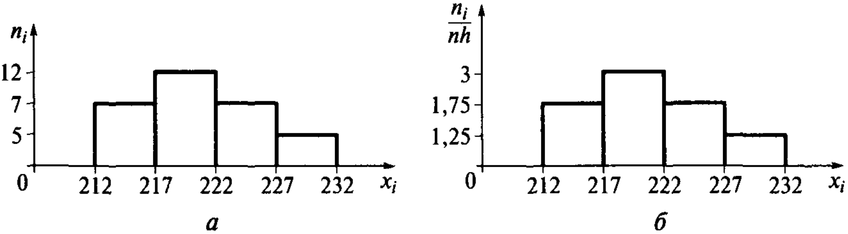


Рис. 3.3

Решение. Составим вариационный ряд, по которому построим гистограмму (рис. 3.3). Имеем $n = 20$, $x_{max} = 232$, $x_{min} = 212$. Применим формулу Стерджесса для подсчета числа интервалов:

$$h = \frac{x_{max} - x_{min}}{k} = \frac{x_{max} - x_{min}}{1 + 3{,}322 \lg n} = \frac{232 - 212}{1 + 3{,}322 \lg 20} \approx 4.$$

Так как график интервального вариационного ряда — гистограмма, вычислим для нее высоты прямоугольников как отношение n_i/h и результаты занесем в таблицу. Тогда интервальный вариационный ряд абсолютных частот имеет вид:

x_i	212—217	217—222	222—227	227—232
n_i	5	7	5	3
n_i/h	1,25	1,75	1,25	0,75

Для вычисления высот прямоугольников при построении гистограммы относительных частот найдем отношения n_i/nh. Интервальный вариационный ряд относительных частот примет вид:

x_i	212—217	217—222	222—227	227—232
n_i/n	0,25	0,35	0,25	0,15
n_i/nh	0,06	0,08	0,06	0,04

Соответствующая гистограмма абсолютных частот представлена на рис. 3.3, *а*, а относительных частот — на рис. 3.3, *б*.

Задача 3.4. Контролер на рынке выявляет отклонение весов в граммах от стандарта на основе выборки. Закон распределения выборки задан вариационным рядом абсолютных частот:

x_i	–2	0	3	5	8
n_i	5	1	7	3	4

Составить закон распределения относительных частот и построить их графики.

Решение. 1) Найдем объем выборки: $n = \sum n_i = 5 + 1 + 7 + 3 + 4 = 20$.

2) Определим относительные частоты по формуле $w_i = n_i/n$: $w_1 = 0{,}25$; $w_2 = 0{,}05$; $w_3 = 0{,}35$; $w_4 = 0{,}15$; $w_5 = 0{,}20$.

3) Составим вариационный ряд относительных частот:

x_i	–2	0	3	5	8
w_i	0,25	0,05	0,35	0,15	0,20

4) Полигон для вариационного ряда распределения абсолютных частот представлен на рис. 3.4, *а*, относительных частот — на рис. 3.4, *б*.

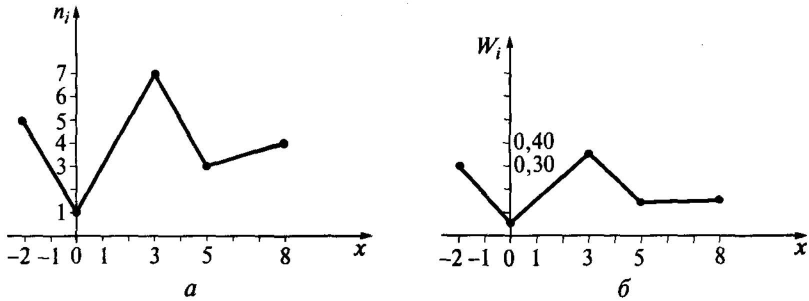


Рис. 3.4

Интервал группировки значительно влияет на вид гистограммы: чем больше h, тем менее различимы особенности распределений.

3.3. Числовые характеристики вариационного ряда

В этом подразделе даны формулы для определения числовых характеристик вариационного ряда: математического ожидания, дисперсии, моды, медианы, среднеквадратического отклонения, коэффициента вариации.

Пусть подлежит обследованию некоторая генеральная совокупность, например, поступление в Федеральный бюджет налогов плательщиков города Тольятти. Такие поступления (согласно налоговому кодексу РФ) включают единый социальный налог, налог на добавленную стоимость, госпошлину и т.д. Ежегодные социологические исследования, т.е. серии выборок объема n, дают некоторые статистические наблюдения за этим процессом. Пусть из генеральной совокупности произведены все возможные выборки равного объема n и для каждой выборки рассчитаны выборочные средние. Например, произведены выборки за прошедшие k лет. По каждому показателю такие исходные данные могут быть представлены в виде последовательности величин X_1, X_2, ..., X_k, где X_i —

результат обследования в i-м году. Распределение, полученное из выборки, называют *выборочным*, а его характеристики — выборочными. Репрезентативность выборки зависит от полноты и достоверности представленной информации.

Поскольку параметры распределения генеральной совокупности являются дискретными (или непрерывными) случайными величинами, то они обладают числовыми характеристиками ДСВ (или НСВ) — математическим ожиданием, дисперсией, модой, медианой, среднеквадратическим отклонением, которые находят по правилам, известным из теории вероятностей. Числовые характеристики генеральной совокупности будем обозначать с сопровождающим индексом: *генеральное среднее* $\bar{x}_г$, *генеральная дисперсия* $D_г$, что соответствует математическому ожиданию $M(X)$ и дисперсии $D(X)$ некоторой непрерывной или дискретной случайной величины некоторого теоретического распределения.

Генеральным средним $\bar{x}_г$ называется среднее арифметическое значений исследуемого признака X генеральной совокупности:

$$\bar{x}_г = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}(x_1 + x_2 + \ldots + x_N).$$

Для ограниченной генеральной совокупности ($N < \infty$) генеральное среднее, очевидно, всегда существует.

Генеральной дисперсией $D_г$ называется среднее арифметическое квадратов отклонений значений признака генеральной совокупности от их среднего значения $\bar{x}_г$:

$$D_г = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_г)^2.$$

Генеральным среднеквадратическим отклонением (стандартом) называется корень квадратный из генеральной дисперсии:

$$\sigma_г = \sqrt{D_г}.$$

Вывод о параметрах *генеральной* совокупности можно сделать на основе изучения числовых характеристик *выборки*.

Пусть из некоторой генеральной совокупности произведены серии выборок объемом n, причем для каждой выборки рассчитаны выборочные средние. Тогда выборочное среднее также является случайной величиной, все возможные значения которой задают распределение выборочной средней. Полученные значения можно представить в виде ряда распределений выборочных средних и найти среднее значение этого распределения $\bar{X}$.

Выборочным средним $\bar{x}_в$ называется среднее арифметическое значений исследуемого признака X выборки, а ***выборочной дис-***

персией $D_в$ — среднее арифметическое квадратов отклонения наблюдаемых значений признака от $\bar{x}_в$.

$$\bar{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \ldots + x_n); \quad D_в = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2.$$

В дальнейшем обозначение числовых характеристик выборки может встречаться как с индексом, так и без него (по умолчанию).

Рассмотрим некоторые числовые характеристики выборки, представленной вариационным рядом.

Для дискретного вариационного ряда ***модой*** Мо называется варианта, которая имеет максимальную статистическую вероятность, ***медианой*** Ме — середина распределения, т.е. такая точка, в которой половина принимаемых значений лежит слева от нее, а половина — справа. Медиану можно найти по формуле

$$\mathrm{Me} = \begin{cases} [x_{n/2} + x_{n/2+1}]/2, & \text{если } n \text{ — четное;} \\ x_{(n+1)/2}, & \text{если } n \text{ — нечетное.} \end{cases} \qquad (3.4)$$

Коэффициентом вариации V называется отношение выборочного среднеквадратического отклонения к выборочному среднему:

$$V_в = \frac{\sigma_в}{\bar{x}_в}.$$

Так как коэффициент вариации — безразмерная величина, то его применяют для сравнения разброса данных тех вариационных рядов, параметры которых имеют различную размерность.

Выборочный метод основан на законе больших чисел и центральной предельной теореме (см. подразд. 2.10.5), согласно кото-

Таблица 3.1

Характеристика	Способ задания вариационного ряда		
	последовательностью	таблицей абсолютных частот	таблицей относительных частот
Среднее значение выборки $\bar{x}_в$	$\frac{1}{n}\sum_{i=1}^{n} x_i$	$\frac{1}{n}\sum_{j=1}^{k} x_j n_j$	$\sum_{j=1}^{k} x_j p_j^*$
Дисперсия выборки $D_в$	$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2 =$ $= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}_в^2$	$\frac{1}{n}\sum_{j=1}^{k} x_j^2 n_j - \bar{x}^2 =$ $= \frac{1}{n}\sum_{j=1}^{k}(x_j - \bar{x})^2 n_j$	$\sum_{j=1}^{k} x_j^2 \frac{n_j}{n} - \bar{x}^2 =$ $= \sum_{j=1}^{k}(x_j - \bar{x})^2 p_j^*$

рым если генеральная совокупность имеет нормальное распределение, то выборочное распределение также подчиняется нормальному закону распределения. Однако при достаточно большом объеме выборки распределение выборочных средних подчиняется нормальному закону распределения независимо от того, какой закон распределения имеет генеральная совокупность.

Формулы для вычисления числовых характеристик случайных величин в математической статистике аналогичны соответствующим формулам теории вероятностей.

Формулы для вычисления основных числовых характеристик выборки приведены в табл. 3.1.

3.4. Статистические оценки параметров распределения

Статистическая оценка параметров генеральной совокупности (теоретического распределения) как совокупность методов, позволяющих делать научно обоснованные выводы о значениях числовых параметров генеральной совокупности по случайной выборке из нее, — одна из главных задач математической статистики.

Пусть некоторая генеральная совокупность исследуется по выборке $\{X_1, X_2, ..., X_n\}$. При исследовании по выборкам, случайно отобранным из генеральной совокупности, можно получить лишь приближенные значения неизвестного параметра Θ, которые служат его *оценкой*. Обозначим через X анализ доходов населения нашего города, тогда такой анализ можно осуществить на основе выборки ограниченного объема, например при $n = 1000$. Тогда числовые характеристики выборки — средний доход и разброс в доходах для лиц, попавших в выборку, можно оценить по формулам соответственно $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ и $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Однако неизвестно, можно ли полученные результаты обобщить и представить как характеристики дохода жителей всего города. Очевидно, что оценки могут изменяться от одной выборки к другой. Обозначим через Θ^* точечную статистическую оценку некоторого параметра Θ теоретического распределения.

Статистической оценкой Θ^* неизвестного параметра теоретического распределения называется функция $f(x, \Theta)$ от наблюдаемых случайных величин выборки. Задача статистического оценивания неизвестных параметров Θ по выборке заключается в построении такой функции от имеющихся данных статистических наблюдений, которая давала бы наиболее точные приближенные значения реальных, не известных исследователю, значений этих параметров.

3.4.1. Виды статистических оценок. Основные требования к точечным оценкам

В этом подразделе рассмотрим виды статистических оценок — точечные и интервальные, а также требования к точечным оценкам — несмещенность, эффективность и состоятельность.

Статистические оценки подразделяют на *точечные* и *интервальные*, в зависимости от способа их представления (соответственно, числом или интервалом).

Так, для нормального закона распределения с плотностью вероятности $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$ параметрами служат математическое ожидание m и среднеквадратическое отклонение σ, а для равномерно распределенной генеральной совокупности с плотностью вероятности $f(x)=\frac{1}{b-a}$ параметрами служат концы интервала a и b.

Точечной называют статистическую оценку параметра Θ теоретического распределения, определяемую *одним* значением параметра $\Theta^*=f(x_1, ..., x_n)$, где $x_1, ..., x_n$ — результаты эмпирических наблюдений над количественным признаком X некоторой выборки. Такие оценки параметров этой совокупности, полученные по различным выборкам, чаще всего отличаются друг от друга. Абсолютную разность $|\Theta^*-\Theta|$ называют *ошибкой выборки* (оценивания).

Для того чтобы статистические оценки давали достоверные представления об оцениваемых параметрах, необходимо выполнение ряда условий (свойств): достоверная оценка статистического параметра Θ должна быть несмещенной, эффективной и состоятельной.

Точечная оценка, математическое ожидание которой равно (не равно) оцениваемому параметру, называется несмещенной (смещенной).

Таким образом, *несмещенной* называют такую статистическую оценку Θ^* параметра Θ, для которой $M\Theta^*=\Theta$.

Разность $M\Theta^*-\Theta$ называется *смещением*, или *систематической ошибкой* оценивания. Для несмещенных оценок систематическая ошибка равна нулю.

Эффективной называют такую статистическую оценку Θ^*, которая при заданном объеме выборки n имеет наименьшую возможную дисперсию: $D[\Theta^*]|_{n=\text{const}}\to\min$. Эффективная оценка имеет наименьший разброс по сравнению с другими несмещенными и состоятельными оценками.

Состоятельной называют такую статистическую оценку Θ^* параметра, которая при $n\to\infty$ стремится по вероятности к оцениваемому параметру Θ, т. е. при увеличении объема выборки n оценка стремится по вероятности к истинному значению неизвестного параметра Θ.

Выполнение условия состоятельности гарантирует отсутствие грубых ошибок в оценке Θ при достаточно больших n.

На практике не всегда удается выполнить все эти требования и получить оценку параметра, вызывающую доверие. Напомним, что стремление по вероятности означает выполнение равенства: при любом $\varepsilon > 0$

$$\lim_{n\to\infty} P(|\Theta^* - \Theta| < \varepsilon) = 1.$$

Это требование согласуется с законом больших чисел: чем больше исходной информации об исследуемом объекте, тем точнее результат свойства состоятельности, проверяемый в первую очередь.

Однако если объем выборки мал, то точечная оценка параметра может привести к серьезным ошибкам.

3.4.2. Точечные оценки

В этом подразделе вы познакомитесь с формулами для вычисления числовых характеристик и научитесь решать задачи, связанные с их применением.

Так как $X_1, X_2, ..., X_n$ — независимые, одинаково распределенные случайные величины, то все они имеют один и тот же закон распределения вероятностей и одинаковые числовые характеристики.

Среднее выборочное $\bar{x}_в$ удовлетворяет всем накладываемым к статистическим оценкам требованиям, т. е. дает несмещенную, эффективную и состоятельную оценку. Действительно:

$$M(\bar{x}_в) = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(x_i) = \frac{1}{n}\sum_{i=1}^{n} \bar{x}_г = \frac{1}{n}(n\bar{x}_г) = \bar{x}_г. \quad (3.5)$$

Это равенство следует из того, что все СВ X_i, соответствующие наблюдаемым значениям x_i, распределены одинаково с математическим ожиданием $M(X) = \bar{x}_г$. Поэтому $\bar{x}_в$ является несмещенной оценкой $\bar{x}_г$.

В то же время эта оценка является *состоятельной*: согласно закону больших чисел, при увеличении n величина $\bar{x}_в$ сходится по вероятности к математическому ожиданию. Говорят, выборочное среднее обладает свойством *статистической устойчивости* (см. подразд. 2.10.4).

Оценим по данным выборки неизвестную генеральную дисперсию $D_г$. Поступим аналогично, т. е. в качестве оценки $D_г$ возьмем $D_в$. Можно доказать, что математическое ожидание $D_в$ равно

$$M[D_в] = \frac{n-1}{n} D_г.$$

Таким образом, $D_в$ оказывается смещенной оценкой генеральной дисперсии, давая *заниженное* значение $D_г$. Это значит, что при малых n ее использование приведет к *систематическим* ошибкам. Для несмещенной оценки $D_г$ достаточно взять величину $\frac{n}{n-1} D_в$, которую называют *исправленной дисперсией* и обозначают s^2. Тогда

$$s^2 = \frac{n}{n-1} D_в = \frac{n}{n-1} \sum_{i=1}^{k} \frac{n_i}{n} (x_i - \bar{x}_в)^2 = \frac{1}{n-1} \sum_{i=1}^{k} n_i (x_i - \bar{x}_в)^2,$$

$$M[s^2] = M\left[\frac{n}{n-1} D_в\right] = D_г = \sigma^2.$$

Таким образом, математическое ожидание исправленной дисперсии действительно равно дисперсии генеральной совокупности и, значит, s^2 — состоятельная оценка генеральной дисперсии.

На практике для оценки генеральной дисперсии применяют исправленную дисперсию при $n \leq 30$. В остальных случаях ($n > 30$), отклонение $D_в$ от $D_г$ малозаметно. Поэтому при больших значениях n ошибкой «смещения» порядка $1/n$ можно пренебречь, так как при $n \to \infty$ коэффициент $\frac{n}{n-1} \to 1$, т. е. s^2 — состоятельная оценка.

Итак, несмещенная оценка дисперсии для выборки, заданной последовательностью значений или таблицей относительных частот, имеет вид

$$s^2 = \frac{n}{n-1} D(X) = \frac{n}{n-1} \sigma^2 = \frac{1}{n-1} \sum_{i=1}^{k} n_i (x_i - \bar{x})^2 =$$

$$= \frac{n}{n-1} \sum_{i=1}^{k} (x_i - \bar{x})^2 p_i^*. \qquad (3.6)$$

Пусть некоторая случайная величина X имеет математическое ожидание $M(X) = m$ и дисперсию $D(X) = \sigma^2$. В ходе эксперимента получена случайная выборка из n независимых испытаний случайной величины X. Тогда справедливы следующие утверждения:

1) среднее выборочное $\bar{x}_в$ служит несмещенной и состоятельной оценкой математического ожидания $M(X)$;

2) если случайная величина X распределена по нормальному закону с параметрами $N(m, \sigma)$, то среднее выборочное $\bar{x}_в$ также рас-

пределено нормально и имеет минимальную дисперсию $D(\bar{x}_в)=\frac{\sigma^2}{n}$, т.е. $M(\bar{x}_в)=m$, $\sigma(\bar{x}_в)=\frac{\sigma}{\sqrt{n}}$. Поэтому среднее выборочное $\bar{x}_в$ — эффективная и состоятельная оценка математического ожидания;

3) выборочная дисперсия $D_в=\sigma_в^2=\frac{1}{n}\sum_{i=1}^{k}n_i(x_i-\bar{x}_в)^2$ является смещенной оценкой генеральной дисперсии σ^2. Несмещенной оценкой генеральной дисперсии σ^2 является «исправленная» дисперсия s^2, для получения которой необходимо умножить σ^2 на так называемую *поправку Бесселя* $\frac{n}{n-1}$. Тогда

$$s^2=\frac{n}{n-1}D_в=\frac{1}{n-1}\sum_{i=1}^{k}n_i(x_i-\bar{x}_в)^2.$$

«Исправленная» выборочная дисперсия s^2 является состоятельной оценкой генеральной дисперсии σ^2;

4) если известно m — математическое ожидание случайной величины X, то выборочная дисперсия $D_в=\sigma_в^2=\frac{1}{n}\sum_{i=1}^{k}n_i(x_i-m)^2$ является несмещенной, состоятельной и эффективной оценкой генеральной дисперсии σ^2;

5) относительная частота $\frac{n_i}{n}$ является несмещенной и состоятельной оценкой вероятности $P(X=x_i)$. Эмпирическая функция распределения $F^*(x)=\frac{n_x}{n}$ — накопленная относительная частота — является несмещенной и состоятельной оценкой теоретической функции распределения $F(x)=P(X<x)$.

Задача 3.5. Найти несмещенные оценки математического ожидания и дисперсии по таблице выборки:

x_i	2	6	12
n_i	3	10	7

Решение. Из таблицы имеем объем выборки $n=20$. Несмещенная оценка математического ожидания есть среднее выборочное $\bar{x}_в$:

$$\bar{x}_в=(2\cdot 3+6\cdot 10+12\cdot 7)/20=7{,}5.$$

Для вычисления несмещенной оценки дисперсии сначала найдем выборочную дисперсию, а затем несмещенную оценку s^2:

$$D(\bar{x}_в) = (4 \cdot 3 + 36 \cdot 10 + 144 \cdot 7)/20 - 7,5^2 = 12,75;$$

$$s^2 = \frac{20}{19} \cdot 12,75 = 13,42.$$

Задача 3.6. Найти несмещенные числовые характеристики выборки, заданной таблицей:

x_i	2	6	12
n_i/n	0,15	0,50	0,35

Решение. Среднее выборочное $\bar{x}_в$ является несмещенной оценкой генерального среднего, а для вычисления несмещенной дисперсии s^2 предварительно вычислим смещенную дисперсию $D(X) = \sigma^2$:

$$\bar{x}_в = 2 \cdot 0,15 + 6 \cdot 0,5 + 12 \cdot 0,35 = 7,5;$$

$$\sigma^2 = (4 \cdot 0,15 + 36 \cdot 0,5 + 144 \cdot 0,35) - (7,5)^2 = 12,75; \; s^2 = \frac{20}{19} \cdot 12,75 = 13,42;$$

$$\sigma = \sqrt{12,75} = 3,57; \; s = \sqrt{13,42} = 3,66.$$

Легко видеть, что задачи 3.5 и 3.6 задают одну и ту же выборку, но в задаче 3.5 она задается таблицей абсолютных частот, а в задаче 3.6 — таблицей относительных частот:

$$n_1/n = 0,15 = 3/20; \; n_2/n = 0,5 = 10/20; \; n_3/n = 0,35 = 7/20.$$

На практике если значение вариант x_i — большие числа, то для облегчения расчетов их представляют в виде суммы некоторого постоянного числа c и условной варианты u_i, как дополнения до x_i, т.е. $x_i = c + u_i$. Это значит, что задан некий новый вариационный ряд для величины U, определенный по выборочным данным u_i. Поскольку выбор c произволен, то лучше взять за c значение, близкое к $\bar{x}_в$. Тогда $\bar{x}_в = c + \sum p_i^* u_i$, а дисперсия не изменится, т.е. $D_в(X) = D_в(U)$, так как по свойствам дисперсии

$$D(C + U) = D(C) + D(U) = D(U),$$

где $C = \text{const}$.

Тогда

$$D_в(X) = D_в(U) = \overline{u^2} - \bar{u}_в^2 = \frac{1}{n}\sum n_i u_i^2 - \frac{1}{n^2}\left(\sum n_i u_i\right)^2.$$

Аналогично вычисляется несмещенная оценка дисперсии:

$$s^2(X) = s^2(U) = \frac{1}{n-1}\left(\sum n_i u_i^2 - \frac{1}{n}\left(\sum n_i u_i\right)^2\right).$$

Если первоначальные варианты представлены десятичными дробями, то их умножают на постоянное число $c = 10^k$, где k — количество десятичных знаков. Тогда условные варианты имеют вид $u_i = cx_i$, т.е. дисперсия увеличилась в c^2 раз, согласно свойству дисперсии. Поэтому $x_i = \frac{u_i}{c}$, а $D_в(X) = \frac{1}{c^2} D_в(U)$.

Аналогично $s^2(X) = \frac{1}{c^2} s^2(U)$.

Задача 3.7. Из генеральной совокупности извлечена выборка. Найти несмещенную оценку генеральной средней и генеральной дисперсии.

x_i	3250	3270	3280
n_i	2	5	3

Решение. 1) Найдем условную варианту и составим для нее ряд распределений.

Пусть $c = 3270$, тогда $u_i = x_i - 3270$.

u_i	–20	0	10
n_i	2	5	3

2) Так как объем выборки $n = 10$, то

$$\bar{u}_в = \frac{1}{n}\sum_{i=1}^{3} u_i n_i = \frac{(-20)\cdot 2 + 0\cdot 5 + 10\cdot 3}{10} = -1;\ \bar{x}_в = c + \bar{u}_в = 3269.$$

3) Найдем выборочную дисперсию для первоначальной варианты с помощью условной варианты:

$$D_в(X) = D_в(U) = \overline{u^2} - [\bar{u}]^2 = ((-20)^2\cdot 2 + 0^2\cdot 5 + 10^2\cdot 3)/10 - (-1)^2 = 109.$$

4) Найдем «несмещенную выборочную дисперсию» — несмещенную оценку генеральной дисперсии: $s^2 = \frac{n}{n-1} D_в = \frac{10}{9}\cdot 109 \approx$ $\approx 121{,}1$.

То, что выбор постоянной c не влияет на значение дисперсии, следует из соответствующего свойства, известного из теории вероятностей. Поэтому выбор постоянной c весьма условен и определяется удобством расчета. Особенно это очевидно при очень малых значениях V: например, если среднеквадратическое отклонение порядка 10^{-7}, а выборочное среднее порядка 10^7 (например, при очень низком отношении сигнал/шум), то затруднительно непосредственно вычислить дисперсию, так как незначительная разница будет меньше погрешности округления (например, на микрокалькуляторе). Поэтому на практике исходят из критерия удобства дальнейших расчетов.

Заметим, что в математической статистике часто используются законы распределения, связанные с нормальными, например, Пирсона ($\chi^2(n)$) и Стъюдента ($\mathrm{St}(n)$) (см. подразд. 2.9). Эти законы распределения нашли широкое применение при обработке статистических данных, поскольку зависят лишь от одного параметра — степени свободы k. Например, если задана выборка, представленная в виде k независимых случайных величин $X_1, X_2, ..., X_k$, то их композиция (сумма) имеет k степеней свободы, так как каждая из этих k величин может изменить свое значение независимо от других. В тех случаях, когда существует зависимость между отдельными величинами из этих k величин (уравнения связи), то, соответственно, число степеней свободы системы уменьшается. Если существует l функционально независимых уравнений связи, то число степеней свободы равно $k - l$. В следующем подразделе будет использован закон распределения Стъюдента для определения интервальных оценок параметров распределения.

3.5. Интервальные оценки параметров распределения

Если точечные оценки вычисляют по выборке небольшого объема, то статистика Θ^* может значительно отличаться от истинного значения параметра Θ и, следовательно, приводить к грубым ошибкам. В таких случаях более точный (более надежный, вызывающий доверие) ответ дают интервальные оценки.

Пусть имеется выборка объемом n и статистическая оценка Θ^* неизвестного параметра Θ. Поскольку статистика Θ^* получена эмпирически (по выборке), она также является случайной величиной. Если неизвестный параметр Θ оценивается с помощью значения Θ^*, полученного статистическим путем, то чем меньше модуль разности $|\Theta - \Theta^*|$, тем точнее Θ^* описывает оцениваемый параметр Θ.

Интервальной называется статистическая оценка, определяемая двумя числовыми значениями — концами исследуемого интервала. Число $\delta > 0$, при котором $|\Theta - \Theta^*| < \delta$, характеризует ***точность*** интервальной оценки.

3.5.1 Доверительная вероятность. Доверительные интервалы

В этом подразделе дадим определения доверительной вероятности (надежности), доверительного интервала, рассмотрим его виды и границы.

Доверительным называется ***интервал*** Δ_Θ, который с заданной вероятностью γ *покрывает* неизвестное значение параметра Θ. Дополнение доверительного интервала до множества всех возможных (реализуемых с ненулевой вероятностью) значений параметра Θ называется *критической областью.* Если критическая область расположена только с одной стороны от доверительного интервала, то доверительный интервал называется *односторонним*: *левосторонним*, если критическая область существует только слева, и *правосторонним* — если только справа. В противном случае доверительный интервал называется *двусторонним.*

Надежностью, или ***доверительной вероятностью***, оценки Θ (с помощью Θ^*) называют вероятность γ, с которой выполняется неравенство

$$|\Theta - \Theta^*| < \delta: \ P(|\Theta - \Theta^*| < \delta) = \gamma. \tag{3.7}$$

Чаще всего доверительную вероятность γ задают заранее (априорно) и на нее накладывают требования быть близкой к единице. Общепринятые значения надежности — 0,95; 0,99; 0,999, которые определяются в зависимости от конкретных условий. Например, надежность $\gamma = 0{,}99$ означает, что мы пренебрегаем вероятностью $\alpha = 1 - \gamma = 0{,}01$ совершить ошибку. Вероятность $\alpha = 1 - \gamma$ называют *вероятностью ошибок*, или *уровнем значимости.*

Пусть $|\Theta - \Theta^*| < \delta$, тогда $-\delta < \Theta - \Theta^* < \delta$, или $\Theta^* - \delta < \Theta < \Theta^* + \delta$. Это означает, что с вероятностью γ можно утверждать, что истинное значение параметра Θ принадлежит интервалу $(\Theta^* - \delta;\ \Theta^* + \delta)$. Точность оценки Θ^* зависит от заданной доверительной вероятности γ: оценка Θ^* тем точнее, чем меньше величина отклонения δ.

Границы (концы) доверительного интервала называют *доверительными границами*, или *критическими значениями.*

Значения границ доверительного интервала зависят от закона распределения параметра Θ^*, т.е. границы различны для различных распределений. Доверительный интервал носит случайный характер и по расположению (относительно Θ^*), и по ширине (поскольку границы зависят от данных выборки). Так как границы доверительного интервала тоже случайная величина, то принято говорить не о вероятности попадания параметра Θ в некоторый построенный доверительный интервал, а о том, что построенный доверительный интервал *покроет* параметр Θ с доверительной вероятностью (надежностью) γ.

Величину отклонения δ, равную половине ширины доверительного интервала, называют *точностью* оценки.

Методы построения доверительных интервалов впервые были разработаны американским статистиком Ю. Нейманом.

Точность оценки δ, доверительная вероятность γ и объем выборки n связаны между собой. Поэтому, зная конкретные значения двух величин, всегда можно вычислить третью по сделанному предположению относительно вида статистики. Обычно в эксперименте непосредственно наблюдаемыми величинами являются n, $\bar{x}_в$ и s.

Сформулируем правила построения доверительных интервалов для отдельных параметров распределения.

Рассмотрим дисперсию среднего выборочного:

$$D(\bar{x}_в) = D\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2\right) = \frac{1}{n^2}D\left(\sum_{i=1}^{n}(x_i - \bar{x}_в)^2\right) = \frac{1}{n^2}nD_г = \frac{D_г}{n}.$$

Тогда стандартное отклонение и «исправленное» стандартное отклонение выборочного среднего равны:

$$\sigma(\bar{x}_в) = \frac{\sigma}{\sqrt{n}};\; s(\bar{x}_в) = \frac{s}{\sqrt{n}}. \tag{3.8}$$

В зависимости от того, известно ли точное значение генеральной дисперсии $\sigma_г^2$ или нет, но найдена ее оценка s^2, рассмотрим задачи определения интервального значения для генерального среднего (математического ожидания соответствующего закона распределения).

Алгоритм построения доверительного интервала имеет вид.

1. Извлечем выборку объемом n из генеральной совокупности с известным распределением $f(x, \Theta)$ случайной величины X.

2. По данным выборки найдем точечную оценку Θ^* неизвестного параметра Θ.

3. Задаем надежность (доверительную вероятность) γ, или уровень значимости α.

4. Определяем границы интервала $(\Theta^* - \delta;\ \Theta^* + \delta)$, используя плотность вероятности, из условия $P(|\Theta - \Theta^*| < \delta) = \int\limits_{\Theta^*-\delta}^{\Theta^*+\delta} f(x,\Theta)dx = \gamma = 1 - \alpha$, причем обычно выбирают $P(X(\Theta) < \Theta^* - \delta) = \frac{\alpha}{2}$; $P(X(\Theta) > \Theta^* + \delta) = \frac{\alpha}{2}$. Полученный интервал $(\Theta^* - \delta;\ \Theta^* + \delta)$ с доверитель-

ной вероятностью $1-\alpha$ *покрывает* неизвестный параметр Θ и является его *интервальной оценкой*. Так как наиболее часто встречается нормальное распределение $N(m;\ \sigma)$, то построим интервальные оценки для параметров нормального распределения.

3.5.2. Доверительные интервалы для оценки математического ожидания нормального распределения

В этом подразделе научимся определять доверительные интервалы для оценки математического ожидания нормального распределения при известном и неизвестном среднеквадратическом отклонении.

Среднеквадратическое отклонение известно. Пусть произведена выборка из генеральной совокупности, подчиненной закону нормального распределения $X\in N(m;\ \sigma)$. Это основное предположение математической статистики основано на центральной предельной теореме (см. подразд. 2.10.5). Пусть известно генеральное среднеквадратическое отклонение σ, но неизвестно математическое ожидание теоретического распределения m (среднее значение $\bar{x}_{\text{г}}$).

В таком случае среднее выборочное $\bar{x}_{\text{в}}$, полученное в ходе эксперимента (см. подразд. 3.4.2), также будет являться случайной величиной $\bar{x}_{\text{в}}\in N((m;\ \sigma/\sqrt{n}\ )$. Тогда «нормализованное» отклонение $\dfrac{x_{\text{в}}-m}{\sigma/\sqrt{n}}\in N(0;1)$ является стандартной нормальной случайной величиной.

Задача состоит в поиске интервальной оценки для m. Построим двусторонний доверительный интервал для m так, чтобы истинное математическое ожидание принадлежало ему с заданной вероятностью (надежностью) γ.

Установить такой интервал для величины $\dfrac{\bar{x}_{\text{в}}-m}{\sigma/\sqrt{n}}$ — значит найти максимальное $u_{\max}$ и минимальное $u_{\min}$ значения этой величины, $u_{\min}=-u_{\max}$ (стандартное нормальное распределение симметрично), которые являются границами критической области:

$$P\left(-u_{\max}<\frac{\bar{x}_{\text{в}}-m}{\sigma/\sqrt{n}}<u_{\max}\right)=\gamma.$$

Поскольку такая вероятность равна $2\Phi(u_{\max})=\gamma$, корень этого уравнения $u_{\max}(\gamma)$ можно найти с помощью таблиц функции Лапласа (см. табл. П. 3).

Тогда с вероятностью γ можно утверждать, что случайная величина $\dfrac{|\bar{x}_{\text{в}}-m|}{\sigma/\sqrt{n}}<u_{\max}(\gamma)$, т.е. искомое генеральное среднее принадлежит интервалу

$$\left(\bar{x}_{\text{в}} - u_{\max}(\gamma)\frac{\sigma}{\sqrt{n}};\ \bar{x}_{\text{в}} + u_{\max}(\gamma)\frac{\sigma}{\sqrt{n}}\right), \tag{3.9}$$

где величина

$$\delta = u_{\max}(\gamma)\frac{\sigma}{\sqrt{n}} \tag{3.10}$$

есть *точность* оценки.

Обратно, по заданному значению отклонения δ можно найти, с какой вероятностью неизвестное генеральное среднее принадлежит интервалу $|\bar{x}_{\text{в}} - m| < \delta$. Для этого нужно вычислить

$$P(|\bar{x}_{\text{в}} - m| < \delta) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) = \gamma. \tag{3.11}$$

Пусть из генеральной совокупности извлечена случайная выборка методом повторного отбора. Из уравнения $\delta = u\frac{\sigma}{\sqrt{n}}$ можно найти *минимальный* объем повторной выборки n, необходимый для того, чтобы доверительный интервал с заданной надежностью γ не превышал наперед заданного значения δ. Оценку требуемого объема выборки проводят по формуле

$$n = \frac{\sigma^2 u_\gamma^2}{\delta^2}. \tag{3.12}$$

Исследуем *точность оценки* $\delta = u_{\max}(\gamma)\frac{\sigma}{\sqrt{n}}$:

1) при возрастании объема выборки n и фиксированном γ величина δ *уменьшается* и, значит, точность оценки *увеличивается*;

2) с *увеличением* надежности оценки γ увеличивается значение аргумента u, так как $\Phi(u)$ монотонно возрастает и, значит, *увеличивается* δ. В таком случае увеличение надежности γ *уменьшает* точность ее оценки δ.

Далее символом $\bar{X}$ будем обозначать выборочное среднее.

Оценку
$$|\bar{X} - m| < t\frac{\sigma}{\sqrt{n}} \tag{3.13}$$

называют *классической* (где t — некий параметр, зависящий от γ и n), поскольку она характеризует наиболее часто встречающиеся законы распределения.

Среднеквадратическое отклонение неизвестно. Пусть известно, что генеральная совокупность подчинена закону нормального распределения $X \in N(m; \sigma)$, где m и σ неизвестны. Построим доверительный интервал, покрывающий с надежностью $\gamma = 1 - \alpha$ истинное значение параметра m.

Для построения доверительного интервала оценки генерального среднего в этом случае используется статистика $T = \frac{\bar{X} - m}{s/\sqrt{n}} = \frac{\bar{X} - m}{\sigma_в/\sqrt{n-1}}$, имеющая распределение Стьюдента с $k = n - 1$ степенями свободы. Это следует из того, что $\frac{\bar{X} - m}{\sigma/\sqrt{n}} \in N(0;1)$ (см. подразд. 3.5.2), а $n\sigma_в^2/\sigma^2 \in \chi^2(n-1)$ (см. подразд. 3.5.3) и из определения распределения Стьюдента (см. подразд. 2.9.2).

Найдем точность классической оценки, построенной для распределения Стьюдента. Пусть неравенство $\left|\frac{\bar{X} - m}{s/\sqrt{n}}\right| < t$ выполняется с надежностью γ:

$$P\left(\left|\frac{\bar{X} - m}{s/\sqrt{n}}\right| < t\right) = \gamma. \qquad (3.14)$$

Найдем параметр t из формулы (3.14). Поскольку $T \in \mathrm{St}(n-1)$, очевидно, что t зависит от γ и n, поэтому обычно пишут $t_{\gamma,n-1}$.

Тогда

$$P\left(\bar{X} - t_{\gamma,n-1}\frac{s}{\sqrt{n}} < m < \bar{X} + t_{\gamma,n-1}\frac{s}{\sqrt{n}}\right) = \gamma. \qquad (3.15)$$

Решив это уравнение относительно m, получим интервал $\left(\bar{X} - t_{\gamma,n-1}\frac{s}{\sqrt{n}}; \bar{X} + t_{\gamma,n-1}\frac{s}{\sqrt{n}}\right)$, который с надежностью γ покрывает неизвестный параметр m.

Величина $t_{\gamma,n-1}$, служащая для определения доверительного интервала случайной величины $T(n-1)$, распределенной согласно критерию Стьюдента с $n - 1$ степенями свободы, называется *коэффициентом Стьюдента*. Его следует находить по заданным значениям n и γ из таблиц «Критические точки распределения Стьюдента» (см. табл. П. 6), которые и представляют собой решения уравнения (3.15).

В итоге получаем следующее выражение *точности* δ доверительного интервала для оценки математического ожидания (генерального среднего), если неизвестна дисперсия:

$$\delta = t_{\gamma,n-1}\frac{s}{\sqrt{n}}. \qquad (3.16)$$

Итак, существует общая формула построения доверительных интервалов для математического ожидания генеральной совокупности:

$$\bar{X} - \delta < m < \bar{X} + \delta, \tag{3.17}$$

где точность доверительного интервала δ в зависимости от известной или неизвестной дисперсии находят по формулам соответственно (3.16) и (3.10).

Задача 3.8. Проведены некоторые испытания, результаты которых занесены в таблицу:

i	1	2	3	4	5
x_i	–25	34	–20	10	21

Известно, что они подчиняются закону нормального распределения с $\sigma = 2$. Найти оценку m^* для математического ожидания m, построить для него 90%-й доверительный интервал.

Решение. 1) Найдем $m^* = (-25 + 34 - 20 + 10 + 21)/5 = 4$ как среднее значений варианты x.

2) По доверительной вероятности определим аргумент функции Лапласа (см. табл. П. 3): $2\Phi(u_{0,9}) = 0,9 \Rightarrow u_{0,9} = 1,65$.

3) Тогда точность оценки равна $\delta = u_\gamma \sigma/\sqrt{n} = 1,65 \cdot 2/\sqrt{5} = 1,47$.

4) Доверительный интервал для m имеет вид $4 - 1,47 < m < 4 +$ $+ 1,47$, или $2,53 < m < 5,47$.

Итак, $m \in (2,53; 5,47)$.

Задача 3.9. Глубина моря измеряется прибором, систематическая ошибка которого равна 0, а случайные ошибки распределяются по нормальному закону со среднеквадратическим отклонением $\sigma = 15$ м. Сколько надо сделать независимых измерений, чтобы определить глубину с ошибками не более 5 м при доверительной вероятности 90 %?

Решение.

По условию задачи имеем $X \in N(m; \sigma)$, где $\sigma = 15$м, $\delta = 5$м, $\gamma = 0,9$. Определим объем n.

1) С заданной надежностью $\gamma = 0,9$ найдем по табл. П. 3 аргумент функции Лапласа $u_\gamma = 1,65$.

2) Зная заданную точность оценки $\delta = \dfrac{\sigma}{\sqrt{n}} = 5$, ищем $\sqrt{n}$. Имеем

$\sqrt{n} = \dfrac{u_\gamma \sigma}{\delta} = \dfrac{1,65 \cdot 15}{5} = 4,95$. Поэтому число испытаний $n \geq 25$.

Задача 3.10. В таблице заданы две выборки температуры t за первые шесть дней января 2004 и 2005 гг. Обозначим через x_1 и x_2 соответственно наблюдаемую температуру за эти годы:

№	1	2	3	4	5	6
x_1	–35	–32	–26	–35	–30	–17
x_2	–31	–27	–28	–35	–40	–31

Найти доверительный интервал для математического ожидания m генеральной совокупности с доверительной вероятностью $\gamma = 0{,}8$.

