Уральский
федеральный
университет
имени первого Президента
России Б. Н. Ельцина
Высшая школа
экономики
и менеджмента

Е. А. ТРОФИМОВА
Н. В. КИСЛЯК
Д. В. ГИЛЁВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИМЕНИ ПЕРВОГО ПРЕЗИДЕНТА РОССИИ Б. Н. ЕЛЬЦИНА

Е. А. Трофимова, Н. В. Кисляк, Д. В. Гилёв

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Рекомендовано
методическим советом Уральского федерального университета
в качестве учебного пособия для студентов вуза,
обучающихся по направлениям подготовки
38.03.01 «Экономика», 38.03.02 «Менеджмент»

Екатеринбург
Издательство Уральского университета
2018

УДК 519.2(075.8)
ББК 22.171я73-1
Т761

Р е ц е н з е н т ы:
М. Ю. Хачай, доктор физико-математических наук,
заведующий отделом математического программирования
(Институт математики и механики им. Н. Н. Красовского УрО РАН);
А. И. Кривоногов, кандидат физико-математических наук,
доцент кафедры прикладной математики
(Уральский государственный архитектурно-художественный университет)

П о д   о б щ е й   р е д а к ц и е й
Е. А. Трофимовой

Трофимова, Е. А.
Т761 Теория вероятностей и математическая статистика : учеб. пособие / Е. А. Трофимова, Н. В. Кисляк, Д. В. Гилёв ; [под общ. ред. Е. А. Трофимовой] ; М-во образования и науки Рос. Федерации, Урал. федер. ун-т. – Екатеринбург : Изд-во Урал. ун-та, 2018. – 160 с.

ISBN 978-5-7996-2317-3

В учебном пособии представлен блок теоретического материала и задачи, как подробно разобранные, так и предназначенные для самостоятельного решения. Каждому математическому понятию дается экономическая интерпретация.

Для студентов, изучающих дисциплину «Теория вероятностей и математическая статистика».

УДК 519.2(075.8)
ББК 22.171я73-1

ISBN 978-5-7996-2317-3

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Данное учебное пособие написано в соответствии с требованиями государственных образовательных стандартов третьего поколения по экономическим специальностям. Включает в себя кратко, но всесторонне изложенный теоретический материал с разобранными на каждую тему практическими заданиями, с объяснением экономического смысла каждого введенного понятия, а также задачи для самостоятельного решения. Данное пособие может быть использовано в качестве основной литературы для проведения лекций и практических занятий.

Предпосылками написания данного пособия являлась необходимость систематизировать накопленный материал при многолетнем прочтении лекций и проведении практических занятий у авторов данного пособия, а также для возможности иметь полностью укомплектованный, учитывающий новые разработки комплект, обеспечивающий дисциплину «Теория вероятностей и математическая статистика». При написании пособия были учтены современные требования и компетенции, предъявляемые к бакалавру экономики. Материал был подобран так, чтобы не только можно было уловить суть предмета, но и понять его назначение в современном мире. Особый уклон был сделан на экономические приложения. Содержание данного пособия не только целиком соответствует рабочей программе по дисциплине, но охватывает даже больше необходимого минимума. Некоторые темы приведены для самостоятельного разбора студентами.

Дисциплина «Теория вероятностей и математическая статистика» является важнейшей частью модуля «Математические методы анализа». Ее прикладная значимость в экономике достаточно велика. На ней зиждется эконометрика, многомерный статистический анализ, нейронные сети, распознавание образов и многие другие научные области. Современный экономист должен уметь использовать аппарат математической статистики на высоком уровне.

1. СЛУЧАЙНЫЕ СОБЫТИЯ

1.1. Предмет теории вероятностей

Можно сказать, что жизнь человека, да и в целом вся окружающая среда состоит из череды некоторых событий. Хотелось бы все эти события не только отслеживать, но и пытаться прогнозировать. При этом стоит понимать, что многие события или, другими словами, явления – случайные, т. е. они могут наступить, а могут не наступить. Например, выиграть в лотерею автомобиль – событие случайное.

Задача любой науки, в том числе и экономической, состоит в выявлении и исследовании закономерностей, которым подчиняются реальные процессы. Найденные закономерности, относящиеся к экономике, имеют не только теоретическую ценность, но и широко применяются на практике.

Выше был приведен пример лотереи. Можно задаться вопросом: зачем нужно изучать степень возможности наступления выигрыша в лотерее автомобиля? Возможно, обычному обывателю и не стоит ничего изучать, а вот организаторам без этого не обойтись. Ведь лотереи проводятся не просто так, а с целью извлечения выгоды (прибыли), а значит, заинтересованность большая. Важно установить, по какой цене нужно продавать билеты лотереи, чтобы получить планируемый доход. Аналогично, страховым компаниям важно понимать степень наступления страхового случая, чтобы сформировать верное страховое вознаграждение (по сути, прибыль компании). Ведь, если завысить страховое вознаграждение, то можно потерять определенную долю рынка, что крайне нежелательно, а если занизить – можно остаться банкротом. То есть смело можно заявить, что теория вероятностей необходима экономистам, чтобы *обезопасить себя от экономических рисков*.

Таким образом, очевидно, что необходимо уметь исследовать случайные явления и находить их закономерности. Этим как раз и занимается теория вероятностей.

При этом в названии пособия есть еще «математическая статистика» – что же это такое и где она применяется? Давайте для начала приведем определения, а затем уже поймем принципиальную разницу между теорией вероятностей и математической статистикой.

Т е о р и я   в е р о я т н о с т е й – это математическая наука, изучающая закономерности случайных явлений.

С л у ч а й н о е   я в л е н и е – это явление с неопределенным исходом, происходящее при неоднократном воспроизведении определенного комплекса условий.

Стоит отметить, что в природе, технике и экономике в каждом явлении присутствует случайность: в спросе на товар, в погодных условиях и прочем.

Существует два подхода к изучению явлений: «детерминистский» и «вероятностный». При первом подходе выделяют основные факторы, характеризующие явление, а при втором – учитывают, помимо основных факторов, второстепенные, которые, если их не учесть, как раз и приводят к случайным возмущениям и искажениям результата.

М а т е м а т и ч е с к а я   с т а т и с т и к а – это раздел математики, изучающий математические методы сбора, систематизации, обработки и интерпретации результатов наблюдений с целью выявления статистических закономерностей.

Таким образом, теория вероятностей изучает закономерности случайных явлений на основе абстрактного описания, а математическая статистика уже на основе этого описания оперирует непосредственно результатами конкретных наблюдений. Можно сказать, что теория вероятностей – это базис, фундаментальная надстройка математической статистики, которая уже применяется в реальной жизни. В связи с этим план изучения таков: сначала понять основные моменты теории вероятностей, а затем на их основе уже рассмотреть инструментарий математической статистики.

Как уже стало понятно, теория вероятностей позволяет находить степень объективной возможности наступления (вероятность) «сложных» событий через «простые», а математическая статистика по наблюдаемым значениям оценивает эту степень либо осуществляет проверку предположений (гипотез) относительно этой степени. Так или иначе, все завязано на событиях, поэтому сейчас настало время перейти к изучению их свойств и операций над ними.

1.2. Пространство элементарных исходов

С л у ч а й н ы м (возможным) называется событие, которое в результате опыта (испытания, эксперимента) может произойти или не произойти.

П р и м е р ы с о б ы т и й

1. Появление герба при подбрасывании монеты.
2. Выигрыш квартиры по билету лотереи.
3. Выход бракованного изделия с конвейера предприятия.

При этом стоит понимать, что событие – это не какое-то происшествие, а возможный исход. События, которыми может закончиться опыт, также называют исходами.

Принято обозначать события большими латинскими буквами: A, B, C.

Буквой Ω обозначается множество всех возможных исходов испытания. Оно бывает конечным или бесконечным (счетным или несчетным). При этом если оно конечное или конечное, но счетное, то множество исходов является дискретным, в случае несчетного – множество исходов непрерывное.

Если при каждом испытании, при котором происходит событие A, происходит и событие B, то говорят, что A в л е ч е т з а с о б о й B (т. е. A содержится в B). Обозначается этот факт так: $A \subset B$.

П р и м е р

Если $A = \{$выпало «2» на кубике$\}$; $B = \{$выпало четное число на кубике$\}$, то $A \subset B$.

Если $A \subset B$ и $B \subset A$, то A и B называются равносильными событиями. Обозначается: $A = B$.

События называются несовместными, если наступление одного из них исключает наступление любого другого. В противном случае события называются совместными.

Примеры

1. Выигрыш по одному билету двух призов в лотерее – события несовместные, а выигрыш двух призов по двум билетам – события совместные.

2. Получение «5» и «4» одновременно за один и тот же экзамен – события несовместные.

Событие называется достоверным, если в результате испытания оно обязательно должно произойти.

Событие называется невозможным, если в результате испытания оно вообще не может произойти. Обозначается такое событие так: $\varnothing$.

События называются равновозможными, если в результате испытания ни одно из этих событий не является объективно более возможным.

Пример

Появление герба или решки при подбрасывании монеты – события равновозможные, если монета «правильная» (т. е. выполнена симметрично).

Несколько событий называются единственно возможными, если в результате испытания обязательно должно произойти хотя бы одно из них.

Пример

Имеются три события, состоящие в том, что в семье из двух детей: $A = \{2 \text{ мальчика}\}$, $B = \{2 \text{ девочки}\}$, $C = \{1 \text{ мальчик, } 1 \text{ девочка}\}$. Эти три события являются единственно возможными.

Несколько событий образуют полную группу (полную систему) событий, если они являются единственно возможными и несовместными исходами испытания. Это означает, что в результате испытания обязательно должно произойти одно и только одно из этих событий.

Два несовместных события, из которых одно должно обязательно произойти, называются п р о т и в о п о л о ж н ы м и. Событие, противоположное исходному событию A, будем обозначать $\overline{A}$.

Стоит отметить, что противоположные события – это частный случай событий, образующих полную группу.

1.3. Операции над событиями и их свойства

Суммой (объединением) двух событий A и B называется новое событие $A + B$ (или: $A \cup B$), состоящее в появлении хотя бы одного из событий A или B в данном опыте.

Произведением (пересечением) двух событий A и B называется новое событие AB (или: $A \cap B$), состоящее в совместном появлении A и B в данном опыте.

Для наглядного понимания представленных двух операций обычно используют так называемые диаграммы Эйлера – Венна.

На рис. 1 изображена диаграмма Эйлера – Венна, прямоугольник на которой – это множество всех исходов, кругами представлены события A и B, а заштрихованная часть – сумма этих событий.

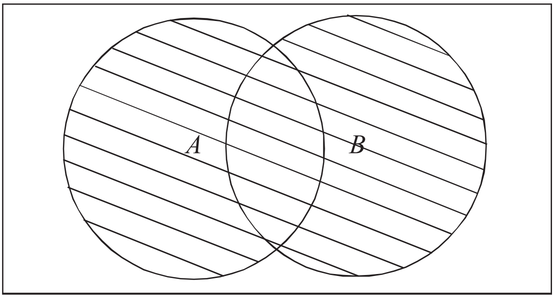

Рис. 1

На рис. 2 также изображена диаграмма Эйлера – Венна, но заштрихованная часть – произведение событий A и B.

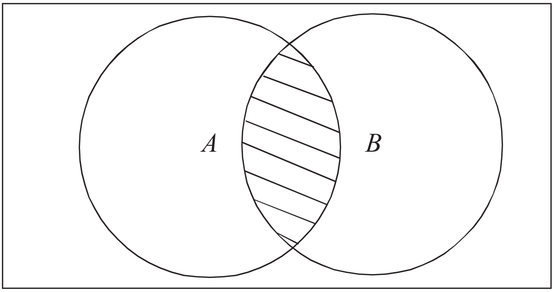

Рис. 2

Далее предлагаем ознакомиться со свойствами операций над произвольными событиями A, B, C:

1) $AB \subseteq A + B$;

2) $A\overline{A} = \varnothing$;

3) $A + \overline{A} = \Omega$;

4) $A + B = B + A$ (коммутативность суммы);

5) $AB = BA$ (коммутативность произведения);

6) $A(BC) = (AB)C = ABC$ (ассоциативность произведения);

7) $A + (B + C) = (A + B) + C = A + B + C$;

8) $(A + B)C = AC + BC$ (дистрибутивность суммы);

9) $AB + C = (A + C)(B + C)$ (дистрибутивность произведения);

10) $A + A = AA = A$ (идемпотентность);

11) $\overline{\overline{A}} = A$ (закон двойного дополнения);

12) $\overline{A + B} = \overline{A}\overline{B}$ (закон де Моргана);

13) $\overline{AB} = \overline{A} + \overline{B}$ (закон де Моргана).

Данные утверждения очевидно выполняются, если для каждого из них изобразить диаграмму Эйлера – Венна. Безусловно, самые старательные читатели могут доказать данные утверждения аналитически, особенно если знакомы с теорией множеств.

В связи с введенными операциями можем переформулировать определение полной группы событий следующим образом.

События A_1, A_2, …, A_n образуют **полную группу событий**, если:

1) они несовместны, т. е. $A_iA_j = \varnothing$ для любых $i = 1, \ldots, n$, $j = 1, \ldots, n$, $i \neq j$;

2) в результате испытания произойдет обязательно одно из событий, т. е. $A_1 + A_2 + \ldots + A_n = \Omega$.

Теперь, когда мы познакомились с событиями, операциями над ними и основными их свойствами, то можем перейти к более важному вопросу, а именно, к определению степени возможности наступления события.

1.4. Классическое определение вероятности

Как было сказано ранее, для практической деятельности важно уметь сравнивать события по степени возможности их наступления. Например, «выигрыш по одному билету» и «выигрыш по каждому из десяти приобретенных билетов» лотереи обладают разной степенью возможности их наступления.

Численная мера степени объективной возможности наступления события называется в е р о я т н о с т ь ю события.

Чтобы измерить эту «вероятность», используют несколько подходов. Начнем с классического.

Пусть исходы некоторого испытания образуют полную группу событий и равновозможны, т. е. единственно возможны, несовместны и равновозможны. Такие исходы называют э л е м е н т а р н ы м и и с х о д а м и, случаями или шансами. При этом говорят, что испытание сводится к с х е м е с л у ч а е в.

Случай называется благоприятным событию A, если появление этого случая ведет за собой появление события A.

Согласно к л а с с и ч е с к о м у о п р е д е л е н и ю вероятность события A равна отношению числа случаев, благоприятных ему, к общему числу случаев, т. е.:

$$p(A) = \frac{m}{n},$$

где $p(A)$ – вероятность события A (другое обозначение: $Pr(A)$ – от англ. *probability*);

m – число случаев, благоприятных событию A;

n – общее число случаев.

П р и м е р

Подбрасывается шестигранная игральная кость. Какова вероятность появления четного числа очков? Ответ на данный вопрос легко получить по классическому определению вероятности. Все, что нам нужно найти, – это число всех возможных и благоприятных исходов. В нашем случае число всех возможных исходов равно шести, а множество благоприятных исходов равно трем (так как четное может выпасть 2, 4, 6). Таким образом, получаем, что $p = 3/6 = 0{,}5$.

Однако не стоит забывать, что классическое определение работает только в *схеме случаев*, иначе его применение приведет к неверному ответу.

П р и м е р (з а д а ч а Д а л а м б е р а)

Монетка подбрасывается 2 раза. Какова вероятность того, что орел выпадет хотя бы один раз? Предлагается следующее решение. Так как монетку бросают 2 раза, то множество всех исходов: 0, 1, 2. Таким образом, число всех возможных исходов равно 3, а число благоприятных исходов равно 2, тогда вероятность $p = 2/3$. Эти рассуждения неверны, ибо орел выпадает один раз чаще, чем два раза или чем вообще не выпадает, т. е. исходы неравновозможны, а значит, классическое определение вероятности неприменимо при таком подходе. Однако если посчитать, сколько же на самом деле возможно исходов, то формула сработает. Действительно, при подбрасывании монетки дважды возможны следующие ситуации: орел – орел, решка – решка, орел – решка, решка – орел. Таким образом, число всех возможных исходов равно 4, а число благоприятных – 3. Получаем ответ согласно классическому определению (здесь оно уже применимо, ибо исходы равновозможны): $p = 3/4$.

Стоит отметить важные с в о й с т в а в е р о я т н о с т е й события A:

1) величина вероятности не может принимать значения большие единицы и меньшие нуля, т. е. $0 \le p(A) \le 1$;

2) вероятность достоверного события равна единице, т. е. $p(\Omega) = 1$;

3) вероятность невозможного события равна нулю, т. е. $p(\varnothing) = 0$.

Для применения классического определения вероятности необходимо знать число благоприятных и всех возможных исходов, в вышеприведенных примерах оно находится простым перечислением, но в большинстве случаев это бывает затруднительно. В этом случае на помощь приходят правила и формулы из комбинаторики, с которыми предлагается ознакомиться читателю в следующем параграфе.

1.5. Правила и формулы комбинаторики

Комбинаторика – это раздел математики, изучающий методы решения на подсчет числа различных комбинаций.

Задачи, в которых производится подсчет всех возможных комбинаций, составленных по некоторому правилу, называются комбинаторными. Раздел математики, занимающийся их решением, называется комбинаторикой.

Правило умножения. Если требуется выполнить одно за другим какие-то k действий, причем первое действие можно выполнить a_1 способами, второе – a_2 способами, и так до k-го действия, которое можно выполнить a_k способами, то все k действий вместе могут быть выполнены $a_1 \cdot a_2 \cdot a_3 \cdot \ldots \cdot a_k$ способами.

Правило сложения. Если два действия взаимно исключают друг друга, причем одно из них можно выполнить m способами, а другое – n способами, то выполнить одно любое из этих действий можно $m + n$ способами.

Это правило легко распространить на любое конечное число действий.

Число перестановок. Имеем n предметов, которые переставляются, тогда число возможных перестановок будет равно $n!$, где $n! = 1 \cdot 2 \cdot \ldots \cdot n$. Отметим, что $n!$ читается как «н-факториал», а по сути это просто сокращенная запись произведения чисел от 1 до n. Также стоит заметить, что договорились считать, что $0! = 1$.

Пример

Порядок выступлений семи участников конкурса определяется жребием. Сколько различных вариантов жеребьевки при этом возможно? Ответ на данный вопрос легко получить, если заметить, что каждый вариант жеребьевки отличается только порядком участников конкурса, т. е. является перестановкой из семи элементов, а значит, различных вариантов жеребьевки будет 7! = 5 040.

Число сочетаний. Сочетанием из n элементов по k называется любое подмножество из k элементов, которые принадлежат множеству, состоящему из n различных элементов. Более простыми словами можно сказать так: если мы имеем n предметов и хотим извлечь из них k предметов (причем порядок извлечения не важен), то количество возможных способов их извлечь есть число сочетаний, которое обозначается C_n^k и находится по формуле $C_n^k = \dfrac{n!}{k!(n-k)!}$.

П р и м е р

В шахматном турнире участвуют 16 человек. Сколько партий должно быть сыграно, чтобы между любыми двумя участниками была сыграна одна партия? Чтобы ответить на данный вопрос, заметим, что мы имеем 16 человек, из них выбирается по два, причем в каком порядке – не важно по условию задачи, а значит, количество партий легко найти с помощью числа сочетаний, а именно:

$$C_{16}^2 = \frac{16!}{2!(16-2)!} = \frac{1 \cdot 2 \cdot ... \cdot 16}{1 \cdot 2 \cdot 1 \cdot 2 \cdot ... \cdot 14} = 120.$$

Число размещений. Размещением из n элементов по k называется упорядоченный набор из k различных элементов из некоторого множества различных n элементов. Число размещений обозначается A_n^k и вычисляется по формуле

$$A_n^k = \frac{n!}{(n-k)!}.$$

П р и м е р

Расписание одного дня состоит из пяти пар. Определить число вариантов расписания при выборе из 11 дисциплин. Ответом является:

$$A_{11}^5 = \frac{11!}{(11-5)!} = 55\,440.$$

Отметим, что можно было решить данную задачу также с помощью правила умножения. Рассуждаем: первой парой возможно поставить одну из 11 дисциплин, второй парой – одну из 10 дисциплин и т. д. Таким образом, число вариантов расписания будет равно: $11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 55\,440$.

Число перестановок с повторениями. Имеем n предметов различных типов: n_1 элементов первого типа, n_2 элементов второго типа, …, n_k элементов k-го типа, которые переставляются, тогда число возможных перестановок будет вычисляться по формуле

$$P(n_1, n_2, ..., n_k) = \frac{n!}{n_1!\, n_2!\, ...\, n_k!}.$$

Число сочетаний с повторениями. Если при неупорядоченном выборе m элементов из n элементы возвращаются обратно (т. е. одни и те же элементы могут вынимать по несколько раз), то полученные выборки есть сочетания с повторениями, которые обозначаются $\overline{C}_n^k$ и вычисляются по формуле $\overline{C}_n^k = C_{n+k-1}^k$.

П р и м е р

В магазине имеется 3 вида пирожных. Сколькими способами можно купить 9 пирожных? Ответом на данный вопрос будет число сочетаний с повторениями: $\overline{C}_9^3 = C_{9+3-1}^3 = C_{11}^3 = \dfrac{11!}{3!\,(11-3)!} = \dfrac{11!}{3! \cdot 8!} = 165.$

Число размещений с повторениями. Если при упорядоченном выборе m элементов из n элементы возвращаются обратно, то полученные выборки называются размещением с повторениями, обозначаются $\overline{A}_n^k$ и вычисляются по формуле $\overline{A}_n^k = n^m$.

П р и м е р

В лифт шестнадцатиэтажного дома вошли 10 жильцов. Сколькими способами могут выйти пассажиры на каждом этаже, начиная со второго? Чтобы ответить на данный вопрос, поймем, что задача сводится к тому, чтобы распределить 10 жильцов по 15-ти этажам, причем возможны повторы, т. е. на одном этаже может выйти несколько человек. А значит, ответ находится по формуле числа размещений с повторениями: $\overline{A}_{15}^{10} = 15^{10}$.

И напоследок отметим несколько важных замечаний про число сочетаний:

$$C_n^0 = C_n^k = 1;\ \ C_n^1 = C_n^{n-1} = n;\ \ C_n^k = C_n^{n-k}.$$

1.6. Подсчет классической вероятности с помощью правил комбинаторики

Важно отметить, что классическое определение (формулу) следует рассматривать не как определение, а как метод вычисления вероятностей для испытаний, сводящихся к схеме случаев.

Для применения классического определения вероятностей необходимо знать число благоприятных и всех возможных исходов, которые как раз и возможно найти с помощью формул и правил комбинаторики. Рассмотрим несколько примеров.

Примеры

1. Слово «экономист» составлено из карточек, на каждой из которых написана одна буква. Карточки перетасовывают и вынимают по одной без возврата. Найти вероятность того, что в результате вынимания снова получится слово «экономист». Ответ легко получается при помощи формулы классической вероятности. Число благоприятных исходов $m = 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 = 2$ находится по правилу умножения, так как буквы «э», «к», «н», «м», «и», «с», «т» можно вытащить только одним способом, а букву «о» – двумя. Число всех исходов есть по сути перестановка девяти карточек, поэтому $n = 9!$, итак имеем: $p = \dfrac{2}{9!} = \dfrac{2}{362\,880} = \dfrac{1}{181\,440}$.

2. Среди 25 студентов, из которых 10 девушек, разыгрываются 3 билета, причем каждый может выиграть только один билет. Какова вероятность того, что среди победителей окажутся две девушки и один юноша? Число всех благоприятных исходов можно найти при помощи числа сочетаний и правила умножения: $m = C_{10}^2 \cdot C_{15}^1$. Число всех возможных исходов $n = C_{25}^3$. Тогда имеем: $p = \dfrac{C_{10}^2 \cdot C_{15}^1}{C_{25}^3}$.

3. Вовочка, набирая пароль от своей страницы в социальной сети, забыл последние три цифры и, помня лишь, что эти цифры различны,

набрал их наудачу. Найти вероятность того, что набраны нужные цифры.

Число благоприятных исходов равно 1, так как существует лишь единственный способ выбрать нужную комбинацию трех цифр. А вот число всех возможных вариантов можно найти с помощью числа размещений, так как порядок расположения цифр в пароле важен, а именно: $n = A_{10}^{3} = \frac{10!}{(10-3)!} = 8 \cdot 9 \cdot 10 = 720$. Тогда $p = \frac{1}{720}$.

4. Из 36 изделий 3 имеют скрытый дефект. Наугад выбрано 5 изделий. Найти вероятность, что среди выбранных изделий 2 имеют скрытый дефект.

Решение: $p = \frac{C_3^2\ C_{33}^3}{C_{36}^5}$. Здесь очень важно понимать, как получился числитель, т. е. как мы нашли число благоприятных исходов. Часто возникает путаница из-за слов в условии, что «среди выбранных 2 изделия имеют скрытый дефект», из-за чего некоторые принимают за число благоприятных исходов C_5^2, что, безусловно, неверно. На самом деле нужно рассуждать так: мы вытаскиваем всего 5 изделий, а вот благоприятно для нас вытащить 2 с дефектом, и так как всего нужно 5, то 3 без дефекта. Далее рассуждаем, сколькими способами можно вытащить 2 изделия с дефектом? Правильно, C_3^2 способами (т. е. 2 изделия с дефектом можно вытащить только среди изделий с дефектами, а таких всего 3). Затем, сколькими способами можно вытащить 3 изделия без дефектов? Аналогично, C_{33}^3 способами (т. е. вытаскиваем 3 изделия из 33-х имеющихся изделий без дефектов). Так как мы одновременно вытаскиваем 2 изделия с дефектами и 3 без таковых, то по правилу умножения из комбинаторики данные способы перемножаются и получается, что $m = C_3^2 C_{33}^3$.

5. Многим известна игра в покер. Для нее берут колоду из 52 карт. Каждый игрок получает по 5 карт. Различают следующие покерные комбинации.

• Флэш-ройяль: самая выигрышная комбинация в покере. Она состоит из пяти карт одной масти, от десятки до туза. Все масти равны в покере.

• Стрит-флэш: пять последовательных карт одной масти.

• Каре: четыре карты одного ранга (например, 4 короля).

• Фул-хаус: 3 равные по значению карты и 2 карты, равные по значению, отличному от значения первых 3-х карт (например, 3 девятки и 2 дамы).

• Флэш: 5 карт одной масти.

• Стрит: 5 карт последовательных значений, но различных мастей.

• Тройка: 3 карты равного значения.

• Две пары: 2 карты одного значения, 2 карты другого значения.

• Пара: 2 карты равного значения.

• Старшая карта: если комбинация в покере не попадает ни в одну из вышеупомянутых категорий, то она оценивается по разряду самой высокой карты.

Вычислить вероятность появления любой покерной комбинации совсем нетрудно: для этого необходимо лишь посчитать, в скольких случаях она возникает, и разделить полученный результат на число способов, которыми можно выбрать 5 карт из 52, т. е., по сути, воспользоваться формулой классической вероятности. Например, стрит-флэш возникает в $13 \cdot 48$ случаях, а 5 карт из 52 можно выбрать C_{52}^{5} способами, поэтому вероятность появления стрит-флэша равна $p = 0{,}000240$. Предлагаем читателю самостоятельно вычислить с точностью до шестого знака вероятности появления оставшихся покерных комбинаций.

1.7. Статистическая и геометрическая вероятности

Было замечено, что при многократном повторении опытов относительная частота появления события в этих опытах стремится к устойчивости. Под относительной частотой появления события понимается отношение $w = M/N$, где N – число опытов; M – число появления события. При увеличении опытов частота появления события будет практически сколь угодно мало отличаться от некоторого постоянного числа, которое и принимается за вероятность события в отдельном опыте. Относительную частоту появления события называют с т а т и с т и ч е с к о й   в е р о я т н о с т ь ю. С возрастанием числа опытов относительная частота стремится к вероятности $P(\Gamma) = 0{,}5$. Относительную частоту при достаточно большом числе опытов можно считать приближенным значению вероятности.

П р и м е р

Отдел технического контроля обнаружил 5 бракованных книг в партии из случайно отобранных 100 книг. Найти относительную частоту появления бракованных книг.

Решение: $w(A) = \dfrac{5}{100} = 0,05.$

Г е о м е т р и ч е с к о й   в е р о я т н о с т ь ю события называется отношение меры области, благоприятствующей появлению события, к мере всей области.

Пусть плоская фигура g составляет часть плоской фигуры G (рис. 3). На фигуру G наудачу брошена точка. Если предположить, что вероятность попадания брошенной точки на фигуру пропорциональна площади этой фигуры и не зависит ни от ее расположения относительно G, ни от формы g, то вероятность попадания точки в фигуру g определяется равенством

$$p = \frac{\text{площадь } g}{\text{площадь } G}.$$

П р и м е р

Два незнакомых мафиози, *А* и *Б*, пытаются убить друг друга. Первый приходит к месту встречи от 12.00 до 13.00, мгновенно и незаметно оставляет мину, которая взорвется через 10 мин, ждет 5 мин и уходит, второй также приходит от 12.00 до 13.00, оставляет свою мину, которая взорвется через 15 мин, ждет 10 мин и уходит. Ответьте на следующие вопросы: 1) Какова вероятность того, что мафиози встретятся?

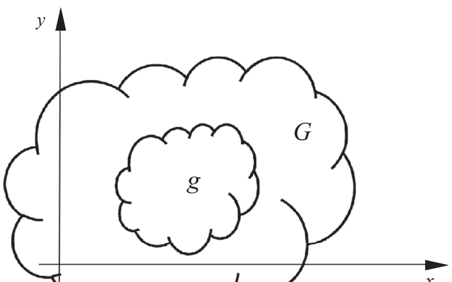

Рис. 3

2) Какова вероятность гибели мафиози *Б*? 3) Какова вероятность гибели мафиози *А*? 4) Какова вероятность гибели обоих мафиози?

Решение. Пусть x – время прихода мафиози *А*, а y – время прихода *Б*. Тогда из условия следует: $y - x < 10$, если $y > x$ (т. е. если *Б* приходит раньше *А*, то он ждет 10 мин, а значит, встретиться они могут только в течение 10 мин, следовательно, разница между приходом *Б* и *А* должна быть менее 10 мин). Аналогично, если $x - y < 5$, то $x > y$. Таким образом, мы формализовали благоприятствующую нам область. Также общее условие состоит в том, что оба мафиози могут прийти с 12.00 до 13.00, т. е. в течение 60 мин, т. е. $0 \le x, y \le 60$. На рис. 4 показаны две эти области. Тогда:

$$p = \frac{S_g}{S_{\text{прямоугольника}}} = \frac{S_{\text{прямоугольника}} - S_{ABC} - S_{DEF}}{S_{\text{прямоугольника}}} =$$

$$= \frac{60 \cdot 60 - 0,5 \cdot 50 \cdot 50 - 0,5 \cdot 55 \cdot 55}{60 \cdot 60} = 0,233.$$

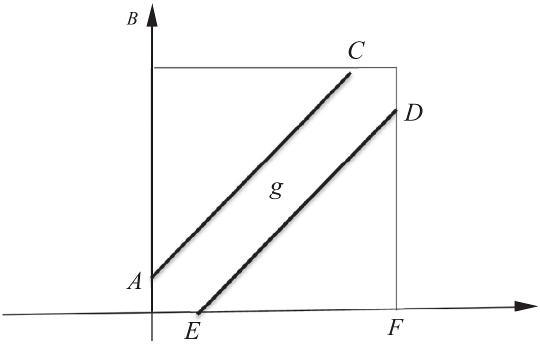

Рис. 4

Пункты 2, 3, 4 предлагаются читателю для самостоятельного разбора.

1.8. Задачи

1.1. Сколько перестановок можно составить из букв слова «линейка»?

1.2. Сколько перестановок можно составить из букв слова «тетрадь»?

1.3. Сколькими способами можно расставить белые фигуры (2 коня, 2 слона, 2 ладьи, ферзь и король) на первой линии шахматной доски?

1.4. В корзине лежат 12 красных и 5 черных чайных роз. Сколькими способами из них можно составить букет, в котором должно быть 3 красных и 2 чайных розы?

1.5. Есть 10 шариков и 4 ящика. В первый ящик сначала положили 2 шарика, затем во второй ящик – 3 шарика, после в третий ящик – 3 и, наконец, в четвертый – 2 шара. Определить, сколькими способами можно было разложить шарики по ящикам.

