ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
Государственное образовательное учреждение высшего профессионального образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Л.И. Лазарева, А.А. Михальчук

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

*Рекомендовано в качестве учебного пособия
Редакционно-издательским советом
Томского политехнического университета*

Второе издание, стереотипное

Издательство
Томского политехнического университета
Томск 2010

УДК 519.21+519.22(075.8)
ББК 22.171+22.172я73
Л171

Лазарева Л.И.

Л171 Теория вероятностей и математическая статистика: учебное пособие / Л.И. Лазарева, А.А. Михальчук; Национальный исследовательский Томский политехнический университет. – 2-е изд., стер. – Томск: Изд-во Томского политехнического университета, 2010. – 144 с.

Учебное пособие включает краткие теоретические сведения; примеры и рекомендации по решению типовых задач; таблицы значений необходимых функций и распределений; таблицу важнейших дискретных и непрерывных распределений случайных величин; наглядные графические иллюстрации, выполненные в пакете «Mathematica – 3.0»; список рекомендуемой литературы. Пособие поможет усвоить теоретический материал и овладеть необходимыми практическими навыками для выполнения контрольных работ.

Пособие подготовлено на кафедре высшей математики и математической физики, соответствует программе дисциплины и предназначено для студентов ИДО, обучающихся по направлениям 140100 «Теплоэнергетика и теплотехника», 140400 «Электроэнергетика и электротехника», 280700 «Техносферная безопасность».

УДК 519.21+519.22(075.8)
ББК 22.171+22.172я73

Рецензент

Кандидат физико-математических наук,
доцент кафедры математики ТУСУРа
Г.С. Голосова

Часть 1

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ

1.1. Предмет и задачи теории вероятностей

Теория вероятностей – математическая наука, позволяющая по вероятностям одних случайных событий находить вероятности других случайных событий, связанных каким-либо образом между собой.

В этом определении есть целый ряд понятий: случайное событие, вероятность случайного события, связь между случайными событиями. Все эти понятия нуждаются в определении и разъяснении. В усвоении этого круга вопросов и состоит первое знакомство с теорией вероятностей.

Теория вероятностей изучает свойства массовых случайных событий, способных многократно повторяться при воспроизведении определенного комплекса условий. Основное свойство любого случайного события независимо от его природы – вероятность его осуществления.

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

Теория вероятностей вначале развивалась как прикладная дисциплина. В связи с этим ее понятия и выводы имели окраску тех областей знаний, в которых они были получены. Лишь постепенно выкристаллизовалось то общее, что присуще вероятностным схемам независимо от области их приложения – массовые случайные события, действия над ними и их вероятности, случайные величины и их числовые характеристики. Большой вклад в развитие теории вероятностей внесли русские ученые. Впервые законченную систему аксиом сформулировал в 1936 г. советский математик академик А.Н. Колмогоров в своей книге «Основные понятия теории вероятностей». Практические приложения способствовали зарождению теории вероятностей, они же питают ее развитие как науки, приводя к появлению все новых ее ветвей и разделов.

Одной из важнейших сфер приложения теории вероятностей время трудно является экономика. В настоящее себе представить исследование и прогнозирование экономических явлений без использования эконометрического моделирования, регрессионного анализа, трендовых и сглаживающей моделей и других методов, опирающихся на теорию вероятностей.

Методы теории вероятностей широко применяются в различных отраслях естествознания и техники: в теории надёжности, теории массового обслуживания, в теоретической физике, геодезии, астрономии, теории стрельбы, теории ошибок наблюдений, теории автоматического управления, общей теории связи и во многих других теоретических и прикладных науках. Все это предопределяет необходимость овладения методами теории вероятностей и математической статистики как инструментом статистического анализа и прогнозирования экономических и технологических явлений и процессов.

1.2. Случайные события и операции над ними

Осуществление каждого отдельного наблюдения, опыта или измерения при изучении эксперимента называют испытанием. Результат испытания назовем ***событием***.

Различают события: достоверные, невозможные и случайные.

Достоверное событие **–** это такое событие, которое всегда происходит в рассматриваемом эксперименте (содержит все точки множества Ω).

Невозможное событие **–** это такое событие, которое никогда не происходит в рассматриваемом эксперименте.

Событие, которое при воспроизведении опыта может наступить, а может и не наступить, называют ***случайным событием***.

События обозначаются большими латинскими буквами A, B, C, ..., невозможное – $\varnothing$, достоверное – Ω.

*ПРИМЕР 1. Монета подбрасывается три раза, описать все возможные события этого опыта.

Решение. Найдем множество всех возможных исходов опыта. Обозначим событие Г – выпадение «герба» вверх, Ц – выпадение «цифры» вверх. Тогда Ω = {ГГГ, ГГЦ, ГЦЦ, ЦЦЦ, ЦЦГ, ЦГГ, ЦГЦ, ГЦГ}. В нашем эксперименте 8 исходов опыта.

Определим операции над событиями, проиллюстрировав их геометрически. Пусть на плоскость, ограниченную прямоугольником, на удачу ставится точка и пусть события A и B состоят в том, что эта точка попадает соответственно в круг A, в круг B (на рис. 1.2.1а – 1.2.1е).

Событие, состоящие в наступлении *хотя бы одного из событий* A или B, будем называть ***суммой*** (*объединением*) событий A и B и обозначать $A+B$ или $A \cup B$ (случай «а» на рис. 1.2.1).

Событие, состоящее в наступлении *обоих событий и* A *и* B, будем называть ***произведением*** (*пересечением*) событий A и B и обозначать A B или $A \cap B$ (случай «б»).

Событие, состоящее в том, что A происходит, а B не происходит, называется ***разностью*** событий A и B и обозначается $A \backslash B$ или $A - B$ (случай «в»).

Событие, обозначаемое через $\overline{A}$, называется ***противоположным*** событию A, если оно происходит тогда и только тогда, когда событие A не происходит (случай «г»).

Событие A ***влечет*** B (обозначается $A \subset B$), если при наступлении события A событие B обязательно происходит (случай «д» соответствует событию $B \subset A$).

Если $A \subset B$ и $B \subset A$, то события A и B называются ***равносильными***, или эквивалентными (записывают $A = B$).

Если наступление события A делает невозможным наступление события B (и наоборот), то событие A и B называются ***несовместными*** или непересекающимися, в этом случае $A \cap B = \varnothing$ (рис. 1.2.1 *е*). Для ***совместных*** событий $A \cap B \neq \varnothing$.

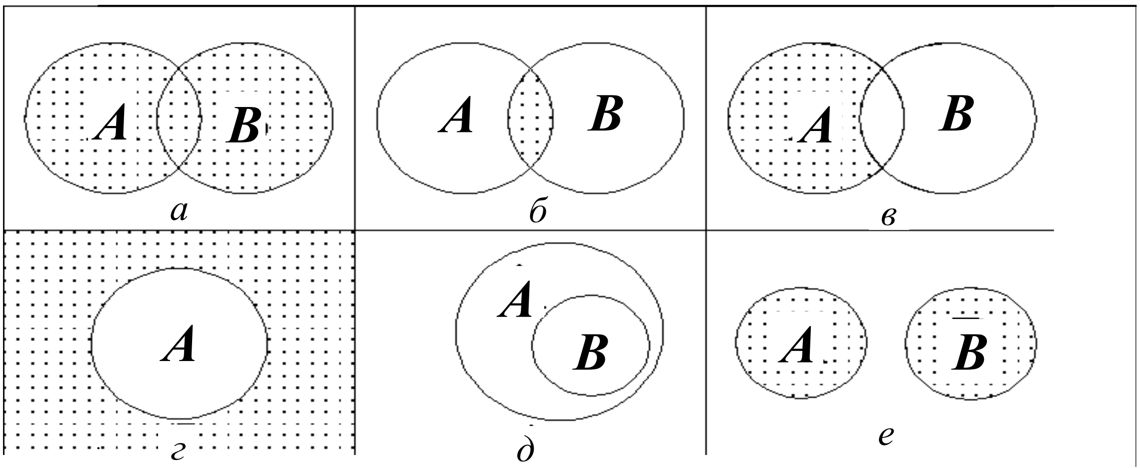


Рис. 1.2.1

***ПРИМЕР 2.** В урне имеются белые и черные шары. При одном испытании (взятии одного шара) событие A (появление белого шара) и событие B (появление черного шара). События A и B являются несовместными событиями в данном испытании.

***ПРИМЕР 3.** Производится один выстрел из орудия. События «разрыв снаряда» и «не разрыв снаряда» – несовместные события, а попадание снаряда в цель и разрыв снаряда – события совместные.

События A_1, A_2, ..., A_k образуют ***полную группу*** событий, если

$$A_1 \cup A_2 \cup \ldots \cup A_k = \Omega = \bigcup_{i=1}^{n} A_i . \tag{1.2.1}$$

Попарно несовместные события ($A_i \cdot A_j = \varnothing$, $\forall\ i \neq j$), образующие полную группу событий называют ***гипотезами***.

*ПРИМЕР 4. Приобретены два билета денежно-вещевой лотереи. Обязательно произойдет одно и только одно из следующих событий: «выигрыш выпал на первый билет и не выпал на второй», «выигрыш не выпал на первый билет и выпал на второй», «выигрыш выпал на оба билета», «на оба билета выигрыш не выпал». Эти события образуют полную группу попарно несовместных событий.

Под ***равновозможными*** событиями принято понимать такие события, что по некоторым условиям «симметрии» опыта нет основания считать, что какое-нибудь из событий более возможно, чем другое.

Элементарными событиями или исходами называют события, удовлетворяющие трем условиям: 1) они попарно несовместны; 2) образуют полную группу и 3) равновозможны.

Операции над событиями удовлетворяют следующим свойствам:

1) $A \cup B = B \cup A$;
2) $A \cap B = B \cap A$;
3) $A \cup A = A$;
4) $A \cap A = A$;
5) $A \cup \Omega = \Omega$;
6) $A \cap \Omega = A$;
7) $A \cup \varnothing = A$;
8) $A \cap \varnothing = \varnothing$;
9) $A \cup (B \cup C) = (A \cup B) \cup C$;
10) $A \cap (B \cap C) = (A \cap C) \cap B$;
11) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
12) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
13) $\overline{\varnothing} = \Omega$;
14) $\overline{\Omega} = \varnothing$;
15) $\overline{A \cup B} = \overline{A} \cap \overline{B}$;
16) $\overline{A \cap B} = \overline{A} \cup \overline{B}$;
17) $\overline{\overline{(A)}} = A$;
18) $A \cap \overline{A} = \varnothing$;
19) $A \cup \overline{A} = \Omega$.

Все перечисленные выше свойства предлагаем проверить геометрически (см. рис. 1.2.1).

*ПРИМЕР 5. Взятая наудачу деталь может оказаться либо первого (событие A), либо второго (событие B), либо третьего (событие C) сорта. Что представляют собой следующие события: $A + B$; $\overline{A+C}$; $A \cdot C$; $A \cdot B$; $A \cdot B + C$?

Решение. По условию задачи события A, B, C – попарно несовместны, т. е. не может деталь быть сразу двух сортов, кроме того события A, B, C – образуют полную группу событий, так как нет не сортовых деталей. $A + B$ – это событие, которое состоится при наступлении хотя бы одного из событий A и B. Следовательно, $A + B$ в нашем случае – деталь первого или второго сорта и $A + B = \overline{C}$. Так как $A + C$ – деталь первого или третьего сорта, то противоположное этому событие $\overline{A+C}$ – деталь второго сорта. ($A \cdot C$) – невозможное событие, поскольку деталь одновременно не может быть и первого и третьего сорта. $A \cdot B + C$ как сумма невозможного события и события C равно C, т. е. $A \cdot B + C = C$ – деталь третьего сорта.

*ПРИМЕР 6. Васильев и Петров договорились в воскресенье пойти на футбол, если Петров купит в субботу билеты, Васильев исправит неудовлетворительную оценку и если в воскресенье не будет дождя.

Решение. Рассмотрим события: A = «Петров в субботу купил билеты на футбол», B = «Васильев исправил неудовлетворительную оценку», C = «в воскресенье нет дождя» и H = «Петров и Васильев в воскресенье пошли на футбол». Ясно, что

$$H = A \cap B \cap C.$$

1.3. Необходимые формулы комбинаторики

При подсчете числа элементарных исходов, составляющих события в классической схеме, часто используется комбинаторика. Сформулируем **основное правило комбинаторики** (правило умножения).

Пусть требуется выполнить одно за другим k действий. Если первое действие можно выполнить n_1 способами, второе действие – n_2 способами, третье действие n_3 способами и так до k-го действия, которое можно выполнить n_k способами, то всю последовательность из k действий вместе можно выполнить $n_1 \times n_2 \times n_3 \times ... \times n_k$ способами.

*ПРИМЕР 1. Сколькими способами N можно собрать слово «мама», имея в азбуке пять букв «а» и три буквы «м»?

Решение. Первую букву слова можно выбрать тремя способами и на каждый вариант первой буквы имеется пять способов выбрать вторую букву. Значит, способов собрать «ма»: $3 \cdot 5 = 15$. Для каждого из них третья буква может быть получена двумя способами (остается только две буквы «м»), а последняя буква – четырьмя способами:

$$N = 3 \cdot 5 \cdot 2 \cdot 4 = 120.$$

Формула для числа перестановок

Она применяется в задачах о перестановках в различных комбинациях нескольких разных объектов, причем в каждой комбинации должны присутствовать все объекты строго по одному разу.

Число таких различных комбинаций (**перестановок)** определяется формулой

$$P_n = n!, \tag{1.3.1}$$

которая непосредственно следует из основного правила комбинаторики.

*ПРИМЕР 2. Сколько существует способов расстановки на полке 6 разных книг?

Решение. На первое место можно поставить любую из 6 книг, для каждого варианта первой книги на второе место может быть поставлена любая из оставшихся 5 книг. Для любой пары первых книг (а всего таких пар $6 \cdot 5 = 30$) на третьем месте может быть одна из 4 книг. Значит, разных троек всего $6 \cdot 5 \cdot 4 = 120$ и т. д. Итак, число перестановок из 6 книг равно $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$.

Формула для числа размещений из *n* элементов по *k*

Если из n разных объектов выбирается по k объектов, то полное число таких различных выборок может быть определено по формуле $\tilde{A}_n^k = n^k$, если выборки отличаются порядком следования объектов, и допускается повторение одного и того же объекта. Это число называют числом размещений с повторениями, оно получается из основного правила комбинаторики, так как на любом из k мест в выборке может быть любой из n объектов.

Если из n разных объектов выбирается по k разных объектов, то с учетом порядка следования полное число разных выборок будет определять формула

$$A_n^k = \frac{n!}{(n-k)!} = n\,(n-1)\,...\,(n-k+1) \qquad (1.3.2)$$

– ***число размещений без повторений***.

Из основного правила эта формула получается на основе следующих рассуждений. На первом месте может быть любой из n объектов, на втором – любой из $(n-1)$ неиспользованных объектов (так как объекты не должны повторяться) и так далее, а на последнем, k-м месте, – любой из неиспользованных $(n-k+1)$ объектов. Заметим, что $A_n^n = P_n$.

*ПРИМЕР 3. Сколько трехзначных чисел можно составить из цифр 1, 2, 3, 4, 5? И сколько из них с неповторяющимися цифрами?

Решение. Если цифры могут повторяться, то на любом месте в числе могут быть любые из пяти цифр. Значит, всего трехзначных чисел получается $5 \cdot 5 \cdot 5 = 5^3 = 125$. Если же цифры не повторяются, то таких чисел

$$A_5^3 = 5 \cdot 4 \cdot 3 = 60.$$

*ПРИМЕР 4. В театре 10 актеров и 8 актрис. Сколькими способами можно распределить роли в спектакле, в котором 6 мужских и 3 женские роли?

Решение. Рассуждаем следующим образом: распределяем мужчин на мужские роли (первое действие). Тогда $n_1 = A_{10}^6$ (важно не только выбрать актеров, но и распределить между ними роли). После этого производим второе действие – распределяем женские роли. Это можно осуществить $n_2 = A_8^3$ способами. Поэтому по принципу произведения

$$B = n_1 \cdot n_2 = \mathrm{A}_{10}^6 \cdot \mathrm{A}_8^3.$$

Формула для числа сочетаний из *n* элементов по *k*

Если в выборках из n объектов по k объектов порядок их следования по условию задачи не имеет значения, то размещения, отличающиеся лишь порядком следования, становятся одинаковыми. Число таких одинаковых выборок по k разных объектов, которые получаются друг из друга перестановкой, равно $k!$

Поэтому, число выборок из n по k без учета порядка следования определяется формулой

$$C_n^k = A_n^k / k! = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\ldots(n-k+1)}{k!} \qquad (1.3.3)$$

– ***число сочетаний без повторений***.

$$\tilde{C}_n^k = \tilde{A}_n^k / k! \qquad (1.3.4)$$

– ***число сочетаний с повторениями***, т. е. число сочетаний меньше числа размещений в $k!$ раз.

*ПРИМЕР 5. Сколько комбинаций из трех монет можно собрать, имея пять разных монет?

Решение. Если учитывать порядок монет в комбинации, то

$$A_5^3 = 5 \cdot 4 \cdot 3 = 60.$$

Если же порядок монет в комбинации не имеет значения, то разных комбинаций:

$$C_5^3 = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10.$$

*ПРИМЕР 6. Найти число B всевозможных заполнений карточки спортлото «6» из «49».

Решение. Генеральная совокупность – числа карточки спортлото (n = 49). Выборка – зачеркнутые 6 чисел. Порядок, в котором вычеркиваются номера, не существенен. Повторов быть не может (в карточке любой номер есть только один раз).

Поэтому $B = C_{49}^6$ = 13 983 816.

Сочетания используются, если важен только состав элементов в выборке.

1.4. Понятие вероятности события

Сравнивать случайные события (т. е. события, которые могут наступить или не наступить) естественно по степени возможности их наступления. С этой целью вводится числовая характеристика этой степени возможности (случайности), называемая вероятностью события. Для события A, вероятность принято обозначать $P(A)$. Существует несколько подходов, поясняющих понятие вероятности. В каждом из этих подходов указываются правила, по которым случайному событию ставится в соответствие положительное число, объективно характеризующее степень возможности появления этого события.

Большинство определений математической вероятности может быть разделено на следующие: статистическое, классическое и геометрическое.

1.4.1. Статистическое определение вероятности

Многочисленными наблюдениями над самыми разнообразными случайными событиями установлен следующий достоверный факт: если над одним и тем же случайным событием в одних и тех же условиях проводить много серий из большого числа испытаний каждая, то наблюдаемая в каждой такой серии частота события будет колебаться от серии к серии в сравнительно узких пределах, будет, как говорят в теории вероятностей, «устойчивой». При этом пределы, в которых будет колебаться устойчивая частота случайного события, будут тем теснее, чем больше число испытаний в каждой серии. Это свидетельствует о наличии статистической закономерности в изучаемом явлении.

Пусть в одних и тех же условиях проведена серия из n^* испытаний, в каждом из которых могло появиться или не появиться интересующее нас событие A. Пусть событие A появилось при этом в m^* испытаниях.

Относительной частотой $P^*(A)$ события A в данной серии испытаний называется отношение m^* (числа испытаний, в которых появилось событие A) к n^* (общему числу проведенных испытаний), т. е.

$$P^*(A) = \frac{m^*}{n^*}. \qquad (1.4.1)$$

Из данного определения следует, что относительная частота случайного события всегда заключена между нулем и единицей:

$$0 \le P^*(A) \le 1. \qquad (1.4.2)$$

Статистической вероятностью $P(A)$ события A называется предел, к которому стремится относительная частота $P^*(A)$ при неограниченном увеличении числа испытаний, т. е.

$$P(A) = \lim_{n\to\infty} P^*(A) = \lim_{n\to\infty} \frac{m^*}{n^*}. \qquad (1.4.3)$$

*****ПРИМЕР 1. Производя 100 выстрелов, стрелок попал в цель 89 раз. Чему равна относительная частота попадания в цель данного стрелка?

Решение. В этой задаче нас интересует событие A – попадание в цель. Произведено $n^* = 100$ выстрелов, причем событие A наступило $m^* = 89$ раз. Поэтому

$$P^*(A) = \frac{89}{100} = 0{,}89.$$

При больших n статистическое определение позволяет в приблизительных расчетах относительную частоту $P^*(A)$ принимать за вероятность события A. Недостатком этого определения вероятности является необходимость проведения большого числа опытов в одинаковых условиях.

1.4.2. Классическое определение вероятности

Классическое определение вероятности сводит понятие вероятности к понятию равновероятности (равновозможности) событий, которое считается основным и не подлежит формальному определению. Это определение применимо в случаях, когда удается выделить полную группу несовместных и равновероятных событий – элементарных исходов. Для примера рассмотрим урну с шарами.

Пусть в урне содержится 7 одинаковых, тщательно перемешанных шаров, причем 2 из них – красных, 1 – синий и 4 – белых. Испытание будет заключаться в том, что из урны наудачу берется один шар. Каждое событие, которое может наступить в проводимом испытании, является элементарным исходом. В данном примере семь элементарных исходов, которые мы обозначим E_1, E_2, ..., E_7. Исходы E_1, E_2 – появление красного шара, E_3 – появление синего шара, E_4, E_5, E_6, E_7 – появление белого шара. В нашем примере события E_1, E_2, ..., E_7 – попарно несовместны. Кроме того, они еще и равновозможны в данном испытании. Пусть событие A заключается в том, что наудачу взятый из урны шар оказался цветным (красным или синим).

Те элементарные исходы, при которых интересующее нас событие A наступает, называют ***исходами, благоприятствующими*** событию A. В нашем примере исходами, благоприятствующими событию A, являются исходы E_1, E_2 и E_3. Разумно в качестве меры возможности появления события A, т. е. вероятности $P(A)$, принять число, равное отношению исходов, благоприятствующих наступлению события A, к числу всех возможных исходов. В нашем примере

$$P(A) = \frac{3}{7}.$$

*Р*ассмотренный пример привел нас к определению вероятности, которое принято называть ***классическим***.

Вероятностью события A называют отношение числа m благоприятствующих этому событию исходов к общему числу n всех ***элементарных исходов***:

$$P(A) = \frac{m}{n}. \tag{1.4.4}$$

Классическое определение вероятности служит хорошей математической моделью тех случайных экспериментов, число исходов которых конечно, а сами исходы – равновозможны.

*ПРИМЕР 2. Бросается игральная кость. Найти вероятность того, что выпадет не более четырех очков.

Решение. Общее число элементарных исходов $n = 6$ (могут выпасть 1, 2, 3, 4, 5, 6). Среди этих исходов благоприятствуют событию A (выпадет не более четырех очков) только четыре исхода $m = 4$. Следовательно искомая вероятность

$$P(A) = \frac{4}{6} = \frac{2}{3}.$$

*ПРИМЕР 3. Какова вероятность, заполняя карточку спортлото «6» из «49» угадать 4 номера?

Решение. Общее число элементарных исходов опыта равно числу способов, которыми можно зачеркнуть 6 номеров из 49, т. е. $n = C_{49}^{6}$. Найдем число исходов, благоприятствующих интересующему нас событию A = {угадано 4 номера}, 4 номера из 6 выигравших можно зачеркнуть C_6^4 способами, при этом остальные два номера должны быть не выигравшими. Зачеркнуть 2 неправильных номера из 43 невыигрышных можно C_{43}^{2} способами. Следовательно, число благоприятствующих исходов $m = C_6^4 \cdot C_{43}^{2}$. Принимая во внимание, что все исходы опыта являются несовместными и равновозможными, находим искомую вероятность по формуле классической вероятности:

$$P(A) = \frac{m}{n} = \frac{C_6^4 \cdot C_{43}^2}{C_{49}^6}.$$

*ПРИМЕР 4. Наудачу взятый телефонный номер состоит из 5 цифр. Как велика вероятность, что в нем: 1) все цифры различные; 2) все цифры нечетные?

Решение

1. Так как на каждом из пяти мест в пятизначном номере может стоять любая из цифр: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, то всех различных пятизначных номеров будет 10^5 (00000 – 1-й, 00001 – 2-й, 00002 – 3-й, ..., 99998 – 99999-й, и, наконец, 99999 – 100 000-й). Номера, у которых все цифры различные, – это размещения из 10 элементов по 5.

Формула для числа *размещений* из n элементов по k:

$$A_n^k = C_n^k\ k\,! = \frac{n!}{(n-k)!} = n\,(n-1)\ldots(n-k+1).$$

Поэтому число благоприятствующих случаев $m = A_{10}^5 = 10\cdot 9\cdot 8\cdot 7\cdot 6$ и искомая вероятность

$$P(A) = \frac{m}{n} = \frac{A_{10}^5}{10^5} = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6}{10^5} = 0{,}3024.$$

2. Из 5 нечетных цифр (1, 3, 5, 7, 9) можно образовать 5^5 различных пятизначных номеров. 5^5 – это число благоприятных исходов m. Так как всех равновозможных случаев $n = 10^5$, то искомая вероятность

$$P(A) = \frac{m}{n} = \frac{5^5}{10^5} = \left(\frac{1}{2}\right)^5 = \frac{1}{32} = 0{,}03125.$$

*ПРИМЕР 5. Полная колода карт (52 листа) делится наугад на две равные пачки по 26 листов. Найти вероятности следующих событий:

A – в каждой из пачек окажется по два туза;

B – в одной из пачек не будет ни одного туза, а в другой – все четыре;

C – в одной из пачек будет один туз, а в другой – три.

Решение. Общее число возможных элементарных исходов испытания равно числу способов, которыми можно извлечь 26 карт из 52 , т. е. числу сочетаний из 52 по 26, $n = C_{52}^{26}$. Число благоприятных событию A случаев $m = C_4^2 \cdot C_{48}^{24}$ (по основному правилу комбинаторики), где первый сомножитель показывает, что два туза из четырех можно взять C_4^2 способами, второй сомножитель показывает, что остальные 24 карты берутся из 48 карт, не содержащих тузов, C_{48}^{24} способами. Искомая вероятность равна отношению числа исходов, благоприятствующих событию A, к общему числу всех исходов:

$$P(A) = \frac{C_4^2 \cdot C_{48}^{24}}{C_{52}^{26}}.$$

Событие B может осуществиться двумя равновозможными способами: либо в первой пачке будут все четыре туза, а во второй – ни одного, либо наоборот:

$$P(B) = 2\frac{C_4^4 \cdot C_{48}^{22}}{C_{52}^{26}}.$$

Аналогично

$$P(C) = 2\frac{C_4^3 \cdot C_{48}^{23}}{C_{52}^{26}}.$$

Заметим, что классическое определение вероятности было введено для случая, когда пространство элементарных событий конечно, а все исходы и испытания равновозможны и несовместны.

1.4.3. Геометрическое определение вероятности

«Классическое» определение вероятности основано на рассмотрении конечной группы равновероятных событий. Для случаев, когда возможно бесконечное множество исходов необходимо перейти к некоторому видоизменению этого определения. В таких случаях вводят понятие геометрической вероятности, т. е. вероятности попадания точки в область (отрезок, часть плоскости, часть тела и т. д.). Пусть, например, на плоскости имеется некоторая область G, и в ней содержится другая область g. Требуется найти вероятность того, что точка, взятая наудачу в области G, попадет в область g.

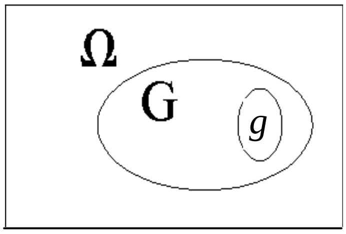


Рис. 1.4.1

При этом выражению «точка, взятая наудачу в области G», придается следующий смысл: эта точка может попасть в любую точку области G.

Пусть пространство элементарных событий состоит из бесконечного числа событий (точек), заполняющих некоторую область (G), а событию A соответствует область (g), содержащаяся в G и состоящая из событий (точек), благоприятствующих A. Тогда вероятностью $P(A)$ события A называется число, вычисляемое по формуле

$$P(A) = \frac{\text{mes} g}{\text{mes} G}. \qquad (1.4.5)$$

(*Геометрическое определение вероятности*)

В этой формуле под символом «**mes *G***» («**mes *g***») понимают длину отрезка, площадь фигуры или объема тела в зависимости от того, где расположена область $G(g)$ на прямой, в плоскости или в пространстве, Тогда вероятность $P(A)$ есть отношение геометрических размеров (мер – «**mes**») областей, т. е. длин отрезков, площадей, объемов.

Геометрическое определение вероятностей является хорошей математической моделью случайных экспериментов, число исходов которых хотя и бесконечно, но все они равноправны.

*ПРИМЕР 6. На отрезке OA длины L числовой оси Ox наудачу поставлена точка $B(x)$. Найти вероятность того, что меньший из отрезков OB и BA имеют длину, большую $L/3$. Предполагается, что вероятность попадания точки на отрезок пропорциональна длине отрезка и не зависит от его расположения на числовой оси.

Решение. Разобьем отрезок OA точками C и D на 3 равные части. Требование задачи будет выполнено, если точка $B(x)$ попадет на отрезок CD длины $L/3$. Искомая вероятность

$$P = (L/3) / L = 1/3.$$

*ПРИМЕР 7. Внутри эллипса $x^2/25 + y^2/16 = 1$ расположен круг $x^2 + y^2 = 9$. Найти вероятность попадания в кольцо, ограниченное эллипсом и кругом, при бросании наудачу точки в эллипс.