Решение. 1) Несмещенную оценку $\bar{X}$ найдем по формуле $\bar{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i$:

$$\bar{x}_1 = -29{,}2 \text{ и } \bar{x}_2 = -32.$$

2) Несмещенную оценку s^2 найдем по формуле $s^2 = \frac{1}{n-1} \times \sum_{i=1}^{n}(x_i - \bar{X})^2$:

x_i	–35	–32	–26	–35	–30	–17	$\sum x_i = -175$
$x_i - \bar{X}$	–5,8	–2,8	3,2	–5,8	–0,8	12,2	
$(x_i - \bar{X})^2$	33,64	7,84	10,24	33,64	0,64	148,84	$(x_i - X)^2 = 234{,}84$

$$s_1^2 = \frac{234{,}84}{5} = 46{,}968; \; s_1 = \sqrt{46{,}968} = 6{,}85;$$

x_i	–31	–27	–28	–35	–40	–31	$\sum x_i = -192$
$x_i - \bar{X}$	1	5	4	–3	–8	1	
$(x_i - \bar{X})^2$	1	25	16	9	64	1	$\sum(x_i - \bar{X})^2 = 116$

$$s_2^2 = \frac{116}{5} = 23{,}2; \; s_2 = \sqrt{23{,}2} = 4{,}8.$$

3) Поскольку генеральная дисперсия неизвестна, но известна ее оценка, для оценки математического ожидания m используем распределение Стьюдента (см. табл. П. 6) и формулу (3.16).

Так как $n_1 = n_2 = 6$, то $t_{\gamma,n_1-1} = t_{\gamma,n_2-1} = t_{08;5} = 1{,}48$, $\bar{x}_1 = -29{,}2$, $s_1 = 6{,}85$, имеем: $\delta_1 = t_{\gamma,n-1} \dfrac{s}{\sqrt{n}} = 1{,}48 \cdot \dfrac{6{,}85}{\sqrt{6}} = 4{,}1$, откуда $-29{,}2 - 4{,}1 < m_1 < -29{,}2 + 4{,}1$.

Поэтому $-33{,}3 < m_1 < -25{,}1$.

Аналогично имеем $\bar{x}_2 = -32$, $s_2 = 4{,}8$, $\delta_2 = 1{,}48 \cdot 4{,}8/\sqrt{6} = 2{,}9$, поэтому $-34{,}9 < m_2 < -29{,}1$. Тогда доверительные интервалы примут вид: $m_1 \in (-33{,}3;\ -25{,}1)$ и $m_2 \in (-34{,}9;\ -29{,}1)$. Сравнив результаты построения доверительных интервалов, можно убедиться, что границы интервала оценки некоторого параметра $\theta = m$ могут изменяться при переходе от одной выборки к другой. Однако с вероятностью 0,8 можно утверждать, что каждый из полученных интервалов *покроет истинное значение* $m_{\text{г}}$.

В прикладных науках, например в строительных дисциплинах, для оценки точности объектов используются таблицы доверительных интервалов, которые приведены в соответствующей справочной литературе.

3.5.3. Доверительный интервал для дисперсии и среднеквадратического отклонения

В этом подразделе вы научитесь определять границы доверительного интервала для дисперсии и среднеквадратического отклонения.

Пусть из некоторой генеральной совокупности значений X, распределенной по нормальному закону $N(m;\ \sigma)$, взята случайная выборка объемом n, для которой вычислены выборочные дисперсии: смещенная $D_{\text{в}}$ и несмещенная (исправленная) s^2. Найдем с заданной надежностью γ интервальные оценки для генеральной дисперсии D и среднеквадратического отклонения σ. В качестве оценки σ примем $s = \sqrt{s^2}$.

Заметим, что так как распределение Пирсона $(\chi)^2$ асимметрично, то правая и левая критические области могут быть взяты с произвольным соотношением вероятностей, поскольку определена только полная вероятность попадания в них: $P\left(\dfrac{(n-1)s^2}{\sigma^2} < a_{\min}\right) + P\left(\dfrac{(n-1)s^2}{\sigma^2} > a_{\max}\right) = \alpha$. Обычно берут обе критические области с одинаковой вероятностью $\alpha/2$, тогда границы доверительного интервала надо искать отдельно:

$$P\left(\frac{(n-1)s^2}{\sigma^2} < a_{\min}\right) = \alpha/2 = P\left(\frac{(n-1)s^2}{\sigma^2} > a_{\max}\right). \qquad (3.18)$$

Тем самым задача сведена к двум односторонним интервалам с половинным уровнем значимости α:

$$\sigma^2_{\min} = \frac{(n-1)s^2}{a_{\max}(\alpha/2, n-1)}; \quad \sigma^2_{\max} = \frac{(n-1)s^2}{a_{\min}(\alpha/2, n-1)}. \tag{3.19}$$

Двусторонний интервал для σ^2 имеет вид:

$$\sigma^2 \in \left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}}; \; \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \right).$$

Отсюда можно получить двусторонний интервал для среднеквадратического отклонения:

$$\frac{\sqrt{n}\sigma_{\text{в}}}{\chi_2} \le \sigma \le \frac{\sqrt{n}\sigma_{\text{в}}}{\chi_1} \quad \text{или} \quad s\sqrt{\frac{n-1}{\chi_2^2}} < \sigma < s\sqrt{\frac{n-1}{\chi_1^2}}, \tag{3.20}$$

где $\chi_1 = \chi_{1-\frac{\alpha}{2}, n-1}$; $\chi_2 = \chi_{\frac{\alpha}{2}, n-1}$.

Очевидно, что границы доверительного интервала не симметричны относительно исправленного значения дисперсии (см. рис. к табл. П. 8).

По формуле (3.18) можно решить также и обратную задачу — по заданному доверительному интервалу генеральной дисперсии определить доверительную вероятность γ.

Концы двустороннего доверительного интервала для дисперсии можно определить и без выполнения арифметических действий по заданному уровню доверия и объему выборки с помощью табл. П. 4. Для этого полученные из табл. П. 4 концы интервала умножают на исправленную дисперсию s^2.

Формулы (3.20) используют для вычисления доверительного интервала $\sigma_{\text{г}}$ при $n \le 30$. При выборках достаточно большого объема ($n > 30$) границы доверительного интервала для генерального среднеквадратического отклонения можно определить по формуле

$$\frac{s\sqrt{2n}}{\sqrt{2n-3} + u_\gamma} \le \sigma \le \frac{s\sqrt{2n}}{\sqrt{2n-3} - u_\gamma}, \tag{3.21}$$

где u_γ — нормированное значение нормальной случайной величины, соответствующее заданной надежности γ и определяемое по табл. П. 3.

Существует и другой способ определения границы доверительного интервала для дисперсии, в основе которого лежит выбор доверительного интервала, симметричного относительно σ^2. Доверительный интервал для среднеквадратического отклонения σ в таком случае можно найти, зная вероятность γ и число степеней свободы n, по формуле

$$s(1-q) < \sigma < s(1+q), \tag{3.22}$$

где $q = q(\gamma, n) = \frac{\delta}{s}$ — некоторое число, которое табулировано и приводится в справочной литературе по математической статистике (см. табл. П. 9). Причем, если $1 - q < 0$, то интервал имеет вид $0 < \sigma < s(1+q)$.

Задача 3.11. Было проверено качество 15 болтов. Предполагая, что ошибка при их изготовлении подчинена нормальному закону распределения, причем выборочное среднеквадратическое отклонение $\sigma_в$ равно 5 мм, определить с надежностью $\gamma = 0{,}95$ доверительный интервал для неизвестного параметра σ.

Решение. I способ. По условию задачи $n = 15 < 30$, поэтому воспользуемся формулами (3.20). Найдем пограничные значения вероятности для $\alpha = 1 - \gamma = 0{,}05$. Тогда имеем

$$P(\chi^2 > \chi_1^2) = 1 - \frac{\alpha}{2} = 1 - \frac{0{,}05}{2} = 0{,}975; \; P(\chi^2 < \chi_2^2) = \frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025.$$

По таблицам χ^2-распределения найдем пограничные значения χ_2^2 при заданной вероятности для $k = n - 1 = 14$ числа степеней свободы: $\chi_1^2 = 5{,}629$ и $\chi_2^2 = 26{,}119$, тогда $\chi_1 = \sqrt{5{,}629} = 2{,}37$ и $\chi_2 = \sqrt{26{,}119} = 5{,}11$. Откуда границы интервала представим в виде двойного неравенства:

$$\frac{\sqrt{15} \cdot 5}{5{,}11} \le \sigma \le \frac{\sqrt{15} \cdot 5}{2{,}37}, \text{ или } 3{,}78 \le \sigma \le 8{,}17 \text{ (мм)}.$$

II способ. Используем три решения табл. П. 4:

1) найдем исправленную дисперсию $s^2 = \frac{15}{14} \cdot 5^2 = 26{,}79$, отсюда $s = 5{,}17$;

2) по табл. П. 4 границы доверительного интервала для σ^2 при $k = 14$ и $\gamma = 0{,}95$; нижняя граница 0,513 и верхняя 2,354;

3) умножив полученные границы на s, найдем доверительный интервал для σ:

$$5{,}17\sqrt{0{,}513} < \sigma < 5{,}17\sqrt{2{,}354},$$

откуда $3{,}7 < \sigma < 7{,}93$.

III способ. Используем формулу (3.20):

1) найдем исправленное среднеквадратическое отклонение $s = \sqrt{s^2} = \sqrt{26,79} = 5,17$;

2) определим значение q из табл. П. 9 по надежности $\gamma = 0,95$ и числу степеней свободы $n = 15$: $q_{15;0,95} = 0,46$;

3) используя формулы (3.20), построим доверительный интервал для σ:

$$5,17(1 - 0,46) < \sigma < 5,17\ (1 + 0,46), \text{ или } 2,79 < \sigma < 7,55.$$

Как видно из вычислений, величина доверительного интервала зависит от способа его построения и дает близкие между собой, но неодинаковые результаты.

Задача 3.12. По данным Гидрометцентра, за последние 25 лет средняя температура в середине октября в нашем регионе имеет среднеквадратическое отклонение $\sigma_в = 8$ °С. Учитывая, что ошибка подчинена нормальному закону распределения, определить с надежностью $\gamma = 0,9$ доверительный интервал для неизвестного параметра σ.

Решение. Так как по условию задачи $n < 30$, то воспользуемся формулами

$$\frac{\sqrt{n}\sigma_в}{\chi_2} \le \sigma \le \frac{\sqrt{n}\sigma_в}{\chi_1}.$$

1) Найдем пограничные значения вероятности для $\alpha = 1 - \gamma = 0,1$:

$$P\left(\chi^2 > \chi^2_{1-\frac{\alpha}{2}, n-1}\right) = 1 - \frac{\alpha}{2} = 1 - \frac{0,1}{2} = 0,95;$$

$$P\left(\chi^2 < \chi^2_{\frac{\alpha}{2}, n-1}\right) = \frac{\alpha}{2} = \frac{0,1}{2} = 0,05.$$

2) По таблицам χ^2-распределения (см. табл. П. 8) определим пограничные значения χ^2 для $k = n - 1 = 24$ числа степеней свободы: зная, что $P\left(\chi^2 > \chi^2_{1-\frac{\alpha}{2}, n-1}\right) = 0,995$ и $P\left(\chi^2 < \chi^2_{\frac{\alpha}{2}, n-1}\right) = 0,05$, имеем $\chi^2_{1-\frac{\alpha}{2}, n-1} = 13,8$ и $\chi^2_{\frac{\alpha}{2}, n-1} = 36,415$.

3) Вычислим χ:

$$\chi_{1-\frac{\alpha}{2}, n-1} = \sqrt{9,886} = 3,714 \quad \text{и} \quad \chi_{\frac{\alpha}{2}, n-1} = \sqrt{36,415} = 6,03.$$

4) Найдем границы интервала $\frac{\sqrt{25}\cdot 8}{6,03} < \sigma < \frac{\sqrt{25}\cdot 8}{3,714}$ или $6,63 < \sigma <$ $< 10,78$.

3.5.4. Доверительный интервал для вероятности успеха в схеме Бернулли

Пусть проводятся независимые повторные испытания, в которых событие A наступает с неизвестной постоянной вероятностью p. Найдем с помощью выборочных испытаний для p точечную и интервальную оценки вероятности.

Точечная оценка вероятности. Согласно методу наибольшего правдоподобия, оценкой вероятности p события A в схеме Бернулли служит относительная частота этого события, т.е. $p^* = \frac{m}{n}$.

Пусть проведена серия из n испытаний, в каждом из которых событие A может произойти с вероятностью p или не произойти с вероятностью $q = 1 - p$. Из этих n испытаний в m случаях событие A произошло. Тогда $p^* = \frac{m}{n}$ — относительная частота, или статистическая вероятность, появления события A. Покажем, что эта относительная частота удовлетворяет всем требованиям, предъявляемым к точечной оценке (см. подразд. 3.4.1), т.е. является состоятельной, эффективной и несмещенной оценкой вероятности. В схеме Бернулли исход отдельного i-го испытания можно описать с помощью случайной величины X:

$$X = \begin{cases} 1, & \text{если событие } A \text{ произошло;} \\ 0, & \text{если событие } A \text{ не произошло.} \end{cases}$$

Тогда в выборке из n испытаний общее число испытаний, в которых событие A произошло, равно математическому ожиданию:

$$m = \sum_{i=1}^{n} X_i.$$

Оценка математического ожидания $M(X) = p$ этой случайной величины равна

$$p^* = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}m = \frac{m}{n}. \tag{3.23}$$

1) Относительная частота успеха является *несмещенной* оценкой, так как ее математическое ожидание равно оцениваемому

параметру. Поскольку A — случайное событие — количество m испытаний, в которых событие A произошло, относительная частота $p^* = \frac{m}{n}$ — тоже случайная величина. Испытания проводились по схеме Бернулли, поэтому $M_m = np$, а дисперсия $D_m = npq$.

Поэтому $M(p^*) = M(m/n) = M(m)/n = np/n = p$.

2) Оценка $\frac{m}{n}$ для p является *состоятельной*, так как для испытания Бернулли справедлива теорема Бернулли (см. подразд. 2.10.4), согласно которой для любого $\varepsilon > 0$ справедливо

$$\lim P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1.$$

3) Можно доказать, что дисперсия $D(p^*) = D(m/n) = \frac{1}{n^2} D(m) = \frac{npq}{n^2} = \frac{pq}{n}$ — минимально возможная дисперсия, т.е. оценка p^* для p является *эффективной*.

Тогда среднеквадратическое отклонение относительной частоты найдем по формуле

$$\sigma_{p^*} = \sqrt{D(P^*)} = \sqrt{\frac{pq}{n}}. \qquad (3.24)$$

Интервальная оценка вероятности. Согласно свойству статистической устойчивости относительной частоты в испытаниях Бернулли (см. подразд. 2.10.4), вероятность отклонения относительной частоты от вероятности в независимых повторных испытаниях при достаточно больших n можно найти с помощью функции Лапласа по формуле:

$$P\left\{\left|\frac{m}{n} - p\right| \le \delta\right\} \approx 2\Phi\left(\delta\sqrt{\frac{n}{pq}}\right) = 2\Phi(u_\gamma) = \gamma,$$

где аргумент функции Лапласа $u = \delta\sqrt{\frac{n}{pq}}$ зависит от **надежности** γ.

Тогда справедливо неравенство:

$$\frac{m}{n} - \delta \le p \le \frac{m}{n} + \delta. \qquad (3.25)$$

Заменив δ на $u_\gamma\sqrt{\frac{pq}{n}}$, q — на $1 - p$, а также учитывая, что точечная оценка неизвестной вероятности есть относительная частота

$\frac{m}{n} = p^*$, получаем, что с вероятностью γ выполняется неравенство:

$$p^* - u_\gamma\sqrt{\frac{p(1-p)}{n}} \le p \le p^* + u_\gamma\sqrt{\frac{p(1-p)}{n}}.$$

Итак, для построения доверительного интервала для вероятности p при поиске аргумента u_γ функции Лапласа необходимо воспользоваться таблицами значений интеграла Лапласа (см. табл. П. 3).

Границы интервала зависят от неизвестной величины p, однако при большом объеме выборки n неизвестное p можно заменить его эмпирическим значением p^*:

$$p^* - u_\gamma\sqrt{\frac{p^*(1-p^*)}{n}} \le p \le p^* + u_\gamma\sqrt{\frac{p^*(1-p^*)}{n}}. \qquad (3.26)$$

Формула (3.26) справедлива, когда $n\frac{m}{n}\left(1-\frac{m}{n}\right) > 9$, т.е. если $\frac{m(n-m)}{n} > 9$.

Полученная формула доверительного интервала позволяет решить еще одну задачу: установить объем выборки n, для которого с надежностью γ точность оценки p^*, полученной по ней для вероятности p, не превосходит заданного значения ε, т.е. $|p^* - p| < \varepsilon$.

Действительно, по формуле доверительного интервала с вероятностью γ выполняется неравенство $\left|p^* - p\right| < u_\gamma \frac{\sqrt{p^*(1-p^*)}}{\sqrt{n}}$.

Это означает, что результат тем точнее, чем больше объем выборки. Нужное значение объема n можно найти из уравнения $u_\gamma \frac{\sqrt{p^*(1-p^*)}}{\sqrt{n}} < \varepsilon$, т.е. $n > \frac{u_\gamma^2}{\varepsilon^2} p^*(1-p^*)$, причем «хороший» результат получается уже для $npq > 9$.

Замечание. Для бесповторной выборки (выборки без возвращения) из генеральной совокупности объемом N длина доверительного интервала с надежностью (уровнем доверия) γ может быть вычислена по формуле:

$$p^* - u_\gamma\sqrt{\frac{m/n(1-m/n)}{n}}\sqrt{1-\frac{n}{N}} \le p \le p^* + u_\gamma\sqrt{\frac{m/n(1-m/n)}{n}}\sqrt{1-\frac{n}{N}}.$$

Такую же поправку (уменьшение дисперсии в $(1 - n/N)$ раз) следует сделать и для случая бесповторной выборки в формулах

Таблица 3.2

Неизвестный параметр	Условия оценки	Вид используемого распределения	Границы интервала	Доверительный интервал
Математическое ожидание m	σ — известно	$\Phi(t)$ — функция Лапласа для нормального распределения	$m \pm \delta$, где $\delta = u_\gamma \dfrac{\sigma}{\sqrt{n}}$	$\bar{x} - \delta \le m \le \bar{X} + \delta$
	σ — неизвестно	$s(t)$ — распределение Стьюдента	$m \pm \delta$, где $\delta = t_{\gamma,n-1} \dfrac{s}{\sqrt{n-1}}$	
Дисперсия σ^2	$n \le 30$ m — известно	χ^2-распределение Пирсона	$\chi_1^2 = \chi^2_{1-\frac{\alpha}{2},n-1}$ $\chi_2^2 = \chi^2_{\frac{\alpha}{2},n-1}$ $\gamma = P(\chi_1^2) - P(\chi_2^2)$	$\dfrac{n\sigma_в^2}{\chi_2^2} \le \sigma^2 \le \dfrac{n\sigma_в^2}{\chi_1^2}$
	$n \le 30$ m — неизвестно	χ^2-распределение	$\chi_1^2 = \chi^2_{1-\frac{\alpha}{2},n-1}$ $\chi_2^2 = \chi^2_{\frac{\alpha}{2},n-1}$ $\gamma = P(\chi_1^2) - P(\chi_2^2)$	$\dfrac{ns^2}{\chi_2^2} \le \sigma^2 \le \dfrac{ns^2}{\chi_1^2}$

Окончание табл. 3.2

Неизвестный параметр	Условия оценки	Вид используемого распределения	Границы интервала	Доверительный интервал
Дисперсия σ^2	$n > 30$	$\Phi(u)$ — функция Лапласа	$\frac{\sqrt{2n}}{\sqrt{2n-3}+u_\gamma} s \le \sigma^2 \le s \frac{\sqrt{2n}}{\sqrt{2n-3}-u_\gamma}$	
Вероятность p	$\frac{m(n-m)}{n} > 9$	$\Phi(u)$ — функция Лапласа	$p \pm \delta$, где $\delta = u_\gamma \sqrt{\frac{\frac{m}{n}\left(1-\frac{m}{n}\right)}{n}}$	$\frac{m}{n} - \delta \le p \le \frac{m}{n} + \delta$

доверительных интервалов для среднего значения нормального распределения.

Задача 3.13. Выборочный статистический опрос 100 студентов показал, что 80 человек из них устраиваются на работу в процессе учебы в вузе. Найти интервальную оценку вероятности того, что случайно выбранный студент совмещает учебу в вузе и работу при условии, что полученный результат допускает ошибку не более чем в 5 % случаев.

Решение. В качестве точечной оценки возьмем частоту (или эмпирическую вероятность): $p^* = \dfrac{80}{100} = 0{,}8.$

Зная вероятность ошибки $\alpha = 0{,}05$ (или доверительную вероятность $\gamma = 1 - \alpha = 0{,}95$), по таблицам значений интеграла Лапласа (см. табл. П 3) найдем значение параметра $u_\gamma = 1{,}96$. Тогда двусторонний доверительный интервал вычисляем по формуле (3.26):

$$0{,}8 - 1{,}96\sqrt{\frac{0{,}8 \cdot 0{,}2}{100}} \le p \le 0{,}8 + 1{,}96\sqrt{\frac{0{,}8 \cdot 0{,}2}{100}}.$$

Итак, доверительный интервал для эмпирической вероятности в нашей задаче имеет вид: $0{,}72 \le p \le 0{,}88$.

Правила построения доверительных интервалов для неизвестных параметров распределения приведены в табл. 3.2, а также в табл. П. 3, П. 6 и П. 8.

3.6. Статистическая проверка статистических гипотез

Статистической проверкой статистической гипотезы называется процедура обоснованного сопоставления сформулированной гипотезы с полученными в ходе эксперимента выборочными данными $x_1, x_2, \ldots, x_n$.

Математика как наука дает возможность изучить некоторое явление, объект или систему с помощью определенной математической модели. Задача исследователя — на основе полученных результатов выдвинуть «гипотезу» (предположение) и проверить, насколько эта модель соответствует опытным данным.

В различных областях знаний — в экономике и медицине, технике и естествознании формулируют *статистические* гипотезы, которые затем проверяют статистическими методами. Современная наука весьма часто пользуется результатами *проверки* статистических гипотез: при прогнозе погоды и стихийных бедствий, анализе политической и экономической жизни, поиске полезных ископаемых и т.д.

3.6.1. Статистические гипотезы. Основные понятия

К *статистическим* будем относить гипотезы, возникающие в ходе некоторых исследований, которые можно проверить с помощью экспериментальных данных.

Типы статистических гипотез. Гипотезы имеют огромное значение во всех областях жизни, так как их главная задача — помочь выбрать правильное решение из двух альтернативных.

Статистическая гипотеза — это утверждение о виде неизвестного распределения или о параметрах известного распределения. Статистические гипотезы проверяются по результатам выборки статистическими методами в ходе эксперимента (эмпирическим путем) с помощью статистических *критериев*.

В тех случаях, когда известен закон, но не известны значения его параметров (дисперсия или математическое ожидание) в конкретной ситуации, статистическую гипотезу называют *параметрической*. Гипотеза о предполагаемой величине параметра этого распределения проверяется статистическими методами. Так, предположения об ожидаемом среднем доходе по акциям или разбросе дохода являются параметрическими гипотезами.

В других случаях, когда закон распределения генеральной совокупности не известен, но есть основания предположить, каков его конкретный вид, выдвигается статистическая гипотеза о виде распределения. Тогда, установив вид распределения, можно делать дальнейшие выводы и принимать решения. В этих случаях гипотезу называют *непараметрической*. Например, можно выдвинуть гипотезу о том, что число дневных продаж в магазине, доход населения или объем выпуска продукции на предприятии подчинены закону нормального распределения.

По содержанию статистических гипотез их можно классифицировать:

1. *Гипотезы о типе вероятностного закона распределения случайной величины*, характеризующего явление или процесс.

Некоторое свойство экономического характера имеет определенный закон распределения, зависящий от некоторых параметров. Проверка статистической гипотезы о законе распределения СВ может установить его с точностью до параметров, характеризующих неизвестный исследователю закон распределения.

2. *Гипотезы об однородности двух или более обрабатываемых выборок*, т.е. некоторых характеристик исследуемой совокупности (гипотезы о равенстве или различии законов распределения СВ, характеризующих изучаемое свойство).

Изучаемое свойство исследуется с помощью двух или более генеральных совокупностей, различающихся некоторыми факторами. Результатом статистического анализа статистической гипотезы такого типа может быть один из двух возможных выводов: исследуемые выборочные характеристики различаются между собой статистически значимо (незначимо), т. е. выборка взята из одной генеральной совокупности (различных генеральных совокупностей).

3. *Гипотезы о свойствах числовых значений параметров исследуемой генеральной совокупности.* С помощью гипотезы такого типа проверяются свойства некоторого числового параметра (среднего, дисперсии и т.д.) о том, что его значение не меньше (не больше) некоторого заданного значения — номинала или находится в заданных пределах.

4. *Гипотезы о вероятностной зависимости двух или более признаков* (*факторов*), характеризующих различные свойства рассматриваемого явления или процесса.

Два или более свойства рассматриваемого экономического процесса вероятностно зависимы. Определенные факторы оказывают влияние на изучаемый процесс и, значит, на его свойства. Эта стохастическая зависимость подчиняется некоторому общему закону. Задача исследователя заключается в определении характера этой функциональной зависимости (например, линейного) между компонентами этого исследуемого многомерного признака.

При сравнении эмпирической и теоретической функций распределения необходимо различать *простые* и *сложные* гипотезы о характере закона: гипотезу, содержащую одно (несколько) предположение(й), называют простой (сложной) гипотезой.

Выдвинутую гипотезу называют *основной*, или *нулевой*, и обозначают H_0. Противоречащую ей гипотезу H_1 называют *альтернативной*, или *конкурирующей*. Выбор альтернативной гипотезы определяется формулировкой решаемой задачи.

Пусть некоторый закон распределения случайной величины X зависит от некоторого параметра θ: (X, θ). Сформулирована некоторая основная гипотеза, например, о величине параметра θ, т.е. H_0: $\theta = \theta_0$, где θ_0 — конкретное значение параметра θ.

Тогда, в зависимости от условия задачи, альтернативная гипотеза H_1, противоположная суждению H_0, может иметь вид H_1: $\theta \neq \theta_0$ (*ненаправленная* гипотеза) или $\theta > \theta_0$, а также $\theta < \theta_0$ (*направленная* гипотеза).

Пример 3.1. При решении вопроса об инвестициях в одну из двух отраслей возникает проблема риска вложений. Предполагается, что распределения ежегодных прибылей на инвестиции подчиняются нормальному закону распределения. Исследуются ожидаемые дисперсии ежегодных прибылей от этих инвестиций. Если предположить, что они взяты из нормально распределенных генеральных совокупностей с равными дисперсиями, то нулевая гипотеза H_0: $\sigma_1^2 = \sigma_2^2$, а в качестве альтернативной может быть выбрана гипотеза H_1: $\sigma_1^2 \neq \sigma_2^2$, т.е. дисперсии различны. В то же время в качестве альтернативной может быть выбрана гипотеза H_1: $\sigma_1^2 > \sigma_2^2$, т.е. дисперсия первой отрасли превышает дисперсию второй.

Статистические критерии. Уровень значимости. Проверка статистической гипотезы осуществляется по данным выборки. Случайную величину K, служащую для проверки нулевой гипотезы, называют ***статистическим критерием***, или *критерием*. Статистический критерий дает возможность по результатам выборки принять либо отвергнуть основную гипотезу H_0.

В то же время под *статистическим критерием* понимают однозначно определенное правило, устанавливающее условие, при котором проверяемая гипотеза отвергается либо не отвергается.

Пример 3.2. Увеличение числа заболевших некоторым заболеванием дает возможность выдвинуть гипотезу о начале эпидемии. Для сравнения доли заболевших в обычных и экстремальных условиях используются статистические данные, на основании которых делается вывод о том, является ли данное массовое заболевание эпидемией. Предполагается, что существует некоторый критерий — *уровень доли заболевших*, критический для этого заболевания, который устанавливается по ранее имевшимся случаям.

Различают три вида критериев:

1. *Параметрические критерии* — критерии значимости, которые служат для проверки гипотез о параметрах распределения генеральной совокупности (например, о значениях m и σ при гипотезе о нормальном распределении).

2. *Критерии согласия* позволяют проверять гипотезы о соответствии распределений генеральной совокупности известной теоретической модели.

3. *Непараметрические критерии* используют в гипотезах, когда не требуется знаний о конкретном виде распределений.

Проверку *параметрических* гипотез проводят на основе *критериев значимости* с помощью табулированных статистик t, χ^2, F и других, *непараметрических* — на основе *критериев согласия*, используя статистики χ^2, Колмогорова—Смирнова и др.

Задача проверки статистических гипотез статистическими методами сводится к исследованию генеральной совокупности по выборке, содержащей n независимых случайных величин X_1, X_2, ..., X_n.

Возможные значения случайной величины X могут быть разделены на два непересекающихся подмножества — критическую область и область принятия гипотезы.

Областью принятия гипотезы, или *областью допустимых значений* $I_{\text{доп}}$, называют совокупность значений критерия, при которых эту гипотезу H_0 принимают.

Критической областью $I_{кр}$ для данного статистического критерия K называют множество значений критерия, при которых нулевую гипотезу H_0 отвергают.

Наблюдаемым значением критерия (*статистикой*) $K_{набл}$ называют такое значение критерия, которое находят по данным выборки.

Границы критической области, отделяющие ее от области принятия гипотезы, называют *критическими точками* и обозначают $k_{кр}$.

Для определения критической области задается уровень значимости α — некая малая вероятность попадания критерия K в критическую область. *Уровень значимости* — вероятность принять H_1, тогда как справедлива H_0, обычно берут $\alpha \in (0; 0,1)$. В соответствии с нашими обозначениями для условной вероятности имеем

$$\alpha = P(H \mid H_0). \tag{3.27}$$

Основной принцип проверки статистических гипотез состоит в следующем: если наблюдаемое значение статистики критерия ($K_{набл}$) попадает (не попадает) в критическую область, то гипотезу $H_0(H_1)$ отвергают (принимают), а гипотезу $H_1(H_0)$ принимают (отвергают) в качестве одного из возможных решений поставленной задачи с формулировкой «гипотеза H_0 противоречит (не противоречит) выборочным данным на уровне значимости α».

В зависимости от содержания альтернативной гипотезы H_1 осуществляется выбор критической области: *левосторонней*, *правосторонней* или *двусторонней*.

Если смысл исследования заключается в доказательстве конкретного изменения наблюдаемого параметра (его уменьшения или увеличения), то говорят об *односторонней* критической области.

Если смысл исследования — выявить различия в изучаемых параметрах, но характер их отклонений от контрольных (или теоретических) не известен, то говорят о *двусторонней* критической области и двусторонних критериях.

Выбор критерия осуществляется до начала эксперимента, но важно учесть, что более точные результаты дают односторонние критерии.

Границы критической области — значения критерия $k_{кр}$ — определяют с помощью уровня значимости α и предположения о характере распределения соответствующей статистики (табл. 3.3).

Однако принятие той или иной гипотезы не дает оснований утверждать, что она верна, так как один положительный результат не может служить основанием для того, чтобы считать некоторое утверждение достоверным. Так, в процессе сбора и обработки экспериментальных данных могли появиться ошибки по различным причинам, мог оказаться недостаточным объем эм-

Таблица 3.3

Критическая область $M_{кр}$	График	Определение критической области	Условие	Значение $k_{кр}$ через α
Левосторонняя	$k_{кр}$ $M(k)$ k	$K < k_{кр}$	$k_{кр} < 0$	$P(K < k_{кр}) = \alpha$
Правосторонняя	$M(k)$ $k_{кр}$ k	$K > k_{кр}$	$k_{кр} > 0$	$P(K > k_{кр}) = \alpha$
Двусторонняя симметричная	$k_{кр1}$ $M(k)$ $k_{кр2}$ k	$K < k_{кр}$, $K > k_{кр2}$ или $\lvert K \rvert > k_{кр}$	$k_{кр} < k_{кр2}$ $M(K) \in (k_{кр1};$ $k_{кр2})$	$P(K < k_{кр1}) =$ $= P(K > k_{кр2}) =$ $= \alpha/2$

пирических данных и т.д. Результаты проверки статистической гипотезы лишь устанавливают на определенном уровне значимости α ее *соответствие* (*несоответствие*) результатам эксперимента.

Алгоритм проверки статистических гипотез. Проверку статистической гипотезы можно осуществить по следующему алгоритму.

1. Сформулировать основную H_0 и альтернативную H_1 гипотезы в вероятностных терминах на основе выборочных данных и в зависимости цели исследования.

2. Выбрать соответствующий уровень значимости критерия; обычно $\alpha = \{0{,}001;\ 0{,}05;\ 0{,}01;\ 0{,}1\}$.

3. Определить (если он не задан) объем выборки n и число степеней свободы k (см. подразд. 2.9.1).

4. Вычислить статистическую характеристику критерия, т.е. наблюдаемую по выборке статистику $K_{набл}$, по формулам в зависимости от характера проверяемой гипотезы.

5. Найти по таблицам соответствующего критерия границу между областью принятия гипотезы и критической областью $I_{кр}$, которая зависит от объема выборки n (степень свободы k) и уровня значимости α.

6. Сформулировать правила проверки гипотезы: гипотеза H_0 не отвергается на уровне значимости α, если наблюдаемое значение статистического критерия данной выборки попадет в область принятия гипотезы. Если же наблюдаемое значение статистического критерия попадает в критическую для H_0 область, то принимается альтернативная гипотеза H_1, так как H_0 противоречит опытным данным.

Заметим, что вероятность принятия гипотезы H_0 основана на ***принципе практической невозможности наступления маловероятных событий*** (см. подразд. 2.10).

Ошибки первого и второго рода. Поскольку результатом исследования гипотезы служит управленческое решение, необходимо в ситуации выявленной неопределенности знать последствия возможных ошибок.

Возможны следующие ошибки:

• отвергнута правильная H_0, а принята неправильная гипотеза H_1 — ошибка первого рода;

• отвергнута правильная альтернативная гипотеза H_1 и принята неправильная нулевая H_0 — ошибка второго рода.

Заметим, что уровень значимости α — вероятность ошибки первого рода. Ошибки первого рода называют *α-риском*. Вероятности допустить такую ошибку соответствует так называемая *«ошибка поставщика»* (*«ложная тревога»*). Обычно (в таблицах для конкретных видов распределений) α задается некоторыми стандартными значениями: 0,05; 0,01; 0,005; 0,001.

Ошибки второго рода принято называть *β-риском*, а вероятность ее допустить обозначают β. Итак, β — вероятность того, что принята гипотеза H_0, если на самом деле справедлива альтернативная гипотеза H_1:

$$\beta = P(H_0 \mid H_1). \tag{3.28}$$

Мощностью критерия называется вероятность попадания критерия в критическую область при условии справедливости конкурирующей гипотезы. Очевидно, что она равна

$$M = P(H_1 \mid H_1) = 1 - \beta. \tag{3.29}$$

Анализ решений в задачах такого вида удобно проводить с помощью таблицы (табл. 3.4).

Таблица 3.4

Принятое решение	Истинное положение	
	H_1 — ложная H_0 — истинная	H_0 — ложная H_1 — истинная
H_0 — отвергнутое H_1 — принятое	$P(H_1 \mid H_0) = \alpha$ α — риск (ложная тревога) *Ошибка первого рода*	$P(H_1 \mid H_1) = 1 - \beta$ *Правильное решение*
H_1 — отвергнутое H_0 — принятое	$P(H_0 \mid H_0) = 1 - \alpha$ *Правильное решение*	$P(H_0 \mid H_1) = \beta$ β — риск (пропуск брака) *Ошибка второго рода*

? Что значит *маловероятные события*? Какой по величине должна быть вероятность, чтобы ее признали малой?

На этот вопрос нельзя ответить однозначно. Все зависит от конкретных решаемых задач.

При уровне значимости $\alpha = 0{,}005$ можно утверждать, что в 1000 выборок одного объема в среднем в пяти случаях допускается ошибка первого рода. Например, забракована партия изделий по результатам дефектов выборки, а во всей генеральной совокупности других бракованных изделий не было. Но если исследуемой системой является надежность технического устройства, например, тормозов автомобиля или реле холодильника, то вероятность отказа 0,001 может оказаться недопустимо высокой, так как если именно ваш единственный из тысячи холодильник или автомобиль выйдет из строя, то это повлечет за собой большие материальные расходы, а порой и неоправданные риски.

В нашем случае, если $\beta = 0{,}05$ — ошибка второго рода, то с вероятностью $p = 0{,}05$ можно утверждать, что в выборке оказалось мало бракованных изделий и всю партию пустили на реализацию. Причем процент дефектных изделий был настолько велик, что потребители и реализаторы пострадали из-за некачественной продукции, а производителям пришлось выплатить компенсацию за некачественный товар. Относительно недавно подобные скандалы пошатнули авторитет солидных автомобильных фирм.

Существуют формулы для расчетов статистик (наблюдаемых значений критерия) для различных гипотез, например, о среднем нормального распределения при неизвестном σ или о дисперсии нормального распределения и т.д., которые лежат в основе так называемой *теории оценок*.

Пусть K — граница критической области для гипотезы H_0 (рис. 3.5). Справа от K расположены ошибки первого рода — попадание в критическую область наблюдаемой статистики с вероятностью α, слева от K — область принятия гипотезы H_0 с вероятностью $1-\alpha$. В то же время эта граница разделяет область принятия альтернативной гипотезы H_1 с вероятностью $1-\beta$ (справа от K), тогда как слева от границы K остаются ошибки второго рода с вероятностью β, которые вычисляются по значениям m_0, m_1, σ и n. Из рисунка видно, что с уменьшением ошибок первого (второго) рода одновременно увеличиваются ошибки второго (перво-

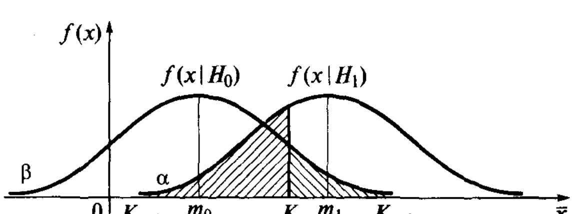


Рис. 3.5

го) рода. При заданном фиксированном α качество критерия для проверки гипотезы измеряется вероятностью отвергнуть H_0, когда справедлива H_1, т.е. мощностью критерия.

Понятно, что для лучшего (наиболее приближенного к действительности) результата нужно, чтобы мощность была более приближена к 1. Однако при заданном объеме выборки одновременно уменьшить вероятности ошибок первого и второго рода невозможно: единственный способ — увеличение выборки до масштабов, сравнимых со всей генеральной совокупностью — сопряжен с техническими и экономическими трудностями. Однако при принятии управленческих решений важно свести их до минимума. Поэтому на практике приходится «из двух зол». выбирать меньшее: пытаться подбирать значения параметров α и β опытным путем в целях минимизации суммарного эффекта от возможных ошибок.

Заметим, что при принятии управленческих решений для одновременного уменьшения ошибок первого и второго рода самым действенным средством является увеличение объема выборки, что согласуется с законом больших чисел (см. подразд. 2.10).

Поясним невозможность одновременного снижения α- и β-риска на примере пенсионной реформы.

Пример 3.3. Проведенные статистические исследования показали, что средний уровень жизни пенсионеров x_0 ниже «прожиточного минимума» $x_k = K$. Для повышения уровня доходов пенсионеров правительство проводит пенсионную реформу. Тогда если уровень жизни после проведения реформы m_1 будет выше «прожиточного минимума» ($m_1 \geq K$), то реформа однозначно признается успешной, если же $m_1 < K$ — неуспешной, а вложение бюджетных денег, затраченных на ее проведение, — неэффективным. Экономисты — проектировщики реформы — рассчитали предполагаемый средний уровень, построили теоретическое распределение доходов пенсионеров после проведения реформы.

Сформулируем основную гипотезу H_0: реформа оставила уровень жизни пенсионеров прежним и альтернативную гипотезу H_1: уровень жизни пенсионеров достиг прожиточного минимума. Соответственно условная плотность вероятности $f(x \mid H_0)$ — производная теоретической функции распределения $F(x \mid H_0)$, построенной по выборочной функции распределения первоначальных статистических данных — описывает распределение доходов при выполнении H_0, а условная плотность вероятности $f(x \mid H_1)$ — это прогнозируемый разработчиками реформы уровень доходов пенсионеров после проведения реформы (см. рис. 3.5). Тогда величина

$\alpha = P(H_1 \mid H_0) = \int\limits_{x_K}^{+\infty} f(t \mid H_0)dt$ — вероятность α-риска — равна площади заштрихованной части фигуры справа от K (I_α), а величина

$\beta = P(H_0 \mid H_1) = \int\limits_{-\infty}^{x_K} f(t \mid H_1)dt$ — вероятность β-риска — площади заштрихованной части фигуры слева от K (I_β).