1.6. Сколько существует двузначных чисел, в которых цифра десятков и цифра единиц различные и нечетные?

1.7. Сколько трехзначных четных чисел можно составить из цифр 0, 1, 2, 3, 4, 5, 6, если цифры могут повторяться?

1.8. Сколько всех четырехзначных чисел можно составить из цифр 1, 5, 6, 7, 8?

1.9. Фрезеровщик Вася произвел за смену 150 деталей, 14 из которых являются бракованными. Начальник смены Федя произвольно выбирает 5 деталей. Сколько существует способов отобрать 2 бракованные детали из 5 выбранных?

1.10. В дежурной части в данный момент имеется 5 офицеров, 20 оперативников и 4 собаки. На вызов срочно требуется 1 офицер, 4 оперативника, 1 собака. Сколькими способами N возможно выбрать сотрудников на срочный вызов?

1.11. Есть 5 сигнальных флажков. Сколькими различными вариантами может быть подан сигнал, если использовать разное количество флажков и в любом порядке? Однако важно понимать, что каждый флаг, поднятый в одном порядке, означает совершенно иное, чем тот же флаг, но поднятый в другом порядке.

1.12. В гардеробе у дамы три кофточки, две юбки и двое туфель. Все вещи по стилю и цвету хорошо сочетаются. Сколько различных вариантов наряда можно составить, комбинируя эти вещи?

1.13. Сколькими способами можно рассадить студентов, нарушающих дисциплину, на первый ряд, состоящий из 4 мест?

1.14. Сколько существует четных двузначных, трехзначных и пятизначных четных чисел?

1.15. Мужчина на шахматную доску случайным образом поставил две ладьи. Какова вероятность, что они не будут бить одна другую?

1.16. При перевозке ящика, в котором содержались 21 стандартная и 10 нестандартных деталей, утеряна одна деталь, причем неизвестно какая. Наудачу извлеченная (после перевозки) из ящика деталь оказалась стандартной. Найти вероятность того, что была утеряна: а) стандартная деталь; б) нестандартная деталь.

1.17. По условиям лотереи «Спортлото 6 из 45» участник лотереи, угадавший 4, 5, 6 видов спорта из отобранных при случайном розыгрыше 6 видов спорта из 45, получает денежный приз. Найти вероятность того, что будут угаданы: а) все 6 цифр; б) 4 цифры.

1.18. В партии 100 изделий, из которых 4 – бракованные. Партия произвольно разделена на две равные части, которые отправлены двум потребителям. Какова вероятность того, что все бракованные изделия достанутся: а) одному потребителю; б) обоим потребителям поровну?

1.19. В магазине было продано 21 из 25 холодильников трех марок, имеющихся в количествах 5, 7 и 13 штук. Полагая, что вероятность быть проданным для холодильника каждой марки одна и та же, найти вероятность того, что остались нераспроданными холодильники: а) одной марки; б) трех разных марок.

1.20. При перевозке ящика, в котором содержались 21 стандартная и 10 нестандартных деталей, утеряна одна деталь, причем неизвестно какая. Наудачу извлеченная (после перевозки) из ящика деталь оказалась стандартной. Найти вероятность того, что была утеряна: а) стандартная деталь; б) нестандартная деталь.

1.21. Куб, все грани которого окрашены, распилен на тысячу кубиков одинакового размера, которые затем тщательно перемешаны. Найти вероятность того, что наудачу извлеченный кубик будет иметь окрашенных граней: а) одну; б) две; в) три.

1.22. Из тщательно перемешанного полного набора 28 костей домино наудачу извлечена кость. Найти вероятность того, что вторую наудачу извлеченную кость можно приставить к первой, если первая кость: а) оказалась дублем; б) не есть дубль.

1.23. В замке на общей оси пять дисков. Каждый диск разделен на шесть секторов, на которых написаны различные буквы. Замок открывается только в том случае, если каждый диск занимает одно определенное положение относительно корпуса замка. Найти вероятность того, что при произвольной установке дисков замок можно будет открыть.

1.24. В ящике 10 одинаковых деталей, помеченных номерами 1, 2, ..., 10. Наудачу извлечены шесть деталей. Найти вероятность того, что среди извлеченных деталей окажутся: а) деталь № 1; б) детали № 1 и № 2.

1.25. Набирая номер телефона, абонент забыл последние три цифры и, помня лишь, что эти цифры различны, набрал их наудачу. Найти вероятность того, что набраны нужные цифры.

1.26. В партии из N деталей имеется n стандартных. Наудачу отобраны m деталей. Найти вероятность того, что среди отобранных деталей ровно k стандартных.

1.27. На складе имеется 15 кинескопов, причем 10 из них изготовлены Львовским заводом. Найти вероятность того, что среди 5-ти взятых наудачу кинескопов окажутся 3 кинескопа Львовского завода.

2. ТЕОРЕМЫ ВЕРОЯТНОСТЕЙ

2.1. Теоремы о произведении и сумме событий

К сожалению, на практике далеко не всегда для подсчета вероятности достаточно формул классической и геометрической вероятностей. Часто приходится прибегать к более сложным конструкциям. Для облегчения подсчета используются несколько теорем вероятностей.

Теорема 1. Если два события A и B несовместны, тогда вероятность суммы двух событий равна сумме вероятностей этих событий.

$$A \cap B = \varnothing = p(A \cap B) = p(A) + p(B).$$

Пример

1) A = {сдать теорию вероятностей на «5»};
B = {сдать теорию вероятностей на «4»};

$$p(A) = 0{,}1;\;\; p(B) = 0{,}3 \Rightarrow P(A + B) = 0{,}1 + 0{,}3 = 0{,}4.$$

2) A = {сдать теорию вероятностей на «5»};
B = {сдать статистику на «5»};

$p(A) = 0{,}6$; $p(B) = 0{,}8 \Rightarrow p(A + B) = 1{,}4$ – нельзя! Так как вероятность не может быть больше 1. Эти события совместны, так как можно сдать и теорию вероятностей, и статистику.

Следствие. Сумма вероятностей событий, образующих полную группу, равна 1, т. е.

$$A_i \cap A_j = \varnothing,\; A_1 + A_2 + \ldots + A_n = \Omega \Rightarrow p(A_1) + \ldots + p(A_n) = 1.$$

Доказательство: $p(A_1 + A_2 + \ldots + A_n) = p(\Omega) = 1$, но так как события независимы, то по теореме 1:

$$p(A_1 + A_2 + \ldots + A_n) = p(A_1) + \ldots + p(A_n).$$

Доказано.

Следствие. Вероятность противоположного события равна 1 – вероятность события A, т. е. $p(\overline{A}) = 1 - p(A)$.

П р и м е р

A = {сдать методы оптимальных решений};

$$p(A) = 0{,}9 \Rightarrow p(\overline{A}) = 0{,}1.$$

Определим понятие у с л о в н о й в е р о я т н о с т и следующим образом: пусть есть события A и B.

Вероятность события B, найденная при условии, что событие A произошло, называется у с л о в н о й в е р о я т н о с т ь ю с о б ы т и я B.

$$P(B \mid A) \text{ или } P_A(B).$$

П р и м е р

Знаем 20 из 25 экзаменационных билетов, экзамен идем сдавать вдвоем.

B = {второй вытянул хороший билет};

A = {первый вытянул хороший билет}.

$$P(A|B) = \frac{19}{24} = 0{,}7917;$$

$$P(A|\overline{B}) = \frac{20}{24} = 0{,}8333.$$

В случае, если билеты возвращаются обратно:

$$P(A|B) = P(A|\overline{B}) = \frac{20}{25} = 0{,}8.$$

Если $P(A)$ не зависит от того, произошло или нет событие B, то события называются н е з а в и с и м ы м и.

$$P(A|B) = P(A|\overline{B}) = P(A).$$

Теорема 2 (у м н о ж е н и е в е р о я т н о с т е й).

$$P(AB) = P(A) \cdot P(B \mid A).$$

Следовательно,

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

Следствие. Если A и B независимы, то вероятность произведения двух событий равна произведению их вероятностей, т. е.:

$$A \text{ и } B \text{ независимы} \Leftrightarrow P(AB) = P(A) \cdot P(B).$$

Доказательство:

$$P(AB) = P(A) \cdot P(B \mid A) = P(A) \cdot P(B);$$

$$P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B).$$

Следовательно, A и B независимы.

2.2. Формула Бернулли

На практике часто приходится сталкиваться с задачами, которые можно представить в виде многократно повторяющихся испытаний при данном комплексе условий, в которых представляет интерес вероятность члена k наступления некоторого события A в n испытаниях. Например, необходимо определить вероятность определенного числа попаданий в мишень при нескольких выстрелах, вероятность определенного числа бракованных изделий в данной партии и т. д.

Итак, пусть вероятность события A равна p в каждом испытании. Проводится n независимых испытаний. Тогда вероятность того, что событие A наступит ровно k раз, обозначается $P_n(k)$ и рассчитывается по *формуле Бернулли*:

$$P_n(k) = C_n^k p^k q^{n-k},$$

где $q = 1 - p$.

При этом важно помнить, что $P_n(n) = p^n$, а $P_n(0) = q^n$.

Пример

В тесте 5 вопросов, на каждый имеется 4 варианта ответа.

1) Какова вероятность, что мы на все 5 вопросов ответим верно?

$$n = 5; \quad p = \frac{1}{4}; \quad q = \frac{3}{4};$$

$$P_5(5) = \left(\frac{1}{4}\right)^5 = \frac{1}{1\,024} = 0,00098.$$

2) Какова вероятность того, что ответим правильно только на 2 вопроса?

$$n = 5;\ \ k = 2;\ \ p = \frac{1}{4};\ \ q = \frac{3}{4};$$

$$P_5(2) = C_5^2\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 = \frac{5!}{2!\,3!}\cdot\frac{1}{16}\cdot\frac{27}{64} = 0,2637.$$

3) Какова вероятность, что мы получим «неудовлетворительно» (т. е. ответим на 2 и менее вопросов)?

$$P_5(k \le 2) = P_5(0) + P_5(1) + P_5(2);$$

$$P_5(0) = q^5 = \left(\frac{3}{4}\right)^5 = 0,237;$$

$$P_5(1) = C_5^1\left(\frac{1}{4}\right)^1\cdot\left(\frac{3}{4}\right)^4 = \frac{5\cdot 81}{4\,256} = 0,3954;$$

$$P_5(k \le 2) = 0,237 + 0,3954 + 0,2637 + 0,8965.$$

Заметим, что при больших n вычисления в формуле трудны! Что делать? Хотелось бы иметь более простые приближенные способы для вычисления $P_n(k)$. Такие формулы называются асимптотическими.

1. Формула Пуассона.

$$P_n(k) \approx P_k(\lambda) = \frac{\lambda^k e^{-\lambda}}{k!},$$

где $\lambda = np$.

Эту формулу применяем только при маленьких p и больших n ($\lambda = np \le 10$).

2. Локальная формула Муавра – Лапласа.

$$P_n(k) = \frac{1}{\sqrt{npq}}\,\varphi(x),$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}$ есть функция Гаусса, $x = \frac{k - np}{\sqrt{npq}}$.

3. Интегральная теорема Муавра – Лапласа.

$$P(k_1,\ k_2) = \Phi(x'') - \Phi(x');$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{-z^2}{2}}\, dz;$$

$$x' = \frac{k_1 - np}{\sqrt{npq}};$$

$$x'' = \frac{k_2 - np}{\sqrt{npq}}.$$

Замечание. Для нахождения значений функций Пуассона, Гаусса и Лапласа используются таблицы, помещенные в приложении данного пособия.

2.3. Полная вероятность

Пусть событие A может произойти в результате одного из событий $H_1, H_2, \ldots, H_n$, образующих полную группу событий.

Известны: вероятности событий, $\sum_i^n p(H_i) = 1$, условные вероятности $p(A \mid H_1), \ldots, p(A \mid H_n)$,тогда

$$P(A) = \sum_i^n P(H_i)\, P(A|H_i) =$$

$$= P(H_1)\, P(A|H_1) + P(H_2)\, P(A|H_2) + \ldots + P(H_n)\, P(A|H_n).$$

Данная формула называется формулой полной вероятности.

Примеры

1. Имеются две урны. В первой 5 белых и 10 черных шаров, во второй 10 белых и 5 черных шаров. Из первой урны наугад вытаскивают 1 шар и перекладывают во вторую урну. Затем из второй урны наудачу вытаскивают 1 шар. Какова вероятность, что вытащенный шар окажется белым?

A – {из второй урны достали белый шар}.
H_1 – {из первой урны во вторую переложили белый шар}.
H_2 – {из первой урны во вторую переложили черный шар}.

Решение:

$$P(H_1) = \frac{5}{15} = \frac{1}{3};$$

$$P(H_2) = \frac{10}{15} = \frac{2}{3};$$

$$P(A|H_1) = \frac{11}{16};$$

$$P(A|H_2) = \frac{10}{16};$$

$$P(A) = \frac{1}{3} \cdot \frac{11}{16} + \frac{2}{3} \cdot \frac{10}{16} = \frac{155}{240} = 0,6458 < \frac{2}{3}.$$

2. На трассе имеется заправка. Мимо проезжают легковые автомобили, грузовые автомобили и автобусы в соотношении 5 : 4 : 1. Известны p (легкового) = 0,1, p (грузового) = 0,05, p (автобуса) = 0,01. Едет ТС, какова вероятность того, что оно зайдет на заправку?

Решение. Что едет? H_1 – легковой, H_2 – грузовой, H_3 – автобус.

$$P(H_1) = \frac{5}{10} = 0,5 \text{ (так как } 5:4:1);$$

$$P(H_2) = \frac{4}{10} = 0,4;$$

$$P(H_3) = \frac{1}{10} = 0,1;$$

$$P(A|H_1) = 0,1;$$

$$P(A|H_2) = 0,05;$$

$$P(A|H_3) = 0,01;$$

$$P(A) = 0,5 \cdot 0,1 + 0,4 \cdot 0,05 + 0,1 \cdot 0,01 = 0,071.$$

2.4. Формула Байеса

Пусть имеем те же условия, что и в формуле полной вероятности. Допустим, что произведено испытание, в результате которого появилось событие A. Поставим своей задачей определить, как изменились вероятности гипотез в связи с тем, что событие A уже наступило. Переоценим вероятность гипотезы H_j после того, как произошло событие A. Сделать это можно при помощи ф о р м у - л ы Б а й е с а:

$$P(H_j|A) = \frac{P(H_j)P(A|H_j)}{P(A)} = \frac{P(H_j)\,P(A|H_j)}{\sum_i^n P(H_i)\,P(A|H_i)}.$$

Значение формулы Байеса состоит в том, чтобы по мере получения новой информации можно было проверить и скорректировать выдвинутые до испытания гипотезы. Такой подход называется байесовским и дает возможность корректировать управленческие решения.

П р и м е р ы

1. Имеются две урны. В первой 5 белых и 10 черных шаров, во второй 10 белых и 5 черных шаров. Из первой урны наугад вытаскивают 1 шар и перекладывают во вторую урну. Затем из второй урны наудачу вытаскивают 1 шар. Пусть из второй урны достаем белый шар, тогда вероятность того, что из первой урны мы переложили во вторую именно белый шар:

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(A)} = \frac{\frac{1}{3}\cdot\frac{11}{16}}{\frac{155}{240}} = 0,3548 > \frac{1}{3}.$$

2. Экономист-аналитик условно подразделяет экономическую ситуацию в стране на «хорошую», «посредственную» и «плохую» и оценивает их вероятности для данного момента времени в 0,15; 0,70 и 0,15 соответственно. Некоторый индекс экономического состояния возрастает с вероятностью 0,60, когда ситуация хорошая; с вероятностью 0,30, когда ситуация посредственная, и с вероятностью 0,10, когда ситуация плохая. Пусть в настоящий момент индекс экономи-

ческого состояния возрос. Чему равна вероятность того, что экономика страны на подъеме?

A – {индекс экономического состояния возрастает};

H_1 – {экономическая ситуация в стране посредственная};

H_2 – {экономическая ситуация в стране хорошая};

H_3 – {экономическая ситуация в стране плохая}.

$$P(H_1) = 0{,}15;\ P(H_2) = 0{,}70;\ P(H_3) = 0{,}15.$$

$$P(H_1) = 0{,}15;\ P(A \mid H_1) = 0{,}6;\ P(A \mid H_2) = 0{,}3; P(A \mid H_3) = 0{,}1.$$

По формуле Байеса

$$P(H_1 \mid A) = \frac{P(H_1)P(A \mid H_1)}{P(H_1)P(A \mid H_1) + P(H_2)P(A \mid H_2) + P(H_2)P(A \mid H_2)} =$$

$$= \frac{0{,}15 \cdot 0{,}6}{0{,}15 \cdot 0{,}6 + 0{,}70 \cdot 0{,}3 + 0{,}15 \cdot 0{,}1} = 0{,}286.$$

2.5. Задачи

2.1. Вероятность того, что студент сдаст первый экзамен, равна 0,9; второй – 0,9; третий – 0,8. Найти вероятность того, что студентом будут сданы: а) только второй экзамен; б) только один экзамен; в) три экзамена; г) по крайней мере два экзамена; д) хотя бы один экзамен.

2.2. Причиной разрыва электрической цепи служит выход из строя элемента K_1 или одновременный выход из строя двух элементов – K_2 и K_3. Элементы могут выйти из строя независимо друг от друга с вероятностями, равными соответственно 0,1; 0,2; 0,3. Какова вероятность разрыва электрической цепи?

2.3. При включении зажигания двигатель начнет работать с вероятностью 0,6. Найти вероятность того, что: а) двигатель начнет работать при третьем включении зажигания; б) для запуска двигателя придется включать зажигание не более трех раз.

2.4. Экспедиция издательства отправила газеты в три почтовых отделения. Вероятность своевременной доставки газет в первое отделение равна 0,95, во второе отделение – 0,9 и в третье – 0,8. Найти вероятность следующих событий: а) только одно отделение получит газеты вовремя; б) хотя бы одно отделение получит газеты с опозданием.

2.5. Мастер обслуживает 4 станка, работающих независимо друг от друга. Вероятность того, что первый станок в течение смены потребует внимания рабочего, равна 0,3, второй – 0,6, третий – 0,4 и четвертый – 0,25. Найти вероятность того, что в течение смены хотя бы один станок не потребует внимания мастера.

2.6. В коробке смешаны электролампы одинакового размера и формы: по 100 Вт – 7 штук, по 75 Вт – 13 штук. Вынуты наудачу 3 лампы. Какова вероятность того, что: а) они одинаковой мощности; б) хотя бы две из них по 100 Вт?

2.7. В коробке 10 красных, 3 синих и 7 желтых карандашей. Наудачу вынимают 3 карандаша. Какова вероятность того, что они все: а) разных цветов; б) одного цвета?

2.8. Брак в продукции завода вследствие дефекта A составляет 4 %, а вследствие дефекта B – 3,5 %. Годная продукция завода составляет 95 %. Найти вероятность того, что: а) среди продукции, не обладающей дефектом A, встретится дефект B; б) среди забракованной по признаку A продукции встретится дефект B.

2.9. Среди 20 поступающих в ремонт часов 8 нуждаются в общей чистке механизма. Какова вероятность того, что среди взятых одновременно наудачу трех часов по крайней мере двое нуждаются в общей чистке механизма?

2.10. Контролер ОТК, проверив качество сшитых 20 пальто, установил,что 16 из них первого сорта, а остальные – второго. Найти вероятность того, что среди взятых наудачу из этой партии трех пальто одно будет второго сорта.

2.11. Среди студентов, собравшихся на семинар по теории вероятностей, выбирают наудачу одного. Пусть событие A заключается в том, что он – юноша. Событие B в том, что он не курит, а событие C в том, что он живет в общежитии. 1) Описать событие ABC; 2) При каком условии будет иметь место тождество $ABC = A$?

3) Когда будет справедливо соотношение $C \subseteq B$? 4) Когда будет верно равенство $A = B$? Будет ли оно иметь место, если все юноши курят?

2.12. Пусть A, B, C – три произвольно выбранных события. Найти выражения для событий, состоящих в том, что из A, B, C: а) произошло только A; б) произошли A и B, но C не произошло; в) все три события произошли; г) произошло хотя бы одно из этих событий; д) произошло хотя бы два события; е) произошло одно и только одно из этих событий; ж) произошло два и только два события; з) ни одно из событий не произошло; и) произошло не более двух событий.

2.13. Вероятность выхода изделия из строя при эксплуатации сроком до одного года равна 0,13, а при эксплуатации сроком до 3-х лет – 0,36. Найти вероятность выхода изделия из строя при эксплуатации сроком от 1 года до 3-х лет.

2.14. На стеллаже библиотеки в случайном порядке расставлено 15 учебников, причем пять из них в переплете. Библиотекарь берет наудачу три учебника. Найти вероятность того, что: а) хотя бы один из взятых учебников окажется в переплете (событие A); б) хотя бы два из взятых учебников окажутся в переплете.

2.15. Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,7, а для второго – 0,8. Найти вероятность того, что при одном залпе в мишень попадает только один из стрелков.

2.16. Вероятность одного попадания в цель при одном залпе из двух орудий равна 0,38. Найти вероятность поражения цели при одном выстреле первым из орудий, если известно, что для второго орудия эта вероятность равна 0,8.

2.17. Из партии изделий товаровед отбирает изделия высшего сорта. Вероятность того, что наудачу взятое изделие окажется высшего сорта, равна 0,8. Найти вероятность того, что из трех проверенных изделий только два изделия высшего сорта.

2.18. Вероятность попадания в мишень стрелком при одном выстреле равна 0,8. Сколько выстрелов должен произвести стрелок, чтобы с вероятностью, меньшей 0,4, можно было ожидать, что не будет ни одного промаха?

2.19. Для сигнализации об аварии установлены два независимо работающих сигнализатора. Вероятность того, что при аварии сигнализатор сработает, равна 0,95 для первого сигнализатора и 0,9 – для второго. Найти вероятность того, что при аварии сработает только один сигнализатор.

2.20. Студент пришел на экзамен, зная лишь 20 из 25 вопросов программы. Экзаменатор задал студенту 3 вопроса. Используя понятие условной вероятности, найти вероятность того, что студент знает все эти вопросы.

2.21. Разыскивая специальную книгу, студент решил обойти три библиотеки. Для каждой библиотеки одинаково вероятно, есть в ее фондах книга или нет. И если книга есть, то одинаково вероятно, занята она другим читателем или нет. Что более вероятно – достанет студент книгу или нет, если известно, что библиотеки комплектуются независимо одна от другой?

2.22. Среди 100 лотерейных билетов есть 5 выигрышных. Найти вероятность того, что 2 наудачу выбранных билета окажутся выигрышными.

2.23. В ящике 10 деталей, среди которых 6 окрашенных. Сборщик наудачу извлекает 4 детали. Найти вероятность того, что все извлеченные детали окажутся окрашенными.

2.24. В урне имеется пять шаров с номерами от 1 до 5. Наудачу по одному извлекают три шара без возвращения. Найти вероятности следующих событий: а) последовательно появятся шары с номерами 1, 4, 5; б) извлеченные шары будут иметь номера 1, 4, 5 независимо от того, в какой последовательности они появились.

2.25. В большой рекламной фирме 21 % работников получают высокую заработную плату. Известно также, что 40 % работников фирмы – женщины, а 6,4 % работников – женщины, получающие высокую заработную плату. Можем ли мы утверждать, что на фирме существует дискриминация женщин в оплате труда?

2.26. По данным переписи населения (1891 г.) Англии и Уэльса установлено: темноглазые отцы и темноглазые сыновья (AB) составили 5 % обследованных лиц, темноглазые отцы и светлоглазые сыновья ($A\overline{B}$) – 7,9 %, светлоглазые отцы и темноглазые сыновья ($\overline{A}B$) – 8,996 %, светлоглазые отцы и светлоглазые сыновья ($\overline{A}\overline{B}$) – 78,2 %. Найти связь между цветом глаз отца и сына.

2.27. Вероятность того, что потребитель увидит рекламу определенного продукта по телевидению, равна 0,04. Вероятность того, что потребитель увидит рекламу того же продукта на рекламном стенде, равна 0,06. Предполагается, что оба события – независимы. Чему равна вероятность того, что потребитель увидит обе рекламы?

2.28. Консультационная фирма претендует на два заказа от двух крупных корпораций. Эксперты фирмы считают, что вероятность получения консультационной работы в корпорации A (событие A) равна 0,45. По предположению экспертов, если фирма получит заказ у корпорации A, то вероятность того, что и корпорация B обратится к ним, равна 0,9. Какова вероятность получения консультационной фирмой обоих заказов?

2.29. Вероятность попадания в мишень при каждом выстреле для первого стрелка равна 0,7, а для второго – 0,8. Оба они делают по одному выстрелу по мишени, а затем каждый из стрелков стреляет еще раз, если при первом сделанном им выстреле он промахнулся. Найти вероятность того, что в мишени ровно 2 пробоины.

2.30. Пакеты акций, имеющихся на рынке ценных бумаг, могут дать доход владельцу с вероятностью 0,5 (для каждого пакета). Сколько пакетов акций различных фирм нужно приобрести, чтобы с вероятностью, не меньшей 0,96875, можно было ожидать доход хотя бы по одному пакету акций?

2.31. В среднем 20 % пакетов акций на аукционах продаются по первоначально заявленной цене. Найти вероятность того, что из 9 пакетов акций в результате торгов по первоначально заявленной цене: 1) не будут проданы 5 пакетов; 2) будет продано: а) менее 2 пакетов; б) не более 2; в) хотя бы 2 пакета.

2.32. Вероятность малому предприятию быть банкротом за время t равна 0,2. Найти вероятность того, что из шести малых предприятий за время t сохранятся: а) два; б) более двух.

2.33. В среднем пятая часть поступающих в продажу автомобилей некомплектны. Найти вероятность того, что среди десяти автомобилей имеют некомплектность: а) три автомобиля; б) менее трех.

2.34. Производится залп из шести орудий по некоторому объекту. Вероятность попадания в объект из каждого орудия равна 0,6. Найти вероятность ликвидации объекта, если для этого необходимо не менее четырех попаданий.

2.35. В среднем по 15 % договоров страховая компания выплачивает страховую сумму. Найти вероятность того, что из десяти договоров с наступлением страхового случая будет связано с выплатой страховой суммы: а) три договора; б) менее двух договоров.

2.36. Два равносильных шахматиста играют в шахматы. Что вероятнее: выиграть две партии из четырех или три партии из шести (ничьи во внимание не принимаются)?

2.37. Известно, что в среднем 95 % выпускаемой продукции удовлетворяют стандарту. Упрощенная схема контроля признает пригодной продукцию с вероятностью 0,98, если она стандартна, и с вероятностью 0,06, если она нестандартна. Определить вероятность того, что: 1) взятое наудачу изделие пройдет упрощенный контроль; 2) изделие стандартное, если оно: а) прошло упрощенный контроль; б) дважды прошло упрощенный контроль.

2.38. Два стрелка независимо друг от друга стреляют по мишени, делая каждый по одному выстрелу. Вероятность попадания в мишень для первого стрелка равна 0,8, для второго – 0,4. После стрельбы в мишени обнаружена одна пробоина. Какова вероятность того, что она принадлежит: а) первому стрелку; б) второму стрелку?

2.39. Радист трижды вызывает корреспондента. Вероятность того, что будет принят первый вызов, равна 0,2, второй – 0,3, третий – 0,4. События, состоящие в том, что данный вызов будет услышан, независимы. Найти вероятность того, что корреспондент услышит вызов радиста.

2.40. Страховая компания разделяет застрахованных по классам риска: I класс – малый риск, II класс – средний, III класс – большой риск. Среди этих клиентов 50 % – первого класса риска, 30 % – второго и 20 % – третьего. Вероятность необходимости выплачивать страховое вознаграждение для первого класса риска равна 0,01, второго – 0,03, третьего – 0,08. Какова вероятность того, что: а) застрахованный получит денежное вознаграждение за период страхования; б) получивший денежное вознаграждение застрахованный относится к группе малого риска?

2.41. В данный район изделия поставляются тремя фирмами в соотношении 5 : 8 : 7. Среди продукции первой фирмы стандартные изделия составляют 90 %, второй – 85 %, третьей – 75 %. Найти вероятность того, что: а) приобретенное изделие окажется нестандартным; б) приобретенное изделие оказалось стандартным. Какова вероятность того, что оно изготовлено третьей фирмой?

2.42. Вся продукция цеха проверяется двумя контролерами, причем первый контролер проверяет 55 % изделий, а второй – остальные. Вероятность того, что первый контролер пропустит нестандартное изделие, равна 0,01, второй – 0,02. Взятое наудачу изделие, маркированное как стандартное, оказалось нестандартным. Найти вероятность того, что это изделие проверялось вторым контролером.

2.43. Вероятность изготовления изделия с браком на данном предприятии равна 0,04. Перед выпуском изделие подвергается упрощенной проверке, которая в случае бездефектного изделия пропускает его с вероятностью 0,96, а в случае изделия с дефектом – с вероятностью 0,05. Определить: а) какая часть изготовленных изделий выходит с предприятия; б) какова вероятность того, что изделие, выдержавшее упрощенную проверку, бракованное?

2.44. Завод выпускает определенного типа изделия; каждое изделие имеет дефект с вероятностью 0,7. После изготовления изделие осматривается последовательно тремя контролерами, каждый из которых обнаруживает дефект с вероятностями 0,8; 0,85; 0,9 соответственно. В случае обнаружения дефекта изделие бракуется и в дальнейшем на осмотр следующему контролеру не передается. Определить вероятность того, что изделие: 1) будет забраковано; 2) будет забраковано: а) вторым контролером; б) всеми контролерами.

2.45. Из полной колоды карт (52 карты) выбирают шесть карт, одну из них смотрят; она оказывается тузом, после чего ее смешивают с остальными выбранными картами. Найти вероятность того, что при втором извлечении карты из этих шести мы снова получим туз.

2.46. В одной урне 5 белых и 6 черных шаров, а в другой – 4 белых и 8 черных шаров. Из первой урны случайным образом вынимают 3 шара и опускают во вторую урну. После этого из второй урны также случайно вынимают 4 шара. Найти вероятность того, что все шары, вынутые из второй урны, белые.

2.47. Завод отправил на базу 10 тыс. стандартных изделий. Среднее число изделий, повреждаемых при транспортировке, составляет 0,02 %. Найти вероятность того, что из 10 тыс. изделий: 1) будет повреждено: а) 3; б) по крайней мере 3; 2) не будет повреждено: а) 9 997; б) хотя бы 9 997.

2.48. По результатам проверок налоговыми инспекциями установлено, что в среднем каждое второе малое предприятие региона имеет нарушение финансовой дисциплины. Найти вероятность того, что из 1 000 зарегистрированных в регионе малых предприятий имеют нарушения финансовой дисциплины: а) 480 предприятий; б) наивероятнейшее число предприятий; в) не менее 480; г) от 480 до 520.

2.49. В страховой компании 10 тыс. клиентов. Страховой взнос каждого клиента составляет 500 руб. При наступлении страхового случая, вероятность которого по имеющимся данным и оценкам экспертов можно считать равной $p = 0{,}005$, страховая компания обязана выплатить клиенту страховую сумму размером 50 тыс. руб. На какую прибыль может рассчитывать страховая компания с надежностью 0,95?

2.50. В банк отправлено 4 тыс. пакетов денежных знаков. Вероятность того, что пакет содержит недостаточное или избыточное число денежных знаков, равна 0,0001. Найти вероятность того, что при проверке будет обнаружено: а) три ошибочно укомплектованных пакета; б) не более трех пакетов.

2.51. Строительная фирма, занимающаяся установкой летних коттеджей, раскладывает рекламные листки по почтовым ящикам. Прежний опыт работы компании показывает, что примерно в одном случае из двух тысяч следует заказ. Найти вероятность того, что при размещении 100 тыс. листков число заказов будет: а) равно 48; б) находиться в границах от 45 до 55.

2.52. Аудиторную работу по теории вероятностей с первого раза успешно выполняют 50 % студентов. Найти вероятность того, что из 400 студентов работу успешно выполнят: а) 180 студентов, б) не менее 180 студентов.