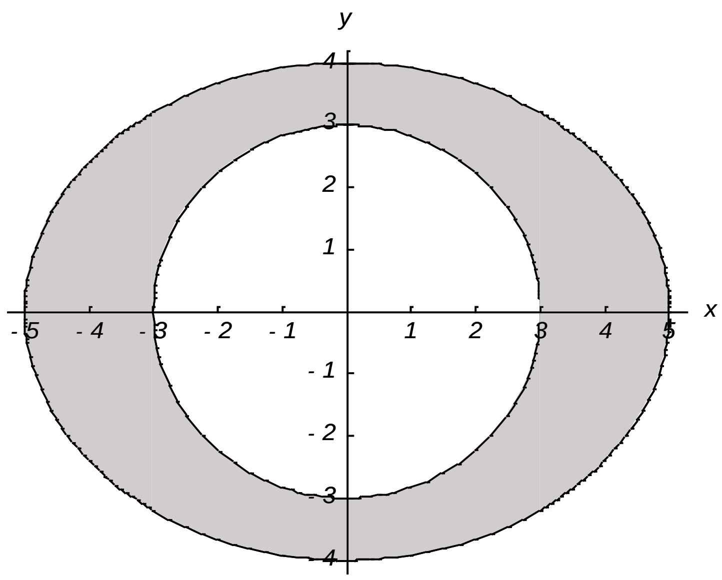


Рис. 1.4.2

Решение. Множество исходов таких и аналогичных им испытаний бесконечно. Оно может быть иллюстрировано геометрически в виде совокупности точек плоской фигуры. Пусть событие A – попадание точки в кольцо.

Тогда $P(A) = \dfrac{S_{\text{кольца}}}{S_{\text{эллипса}}}$, где $S_{\text{кольца}} = S_{\text{эллипса}} - S_{\text{круга}} = \pi\, ab - \pi\, r^2$.

Так как $a = 5$, $b = 4$, $r = 3$, то

$$P(A) = \frac{20\cdot\pi - 9\cdot\pi}{20\cdot\pi} = \frac{11}{20}.$$

*ПРИМЕР 8. (Задача о встрече). Студент (событие A) и студентка (событие B) условились встретиться в определенном месте между 19 ч и 19 ч 50 мин. Пришедший первым ждет другого в течение 10 мин, после чего уходит. Чему равна вероятность их встречи, если каждый из них обязательно придет в течение указанных 50 мин наудачу и независимо от другого?

Решение. Обозначим момент прихода студента A через x, а студентки B – через y. Для того, чтобы встреча произошла, необходимо и достаточно, чтобы $|x - y| \le 10$. Изобразим x и y как декартовы координаты на плоскости, а в качестве единицы масштаба выберем одну минуту. Возможные исходы изобразятся точками квадрата со стороной 50, а исходы, благоприятствующие встрече – точками заштрихованной области (рис. 1.4.3), которая получится при изображении решения неравенства $|x - y| \le 10$:

$$x - y \le 10, \quad x - y \ge -10 \qquad \text{или} \qquad x - 10 \le y, \qquad y \le x + 10.$$

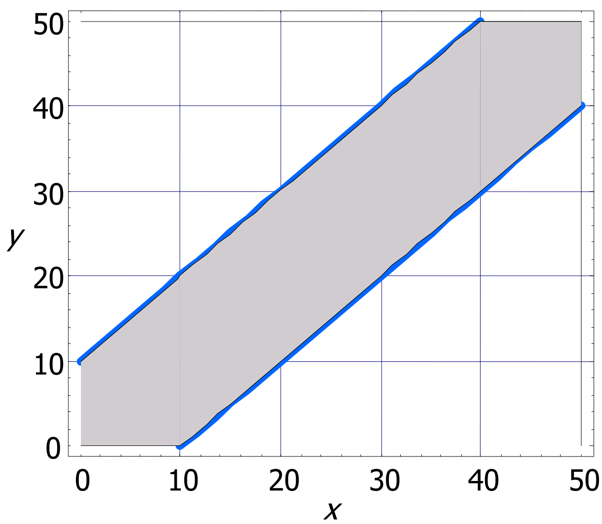


Рис. 1.4.3

Искомая вероятность равна отношению площади заштрихованной фигуры к площади всего квадрата:

$$P(A) = \frac{S_{\text{квадрата}} - 2S_{\text{треугольника}}}{S_{\text{квадрата}}} = \frac{50^2 - 40^2}{50^2} = 0{,}36.$$

*****ПРИМЕР 8′. (Еще одна задача о «встрече»). В сигнализатор поступают сигналы от двух устройств, причем поступление каждого из сигналов равновозможно в любой момент промежутка времени длительностью T = 50 с. Моменты поступления сигналов независимы один от другого. Сигнализатор срабатывает, если разность между моментами поступления сигналов меньше t = 10 с. Найти вероятность того, что сигнализатор сработает за время T, если каждое из устройств пошлет по одному сигналу.

Решение. По постановке данная задача аналогична предыдущей. Используя аналогичную схему рассуждений с привлечением геометрической интерпретации на рис. 1.4.3, можно прийти к аналогичному выводу: искомая вероятность

$$P = \frac{T^2 - (T-t)^2}{T^2} = \frac{50^2 - 40^2}{50^2} = 0{,}36.$$

Пользуясь классическим или геометрическим определением вероятности, легко получить следующие ее свойства:

1. Вероятность любого события A неотрицательна и не превосходит единицу:

$$0 \le P(A) \le 1, \text{ т. к. } m \le n, \text{ «mes } g\text{»} \le \text{«mes } G\text{».}$$

2. Вероятность достоверного события Ω равна единице:

$$P(\Omega) = 1, \text{ т. к. } m = n, \text{ «mes } g\text{»} = \text{«mes } G\text{».}$$

3. Вероятность невозможного события $\varnothing$ равна нулю:

$$P(\varnothing) = 0, \text{ т. к. } m = 0, \text{ «mes } g\text{»} = 0.$$

4. Вероятность противоположного событию A события $\overline{A}$ вычисляется по формуле

$$P(\overline{A}) = 1 - P(A), \quad \text{т. к.} \quad \frac{n}{n} - \frac{m}{n} = \frac{n-m}{n} = \frac{\text{mes}G - \text{mes}g}{\text{mes}G}.$$

1.5. Теоремы сложения и умножения вероятностей. Условная вероятность

1.5.1. Теорема сложения вероятностей событий

Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления:

$$P(A+B) = P(A) + P(B) - P(AB). \qquad (1.5.1)$$

Теорема может быть обобщена на любое конечное число совместных событий, например, в случае трех совместных событий она имеет вид:

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

*ПРИМЕР 1. Найти вероятность того, что наудачу взятое двузначное число окажется кратным либо 2, либо 5, либо тому и другому одновременно.

Решение. Пусть A – событие, которое состоится, если наудачу взятое двузначное число кратно 2, а B – событие, которое состоится, если это число кратно 5. Надо найти $P(A + B)$. Так как A и B – события совместные, то:

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

Двузначные числа – это 10, 11, ..., 98, 99. Всех их – 90 элементарных исходов. Очевидно, 45 из них кратны 2 (благоприятствуют наступлению A), 18 кратны 5 (благоприятствуют наступлению B) и, наконец, 9 кратны и 2, и 5 одновременно (благоприятствуют наступлению $A \cdot B$). Далее по классическому определению вероятности:

$$P(A) = 45/90 = 0{,}5;$$
$$P(B) = 18/90 = 0{,}2;$$
$$P(A \cdot B) = 9/90 = 0{,}1$$

и, следовательно:

$$P(A+B) = 0{,}5 + 0{,}2 - 0{,}1 = 0{,}6.$$

В частном случае для несовместных событий A и B (т. е. когда $A \cdot B = \varnothing$ и $P(A \cdot B) = P(\varnothing) = 0$), теорема сложения имеет вид:

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

*ПРИМЕР 2. В урне 10 белых, 15 черных, 20 синих и 25 красных шаров. Вынули один шар. Найти вероятность того, что вынутый шар: белый; черный; синий; красный; белый или черный; синий или красный; белый, черный или синий.

Решение. Полное число элементарных исходов

$$n = 10 + 15 + 20 + 25 = 70,$$

по классическому определению вероятности:

$$P(\text{Б}) = 10/70 = 1/7; \quad P(\text{Ч}) = 15/70 = 3/14;$$
$$P(\text{С}) = 20/70 = 2/7; \quad P(\text{К}) = 25/70 = 5/14.$$

Применяя теорему сложения вероятностей и учитывая, что нет двухцветных шаров, получим

$$P(\text{Б} + \text{Ч}) = P(\text{Б}) + P(\text{Ч}) = 1/7 + 3/14 = 5/14;$$
$$P(\text{С} + \text{К}) = P(\text{С}) + P(\text{К}) = 2/7 + 5/14 = 9/14;$$
$$P(\text{Б} + \text{Ч} + \text{С}) = 1 - P(\text{К}) = 1 - 5/14 = 9/14.$$

Вероятность события A, вычисленная при условии того, что имело место другое событие B, называется ***условной вероятностью*** события A и обозначается $P(A / B)$ или $P_B(A)$.

События A и B ***независимы***, если справедливо равенство

$$P(A/B) = P(A). \tag{1.5.2}$$

Это легко доказать: поскольку события A и B независимы, то

$$\frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A). \tag{1.5.3}$$

1.5.2. Теорема умножения вероятностей

Вероятность совместного появления двух событий равна произведению вероятности одного из них на условную вероятность другого, вычисленную в предположении, что первое событие уже наступило:

$$P(AB) = P(A) \cdot P(B/A) = P(B) \cdot P(A / B). \tag{1.5.4}$$

В частности, для независимых событий

$$P(AB) = P(A) \cdot P(B) = P(B) \cdot P(A), \tag{1.5.5}$$

т. е. вероятность совместного появления двух независимых событий равна произведению вероятностей этих событий.

Условная вероятность события A_k, определенная в предположении, что осуществились события $A_1, A_2, ..., A_{k-1}$, обозначается

$$P(A_k / A_1 A_2 \ldots A_{k-1}). \qquad (1.5.6)$$

Вероятность произведения нескольких событий равна произведению вероятностей этих событий, причем вероятность каждого следующего по порядку события вычисляется при условии, что все предыдущие имели место:

$$P(A_1 \cdot A_2 \cdot \ldots \cdot A_k) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1A_2) \ldots P(A_k / A_1A_2 \ldots A_k).$$

*ПРИМЕР 3. Три стрелка независимо друг от друга стреляют по цели. Вероятность попадания в цель для первого стрелка равна 0,75, для второго – 0,8, для третьего – 0,9. Определить вероятность того, что все три стрелка одновременно попадут в цель.

Решение. Пусть событие A – попал первый стрелок, событие B – попал второй стрелок. По теореме умножения для независимых событий

$$P(A \cdot B \cdot C) = P(A) \cdot P(B) \cdot P(C) = 0{,}75 \cdot 0{,}8 \cdot 0{,}9 = 0{,}54.$$

*ПРИМЕР 4. У сборщика имеется 3 конусных и 7 эллиптических валиков. Сборщик взял один валик, а затем второй. Найти вероятность того, что первый из взятых валиков – конусный, а второй – эллиптический.

Решение. Вероятность того, что первый валик окажется конусным (событие A),

$$P(A) = 3/10.$$

Вероятность того, что второй валик окажется эллиптическим (событие B), вычисленная в предположении, что первый валик – конусный, т. е. условная вероятность

$$P(B/A) = 7/9.$$

По теореме умножения, искомая вероятность

$$P(AB) = P(A) \cdot P(B/A) = (3/10) \cdot (7/9) = 7/30.$$

*ПРИМЕР 5. В цехе работают семь мужчин и три женщины. По табельным номерам наудачу отобраны три человека. Найти вероятность того, что все отобранные лица окажутся мужчинами.

Решение. Введем обозначения событий: A – первым отобран мужчина; B – вторым отобран мужчина, C – третьим отобран мужчина.

Вероятность того, что первым отобран мужчина, $P(A) = 7/10$. Вероятность того, что вторым отобран мужчина, при условии, что первым уже отобран мужчина, т. е. условная вероятность события B, следующая:

$$P(B/A) = 6/9 = 2/3.$$

Вероятность того, что третьим будет отобран мужчина, при условии, что уже отобраны двое мужчин, т. е. условная вероятность события C, такова:

$$P(C/AB) = 5/8.$$

Искомая вероятность того, что три отобранных лица окажутся мужчинами:

$$P(ABC) = P(A) \cdot P(B/A) \cdot P(C/AB) = 7/10 \cdot 2/3 \cdot 5/8 = 7/24.$$

При числе слагаемых событий больше трех для нахождения вероятности события

$$A_1 + A_2 + \ldots + A_n = \sum_{i=1}^{n} Ai \qquad (1.5.7)$$

проще пользоваться формулой

$$P(\sum_{i=1}^{n} Ai) = 1 - P(\overline{A}_1 \overline{A}_2 \ldots \overline{A}_n), \qquad (1.5.8)$$

которая вытекает непосредственно из приведенного выше свойства операций над событиями:

$$\overline{A_1+A_2+\ldots+A_n} = \overline{A}_1 \overline{A}_2 \ldots \overline{A}_n.$$

В частном случае вероятность того, что произойдет по крайней мере одно из этих событий $A_1, A_2, \ldots, A_k$, определяется по формуле

$$P(A_1 + A_2 + \ldots + A_k) = 1 - P(\overline{A}_1) \cdot P(\overline{A}_2) \ldots P(\overline{A}_k). \qquad (1.5.9)$$

*ПРИМЕР 6. Три стрелка стреляют в цель независимо друг от друга. Вероятность попадания в цель для первого стрелка равна 0,6, для второго – 0,7 и для третьего – 0,75.

1. Найти вероятность по крайней мере одного попадания в цель, если каждый стрелок сделает по одному выстрелу.

2. Найти вероятность того, что будет одно и только одно попадание в цель.

Решение. Пусть A, B, C – *события*, состоящие в том, что соответственно в цель попал первый, второй, третий стрелок. Из условия задачи следует, что $P(A) = 0{,}6$; $P(B) = 0{,}7$; $P(C) = 0{,}75$.

1. Требуется найти вероятность по крайней мере одного попадания в цель, т. е. $P(A + B + C)$.

Так как событие $\overline{A} \cdot \overline{B} \cdot \overline{C}$ – все промахнулись, имеем

$$P(A + B + C) = 1 - P(\overline{A}) \cdot P(\overline{B}) \cdot P(\overline{C}) =$$

$$=1 - (1 - 0,6) \cdot (1 - 0,7) \cdot (1 - 0,75) = 1 - 0,4 \cdot 0,3 \cdot 0,25 = 0,97.$$

2. Подчеркнем, что «хотя бы одно попадание» и «одно попадание» – разные события.

Первое включает второе, значит, его вероятность больше. В нашей задаче одно и только (одно попадание) – это событие

$$I = A \cdot \overline{B} \cdot \overline{C} + B \cdot \overline{A} \cdot \overline{C} + C \cdot \overline{A} \cdot \overline{B}.$$

Его вероятность из-за независимости стрельбы и несовместности слагаемых может быть посчитана по формуле

$$P(I) = P(A) \cdot P(\overline{B}) \cdot P(\overline{C}) + P(B) \cdot P(\overline{A}) \cdot P(\overline{C}) + P(C) \cdot P(\overline{A}) \cdot P(\overline{B}) =$$

$$= 0,6 \cdot (1 - 0,7) \cdot (1 - 0,75) + 0,7 \cdot (1 - 0,6) \cdot (1 - 0,75) +$$

$$+ 0,75 \cdot (1 - 0,6) \cdot (1 - 0,7) = 0,205.$$

1.6. Формула полной вероятности

Пусть событие A может произойти с одним из несовместных событий H_1, H_2, ..., H_n, образующих полную группу событий, называемых гипотезами. Предположим, что известны вероятности гипотез

$$P(H_1), P(H_2), ..., P(H_n)$$

и условные вероятности

$$P(A/H_1), P(A/H_2), ..., P(A/H_n).$$

Требуется найти $P(A)$. В нашем случае $H_1 + H_2 + ... + H_n = \Omega$ есть событие достоверное. Можем записать, что

$$A = A\,\Omega = A\,(H_1 + H_2 + ... + H_n) = AH_1 + AH_2 + ... + AH_n.$$

Так как события H_i несовместны, то несовместны и события AH_i, поэтому по правилу сложения вероятностей несовместных событий и правилу умножения вероятностей, из последнего соотношения получаем

$$P(A) = \sum_{i=1}^{n} P(A \cdot H_i) = \sum_{i=1}^{n} P(H_i) P(A / H_i). \qquad (1.6.1)$$

Полученное выражение называется ***формулой полной вероятности***.

***ПРИМЕР** 1. Радиолампа может принадлежать к одной из трех партий с вероятностями $p_1 = 0,25$; $p_2 = 0,35$ и $p_3 = 0,40$. Вероятности того, что лампа проработает заданное число часов для этих партий, равны соответственно 0,1; 0,2 и 0,3. Определить вероятность того, что случайно взятая лампа проработает заданное число часов.

Решение. Введем обозначения событий: A – лампа проработает заданное число часов; события H_1, H_2, H_3 – лампа принадлежит соответственно первой, второй или третьей партии. По условию задачи

$$P(H_1) = p_1 = 0{,}25; \quad P(H_2) = p_2 = 0{,}35; \quad P(H_3) = p_3 = 0{,}40;$$

$$P(A/H_1) = 0{,}1; \quad P(A/H_2) = 0{,}2; \quad P(A/H_3) = 0{,}3.$$

По формуле полной вероятности находим

$$P(A) = P(H_1)\cdot P(A/H_1) + P(H_2)\cdot P(A/H_2) + P(H_3)\cdot (P(A/H_3) =$$
$$= 0{,}25\cdot 0{,}1 + 0{,}35\cdot 0{,}2 + 0{,}40\cdot 0{,}3 = 0{,}215.$$

1.7. Формула Байеса

Пусть имеет место равенство: $A = AH_1 + AH_2 + \ldots + AH_n$, где H_1, H_2, ..., H_n – полная группа несовместных событий. Предположим , что произведен эксперимент, в результате которого событие A наступило. Эта дополнительная информация позволяет произвести переоценку вероятностей гипотез H_i, вычислив $P(H_i/A)$.

По теореме умножения вероятностей

$$P(A\,H_i) = P(A)\cdot P(H_i/A) = P(H_i)\cdot P(A/H_i).$$

Откуда

$$P(H_i/A) = \frac{P(H_i)\cdot P(A/H_i)}{P(A)},$$

или, вычислив $P(A)$ по формуле полной вероятности (1.6.1), получим

$$P(H_i/A) = \frac{P(H_i)\cdot P(A/H_i)}{\sum_{k=1}^{n} P(H_k)\cdot P(A/H_k)}. \qquad (1.7.1)$$

Эту формулу обычно называют ***формулой Байеса***. Она дает выражение после опытных вероятностей через до опытные вероятности. Так как H_1, H_2, ..., H_n называют гипотезами, поэтому найденную формулу еще называют ***формулой переоценки гипотез***.

*ПРИМЕР 1. 40 % приборов собирается из высококачественных деталей, остальные – из деталей обычного качества. В первом случае надежность прибора (вероятность безотказной работы за время T) равна 0,90, а если прибор собрали из обычных деталей, его надежность 0,60. Прибор в течение времени T работал безотказно. Чему равна вероятность того, что он собран из высококачественных деталей?

Решение. Возможны две гипотезы: H_1 – прибор собран из высококачественных деталей; H_2 – прибор собран из обычных деталей. Вероятность того, что прибор собран из высококачественных деталей равна $P(H_1) = 0{,}4$ и из обычных – $P(H_2) = 0{,}6$.

Пусть A – событие, состоящее в том, что прибор безотказно работал в течение времени T. Тогда вероятность того, что прибор безотказно проработавший в течение времени T, был собран из высококачественных деталей равна $P(A/H_1) = 0{,}9$ и из обычных – $P(A/H_2) = 0{,}6$. По условию задачи прибор в течение времени T работал безотказно.

Нас интересует вероятность $P(H_1/A)$, что этот прибор собран из высококачественных деталей. По формуле Байеса получаем

$$P(H_1 / A) = P(H_1) \cdot P(A / H_1)/\{P(H_1) \cdot P(A / H_1) + P(H_2) \cdot P(A / H_2)\} =$$
$$= 0{,}4 \cdot 0{,}9 / (0{,}4 \cdot 0{,}9 + 0{,}6 \cdot 0{,}6\ ) = 0{,}5.$$

*ПРИМЕР 2. На предприятии изготавливаются изделия определенного вида на трех машинах. Первая машина изготавливает 25 %, вторая 35 %, третья – 40 % всех изделий. В их продукции брак составляет соответственно 5 %, 4 %, 2 %.

а) Какова вероятность того, что случайно выбранное изделие окажется бракованным?

б) Случайно выбранное изделие оказалось бракованным.

Какова вероятность того, что оно было сделано первой, второй, третьей машиной?

Решение. а) Интересующее нас событие обозначим A – выбор дефектного изделия. Заметим, что классическим определением вероятности нельзя воспользоваться, так как число благоприятствующих исходов нельзя найти, потому что не знаем, на какой именно машине было изготовлено изделие. Относительно этого можно сделать три предположения (выдвинуть гипотезы):

H_1 – случайно выбранное изделие изготовлено на 1-й машине.

H_2 – случайно выбранное изделие изготовлено на 2-й машине.

H_3 – случайно выбранное изделие изготовлено на 3-й машине.

Эти события являются попарно не совместными и образуют полную группу событий, так как одно из них обязательно происходит. Следовательно, эти события удовлетворяют условиям, налагаемым на гипотезы.

Ответ на вопрос: «какова вероятность того, что случайно выбранное изделие окажется бракованным?» будем искать по формуле полной вероятности:

$$P(A)=\sum_{i=1}^{3}P(A/H_i)P(H_i)=$$
$$P(A/H_1)P(H_1)+P(A/H_2)P(H_2)+P(A/H_3)P(H_3),$$

где $P(A/H_1)=\dfrac{5}{100}$ – вероятность того, что случайно выбранное изделие окажется бракованным, если оно изготовлено на первой машине;

$P(A/H_2)=\dfrac{4}{100}$ – вероятность того, что случайно выбранное изделие окажется бракованным, если оно изготовлено на второй машине;

$P(A/H_3)=\dfrac{2}{100}$ – вероятность того, что случайно выбранное изделие окажется бракованным, если оно изготовлено на третьей машине;

$P(H_1)=\dfrac{25}{100}$ – вероятность того, что изделие изготовлено на первой машине (до опыта);

$P(H_2)=\dfrac{35}{100}$ – вероятность того, что изделие изготовлено на второй машин;

$P(H_3)=\dfrac{40}{100}$ – вероятность того, что изделие изготовлено на третьей машине.

$$P(A) = 0{,}05 \cdot 0{,}25 + 0{,}04 \cdot 0{,}35 + 0{,}02 \cdot 0{,}4 \approx 0{,}0345.$$

Ответим на второй вопрос задачи. Для этого требуется «переоценить» гипотезы, т. е. найти вероятности гипотез после того, как событие A уже произошло. В этом случае вероятность вычисляется по формуле Байеса. Найдем вероятность того, что бракованное изделие произведено на первой машине (после опыта):

$$P(H_1/A)=\frac{P(H_1)\cdot P(A/H_1)}{P(A)}=\frac{0{,}25\cdot 0{,}05}{0{,}0345}=\frac{125}{345}.$$

Вероятность того, что бракованное изделие произведено на второй машине:

$$P(H_2/A)=\frac{P(H_2)\cdot P(A/H_2)}{P(A)}=\frac{0{,}35\cdot 0{,}04}{0{,}0345}=\frac{140}{345}.$$

Вероятность того, что бракованное изделие произведено на третьей машине:

$$P(H_3/A)=\frac{P(H_3)\cdot P(A/H_3)}{P(A)}=\frac{0{,}02\cdot 0{,}4}{0{,}0345}=\frac{80}{345}.$$

1.8. Формула Бернулли и приближенная формула Пуассона. Наивероятнейшее число наступления событий

Часто встречаются задачи, в которых один и тот же опыт повторяется неоднократно. В результате каждого опыта может появиться или не появиться некоторое событие A с вероятностью $P(A) = p$. Нас будет интересовать не результат отдельного опыта, а общее число наступлений события A в серии n опытов (испытаний).

Например, при проверке партии радиоламп важно общее количество годных радиоламп, а не результат проверки каждой радиолампы в отдельности. Такую схему испытаний принято называть схемой Бернулли, при этом должны выполнены условия: 1) опыты независимы, т. е. результат каждого опыта не оказывает влияния на другие; 2) вероятность $P(A) = p$ наступления события A в каждом опыте одна и та же.

Тогда вероятность того , что событие A появится ровно m раз в n испытаниях, выражается формулой Бернулли:

$$P_n(m) = C_n^m\, p^m\, q^{n-m}, \tag{1.8.1}$$

где $$q = 1-p,\ C_n^m = \frac{n!}{m!(n-m)!}=\frac{n(n-1)\ldots(n-m+1)}{m!}. \tag{1.8.2}$$

В связи с тем, что вероятности $P_n(m)$ по форме представляют собой члены разложения бинома $(q + p)^n$, распределение вероятностей по схеме Бернулли называется *биномиальным распределением*.

Число $\boldsymbol{m_0}$ называется ***наивероятнейшим числом наступления события A*** в $\boldsymbol{n}$ испытаниях, если значение $P_n(m)$ при $m = m_0$ достигает наибольшей величины, т. е.

$$P_n(m_0) \geq P_n(m_i).$$

Если $p \neq 0$ и $p \neq 1$, то число m_0 можно определить из двойного неравенства

$$np - q \le m_0 \le np + p. \qquad (1.8.3)$$

Разность граничных значений в этом двойном неравенстве равна 1. Если $np + p$ не является целым числом, то двойное неравенство определяет лишь одно наивероятнейшее значение m_0. Если же $np + p$ – целое число, то имеется два наивероятнейших значения:

$$m'_0 = np - q \text{ и } m''_0 = np + p. \qquad (1.8.4)$$

*ПРИМЕР 1. Что вероятнее, выиграть у равносильного противника (ничейный исход партии исключен): три партии из четырех или пять из восьми?

Решение. Так как противники равносильные, то вероятности выигрыша и проигрыша каждой партии одинаковы и равны $p = q = 1/2$.

Вероятность выиграть три партии из четырех определим по формуле Бернулли:

$$P_4(3) = C_4^3\, p^3 q^1 = C_4^1 \cdot \frac{1}{2^4} = \frac{1}{4}.$$

Аналогично считаем вероятность выиграть пять партий из восьми:

$$P_8(5) = C_8^5 p^5 q^3 = C_8^3 \cdot \frac{1}{2^8} = \frac{7}{32}.$$

Так как $\frac{1}{4} > \frac{7}{32}$, то вероятнее выиграть три партии из четырех.

*ПРИМЕР 2. Найти наивероятнейшее число годных деталей среди 19 проверяемых, если вероятность детали быть годной равна 0,9.

Решение. В нашем случае $p = 0{,}9$, $q = 0{,}1$, $n = 19$. Число $np - q = 0{,}9 \cdot 19 - 0{,}1 = 17{,}1 - 0{,}1 = 17$ оказалось целым. Поэтому наивероятнейшее число годных деталей равно 17 или 18.

В случае, когда n велико, а p мало (обычно $p < 0{,}1$; $npq \le 9$), вместо сложной формулы Бернулли применяют приближенную **формулу Пуассона**:

$$P_n(m) \approx \frac{a^m e^{-a}}{m!}, \; a = np.$$

Формула Пуассона используется в задачах, относящихся к редким событиям.

1.9. Локальная и интегральная теоремы Муавра – Лапласа

При больших значениях n и m вычисление вероятностей $P_n(m)$ по формуле Бернулли представляет значительные трудности, так как возникают факториалы больших чисел и большие степени p и q. Для таких случаев найдены приближенные формулы, позволяющие с достаточной точностью определить эти вероятности.

1.9.1. Локальная теорема Лапласа

Вероятность того, что в n независимых испытаниях, в каждом из которых вероятность появления события равна p $(0 \le p \le 1)$, событие наступит ровно m раз, приближенно равна (тем точнее, чем больше n):

$$P_n(m) \cong \frac{1}{\sqrt{npq}}\varphi(x), \qquad (1.9.1)$$

где

$$(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}; \; x = \frac{m - np}{\sqrt{npq}}. \qquad (1.9.2)$$

Функция $\varphi(x)$ табулирована и таблица приведена в данном пособии (см. прил. 2). Заметим, что $\varphi(x) = \varphi(-x)$.

Когда требуется найти вероятность того, что число m наступления события A содержится в заданных пределах $m_1 \le m \le m_2$, при небольших n, можно воспользоваться формулой Бернулли:

$$P_n(m_1 \le m \le m_2) = \sum_{m=m_1}^{m_2} P_n(m). \qquad (1.9.3)$$

При больших n вероятность $P_n(m_1 \le m \le m_2)$ можно приблизительно вычислить, пользуясь следующей теоремой.

1.9.2. Интегральная теорема Муавра – Лапласа

Вероятность того, что в n независимых испытаниях, в каждом из которых вероятность появления события равна p $(0 \le p \le 1)$,событие наступит не менее m_1 раз и не более m_2 раз, приблизительно равна

$$P_n(m_1 \le m \le m_2) \cong \Phi(x'') - \Phi(x'), \qquad (1.9.4)$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ – функция Лапласа;

$$x' = \frac{m_1 - np}{\sqrt{npq}}; \quad x'' = \frac{m_2 - np}{\sqrt{npq}}. \tag{1.9.5}$$

Для функции Лапласа имеются таблицы, которые приведены в данном пособии (см. прил. 3). Для отрицательных значений x используют эту же таблицу, учитывая, что функция Лапласа нечетная: $\Phi(-x) = -\Phi(x)$.

*ПРИМЕР 1. Найти вероятность того, что событие A наступит ровно 70 раз в 243 испытаниях, если вероятность появления этого события в каждом испытании равна 0,25.

Решение. По условию $n = 243$; $m = 70$; $p = 0{,}25$; $q = 0{,}75$. Так как $n = 243$ – достаточно большое число, воспользуемся локальной теоремой Лапласа (1.9.1):

$$P_n(m) \cong \frac{1}{\sqrt{npq}} \varphi(x).$$

Найдем значение x:

$$x = \frac{m - np}{\sqrt{npq}} = \frac{70 - 243 \cdot 0{,}25}{\sqrt{243 \cdot 0{,}25 \cdot 0{,}75}} = \frac{9{,}25}{6{,}75} = 1{,}37.$$

По таблицам находим $\varphi(1{,}37) = 0{,}1561$. Искомая вероятность

$$P_{243}(70) = \frac{0{,}1561}{6{,}75} = 0{,}0231.$$

*ПРИМЕР 2. Вероятность появления события в каждом из 100 независимых испытаний постоянна и равна $p = 0{,}8$. Найти вероятность того, что событие появится не менее 75 раз и не более 90 раз.