Соответственно при сформулированных (фиксированных) оценках $f(x|H_0)$ и $f(x|H_1)$ увеличение критерия K приводит к увеличению вероятности β-риска, а уменьшение критерия K — к увеличению вероятности α-риска. Таким образом, приходится минимизировать возможные потери от совокупности всех рисков.

На бытовом уровне ошибки второго рода (ошибки потребителя) могут иметь более трагические последствия, чем ошибки первого рода. Говорят: ошибки первого рода — «ошибки поставщика» — ошибки осторожных людей, ошибки второго рода — пропуск брака — порой лихачество. Например, принимая решение о продаже партии автомобилей, в которой были замечены и не устранены дефекты, поставщик подвергает потребителя риску с вероятностью α, последствия которого могут привести к человеческим жертвам.

3.6.2. Гипотезы о законе распределения

Ранее предполагалось, что закон распределения либо является нормальным, либо близок к нему. Неизвестными же были отдельные параметры распределения: генеральное среднее и дисперсия. В тех случаях, когда закон распределения неизвестен, указать его вид помогает гистограмма. Например, если ее вид напоминает кривую Гаусса, то можно высказать гипотезу о нормальном законе распределения исследуемой случайной величины.

Критерий согласия Пирсона (χ^2-критерий). Рассмотрим простые гипотезы о степени близости эмпирического закона распределения теоретическому, т.е. соответствия опытных данных и теоретической модели. ***Критериями согласия*** называют критерии, в которых гипотеза определяет закон распределения либо полностью, либо с точностью до небольшого числа параметров.

Причины расхождения результатов эксперимента и теоретических характеристик могут быть вызваны малым объемом выборки, неудачным способом группировки наблюдений, неверными вычислениями теоретических частот, ошибками в выборе гипотезы о виде распределения генеральной совокупности и др.

Рассмотрим универсальный критерий согласия — *критерий Пирсона* χ^2 (для простой гипотезы), так как он используется и при независимых повторных испытаниях для выборки дискретного распределения и в более сложных случаях для произвольных выборок.

Проверка гипотезы H_0 о том, что эмпирическая частота мало отличается от соответствующей теоретической частоты, осуществляется с помощью величины χ^2 — меры расхождения между ними. Пусть p_i — вероятность принятия значения x_i; n_i — эмпирическая частота для соответствующих значений x_i. Тогда

$$\chi^2_{\text{набл}} = \sum_{i=1}^{s} \frac{(n_i - np_i)^2}{np_i}, \tag{3.30}$$

где n — объем выборки; s — число вариант выборки.

Числа n_i и np_i называют соответственно эмпирическими и теоретическими частотами, причем $\sum p_i = 1$.

Для произвольных выборок, когда распределение непрерывно или число различных вариант достаточно большое, все пространство наблюдаемых вариант делят на конечное число непересекающихся областей, в каждой из которых подсчитывают наблюдаемую частоту и теоретическую (гипотетическую) вероятность. В таком случае выборку группируют с учетом, что $n_i > 5$. Такая статистика имеет распределение $\chi^2(s - r - 1)$ с k степенями свободы, где s — число различных интервалов группировки выборки; r — число неизвестных параметров предполагаемого теоретического распределения, оцениваемых по данным выборки (независимых от вида закона распределения, которому подчинена генеральная совокупность).

Для применения критерия согласия Пирсона необходимо:

- вычислить значение статистики по формуле $\chi^2_{\text{набл}} = \sum_{i=1}^{s} \frac{(n_i - np_i)^2}{np_i}$;
- по табл. П. 8 распределения Пирсона найти критические значения $\chi^2_{\alpha,k}$, где $k = s - r - 1$ — число степеней свободы и α — выбранный уровень значимости. Это значит, что строится правосторонний интервал;
- если $\chi^2_{\text{набл}} > \chi^2_{\alpha,k}$, то гипотеза H_0 отвергается; если $\chi^2 \le \chi^2_{\alpha,k}$, то гипотеза H_0 принимается, т.е. чем больше отклонение, тем меньше согласованы теоретическое и эмпирическое распределения. Поэтому принято использовать только правостороннюю критическую область.

Расчетная таблица имеет вид:

Интервалы $[x_i; x_{i+1}]$	Середины i-го интервала x_i	Эмпирические частоты n_i	Вероятности p_i	Теоретические частоты np_i	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$

Задача 3.14. В известном опыте Бюффона (см. задачу 2.32, п. 2.10.4) определить, согласуются ли экспериментальные данные с теоретической вероятностью выпадения орла.

Решение. Данные эксперимента — это число опытов $n = 4040$, число выпадений орла $m = 2048$. Тогда эмпирическая частота $p_i = 0{,}507$.

Составим ряд распределений, приняв случайную величину — выпадение решки за $x_1 = 0$, а случайную величину — выпадение орла за $x_2 = 1$. Соответствующие вероятности занесем в таблицу:

x_i	n_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
0	1992	2020	–28	784	0,388
1	2048	2020	28	784	0,388

Тогда статистика $\chi^2_{набл} = 0{,}388 + 0{,}388 = 0{,}776$, а число степеней свободы найдем по формуле $k = s - r - 1$. Так как случайная величина X принимает всего два значения ($s = 2$), а неизвестных параметров нет, то $r = 0$. Поэтому число степеней свободы равно $k = 2 - 0 - 1 = 1$. По табл. П. 8 находим значение $\chi^2_{0,05;1} = 3{,}84$ при уровне значимости $\alpha = 0{,}05$. Это значит, что найденное значение $\chi^2_{набл} = 0{,}776$ меньше критерия $\chi^2_{0,05;1}$, и поэтому можно сделать вывод: гипотеза о равновероятном выпадении орла и решки в опытах Бюффона хорошо согласуется с теоретической вероятностью.

Задача 3.15. По таблице эмпирического распределения изменения (в %) темпа роста акций проверим гипотезу о нормальном распределении выборки:

Интервалы	(–3; –1)	(–1; 0)	(0; 1)	(1; 3)	Итого
m_i	7	14	18	11	50
P_i	0,14	0,28	0,36	0,22	1,0

Решение. Гипотезу о нормальном распределении проверим по критерию χ^2, используя расчетную таблицу эмпирического распределения для формулы $\chi^2_{набл} = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$:

Интервалы	n_i	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
(–3;–1)	7	0,157	7,85	–0,85	0,7225	0,092
(–1;0)	14	0,341	17,05	–3,05	9,3025	0,546
(0;1)	18	0,341	17,05	–0,95	0,9025	0,053
(1;3)	11	0,157	7,85	–3,15	9,9225	1,264

Итого: 1,955

$\chi^2_{набл} = 1{,}955.$

Находим $\chi^2_{\text{табл}}$ по числу степеней свободы $k = s - r - 1$. Поскольку $s = 4$ — число интервалов, $r = 2$ — число параметров теоретического распределения, то $k = 4 - 2 - 1 = 1$; для $\alpha = 0{,}05$ имеем $\chi^2_{\text{табл}}(0{,}05; 1) = 3{,}841$. Так как $1{,}955 < 3{,}841$, то $\chi^2_{\text{набл}} < \chi^2_{\text{табл}}$, т.е. гипотеза о нормальном распределении подтверждается.

3.6.3. Статистические гипотезы о числовом значении генерального среднего выборочного

В этом подразделе приведем сводную таблицу по определению границ критической области для различных статистических гипотез и примеры решения задач.

Установление двусторонней критической области на уровне значимости α для проверки гипотезы соответствует отысканию соответствующего доверительного интервала с надежностью $\gamma = 1 - \alpha$.

Если для некоторого параметра был найден доверительный интервал (в задачах на проверку гипотез) и основная гипотеза не проходит, то сразу можно указать эмпирическую оценку параметра. В качестве значения параметра альтернативной гипотезы может быть использовано эмпирическое значение параметра (например, вместо m — $\bar{x}$).

Действительно, если при проверке гипотезы H_0: $\bar{x} = m_0$ при H_1: $\bar{x} \neq m_0$ было с вероятностью α наложено требование о том, что статистический критерий попадет в двустороннюю критическую область, то с вероятностью $\gamma = 1 - \alpha$ можно утверждать, что он попадет в область принятия гипотезы. Тогда с надежностью $\gamma = 1 - \alpha$ выполняется неравенство $-u_{\text{кр}} < \dfrac{\bar{x} - m}{\frac{\sigma}{\sqrt{n}}} < u_{\text{кр}}$. Откуда с надежностью α получаем равносильное неравенство $\bar{x} - u_{\text{кр}}\dfrac{\sigma}{\sqrt{n}} < m < \bar{x} + u_{\text{кр}}\dfrac{\sigma}{\sqrt{n}}$, где $u_{\text{кр}}$ — корень уравнения $\Phi(u_{\text{кр}}) = \gamma/2$, что соответствует формуле (3.11) (см. подразд. 3.5.2).

Систематизируем условия применения различных статистических гипотез в табл. 3.5.

Задача 3.16. Результаты исследований в течение 35 лет показали, что среднее изменение доходности векселей равно 5,5 %. Полагая, что изменение доходности подчиняется нормальному закону распределения с $\sigma = 2\%$, на уровне значимости $\alpha = 0{,}05$, решить:

а) можно ли принять 6 % в качестве нормативного процента (математического ожидания) изменения доходности;

б) можно ли принять за норматив 6,5 %?

Таблица 3.5

№ п/п	Тип гипотезы H_0	Условия	Границы критической области на уровне значимости α	Статистика наблюдений
1.	О числовом значении генерального среднего $M(\bar{X}) = m_0$, или $m = m_0$	σ — известно для $N(m, \sigma)$	Функция Лапласа $\Phi(u_{\text{кр}}) = \alpha$	$U = \dfrac{\bar{x} - m_0}{\dfrac{\sigma}{\sqrt{n}}}$
		σ — неизвестно для $N(m, \sigma)$	Распределение Стьюдента с $k = n - 1$ степенями свободы	$T = \dfrac{\bar{x} - m_0}{\dfrac{s}{\sqrt{n-1}}}$
2.	О числовом значении дисперсии $M(s^2) = \sigma_0^2$, или $\sigma^2 = \sigma_0^2$	σ_0^2 — гипотетическое значение для $N(m, \sigma)$	Распределение Пирсона с $k = n - 1$ степенями свободы	$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$; $\chi^2 = \dfrac{nD_{\text{в}}}{\sigma_0^2}$
3.	Сравнение средних двух совокупностей $M(\bar{X}) = M(\bar{Y})$	$X \sim N(m_1, \sigma_1)$ $Y \sim N(m_2, \sigma_2)$ σ_1^2, σ_2^2 — известны	Функция Лапласа $\Phi(z)$ (Z — нормированная НСВ)	$Z = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{D(X)}{n} + \dfrac{D(Y)}{m}}}$
		$\sigma_1^2 = \sigma_2^2 = \sigma^2$ — неизвестны, малые независимые выборки объемами n и m	Распределение Стьюдента с $k = n_1 + n_2 - 2$ степенями свободы	$T = \dfrac{(\bar{X} - \bar{Y})\sqrt{\dfrac{nm(n+m-2)}{n+m}}}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}}$
4.	Сравнение относительной частоты с гипотетической вероятностью $p = p_0$	$p^* = \dfrac{m}{n}$ — относительная частота	Функция Лапласа $\Phi(u)$	$U = \dfrac{p^* - p_0}{\sqrt{p_0 q_0 / n}}$
5.	О законе распределения (критерий согласия гипотезы о теоретическом распределении с опытными данными)	r — число параметров теоретического распределения, вычисленных по выборке	Распределение χ^2 с $k = s - r - 1$ степенями свободы, где s — число интервалов группировки	$\chi^2 = \sum\limits_{i=1}^{s} \dfrac{(n_i - np_i)^2}{np_i}$

Решение. При проверке гипотезы о *числовом значении генерального среднего* $\bar{X}_г$ при *известной* дисперсии применяют статистику $U = \dfrac{\bar{X} - m_0}{\sigma/\sqrt{n}}$.

а) По условию задачи нулевая гипотеза H_0: $m_0 = 6\,\%$. Так как $\bar{x}_в = 5{,}5\,\%$, то в качестве альтернативной гипотезы возьмем гипотезу H_1: $m < 6\,\%$, которой соответствует левосторонняя критическая область с интервалом $(-\infty;\ k_{кр})$. По таблицам интеграла Лапласа (см. табл. П. 3) найдем аргумент для функции Лапласа из уравнения $\Phi(x) = \dfrac{1-2\alpha}{2}$. Зная, что $\Phi(x) = 0{,}45$, получим $u_{кр} = k_{кр} = -1{,}65$, т.е. левосторонняя критическая область лежит в интервале $(-\infty;\ -1{,}65)$.

По формуле критерия определяем $U_{набл} = \dfrac{\bar{x}_в - m_0}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{5{,}5-6}{\dfrac{2}{\sqrt{35}}} =$ $= -1{,}48$. Поскольку $-1{,}48 \notin (-\infty;\ -1{,}65)$, т.е. $U_{набл} > k_{кр}$, нет оснований отвергать нулевую гипотезу H_0. Значит, в качестве нормативного процента можно принять $6\,\%$.

б) По условию задачи в этом случае нулевая гипотеза H_0: $m_0 =$ $= 5\,\%$, а альтернативная $H_1: m < 6{,}5\,\%$. Тогда $U_{набл} = \dfrac{5{,}5-6{,}5}{\dfrac{2}{\sqrt{35}}} = -2{,}96$, причем $U_{набл}$ попадает в критическую область $-2{,}96 \in (-\infty;\ -1{,}65)$, где $U_{набл} < k_{кр}$. Итак, нулевую гипотезу H_0 отвергаем, а альтернативную принимаем, так как она не противоречит данным эксперимента. В этом случае с вероятностью $\alpha = 0{,}05$ можно совершить ошибку первого рода.

Задача 3.17. Анализ ежедневного объема продаж за I квартал текущего года показал, что для 42 торговых точек Автозаводского района средний ежедневный объем продаж составляет 64 тыс. р., а для 20 торговых точек Центрального района — 62 тыс. р., при среднеквадратических отклонениях, соответственно равных 4 и 5 тыс. р. Существует ли различие между ежедневным объемом продаж в Автозаводском и Центральном районах на 5%-м уровне значимости?

Решение. Имеем: $n = 42$; $m = 20$; $\bar{X} = 64$; $\sigma_x^2 = 16$; $\bar{Y} = 16$; $\sigma_y^2 = 25$.

Предположим, что ежедневный объем продаж подчинен нормальному закону распределения, для которого неизвестны математическое ожидание и среднеквадратическое отклонение. Проверяется гипотеза о сравнении средних двух генеральных совокупностей $M(\bar{X}) = M(\bar{Y})$. Тогда H_0: $M(X) = M(Y)$, при альтернативной гипотезе H_1: $M(X) \neq M(Y)$.

Вычисляем наблюдаемое значение статистики по формуле

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{n\sigma_x^2 + m\sigma_y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}};$$

$$T_{\text{набл}} = \frac{64-62}{\sqrt{42 \cdot 16 + 20 \cdot 25}} \sqrt{\frac{42 \cdot 20(42+20-2)}{42+20}} = 4{,}348.$$

Критическая область для проверки гипотезы $|T| > t_{N;\alpha}$ $N = n_x + n_y - 2 = 60$ при $\alpha = 0{,}05$.

Значение $t_{\text{кр}} = t_{60;0,05}$ находим по таблице критических значений распределения Стъюдента, α в верхней строке: $t_{60;0,05} = 2{,}00$. Поскольку $|4{,}348| > 2{,}00$, основная гипотеза H_0 отвергается и принимается альтернативная гипотеза H_1.

Таким образом, различие между ежедневным объемом продаж в Автозаводском и Центральном районах на 5%-м уровне значимости несущественно.

Задача 3.18. Точность работы программы проверяют по дисперсии контролируемого количества символов в коде, которая не должна превышать 0,1. По выборке из 15 сообщений вычислена исправленная оценка дисперсии 0,22. При уровне значимости 0,05 проверить, обеспечивает ли программа необходимую точность.

Решение. Имеем: $n = 15$; $s^2 = 0{,}22$; $\alpha = 0{,}05$; $\sigma^2 < 0{,}1$.

1) Сформулируем *гипотезу о числовом значении дисперсии*:

• H_0 — программа обеспечивает необходимую точность $\sigma^2 = 0{,}1$;

• H_1 — программа не обеспечивает необходимую точность $\sigma^2 > 0{,}1$.

2) Определим статистику χ^2: $\chi^2_{\text{набл}} = \dfrac{(n-1)s^2}{\sigma^2} = \dfrac{14 \cdot 0{,}22}{0{,}1} = 30{,}8.$

3) Найдем критическую точку — границы критической области: $\chi^2_{0,05;14} = 23{,}7$.

4) Поскольку $30{,}8 > 23{,}7$, $\chi^2 > \chi^2_{0,05;14}$.

Следовательно, принимаем гипотезу H_1, так как H_0 противоречит опытным данным.

Вывод: программа не обеспечивает необходимую точность.

3.7. Метод статистических испытаний. Метод Монте-Карло

Ранее рассматривались задачи, в которых удавалось аналитически выразить зависимость между исходными данными и неким конечным неизвестным результатом (т.е. *формализовать* задачу). На практике это не всегда возможно. В тех случаях, когда не удается найти формулы, выражающие зависимость между данными и условиями задачи, применяют *метод статистических испытаний*.

Под ***статистическими*** понимают испытания, предполагающие многократные повторения однотипных явлений. Результат отдельно взятого испытания случаен и сам по себе не представляет особого интереса для исследования. Однако результат большого числа испытаний обладает *статистической устойчивостью* (см. подразд. 2.9), характеризующей с количественной стороны исследуемое явление.

Смысл метода статистических испытаний заключается в том, что вместо аналитического решения некоторой задачи проводят соответствующие экспериментальные исследования либо сами испытания заменяют аналогичными, имеющими сходную вероятностную сущность и воспроизводящими физический процесс с помощью вероятностной математической модели. Отдельные такие воспроизведения называют *испытаниями* функционирующей системы.

Процесс отыскания возможных значений случайных чисел называется ***моделированием, или «разыгрыванием», случайной величины***. В основе метода статистического моделирования лежит закон больших чисел (см. подразд. 2.10), который обосновывает сходимость по вероятности средних значений результатов многократных испытаний. Согласно закону больших чисел, чем больше испытаний, тем точнее результат вычисления неизвестной величины.

Численный метод решения задачи с помощью моделирования случайных чисел называют также ***методом Монте-Карло*** (по имени города в Княжестве Монако, прославившемся в игорном бизнесе, так как именно в казино можно встретить рулетку — простейший прибор для генерирования случайных чисел).

Официальной датой рождения метода Монте-Карло принято считать 1949 г. — год появления статьи Н. Метрополиса и С. Улама «Метод Монте-Карло». Но теоретические основы метода были известны и ранее. В 40-е годы XX в. этот метод и его приложения уже использовали Дж. Нейман, Г. Канн и др.

Методом Монте-Карло:

- решают большой круг любых математических и статических задач как общего, так и вероятностного характера;
- решают системы алгебраических уравнений высшего порядка;
- исследуют различные сложные системы, в том числе многомерные;
- моделируют естественные процессы вероятностного характера путем имитации, свободного пробега, используя статистические весы.

Метод Монте-Карло полезен и эффективен для вероятностных расчетов в различных прикладных задачах, которые трудно или невозможно решить аналитически. Метод статистических испытаний универсален, так как не ограничен рамками сложных формул и условий. Поэтому его используют для проверки степени точности построенных аналитических моделей, применяемых в конкретных ситуациях в различных предметных областях: теории массового обслуживания, теории игр, статистической физике, экономике и т.д.

Метод Монте-Карло позволяет ответить на два основных вопроса:

1) как выбрать наиболее удобную величину X для расчетов в конкретной задаче;

2) как находить значения x_1, x_2, ... любой произвольно полученной случайной величины X?

В большинстве случаев метод Монте-Карло заключается в расчете математического ожидания (среднего) для случайной величины X с помощью методов случайного поиска и стохастических приближений.

В процессе решения задач методом Монте-Карло необходимо:

- разработать и построить структурную схему процесса, выявив основные взаимосвязи;
- формально описать процесс;
- смоделировать случайное явление (случайное событие, случайную величину), характеризующее функционирование изучаемой системы;
- смоделировать функционирование изучаемой системы, т.е. воспроизведение процесса, согласно построенной структурной схеме и формальному описанию;
- собрать результаты многократного моделирования для исследования методами математической статистики, проанализировать и обобщить результаты.

Поскольку статистические методы при изучении вероятностных систем сопровождаются ошибками эксперимента, полученные выводы подвергаются статистической проверке.

Итак, изучаемые вероятностные системы характеризуются некоторыми случайными явлениями, которые описываются случайными событиями, случайными величинами и случайными функциями. Моделирование случайных событий и случайных функций осуществляется с помощью случайных величин.

3.7.1. Моделирование случайных величин

Моделирование (разыгрывание) реальных процессов происходит с помощью некоторым образом полученных *случайных чисел.*

Для моделирования случайных величин необходимо знать законы их распределения. Обычно в качестве стандарта выбирают равномерно распределенную НСВ на интервале [0; 1], а затем подбирают такую функцию от этой случайной величины, которая распределена по требуемому закону. Существует три способа получения последовательности чисел, равномерно распределенной на интервале [0; 1]:

1) с помощью генератора случайных чисел;
2) с помощью таблицы случайных чисел;
3) методом псевдослучайных чисел.

Таким примитивным *генератором случайных чисел* могут являться также обычные монеты или игральная кость, лототрон для игры «Спортлото» или рулетка в казино.

Физическим генератором случайных чисел, или *датчиком* случайных чисел, называют специальное приспособление на ЭВМ, в котором результаты некоторого физического процесса случайным образом преобразуются в двоичные числа. В качестве случайного физического процесса могут использоваться собственные шумы, например меняющееся случайным образом напряжение или радиоактивный распад. К недостаткам этого способа получения случайных чисел относят трудности, связанные:

- с проверкой качества вырабатываемых сигналов;
- с невоспроизводимостью, а значит, и невозможностью их повторения при случайном сбое.

В тех случаях, когда используются *таблицы случайных чисел* (ТСЧ), отдельные цифры имитируют значения ДСВ с равномерным распределением. Подразумевается, что появление каждой из десяти цифр в таблице случайных чисел равновероятно. Наиболее мощная из известных таблиц случайных чисел включает миллион цифр. Однако составление таких таблиц сопряжено со значительными трудностями.

Обозначим через R непрерывную случайную величину, равномерно распределенную на интервале (0,1). Возможные значения r НСВ R, равномерно распределенной на интервале (0,1), называют *случайными числами*.

Так как каждая цифра из известных десяти распределена с одинаковой вероятностью, ряд *распределений принимает вид*:

R	0	1	...	9
p	0,1	0,1		0,1

Если случайная величина X задана двоичной случайной цифрой, то ее распределение имеет вид:

R	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$

Компьютерное моделирование случайных процессов не обходится без набора случайных чисел, которые *табулируются*. ЭВМ в соответствии с программой создает таблицу случайных чисел, а затем случайным образом выбирает отдельные значения (говорят: «*проигрывает* статистические испытания»).

ЭВМ дает возможность также генерировать значения равновероятно распределенной случайной величины по специально разработанным рекуррентным формулам и получать так называемые *псевдослучайные* числа, которые применяют вместо случайных чисел. Один из методов получения псевдослучайных чисел, разработанный Дж. Нейманом, — «*Метод срединных квадратов*», согласно которому для получения псевдослучайных чисел берут любое число, например $r_0 = 0{,}8375$, и возводят его в квадрат: $r_0^2 = 0{,}70\underline{1406}25$. Из полученного квадрата выбирают четыре средние цифры (представляя их как десятичную правильную дробь) $r_1 = 0{,}1406$, которые также возводят в квадрат $r_1^2 = 0{,}01\underline{1976}83$, затем вновь выбирают четыре средние цифры $r_2 = 0{,}1976$ и также возводят их в квадрат $r_2^2 = 0{,}95\underline{4138}24$. В таком случае последовательность $r_1, r_2, \ldots$ — последовательность средних квадратов, как и случайные числа, получена случайным образом. Несмотря на то что этот алгоритм получения ТСЧ не дает желаемого равномерного распределения, описанный метод обладает рядом достоинств, среди которых, в частности, возможность их воспроизведения. Так как в действительности при использовании некоторого алгоритма образования ТСЧ нельзя говорить о равномерном распределении НСВ R, ее заменяют квазиравномерной случайной величиной R^*.

На практике при решении задач методом Монте-Карло производят n испытаний для получения i возможных значений X, затем находят их среднее (арифметическое) $\bar{X} = \frac{1}{n}\sum X_i$, которое принимают в качестве приближенного значения a^* некоторого числа a: $a \approx a^* = \bar{x}$. Очевидно, имеется множество таких значений X, для которых $a^* = \bar{x}$.

Существует основное соотношение, связывающее случайные числа с заданным законом распределения и случайные числа с равномерным законом распределения на интервале $[0; 1]$. Идея заключается в том, что для преобразования последовательности случайных чисел с равномерным законом распределения на интервале $[0; 1]$ в последовательность случайных чисел с заданным законом распределения $F(x)$ из множества случайных чисел с равномерным законом распределения на интервале $[0;1]$ надо выбрать случайное число a и решить относительно x уравнение:

$$F(x) = a.$$

Найденное решение является случайным числом из множества случайных чисел, имеющих функцию распределения $F(x)$.

В тех случаях, когда НСВ задана не функцией распределения $F(x)$, а плотностью вероятности $f(x)$, необходимо решить относительно x уравнение

$$\int_{-\infty}^{x} f(t)dt = a.$$

Оценка погрешности метода Монте-Карло

В ходе статистических испытаний важно оценить точность смоделированного процесса, т.е. найти величину ошибки.

Искомая верхняя граница δ есть *точность оценки* математического ожидания по выборочной средней, которая участвует в построении доверительных интервалов (см. подразд. 3.5).

Возможны различные случаи.

1. *Случайная величина X распределена нормально $X \in N(m; \sigma)$ с известным среднеквадратическим отклонением σ.*

Это означает, что с надежностью γ верхняя граница ошибки равна

$$\delta = t\frac{\sigma}{\sqrt{n}}, \qquad (3.31)$$

где n — число испытаний, т.е. разыгранных значений X; t — значение аргумента функции Лапласа, при котором

$$\Phi(t) = \frac{\gamma}{2}. \qquad (3.32)$$

Задача 3.19. С надежностью $\gamma = 0,95$ найти верхнюю границу ошибки для оценки математического ожидания случайной величины $X \in N(m; 0,6)$, если было разыграно 25 возможных значений n.

Решение. Имеем $\gamma = 0,95$; $y = 0,6$; $n = 25$, тогда

1) $\Phi(t) = \frac{\gamma}{2} = \frac{0,95}{2} = 0,475.$

2) Из табл. П. 3 по значению функции Лапласа, равному 0,475, найдем аргумент $t = 1,96$.

3) По формуле (3.31) найдем верхнюю границу ошибки δ:

$$\delta = \frac{1,96 \cdot 0,6}{\sqrt{25}} = 0,2352.$$

2. *Случайная величина X распределена* ***нормально, но ее средне-*** *квадратическое отклонение σ неизвестно.*

В таком случае для оценки математического ожидания принимают значение исправленного среднеквадратического отклонения как несмещенную оценку:

$$\delta = t_\gamma \frac{S}{\sqrt{n}}, \qquad (3.33)$$

где t_γ — находят с учетом γ как аргумент в распределении Стьюдента из табл. П. 6.

Задача 3.20. Найти верхнюю границу ошибки с надежностью $\gamma = 0{,}95$, если для оценки значения математического ожидания было разыграно 100 возможных значений X и по этим данным найдено «исправленное» среднеквадратическое отклонение $S = 0{,}4$.

Решение. Из условия задачи имеем $\gamma = 0{,}95$; $n = 100$; $S = 0{,}4$.

По значению $\gamma = 0{,}95$ найдем $t_\gamma = 1{,}984$; по формуле $\delta = \frac{t_\gamma S}{\sqrt{n}}$ определим верхнюю границу $\delta = 0{,}07936$.

3. *Случайная величина X распределена по закону, отличному от нормального.*

Как известно, при большом числе испытаний ($n > 30$), согласно закону больших чисел, любое распределение стремится к нормальному. В таком случае задача поиска верхней границы ошибки сводится к задаче 3.19 (по известному σ) или к задаче 3.20 (по неизвестному σ), где требуется найти оценку математического ожидания δ.

Причем, чем больше n ($n \to \infty$), тем меньше различие между результатами применения формул (распределение Стьюдента стремится к нормальному).

Оказывается, что разыгрывание событий можно свести к разыгрыванию ДСВ.

3.7.2. Случайные числа. Разыгрывание дискретных и непрерывных случайных величин

В этом подразделе вы научитесь разыгрывать полную группу событий, т.е. разыгрывать испытания, в каждом из которых наступает одно из событий полной группы, вероятности которых известны. Разыгрывание полной группы событий сводится к разыгрыванию ДСВ. При разыгрывании НСВ применяется метод обратных функций.

Разыгрывание дискретных случайных величин. Пусть R — ДСВ, равномерно распределенная в интервале (0; 1).

Как уже отмечалось, на практике вместо равномерно распределенной случайной величины R, имеющей бесконечное число десятичных знаков, используют *квазиравномерную* случайную величину R^*, возможные значения которой имеют конечное число знаков.

Поэтому разыгрываемая величина имеет не точное, а приближенно заданное распределение.

Правило 1. Для того чтобы разыграть ДСВ, заданную законом распределения

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

необходимо:

1) разбить интервал $(0; 1)$ на n частичных интервалов: $\Delta_1 = (0; p_1)$, $\Delta_2 = (p_1; p_1 + p_2)$, ..., $\Delta_n = (p_1 + ... + p_{n-1}; 1)$;

2) выбрать случайное число r_j (например, из ТСЧ).

Если $r_j \in \Delta_j$, то разыгрываемая ДСВ приняла возможное значение x_j.

Разыгрывание полной группы событий

Правило 2. При разыгрывании полной группы несовместных событий $A_1, A_2, ..., A_n$ с вероятностью $p_1, p_2, ..., p_n$ можно ограничиться разыгрыванием ДСВ X с законом распределения

X	1	2	3	...	n
P	p_1	p_2	p_3	...	p_n

по известному правилу.

Если в испытании величина X приняла возможное значение $X_i = i$, то наступило событие A.

Это правило справедливо, поскольку число n возможных значений X совпадает с числом возможных событий полной группы событий и вероятности возможных значений x_i и соответствующим им событий A_i равны между собой: $P(X = x_i) = P(A_i) = p_i$.

Задача 3.21. По заданным вероятностям четырех событий, образующих полную группу событий, $p_1 = P(A_1) = 0{,}18$; $p_2 = P(A_2) = 0{,}25$; $p_3 = P(A_3) = 0{,}36$; $p_4 = P(A_4) = 0{,}21$, разыграть пять испытаний, в каждом из которых должно появиться одно из этих четырех событий.

Решение. 1) По правилу разыгрывания полной группы событий надо разыграть ДСВ, имеющую закон распределения

X	1	2	3	4
P	0,18	0,25	0,36	0,21

2) По правилу разыгрывания ДСВ разобьем интервал $(0; 1)$ на четыре частных интервала: $\Delta_1 = (0; 0{,}18)$; $\Delta_2 = (0{,}18; 0{,}43)$; $\Delta_3 = (0{,}43; 0{,}79)$; $\Delta_4 = (0{,}79; 1)$.

3) Выберем из ТСЧ пять случайных чисел: 0,56; 0,82; 0,78; 0,75; 0,86.

4) Так как число $0{,}56 \in \Delta_3$, наступило событие A_3, числу $0{,}82 \in \Delta_4$ соответствует событие A_4, аналогично, числам $\{0{,}78;\ 0{,}75\} \in \Delta_3$ соответствует событие A_3, а числу $0{,}86 \in \Delta_4$ — событие A_4.

Поэтому по заданному условию установим последовательность событий A_3, A_4, A_3, A_3, A_4.

Задача 3.22. События A и B независимы и совместны. Разыграть пять испытаний, в каждом из которых $P(A) = 0{,}7$; $P(B) = 0{,}1$.

Решение. Из условия следует, что вероятности противоположных событий $P(\bar{A}) = 0{,}3$; $P(\bar{B}) = 0{,}9$. События A и B дают четыре варианта исходов:

$M_1 = AB$, так как A и B независимы, $P(AB) = 0{,}7 \cdot 0{,}1 = 0{,}07$;

$M_2 = A\bar{B}$, поэтому $P(A\bar{B}) = 0{,}7 \cdot 0{,}9 = 0{,}63$;

$M_3 = \bar{A}\bar{B}$, следовательно, $P(\bar{A}\bar{B}) = 0{,}3 \cdot 0{,}9 = 0{,}27$;

$M_4 = \bar{A}B$, аналогично $P(\bar{A}B) = 0{,}3 \cdot 0{,}1 = 0{,}03$.

Итак, разыгрывается полная группа событий: M_1 — с вероятностью $p_1 = 0{,}07$; M_2 — с вероятностью $p_2 = 0{,}63$; M_3 — с вероятностью $p_3 = 0{,}27$ и M_4 — с вероятностью $p_4 = 0{,}03$.

Построим частичные интервалы: $\Delta_1 = (0;0{,}07)$; $\Delta_2 = (0{,}07;0{,}7)$; $\Delta_3 = (0{,}7;0{,}97)$; $\Delta_4 = (0{,}97;1)$.

Задача свелась к разыгрыванию пяти случайных чисел и установлению их принадлежности конкретному частичному интервалу. Пусть r_i — числа, взятые случайным образом из ТСЧ: например, $r_1 = 0{,}23$; $r_2 = 0{,}018$; $r_3 = 0{,}93$; $r_4 = 0{,}64$; $r_5 = 0{,}48$.

Случайное число $r_1 = 0{,}23 \in \Delta_2$, значит, наступило событие $M_2 = A\bar{B}$;

случайное число $r_2 = 0{,}018 \in \Delta_1$, значит, наступило событие $M_1 = AB$;

случайное число $r_3 = 0{,}93 \in \Delta_3$, значит, наступило событие $M_3 = \bar{A}\bar{B}$;

случайному числу $r_4 = 0{,}67$ соответствует событие $M_4 = \bar{A}B$;

случайное число $r_5 = 0{,}48$, следовательно, $M_2 = A\bar{B}$.

Поэтому последовательность исходов, соответствующая условию этой задачи, имеет вид $A\bar{B}$, AB, $\bar{A}\bar{B}$, $\bar{A}B$, $A\bar{B}$.

Задача 3.23 (*случайный выбор одной из альтернатив, каждая из которых имеет свой вес*). Ряд практических задач связан с выбором одного из возможных вариантов, каждый из которых имеет определенный вес. Подобные задачи можно решить описанным методом (см. задачу 3.22), заменив вероятность на величину соответствующего веса по формуле $p^*(x_i) = \dfrac{q(x_i)}{\sum_i q(x_i)}$.

Идея решения задачи основана на попадании случайной точки в один из интервалов, каждый из которых пропорционален величине $p^*(x_i)$.

Пусть, например, имеются четыре альтернативы x_1, x_2, x_3, x_4, вес каждой из которых $q(x_1) = 9$, $q(x_2) = 12$, $q(x_3) = 7$, $q(x_4) = 5$. Выбрать альтернативу при выпадении случайного числа $r = 0{,}53$.

Решение. Рассмотрим полную группу событий, состоящую из четырех событий, вероятность заменим относительным весом соответствующей альтернативы по формуле $p^*(x_i) = \dfrac{q(x_i)}{\sum_{i=1}^{4} q(x_i)}$.

1) По правилу разыгрывания полной группы событий надо разыграть ДСВ, имеющую закон распределения:

i	1	2	3	4
$p^*(x_i)$	9/33	12/33	7/33	5/33

2) По правилу разыгрывания ДСВ разобьем интервал (0; 1) на четыре частных интервала: $\Delta_1 = (0; 0{,}27)$; $\Delta_2 = (0{,}27; 0{,}63)$; $\Delta_3 = (0{,}63; 0{,}84)$; $\Delta_4 = (0{,}84; 1)$.

3) Так как случайное число $r = 0{,}53 \in \Delta_2 = (0{,}27; 0{,}63)$, выбирается вторая альтернатива.

Разыгрывание непрерывных случайных величин (*метод обратных функций*). Пусть необходимо разыграть НСВ X. Это значит, что надо получить последовательность ее возможных значений x_i ($i = 1, 2, \ldots$) по функции распределения $F(x)$ или по плотности вероятности $f(x)$.

Правило 1. *Если r_1 — случайное число, то возможное значение x_i разыгрываемой НСВ X с заданной функцией распределения $F(x)$, соответствующее r_i, является корнем уравнения $F(x) = r_i$.*

Это означает, что вероятность попадания возможного значения X в интервал (c, d) равна приращению функции распределения $F(x)$ в этом интервале.

Правило 2. *Если известна плотность вероятности $f(x)$, то надо выбрать случайное число r_i и решить относительно x_i уравнение: $\int_{-\infty}^{x_i} f(x)\,dx = r_i$, или уравнение $\int_{a}^{x_i} f(x)\,dx = a$, где a — наименьшее конечное значение X.*

Задача 3.24 (*вычисление определенного интеграла*). Пусть необходимо вычислить определенный интеграл $\int_0^1 f(x)\,dx$.

Решение. Наложим условие $0 \le f(x) \le 1$, $\forall x \in [0; 1]$. Как известно, величина определенного интеграла равна площади графика, ограниченного кривой $y = f(x)$, осью OX и прямыми $x = 0$ и $x = 1$ (рис. 3.6).

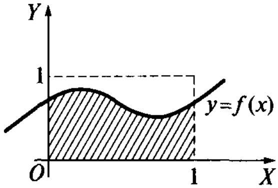


Рис. 3.6

Если провести статистические испытания количества точек, попавших в область N_0, и сравнить с количеством точек всего квадрата N, то геометрическая вероятность события A — попадания некоторой точки (x, y) в область подграфика — может быть вычислена по формуле $P(A) = \dfrac{\operatorname{mes} g}{\operatorname{mes} G}$ и примерно равна статистической вероятности или частоте $P(A) \approx \dfrac{N_0}{N}$.

Так как в данном случае мерой областей являются площади соответствующих фигур, то $\operatorname{mes} G = S_G = 1$, $\operatorname{mes} g = S_g = \int\limits_0^1 f(x)\,dx$,

$$P(A) = \frac{\operatorname{mes} g}{\operatorname{mes} G} = \frac{S_g}{S_G} = \int\limits_0^1 f(x)\,dx = \frac{N_0}{N}. \tag{3.34}$$

Поэтому через случайно выбранное число $\dfrac{N_0}{N}$ можно найти частоту и приближенное значение определенного интеграла $\int\limits_0^1 f(x)\,dx$. Вообще говоря, вероятность попадания точки в область g не зависит от расположения квадрата и его размеров.

Задача 3.25. Вычислить значение числа π методом статистических испытаний.

Решение. Как известно, числом π называют отношение длины окружности $C = 2\pi r$ к диаметру $d = 2r$.

В процессе решения этой задачи можно проследить механизм реализации метода статистических испытаний. Разыграем значения равномерного распределения, каждое из которых с одинаковой вероятностью может находиться в любой точке квадрата, в который вписана четверть круга радиусом, равным стороне квадрата.

Обозначим через A событие «попадание в круг некоторой точки с координатами (x, y)».

Возьмем круг с центром в начале координат и радиусом $r = 1$. Тогда, согласно геометрическому определению вероятности события A, имеем (см. подразд. 1.4.2.):

$P(A) = \dfrac{S_{\text{кр}}}{S_{\text{кв}}} = \dfrac{\operatorname{mes} g}{\operatorname{mes} G}$, где g — площадь круга ($S_{\text{кр}}$); G — площадь квадрата ($S_{\text{кв}}$).

Тогда если $S_{кв} = 1$, а координаты точек находятся на интервале $[0; 1]$, то $\frac{1}{4}S_{кр} = \frac{1}{4}\pi r^2 = \frac{\pi}{4}$ при $r = 1$ (рис. 3.7).

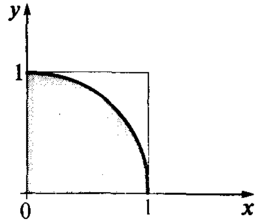


Рис. 3.7

Пусть разыгрывается всего N точек, среди которых N_0 попадают в четверть круга. Тогда за отношение площадей фигур можно принять отношение числа точек, попавших в четверть круга, к общему числу разыгранных точек. Поэтому $\frac{S_{кр}}{S_{кв}} = \frac{\pi}{4} = P(A)$, т.е. $\pi \approx 4P(A) = 4\frac{N_0}{N}$.

Задача 3.26 (***разыгрывание равномерно распределенных случайных величин***). Известно, что если случайная величина X равномерно распределена на отрезке $[a,b]$, то ее функция плотности вероятности $f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b]; \\ 0, & x \notin [a,b]. \end{cases}$

Пусть генерируется некоторое число $r \in (0; 1)$. Тогда составим уравнение:

$$\int_a^x \frac{1}{b-a}dx = r, \text{ откуда } x = a + r(b-a).$$

Задача 3.27 (***разыгрывание нормальных случайных величин***). Известно, что числовые характеристики равномерно распределенной случайной величины можно найти по формулам: $M(X) = \frac{a+b}{2}$, $D(X) = \frac{(b-a)^2}{12}$ и $\sigma = \frac{b-a}{\sqrt{12}}$.