2.53. При обследовании уставных фондов банков установлено, что пятая часть банков имеют уставный фонд свыше 100 млн руб. Найти вероятность того, что среди 1 800 банков имеют уставный фонд свыше 100 млн руб.: а) не менее 300; б) от 300 до 400 включительно.

2.54. Два баскетболиста делают по 3 броска мячом в корзину. Вероятности попадания мяча в корзину при каждом броске равны соответственно 0,6 и 0,7. Найти вероятность того, что: а) у обоих будет одинаковое количество попаданий; б) у первого баскетболиста будет больше попаданий, чем у второго.

2.55. Брокер может приобрести акции одной из трех компаний: A, B, C. Риск прогореть при покупке акций компании A составляет 50 %, B – 40 %, C – 20 %. Брокер решает вложить все деньги в акции случайно выбранной компании. Какова вероятность того, что брокер прогорит?

2.56. Имеются три одинаковые с виду урны. В первой 6 белых шаров и 4 черных; во второй – 8 белых и 7 черных; в третьей – только белые шары. Некто подходит наугад к одной из урн и вынимает из нее один шар. Найти вероятность того, что этот шар белый.

2.57. Торты «Прага», «Медовик» и «Сливочный» продаются в соотношении 5 : 4 : 1. Вероятность того, что торт «Прага» будет продан в течение дня, равна 0,6; «Медовик» – 0,35; «Сливочный» – 0,40. Найти вероятность того, что все торты «Прага» будут проданы.

2.58. Прибор состоит из двух узлов. Работа каждого узла необходима для работы прибора в целом. Надежность (вероятность безотказной работы в течение времени t) первого узла равна 0,9, второго – 0,8. Прибор испытывался в течение времени t, в результате чего обнаружено, что он вышел из строя (отказал). Найти вероятность того, что отказал только первый узел, а второй исправен.

2.59. Исследуется динамика курсов валют A и B по отношению к валюте C. Статистика торгов на валютной бирже свидетельствует, что при возрастании курса валюты B курс валюты A растет в 80 % случаев, при снижении курса валюты B курс валюты A растет в 25 % случаев, при неизменности курса валюты B курс валюты A растет в 50 % случаев. Предполагая, что варианты изменения курса валюты B имеют одинаковую вероятность, определить вероятности соответствующих изменений при условии, что на последних торгах курс валюты A вырос.

2.60. Инвестор решил вложить поровну средств в три предприятия при условии возврата ему каждым предприятием через определенный срок 150 % от вложенной суммы. Вероятность банкротства каждого из предприятий 0,2. Найти вероятность того, что по истечении срока кредитования инвестор получит обратно, по крайней мере, вложенную сумму.

2.61. В банке, осуществляющем кредитование населения, 1 тыс. клиентов. Каждому из клиентов выдается кредит 500 тыс. ден. ед. при условии возврата 110 % от этой суммы. Вероятность невозврата кредита каждым из клиентов в среднем составляет $p = 0{,}01$. Какая прибыль гарантирована банку с вероятностью: а) 0,8; б) 0,995?

3. ОДНОМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

3.1. Понятие случайной величины

Случайная величина – величина, которая в результате испытания примет значение, наперед неизвестное и зависящее от причин, все из которых заранее не могут быть учтены.

X, Y, Z – случайные величины.

Примеры

1. X – оценка на экзамене 2, 3, 4, 5.
2. X – число очков при бросании кубика 1, 2, 3, 4, 5, 6.
3. X – число мальчиков из 100 детей, $0 \le x \le \infty$.
4. X – $t°$ (–273; +∞).
5. X – курс доллара (0; +∞).
6. X – зарплата (0; +∞).

Дискретная случайная величина – это случайная величина, которая принимает набор отдельных изолированных значений (конечное (счетное) или бесконечное число).

Непрерывная случайная величина – это случайная величина, набор значений которой целиком заполняет некоторый интервал.

3.2. Определение и примеры дискретной случайной величины

Пусть в результате испытания дискретная случайная величина X может принять одно из следующих значений: x_1, x_2, x_3, ..., x_n;

p_i – вероятность того, что случайная величина примет значение x_i:

$$p_i = p(x = x_i),\ \sum p_i = 1.$$

Перечисление возможных значений случайной величины соответствующими вероятностями называется **закон распределения дискретной случайной величины**. Оформляется обычно в виде таблицы.

X	x_1	x_2	x_3	…	x_n
p	p_1	p_2	p_3	…	p_n

Приведем несколько основных дискретных случайных величин.

1. Закон распределения случайной величины – оценка на экзамене.

X	2	3	4	5
p	0,2	0,4	0,3	0,1

2. Распределение Бернулли.

X	0	1
p	q	p

где $q = 1 - p$.

Случайная величина имеет распределение Бернулли, если она принимает всего два значения: 0 и 1 с вероятностями p и $q = 1 - p$. Таким образом, $P(x = 1) = p$; $P(x = 0) = q$.

Принято говорить, что событие $\{X = 1\}$ соответствует «успеху», а событие $\{X = 0\}$ – «неудаче».

3. Распределение случайной величины, вероятности каждого значения которой одинаковы.

X – число очков при бросании кости.

X	1	2	3	4	5	6
p	1/6	1/6	1/6	1/6	1/6	1/6

Для n-величины $P(x=k)=\dfrac{1}{n},\ k=\overline{1,n}.$

X	1	2	3	4	...	n
p	$1/n$	$1/n$	$1/n$	$1/n$	...	$1/n$

4. Геометрическое распределение.

Пусть вероятность попадания в одном выстреле равна 0,7 и она одинакова. Стреляют до первого попадания в цель. Составить закон распределения числа выстрелов.

X	1	2	3	4	...	k
p	0,7	0,21	0,063	0,00189	...	$0{,}3^{k-1}\cdot 0{,}7$

$$P(X=1)=0{,}7;$$

$$P(X=2)=0{,}3\cdot 0{,}7=0{,}21;$$

$$P(X=3)=0{,}3\cdot 0{,}3\cdot 0{,}7=0{,}063;$$

$$P(X=4)=0{,}3^3\cdot 0{,}7=0{,}00189;$$

$$P(X=k)=0{,}3^{k-1}\cdot 0{,}7;$$

$$\sum P_k=\frac{0{,}7}{1-0{,}3}=1,$$

что и требовалось доказать (геометрическая прогрессия).

5. Пусть имеется N объектов, из них M обладает нужным свойством ($M<N$). Извлекаем n объектов. Тогда говорят, что k – число объектов из n объектов, обладающих нужным свойством, имеет гипергеометрическое распределение со следующим законом распределения:

$$P(x=k)=\frac{C_M^k\ C_{N-M}^{n-k}}{C_N^n}.$$

6. Пусть вероятность попадания в мишень первого стрелка – 0,7, второго – 0,8, третьего – 0,9. Эти вероятности независимы. Стреляют 1 раз. Составить закон распределения числа попаданий в цель.

X	0	1	2	3
p	0,06	0,092	0,398	0,504

$$P(X=0)=0{,}3\cdot 0{,}2\cdot 0{,}1=0{,}06;$$

$$P(X=1)=0{,}7\cdot 0{,}2\cdot 0{,}1+0{,}3\cdot 0{,}8\cdot 0{,}1+0{,}3\cdot 0{,}2\cdot 0{,}9=0{,}21;$$

$$P(X=3)=0{,}3\cdot 0{,}3\cdot 0{,}7=0{,}063;$$

$$P(X=4)=0{,}3^3\cdot 0{,}7=0{,}00189.$$

7. Биноминальное распределение.

Пусть вероятность попадания в 1 выстреле 0,7. Стреляют 4 раза. Составить закон распределения числа попадания в цель.

X	0	1	2	3	4
p	0,0081	0,0756	0,2646	0,4116	0,2401

$$P(X=0)=0{,}3^4=0{,}0081;$$

$$P(X=4)=0{,}7^4=0{,}2401;$$

$$P(X=1)=C_4^1\cdot 0{,}7\cdot 0{,}3^3=0{,}0756;$$

$$P(X=2)=C_4^2\cdot 0{,}7^2\cdot 0{,}3^2=0{,}2646;$$

$$P(X=3)=C_4^3\cdot 0{,}7^3\cdot 0{,}3^1=0{,}4116.$$

Формула Бернулли: $P_n(k)=C_n^k p^k q^{n-k}$, где $q=1-p$.

Пусть проводится n независимых испытаний. Вероятность успеха появления событий A в каждом испытании постоянна и равна p. Тогда говорят, что случайная величина X – число появления события A в n испытаниях имеет биноминальное распределение со следующим законом распределения:

$$P_n(k)=C_n^k p^k q^{n-k},$$

где $k=\overline{0,n}$.

$$X\sim B_i(p);$$

$$X\sim B_i(n,p);$$

$$\sum P_k = \sum C_n^k p^k q^{n-k} = 1;$$

$$q^n + C_n^1 pq^{n-1} + C_n^2 p^2 q^{n-2} + \ldots + p^n = (p + q)^n = 1$$

(формула бинома Ньютона).

8. Распределение Пуассона.

$$X \sim \text{Poisson}\ (\lambda).$$

$$p(x = \lambda) = \frac{\lambda^k e^{-\lambda}}{k!},\ k = \overline{0, +\infty},\ \lambda > 0,\ \sum p_k = 1.$$

3.3. Арифметические операции двух случайных величин

Имеем две дискретные случайные величины X и Y:

X	X_1	X_2	X_3	…	X_{n_x}
p	p_{x_1}	p_{x_2}	p_{x_3}	…	$p_{x_{n_x}}$

Y	Y_1	Y_2	Y_3	…	Y_{n_y}
p	p_{y_1}	p_{y_2}	p_{y_3}	…	$p_{y_{n_y}}$

Тогда случайная величина $X + Y$ – с у м м а двух случайных величин X и Y имеет закон распределения:

$X + Y$	$x_1 + y_1$	$x_1 + y_2$	...	$x_1 + y_{n_y}$	$x_2 + y_1$	...	$x_{n_x} + y_{n_y}$
p	p_{11}	p_{12}	...	$p_{1_{n_y}}$	p_{21}	...	$p_{n_x} p_{n_y}$

где $p_{ij} = p((X = x_i)\,(Y = y_j))$, которая в случае независимости случайных величин равна $p_{x_i} p_{y_j}$.

Аналогично случайная величина $X \cdot Y$ – произведение двух случайных величин X и Y имеет закон распределения:

XY	x_1y_1	x_1y_2	...	$x_1y_{n_y}$	x_2y_1	...	$x_{n_x}y_{n_y}$
p	p_{11}	p_{12}	...	$p_{1_{n_y}}$	p_{21}	...	$p_{n_x}p_{n_y}$

где $p_{ij} = p((X = x_i)(Y = y_j))$, которая в случае независимости случайных величин равна $p_{x_i}p_{y_j}$.

3.4. Числовые характеристики дискретной случайной величины

Если мы имеем просто список оценок за экзамен или зачет, то существует их распределение (так как это случайная величина). Если мы хотим узнать, какая оценка чаще бывает за экзамен, то проще это сделать, узнав, какая оценка в среднем чаще встречается. Таким средним значением для случайной величины является математическое ожидание.

$$EX \text{ или } E(X)$$

или

$$M_x \text{ или } M(x).$$

$$EX = \sum x_i p_i.$$

Пример

Пусть X – случайная величина распределения оценок на экзамене.

X	2	3	4	5
p	0,2	0,4	0,3	0,1

$E_x = 2 \cdot 0,2 + 3 \cdot 0,4 + 4 \cdot 0,3 + 5 \cdot 0,1 = 3,3$ – самая ожидаемая оценка на экзамене.

Свойства математического ожидания:

1. EX – величина детерминированная.
2. $EC = c$ – математическое ожидание числа есть число:

$$E(EX) = EX, \text{ так как } EX \text{ – число.}$$

3. Число можно выносить за знак математического ожидания:

$$E(cX) = cE(X).$$

4. $E(X + Y) = E_x + E_y$, $E(X - Y) = E_x - E_y$.
5. Если x и y независимы, то $E_{XY} = E_X \cdot E_X$. В общем случае это неверно.

Дисперсия случайной величины

$Var X = E(X - E_X)^2$, $X - E_X$ – отклонения от математического ожидания.

$E(X - E_X) = E_X - E(E_X) = E_X - E_X = 0$ для любой случайной величины.

Приведем способы нахождения дисперсии.

1-й способ:

$$var X = E(X - E_X)^2 = E(X^2 - 2XE_X + (E_X)^2) =$$
$$= E_{X^2} - E(2XE_X) + E(E_X)^2 = E_{X^2} - 2E_XE_X + (E_X)^2 = E_{X^2} - (E_X)^2.$$

2-й способ:

$$varX = \sum (x_i - Ex_i)^2 p_i = \sum x_i^2 p_i - (E_X)^2.$$

П р и м е р

Дана случайная величина X, заданная следующим законом:

X	2	3	4	5
p	0,2	0,4	0,3	0,1

Тогда $var X = 2^2 \cdot 0{,}2 + 3^2 \cdot 0{,}4 + 4^2 \cdot 0{,}3 + 5^2 \cdot 0{,}1 - (3{,}3)^2 = 0{,}81.$

Свойства дисперсии:

1. $Var X$ – величина детерминированная.
2. $Var C = 0$.

3. $Var(X + C) = varX$.

4. $Var(CX) = C^2varX$.

5. Если X, Y – независимые, то $var(X + Y) = varX + varY$,

$$var(X - Y) = varX + (-1)^2varY = varX + varY.$$

Среднее квадратическое отклонение

Дисперсия имеет размерность квадрата случайной величины, что не всегда удобно, поэтому принято использовать квадратный корень из дисперсии, который назвали средним квадратическим отклонением:

$$\sigma_x = \sqrt{varX}.$$

Примеры

1. $\sigma_x = \sqrt{0,81} = 0,9$.

2. $Z = 8X - 5Y + 7$. Найти $varX$, если X и Y независимые, и $varX = 1,5$ и $varY = 1$.

Решение: $varZ = 64varX + 25varY = 64 \cdot 1,5 + 25 \cdot 1 = 12,1$.

3. Пусть ежедневные расходы на обслуживание и рекламу автомобилей в автосалоне составляют в среднем 120 тыс. ден. ед., а число продаж X автомашин в течение дня подчиняется закону распределения:

X	0	1	2	3	4	5	6	7	8	9
p	0,25	0,2	0,1	0,1	0,1	0,1	0,05	0,05	0,025	0,025

Найти математическое ожидание ежедневной прибыли при цене машины 150 тыс. ден. ед.

Решение. Ежедневная прибыль подсчитывается по формуле

$$П = (150X - 120).$$

Искомая характеристика $E(П)$ находится с использованием указанных ранее свойств математического ожидания (в тысячах денежных единиц):

$$E(П) = E(150X - 120) = 150E(X) - 120 =$$
$$= 150 \cdot 2,675 - 120 = 281,25.$$

Теперь приведем интерпретацию математического ожидания и дисперсии в ф и н а н с о в о м   а н а л и з е.

Известно распределение доходности X некоторого актива (например, акции), т. е. известны значения доходности x_i и соответствующие их вероятности p_i за рассматриваемый промежуток времени. Тогда математическое ожидание отражает среднюю (прогнозную) доходность актива, а дисперсия или среднее квадратическое отклонение – меру отклонения, колебания доходности от ожидаемого среднего значения, т. е. риск данного актива.

3.5. Числовые характеристики некоторых дискретных случайных величин

1. Распределение Бернулли.

X	0	1
p	q	p

$$EX = p, \text{ так как } EX = 0 \cdot q + 1 \cdot p = p.$$

$$Var\,X = pq = p - p^2 = p(1 - p) = pq.$$

2. Геометрическое распределение.

$$P(X = k) = pq^{k-1},\ k = 1, 2, \ldots$$

$$EX = \frac{1}{p},\ varX = \frac{q}{p^2}.$$

3. Биноминальное распределение.

$$P_n(k) = C_n^k p^k q^{n-k}, \text{ где } k = \overline{0, n},$$

$$EX = np,\ varX = npq.$$

4. Распределение Пуассона.

$$p(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},\ k = \overline{0, +\infty},\ \lambda > 0.$$

$$EX = \lambda,\ var\,X = \lambda^2.$$

3.6. Непрерывные случайные величины

Ранее мы рассматривали случайные величины, которые принимают изолированные или дискретные значения. Кроме таких случайных величин встречаются и другие случайные величины, например, которые принимают значения из некоторого интервала. Для описания таких случайных величин вводят понятие функции распределения случайной величины.

Пусть X – некоторая случайная величина.

Функцию $F_X(x_0) = P(x < x_0)$ называют функцией распределения вероятностей случайной величины (интегральной функцией). В дальнейшем, если понятно, о какой функции распределения идет речь, мы будем обозначать эту функцию $F(x)$. Функция распределения содержит всю вероятностную информацию о случайной величине.

Пример

Дискретная случайная величина:

X	1	2	5	8
p	0,2	0,4	0,25	0,15

График интегральной функции распределения приведен на рис. 5.

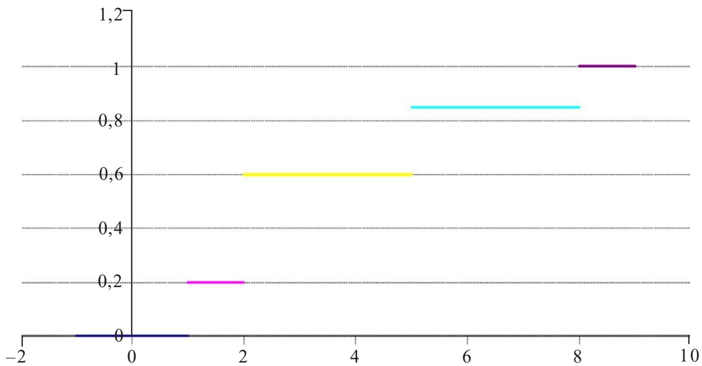

Рис. 5

Свойства функции распределения случайной величины X:

1. $0 \le F(x) \le 1$.

2. $F(x)$ – неубывающая функция, т. е. если $x_2 \ge x_1$, то $F(x_1) \ge F(x_2)$.

Покажем это. Пусть $x_2 > x_1$, тогда

$$P(x < x_2) = P(x < x_1) + P(x_1 \le x < x_2);$$

$$P(X < x_2) - P(X < x_1) = P(x_1 \le X < x_2);$$

$$F(x_2) - F(x_1) = P(x_1 \le x < x_2) \ge 0.$$

Следствие. Вероятность того, что случайная величина примет значения из интервала $[a, b)$, равна $P(a \le x < b) = F(b) - F(a)$.

Пример

$$F(x) = \begin{cases} 0, x \le -1; \\ \frac{1}{4}x + \frac{1}{4}, -1 < x \le 3; \\ 1, \ x > 3, \end{cases}$$

тогда $P(0 \le x < 2) = F(2) - F(0) = 0{,}5$.

3. $F(x) \to 0$, если $x \to -\infty$;

$F(x) \to 1$, если $x \to +\infty$.

Если случайная величина может принимать значения только из интервала (a, b), то

$$F(x) = 0, \text{ если } x < a \text{ и } F(x) = 1, \text{ если } x > b.$$

X – непрерывная случайная величина, если ее функция распределения кусочно-дифференцируема.

Утверждение. Вероятность того, что непрерывная случайная величина X примет определенное значение, равна нулю.

Доказательство. Действительно,

$$P(x_1 \le x \le x_1 + \Delta x) = F(x_1 + \Delta x) - F(x_1); \ \Delta x \to 0,$$

тогда, поскольку X – непрерывная случайная величина, то $P(X = x_1) = F(x_1) - F(x_1) = 0$. Утверждение доказано.

Для непрерывных случайных величин можно ввести, кроме функции распределения случайной величины, еще и плотность распределения.

П л о т н о с т ь р а с п р е д е л е н и я случайной величины X – функция $f_X(x) = F'_X(x)$.

Если в дальнейшем будет понятно, о какой случайной величине идет речь, мы будем обозначать плотность распределения случайной величины X просто $f(x)$.

Теорема. $P(a \le x \le b) = \int_a^b f(x)dx.$

$$P(a \le X < b) = F(b) - F(a) = \int_a^b F'(x)dx = \int_a^b f(x)dx.$$

Утверждение. Зная плотность распределения случайной величины, можно найти функцию распределения.

$F(x_0) = \int_{-\infty}^{x_0} f(x)dx$, так как $F(x_0) = P(x < x_0) = P(-\infty < x < x_0) =$

$= \int_{-\infty}^{x_0} f(x)dx.$

Свойства дифференциальной функции распределения:

1. $f(x) \ge 0$. Следует из того факта, что функция распределения неубывающая.

2. $\int_{-\infty}^{+\infty} f(x)dx = 1.$ Поскольку $\int_{-\infty}^{+\infty} f(x)dx = P(-\infty < x < +\infty) = 1.$

Если случайная величина принимает значения только из интервала (a, b), то $\int_a^b f(x)dx = 1.$

Плотность распределения называют еще законом распределения непрерывной случайной величины по аналогии с дискретными случайными величинами.

3.7. Числовые характеристики непрерывных случайных величин

Так же как и в случае дискретных случайных величин, полезно рассматривать некоторые характеристики случайных величин, которые описывают нашу случайную величину «в среднем». Необходимые характеристики приведены в следующей таблице.

Случайная величина определена на всей числовой оси (интегралы сходятся абсолютно)	Случайная величина принимает значения только из некоторого интервала (a, b)
$E(X)=\int\limits_{-\infty}^{+\infty} xf(x)dx$	$E(X)=\int\limits_{a}^{b} xf(x)dx$
$Var(X)=\int\limits_{-\infty}^{+\infty}[x-E(X)]^2 f(x)dx=$	$Var(X)=\int\limits_{a}^{b}[x-E(X)]^2 f(x)dx=$
$=\int\limits_{-\infty}^{+\infty} x^2 f(x)dx-[E(X)]^2$	$=\int\limits_{a}^{b} x^2 f(x)dx-[E(X)]^2$

$$\sigma(X)=\sqrt{var(X)}.$$

Все свойства, которые мы рассматривали для дискретных случайных величин, остаются в силе.

П р и м е р ы

1. Дана плотность вероятности $y = f(x)$ некоторой случайной величины X. Требуется:

1) Определить, чему равен параметр a.

2) Вычислить математическое ожидание, дисперсию и среднеквадратичное отклонение X.

3) Найти вероятность попадания случайной величины X в интервал [–0,5; 0,5].

4) Построить функцию распределения X.

5) Построить графики функции и плотности распределения.

$$f(x)=\begin{cases}0, x\le -1;\\ ax^4,\ \ -1<x\le 1;\\ 0,\ \ x>1.\end{cases}$$

1) Параметр a найдем из условия, которому должна удовлетворять любая плотность распределения: $\int\limits_{-\infty}^{+\infty} f(x)dx = 1.$

$$\int\limits_{-\infty}^{+\infty} f(x)dx = \int\limits_{-1}^{1} ax^4 dx = a\frac{x^5}{5}\Bigg|_{-1}^{1} = a\left(\frac{1}{5}+\frac{1}{5}\right) = \frac{2}{5}a. \text{ Откуда } a = \frac{5}{2}.$$

2) $E_X = \int\limits_{-\infty}^{+\infty} xf(x)dx$ – математическое ожидание;

$$varX = \int\limits_{-\infty}^{+\infty} [x - E_X]^2 f(x)dx = \int\limits_{-\infty}^{+\infty} x^2 f(x)dx - [E_X]^2 \text{ – дисперсия,}$$

$\sigma(X) = \sqrt{varX}$ – среднеквадратическое отклонение.

$$E_X = \int\limits_{-1}^{1} x\frac{5}{2}x^4 dx = \frac{5}{2}\int\limits_{-1}^{1} x^5 dx = \frac{5}{2}\cdot\frac{x^6}{6}\Bigg|_{-1}^{1} = \frac{5}{2}\left(\frac{1}{6}-\frac{1}{6}\right) = 0.$$

$$VarX = \int\limits_{-1}^{1} x^2\frac{5}{2}x^4 dx - 0^2 = \frac{5}{7},\ \sigma(X) = \sqrt{\frac{5}{7}}.$$

3) Вероятность попадания случайной величины в заданный интервал найдем по формуле $P(a \le x \le b) = \int\limits_{a}^{b} f(x)dx.$

$$P(-0,5 \le x \le 0,5) = \int\limits_{-0,5}^{0,5} \frac{5}{2}x^4 dx = \frac{5}{2}\cdot\frac{x^5}{5}\Bigg|_{-0,5}^{0,5} = \frac{1}{2}\left(\frac{1}{32}+\frac{1}{32}\right) = \frac{1}{32}.$$

4) Найдем функцию распределения X $F(x_0) = \int\limits_{-\infty}^{x_0} f(x)dx.$

$$f(x) = \begin{cases} 0, x \le -1; \\ \frac{5}{2}x^4,\ -1 < x \le 1; \\ 0,\ x > 1. \end{cases}$$

$$F(x) = \begin{cases} 0, x \le -1 \\ \int\limits_{-1}^{x} \frac{5}{2}x^4 dx,\ -1 < x \le 1 \\ 1,\ x > 1 \end{cases} = \begin{cases} 0, x \le -1 \\ \frac{x^5}{2}+\frac{1}{2},\ -1 < x \le 1. \\ 1,\ x > 1 \end{cases}$$

5) Построим графики функции и плотности распределения случайной величины X.

График плотности распределения X изображен на рис. 6.

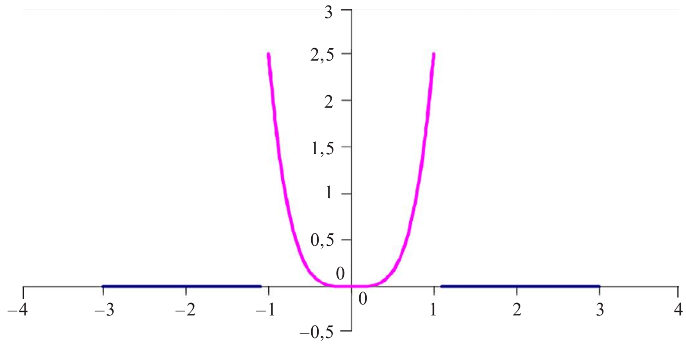

Рис. 6

График функции распределения X показан на рис. 7.

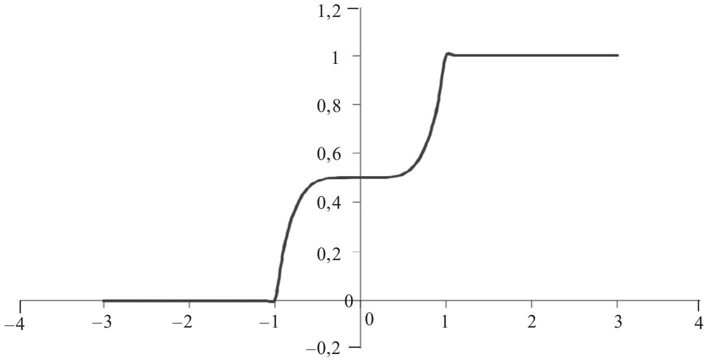

Рис. 7

2. Известно, что в партии из 20 телефонных аппаратов 5 недействующих. Случайным образом из этой партии взято 4 аппарата. Построить закон распределения случайной величины X – числа недействующих аппаратов из отобранных. Найти дисперсию этой величины. В каких единицах она измеряется? Построить интегральную функцию распределения случайной величины X, многоугольник распределения.

Случайная величина X – число недействующих аппаратов из отобранных четырех. Данная случайная величина – дискретная, принимающая следующие возможные значения:

$x_1 = 0$ – среди отобранных аппаратов все работающие;

$x_2 = 1$ – среди отобранных аппаратов только один недействующий;

$x_3 = 2$ – среди отобранных аппаратов ровно два недействующих;

$x_4 = 3$ – среди отобранных аппаратов ровно три недействующих;

$x_5 = 4$ – все четыре отобранных аппарата недействующие.

Найдем вероятности, с которыми случайная величина X принимает свои значения, для чего воспользуемся классическим определением вероятности.

Для всех пяти случаев элементарными исходами являются любые комбинации четырех телефонов из 20 телефонов. Число элементарных исходов: $n = C_{20}^4 = 4\,845$.

Благоприятные исходы:

1) набор из четырех работающих телефонов: $m_1 = C_{15}^4 = 1\,365$;

2) набор из трех работающих телефонов и одного неработающего: $m_2 = C_{15}^3 C_5^1 = 2\,275$;

3) набор из двух работающих телефонов и двух неработающих: $m_3 = C_{15}^2 C_5^2 = 1\,050$;

4) набор из одного работающего телефона и трех неработающих: $m_4 = C_{15}^1 C_5^3 = 150$;

5) набор из четырех неработающих телефонов: $m_5 = C_5^4 = 5$.

Итак:

$$p_1 = P(X = x_1) = \frac{m_1}{n} = \frac{1\,365}{4\,845} = 0,2817;$$

$$p_2 = P(X = x_2) = \frac{m_2}{n} = \frac{2\,275}{4\,845} = 0,4696;$$

$$p_3 = P(X = x_3) = \frac{m_3}{n} = \frac{1\,050}{4\,845} = 0,2167;$$

$$p_4 = P(X = x_4) = \frac{m_4}{n} = \frac{150}{4\,845} = 0,031;$$

$$p_5 = P(X = x_5) = \frac{m_5}{n} = \frac{5}{4\,845} = 0,001.$$

Закон распределения случайной величины – перечень возможных значений случайной величины с соответствующими вероятностями.

Закон распределения случайной величины X – числа неработающих телефонных аппаратов из отобранных четырех:

X	0	1	2	3	4	...	$p_1 + p_2 + \ldots + p_n = 1$
p	0,2817	0,4696	0,2167	0,031	0,001	...	

Тогда $E(X) = p_1x_1 + \ldots + p_nx_n = 0 \cdot 0{,}2817 + 1 \cdot 0{,}4696 + 2 \cdot 0{,}2167 + 3 \cdot 0{,}031 + 4 \cdot 0{,}001 = 1$.

Найдем дисперсию случайной величины, воспользовавшись следующей формулой:

$$varX = E(X^2) - (E_X)^2 = p_1x_1^2 + \ldots + p_nx_n^2 - (E_X)^2 =$$

$$0 \cdot 0{,}2718 + 1 \cdot 0{,}4696 + 4 \cdot 0{,}2167 + 9 \cdot 0{,}031 + 16 \cdot 0{,}001 - 1 = 0{,}63.$$

Найдем интегральную функцию распределения случайной величины.

$Fx(x_0) = P(x < x_0)$ – функция распределения случайной величины X.

$$F_X(x) = \begin{cases} 0, x \le 0; \\ P(X = 0), 0 < x \le 1; \\ P(X = 0) + P(X = 1), 1 < x \le 2; \\ P(X = 0) + P(X = 1) + P(X = 2), 2 < x \le 3; \\ P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3), 3 < x \le 4; \\ P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4),\ x > 4. \end{cases}$$

$$F_X(x) = \begin{cases} 0, x \le 0; \\ 0, 0 < x \le 1; \\ 0{,}2817,\ 1 < x \le 2; \\ 0{,}7513,\ 2 < x \le 3; \\ 0{,}968,\ 3 < x \le 4; \\ 1,\ x > 4. \end{cases}$$

М н о г о у г о л ь н и к   р а с п р е д е л е н и я (рис. 8) – ломаная линия, соединяющая точки с координатами (x_i, p_i).

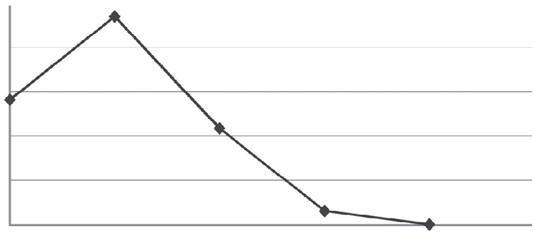

Рис. 8

Помимо математического ожидания, дисперсии и среднего квадратического отклонения рассматривают еще несколько характеристик случайных величин.

Мода распределения (рис. 9) – значение случайной величины, плотность распределения в которой принимает наибольшее значение:

$$MoX = arg\max_x f_X(x).$$

Если у распределения две моды, такое распределение называется бимодальным.