Решение. Воспользуемся интегральной теоремой Лапласа (1.9.4):

$$P_n(m_1 \le m \le m_2) = \sum_{m=75}^{90} P(m) \cong \Phi(x'') - \Phi(x').$$

По условию $n = 100$; $p = 0{,}8$; $q = 0{,}2$; $m_1 = 75$; $m_2 = 90$. Вычислим x' и x'':

$$x' = \frac{m_1 - np}{\sqrt{npq}} = \frac{75 - 100 \cdot 0{,}8}{\sqrt{100 \cdot 0{,}8 \cdot 0{,}2}} = -1{,}25;$$

$$x'' = \frac{m_2 - np}{\sqrt{npq}} = \frac{90 - 100 \cdot 0{,}8}{\sqrt{100 \cdot 0{,}8 \cdot 0{,}2}} = 2{,}5.$$

Учитывая, что функция Лапласа нечетная, т. е. $\Phi(-x) = -\Phi(x)$, получим

$$P_{100}(75; 90) = \Phi(2{,}5) - \Phi(-1{,}25) = \Phi(2{,}5) + \Phi(1{,}25).$$

По таблице найдем

$$\Phi(2{,}5) = 0{,}4938; \qquad \Phi(1{,}25) = 0{,}3944.$$

Искомая вероятность: $P_{100}(75; 90) = 0{,}4938 + 0{,}3944 = 0{,}8882$.

1.10. Случайные величины, законы их распределения

Случайной величиной X называется величина, которая в результате опыта может принять то или иное значение x_i. Выпадение некоторого значения случайной величины X – есть случайное событие: $X = x_i$.

Среди случайных величин выделяют прерывные (дискретные) и непрерывные случайные величины. Например, номер повстречавшегося автомобиля, рост случайного попутчика.

1.10.1. Ряд, многоугольник и функция распределения дискретной случайной величины

Дискретной называют случайную величину, которая может принимать отдельные, изолированные значения с определенными вероятностями.

*ПРИМЕР 1. Число появлений герба при трех бросаниях монеты. Возможные значения: 0, 1, 2, 3, их вероятности равны соответственно:

$$P(0) = \frac{1}{2^3} = \frac{1}{8}; \qquad P(1) = \frac{3}{8}; \quad P(2) = \frac{3}{8}; \quad P(3) = \frac{1}{8}.$$

*ПРИМЕР 2. Число отказавших элементов в приборе, состоящем из пяти элементов. Возможные значения: 0, 1, 2, 3, 4, 5; их вероятности зависят от надежности каждого из элементов.

Дискретная случайная величина X может быть задана рядом распределения или функцией распределения (интегральным законом распределения).

Рядом распределения называется совокупность всех возможных значений x_i и соответствующих им вероятностей $p_i = P\ (X = x_i)$, он может быть задан в виде таблицы:

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

При этом вероятности p_i удовлетворяют условию

$$\sum_{i=1}^{n} p_i = 1, \text{ потому что } \sum_{i=1}^{n} (X = x_i) = \Omega,$$

где число возможных значений n может быть конечным или бесконечным.

Графическое изображение ряда распределения ***называется многоугольником распределения***. Для его построения возможные значения случайной величины (x_i) откладываются по оси абсцисс, а вероятности p_i – по оси ординат; точки A_i с координатами (x_i, p_i) соединяются ломаными линиями.

Функцией распределения случайной величины X называется функция $F(x)$, значение которой в точке x равно вероятности того, что случайная величина X будет меньше этого значения x, т. е.

$$F(x) = P\,(X < x).$$

Функция $\boldsymbol{F(x)}$ для **дискретной случайной величины** вычисляется по формуле

$$F(x) = \sum_{x_i < x} p_i\,, \qquad (1.10.1)$$

где суммирование ведется по всем значениям i, для которых $x_i < x$.

*ПРИМЕР 3. Из партии, содержащей 100 изделий, среди которых имеется 10 дефектных, выбраны случайным образом пять изделий для проверки их качества. Построить ряд распределений случайного числа X дефектных изделий, содержащихся в выборке.

Решение. Так как в выборке число дефектных изделий может быть любым целым числом в пределах от 0 до 5 включительно, то возможные значения x_i случайной величины X равны:

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = 2,$$

$$x_4 = 3, \quad x_5 = 4, \quad x_6 = 5.$$

Вероятность $P(X = k)$ того, что в выборке окажется ровно k (k = 0, 1, 2, 3, 4, 5) дефектных изделий , равна

$$P(X = k) = \frac{C_{10}^{k} \cdot C_{90}^{5-k}}{C_{100}^{5}}.$$

В результате расчетов по данной формуле с точностью 0,001 получим

$$p_1 = P(X=0) \cong 0{,}583;\ p_2 = P(X=1) \cong 0{,}340;\ p_3 = P(X=2) \cong 0{,}070;$$

$$p_4 = P(X=3) \cong 0{,}007;\quad p_5 = P(X=4) \cong 0;\quad p_6 = P(X=5) \cong 0.$$

Используя для проверки равенство $\sum_{k=1}^{6} p_k = 1$, убеждаемся, что расчеты и округление произведены правильно (см. табл.).

x_i	0	1	2	3	4	5
p_i	0,583	0,340	0,070	0,007	0	0

*ПРИМЕР 4. Дан ряд распределения случайной величины X:

x_i	10	20	30	40	50
p_i	0,2	0,3	0,35	0,1	0,05

Найти функцию распределения вероятности $F(x)$ этой случайной величины и построить ее.

Решение. Если $x \le 10$, то $F(x) = P(X < x) = 0$;
если $10 < x \le 20$, то $F(x) = P(X < x) = 0{,}2$;
если $20 < x \le 30$, то $F(x) = P(X < x) = 0{,}2 + 0{,}3 = 0{,}5$;
если $30 < x \le 40$, то $F(x) = P(X < x) = 0{,}2 + 0{,}3 + 0{,}35 = 0{,}85$;
если $40 < x \le 50$, то $F(x) = P(X < x) = 0{,}2 + 0{,}3 + 0{,}35 + 0{,}1 = 0{,}95$;
если $x > 50$, то $F(x) = P(X < x) = 0{,}2 + 0{,}3 + 0{,}35 + 0{,}1 + 0{,}05 = 1$.

Построим многоугольник распределения и полученную функцию распределения вероятности (рис. 1.10.1).

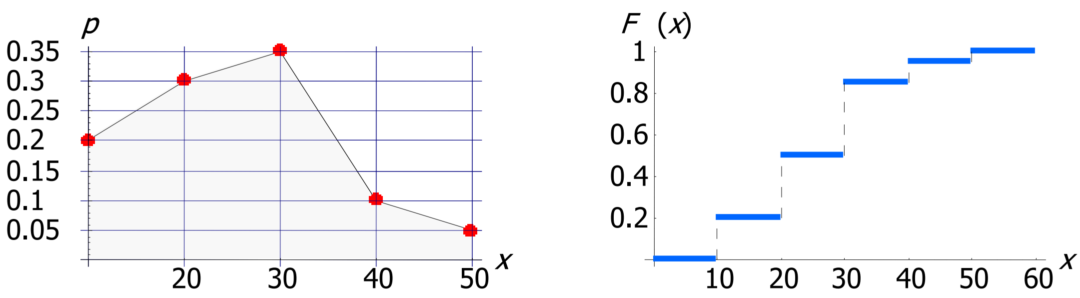


Рис. 1.10.1

1.10.2. Функция распределения и плотность вероятности непрерывной случайной величины

Непрерывной называют случайную величину, возможные значения которой непрерывно заполняют некоторые конечные или бесконечные промежутки.

*ПРИМЕР 5. Момент выхода прибора из строя.

*ПРИМЕР 6. Расстояние от точки попадания до центра мишени.

Случайная величина называется непрерывной, если существует неотрицательная функция $f(x)$, удовлетворяющая при любых x равенству

$$F(x) = \int_{-\infty}^{X} f(x)dx. \qquad (1.10.2)$$

Функция $f(x)$ называется ***плотностью вероятности***:

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x)}{\Delta x}. \qquad (1.10.3)$$

Непрерывная случайная величина задается либо ***функцией распределения F*(*x*)** (интегральным законом распределения), либо ***плотностью вероятности f*(*x*)** (дифференциальным законом распределения).

Функция распределения $F(x) = P(X < x)$, где x – произвольное действительное число, дает вероятность того, что случайная величина меньше x.

Функция распределения $F(x)$ имеет следующие свойства:

1) $P(a \le X < b) = F(b) - F(a)$; (1.10.4)

2) $F(x_1) \le F(x_2)$, если $x_1 < x_2$; (1.10.5)

3) $\lim\limits_{x \to +\infty} F(x) = 1$; (1.10.6)

4) $\lim\limits_{x \to -\infty} F(x) = 0$. (1.10.7)

Плотность вероятности $f(x)$ (дифференциальный закон распределения) обладает следующими основными свойствами:

1) $f(x) \ge 0$. (1.10.8)

Геометрически это означает, что вся кривая распределения лежит не ниже оси абсцисс;

2) $f(x) = \dfrac{dF(x)}{dx} = F'(x);$ (1.10.9)

3) $\int\limits_{-\infty}^{\infty} f(x)dx = 1.$ (1.10.10)

Геометрически это означает, что полная площадь, ограниченная кривой распределения и осью абсцисс равна единице;

4) $P(a \le X < b) = \int\limits_{a}^{b} f(x)dx.$ (1.10.11)

Геометрически вероятность попадания величины X на участок (a, b) равна площади кривой распределения, опирающейся на этот участок.

Функция распределения связана с плотностью следующими формулами:

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt, \quad (1.10.12)$$

$$f(x) = \frac{dF(x)}{dx} = F'(x). \quad (1.10.13)$$

*ПРИМЕР 7. Задана плотность распределения $f(x)$ случайной величины X:

$$f(x) = \begin{cases} A\cos x, & |x| \le \dfrac{\pi}{2}; \\ 0, & |x| > \dfrac{\pi}{2}. \end{cases}$$

Требуется:
1) найти коэффициент A;
2) построить график плотности распределения $f(x)$;
3) найти функцию распределения $F(x)$ и построить ее график;
4) найти вероятность попадания величины X на участок от 0 до $\dfrac{\pi}{4}$.

Решение

1. Для определения коэффициента A воспользуемся свойством (1.10.10) плотности распределения:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} A\cos x\,dx = 2A = 1,$$

откуда $A = \frac{1}{2}$.

2. График плотности $f(x)$ представлен на рис. 1.10.2.

3. По формуле (1.10.12) $F(x) = \int_{-\pi/2}^{x} f(t)dt$ получаем выражение функции распределения

$$F(x) = \begin{cases} 0, & x < -\frac{\pi}{2}; \\ \frac{1}{2}(\sin x + 1), & |x| \le \frac{\pi}{2}; \\ 1, & x > \frac{\pi}{2}. \end{cases}$$

График функции $F(x)$ изображен на рис. 1.10.3.

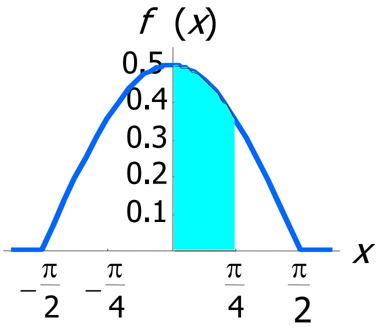


Рис. 1.10.2

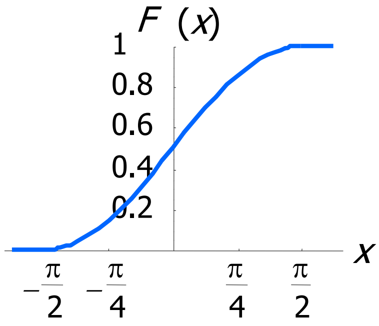


Рис. 1.10.3

4. По формуле (1.10.11) находим вероятность попадания величины X на участок от 0 до $\frac{\pi}{4}$:

$$P(a \le X \le b) = \int_a^b f(x)dx,$$

$$P(0 \le X < \frac{\pi}{4}) = \frac{1}{2}\int_0^{\frac{\pi}{4}} \cos x dx = \frac{1}{2}(\sin\frac{\pi}{4} - \sin 0) = \frac{\sqrt{2}}{4},$$

либо по формуле (1.10.4):

$$P(a \le X < b) = F(b) - F(a) = \frac{1}{2}(\sin\frac{\pi}{4} + 1) - \frac{1}{2}(\sin 0 + 1) = \frac{\sqrt{2}}{4}.$$

1.11. Основные характеристики (параметры распределения) случайной величины

Свойства случайной величины могут характеризоваться различными параметрами. Важнейшие из них – ***математическое ожидание*** случайной величины, которое обозначается через $\boldsymbol{M(X)}$, и ***дисперсия*** $\boldsymbol{D(X) = \sigma^2(X)}$, корень квадратный из которой $\boldsymbol{\sigma(X)}$ называют ***средне-квадратическим отклонением*** или стандартом.

1.11.1. Математическое ожидание

Математическим ожиданием M(X) (*средним по распределению*) **дискретной** (прерывной) случайной величины X называют сумму произведений всех возможных значений случайной величины на соответствующие им вероятности.

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i . \qquad (1.11.1)$$

Если выражение математического ожидания распространить на дискретные величины, имеющие бесконечное число возможных значений x_i, то

$$M(X) = \lim_{n\to\infty} \sum_{i=1}^{n} x_i p_i , \qquad (1.11.2)$$

причем предполагается, что этот ряд сходится абсолютно.

Учитывая, предыдущие записи, и что $\sum_{i=1}^{n} p_i = 1$, иногда пишут:

$$M(X) = \sum_{i=1}^{n} x_i p_i \Big/ \sum_{i=1}^{n} p_i . \qquad (1.11.3)$$

Эта запись позволяет дать ***механическую интерпретацию математического ожидания***: $M(X)$ – абсцисса центра тяжести системы точек, абсциссы которых равны возможным значениям случайной величины, а массы, помещенные в эти точки, равны соответствующим вероятностям.

Математическим ожиданием непрерывной случайной величины X называется интеграл:

$$M(X) = \int_{-\infty}^{\infty} x(x)dx, \tag{1.11.4}$$

причем предполагается, что интеграл сходится абсолютно; здесь $f(x)$ – плотность вероятности распределения случайной величины X.

Математическое ожидание $M(X)$ можно понимать как «теоретическое среднее значение случайной величины».

Рассмотрим ***свойства математического ожидания***:

1. Математическое ожидание имеет ту же размерность, что и сама случайная величина.

2. Математическое ожидание может быть как положительным, так и отрицательным числом.

3. Математическое ожидание постоянной величины C равно этой постоянной, т. е.

$$M(C) = C.$$

4. Математическое ожидание суммы нескольких случайных величин равно сумме математических ожиданий этих величин, т. е.

$$M(X + Y + ... + W) = M(X) + M(Y) + ... + M(W).$$

5. Математическое ожидание произведения двух или нескольких взаимно независимых случайных величин равно произведению математических ожиданий этих величин, т. е.

$$M(XY) = M(X) \cdot M(Y).$$

6. Математическое ожидание произведения случайной величины на постоянную C равно произведению математического ожидания случайной величины:

$$M(CX) = C \cdot M(X).$$

1.11.2. Дисперсия

Если математическое ожидание случайной величины дает нам «ее среднее значение» или точку на координатной прямой, «вокруг которой разбросаны» значения рассматриваемой случайной величины, то ***дисперсия характеризует «степень разброса» значений случайной величины около ее среднего.***

***Дисперсией* D(X)** случайной величины X называется математическое ожидание квадрата отклонения значения случайной величины от ее математического ожидания, т. е.

$$D(X) = M\{[X - M(X)]^2\}.$$

Дисперсию удобно вычислять по формуле

$$D(X) = M(X^2) - [M(X)]^2. \qquad (1.11.5)$$

Для ***дискретной*** случайной величины X формула (1.11.5) дает

$$D(X) = \sum_{i=1}^{n} (x_i)^2 p_i - \{M(X)\}^2. \qquad (1.11.6)$$

Для ***непрерывной*** случайной величины X:

$$D(X) = \int_{-\infty}^{\infty} \{x - M(X)\}^2 f(x) dx. \qquad (1.11.7)$$

Дисперсия имеет размерность, равную квадрату размерности случайной величины.

Рассмотрим ***свойства дисперсии***:

1. Дисперсия постоянной величины всегда равна нулю:

$$D(C) = 0.$$

2. Постоянный множитель можно выносить за знак дисперсии, предварительно возведя его в квадрат:

$$D(CX) = C^2 D(X).$$

3. Дисперсия алгебраической суммы двух ***независимых*** случайных величин равна сумме их дисперсией:

$$D(X + Y) = D(X) + D(Y).$$

Положительный корень из дисперсии называется ***среднеквадратичным отклонением*** и обозначается $\sigma = +\sqrt{D(X)}$.

Случайная величина называется ***центрированной***, если $M(X) = 0$.

Случайная величина называется ***стандартизированной***, если $M(X) = 0$ и $\sigma = 1$.

Среднее квадратическое отклонение имеет ту же размерность, что и случайная величина.

*ПРИМЕР 1. Найти математическое ожидание $M(X)$, дисперсию $D(X)$ и среднее квадратическое отклонение $\sigma(X)$ дискретной случайной величины X, заданной законом распределения

X	–5	2	3	4
p	0,4	0,3	0,1	0,2

Решение. Найдем математическое ожидание X по формуле (1.11.1):

$$M(X) = -5 \cdot 0{,}4 + 2 \cdot 0{,}3 + 3 \cdot 0{,}1 + 4 \cdot 0{,}2 = -0{,}3.$$

Дисперсию можно вычислить исходя, из ее определения, однако мы воспользуемся формулой (1.11.6)

$$D(X) = M(X^2) - [M(X)]^2,$$

которая быстрее ведет к цели.

Напишем закон распределения X^2:

X^2	25	4	9	16
p	0,4	0,3	0,1	0,2

Найдем математическое ожидание X^2:

$$M(X^2) = 25 \cdot 0{,}4 + 4 \cdot 0{,}3 + 9 \cdot 0{,}1 + 16 \cdot 0{,}2 = 15{,}3.$$

Найдем искомую дисперсию по формуле (1.11.6):

$$D(X) = M(X^2) - [M(X)]^2 = 15{,}3 - (-0{,}3)^2 = 15{,}21.$$

Тогда искомое среднее квадратическое отклонение будет:

$$\sigma(X) = \sqrt{D(X)} = \sqrt{15{,}21} = 3{,}9.$$

*ПРИМЕР 2. Найти дисперсию непрерывной случайной величины X, заданной функцией распределения

$$F(x) = \begin{cases} 0, & x < -2; \\ \dfrac{x}{4} + \dfrac{1}{2}, & |x| \le 2; \\ 1, & x > 2. \end{cases}$$

Решение. Найдем плотность распределения по формуле (1.10.13):

$$f(x) = F'(x) = \begin{cases} \frac{1}{4}, & |x| \le 2; \\ 0, & |x| > 2. \end{cases}$$

Функции показаны на рис. 1.10.5.

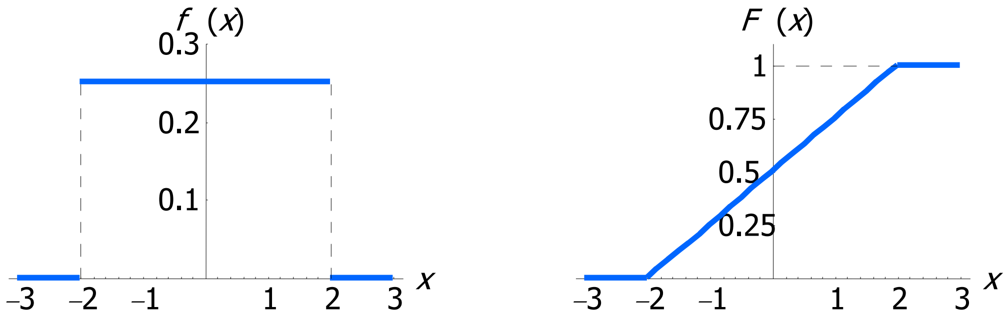


Рис. 1.10.5

Найдем математическое ожидание по формуле (1.11.4):

$$M(X) = \int_{-2}^{2} x\, f(x)dx = \int_{-2}^{2} x \frac{1}{4} dx = 0$$

(подынтегральная функция нечетная, пределы интегрирования симметричны относительно начала координат).

Найдем искомую дисперсию по формуле (1.11.7), учитывая, что $M(X) = 0$:

$$D(X) = \int_{-2}^{2} x - \{M(X)\}^2 f(x)dx = \int_{-2}^{2} x^2 \frac{1}{4} dx = \frac{2}{4}\int_{0}^{2} x^2 dx = \frac{4}{3}.$$

1.11.3. Начальные и центральные моменты

В общем случае свойства случайной величины могут характеризоваться различными начальными и центральными моментами.

Начальным моментом k-го порядка называется число α_k, определяемое формулой:

$$\alpha_k = M(X^k) = \begin{cases} \sum_i x_i^k p_i; \\ \int_{-\infty}^{+\infty} x^k f(x)dx, \end{cases} \tag{1.11.8}$$

где $M(X^k)$ – математическое ожидание k-й степени случайной величины X^k (соответственно, для случайных величин дискретного и непрерывного типа).

Центральным моментом k-го порядка называется число μ_k, определяемое формулой

$$\mu_k = M[(X-\alpha_1)^k] = \begin{cases} \sum_i (x_i - \alpha_1)^k p_i; \\ \int_{-\infty}^{+\infty} (x-\alpha_1)^k f(x)dx. \end{cases} \tag{1.11.9}$$

Из определений моментов, в частности, следует:

$$\begin{gathered} \alpha_0 = \mu_0 = 1; \\ \alpha_1 = M(X); \\ D(X) = \mu_2 = \alpha_2 - \alpha_1^2. \end{gathered} \tag{1.11.10}$$

Коэффициентом асимметрии **(или *скошенности*)** распределения называется величина

$$A = \frac{\mu_3}{\sigma^3}. \tag{1.11.11}$$

Коэффициентом эксцесса **(или *островершинности*)** распределения называется величина

$$E = \frac{\mu_4}{\sigma^4} - 3. \tag{1.11.12}$$

Квантилью порядка p распределения случайной величины X непрерывного типа называется действительное число t_p, удовлетворяющее уравнению

$$P(X < t_p) = p. \tag{1.11.13}$$

Критической точкой порядка p распределения случайной величины X непрерывного типа называется действительное число k_p, удовлетворяющее уравнению

$$P(X \geq k_p) = p. \tag{1.11.14}$$

Квантиль и критическая точка одного и того же распределения связаны соотношением:

$$k_p = t_{1-p}. \tag{1.11.15}$$

1.12. Некоторые частные распределения

В предыдущих пунктах 1.10–1.11 были рассмотрены общие формы законов распределения и числовые характеристики дискретных и непрерывных случайных величин. Рассмотрим некоторые важные для практики распределения случайных величин и соответствующие им числовые характеристики.

1.12.1. Биномиальное распределение

Дискретная случайная величина *X*, которая принимает значение с вероятностью (см. п. 1.8 и формулы (1.8.1) –(1.8.2))

$$P_n(X = m) = P_n(m) = C_n^m p^m q^{n-m}, \tag{1.12.1}$$

$$m = 0, 1, \ldots, n; \quad 0 \leq p \leq 1; \quad q = 1 - p$$

называется распределенной по *биномиальному закону*.

На практике биномиальное распределение возникает при следующих условиях. Пусть проводятся n независимых испытаний (опытов), в каждом из которых событие *A* («успех» опыта) может наступить с одной и той же вероятностью p, тогда число наступлений события *A* в n испытаниях и есть случайная величина *X*.

Ряд биномиального распределения случайной величины *X* будет иметь вид:

X	0	1	2	...	n
$P_n(m)$	q^n	npq^{n-1}	$\frac{n(n-1)}{1\cdot 2}p^2q^{n-2}$	...	p^n

Многоугольники биномиального распределения для частных случаев n = 20; p = 0,1; 0,5; 0,9 приведены на рис. 1.12.1.

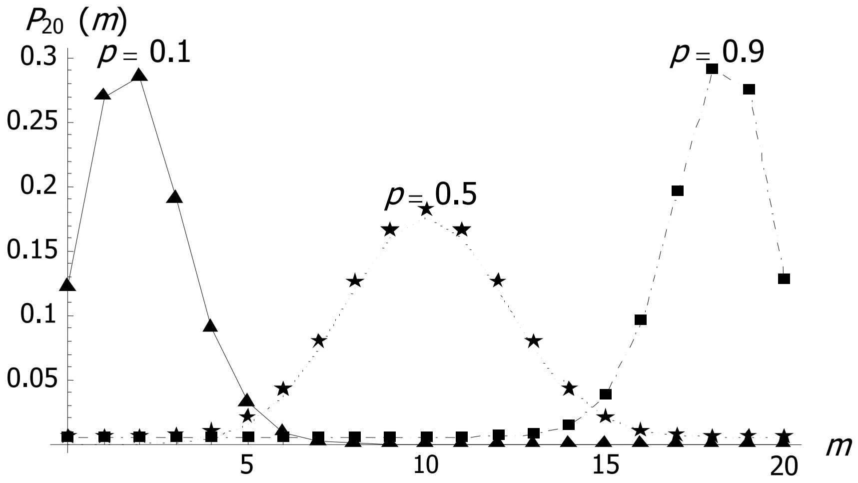


Рис. 1.12.1

Основные характеристики биномиального распределения, вычисленные по формулам (1.11.8)–(1.11.12), примут вид:

$$M(X) = np; \quad D(X) = npq; \quad A = \frac{q-p}{\sqrt{npq}}; \quad E = \frac{1-6pq}{npq}. \qquad (1.12.2)$$

*ПРИМЕР 1. Передается n = 5 сообщений по каналу связи. Каждое сообщение с вероятностью p = 0,3 независимо от других искажается. Случайная величина X – число искаженных сообщений. Построить ее ряд распределения, многоугольник распределения, найти ее числовые характеристики. Найти вероятность того, что будет искажено не менее двух сообщений.

Решение. Случайная величина X – число искаженных сообщений – распределена по биномиальному закону (под «опытом» разумеется передача сообщения, а под «успехом» – его искажение).

По формуле (1.12.1) находим

$P_5(0) = C_5^0 0{,}3^0 0{,}7^{5-0} = 1\cdot 1\cdot 0{,}7^5 = 0{,}16807;$

$P_5(1) = C_5^1 0{,}3^1 0{,}7^{5-1} = 5\cdot 0{,}3\cdot 0{,}2401 = 0{,}36015;$

$P_5(2) = C_5^2 0{,}3^2 0{,}7^{5-2} = 10\cdot 0{,}09\cdot 0{,}343 = 0{,}30870;$

$P_5(3) = C_5^3 0{,}3^3 0{,}7^{5-3} = 0{,}13230;$

$P_5(4) = C_5^4 0{,}3^4 0{,}7^{5-4} = 0{,}02835;$

$P_5(5) = C_5^5 0{,}3^5 0{,}7^{5-5} = 0{,}00243.$

Приближенно ряд распределения будет иметь вид:

$X = m$	0	1	2	3	4	5
$P_5(m)$	0,168	0,360	0,309	0,133	0,028	0,002

Изобразим многоугольник распределения (см. рис. 1.12.2).

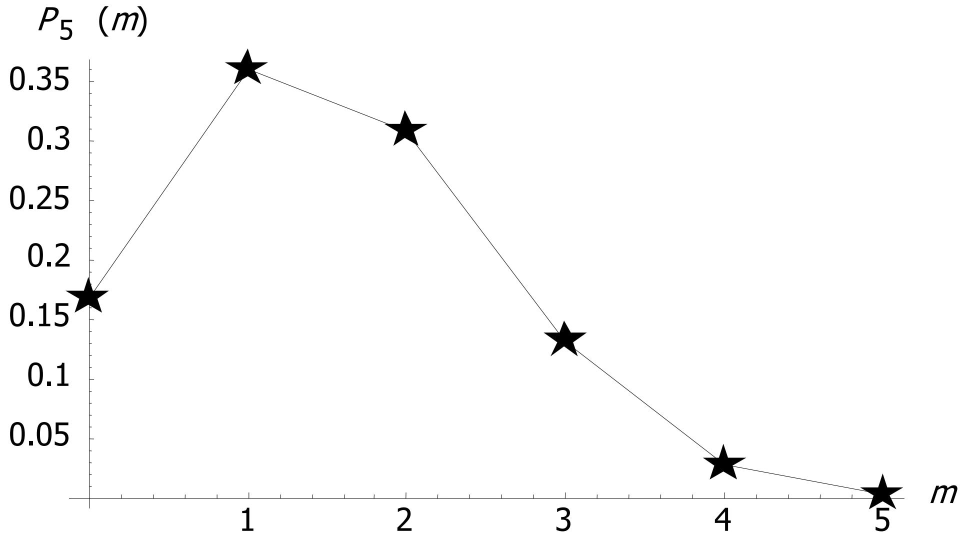


Рис. 1.12.2

Вычислим основные характеристики биномиального распределения по формулам (1.12.2):

$$M(X) = 1{,}5; \quad D(X) = 1{,}05; \quad A = 0{,}388; \quad E = –0.248.$$

В частности, коэффициент асимметрии $A = 0{,}388 > 0$ указывает на скошенность многоугольника распределения влево, а коэффициент эксцесса $E = –0{,}248 < 0$ подчеркивает туповершинность многоугольника распределения.