Тогда для случайной величины R, равномерно распределенной на отрезке $[0,1]$, справедливы характеристики:

$$f(R) = \begin{cases} 1 & \text{при } X \in [0;\ 1]; \\ 0 & \text{при } X \notin [0;\ 1]; \end{cases}$$

$$F(R) = \begin{cases} 0 & \text{при } X < 0; \\ X & \text{при } X \in [0;\ 1]; \\ 1 & \text{при } X > 1, \end{cases}$$

а также $M(R) = \frac{1}{2}$; $D(R) = \frac{1}{12}$ и $\sigma = \sqrt{\frac{1}{12}}$.

Составим сумму n независимых равномерно распределенных на интервале [0; 1] случайных величин R_i, т.е. $\sum_{i=1}^{n} R_i$.

Если $M(R) = \frac{1}{2}$, для n случайных величин $M(\sum_{i=1}^{n} R_i) = \frac{n}{2}$. Аналогично, если $D(R) = \frac{1}{12}$, то для n случайных величин $D(\sum_{i=1}^{n} R_i) = \frac{n}{12}$; $\sigma(\sum_{i=1}^{n} R_i) = \sqrt{\frac{n}{12}}$.

Если привести рассматриваемую величину R к стандартному виду, пронормировав ее по формуле $\frac{x-m}{\sigma}$, то $X_i = \frac{\sum_{i=1}^{n} R_i - \frac{n}{12}}{\sqrt{\frac{n}{12}}}$.

Согласно центральной предельной теореме, при $n \to \infty$ распределение такой стандартной величины стремится к нормальному $N(0; 1)$. Тогда при $n = 12$ имеем $X_i = \sum_{i=1}^{12} R_i - 6$.

Это означает, что для разыгрывания возможных значений X_i нормальной случайной величины $X \in N(0; 1)$ надо выбрать случайным образом 12 чисел из ТСЧ и найти их сумму $\sum_{i=1}^{n} R_i$. Затем найти значение $X_i = \sum_{i=1}^{12} R_i - 6$.

Задача 3.28 (*задача Бюффона*). Произвольным образом иглу длиной $2l$ бросают на плоскость, разлинованную параллельными прямыми, расстояние между которыми $2d$. Какова вероятность того, что игла пересечет одну из этих параллельных прямых?

Решение. Расположим иглу на полосе между параллельными прямыми. Совместим с одной из параллельных прямых ось OX прямоугольной системы координат. Тогда положение иглы можно описать с помощью координат $(x; y)$ и угла поворота α к положительному направлению оси OX, где $(x; y)$ — координаты середины иглы. Рассмотрим часть плоскости между параллельными прямыми, где положение иглы зависит лишь от двух переменных y и α. Координата y в силу симметрии принимает значения $y \in [0; \alpha]$, $\alpha \in [-\frac{\pi}{2}; \frac{\pi}{2}]$. Поскольку игла произвольно падает на плоскость, то ее координаты $(y; \alpha)$ — случайные числа, равномерно распределенные на отрезках $y \in [0; d]$, $\alpha \in [0; \frac{\pi}{2}]$.

Тогда любая точка прямоугольника с равной вероятностью может оказаться серединой брошенной иглы.

Пусть $d \geq l$. Игла с центром в точке $(x; y)$ и углом наклона α к оси OX пересекает прямую тогда и только тогда, когда $y \leq l \sin\alpha$. Этому условию удовлетворяют только точки плоскости, расположенные ниже синусоиды.

Тогда вероятность пересечения иглы с одной из параллельных прямых может быть найдена с помощью геометрического определения вероятности (см. подразд. 1.5) как отношение площади под синусоидой $S_{\sin}$ и площади всего прямоугольника $S_{\text{пр}}$. Найдем площади по формулам: $S_{\text{пр}} = \frac{\pi}{2}d$ и $S_{\sin} = \int\limits_0^{\pi/2} l \sin\alpha \, d\alpha = -l\cos\alpha\Big|_0^{\pi/2} = l.$

Тогда вероятность события A равна дроби $P(A) = \dfrac{S_{\sin\alpha}}{S_{\text{пр}}} = \dfrac{l}{\frac{\pi}{2}d} = \dfrac{2l}{\pi d}.$

Согласно закону больших чисел в форме Бернулли, частота появления события в n испытаниях сходится по вероятности к вероятности этого события (при $n \to \infty$). Отсюда следует, что с помощью достаточно большой серии испытаний можно оценить с достаточно высокой точностью вероятность пересечения иглой одной из параллельных прямых.

Описанный метод позволяет вычислить число π. Действительно, так как $P(A) \to 1$, то $\dfrac{2l}{\pi d} \approx 1$. Откуда $\pi \approx \dfrac{2l}{d}$.

3.8. Основы вероятностной теории информации

Существуют различные подходы к анализу информации:

- *объемный подход* (связан с подсчетом количества информации в двоичной системе счисления с учетом только количества символов);
- *вероятностный подход* (учитывает частоту появления символа в тексте).

Сигнал есть сообщение, передаваемое с помощью некоторого носителя. Если источник вырабатывает сигнал, параметр которого — непрерывная функция времени, то он называется *непрерывным*. Непрерывна и информация, передаваемая такими сигналами.

Сигнал называется *дискретным*, если соответствующие ему числовые значения существуют в конечный момент времени и представляют собой конечную последовательность символов (например, 0,1).

Процедура преобразования непрерывных сигналов в дискретные называется ***дискретизацией***. Для дискретизации выбирают из алфавита конкретные сигналы, которые характеризуют все целостное сообщение.

Измерение информации. Математическое представление об информации связано с ее измерением. Существует несколько способов измерения количества информации.

Вообще под ***количеством информации*** принято понимать числовую характеристику сигнала, которая не зависит от его формы и содержания и соответствует степени неопределенности, исчезающей после получения сообщения в виде этого сигнала.

Информация, переданная в виде последовательности сигналов (знаков), представляет собой определенный *порядок*, некоторую структурную определенность. Но порядок не существует сам по себе, а всегда сопоставим с *беспорядком* — отсутствием структуры.

Мерой беспорядка любой системы является ***энтропия***. Понятие энтропии как меры беспорядка системы заимствовано из термодинамики и стало одним из основных не только в теории информации, но и в современном философском осмыслении открытых биосоциальных систем.

Формула Хартли. В передачу сообщения входит процесс выбора, который осуществляет как отправитель, так и получатель сообщения. Этот процесс выбора можно измерить количественно.

Пусть необходимо передать одно из восьми возможных сообщений: A, B, C, D, E, F, I, H.

Тогда один двоичный сигнал позволяет выбрать одно из двух сообщений: $2 = 2^1$; два двоичных сигнала — одно из четырех сообщений: $4 = 2^2$; три сигнала — одно из восьми сообщений: $8 = 2^3$; шесть сигналов — одно из 64 сообщений: $64 = 2^6$; m двоичных сигналов — одно из 2^m сообщений. И наоборот, для набора в $N = 2^m$ сообщений нужно иметь не менее m сигналов. Важно знать, что одно из 32 сообщений выбирается с помощью пяти сигналов. Поэтому для распознавания определенного сообщения из N возможных сообщений требуется $m = \log_2 N$ сигналов.

Так, при $N = 100$ требуется

$$m = \log_2 100 = \log_2 10^2 = 2\log_2 10 = 2\frac{1}{\lg 2} \approx 2 \cdot 3{,}32 \approx 6{,}64 \approx 7 \text{ знаков.}$$

Таким образом, для выбора одного из N вариантов нужна информация в J бит:

$$J = \log_2 N. \tag{3.34}$$

Эту формулу в 1928 г. предложил американский математик Р. Хартли. В качестве единицы измерения используется бит. Таким образом, для выбора из восьми знаков потребуется 3 бита.

Далее для меньшей громоздкости логарифм по основанию 2 будем обозначать log.

Формула Шеннона. Если сообщение появляется с одинаковой частотой, то полученные выводы единственно возможные. Если относительная частота сообщений одинакова, то для одного из четырех сообщений требуется $J = \log 4 = 2$ бита. Если частота появления отдельных сообщений различна, то может быть использовано и меньшее количество сигналов.

В естественном языке, например русском, учет вероятности (частоты) появления слов различной длины заключается в том, что с большей вероятностью появляются короткие слова.

Для расчета среднего количества информации найдем математическое ожидание дискретной случайной величины J и обозначим эту сумму через H. Тогда

$$H = M(J) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} = -\sum_{i=1}^{n} p_i \log p_i. \tag{3.35}$$

Эта формула дает среднее значение количества информации, приходящееся на один символ алфавита, — знаменитая формула К. Шеннона.

Функция H достигает максимума, когда частота (статистическая вероятность) для всех n сообщений одинакова и, следовательно, равна $p_i = \dfrac{1}{n}$, в этом случае

$$H_{\max} = \sum_{i=1}^{n} \frac{1}{n} \log n = n \frac{1}{n} \log n = \log n.$$

Таким образом, формула Хартли является частным случаем формулы Шеннона, когда исходы равновероятны. В тех случаях, когда мы не обладаем всей информацией о статистической вероятности исходов, а условно считаем их равновероятными, также будем применять для подсчетов формулу Хартли. Однако чтобы отличать такое ее применение от непосредственного, будем использовать понятие «частной» информации. Таким образом, подчеркивается тот частный случай, когда принимается условие равновероятных исходов: частную информацию будем находить по формуле $J_{x_i} = \log(1/p_i)$.

Отметим частные случаи формулы Хартли: если можно передать только один сигнал ($n = 1$), то $H = \log 1 = 0$. Таким образом, видим, что подобные сообщения не содержат информации вообще. Например, если каждое утро восходит солнце (этот сигнал является фоновым), то, увидев светящее солнце, мы ничего нового не узнаем. Если с равной вероятностью могут передаваться два сигнала ($n = 2$), то $H = \log 2 = 1$ (бит). Это также понятно, поскольку формула Хартли как раз и строилась в предположении о двоичности бита и мы лишь проверили ее действие на себе и, тем самым, непротиворечивость. Так, нетикающие часы свидетельствуют о том, что они остановились, тикающие — что идут.

С вероятностной точки зрения рассмотрим H_1 — количество энтропии до принятия сообщения и H_2 — количество энтропии после принятия сообщения. Тогда $J = H_1 - H_2$.

Пример 3.4. Эксперимент состоит в бросании игральной кости, где J — информация о результате бросания. Поскольку по окончании эксперимента известен результат бросания, то неопределенность снята: $H_2 = 0$, поэтому информация равна энтропии системы $J = H$.

Задача 3.29. Найти количество информации, имеющееся при получении сообщения на русском языке.

Решение. Поскольку в русском языке 33 буквы и один пробел между словами, то необходимо 34 места для символов. Тогда по формуле Хартли имеем $H = \log 34 = 5{,}09$ бит.

Так как в словах русского языка буквы встречаются с разной частотой, необходим вероятностный подход, т.е. учет статистической вероятности появления буквы в тексте. Анализ больших текстов дает такие численные значения статистической вероятности (частоты) появления букв в тексте.

Если учитывать частоту появления каждой буквы в тексте, то по формуле Шеннона $H = 4{,}35$ бит — для русского текста и $H = \log 27 = 4{,}76$ бит — для всех языков, основанных на латинском алфавите, в котором 26 букв.

Задача 3.30. На одной из клеток шахматной доски стоит фигура. Будем считать, что все положения этой фигуры на шахматной доске равновероятны. Какое количество информации несет сообщение:

а) о точном месте нахождения этой фигуры;

б) о том, что фигура находится в одной из угловых клеток доски?

Решение. а) Неопределенность информации, т.е. энтропия этой системы A, с N равновероятными состояниями, по формуле Хартли равна $\log N$, поэтому информация $J = H(A) = \log 64 = \log 2^6 = 6$ бит, т.е. сообщение о том, что фигура находится, например, в клетке D5, несет в себе 6 бит информации.

б) Так как на доске четыре угловые клетки, а вероятность находиться в каждой из них $\frac{1}{64}$, то вероятность попасть на угол $p = \frac{4}{64} = \frac{1}{16}$. Тогда частная информация сообщения $J_A = \log\frac{1}{p(A)} = \log\frac{1}{1/16} = 4$ бит.

Задача 3.31. Какую частную информацию несет сообщение «В этом году я хожу в гости по последним субботам каждого месяца»?

Решение. Так как в гости можно пойти в любую субботу года, подсчитаем число возможных суббот: 12. Таким образом, вероятность пойти в гости в последнюю субботу $P = \frac{1}{12}$, а частная информация $J_{A_i} = \log\frac{1}{p_i} = \log\frac{1}{1/12} = \log 12 \approx 3{,}58$ бит.

Задача 3.32. Мой новый знакомый сообщил: «На этой неделе мой день рождения».

а) Какое количество информации заключено в сообщении?

б) Каким минимальным числом уточняющих вопросов (с ответами «Да» и «Нет») можно узнать точную дату дня рождения знакомого?

Решение. а) Поскольку любой день недели равновероятен, $P_1 = P_2 = ... = P_7 = \frac{1}{7}$. Тогда частная информация $J_A = \log\frac{1}{1/7} = \log 7 \approx 2{,}8 \approx 3$ бит.

б) Поскольку информация, заключенная в сообщении $I_A \approx 3$, необходимо минимум три вопроса для выяснения точной даты дня рождения. Представим все дни недели в виде отрезка с семью делениями.

Принцип постановки уточняющего вопроса (с ответами «Да» и «Нет») основан на «методе половинного деления». Пусть неизвестная дата дня рождения приходится на субботу, тогда возможен такой вариант диалога:

Вопрос 1: «Твой день рождения до четверга?»

Ответ 1: «Нет».

Вопрос 2: «Твой день рождения после пятницы?»

Ответ 2: «Да».

Вопрос 3: «Твой день рождения в воскресенье?»

Ответ 3: «Нет».

Вывод. День рождения в субботу.

Таким образом, если о системе известны определенные начальные условия, то можно измерить количество информации, заключенной в сообщении об этой системе.

Формула Хартли $J = \log_2 N$ позволяет установить важное свойство ***аддитивности информации*** для определения количества информации, соответствующего двум сообщениям: если $J_1 = \log_2 N_1$ и $J_2 = \log_2 N_2$, то $J_1 + J_2 = \log_2 N_1 + \log_2 N_2 = \log_2 N_1 N_2$.

Задача 3.33. Найти количество информации, которое можно получить при ответе на вопрос: при бросании игрального кубика выпала цифра «5».

Решение. Вероятность выпадения числа «5» на кубике $p = \frac{1}{6}$, а выпадения другой цифры $\bar{p} = \frac{5}{6}$. Тогда по формуле Шеннона имеем $J = -\frac{1}{6}\log_2\frac{1}{6} - \frac{5}{6}\log_2\frac{5}{6} = 0,65$ бит.

Контрольные вопросы

3.1. В чем заключается метод суперпозиций разыгрывания НСВ?
3.2. Что называется генеральной совокупностью объектов?
3.3. Чем выборка отличается от генеральной совокупности объектов?
3.4. Какими качествами должна обладать выборка?
3.5. Чем характеризуются статические распределения выборки?
3.6. Чем вариационный ряд отличается от ряда распределения ДСВ?
3.7. В чем отличие эмпирической функции распределения от теоретической?
3.8. Как построить полигон частот и для чего он используется?
3.9. Как построить полигон относительных частот?
3.10. Как построить гистограмму частот?
3.11. Каковы статистические оценки параметра распределения?
3.12. Чем характеризуется интервальная оценка?
3.13. Как вычислить генеральное (выборочное) среднее и его оценку?
3.14. Как вычислить генеральную (выборочную) дисперсию?
3.15. Почему вводится исправленная дисперсия и чем она отличается от обычной?
3.16. Чем отличается точечная оценка от интервальной?
3.17. Как найти доверительный интервал для оценки математического ожидания нормального распределения при известном среднеквадратическом отклонении?
3.18. Как оценить точность измерений при биномиальном распределении ДСВ по относительной частоте?
3.19. В чем заключается выборочный метод?
3.20. Какие гипотезы называются статистическими?
3.21. Какую гипотезу называют альтернативной?
3.22. В чем различие ошибок первого и второго рода?
3.23. Что называют областью принятия гипотезы?
3.24. Какими бывают критические области?
3.25. В чем заключается принцип проверки статистических гипотез?
3.26. В чем суть метода Монте-Карло?
3.27. В чем заключается разыгрывание ДСВ?
3.28. Как разыграть НСВ методом обратных функций?

Контрольные задания

Задание 3.1.

1. Выберите проблему исследования методами математической статистики. Работу можно проводить как индивидуально, так и в микрогруппах («научных обществах»). Найдите по выбранной проблеме статистические данные либо в периодической печати, либо с помощью социологического опроса, взяв за основу изменение некоторой варианты.

2. Рассмотрите выборку из полученных статистических данных и в каждом проекте выполните следующие задания:

а) постройте вариационный ряд, эмпирическую функцию распределения и ее график — кумуляту;

б) постройте полигон и гистограмму;

в) вычислите точечную и интервальную оценки параметров распределения;

г) вычислите точечную несмещенную оценку для дисперсии;

д) найдите интервал, в который с заданной вероятностью попадает СВ, распределенная нормально или по Стьюденту, с помощью статистических таблиц;

е) вычислите доверительный интервал для математического ожидания m нормального распределения;

ж) вычислите доверительный интервал для генеральной дисперсии D и среднеквадратического отклонения σ;

з) проверьте статистическую гипотезу о числовом значении среднего, если выборка производится из одной совокупности.

3. Оформите свой проект по правилам оформления исследовательских работ. Для этого необходимо:

- аргументировать актуальность темы;
- определить проблему исследования, объект, предмет исследования;
- определить цели и задачи исследования;
- сформулировать гипотезу исследования;
- выбрать методы исследования;
- обработать эмпирические данные;
- сформулировать аргументированный вывод по проблеме исследования;
- провести презентацию проекта.

Задание 3.2. «*Метод Монте-Карло*».

Найдите значение числа π четырьмя способами, моделируя попадание чисел на указанный интервал:

1) механическим путем; для этого рассыпать на квадратном листе бумаги крупу (рис или гречку), подсчитать отношение числа зерен в четверти единичного круга, радиус которого совпадает со стороной квадрата, к числу зерен во всем квадрате. Полученное отношение останется лишь увеличить в четыре раза (см. задачу 3.25);

2) с помощью таблицы случайных чисел; для этого случайным образом формируются координаты точек как пары чисел, взятых из ТСЧ, затем точки строятся все в том же квадрате, содержащем четверть единичного круга;

3) с помощью ЭВМ как генератора случайных чисел по описанному ранее плану;

4) с помощью иглы Бюффона (см. задачу 3.28).

Для обеспечения действия закона больших чисел необходимо большое число испытаний, что довольно трудно осуществить в одиночку.

Поэтому данное задание также рекомендуется выполнять в микрогруппе — «научном обществе» и представить результат общих наблюдений, обобщив полученные данные.

Задачи для самостоятельного решения

3.1. Постройте вариационный ряд, полигон частот, полигон относительных частот и график функции распределения по данным выборки:

1) 2, 4, 2, 4, 3, 3, 3, 2, 0, 6, 1, 2, 3, 2, 2, 4, 5, 6, 6, 1, 1, 2, 3, 6;

2) 5, 8, 7, 6, 7, 8, 9, 10, 5, 8, 7, 5, 6, 6, 9, 7, 8, 10, 9, 8, 5, 6, 8;

3) 10,12, 15, 13, 16, 17, 18, 12, 10, 15, 13, 15, 17, 18, 15, 16, 17, 17, 18, 15;

4) 16, 20, 23, 18, 24, 18, 23, 25, 24, 24, 25, 18, 16, 16, 23, 25;

5) 32, 35, 38, 33, 39, 35, 36, 36, 38, 35, 32, 33, 33, 33, 39, 39;

6) 46, 45, 48, 45, 43, 43, 43, 48, 45, 44, 44, 48, 48, 44, 45, 43, 44, 48.

3.2. По данным выборки постройте гистограмму частот, гистограмму относительных частот, эмпирическую функцию распределения и ее график — кумуляту:

1) 23,5; 26,4; 48,6; 35,8; 32,9; 41,1; 33,3; 46,3; 49,9; 34,1; 45,2; 34,5; 42,4; 47,3; 32,4; 33,3; 34,4; 30,8; 43,7; 46,9; 41,3; 34,6;

2) 50,5; 65,4; 51,6; 69,8; 65,9; 57,1; 67,3; 64,3; 54,9; 56,1; 61,2; 67,5; 64,4; 63,3; 62,4; 60,3; 69,4; 55,8; 53,7; 58,9; 57,3; 50,6;

3) 45,4; 51,4; 56,5; 47,8; 53,5; 47,2; 49,7; 48,3; 45,9; 51,3; 54,9; 54,8; 56,3; 56,4; 53,4; 53,9; 45,8; 57,4; 54,8; 48,7; 46,3; 49,6; 58,6; 54,7;

4) 30,8; 28,7; 36,5; 28,4; 27,5; 36,5; 34,2; 39,6; 30,8; 32,7; 28,7; 25,5; 26,1; 35,1; 38,1; 39,2; 37,0; 34,2; 36,1; 26,1; 28,9; 25,2; 32,5; 37,8; 34,2; 36,5; 37,3; 38,1; 29,0; 30,2;

5) 82,5; 79,8; 76,9; 74,8; 84,7; 85,2; 80,9; 80,7; 76,9; 75,8; 85,7; 82,5; 82,4; 75,9; 79,6; 83,6; 89,5; 84,7; 76,9; 78,6; 79,5; 89,4; 82,2; 86,7; 78,9;

6) 98,6; 87,6; 94,7; 86,5; 85,9; 82,3; 85,6; 83,9; 89,0; 96,8; 95,8; 95,9; 94,8; 84,9; 89,5; 83,9; 86,5; 87,9; 82,0; 84,8; 95,7; 84,3; 84,9; 82,5; 85,7.

3.3. Запишите статистические распределения дискретного признака X и найдите его числовые характеристики:

1) сведения о числе пропущенных уроков по математической статистике у 25 студентов третьего курса имеют вид: 4, 3, 6, 0, 0, 0, 5, 0, 2, 2, 4, 5, 3, 0, 0, 2, 4, 5, 4, 5, 5, 6, 0, 0, 0;

2) в колледже проводилось тестирование по теории вероятностей, содержащее 60 вопросов. Данные о результатах тестирования группы из 25 студентов имеют вид: 44, 35, 56, 60, 50, 48, 55, 60, 52, 52, 54, 45, 43, 60, 40, 52, 54, 56, 49, 59, 58, 56, 50, 60, 60;

3) для практического занятия по математической статистике студенты провели исследование о затратах времени на ежедневную подготовку к следующему учебному дню среди однокурсников. Статистические данные, выраженные в часах, имели вид: 3,4; 4,3; 2,5; 2,7; 1,9; 2,5; 3,5; 2,8; 3,4; 2,5; 2,5; 3,5; 2,8; 3,8; 1,9; 2,3; 1,9; 2,7; 4,0; 2,8;

4) для практического занятия по математической статистике студенты провели исследование, выясняя число клиентов сберкассы в период с 18 до 19 ч. Полученные статистические данные за октябрь имеют вид: 12, 16, 24, 15, 21, 18, 21, 16, 19, 32, 28, 27, 29, 34, 28, 17, 15, 16, 20, 21, 24, 16, 14, 18, 25, 21;

5) наблюдения за числом посетителей сайта колледжа за последние 25 дней дали следующие результаты: 22, 12, 26, 24, 15, 11, 28, 21, 16, 29, 32, 28, 37, 29, 34, 28, 37, 15, 16, 24, 16, 14, 18, 25, 27;

6) на контрольной работе по математике отобрали случайным образом 16 студентов и провели хронометраж временны́х затрат для выполнения одного задания. Статистические наблюдения, выраженные в минутах, имели вид: 2,4; 4,3; 2,3; 1,7; 1,9; 2,0; 3,1; 1,8; 2,5; 3,5; 2,8; 1,9; 2,3; 1,9; 2,7; 2,0;

7) Проведите собственное исследование и решите аналогичную задачу.

3.4. По данному распределению выборки найдите $\bar{x}_в$, $D_в$, $\sigma_в$ и дайте оценки генеральных совокупностей:

1)

X_i	1	4	8	9
n_i	5	10	15	20

2)

X_i	1	3	5	7
n_i	8	12	16	14

3)

X_i	1	2	3	5
n_i	5	15	10	20

4)

X_i	2	4	6	8
n_i	15	10	20	5

5)

X_i	1	2	3	4
n_i	8	16	12	14

6)

X_i	5	6	8	9
n_i	3	21	29	17

3.5. По выборке объемом n найдена смещенная оценка выборочной дисперсии. Найдите несмещенную оценку генеральной совокупности $D_г$ и $\sigma_г$:

1) $n = 41$, $D_в = 3$; 2) $n = 31$, $D_в = 5$; 3) $n = 51$, $D_в = 3$;

4) $n = 39$, $D_в = 5$; 5) $n = 42$, $D_в = 6$; 6) $n = 49$, $D_в = 8$.

3.6. По известным n и γ с помощью функции Лапласа найдите доверительный интервал для оценки неизвестного математического ожидания m с заданной надежностью, предполагая, что измеряемая величина распределена нормально:

1) $\sigma = 2$; $\bar{x}_в = 5{,}4$; $n = 100$; $\gamma = 0{,}99$; 2) $\sigma = 3$; $\bar{x}_в = 6{,}1$; $n = 50$, $\gamma = 0{,}95$;

3) $\sigma = 5$; $\bar{x}_в = 7{,}3$; $n = 80$; $\gamma = 0{,}999$; 4) $\sigma = 4$; $\bar{x}_в = 9{,}2$; $n = 100$; $\gamma = 0{,}95$;

5) $\sigma = 7$; $\bar{x}_в = 9{,}8$; $n = 80$; $\gamma = 0{,}999$; 6) $\sigma = 9$; $\bar{x}_в = 5{,}7$; $n = 100$; $\gamma = 0{,}99$.

3.7. По известным n, γ и S и неизвестным m и σ найдите доверительные интервалы с помощью распределения Стьюдента, предполагая, что измеряемая величина распределена нормально.

1) $S = 1{,}5$; $\bar{x}_в = 18{,}6$; $n = 15$; $\gamma = 0{,}99$; 2) $S = 1{,}6$; $\bar{x}_в = 16{,}2$; $n = 12$; $\gamma = 0{,}95$;

3) $S = 1{,}8$; $K_\gamma = 15{,}6$; $n = 18$; $\gamma = 0{,}999$; 4) $S = 2{,}4$; $\bar{x}_в = 4{,}2$; $n = 20$; $\gamma = 0{,}99$;

5) $S = 2{,}6$; $\bar{x}_в = 8{,}2$; $n = 24$; $\gamma = 0{,}95$; 6) $S = 2{,}8$; $\bar{x}_в = 12{,}6$; $n = 25$; $\gamma = 0{,}999$.

3.8. Перечислите все возможные значения вариант случайной величины и их вероятности. Постройте распределение частот возможных результатов, полученных в этой задаче:

1) три человека отвечают на вопрос в форме «Да» и «Нет». Перечислите все возможные варианты положительных ответов;

2) при подбрасывании двух игральных костей сумма цифр;

3) при подбрасывании двух игральных костей разность цифр;

4) при подбрасывании двух монет выпадение орла;

5) при подбрасывании трех монет выпадение орла;

6) при подбрасывании двух предметов: игральной кости и монеты выпадение четной цифры и орла одновременно.

3.9. Найдите несмещенные оценки: выборочное среднее, дисперсию, среднеквадратическое отклонение по следующим данным:

1) в целях определения времени, затрачиваемого на обработку детали, исследована выборочно производительность труда 100 рабочих авиационного завода. Результаты обследования приведены в таблице:

Время обработки, мин	2,6—3,2	3,2—3,8	3,8—4,4	4,4—5,0	5,0—5,6
Количество рабочих	12	34	30	15	9

2) проводилось выборочное обследование продуктивности коров на молочных фермах Северо-Западного экономического региона РФ. Получены следующие результаты:

Надой за год, л	3000—3400	3400—3800	3800—4200	4200—4600	4600—5000
Количество коров	43	71	102	64	27

3) изучалось распределение населения одного из городов РФ по среднедушевому совокупному доходу в 1992 г. Результаты исследований:

Месячный доход, р.	Менее 1000	1000—2000	2000—3000	3000—4000	4000—5000	5000—6000	6000—7000	Более 7000
Количество человек	70	326	342	250	120	80	26	6

4) исследовалось время безотказной работы 500 бытовых ультразвуковых стиральных машин «Пони» в течение гарантийного срока. Для этого фиксировалось время (в днях) с момента приобретения до момента первого обращения покупателя в гарантийную мастерскую. Были зафиксированы следующие результаты:

Время безотказной работы, дни	0—60	60—120	120—180	180—240	240—300	300—360
Число отказов	16	44	121	185	112	22

5) в универсаме проводилась контрольная проверка массы расфасованных товаров, которая дала следующие результаты:

Ошибка взвешивания, г	–15...–10	–10...–5	–5...0	0...5	5...10	10...15
Число наблюдений, попавших в данный интервал	25	44	112	67	32	20

6) на СТОА (станция технического обслуживания автомобилей) исследовались затраты времени на ремонт карбюратора. Были зафиксированы следующие результаты:

Затраты времени, мин	10—20	20—30	30—40	40—50	50—60	Более 60
Число наблюдений, попавших в данный интервал	8	12	43	47	24	16

3.10. Вычислите доверительный интервал для генеральной дисперсии D и среднеквадратического отклонения σ по данным следующим выборок:

1) выборочное обследование возраста женщин, вступающих в брак в нашем городе в течение года, дало следующие результаты:

Возраст, лет	15—21	21—27	27—33	33—39	39—45	45—51	51—57	57—63
Количество женщин	22	320	380	340	250	112	58	18

2) в целях улучшения режима работы овощной базы проводился хронометраж времени, затраченного автомобилями-самосвалами на погрузку, данные которого представлены в таблице:

Время пребывания, мин	30—35	35—40	40—45	45—50	50—55	55—60
Число автомобилей	28	42	64	36	22	8

3) проводилось выборочное обследование затрат времени покупателей крупного универсама на очередь в кассу. Были получены следующие данные:

Время ожидания, мин	0—1	1—2	2—3	3—4	4—5	5—6
Количество покупателей	22	96	146	84	42	10

4) распределение роста студентов нашего факультета представлено в таблице:

Рост, см	154—160	160—166	166—172	172—178	178—184	184—190	190—196
Количество студентов	10	45	55	62	18	8	2

5) выборочные сведения о выполнении норм выработки рабочими приведены в таблице:

Процент выполнения	90—100	100—110	110—120	120—130	130—140
Количество рабочих	20	160	120	70	30

6) распределение семей по размеру жилой площади, приходящейся на одного человека, в некоторой организации составляет:

№ п/п	Площадь, приходящаяся на одного человека, м^2	Число семей с данным размером площади
1	3—5	5
2	5—7	25
3	7—9	55
4	9—11	40
5	11—13	25

3.11. По результатам, полученным в задаче 3.8 (1), ответьте на вопросы: какова вероятность получения суммы (односторонняя вероятность):

1) равной 9 и более очкам;
2) не менее 10 очков;
3) 3 и менее очков;
4) не более 4 очков;
5) не менее 11 очков;
6) не более 2 очков?

3.12. По результатам, полученным в задаче 3.8 (1), ответьте на вопросы: какова вероятность появления результата (двусторонняя вероятность):

1) 11 и более или трех и менее очков;
2) 10 и более или не менее двух очков;
3) столь же редкого, что и сумма в 5 и менее очков;

4) столь же редкого, что и сумма в 11 очков;

5) от 8 до 11 очков;

6) от 5 до 9 очков?

3.13. Вероятно ли, чтобы после двенадцати подбрасываний монет получен результат, столь же редкий, как и выпадание:

1) трех и менее орлов;

2) 11 и более орлов;

3) 4 и менее орлов;

4) 10 и более орлов;

5) от 5 до 8 орлов;

6) от 1 до 3 орлов?

3.14. Сформулируйте нулевую и альтернативную гипотезы по исследуемой проблеме, если изучается проблема:

1) получения высшего образования выпускниками колледжа;

2) возможности свободного устройства на работу выпускников колледжа;

3) дискриминации женщин при поступлении на машиностроительный факультет колледжа;

4) отказа в приеме на работу в связи с отсутствием стажа работы;

5) устройства на работу выпускников колледжа по полученной специальности;

6) устройства на работу по специальности в процессе обучения в колледже.

ГРЕЧЕСКИЙ АЛФАВИТ

Α, α — альфа
Β, β — бета
Γ, γ — гамма
Δ, δ — дельта
Ε, ε — эпсилон
Ζ, ζ — дзета
Η, η — эта
Θ, θ — тета
Ι, ι — йота
Κ, κ — каппа
Λ, λ — ламбда
Μ, μ — мю

Ν, ν — ню
Ξ, ξ — кси
Ο, ο — омикрон
Π, π — пи
Ρ, ρ — ро
Σ, σ — сигма
Τ, τ — тау
Υ, υ — ипсилон
Φ, φ — фи
Χ, χ — хи
Ψ, ψ — пси
Ω, ω — омега

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ 1

Статистико-математические таблицы

Таблица П. 1

Значения функции $P(m, \lambda) = \frac{\lambda^m}{m!} e^{-\lambda}$ (распределение Пуассона)

m	λ									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,90484	81873	74082	67032	60653	54881	49659	44933	40657	36788
1	09048	16375	22225	26813	30327	32929	34761	35946	36591	36788
2	00452	01637	03334	05363	07582	09879	12166	14379	16466	18394
3	00015	00109	00333	00715	01264	01976	02839	03834	04940	06131
4		00005	00025	00072	00158	00296	00497	00767	01111	01533
5			00002	00006	00016	00036	00070	00123	00200	00307
6					00001	00004	00008	00016	00030	00051
7							00001	00002	00004	00007
8										00001

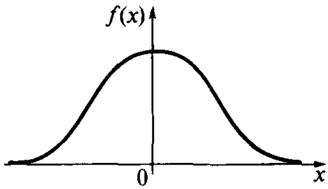
m	λ									
	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0
0	0,22313	13534	08208	04979	03020	01832	01111	00674	00409	00248
1	33470	27067	20521	14936	10569	07326	04999	03369	02248	01487
2	25102	27067	25652	22404	18496	14653	11248	08422	06181	04462
3	12551	18045	21376	22404	21579	19537	16872	14037	11332	08924
4	04707	09022	13360	16803	18881	19537	18981	17547	15582	13385

Окончание табл. П. 1

m	λ									
	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0
5	01412	03609	06680	10082	13217	15629	17083	17547	17140	16062
6	00353	01203	02783	05041	07710	10420	12812	14622	15712	16062
7	00076	00344	00994	02160	03855	05954	08236	10444	12345	13768
8	00014	00086	00311	00810	01687	02977	04633	06528	08487	10326
9	00002	00019	00086	00270	00656	01323	02316	03627	05187	06884
10		00004	00022	00081	00230	00529	01042	01813	02853	04130
11		00001	00005	00022	00073	00192	00426	00824	01426	02253
12			00001	00006	00021	00064	00160	00343	00654	01126
13				00001	00006	00020	00055	00132	00277	00520
14					00001	00006	00018	00047	00109	00223
15						*00002*	*00005*	*00016*	*00040*	*00089*
16							00002	00005	00014	00033
17								00001	00004	00012
18									00001	00004
19										00001

Таблица П. 2

Значения функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$



x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29431	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1,7	09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05339	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337

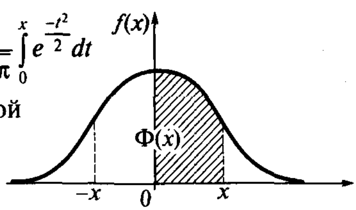
x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014
x	0		2		4		6		8	
4,0	0,00013	38	0000	589	0000	249	0000	101	00000	40
5,0	00000	15								

Таблица П. 3

Значение интеграла Лапласа $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{\frac{-t^2}{2}} dt$

При $x > 5$ можно пользоваться формулой

$$\Phi(x) \approx \frac{1}{2} - \frac{e^{-x^2}}{2x\sqrt{\pi}}.$$



x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524

x	0	1	2	3	4	5	6	7	8	9
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41308	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997

Окончание табл. П. 3

Десятые доли x					
x	0	2	4	6	8
4,0	0,4999683	4999867	4999946	4999979	4999992
5,0	4999997				

Таблица П. 4

Границы доверительных интервалов для дисперсии в зависимости от числа степеней свободы и надежности

Для получения несмещенных доверительных границ интервалов для оценки дисперсии σ^2 при нормальном распределении случайных величин в выборке находят несмещенную оценку дисперсии s^2, затем эту оценку умножают на коэффициенты, взятые из таблицы.

Пример. Пусть по выборке $n = 10$ при неизвестном среднем $s^2 = 2{,}87$. Число степеней свободы $k = 9$, поэтому границы доверительного интервала надежности 0,90 таковы: нижняя — $0{,}498 \cdot 2{,}87 = 1{,}429$, верхняя — $2{,}481 \cdot 2{,}87 = 7{,}120$.

Число степеней свободы n	Надежность					
	0,90		0,95		0,99	
	Нижняя граница	Верхняя граница	Нижняя граница	Верхняя граница	Нижняя граница	Верхняя граница
5	0,402	3,691	0,348	5,054	0,270	9,927
6	0,432	3,201	0,377	4,211	0,297	7,569
7	0,457	2,876	0,402	3,679	0,319	6,239
8	0,479	2,651	0,424	3,314	0,341	5,304
9	0,498	2,481	0,443	3,048	0,360	4,680
10	0,514	2,348	0,460	2,844	0,376	4,243
11	0,529	2,242	0,475	2,683	0,391	33,897
12	0,543	2,153	0,489	2,553	0,405	3,635
13	0,555	2,080	0,502	2,440	0,417	3,424
14	0,566	2,018	0,513	2,354	0,429	3,243
15	0,576	1,965	0,525	2,272	0,440	3,091
16	0,586	1,918	0,534	2,208	0,450	2,961
17	0,594	11,876	0,544	2,147	0,461	2,833
18	0,602	1,839	0,552	2,095	0,470	2,739

Число степеней свободы n	Надежность					
	0,90		0,95		0,99	
	Нижняя граница	Верхняя граница	Нижняя граница	Верхняя граница	Нижняя граница	Верхняя граница
19	0,610	1,805	0,560	2,050	0,477	2,664
20	0,617	1,776	0,568	2,006	0,487	2,573
30	0,671	1,584	0,626	1,743	0,548	2,121
39	0,703	1,490	0,660	1,618	0,586	1,914
1000	0,931		0,918	1,094		

Таблица П. 5

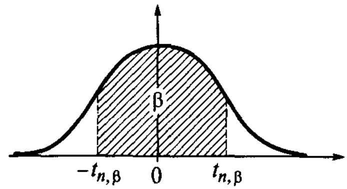
Значение K_β нормального распределения

Обозначения: α — ошибка; γ — уровень доверия; K_γ — решения уравнения $\frac{1}{\sqrt{2\pi}}\int\limits_{-K_\gamma}^{+K_\gamma} e^{-\frac{x^2}{2}}dx = \gamma$ (значения γ в первом столбце, соответствующая ошибка $\alpha = 1 - \gamma$ во втором столбце, поэтому, если задана ошибка α, то надо искать ее во втором столбце). Если нужно построить одностороннюю область, т.е. решить уравнение вида $\tilde{\alpha} = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{-K_{\tilde{\alpha}}} e^{-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int\limits_{+K_{\tilde{\alpha}}}^{\infty} e^{-\frac{x^2}{2}}dx$, то $\tilde{\alpha} = 1 - \gamma$ необходимо искать в третьем столбце.

γ	$\alpha = 1 - \gamma$	$\tilde{\alpha} = \alpha/2$	K_γ
0,9	0,1	0,05	1,65
0,95	0,05	0,025	1,96
0,98	0,02	0,01	2,3
0,99	0,01	0,005	2,58
0,9975	0,0025	0,00125	3,02

Таблица П. 6

Значение $t_{n,\beta}$ распределения Стьюдента



Обозначения: α — уровень значимости (ошибка); β— уровень доверия. В таблице заданы $t_{n,\beta}$ — решения уравнения $\beta = \int_{-t_{n,\beta}}^{+t_{n,\beta}} f_n(x)dx$ — в зависимости от числа степеней свободы n и уровня значимость (α)/доверие (β).