Медиана распределения – значение случайной величины, для которого вероятность попасть слева и справа от него одинакова (равна 0,5).

Медиану можно определить как по плотности распределения, так и по функции распределения.

По плотности распределения: $\int_{-\infty}^{MeX} f(x)dx = \int_{MeX}^{\infty} f(x)dx = 0,5.$

По функции распределения: $F_X(MeX) = 0,5.$

Для симметричных распределений, если математическое ожидание существует, то оно совпадает с медианой распределения.

Начальный момент распределения k-го порядка обозначается v_k и определяется следующим образом:

$$v_k = EX^k,\ k = 1,\ 2,\ \ldots$$

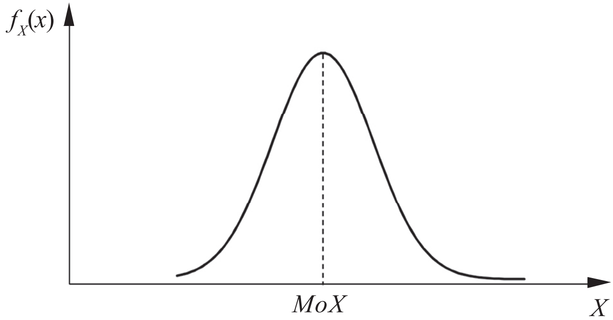

Рис. 9

Центральный момент k-го порядка обозначается μ_k и определяется следующим образом:

$$\mu_k = E(X - EX)^k,\ k = 1,\ 2,\ \ldots$$

Дисперсия – центральный момент второго порядка. Центральные моменты выражаются через начальные распределения. Например, $\mu_2 = E(X - EX)^2 = EX^2 - (EX)^2 = v^2 - v_1^2$.

Для симметричных распределений центральные моменты нечетного порядка равны нулю.

Особое значение имеют центральные моменты третьего и четвертого порядка.

Асимметрия – нормированный центральный момент третьего порядка:

$$As = \frac{\mu_3}{\sigma^3} = \frac{E(X - EX)^3}{\sigma^3}.$$

Для симметричных распределений асимметрия равна нулю, для несимметричных отлична от нуля.

Эксцесс распределения – нормированный центральный момент четвертого порядка:

$$K = \frac{\mu_4}{\sigma^4} = \frac{E(X - EX)^4}{\sigma^4}.$$

Эксцесс служит для сравнения какого-либо распределения с нормальным распределением. Для нормального распределения эксцесс равен трем. Он больше трех, если пик распределения около математического ожидания острый, и меньше трех, если пик очень гладкий.

К в а н т и л ь р а с п р е д е л е н и я: $x_q - q \cdot 100\,\%$ квантиль, если вероятность попасть случайной величины слева от x_q равна q:

$$P(X < x_q) = q.$$

Медиана – это 50 %-ный квантиль. В статистике возникают также квартили и децили. Квантиль можно найти при помощи плотности распределения и функции распределения.

По плотности распределения: $\int\limits_{-\infty}^{x_q} f(x)dx = q.$

По функции распределения: $F_X(x_q) = q.$

3.8. Основные распределения непрерывных случайных величин

Теперь познакомимся с некоторыми специальными распределениями, которые пригодятся нам в дальнейшем.

1. Равномерно распределенная случайная величина.

Пусть C – неизвестное число зависит от величины интервала, на котором определена случайная величина. Тогда плотность распределения имеет вид:

$$f(x) = \begin{cases} 0, & x \notin (a, b); \\ C, & x \in (a, b). \end{cases}$$

Константа C находится из условия: $\int\limits_a^b f(x)dx = 1.$ Получим: $\int\limits_a^b Cdx = 1$, откуда $C = \dfrac{1}{b-a}$.

Таким образом, плотность распределения равномерно распределенной случайной величины:

$$f(x)=\begin{cases}0, & x\notin(a,b);\\ \dfrac{1}{b-a}, & x\in(a,b).\end{cases}$$

График плотности распределения этой случайной величины представлен на рис. 10.

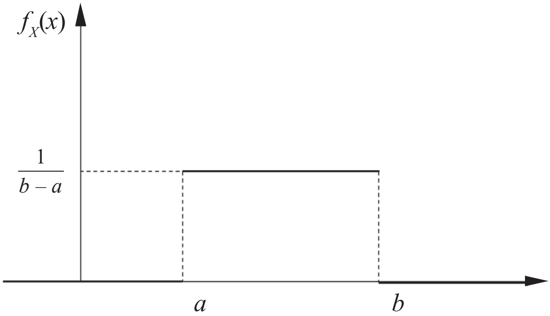

Рис. 10

Функция распределения равномерно распределенной случайной величины:

$$F(x)=\begin{cases}0, & x<a;\\ \dfrac{x-a}{b-a}, & x\in(a,b);\\ 1, & x>b.\end{cases}$$

График функции распределения изображен на рис. 11.

Математическое ожидание: $E(X)=\dfrac{a+b}{2}$.

Дисперсия: $var(X)=\dfrac{(b-a)^2}{12}$.

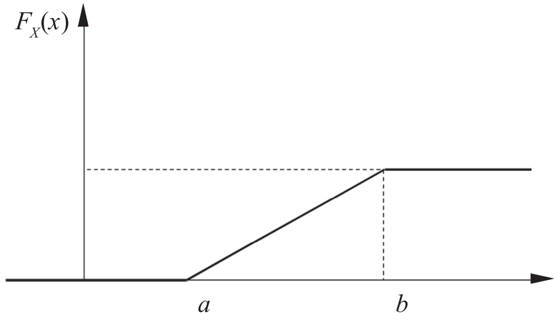

Рис. 11

2. Нормальное распределение.

Случайная величина X называется нормально распределенной случайной величиной с параметрами m и σ, если плотность ее распределения имеет следующий вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Математическое ожидание: $E(X) = m$.

Дисперсия: $var(X) = \sigma^2$.

Среднеквадратичное отклонение: $\sigma(X) = \sqrt{var(X)} = \sigma$.

График плотности нормального распределения называется *нормальной кривой* (кривой Гаусса) и изображен на рис. 12.

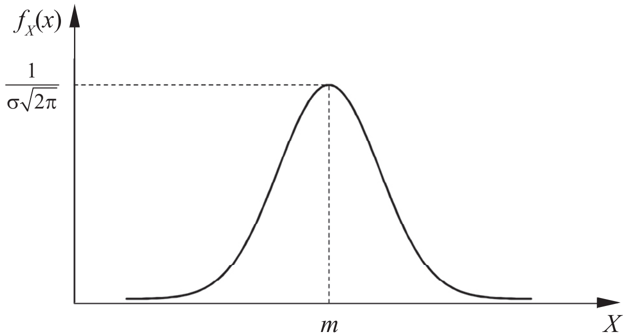

Рис. 12

С уменьшением σ кривая становится круче, с увеличением σ кривая выглаживается (рис. 13). При изменении m кривая не меняет формы, сдвигаясь вправо или влево (рис. 14).

При $x = m$ $y = \dfrac{1}{\sigma\sqrt{2\pi}}$.

Точки перегиба: $(m - \sigma, \dfrac{1}{\sqrt{2\pi e}})$; $(m + \sigma, \dfrac{1}{\sqrt{2\pi e}})$.

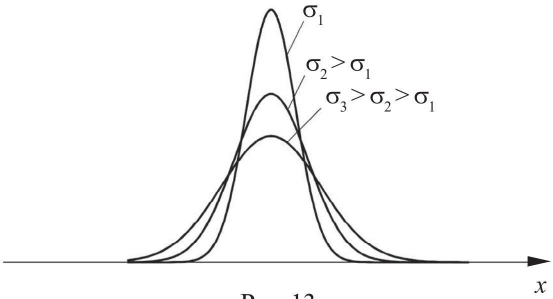

Рис. 13

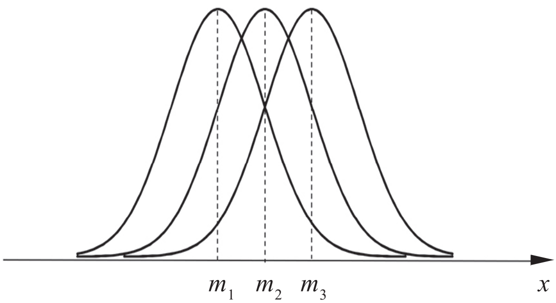

Рис. 14

Н о р м и р о в а н н о е   н о р м а л ь н о е   р а с п р е д е л е н и е – нормальное распределение с параметрами $m = 0$ и $\sigma = 1$. Оно имеет важное значение в теории вероятностей и, главным образом, в статистике.

Плотность распределения имеет следующий вид:

$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – эта функция затабулирована, т. е. имеются таблицы.

График плотности нормированного нормального распределения представлен на рис. 15.

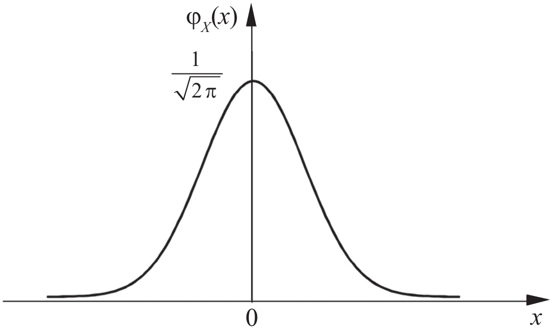

Рис. 15

Функция нормированного нормального распределения:

$$F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$ –таблицы для этой функции также имеются.

Утверждение. Пусть у нас имеется нормально распределенная случайная величина X с параметрами m и σ. Тогда случайная величина $\frac{X - m}{\sigma}$ имеет нормированное нормальное распределение.

Следующий факт предлагается читателю для самостоятельного доказательства:

$$F(x) = F_0\left(\frac{x - m}{\sigma}\right).$$

Функция Лапласа:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{z^2}{2}} dz.$$

Эта функция нечетная, т. е. $\Phi(-x) = -\Phi(x)$.

Таблицы этой функции имеются во всех учебниках по теории вероятностей и математической статистике.

Вероятность того, что нормированная нормальная случайная величина примет значения из интервала $(0,\ x_0)$, где $x_0 > 0$, равна следующему:

$$P(0 \le x \le x_0) = \int_0^{x_0} \varphi(x)dx = \Phi(x_0).$$

Замечание. Поскольку $\int_{-\infty}^{+\infty} \varphi(x)dx = 1$ и так как $\varphi(x)$ – симметрична относительно нуля, то $\int_{-\infty}^{0} \varphi(x)dx = \frac{1}{2}$, тогда $F_0(x) = 0{,}5 + \Phi(x)$.

Вероятность попадания в заданный интервал нормальной случайной величины

Пусть X – нормально распределенная случайная величина с параметрами m и σ.

$$P(A < X < B) = \int_A^B f(x)dx = \frac{1}{\sigma\sqrt{2\pi}}\int_A^B e^{-\frac{(x-m)^2}{2\sigma^2}}dx = \left[z = \frac{x-m}{\sigma}\right] =$$

$$= \frac{1}{\sqrt{2\pi}}\int_{\frac{A-m}{\sigma}}^{\frac{B-m}{\sigma}} e^{-\frac{z^2}{2}}dz = \Phi\left(\frac{B-m}{\sigma}\right) - \Phi\left(\frac{A-m}{\sigma}\right).$$

П р и м е р

Магазин продает мужские костюмы. По данным статистического исследования известно, что распределение по размерам является нормальным с математическим ожиданием и средним квадратическим отклонением, равными 48 и 2 соответственно. Определить процент спроса на 50-й размер при условии разброса значений этой величины в интервале (49, 51).

Решение. $P(49 < X < 51) = \Phi((51 - 48)/2) - \Phi((49 - 48)/2) = \Phi(1{,}5) - \Phi(0{,}5) = 0{,}4332 - 0{,}1915 = 0{,}2417$. Следовательно, спрос на 50-й размер костюмов составит около 24 %, и магазину нужно предусмотреть это в общем объеме закупки.

Вычисление вероятности заданного отклонения

Пусть X – нормально распределенная случайная величина с параметрами m и σ. Посчитаем вероятность того, что отклонение этой величины по абсолютному значению будет меньше заданного положительного числа δ.

$$P(|x-m|<\delta)=P(m-\delta<x<m+\delta)=$$
$$=\Phi\left(\frac{m+\delta-m}{\sigma}\right)-\Phi\left(\frac{m-\delta-m}{\sigma}\right)=2\Phi\left(\frac{\delta}{\sigma}\right),$$

поскольку функция Лапласа – нечетная функция (здесь мы воспользовались предыдущим результатом). Это выражение нам понадобится для построения доверительных интервалов.

Правило трех сигм

Правило трех сигм можно выразить следующим образом:

$$\delta = \sigma t.$$
$$P(|x-m|<\sigma t)=2\Phi(t);$$
$$t=3: P(|x-m|<3\sigma)=2\Phi(3)=2\cdot 0,49865=0,9973;$$
$$P(|x-m|>3\sigma)=0,0027,$$

т. е. практически невозможное событие.

На практике, если распределение случайной величины неизвестно, но правило трех сигм выполняется, то считают, что распределение нормальное; если не выполняется, то распределение нормальным не является.

3. Распределение χ^2.

Пусть X_1, ..., X_n – нормированные нормальные случайные величины, тогда $\chi^2=\sum_{i=1}^{n}X_i^2$ распределена по закону χ^2 с n степенями свободы. Если выполняется соотношение $X_n = X_1 + ... + X_{n-1}$, то число степеней свободы равно $n-1$.

Плотность распределения:

$$f_{\chi^2}(x) = \begin{cases} 0, \ x \le 0; \\ \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}, \ x > 0, \end{cases}$$

где $\Gamma(x) = \int\limits_0^\infty t^{x-1} e^{-t} dt$ – гамма-функция.

Графики плотности распределения χ^2 при $k = 5$, $k = 15$ и $k = 10$ изображены на рис. 16.

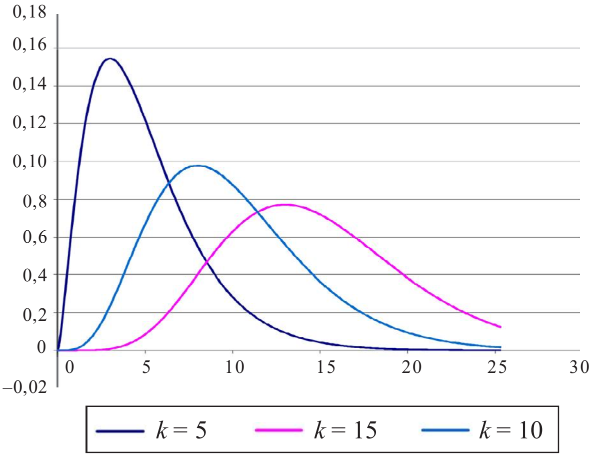

Рис. 16

4. Распределение Стьюдента.

Пусть Z – нормированная нормальная случайная величина, $\chi^2\ (k)$ – независимая с Z случайная величина, распределенная по закону χ^2 с k степенями свободы. Тогда случайная величина $T = \dfrac{Z}{\sqrt{\dfrac{\chi^2(k)}{k}}} = t(k)$ имеет t-распределение или распределение

Стьюдента с k степенями свободы. С возрастанием k распределение Стьюдента быстро приближается к нормированному нормальному распределению. Уже для k = 30 распределение Стьюдента становится почти нормированным нормальным.

Плотность распределения Стьюдента с k степенями свободы:

$$f_{t(k)}(x)=\frac{\Gamma(\frac{k}{2})}{\sqrt{\pi(k-1)}\cdot\Gamma(\frac{k-1}{2})}(1+\frac{x^2}{k-1})^{-\frac{1}{2}k},\ -\infty<x<+\infty.$$

Графики плотности распределения Стьюдента при $k = 5$, $k = 20$ и стандартного нормального распределения представлены на рис. 17.

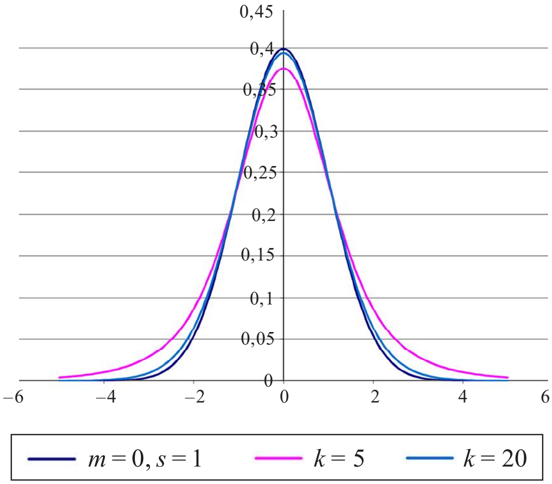

Рис. 17

$$E(t(k))=0,\ var(t(k))=\frac{k}{k-2}.$$

5. Распределение Фишера.

Пусть $\chi_1^2(k_1)$ и $\chi_2^2(k_2)$ – независимые случайные величины, распределенные по закону χ^2 с числом степеней свободы k_1 и k_2 соответственно. Тогда случайная величина $F(k_1,k_2)=\dfrac{\chi_1^2(k_1)\Big/k_1}{\chi_2^2(k_2)\Big/k_2}$ распределена по закону Фишера с числом степеней свободы числителя k_1 и знаменателя k_2.

Плотность распределения Фишера:

$$f_F(x)=\begin{cases}0, & x\le 0;\\ C_0\dfrac{x^{\frac{k_1-1}{2}}}{(k_2+k_1x)^{\frac{k_1+k_2}{2}}}, & x>0,\end{cases} \quad \text{где} \quad C_0=\frac{\Gamma(\frac{k_1+k_2}{2})k_1^{\frac{k_1}{2}}k_2^{\frac{k_2}{2}}}{\Gamma(\frac{k_1}{2})\Gamma(\frac{k_2}{2})}.$$

График плотности распределения Фишера изображен на рис. 18.

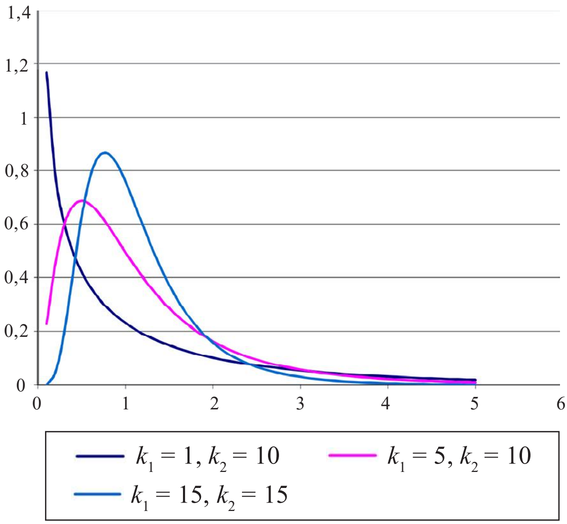

Рис. 18

3.9. Задачи

3.1. Дан ряд распределения случайной величины:

x_i	1	4	5	7
p_i	0,4	0,1	0,3	0,2

Найти и изобразить графически ее функцию распределения.

3.2. Известно, что случайная величина *X* имеет распределение:

–2	–1	0	1	2
0,1	0,3	0,2	?	0,2

(с одной недостающей вероятностью). 1) Построить график функции распределения случайной величины, вычислить ее математическое ожидание, дисперсию и среднеквадратическое отклонение. 2) Найти закон распределения случайной величины $Y = |X|$. Построить график распределения *Y* и найти ее математическое ожидание и дисперсию.

3.3. Случайная величина *X* принимает три значения: –1, 0, 1. Составить ряд ее распределения, если $E(X) = 0$, $var(X) = 0{,}5$.

3.4. По заданной функции распределения

$$F(x) = \begin{cases} 0 \text{ при } x \le -2; \\ \dfrac{1}{3} \text{ при } -2 < x \le 3; \\ \dfrac{5}{6} \text{ при } 3 < x \le 5; \\ 1 \text{ при } x > 5 \end{cases}$$

случайной величины *X* найти математическое ожидание и дисперсию.

3.5. Трое студентов сдают экзамен по математике на «отлично» (независимо друг от друга) с вероятностями 0,9; 0,8; 07. Пусть *X* – общее число полученных ими отличных оценок. Вычислить математическое ожидание и дисперсию случайной величины *X*.

3.6. Из урны, содержащей 10 белых и 15 черных шаров, наугад одновременно извлекают 8 шаров. Сколько в среднем будет из них белых шаров?

3.7. Клиенты банка, не связанные друг с другом, не возвращают кредиты в срок с вероятностью 0,1. Составить закон распределения числа возвращенных кредитов в срок из пяти выданных. Найти математическое ожидание, дисперсию и среднеквадратическое отклонение.

3.8. Вероятность выигрыша по облигации займа за все время его действия равна 0,1. Составить закон распределения числа выигравших облигаций среди приобретенных 19. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение и моду этой случайной величины.

3.9. Компания рассматривает проект строительства четырех домов в разных местах. Средства для строительства дают сами будущие жильцы. Вероятность набрать необходимые средства для постройки дома оценивается в 0,8 (собственно, речь идет об агитации будущих жильцов). Каждый построенный дом окупает 1/3 всех затрат компании по проекту. Найти распределение прибыли компании (через сумму затрат), вычислить ожидаемую прибыль.

3.10. В лотерее разыгрываются: автомобиль стоимостью 5 тыс. ден. ед., 4 телевизора стоимостью 250 ден. ед., 5 видеомагнитофонов стоимостью 200 ден. ед. Всего продается 1 тыс. билетов по 7 ден. ед. Составить закон распределения чистого выигрыша, полученного участником лотереи, купившим один билет.

3.11. Найти математическое ожидание случайной величины $z = 8X - 5Y + 7$, если известно, что $E(X) = 3$, $E(Y) = 2$.

3.12. Дискретная случайная величина X задана рядом распределения:

x_i	1	2	3	4	5
p_i	0,2	0,3	0,3	0,1	0,1

Найти условную вероятность события $X < 5$ при условии, что $X > 2$.

3.13. Сделано два высокорисковых вклада: 10 тыс. руб. в компанию A и 15 тыс. руб. – в компанию B. Компания A обещает 50 % годовых, но может «лопнуть» с вероятностью 0,2. Компания B обещает 40 % годовых, но может «лопнуть» с вероятностью 0,15. Составить закон распределения случайной величины – общей суммы прибыли (убытка), полученной от двух компаний через год, и найти ее математическое ожидание.

3.14. Случайные величины X и Y независимы и имеют один и тот же закон распределения:

x_i	1	2	4
p_i	0,2	0,3	0,5

Составить закон распределения случайных величин $2X$ и $X + Y$. Убедиться в том, что $2X$ не равно $X + Y$, но $E(2X) = E(X + Y)$.

3.15. Пусть X, Y, Z – случайные величины: X – выручка фирмы, Y – ее затраты, $Z = X - Y$ – прибыль. Найти распределение прибыли Z, если затраты и выручка независимы и заданы распределениями:

x_i	3	4	5
p_i	1/3	1/3	1/3

y_j	1	2
p_j	1/2	1/2

3.16. Охотник, имеющий 4 патрона, стреляет по дичи до первого попадания или до израсходования всех патронов. Вероятность попадания при первом выстреле равна 0,6, при каждом последующем – уменьшается на 0,1. Необходимо: а) составить закон распределения числа патронов, израсходованных охотником; б) найти математическое ожидание и дисперсию этой случайной величины.

3.17. Имеются 4 ключа, из которых только один подходит к замку. Составить закон распределения числа попыток открывания замка, если испробованный ключ в последующих попытках не участвует. Найти математическое ожидание, дисперсию и среднеквадратическое отклонение этой случайной величины.

3.18. При каком значении c функция

$$F(x) = \begin{cases} 0 \text{ при } x \le 1, \\ 1 - \dfrac{c}{x} \text{ при } x > 1 \end{cases}$$

служит функцией распределения некоторой случайной величины X? Какой в этом случае будет плотность вероятности величины и чему равно математическое ожидание?

3.19. Существует ли значение C такое, что функция $f(x)$ служит плотностью вероятности? Если существует, то указать его и вычислить математическое ожидание и дисперсию.

$$1)\ f(x) = \begin{cases} C \text{ при } 0 < x < 5, \\ 0 \text{ при } x < 0 \text{ и } x > 5; \end{cases}$$

$$2)\ f(x) = \begin{cases} C(1 - |x|) \text{ при } |x| < 1, \\ 0 \text{ при } |x| > 1; \end{cases}$$

$$3)\ f(x) = \begin{cases} Ce^{-2x} \text{ при } x > 0, \\ 0 \text{ при } x < 0; \end{cases}$$

$$4)\ f(x) = Ce^{-|x|} \text{ при } -\infty < x < +\infty;$$

$$5)\ f(x) = \begin{cases} Ce^{-\frac{x^2}{2}} \text{ при } x > 0, \\ 0 \text{ при } x < 0. \end{cases}$$

3.20. Для случайной величины $X \sim R(0,4)$ вычислить:
1) $P(X < EX)$; 2) $P(X > \sqrt{var\ X})$; 3) $P(-5 \le X \le 5)$.

3.21. Известно, что случайная величина X имеет равномерное распределение в интервале (a, b), причем $E(X) = varX = 3$. Найти числа a и b.

3.22. Поезда метрополитена идут регулярно с интервалом 2 мин. Пассажир выходит на платформу в случайный момент времени. Какова вероятность того, что ждать пассажиру придется не больше полминуты. Найти математическое ожидание и среднее квадратическое отклонение случайной величины X – времени ожидания поезда.

3.23. Доказать, что если $X \sim \exp(\lambda)$, то при $0 < a < b$ $P(a < X < b) = e^{-\lambda a} - e^{-\lambda b}$.

3.24. Установлено, что время ремонта телевизоров есть случайная величина X, распределенная по показательному закону. Определить вероятность того, что на ремонт телевизора потребуется не менее 20 дней, если среднее время ремонта телевизоров составляет 15 дней. Найти плотность вероятности, функцию распределения и среднее квадратическое отклонение случайной величины X.

3.25. Среднее время безотказной работы прибора равно 80 ч. Полагая, что время безотказной работы прибора имеет показательный закон распределения, найти: а) выражение его плотности вероятности и функции распределения; б) вероятность того, что в течение 100 ч прибор не выйдет из строя.

3.26. Вычислить вероятности попадания случайной величины $X \sim N(1, 4)$ в промежутки: $(-3; 1)$; $(-\infty; -2)$; $(3; \infty)$.

3.27. Для случайной величины $X \sim N(-1, 1)$ записать функцию распределения и плотность распределения, а также, используя соответствующие таблицы, найти x из условия: а) $P(x < X < 1) = 0{,}8$; б) $P(0 < X < x) = 0{,}8$; в) $P(-1 - x < X < -1 + x) = 0{,}8$.

3.28. Для случайной величины $X \sim N(-2, 9)$ вычислить $E((3 - X)(X + 5))$.

3.29. Текущая цена акции может быть смоделирована с помощью нормального закона распределения с математическим ожиданием 15 ден. ед. и средним квадратическим отклонением 0,2 ден. ед. Найти вероятность того, что цена акции: а) не выше 15,3 ден. ед.; б) не ниже 15,4 ден. ед.; в) от 14,9 до 15,3 ден. ед.

3.30. С помощью правила трех сигм найти границы, в которых будет находиться текущая цена акции.

3.31. Цена некой ценной бумаги нормально распределена. В течение последнего года 20 % рабочих дней она была ниже 88 ден. ед., а 75 % – выше 90 ден. ед. Найти: а) математическое ожидание и среднее квадратическое отклонение стоимости ценной бумаги; б) вероятность того, что в день покупки цена будет заключена в пределах от 83 до 96 ден. ед.; в) с надежностью 0,95 определить максимальное отклонение стоимости первой ценной бумаги от среднего (прогнозного) значения (по абсолютной величине).

3.32. Квантиль уровня 0,15 нормально распределенной случайной величины X равен 12, а квантиль уровня 0,6 равен 16. Найти математическое ожидание и среднее квадратическое отклонение случайной величины.

3.33. 20 %-ная точка нормально распределенной случайной величины равна 50, а 40 %-ная точка равна 35. Найти вероятность того, что случайная величина примет значение в интервале (25; 45).

3.34. Фирма, занимающаяся продажей товаров по каталогу, ежемесячно получает по почте заказы. Число этих заказов есть нормально распределенная случайная величина со средним квадратическим отклонением, равным 560, и неизвестным математическим ожиданием a. В 90 % случаев число ежемесячных заказов превышает 12 439. Найти среднее число заказов, получаемых фирмой за месяц. Какова вероятность того, что число ежемесячных заказов: а) менее 13 718; б) более 12 598?

3.35. Дан закон распределения дискретной случайной величины:

x_i	110	120	130	140	150
p_i	0,1	0,2	0,3	0,2	0,2

Вычислить ее математическое ожидание, дисперсию, среднеквадратическое отклонение. Найти функцию распределения, построить ее график. Построить многоугольник распределения. Найти вероятность того, что случайная величина примет значение в промежутке (120; 160).

3.36. Известна функция распределения дискретной случайной величины. Найти ее закон распределения и записать его в виде таблицы.

$$F(x)=\begin{cases} 0 \text{ при } x \le 3; \\ 0,3 \text{ при } 3 < x \le 6; \\ 0,7 \text{ при } 6 < x \le 9; \\ 1 \text{ при } x > 9. \end{cases}$$

3.37. В среднем по 10 % договоров страховая компания выплачивает страховые суммы в связи с наступлением страхового случая. Составить закон распределения числа таких договоров среди наудачу выбранных четырех. Вычислить математическое ожидание и дисперсию этой случайной величины.

3.38. Даны законы распределения двух независимых случайных величин X и Y:

x_i	0	1	3
p_i	0,2	0,5	?

y_j	2	3
p_j	0,4	?

Найти вероятности, с которыми случайные величины принимают значение 3, а затем составить закон распределения случайной величины $3X - 2Y$ и проверить выполнение свойств математических ожиданий и дисперсий:

$$E(3X - 2Y) = 3E(X) - 2E(Y),\ var(3X - 2Y) = 9var(X) + 4var(Y).$$

3.39. Распределение дискретной случайной величины X задано формулой $p(X = k) = Ck^2$, где $k = 1, 2, 3, 4, 5$. Найти: а) константу C; б) вероятность события $|X-2| \le 1$.

3.40. Случайная величина X задана функцией распределения

$$F(x)=\begin{cases}0 \text{ при } x\le -1;\\ Cx+0,75 \text{ при } -1<x\le 1/3;\\ 1 \text{ при } x>1/3;\\ 1 \text{ при } x>9.\end{cases}$$

Найти: 1) плотность распределения вероятностей; 2) неизвестный параметр C; 3) вероятность того, что в результате одного из испытаний случайная величина X примет значение, заключенное в интервале (–0,5; 1); 4) математическое ожидание и дисперсию.

3.41. Для случайной величины $X \sim N(2, 5)$, используя таблицу значений функции Лапласа, вычислить: 1) $P(1 < X < 3)$; 2) $P(X < 2)$; 3) $P(X \le 3)$; 4) $P(X > 2,5)$; 5) $P(|X| < 2)$; 6) $P(|X \ge 1)$.

3.42. Страховщик использует модель нормального распределения для анализа вероятных выплат по страховому портфелю. Среднее значение страховой выплаты – 980 ден. ед., стандартное отклонение – 120 ден. ед.

1) Найти вероятность того, что размер страховой выплаты составит: а) более 1 250 ден. ед.; б) меньше 850 ден. ед.; в) больше 700 ден. ед. и меньше 1 200 ден. ед.; г) отклонится от среднего значения страховой выплаты меньше чем на 50 ден. ед.; д) отклонится от среднего значения больше чем на 50 ден. ед.

2) Найти интервал, в котором отклонение страховой выплаты от среднего значения не превысит трехкратного стандартного отклонения (трех сигм).

3) С вероятностью 0,899 определить интервал, в котором будет находиться размер страховой выплаты. Какова при этом условии максимальная величина отклонения страховой выплаты от среднего значения?

3.43. Вероятность гибели объекта по договору страхования – 0,001. Какова вероятность того, что в 2 тыс. страховых договоров гибель объекта произойдет соответственно не менее чем в двух и не более чем в четырех договорах страхования?