Наконец, найдем вероятность того, что будет искажено не менее двух сообщений:

$$P\{X \geq 2\} = 1 - P\{X < 2\} = 1 - (P_5(0) + P_5(1)) = 0{,}472.$$

1.12.2. Распределение Пуассона

Случайная величина X называется распределенной по закону Пуассона с параметром $a > 0$, если ее возможные значения равны 0, 1, …, а соответствующие вероятности определяются формулой

$$P_m \approx \frac{a^m e^{-a}}{m!}. \qquad (1.12.3)$$

Для распределения Пуассона характерно совпадение математического ожидания и дисперсии, причем

$$M(X) = D(X) = a.$$

Распределение Пуассона получается из биномиального распределения предельным переходом $n \to \infty$, $p \to 0$ при условии $np = a = \text{const}$ и в этом случае интерпретируется как *закон редких событий.* Этим распределением можно пользоваться приближенно, если производится большое число независимых опытов, в каждом из которых событие происходит с малой вероятностью. Пуассоновскому распределению подчиняется так называемый «поток событий», т. е. последовательность однородных событий, наступающих одно за другим в случайные моменты времени. Примеры: поток вызовов на телефонной станции, на пункт неотложной медицинской помощи, поток сбоев на ЭВМ, поток клиентов на предприятие бытового обслуживания. Поток называется *простейшим* (пуассоновским), если он обладает свойствами:

- *стационарности*, т. е. вероятность попадания того или иного числа событий на любой участок времени длиной τ не зависит от начала этого участка на оси времени, а зависит только от его длины τ;
- *ординарности*, т. е. события возникают поодиночке, а не группами по два, по три и т. д. или вероятность попадания двух и более событий на малый участок *времени* длиной τ пренебрежимо мала по сравнению с вероятностью попадания на него одного события;
- *отсутствия последствий*, т. е. «будущее» потока не зависит от его прошлого или вероятность попадания того или иного числа событий на заданный участок оси времени не зависит от того, сколько событий попало на любой другой не пересекающийся с ним участок.

Число событий X, попадающих на любой участок времени длиной τ, распределено по закону Пуассона. При этом, $\lambda = a/\tau$ определяет среднее число событий, происходящих за единицу времени, т. е. интенсивность потока.

Закон распределения Пуассона может быть записан в виде таблицы:

$X = m$	0	1	2	...	m	...
P_m	e^{-a}	ae^{-a}	$\dfrac{a^2e^{-a}}{2!}$	...	$\dfrac{a^me^{-a}}{m!}$	...

Многоугольники распределения для случайной величины, распределенной по закону Пуассона с параметрами a = 0,5; a = 2; a = 5 приведены на рис. 1.12.3.

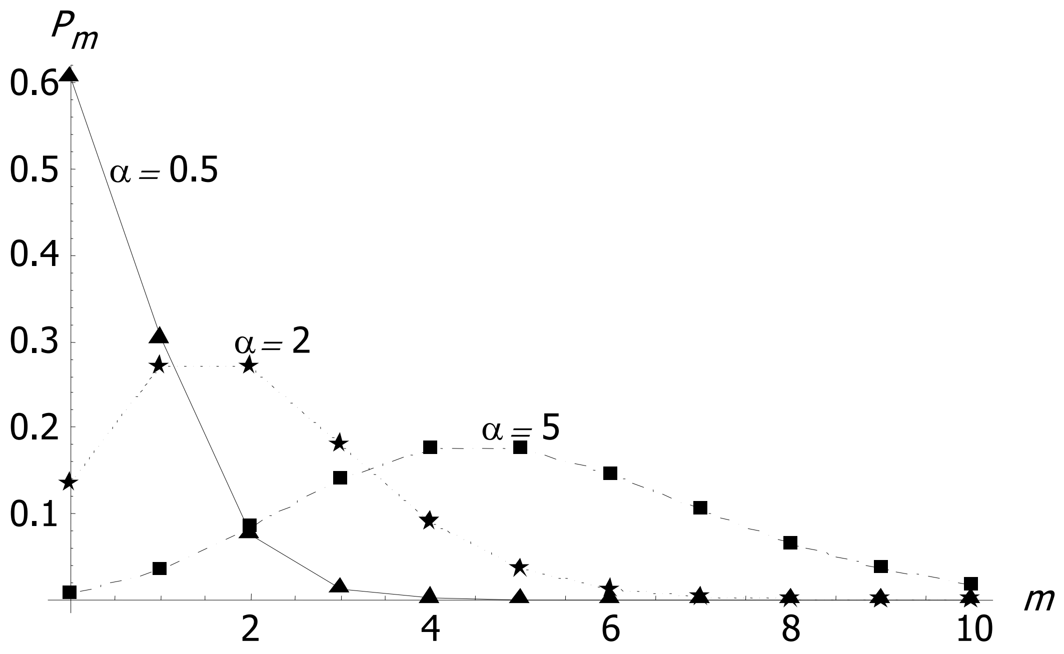


Рис. 1.12.3

∗ПРИМЕР 2. Среднее число вызовов, поступающих на АТС в одну минуту, равно 2. Найти вероятность того, что за 5 минут поступит 1) 2 вызова; 2) менее двух вызовов; 3) не менее двух вызовов. Поток вызовов предполагается простейшим.

Решение. По условию $\lambda = 2$, $\tau = 5$, $m = 2$. Таким образом, $a = 10$ и с учетом формулы (1.12.3) находим:

1) $P_2 \approx \dfrac{10^2 e^{-10}}{2!} \approx 0{,}00225$;

2) $P(m < 2) = P_0 + P_1 = \dfrac{10^0 e^{-10}}{0!} + \dfrac{10^1 e^{-10}}{1!} \approx 0{,}000495$;

3) $P(m \geq 2) = 1 - P(m < 2) \approx 1 - 0{,}000495 \approx 0{,}999505$.

Таким образом, если первые два события практически невозможны, то последнее – практически достоверно.

1.12.3. Геометрическое распределение

Дискретная случайная величина X, которая принимает значение m с вероятностью

$$P\,(X = m) = P_m = p\ q^m, \qquad (1.12.4)$$

$$m = 0, 1, \ldots, n; \quad 0 \leq p \leq 1;\ q = 1 - p,$$

называется распределенной по *геометрическому закону*.

Вероятности P_m для последовательных значений m образуют геометрическую прогрессию с первым членом p и знаменателем q.

На практике геометрическое распределение возникает при следующих условиях. Пусть проводятся ряд независимых испытаний (опытов) с целью получения какого – то результата («успеха») A; при каждой попытке «успех» достигается с одной и той же вероятностью p, тогда число «безуспешных» попыток (до первой попытки, в которой наступает результат A) и есть случайная величина X.

Ряд геометрического распределения случайной величины X будет иметь вид:

$X = m$	0	1	2	...	m	...
P_m	p	pq	pq^2	...	pq^m	...

Многоугольники распределения для случайной величины, распределенной по геометрическому закону, с параметрами p = 0,2; p = 0,5; p = 0,8 приведены на рис. 1.12.4.

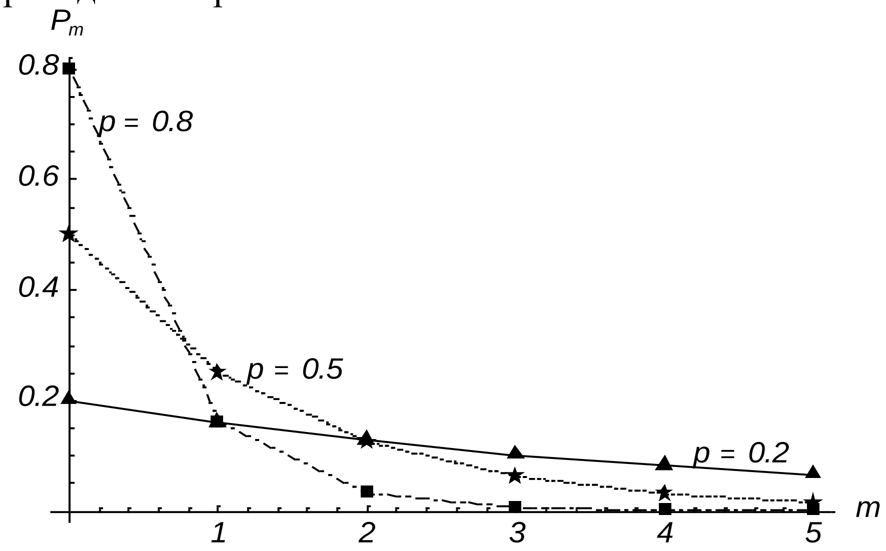


Рис. 1.12.4

Основные числовые характеристики геометрического распределения выражаются по формулам (1.11.2), (1.11.5) через параметр распределения в виде:

$$M(X) = \frac{1-p}{p}; \quad D(X) = \frac{1-p}{p^2}. \qquad (1.12.5)$$

На практике приходится рассматривать случайную величину $Y = X + 1$ – число попыток до первого «успеха», включая удавшуюся.

Ряд распределения случайной величины Y будет иметь вид:

$Y = m$	1	2	3	...	m	...
P_m	p	pq	pq^2	...	pq^{m-1}	...

Основные числовые характеристики распределения случайной величины Y будут иметь вид согласно формулам (1.11.2), (1.11.5):

$$M(Y) = \frac{1}{p}, \quad D(Y) = \frac{1-p}{p^2}. \tag{1.12.6}$$

*ПРИМЕР 3. Из орудия производится стрельба по цели до первого попадания. Вероятность попадания в цель p = 0,6. Найти вероятность того, что попадание произойдет при третьем выстреле.

Решение. В данном случае Y – число выстрелов до первого попадания. Искомая вероятность вычисляется по формуле

$$P\ (Y = m) = p\ \ q^{m-1} = 0{,}6 \cdot 0{,}4^{3-1} = 0{,}096.$$

1.12.4. Гипергеометрическое распределение

Случайная величина X имеет гипергеометрическое распределение с параметрами N = 2, 3, …; M = 1, 2, …< N; n = 1, 2, …< N , если она принимает конечное множество натуральных значений X = m, соответственно с вероятностями

$$P\{X = m\} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \tag{1.12.7}$$

где m = 0, 1, 2,…, min(M, n).

На языке «схемы урн», гипергеометрическое распределение возникает при следующих условиях. В урне находится N шаров, из которых ровно M белые. Пусть одновременно (или один за другим без возврата) извлечено n шаров. Тогда вероятность того, что среди этих извлеченных шаров будет m белых шаров, определяется по формуле (1.12.7).

На практике гипергеометрическое распределение применяется при решении задач, связанных с контролем продукции.

*ПРИМЕР 4. Из партии, содержащей N изделий, среди которых имеется M стандартных, выбраны случайным образом n изделий для проверки их качества. Построить ряд распределений случайного числа $X = m$ стандартных изделий, содержащихся в выборке.

Решение. Искомая вероятность вычисляется по формуле (1.12.7).

Основные числовые характеристики случайной величины, распределенной по закону (1.12.7), имеют вид

$$M(X)=\frac{nM}{N};\ D(X)=\frac{nM(N-M)(N-n)}{(N-1)N^2}. \qquad (1.12.8)$$

В данном случае интерес представляют также числовые характеристики, определяющие степень асимметричности многоугольника распределения:

$$A=\frac{\sqrt{-1+N}(-2M+N)(-2n+N)}{(-2+N)\sqrt{Mn(-M+N)(-n+N)}}$$

и степень островершинности многоугольника распределения:

$$E=\frac{-3+((-1+N)N^2(N(1+N)-6n((-n+N)+ \\ +\frac{1}{N^2}(3M(-M+N)(-n^2N+(-2+n)N^2+6n(-n+N)))}{(Mn(-3+N)(-2+N)(-M+N)(-n+N))}.$$

*ПРИМЕР 5. Имеется 7 радиоламп, среди которых 3 неисправные, на вид неотличимые от новых. Наугад берутся 4 радиолампы и вставляются в 4 патрона. Найти ряд распределения и построить многоугольник распределения числа радиоламп X, которые будут работать. Найти его числовые характеристики.

Решение. Величина X имеет гипергеометрическое распределение с параметрами $N = 7$; $n = 4$; $M = 4$. По формуле (1.12.7) находим:

$$P\{X=1\}=\frac{C_4^1C_3^3}{C_7^4}=4/35; \quad P\{X=2\}=\frac{C_4^2C_3^2}{C_7^4}=18/35;$$

$$P\{X=3\}=\frac{C_4^3C_3^1}{C_7^4}=12/35; \quad P\{X=4\}=\frac{C_4^4C_3^0}{C_7^4}=1/35.$$

Ряд распределения случайной величины X будет иметь приближенный вид:

$X = m$	1	2	3	4
P_m	0,114	0,514	0,343	0,029

Основные числовые характеристики случайной величины X вычислим по формулам (1.12.8): $M(X) \approx 2{,}286$; $D(X) \approx 0{,}490$; $A \approx 0{,}041$; $E \approx -0{,}283$.

В частности, коэффициент асимметрии $A = 0{,}041 > 0$ указывает на скошенность многоугольника распределения влево, а коэффициент эксцесса $E = -0{,}283 < 0$ подчеркивает туповершинность многоугольника распределения, который изображен на рис. 1.12.5.

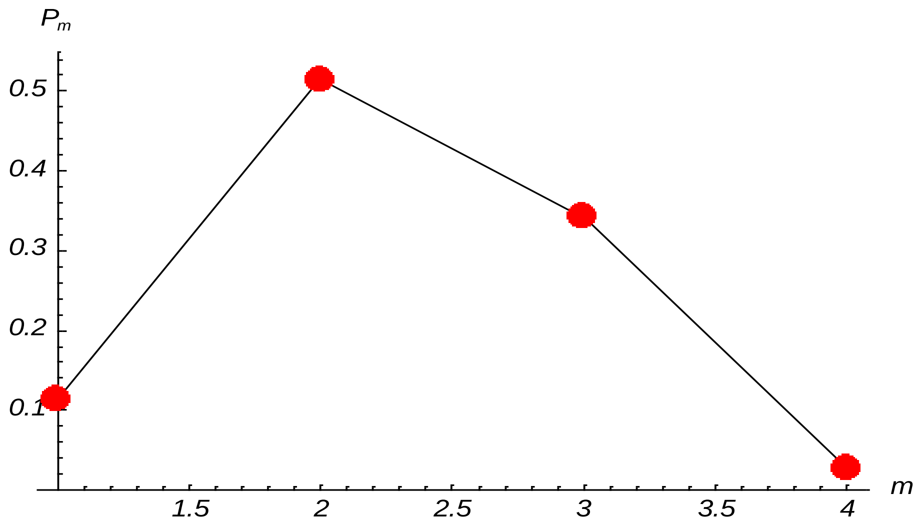


Рис. 1.12.5

1.12.5. Равномерное распределение

Непрерывная случайная величина X называется ***распределенной равномерно*** на отрезке $[a,b]$, если ее плотность распределения вероятностей постоянна на данном отрезке:

$$f(x) = \begin{cases} 0, & x \notin [a,b]; \\ \dfrac{1}{b-a}, & x \in [a,b]. \end{cases} \tag{1.12.9}$$

Функция распределения в этом случае согласно (1.10.2), примет вид:

$$F(x) = \begin{cases} 0, & x < a; \\ \dfrac{x-a}{b-a}, & a \le x < b; \\ 1, & x \ge b. \end{cases} \tag{1.12.10}$$

Это распределение реализуется, например, в эксперименте, в котором наудачу ставится точка на отрезок $[a,b]$, при этом случайная величина X – абсцисса поставленной точки.

Примером равномерно распределенной непрерывной случайной величины X является ошибка при округлении отсчета до ближайшего целого деления шкалы измерительного прибора, проградуированной в некоторых единицах.

Основные характеристики равномерного распределения:

$$M(X)=\frac{a+b}{2},\quad D(X)=\frac{(b-a)^2}{12},\quad A=0, E=-1{,}2. \qquad (1.12.11)$$

Как и следовало ожидать, график плотности равномерного распределения симметричен ($A = 0$) относительно $x = M(X)$ и туповершинен ($E < 0$).

*ПРИМЕР 6. См. ПРИМЕР 2. п.1.11.2.

1.12.6. Экспоненциальное (показательное) распределение

Случайная величина X распределена по *экспоненциальному закону с параметром* $\lambda>0$. Если ее плотность распределения вероятностей задается формулой:

$$f(x)=\begin{cases}0, & x<0;\\ \lambda e^{-\lambda x}, & x\ge 0.\end{cases} \qquad (1.12.12)$$

Функция распределения показательного закона:

$$F(x)=\begin{cases}0, & \text{при } x<0;\\ 1-e^{-\lambda x}, & \text{при } x\ge 0.\end{cases} \qquad (1.12.13)$$

Типичные примеры, где реализуется экспоненциальное распределение – теория обслуживания, при этом X – например, время ожидания при техническом обслуживании, и теория надежности, здесь X – например, срок службы радиоэлектронной аппаратуры.

Показательное распределение тесно связано с простейшим (пуассоновским) потоком событий (см. п. 1.12.2): интервал времени T между двумя соседними событиями в простейшем потоке имеет показательное распределение с параметром, равным интенсивности потока:

$$f(t)=\lambda e^{-\lambda t}\ (t>0).$$

Основные характеристики показательного распределения:

$$M(X)=\frac{1}{\lambda},\quad D(X)=\frac{1}{\lambda^2}. \qquad (1.12.14)$$

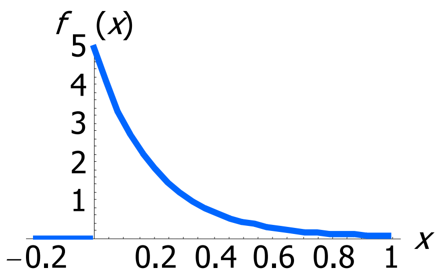
*ПРИМЕР 7. Время безотказной работы ЭВМ – случайная величина T, имеющая показательное распределение с параметром $\lambda = 5$ (физический смысл величины λ – среднее число отказов в единицу времени, не считая простоев ЭВМ для ремонта). Известно, что ЭВМ уже проработала без отказов время τ. Найти при этих условиях плотность и функцию распределения времени, которое проработает ЭВМ после момента τ до ближайшего отказа.

Решение. Так как простейший поток отказов не имеет последствия, вероятность появления хотя бы одного отказа на участке от τ до $\tau + t$ не зависит от того, появлялись ли отказы ранее момента τ. Следовательно, подставив $\lambda = 5$ в соотношение (1.12.12) и в (1.12.13), получим

$$f(x) = \begin{cases} 0, & x < 0; \\ 5e^{-5x}, & x \ge 0. \end{cases}$$

$$F(x) = \begin{cases} 0, & \text{при} \quad x < 0; \\ 1 - e^{-5x}, & \text{при} \quad x \ge 0. \end{cases}$$

Графики плотности и функции полученного показательного распределения изображены на рис. 1.12.6.



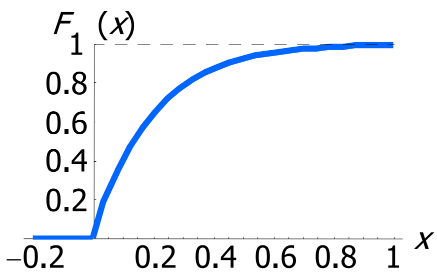


Рис. 1.12.6

1.12.7. Нормальное распределение ($N(a, \sigma)$)

Непрерывная случайная величина имеет *нормальное распределение с параметрами* $a \in R$ и $\sigma > 0$, если плотность распределения вероятностей имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \qquad (1.12.15)$$

где a – математическое ожидание; σ – среднее квадратическое отклонение X;

$$a = M[X], \qquad \sigma = +\sqrt{D[X]}. \tag{1.12.16}$$

Если случайная величина распределена по закону $N(0,1)$, то она называется *стандартизированной нормальной величиной*. Функция распределения для нее имеет вид:

$$F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt\,. \tag{1.12.17}$$

Графики плотности и функции нормального распределения изображены на рис. 1.12.7.

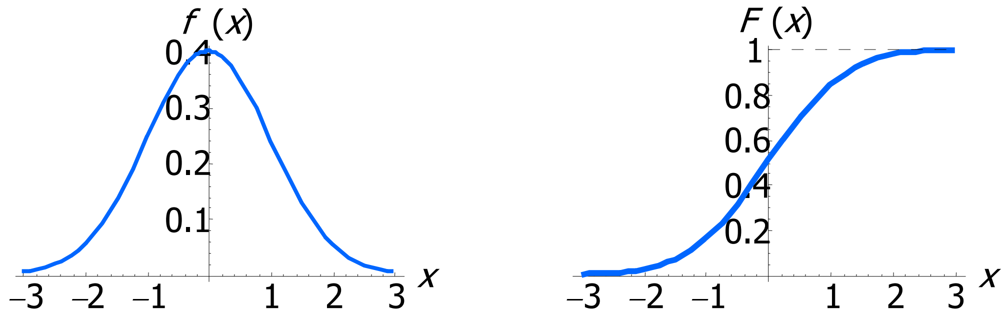


Рис. 1.12.7

Функция нормального распределения (1.12.17) связана с функцией Лапласа (см. п. 1.9.2), имеющей вид:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-t^2/2} dt\,,$$

соотношением $F_0(x) = \Phi(x) + 0{,}5$. Для функции Лапласа справедливо соотношение

$$\Phi(-x) = -\Phi(x).$$

С помощью функции Лапласа можно вычислять интервальные вероятности для нормального распределения $N(a,\sigma)$:

$$P\{x_1 \le X \le x_2\} = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right). \tag{1.12.18}$$

Значения функций $f(x)$ и $\Phi(x)$ можно вычислить с помощью таблиц (см. прил. 2, 3).

Через функцию Лапласа выражается и функция нормального распределения в общем случае $N(a,\sigma)$:

$$F(x) = \Phi\left(\frac{x-a}{\sigma}\right) + 0{,}5. \qquad (1.12.19)$$

Центральные моменты нормального распределения удовлетворяют рекуррентному соотношению

$$\mu_{n+2} = (n+1)\sigma^2\mu_n, \qquad n = 1,\ 2,\ \ldots, \qquad (1.12.20)$$

откуда, в частности, следует, что все центральные моменты нечетного порядка равны нулю, так как $\mu_1 = 0$ и, таким образом, согласно (1.11.11), $A = \dfrac{\mu_3}{\sigma^3} = 0$. С учетом (1.12.19), $\mu_4 = 3\sigma^2\mu_2 = 3\sigma^4$ и, согласно (1.11.12), имеем: $E = \dfrac{\mu_4}{\sigma^4} - 3 = 0$. В этом смысле кривая плотности нормального распределения является эталонной ($A = 0$, $E = 0$), с которой сравнивают $f(x)$ других распределений при одинаковых $M(X)$ и $D(X)$. Причем на фоне кривой плотности нормального распределения график плотности распределения $f(x)$ деформирован (асимметричен) влево, если $A > 0$, и вправо, если $A < 0$; остроконечен (вытянут вверх), если $E > 0$, и тупоконечен, если $E < 0$.

Нормальное распределение непрерывной случайной величины имеет очень широкое распространение в случайных явлениях природы, так как ему подчиняется распределение случайной величины, представленной в виде суммы слабо зависимых случайных величин, сравнимых по порядку их влияния на рассеивание суммы. На практике очень часто встречаются случайные величины, образующиеся именно в результате суммирования многих случайных слагаемых, сравнимых по степени своего влияния на рассеивание суммы. В частности, в очень многих случаях из практики – ошибки измерения, ошибки стрельбы, ошибки выполнения команд автоматизированных устройств, ошибки вывода космического корабля в заданную точку пространства, ошибки параметров элементов вычислительной техники, а также множество других «ошибок», сопровождающих целенаправленную деятельность человека, в очень большой доле случаев имеют распределение по закону, близкому к нормальному закону распределения.

*ПРИМЕР 8. Плотность распределения задана законом

$$f(x)=\frac{1}{\sqrt{7\pi}}e^{-\frac{(x+1)^2}{7}}.$$

Определить вид распределения найти функцию распределения, $M(X)$ и $D(X)$.

Решение. Сравнивая заданную плотность распределения с (1.12.15), делаем вывод, что у нас задан нормальный закон распределения с $a=-1$, значит, $M(X)=-1$ – математическое ожидание, $\sigma^2=D(X)=7/2$ – искомая дисперсия. Следовательно, функция данного нормального распределения определяется по формуле (1.12.19)

$$F(x) = \Phi\left(\frac{x+1}{\sqrt{7/2}}\right) + 0{,}5.$$

Графики плотности и функции нормального распределения с параметрами $a=-1$ и $\sigma=\sqrt{7/2}$ изображены на рис. 1.12.8.

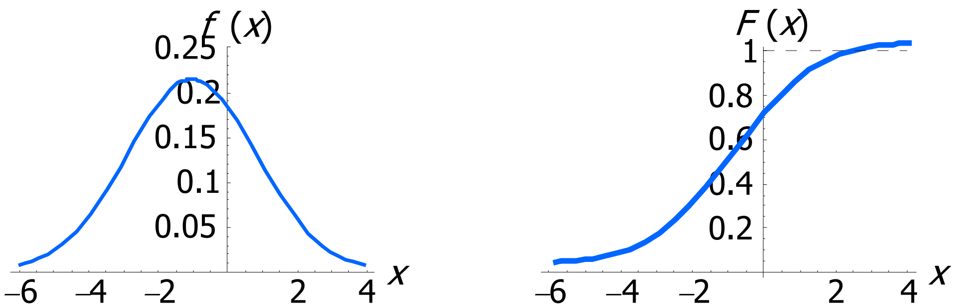


Рис. 1.12.8

*ПРИМЕР 9. Имеется случайная величина, распределенная по нормальному закону с параметрами a и σ. Найти вероятность того, что случайная величина отклонится от своего математического ожидания a больше, чем на 3σ.

Решение. С учетом (1.12.18) будем иметь

$$P\{|X-a|>3\sigma\}=1-P\{|X-a|<3\sigma\}=1-2\Phi\left(\frac{3\sigma}{\sigma}\right)=1-2\Phi(3),$$

где $\Phi(3)$ вычисляется по таблице функции Лапласа (прил. 3): $\Phi(3)\approx 0{,}49865$.

В итоге $P\{|X-a|>3\sigma\}=1-2\Phi(3)\approx 1-0{,}9973\approx 0{,}0027$.

1.13. Системы случайных величин

Функцией распределения системы двух случайных величин $\{X,Y\}$ называется неслучайная функция двух действительных переменных, определяемая как вероятность совместного выполнения двух неравенств:

$$F_{X,Y}(x, y) = P\{X < x, Y < y\} \tag{1.13.1}$$

и удовлетворяющая следующим свойствам:

1) $\lim\limits_{x\to-\infty} F_{X,Y}(x, y) = \lim\limits_{y\to-\infty} F_{X,Y}(x, y) = 0, \quad \lim\limits_{x,y\to+\infty} F_{X,Y}(x, y) = 1;$

2) $\lim\limits_{x\to+\infty} F_{X,Y}(x, y) = F_Y(y), \quad \lim\limits_{y\to+\infty} F_{X,Y}(x, y) = F_X(x);$ (1.13.2)

3) $F_{X,Y}(x, y)$ – неубывающая функция своих аргументов.

Систему двух случайных величин называют ***непрерывно распределенной***, если их функция распределения непрерывна на всей плоскости и существует такая неотрицательная интегрируемая функция $f_{X,Y}(x, y)$, называемая ***плотностью распределения вероятностей*** $\{X,Y\}$, что, согласно (1.13.1) – (1.13.2):

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} d\xi \int_{-\infty}^{y} f_{X,Y}(\xi,\eta)d\eta; \tag{1.13.3}$$

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}; \tag{1.13.4}$$

$$\int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} f_{X,Y}(\xi,\eta)d\eta = 1, \tag{1.13.5}$$

$$f_{X,Y}(x, y) \ge 0.$$

Плотности распределения вероятностей по каждой переменной выражаются в виде:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,\eta)d\eta, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(\xi, y)d\xi. \tag{1.13.6}$$

Тогда, согласно (1.13.2), (1.13.3):

$$F_X(x) = \int_{-\infty}^{x} f_X(\eta)d\eta, \qquad F_Y(y) = \int_{-\infty}^{y} f_Y(\eta)d\eta. \tag{1.13.7}$$

Для системы двух случайных величин {X,Y} вводятся следующие числовые характеристики.

Начальный момент порядка k + s:

$$\alpha_{k,s} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^k y^s f_{X,Y}(x,y)dxdy. \tag{1.13.8}$$

В частности, $(m_X, m_Y) = (\alpha_{1,0}, \alpha_{0,1})$ называется ***математическим ожиданием*** $\{X,Y\}$ или центром рассеивания.

Центральный момент порядка k + s:

$$\mu_{k,s} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-m_X)^k(y-m_Y)^s f_{X,Y}(x,y)dxdy. \qquad (1.13.9)$$

В частности, $\mu_{2,0} = D_X$, $\mu_{0,2} = D_Y$ – ***дисперсии***, $\mu_{1,1} = K_{XY}$ – ***ковариация (корреляционный момент)***. Нормированная ковариация

$$\rho_{XY} = \frac{K_{XY}}{\sigma_X\sigma_Y} \qquad (1.13.10)$$

называется ***коэффициентом корреляции*** системы двух случайных величин. Здесь $\sigma_X = \sqrt{D_X}$, $\sigma_Y = \sqrt{D_Y}$ – среднеквадратичные отклонения. Корреляционный коэффициент удовлетворяет условию $|\rho_{XY}| \le 1$ и определяет степень линейной зависимости между X и Y.

*ПРИМЕР 1. Найти функцию распределения двумерной случайной величины по данной плотности совместного распределения

$$f(x,y) = \frac{1}{\pi^2(1+x^2)(1+y^2)},$$

график которой изображен на рис. 1.13.1.