Число степеней свободы n	Уровень значимости α = 1 – β (двусторонняя критическая область)			
	0,10/0,90	0,05/0,95	0,02/0,98	0,01/0,99
1	6,31	12,7	31,82	63,7
2	2,92	4,30	6,97	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,6
5	2,01	2,57	3,37	4,03
6	1,94	2,45	3,14	3,71
7	1,89	2,36	3,00	3,5
8	1,86	2,31	2,9	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
11	1,80	2,220	2,72	3,11
12	1,78	2,18	2,68	3,05
13	1,77	2,16	2,65	3,01
14	1,76	2,14	2,62	2,98
15	1,75	2,13	2,60	2,95

Число степеней свободы n	Уровень значимости $\alpha = 1 - \beta$ (двусторонняя критическая область)			
	0,10/0,90	0,05/0,95	0,02/0,98	0,01/0,99
16	1,75	2,12	2,58	2,92
17	1,74	2,11	2,57	2,90
18	1,73	2,10	2,55	2,88
19	1,73	2,09	2,54	2,86
20	1,73	2,09	2,53	2,85
30	1,70	2,04	2,46	2,75
40	1,68	2,02	2,42	2,70
60	1,67	2,00	2,39	2,66
120	1,66	1,98	2,36	2,62
∞	1,64	1,96	2,33	2,58
	0,05/0,95	0,025/0,975	0,01/0,99	0,005/0,995

Уровень значимости $\alpha = 1 - \beta$ (односторонняя критическая область).

Таблица П. 7

Критическое значение для F-распределения при $\alpha = 0{,}05$

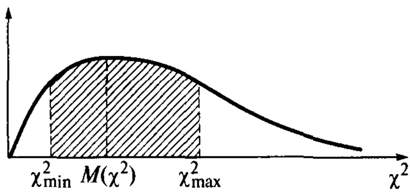
Обозначения: k_2 — число степеней свободы знаменателя; k_1 — число степеней свободы числителя

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,0	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Таблица П. 8

Критические точки $\chi^2_{\alpha,k}$ распределения Пирсона (χ^2)



n	α					
	0,99	0,98	0,95	0,90	0,80	0,20
1	0,0001	0,0006	0,0039	0,0158	0,0642	1,642
2	0,0201	0,0404	0,13	0,211	0,446	3,219
3	0,115	0,185	0,352	0,584	1,005	4,642
4	0,297	0,429	0,711	1,064	1,649	5,989
5	0,554	0,752	1,145	1,610	2,343	7,289
6	0,872	1,134	1,635	2,204	3,070	8,558
7	1,239	1,564	2,167	2,833	3,822	9,803
8	1,646	2,032	2,733	3,490	4,594	11,030
9	2,088	2,532	3,325	4,168	5,380	12,242
12	2,558	3,059	3,940	4,865	6,179	13,442
11	3,053	3,609	4,575	5,578	6,989	14,631
12	3,571	4,178	5,226	6,304	7,807	15,812
13	4,107	4,765	5,892	7,042	8,634	16,985

Продолжение табл. П. 8

n	α					
	0,99	0,98	0,95	0,90	0,80	0,20
14	4,660	5,368	6,571	7,790	9,467	18,151
15	5,229	5,985	7,262	8,547	10,307	19,311
16	5,812	6,614	7,962	9,312	11,152	20,465
17	6,408	7,255	8,672	10,085	12,002	21,615
18	7,015	7,906	9,390	10,865	12,857	22,760
19	7,633	8,567	10,117	11,651	13,716	23,900
20	8,260	9,237	10,851	12,443	14,578	25,038
21	8,897	9,915	11,591	13,240	15,445	26,171
22	9,542	10,600	12,338	14,041	16,314	27,301
23	10,196	11,293	13,091	14,848	17,187	28,429
24	10,856	11,292	13,848	15,659	18,062	29,553
25	11,524	12,697	14,611	16,473	18,940	30,675
26	12,198	13,409	15,379	17,292	19,820	31,795
27	12,879	14,125	16,151	18,114	20,703	31,912
28	13,565	13,847	16,928	18,939	21,588	34,027
29	14,256	15,574	17,708	19,768	22,475	35,139
30	14,953	16,306	18,493	20,599	23,364	36,250

Продолжение табл. П. 8

n	α				
	0,10	0,05	0,02	0,01	0,001
1	2,706	3,841	5,412	6,635	10,827
2	4,605	5,991	7,824	9,210	13,815
3	6,251	7,815	9,837	11,341	16,268
4	7,779	9,488	11,668	13,277	18,465
5	9,236	11,070	13,388	15,086	20,517
6	10,645	12,592	15,033	16,812	22,457
7	12,017	14,067	16,622	18,475	24,322
8	13,362	15,507	18,168	20,090	26,125
9	14,684	16,919	19,679	21,666	27,877

Окончание табл. П. 8

n	α				
	0,10	0,05	0,02	0,01	0,001
10	15,987	18,307	21,161	23,209	29,588
11	17,275	19,675	26,618	24,725	31,264
12	18,549	21,026	24,054	26,217	32,909
13	19,812	22,362	25,472	27,688	34,528
14	21,064	23,685	26,873	29,141	36,123
15	22,307	24,996	28,259	30,578	37,697
16	23,542	26,296	29,633	32,000	39,252
17	24,769	27,587	30,995	33,409	40,790
18	25,989	28,869	32,346	34,805	42,312
19	27,204	30,144	33,687	36,191	43,820
20	28,412	34,410	35,020	37,566	45,315
21	29,615	32,671	36,343	38,932	46,797
22	30,813	33,924	37,659	40,289	48,268
23	32,007	35,172	38,968	41,638	49,728
24	33,169	36,415	40,270	42,980	51,179
25	34,382	37,652	41,566	44,314	52,620
26	35,563	38,885	42,856	45,642	54,052
27	36,741	40,113	44,140	46,963	55,476
28	37,916	41,337	45,419	48,278	56,893
29	39,087	42,557	46,693	49,588	58,302
30	40,256	43,773	47,962	50,892	59,703

Таблица П. 9

Значения $q = q(\gamma, n)$

n	γ		n	γ	
	0,95	0,99		0,95	0,99
5	2,78	4,60	20	0,95	0,99
6	2,57	4,03	25	2,093	2,861
7	2,45	3,71	30	2,064	2,797

n	γ		n	γ	
	0,95	0,99		0,95	0,99
8	2,37	3,50	35	2,045	2,756
9	2,31	3,36	40	2,032	2,720
10	2,26	3,25	45	2,023	2,708
11	2,23	3,17	50	2,016	2,692
12	2,20	3,11	60	2,009	2,679
13	2,18	3,06	70	2,001	2,662
14	2,16	3,01	80	1,996	2,649
15	2,15	2,98	90	1,991	2,640
16	2,13	2,95	100	1,987	2,633
17	2,12	2,92	120	1,984	2,627
18	2,11	2,90	∞	1,980	2,617
19	2,10	2,88			

Таблица П. 10

Таблица случайных чисел

1009732533	7652013586	3467354876	3754204805	6489474296	2460524037
0842268953	1964509303	2320902560	9901902529	0937670715	3631131165
1280799970	8015736147	6403236653	8095909917	3929274945	6606574717
2063610402	0082291665	3106010805	1595334764	3508033606	8526977602
8867674397	0443627659	6357232135	9895116877	1217176833	7379645753
3407246850	3669736170	6581339885	4557182406	3530342614	8679907439
0205165692	6866574818	7305385247	0352964778	3580834282	6093520344
1119929170	9852017767	1490568607	2340309732	1180505431	3980827732
1862388579	8345299634	0628898083	8349125624	8868540200	8650758401
3527388435	9959467648	8751764969	2210940558	6097093433	5050073998
5072268248	2940524201	5277567851	1374670078	1847540610	6871177617
3676667951	9036476493	2960911062	9182608928	9378561366	2347834113
6548117674	1746850950	5804776974	8012425635	1772708015	4531622374
7435099817	7740277214	4323600210	6991626803	6625229148	3693667203
0989320505	1422568514	4642756788	7303957186	4021816544	9149914523
2111578253	1438553763	8033694598	4552164237	9628602655	4410461949
7662113990	9440056418	1255073742	9629778822	5438214598	6360649329
6847927686	4616283554	9475089923	2694036858	7029734135	5314033340
8515747954	3297926575	5765040881	1110022040	1286074697	9664489439
1650534484	4021952563	4365177082	3708920048	6119690446	2645747774

4205082341	1547445266	9527079953	2222206413	9455728573	6769654367
2870725815	4248116213	9734408721	0720731790	2352378317	7320869637
5192433729	6539459593	4258260527	5936283848	8239610118	3321159466
5462244431	9119042592	9292745973	1686846767	0307112059	2570146670
6879359141	6262522966	5522825626	0449352494	7524633824	4586251025
4605685236	0139092286	7728144077	6196279335	6533712472	3217900597
5469282391	2328729529	6923461406	7797450024	9010339333	1956541430
1302124892	7856520106	4515514938	9391083647	7061342941	9486431994
6737925241	0556707007	6674317157	2011745204	1595660000	1874392423
0175875379	4041921585	6667436806	1947607246	4366794543	5904790033
3616810851	3488881553	0154035456	8539411838	9808624626	4524026404
9716896338	3318516232	4194150949	8496285207	8095100406	9636270774
2062669541	7975249140	7196126296	0501451176	1663332537	9614506571
1499906696	3909473407	3544131880	8943548581	8869541994	3754673043
2015123367	2501625296	9462461171	6985102591	7485220539	0036759579
3101024674	0545561427	7793691936	7402943902	7755732270	9779017119
5417845611	8099337143	0533512969	1166449663	5207964627	5936171539
4632477929	3124964710	0229536870	6907494138	6763791976	3556404401
5252756021	6061451748	0918826097	5612719255	3604090324	9004565497
0997333440	8846123356	7318950207	5232307546	1502009994	7576676490
1051621615	0184876938	5401644056	8253952704	4220863046	3369673745
5196150654	9493881997	9187076150	4767726269	6229064464	2712467018
2097181749	9042912272	9537505871	6628131003	0068227398	2071453295
6427858044	0835869910	7854242785	6847664659	2830602264	8133310591
6647664659	2630602264	8133310591	4136182760	5384086233	8159413628
9382343178	9175753741	6161362269	0770617813	8941592694	0039756391
1366156673	0461897553	3122308420	4051007693	3260460475	9411901640
5121590290	2646668795	7776220791	5026390212	5576176514	6346347055
5026390112	5576176514	6348347055	1260717646	4894987230	6945413740

Примечание. Приведенные в таблице цифры можно рассматривать как реализации независимых и одинаково распределенных случайных величин, принимающих значения 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 с одной и той же вероятностью, равной *0,1*. Таблица случайных чисел заимствована из книги: *Чистяков В. И.* Курс теории вероятностей. — М.: Наука, 1978.

ПРИЛОЖЕНИЕ 2

Парадоксы теории вероятностей

Парадокс игры в кости

История парадокса. С давних времен до середины Средних веков игра в кости была одним из самых популярных и доступных развлечений — азартных игр. Понятие «азарт» произошло от арабского alzar, что означает «игральная кость». Согласно греческой легенде, игру в кости впервые предложил Паламедей для развлечения греческих солдат, ожидавших знаменитой Троянской битвы.

Самой ранней книгой по теории вероятностей считается «Книга об игре в кости» итальянского математика Джероламо Кардано (1501—1576).

В начале XVII в. знаменитый Галилео Галилей (1564—1642) написал на тему парадокса игры в кости трактат «Об открытиях, совершенных при игре в кости».

Суть парадокса. Назовем игральную кость *правильной*, если сумма очков противоположных граней равна 7. Тогда падение кости (кубика) на грань с цифрой «2» одновременно означает выпадание пятерки. Очевидно, что при бросании симметричного кубика равновероятны выпадания любой из граней 1, 2, 3, 4, 5 или 6.

При выбрасывании двух игральных костей сумма выпавших чисел на гранях заключается между числами 2 и 12. В этом случае каждое из чисел 9 и 10 можно получить двумя способами: $9 = 3 + 6 = 4 + 5$, а $10 = 4 + 6 = 5 + 5$. С другой стороны, при выбрасывании трех костей и 9, и 10 могут получаться шестью способами.

? Почему при выбрасывании двух костей 9 встречается чаще, чем 10, а при выбрасывании трех костей 10 выпадает чаще, чем 9?

Объяснение парадокса. Для того чтобы все исходы были равновозможны, необходимо учесть порядок выпадания чисел (например, на первой или на второй кости появилось некоторое число). Тогда при бросании двух кубиков сумма, равная 9, выпадает в наборах: $9 = 6 + 3 = 3 + 6 = 5 + 4 = 4 + 5$, т.е. четыре раза, а сумма, равная 10, выпадает три раза: $10 = 6 + 4 = 4 + 6 = 5 + 5$. При выбрасывании двух костей всего могут быть $6^2 = 36$ различных равновозможных исхода. Тогда вероятность появления суммы, равной 9 (событие A), $P(A) = 4/36$, а вероятность появления суммы, равной 10 (событие B), $P(B) = 3/36$. Таким образом, $P(A) > P(B)$.

При выбрасывании трех костей число равновозможных различных исходов равно $6^3 = 216$. При этом сумма, равная 9, может появиться в 25 случаях, т.е. $P(A) = 25/216$, а сумма, равная 10, — в 26 случаях, т.е. $P(B) = 26/216$, т.е. $P(A) < P(B)$.

Несмотря на то, что задача о костях достаточно проста, в средние века она вызывала немало ошибок даже у таких выдающихся математиков, как Г.-В.Лейбниц (1646—1716) и Ж.Д'Аламбер (1717—1783).

В рассмотренной задаче математиков побуждало заниматься решением этого парадокса противоречие.

Противоречие: при выбрасывании двух костей шансы иметь сумму, равную 9, больше, чем сумму, равную 10, а при выбрасывании трех костей наоборот. Между тем для выпадения этих сумм необходимо иметь один из двух наборов: $9 = 3 + 6 = 4 + 5$, $10 = 4 + 6 = 5 + 5$.

Парадокс де Мере

История парадокса. Согласно рассказам Лейбница, известный французский игрок XVII в. Шевалье де Мере предложил одному из самых знаменитых ученых того периода (да и вообще, в мировой науке) Б. Паскалю решить две задачи, связанные с азартными играми.

Решение этих задач Паскаль обсуждал в письмах, адресованных в Тулузу другому выдающемуся французскому ученому П. Ферма. Эти знаменитые письма сохранились и стали достоянием истории. Так как оба ученых независимо друг от друга получили одинаковые результаты, то широко известны слова Паскаля по поводу решения этих задач: «Я вижу, что истина одинакова и в Тулузе, и в Париже».

Суть парадокса. При четырех подбросах одной игральной кости вероятность того, что единица выпадет по крайней мере один раз, *больше* $\frac{1}{2}$.

При двадцати четырех подбросах двух игральных костей вероятность выпадания одновременно двух единиц хотя бы однажды *меньше* $\frac{1}{2}$.

? Почему, несмотря на то что вероятность выпадания одной единицы в 6 раз больше, чем вероятность выпадания двух единиц, а число 24 тоже в 6 раз больше, чем 4, ожидаемая пропорциональность не проявилась в результатах подсчета вероятностей?

Объяснение парадокса. Число равновозможных исходов при условии, что игральную кость бросали по меньшей мере k раз, равно $\hat{A}_6^k = 6^k$. Пусть событие B — выпадание по меньшей мере один раз единицы при подбрасывании одной игральной кости. Число благоприятных исходов для B можно найти, «убрав» из общего числа исходов те, когда выпадают остальные пять цифр, т.е. $\hat{A}_6^k - \hat{A}_5^k = 6^k - 5^k$.

Тогда имеем $P(B) = \frac{6^k - 5^k}{6^k} = 1 - \left(\frac{5}{6}\right)^k$, т.е. при $k=4$ вероятность выпадания по крайней мере одной единицы больше половины: $P(B) = 1 - \left(\frac{5}{6}\right)^k > \frac{1}{2}$.

Действительно, $P(B) = 1 - \frac{625}{1296} = \frac{671}{1296} \approx 0{,}52 > \frac{1}{2}$.

Пусть событие C — выпадание единицы при k подбросах двух игральных костей. Число равновозможных исходов при k подбрасываниях двух костей равно 36^k, число благоприятных исходов: $36^k - 35^k$. Тогда $P(C) = \frac{36^k - 35^k}{36^k} = 1 - \left(\frac{35}{36}\right)^k$.

При $k=24$ имеем вероятность выпадания по крайней мере одной единицы меньше половины: $P(C) = 1 - \left(\frac{35}{36}\right)^k < \frac{1}{2}$.

Однако при $k = 25$ вероятность выпадания по крайней мере одной единицы уже больше половины: $P(C) = 1 - \left(\frac{35}{36}\right)^k > \frac{1}{2}$, что противоречит правилу пропорциональности, согласно которому «Если вероятность в 6 раз уменьшается, то критическое значение в 6 раз увеличивается». Тогда ожидаемое соотношение 4 : 6 = 24 : 36 нарушается.

Другой известный математик А.Муавр в книге «Доктрина шансов» (1718) доказал, что, вообще говоря, можно рассматривать «правило пропорциональности» критических значений как «асимптотически верное», а ошибка от его применения увеличивается с ростом p. Рассмотрим решение парадокса с точки зрения Муавра. Пусть p — вероятность некоторого события, например, выбросить единицу при подбрасывании одного кубика, равна $p=\frac{1}{6}$. Тогда критическое значение $k=x$ можно найти из показательного уравнения $(1-p)^x=\frac{1}{2}$, откуда $x=\log_{1-p}\frac{1}{2}=-\log_{1-p}2$. Критическое значение k является наименьшим целым числом, бóльшим x, и имеет вид:

$$x=-\frac{\ln 2}{\ln(1-p)}=\frac{\ln 2}{(p+\frac{p^2}{2}+\ldots)}.$$

Очевидно, что если p^2 пренебрежимо мaлó, то p убывает почти пропорционально возрастанию критического значения при $\frac{p^2}{2}+\frac{p^3}{3}+\ldots\to 0$. Так, Муавр использовал приближенную формулу $x\approx\frac{\ln 2}{p}\approx\frac{0,69}{p}$ в связи с Лондонской лотереей. Тогда при $p=\frac{1}{32}$ $x=22,135\ldots$, а по формуле $n=\frac{p^2}{2}$ приближенное значение $x\approx 22,08$. Таким образом можно сделать вывод о том, что парадокс де Мере возникает из-за того, что при $p=\frac{1}{6}$ слагаемым $\frac{p^2}{2}$ и другими слагаемыми в знаменателе формулы нельзя пренебречь. Решением парадокса де Мере является вывод о том, что «правило пропорциональности критических значений» оказывается асимптотически верным, т.е. с ростом p увеличивается ошибка при его применении.

Противоречие: вероятность выпадания одной единицы в 6 раз больше, чем вероятность выпадания двух единиц, причем число 24 тоже в 6 раз больше, чем 4, а ожидаемая пропорциональность не проявилась в результатах: «критическое значение» для одной игральной кости равно 4, а для двух костей — 25.

Заметим, что существуют «случайные величины», которые подчиняются «правилу пропорциональности». Так, в ядерной физике такое критическое значение называется «период полураспада», и он обратно пропорционален постоянной распада, которая соответствует p.

В теории вероятностей (и не только в ней) очень важно точно сформулировать вопрос, от которого в корне может измениться вывод.

Рассмотрим задачу, в которой случайная величина X — число выбрасываний игрального кубика до первого появления единицы. Вероятность того, что X примет значение, равное k, вычисляется по формуле $P(X=k)=\frac{(5/6)^{k-1}}{6}$. Полученная ДСВ X соответствует геометрическому распределению с вероятностью $p=1/6$, поэтому математическое ожидание величины X вычисляется по формуле $M(X)=1/p=6$. Итак, математическое ожидание величины X — число выбрасываний игрального кубика до первого появления единицы, подчиняется «правилу пропорциональности»: для получения двух единиц в среднем потребуется в шесть раз больше выбрасываний, чем для получения одной единицы.

Парадокс раздела ставки

История парадокса. Впервые парадокс был опубликован в 1494 г. в Венеции в обзоре средневековой математики в книге Л. Пачоли (ок. 1445—ок. 1514) «Сумма знаний по арифметике, геометрии, отношениям и пропорциональности». Однако известно, что задачу о разделе ставки решали еще в конце XIV в. арабские ученые. Самостоятельно и независимо друг от друга нашли верное решение задачи о разделе ставки в XVII в. знаменитые французские математики П. Ферма и Б. Паскаль. Есть мнение, что именно с решения этой задачи (1654) началось зарождение и развитие теории вероятности как ветви математики.

Суть парадокса. Два противника играют в некоторую безобидную игру, т.е. имеют одинаковые шансы победить в ней. Победителем в этой игре считается тот, кто одержал шесть побед. Однако игру прервали раньше (например, при счете 5 : 3).

? Как правильно разделить ожидаемую награду — некоторый приз за победу?

За весь период решения этой задачи предлагались разные варианты ответов. Н. Тарталья (ок. 1499—1557) предложил разделить приз в отношении 2 : 1.

Объяснение парадокса. Поэтому первый игрок выиграл на две партии больше, чем второй, а это составляет третью часть общего числа возможных побед, то победитель получает 1/3 часть приза, а остаток делится пополам.

Некоторые математики считали, что ставку надо делить в отношении 5 : 3, другие — что верный ответ 7 : 1.

Решение задачи. Для победы в игре первому игроку достаточно одной партии, а второму надо выиграть все оставшиеся три партии. В каждой из партий возможны лишь два исхода, поэтому всего будет $2^3=2\cdot2\cdot2=8$ различных вариантов, или исходов (рассматриваются все варианты, включая те, когда игра продолжается после победы первого игрока).

Так как при восьми вариантах продолжения игры второй игрок получает приз только в случае, если он выиграет все три партии, то у него 1 шанс из 8. Тогда остальные 7 шансов на победу имеет первый игрок. Поэтому правильный, т.е. справедливый, раздел ставки в отношении 7:1. Таким образом, 7/8 и 1/8 — шансы победить в игре соответственно первого и второго игроков, при остановке игры на счете 5:3.

В общем случае, если первому игроку для победы осталось сыграть n партий, а второму m партий, то шансы на получение приза первым игроком можно найти по формуле $2^{-n-m+1}\sum\limits_{i=n}^{n+m-1} C_i^{n+m-1}$, где число фиктивных партий в игре $m+n-1$, из всех 2^{n+m-1} равновозможных исходов.

? В чем же парадокс раздела ставки, в чем противоречие?

В рассмотренной задаче *раздела ставки* мы не смогли сформулировать противоречие, т.е. сама задача по сути не является парадоксом. Скорее можно назвать парадоксальным процесс ее решения, когда авторитетные математики предлагали несколько различных вариантов ее решения. Но в науке это типичная ситуация: при столкновении различных мнений побеждают математически и логически верно выстроенные доказательства.

Парадокс независимости

История парадокса. Обозначение вероятности события A символом $P(A)$ вошло в употребление в связи с тем, что у многих европейских народов с этой буквы начинается слово «вероятность» (например, в латинском — *probabilitas*).

Напомним, что по определению для независимых событий A и B справедлива симметричная формула: $P(AB) = P(A)P(B)$, т.е. условная вероятность A равна безусловной $P(A|B) = P(A)$.

С.Н. Бернштейн обратил внимание на парадокс, связанный с нарушением симметричности формулы произведения независимых событий.

Суть парадокса. Пусть бросают две правильные монеты. Обозначим:

событие A — на первой монете выпал орел;

событие B — на второй монете выпал орел;

событие C — на одной и только на одной монете выпал орел.

? Почему, несмотря на то что события A, B, C попарно независимы, любые два из них однозначно определяют третье?

Объяснение парадокса. 1) События A и B независимы, так как выпадание орла на первой монете никак не повлияло на выпадание орла на второй монете (говорят: «монета не имеет памяти»).

2) Пары событий (A; C) и (B; C) также независимы. Действительно,

$$P(AC) = P(A)P(C) = \frac{1}{4}.$$

Любая пара событий из тройки (A, B, C) определяет третье, так как каждое из них происходит тогда и только тогда, когда происходит лишь одно из двух других.

Этот парадокс подтверждает, что попарная независимость событий не означает их независимость в совокупности. Поэтому для определения независимости в совокупности накладывается требование мультипликативности.

События A_1, A_2, ..., A_n из некоторой совокупности независимы, если для любого конечного набора событий из этой совокупности справедливо свойство мультипликативности: $P(A_1, A_2, ..., A_n) = P(A_1)P(A_2) \dots P(A_n)$, т.е. вероятность осуществления совместного события равна произведению индивидуальных (маргинальных) вероятностей.

Заметим, что если события A_1, A_2, ..., A_n не являются независимыми, то можно утверждать, что выполняется неравенство:

$$-\left(1-\frac{1}{n}\right)^n \le P(A_1, A_2, \dots, A_3) - P(A_1)P(A_2)\dots P(A_n) \le (n-1)n^{-\frac{n}{n-1}}.$$

Рассмотрим задачу. Петру родители пообещали подарить на день рождения сотовый телефон, если сын дважды подряд выиграет у кого-либо из них в шахматы. Петру предложено самому выбрать порядок игр: «отец—мать—отец» или «мать—отец—мать». Пусть Петр победит отца с вероятностью p, а мать — с вероятностью q. Зная, что отец играет в шахматы лучше матери, Петр продумывает варианты стратегии. Если Петр выбирает первую стратегию, то он должен одержать победу либо в первой паре игр (с вероятностью pq), либо во второй паре — с вероятностью qp. Так как события A_1 и A_2 совместные, то $P_1 = P(A_1 \cup A_2) = pq + qp - pqp$ в первом случае и $P_2(A_1 \cap A_2) = qp + pq - qpq$ во втором случае. Но $pq + qp - pqp > qp + pq - qpq$. Итак, вероятность победы в первой стратегии выше, чем во второй, так как $p < q$.

Парадокс раздачи подарков, телефонных вызовов, лотерей, дней рождения

По капельке — море,
по былинке — стог.

Пословица

История парадокса. Одна из первых книг по теории вероятностей была опубликована в Париже в 1708 г. Автор книги Р. де Монмор включил в нее задачи на вычисление вероятности по классическому определению $P(A) = \frac{m}{n}$, где число благоприятных и равновозможных исходов вычислялось с помощью комбинаторных формул. «Парадокс раздачи подарков» — одна из парадоксальных задач, обсуждаемых на страницах этой книги.

Задача 1. Парадокс раздачи подарков

Суть парадокса. П. де Монмор предложил следующий вариант этой задачи. Компания друзей решила так себя поздравить с Новым годом: каждый приносит подарок, затем все подарки складывают вместе и перемешивают, потом случайным образом распределяют их между участниками. При таком распределении подарков подразумевается, что вероятность получения собственноручно принесенного подарка в большой компании друзей ничтожно мала.

? Почему вероятность получения собственноручно принесенного подарка («совпадения») намного больше вероятности того, что совпадений нет (кроме случая, когда группа минимальна и $n = 2$, а вероятность отсутствия совпадений равна 1/2)?

Объяснение парадокса. Рассмотрим задачи на вычисление вероятности события по классическому определению с применением комбинаторных формул.

Пусть компания из n друзей собрала для раздачи n подарков.

Подарки могут быть розданы участникам игры $n!$ различными способами, т. е. число всех равновозможных исходов $n!$ Пусть событие A заключается в том, что никто из друзей не получит свой подарок. Число исходов, в которых никто из друзей не получит свой подарок, равно

$$C_n^0 n! - C_n^1 (n-1)! + C_n^2 (n-2)! - C_n^3 (n-3)! \dots (-1)^n C_n^n 0!$$

Тогда вероятность наступления события A можно вычислить по формуле

$$p_n = P(A) = \frac{1 \cdot n!}{2} - \frac{n!(n-1)!}{n!} + \frac{n!(n-2)!}{2(n-2)^* n!} - \dots + \frac{(-1)^n}{n!}, \text{ или}$$

$$p_n = P(A) = \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!}.$$

При $n > 2$ вероятность того, что никто из друзей не получит свой подарок $P(A) < \frac{1}{2}$ (так как ряд сходится). Так, при $n \geq 6$ для компании из не менее шестерых друзей, учитывая сходимость ряда, имеем

$$P_6 = P(A_6) \approx \frac{1}{e} \approx \frac{1}{3} \approx 0,37 < 0,5.$$

Вероятность P_n сходится к e^{-1} с увеличением n. Если $n \geq 6$, то $P_n = e^{-1}$ с точностью до четырех знаков после запятой, т. е. $p_n \approx \frac{1}{e} \approx 0,3679$. При $n = k$: $P(A) = \frac{e^{-1}}{k!}$.

Вероятность того, что конкретный человек получит принесенный им подарок (вероятность совпадения), равна $P_1 = \frac{1}{n}$. При увеличении n, несмотря на то, что отдельно взятые вероятности ничтожны, т.е. $P_n = \frac{1}{n} \to 0$, вероятность противоположного события, что произойдет по крайней мере одно совпадение, равна $P(\bar{A}) = \frac{2}{3}$.

Противоречие: несмотря на то что вероятность получения собственноручно принесенного подарка в большой компании друзей ничтожно мала, вероятность того, что произойдет по крайней мере одно совпадение, равна $P(\bar{A}) = \frac{2}{3}$, т.е. больше половины.

Задача 2. Парадокс лотереи (формула Пуассона)

В новогодние праздники n одноклассников устроили лотерею: принесли n призов, которые распределили между собой так, что у всех одинаковая вероятность получить любой из призов независимо от других, т.е. может оказаться, что кому-то достанется несколько призов, а кто-то не получит их вообще.

Тогда общее число исходов распределения призов возможно $A_n^n = n^n$ способами. Обозначим через A событие — одному из одноклассников не достанется приз. Значит, эти n призов распределены между остальными $n - 1$ одноклассниками $A_{n-1}^n = (n-1)^n$ способами. Поэтому вероятность события A можно найти по формуле $q_n = P(A) = \frac{(n-1)^n}{n^n} = \left(1 - \frac{1}{n}\right)^n$. При $n \to \infty$ последовательность $\left(1 - \frac{1}{n}\right)^n$ сходится к e^{-1} (второй замечательный предел). В итоге вероятность того, что конкретный одноклассник получит ровно k призов, сходится к $\frac{e^{-1}}{k!}$ при $n \to \infty$.

Пусть число людей (n) и число призов (m) не совпадают ($n \neq m$). Тогда вероятность того, что призов не достанется, равна $q_n = P(A) = \left(1 - \frac{1}{n}\right)^m$.

Обозначим $\frac{m}{n} = \lambda$, где параметр λ — среднее число призов, приходящихся на одного человека, причем $\lambda > 0$. Тогда вероятность q_n того, что конкретный человек не получит призов, сходится к $e^{-\lambda}$. В итоге вероятность того, что конкретный одноклассник получит ровно k призов, вычисляется по формуле $p_k = P(n = k) = \frac{\lambda^k e^{-\lambda}}{k!}$.

Случайная величина, рассмотренная в этой задаче, которая принимает лишь неотрицательные целочисленные значения $n = k$ с вероятностью p_k, имеет *распределение Пуассона* («закон редких явлений»).

Поэтому в парадоксе «раздачи подарков» получаем, что число друзей, которым достанутся собственноручно принесенные подарки, тоже подчиняется распределению Пуассона с параметром $\lambda = 1$ (так как вероятность получить собственный подарок для одного человека $p = \frac{1}{n}$, для n человек $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} \frac{1}{n} = 1$ при любом числе n).

Исторически понятие «закон редких явлений» впервые появилось в работе «Исследования о вероятности судебных приговоров по уголовным и гражданским делам» в 1837 г. французского ученого С. Пуассона (1781—1840). Книга была посвящена приложениям теории вероятностей к судебной практике.

Известно, что вероятность числа успехов при биномиальных распределениях вычисляется по формуле Бернулли:

$$P_n = C_n^k p^k (1-p)^{n-k}.$$

Прилагательное «биномиальное» указывает на то, что P_k есть k-й член в разложении бинома $(p+q)^n$.

Распределение Пуассона является приближенным для биноминального распределения при малых значениях вероятности. После опубликования материалов об этом распределении «забыли» на 20 лет.

Но, начиная с 1894 г., это распределение стали применять при изучении трагических случаев смерти солдат в кавалерии германской армии. В соответствии с проведенными 280 наблюдениями, за 20 лет в 14 корпусах армии погибли 196 солдат от удара копытом лошади. Отсюда $\lambda = \frac{m}{n} = \frac{196}{280} = 0,7.$

Если для изучения этой трагической статистики использовать распределение Пуассона, то ряд распределений имеет вид при $\lambda = 0,7$ и $n = 280$:

$n = k$	0	1	2	...
P_k	$\frac{139}{280} \approx 0,486$	$\frac{97}{280} \approx 0,346$	$\frac{34}{280} \approx 0,121$	

В сравнении с теоретическим, статистика фактических распределений числа смертельных исходов имела вид

$n = k$	0	1	2	...
W_k	$\frac{140}{280} \approx 0,5$	$\frac{91}{280} \approx 0,325$	$\frac{32}{280} \approx 0,14$	

Л. Борткевич назвал этот закон (в 1898 г.) «закон малых чисел», так как при $n \to \infty$ $P_n \to 0$.

С помощью закона редких явлений можно приближенно описать количество товара определенного вида (неходового), проданного в магазине в течение дня, число забастовок и войн за год, число телефонных соединений с конкретным номером за день и т. д.

Если N — число телефонных линий, по которым может произойти телефонное соединение, то число занятых в данный момент линий имеет распределение Пуассона.

В 1906 г. датский математик А. К. Эрланг (1878—1929) установил, что лучшее приближение получается из усеченного пуассоновского распределения по формуле $e_k = \frac{c\lambda^k}{k!}$, где $k = \overline{0; N}$; $c = \left(\sum_{k=0}^{N} \frac{\lambda^k}{k!} \right)^{-1}$.

Этот вид распределений назвали распределениями Эрланга и используют при изучении случайных процессов.

Задача 3. Парадокс дня рождения

Суть парадокса. Пусть у меня не более 365 приятелей. Среди них, скорее всего, найдутся, по меньшей мере двое, у которых дни рождения совпадают.

Найдем число таких людей, среди которых с вероятностью 0,99 найдутся двое, у которых дни рождения совпадают. Для этого, оказывается, достаточно всего 55 человек. Среди 68 человек с вероятностью 0,999, по меньшей мере, у двоих дни рождения совпадают.

? Почему вероятность изменилась незначительно, всего на 0,1 %, а соответствующее число людей увеличилось на 13 человек, т. е. на 25 %?

Объяснение парадокса. Эта парадоксальная ситуация иллюстрирует одну из главных причин — почему теория вероятностей применяется так широко.

Пусть n — число дней в году, а x — число моих друзей, причем $x < n$. Найдем вероятность того, что ни у какой пары друзей дни рождения не совпадут (событие A). Число равновозможных исходов найдем по формуле $\hat{A}_n^x = n^x$, так как и в первый, и во второй, ... и в последний n-й день года могли родиться все x человек рассматриваемого множества.

Число благоприятных исходов есть число возможных размещений из n по x, т. е. $A_n^x = \frac{n!}{(n-x)!} = n(n-1)...(n-x+1)$.

Тогда $P(A) = \frac{A_n^x}{\hat{A}_n^x} = \frac{n(n-1)...(n-x+1)}{n^x}$.

Если p — вероятность того, что среди x друзей найдутся такие, у кого дни рождения совпадают, то $1 - p = \frac{n(n-1)(n-2)...(n-x+1)}{n^x}$.

Откуда найдем $x \approx \sqrt{2n \ln (1-p)^{-1}}$.

С другой стороны, если $p = 1$, то $x = n + 1$.

Обобщим и сделаем выводы: если в группе из x человек у k человек дни рождения совпадают с вероятностью p, то $x = cn^{\frac{(k-1)}{k}}$, где c — постоянная, вычисляемая по формуле

$$c = \left[k! \ln (1-p)^{-1}\right]^{\frac{1}{k}}.$$

Санкт-Петербургский парадокс

Все больше и больше ученых считали, что теория вероятностей — это не что иное, как путеводитель по жизни, здравый смысл, выраженный в числах.

Габор Секей

История парадокса. В XVII в. многие крупные научные журналы довольно часто публиковали статьи по «модной» в тот период науке — теории вероятностей. Так, в начале XVII в. Петербургская Академия наук опубликовала статью Д. Бернулли. Однако автор статьи приводил такие математические вычисления, которые противоречили здравому смыслу. Проблему, получившую название Санкт-Петербургского парадокса, первым стал изучать двоюродный брат Д. Бернулли Н. Бернулли, написав свое решение в письме П. де Монмору в 1713 г.

Суть парадокса. Назовем *петербургской игрой* такое подбрасывание правильной монеты, при котором победа наступает при выпадании решки. Причем, если решка выпадает при k-м бросании, то игрок получает 2^k условных денежных единиц из банка. Поэтому с каждым новым подбрасыванием монеты выигрыш удваивается.

? Сколько следует заплатить игроку за участие в игре, чтобы игра стала справедливой, т. е. чтобы математическое ожидание выигрыша (среднее значение) было равно взносу за игру? Парадокс заключается в том, что это требование невыполнимо при любой (конечной) сумме, которую может заплатить игрок.

Объяснение парадокса. Рассмотрим потери банка. Так как вероятность окончания игры при k-м бросании равна $\frac{1}{2^k}$, а игрок в этом случае получает 2^k усл. ед., математическое ожидание потерь банка составляет бесконечно большое число.

Ряд распределений случайной величины k бросаний (или случайной величины 2^k — потерь банка) при $k \to \infty$ имеет вид, представленный таблицей:

k	1	2	3	4	...
2^k	2	4	8	16	...
p_k	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...

Тогда банк в среднем должен заплатить $M(X)=\frac{1}{2}\cdot 2+\frac{1}{4}\cdot 4+\frac{1}{8}\cdot 8+...=$ $=1+1+...$ усл. ед., что соответствует бесконечно большому числу условных единиц, т.е. $M(X)$ не существует.

Тогда известные математики Ж. Бюффон и Г. Крамер предложили исходить из предположения об ограниченности ресурсов, которыми пополняется банк.

Пусть в банке 1 млн усл. ед. Тогда можно указать конечное математическое ожидание выигрыша. Составим новую сумму, заменив в ней 2^{20} на 10^6. Имеем:

$$M(X)=\frac{1}{2}\cdot 2+\frac{1}{4}\cdot 4+...+\frac{1}{2^{10}}\cdot 2^{10}+\left(\frac{1}{2^{20}}+\frac{1}{2^{21}}+...\right)\cdot 10^6=$$
$$=19+1{,}9...\approx 21 \text{ усл. ед.}$$

(При вычислении использовали неравенство $2^{20}>10^6$.) Тогда, если вступительный взнос игрока равен 21 усл. ед., то игра становится выгодной для банка.

Однако В. Феллер предложил вариант взноса, когда петербургская игра становится справедливой. Пусть игрок участвовал в n играх. Будем считать игру безобидной, если отношение суммарного выигрыша N_n к суммарному вступительному взносу R_n при $n\to\infty$ сходится к единице, т.е. если $\forall\varepsilon>0$, то при $n\to\infty$ выполняется $P\left\{\left|\frac{N_n}{R_n}-1\right|<\varepsilon\right\}\to 1$.

Так, В. Феллер доказал, что приняв $R_n=n\log_2 n$, игра может стать безобидной. Последнее соотношение свидетельствует об устойчивости величины N_n. Из условия игры она не может быть справедливой для $R_n=cn$, где c — произвольная конечная константа.

Согласно теореме В. Феллера, если вступительный взнос будет зависеть от числа игр, в которых участвовал игрок, то петербургский парадокс разрешен.

Так Ж. Бюффон после проведения 2084 игр установил, что игра становится справедливой, если вступительный взнос составляет около 10 усл. ед.

Парадокс смертности населения

История парадокса. Проблемы страховых компаний при страховании морских судов, а также людей, возникшие в связи с развитием мореходства в XVII в., послужили причиной математических исследований во вопросу продолжительности жизни населения. Э. Галлей (открывший известную комету, названную в его честь) опубликовал в 1693 г. статью, с которой берет начало математическая теория страхования жизни. На парадокс смертности населения впервые обратил внимание математик Ж. Д'Аламбер.

Суть парадокса. Согласно таблицам смертности Галлея, средняя продолжительность жизни составляла в те времена 26 лет.

? Однако почему равновероятны события: «прожить до восьми лет» и «прожить больше восьми лет»?

Объяснение парадокса. Согласно таблицам смертности Галлея, шансы умереть до восьми лет и прожить больше восьми лет равны. Если человек прожил больше 8 лет, то его жизнь может продолжаться еще несколько десятилетий. Поэтому не удивительно, что в тот период математическое ожидание (средняя продолжительность жизни) составляла 26 лет.

Поскольку продолжительность жизни непрерывная случайная величина, то $M(X)=\int\limits_0^\infty xf(x)\,dx$. С другой стороны, вероятную продолжительность жизни m , где m — медиана, т.е. возраст, до которого доживает человек с вероятностью 50 %, можно найти, решив уравнение $F(m)=\frac{1}{2}$. Тогда за период времени m вымирает половина всего населения. Напомним, что $M(X)$, вообще говоря, не должно быть равно m, так как $M(X)$ — математическое ожидание продолжительности жизни, а m — медиана.

Гибель человека может быть перенесена на гибель некоторого объекта (или системы) в широком смысле этого понятия: распад атомов, амортизация промышленного оборудования, т.е. гибель как смерть объекта исследования.