3.44. Определить ожидаемый размер средней страховой выплаты по страховому портфелю, если предполагается, что размеры страховых выплат являются случайными величинами, распределенными по нормальному закону с дисперсией 22 500. Известно, что 28 % страховых выплат по портфелю имеют размер более 1 тыс. ден. ед.

3.45. Размер страховой выплаты по портфелю договоров имущественного страхования подчиняется нормальному закону распределения: ожидаемая средняя выплата – 375 ден. ед., стандартное отклонение – 25 ден. ед. Найти вероятность того, что размер страховой выплаты составит: а) от 300 до 425 ден. ед.; б) не более 450 ден. ед.; в) больше 300 ден. ед.

3.46. Страховщик полагает, что убытки от огневых рисков подчиняются экспоненциальному распределению, которое характеризуется функцией распределения: $F_Y(x) = 1 - e^{-\lambda x}$, $x > 0$. Требуется найти плотность распределения.

3.47. Среднее значение ущерба от огневых рисков по объекту за год оценивается страховщиком в размере 50 ден. ед. Определить: а) вероятность предъявления иска по страховой выплате в сумме 70 ден. ед.; б) вероятность того, что предъявленный иск будет не более 70 ден. ед.

3.48. Найти ожидаемую доходность и стандартное отклонение доходности портфеля, состоящего из 30 % акций компании A и 70 % акций компании B, если их доходности независимы и равны, соответственно, 25 % и 10 %, а стандартные отклонения – 10 % и 5 %.

4. МНОГОМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

4.1. Функция распределения многомерной случайной величины

Очень часто результаты испытания характеризуются не одной случайной величиной, а некоторой системой случайных величин $X_1, X_2, \ldots, X_n$, которую называют многомерной случайной величиной или случайным вектором $X = (X_1, \ldots, X_n)$.

П р и м е р ы

1. Успеваемость выпускника вуза характеризуется системой n случайных величин $X_1, X_2, \ldots, X_n$ – оценками по различным дисциплинам, выставленными в дипломе.

2. Погода в данном месте в определенное время суток может быть охарактеризована системой случайных величин: $X_1 – t$, °C; X_2 – влажность; X_3 – скорость ветра; X_4 – давление.

Случайные величины, входящие в систему, могут быть как дискретными, так и непрерывными.

С о в м е с т н а я ф у н к ц и я р а с п р е д е л е н и я вектора случайной величины – это совместная вероятность того, что:

$$F_{x_1, x_2, \ldots, x_n}(x_1, x_2, \ldots, x_k) = p((X_1 < x_1) \cdot (X_2 < x_2) \cdot (X_k < x_k)).$$

Если это двумерная случайная величина, т. е. (X, Y), то

$$F_{XY}(x, y) = P((X < x) \cdot (Y < y)).$$

Геометрическая интерпретация распределения $F_{XY}(x, y)$ представлена на рис. 19 и означает вероятность попадания случайной точки (x, y) в заштрихованную область – бесконечный квадрат, лежащий левее и ниже точки (x, y).

Свойства функции распределения:

1. Совместная функция распределения лежит в промежутке от 0 до 1, т. е.

$$0 \le F_{x_1, x_2, ..., x_n}(x_1, x_2, ..., x_k) \le 1.$$

2. Неубывающая по всем аргументам.

3. $F(-\infty) = \lim\limits_{x \to \infty} F(x),$ тогда

$$F_X(-\infty, ..., -\infty) = 0 \text{ или } F_X(-\infty, x_2, ..., x_k) = 0.$$

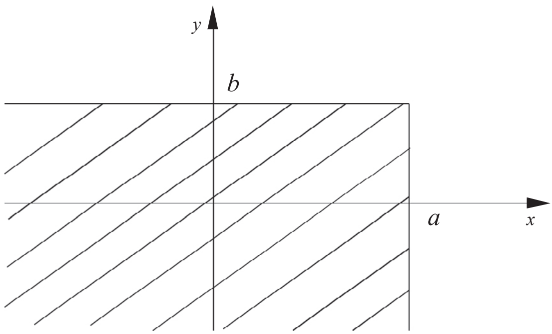

Рис. 19

4. $F_X(+\infty, +\infty, ..., +\infty) = 1.$

5. Если устремить к бесконечности какие-то аргументы (не все), то мы получим функцию распределения подсистемы из оставшихся аргументов:

$$F_X(+\infty, +\infty, ..., +\infty, x_k) = F_{X_k}(x_k).$$

Если $k = 5$, то

$$F_{X_1, X_2, X_3, X_4, X_5}(x_1, x_2, +\infty, +\infty, x_5) = F_{X_1, X_4, X_5}(x_1, x_4, x_5).$$

6. Если случайные величины $x_1, x_2, \ldots, x_k$ независимы, то совместная величина:

$$F_{X_1, ..., X_k}(x_1, ..., x_k) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot ... \cdot F_{X_k}(x_k).$$

4.2. Двумерное дискретное распределение

Закон распределения:

$$p_{ij} = p((X = x_i)(Y = y_j)).$$

$X \backslash Y$	Y_1	Y_2	...	Y_m
X_1	p_{11}	p_{12}	…	p_{1m}
X_2	p_{21}	p_{23}	…	p_{2m}
...	…	…	…	…
X_n	p_{n_1}	…	…	p_{nm}

Если X и Y независимы, то:

$$p_{ij} = p((X = x_i)(Y = y_j)) = p(X = x_i) \cdot p(Y = y_j) = p_{xi} \cdot p_{yj}...$$

$$\sum p_{ij} = 1;$$

$$p(X = x_i) = p_{x_i} = \sum_{j=1}^{m} p_{ij};\ p(Y = y_j) = p_{y_j} = \sum_{i=1}^{m} p_{ij}.$$

П р и м е р

Пол (X) / оценка (Y)	2	3	4	5
Ж-0	0,05	0,15	0,2	0,07
М-1	0,15	0,25	0,1	0,03

Найдем закон распределения одновременных случайных величин X и Y:

X	0	1
p	0,47	0,53

X	2	3	4	5
p	0,2	0,4	0,3	0,1

$$EY = 3{,}3;$$

$$varY = 0{,}81.$$

$$0{,}47 \cdot 0{,}2 \neq 0{,}05 \Rightarrow \text{случайные величины } X \text{ и } Y \text{ зависимы.}$$

Условные законы распределения:

$$p_{y_j|x_i} = p(Y = y_j \mid X = x_i) = \frac{p(Y = y_j) \cdot p(X = x_i)}{p(X = x_i)} = \frac{p_{ij}}{p_{xi}};$$

$$p_{y_j|x_i} = p(Y = y_j \mid X = x_i) = \frac{p_{ij}}{p_{y_j}}.$$

П р и м е р

В предыдущем примере найдем $Y|X = 0$ и $Y|X = 1$.

$Y\|X$	2	3	4	5
p	0,1064	0,3191	0,4255	0,1489

$$E\left[Y \mid X = 0\right] = \sum p_{Y|X=0} \cdot y = 3{,}61166,\ cov(X,Y) = cov(X,Y).$$

$Y\|X = 1$	2	3	4	5
p	0,283	0,4717	0,1887	0,0566

$$E\left[Y \mid X = 1\right] = \sum p_{Y|X=1} \cdot y = 3{,}0189.$$

Утверждение. Если X и Y независимы, то условные законы распределения совпадают с безусловными.

4.3. Условное математическое ожидание в условных законах распределения

К о р р е л я ц и я – вид статистической зависимости, при которой математическое ожидание первой случайной величины зависит от того, какое математическое ожидание имеет другая случайная величина.

Закон итерации:

$$E\left[E\left[Y \mid X\right]\right]=E_Y.$$

Если $var\left[Y \mid X=0\right] \neq var\left[Y \mid X=1\right]$, то присутствует г е т е р о с к е д а с т и ч н о с т ь, в ином случае – г о м о с к е д а с т и ч н о с т ь.

Математическое ожидание и дисперсия случайных величин X и Y недостаточно полно характеризуют двумерную случайную величину, так как не выражают степень зависимости ее составляющих X и Y. Эту роль выполняет к о в а р и а ц и я:

$$cov(X,Y)=E\left[(X-E_X)(Y-E_Y)\right].$$

Утверждение. $Cov(X,Y)=E\left[(X-E_X)(Y-E_Y)\right]=E(XY)-E_X \cdot E_Y.$

Доказательство:

$$E\left[(X-EX)(Y-EY)\right]=E(XY-XEY-YEX+EXEY=$$
$$=E(XY)-E(XE_Y)-E(YEX)+E(EXEY)=$$
$$=E(XY)-EXEY-EXEY+EXEY=E(XY)-EX \cdot EY.$$

Утверждение доказано.

Утверждение. $Cov(X,X)=var(X).$

Ковариация показывает наличие линейной корреляционной связи между X и Y. На рис. 20, так как больше точек там, где производная положительна, следовательно, $cov(X, Y) > 0$. На рис. 21, так как больше точек там, где производная отрицательна, следовательно, $cov(X, Y) < 0$.

На рис. 22 и 23 представлены случаи, когда $cov(X, Y) = 0$.

Свойства ковариации:

1. $Cov(X, Y)$ – величина детерминированная.

$Cov(X, Y) > 0$ значит, что функция $E[Y|X]$ возрастает (с ростом X, Y, в среднем, растет).

$Cov(X, Y) < 0$ значит, что функция $E[Y|X]$ убывает (с ростом X, Y, в среднем, убывает).

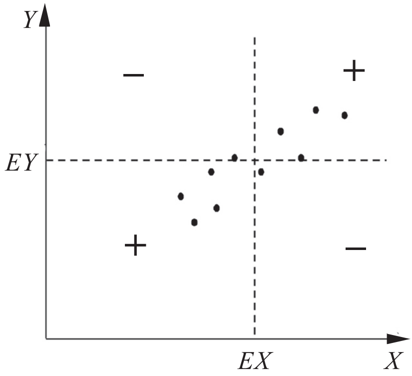

Рис. 20

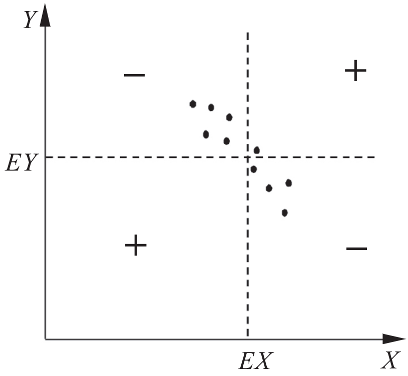

Рис. 21

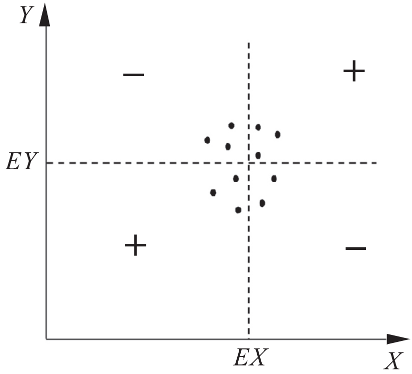

Рис. 22

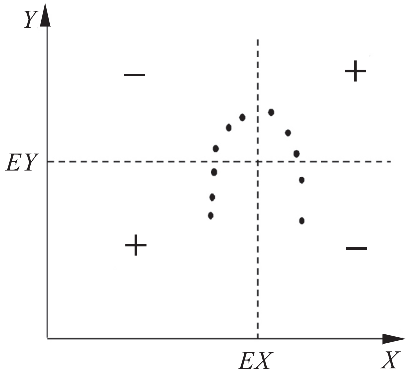

Рис. 23

Утверждение. Если X и Y независимы, то $cov(X, Y) = 0$. Отметим, что обратное верно не всегда.

Утверждение. X и Y независимы и нормальны $\Leftrightarrow cov(X, Y) = 0$.

2. $Cov(X, X) = var(X)$.
3. $Cov(X, Y) = cov(Y, X)$.
4. $Cov(aX + bY, Z) = acov(X, Z) + bcov(Y, Z)$, где X, Y, Z – случайные величины.

С помощью ковариации можно дополнить и уточнить некоторые свойства дисперсии.

Утверждение. $Var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$;

$$var(X-Y) = var(X) + var(Y) - 2cov(X,Y).$$

Введем понятие к о э ф ф и ц и е н т а   к о р р е л я ц и и:

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}.$$

Свойства коэффициента корреляции:

1. $|Corr(X,Y)| \leq 1$.
2. $\text{Sign}(corr(X,Y) = \text{sign}(cov(X,Y))$.

Утверждение. Если $corr(X,Y) \leq 0 \Rightarrow cov(X,Y) < 0 \Rightarrow E(Y|X)$ убывает; если $corr(X,Y) > 0 \Rightarrow cov(X,Y) > 0 \Rightarrow E(Y|X)$ возрастает.

Утверждение. Если X и Y независимы, то $corr(X, Y) = 0$.

Утверждение. Если X и Y независимы и нормальны, то $\Leftrightarrow$ $corr(X, Y) = 0$.

Если $corr(X, Y) = 0$ и $cov(X, Y) = 0$, то переменные не коррелируют линейно (отсутствует линейная зависимость, однако может наблюдаться другой вид зависимости, например, параболоидный).

3. $|Corr(X,Y)| = 1 \Leftrightarrow X, Y$ линейно зависимы: $Y = aX + b$.

Причем, если $b > 0 \Rightarrow corr(X, Y) = 1$, а если $b < 0 \Rightarrow corr(X, Y) = –1$.

4. От линейного преобразования может поменяться только знак линейной корреляции:

$$|corr(a+bX, Y)| = |corr(X,Y)|.$$

Рассмотрим $U = \dfrac{X - E_X}{\sigma_X}$; $V = \dfrac{Y - E_Y}{\sigma_Y}$ – нормированные переменные.

Утверждение. Нормирование не влияет на корреляцию, т. е. $corr(X, Y) = corr(U, V)$.

М а т р и ц а   к о в а р и а ц и й выглядит следующим образом:

$$\sum=\left\{cov\left(X_i, Y_j\right)\right\}_{i,\,j=1}^{k}=\begin{pmatrix} cov_{X_1,X_1} & cov_{X_1,X_2} & \dots & cov_{X_1,X_k} \\ cov_{X_2,X_1} & cov_{X_2,X_2} & \dots & cov_{X_2,X_k} \\ \dots & \dots & \dots & \dots \\ cov_{X_k,X_1} & cov_{X_k,X_2} & \dots & cov_{X_k,X_k} \end{pmatrix}=$$

$$=\begin{pmatrix} var_{X_1} & cov_{X_1,X_2} & \dots & cov_{X_1,X_k} \\ cov_{X_2,X_1} & var_{X_2} & \dots & cov_{X_2,X_k} \\ \dots & \dots & \dots & \dots \\ cov_{X_k,X_1} & cov_{X_k,X_2} & \dots & var_{X_K} \end{pmatrix}.$$

Свойства матрицы ковариаций:

1. Симметричная, положительно определенная матрица.
2. $Y = a + HX$, тогда $E_Y = a + H \cdot E(X)$.

П р и м е р

$$\sum=\left\{cov(X_i, Y_j)\right\}_{i,j=1}^{k}=\begin{pmatrix} 4 & 1 & 0,5 \\ 1 & 16 & 0 \\ 0,5 & 0 & 9 \end{pmatrix}.$$

Требуется найти $var(2X-5Y)$ и $cov(X+2Y; -3Y+4Z)$.

Решение:

$$var(2X-5Y)=var(2X)+var(5Y)-2cov(2X;5Y)=$$
$$=4var(X)+25var(Y)-20cov(X;Y)=4\cdot 4+25\cdot 16-20\cdot 1=396.$$

Найти $cov(X+2Y; -3Y+4Z)$ предлагается читателю самостоятельно.

4.4. Двумерная непрерывная случайная величина

Двумерная непрерывная случайная величина (X, Y) задается совместной функцией распределения $F_{XY}(x, y) = p((X < x) \cdot (Y < y))$. Тогда совместная функция плотности имеет вид:

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}.$$

Свойства совместной функции плотности двумерной непрерывной величины:

1. $f_{XY}(x, y) \geq 0$.

2. $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{XY}(x, y)\, dxdy = 1$.

3. X и Y – независимы $\Leftrightarrow$ $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$.

4. Зная совместную плотность $f_{XY}(x, y)$, можно найти индивидуальные плотности распределения и математические ожидания случайных величин X и Y:

$$f_X(x) = \int\limits_{-\infty}^{+\infty} f_{XY}(x, y)dy; \qquad f_Y(y) = \int\limits_{-\infty}^{+\infty} f_{XY}(x, y)dx;$$

$$EX = \iint\limits_D x \cdot f_{XY}(x, y)dxdy; \qquad EY = \iint\limits_D y \cdot f_{XY}(x, y)dxdy.$$

Условные законы распределения выглядят следующим образом:

$$f_{Y|X}(x, y) = \frac{f_{X,Y}(x, y)}{f_X(x)};$$

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)};$$

$$f(x) = E[Y \mid X] = \int\limits_{-\infty}^{+\infty} y \cdot f_{Y|X}(x, y)dy.$$

Утверждение. Закон итерационных математических ожиданий будет иметь вид:

$$E[E[Y \mid X]] = E_Y.$$

Утверждение. $Cov(X, Y) = E((X - E_X)\ (Y - E_Y)) = E(XY) - E_X - E_Y$,

где $E(x, y) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xyf_{XY}(x, y)\, dxdy.$

4.5. Многомерное нормальное распределение

Говорят, что многомерная случайная величина имеет многомерное нормальное распределение, если существует вектор $m = (m_1, m_2, ..., m_k)'$ и матрица $\sum$ размерностью $k \times k$ такие, что совместная площадь распределения имеет следующий вид:

$$f_{X_1, X_2, ..., X_k}(x_1, ..., x_k) =$$

$$= \frac{1}{(2\pi)^{k/2} \cdot (\det \Sigma)^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x - m)^1 \Sigma^{-1}(x - m)\right\}.$$

Свойства многомерного нормального распределения:

1. Любая компонента нормального вектора имеет нормальное распределение:

$$X_i \sim N(m_i, \sigma_{ii}).$$

2. Условное распределение $X_i|X_j$ также имеет нормальное распределение:

$$X_i|X_j \sim N(...).$$

3. Любое линейное преобразование нормального вектора также является нормальным вектором:

$$Y \sim N(m, \sum),$$

$$Y = b + HX;$$

$$Y \sim N(b + Hm, H\sum H^T).$$

4. Если два нормальных случайных вектора не коррелируют, то они независимы.

5. $(X - m)'\Sigma^{-1}(X - m) \sim \chi^2(k)$.

Пример

$$X \sim N\left(\begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix}; \begin{pmatrix} 4 & 1 & 0,5 \\ 1 & 16 & 0 \\ 0,5 & 0 & 9 \end{pmatrix}\right).$$

1) Выясним, являются ли независимыми $2X_1 + X_2$ и $X_3 - 4X_1$, где X_1, X_2, X_3 – нормальные векторы, так как составляющие $X \sim N$;

$2X_1 + X_2$ – нормальные, так как представляют собой линейную комбинацию нормальных векторов.

$$Cov(2X_1 + X_2; X_3 - 4X_1) =$$
$$= cov(2X_1; X_3) + cov(2X_1; -4X_1) + cov(X_2; X_3) + cov(X_2; -4X_1) =$$
$$= 2cov(X_1; X_3) - 8cov(X_1; X_1) + cov(X_2; X_3) - 4cov(X_2; X_1) =$$
$$= 2 \cdot 0,5 - 8 \cdot 4 + 0 - 4 \cdot (-1) = -27 \neq 0.$$

Следовательно, $2X_1 + X_2$ и $X_3 - 4X_1$ зависимы.

2) Найдем $p(X_1 - 2X_2 > 0)$.

$(X_1 - 2X_2) \sim N(-1; 44)$ – это верно, так как $E(X_1 - 2X_2) = E_{X_1} - 2E_{X_2} =$ $= 3 - 2 \cdot 2 = -1$;

$$var(X_1 - 2X_2) = varX_1 - 4varX_2 - 4cov(X_1; X_2) = 44;$$

$$p(X_1 - 2X_2 > 0) = \Phi(-\infty) - \Phi\left(\frac{0-(-1)}{\sqrt{44}}\right) = 0,5 - 0,0596 = 0,04404.$$

4.6. Задачи

4.1. Известно совместное распределение вероятностей случайных величин X и Y:

X \ Y	–1	0	1
–1	1/6	1/12	7/24
1	1/8	1/3	0

1) Найти частные распределения этих величин, их математические ожидания и дисперсии.

2) Решить, какими являются величины X и Y – зависимыми или независимыми.

3) Составить ковариационную и корреляционную матрицы случайного вектора (X, Y), выяснить, какими являются случайные величины X и Y – коррелированными или некоррелированными.

4) Описать регрессии $H(X)$ и $G(Y)$ величины Y на X и величины X на Y.

5) Построить наилучшие в среднем квадратическом оценки $\hat{Y} = H(X)$ и $\hat{X} = G(Y)$.

6) Проверить по данной задаче формулу полного математического ожидания.

4.2. Предположим, что совместная плотность распределения случайных величии следующая:

$$f(x,y) = \begin{cases} ce^{-x-y}, & 0 < x < \infty \text{ и } 0 < y < \infty, \\ 0, \text{ в противном случае.} \end{cases}$$

1) Найти константу c.

2) Являются ли эти величины некоррелированными? независимыми?

3) Найти вероятность $P(X < Y)$.

4.3. Дана матрица ковариаций системы трех случайных величин X, Y, Z:

$$\begin{pmatrix} 2 & -1 & 0{,}5 \\ -1 & 4 & -0{,}3 \\ 0{,}5 & -0{,}3 & 3 \end{pmatrix}.$$

1) Чему равна дисперсия случайной величины $4X - 2Z + 3$?

2) Найти $cov(2X + 3Y, -X + 4Z)$.

3) По данной матрице построить матрицу корреляций.

4.4. Задано распределение вероятностей дискретной двумерной случайной величины (X, Y):

$Y \backslash X$	-1	0	1
1	0,15	0,3	0,3
2	0,1	0,05	0,1

Здесь случайная величина X описывает доход инвестиционной компании на рынке акций, а случайная величина Y – доход на рынке облигаций. Составить ряды распределения ее компонент X и Y, а также условный закон распределения компоненты X при условии $Y = 2$. Выяснить, зависимы ли компоненты X и Y. Найти закон распределения суммарного дохода компании $X + Y$.

4.5. Задано распределение вероятностей дискретной двумерной случайной величины (X, Y):

$Y \backslash X$	-1	0	1
0	0	0,1	0,4
1	0,2	0,2	0,1

Составить ряды распределения ее компонент X и Y. Определить вероятность $P(X < Y)$.

4.6. Ожидаемая доходность первого актива равна 8 % со средним квадратичным отклонением 7 %, ожидаемая доходность второго актива равна 11 % со средним квадратичным отклонением 10 %. Коэффициент корреляции между этими активами составляет 0,7. Найти ожидаемую доходность и среднее квадратичное отклонение портфеля, состоящего на 35 % из первого актива и на 65 % – из второго.

4.7. Система случайных величин (X, Y) подчинена закону распределения с плотностью $f(x, y)$, где $f(x, y) = a(x + y)$ в области D и $f(x, y) = 0$ вне этой области. Область D – квадрат, ограниченный прямыми $x = 0$, $x = 3$, $y = 0$, $y = 0$. Требуется:

1) определить коэффициент a;

2) вычислить вероятность попадания случайной точки (X, Y) в квадрат Q, ограниченный прямыми $x = 1$, $x = 2$, $y = 1$, $y = 2$;

3) найти математические ожидания m_x и m_y;

4) рассчитать средние квадратичные отклонения σ_x и σ_y.

4.8. Система случайных величин (X, Y) подчинена закону распределения с плотностью $f(x, y)$, где $f(x, y) = a\sin(x + y)$ в области D и $f(x, y) = 0$ вне этой области.

Область D определяется неравенствами $0 \le x \le \frac{\pi}{2}$, $0 \le y \le \frac{\pi}{2}$. Найти:

1) коэффициент a;

2) математические ожидания m_x и m_y;

3) средние квадратичные отклонения σ_x и σ_y;

4) коэффициент корреляции r_{xy}.

5. ЗАКОН БОЛЬШИХ ЧИСЕЛ И ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

5.1. Закон больших чисел

Под законом больших чисел в широком смысле понимается общий принцип, согласно которому, по Колмогорову, совокупное действие случайных факторов приводит к результату, почти не зависящему от случая.

Под законом больших чисел в узком смысле понимают ряд математических теорем, в каждой из которых для тех или иных совокупностей условий устанавливается факт приближения средних характеристик большого числа испытаний к некоторым определенным постоянным.

Имеем некоторую последовательность x_n: $\{x_n\}_{n=1}^{\infty}$.

Говорят, что x_n сходится к x по вероятности: $x_n \xrightarrow[n\to\infty]{P} x$, если $\forall\varepsilon > 0, \lim\limits_{n\to\infty} x_n = x$.

Это записывается так: $p\lim\limits_{n\to\infty} x_n = x$.

Говорят, что последовательность случайных величин x_n сходится по распределению к случайной величине x, если сходятся их функции распределения, т. е.:

$$\{x_n\}_{n=1}^{\infty},\ x_n \to x, \text{ если } F_{x_n}(x_n) \xrightarrow[n\to\infty]{} F_x(x).$$

Лемма 1. Неравенство Маркова.

Пусть X – случайная величина, принимающая неотрицательные значения, у которой есть математическое ожидание EX.

Тогда $p(x \geq A) \leq \dfrac{EX}{A}$ для любого $A > 0$.

Доказательство для дискретной случайной величины:

X	X_1	X_2	...	X_n
P	P_1	P_2	...	P_n

Для любого $A > 0$ предположим, что A или равно X_n, или принадлежит $X_{n-1} \le A \le X_n$.

$$EX = x_1 p_1 + \ldots + x_{k-1} p_{k-1} + x_k p_k + \ldots + x_{k+1} p_{k+1} \ge$$
$$\ge x_k p_k + x_{k+1} p_{k+1} + \ldots \ge A p_k + A p_{k+1} + \ldots = A(p_k + p_{k+1}) = A(p(x \ge A);$$

$$EX \ge A(p(x \ge A)) => p(x \ge A) \le \frac{EX}{A}.$$ Лемма доказана.

Следствие. Неравенство Маркова в другом виде.

$$p(x < A) \ge 1 - \frac{EX}{A} \text{ для любого } A > 0.$$

Лемма 2. Неравенство Чебышева.

X – случайная величина с математическим ожиданием EX и дисперсией $var(X)$.

Тогда для любого $\varepsilon > 0, p \mid x - E_x \mid < \varepsilon \ge 1 - \frac{var(x)}{\varepsilon^2}.$

Доказательство. Применим неравенство Маркова к величине $(x - E_x)^2$; $A = \varepsilon^2$.

$$p((x - E_x)^2 < \varepsilon^2) \ge 1 - \frac{E(x - EX)^2}{\varepsilon^2} \Rightarrow p(|x - E_x| < \varepsilon) \ge 1 - \frac{var\,(x)}{\varepsilon^2}.$$

Лемма доказана.

В приведенной лемме 2 дается оценка вероятности того, что отклонение случайной величины от своего математического ожидания не превысит по абсолютному значению положительную величину ε. Если ε достаточно мало, то мы, таким образом, оценили вероятность того, что случайная величина X примет значения, достаточно близкие к математическому ожиданию. Для практики это неравенство имеет ограниченное значение, поскольку эта оценка груба. Однако велико теоретическое значение неравенства Чебышева. Оно используется для доказательства следующего утверждения, именуемого теоремой Чебышева.

Теорема Чебышева. Закон больших чисел.

Пусть имеется последовательность случайных величин $\{x_i\}_{i=1}^{n}$ – последовательность независимых случайных величин, каждая из которых имеет математическое ожидание и дисперсию и которая равномерно ограничена, т. е.:

$$\exists EX_i = m_i; var(x_i) = \sigma_i^2;$$

$$\exists c : \forall i;\ var(x_i) \le c.$$

Тогда их среднее арифметическое сходится по вероятности к среднему арифметическому их математических ожиданий, т. е.:

$$p\lim_{n\to\infty}\frac{n}{x} = m$$

или

$$p\lim_{n\to\infty}\left(\frac{x_1 + ... + x_n}{n} - \frac{EX_1 + ... + EX_n}{n}\right) = 0.$$

Доказательство. Применим неравенство Чебышева к случайной величине $\overline{x} = \frac{x_1 + ... + x_n}{n}$.

Необходимо доказать, что

$$\lim_{n\to\infty} p\left(\left|\frac{x_1 + ... + x_n}{n} - \frac{EX_1 + ... + EX_n}{n}\right| < E\right) = 1.$$

1) $E\left(\frac{x_1 + ... + x_n}{n}\right) = \frac{E}{n}(x_1 + ... + x_n) = \frac{1}{n}\sum_{i=1}^{n} E_i = \frac{EX_1 + ... + EX_n}{n}.$

2) $Var\left(\frac{x_1 + ... + x_n}{n}\right) = \frac{1}{n^2}(varX_1 + ... + varX_n)) \le \frac{C + C + ... + C}{n^2} =$

$= \frac{C \cdot n}{n^2} = \frac{C}{n};$

$$p\left(\left|\frac{x_1+\ldots+x_n}{n}-\frac{EX_1+\ldots+EX_n}{n}\right|<\varepsilon\right)\geq 1-\frac{var\left(\frac{x_1+\ldots+x_n}{n}\right)}{\varepsilon^2}\geq 1-\frac{\frac{C}{n}}{\varepsilon^2};$$

$$1\geq p\left(\left|\frac{x_1+\ldots+x_n}{n}-\frac{EX_1+\ldots+EX_n}{n}\right|<\varepsilon\right)\geq 1-\frac{C}{n\varepsilon^2},$$

следовательно, $\frac{x_1+\ldots+x_n}{n}$ сходится по вероятности к $\frac{EX_1+\ldots+EX_n}{n}$. Теорема доказана.

Замечание. Одинаково распределенные взаимно независимые случайные величины $\{x_1, \ldots, x_n\}$будем обозначать *idd* (m, σ^2), что является сокращением от английского *independent identically distributed*.

Следствие. Пусть x_n – последовательность, такая, что $x_i \sim (m, \sigma^2)$, тогда

$$\frac{EX_1+\ldots+EX_n}{n}=\frac{m+\ldots+m}{n}=\frac{n\cdot m}{n}=m=EX,$$

и по теореме Чебышева $p > p$ (состоятельность выборочного среднего).

Следствие. Частота события A в серии из n независимых испытаний, в каждом из которых оно может произойти с вероятностями $p_1, \ldots, p_n$ соответственно при неограниченном увеличении числа n, сходится по вероятности к среднему арифметическому вероятности события в отдельных испытаниях, т. е.

$$\lim_{n\to\infty} P\left(\left|\frac{k}{n}-\frac{p_1+\ldots+p_n}{n}\right|\leq\varepsilon\right)=1.$$

Стоит отметить, что закон больших чисел в экономике и социально-экономической статистике – это проявление одного из важнейших объективных законов.

П р и м е р

Пусть необходимо дать оценку доходов населения страны. Возьмем 15 наблюдений, у 10-ти респондентов доходы примерно составляют 30 тыс. ден. ед., а у 5-ти – 150 тыс. ден. ед., тогда средний доход составит 70 тыс. ден. ед. И это вовсе не отражает реальный уровень дохода. Если же рассмотреть 200 наблюдений, в которых у 180 человек доходы 20 тыс. ден. ед., а у 20-ти – 120 тыс. ден. ед., то средний доход будет 30 тыс. Полученный результат отражает наиболее адекватную картину доходов. При увеличении числа наблюдений среднее будет стремиться к истинному значению.

Обсудим закон больших чисел применительно к страхованию. Например, страховой компании необходимо установить размер страхового взноса, который должен уплачивать страхователь, при этом страховая компания должна выплатить при наступлении страхового случая определенную страховую сумму. Рассматривая частоту и убытки страхователя при наступлении страхового случая как величины случайной и обладая известной статистикой таких случаев, можно определить среднее число или средний убыток при наступлении страховых случаев, которые на основании закона больших чисел с большой степенью уверенности можно считать величиной почти неслучайной, тогда на основании этих данных и предполагаемой страховой суммы определяется размер страхового взноса. Без учета закона больших чисел возможны существенные убытки компании (при занижении размера страхового взноса) либо потеря привлекательности страховых услуг (при завышении размера взноса).