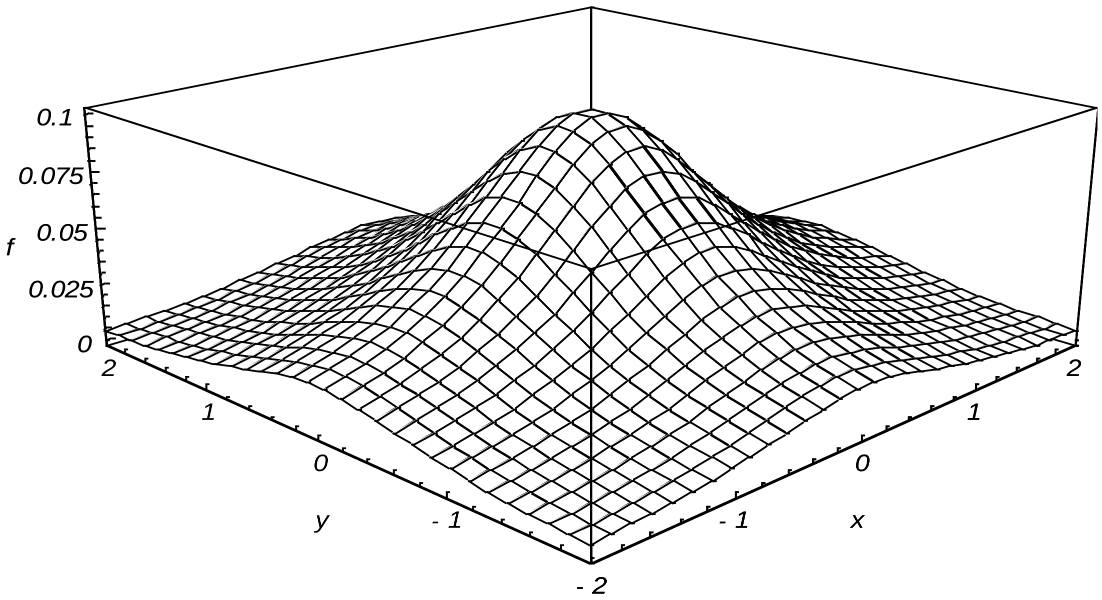


Рис. 1.13.1

Решение. Воспользуемся формулой (1.13.3):

$$F_{X,Y}(x,y)=\frac{1}{\pi^2}\int_{-\infty}^{y}\left(\frac{1}{1+y^2}\int_{-\infty}^{x}\frac{dx}{1+x^2}\right)dy=\frac{1}{\pi^2}\int_{-\infty}^{y}\frac{1}{1+y^2}\left(\operatorname{arctg}x+\frac{\pi}{2}\right)dy=$$

$$=\left(\frac{1}{\pi}\operatorname{arctg}x+\frac{1}{2}\right)\frac{1}{\pi}\int_{-\infty}^{y}\frac{dy}{1+y^2}=\left(\frac{1}{\pi}\operatorname{arctg}x+\frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg}y+\frac{1}{2}\right).$$

График построенной функции распределения изображен на рис. 1.13.2.

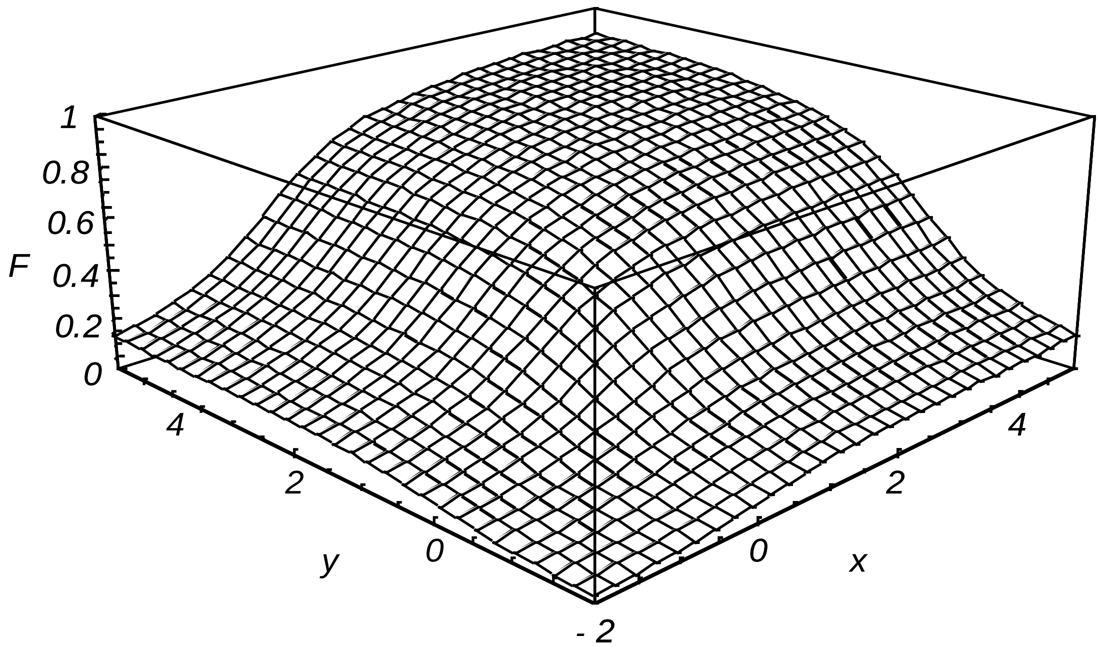


Рис. 1.13.2

*ПРИМЕР 2. Дана плотность вероятности системы случайных величин:

$$f(x,y)=\begin{cases}a\sin(x+y) & \text{при} \quad 0\le x\le \pi/2; \quad 0\le y\le \pi/2;\\ 0 & \text{при любых других } x \text{ и } y.\end{cases}$$

Определить коэффициент a, функцию распределения системы, математические ожидания и дисперсии системы случайных величин и их корреляционный момент. Найти одномерные законы распределения каждой из величин.

Решение. На основании свойства нормировки (1.13.5) определяем коэффициент a:

$$a = 1\Big/\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} \sin(x+y)dydx,$$

где $\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} \sin(x+y)dydx = -\int_0^{\frac{\pi}{2}} \cos(x+y)\Big|_0^{\frac{\pi}{2}} dx = -\int_0^{\frac{\pi}{2}}(\cos(x+\frac{\pi}{2}) - \cos x)dx = 2.$ Тогда a = 0,5, и плотность вероятности системы случайных величин $f(x, y)$ принимает конкретный вид:

$$f(x,y) = \begin{cases} 0{,}5\sin(x+y) & \text{при} \quad 0 \le x \le \pi/2; \quad 0 \le y \le \pi/2; \\ 0 & \text{при любых других } x \text{ и } y. \end{cases}$$

Строим график $f(x, y)$ (рис. 1.13.3).

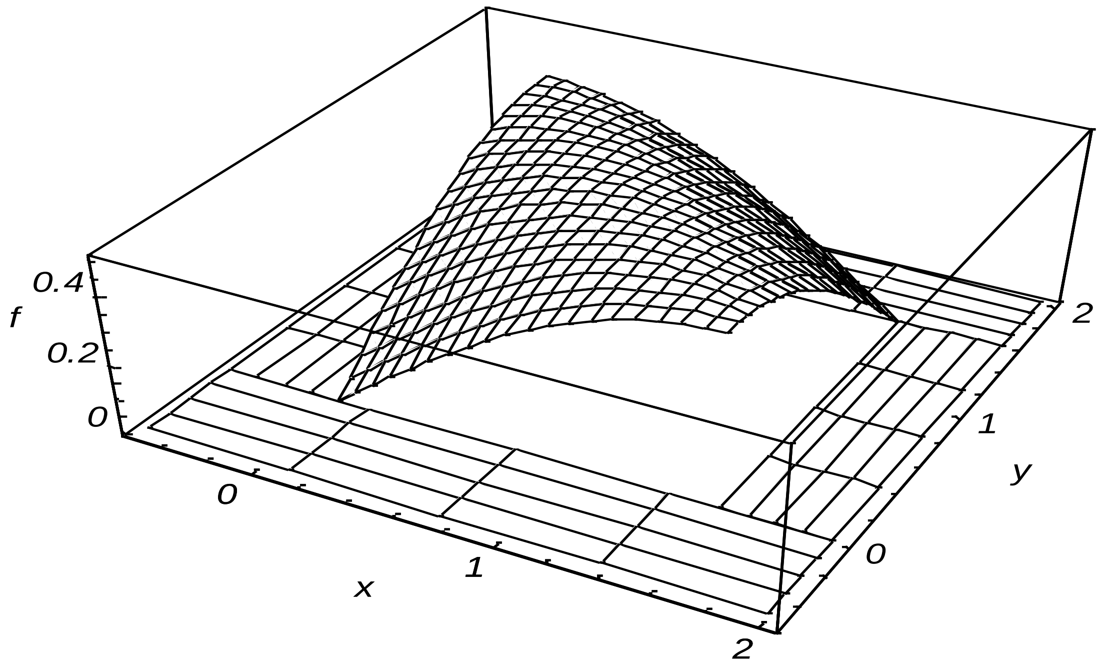


Рис. 1.13.3

Согласно (1.13.3), определяем функцию распределения вероятности системы случайных величин $F_{X,Y}(x, y)$:

$$F(x,y) = \int_0^x\int_0^y 0{,}5\sin(\xi+\eta)d\xi d\eta = -0{,}5\int_0^y[\cos(x+\eta) - \cos\eta]d\eta =$$

$$= 0{,}5\sin x + 0{,}5\sin y - 0{,}5\sin(x + y).$$

Строим график $F_{X,Y}(x, y)$ (см. рис. 1.13.4).

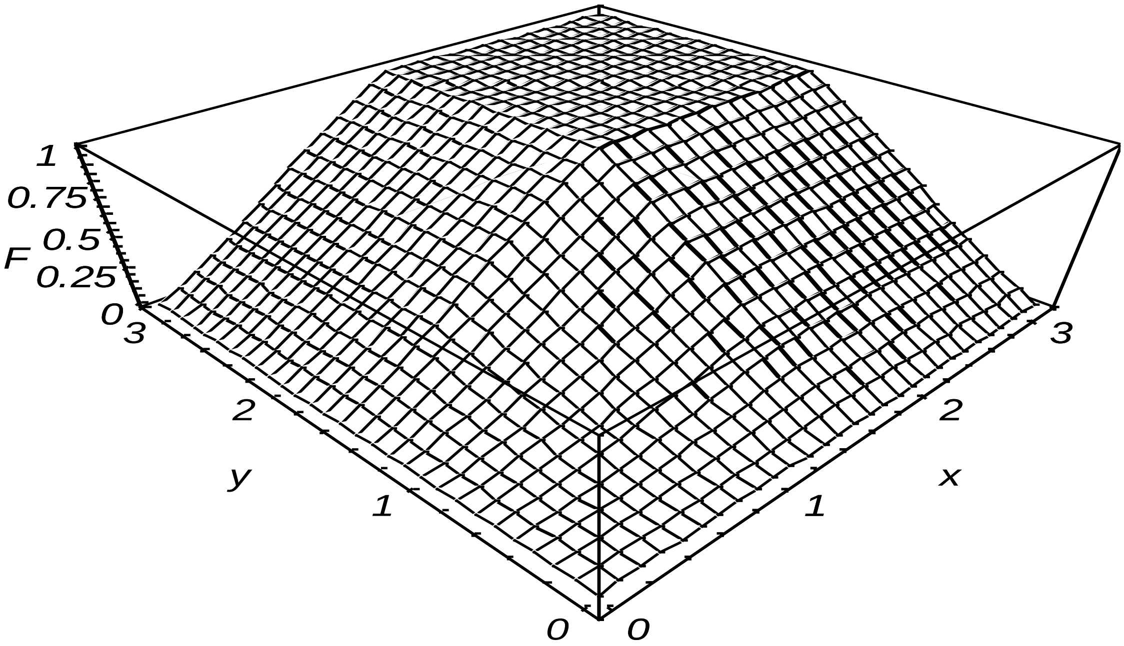


Рис. 1.13.4

По формулам (1.13.6) определяем плотности распределения вероятностей по каждой переменной:

$$f_X(x)=\int_{-\infty}^{+\infty} f_{X,Y}(x,\eta)d\eta\,,$$

$$\int_0^{\frac{\pi}{2}} f_{X,Y}(x,\eta)d\eta=\int_0^{\frac{\pi}{2}} 0{,}5\sin(x+\eta)d\eta=-0{,}5\cos(x+\eta)\,\Big|_0^{\frac{\pi}{2}}=0{,}5\cos x+0{,}5\sin x.$$

Аналогично

$$f_Y(y)=\int_{-\infty}^{+\infty} f_{X,Y}(\xi,y)d\xi=0{,}5\cos y+0{,}5\sin y.$$

Далее определяем числовые характеристики системы случайных величин (ξ, η). По формулам (1.13.8) вычисляем математические ожидания:

$$m_x=\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} x\;\; f_{X,Y}(x,y)dxdy=0{,}785398.$$

$$m_Y=\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} y\;\; f_{X,Y}(x,y)dxdy=0{,}785398.$$

А по формулам (1.13.9) – дисперсии (среднеквадратичные отклонения) и корреляционный момент:

$$D_X=\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}}(x-m_X)^2 f_{X,Y}(x,y)dxdy=0{,}187647,\ \sigma_X=\sqrt{D_X}=0{,}433182;$$

$$D_Y=\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}}(y-m_y)^2 f_{X,Y}(x,y)dxdy=0{,}187647,\ \sigma_Y=\sqrt{D_Y}=0{,}433182;$$

$$K_{XY}=\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}}(x-m_X)\ (y-m_Y)\ f_{X,Y}(x,y)dxdy=-0{,}0460539.$$

Наконец, согласно (1.13.10), можно найти коэффициент корреляции

$$\rho_{XY}=\frac{K_{XY}}{\sigma_X\sigma_Y}=-0{,}245429.$$

Систему двух случайных величин $\{X,Y\}$ называют ***дискретно распределенной***, если множество возможных значений $\{x_i,\ y_j\}$ счетное и задана соответствующая каждой паре вероятность $p_{ij}=P\{X=x_i,\ Y=y_j\}$, удовлетворяющая условию

$$\sum_i\sum_j p_{ij}=1, \qquad (1.13.11)$$

где суммирование распространяется на все возможные значения индексов i и j. В случае конечного числа возможных значений строят таблицу распределения системы двух случайных величин $\{X,Y\}$.

Y	X			
	x_1	x_2	…	x_n
y_1	p_{11}	p_{12}	…	p_{1n}
y_2	p_{21}	p_{22}	…	p_{2n}
…	…	…	…	…
y_m	p_{m1}	p_{m2}	…	p_{mn}

Одномерные законы распределения отдельных компонент выражаются формулами:

$$p_i=P\{X=x_i\}=\sum_j p_{ij},\quad p_j=P\{Y=y_j\}=\sum_i p_{ij}. \qquad (1.13.12)$$

***Начальный момент порядка k + s*:**

$$\alpha_{k,s}=\sum_i\sum_j x_i^k y_j^s p_{ij}. \qquad (1.13.13)$$

Центральный момент порядка k + s:

$$\mu_{k,s} = \sum_i \sum_j (x_i - m_X)^k (y_j - m_Y)^s p_{ij} . \qquad (1.13.14)$$

В частности, $m_X = \sum_i \sum_j x_i \, p_{ij} = \sum_i x_i p_i$,

$$m_Y = \sum_i \sum_j y_j p_{ij} = \sum_j y_j p_j , \qquad (1.13.15)$$

$$D_X = \sum_i \sum_j (x_i - m_X)^2 p_{ij} = \sum_i (x_i - m_X)^2 p_i = \sum_i x_i^2 p_i - m_X^2 ;$$

$$D_Y = \sum_i \sum_j (y_j - m_Y)^2 p_{ij} = \sum_j (y_j - m_Y)^2 p_j = \sum_j y_j^2 p_j - m_Y^2 ;$$

$$K_{XY} = \sum_i \sum_j (x_i - m_X) \ (y_j - m_Y) \ p_{ij} = \sum_i \sum_j x_i y_j p_{ij} - m_X m_Y .$$

*ПРИМЕР 3. Проводится проверка качества партии изделий. Из отобранных 5 изделий ξ изделий отвечают стандарту; среди них η (η ≤ 3) – высшего качества. Совместное распределение системы (ξ, η) задано двумерной таблицей.

Таблица 1

x_i y_i	0	1	2	3	4	5
0	0,02	0,08	0,1	0,06	0,04	0,02
1	0	0,03	0,07	0,09	0,1	0,12
2	0	0	0,02	0,04	0,06	0,08
3	0	0	0	0,01	0,02	0,04

Найти одномерные законы распределения каждой из величин ξ и η. Определить математические ожидания, среднеквадратичные отклонения величин ξ и η и их коэффициент корреляции.

Решение. Строим график вероятностей распределения (гистограмму системы (ξ, η)) (см. рис. 1.13.5).

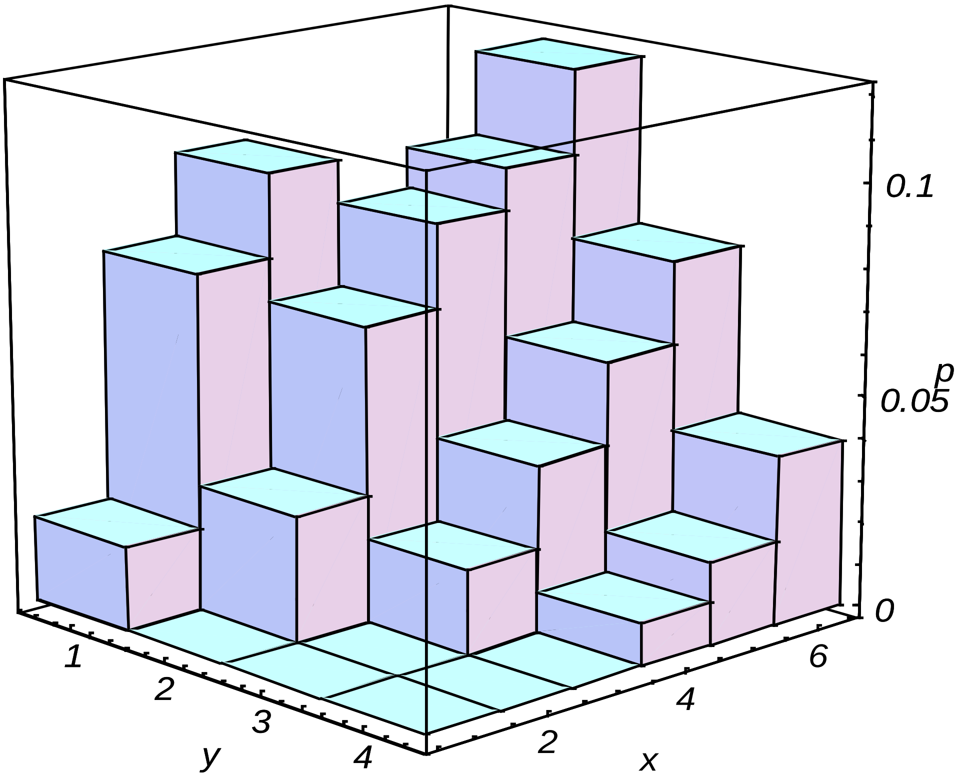


Рис. 1.13.5. Гистограмма

Возможен плоский вариант гистограмм (наложение построчных распределений) (рис. 1.13.6).

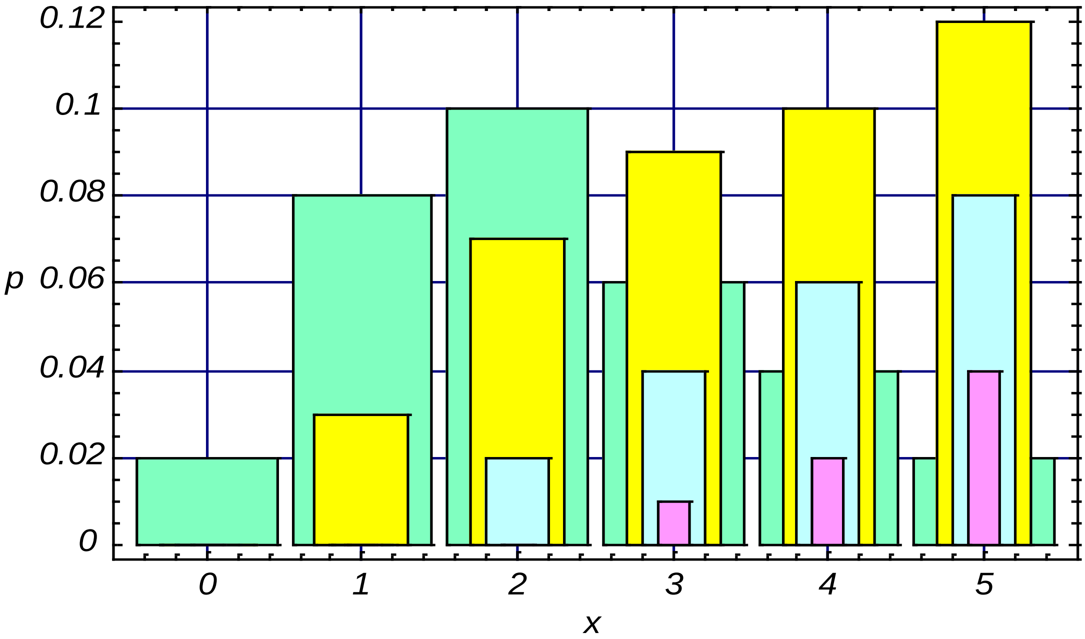


Рис. 1.13.6

Согласно (1.13.11), можно осуществить контроль вероятностей распределения:

$$\sum_{m=1}^{4}\sum_{n=1}^{6} p_{mn} = 1.$$

Функция распределения вектора (X,Y) определяется:

$$F_{X,Y}(x, y) = \sum_{m=1}^{4}\sum_{n=1}^{6} p_{mn},$$

где суммирование распространяется на все m, для которых $y_m < y$, а все n принимают все такие значения, для которых $x_n < x$.

Можно построить функцию распределения, которая в матричной форме имеет вид:

$$\begin{vmatrix} 0{,}02 & 0{,}1 & 0{,}2 & 0{,}26 & 0{,}3 & 0{,}32 \\ 0{,}02 & 0{,}13 & 0{,}3 & 0{,}45 & 0{,}59 & 0{,}73 \\ 0{,}2 & 0{,}13 & 0{,}32 & 0{,}51 & 0{,}71 & 0{,}93 \\ 0{,}02 & 0{,}13 & 0{,}32 & 0{,}52 & 0{,}74 & 1 \end{vmatrix}.$$

Строим график функции распределения (рис. 1.13.7).

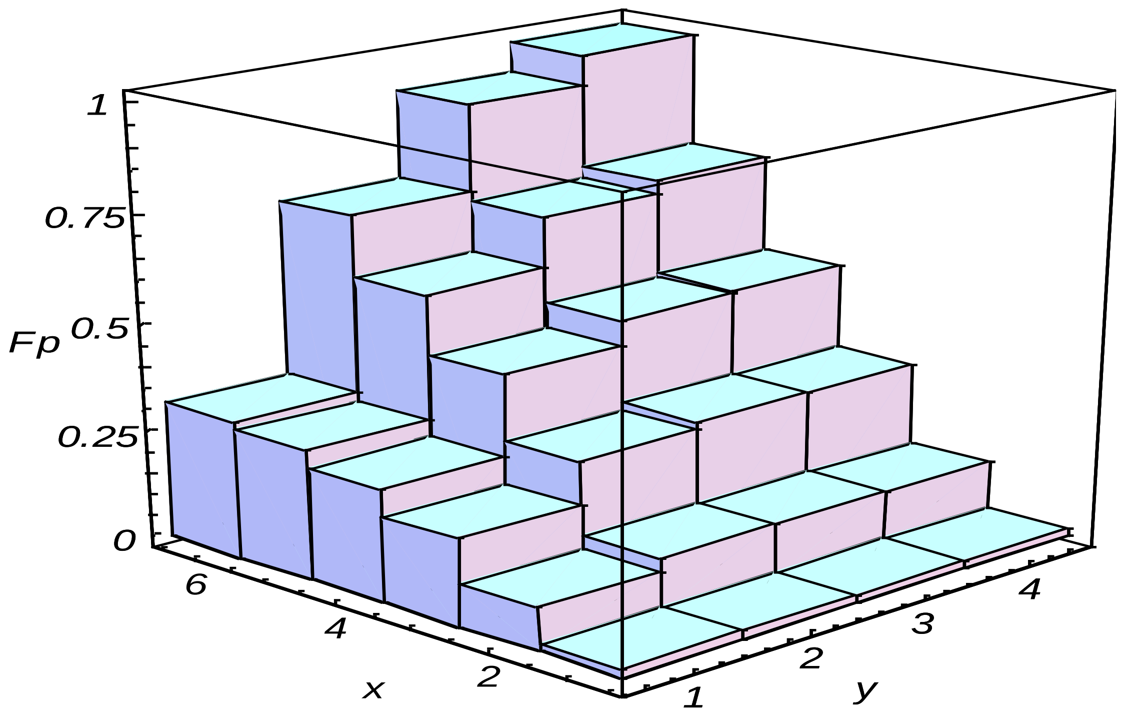


Рис. 1.13.7. График функции распределения

Найдем одномерные законы распределения по формулам (1.13.12) каждой из величин ξ и η суммируя по столбцам, а затем по строкам элементы матрицы распределения, и построим гистограммы одномерных законов распределения каждой из величин (табл. 2; рис. 1.13.8).

Таблица 2

X	0	1	2	3	4	5
P_i	0,02	0,11	0,19	0,2	0,22	0,26

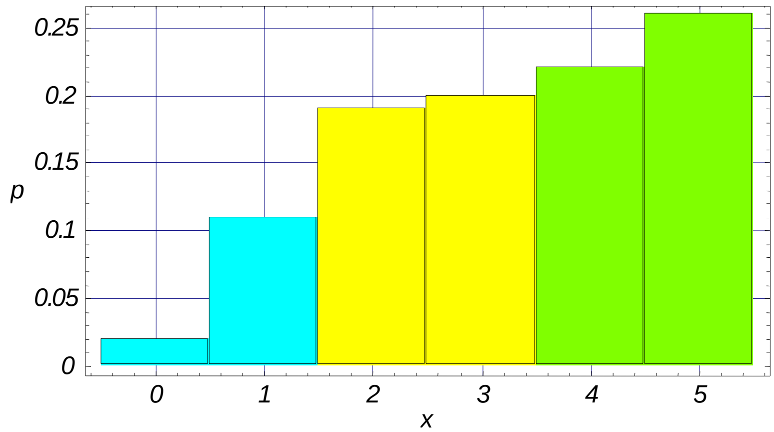


Рис. 1.13.8. Ряд распределения случайной величины X

Аналогично по Y (табл. 3; рис. 1.13.9).

Таблица 3

Y	0	1	2	3
p_i	0,32	0,41	0,2	0,07

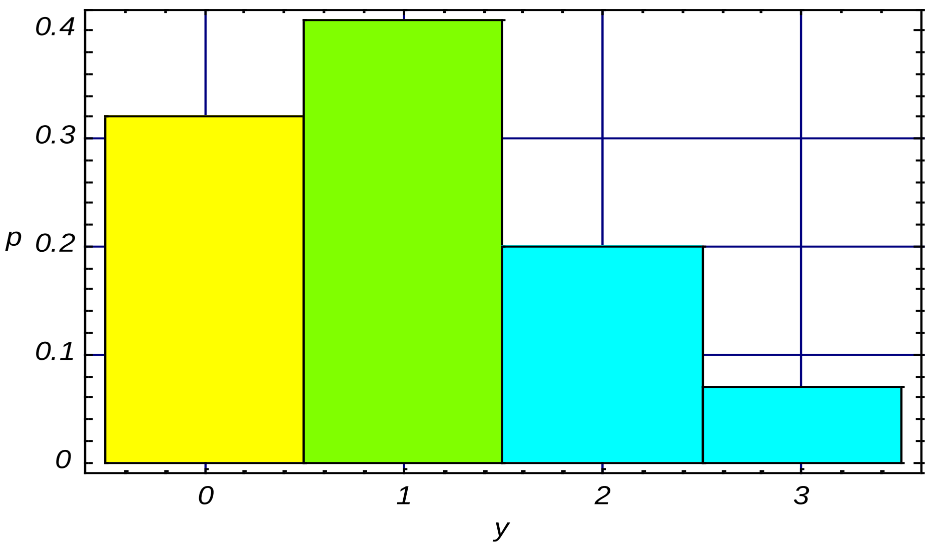


Рис. 1.13.9. Ряд распределения случайной величины Y

Числовые характеристики найдем в соответствии с (1.13.15), имея данные табл. 2 и 3.

Математические ожидания:

$$m_X = \sum_i \sum_j x_i\, p_{ij} = \sum_i x_i p_i = 0 \cdot 0{,}02 + 1 \cdot 0{,}11 + 2 \cdot 0{,}19 +$$

$$+ 3 \cdot 0{,}2 + 4 \cdot 0{,}22 + 5 \cdot 0{,}26 = 3{,}27;$$

$$m_Y = \sum_i \sum_j y_j p_{ij} = \sum_j y_j p_j = 0 \cdot 0{,}32 + 1 \cdot 0{,}041 + 2 \cdot 0{,}2 + 3 \cdot 0{,}07 = 1{,}02.$$

Дисперсии и среднеквадратичные отклонения:

$$D_X = \sum_i \sum_j (x_i - m_X)^2 p_{ij} = \sum_i (x_i - m_X)^2 p_i = \sum_i x_i^2 p_i - m_X^2 = 1{,}9971;$$

$$\sigma_X = \sqrt{D_X} = 1{,}41319;$$

$$D_Y = \sum_i \sum_j (y_j - m_Y)^2 p_{ij} = \sum_j (y_j - m_Y)^2 p_j = \sum_j y_j^2 p_j - m_Y^2 = 0{,}7996;$$

$$\sigma_Y = \sqrt{D_Y} = 0{,}894204;$$

$$K_{XY} = \sum_i \sum_j (x_i - m_X)\ (y_j - m_Y)\ \ p_{ij} = \sum_i \sum_j x_i y_j p_{ij} - m_X m_Y = 0{,}63.$$

Коэффициент корреляции системы двух случайных величин:

$$\rho_{XY} = \frac{K_{XY}}{\sigma_X \sigma_Y} = 0{,}502185.$$

1.14. Практическое приложение к первой части

1.14.1. Случайные события

*ПРИМЕР 1. Дана система из трех блоков a_1, a_2, b (рис. 1.1), где a_1 и a_2 – дублирующие блоки (схема последовательного соединения блока b с подсистемой a параллельно соединенных блоков a_1 и a_2).

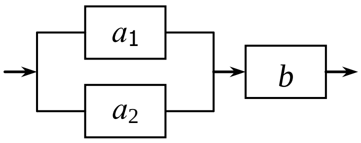


Рис. 1.1

Записать события, состоящие в том, что система: 1) исправна; 2) неисправна.