Назовем объект безвозрастным, если вероятность его существования в течение определенного продолжительного временно́го промежутка не зависит от того времени, которое объект уже существовал. Заметим, что человек не обладает такими качествами. Чем больше он живет, тем более вероятна его последующая гибель.

Пусть средняя продолжительность существования безвозрастного объекта равна T. Тогда вероятность того, что он за следующий период времени x будет продолжать существовать, равна $e^{\frac{-x}{T}}$, где $x>0$.

Как известно, скорость распада радиоактивных частиц пропорциональна числу нераспавшихся частиц, где коэффициент пропорциональности λ есть постоянная распада, т.е. $T=\frac{1}{\lambda}$. Тогда продолжительность существования радиоактивных частиц можно описать с помощью показательного распределения с параметром λ. Напомним, что продолжительность существования полураспада имеет плотность вероятности

$f(x) = \lambda e^{-\lambda x}$. Так как мы ищем период *полураспада* безвозрастного объекта, решаем уравнение $e^{-\lambda x} = \frac{1}{2}$, откуда имеем $x = \frac{\ln 2}{\lambda}$.

Понятие периода полураспада безвозрастного объекта широко распространено в различных областях знаний. Так, за радиоуглеродный метод, который основан на этих идеях и применяется для хронологического датирования археологических находок, его автору — американскому химику У. Либби в 1960 г. была присуждена Нобелевская премия.

Аналогичные идеи в лингвистике предложил применять М. Свадеш в 1950 г. Согласно его предположению, слова можно рассматривать как «атомы речи» — тоже своего рода безвозрастные объекты. Тогда период полураспада древнего базового словаря языков составляет 2 тыс. лет. Благодаря этой идее, в лингвистике устанавливают дату, когда исторически произошло разделение языков (например, латинский и санскрит, русский и украинский, венгерский и финский и т.д.). Этот метод получил название «лексико-статистика» или «глоттохронология». В частности, для определения даты разделения языков достаточно знать, какая часть базового словаря сохранилась в обоих словарях по настоящее время. Так, было установлено, что базовая часть венгерского и финского словаря составляет соответственно 21 и 27 %. Отсюда был сделан вывод о том, что эти языки разделились около 4—5 тыс. лет назад.

Пусть параметр λ — постоянная распада. Тогда по формуле Пуассона можно найти вероятность распада ровно k частиц за время t: $P(x = k) = (\lambda e)^k \frac{e^{-\lambda t}}{k!}$, т.е. для случайной величины X — числа распавшихся частиц — математическое ожидание $M(X) = \lambda t$.

? Можно ли с помощью описанного метода оценить общее количество людей, когда-либо живших на Земле?

Благодаря вероятностным методам удалось установить, что в настоящее время на земле живет около 9 % общего количества людей, когда-либо живших на Земле.

Парадокс закона больших чисел Бернулли

История парадокса. Я. Бернулли в книге «Искусство предположений», опубликованной уже после его смерти в 1713 г., доказал одну из теорем, которую впоследствии (1837 г.) Пуассон назвал «законом больших чисел». Парадокс связан с теоремой, согласно которой при выбрасывании правильной монеты $\exists \varepsilon > 0,\ \delta > 0,\ P\left\{\left|\frac{k}{n} - \frac{1}{2}\right| < \varepsilon\right\} > 1 - \delta$. Таким образом, при достаточно больших значениях n, зависящих от ε и δ, модуль разности относительной частоты выпадания орла k/n и $1/2$ меньше ε с вероятностью, превосходящей величину $1 - \delta$.

Суть парадокса. Известно, что при n-кратном выбрасывании правильной монеты орел выпадает ровно k раз, и с увеличением числа выбрасываний при достаточно больших значениях n, согласно формуле Бернулли, относительная частота выпадания орла k/n стремится к $1/2$. Поэтому есть надежда, что если многократно выпадал орел, то шансы выпадания решки при достаточно больших значениях n тоже возрастут.

? Почему шансы выпадания орла для монеты остаются равными $\frac{1}{2}$, даже если в эксперименте подряд тысячу раз выпала решка?

Объяснение парадокса. Согласно закону Бернулли, при большом числе бросаний правильной монеты, выпадание орла и решки приблизительно равновероятны, т.е. отношение числа выпаданий орла к общему числу бросаний (с вероятностью, близкой к единице) *приблизительно* равно $\frac{1}{2}$. Важен глубинный смысл слова «приблизительно»: разность логарифмов этих чисел при увеличении числа бросаний стремится к нулю.

Пусть монету бросают n раз.

? Какой максимальной длины может быть серия, состоящая из одних орлов?

При 100 бросаниях может выпасть подряд шесть-семь орлов, при $n = 1000$ — девять-десять орлов подряд, при $n = 1000000$ — 19—20 орлов.

n	100	1 000	1 000 000
e	6—7	9—10	19—20

П. Эрдеш и А. Реньи доказали следующую теорему: при бросании монеты n раз при $n \to \infty$, серия из орлов длиной $\log_2 n$ наблюдается с вероятностью, стремящейся к 1.

Эта теорема может быть использована при анализе информации, представленной последовательностью из двух символов, когда необходимо установить, является она случайной или нет.

Закон больших чисел Бернулли можно сформулировать, используя понятие сходимости по вероятности: последовательность случайных величин X_1, X_2, ... сходится по вероятности к случайной величине X, если вероятность события $|X_n - X| > \xi$ сходится к нулю для $\forall\varepsilon > 0$. Согласно закону Бернулли, относительная частота $\frac{k}{n}$ события сходится по вероятности к вероятности этого события.

Эта теорема была доказана в подразд. 2.10.4 с помощью неравенства $P(|X - M(X)| > \xi) \le \frac{D^2}{\xi^2}$, которое одновременно в одном и том же французском журнале в 1867 г. опубликовали независимо друг от друга русский математик П. Л. Чебышев и француз Ж. Бьенеме: $P(|X - M(X)| > \xi) \le$

$\leq \frac{D^2}{\xi^2}$, и которое называется «неравенство Чебышева» (см. подразд. 2.12.2).

Парадокс Муавра

История парадокса. Одним из самых важных открытий в теории вероятностей (да и вообще в математике) является открытие *нормального распределения* А. Муавром. Ученому миру Муавр сообщил об этом открытии в третьем издании книги «Доктрина шансов» в 1756 г., хотя некоторых друзей посвятил в особенности нормального распределения еще в 1733 г.

Суть парадокса. Согласно закону больших чисел Бернулли, вероятности появления орла и решки приблизительно равны, а *их отношение при увеличении числа испытаний стремится к 1.* Однако, одновременно, вероятность того, что *число орлов в точности равно числу решек, стремится к нулю.*

Пусть m — число выпадания орлов при n выбрасываниях правильной монеты. Тогда вероятность выпадания герба $p = \frac{m}{n}$.

В общем случае при выбрасывании монеты $2n$ раз вероятность того, что герб выпадет ровно n раз, равна $P_{2n}(n) = C_{2n}^n \left(\frac{1}{2}\right)^{2n}$. Для достаточно больших значений n вероятность $P_{2n}(n) \approx \frac{1}{\sqrt{\pi n}}$, т.е. с ростом n вероятность стремится к нулю.

? Почему вероятность того, что число орлов приближенно равно числу решек, стремится к единице, а вероятность того, что число гербов в точности совпадает с числом решек, стремится к нулю?

Объяснение парадокса. Объяснение этой парадоксальной ситуации нашел А. Муавр. Пусть $\frac{H_n}{T_n}$ — число выпадания орлов (решек) при n бросаниях монеты. Тогда, согласно закону больших чисел Бернулли (см. подразд. 2.10.4), вероятность того, что разность $|H_n - T_n| \to 1$ при $n \to \infty$ становится пренебрежимо малой по сравнению с n. Но если модуль разности $|H_n - T_n|$ пренебрежимо мал по сравнению с n, то этого нельзя сказать о $\sqrt{n}$.

Так он вычислил, что при $n = 3600$ вероятность того, что $|H_n - T_n|$ не превосходит 60, равна $P\{|H_n - T_n| < 60\} = 0{,}682688\ldots$.

Обозначим через $A_n(x)$ вероятность того, что $P_n(|H_n - T_n| < x\sqrt{n})$ для $\forall\varepsilon > 0$. С ростом n, согласно Муавру, $A_n(x) \to A(x)$, причем $0 < A(x) < 1$.

При увелечении x от 0 до ∞, $A(x)$ возрастает от 0 до 1. Дело в том, что согласно Муавру, $A(x)$ имеет вид $A(x) = \sqrt{\frac{\pi}{2}}\int_0^x e^{-\frac{u^2}{2}}du$.

Тогда при $x \to 0$ последнюю формулу можно записать в виде $\lim\limits_{x\to 0} P(|H_n - T_n| < x\sqrt{n}) = A(x)$, или при $x>0$ $\lim\limits_{x\to 0} P(H_n - T < x\sqrt{n}) = \Phi(x)$, где $\Phi(x)$ — функция Лапласа. Муавр доказал, что величина $|H_n - T_n|\sqrt{n}$ при достаточно больших значениях n подчиняется стандартному нормальному закону распределения.

Но так как $H_n + T_n = n$, то полученный вывод можно переформулировать: $\lim\limits_{n\to\infty} P\left(H_n < \frac{n}{2} + \frac{x\sqrt{n}}{2}\right) = \Phi(x)$.

Кроме случая с правильной монетой, Муавр рассмотрел вариант неправильной монеты, т.е. со смещенным центром, и вероятности выпадания орла и решки различны: $p \neq 1-p$. В этом случае справедлива формула $\lim\limits_{n\to\infty} P(H_n < np + x\sqrt{np(1-p)}) = \Phi(x)$, которая была использована в подразд. 1.12 («интегральная», или «предельная», теорема Муавра—Лапласа). Эта теорема широко применяется в экономике, например при планировании.

Задача. Пусть автохозяйство обслуживает 300 машин, из них 70 % находится в рабочем состоянии, а 30 % — ремонтируется. Тогда в среднем необходимо обеспечить 210 автомобилей водителями. Но может возникнуть ситуация, когда все 300 машин автохозяйства могут работать.

Какое количество водителей должно быть в этом автопарке, чтобы обеспечить эксплуатацию 300 машин с вероятностью 99,9 %?

Будем понимать под H_n число работающих машин из всех $n = 300$. Вероятность работы одной машины $p = 0{,}7$.

Тогда по табл. П. 3 найдем значение интеграла Лапласа при $x = 3$: $\Phi(3) = 0{,}9987$. Определим соответствующее значение выражения $np + x\sqrt{np(1-p)} = 300 \cdot 0{,}7 + 3\sqrt{300 \cdot 0{,}7 \cdot 0{,}3} = 210 + 3\sqrt{63} \approx 234$ машины.

Парадокс Байеса

История парадокса. Знаменитая формула Байеса (см. подразд. 1.9), опубликованная после смерти Т. Байеса, дает возможность найти «апостериорные» вероятности H_i (после того как событие A произойдет) по «априорным» вероятностям (до того, как событие A произошло), т.е. по вероятностям гипотез H_i найти $P(H_i)$.

Если рассматривать события H_i как причины, а события A, при условии, что H_i произошло, как следствие, то теорема Байеса есть теорема о вероятности причин.

Однако чаще всего априорные вероятности заранее неизвестны, и тогда гипотезы H_i рассматривают как равновероятные. Очевидно, такой подход не всегда справедлив. Сам Байес применял свою теорему для равномерного на интервале $(0,1)$ распределения априорных вероятностей. Согласно его теореме, если в $n+m$ наблюдениях событие, имеющее некоторую неизвестную вероятность p, произошло n раз, то вероятность того, что p принадлежит интервалу $(a,b) \subset (0,1)$ равна отношению определенных интегралов $\int_a^b x^n(1-x)^m dx \Big/ \int_0^1 x^n(1-x)^m dx$.

Идея Байеса заключалась в том, что если нет предварительной информации о параметре p, то априорная плотность вероятности параметра p равномерна на всем интервале $(0,1)$. Так, если $m=0$, $n=1$, $a=\frac{1}{2}$, $b=1$, то, согласно этой формуле, шансы того, что искомая вероятность $p>\frac{1}{2}$, равны $\frac{3}{4}$.

Отсутствие априорной информации о распределении вначале не убедило математиков в возможности применять формулу Байеса.

Однако в середине XX в. такой подход получил широкое распространение. Математики увидели в нем возможность последовательного применения формулы Байеса для пересчета апостериорных вероятностей после каждого наблюдения для того, чтобы на следующем этапе использовать их в качестве априорных.

Суть парадокса. Пусть возможные значения СВ X — целые числа, а распределение вероятностей p для СВ X принадлежит отрезку $[a,b]$. Поскольку независимые наблюдения X_1, X_2, X_3, ... получены из неизвестного распределения X (с неизвестным параметром p), т.е. X_i имеют то же распределение, что и X, то можно сказать, что последовательность апостериорных распределений, вычисляемых по исходному равномерному априорному распределению концентрируется около истинного значения p. Однако парадокс заключается в том, что апостериорное значение p не всегда совпадает с истинным.

Например, истинное значение вероятности $p=\frac{1}{4}$, а на самом деле после опыта пересчет апостериорной вероятности по формуле Байеса дает результат в окрестности $p=\frac{3}{4}$.

? Почему, несмотря на то что наибольшее значение функции апостериорной плотности вероятности «должно» находиться в окрестности истинного значения $p=\frac{1}{4}$, ожидаемый результат не получен?

Объяснения парадокса. На самом деле противоречия нет и функция апостериорных плотностей может концентрироваться около $p=\frac{3}{4}$ в том случае, если она при $p=\frac{1}{4}$ имеет максимум, а затем убывает, оставаясь при $p=\frac{3}{4}$ достаточно высокой. Однако если число возможных значений СВ X конечно, то такая ситуация невозможна.

Если же под X понимаем любые целые числа, то несмотря на свою парадоксальность, ситуация возможна.

Пусть на некотором отрезке $\left[\frac{1}{8}, \frac{7}{8}\right]$ равномерно распределены априорные вероятности p. Пусть также на этом отрезке задана некоторая функция $f(p)$, которая принимает на нем целочисленные значения за исключением точек $p=\frac{1}{4}$ и $p=\frac{3}{4}$, в которых $f\left(\frac{1}{4}\right)=f\left(\frac{3}{4}\right)=+\infty$.

Пусть распределение СВ X задается формулой $P(X=i)=c(1-p)p^i$, где $i=0, 1, 2, ..., f(p)$, причем c зависит от p и является постоянной, для которой справедливо $\sum\limits_{i-1}^{f(p)} c(1-p)p^i=1$. В этом случае при соответствующем выборе $f(p)$ осуществима эта парадоксальная ситуация.

Парадокс событий, происходящих почти наверняка

Пусть события A и B происходят с вероятностью 0,99 и 0,9999 соответственно. Хотя $0{,}99 \neq 0{,}9999$, можно утверждать, что эти события происходят практически наверняка, т. е. их вероятности одинаковы.

Но в отдельных случаях разница в вероятностях ощутима. Рассмотрим независимые события, которые могут происходить в любой день этого года с вероятностью $p=0{,}99$. Тогда вероятность того, что эти события будут происходить ежедневно, меньше, чем $p=0{,}03$, а если вероятность события, происходящего в единичном случае, $p=0{,}9999$, то вероятность того, что это событие будет происходить ежедневно, $p=0{,}97$.

Парадокс вероятности и относительной частоты

Известный математик Д. Пойа в книге «Математика и правдоподобные рассуждения» приводит такой пример: «Вы очень больны» — сообщил доктор Тел пациенту. — «Из десяти человек с такой болезнью выживает только один». Заметив, как был испуган пациент, доктор его успокаивал: «Но Вам очень повезло, что Вы пришли ко мне, сэр. У меня уже умерли от этой болезни девять пациентов, так что Вы выживите».

Алгоритмы решения ключевых задач

Комплексные умения и алгоритмы к гл. 1 «Основные понятия и теоремы теории вероятностей»

№ п/п	Умения	Алгоритмы
1	Вычисление числа соединений — вариантов различных подмножеств (выборок) для конечных множеств	1. Установить количество элементов n всего множества и количество элементов m его подмножества. 2. Определить, влияет ли порядок расположения элементов в подмножестве на число вариантов различных подмножеств, состоящих из этих m элементов. 3. Выбрать, в зависимости от конкретного случая, комбинаторную операцию: а) если число комбинаций всего множества зависит от порядка расположения элементов в нем и нет повторяющихся элементов, то *перестановки без повторений* $P_n = n!$; б) если число комбинаций всего множества зависит от порядка расположения элементов в нем и есть повторяющиеся элементы, то *перестановки с повторениями* $\tilde{P}_{n_1, \ldots, n_k} = \dfrac{(n_1 + n_2 + \ldots + n_k)!}{n_1!\, n_2! \ldots n_k!}$; в) если число комбинаций в подмножестве (выборке) зависит от порядка расположения элементов в нем и нет повторяющихся элементов, то *размещения без повторений* $A_n^m = \dfrac{n!}{(n-m)!}$; г) если число комбинаций в подмножестве (выборке) зависит от порядка расположения элементов в нем и элементы повторяются, то *размещения с повторениями* $\tilde{A}_n^m = n^m$; д) если число комбинаций в подмножестве (выборке) не зависит от порядка расположения элементов в нем и нет повторяющихся элементов, то *сочетания без повторений* $C_n^m = \dfrac{n!}{m!(n-m)!}$; е) если число комбинаций в подмножестве (выборке) не зависит от порядка расположения элементов в нем и есть повторя-

№ п/п	Умения	Алгоритмы
		ющиеся элементы, то *сочетания с повторениями* $\hat{C}_n^m = C_{n+m-1}^m$
2	Вычисление вероятностей событий по определению	1. Ввести обозначения для заданных величин и вопроса задачи. 2. Выбрать формулу вероятности, соответствующую данному случаю: а)*классическое определение*: если задано общее число равновозможных исходов n и число исходов m, благоприятствующих событию A (которые можно сосчитать), то находим вероятность по формуле $P(A)=\dfrac{m}{n}$; б) *геометрическое определение*: если все возможные исходы можно изобразить с помощью геометрической фигуры (отрезок, круг, полоса, куб и др.– как полное пространство элементарных событий Ω), то надо: • нарисовать эту фигуру, соответствующую полному пространству элементарных исходов Ω; • внутри нее нарисовать фигуру, соответствующую исходам, благоприятствующим событию A; • вычислить площади фигур A и Ω; • найти вероятность как отношение этих площадей по формуле $P(A)=\dfrac{S(A)}{S(\Omega)}$
3	Вычисление вероятностей событий по известным вероятностям других событий, с ними связанных	1. Обозначить все события, указанные в задаче. Известные вероятности представить в виде дроби. 2. Установить связи между событиями. 3. Вычислить требуемые вероятности, используя теоремы сложения и умножения вероятностей, а также формулу для вычисления противоположного события. Если надо вычислить вероятность того, что событие произойдет: а) *не менее, чем k раз*, то находят $P(m \geq$ $\geq k)=\sum\limits_{m=k}^{n} p_n(m)$;

Продолжение табл.

№ п/п	Умения	Алгоритмы
		б) *«хотя бы один раз»* или *«не менее одного раза»*, то определяют $P(0 < m \le n) = 1 - P_n(0)$ (событие, противоположное тому, что A не произошло ни разу); *«хотя бы два раза»* — $P_n(2 \le m \le n) = 1 - (P_n(0) + P_n(1))$; в) *не более чем k раз*, то вычисляют $P(m \le$ $\le k) = \sum\limits_{m=0}^{k} p_n(m)$; г) *более, чем k раз*, то находят $P(m > k) =$ $= \sum\limits_{m=k+1}^{n} p_n(m)$; д) *менее, чем k раз*, то вычисляют $P(m < k) =$ $= \sum\limits_{m=0}^{k=1} p_n(m)$
4	Вычисление вероятностей событий в зависимости от числа различных подмножеств конечных множеств (различных соединений)	1. Обозначить все события, указанные в задаче. 2. Вычислить число равновозможных исходов n и число исходов m, благоприятствующих событию A, пользуясь комбинаторными операциями по алгоритму 1, а также правилами суммы и произведения. 3. Найти формулу вероятности для данного случая, пользуясь *классическим определением* по формуле $P(A) = \dfrac{m}{n}$
5	Вычисление вероятности события A по формуле полной вероятности. Вычисление вероятности одной из гипотез по формуле Байеса	1. Дать описание всех гипотез $H_1, H_2, \ldots, H_n$, на которые можно разбить пространство элементарных исходов и события A. 2. Вычислить вероятность каждой гипотезы $P(H_1), P(H_2), \ldots, P(H_n)$. 3. Вычислить условную вероятность события A по каждой гипотезе $P(A\|H_1), P(A\|H_2), \ldots,$ $P(A\|H_n)$. 4. Вычислить вероятность события A по формуле полной вероятности: $$P(A) = \sum P(H_i)/P(A\|H_i).$$ 5. Вычислить вероятность гипотезы H_i при условии, что событие A произошло, по формуле Байеса:

№ п/п	Умения	Алгоритмы
		$$P(H_i \vert A) = \frac{P(H_i)P(A \vert H_i)}{P(A)}$$
6	Вычисление вероятностей числа m успехов в n независимых повторных испытаниях (биномиальные распределения) по формуле Бернулли, если надо найти точное значение m, где $n < 10$	1. Ввести обозначения для заданных величин: числа испытаний, числа успехов, вероятности наступления события A, и выписать их значения. Выписать формулу для искомой вероятности, придерживаясь общепринятых обозначений: • n — число испытаний при $n < 10$; • m — число успехов наступлений события A; • p — вероятность наступления события A в единичном испытании; • $q = 1 - p$ (вероятность неудачи); • $P_n(m)$ — вероятность наступления события A m раз в n испытаниях. В колонке «Конкретное соответствие» выписать заданные в задаче значения n, m и p. 2. Сосчитать вероятность. Если требуется найти вероятность того, что событие произошло: а) *ровно m раз*, то надо пользоваться формулой Бернулли для биномиальных распределений: $$P_n(m) = C_n^m p^m q^{n-m} = \frac{n!}{m!(n-m)!} p^m q^{n-m};$$ б) *не менее, чем k раз*, то находят $P(m \geq k) = \sum_{m=k}^{n} p_n(m)$; по алгоритмам 3а и 5а; в) «*хотя бы один раз*» или «*не менее одного раза*», то определяют $P(0 < m \leq n) = 1 - P_n(0)$ (событие, противоположное тому, что A не произошло ни разу); г) *хотя бы два раза*, то находят $P_n(2 \leq m \leq n) = 1 - [P_n(0) + P_n(1)]$ и т.д., по алгоритмам 3б и 5а
7	Вычисление вероятностей числа m успехов в независимых повторных испыта-	Ввести обозначения для заданных величин, используя алгоритм 5. Вычислить вероятность по формуле Пуассона

№ п/п	Умения	Алгоритмы
	ниях n (биномиальные распределения) по формуле Пуассона, если вероятность p наступления события A мала, а n велико и $\lambda = np < 10$	а) $p_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}$, б) $p_n(m \le k) = e^{-\lambda} \sum_{m=0}^{k} \frac{\lambda^m}{m!}$, используя для вычислений табл. 1, а также алгоритмы 3а и 3б. Вычисления можно выполнить на МК
8	Вычисление вероятностей для числа m успехов в n независимых повторных испытаниях, если n велико и $np > 10$, когда надо найти для m: а) конкретное значение вероятности; б) вероятность попадания в интервал $[m_1, m_2]$	1. Ввести обозначения для заданных величин, используя алгоритм 5. 2. а) Вычислить вероятность по формуле Муавра—Лапласа: $P_n(m) \approx \frac{1}{\sqrt{npq}} \varphi(x)$, где $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$; $x = \frac{m-np}{\sqrt{npq}}$, функцию $\varphi(x)$ определяют по табл. П. 2, причем $\varphi(x) = \varphi(-x)$; б) вычислить вероятность, используя интегральную формулу Лапласа: $P_n(m_1 \le m \le m_2) = \Phi(x_2) - \Phi(x_1)$, где $x_1 = \frac{m_1 - np}{\sqrt{npq}}$; $x_2 = \frac{m_2 - np}{\sqrt{npq}}$; $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ и $\Phi(x)$ — функция Лапласа (см. табл. П. 3), причем $\Phi(-x) = -\Phi(x)$

Задачи на применение алгоритмов

Алгоритм 1.

Вычисление числа соединений — вариантов различных подмножеств (выборок) для конечных множеств

Задача 1а. В футбольном турнире участвовали команды пяти факультетов. Найти число вариантов возможного распределения мест между ними.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Установить количество элементов n всего множества и количество элементов m его подмножества	Множество состоит из пяти элементов $n = 5$, подмножество в условии не рассматривается
2	Определить, влияет ли порядок расположения элементов в подмножестве на число вариантов различных подмножеств	В задаче требуется найти различные варианты распределения мест между командами, т. е. порядок расположения элементов важен
3	Выбрать в зависимости от конкретного случая комбинаторную операцию: а) если число комбинаций всего множества зависит от порядка расположения элементов в нем и нет повторяющихся элементов, то использовать формулу *перестановок без повторений* $P_n = n!$	Так как число комбинаций всего множества зависит от порядка расположения элементов в нем и нет повторяющихся элементов, выбираем формулу перестановок без повторений $P_n = n!$ Имеем: $P_5 = 5! = 120$ вариантов

Задача 1б. Найти число вариантов распределения призовых мест на футбольном турнире между пятью командами факультета.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Установить количество элементов n всего множества и количество элементов m его подмножества	Множество состоит из пяти элементов $n = 5$, подмножество призовых мест $m = 3$
2	Определить, влияет ли порядок расположения элементов в подмножестве на число вариантов различных подмножеств, состоящих из этих m элементов	В задаче требуется найти различные варианты распределения призовых мест между командами, т. е. порядок расположения элементов в подмножестве призовых мест важен

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
3	Выбрать в зависимости от конкретного случая комбинаторную операцию: в) если число комбинаций в подмножестве (выборке) зависит от порядка расположения элементов в нем и нет повторяющихся элементов, то использовать формулу *размещений без повторений* $A_n^m = \dfrac{n!}{(n-m)!}$	Так как число комбинаций в подмножестве зависит от порядка расположения элементов в нем и нет повторяющихся элементов, выбираем формулу размещений без повторений $A_5^3 = \dfrac{5!}{(5-3)!} = \dfrac{1\cdot 2\cdot 3\cdot 4\cdot 5}{2!} = 60$ вариантов

Задача 1в. Сколько игр будет проведено в футбольном турнире на первенство факультета по футболу, если в нем участвуют пять команд и каждая проводит с каждым из соперников по одной игре?

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Установить количество элементов n всего множества и количество элементов m его подмножества	Множество состоит из пяти элементов $n = 5$, подмножество команд, участвующих в одной игре, — из двух элементов (два противника), т.е. $m = 2$
2	Определить, влияет ли порядок расположения элементов в подмножестве на число вариантов различных подмножеств, состоящих из этих m элементов	В задаче требуется найти различные варианты составления плана проведения такого турнира, но порядок расположения элементов в подмножестве не важен (два противника в каждой игре равноправны)
3	Выбрать в зависимости от конкретного случая комбинаторную операцию: д) если число комбинаций в подмножестве (выборке) не зависит от порядка расположе-	Так как число комбинаций в подмножестве не зависит от порядка расположения элементов в нем и нет повторяющихся элементов, выбираем формулу *сочетаний* без повторений $C_5^3 = \dfrac{5!}{3!(5-3)!} = \dfrac{1\cdot 2\cdot 3\cdot 4\cdot 5}{3!2!} = 10$ вариантов

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
	ния элементов в нем и нет повторяющихся элементов, то использовать формулу *сочетаний без повторений* $C_n^m = \frac{n!}{m!(n-m)!}$	

Задача 1г. На футбольном турнире каждой команде присваивается число — двоичный номер, состоящий из цифр 1 (победа) и 0 (поражение), заработанных в каждой проведенной игре. Найти число возможных двоичных номеров этого турнира между пятью командами факультета.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Установить количество элементов n всего множества и количество элементов m его подмножества	Множество «мест» для записи двоичного числа состоит из двух элементов $n = 2$, «подмножество» мест для записи результатов игр (победа или поражение) $m = 10$
2	Определить, влияет ли порядок расположения элементов в подмножестве на число вариантов различных подмножеств, состоящих из этих m элементов	В задаче требуется найти различные варианты распределения нулей и единиц между командами, т. е.порядок расположения элементов в подмножестве важен
3	Выбрать в зависимости от конкретного случая комбинаторную операцию: если число комбинаций в подмножестве (выборке) зависит от порядка расположения элементов в нем и есть повторяющиеся элементы, то использовать формулу *размещений с повторениями* $\bar{A}_n^m = n^m$	Так как число комбинаций в подмножестве зависит от порядка расположения элементов в нем и есть повторяющиеся элементы, выбираем формулу размещений с повторениями $\bar{A}_2^{10} = 2^{10} = 1\,024$ варианта. *Замечание.* Эту задачу можно было решить, пользуясь лишь правилом произведения: на каждом из возможных десяти мест, предназначенных для записи результата игры, может быть одна из двух цифр — ***и*** 1, ***и*** 0. Поэтому количество цифр 2 необходимо умножить десять раз по числу проведенных игр, т. е. $\bar{A}_2^{10} = 2^{10} = 1\,024$ варианта

Задача 1д. Сколько различных слов можно составить из букв слова «барабан»?

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Установить количество элементов n всего множества и количество элементов m его подмножества	Множество состоит из семи элементов $n = 7$, подмножество в условии не рассматривается
2	Определить, влияет ли порядок расположения элементов в подмножестве на число вариантов различных подмножеств	В задаче требуется найти различные варианты распределения данных букв между семью местами, для них предназначенных, т. е. порядок расположения элементов важен
3	Выбрать, в зависимости от конкретного случая, комбинаторную операцию: если число комбинаций всего множества зависит от порядка расположения элементов в нем и есть повторяющиеся элементы, то использовать формулу *перестановок с повторениями* $\bar{P}_{n_1, \ldots, n_k} = \frac{(n_1 + n_2 + \ldots + n_k)!}{n_1!\, n_2! \ldots n_k!}$	Так как число комбинаций всего множества зависит от порядка расположения элементов в нем и есть повторяющиеся элементы, выбираем формулу перестановок с повторениями. Сосчитаем число повторяющихся букв: $n_1 = 3$ (для буквы а), $n_2 = 2$ (для буквы б), $n_3 = n_4 = 1$ (для букв р и н), т. е. $\bar{P}_{3,2,1,1} = \frac{(3+2+1+1)!}{3!2!1!1!} = \frac{7!}{3!2!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2 \cdot 3} = 420$ вариантов

Алгоритм 2.
Вычисление вероятности событий по определению

Задача 2а. Студент знает ответы на 18 вопросов зачета из 30. Какова вероятность того, что ему достанется на зачете известный вопрос?

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин и вопроса задачи	n — число всех вопросов; m — число «знакомых» вопросов; событие A — вопрос «знакомый»; $n = 30$, $m = 20$. Найти $p(A)$

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
2	Найти формулу вероятности для данного случая: а) *классическое* определение $P(A)=\frac{m}{n}$; б) *геометрическое* определение $P(A)=\frac{S(A)}{S(\Omega)}$	Задано общее число равновозможных событий n и число благоприятных событий m, следовательно, нужна формула классического определения вероятности: $P(A)=\frac{18}{30}=0,6$

Задача 2б. На квадратном дачном участке находится огород, также в форме квадрата, сторона которого вдвое меньше стороны дачного участка. Найти вероятность того, что капля долгожданного дождя попадет в огород.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин	a — длина стороны дачного участка; $a/2$ — длина стороны огорода; $S(\Omega)$ — площадь дачи, где Ω — пространство элементарных исходов; $S(A)$ — площадь огорода, где событие A — попадание капли дождя на огород — благоприятные исходы, тогда $S(\Omega)=a^2$; $S(A)=\frac{a^2}{4}$. Найти $P(A)$
2	Изобразить с помощью геометрических фигур полное пространство элементарных событий Ω и фигуру, соответствующую благоприятным исходам	1. Нарисовать фигуру (отрезок, круг, полоса, куб и др.), соответствующую полному пространству событий Ω. 2. Нарисовать внутри нее фигуру, соответствующую исходам, благоприятным для события A. В данном случае квадрат A расположен внутри квадрата Ω
3	Найти формулу вероятности для данного случая: а) *классическое* определение $P(A)=\frac{m}{n}$; б) *геометрическое* определение $P(A)=\frac{S(A)}{S(\Omega)}$.	Так как события описываются с помощью геометрических фигур, нужна формула б. Надо вычислить площади фигур A и Ω и найти вероятность как отношение этих площадей: $P(A)=\frac{S(A)}{S(\Omega)}=\frac{\frac{a^2}{4}}{a^2}=\frac{1}{4}=0,25$

Алгоритм 3.
Вычисление вероятностей событий по известным вероятностям других событий, с ними связанных

Задача 3. Стрелок производит три выстрела по мишени. Вероятности попадания при первом, втором и третьем выстрелах соответственно равны 0,4; 0,5 и 0,7. Найти вероятность того, что в результате этих выстрелов окажется: а) одно попадание в мишень; б) хотя бы одно попадание; в) не более одного попадания.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин	A_1 — попадание при первом выстреле; A_2 — попадание при втором выстреле; A_3 — попадание при третьем выстреле; B — одно попадание в мишень; C — хотя бы одно попадание в мишень; D — не более одного попадания. $P(A_1) = 0{,}4$; $P(A_2) = 0{,}5$; $P(A_3) = 0{,}7$. Найти: а) $p(B)$; б) $p(C)$; в) $p(D)$
2	Найти формулу для этого случая	Надо найти вероятность событий по вероятностям событий, связанных с первыми — (см. подразд. 1.2). Так как события A_1, A_2, A_3 независимые, но совместные, имеем: $B = A_1\bar{A}_2\bar{A}_3 + \bar{A}_1A_2\bar{A}_3 + \bar{A}_1\bar{A}_2A_3$; а) $p(B) = p(A_1)p(\bar{A}_2)p(\bar{A}_3) +$ $+p(\bar{A}_1)p(A_2)p(\bar{A}_3) + p(\bar{A}_1)p(\bar{A}_2)p(A_3) =$ $= 0{,}4 \cdot 0{,}5 \cdot 0{,}3 + 0{,}6 \cdot 0{,}5 \cdot 0{,}3 +$ $+ 0{,}6 \cdot 0{,}5 \cdot 0{,}7 = 0{,}36$; $\bar{C} = \bar{A}_1\bar{A}_2\bar{A}_3$; б) $p(C) = 1 - p(\bar{C}) = 1 - p(\bar{A}_1\bar{A}_2\bar{A}_3) =$ $= 1 - p(\bar{A}_1)p(\bar{A}_2)p(\bar{A}_3) = 1 - 0{,}6 \cdot 0{,}5 \cdot 0{,}3 =$ $= 0{,}91$; $D = C + B$; в) $P(D) = p(\bar{C}) + P(B) = 0{,}01 + 0{,}36 = 0{,}37$

Алгоритм 4.
Вычисление вероятностей событий с помощью соединений

Задача 4а. Имеется некоторое собрание сочинений из шести томов. На верхней полке умещаются только четыре тома. Эти четыре тома берут из шести томов случайным образом и расставляют на верхней полке в случайном порядке. Какова вероятность того, что тома расположатся в порядке 1, 2, 3, 4 или 4, 3, 2, 1?

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Обозначить все события, указанные в задаче	Событие B — тома расположатся в порядке 1, 2, 3, 4 или 4, 3, 2, 1
2	Вычислить число n всех равновозможных исходов	Число всех равновозможных исходов есть размещение, равное $n = A_n^k = \frac{n!}{(n-k)!}$; $n = A_6^2 = \frac{6!}{2!} = 360$
3	Вычислить число всех исходов m, благоприятствующих событию A	Количество всех исходов m, благоприятствующих событию A, есть размещение $m = 2$, так как в условии указаны лишь два возможных варианта
4	Найти формулу вероятности для данного случая, пользуясь *классическим определением* вероятности по формуле $P(A) = \frac{m}{n}$	Пользуясь классическим определением, имеем $P(B)\frac{2}{A_6^2} = \frac{2}{360} = \frac{1}{180}$

Задача 4б. Имеется некоторое собрание сочинений из шести томов. На верхней полке умещаются только четыре тома, которые берут из шести томов случайным образом и расставляют их на верхней полке. Какова вероятность того, что для размещения на верхней полке будут выбраны тома 1, 2, 3, 4?

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Обозначить все события, указанные в задаче	Событие B — «для размещения на верхней полке будут выбраны тома 1, 2, 3, 4»
2	Вычислить число n всех равновозможных исходов	Количество всех равновозможных исходов есть сочетание (порядок расположения томов не важен), равное $n = C_n^k = C_6^4 = \frac{6!}{2!4!} = \frac{5 \cdot 6}{2} = 15$
3	Вычислить число m всех исходов, благоприятствующих событию A	Количество всех исходов m, благоприятствующих событию A, равно $m = 1$, поскольку возможен единственный вариант

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
4	Найти формулу вероятности для данного случая, пользуясь *классическим определением* вероятности по формуле $P(A)=\frac{m}{n}$	Пользуясь классическим определением, имеем $$P(B)=\frac{1}{C_6^4}=\frac{1}{15}$$

Задача 4в. Из партии в 20 деталей, среди которых шесть дефектных, наугад берут три детали. Найти вероятность того, что одна из трех деталей с дефектом.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Обозначить все события, указанные в задаче	Событие A — «одна деталь с дефектом»
2	Вычислить число n всех равновозможных исходов	Число всех равновозможных исходов — сочетание (порядок не важен): $$n=C_{20}^3=\frac{20!}{17!3!}=\frac{18\cdot 19\cdot 20}{1\cdot 2\cdot 3}=1\,140$$
3	Вычислить число m всех исходов, благоприятствующих событию A	Так как по условию задачи только одна из трех деталей с дефектом, значит две другие без дефекта, поэтому количество всех исходов m, благоприятствующих событию A, есть произведение сочетаний (порядок не важен): $$m=C_6^1C_{14}^2=\frac{6!}{5!1!}\frac{14!}{12!2!}=\frac{6!}{5!}\frac{13\cdot 14}{2!}=$$ $$=6\cdot\frac{13\cdot 14}{2}=546$$
4	Найти формулу вероятности для данного случая, пользуясь *классическим определением* вероятности по формуле гипергеометрических распределений: $$P(A)=\frac{m}{n}=\frac{C_T^tC_{R-T}^{r-t}}{C_R^t}$$	Пользуясь классическим определением, имеем формулу числа успехов гипергеометрических распределений: $$P(A)=\frac{C_6^1C_{14}^2}{C_{20}^3}=\frac{546}{1\,140}=0{,}48$$

Алгоритм 5.

Вычисление вероятности события A по формуле полной вероятности. Вычисление вероятности одной из гипотез по формуле Байеса

Задача 5. В группе из 10 студентов, пришедших на экзамен, трое подготовлены отлично, четверо — хорошо, двое — удовлетворительно и 1 — плохо. Имеется 20 вопросов, причем: отлично подготовленный студент может ответить на все, хорошо подготовленный — на 16, удовлетворительно подготовленный — на 10 и плохо подготовленный — на 5. Найти вероятность того, что случайно выбранный студент:

а) сможет ответить на доставшийся ему вопрос;

б) студент плохо подготовлен и ему просто повезло с вопросом.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин и вычислить верятности по *классической формуле* $P(A) = m/n$, учитывая, что $\sum_{i=1}^{n} P(H_i) = 1$	1. Дать описание всех гипотез H_1, H_2, ..., H_n, на которые можно разбить пространство элементарных исходов, и события A: H_1 — студент-отличник; H_2 — студент учится на «хорошо»; H_3 — студент учится удовлетворительно; H_4 — студент плохо подготовлен; A — вопрос «знакомый». 2. Вычислить вероятность каждой гипотезы $P(H_1)$, $P(H_2)$, ..., $P(H_n)$: $P(H_1) = 0{,}3$ (3 из 10); $P(H_2) = 0{,}4$ (4 из 10); $P(H_3) = 0{,}2$ (2 из 10); $P(H_4) = 0{,}1$ (1 из 10). 3. Вычислить условную вероятность события A по каждой гипотезе $P(A\|H_1)$, $P(A\|H_2)$, ..., $P(A\|H_n)$: $P(A\|H_1) = 1$; $P(A\|H_2) = 16/20 = 0{,}8$; $P(A\|H_3) = 10/20 = 0{,}5$; $P(A\|H_4) = 5/20 = 0{,}25$. Найти: а) $P(A)$ и б) $P(H_4\|A)$
2	Найти формулу для этого случая	Пространство элементарных событий разбито на четыре непересекающиеся области, поэтому пользуемся формулой полной вероятности для вычисления $P(A)$: $$P(A) = \sum_{i=1}^{n} P(H_i)P(A\|H_i)$$

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
		и формулой Байеса для вычисления $P(H_4 \mid A)$: $$P(H_k \mid A) = \frac{P(H_k)P(A \mid H_k)}{\sum_{i=1}^{n} P(H_i)P(A \mid H_i)}.$$ а) $P(A) = 0{,}3 \cdot 1 + 0{,}4 \cdot 0{,}8 + 0{,}2 \cdot 0{,}5 + 0{,}1 \cdot 0{,}25 = 0{,}745$; б) $P(H_4 \mid A) = 0{,}1 \cdot 0{,}25/0{,}745 = 0{,}034$

Алгоритм 6.