5.2. Центральная предельная теорема

Закон больших чисел устанавливает факт приближения среднего большого числа независимых случайных величин к некоторой детерминированной величине. Однако закономерности, возникающие при суммарном воздействии большого числа случайных величин, на этом не ограничиваются. Оказывается, что при совокупности некоторых достаточно общих условий совокупное действие случайных величин приводит к определенному, а именно, нормальному распределению.

Теорема. Центральная предельная теорема (теорема Ляпунова).

Пусть $x_1, \ldots, x_n$ – последовательность случайных независимых величин, каждая из которых имеет математическое ожидание, дисперсию, $\mu_3 = E\,|\,x_i - E_{x_i}\,|^3$ и $\lim\limits_{n\to\infty} \dfrac{\sum\limits_{i=1}^{n} \mu_{3i}}{\left(\sum\limits_{i=1}^{n} \sigma^2{}_i\right)^{3/2}} = 0$ (т. е. удельный вклад каждого слагаемого стремится к 0).

Тогда $\sum\limits_{i=1}^{n} x_i \xrightarrow[n \to \infty]{d} N.$

П р и м е р

Потребление электроэнергии для бытовых нужд за месяц в каждой квартире многоквартирного дома можно представить в виде n различных случайных величин. Если потребление электроэнергии в каждой квартире по своему значению резко не выделяется среди остальных, то на основании Центральной предельной теоремы можно считать, что потребление электроэнергии всего дома, т. е. сумма n независимых случайных величин, будет случайной величиной, имеющей приближенный нормальный закон распределения. Если в доме поместился некоторый вычислительный центр с высокопотребляемой электротехникой, то нарушается условие Центральной предельной теоремы, и тогда распределение не является нормальным.

5.3. Задачи

5.1. Страховая компания заключила 16 тыс. договоров. В среднем страховой случай наступает у одного человека из 10. Пусть S – количество наступивших страховых случаев. Найдите $P(S > 1\,800)$, $P(1\,550 < S < 1\,650)$, $P(S < 2\,000)$.

5.2. В среднем 20 % покупателей супермаркета делают покупку на сумму свыше 500 руб. Какова вероятность того, что из 200 покупателей менее 81 % сделают покупку на сумму не более 500 руб.?

5.3. Портфель страховой компании состоит из 1 тыс. договоров, заключенных 1 января и действующих в течение года. При наступлении страхового случая по каждому из договоров компания обязуется выплатить 30 тыс. руб. Вероятность наступления страхового события по каждому из договоров предполагается равной 0,05 и не зависящей от наступления страховых событий по другим контрактам. Каков должен быть совокупный размер резерва страховой компании для того, чтобы с вероятностью 0,95 она могла бы удовлетворить требования, возникающие по указанным договорам?

5.4. На лекции по теории вероятностей присутствует 200 человек. Вероятность того, что день рождения случайно выбранного студента приходится на определенный день года, составляет $\frac{1}{365}$.

Найти вероятность того, что один человек из присутствующих родился 1 января и два человека родились 8 марта (замечание: применить распределение Пуассона).

5.5. В городе работают 1 тыс. коммерческих банков, из которых 330 допускают нарушения налогового законодательства. Определить число банков, которые должна отобрать для проверки налоговая инспекция, чтобы с вероятностью, не меньшей 0,99, среди них оказался хотя бы один нарушитель законодательства.

6. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

6.1. Выборочный метод математической статистики

Математическая статистика является частью общей прикладной математической дисциплины «Теория вероятностей и математическая статистика», однако задачи, решаемые ею, носят специфический характер. Если теория вероятностей исследует явления, полностью заданные их моделью, то в математической статистике вероятностная модель определена с точностью до неизвестных параметров. Отсутствие сведений о параметрах компенсируется «пробными» испытаниями, на основе которых и восстанавливается недостающая информация.

Задачи математической статистики:

1. Определение способов сбора и группировки статистической информации.

2. Разработка методов анализа статистических данных, интерпретация и формирование выводов.

Изучение закономерностей объектов достаточно большой совокупности методами математической статистики основано на использовании статистических данных для некоторой конечной части рассматриваемых объектов.

Допустим, у нас есть некоторая совокупность однородных объектов и нас интересует некоторый количественный или качественный признак, характеризующий эти объекты, например, размер деталей, в магазине – вес расфасованных продуктов. Данный признак мы будем интерпретировать как случайную величину, значение которой меняется от объекта к объекту. Иногда проводят сплошное обследование – обследуют каждый объект совокупности относительно признака, которым интересуются. Но не всегда это возможно. Обычно из всей совокупности объектов случайным образом отбирают ограниченное число объектов, которые и подвергают изучению.

Генеральная совокупность – это все мыслимые значения (измерения, наблюдения), описывающие поведение исследуемого объекта или явления.

Выборка – ограниченный набор реально наблюдаемых выборочных из генеральной совокупности значений, описывающих исследуемый объект или явление.

Количество этих значений называют объемом выборки.

Понятие генеральной совокупности аналогично понятию случайной величины.

Выборку можно рассматривать как некоторый эмпирический аналог генеральной совокупности.

Пример

Количество зарегистрированных малых предприятий торговли питания в городе Екатеринбурге равно 2 531. Для исследования предприятий по объему товарооборота взято 136.

В данном случае 2 531 – объем генеральной совокупности, а 136 – объем выборки.

Сущность выборочного метода состоит в том, чтобы по некоторой части генеральной совокупности (т. е. по выборке) выносить данные о ее свойствах в целом.

Достоинства выборочного метода:

– позволяет существенно экономить затраты ресурсов;

– является единственно возможным в случае бесконечной генеральной совокупности;

– при тех же затратах ресурсов дает возможность проведения углубленного исследования за счет расширения программы исследования;

– позволяет снизить ошибки регистрации.

Недостатки выборочного метода:

– ошибки исследования, называемые ошибками репрезентативности.

Однако неизбежные ошибки могут быть и заранее оценены с помощью правильной организации выборки и сведены к практически незначительным величинам.

Чтобы по данным выборки иметь возможность судить о генеральной совокупности, она должна быть отобрана случайно.

Выборка называется репрезентативной (представительной), если она достаточно хорошо воспроизводит генеральную совокупность.

Различают несколько видов выборки:

x_i – значение признака (т. е. случайной величины X);

N, n – объемы генеральной совокупности и выборки.

Средние арифметические распределения признака в генеральной и выборочной совокупности называются генеральным и выборочным средним соответственно. Дисперсии – генеральной и выборочной дисперсией.

Важнейшей задачей выборочного метода является оценка параметров (характеристик) генеральной совокупности по данным выборки.

Теоретическую основу применимости выборочного метода составляет закон больших чисел, согласно которому при неограниченном увеличении выборки практически достоверно, что случайные выборочные характеристики как угодно близко приближаются по вероятности к ограниченным параметрам генеральной совокупности.

Оценка неизвестных параметров переменной происходит на основании анализа материала наблюдения. Однако прежде чем приступить к оцениванию, производят предварительную обработку материала наблюдения – составляют вариационный ряд и рассчитывают некоторые описательные статистики этого ряда, которые будут анализироваться дальше.

6.2. Применение математической статистики

Зачем нужна математическая статистика? Рассмотрим несколько областей, в которых она применяется.

• *Маркетинг*. Изучение окупаемости рекламы, исследование рынка, исследование и анализ целевых аудиторий и потребительских предпочтений, выстраивание прогнозов спроса и предложения.

• *Бизнес*. При разработке бизнес-плана потребуется все детально рассчитывать: за сколько вы сможете купить или продать тот или иной товар или услугу. Необходимо построить (смоделировать) несколько сценариев: оптимистичный, средний, пессимистичный. Затем в каждом из сценариев есть разветвления. Таким образом, можно построить некое подобие дерева решений, где можно в каждой ситуации просчитать ожидаемую прибыль к тому или иному месяцу.

• *Банковское дело*. Построение разумной стратегии по выдаче кредитов. Возникает случайная величина: будет возвращен кредит или нет. Чтобы определить, кому выдать кредит, а кому – нет, банк анализирует статистическую информацию. Сюда входит и кредитная история самого человека, и процент вернувших кредит в срок и т. д. Этот анализ проводится методами теории вероятностей и математической статистики.

П р и м е р

Банк выдает кредиты по 1 млн руб. сроком на 1 год. Известно, что в среднем вероятность невозврата кредита равна 1 %. Какую процентную ставку должен установить банк, чтобы в среднем иметь прибыль?

Обозначим ставку P = 100 %. Прибыль – величина случайная.

P	–1 млн
0,99	0,01

$EX = 0{,}99p - 0{,}01 > 0 \Rightarrow p > 1/99 \Rightarrow$ ставка должна быть 100/99, т. е. несколько больше одного процента.

• *Страхование*. Наступление страхового случая или избежание такового – величина случайная. Страховая компания анализирует статистические данные по наступлению страхового случая, в котором они наступили. Таким образом, можно оценить вероятность наступления страхового случая и назначить для него страховой взнос.

П р и м е р

Пусть страховая компания заключает договоры страхования сроком на 1 год на S рублей каждый. Страховой случай происходит с вероятностью p и не происходит с вероятностью $q = 1 - p$. Таким образом, имеем закон распределения случайной величины X – количество страховых случаев у одного страхователя (0 – страховой случай не наступил).

0	1
P	q

Следовательно, $EX = p$; $varX = pq$.

Случайная величина $X = x_1, ..., x_n$ – количество страховых случаев у страхователей имеет математическое ожидание $EX = np$ и $varX = npq$.

В силу ЦПТ: X – нормально распределенная случайная величина.

В среднем страховая компания должна будет выплатить npS страховых возмещений, т. е. если с каждого брать по pS страхового взноса, то у компании будет нулевой баланс. Реальная страховая ставка: $p > p$.

Таким образом, нам нужно исследовать поведение тех или иных объектов или явлений. А оно осуществляется на основе изучения статистических данных – наблюдений и измерений.

6.3. Вариационные ряды и их характеристики

Установление статистических закономерностей, присущих массовым случайным явлениям, основано на изучении статистических данных – сведений о том, какие значения принял в результате наблюдений интересующий нас признак X.

Рассмотрим X – числовую характеристику совокупности объектов.

П р и м е р

Необходимо изучить распределение размеров обуви, проданной в интересующем магазине, с целью обеспечить нужное количество обуви каждого размера. Получены следующие данные о размерах проданной в магазине за сутки обуви (женской): 35, 35, 36, 36, …, 42, 42. Всего продано 100 штук.

Рассмотрение и осмысление данных, представленных в таком виде, практически невозможно из-за обилия числовой информации. Поэтому проводят группировку представленной совокупности чисел.

Различные значения признака X, наблюдавшиеся у объектов, называются **в а р и а н т а м и**, а их количество – **ч а с т о т а м и**.

Сгруппированный ряд представляют в виде таблицы.

П р и м е р

За смену продано 100 пар обуви.

x_i – варианты	36	37	38	39	40	41	42
n_i – частоты	2	6	13	20	25	21	13
m_i – частности	0,02	0,06	0,13	0,2	0,25	0,21	0,13

где $\sum_{i=1}^{N} m_i = 1;\ \sum_{i=1}^{N} n_i = N;\ m_i = \frac{n_i}{N},\ N = 100.$

Частоты показывают, сколько раз встречаются наблюдения, у которых значение признака X равно данной варианте.

В а р и а ц и о н н ы м р я д о м называется ранжированный в порядке возрастания или убывания ряд вариант с соответствующими весами (частотами или частностями).

Вариационный ряд можно определить для дискретных и непрерывных величин X. В последнем случае проводят интервальную группировку ряда. Число интервалов рекомендуется брать по формуле Стерджеса: $n = 1 + 3{,}322 \cdot \lg N$;

$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \cdot \lg N}$, где $x_{\max} - x_{\min}$ – разность между наибольшим и наименьшим значением признака. За начало первого интервала рекомендуют брать величину $x_{нач} = x_{\min} - h/2$. Частота показывает число членов совокупности, у которых признак X принимает значения в границах интервалов.

П р и м е р

Урожайность, ц/га	(8–12)	(12–16)	(16–20)	(20–25)	(25–40)
Количество хозяйств	10	14	30	26	20
Частности	0,1	0,14	0,3	0,26	0,2

$N = 100.$

В этом случае n_i – плотность распределения, а m_i – относительная плотность распределения вариационного ряда.

Полученный вариационный ряд позволяет выявить закономерности изменчивости признака, закономерности распределения обуви по размеру проданных пар и участков по урожайности, что сделать по первичным, несгруппированным данным оказалось затруднительно.

Наряду с понятием частот и относительных частот для описания вариационного ряда используются накопленные частоты и накопленные относительные частоты.

Графическое представление вариационных рядов

Представление вариационного ряда в виде таблицы не всегда удобно. Поэтому используют различные способы графического представления вариационных рядов.

Полигон частот – ломаная, соединяющая точки (x_i, n_i).

Полигон относительных частот – ломаная, соединяющая точки (x_i, m_i).

В случае непрерывного признака X целесообразно строить различные гистограммы.

Гистограмма частот – содержит столбики с основанием – интервалом, высотой – плотностью вариационного ряда, деленной на величину соответствующего интервала. Площадь столбиков – число наблюдений N.

Гистограмма относительных частот – содержит столбики с основанием – интервалом, высотой – относительной плотностью вариационного ряда, деленной на величину соответствующего интервала. Площадь столбиков – 1.

Эмпирической функцией распределения называется функция $F_N(x)$, выражающая для каждого x долю значений параметра общего объема N, для которых рассматриваемый признак меньше x.

$$F_N(x) = \frac{n(x)}{N},$$

где $n(x) = \sum_{x_0 < x} n_i$.

Вариационный ряд является статистическим аналогом (реализацией) распределения признака (случайной величины X). В этом смысле полигон или гистограмма аналогичен кривой распределения, а эмпирическая функция распределения – функции распределения случайной величины X.

Вариационный ряд содержит достаточно полную информацию об изменчивости признака X. Однако на практике часто оказывается, что этого недостаточно и необходимо найти некоторые сводные характеристики вариационных рядов: средних, центральной тенденции и изменчивости (показателей вариации), расчет которых представляет собой следующий этап после группировки и обработки данных наблюдений.

Средние величины характеризуют значения признака, вокруг которого концентрируются наблюдения. Наиболее распространенной среди средних величин является средняя арифметическая.

Средняя арифметическая вариационного ряда и ее свойства

$$\overline{x} = \frac{\sum_{i=1}^{n} n_i x_i}{N} = \sum_{i=1}^{n} m_i x_i,$$

где x_i – i-я варианта, если признак X – дискретный, и середина i-го интервала, если X – непрерывный признак.

Свойства средней арифметической:

1. $\overline{x \pm c} = \overline{x} \pm c.$
2. $\overline{cx} = c\overline{x}.$
3. $\dfrac{\sum_{i=1}^{n} n_i (x_i - \overline{x})}{N} = \sum_{i=1}^{n} m_i (x_i - \overline{x}) = 0.$
4. $\overline{(x \pm y)} = \overline{x} \pm \overline{y}.$

5. Если признак X разбит на группы, то имеются групповая и общая средние.

Середина интервала	Q_1	...	Q_j	...	Q_m
x_1	s_{11}		s_{1j}		s_{1m}
…					
x_i	s_{i1}		s_{ij}		s_{im}
…					
x_n	s_{n1}		s_{nj}		s_{nm}
Итого	N_1		N_j		N_m
	$\overline{x}_1$		$\overline{x}_j$		$\overline{x}_m$

где s_{ij} – частота появления i-го наблюдения в j-й группе;

$N_j = \sum_{i=1}^{n} s_{ij}$ – число наблюдений в j-й группе $\left(\sum_{j=1}^{m} N_j = N\right)$;

$\overline{x}_j = \dfrac{\sum_{i=1}^{n} s_{ij} x_i}{N_j}$ – средняя j-й группы;

$\overline{x} = \dfrac{\sum_{j=1}^{m} N_j \overline{x}_j}{N}$ – общая средняя.

Кроме средней арифметической также вычисляют моду и медиану.

М е д и а н о й вариационного ряда называется значение признака, приходящегося на середину ранжированного ряда наблюдений.

На медиану не влияют изменения крайних членов вариационного ряда. Медиана как показатель среднего предпочтительнее средней арифметической для ряда, у которого крайние варианты по сравнению с большими оказались чрезмерно большими или малыми.

М о д о й вариационного ряда называется варианта, которой соответствует наибольшая частота (модальный интервал).

Особенность моды как показателя среднего заключается в том, что она не изменяется при изменении крайних членов ряда, т. е. обладает устойчивостью к вариации признака.

Кроме средних величин для вариационного ряда рассчитывают еще показатели вариации.

Выборочная дисперсия и ее свойства

Дисперсией вариационного ряда называется величина

$$D_{\text{В}}(x) = \sigma^2(X) = \frac{\sum_{i=1}^{n} n_i (x_i - \overline{x})^2}{N} = \sum_{i=1}^{n} m_i (x_i - \overline{x})^2 = \frac{\sum_{i=1}^{n} n_i x_i^2}{N} - \overline{x}^2,$$

где у нас по-прежнему x_i – i-я варианта, если признак X – дискретный, и середина i-го интервала, если X – непрерывный признак.

$\sigma_{\text{В}}(X) = \sqrt{D_{\text{В}}(X)}$ – среднеквадратичное отклонение вариационного ряда.

Свойства выборочной дисперсии:

1. $D_{\text{В}}(CX) = C^2 D_{\text{В}}(X)$.
2. $D_{\text{В}}(X \pm C) = D_{\text{В}}(X)$.

Понятие групповой, межгрупповой, внутригрупповой и общей дисперсии

Допустим, что все значения признака X разбиты на ряд групп:

Середина интервала	Q_1	…	Q_j	…	Q_m
x_1	s_{11}		s_{1j}		s_{1m}
…					
x_i	s_{i1}		s_{ij}		s_{im}
…					
x_n	s_{n1}		s_{nj}		s_{nm}
Итого	N_1		N_j		N_m
	$\overline{x}_1$		$\overline{x}_j$		$\overline{x}_m$
	D_1		D_j		D_m

где s_{ij} – частота появления i-го наблюдения в j-й группе;

$N_j = \sum_{i=1}^{n} s_{ij}$ – число наблюдений в j-й группе $\left(\sum_{j=1}^{m} N_j = N \right)$;

$\bar{x}_j = \frac{\sum_{i=1}^{n} s_{ij} x_i}{N_j}$ – средняя j-й группы;

$D_j = \frac{\sum s_{ij} (x_i - \bar{x}_j)^2}{N_j}$ – групповая дисперсия, характеризует рассеяние признака X внутри j-й группы;

$D_{\text{ВГ}} = \frac{\sum_{j=1}^{m} N_j D_j}{N}$ – среднеарифметическое групповых дисперсий, взвешенных по объему группы – внутригрупповая дисперсия;

$D_{\text{МГ}} = \frac{\sum_{j=1}^{m} N_j (\bar{x}_j - \bar{x})^2}{N}$ – межгрупповая дисперсия, характеризует рассеяние групповых средних вокруг общей средней;

$D_{\text{общ}} = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} s_{ij} (x_i - \bar{x})^2}{N}$ – общая дисперсия, характеризует рассеяние всей совокупности вокруг общей средней.

Утверждение. $D_{\text{общ}} = D_{\text{ВГ}} + D_{\text{МГ}}$.

Для вариационного ряда рассчитывают начальные и центральные моменты, частными случаями которых являются средняя арифметическая и дисперсия. В число таких показателей входят асимметрия и эксцесс.

6.4. Оценивание распределения случайных величин

Пусть имеется некоторая случайная величина (одномерная или многомерная), распределение которой неизвестно или известно с точностью до некоторого параметра θ. Имеется некоторая реализация этой случайной величины $X_1, X_2, \ldots, X_N$ (выборка). На основании имеющейся выборки мы должны оценить распределение случайной величины.

Для нахождения оценки характеристики используют два подхода:

1. Если распределение случайной величины нам известно с точностью до некоторого параметра $f_X(x, \theta)$, то оценив по выборке значение этого параметра, мы получим оценку распределения. $\widehat{f_X}(x,\theta) = f_X(x,\hat{\theta})$. Такой подход называется параметрическим.

Примеры

1) $X \sim N(m; \sigma)$;

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-m)^2}{2\sigma^2}}, \text{ тогда } \theta = \begin{pmatrix} m \\ q \end{pmatrix}.$$

2) $X \sim \exp(\lambda)$, $\theta = \begin{pmatrix} \theta_1 \\ \ldots \\ \theta_k \end{pmatrix}$ – неизвестен.

3) $X \sim bi(p)$;

$p(x=k) = C_n^k p^k q^{n-k} \Rightarrow \theta = p.$

4) $X \sim U(a; b)$;

$$f_x(x) = \frac{1}{b-a}, \quad \theta = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Таким образом, имеем $f_x(x, \theta)$, где θ – неизвестный параметр, и мы этот неизвестный параметр оцениваем, т. е. получаем оценку (приближенное значение) этого параметра, которое называем θ. Тогда имеем известную $f_X(x, \theta)$.

2. Непараметрический. Вид функции распределения нам неизвестен. Тогда составим алгоритм расчета значения функции в каждой точке.

Итак, работаем в рамках первого подхода.

$f_X(x, \theta)$ знаем, но параметр не знаем.

$$\theta = \begin{pmatrix} \theta_1 \\ \ldots \\ \theta_k \end{pmatrix}$$ – вектор неизвестных параметров (из k неизвестных).

Как оценить эти k параметров? Существуют различные виды статистического оценивания параметров.

Виды статистического оценивания параметров:

1. Точечное.

$\theta = \varphi(x_1, x_2, \ldots, x_n)$.

Вообще говоря, $\theta \neq \theta$, но мы надеемся, что $\theta \approx \theta$.

2. Интервальное.

Строим интервал, который с заданной вероятностью γ покрывает истинное значение оцениваемого параметра.

Величину γ называют уровнем надежности (заданная вероятность накрытия интервалом), т. е. $p(\theta \in (\theta_1; \theta_2)) = \gamma$.

Точечное статистическое оценивание неизвестного параметра

1. Метод математических ожиданий/метод аналогий.

Выборочная средняя $\overline{x} = \dfrac{\sum x_i}{n}$ нам известна.

$\theta = Ex$ – неизвестная нам величина.

$\hat{\theta} = \dfrac{\sum x_i}{n}$ – оценка математического ожидания (нам известна).

$\hat{\theta} = \dfrac{\sum x^2{}_i}{n}$ – известная нам оценка, тогда как $\theta = EX^2$ нам неизвестна.

2. Метод моментов (метод Пирсона).

Пусть $\theta = \begin{pmatrix} \theta_1 \\ \ldots \\ \theta_k \end{pmatrix}$ – вектор неизвестных параметров.

$\nu_n = EX^m$ – начальный момент m порядка.

$$\begin{bmatrix} Ex = g_1(\theta) \\ Ex^2 = g_2(\theta) \\ \ldots \\ Ex^m = g_m(\theta) \end{bmatrix}$$ – система условий на момент распределения.

$$\Rightarrow \begin{bmatrix} \overline{x} = \dfrac{\sum x_i}{n} = g_1(\theta) \\ \dfrac{\sum x_i^2}{n} = g_2(\theta) \\ \ldots \\ \dfrac{\sum x_i^m}{n} = g_m(\theta) \end{bmatrix}$$

П р и м е р ы

1) Пусть $X \sim U(0; \theta)$.

Методом моментов найдем оценку параметров θ.

$$E_x = \frac{\theta}{2}; \ \overline{x} = \frac{\theta}{2} => \hat{\theta}_{\text{ММ}} = 2\overline{x}.$$

Заметим, что сколько неизвестных параметров, столько моментов мы используем.

2) Пусть $X \sim \exp(\lambda), \theta = \lambda$.

Тогда $E_x = \dfrac{1}{\lambda}; \ \overline{x} = \dfrac{1}{\lambda} => \hat{\lambda} = \dfrac{1}{\overline{x}}$.

3) Пусть $X \sim N\ (m; \sigma^2); \ \theta = \dfrac{m}{\sigma^2}$.

$E_x = m, E_{x^2} = m^2 + \sigma^2$, так как $var(X) = E(X - EX)^2 = EX^2 - (EX)^2$;

$$\sigma^2 = E_{x^2} - (E_x)^2 => E_{x^2} = \sigma^2 + (E_x)^2 = \sigma^2 + m^2;$$

$$\overline{x} = m, \ \overline{x} = m;$$

$$x_i - C_i(\theta) \forall i = \overline{1, n};$$

$$\frac{\sum x_i^2}{n} = m^2 + \sigma^2;$$

$$\widehat{\sigma^2} = \frac{\sum x_i^2}{n} - \overline{x}^2;$$

$$d_X^2 = \frac{\sum (x_i - \overline{x})^2}{n} = \frac{\sum x_i^2}{n} - \overline{x}.$$

3. Метод минимальных расстояний.

x_i сопоставим с $C_i(\theta)$ – функция цели.

Рассмотрим $x_i - C_i(\theta) \forall i = \overline{1, n};$

$C(\theta) = \sum_{i=1}^{n} |x_i - C_i(\theta)|$ – метод наименьших модулей;

$C(\theta) = \sum_{i=1}^{n} (x_i - C_i(\theta)^2)$ – метод наименьших квадратов (МНК).

$$\hat{\theta}_{MD} = \arg\min C(\theta).$$

П р и м е р

Неизвестный параметр – математическое ожидание, т. е. $\theta = EX$.

$$C_i(\theta) = EX = \theta;$$

$$\text{МНК: } \sum_{i=1}^{n} (x_i - \theta)^2 \to \min.$$

Находим необходимое условие существования экстремума:

$$C'_\theta = \frac{\partial C}{\partial \theta} = -2\sum (x_i - \theta) = \theta;$$

$$\hat{\theta} \approx \theta.$$

4. Метод максимального правдоподобия (Maximum Likelihood).

Этот метод среди всех возможных параметров выбирает тот, для которого вероятность получить данную выборку максимальна.

Основу метода максимального правдоподобия (ML) составляет функция правдоподобия.

$$L(\theta) = f_{X_1, X_2, ..., X_n)}(x_1, x_2, ..., x_n, \theta) = \prod f_{x_i}(x_i, \theta);$$

$$L(\theta) \to \max;$$

$$\theta_{ML} = \arg\max L(\theta).$$

П р и м е р

X	-1	0	1
p	θ	$1{,}2 - 2\theta$	$\theta - 0{,}2$

Необходимо найти θ.

Выборка $x_1 = 0$; $x_2 = 0$ независима.

$$L(\theta) = p((x_1 = 0)\,(x_2 = 1)) = p(x_1 = 0) \cdot p(x_2 = 1) =$$
$$= (1{,}2 - 2\theta)\,(\theta - 0{,}2) \to \max.$$

Необходимые условия экстремума таковы:

$$L'_\theta = 1{,}2\theta - 2\theta^2 + 0{,}4\theta - 0{,}24 = 0;$$
$$\Rightarrow 1{,}6 - 4\theta = 0 \Rightarrow \theta_{ML} = 0{,}4;$$
$$L''_\theta = -4 < 0 \Rightarrow \theta_{ML} = 0{,}4.$$

Таким образом, получили оценку методом максимального правдоподобия, а теперь применим метод моментов (ММ):

$$EX = -\theta + \theta - 0{,}2 = -0{,}2;$$

$$EX^2 = \theta + \theta - 0{,}2 = 2\theta - 0{,}2;$$

$$\frac{\sum x_i^2}{n} = 2\theta - 0,2;$$

$$\hat{\theta}_{MM} = \frac{\sum x_i^2}{2n} + 0,1;$$

$$n = 2 \Rightarrow \hat{\theta}_{MM} = \frac{0^2 + 1^2}{4} + 0,1 = 0,25 + 0,1 = 0,35.$$

Оценки неизвестного параметра методом максимального правдоподобия и методом моментов оказались разные.

Нахождение оценки θ_n упрощается, если максимизировать не саму функцию L, а $\ln L$, поскольку максимум обеих функций достигается при одном и том же значении θ.

$$L(\theta) = f_{X_1, X_2, \ldots, X_N}\left(x_1, \ldots, x_N, \theta\right) = \prod_{i=1}^{N} f\left(x_i, \theta\right);$$

$$L(\theta) \to \max;$$

$$\theta_{ML} = \arg\max L(\theta).$$

Необходимое условие: $L'(\theta) = 0$.

Рассмотрим $l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f_x(x_i, \theta);$

$$\theta_{ML} = \arg\max l(\theta),$$

$l'_\theta = \dfrac{\partial l(\theta)}{\partial \theta} = 0$, а затем надо отобрать то решение, которое обращает функцию $\ln L$ в max, т. е. воспользоваться достаточным условием существования экстремума (матрица Гессе).

Пример

Найти оценку методом максимального правдоподобия.

$$x \sim N(m, \sigma^2);\ \theta = \begin{pmatrix} m \\ \sigma^2 \end{pmatrix};\ f_x(x) = \frac{1}{\sqrt{\sigma^2}\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}};$$

$$\ln f_x(x) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln(2\pi) - \frac{(x-m)^2}{2\sigma^2};$$

$$l(\theta) = \sum_{i=1}^{n} \ln f_{x_i}(x_i, \theta) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{\sum (x_i - m)^2}{2\sigma^2};$$

$$l'_m = 2\frac{\sum (x_i - m)}{2\sigma^2} = 0;$$

$$l'_{\sigma^2} = -\frac{m}{2\sigma^2} + \frac{\sum (x_i - m)^2}{2\sigma^4} = 0;$$

$$\frac{m}{2\sigma^2} = \frac{\sum (x_i - \overline{x})^2}{2\sigma^4} \Rightarrow \frac{n\sigma^2}{2\sigma^4} = \frac{\sum (x_i - \overline{x})^2}{2\sigma^4};$$

$$\sigma^2 : \frac{n\sigma^2}{2\sigma^4} = \frac{\sum (x_i - \overline{x})^2}{2\sigma^4};$$

$$\widehat{\sigma^2}_{ML} = \frac{\sum (x_i - \overline{x})^2}{n};$$

$$
\begin{bmatrix}
\sum (x_i - m) = 0 \\
\sum x_i - \sum m = 0 \\
\sum x_i - nm = 0 \\
\widehat{m^2}_{ML} = \dfrac{\sum x_i}{n}
\end{bmatrix};
$$

$$
l''_{mm} = -\frac{n}{\sigma^2} < 0;\ \ l''_{\sigma^2\sigma^2} = \frac{n}{2\sigma^4} - \frac{2\sum (x_i - m)^2}{2\sigma^6};
$$

$$
l''_{m\sigma^2} = -\frac{\sum (x_i - m)}{\sigma^4} = 0,\ \text{так как}\ \sum (x_i - \widehat{m}) = \sum (x_i - \overline{x}) =
$$

$$
= \sum x_i - \sum \overline{x} = \sum x_i - n\overline{x} = \sum x_i - \sum x_i = 0.
$$

Проверим знакоопределенность матрицы Гессе:

$$
l''_{\sigma^2\sigma^2} = \frac{n}{2\left[\dfrac{\sum (x_i - \overline{x})^2}{n}\right]} - \frac{\sum (x_i - \overline{x})^2}{\left[\dfrac{\sum (x_i - \overline{x})^2}{n}\right]^3} =
$$

$$
= \frac{n^3}{2(\sum (x_i - \overline{x})^2)^2} - \frac{n^3}{(\sum (x_i - \overline{x})^2)^2} = -\frac{n^3}{2(\sum (x_i - \overline{x})^2)^2} < 0;
$$

$$
H = \begin{pmatrix} -\dfrac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\dfrac{n^3}{(\sum (x_i - \overline{x})^2)^2} \end{pmatrix} \text{ – матрица Гессе.}
$$

Значит, действительно, найденная точка является точкой максимума и полученная оценка является оценкой максимального правдоподобия.

6.5. Свойства статистических оценок

Свойства статистических оценок различают на больших и конечных выборках.

Свойства оценок на больших выборках ($n \to \infty$):

1. Состоятельность.
2. Асимптотическая нормальность.
3. Асимптотическая несмещенность.
4. Асимптотическая эффективность.