Решение. Введем обозначения: A_1 – событие, состоящее в том, что блок a_1 исправен; A_2 – событие, состоящее в том, что блок a_2 исправен; B – событие, состоящее в том, что блок b исправен; S – событие, состоящее в том, что система исправна; S_a – событие, состоящее в том, что подсистема a исправна. Разобьем систему на две подсистемы a и b. Подсистема a дублирующих блоков исправна в том случае, когда исправен хотя бы один из блоков a_k (k = 1,2), т. е. $S_a = A_1 + A_2$. Для исправности системы необходима исправность подсистемы a и блока b, т. е. $S = B \cdot S_a$. Таким образом, $S = B \cdot (A_1 + A_2)$. Аналогично $\overline{S} = A_1 \cdot A_2 + B$.

*ПРИМЕР 2. Система S состоит из двух независимых подсистем S_a и S_b (рис. 1.2). Неисправность хотя бы одной подсистемы ведет к неисправности всей системы (подсистемы соединены последовательно). Каждая подсистема состоит из двух независимых дублирующих блоков a_k и b_k (k = 1,2) (схема параллельного подсоединения блоков в подсистемах).

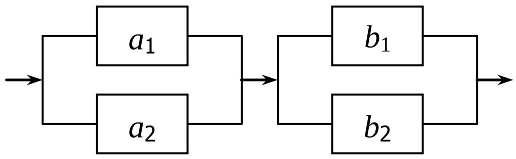


Рис. 1.2

Найти надежность системы – вероятность того, что система будет исправна в течение некоторого времени, если известны надежности блоков $P(a_k)$ = 0,8; $P(b_k)$ = 0,9.

Решение. Введем обозначения: A_k – событие, состоящее в том, что блок a_k исправен; B_k – событие, состоящее в том, что блок b_k исправен; S – событие, состоящее в том, что система исправна; S_a – событие, состоящее в том, что подсистема a исправна; S_b – событие, состоящее в том, что подсистема b исправна. Разобьем систему на две подсистемы a и b. Подсистема a дублирующих блоков исправна в том случае, когда исправен хотя бы один из блоков a_k (k = 1,2), т. е. $S_a = A_1 + A_2$ – сумма двух совместных независимых событий. Следовательно $P(S_a) = P(a_1) + P(a_2) - P(a_1) \cdot P(a_2) = 0{,}8 + 0{,}8 - 0{,}64 = 0{,}96$. Аналогично $S_b = B_1 + B_2$. Следовательно,

$$P(S_b) = P(b_1) + P(b_2) - P(b_1) \cdot P(b_2) = 0{,}9 + 0{,}9 - 0{,}81 = 0{,}99.$$

Для исправности системы необходима исправность подсистем a и b, т. е. $S = S_a \cdot S_b$. Таким образом, $P(S) = P(S_a) \cdot P(S_b) = 0{,}9504$.

****УПРАЖНЕНИЯ**

1. Пусть событие A – «появление 6 очков при бросании игральной кости», событие B – «появление 5 очков при бросании игральной кости», событие C – «появление 4 очков при бросании игральной кости». В чем состоит событие: $A \cup B \cup C$?

Ответ: «появилось не меньше 4 очков».

2. Пусть событие A_1 – «поражение мишени одним выстрелом», событие A_2 – «поражение мишени двумя выстрелами», событие A_3 – «поражение мишени тремя выстрелами», …, A_{100} – «поражение мишени сотней выстрелов». В чем состоит событие $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_{100}$?

Ответ: мишень поражена не больше чем 100 выстрелами.

3. Пусть A, B и C – случайные события, выраженные элементарными событиями одного и того же пространства элементарных событий. Запишите такие события:

а) произошло только B;
б) произошло одно и только одно из данных событий;
в) произошли два и только два из данных событий;
г) произошли все три события;
д) произошло хотя бы одно из данных событий;

Ответ: а) $\overline{A} \cap B \cap \overline{C}$;
б) $(A \cap \overline{B} \cap \overline{C}) \cup (\overline{A} \cap B \cap \overline{C}) \cup (\overline{A} \cap \overline{B} \cap C)$;
в) $(A \cap B \cap \overline{C}) \cup (A \cap \overline{B} \cap C) \cup (\overline{A} \cap B \cap C)$;
г) $A \cap B \cap C$;
д) $A \cup B \cup C$.

1.14.2. Комбинаторика

*ПРИМЕР 3. Сколькими способами можно расположить на шахматной доске 8 ладей так, чтобы они не могли взять друг друга?

Решение. Ясно, что в этом случае на каждой горизонтали и каждой вертикали шахматной доски может быть расположено только по одной ладье. Число возможных позиций – число перестановок из 8 элементов:

$$P_8 = 8! = 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 40\,320.$$

*ПРИМЕР 4. На тренировках занимаются 12 баскетболистов. Сколько может быть образовано тренером разных стартовых пятерок?

Решение. Так как при составлении стартовой пятерки тренера интересует только состав пятерки, то достаточно определить число сочетаний из 12 элементов по 5:

$$C_{12}^{5} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 792.$$

****УПРАЖНЕНИЯ**

1. Сколько разных пятизначных чисел можно составить из цифр 1, 2, 3, 4 и 5 при условии, что ни одна цифра не повторяется?

Ответ: 120.

2. Игрок сначала бросает белую игральную кость, потом черную. Сколько может быть случаев, когда число очков, появившихся на белой кости, больше числа очков, появившихся на черной кости?

Ответ: 15.

3. В кружке юных математиков 25 членов. Необходимо избрать председателя кружка, его заместителя, редактора стенгазеты и секретаря. Сколькими способами можно образовать эту руководящую четверку, если одно лицо может занимать только один пост?

Ответ: 303 600.

4. В одной арабской сказке речь идет о такой задаче. Вокруг костра сидят 12 разбойников. Каждый из них смертельно ненавидит двух ближайших соседей. С целью спрятать награбленное необходимо выделить 5 разбойников. Сколькими способами атаман может назначить пятерых так, чтобы между ними не было распрей?

Ответ: 36.

5. В колоде 32 карты. Раздаются 3 карты. Сколько может быть случаев появления одного туза среди розданных карт?

Ответ: 1 512.

1.14.3. Вероятность события

*ПРИМЕР 5. Таня и Ваня договорились встречать Новый год в компании из 10 человек. Они оба очень хотели сидеть за праздничным столом рядом. Какова вероятность исполнения их желания, если среди их друзей принято места распределять путем жребия?

Решение. Ясно, что 10 человек могут усесться за стол 10! разными способами. Сколько же из этих n =10! равновозможных способов благоприятны для Тани и Вани? Таня и Ваня, сидя рядом, могут занять 20 разных позиций. В то же время восьмерка их друзей может сесть за стол 8! разными способами, поэтому m = 20 · 8!. Следовательно,

$$P(\text{"исполнение желания Тани и Вани"}) = \frac{20 \cdot 8!}{10!} = \frac{2}{9}.$$

****УПРАЖНЕНИЯ**

1. На первом этаже семиэтажного дома в лифт зашли 3 человека. Вероятность выхода каждого из лифта на любом этаже одинакова. Найдите вероятность событий:

A – «все вышли из лифта на четвертом этаже»;

B – «все вышли из лифта на одном этаже»;

Ответ: $P(A) = \dfrac{1}{216}$; $P(B) = \dfrac{1}{36}$.

2. Номер телефона состоит из 5 цифр. Какова вероятность того, что все цифры наугад набранного номера разные?

Ответ**:** 0,3024.

3. Замок содержит на общей оси 4 диска, каждый из которых разделен на 6 секторов, отмеченных цифрами. Замок может быть открыт только в том случае, если все диски занимают определенные положения относительно корпуса замка, их цифры образуют определенное число, составляющее «секрет» замка. Какова вероятность открыть замок, установив произвольную комбинацию цифр?

Ответ: 0,00077.

4. В партии, состоящей из k изделий, имеется l дефектных. Из партии выбирается для контроля r изделий. Найти вероятность p того, что из них ровно s будут дефектными.

Ответ: $P = \dfrac{C_l^s\, C_{k-l}^{r-s}}{C_k^r}$.

1.14.4. Теоремы сложения и умножения вероятностей

*ПРИМЕР 6. На военных учениях летчик получил задание «уничтожить» 3 рядом расположенных склада боеприпасов противника. На борту самолета одна бомба. Вероятность попадания в первый склад примерно равна 0,01, во второй – 0,008, в третий – 0,025.

Любое попадание в результате детонации вызовет взрыв и остальных складов. Какова вероятность того, что склады противника будут уничтожены?

Решение. Обозначим события: A – «склады уничтожены»; A_1 – «попадание в первый склад»; A_2 – «попадание во второй склад»; A_3 – «попадание в третий склад».

Для уничтожения складов достаточно попадания в один из упомянутых трех складов. Поэтому:

$$A = A_1 \cup A_2 \cup A_3,$$
$$P(A) = P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) =$$
$$= 0{,}01 + 0{,}008 + 0{,}025 = 0{,}043.$$

*ПРИМЕР 7. Подбрасываем две монеты. Какова вероятность выпадения хотя бы одного герба?

Решение. Обозначим события:

A – «появление герба при подбрасывании первой монеты»;

B – «появление герба при подбрасывании второй монеты»;

Нам надо определить вероятность события $C = A \cup B$:

$$P(C) = P(A \cup B) = P(A) + P(B) – P(A \cap B).$$

Ясно, что $P(A) = \dfrac{1}{2}$, $P(B) = \dfrac{1}{2}$, $P(A \cap B) = \dfrac{1}{4}$.

Отсюда $P(C) = \dfrac{1}{2} + \dfrac{1}{2} - \dfrac{1}{4} = \dfrac{3}{4}$.

*ПРИМЕР 8. Задача, показывающая, что интуитивное понимание независимости событий тоже приводит к формальному соотношению:

$$P(A \cap B) = P(A) \cdot P(B). \tag{1.14.1}$$

Пусть точка M наудачу бросается в прямоугольник с размерами X и $У$, стороны которого параллельны координатным осям. Какова вероятность того, что M попадет в прямоугольник с размерами x и y, стороны которого тоже параллельны координатным осям?

Решение. Обозначим события: A – «точка M попала в полосу шириной x». Разумеется, $P(A) = \frac{x}{X}$, B – «M попала в полосу шириной y». Разумеется, $P(B) = \frac{y}{У}$. По формуле (1.14.1) $P = P(A \cap B)$, где P – «M попала в маленький прямоугольник», но $P = \frac{xy}{XУ} = \frac{x}{X} \cdot \frac{y}{У} = P(A) \cdot P(B)$.

*ПРИМЕР 9. Зашедший в магазин мужчина что-нибудь покупает с вероятностью 0,1, а зашедшая женщина – с вероятностью 0,6.

У прилавка один мужчина и две женщины. Какова вероятность того, что по крайней мере одно лицо что-нибудь купит?

Решение. Обозначим события: A – «покупку сделает мужчина». B_1 – «покупку сделает первая женщина», B_2 – «покупку сделает вторая женщина». Если событие C – «по крайней мере одно лицо что-нибудь купит», то $C = A \cup B_1 \cup B_2$. По формуле (1.5.1) имеем

$$P(C) = P(A) + P(B_1) + P(B_2) - P(A) \cdot P(B_2) - P(A) \cdot P(B_1) - $$
$$- P(B_1) \cdot P(B_2) + P(A) \cdot P(B_1) \cdot P(B_2);$$

$$P(C) = 0{,}1 + 0{,}6 + 0{,}6 - 0{,}1 \cdot 0{,}6 -$$
$$- 0{,}1 \cdot 0{,}6 - 0{,}6 \cdot 0{,}6 + 0{,}1 \cdot 0{,}6 \cdot 0{,}6 = 0{,}856.$$

*ПРИМЕР 10. Из полной колоды карт (52 листа) вынимаются сразу четыре карты. Найти вероятность того, что все эти четыре карты будут разных мастей.

Решение. Первая карта может быть какой угодно масти; вторая должна быть не такой, как первая; третья – не такой, как первая и вторая; четвертая – не такой, как три первые. Искомая вероятность равна

$$P = 1 \cdot \frac{39}{51} \cdot \frac{26}{50} \cdot \frac{13}{49} \approx 0{,}106.$$

*ПРИМЕР 11. Та же задача, но каждая карта после вынимания возвращается в колоду.

Ответ: $P = 1 \cdot \frac{39}{52} \cdot \frac{26}{52} \cdot \frac{13}{52} \approx 0{,}094.$

****УПРАЖНЕНИЯ**

1. В ящике 10 белых и 8 красных шариков. Одновременно наугад вынимают 2 шарика. Какова вероятность того, что они разных цветов?

Ответ: $\frac{80}{153}$.

2. В ящике 7 белых и 9 черных шариков. Наугад вынимают один шарик, рассматривают его на свету и кладут обратно в ящик. Опять наугад вынимают один шарик. Какова вероятность, что оба шарика белые?

Ответ: $\frac{49}{256}$.

3. Каждая буква слова «комбинаторика» написана на отдельной карточке, которые тщательно перемешиваются. Последовательно извлекаются четыре карточки. Какова вероятность получить слово «ромб»?

Ответ: $\frac{1}{8580}$.

1.14.5. Формула полной вероятности. Формула Байеса

*ПРИМЕР 12. В группе 40 стрелков, из которых 10 человек стреляют отлично, 20 – хорошо, 6 – удовлетворительно, 4 – плохо. Вероятности попадания в цель для них соответственно 0,9; 0,8; 0,7; 0,4. На линию огня случайным образом вызывают одного, который производит один выстрел в цель. Найти вероятность попадания при этом.

Решение. Введем обозначения: A – событие, состоящее в том, что при выстреле будет попадание. Вводим гипотезы: H_1 – событие, состоящее в том, что на линию огня вызван стрелок, стреляющий отлично; H_2 – событие, состоящее в том, что на линию огня вызван стрелок, стреляющий хорошо, H_3 – событие, состоящее в том, что на линию огня вызван стрелок, стреляющий удовлетворительно, H_4 – событие, состоящее в том, что на линию огня вызван стрелок, стреляющий плохо. По условию задачи

$$P(H_1) = p_1 = \frac{10}{40}; \qquad P(H_2) = p_2 = \frac{20}{40};$$

$$P(H_3) = p_3 = \frac{6}{40}; \qquad P(H_4) = p_4 = \frac{4}{40};$$

$$P(A/H_1) = 0{,}9; \qquad P(A/H_2) = 0{,}8;$$

$$P(A/H_3) = 0{,}7; \qquad P(A/H_4) = 0{,}4.$$

По формуле полной вероятности находим

$$P(A) = P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) +$$
$$+ P(H_3) \cdot (P(A/H_3) + P(H_4) \cdot P(A/H_4) =$$
$$= 0{,}9 \cdot \frac{1}{4} + 0{,}8 \cdot \frac{1}{2} + 0{,}7 \cdot \frac{3}{20} + 0{,}4 \cdot \frac{1}{10} = 0{,}770.$$

*ПРИМЕР 13. В урну, содержащую два шара, опущен белый шар, после чего из нее наудачу извлечен один шар. Найти вероятность того, что извлеченный шар окажется белым, если равновозможны все возможные предположения о первоначальном составе шаров (по цвету).

Решение. Введем обозначения: A – событие – извлечен белый шар. Возможны следующие предположения (гипотезы) о первоначальном составе шаров: B_1 – белых шаров нет, B_2 – один белый шар, B_3 – два белых шара. Поскольку равновозможны все возможные предположения о первоначальном составе шаров (по цвету), получим

$$P(B_1) = P(B_2) = P(B_3) = 1/3.$$

$$P(A/B_1) = 1/3; \quad P(A/B_2) = 2/3; \quad P(A/B_3) = 3/3.$$

По формуле полной вероятности находим

$$P(A) = P(B_1) \cdot P(A/B_1) + P(B_2) \cdot P(A/B_2) + P(B_3) \cdot (P(A/B_3) =$$
$$= \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{3}{3} = \frac{2}{3}.$$

*ПРИМЕР 14. Имеется 12 урн, из них в 6-ти урнах (состав H_1) по 3 белых и 4 черных шара, в 3-х (состав H_2) – по 2 белых и 8 черных шаров, в 2-х урнах (состав H_3) – по 6 белых и 1 черному и в 1-й урне (состав H_4) – 4 белых и 3 черных шара.

1. Чему равна вероятность того, что случайно взятый шар из наугад выбранной урны оказался белым?

2. Из наугад выбранной урны взят шар. Чему равна вероятность того, что шар взят из урны состава H_3, если он оказался белым?

Решение

1. Интересующее нас событие обозначим A – выбор белого шара. Заметим, что классическим определением вероятности нельзя воспользоваться, так как число благоприятствующих исходов нельзя найти, потому что неизвестно, из какой именно урны был взят шар. Относительно этого сделано четыре предположения (выдвинуты гипотезы):

H_1 – случайно выбранный шар взят из урны состава один;

H_2 – случайно выбранный шар взят из урны состава два;

H_3 – случайно выбранный шар взят из урны состава три;

H_4 – случайно выбранный шар взят из урны состава четыре.

Эти события являются попарно несовместными и образуют полную группу событий, так как одно из них обязательно происходит. Следовательно, эти события удовлетворяют условиям, налагаемым на гипотезы.

Ответ на вопрос: «Какова вероятность того, что случайно выбранный шар окажется белым?» – будем искать по формуле полной вероятности.

$$P(A)=\sum_{i=1}^{4}P(A/H_i)P(H_i)=$$
$$=P(A/H_1)P(H_1)+P(A/H_2)P(H_2)+$$
$$+P(A/H_3)P(H_3)+P(A/H_4)P(H_4),$$

где $P(A/H_1)=\dfrac{3}{3+4}=\dfrac{3}{7}$ – вероятность того, что белый шар взят из урны состава один;

$P(A/H_2)=\dfrac{2}{2+8}=\dfrac{1}{5}$ – вероятность того, что белый шар взят из урны состава два;

$P(A/H_3)=\dfrac{6}{6+1}=\dfrac{6}{7}$ – вероятность того, что белый шар взят из урны состава три;

$P(A/H_4)=\dfrac{4}{4+3}=\dfrac{4}{7}$ – вероятность того, что белый шар взят из урны состава четыре;

$P(H_1) = \dfrac{6}{12}=\dfrac{1}{2}$ – вероятность того, что случайно выбранный шар взят из урны состава один;

$P(H_2) = \dfrac{3}{12}=\dfrac{1}{4}$ – вероятность того, что случайно выбранный шар взят из урны состава два;

$P(H_3) = \dfrac{2}{12}=\dfrac{1}{6}$ – вероятность того, что случайно выбранный шар взят из урны состава три;

$P(H_4) = \dfrac{1}{12}$ – вероятность того, что случайно выбранный шар взят из урны состава четыре.

$$P(A)=\frac{1}{2}\cdot\frac{3}{7}+\frac{1}{4}\cdot\frac{1}{5}+\frac{1}{6}\cdot\frac{6}{7}+\frac{1}{12}\cdot\frac{4}{7}\approx 0{,}4595.$$

2. Ответим на второй вопрос задачи. Для этого требуется «переоценить» гипотезы, т. е. найти вероятности гипотез после того, как событие A уже произошло. В этом случае вероятность вычисляется по формуле Байеса.

Найдем вероятность того, что шар взят из урны состава H_3, если он оказался белым:

$$P(H_3/A)=\frac{P(H_3)\cdot P(A/H_3)}{P(A)}=\frac{\frac{1}{6}\cdot\frac{6}{7}}{\frac{3}{14}+\frac{1}{20}+\frac{1}{7}+\frac{1}{21}}\approx 0{,}314.$$

****УПРАЖНЕНИЯ**

1. В трех урнах находятся белые и черные шары: в первой – 2 белых и 4 черных, во второй – 3 белых и 5 черных, в третьей – 4 белых и 6 черных. Из первой урны взяли наудачу два шара и переложили во вторую. После этого взяли 2 шара из второй урны и преложили в третью. Наконец, из третьей урны в первую переложили 2 шара.

Определить, чему равна вероятность того, что: 1) состав шаров во всех урнах не изменится? 2) состав во всех урнах изменится?

Ответ: 1) $\frac{173}{495}$; 2) 0,094.

2. Литье в болванках поступает из двух заготовительных цехов: 70 % из первого и 30 % из второго. При этом материал первого цеха имеет 10 % брака, а второго – 20 %. Найти вероятность того, что одна взятая наугад болванка без дефектов?

Ответ: 0,87.

3. Некто, заблудившись в лесу, вышел на поляну, откуда вело 5 дорог. Известно, что вероятности выхода из леса за час для различных дорог равны соответственно 0,6; 0,3; 0,2; 0,1; 0,1. Чему равна вероятность того, что заблудившийся пошел по первой дороге, если известно, что он вышел из леса через час?

Ответ: $\frac{6}{13}$.

4. Из 10 учеников, которые пришли на экзамен по математике, трое подготовились отлично, четверо – хорошо, двое – удовлетворительно, а один совсем не готовился – понадеялся на то, что все помнит. В билетах 20 вопросов. Отлично подготовившиеся ученики могут ответить на все 20 вопросов, хорошо – на 16 вопросов, удовлетворительно – на 10 и не подготовившийся – на 5 вопросов. Каждый ученик получает наугад 3 вопроса из 20. Приглашенный первым ученик ответил на все три вопроса. Какова вероятность того, что он отличник?

Ответ: 0,58.

1.14.6. Формула Бернулли и приближенная формула Пуассона. Наивероятнейшее число наступления событий

*ПРИМЕР 15. Вероятность того, что расход электроэнергии в продолжение одних суток не превысит установленной нормы, равна p = 0,75. Найти вероятность того, что в ближайшие 6 суток расход электроэнергии в течение 4 суток не превысит нормы.

Решение. Вероятность нормального расхода электроэнергии в продолжение каждых из 6 суток постоянна и равна p = 0,75. Следовательно, вероятность перерасхода электроэнергии в каждые сутки также постоянна и равна $q = 1 - p = 1 - 0{,}75 = 0{,}25$. Искомая вероятность находится по формуле Бернулли:

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

$$P_6(4) = C_6^4\, p^4\, q^2 = C_6^2\, p^4\, q^2 = \frac{6\cdot 5}{1\cdot 2}(0{,}75)^4\cdot(0{,}25)^2 = 0{,}30.$$

*ПРИМЕР 16. Подводная лодка атакует крейсер, выпуская по нему одну за другой 4 торпеды. Вероятность попадания каждой торпедой примерно равна $\frac{3}{4}$. Любая из торпед с одинаковой вероятностью может пробить один из 10 отсеков крейсера, которые в результате попадания наполняются водой. При заполнении хотя бы двух отсеков крейсер тонет. Вычислить вероятность гибели крейсера.

Решение. Эту задачу удобно решать по формуле полной вероятности с гипотезами A_i – в корабль попало i (i =1, 2, 3, 4) торпед.

Обозначим события:

A_1 – «попадание одной торпедой»; A_2 – «попадание двумя торпедами»; A_3 – «попадание тремя торпедами»; A_4 – «попадание четырьмя торпедами»; A – «крейсер потоплен».

Согласно формуле Бернулли, имеем

$$P(A_1) = C_4^1\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^3 = \frac{3}{64}; \qquad P(A_2) = C_4^2\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \frac{27}{128};$$

$$P(A_3) = C_4^3\left(\frac{3}{4}\right)^3\frac{1}{4} = \frac{27}{64}; \qquad P(A_4) = C_4^4\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^0 = \frac{81}{256};$$

$$P(A/A_1) = 0;\; P(A/A_2) = 1 - \frac{1}{10} = \frac{9}{10};$$

$$P(A/A_3) = 1 - \frac{1}{10^2} = \frac{99}{100}.\; P(A/A_4) = 1 - \frac{1}{10^3} = \frac{999}{1000}.$$

По формуле полной вероятности

$$P(A) = \frac{3}{64}\cdot 0 + \frac{27}{128}\cdot\frac{9}{10} + \frac{27}{64}\cdot\frac{99}{100} + \frac{81}{256}\cdot\frac{999}{1000} \approx 0{,}9237.$$

У крейсера противника мало шансов на спасение!

*ПРИМЕР 17. Прядильщица обслуживает 1 000 веретен. Вероятность обрыва нити на одном веретене в течение 1 минуты равна 0,002. Найти вероятность того, что в течение 1 минуты обрыв произойдет *более чем на трех* веретенах.

Решение. Если нити назвать «успехом», то в данном примере требуется найти вероятность того, что в n =1 000 испытаниях Бернулли число успехов m_n превысит 3. Вероятность успеха в каждом испытании p = 0,002, а вероятность неуспеха q = 0,998.

В случае, когда n велико, а p мало (обычно $p < 0{,}1$; $npq \le 9$), вместо сложной формулы Бернулли применяют приближенную ***формулу Пуассона***:

$$P_n(m) \approx \frac{a^m e^{-a}}{m!}, \quad a = np.$$

Находим

$$P(m_n > 3) = 1 - e^{-2}\frac{2^0}{0!} - e^{-2}\frac{2}{1!} - e^{-2}\frac{2^2}{2!} - e^{-2}\frac{2^3}{3!} =$$
$$= 1 – 0{,}1353 – 0{,}2707 – 0{,}2707 – 0{,}1805 = 0{,}1428.$$

*ПРИМЕР 18. Какова вероятность наступления события A в каждом испытании, если наивероятнейшее число наступлений события A в 120 испытаниях равно 32?

Решение. Согласно неравенству (1.8.3) $np - q \le m_0 \le np + p$, имеем

$$n=120, \quad m_0 = 32,$$

тогда

$$120p - (1 - p) \le 32; \quad 120p + p \ge 32.$$

Решая полученную систему, находим, что $\dfrac{32}{121} \le p \le \dfrac{33}{121}$.

****УПРАЖНЕНИЯ**

1. Игральная кость подброшена 10 раз. Найти вероятность выпадения единицы 7 раз.

Ответ: $2{,}537 \cdot 10^{-4}$.

2. Вероятность выигрыша по облигации займа за все время его действия равна 0,25. Какова вероятность того, что некто, приобретая 8 облигаций, выиграет по 6 из них?

Ответ: 0,0038.

3. Случайно встреченное лицо с вероятностью 0,2 может оказаться брюнетом, с вероятностью 0,3 – блондином, с вероятностью 0,4 – шатеном и с вероятностью 0,1 – рыжим. Какова вероятность того, что среди пяти случайно встреченных лиц: а) не менее четырех блондинов; б) два блондина и три шатена; в) хотя бы один рыжий?

Ответ: а) 0,03078; б) 0,07; в) 0,3439.

4. Производится 21 выстрел по цели, вероятность попадания в которую при одном выстреле равна 0,25. Найти наивероятнейшее число попаданий в цель.

Ответ: 5.

5. В наблюдениях Резерфорда и Гейгера радиоактивное вещество за промежуток времени 7,5 с испускало в среднем 3,87 α-частицы. Найти вероятность тог, что за 1 с это вещество испустит хотя бы одну α-частицу.

Ответ: $1 - e^{-0,516} \approx 0,4043$.

1.14.7. Локальная и интегральная теоремы Муавра – Лапласа

*ПРИМЕР 19. Вероятность изготовления детали высшего сорта на данном станке равна 0.4. Найти вероятность того, что среди наудачу взятых 26 деталей половина окажется высшего сорта.

Решение. Требуется найти вероятность $P_{26}(13)$. По условию $n = 26$; $m = 13$; $p = 0,4$; $q = 1 - 0,4 = 0,6$. Так как $n = 26$ – достаточно большое число, воспользуемся локальной теоремой Лапласа (1.9.1):

$$P_n(m) \cong \frac{1}{\sqrt{npq}} \varphi(x).$$

Необходимые вычисления производим по следующей схеме:

$$np = 26 \cdot 0,4 = 10,4; \qquad npq = 10,4 \cdot 0,6 = 6,24;$$

$$\sqrt{npq} = \sqrt{6,24} = 2,50; \qquad m - np = 2,6.$$

Найдем значение x:

$$x = \frac{m - np}{\sqrt{npq}} = \frac{2,60}{2,50} = 1,04.$$

По таблицам находим φ (1,04) = 0,2323. Искомая вероятность

$$P_{26}(13) = \frac{0,2323}{2,50} = 0,093.$$

*ПРИМЕР 20. Вероятность получения по лотерее проигрышного билета равна 0,1. Какова вероятность того, что среди 500 наугад купленных билетов не менее 48 и не более 55 безвыигрышных?

Решение. Воспользуемся интегральной теоремой Лапласа (1.9.4):

$$P_n(m_1 \le m \le m_2) = \sum_{m=48}^{55} P(m) \cong \Phi(x'') - \Phi(x').$$

По условию n =500; p=0,1; q=0,9; m_1 = 48; m_2 =55. Вычислим x' и x'':

$$x' = \frac{m_1 - np}{\sqrt{npq}} = \frac{48 - 500 \cdot 0,1}{\sqrt{500 \cdot 0,1 \cdot 0,9}} \approx -0,298;$$

$$x'' = \frac{m_2 - np}{\sqrt{npq}} = \frac{55 - 500 \cdot 0,1}{\sqrt{500 \cdot 0,1 \cdot 0,9}} \approx 0,745;$$

Учитывая, что функция Лапласа нечетная, т. е. $\Phi(-x) = -\Phi(x)$, получим

$$P_{500}(48;\, 55) = \Phi(0,745) + \Phi(0,298) = 0,3913.$$

*ПРИМЕР 21. С конвейера сходит в среднем 85 % изделий первого сорта. Сколько изделий необходимо взять, чтобы с вероятностью 0,997 отклонение частости изделий первого сорта в них от 0,85 по абсолютной величине не превосходило 0,01?

Решение. Здесь p = 0,85; q=0,15; ε = 0,01; P = 0,997; n – ?