Вычисление вероятностей числа успехов в независимых повторных испытаниях по формуле Бернулли

Задача 6. Вероятность того, что отремонтированный телевизор выдержит нормативную нагрузку, равна 0,9. Найти вероятность того, что из семи телевизоров, находящихся в ремонте, испытания выдержат: а) ровно пять; б) не менее пяти; в) хотя бы один; г) не более пяти.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин	n — число испытаний; m — число телевизоров, выдержавших испытания; p — вероятность выдержать испытания: $p = 0{,}9$; $q = 1 - p = 0{,}1$; $n = 7$. Найти: а) $p_7(5)$; б) $p_7\,(m \le 5 \le 7)$; в) $p_7\,(m \ge 1)$; г) $p_7\,(0 \le m \le 5)$
2	Сосчитать требуемую вероятность по *формуле Бернулли* $$p_n(m) = C_n^m p^m q^{n-m} = \frac{n!}{m!(n-m)!}p^m q^{n-m}$$	Так как $n < 10$, нужно воспользоваться формулой Бернулли $а) p_7(5) = C_7^5 \cdot 0{,}9^5 \cdot 0{,}1^2 = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} 0.9^2 \cdot 0{,}1^2 = 0{,}124;$ $б) p_7(5 \le m \le 7) = p_7(5) + p_7(6) + p_7(7) = C_7^5 \cdot 0{,}9^5 \cdot 0{,}1^2 + C_7^6 \cdot 0{,}9^6 \cdot 0{,}1^1 + C_7^7 \cdot 0{,}9^7 \cdot 0{,}1^0 = 0{,}974;$ $в) p_7(m \ge 1) = 1 - p_7(0) = 1 - 0{,}1^7 \approx 0{,}999;$ $г) p_7(0 \le m \le 5) = 1 - p(5 < m < 7) = 1 - p_7(m = 6) - p_7(m = 7) = 1 - 0{,}850 = 0{,}150$

Алгоритм 7.

Вычисление вероятностей числа успехов в независимых повторных испытаниях по формуле Пуассона

Задача 7. На факультете учатся 400 студентов. Найти вероятность того, что первое апреля является днем рождения: а) пяти студентов; б) менее пяти студентов; в) не менее пяти студентов; г) хотя бы одного студента.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин	n — число испытуемых (число равновозможных исходов); m — число студентов, родившихся 1 апреля (число благоприятных исходов); p — вероятность того, что студент родился 1 апреля: $p = 1/365$; $n = 300$. Найти: а) $p_{400}(5)$; б) $p_{400}(m < 5)$; в) $p_{400}(m \geq 5)$; г) $p_{400}(m \geq 1)$
2	Сосчитать требуемую вероятность *по формуле Пуассона* $p_n(m) \approx \dfrac{\lambda^m}{m!} e^{-\lambda}$ (см. табл. П.1)	Так как n велико, а p мало́ и $\lambda = np = 0{,}8$, нужно воспользоваться формулой Пуассона: а) $p_{400}(5) \approx \dfrac{0{,}8^5}{5!} e^{-0{,}8} \approx 0{,}0123$; б) $p_{400}(m < 5) \approx p_{400}(m = 1) + p_{400}(m = 2) + p_{400}(m = 3) + p_{400}(m = 4) + p_{400}(m = 5) \approx 0{,}4493 + 0{,}35946 + 0{,}14379 + 0{,}03834 + 0{,}00767 \approx 0{,}997$; в) событие $m \geq 5$ противоположное для события $m < 5$, поэтому $p_{400}(m \geq 5) \approx 1 - p_{400}(m < 5)$, тогда $p_{400}(m \geq 5) \approx 1 - 0{,}997 \approx 0{,}003$; г) событие $m \geq 1$ противоположное для события $m < 1$, поэтому $p_{400}(m \geq 1) \approx 1 - p_{400}(m = 0) \approx 1 - 0{,}449 \approx 0{,}551$

Алгоритм 8.

Вычисление вероятности числа m успехов для n независимых повторных испытаний, если n велико и $np > 10$, когда надо найти:

а) конкретное значение вероятности для m (по формуле Муавра—Лапласа);

б) вероятности попадания в интервал $[m_1, m_2]$ (по формуле Лапласа)

Задача 8. В первые классы школы должны быть приняты 200 детей. Вероятность появления среди принятых детей мальчика равна 0,515. Найти вероятность того, что среди них: а) девочек и **мальчиков будет поровну;** б) мальчиков будет меньше, чем девочек.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин	n — число детей; m — число мальчиков; p — вероятность того, что ученик мальчик ($p = 0{,}515$); $q = 0{,}485$; $n = 200$. Найти: а) $p_{200}(100)$; б) $p_{200}(m < n - m) =$ $= p_{200}(m < n/2) = p_{200}(m < 100)$
2	Вычислить требуемую вероятность, используя: а) формулу Муавра—Лапласа, когда надо найти конкретное значение вероятности для m; б) интегральную формулу Лапласа, когда надо найти вероятность попадания в интервал $[m_1, m_2]$	а) Так как n велико, а p мaлó и $np > 10$, нужно воспользоваться локальной теоремой Муавра—Лапласа, по которой $p_n(m)$ можно вычислить по формуле: $$p_n(m) \approx \frac{1}{\sqrt{npq}}\varphi(x),\ x = \frac{m-np}{\sqrt{npq}}$$ (значения $\varphi(x)$ см. в табл. П. 2), $np = 103$; $npq = 49{,}995$; $\sqrt{npq} = 7{,}068$; $m - np = 100 -$ $- 103 = -3$; $x = \dfrac{m-np}{\sqrt{npq}} = -0{,}4245$; $\varphi(-0{,}4245) = \varphi(0{,}4225) = 0{,}364$; $$p_{200}(100) \approx \frac{\varphi(0{,}4225)}{\sqrt{npq}} = \frac{0{,}364}{7{,}068} = 0{,}05;$$ б) Для вычисления значения $p_{200}(m <$ $< 100)$ используют интегральную теорему Лапласа: $p_n(m_1 \le m \le m_2) = \Phi(x_2) - \Phi(x_1)$, где $x_1 = \dfrac{m_1 - np}{\sqrt{npq}}$; $x_2 = \dfrac{m_2 - np}{\sqrt{npq}}$; $\Phi(x) =$ $= \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^x e^{-\frac{t^2}{2}}dt;$ $\Phi(x)$ — функция Лапласа (см. табл. П. 3), причем $\Phi(-x) = -\Phi(x)$. $$p_{200}(-\infty < m < 100) = \Phi\left(\frac{100 - 100\cdot 0{,}515}{\sqrt{100\cdot 0{,}515\cdot 0{,}485}}\right) -$$ $- \Phi(-\infty) = \Phi(-0{,}43) + 0{,}5 = 0{,}5 - \Phi(0{,}43) =$ $= 0{,}5 - 0{,}17 = 0{,}33.$

№ п/п	Умения	Алгоритмы
9	Составление ряда распределений и определение числовых характеристик для подсчета вероятностей числа успехов по схеме Бернулли (биномиальные распределения)	1. Ввести обозначения для заданных величин: числа испытаний, числа успехов, вероятности наступления события A, и выписать их значения. Выписать формулу Бернулли для искомой вероятности (по алгоритму 6). 2. Сосчитать требуемую вероятность, выбрав соответствующую содержанию задачи формулу Бернулли. 3. Найти числовые характеристики ДСВ по формулам $M(X) = np$, $D(X) = npq$ и $\sigma = \sqrt{D}$. 4. Составить ряд распределений случайной величины X — числа возможных образцов. 5. Составить функцию распределения случайной величины X — числа возможных образцов. 6. Построить график функции распределения ДСВ, учитывая значение накопительной вероятности на каждом интервале
10	Составление ряда распределений и нахождение числовых характеристик для определения вероятностей числа успехов в k-м испытании (геометрические распределения)	1. Ввести обозначения для заданных величин: числа испытаний, числа успехов, вероятности наступления события A и выписать их значения. Выписать формулу числа успехов геометрических распределений для искомой вероятности, т.е. вероятности того, что событие впервые произойдет в k-м испытании. 2. Сосчитать требуемую вероятность, выбрав соответствующую содержанию задачи формулу числа успехов в k-м испытании, т.е. геометрических распределений $P(X = k) = q^{k-1}p$. 3. Найти числовые характеристики ДСВ по формулам $M(X) = \frac{1}{p}$; $D(X) = \frac{q}{p^2}$ и $\sigma = \sqrt{D}$. 4. Составить ряд распределений случайной величины X — числа возможных образцов. 5. Составить функцию распределения случайной величины X — числа возможных образцов.

№ п/п	Умения	Алгоритмы
		6. Построить график функции распределения ДСВ, учитывая значение накопительной вероятности на каждом интервале
11	Составление ряда распределений и нахождение числовых характеристик для определения вероятностей числа успехов гипергеометрических распределений	1. Ввести обозначения для заданных величин: числа испытаний, числа успехов, вероятности наступления события A и выписать их значения. Выписать формулу для искомой вероятности: если необходимо установить вероятность появления отдельных долей подмножеств, то используют гипергеометрические распределения. 2. Найти требуемую вероятность, выбрав соответствующую содержанию задачи формулу числа успехов для гипергеометрических распределений по алгоритму 4 (п. 3). 3. Найти числовые характеристики ДСВ по общим формулам $M(X)=\sum_{i=1}^{n} x_i p_i$; $D(X)=$ $=\sum_{i=1}^{n} x_i^2 p_i - [M(X)]^2$ и $\sigma=\sqrt{D}$. 4. Составить ряд распределений случайной величины X— числа возможных образцов. 5. Составить функцию распределения случайной величины X — числа возможных образцов. 6. Построить график функции распределения ДСВ, учитывая значение накопительной вероятности на каждом интервале
12	Определение числовых характеристик ДСВ $Z=f(X,Y)$ и вероятности попадания в интервал случайной величины $Z=f(X,Y)$	1. Составить ряд распределений для одинаково распределенных случайных величин X, Y, $Z=f(X,Y)$. 2. Вычислить математическое ожидание по формуле $M(Z)=\sum_i z_i p_i$. 3. Вычислить дисперсию по формуле $D(Z)=\sum_i z_i^2 p_i - M^2(Z)$ и среднеквадратическое отклонение $\sigma(Z)=\sqrt{D(Z)}$. 4. Вычислить вероятности попадания в интервал случайной величины $Z=f(X,Y)$

№ п/п	Умения	Алгоритмы
13	Определение числовых характеристик НСВ и вероятности попадания НСВ X в интервал $P(a < X < b)$	1. Записать функцию плотности вероятности $f(x) = F(x)$. 2. Вычислить математическое ожидание на указанном отрезке по формуле: $M(X) = \int_a^b xf(x)dx$. 3. Вычислить дисперсию на указанном отрезке по формуле $D(X) = \int_a^b x^2 f(x)dx - [M(X)]^2$ и среднеквадратическое отклонение по формуле $\sigma = \sqrt{D}$. 4. Найти моду, исследовав на экстремум функцию $f(x)$ с помощью производной $f(x)$. 5. Найти медиану, решив уравнение $P\{x < \text{Me}\} = P\{x > \text{Me}\} = \frac{1}{2}$. 6. Вычислить вероятность попадания в интервал $(a; b)$ по формуле: $P(a < X < b) = \Phi(b) - \Phi(a)$. 7. Построить график функции плотности вероятности $f(X)$ и функции распределения $F(X)$
14	Определение числовых характеристик НСВ, равномерно распределенной на отрезке $[a, b]$, и вероятности попадания НСВ X в интервал $P(a < X < b)$	1. Записать функцию плотности вероятности $f(x)$, определив границы интервала a и b. 2. Вычислить математическое ожидание по формуле для равномерных распределений $M(X) = \frac{a+b}{2}$. 3. Вычислить дисперсию по формуле для равномерных распределений $D(X) = \frac{(b-a)^2}{12}$. 4. Записать функцию распределения вероятности $F(x)$, имеющую вид $F(x) = \begin{cases} 0 \text{ при } x < 0; \\ \frac{1}{b-a}x \text{ при } x \in [a; b); \\ 1 \text{ при } x \geq 1. \end{cases}$

№ п/п	Умения	Алгоритмы
		5. Вычислить вероятности попадания НСВ в интервал $P(a < X < b)$ по алгоритму 13 (п. 6).
15	Определение числовых характеристик НСВ, имеющей показательное распределение на отрезке $[a, b]$	1. Записать функцию плотности вероятности показательного распределения для заданного значения λ по формуле $$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ \lambda e^{-\lambda x} & \text{при } x \geq 0. \end{cases}$$ 2. Записать функцию распределения для заданного значения λ по формуле $$F(x) = \begin{cases} 0 & \text{при } x < 0; \\ 1 - e^{-0{,}0002x} & \text{при } x \geq 0. \end{cases}$$ 3. Вычислить математическое ожидание по формуле $M(X) = \frac{1}{\lambda}$. 4. Вычислить дисперсию и среднеквадратическое отклонение по формулам $D(X) = \frac{1}{\lambda^2}$ и $\sigma = \sqrt{D} = \frac{1}{\lambda}$. 5. Вычислить вероятность попадания в интервал по формуле $P(a < X < b) = e^{-ax} - e^{-bx}$
16	Определение числовых характеристик и вероятности попадания нормально распределенной НСВ X в интервал $P(a < X < b)$	1. Записать математическое ожидание m и среднеквадратическое отклонение σ для нормально распределенной НСВ по закону $N(m, \sigma)$. 2. Записать функцию плотности вероятности $f(x)$ для нормально распределенной НСВ X. 3. Записать функцию распределения вероятности для нормально распределенной НСВ X. 4. Вычислить $P(a < x < b)$, используя табл. П. 3 по формуле $$P(a < X < b) = \Phi\left(\frac{b - m}{\sigma}\right) - \Phi\left(\frac{a - m}{\sigma}\right).$$

№ п/п	Умения	Алгоритмы
		5. Вычислить $P(-\delta < x - m < \delta)$ или $P(\|x - m\| < < \delta)$ по формуле $P(\|x - a\| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$, используя табл. П. 3
17	Оценка вероятности того, что НСВ X отличается от среднего на величину ε с помощью неравенства Чебышева	1. Записать условие символически: математическое ожидание m_X, среднеквадратическое отклонение σ, дисперсия $D(X)$. 2. Подставить значения m_X, $D(X)$ и ε в неравенство Чебышева $P(\|X - m_X\| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$, если надо оценить вероятность того, что НСВ отличается от среднего на величину, больше, чем ε. 3. Подставить значения m_X, $D(X)$и ε в неравенство Чебышева $P(\|X - m_X\| < \varepsilon) \geq 1 - -\frac{D(X)}{\varepsilon^2}$, если надо оценить вероятность того, что НСВ отличается от среднего на величину, меньше, чем ε

Задачи на применение алгоритмов

Алгоритм 9.

Составление ряда распределений и определение числовых характеристик для подсчета вероятностей числа успехов по схеме Бернулли (биномиальные распределения)

Задача 9. Вероятность того, что образец бетона выдержит нормативную нагрузку, равна 0,9. Случайная величина X — число образцов, которые выдержат испытания. Составить ряд распределения, найти функцию распределения ДСВ X, построить ее график и найти все числовые характеристики, если в нашем распоряжении пять образцов.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин	n — число испытаний; m — число образцов, выдержавших испытания; p — вероятность выдержать испытание $p = 0{,}9$; $q = 1 - p = 0{,}1$; $n = 7$. Найти: $p_5(5)$, $p_5(4)$, $p_5(3)$, $p_5(2)$, $p_5(1)$, $p_5(0)$

<table>
<tr><th>№ п/п</th><th>Алгоритмы</th><th>Конкретное соответствие задания заданному алгоритму</th></tr>
<tr><td>2</td><td>Сосчитать требуемую вероятность, выбрав соответствующую содержанию задачи формулу Бернулли</td><td>Так как $n < 10$, нужно воспользоваться формулой Бернулли

$$p_5(5) = C_5^5 \cdot 0,9^5 \cdot 0,1^0 = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} \cdot 0,59 \cdot 1 = 0,59;$$

$$p_5(4) = C_5^4 \cdot 0,9^4 \cdot 0,1^1 = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 3 \cdot 4} \cdot 0,656 \cdot 0,1 = \\ = 0,328;$$

$$p_5(3) = C_5^3 \cdot 0,9^3 \cdot 0,1^2 = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2} \cdot 0,729 \cdot 0,01 = \\ = 0,0729;$$

$$p_5(2) = C_5^2 \cdot 0,9^2 \cdot 0,1^3 = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 1 \cdot 2 \cdot 3} \cdot 0,81 \cdot 0,001 = \\ = 0,0081;$$

$$p_5(1) = C_5^1 \cdot 0,9^1 \cdot 0,1^4 = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 1 \cdot 2 \cdot 3 \cdot 4} = 0,00045;$$

$$p_5(0) = C_5^0 \cdot 0,9^0 \cdot 0,1^5 = 0,00001$$</td></tr>
<tr><td>3</td><td>Найти числовые характеристики ДСВ по формулам $M(X) = np$; $D(X) = npq$</td><td>$M(X) = np = 5 \cdot 0,9 = 0,45;$
$D(X) = npq = 5 \cdot 0,9 \cdot 0,1 = 0,045;$
$\sigma = \sqrt{D} = \sqrt{0,045} = 0,212$</td></tr>
<tr><td>4</td><td>Составить ряд распределений СВ X — числа возможных образцов</td><td><table><tr><td>x</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>P</td><td>0,00001</td><td>0,00045</td><td>0,0081</td><td>0,0729</td><td>0,328</td><td>0,59</td></tr></table></td></tr>
<tr><td>5</td><td>Составить функцию распределения СВ X — числа возможных образцов</td><td>

$$F(z) = \begin{cases} 0 & \text{при } z \le 0; \\ 0,00001 & \text{при } 0 < z \le 1; \\ 0,00046 & \text{при } 1 < z \le 2; \\ 0,00856 & \text{при } 2 < z \le 3; \\ 0,08146 & \text{при } 3 < z \le 4; \\ 0,40946 & \text{при } 4 < z \le 5; \\ 1 & \text{при } z > 5 \end{cases}$$</td></tr>
</table>

Алгоритм 10.

Составление ряда распределений и нахождение числовых характеристик для определения вероятностей числа успехов в k-м испытании (геометрические распределения)

Задача 10. Вероятность того, что образец бетона выдержит нормативную нагрузку, равна 0,9. Случайная величина X — число возможных испытаний до появления первого бракованного образца. Составить ряд распределения, найти функцию распределения ДСВ X, построить ее график и найти все числовые характеристики (ограничиться тремя—пятью испытаниями).

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин	n — число испытаний; p — вероятность выдержать испытание: $p = 0,9$; $q = 1 - p = 0,1$. Найти: $p_5(5)$, $p_5(4)$, $p_5(3)$, $p_5(2)$, $p_5(1)$, $p_5(0)$
2	Сосчитать требуемую вероятность, выбрав соответствующую содержанию задачи формулу	Так как случайная величина X — число возможных испытаний до появления первого бракованного образца, воспользуемся геометрической вероятностью: $P(5) = p^4 \cdot q = 0,9^4 \cdot 0,1 = 0,656 \cdot 0,1 = 0,0656;$ $P(4) = p^3 \cdot q = 0,9^3 \cdot 0,1 = 0,729 \cdot 0,1 = 0,0729;$ $P(3) = p^2 \cdot q = 0,9^2 \cdot 0,1 = 0,81 \cdot 0,1 = 0,081;$ $P(2) = p \cdot q = 0,9 \cdot 0,1 = 0,09;$ $P(1) = q = 0,1$
3	Найти числовые характеристики ДСВ по формулам: $M(X) = \frac{1}{p}$; $D(X) = \frac{q}{p^2}$; $\sigma = \sqrt{D}$	$M(X) = \frac{1}{p} = \frac{1}{0,9} = 1,11;$ $D(X) = \frac{q}{p^2} = \frac{0,1}{0,81} = 0,123;$ $\sigma = \sqrt{D} = \sqrt{0,123} = 0,351$
4	Составить ряд распределений СВ X — числа возможных образцов	<table><tr><td>x</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>P</td><td>0,1</td><td>0,09</td><td>0,0081</td><td>0,0729</td><td>0,0656</td></tr></table>

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
5	Составить функцию распределения случайной величины X — числа возможных образцов и построить график	$F(x)=\begin{cases}0 & \text{при } x\le 1;\\ 0{,}1 & \text{при } 1<x\le 2;\\ 0{,}19 & \text{при } 2<x\le 3;\\ 0{,}1981 & \text{при } 3<x\le 4;\\ 0{,}271 & \text{при } 4<x\le 5;\\ 0{,}3366 & \text{при } x>5;\\ \ldots & .\end{cases}$

Алгоритм 11.

Составление ряда распределений и нахождение числовых характеристик для определения вероятностей числа успехов гипергеометрических распределений

Задача 11. В партии автомобилей «ВАЗ», состоящей из 10 машин, 7 машин — ВАЗ-2112. Сегодня двое покупателей планируют приобрести автомобили этой модели. Составить закон распределения числа проданных ВАЗ-2112 и найти его числовые характеристики.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Ввести обозначения для заданных величин	Число проданных ВАЗ-2112 есть ДСВ. Обозначим ее через X. Для подсчета вероятностей этих значений используем гипергеометрическое распределение, основанное на биномиальных коэффициентах. Найти: $P(0)$; $P(1)$; $P(2)$
2	Сосчитать требуемую вероятность, выбрав соответствующую содержанию задачи формулу	$P(x=0)=\dfrac{C_7^0C_3^2}{C_{10}^2}=\dfrac{1}{15}$; $P(x=1)=\dfrac{C_7^1C_3^1}{C_{10}^2}=\dfrac{7}{15}$; $P(x=2)=\dfrac{C_7^2C_3^0}{C_{10}^2}=\dfrac{7}{15}$
3	Найти числовые характеристики	$M(X)=0\cdot\dfrac{1}{15}+1\cdot\dfrac{7}{15}+2\cdot\dfrac{7}{15}=\dfrac{7}{15}\approx 1{,}4;$ $M(X^2)=0\cdot\dfrac{1}{15}+1\cdot\dfrac{7}{15}+4\cdot\dfrac{7}{15}=\dfrac{7}{3}=2\dfrac{1}{3}\approx 2{,}33;$ $D(X)\approx 2{,}33-1{,}96=0{,}37;$ $\sigma=\sqrt{0{,}37}\approx 0{,}6$

<table>
<tr><th>№ п/п</th><th>Алгоритмы</th><th>Конкретное соответствие задания заданному алгоритму</th></tr>
<tr><td>4</td><td>Составить ряд распределений случайной величины X — числа возможных образцов</td><td>
<table>
<tr><td>x_i</td><td>0</td><td>1</td><td>2</td></tr>
<tr><td>P_i</td><td>$\frac{1}{15}$</td><td>$\frac{7}{15}$</td><td>$\frac{7}{15}$</td></tr>
</table>
</td></tr>
<tr><td>5</td><td>Составить функцию распределения случайной величины X — числа возможных образцов и построить график</td><td>

$$F(x) = \begin{cases} 0 & \text{при } t \le 0; \\ 1/15 & \text{при } 0 < t \le 1; \\ 8/15 & \text{при } 1 < t \le 2; \\ 1 & \text{при } t > 2 \end{cases}$$

</td></tr>
</table>

Алгоритм 12.
Определение числовых характеристик ДСВ $Z = f(X, Y)$ и вероятности попадания в интервал случайной величины $Z = f(X, Y)$

Задача 12. Случайная величина X задана рядом распределений:

x_i	–2	–1	0	1	2
p_i	0,4	0,2	0,1	0,1	0,2

Найти закон распределения случайной величины $Z = 3x^2 - 4$ и ее числовые характеристики: математическое ожидание $M(Z)$, дисперсию $D(Z)$, среднеквадратичное отклонение $\sigma(Z)$, а также вероятности $P(Z < 0)$; $P(Z > 0)$; $P(-1 < Z < 3)$.

Решение.

<table>
<tr><th>№ п/п</th><th>Алгоритмы</th><th>Конкретное соответствие данной ситуации предложенному алгоритму</th></tr>
<tr><td>1</td><td>Составить ряд распределений для одинаково распределенных СВ:
$Y_1 = X^2$;
$Y_2 = 3X^2$;
$Z = 3X^2 - 4$</td><td>
<table>
<tr><td>X^2</td><td>0</td><td>1</td><td>4</td></tr>
<tr><td>p</td><td>0,1</td><td>0,3</td><td>0,6</td></tr>
</table>
,
<table>
<tr><td>$3X^2$</td><td>0</td><td>3</td><td>12</td></tr>
<tr><td>p</td><td>0,1</td><td>0,3</td><td>0,6</td></tr>
</table>
.
Таким образом,
<table>
<tr><td>$3X^2 - 4$</td><td>–1</td><td>1</td><td>8</td></tr>
<tr><td>p</td><td>0,1</td><td>0,3</td><td>0,6</td></tr>
</table>
</td></tr>
</table>

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
2	Вычислить математическое ожидание $M(Z) = \sum_i z_i p_i$	$M(Z) = (-1) \cdot 0{,}1 + 1 \cdot 0{,}3 + 8 \cdot 0{,}6 = 5$. Другой способ: если известно $M(X)$, то $M(aX + b) = aM(X) + b$. Тогда $M(X^2) = 0 \cdot 0{,}1 + 1 \cdot 0{,}3 + 4 \cdot 0{,}6 = 2{,}7$, тогда $M(aX^2 + b) = M(3X^2 - 4) = 3 \cdot 2{,}7 - 4 = 4{,}1$
3	Вычислить дисперсию $D(Z) = \sum_i z_i^2 p_i - M(Z^2)$ и среднеквадратическое отклонение $\sigma(Z)$	$D(Z) = (-1)^2 \cdot 0{,}1 + 1^2 \cdot 0{,}3 + 8^2 \cdot 0{,}6 - 5^2 = 0{,}1 + 0{,}3 + 38{,}4 - 25 = 13{,}8$; $\sigma(Z) = \sqrt{13{,}8} = 3{,}7$. Другой способ: $D(X^2) = 0^2 \cdot 0{,}1 + 1^2 \cdot 0{,}3 + 4^2 \cdot 0{,}6 - 2{,}7^2 = 2{,}6$; $D(aX + b) = a^2 DX + b$; $D(Z) = D(3X^2 - 4) = D(3X^2) - 4 = 9D(X^2) - 4 = 9 \cdot 2{,}6 - 4 = 19{,}5$
4	Вычислить вероятности попадания случайной величины $Z = f(X, Y)$ в интервал $P(Z < 0)$, $P(Z > 0)$, $P(-1 < Z < 3)$	$P(Z < 0) = P(z = -1) = 0{,}1$; $P(Z > 0) = P(z = 1) + P(z = 8) = 0{,}3 + 0{,}6 = 0{,}9$; $P(-1 < Z < 3) = P(z = 1) = 0{,}3$

Алгоритм 13.
Определение числовых характеристик НСВ и вероятности попадания НСВ X в интервал $P(a < X < b)$

Задача 13а. Функция плотности случайной величины X имеет вид:

$$f(x) = \begin{cases} 0 & \text{при } x < 0;\ x > \dfrac{1}{\sqrt{5}}; \\ 10x & \text{при } 0 \le x \le \dfrac{1}{\sqrt{50}}. \end{cases}$$

Найти математическое ожидание $M(X)$, дисперсию $D(X)$, а также вероятность попадания СВ X в интервал $P(0 < X < 0{,}1)$.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Записать функцию плотности вероятности $f(x) = F(x)$. Вычислить математическое ожидание на указанном отрезке	Функция плотности вероятности $f(x)$ задана. Найдем математическое ожидание $M(X)=\int\limits_{\infty}^{-\infty} xf(x)dx=\int\limits_{0}^{\frac{1}{\sqrt{5}}} x\cdot 10xdx=10\int\limits_{0}^{\frac{1}{\sqrt{5}}} x^2dx=$ $=\frac{10}{3}x^3\Big\vert_0^{\frac{1}{\sqrt{5}}}=\frac{10}{3}\cdot\frac{1}{5\sqrt{5}}=\frac{2}{3\sqrt{5}}=0,3$
2	Вычислить дисперсию $D(X)$ и среднеквадратическое отклонение $\sigma(X)$	$D(X)=\int\limits_{-\infty}^{+\infty} x^2f(x)dx-[M(X)]^2=\int\limits_{0}^{\frac{1}{\sqrt{5}}} x^2\cdot 10xdx-$ $-\left(\frac{2}{3\sqrt{5}}\right)^2=10\int\limits_{0}^{\frac{1}{\sqrt{5}}} x^3dx-\frac{4}{45}=\frac{10}{4}x^4\Big\vert_0^{\frac{1}{\sqrt{5}}}-\frac{4}{45}=$ $=\frac{5}{2}\cdot\frac{1}{25}-\frac{4}{45}=\frac{1}{90}\approx 0,01;\ \sigma(X)\approx 0,11$
3	Вычислить $P(0 < X < 0,2)$	$P(0<X<0,2)=\int\limits_{0}^{0,2} 10x\,dx=5x^2\Big\vert_0^{0,2}=0,2$

Задача 136. НСВ X задана функцией распределения

$$F(x)=\begin{cases}0 & \text{при } x\le -2;\\ \frac{x}{5}+\frac{2}{5} & \text{при } -2<x\le 3;\\ 1 & \text{при } x>3.\end{cases}$$

Найти:

1) плотность вероятности $f(x)$;

2) числовые характеристики НСВ X;

3) вероятность того, что в результате испытаний СВ X примет значение в интервале$(0;2)$;

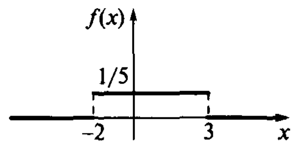
4) графики функции распределения и плотности вероятности.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Записать функцию плотности вероятности $f(x) = F(x)$	$f(x)=\left(\frac{x}{5}+\frac{2}{5}\right)'=\frac{1}{5};\ f(x)=\begin{cases}\frac{1}{5} & \text{при } x\in(-2;3];\\ 0 & \text{при } x\notin(-2;3]\end{cases}$

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
2	Вычислить математическое ожидание на указанном отрезке $M(X)=\int\limits_a^b xf(x)dx$	$M(X)=\int\limits_{-2}^{3} x\frac{1}{5}dx=\frac{1}{5}\cdot\frac{x^2}{2}\Big\|_{-2}^{3}=\frac{9-4}{10}=\frac{5}{10}=\frac{1}{2}$
3	Вычислить дисперсию $D(X)$ и среднеквадратическое отклонение $\sigma(X)$	$D(X)=\int\limits_{-\infty}^{+\infty} x^2f(x)dx-M[(X)]^2=\int\limits_{-2}^{3} x^2\cdot\frac{1}{4}dx-$ $-\left(\frac{1}{2}\right)^2=\frac{1}{5}\int\limits_{-2}^{3} x^2dx-\frac{1}{9}=\frac{x^3}{15}\Big\|_{-2}^{3}-\frac{1}{4}=$ $=\frac{21}{15}-\frac{1}{4}=\frac{69}{60};\ \sigma(X)=\sqrt{1,21}=1,1$
4	Найти моду Мо, исследовав на экстремум функцию $f(x)$	$f'(x)=\left(\frac{1}{5}\right)'=0$, поэтому моды нет
5	Найти медиану Me, решив уравнение $P(x<\text{Me})=$ $=P(x>\text{Me})=\frac{1}{2}$	$P(x<\text{Me})=\int\limits_{-2}^{\text{Me}}\frac{1}{5}dx=\frac{1}{5}(\text{Me}+2).$ Для поиска медианы решим уравнение $P(x<$ $<\text{Me})=\frac{1}{2}$, значит $\text{Me}=2$
6	Вычислить вероятность попадания в интервал $(a; b)$ по формуле $P(a<X<b)=$ $=F(b)-F(a)$	Найти вероятность попадания в интервал $(a; b)$: $P(0<X<2)=\frac{2-0}{5}=\frac{2}{5}$
7	Построить график функции распределения	Функция распределения $F(x)=\begin{cases}0 & \text{при } x\le -2;\\ \frac{x}{5}+\frac{2}{5} & \text{при } -2<x\le 3;\\ 1 & \text{при } x>3\end{cases}$

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
8	Построить график плотности вероятности	$f(x)=\begin{cases}\frac{1}{5} & \text{при} \quad x\in(-2;3]; \\ 0 & \text{при} \quad x\notin(-2;3]\end{cases}$ 

Алгоритм 14.

Определение числовых характеристик НСВ, равномерно распределенной на отрезке [a, b], и вероятности попадания НСВ X в интервал $P(a < X < b)$

Задача 14. Случайная величина X — время ожидания дождя в течение суток — имеет равномерное распределение на отрезке [0, 20]. Найти математическое ожидание $M(X)$, дисперсию $D(X)$, функцию распределения, а также вероятности $P(X < 5)$ и $P(X > 3)$.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Записать функцию плотности вероятности $f(x)$, определив границы интервала a и b	$f(x)=\begin{cases}0 & \text{при} \quad x\notin[0;20]; \\ \frac{1}{20} & \text{при} \quad x\in[0;20],\end{cases}$ где $a=0$ и $b=20$
2	Вычислить математическое ожидание $M(X)$ по формуле для равномерных распределений $M(X)=\frac{a+b}{2}$	$M(X)=\int\limits_{-\infty}^{+\infty} xf(x)dx=\int\limits_{0}^{20}\frac{1}{20}x\,dx=\frac{1}{40}x^2\Big\vert_0^{20}=10,$ или по формуле для равномерных распределений $M(X)=\frac{a+b}{2}=\frac{0+20}{2}=10$
3	Вычислить дисперсию $D(X)$ по формуле для равномерных распределений	$D(X)=\int\limits_{-\infty}^{+\infty} x^2 f(x)dx-[M(X)]^2=\int\limits_{0}^{20}\frac{1}{20}x^2dx-10^2=$ $=\frac{1}{20}\cdot\frac{x^3}{3}\Big\vert_0^{20}-100=\frac{400}{3}-100=\frac{100}{3}=33,3,$ или

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
	$D(X)=\frac{(b-a)^2}{12}$	$D(X)=\frac{(b-a)^2}{12}=\frac{(20-0)^2}{12}=\frac{400}{12}=\frac{100}{3}=33,3$
4	Записать функцию распределения вероятности $F(X)$, имеющую вид $F(x)=\begin{cases}0 \text{ при } x<0;\\ \frac{1}{b-a}x \text{ при } x\in[a;b);\\ 1 \text{ при } x\geq 1\end{cases}$	$F(x)=\int\limits_{-\infty}^{+\infty} f(t)dt=\int\limits_{0}^{x}\frac{1}{20}dt=\frac{1}{20}x$, откуда $F(x)=\begin{cases}0 & \text{при } x<0;\\ \frac{1}{20}x & \text{при } x\in[0;20);\\ 1 & \text{при } x\geq 1\end{cases}$
5	Вычислить вероятности попадания НСВ в интервал $P(a<X<b)$ по алгоритму 13 (п. 6)	$P(X<5)=\int\limits_{0}^{5}\frac{1}{20}dx=\frac{5}{20}=\frac{1}{4}=0,25;$ $P(X>3)=\int\limits_{3}^{20}\frac{1}{20}dx=\frac{17}{20}=0,85$

Алгоритм 15.

Определение числовых характеристик НСВ, имеющей показательное распределение на отрезке [*a*, *b*]

Задача 15. Вероятность безотказной работы прибора в течение x часов равна $e^{-0,0002X}$ (X — момент отказа прибора). Найти математическое ожидание $M(X)$ — среднюю наработку на отказ T и вероятность безотказной работы прибора в течение 500 ч.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Записать функцию плотности вероятности $f(x)$ показательного распределения по формуле	$f(x)=\begin{cases}0 & \text{при } x<0;\\ 0,0002e^{-0,0002x} & \text{при } x\geq 0\end{cases}$

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
	$f(x)=\begin{cases}0 & \text{при } x<0;\\ \lambda e^{-\lambda x} & \text{при } x\geq 0\end{cases}$ для заданного значения $\lambda = 0{,}0002$	
2	Записать функцию распределения $F(x)$ показательного распределения $F(x)=\begin{cases}0 & \text{при } x<0;\\ 1-e^{-0{,}0002x} & \text{при } x\geq 0\end{cases}$	$F(x)=\begin{cases}0 & \text{при } x<0;\\ 1-e^{-0{,}0002x} & \text{при } x\geq 0\end{cases}$
3	Вычислить математическое ожидание	$M(X)=\dfrac{1}{\lambda}=\dfrac{1}{0{,}0002}=5000$ (ч)
4	Вычислить дисперсию и среднеквадратическое отклонение	$D(X)=\dfrac{1}{\lambda^2}=\dfrac{1}{0{,}0002^2}=25\,000\,000;\ \sigma=\dfrac{1}{\lambda}=5000$
5	Вычислить вероятность попадания в интервал по формуле $P(a<X<b)=e^{-ax}-e^{-bx}$	$P(X>500)=e^{-0{,}0002\cdot 500}=e^{-0{,}1}=0{,}905$

Алгоритм 16.

Определение числовых характеристик и вероятности попадания нормально распределенной НСВ X в интервал $P(a<X<b)$

Задача 16а. Случайная величина X имеет нормальное распределение $N(m,\sigma)=N(3,2)$. Найти числовые характеристики, функцию распределения, плотность вероятности НСВ X, а также вероятности попадания НСВ X в интервалы $P(-1<X<1)$, $P(-2<X-3<2)$, $P(-4<X-3<6)$.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Записать математическое ожидание $M(X)$, дисперсию $D(X)$ и среднеквадратическое отклонение y для нормально распределенной СВ	$M(X) = 3$; $D(X) = 2^2 = 4$; $\sigma = 2$
2	Записать функцию распределения $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \times \times \int\limits_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt$ и плотность вероятности НСВ X $f(X) = \frac{1}{\sigma\sqrt{2\pi}} \times \times e^{-\frac{(x-m)^2}{2\sigma^2}}$	$F(x) = \frac{1}{2\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t-3)^2}{8}} dt;$ $f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-3)^2}{8}}$
3	Вычислить $P(-1 < < x < 1)$, используя табл. П. 3, по формуле $P(a < x < < b) = \Phi\left(\frac{b-m}{\sigma}\right) - -\Phi\left(\frac{a-m}{\sigma}\right)$	$P(-1 < X < 1) = \Phi\left(\frac{1-3}{2}\right) - \Phi\left(\frac{-1-3}{2}\right) = \Phi(-1) - -\Phi(-2) = -0{,}34338 - (-0{,}4772) = 0{,}1334$
4	Вычислить $P(-2 < < x - 3 < 2)$ или $P(\|x-3\| < 2)$ по формуле $P(\|x-a\| < < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$, используя табл. П. 3	$P(-2 < x - 3 < 2) = P(\|x-3\| < 2) = 2\Phi\left(\frac{2}{2}\right) = = 2\Phi(1) = 0{,}6826$
5	Вычислить $P(-4 < < X - 3 < 6)$, ис-	$P(-4 < X - 3 < 6) = P(-1 < x < 9) =$

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
	пользуя табл. П. 3, по формуле $P(a<x<b)=$ $\Phi\left(\frac{b-m}{\sigma}\right)-$ $-\Phi\left(\frac{a-m}{\sigma}\right)$	$=\Phi\left(\frac{9-3}{2}\right)-\Phi\left(\frac{-1-3}{2}\right)=\Phi(3)-\Phi(-2)=$ $=0,9986-0,0228=0,9758$
6	Вычислить $P(-6<$ $<X-3<6)$ с помощью правила «трех сигм»	$P(-6<X-3<6)=P\left(\frac{-6}{2}<\frac{X-3}{2}<\frac{6}{2}\right)=$ $\Phi(3)-\Phi(-3)=0,9986-0,0014=0,9972$

Задание 16б. Масса цемента, упакованного автоматом в бумажный мешок, есть случайная нормально распределенная величина с математическим ожиданием $m=50$ кг и среднеквадратичным отклонением $\sigma_x=2$ кг (заданный стандарт $x=50\pm2$ кг). Найти вероятность того, что: а) случайно выбранный мешок будет содержать не менее 48 кг цемента; б) партия из 100 мешков будет содержать не более 5040 кг.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Записать условие символически: математическое ожидание $M(X)$, среднеквадратическое отклонение σ_x, дисперсия $D(X)$	а) $M(X)=m=50$; $\sigma=2$; $D(X)=4$; б) пусть СВ Y — масса 100 мешков цемента, тогда $Y=\sum_{i=1}^{100}X_i$, $M(Y)=5000$; $D(Y)=100D(X)$; $\sigma_y=10\sigma_x=20$. Найти а) $P(X\geq48)$, б) $P(Y<5040)$
2	Вычислить вероятности попадания в интервал с помощью таблиц нормального распределения (функция Лапласа, табл. П. 3)	а) $P(48\leq X<\infty)=\Phi\left(\frac{\infty-50}{2}\right)-\Phi\left(\frac{48-50}{2}\right)=$ $=\Phi(\infty)-\Phi(-1)=0,5+0,3413=0,8413$; б) $P(0\leq Y<5040)=\Phi\left(\frac{5040-5000}{20}\right)-$ $-\Phi\left(\frac{0-5000}{20}\right)=\Phi(2)-\Phi(-250)=$ $=\Phi(2)+\Phi(250)=0,4772+0,5=0,9772$ *Замечание.* Использовали свойство нечетной функции Лапласа: $\Phi(-x)=-\Phi(x)$

Алгоритм 17.