Свойства оценок на конечных выборках:

1. Асимптотическая несмещенность.
2. Асимптотическая эффективность.

Имеем θ, θ, где θ ≠ θ. В общем случае надеемся, что θ ≈ θ, θ – число, θ – случайная величина.

1. Состоятельность.

Оценка θ параметра θ называется *состоятельной*, если она удовлетворяет закону больших чисел, т. е. сходится по вероятности к оцениваемому параметру:

$$\lim_{x\to\infty} p(|\hat{\theta}-\theta|<E)=1 \text{ или } p\lim_{x\to\infty}\hat{\theta}=\theta.$$

То есть, если оценка состоятельна, то практически достоверно, что при достаточно большом n θ = θ.

Отметим, что оценки ММ и ML состоятельны, а МНК – нет.

2. Несмещенность.

Оценка θ параметра θ называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру, т. е. $E\theta = \theta$.

Если $E\theta \neq \theta$, то смещенная оценка систематически завышает/занижает истинное значение оцениваемого параметра.

$Bias(\theta) = \theta - E\theta$ – величина смещения.

Пусть $x_i \sim iid(m, \sigma^2)$, $EX = m$.

$$E\overline{x} = E\left(\frac{\sum x_i}{n}\right) = \frac{E\sum x}{n} = \frac{\sum Ex_i}{n} = \frac{\sum m}{n} = \frac{nm}{n} = m = EX$$ – выборочное среднее есть несмещенная оценка.

$$Var(\hat{\theta}_1) > var(\hat{\theta}_2).$$

Следовательно:

$d_x^2 = \dfrac{\sum (x_i - \overline{x})^2}{n}$ – смещенная, но асимптотически несмещенная оценка;

$S_x^2 = \dfrac{\sum (x_i - \overline{x})^2}{n-1}$ – несмещенная оценка, которая в данном случае называется исправленной выборочной дисперсией.

Отметим, что из двух несмещенных оценок выбирают ту, дисперсия у которой меньше.

Рассмотренные нами методы оценки не гарантируют несмещенности.

3. Эффективность.

Эффективной в классе несмещенных оценок называется оценка, обладающая минимальной дисперсией среди всех несмещенных оценок.

$$\hat{\theta}_1, \hat{\theta}_2;\ E_{\hat{\theta}_1} = \theta;\ E_{\hat{\theta}_2} = \theta;$$

$var(\theta_1) > var(\theta_1) \Rightarrow$ оценка θ_2 более эффективна, чем θ_1.

Обращаем внимание читателей, что по дисперсии сравниваются только несмещенные оценки.

$$\overline{x} = \widehat{Ex};\ \ var\overline{x} = \frac{varX}{n};$$

$$var\overline{x} = \frac{1}{n^2}\sum_1^n var(x_i) = \frac{1}{n^2}\sum_1^n \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} = \frac{varX}{n}.$$

Информацией Фишера о неизвестном параметре

$$\theta = \begin{pmatrix} \theta_1 \\ \cdot \\ \cdot \\ \theta_k \end{pmatrix} \text{ называется } I(\theta) = -E\left(\frac{\partial^2 \ln f(x,\theta)}{\partial\theta\partial\theta'}\right).$$

Если $\theta = (\theta)$, то $I(\theta) = -E\left(\dfrac{\partial^2 \ln f(x,\theta)}{(\partial\theta)^2}\right)$.

Теорема Рао – Фреше – Крамера. Неравенство Рао – Крамера. Пусть выполняется некое условие регулярности, а именно:
1. $I(\theta)$ существует и положительно.
2. Область определения случайной величины X не зависит от θ.
Тогда матрица $[n \cdot I(\theta)]^{-1}$ дает нижнюю границу дисперсий:
– состоятельных;
– асимптотически нормальных оценок θ.
В частности, если θ – скаляр, то $var(\hat{\theta}) \geq \dfrac{1}{n \cdot I(\theta)}$.

Если дисперсия несмещенной оценки достигает границы, указанной неравенством Рао – Крамера, значит, она эффективна, т. е. обладает наименьшей дисперсией в классе несмещенных оценок.

П р и м е р

$f_x(x,\theta) = \dfrac{1}{\theta} e^{-\frac{1}{\theta}x}, \; x \geq 0$. Требуется найти оценку методом максимального правдоподобия; проверить состоятельность, несмещенность и эффективность.

Решение. Составим $l(\theta) = \prod\limits_{i=1}^{n} \dfrac{1}{\theta} e^{-\frac{1}{\theta}x_i} \to \max$.

Тогда:

$$\ln f_x(x,\theta) = -\ln\theta - \frac{1}{\theta}x_i;$$

$$\sum \ln f_x(x,\theta) = \sum(-\ln\theta - \frac{1}{\theta}x_i);$$

$$l(\theta) = -n\ln\theta - \frac{1}{\theta}\sum x_i \to \max.$$

Необходимое условие существования экстремума:

$$l'_\theta = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum x_i = 0.$$

Проверим:

$$\frac{-n\theta}{\theta^2} + \frac{1}{\theta^2}\sum x_i \cdot \theta^2 => -n\theta + \sum x_i = 0 =>$$

$$=> \hat{\theta}_{ML} = \frac{\sum x_i}{n} = \overline{x}.$$

При этом нужно еще доказать, что это точка максимума, т. е. проверить достаточность условия существования экстремума, но мы упустим этот момент, оставляя его для самостоятельной работы читателя.

Таким образом, оценка неизвестного параметра найдена. Теперь проверим свойства найденной оценки.

Состоятельность: $p\lim_{x\to\infty}\hat{\theta} = p\lim_{x\to\infty}\overline{x} = E_x = \theta;\ x_i \sim iid.$

$$E_x = \int_0^{+\infty} x^2 \frac{1}{\theta} e^{\frac{-1}{\theta}x} dx = \theta \Rightarrow \text{ оценка состоятельна.}$$

$$VarX = \int_0^{+\infty} x^2 \frac{1}{\theta} e^{\frac{-1}{\theta}x} dx - (E_x)^2 = \theta^2.$$

Несмещенность:

$E\hat{\theta} = E\,\overline{x} = E\left[\frac{\sum x_i}{n}\right] = \frac{\sum Ex_i}{n} = \frac{\sum Ex}{n} = \frac{nEx}{n} = Ex = \theta$ => оценка несмещенная.

Эффективность:

$$var(\hat{\theta}_{ML}) = var\left[\frac{\sum x_i}{n}\right] = \frac{1}{n^2} var(\sum x_i) = \frac{1}{n^2}\sum var(x_i) =$$

$$= \frac{1}{n^2}\theta^2 = \frac{n\theta^2}{n^2} = \theta^2 / n => var(\hat{\theta}_{ML}) = \frac{\theta^2}{n};$$

$$MSE(\hat{\theta}) = E(\theta - E\hat{\theta})^2 = var\hat{\theta} + (\theta - E\hat{\theta})^2 = var\hat{\theta} + bias^2(\hat{\theta}).$$

Далее посмотрим, совпадет ли дисперсия оценки с нижней границей, полученной из неравенства Рао – Фреше – Крамера.

Ищем нижнюю границу из теоремы Рао – Фреше – Крамера, для этого находим вторую производную, а затем – информацию Фишера $I(\theta)$.

$$\frac{\partial \ln f_x(x,\theta)}{\partial\theta} = -\frac{1}{\theta} + \frac{1}{\theta^2}x_i - \text{это производная от } \ln\left(\frac{1}{\theta}e^{-\frac{1}{\theta}x}\right) = -\ln\theta - \frac{1}{\theta}x_i;$$

$$\frac{\partial^2 \ln f_x(x,\theta)}{\partial\theta^2} = \frac{1}{\theta^2} + \frac{2}{\theta^3}x_i;$$

$$\ln\left(\frac{1}{\theta}e^{-\frac{1}{\theta}x}\right) = -\ln\theta - \frac{1}{\theta}x_i;$$

$$I(\theta) = -E\left(\frac{1}{\theta^2} + \frac{2}{\theta^3}x_i\right) = -\left(\frac{1}{\theta^2} + \frac{2}{\theta^3}E_x\right) =$$

$$= -\left(\frac{1}{\theta^2} + \frac{2}{\theta^2}\right) = -\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2} = I(\theta).$$

Из неравенства Рао – Крамера:

$$var(\hat{\theta}) \geq \frac{1}{n \cdot I(\theta)}.$$

А у нас:

$\dfrac{1}{n \cdot I(\theta)} = \dfrac{1}{n \cdot \dfrac{1}{\theta^2}} = \dfrac{\theta^2}{n} \Rightarrow var(\hat{\theta})$ => дисперсия оценки совпадает с нижней границей Рао – Фреше – Крамера => наша оценка эффективна.

Теорема. $\overline{x}$ является состоятельной, несмещенной и эффективной оценкой математического ожидания случайной величины.

Утверждение.

1. Метод математических ожиданий: оценка, полученная этим методом, состоятельная и асимптотически нормальная.
2. Метод моментов: оценка состоятельная, асимптотически нормальная, но не гарантирует несмещенность.
3. Метод минимальных расстояний: выполнений свойств не гарантируется, все свойства необходимо проверять для каждой оценки индивидуально.
4. Метод максимального правдоподобия: оценка – состоятельная, асимптотически нормальная, асимптотически несмещенная, асимптотически эффективная (в пределе этот метод самый лучший).

$$\sqrt{n}(\hat{\theta}_{ML} - \theta) \xrightarrow[n\to\infty]{d} N(0; I^{-1}(\theta)).$$

Как сравнивать смещенные и несмещенные оценки?

$$MSE(\hat{\theta}) = E(\theta - E\hat{\theta})^2 = var\hat{\theta} + (\theta - E\hat{\theta})^2 = var\hat{\theta} + bias^2(\hat{\theta}).$$

Та оценка, у которой MSE (Mean Square Error) меньше, – точнее!

С теорией статистического оценивания параметров тесно связана проверка статистических гипотез. Она используется всякий раз, когда необходимо сделать обоснованный вывод о преимуществах того или иного способа инвестиций, об эффективности метода управления, уровне доходности ценных бумаг, о значимости математической модели.

6.6. Общая схема проверки статистических гипотез

Статистическая гипотеза – гипотеза о виде распределения случайной величины или о параметрах распределения.

Имеем: X – случайная величина; $f_x(x;\ \theta)$; θ – неизвестный параметр.

Выдвинутую гипотезу называют нулевой и обозначают H_0.

Альтернативная гипотеза – это гипотеза, противоречащая нулевой. Ее обозначают H_1 или H_a.

Нулевую гипотезу формулируют так: «нет различий», «=».

Альтернативную так: «различия», «≠».

Для проверки гипотезы используется выборка. Поскольку выборка действительность отражает не полностью, то при принятии решений возможны ошибки.

H_0	Принимаем H_0	Не принимаем H_0
Верная	ОК	Ошибка I рода (α)
Неверная	Ошибка II рода (β)	ОК

$$p(\text{I рода}) = p(H_a \mid \text{H}_0 \text{ верна}) = \alpha \text{ – уровень значимости.}$$

Пример

α – вероятность вынесения судом обвинительного приговора при том, что подсудимый невиновен.

$$p(\text{II рода}) = p(H_0 \mid H_a \text{ верна}) = \beta.$$

П р и м е р

β – вероятность вынесения оправдательного приговора при том, что подсудимый виновен.

$$\mu = 1 - \beta \text{ – мощность критерия.}$$

При этом, чем больше α, тем меньше β, и наоборот.

Поэтому вероятность ошибки I рода выбираем сами.

Установим $\alpha = 0{,}05$; $\beta \to \min$.

Для проверки гипотез используется специальным образом подобранная случайная величина, называемая с т а т и с т и ч е с к и м к р и т е р и е м.

Закон распределения критерия при справедливости нулевой гипотезы известен точно или приблизительно (асимптотически).

Наиболее часто используемые случайные величины:

1. Стандартное нормальное распределение.
2. Распределение Стьюдента.
3. Распределение Фишера.
4. Распределение χ^2.

Все возможные значения критерия разбиваются на две непересекающиеся области: на критическую область и область принятия гипотезы.

К р и т и ч е с к а я о б л а с т ь – это область таких значений критерия, попадание в которую критерием маловероятно ($p = \alpha$) при справедливости нулевой гипотезы.

О б л а с т ь п р и н я т и я г и п о т е з ы – это область, попадание в которую критерием вероятно при справедливости нулевой гипотезы.

Если наблюдаемое значение критерия попало в критическую область, значит, гипотеза неверна и отвергаем ее.

Если наблюдаемое значение критерия не попало в критическую область (попало в область принятия гипотезы), мы говорим, что нет основания отвергнуть нулевую гипотезу.

Разделяют критическую область и область принятия гипотезы (критические точки).

$$k_{\text{набл}} = k \text{ (выборка),}$$

$$w_{кр} \text{ – критическая область; } k \text{ – критерий.}$$

$p(k \in w_{кр} \mid H_0) = \alpha$ – вероятность попасть в критическую область при выполнении нулевой гипотезы.

Вывод: $k \in w_{кр}$ – H_0 отвергаем;

$k \notin w_{кр}$ – нет оснований отвергнуть H_0 (так как гипотеза – это не установленный факт, а лишь как достаточно правдоподобное, не противоречащее выборке утверждение).

6.7. Проверка гипотезы о виде распределения случайной величины

Имеем некоторую случайную величину X. Предполагаем, что она имеет распределение A. То есть формулируем гипотезы:

$$H_0: X \sim A;$$
$$H_a: X \nsim A.$$

Для проверки гипотезы H_0 выбирают некоторую случайную величину u, характеризующую степень расхождения теоретического и эмпирического распределения. Проверку проводят с помощью критерия согласия Пирсона.

Условия применения критерия Пирсона:

1. Объем выборки большой.
2. Данные должны быть сгруппированы.
3. Частота каждого интервала (каждой варианты) должна быть достаточно большой.

Частоты интервала – эмпирические частоты. Теоретическая частота – это частота, которую бы имел в среднем данный интервал (варианта), если бы случайная величина X имела предполагаемое распределение.

Обозначим: n_i – эмпирическая частота, n_i' – теоретическая частота.

Тогда $n_i' = p_i n$, где:

$$p_i = p(X = x_i \mid H_0) \text{, если } X \text{ – дискретная;}$$

$$p_i = p(x_i \in (x_{i-1};\, x_i) \,|\, H_0), \text{ если } X \text{ – непрерывная.}$$

Иногда распределение A параметрическое, и чтобы посчитать p_i, нам нужно оценить неизвестные параметры:

$$\text{если } n_i \approx n_i' \Rightarrow H_0;$$

$$\text{если } n_i \not\approx n_i' \Rightarrow \text{отвергаем } H_0;$$

$$u = \sum \frac{(n_i - n_i')^2}{n_i'} \sim \chi^2(m - r - 1),$$

где u называют статистикой Пирсона.

При справедливости H_0 u имеет распределение χ^2 с количеством степеней свободы $m - r - 1$, где r – число параметров распределения A, которые мы оцениваем для того, чтобы посчитать p_i; m – число интервалов.

Таким образом, проверка сводится к следующему:

1. Считаем $u_{\text{набл}}$ – наблюдаемое значение критерия.

2. Для выбранного уровня значимости по таблице χ^2 ищем критическое значение $u_{\text{кр}}$.

3. Делаем вывод.

Рассматриваем $w_{\text{кр}} = \{u > u_{\text{кр}}\}$ (на рис. 24) при $p(u > u_{\text{кр}} \,|\, H_0) = \alpha$ – уровень значимости, тогда $u_{\text{кр}} = \chi^2_{\text{кр}}\,(\alpha, m - r - 1)$, $m - r - 1$ – количество степеней свободы.

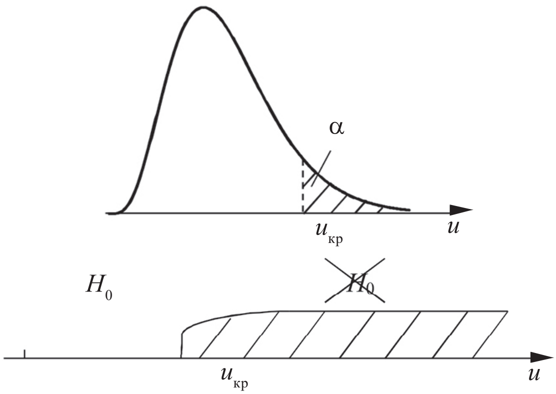

Рис. 24

Вывод: $u_{\text{набл}} > u_{\text{кр}} \Rightarrow$ отвергаем H_0, $u_{\text{набл}} < u_{\text{кр}}$, следовательно, нет оснований отвергнуть H_0.

Пример

	n_i	n'_i
Орел	40	50
Решка	60	50

$$u_{\text{набл}} = \frac{(40-50)^2}{50} + \frac{(60-50)^2}{50} = \frac{2 \cdot 100}{50} = 4;$$

$u_{\text{кр}} = 3{,}84$ ($u_{\text{кр}}$ $(0{,}05; 2-0-1) = u_{\text{кр}}$ $(0{,}05; 1) = 3{,}84$.

Вывод: $u_{\text{набл}} = 4 > u_{\text{кр}} = 3{,}84 \Rightarrow$ отвергаем H_0.

Особое внимание стоит уделить тестированию нормальности:

$$H_0: x \sim N;$$
$$H_a: x \nsim N.$$

Критерий Харке – Бера применяется для проверки нормальности распределения. Он проверяет не совпадение всего распределения с нормальным, а только моментов распределения. Как известно, для нормального распределения асимметрия равна нулю, а эксцесс – трем. Критерий проверяет незначимость отличия выборочных моментов распределения от теоретических для нормального распределения.

$$H_0: As = 0, K = 3;$$
$$H_a: As \neq 0 \text{ или } K \neq 3.$$

Для проверки гипотезы используется следующий статистический критерий:

$$JB = \frac{N}{6}\left(As^2 + \frac{1}{4}(K-3)^2\right) \sim \chi^2_2.$$

В случае справедливости нулевой гипотезы данный критерий имеет асимптотическое распределение χ^2 с двумя степенями свободы. Мы отвергаем нулевую гипотезу, если наблюдаемое значение критерия больше критической точки распределения χ^2 с двумя степенями свободы для выбранного уровня значимости: $JB_{набл} > \chi^2(\alpha, 2) \Rightarrow$ отвергаем H_0.

Критерий Харке – Бера обладает малой мощностью, т. е. он часто распределения, не являющиеся нормальными, относит к нормальному распределению.

6.8. Проверка нормальности из графического анализа гистограмм

1. Проверка гипотезы о равенстве математического ожидания нормально распределенной случайной величины некоторому числу.

$$X \sim N(m_x, \sigma_x).$$

H_0: $m_x = m_0$, m_0 – некоторое известное число.

1.1. σ_x – известно.

Имеем x_i: $x_1, x_2, \ldots, x_n \to \bar{X}$.

Вспомним факты относительно выборочного среднего X:

1) $\bar{x} = \dfrac{\sum x_i}{n}$, $\bar{X} \sim N$;

2) $E_{\bar{x}} = E_x = m_0$, если верно H_0;

3) $var\bar{X} = \dfrac{varX}{n} = \dfrac{\sigma_x^2}{n} \Rightarrow \sigma_{\bar{x}} = \dfrac{\sigma_x}{\sqrt{n}}$.

Тогда при нулевой гипотезе имеем:

$$\bar{x} \sim N\left(m_0, \frac{\sigma_x^2}{n}\right);$$

$$\bar{x} \sim N\left(m_0, \frac{\sigma_x}{n}\right);$$

$$\frac{\bar{x} - m_0}{\sigma_x / \sqrt{n}} \sim N(0,1).$$

Рассмотрим:

$$z = \frac{\left(\overline{x} - m_0\right)\sqrt{n}}{\sigma_x} \sim N(0,1).$$

Данную величину называют z-статистикой.

а) H_a: $m_x \neq m_0$.
Рассматриваем $w_{кр} = \{|z| > z_{кр}\}$ при $P(|z| > z_{кр} \mid H_0) = \alpha$ (рис. 25).

Ищем $z_{кр}$: $\Phi(z_{кр}) = \frac{1}{2} - \frac{\alpha}{2}$.

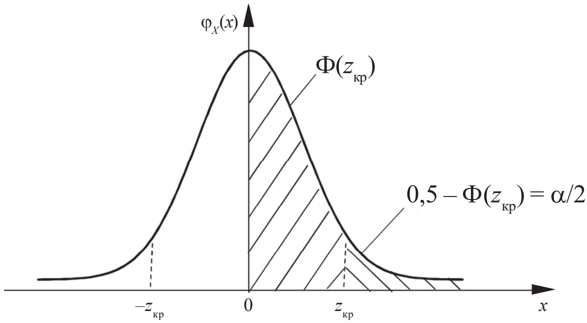

Рис. 25

Вывод: если $|z_{набл}| > z_{кр} \Rightarrow H_0$ отвергаем; если $|z_{набл}| < z_{кр} \Rightarrow$ нет оснований отвергнуть H_0.

П р и м е р
Проверить, что средний рост равен 170 см.

$$H_0\colon m_x = 170;$$

$$H_a\colon m_x \neq 170;$$

$$\alpha = 0{,}05 \Rightarrow z_{кр} = 1{,}96;$$

$$\sigma_x = 6 \text{ см}, n = 100, \ \overline{x} = 169 \text{ см};$$

$$z = \frac{(169 - 170)\sqrt{100}}{6} = -1{,}67.$$

Вывод:
$|z_{набл}| < z_{кр} \Rightarrow$ нет оснований отвергнуть H_0.

Здесь можно задать закономерный вопрос: какие m_0 не отвергнутся нашей проверкой?

$$|z_{\text{набл}}| < z_{\text{кр}};$$

$$z = \frac{(\overline{x} - m_0)\sqrt{n}}{\sigma_x} < z_{\text{кр}}$$ – решим неравенство относительно m_0;

$$\left|\overline{x} - m_0\right|\sqrt{n} < z_{\text{кр}}\sigma_x;$$

$$\left|\overline{x} - m_0\right| < \frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};$$

$$-\overline{x} - \frac{z_{\text{кр}}\sigma_x}{\sqrt{n}} \le -m_0 \le -\overline{x} + \frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};$$

$$\overline{x} - \frac{z_{\text{кр}}\sigma_x}{\sqrt{n}} \le m_0 \le \overline{x} + \frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};$$

$$\left(\overline{x} - \frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};\ \overline{x} + \frac{z_{\text{кр}}\sigma_x}{\sqrt{n}}\right);$$

$$CI(\gamma, m_x) = \left(\overline{x} - \frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};\ \overline{x} + \frac{z_{\text{кр}}\sigma_x}{\sqrt{n}}\right),$$

где $\gamma = 1 - \alpha$ – уровень надежности.

CI – Confidence Interval (д о в е р и т е л ь н ы й   и н т е р в а л) – интервал, который с заданной надежностью γ накрывает истинное значение математического ожидания случайной величины X.

б) H_a: $m_x > m_0$.

$$w_{\text{кр}} = \{z < z_{\text{кр}}\};$$

$$P(z < z_{\text{кр}} \mid H_0) = \alpha;$$

$$z_{\text{кр}}:\ \Phi(z_{\text{кр}}) = \frac{1}{2} - \alpha;$$

$$\alpha = 0{,}05;\ z_{\text{кр}}^{0{,}05} = -1{,}65.$$

Вывод:
если $z_{\text{набл}} > z_{\text{кр}} \Rightarrow H_0$ отвергаем;
если $z_{\text{набл}} < z_{\text{кр}} \Rightarrow$ нет оснований отвергнуть H_0.

в) H_a: $m_x < m_0$.

$$w_{\text{кр}} = \{z < z_{\text{кр}}\};$$

$$P(z < z_{\text{кр}} \mid H_0) = \alpha;$$

$$z_{\text{кр}}:\ \Phi(z_{\text{кр}}) = \frac{1}{2} - \alpha;$$

$$\alpha = 0{,}05;\ z_{\text{кр}}^{0,05} = -1{,}65.$$

Вывод:
если $z_{\text{набл}} < z_{\text{кр}} \Rightarrow H_0$ отвергаем;
если $z_{\text{набл}} > z_{\text{кр}} \Rightarrow$ нет оснований отвергнуть H_0.

1.2. σ_x – неизвестно.

$$x_i\!: x_1, x_2, \ldots, x_n \to \bar{X}, s_x.$$

Отметим: факты справедливы, только если $x \sim N$:

1) $\bar{x} \sim N\left(m_0, \dfrac{\sigma_x^2}{n}\right)$;

2) $\dfrac{s_x^2(n-1)}{\sigma_x^2} \sim \chi^2(n-1)$;

3) $\bar{x}$ и $\bar{s}_X$ независимы.

Рассмотрим распределение Стьюдента:

$$t_{\text{набл}} = \frac{N(0,1)}{\sqrt{\dfrac{\chi^2(m)}{m}}} = \frac{\dfrac{(\bar{x}-m_0)\sqrt{n}}{\sigma_x}}{\sqrt{\dfrac{s_x^2(n-1)}{\sigma_x^2(n-1)}}} = \frac{(\bar{x}-m_0)\sqrt{n}}{s_x} \sim t(n-1);$$

$$t_{\text{набл}} = \frac{(\bar{x}-m_0)\sqrt{n}}{s_x} \sim t(n-1).$$

а) H_a: $m_x \neq m_0$.

$$w_{кр} = \{|t| > t_{кр}\}.$$

Критические значение распределения Стьюдента $t_{кр}^{дв}(\alpha, n-1)$ находим из таблиц распределения критических точек распределения Стьюдента для выбранного уровня значимости α, числа степеней свободы $n-1$ и двусторонней критической области. Также данное значение можно получить посредством программы Microsoft Excel, используя предопределенную функцию СТЬЮДРАСПОБР.

После того, как найдем наблюдаемое и критическое значения критерия, то делаем вывод:

$$\text{если } |t_{набл}| > t_{кр} \Rightarrow H_0 \text{ отвергаем;}$$
$$\text{если } |t_{набл}| < t_{кр} \Rightarrow \text{нет оснований отвергнуть } H_0.$$

Пример

Рассмотрим вес людей, больных гипертонией. Можно ли утверждать, что средний вес людей, больных гипертонией, равен 90 кг, если известно: $n = 50$, $\overline{x} = 88$ кг, $s_x = 6$ кг?

Решение:

$$H_0:\ m_x = m_0 = 90;$$
$$H_a:\ m_x \neq 90;$$
$$t = \frac{(88 - 90)\sqrt{50}}{6} = -2,36;$$
$$t_{кр}^{дв} = (0,05;\ 49) = 2,01;$$
$$|t_н| = 2,36 > t_{кр} = 2,01 \Rightarrow \text{отвергаем } H_0.$$

Делаем вывод, что нельзя утверждать, что средний рост больных гипертонией равен 90 кг.

Найдем m_0, которые не отвергнутся нашей проверкой:

$$|t| < t_{кр};$$
$$\frac{|\overline{x} - m_0|\sqrt{n}}{s_x} < t_{кр};$$

$$\left|\overline{x} - m_0\right| < \frac{t_{кр} s_x}{\sqrt{n}};$$

$$-\frac{t_{кр} s_x}{\sqrt{n}} - \overline{x} < -m_0 < \frac{t_{кр} s_x}{\sqrt{n}} - \overline{x};$$

$$\overline{x} - \frac{t_{кр} s_x}{\sqrt{n}} < m_0 < \frac{t_{кр} s_x}{\sqrt{n}} + \overline{x};$$

$$CI(\gamma, m_x) = \left(\overline{x} - \frac{t_{кр} s_x}{\sqrt{n}};\ \overline{x} + \frac{t_{кр} s_x}{\sqrt{n}} \right).$$

Таким образом, мы построили доверительный интервал для m_0 нормально распределенной случайной величины с уровнем надежности γ в случае, когда дисперсия неизвестна.

б) H_a: $m_x > m_0$.

Тогда рассматриваем $w_{кр} = \{t > t_{кр}\}$ при $P(t > t_{кр} \mid H_0) = \alpha$ (рис. 26).

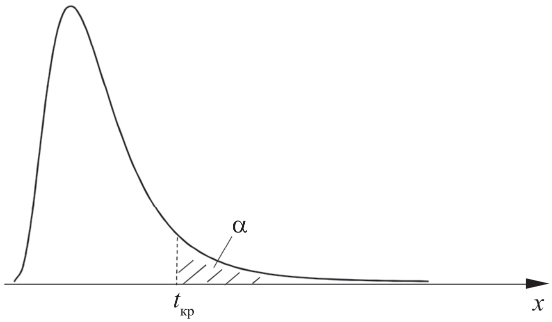

Рис. 26

$t_{кр}^{одн}(\alpha; n-1)$ находим в таблицах критических точек распределения Стьюдента с уровнем значимости α, числом степеней свободы $n - 1$ и односторонней критической областью или с помощью функции СТЬЮДЕНТ.ОБР.ПХ в программе Microsoft Excel, используя равенство:

$$t_{кр}^{одн}(\alpha; n-1) = t_{кр}^{дв}(2\alpha; n-1).$$

После нахождения критического и наблюдаемого значений критерия делаем вывод:

$$\text{если } t_{\text{набл}} > t_{\text{кр}} \Rightarrow H_0 \text{ отвергаем;}$$

$$\text{если } t_{\text{набл}} < t_{\text{кр}} \Rightarrow \text{нет оснований отвергнуть } H_0.$$

в) H_a: $m_x < m_0$.

Рассматриваем $w_{\text{кр}} = \{t < t_{\text{кр}}\}$ при $P(t < t_{\text{кр}} \mid H_0) = \alpha$, тогда $t_{\text{кр}} = -t_{\text{кр}}^{\text{одн}}\ (\alpha; n-1)$.

Вывод:

$$\text{если } t_{\text{набл}} < t_{\text{кр}} \Rightarrow H_0 \text{ отвергаем;}$$

$$\text{если } t_{\text{набл}} > t_{\text{кр}} \Rightarrow \text{нет оснований отвергнуть } H_0.$$

2. Проверка гипотезы о равенстве математических ожиданий двух нормально распределенных случайных величин.

Пусть $X \sim N\,(m_x; \sigma_x^2)$; $Y \sim N\,(m_y; \sigma_y^2)$.

Предполагаем, что $\sigma_x = \sigma_y$. Тогда:

$$H_0: m_x = m_y;$$

$$H_a: m_x \neq m_y.$$

Имеем:

X: $x_1, x_2, \ldots, x_{N_x} \rightarrow$ считаем $\overline{x}, s_x^2$ (исправленная дисперсия).

Y: $y_1, y_2, \ldots, y_{N_y} \rightarrow$ считаем $\overline{y}, s_y^2$ (исправленная дисперсия).

Тогда:

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{s_x^2(n_x - 1) + s_y^2(n_y - 1)}} \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}} \sim t(n_x + n_y - 2);$$

$$w_{\text{кр}} = \{|t| > t_{\text{кр}}\};$$

$$P(|t| > t_{\text{кр}} \mid H_0) = \alpha;$$

$t_{\text{кр}}^{\text{дв}}(\alpha; n_x + n_y - 2)$ ищем в таблице критических точек распределения Стьюдента со степенью свободы $n_x + n_y - 2$ и уровнем значимости α.

И далее делаем вывод:

если $|t_{\text{набл}}| > t_{\text{кр}} \Rightarrow H_0$ отвергаем;

если $|t_{\text{набл}}| < t_{\text{кр}} \Rightarrow$ нет оснований отвергнуть H_0.

3. Проверка гипотезы о равенстве дисперсий двух нормально распределенных совокупностей.

Пусть $X \sim N\,(m_x;\, \sigma_x)$; $Y \sim N\,(m_y;\, \sigma_y)$;

$$H_0\colon \sigma_x^2 = \sigma_y^2;$$

$$H_a\colon \sigma^2_{\text{больш}} = \sigma^2_{\text{меньш}},$$

где $\sigma^2_{\text{больш}}$ – это X или Y в зависимости от того, чья исправленная дисперсия больше.

$$F_{\text{набл}} = \frac{s^2_{\text{больш}}}{s^2_{\text{меньш}}} \sim F(n_{\text{больш}} - 1, n_{\text{меньш}} - 1);$$

$n_{\text{больш}} - 1$ – степень свободы числителя;

$n_{\text{меньш}} - 1$ – степень свободы знаменателя;

$n_{\text{больш}}$ – объем выборки, по которой вычислена большая из исправленных дисперсий.