Из теоремы Бернулли и интегральной теоремы Лапласа имеем приближенное равенство

$$P\left(\left|\frac{m}{n} - p\right| \langle \varepsilon\right) \approx \Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

Так как в последнем равенстве известна вероятность P, стоящая слева, то сначала решим уравнение $\Phi(t) = P$. Пусть t_p – корень уравнения, тогда $\varepsilon\sqrt{\dfrac{n}{pq}} \approx t_p$, $\varepsilon^2 \dfrac{n}{pq} \approx t_p^{\ 2}$ и, следовательно , $n \approx \dfrac{nqt^2 p}{\varepsilon^2}$. Для нашего случая $t_{0,997}$ =3, поэтому

$$n \approx \frac{nqt^2 p}{\varepsilon^2} = \frac{0{,}85 \cdot 0{,}15 \cdot 3^2}{(0.01)^2} = \frac{1{,}1475}{0{,}0001} = 11\ 475.$$

****УПРАЖНЕНИЯ**

1. Вероятность встретить на улице своего учителя, допустим, равна 0,002. Какова вероятность того, что среди 1 200 случайных прохожих вы встретите не более трех своих учителей?

Ответ: 0,7389.

2. Найти вероятность того, что число мальчиков среди 1 000 новорожденных больше 480, но меньше 540 (вероятность рождения мальчика принять равной 0,515).

Ответ: 0,929.

3. Вероятность выпуска радиолампы с дефектом равна 0,03. Найти максимально возможное отклонение ε частости от 0,03 среди 2 000 радиоламп, чтобы вероятность получить отклонение, по абсолютной величине меньше ε, была равна 0,999.

Ответ: 0,013.

1.14.8. Случайные величины, законы их распределения. Основные характеристики случайной величины

*ПРИМЕР 22. Мишень установлена так, что может вращаться вокруг оси (ось в центре). Мишень разбита на 8 секторов. При достаточно большой угловой скорости вращения стрелок не в состоянии различить цифры, выписанные по одной на секторах. Он вынужден стрелять наугад. При попадании в сектор 1 стрелок выигрывает 1 руб., в сектор 2 – 2 руб., в сектор 3 – 3 руб., в сектор 4 – 4 руб. и т. д., в сектор 8 – 8 руб. Стоит ли ему участвовать в такой игре, если за право стрелять один раз надо платить 5 руб.?

Решение. Поскольку мишень вращается, то способности стрелка здесь не имеют никакого значения: попадание – чистая случайность. Случайная величина ξ выражает возможные выигрыши. Она может принимать значения 1, 2, 3, 4, 5, 6, 7, 8.

Так как все секторы одинаковые, то каждое из этих значений – случайная величина – принимает с одинаковой вероятностью $\frac{1}{8}$. Значит, по формуле (1.11.1)

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i .$$

$$M(\xi) = 1\cdot\frac{1}{8} + 2\cdot\frac{1}{8} + 3\cdot\frac{1}{8} + 4\cdot\frac{1}{8} + 5\cdot\frac{1}{8} + 6\cdot\frac{1}{8} + 7\cdot\frac{1}{8} + 8\cdot\frac{1}{8} = 4{,}5 \text{ (руб.)}.$$

Итак, математическое ожидание выигрыша 4,5 руб., а стоимость выстрела 5 руб. Стрелять много раз явно не выгодно. На основании подобных расчетов в капиталистических странах организуются разнообразные азартные игры, приводящие игроков к разорению.

*ПРИМЕР 23. Дискретная случайная величина X задана законом распределения:

x_i	1	3
p_i	0,4	0,6

Найти начальные моменты первого, второго и третьего порядков.

Решение. Найдем начальный момент первого порядка (математическое ожидание X) по формулам (1.11.1), (1.118):

$$\alpha_1 = M(X) = 1\cdot 0{,}4 + 3\cdot 0{,}6 = 2{,}2.$$

Напишем закон распределения X^2:

X^2	1	9
p	0,4	0,6

Найдем начальный момент второго порядка (математическое ожидание X^2):

$$\alpha_2 = M(X^2) = 1\cdot 0{,}4 + 9\cdot 0{,}6 = 5{,}8.$$

Напишем закон распределения величины X^3:

X^3	1	27
p	0,4	0,6

Найдем начальный момент третьего порядка (математическое ожидание X^3):

$$\alpha_3 = M(X^3) = 1\cdot 0{,}4 + 27\cdot 0{,}6 = 16{,}6.$$

*ПРИМЕР 24. Случайная величина ξ распределена на промежутке [0, 2) по закону Симпсона:

$$f(x) = \begin{cases} x & \text{при} \quad 0 \le x < 1; \\ 2 - x & \text{при} \quad 1 \le x < 2. \end{cases}$$

Доказать, что

$$P(0{,}5 < \xi < 1{,}5) > \frac{1}{3}.$$

Решение. Для решения будем применять неравенство Чебышева (оно позволяет оценить вероятность отклонения случайной величины ξ от ее математического ожидания). Следовательно, нам необходимо знать $M(\xi)$, $D(\xi)$. По формуле (1.11.4)

$$M(\xi) = \int_0^1 x\cdot x\cdot dx + \int_1^2 x(2-x)\,dx = 1.$$

$$D(\xi) = \int_a^b x^2 f(x)dx - \left(\int_a^b xf(x)dx\right)^2 = \int_0^1 x^2\cdot x\cdot dx + \int_1^2 x^2(2-x)dx - 1^2 = \frac{1}{6}.$$

По неравенству Чебышева $P\left(|\xi - M(X)| \ge \varepsilon\right) \le \dfrac{D(\xi)}{\varepsilon^2}$, имеем

$$P\left(|\xi - 1| \langle\ 0{,}5\right) = P(1 - 0{,}5 \le \xi \le 1 + 0{,}5) = P(0{,}5 \le \xi \le 0{,}5) \ge 1 - \frac{1/6}{(0{,}5)^2} = \frac{1}{3.}$$

**УПРАЖНЕНИЯ

1. У дежурного гостиницы в кармане 8 различных ключей от разных комнат. Вынув наугад ключ, он пробует открыть дверь ближайшей комнаты. Сколько раз в среднем ему придется пробовать открывать эту комнату, если:

1) проверенный ключ кладется обратно в карман;

2) проверенный ключ не кладется обратно в карман?

Ответ: 8; 4-5.

2. У охотника 4 патрона. Он стреляет по зайцу, пока не попадет или пока не кончатся патроны. Найдите математическое ожидание количества выстрелов, если вероятность попадания примерно 0,025.

Ответ: 2,734.

3. Производятся три независимых опыта, в каждом из которых событие A появляется с вероятностью 0,4. Рассматривается случайная величина X – число появлений события A в трех опытах. Построить ряд распределения и функцию распределения случайной величины X. Найти ее математическое ожидание $M(X)$, дисперсию $D(X)$, среднее квадратическое отклонение $\sigma(X)$.

Ответ: Ряд распределения

x_i	0	1	2	3
p_i	0,216	0,432	0,288	0,064

$$M(X) = 1{,}2; \quad D(X) = 0{,}72; \quad \sigma(X) = 0{,}85.$$

4. Случайная величина ξ распределена на промежутке $[0, 2)$ с плотностью $f(x) = 1 - \dfrac{x}{2}$. Найти $M(\xi)$.

Ответ: $\dfrac{2}{3}$.

5. Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при} \quad x = 0; \\ (1 - \cos x)/2 & \text{при} \quad 0 \angle x \le \pi; \\ 1 & \text{при} \quad x > \pi. \end{cases}$$

Найти плотность распределения.

Ответ: $f(x) = (\sin x)/2$ в интервале $(0, \pi)$; вне этого интервала $f(x)=0$.

6. Случайная величина X распределена по нормальному закону (1.12.15) с $a = 0$ и $\sigma = 1$. Найти плотность распределения вероятностей случайной величины $У=X^{-1}$. Построить график полученной случайной величины.

Ответ: $\left[y^2 \sqrt{2\pi} e^{\frac{1}{y^2}} \right]^{-1}$.

7. Ребро куба x измерено приближенно, причем $a \le x \le b$. Рассматривая длину ребра куба как случайную величину ξ, распределенную равномерно в интервале (a, b), найти математическое ожидание и дисперсию объема куба.

Ответ: $M\xi^3 = \dfrac{(b+a)(b^2+a^2)}{4}$; $\quad D\xi^3 = \dfrac{b^7 - a^7}{7(b-a)} - \left[\dfrac{(b+a)(b^2+a^2)}{4} \right]^2$.

8. (распределение Коши) Функция распределения случайной величины ξ задана формулой $F(x) = A + B\mathrm{arctg}(x)$, $(-\infty < x < +\infty)$.

Найти: 1) постоянные A и B; 2) плотность распределения вероятности $f(x)$; 3) вероятность того, что величина ξ попадет в отрезок $[-1, 1]$; 4) математическое ожидание и дисперсию; 5) моду и медиану случайной величины ξ.

Ответ: 1) $A = \dfrac{1}{2}$; $B = \dfrac{1}{\pi}$; 2) $f(x) = \dfrac{1}{\pi(1+x^2)}$; 3) $\dfrac{1}{2}$; 4) $M(\xi)$ и $D(\xi)$ не существуют, так как определяющие их интегралы расходятся (найти убедиться); 5) $M_0 = M_c = 0$.

1.14.9. Системы случайных величин

*ПРИМЕР 25. Система непрерывных случайных величин $(X,У)$ задана в круге $x^2+y^2 \le R^2$ плотностью распределения $f(x,y) = \dfrac{2}{\pi R^4}(x^2+y^2)$. Найти математическое ожидание случайной величины $Z=X^2 \cdot Y^2$.

Решение. Применяя формулу для нахождения математического ожидания, получим

$$m = \frac{2}{\pi R^4} \iint\limits_{(D)} x^2 y^2 (x^2 + y^2) dxdy,$$

где D – круг $x^2 + y^2 \le R^2$.

Перейдем к полярной системе координат: $x = r\cos\varphi$, $y = r\sin\varphi$:

$$m = \frac{2}{\pi R^4} \int\limits_0^R r^7 dr \int\limits_0^{2\pi} \cos^2\varphi \sin^2\varphi\, d\varphi = \frac{R^4}{16\pi} \int\limits_0^{2\pi} \sin^2 2\varphi\, d\varphi = \frac{R^4}{16\pi} \int \frac{1-\cos 4\varphi}{2} d\varphi.$$

****УПРАЖНЕНИЯ**

1. Система случайных величин (X, $У$) задана плотностью распределения вероятностей

$$f(x, y) = \begin{cases} c\ (x + y) & \text{при } 0 \le y \le x \le 1; \\ 0 & \text{в остальных точках}. \end{cases}$$

Найти константу c и $P(X + У < 1)$.

Ответ: $c = 2$; $P(X + У < 1) = \frac{1}{3}$.

2. Найти вероятность того, что в результате испытания составляющая X двумерной случайной величины $(X, У)$ примет значение $X < 2$ и при этом составляющая $У$ примет значение $У < 3$, если известна функция распределения системы

$$F_{X,Y}(x, y) = \left(\frac{1}{\pi}\mathrm{arctg}\frac{x}{2} + \frac{1}{2}\right) \left(\frac{1}{\pi}\mathrm{arctg}\frac{y}{3} + \frac{1}{2}\right).$$

Ответ: $\frac{9}{16}$.

Часть 2

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

2.1. Предмет и задачи математической статистики

Как известно, теория вероятностей изучает математические модели случайных явлений, при этом сама математическая модель считается заданной. Так, например, если изучается некоторое случайное событие A, то известно $P(A)$. Если же речь идёт о случайной величине X, то известен закон распределения вероятностей в какой-либо форме и, как следствие, числовые характеристики исследуемой случайной величины. В практических задачах эти характеристики, как правило, неизвестны, но имеются некоторые экспериментальные данные о событии или случайной величине. Требуется на основании этих данных построить подходящую вероятностную модель изучаемого явления, т. е. приближенно оценить неизвестные закон распределения и числовые характеристики исследуемой случайной величины на основе экспериментальных данных. Это и является задачей математической статистики.

Математической статистикой называется наука, занимающаяся методами обработки экспериментальных данных, полученных в результате наблюдений над случайными явлениями. При этом решаются следующие в порядке сложности и важности задачи:

- описание явлений, т. е. упорядочение поступившего статистического материала, представление его в наиболее удобном для обозрения и анализа виде (таблицы, графики);
- анализ и прогноз, т. е. приближенная оценка на основании статистических данных интересующих нас характеристик, например, математического ожидания и дисперсии наблюдаемой случайной величины и определение погрешности этих оценок;
- выработка оптимальных решений, т. е., например, определение числа опытов n, достаточного для того, чтобы ошибка от замены теоретических числовых характеристик их экспериментальными оценками была не более заданной. В связи с этим возникает задача проверки правдоподобия гипотез о параметрах распределения и о законах распределения случайной величины, решением которой является возможность сделать один из выводов:

– отбросить гипотезу, как противоречащую опытным данным;
– принять гипотезу, считать ее приемлемой.

Математическая статистика, опираясь на размышляющий, оценивающий, сопоставляющий человеческий разум, помогает экспериментатору лучше разобраться в опытных данных, полученных в результате наблюдений над случайными явлениями; оценить, значимы или не значимы наблюдаемые факты; принять или отбросить те или иные гипотезы о природе случайных явлений.

2.2. Выборочный метод

2.2.1. Построение гистограммы и эмпирической функции распределения по выборке

Полный набор всех возможных N значений дискретной случайной величины X называют ***генеральной совокупностью***. Однако в реальных условиях нельзя рассчитывать на такую подробную информацию. Часть генеральной совокупности из n элементов, отобранных случайным образом, называется ***выборкой***, при этом число n называют ***объемом выборки***. Различают выборки малого объема ($n < 30$) и большого объема ($n > 30$).

В начале на основе результатов эксперимента строят ***простой статистический ряд*** – таблицу, состоящую из двух строк, в первой – порядковый номер измерения, во второй – его результат:

i	1	2	…	n
x_i	x_1	x_2	…	x_n

Для визуальной оценки распределения случайной величины производят группировку данных. Вначале x_k располагают в порядке возрастания, затем интервал наблюдаемых значений случайной величины разбивают на k последовательных непересекающихся частичных интервалов $\tilde{x}_0 \div \tilde{x}_1$, $\tilde{x}_1 \div \tilde{x}_2$, … $\tilde{x}_j \div \tilde{x}_{j+1}$, …, $\tilde{x}_{k-1} \div \tilde{x}_k$, далее подсчитывают числа n_j – количество x_k, попавших в j-й интервал. Полученный таким образом ***группированный статистический ряд*** отражают таблицами вида:

$\tilde{x}_j \div \tilde{x}_{i+1}$	$\tilde{x}_0 \div \tilde{x}_1$	$\tilde{x}_1 \div \tilde{x}_2$	. . .	$\tilde{x}_{k-1} \div \tilde{x}_k$
n_i	n_1	N_2	. . .	n_k

или, подсчитывая частоты $p_j = \dfrac{n_j}{n}$,

$\tilde{x}_j \div \tilde{x}_{j+1}$	$\tilde{x}_0 \div \tilde{x}_1$	$\tilde{x}_1 \div \tilde{x}_2$	...	$\tilde{x}_{k-1} \div \tilde{x}_k$
p_j	p_1	p_2	...	p_k

или, определяя середину j-го интервала $\bar{\bar{x}}_j = \tilde{x}_j + 0{,}5\Delta_j$, где $\Delta_j = \tilde{x}_{j+1} - \tilde{x}_j$ – длина j-го интервала, получим ряд распределения в виде

$\bar{\bar{x}}_j$	$\bar{\bar{x}}_1$	$\bar{\bar{x}}_2$	...	$\bar{\bar{x}}_k$
p_j	p_1	p_2	...	p_k

При этом частоты p_i удовлетворяют условию $\sum_{i=1}^{k} p_i = 1$. Деля частоту p_j на длину соответствующего интервала Δ_j, получим таблицу плотностей частоты f_j. Откладывая по оси абсцисс интервалы $\tilde{x}_j \div \tilde{x}_{j+1}$, и надстраивая на каждом интервале, как на основании, прямоугольник высотой f_j, т. е. площадью p_j, получим ступенчатую фигуру – ***гистограмму*** частот – статистический аналог кривой плотности распределения. Еще более точной оценкой кривой плотности распределения является ***полигон*** частот – ломаная, отрезки которой соединяют точки ($\bar{\bar{x}}_j$, f_j). В итоге ряд распределения можно представить таблицей вида

$\bar{\bar{x}}_j$	$\bar{\bar{x}}_1$	$\bar{\bar{x}}_2$	...	$\bar{\bar{x}}_k$
f_j	f_1	f_2	...	f_k

Иногда обходятся без группировки, но, располагая x_i в порядке возрастания, затем подсчитывают n_i – число повторов каждого значения случайной величины x_i и определяют в итоге частоту $p_i = \dfrac{n_i}{n}$, что приводит к ряду распределения вида

x_i	x_1	x_2	…	x_m
p_i	p_1	p_2	…	p_m

Здесь m – число разных значений наблюдаемой случайной величины.

Другим способом графического представления эмпирического закона распределения является *эмпирическая функция распределения* – оценка функции распределения дискретной случайной величины X, вычисляемая по формуле

$$F(x)=\sum_{x_i<x} p_i \qquad (2.2.1)$$

и являющаяся разрывной ступенчатой, равной нулю левее наименьшего наблюдаемого значения, испытывающей скачок величиной p_j при переходе через левую границу i-го интервала и в итоге достигающей единицы правее наибольшего наблюдаемого значения.

2.2.2. Точечные оценки математического ожидания и дисперсии

Характеристики случайной величины, построенные на основании выборочных данных, называются ***выборочными*** или ***точечными оценками***. Свойства случайной величины могут характеризоваться различными начальными и центральными моментами, вычисляемыми в случае дискретной случайной величины по следующим формулам.

Начальный момент порядка ***k***:

$$\alpha_k=\sum_i \overset{=}{x}{}_i^k p_i=\sum_{j=1}^m x_j^k p_j\,. \qquad (2.2.2)$$

Центральный момент порядка ***k***:

$$\mu_k=\sum_i (\overset{=}{x}_i-\alpha_1)^k p_i = \sum_{j=1}^m (x_j-\alpha_1)^k p_j\,. \qquad (2.2.3)$$

Важнейшие из них – ***математическое ожидание*** $\boldsymbol{M(X) = m_X}$ и ***дисперсия*** $\boldsymbol{D(X) = \sigma^2 (X)}$, где через $\boldsymbol{\sigma}$ обозначено ***среднеквадратическое отклонение,*** – являются частными случаями моментов:

$$\bar{x}=\alpha_1,\quad \overline{D}=\mu_2,\quad \bar{\sigma}=\sqrt{\overline{D}}\,. \qquad (2.2.4)$$

Выделяют также несмещенную выборочную дисперсию:

$$S^2=\frac{n}{n-1}\overline{D}\,. \qquad (2.2.5)$$

Если ***выборочное*** математическое ожидание случайной величины дает нам «ее среднее значение» или точку на координатной прямой, «вокруг которой разбросаны» значения рассматриваемой случайной величины, то ***выборочная*** дисперсия характеризует «степень разброса» значений

Используются также оценки коэффициента асимметрии:

$$A=\frac{\mu_3}{\sigma^3} \qquad (2.2.6)$$

и коэффициента эксцесса (островершинности):

$$E = \frac{\mu_4}{\sigma^4} - 3 \qquad (2.2.7)$$

как степени отклонения полигона частот от плотности нормального распределения непрерывной случайной величины, для которой они равны нулю.

*ПРИМЕР 1. По выборке объема n = 50 построен ряд распределения:

x_i	–2,25	–1,75	–1,25	–0,75	–0,25	0,25	0,75	1,25	1,75	2,25
p_i	0,02	0,04	0,11	0,18	0,27	0,16	0,10	0,07	0,03	0,02

Построить гистограмму, полигон и эмпирическую функцию распределения, а также найти выборочные математическое ожидание, дисперсию, среднеквадратичное отклонение, асимметрию и эксцесс.

Решение. Можно проконтролировать условие $\sum_{i=1}^{10} p_i = 1$. По формуле (2.2.2) находим

$$\alpha_1 = \sum_{j=1}^{10} x_j p_j = x_1p_1 + x_2p_2 + \ldots + x_{10}p_{10} =$$

$$= (-2{,}25)\cdot 0{,}02 + (-1{,}75)\cdot 0{,}04 + \ldots + 2{,}25\cdot 0{,}02 = -0{,}155;$$

тогда по формуле (2.2.3) определим

$$\mu_2 = \sum_{j=1}^{10}(x_j - \alpha_1)^2 p_j = 0{,}858475;$$

$$\mu_3 = \sum_{j=1}^{10}(x_j - \alpha_1)^3 p_j = 0{,}235727;$$

$$\mu_4 = \sum_{j=1}^{10}(x_j - \alpha_1)^4 p_j = 2{,}23323;$$

наконец, по формулам (2.2.4) – (2.2.7) находим окончательно выборочные числовые характеристики: математическое ожидание, дисперсию и среднеквадратичное отклонение (смещенные и несмещенные), асимметрию и эксцесс:

$$\overline{x} = \alpha_1 = -0{,}155;$$

$$\overline{D} = \mu_2 = 0{,}858475; \qquad \overline{\sigma} = \sqrt{\overline{D}} = 0{,}92654;$$

$$S^2 = \frac{n}{n-1}\overline{D} = 0{,}875995; \qquad S = 0{,}935946;$$

$$A = \frac{\mu_3}{\sigma^3} = 0{,}29636; \qquad E = \frac{\mu_4}{\sigma^4} - 3 = 0{,}03025.$$

Построим гистограмму. Для этого построим массив $\{x_i, f_i\}$, деля частоту p_i на длину соответствующего интервала $\Delta_i = 0{,}5$; в результате получим таблицу плотностей частоты f_i.

x_i	–2,25	–1,75	–1,25	–0,75	–0,25	0,25	0,75	1,25	1,75	2,25
f_i	0,04	0,08	0,22	0,36	0,54	0,32	0,20	0,14	0,06	0,04

Откладывая на оси абсцисс интервалы с центрами в точках x_i и шириной $\Delta_i = 0{,}5$, и надстраивая на каждом интервале, как на основании, прямоугольник высотой f_j, т. е. площадью p_j, получим ступенчатую фигуру – гистограмму – статистический аналог кривой плотности распределения. Здесь и ниже точками выделен массив $\{x_i, f_i\}$ (рис. 2.2.1).

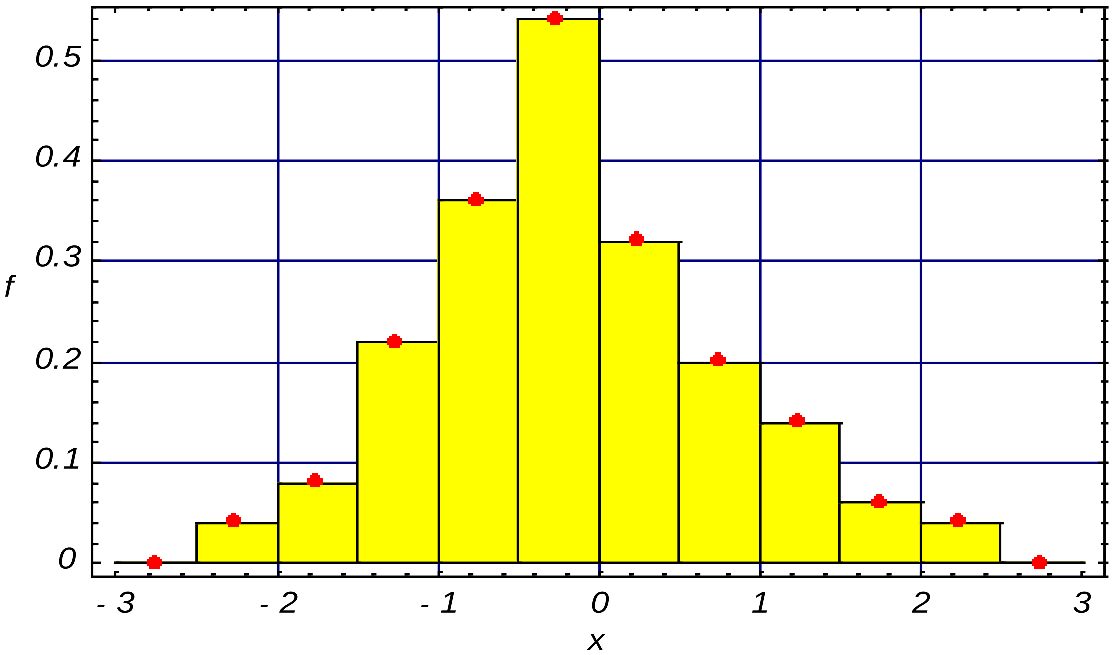


Рис. 2.2.1. Гистограмма

Построим полигон частот – еще более точную оценку кривой плотности распределения – ломаную, отрезки которой соединяют точки (x_i, f_i) (см. рис. 2.2.2).

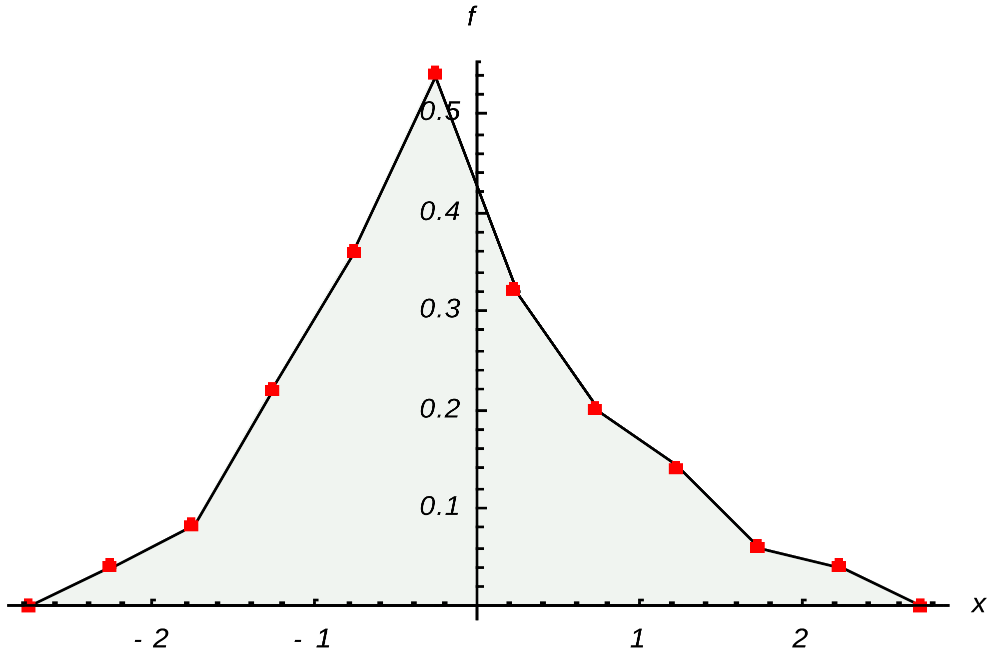


Рис. 2.2.2. Полигон

Если через точки массива $\{x_i, f_i\}$ провести плавную кривую, то мы получим самое точное приближение кривой плотности распределения $f(x)$ (рис. 2.2.3).

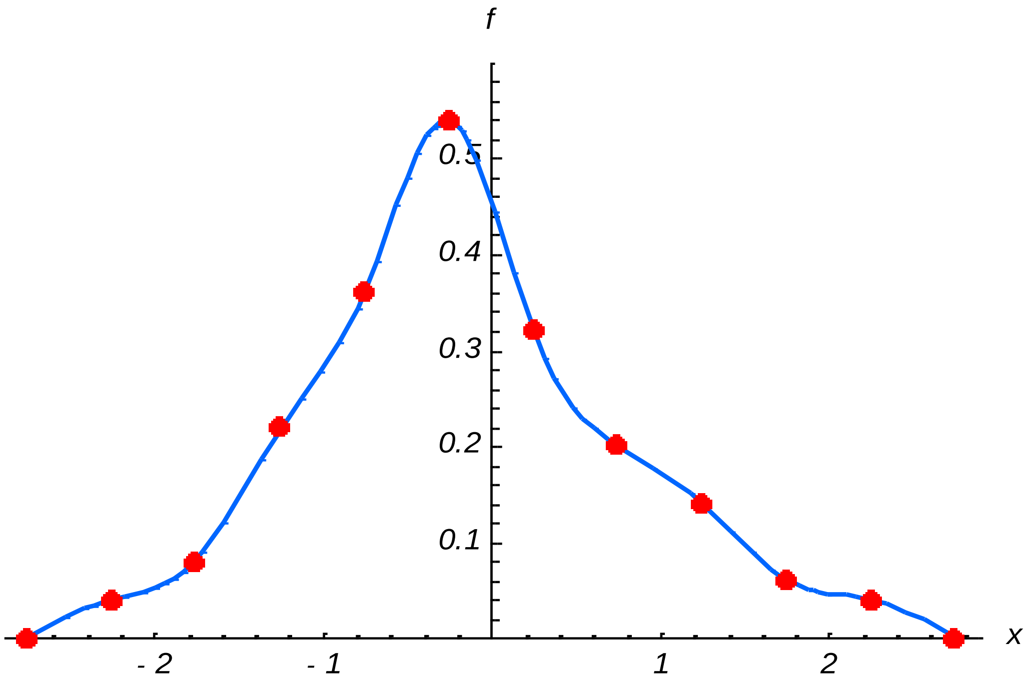


Рис. 2.2.3. Эмпирическая плотность распределения

В связи с этим представляет интерес сопоставление полученного эмпирического распределения с теоретическим распределением по нормальному закону $N(m_x, \sigma)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-m_x)^2}{2\sigma^2}\right],$$

график плотности которого для $m_x = 0$ и $\sigma = 1$ изображен на рис. 2.2.4 тонкой линией.