ка вероятности того, что НСВ отличается от среднего на величину ε, с помощью неравенства Чебышева

дача 17а. Масса мужчины — случайная величина со средним 80 кг дерсией 50 кг^2. Оценить с помощью неравенства Чебышева вероть того, что масса случайно встреченного мужчины отличается еднего на величину а) больше, чем 10 кг, б) не больше, чем

шение.

Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
Записать условие символически: математическое ожидание m_X, дисперсия $D(X)$, оценка отклонения ε	$m_X = 80$; $D(X) = 50$; а) $\varepsilon_1 = 10$; б) $\varepsilon_2 = 5$
Подставить значения m_X, $D(X)$ и ε в неравенство Чебышева $P(\lvert X - m_X\rvert \geq \varepsilon) \leq \dfrac{D(X)}{\varepsilon^2}$, так как надо оценить вероятность того, что НСВ отличается от среднего на величину, больше, чем ε	а) $P(\lvert X - 80\rvert \geq 10) \leq \dfrac{50}{100} = 0,5$
Подставить значения m_X, $D(X)$ и ε в неравенство Чебышева $P(\lvert X - m_X\rvert < \varepsilon) \geq 1 - \dfrac{D(X)}{\varepsilon^2}$, так как надо оценить вероятность того, что НСВ отличается от среднего на величину, меньше, чем ε	б) $P(\lvert X - 80\rvert < 15) \geq 1 - \dfrac{50}{225} = 1 - 0,22 = 0,78$

Комплексные умения и алгоритмы к гл. 3 «Элементы математической статитики»

<table>
<tr><th>№ п/п</th><th>Умения</th><th>Алгоритмы</th></tr>
<tr><td>18</td><td>Построение вариационного ряда, эмпирической функции распределения и ее графика — кумуляты</td><td>
1. Для построения точечного вариационного ряда расположить заданные значения варианты x_i в порядке возрастания, одинаковые значения объединить и найти их соответствующие частоты.
Точечный вариационный ряд абсолютных частот примет вид:
<table>
<tr><td>x_i</td><td>x_1</td><td>...</td><td>x_k</td></tr>
<tr><td>n_i</td><td>n_1</td><td>...</td><td>n_k</td></tr>
</table>
2. Найти их соответствующие относительные частоты (статистические вероятности) $p_i^* = \frac{n_i}{n}$, где $1 \le i \le k$; причем $\sum_{k=1}^{n} n_k = n$; $\sum_{k=1}^{n} \frac{n_k}{n} = 1$.
Вариационный ряд относительных частот примет вид:
<table>
<tr><td>x_i</td><td>x_1</td><td>...</td><td>x_n</td></tr>
<tr><td>p_i^*</td><td>p_1^*</td><td>...</td><td>p_n^*</td></tr>
</table>
3. Составить эмпирическую функцию распределения, для чего найти накопительные частоты (плотность вероятности) на каждом интервале, просуммировав их по формуле:

$$F^*(x) = \begin{cases} 0 & \text{при } x < x_1; \\ \frac{n_1}{n} & \text{при } x_1 < x \le x_2; \\ \frac{(n_1 + n_2)}{n} & \text{при } x_2 < x \le x_3; \\ \frac{(n_1 + n_2 + n_3)}{n} & \text{при } x_3 < x \le x_4; \\ \ldots & \ldots \\ \frac{(n_1 + n_2 + \ldots + n_{k-1})}{n} & \text{при } x_{k-1} < x \le x_k; \\ 1 & \text{при } x > x_k. \end{cases}$$
</td></tr>
</table>

№ п/п	Умения	Алгоритмы
		4. Для построения кумуляты начертить декартову систему координат и отложить на оси абсцисс все возможные значения варианты $x_1, x_2, ..., x_k$, а на оси ординат — соответствующие значения функции $F^*(x) = \frac{1}{n}\sum_{i=1}^{k} n_i$. 5. Построить график эмпирической функции распределения, состоящий из отрезков с ординатами p_i, построенными на соответствующих интервалах $1 \le i \le k$ (для дискретных вариант) или соединить отрезками полученные точки (для непрерывных вариант)
19	Построение полигона и гистограммы	1. Для построения вариационного ряда расположить заданные значения варианты x_i в порядке возрастания, одинаковые значения объединить и найти их соответствующие частоты. 2. Одинаковые значения объединить и найти их соответствующие частоты (статистические вероятности) $p_i = \frac{n_i}{n}$, где $1 \le i \le k$. Вариационный ряд относительных частот примет вид: $\begin{array}{\|c\|c\|c\|c\|} \hline x_i & x_1 & \ldots & x_n \\ \hline p_i & p_1 & \ldots & p_n \\ \hline \end{array}$ 3. Отложить на оси ординат абсолютные частоты $n_1, n_2, ..., n_k$ или относительные частоты $p_1^*, p_2^*, ..., p_k^*$. 4. Для построения графика полигона относительных (абсолютных) частот необходимо соединить полученные точки с координатами $(x_1; p_1^*), (x_2; p_2^*), ..., (x_n; p_k^*)$ $[(x_1; n_1), (x_2; n_2), ..., (x_n; n_k)]$ отрезками прямых. 5. Для построения интервального вариационного ряда найти «размах» выборки — ее границы, т. е. x_{max} и x_{min}, $h = \frac{x_{max} - x_{min}}{k}$; k — число интервалов так, чтобы в каждом было не менее пяти значений варианты. При подсчете частоты признака в интервал включа-

№ п/п	Умения	Алгоритмы
		ют только начало интервала, поскольку конец является началом следующего интервала. В случае, если признак встретился не более пяти раз ($n_i \le 5$), надо объединить соседние интервалы. 6. Для построения гистограммы на оси OX откладывают полученные интервалы. Гистограмма состоит из прямоугольников, построенных на этих интервалах, высотами которых являются соответствующие этим интервалам значения частот (абсолютных или относительных). Для составления вариационного ряда нужно: 1) найти минимальное и максимальное значения выборки $x_{\min}$ и $x_{\max}$; 2) в первой строке таблицы записать варианты данной генеральной совокупности (выборки) в порядке возрастания; 3) во второй строке записать значение частоты, соответствующей данной варианте
20	Вычисление точечной оценки параметров распределения по выборке	1. Построить вариационный ряд по алгоритму 18. 2. Вычислить числовые характеристики: а) выборочное среднее (*несмещенная оценка*): $\bar{x} = \sum\limits_{j=1}^{k} x_j \frac{n_j}{n} = \frac{1}{n}\sum\limits_{j=1}^{k} x_j n_j$; б) выборочную дисперсию (*смещенная оценка при* $n > 30$): $D = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^2 - \bar{x}^2$ — для выборки, заданной вариационным рядом; $D = \frac{1}{n}\sum\limits_{j=1}^{k} x_j^2 n_j - \bar{x}^2 = \sum\limits_{j=1}^{k}(x_j - \bar{x})^2 \frac{n_j}{n}$ или $D = \sum\limits_{j=1}^{k} x_j^2 \frac{n_j}{n} - \bar{x}^2$ — для выборки, заданной таблицей; в) выборочное среднеквадратическое отклонение (*смещенная оценка при* $n > 30$) $\sigma = \sqrt{D}$

№ п/п	Умения	Алгоритмы
21	Вычисление точечной несмещенной оценки для дисперсии	1. Вычислить смещенную точечную оценку для дисперсии по формулам алгоритма 20. 2. Вычислить несмещенную точечную оценку для дисперсии по формуле: $s^2 = \frac{n}{n-1}D$. 3. Вычислить несмещенную точечную оценку для среднеквадратического отклонения $\sigma = \sqrt{s^2}$
22	Нахождение с помощью статистических таблиц интервала, в который с заданной вероятностью попадает случайная величина, распределенная нормально или по Стъюденту	1. Выписать заданные в условии задачи значения или найти их по алгоритмам 20 и 21, формулу для интервала, а также используемые табличные значения, указав номер таблицы. 2. а) Пусть величина распределена нормально $X \sim N(m, \sigma)$ при $\gamma \to 1$. Использовать формулу $P\left\{\lvert\bar{x} - m\rvert < t\frac{\sigma}{\sqrt{n}}\right\} = \gamma$ или $P\left(\bar{x} - t\frac{\sigma}{\sqrt{n}} < m < \bar{x} + t\frac{\sigma}{\sqrt{n}}\right) = \gamma$, где t_γ определяют по табл. П. 5. Так, для $\gamma = 0{,}95$ эта формула обращается в правило «*двух сигм*» ($t_{0,95} = 2$), а для $\gamma = 0{,}997$ — в правило «*трех сигм*» ($t_{0,997} = 3$): $p = (\lvert\xi - a\rvert \le 2\sigma) = 2\Phi(2) = 0{,}954;$ $p = (\lvert\xi - a\rvert \le 3\sigma) = 2\Phi(3) = 0{,}997.$ б) Пусть величина распределена нормально — $N(m, \sigma)$ при $\alpha \to 0$. Найти значения α по табл. П. 5, отыскивая α в третьем столбце, или по одной из соответствующих данному α формул: $\alpha = P\{\xi < m - 1{,}65\sigma\} = P\{\xi > m + 1{,}65\sigma\} = 0{,}05;$ $\alpha = P\{\xi < m - 2\sigma\} = P\{\xi > m + 2\sigma\} = 0{,}025;$ $\alpha = P\{\xi < m - 2{,}58\sigma\} = P\{\xi > m + 2{,}58\sigma\} = 0{,}005.$ в) Величина задана распределенной по Стъюденту при $\gamma \to 1$. Найти $t_{n,\gamma}$ из формулы $\int_{-t_{n,\gamma}}^{+t_{n,\gamma}} f_n(x)dx = \gamma$ по табл. П. 6, отыскивая нужное γ в верхней

№ п/п	Умения	Алгоритмы
		строке таблицы, а $t_{n,\gamma}$ — в строке, соответствующей данному n. г) Величина задана распределенной по Стьюденту при $\alpha \to 0$. Использовать формулу $\alpha = \int\limits_{-\infty}^{-t_{n,\alpha}} f_n(x)dx = \int\limits_{+t_{n,\alpha}}^{\infty} f_n(x)dx$ и найти $t_{n,\alpha}$ с помощью табл. П. 6, отыскивая нужное α в нижней строке таблицы, а $t_{n,\alpha}$ — в строке, соответствующей данному n. 3. Выписать полученный интервал
23	Вычисление доверительных интервалов для математического ожидания m нормального распределения	1. Сосчитать выборочное среднее, выборочное среднеквадратическое отклонение (если не известно истинное). 2. Выписать нужную формулу доверительного интервала для математического ожидания m нормального распределения с уровнем доверия γ для случая, когда среднеквадратическое отклонение σ распределения: а) *известно*: $\bar{x} - t_\gamma \frac{\sigma}{\sqrt{n}} < m < \bar{x} + t_\gamma \frac{\sigma}{\sqrt{n}}$, где t_γ находят по табл. П. 5 нормального распределения по заданному уровню доверия γ; б) *неизвестно*: использовать вместо σ его эмпирическую оценку s с заменой в формуле доверительного интервала n на $n - 1$ и вместо значений нормального распределения — значения $t_{n-1,\gamma}$ распределения Стьюдента с $n - 1$ степенями свободы (см. табл. П. 6): $\bar{x} - t_{n-1,\gamma} \frac{D}{\sqrt{n-1}} < m < \bar{x} + t_{n-1,\gamma} \frac{D}{\sqrt{n-1}}$, где $t_{n-1,\gamma}$ находят по табл. П. 6 или через несмещенную оценку s: $\bar{x} - t_{n-1,\gamma} \frac{s}{\sqrt{n}} < m < \bar{x} + t_\gamma \frac{s}{\sqrt{n}}$. 3. Найти точность оценки по формуле $\delta = t_\gamma \frac{s}{\sqrt{n}}$.

№ п/п	Умения	Алгоритмы
		4. Выписать полученный доверительный интервал по табл. П. 5, П. 6 или П. 7, вычислив границы требуемого в задании интервала по формуле $\bar{x} - \delta < m < \bar{x} + \delta$
24	Вычисление доверительных интервалов для генеральной дисперсии *D* и среднеквадратического отклонения σ	1. Выписать формулу, заданные в условиях задачи значения. 2. Найти пограничные значения вероятности для $\alpha = 1 - \gamma = 0,1$. 3. Найти по таблицам χ^2-распределения пограничные значения χ^2 для $\nu = n - 1$ числа степеней свободы, зная значения $P(\chi^2 \geq \chi_1^2)$ и $P(\chi^2 \leq \chi_2^2)$. 4. Найти значения для χ как $\chi_1 = \sqrt{\chi_1^2}$ и $\chi_2 = \sqrt{\chi_2^2}$. 5. Вычислить границы интервала по формуле $\frac{\sqrt{n}S}{\chi_2} \leq \sigma \leq \frac{\sqrt{n}S}{\chi_1}$. 6. Выписать полученный доверительный интервал
25	Вычисление с помощью таблиц нормального распределения доверительного интервала для вероятности *p* наступления события *A*	1. Вычислить оценку p^* для *p*. 2. Найти доверительный интервал для *p* по формуле $p^* - t_\gamma\sqrt{\frac{p^*(1-p^*)}{n}} \leq p \leq p^* + t_\gamma\sqrt{\frac{p^*(1-p^*)}{n}}$, где $t_\gamma = 1,96$ для уровня доверия 95 % и $t_\gamma = 3$ для уровня доверия 99,7 %. 3. Для проверки гипотезы сформулировать вывод из эксперимента, провести вычисления с доверительным интервалом

Задачи на применение алгоритмов

Алгоритм 18.

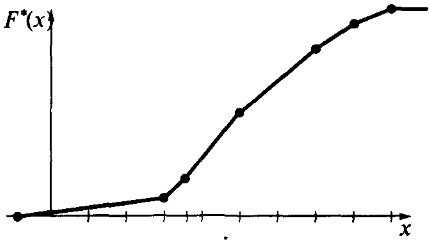
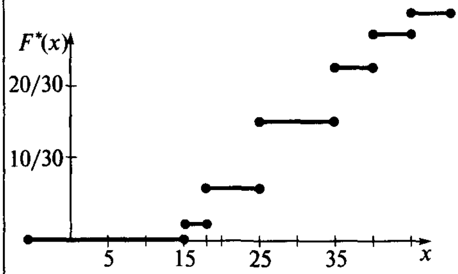
Построение вариационного ряда, эмпирической функции распределения и ее графика — кумуляты

Задача 18. Построить вариационный ряд, эмпирическую функцию распределения и ее график — кумуляту некоторой выборки для $15 \leq X \leq 64$

распределения времени на подготовку к экзамену по математике (мин): 15; 18; 45; 15; 25; 35; 40; 25; 35; 25; 35; 35; 45; 45; 25; 18; 25; 35; 35; 18; 40; 18; 40; 50; 35; 25; 25; 35; 40; 25.

<table>
<tr><th>№ п/п</th><th>Алгоритмы</th><th>Конкретное соответствие данной ситуации предложенному алгоритму</th></tr>
<tr><td>1</td><td>Для построения вариационного ряда расположить заданные значения варианты x_i в порядке возрастания, одинаковые значения объединить и найти их соответствующие частоты (статистические вероятности)</td><td>Расположим заданные значения варианты x_i в порядке возрастания: 15, 15, 18, 18, 18, 18, 25, 25, 25, 25, 25, 25, 25, 25, 35, 35, 35, 35, 35, 35, 35, 35, 40, 40, 40, 40, 40, 45, 45, 45. Вариационный ряд абсолютных частот примет вид:
<table>
<tr><td>x_i</td><td>15</td><td>18</td><td>25</td><td>35</td><td>40</td><td>45</td></tr>
<tr><td>n_i</td><td>2</td><td>4</td><td>8</td><td>8</td><td>5</td><td>3</td></tr>
</table></td></tr>
<tr><td>2</td><td>Найти соответствующие относительные частоты (статистические вероятности) $p_i = \frac{n_i}{n}$, где $1 \le i \le k$. Вариационный ряд относительных частот примет вид:
<table>
<tr><td>x_i</td><td>x_1</td><td>...</td><td>x_n</td></tr>
<tr><td>p_i</td><td>p_1</td><td>...</td><td>p_n</td></tr>
</table></td><td>Вариационный ряд относительных частот примет вид:
<table>
<tr><td>x_i</td><td>15</td><td>18</td><td>25</td><td>35</td><td>40</td><td>45</td></tr>
<tr><td>p_i</td><td>$\frac{2}{30}$</td><td>$\frac{4}{30}$</td><td>$\frac{8}{30}$</td><td>$\frac{8}{30}$</td><td>$\frac{5}{30}$</td><td>$\frac{3}{30}$</td></tr>
</table></td></tr>
<tr><td>3</td><td>Составить эмпирическую функцию распределения, для чего найти накопительные частоты (плотность вероятности) на каждом интервале, просуммировав их по формуле
$$F^*(x) = \begin{cases} 0; \\ n_1/n; \\ (n_1+n_2)/n; \\ (n_1+n_2+n_3)/n; \\ \dots \\ (n_1+n_2+\dots+n_{k-1})/n; \\ 1 \end{cases}$$</td><td>Составим эмпирическую функцию распределения:
$$F^*(x) = \begin{cases} 0 & \text{при } x \le 15; \\ \frac{2}{30} & \text{при } 15 < x \le 18; \\ \frac{6}{30} & \text{при } 18 < x \le 25; \\ \frac{14}{30} & \text{при } 25 < x \le 35; \\ \frac{22}{30} & \text{при } 35 < x \le 40; \\ \frac{27}{30} & \text{при } 40 < x \le 45; \\ 1 & \text{при } x > 45 \end{cases}$$</td></tr>
</table>

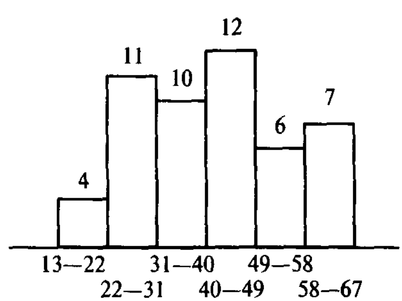
Окончание табл.

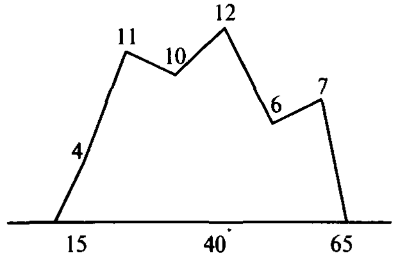
№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
4	Для построения кумуляты начертить декартову систему координат и отложить *на оси абсцисс все возможные* значения варианты $x_1, x_2, ..., x_k$, а на оси ординат — соответствующие значения функции $F^*(x)$	
5	Построить график эмпирической функции распределения, состоящий из отрезков с ординатами p_i, построенными на соответствующих интервалах $1 \leq i \leq k$ (для дискретных вариант) или соединить отрезками полученные точки (для непрерывных вариант)	

Алгоритм 19.
Построение полигона и гистограммы

Задача 19. *Построить гистограмму и полигон частот некоторой выборки* $15 \leq X \leq 64$ для распределения времени на сдачу экзамена по математике (мин).

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Для построения вариационного ряда расположить заданные значения варианты x_i в порядке возрастания	Расположим заданные значения варианты x_i в порядке возрастания 15,16, 18, 19, 23, 24, 25, 26, 26, 27, 27, 27, 28, 28, 28, 31, 32, 33, 34, 35, 36, 36, 36, 37, 37, 39, 40, 40, 41, 43, 45, 45, 46, 47, 48, 50, 52,53, 54,55, 51, 60, 61, 63, 64
2	Найти число вариант (объем выборки)	$n = 50$
3	Формула Стерджесса: $k = 1 + 3{,}322 \lg n$	$k \approx 1 + 3{,}322 \lg n = 6{,}64$ Возьмем, например, $k = 6$

<table>
<tr><th>№ п/п</th><th>Алгоритмы</th><th>Конкретное соответствие данной ситуации предложенному алгоритму</th></tr>
<tr><td>4</td><td>Найти минимальное и максимальное значения выборки — $x_{\min}$ и $x_{\max}$ и размах выборки $x_{\max} - x_{\min}$</td><td>$x_{\max} = 64$
$x_{\min} = 15$
$x_{\min} - x_{\max} = 49$</td></tr>
<tr><td>5</td><td>Найти удобное значение интервала $h = \dfrac{x_{\max} - x_{\min}}{k}$</td><td>$h = \dfrac{49}{6} \approx 8{,}17$, поэтому удобно взять $h = 9$, тогда все интервалы покроют 54 единицы</td></tr>
<tr><td>6</td><td>Выбрать удобную нижнюю границу интервала, исходя из примерной формулы $x_{\text{нижн}} = x_{\min} - \dfrac{kh - n}{2}$</td><td>$x_{\text{нижн}} \approx 15 - \dfrac{9 \cdot 6 - 49}{2} = 12{,}5$
$x_{\text{нижн}} = 13$</td></tr>
<tr><td>7</td><td>Составить интервалы $\Delta_i = [x_{\text{нижн}} + (i-1)\cdot h;\ x_{\text{нижн}} + i \cdot h)$, $i = 1 \ldots k$
Вариационный ряд относительных частот примет вид:
<table>
<tr><td>Δ_i</td><td>Δ_1</td><td>…</td><td>Δ_k</td></tr>
<tr><td>n_i</td><td>n_1</td><td>…</td><td>n_k</td></tr>
</table></td><td>Вариационный ряд относительных частот примет вид:
<table>
<tr><td>Δ_i</td><td>13—12</td><td>22—31</td><td>31—40</td><td>40—49</td><td>49—58</td><td>58—67</td></tr>
<tr><td>n_i</td><td>4</td><td>11</td><td>10</td><td>12</td><td>6</td><td>7</td></tr>
</table></td></tr>
<tr><td>8</td><td>Для построения гистограммы на оси OX откладывают полученные интервалы. Гистограмма состоит из прямоугольников, построенных на этих интервалах, высотами которых являются соответствующие этим интервалам значения частот (абсолютных или относительных).
Отложить на оси ординат абсолютные $n_1, n_2, \ldots, n_k$ (или относительные частоты $p_1^*, p_2^*, \ldots, p_k^*$)</td><td>
</td></tr>
<tr><td>9</td><td>Для построения графика полигона относительных (абсолютных) частот необ-</td><td></td></tr>
</table>

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
	ходимо соединить середины интервалов $(x_{\text{нижн}} + (i - 1) \cdot h;\ n_i)$, $i = 1 \ldots k$ отрезками прямых	

Алгоритм 20.

Вычисление точечной оценки параметров распределения по выборке

Задача 20. Процент выполнения плана по уборке урожая характеризуется данными из таблицы:

Процент выполнения плана	90—100	100—110	110—120	120—130	130—140	
Середина интервала	95	105	115	125	135	
Количество рабочих	10	160	100	60	20	$\sum = 350$

Найти смещенную и несмещенную оценку для дисперсии выполнения норм выработки и 95%-й доверительный интервал для генерального среднеквадратического отклонения. Проверить гипотезу о том, что среднеквадратическое отклонение выполнения плана равно 10 (уровень значимости 0,05).

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Построить вариационный ряд по алгоритму 18	Интервальный вариационный ряд задан, поэтому найдем середины интервалов и построим точечный вариационный ряд (указан в условии задачи)
2	Вычислить: а) выборочное среднее (*несмещенная оценка*): $\bar{x} = \sum_{j=1}^{k} x_j \frac{n_j}{n};$	а) Данные представлены таблицей и сгруппированы, поэтому, чтобы получить выборочное среднее, воспользуемся второй формулой: $\bar{x} = \frac{1}{350} = (95 \cdot 10 + 105 \cdot 160 + 115 \cdot 100 +$

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
2	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i n_i = \sum_{j=1}^{k} x_j p_j^*$ — для выборки, заданной таблицей группировки; б) смещенную точечную оценку для дисперсии по формулам: $D = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$ — для выборки, заданной вариационным рядом; $D = \frac{1}{n}\sum_{j=1}^{k} x_j^2 n_j - \bar{x}^2 = \sum_{j=1}^{k}(x_j - \bar{x})^2 \frac{n_j}{n}$ или $D = \sum_{j=1}^{k} x_j^2 \frac{n_j}{n} - \bar{x}^2$ — для выборки, заданной таблицей; в) выборочное среднеквадратическое отклонение (*смещенная оценка*) $\sigma = \sqrt{D}$	$+ 125 \cdot 60 + 135 \cdot 20) = \frac{1}{350} 39\,450 = 112{,}71$ б) $D = \frac{1}{350}(95^2 \cdot 10 + 105^2 \cdot 160 + 115^2 \cdot 100 + 125^2 \cdot 60 + 135^2 \cdot 20) - 112{,}71^2 = \frac{1}{350} \cdot 4\,478\,750 - 12\,703{,}544 = 92{,}884$; в) $\sigma = \sqrt{92{,}884} = 9{,}637$

Алгоритм 21.

Вычисление точечной несмещенной оценки для дисперсии

Задача 21. По условию задачи 20 найти точечную несмещенную оценку для дисперсии.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Вычислить смещенную точечную оценку для дисперсии по формулам алгоритма 20	$D = \frac{1}{350}(95^2 \cdot 10 + 105^2 \cdot 160 + 115^2 \cdot 100 + 125^2 \cdot 60 + 135^2 \cdot 20) - 112{,}71^2 = 92{,}884$

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
2	Вычислить несмещенную точечную оценку для дисперсии по формуле $s^2 = \frac{n}{n-1} D$	$s^2 = \frac{350}{349} \cdot 92,884 = 93,15$
3	Вычислить несмещенную точечную оценку для среднеквадратического отклонения $\sigma = \sqrt{s^2}$	$\sigma = \sqrt{93,15} = 9,65$ В этой задаче несмещенную оценку можно было не вычислять, так как значение n — объем выборки достаточно большой ($n > 30$)

Алгоритм 22.

Нахождение с помощью статистических таблиц интервала, в который с заданной вероятностью попадает СВ, распределенная нормально или по Стьюденту

Задача 22а. Найти симметричный относительно среднего значения интервал, в который величина $X \sim N(4,3)$ попадает с вероятностью 0,95.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Выписать заданные в условии задачи значения, формулу, а также используемые табличные значения, указав номер таблицы (см. прил. 1)	Дано: $X \sim N(4, 3)$, т. е. $m = 4$, $\sigma = 3$, $\gamma = 0,95$. Воспользуемся формулой (2а) алгоритма 22: $P(\lvert \bar{x} - m \rvert \le 1,96\sigma) = 2\Phi(t_\gamma) = \gamma$ ($t_{0,95} = 1,96$, см. табл. П. 5), имеем $P(\lvert \bar{x} - m \rvert \le 1,96\sigma) = 2\Phi(1,96) = 0,95$
2	Выписать полученный интервал, если σ *известно*: $\bar{x} - t_\gamma \frac{\sigma}{\sqrt{n}} < m < \bar{x} + t_\gamma \frac{\sigma}{\sqrt{n}}$	Интервал, в который $N(4, 3)$ попадает с вероятностью 0,95, имеет вид $m \in [4 - 1,96 \cdot 3; 4 + 1,96 \cdot 3]$ т.е. $m \in [-2; 10]$

Задача 22б. По результатам десяти котировок выявлено, что средний темп роста акций ОАО «Интеграл» составляет 105,43 %. Предполагая, что ошибка наблюдений распределена по нормальному закону со средним квадратическим отклонением 1 %, определить с надежностью 0,95 интервальную оценку для генеральной средней.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Выписать заданные в условии задачи значения, формулу, при условии, что σ *известно*: $\bar{x} - t_\gamma \frac{\sigma}{\sqrt{n}} < m < \bar{x} + t_\gamma \frac{\sigma}{\sqrt{n}}$, где t_γ определяют по табл. П. 5 нормального распределения по заданному уровню доверия	Имеем $X \sim N(m,1)$, т.е. $\bar{x} = 105{,}43$; $\sigma = 1$; $n = 10$; $\gamma = 0{,}95$. Воспользуемся формулой (2а) из алгоритма 22 для вычисления доверительного интервала математического ожидания m нормального закона распределения по средней выборочной: $P(\lvert\bar{x} - m\rvert \le 1{,}96\sigma) = 2\Phi(t_\gamma) = \gamma$, где $t_{0{,}95} = 1{,}96$ (см. табл. П. 5). Имеем $P(\lvert\bar{x} - m\rvert \le 1{,}96\sigma) = 2\Phi(1{,}96) = 0{,}95$
2	Выписать полученный интервал	Интервал, в который $N(m,1)$ попадает с вероятностью 0,95, имеет вид: $105{,}43 - 1{,}96\frac{1}{\sqrt{10}} < m < 105{,}43 + 1{,}96\frac{1}{\sqrt{10}}$, или $105{,}43 - 0{,}62 < m < 105{,}43 + 0{,}62$, т.е. $104{,}81 < m < 106{,}05$. Окончательно, $m \in (104{,}81;\ 106{,}05)$

Задача 22в. Найти односторонний интервал $(-\infty, -t]$, в который величина ξ, имеющая распределение Стъюдента с девятью степенями свободы, попадает с вероятностью 0,05, а также односторонний интервал $[t, \infty)$.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Выписать формулу, заданные в условиях задачи значения, а также используемые табличные значения, указав номер таблицы	Имеем: $\alpha = 0{,}05$; $X \sim t_9$ (распределение Стьюдента с девятью степенями свободы); $\alpha = 0{,}05$, $n = 9$. Воспользуемся формулой (2г) алгоритма 22 для вычисления доверительного интервала математического ожидания m распределения Стьюдента по средней выборочной: $$\alpha = \int_{-\infty}^{-t_{n,\alpha}} f_n(x)dx = \int_{+t_{n,\alpha}}^{\infty} f_n(x)dx.$$ Надо найти $t_{n,\alpha}$ по табл. П. 6, отыскивая нужное α в нижней строке таблицы, а $t_{n,\alpha}$ — в строке, соответствующей данному n. Из табл. П. 6 при $\alpha = 0{,}05$ и $n = 9$ имеем $t = 1{,}83$

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
2	Выписать полученный интервал	Существует два односторонних интервала, вероятность попадания в каждый из которых величины, распределенной по Стъюденту с девятью степенями свободы, равна 0,05: $(-\infty, -1{,}83]$ и $[1{,}83, \infty)$

Алгоритм 23.
Вычисление доверительных интервалов для математического ожидания m нормального распределения

Задача 23. Для случайно отобранных семи выпускников нашего колледжа стаж работы по специальности оказался равным: 10, 3, 5, 12, 11, 7, 9. Чему равен для них средний стаж и чему равен разброс (среднеквадратическое отклонение)? Найти 95%-й доверительный интервал для генерального среднего.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Сосчитать выборочное среднее, выборочное среднеквадратическое отклонение (если не известно истинное). Вычислить по алгоритму 19: а) выборочное среднее (*несмещенная оценка*): $\bar{x} = \sum_{j=1}^{k} x_j \frac{n_j}{n}$; б) выборочную дисперсию (*смещенная оценка*); $D = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$ — для выборки, заданной вариационным рядом	Построим вариационный ряд по алгоритму 18: x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 3 | 5 | 7 | 9 | 10 | 11 | 12 Найдем несмещенную оценку среднего стажа: $\bar{x} = \frac{10+3+5+12+11+7+9}{7} = 8{,}1$ [года], а также D и σ — смещенные оценки отклонений от среднего стажа: $D^* = 1/7(10^2 + 3^2 + 5^2 + 12^2 + 11^2 + 7^2 + 9^2) - 8{,}1 = 75{,}57 - 65{,}61 = 9{,}96$, значит, $\sigma^* = \sqrt{9{,}96} = 3{,}16$ [года]

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
2	Вычислить по алгоритму 21 несмещенные точечные оценки для дисперсии по формуле: $s^2 = \frac{n}{n-1} D$ и для среднеквадратического отклонения по формуле $s = \sqrt{s^2}$	$s^2 = \frac{7}{6} \cdot 9,96 = 11,62;$ $s = \sqrt{11,62} = 3,4$
3	Выписать нужную формулу доверительного интервала для математического ожидания m нормального распределения с уровнем доверия γ для случая, когда среднеквадратическое отклонение распределения σ *неизвестно*: вместо значений нормального распределения использовать значения $t_{n-1,\gamma}$ распределения Стъюдента с $n-1$ степенями свободы (см. табл. П. 6) $\bar{x} - t_{n-1,\gamma}\frac{\sigma}{\sqrt{n-1}} < m <$ $< \bar{x} + t_{n-1,\gamma}\frac{\sigma}{\sqrt{n-1}}$, где $t_{n-1,\gamma}$ определяют по табл. П. 6 или через несмещенную оценку s: $\bar{x} - t_{n-1,\gamma}\frac{s}{\sqrt{n}} < m <$ $< \bar{x} + t_{n-1,\gamma}\frac{s}{\sqrt{n}}$	Имеем при σ *неизвестном*: $\bar{x} - t_{n-1,\gamma}\frac{s}{\sqrt{n}} < m < \bar{x} + t_{n-1,\gamma}\frac{s}{\sqrt{n}}$, для $t_{n-1,\gamma} =$ $= t_{6,0.95} = 2,45$, где число степеней свободы $n-1 = 7-1 = 6$, а уровень доверия $\gamma =$ $= 0,95$. Найдем соответствующий доверительный интервал через несмещенную оценку s по формуле $\bar{x} - t_{n-1,\gamma}\frac{s}{\sqrt{n}} < m < \bar{x} + t_{n-1,\gamma}\frac{s}{\sqrt{n}}$: $8,1 - 2,45 \cdot \frac{3,4}{\sqrt{7}} < m < 8,1 + 2,45 \cdot \frac{3,4}{\sqrt{7}};$ $8,1 - 3,14 < m < 8,1 + 3,14$

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
4	Вычислить границы доверительного интервала для генеральной дисперсии по табл. П. 6	Найдем соответствующий доверительный интервал: $4{,}96 < m < 11{,}24$

Алгоритм 24.

Вычисление доверительных интервалов для генеральной дисперсии D и среднеквадратического отклонения σ

Задача 24. По данным Гидрометцентра за последние 25 лет среднесуточная температура в середине октября имеет среднеквадратическое отклонение $S = 8$ °C. Учитывая, что ошибка подчинена нормальному закону распределения, определить с надежностью $\gamma = 0{,}9$ доверительный интервал для неизвестного параметра σ.

Решение.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
1	Выписать формулу, заданные в условиях задачи значения	По условию задачи $n = 25 < 30$, поэтому воспользуемся формулами $\frac{\sqrt{n}S}{\chi_2} \le \sigma \le \frac{\sqrt{n}S}{\chi_1}$ для $S = 8$
2	Найти пограничные значения вероятности, зная, что $\alpha = 1 - \gamma$	Для $\alpha = 1 - \gamma = 0{,}1$ имеем: $P(\chi^2 \ge \chi_1^2) = 1 - \frac{\alpha}{2} = 1 - \frac{0{,}1}{2} = 0{,}995$; $P(\chi^2 \le \chi_2^2) = \frac{\alpha}{2} = \frac{0{,}1}{2} = 0{,}05$
3	Найти по таблицам χ^2-распределения пограничные значения χ^2 для $\nu = n - 1$ числа степеней свободы, зная значения $P(\chi^2 \ge \chi_1^2)$ и $P(\chi^2 \le \chi_2^2)$	Для $\nu = n - 1 = 24$ числа степеней свободы, зная, что $P(\chi^2 \ge \chi_1^2) = 0{,}995$ и $P(\chi^2 \le \chi_2^2) = 0{,}05$, имеем $\chi_1^2 = 9{,}886$ и $\chi_2^2 = 36{,}415$
4	Найти значения для χ как $\chi_1 = \sqrt{\chi_1^2}$ и $\chi_2 = \sqrt{\chi_2^2}$	$\chi_1 = \sqrt{9{,}886} = 3{,}14$ и $\chi_2 = \sqrt{36{,}415} = 6{,}03$

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие данной ситуации предложенному алгоритму
5	Найти границы интервала по формуле $\frac{\sqrt{n}S}{\chi_2} \le \sigma \le \frac{\sqrt{n}S}{\chi_1}$	$\frac{\sqrt{25}\cdot 8}{6,03} \le \sigma \le \frac{\sqrt{25}\cdot 8}{3,14}$
6	Выписать полученный интервал	$6,63 \le \sigma \le 12,73$

Алгоритм 25.

Вычисление с помощью таблиц нормального распределения доверительного интервала для вероятности *p* наступления события *A*

Задача 25. Выборочная проверка показала, что из 100 выпускников 87 человек удовлетворяют требованиям стандарта по математике. Преподаватели хотят быть уверены на 95 %, что ошибаются в оценке процента учащихся, не усвоивших программу. В каких пределах он находится? Каков должен быть объем выборки, чтобы оценить процент брака с точностью до 0,1?

Решение.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
1	Вычислить оценку p^* для p	Имеем: $n = 100$; $m = 100 - 87 = 13$; $p^* = 0,13$
2	Найти доверительный интервал для p, используя формулу для повторной выборки с возвратом	$p^* - t_\gamma\sqrt{\frac{p^*(1-p^*)}{n}} \le p \le p^* + t_\gamma\sqrt{\frac{p^*-(1-p^*)}{n}}$, при $t_\gamma = 1,96$ имеем $0,13 - 1,96\sqrt{\frac{0,13\cdot 0,87}{100}} \le$ $\le p \le 0,13 + 1,96\sqrt{\frac{0,13\cdot 0,87}{100}}$. Подставив n и p, получаем $0,06 \le p \le 0,2$
3	Для проверки гипотезы сформулировать вывод из эксперимента, провести вычисле-	Неравенство $\lvert p - p^*\rvert \le 1,96\sqrt{\frac{p^*(1-p^*)}{n}}$ выполняется с доверительной вероятностью

Окончание табл.

№ п/п	Алгоритмы	Конкретное соответствие задания заданному алгоритму
	ния с доверительным интервалом	0,95 при $t_\gamma = 1{,}96$. Требуемая точность составляет 0,1, поэтому $1{,}96\sqrt{\dfrac{p^*(1-p^*)}{N}} = 0{,}1 \Rightarrow$ $N = \dfrac{p^*(1-p^*)\cdot(1{,}96)^2}{0{,}1^2} = 384{,}16\cdot 0{,}13\cdot 0{,}87 =$ $= 43{,}45 \approx 44.$ Таким образом, необходимо проверить знания 44 выпускников

СПИСОК ЛИТЕРАТУРЫ

Основная литература

Гмурман В. Е. Теория вероятностей и математическая статистика. — М.: Высшая школа, 2002.

Гмурман В. Е. Руководство к решению задач по теории вероятностей. — М.: Высшая школа, 2001.

Горелова Г. В. Теория вероятностей и математическая статистика. В примерах и задачах с применением EXCEL/Г. В. Горелова, И. А. Кацко. — Ростов-на-Дону: изд-во «Феникс», 2002.

Калинина В. Н. Математическая статистика / В. Н. Калинина, В. Ф. Панкин. — М.: Высшая школа, 2001.

Красс М. С. Основы математики и ее приложения в экономическом образовании / М. С. Красс, Б. П. Чупрынов. — М.: изд-во «Дело», 2002.

Теория статистики с основами теории вероятностей / И. И. Елисеева [и др.]. — М.: изд-во «Юнити», 2001.

Дополнительная литература

Прикладная статистика. Основы моделирования и первичная обработка данных / С. В. Айвазян [и др.]. — М.: Финансы и статистика, 1983.

Вентцель Е. С. Теория вероятности. — М.: Издательский центр «Академия», 2004.

Гнеденко Б. В. Курс теории вероятностей. — М.: Физматлит, 1961.

Данко П. Е. Высшая математика в упражнениях и задачах: в 2 т. — Т. 2 / П. Е. Данко, А. Г. Попов, Т. Я. Кожевникова. — М.: Высшая школа, 1996.

Колде Я. К. Практикум по теории вероятностей и математической статистике. — М.: Высшая школа, 1991.

Колемаев В. А. Теория вероятностей и математическая статистика / В. А. Колемаев, В. И. Калинина. — М.: ИНФРА-М, 1997.

Ивченко Г. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев — М.: Высшая школа, 1984.

Мацкевич И. П. Сборник задач и упражнений по высшей математике. Теория вероятностей и математическая статистика / И. П. Мацкевич, Г. П. Свирид, Г. М. Булдык. — Минск: Вышэйшая школа, 1996.

Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров. — М.: Инфра-М, 1995.

ОГЛАВЛЕНИЕ