Рассмотрим $w_{\text{кр}} = \{F > F_{\text{кр}}\}$ при $P(F > F_{\text{кр}} \mid H_0) = \alpha$ (рис. 27), тогда $F_{\text{кр}}(n_{\text{больш}} - 1,\, n_{\text{меньш}} - 1)$;

$F_{\text{кр}}$ находим в таблицах критических точек распределения Фишера для выбранного уровня значимости α и числа степеней свободы числителя: $n_{\text{больш}} - 1$ и знаменателя: $n_{\text{меньш}} - 1$ или с помощью FРАСПОБР в программе Microsoft Excel.

Вывод:

если $F_{\text{набл}} < F_{\text{кр}} \Rightarrow H_0$ отвергаем;

если $F_{\text{набл}} > F_{\text{кр}} \Rightarrow$ нет оснований отвергнуть H_0.

П р и м е р

Можно ли достоверно утверждать, что средний рост мужчин и женщин одинаковый?

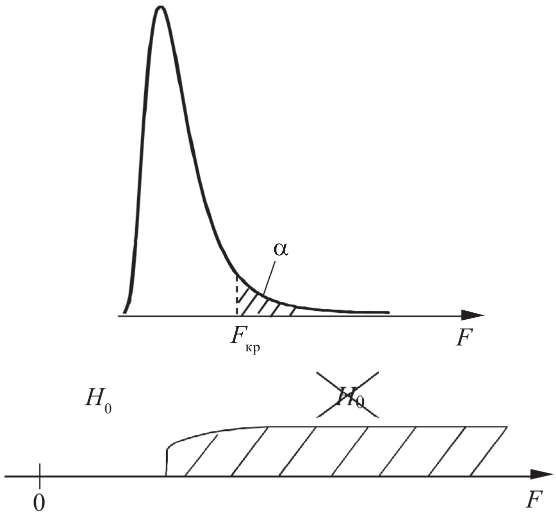

Рис. 27

$$\overline{x} = 168 \text{ см}, \; s_x^2 = 36 \text{ см}^2, \; n_x = 100;$$
$$\overline{y} = 171 \text{ см}, \; s_y^2 = 40 \text{ см}^2, \; n_y = 50.$$

Решение. Для начала проверим равенство дисперсий по Фишеру, чтобы можно было проверить равенство математических ожиданий по Стьюденту.

$$H_0\colon \sigma_x^2 = \sigma_y^2;$$
$$H_a\colon \sigma_y^2 > \sigma_x^2, \text{ так как } s_y^2 > s_x^2;$$
$$F_{набл} = \frac{40}{36} = 1{,}11; \;\; F_{кр}(0{,}05;\, 99;\, 49) = 1{,}47;$$
$$F_{набл} = 1{,}11 < F_{кр} = 1{,}47 \Rightarrow$$

нет оснований отвергнуть H_0, а значит, можем сделать проверку гипотезы с помощью критерия Стьюдента. Если бы отвергли, то тогда использовали бы критерий Стьюдента для неравных дисперсий, но в данном пособии таких критериев мы рассматривать не будем.

$$H_0\colon m_x = m_y;$$
$$H_a\colon m_x \neq m_y.$$

$$t_{\text{набл}} = \frac{168 - 171}{\sqrt{36 \cdot 99 - 40 \cdot 49}} \sqrt{\frac{100 \cdot 50 \cdot (100 + 50 - 2)}{100 + 50}} = -2,84;$$

$$t_{\text{кр}}^{\text{дв}}\ (0,05; 148) = 1,99.$$

Вывод:

$|t_{\text{набл}}| = 2,84 > t_{\text{кр}} = 1,99 \Rightarrow H_0$ отвергаем, т. е. нельзя сказать, что средний рост одинаковый.

6.9. Задачи

6.1. Дан ряд распределения случайной величины:

X	-1	0	2
P	θ	$1,1 - 2\theta$	$\theta - 0,1$

Имеется независимая выборка $X_1 = 0$, $X_2 = -1$. Найти оценку неизвестного параметра методом максимального правдоподобия, методом моментов, если возможно, проверить состоятельность и несмещенность. Найти дисперсию оценки метода моментов.

6.2. Асимметричная монетка подбрасывается N раз, k раз выпадает орел. Оценить вероятность выпадения орла, если возможно, проверить состоятельность и несмещенность.

6.3. Методом моментов построить оценку параметра следующего распределения: $f(x, \theta) = \dfrac{2x}{\theta^2}$, $x \in [0, \theta]$. Проверить состоятельность, несмещенность, найти дисперсию оценки. Можно ли здесь применить неравенство Рао – Фреше – Крамера для проверки эффективности оценки?

6.4. Методом моментов и методом максимального правдоподобия построить оценку параметра следующего распределения: $f(x, \theta) = \theta x^{\theta-1}$, $x > 0$. Проверить состоятельность, несмещенность, найти дисперсию оценки. Можно ли здесь применить неравенство Рао – Фреше – Крамера для проверки эффективности оценки?

6.5. Пусть p_0, p_1, и p_2 – доли семей студентов, у которых 0, 1 и 2 ребенка соответственно (в нашей популяции нет семей студентов с тремя и более детьми). Известно, что $p_1 = 3p_2$. У четырех случайно опрошенных семей оказалось следующее количество детей: 1, 2, 0, 1. Найти оценки максимального правдоподобия и метода моментов величин p_0, p_1 и p_2.

6.6. Закон распределения случайной величины задан таблицей:

X	-1	0	1
P	θ	2θ	$1-3\theta$

1) Имеется три наблюдения случайной величины: $X_1 = 0$, $X_2 = 1$, $X_3 = 0$. Найти ML- и MM-оценки неизвестного параметра.

2) Решить эту задачу для произвольной случайной выборки $X_1, \ldots, X_n$.

3) Для метода моментов проверить состоятельность и несмещенность, найти дисперсию оценки.

6.7. Построить там, где возможно, оценки ML и MM неизвестных параметров по случайной выборке $X_1, \ldots, X_n$ и везде, где можно, проверить состоятельность, несмещенность, найти информацию Фишера о неизвестном параметре и дисперсию оценок:

а) неизвестного параметра p геометрического распределения;

б) неизвестного параметра p биномиального распределения;

в) $f(x, \theta) = \dfrac{2x}{\theta^2}$, при $x \in (0, \theta)$;

г) $f(x, \theta) = \dfrac{\theta(\ln^{\theta-1} x)}{x}$, при $x \in (1, e)$.

6.8. По случайной выборке $X_1, \ldots, X_n$ из нормального распределения $N(\mu_1, \mu_2 - \mu_1^2)$ методом моментов оценить параметры μ_1, μ_2. Дать определения несмещенности и состоятельности и проверить выполнение этих свойств для оценки μ_1.

6.9. По случайной выборке $X_1, \ldots, X_n$ из нормального распределения $N(\theta, 1)$ методом максимального правдоподобия оценить параметр θ. Будет ли найденная оценка эффективной (ответ обосновать)?

6.10. Имеется квадрат со стороной a, нам необходимо оценить площадь этого квадрата. Для этого мы проводим n независимых замеров стороны квадрата и получаем выборку длин сторон X_1, …, X_n, $X_i \sim N\ (a;\ \sigma^2)$. Предлагаются две оценки площади квадрата:

а) $\dfrac{\sum X_i^2}{N}$ – средний квадрат;

б) $\bar{X}^2$ – квадрат среднего.

Какая оценка предпочтительнее?

6.11. Имеются две оценки дисперсии случайной величины σ^2, построенные по независимой выборке: выборочная дисперсия d_X^2 и исправленная выборочная дисперсия s_X^2:

а) показать, что d_X^2 смещена, а s_X^2 нет;

б) но $var(d_X^2) < var(s_X^2)$;

в) сравнить $MSE(d_X^2)$ и $MSE(s_X^2)$.

Являются ли эти оценки состоятельными?

6.12. Пусть X – случайная величина с неизвестным математическим ожиданием m и дисперсией σ^2. Для оценки параметра m независимо были сделаны две независимые выборки. По одной было посчитано выборочное среднее $\bar{X}_1$, объем выборки составил N_1, по второй – $\bar{X}_2$, объем выборки N_2. На основании этих двух оценок предложены два способа получить оценку m:

$$\widehat{m}_1 = \frac{\bar{X}_2 + \bar{X}_2}{2};$$

$$\widehat{m}_2 = \frac{N_1\bar{X}_2 + N_2\bar{X}_2}{N_1 + N_2}.$$

Какой способ лучше?

6.13. У 20 студентов, сдававших экзамен по теории вероятностей, сердце билось в среднем со скоростью 96 ударов в минуту. Стандартное отклонение выборки было равно 5 ударам в минуту. Найти 95 %-ный доверительный интервал для генерального среднего.

6.14. Средний доход фирмы в день составлял 1 020 единиц. После реорганизации выборочный средний доход в день за 30 рабочих дней составил 1 070 единиц с выборочным средним квадратическим отклонением (исправленным) 90 единиц. Можно ли утверждать (на уровне значимости 5 %), что реорганизация привела к увеличению среднего дохода?

6.15. В книжном магазине проведено исследование продаж фантастического романа писателя Бурьяненко «Танцы в пустоте». В течение 25 рабочих дней роман продавался ежедневно в среднем по 64 экземпляра со средним квадратическим отклонением 10 экземпляров. Можно ли утверждать на уровне значимости 5 %, что этот роман расходится хуже, чем предыдущий роман автора «Звездная жуть», если тот расходился в среднем по 70 экземпляров в день?

6.16. Инвестиция 1 рассчитана на 12 лет, дисперсия ежегодных прибылей составляет 20 %. Инвестиция 2 рассчитана на 10 лет, дисперсия ежегодных прибылей составляет 30 %. Предполагается, что распределение ежегодных прибылей на инвестиции подчиняется нормальному закону распределения. Равны ли риски 1 и 2? Доверительная вероятность 95 %.

6.17. Производительность двух моторных заводов, выпускающих дизельные двигатели, характеризуется следующими данными:

1-й завод	72	84	69	74	82	67	75	86	68	61
2-й завод	55	65	73	66	58	71	77	68	68	59

Можно ли считать одинаковыми производительности дизельных двигателей на обоих заводах при уровне значимости 0,05?

6.18. Проведено исследование розничного товарооборота продовольственных магазинов в двух районах области (по 50 магазинов в каждом). Априори известны средние значения товарооборота 78,9 и 78,68 тыс. руб. Полученные в результате оценки среднеквадратических отклонений в первом и во втором районах области соответственно равны 7,22 и 7,79 тыс. руб. Можно ли считать, что разброс розничного товарооборота магазинов в районах неодинаков при уровне значимости 0,05?

6.19. Менеджер банка утверждает, что размер кредита, выдаваемого клиентам, составляет в среднем 4 800$ со стандартным отклонением 800$. В выборке из 25 клиентов, бравших кредиты, средний размер кредита оказался равен 4 235$. При $\alpha = 0{,}10$ есть ли достаточные основания опровергать утверждение менеджера?

6.20. В таблице представлены результаты наблюдений: X – число сделок на фондовой бирже за квартал, $N = 400$ (инвесторов).

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	146	97	73	34	23	10	6	3	4	2	2

Найти: выборочное среднее, выборочную дисперсию, выборочное среднеквадратическое отклонение, исправленную выборочную дисперсию, исправленное выборочное среднеквадратическое отклонение, моду, медиану, асимметрию, эксцесс.

6.21. В таблице представлены результаты наблюдений: X – месячный доход жителя региона (в руб.), $N = 1\ 000$ (жителей).

x_i	Менее 500	500–1 000	1 000–1 500	1 500–2 000	2 000–2 500	Свыше 2 500
n_i	58	96	239	328	147	132

Найти: выборочное среднее, выборочную дисперсию, выборочное среднеквадратическое отклонение, исправленную выборочную дисперсию, исправленное выборочное среднеквадратическое отклонение, моду, медиану, асимметрию, эксцесс.

6.22. Имеется монетка. Мы предполагаем, что монетка «жульническая» и орел выпадает в три раза чаще, чем решка. Для проверки этой гипотезы предложена следующая процедура: бросаем монетку 3 раза и считаем, что монетка жульническая, если орел выпал 3 раза. Найти вероятность ошибок I и II рода.

6.23. В таблице представлены данные о годовых доходах и расходах на личное потребление (в долларах) для 10 семей. Найти выборочную ковариацию.

Годовой доход	Расходы на личное потребление
2 508	2 406
2 572	2 464
2 408	2 336
2 522	2 281
2 700	2 641
2 531	2 385
2 390	2 297
2 595	2 416
2 524	2 460
2 685	2 448

6.24. Имеется равномерное распределение с неизвестным параметром k:

$$f(x)=\begin{cases}\dfrac{1}{k}, x\in[0;k];\\ 0, x\notin[0;k].\end{cases}$$

Нужно проверить нулевую гипотезу H_0: $k = 1$ при альтернативной гипотезе H_a: $k = 2$. Для проверки имеется лишь одно наблюдение x_1. Предложены два способа: а) H_0 отвергается при $x_1 \geq \dfrac{1}{2}$; б) H_0 отвергается при $1 \leq x_1 \leq 2$. Найти вероятность ошибок I и II рода.

6.25. Из генеральной совокупности извлечена выборка объема $n = 12$:

Варианта x_i	–0,5	–0,4	–0,2	0	0,2	0,6	0,8	1	1,2	1,5
Частота n_i	1	2	1	1	1	1	1	1	2	1

Оценить с надежностью 0,95 математическое ожидание m нормально распределенного признака генеральной совокупности с помощью доверительного интервала. Решить задачу для надежности 0,9 и 0,99.

6.26. За последние 5 лет годовой рост цены актива *A* составлял в среднем 20 % со средним квадратическим отклонением (исправленным) 5 %. Построить доверительный интервал с вероятностью 95 % для цены актива в конце следующего года, если в начале года она равна 100 ден. ед.

6.27. Из генеральной совокупности извлечена выборка объема $n = 10$:

Варианта x_i	–2	1	2	3	4	5
Частота n_i	2	1	2	2	2	1

Оценить с надежностью 0,95 математическое ожидание m нормально распределенного признака генеральной совокупности по выборочному среднему при помощи доверительного интервала.

6.28. За последние 9 лет годовой рост цены актива *A* составлял в среднем 22 % со средним квадратическим отклонением (исправленным) 6 %. Построить доверительный интервал с вероятностью 90 % для средней цены актива в конце следующего года, если в начале она была равна 200 ден. ед.

6.29. Для отрасли, включающей 1 200 фирм, составлена случайная выборка из 19 фирм. По выборке оказалось, что в фирме в среднем работают $\overline{X} = 77{,}5$ человека при среднем квадратичном отклонении $S_x = 25$ человек. Пользуясь 95 %-ным доверительным интервалом, оценить среднее число работающих в фирме по всей отрасли и общее число работающих в отрасли. Предполагается, что количество работников фирмы имеет нормальное распределение.

6.30. Бухгалтер компании решил предпринять выборочную проверку и выбрал 18 из 1 200 компонент, продававшихся в прошлом месяце. Стоимость отобранных компонент 82; 30; 98; 116; 80; 150; 200; 88; 70; 90; 160; 100; 86; 76; 90; 140; 76; 68 (ден. ед.). Найти оценку средней стоимости всех компонент и построить для нее доверительный интервал с надежностью 0,95.

ПРИЛОЖЕНИЕ

Таблица 1

Критические значения *t*-критерия Стьюдента при различных уровнях значимости α

Число степеней свободы k	α			Число степеней свободы k	α		
	0,05	0,01	0,01		0,05	0,01	0,01
1	12,71	63,66	64,60	18	2,10	2,88	3,92
2	4,30	9,92	31,60	19	2,09	2,86	3,88
3	3,18	5,84	12,92	20	2,09	2,85	3,85
4	2,78	4,60	8,61	21	2,08	2,83	3,82
5	2,57	4,03	6,87	22	2,07	2,82	3,79
6	2,45	3,71	5,96	23	2,07	2,81	3,77
7	2,37	3,50	5,41	24	2,06	2,80	3,75
8	2,31	3,36	5,04	25	2,06	2,79	3,73
9	2,26	3,25	4,78	26	2,06	2,78	3,71
10	2,23	3,17	4,59	27	2,05	2,77	3,69
11	2,20	3,11	4,44	28	2,05	2,76	3,67
12	2,18	3,05	4,32	29	2,05	2,76	3,66
13	2,16	3,01	4,22	30	2,04	2,75	3,65
14	2,14	2,98	4,14	40	2,02	2,70	3,55
15	2,13	2,95	4,07	60	2,00	2,66	3,46
16	2,12	2,92	4,02	120	1,98	2,62	3,37
17	2,11	2,90	3,97	∞	1,96	2,58	3,29

Таблица 2

Критические значения критерия χ_2^2 при различных уровнях значимости α и степенью свободы *k*

α \ k	.995	.990	.950	.900	.500	.100	.050	.025	.010	.005
1	0.00004	0.00016	0.00393	0.01579	0.45494	2.70554	3.84146	5.02389	6.63490	7.87944
2	0.01003	0.02010	0.10259	0.21072	1.38629	4.60517	5.99146	7.37776	9.21034	10.59663
3	0.07172	0.11483	0.35185	0.58437	2.36597	6.25139	7.81473	9.34840	11.34487	12.83816
4	0.20699	0.29711	0.71072	1.06362	3.35669	7.77944	9.48773	11.14329	13.27670	14.86026
5	0.41174	0.55430	1.14548	1.61031	4.35146	9.23636	11.07050	12.83250	15.08627	16.74960
6	0.67573	0.87209	1.63538	2.20413	5.34812	10.64464	12.59159	14.44938	16.81189	18.54758
7	0.98926	1.23904	2.16735	2.83311	6.34581	12.01704	14.06714	16.01276	18.47531	20.27774
8	1.34441	1.64650	2.73264	3.48954	7.34412	13.36157	15.50731	17.53455	20.09024	21.95495
9	1.73493	2.08790	3.32511	4.16816	8.34283	14.68366	16.91898	19.02277	21.66599	23.58935
10	2.15586	2.55821	3.94030	4.86518	9.34182	15.98718	18.30704	20.48318	23.20925	25.18818
11	2.60322	3.05348	4.57481	5.57778	10.34100	17.27501	19.67514	21.92005	24.72497	26.75685
12	3.07382	3.57057	5.22603	6.30380	11.34032	18.54935	21.02607	23.33666	26.21697	28.29952
13	3.56503	4.10692	5.89186	7.04150	12.33976	19.81193	22.36203	24.73560	27.68825	29.81947
14	4.07467	4.66043	6.57063	7.78953	13.33927	21.06414	23.68479	26.11895	29.14124	31.31935
15	4.60092	5.22935	7.26094	8.54676	14.33886	22.30713	24.99579	27.48839	30.57791	32.80132

Окончание табл. 2

k \ α	.995	.990	.950	.900	.500	.100	.050	.025	.010	.005
16	5.14221	5.81221	7.96165	9.31224	15.33850	23.54183	26.29623	28.84535	31.99993	34.26719
17	5.69722	6.40776	8.67176	10.08519	16.33818	24.76904	27.58711	30.19101	33.40866	35.71847
18	6.26480	7.01491	9.39046	10.86494	17.33790	25.98942	28.86930	31.52638	34.80531	37.15645
19	6.84397	7.63273	10.11701	11.65091	18.33765	27.20357	30.14353	32.85233	36.19087	38.58226
20	7.43384	8.26040	10.85081	12.44261	19.33743	28.41198	31.41043	34.16961	37.56623	39.99685
21	8.03365	8.89720	11.59131	13.23960	20.33723	29.61509	32.67057	35.47888	38.93217	41.40106
22	8.64272	9.54249	12.33801	14.04149	21.33704	30.81328	33.92444	36.78071	40.28936	42.79565
23	9.26042	10.19572	13.09051	14.84796	22.33688	32.00690	35.17246	38.07563	41.63840	44.18128
24	9.88623	10.85636	13.84843	15.65868	23.33673	33.19624	36.41503	39.36408	42.97982	45.55851
25	10.51965	11.52398	14.61141	16.47341	24.33659	34.38159	37.65248	40.64647	44.31410	46.92789
26	11.16024	12.19815	15.37916	17.29188	25.33646	35.56317	38.88514	41.92317	45.64168	48.28988
27	11.80759	12.87850	16.15140	18.11390	26.33634	36.74122	40.11327	43.19451	46.96294	49.64492
28	12.46134	13.56471	16.92788	18.93924	27.33623	37.91592	41.33714	44.46079	48.27824	50.99338
29	13.12115	14.25645	17.70837	19.76774	28.33613	39.08747	42.55697	45.72229	49.58788	52.33562
30	13.78672	14.95346	18.49266	20.59923	29.33603	40.25602	43.77297	46.97924	50.89218	53.67196

Таблица 3

Таблица значений функции Лапласа

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319

Окончание табл. 3

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Критические значения критерия Фишера при уровне

k_2 \ k_1	1	2	3	4	5	6	7	8	9
1	161,4	199,50	215,70	224,583	230,161	233,986	236,768	238,882	240,543
2	18,51	19,000	19,164	19,2468	19,2964	19,3295	19,3532	19,3710	19,3848
3	10,12	9,5521	9,2766	9,1172	9,0135	8,9406	8,8867	8,8452	8,8123
4	7,708	6,9443	6,5914	6,3882	6,2561	6,1631	6,0942	6,0410	5,9988
5	6,607	5,7861	5,4095	5,1922	5,0503	4,9503	4,8759	4,8183	4,7725
6	5,987	5,1433	4,7571	4,5337	4,3874	4,2839	4,2067	4,1468	4,0990
7	5,591	4,7374	4,3468	4,1203	3,9715	3,8660	3,7870	3,7257	3,6767
8	5,317	4,4590	4,0662	3,8379	3,6875	3,5806	3,5005	3,4381	3,3881
9	5,117	4,2565	3,8625	3,6331	3,4817	3,3738	3,2927	3,2296	3,1789
10	4,964	4,1028	3,7083	3,4780	3,3258	3,2172	3,1355	3,0717	3,0204
11	4,844	3,9823	3,5874	3,3567	3,2039	3,0946	3,0123	2,9480	2,8962
12	4,747	3,8853	3,4903	3,2592	3,1059	2,9961	2,9134	2,8486	2,7964
13	4,667	3,8056	3,4105	3,1791	3,0254	2,9153	2,8321	2,7669	2,7144
14	4,600	3,7389	3,3439	3,1122	2,9582	2,8477	2,7642	2,6987	2,6458
15	4,543	3,6823	3,2874	3,0556	2,9013	2,7905	2,7066	2,6408	2,5876
16	4,494	3,6337	3,2389	3,0069	2,8524	2,7413	2,6572	2,5911	2,5377
17	4,451	3,5915	3,1968	2,9647	2,8100	2,6987	2,6143	2,5480	2,4943
18	4,413	3,5546	3,1599	2,9277	2,7729	2,6613	2,5767	2,5102	2,4563
19	4,380	3,5219	3,1274	2,8951	2,7401	2,6283	2,5435	2,4768	2,4227
20	4,351	3,4928	3,0984	2,8661	2,7109	2,5990	2,5140	2,4471	2,3928
21	4,324	3,4668	3,0725	2,8401	2,6848	2,5727	2,4876	2,4205	2,3660
22	4,300	3,4434	3,0491	2,8167	2,6613	2,5491	2,4638	2,3965	2,3419
23	4,279	3,4221	3,0280	2,7955	2,6400	2,5277	2,4422	2,3748	2,3201

Таблица 4

значимости 0,05 и степенью свободы k_1 и k_2

10	15	20	24	30	40	60	120	∞
241,881	245,949	248,013	249,051	250,095	251,143	252,195	253,252	254,314
19,3959	19,4291	19,4458	19,4541	19,4624	19,4707	19,4791	19,4874	19,4957
8,7855	8,7029	8,6602	8,6385	8,6166	8,5944	8,5720	8,5494	8,5264
5,9644	5,8578	5,8025	5,7744	5,7459	5,7170	5,6877	5,6581	5,6281
4,7351	4,6188	4,5581	4,5272	4,4957	4,4638	4,4314	4,3985	4,3650
4,0600	3,9381	3,8742	3,8415	3,8082	3,7743	3,7398	3,7047	3,6689
3,6365	3,5107	3,4445	3,4105	3,3758	3,3404	3,3043	3,2674	3,2298
3,3472	3,2184	3,1503	3,1152	3,0794	3,0428	3,0053	2,9669	2,9276
3,1373	3,0061	2,9365	2,9005	2,8637	2,8259	2,7872	2,7475	2,7067
2,9782	2,8450	2,7740	2,7372	2,6996	2,6609	2,6211	2,5801	2,5379
2,8536	2,7186	2,6464	2,6090	2,5705	2,5309	2,4901	2,4480	2,4045
2,7534	2,6169	2,5436	2,5055	2,4663	2,4259	2,3842	2,3410	2,2962
2,6710	2,5331	2,4589	2,4202	2,3803	2,3392	2,2966	2,2524	2,2064
2,6022	2,4630	2,3879	2,3487	2,3082	2,2664	2,2229	2,1778	2,1307
2,5437	2,4034	2,3275	2,2878	2,2468	2,2043	2,1601	2,1141	2,0658
2,4935	2,3522	2,2756	2,2354	2,1938	2,1507	2,1058	2,0589	2,0096
2,4499	2,3077	2,2304	2,1898	2,1477	2,1040	2,0584	2,0107	1,9604
2,4117	2,2686	2,1906	2,1497	2,1071	2,0629	2,0166	1,9681	1,9168
2,3779	2,2341	2,1555	2,1141	2,0712	2,0264	1,9795	1,9302	1,8780
2,3479	2,2033	2,1242	2,0825	2,0391	1,9938	1,9464	1,8963	1,8432
2,3210	2,1757	2,0960	2,0540	2,0102	1,9645	1,9165	1,8657	1,8117
2,2967	2,1508	2,0707	2,0283	1,9842	1,9380	1,8894	1,8380	1,7831
2,2747	2,1282	2,0476	2,0050	1,9605	1,9139	1,8648	1,8128	1,7570

k_2 \ k_1	1	2	3	4	5	6	7	8	9
24	4,259	3,4028	3,0088	2,7763	2,6207	2,5082	2,4226	2,3551	2,3002
25	4,241	3,3852	2,9912	2,7587	2,6030	2,4904	2,4047	2,3371	2,2821
26	4,225	3,3690	2,9752	2,7426	2,5868	2,4741	2,3883	2,3205	2,2655
27	4,210	3,3541	2,9604	2,7278	2,5719	2,4591	2,3732	2,3053	2,2501
28	4,196	3,3404	2,9467	2,7141	2,5581	2,4453	2,3593	2,2913	2,2360
29	4,183	3,3277	2,9340	2,7014	2,5454	2,4324	2,3463	2,2783	2,2229
30	4,170	3,3158	2,9223	2,6896	2,5336	2,4205	2,3343	2,2662	2,2107
40	4,084	3,2317	2,8387	2,6060	2,4495	2,3359	2,2490	2,1802	2,1240
60	4,001	3,1504	2,7581	2,5252	2,3683	2,2541	2,1665	2,0970	2,0401
120	3,920	3,0718	2,6802	2,4472	2,2899	2,1750	2,0868	2,0164	1,9588
∞	3,8415	2,9957	2,6049	2,3719	2,2141	2,0986	2,0096	1,9384	1,8799

Окончание табл. 4

10	15	20	24	30	40	60	120	∞
2,2547	2,1077	2,0267	1,9838	1,9390	1,8920	1,8424	1,7896	1,7330
2,2365	2,0889	2,0075	1,9643	1,9192	1,8718	1,8217	1,7684	1,7110
2,2197	2,0716	1,9898	1,9464	1,9010	1,8533	1,8027	1,7488	1,6906
2,2043	2,0558	1,9736	1,9299	1,8842	1,8361	1,7851	1,7306	1,6717
2,1900	2,0411	1,9586	1,9147	1,8687	1,8203	1,7689	1,7138	1,6541
2,1768	2,0275	1,9446	1,9005	1,8543	1,8055	1,7537	1,6981	1,6376
2,1646	2,0148	1,9317	1,8874	1,8409	1,7918	1,7396	1,6835	1,6223
2,0772	1,9245	1,8389	1,7929	1,7444	1,6928	1,6373	1,5766	1,5089
1,9926	1,8364	1,7480	1,7001	1,6491	1,5943	1,5343	1,4673	1,3893
1,9105	1,7505	1,6587	1,6084	1,5543	1,4952	1,4290	1,3519	1,2539
1,8307	1,6664	1,5705	1,5173	1,4591	1,3940	1,3180	1,2214	1,0000

Т а б л и ц а 5

Значения $P_k(\lambda) = \dfrac{\lambda^k e^{-\lambda}}{k!}$ (распределение Пуассона)

k	$\lambda = 0{,}1$	$\lambda = 0{,}2$	$\lambda = 0{,}3$	$\lambda = 0{,}4$	$\lambda = 0{,}5$	$\lambda = 0{,}6$	$\lambda = 0{,}7$	$\lambda = 0{,}8$	$\lambda = 0{,}9$	
0	0,9018	0,8187	0,7408	0,6703	0,6065	0,5458	0,4966	0,4493	0,4066	
1	0,0905	0,1638	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659	
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647	
3	0,0002	0,0019	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494	
4		0,0001	0,0002	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111	
5				0,0001	0,0002	0,0004	0,0007	0,0012	0,0020	
6							0,0001	0,0002	0,0003	

k	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0000
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023
3	0,0613	0,1804	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337	0,0189
5	0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0,0378
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631

Окончание табл. 5

k	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
7	0,0001	0,0037	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1171	0,0901
8		0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318	0,1126
9		0,0002	0,0027	0,0132	0,0363	0,0688	0,1014	0,1241	0,1318	0,1251
10			0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186	0,1251
11			0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970	0,1137
12			0,0001	0,0006	0,0034	0,0126	0,0263	0,0481	0,0728	0,0948
13				0,0002	0,0013	0,0052	0,0142	0,0296	0,0504	0,0729
14				0,0001	0,0005	0,0022	0,0071	0,0169	0,0324	0,0521
15					0,0002	0,0009	0,0033	0,0090	0,0194	0,0347
16						0,0003	0,0014	0,0045	0,0109	0,0217
17						0,0001	0,0006	0,0021	0,0058	0,0128
18							0,0002	0,0009	0,0029	0,0071
19							0,0001	0,0004	0,0014	0,0037
20								0,0002	0,0006	0,0019
21								0,0001	0,0003	0,0009
22									0,0001	0,0004
23										0,0002
24										0,0001

Т а б л и ц а 6

Значения функции Гаусса $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	268
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	220
1,1	2179	2155	2131	2107	2083	2059	2036	2012	198	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518

Продолжение табл. 6

	0	1	2	3	4	5	6	7	8	9
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061

Окончание табл. 6

	0	1	2	3	4	5	6	7	8	9
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Учебное издание

Трофимова Елена Александровна
Кисляк Надежда Валерьевна
Гилёв Денис Викторович

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Заведующий редакцией *М. А. Овечкина*
Редактор *С. Г. Галинова*
Корректор *С. Г. Галинова*
Компьютерная верстка *Г. Б. Головина*

Подписано в печать 07.05.18. Формат 60×84/16.
Бумага офсетная. Цифровая печать.
Уч.-изд. л. 7,5. Усл. печ. л. 9,3. Тираж 50 экз. Заказ 88.

Издательство Уральского университета
Редакционно-издательский отдел ИПЦ УрФУ
620083, Екатеринбург, ул. Тургенева, 4
Тел.: +7 (343) 389-94-79, 350-43-28
E-mail: rio.marina.ovechkina@mail.ru

Отпечатано в Издательско-полиграфическом центре УрФУ
620083, Екатеринбург, ул. Тургенева, 4
Тел.: +7 (343) 358-93-06, 350-58-20, 350-90-13
Факс: +7 (343) 358-93-06
http://print.urfu.ru

ЛИДЕР УНИВЕРСИТЕТСКОГО КНИГОИЗДАНИЯ
УНИВЕРСИТЕТСКАЯ КНИГА
2015

9785799623173