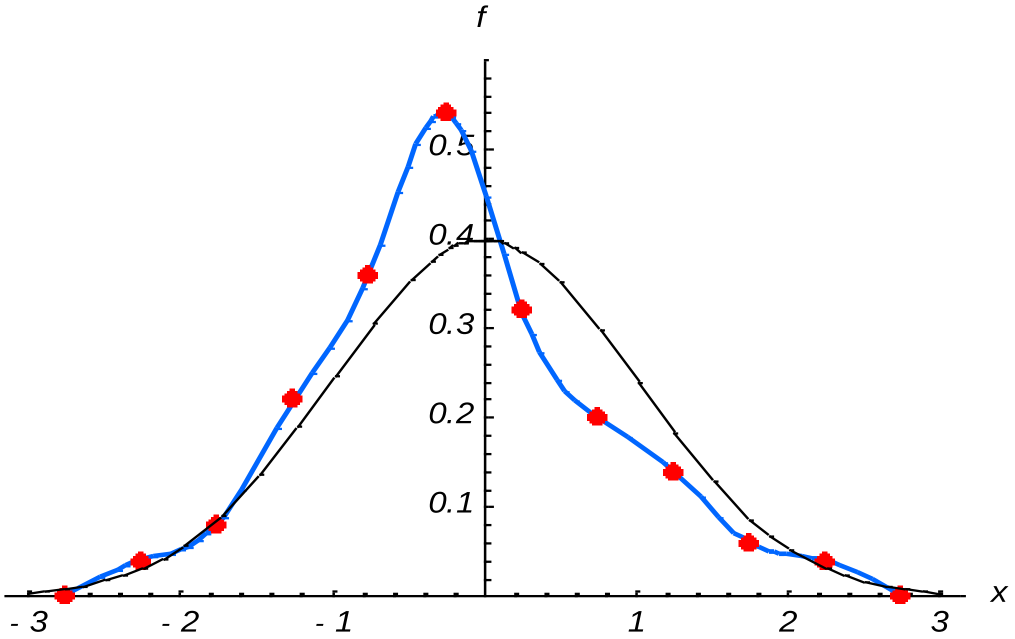


Рис. 2.2.4

Близость полученного эмпирического распределения и теоретического нормального распределения $N(0, 1)$ подтверждается не только графически на рис. 2.2.4, но и сравнением числовых характеристик:

Числовые характеристики	Распределения	
	Эмпирическое	Нормальное
m	–0,155	0
D	0,858475	1
σ	0,92654	1
A	0,29636	0
E	0,03025	0

Причем на фоне кривой плотности нормального распределения график эмпирической плотности распределения деформирован влево (выборочное математическое ожидание $\bar{x} = -0{,}155 < 0$, выборочная асимметрия $A = 0{,}29636 > 0$), сужен (выборочная дисперсия $D = 0{,}858475 < 1$) и вытянут вверх (выборочный эксцесс $E = 0{,}03025 > 0$).

Строим эмпирическую функцию распределения по формуле (2.2.1); в результате получим интегральный ряд распределения:

x_i	–2,25	–1,75	–1,25	–0,75	–0,25	0,25	0,75	1,25	1,75	2,25
$F(x)$	0,02	0,06	0,17	0,35	0,62	0,78	0,88	0,95	0,98	1,00

Эмпирическая функция распределения является разрывной ступенчатой функцией, равной нулю левее левой границы $x = -2{,}5$ интервала наименьшего наблюдаемого значения, испытывающей скачок величиной p_j при переходе через левую границу i-го интервала и в итоге достигающей единицы на последнем интервале наибольшего наблюдаемого значения (рис. 2.25).

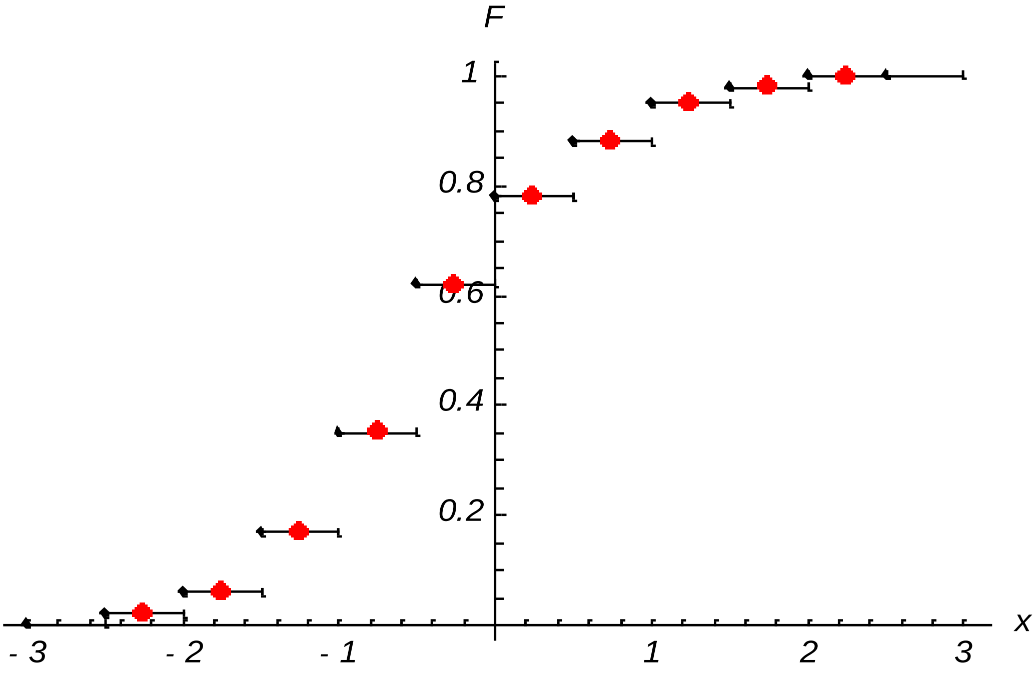


Рис. 2.2.5

На рис. 2.2.5 жирными точками выделен массив $\{x_i, F(x_i)\}$.

Проводя через точки массива $\{x_i, F(x_i)\}$ плавную кривую, получим аналог кривой функции распределения (2.2.6).

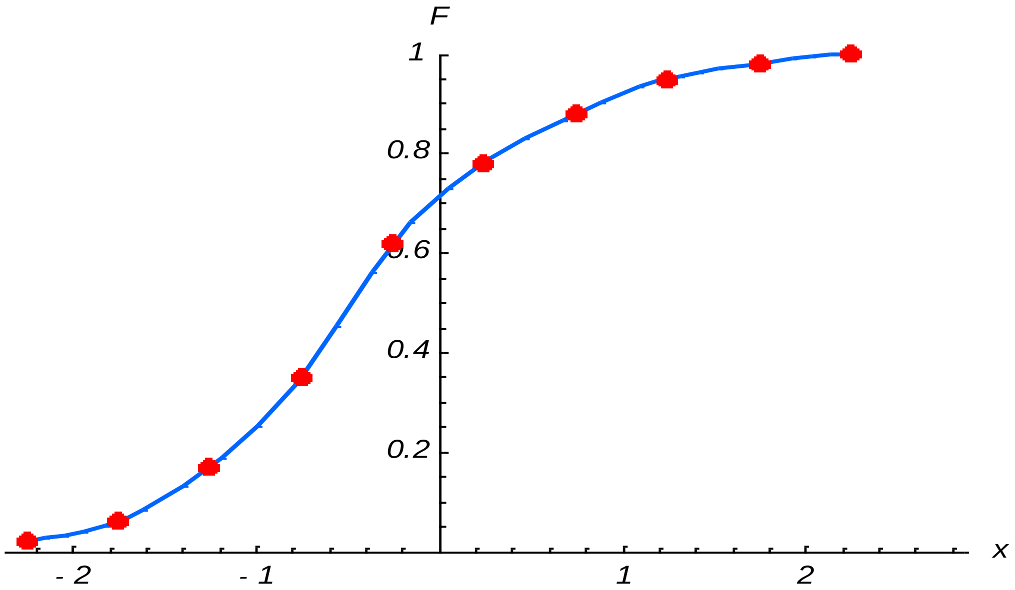


Рис. 2.2.6

2.3. Интервальные оценки

Рассмотрим результаты эксперимента в форме ***простого статистического ряда*** – таблицы, состоящей из двух строк, в первой – порядковый номер измерения, во второй – его результат:

i	1	2	…	n
x_i	x_1	x_2	…	x_n

В предыдущем параграфе был рассмотрен вопрос о нахождении выборочных числовых характеристик случайной величины – приближенных значений параметров распределения. Чтобы охарактеризовать погрешность этих значений, нужно указать граничные значения, за которые не выходит оцениваемый параметр. Поскольку все расчёты производятся на основании случайных результатов опыта, то и граничные значения также случайные величины. Таким образом, речь идёт о построении интервала со случайными границами, который с заданной вероятностью содержал бы неизвестное значение параметра распределения.

Для определения погрешности полученных значений используют **интервальные оценки**, применяя понятие «**доверительного интервала**» – интервала, внутри которого параметр, как ожидается, найдется с некоторой доверительной вероятностью (надежностью) β. Иногда вместо β используют величину $\alpha = 1 - \beta$, называемую уровнем значимости.

2.3.1. Доверительный интервал для математического ожидания

Рассмотрим нахождение доверительного интервала для математического ожидания m_x нормально распределенной случайной величины. Ширина 2ε такого интервала ($\bar{x}$ – ε, $\bar{x}$ + ε), обладающего симметрией относительно $\bar{x}$ – выборочного значения m_x, находится из условия

$$P(|x-\bar{x}|<\varepsilon)=\beta, \qquad (2.3.1)$$

причем сама вероятность $P(|x-\bar{x}|<\varepsilon)$ определяется законом нормального распределения, если известна дисперсия $D = \sigma^2$, и законом распределения Стьюдента со степенью свободы $k = n - 1$, если дисперсия не известна, а лишь подсчитано ее несмещенное значение $\tilde{D}_X = s^2$. На рис. 2.3.1 приведены сравнительные графики плотностей распределения:

- нормального с $m_x = 0$, $\sigma = 1$ (пунктирная линия) – $f(x)$ для $N(0,1)$;
- Стьюдента с $k = 5$ (тонкая сплошная линия) – $f_{\mathrm{St}}(x)$ для St(5);
- Стьюдента с $k = 1$ (толстая сплошная линия) – $f_{\mathrm{St}}(x)$ для St(1).

Как видно из графиков, с увеличением степени свободы k, т. е. с увеличением объема выборки, распределение Стьюдента стремится к нормальному.

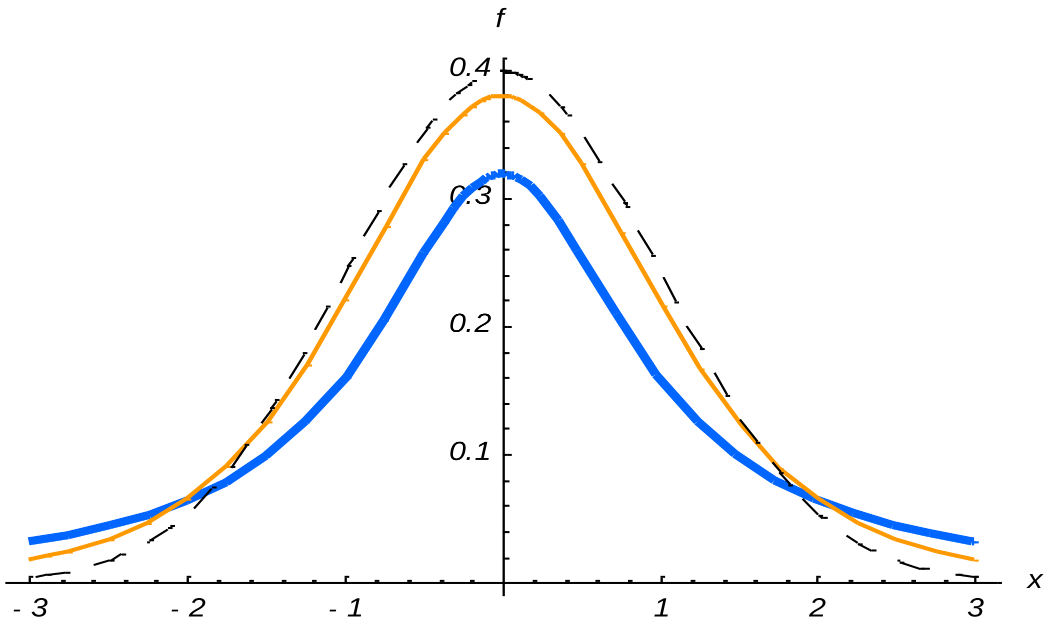


Рис. 2.3.1

Рассмотрим нахождение доверительного интервала для математического ожидания, если известна дисперсия $D = \sigma^2$. В этом случае вероятность β покрытия математического ожидания m_X доверительным интервалом ($\bar{x}$ – ε, $\bar{x}$+ε), определенная в (2.3.1), вычисляется, согласно закону нормального распределения, по формуле

$$\beta = 2\Phi(t), \qquad (2.3.2)$$

где $\Phi(t) = \int_0^t f(x)dx = \frac{1}{\sqrt{2\pi}}\int_0^t \exp(-\frac{x^2}{2})dx$ – функция Лапласа, $t = \frac{\varepsilon\sqrt{n}}{\sigma}$.

Таким образом, для отыскания границ доверительного интервала сначала по таблице для функции Лапласа (см. прил. 3) находим то значение t, для которого $\Phi(t)$ = 0,5β, а затем из условия $t = \frac{\varepsilon\sqrt{n}}{\sigma}$ находим $\varepsilon = \frac{t\sigma}{\sqrt{n}}$. В результате определяется доверительный интервал ($\bar{x} - \frac{t\sigma}{\sqrt{n}}$, $\bar{x} + \frac{t\sigma}{\sqrt{n}}$) такой, что с вероятностью β выполняется неравенство

$$\bar{x} - \frac{t\sigma}{\sqrt{n}} < m_x < \bar{x} + \frac{t\sigma}{\sqrt{n}}.$$

*ПРИМЕР 1. Найти доверительный интервал с надежностью β = 0,90 неизвестного математического ожидания нормальной случайной величины X, если σ = 2, $\bar{x}$ = 20,9, n = 16.

Решение. В нашем случае $\Phi(t)$ = 0,5β = 0,45. По таблице для функции Лапласа находим соответствующее значение t =1,645. На рис. 2.3.2 затемненной областью на фоне графика плотности нормального распределения выделена площадь, численно равнаяβ, согласно (2.3.2):

$$\beta = 2\Phi(t) = 2\int_0^t f(x)dx = 2\frac{1}{\sqrt{2\pi}}\int_0^t \exp(-\frac{x^2}{2})dx.$$

На рис. 2.3.3 на фоне графика функции Лапласа выделена точка (t, $\Phi(t)$).

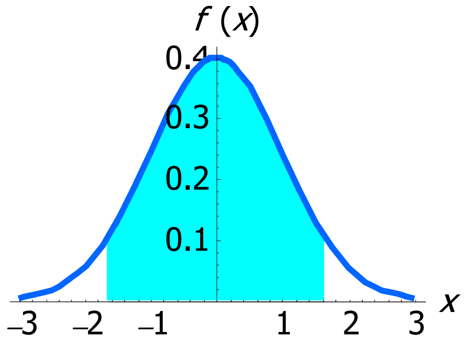


Рис. 2.3.2

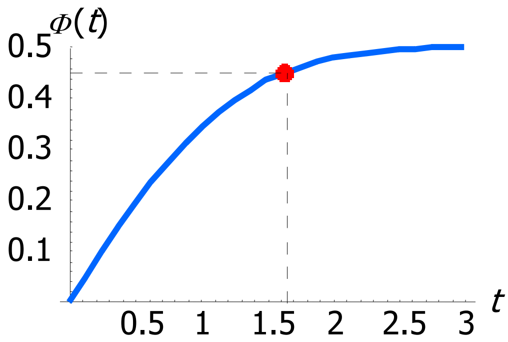


Рис. 2.3.3

Следовательно, $\varepsilon = \frac{t\sigma}{\sqrt{n}} = \frac{1{,}645\cdot 2}{\sqrt{16}} = 0{,}8225$. Таким образом, с вероятностью β = 0,90 интервал (20,0775; 21,7225) покрывает математического ожидания m_x.

Рассмотрим нахождение доверительного интервала для математического ожидания, если дисперсия не известна, а лишь подсчитано ее несмещенное значение $\tilde{D}_X = s^2$. В этом случае вероятность β покрытия математического ожидания m_x доверительным интервалом ($\bar{x} - \varepsilon$, $\bar{x} + \varepsilon$), определенная в (2.3.1), вычисляется, согласно закону распределения Стьюдента со степенью свободы $k = n - 1$, по аналогичной предыдущему случаю формуле. Имеются таблицы (см. прил. 4), позволяющие по значениям k и $\alpha = 1 - \beta$, найти соответствующее значение t_β, а из условия $t_\beta = \frac{\varepsilon\sqrt{n}}{s}$ найти $\varepsilon = \frac{t_\beta s}{\sqrt{n}}$. В результате строится доверительный интервал ($\bar{x} - \frac{t_\beta s}{\sqrt{n}}$, $\bar{x} + \frac{t_\beta s}{\sqrt{n}}$), содержащий параметр m_x с вероятностью β.

*ПРИМЕР 2. В примере предыдущего параграфа по данным выборки объема n = 50 были найдены выборочные $\bar{x}$ = –0,155 и s = 0,936. Найти доверительный интервал с надежностью β = 0.95 неизвестного математического ожидания нормальной случайной величины X.

Решение. В данном случае при n = 50 и β = 0,95 по таблице распределения Стьюдента (см. прил. 4) для k = 49 и $\alpha = 1 - \beta$ = 0,05 находим t_β = 2,009. На рис. 2.3.4 затемненной областью на фоне графика плотности распределения Стьюдента выделена площадь, численно равная β согласно формуле, аналогичной (2.3.2):

$$\beta = 2\Phi_{St}(t) = 2\int_0^t f_{St}(x)dx.$$

На рис. 2.3.5 на фоне графика функции $\Phi_{St}(t)$, аналогичной функции Лапласа, для распределения Стьюдента выделена точка (t, $\Phi_{St}(t)$).

Вычисляем далее по формуле $\varepsilon = \frac{t_\beta s}{\sqrt{n}} = \frac{2{,}009\cdot 0{,}936}{\sqrt{50}} \approx 0{,}266$ и записываем доверительный интервал (–0,155 – 0,266; –0,155 + 0,266). Таким образом, с вероятностью β = 0,95 справедливо неравенство –0,421 < m_x < 0,111.

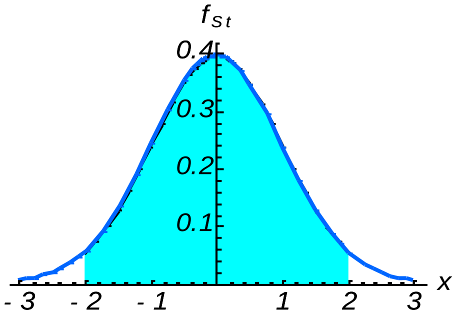


Рис. 2.3.4

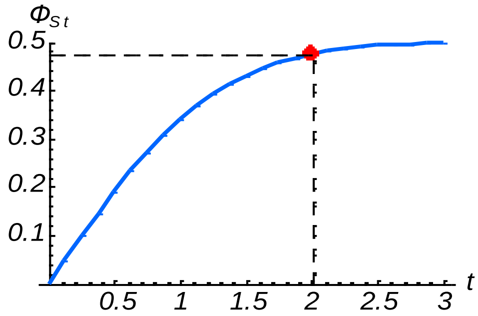


Рис. 2.3.5

2.3.2. Доверительный интервал для дисперсии

Можно аналогичным образом вычислять доверительный интервал для дисперсии $D = \sigma^2$, нормально распределенной случайной величины из условия:

$$P\left(\frac{(n-1)s^2}{z_2^2} < \sigma^2 < \frac{(n-1)s^2}{z_1^2}\right) = \beta = 1 - \alpha,$$

причем сама вероятность определяется законом χ^2 (хи – квадрат) – распределения со степенью свободы $k = n - 1$ таким образом, что

$$P(\chi^2 > z_2^2) = \frac{\alpha}{2} = \alpha_2; \qquad P(\chi^2 > z_1^2) = 1 - \frac{\alpha}{2} = \alpha_1.$$

С помощью таблиц χ^2-распределения (см. прил. 5) по вычисленным k и α_2 находят z_2^2, а по паре k и α_1 находят z_1^2. Для k >30 значение z^2 находят уже не из таблиц χ^2-распределения, а вычисляют по формуле

$$z^2 = \frac{1}{2}(\sqrt{2k-1} + t)^2.$$

Здесь t, определяемое равенством $\Phi(t) = \frac{1}{2}(1 - 2\alpha)$, вычисляется по таблице функции Лапласа (см. прил. 3).

Таким образом, $\left(\frac{(n-1)s^2}{z_2^2}; \frac{(n-1)s^2}{z_1^2}\right)$ есть доверительный интервал для σ^2 с надежностью $1 - \alpha$.

На рис. 2.3.6 приведены сравнительные графики плотностей распределения:

- нормального с $\mu = 0$, $\sigma = 1$ (пунктирная линия) – $f(x)$ для $N(0,1)$;
- χ^2-распределения с $k = 1$ (тонкая сплошная линия) – $f_{\text{Ch}}(x)$ для Ch(1);
- χ^2-распределения с $k = 10$ (толстая сплошная линия) – $f_{\text{Ch}}(x)$ для Ch(10).

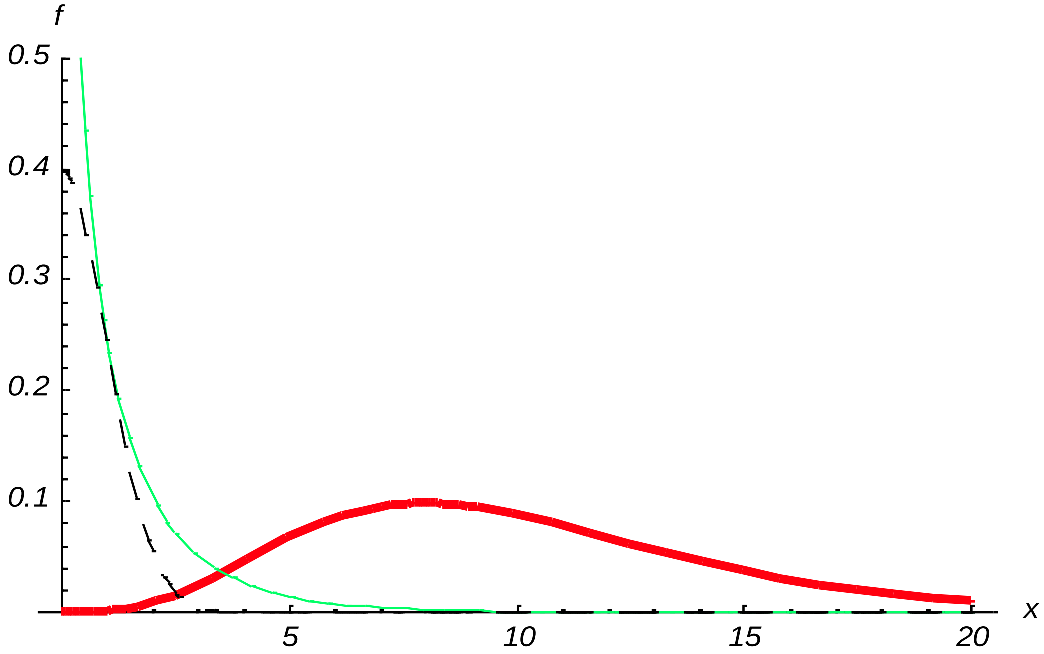


Рис. 2.3.6

*ПРИМЕР 3. По данным выборки объема $n = 20$ была найдена выборочная несмещенная дисперсия $S^2 = 0{,}876$; найти доверительный интервал, содержащий с надежностью $\beta = 0{,}90$ неизвестную дисперсию σ^2 нормальной случайной величины X.

Решение. В данном случае $n = 20$ и $\alpha = 1 - \beta = 1 - 0{,}90 = 0{,}10$, следовательно, $k = n - 1 = 19$, $\alpha_2 = \alpha/2 = 0{,}05$ и $\alpha_1 = 1 - \alpha/2 = 0{,}95$; с помощью таблиц χ^2-распределения (см. прил. 5) по $k = 19$ и $\alpha_2 = 0{,}05$ находим $z_2^2 = 30{,}1$, а по $k = 19$ и $\alpha_1 = 0{,}95$ находим $z_1^2 = 10{,}1$. На рис. 2.3.7 затемненной областью на фоне графика плотности χ^2-распределения выделена площадь, численно равная β, согласно формуле, аналогичной (2.3.2):

$$\beta = F_{\text{Ch}}(z_2^2) - F_{\text{Ch}}(z_1^2) = \int_{z_1^2}^{z_2^2} f_{\text{Ch}}(x)dx.$$

На рис. 2.3.8 на фоне графика функции χ^2-распределения $F_{\text{Ch}}(\chi^2)$ выделены точки $(z_1^2, F_{\text{Ch}}(z_1^2))$ и $(z_2^2, F_{\text{Ch}}(z_2^2))$.

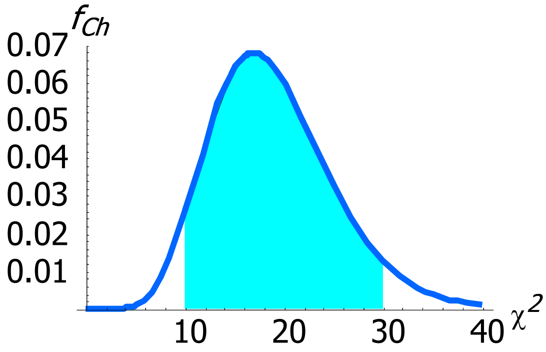


Рис. 2.3.7

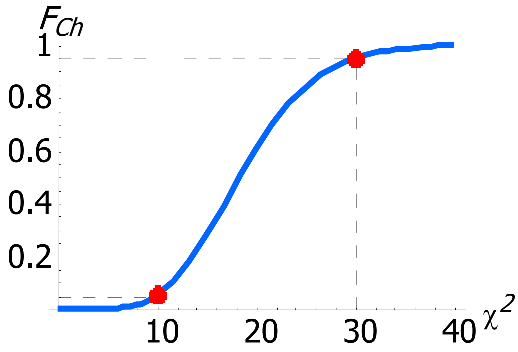


Рис. 2.3.8

На рис. 2.3.9 и рис. 2.3.10 затемненной областью на фоне графика плотности χ^2-распределения выделены площади, численно равные $\alpha_1 = 0,95$ и $\alpha_2 = 0,05$ соответственно, согласно формулам

$$P(\chi^2 > z_2^2) = \frac{\alpha}{2} = \alpha_2, \qquad P(\chi^2 > z_1^2) = 1 - \frac{\alpha}{2} = \alpha_1.$$

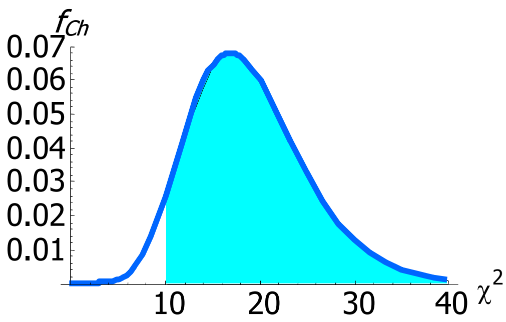


Рис. 2.3.9

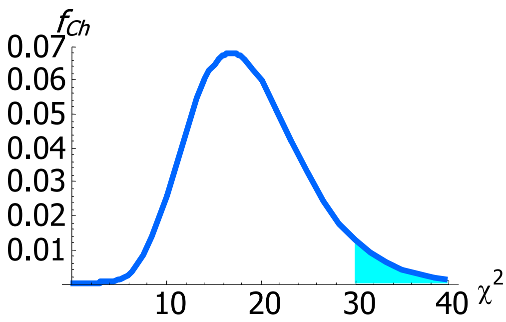


Рис. 2.3.10

Наконец, определяем границы доверительного интервала:

$$\frac{(n-1)s^2}{z_2^2} = \frac{19 \cdot 0,876}{30,1} \approx 0,553; \qquad \frac{(n-1)s^2}{z_1^2} = \frac{19 \cdot 0,876}{10,1} \approx 1,648.$$

2.4. Корреляционно-регрессионный анализ. Остаточная дисперсия

Для многих явлений в природе и технике типичны случайные зависимости. Случайные величины находятся в корреляционной зависимости, если каждому значению одной из них соответствует некоторое распределение другой, что математически отражается в уравнении регрессии одной случайной величины на другую.

По результатам эксперимента сначала оформляется таблица наблюдений системы дискретных случайных величин (X, Y) – ***матрица распределения*** – прямоугольная таблица, в которой записаны наблюдаемые значений для X: $\{x_1, x_2, \ldots, x_n\}$, для Y: $\{y_1, y_2, \ldots, y_m\}$, и соответствующая каждой паре $\{x_i, y_k\}$ вероятность $p_{ik} = P\{X = x, Y = y_k\}$, удовлетворяющая условию

$$\sum_k \sum_i p_{ki} = 1. \tag{2.4.1}$$

Таблица 1

Y	X			
	x_1	x_2	…	x_n
y_1	p_{11}	p_{12}	…	p_{1n}
y_2	p_{21}	p_{22}	…	p_{2n}
…	…	…	…	…
y_m	p_{m1}	p_{m2}	…	p_{mn}

К подобной матрице распределения можно прийти либо в случае повторяющихся наблюдаемых значений (X, Y), либо посредством построения группированных распределений по аналогии с одномерным случаем в п. 2.2; в последнем случае $\{x_i, y_k\}$ – центры соответствующих интервалов.

Одномерные законы распределения отдельных компонент выражаются формулами

$$p_i = P\{X = x_i\} = \sum_{k=1}^{m} p_{ki}\,, \quad p_k = P\{Y = y_k\} = \sum_{i=1}^{n} p_{ki}\,. \tag{2.4.2}$$

После чего можно привести более полный вариант табл. 1, расширенный одномерными законами распределения:

Таблица 1*

Y	X				p_y
	x_1	x_2	…	x_n	
y_1	p_{11}	p_{12}	…	p_{1n}	$\sum_{i=1}^{n} p_{1i}$
y_2	p_{21}	p_{22}	…	p_{2n}	$\sum_{i=1}^{n} p_{2i}$
…	…	…	…	…	…
y_m	p_{m1}	p_{m2}	…	p_{mn}	$\sum_{i=1}^{n} p_{mi}$
P_x	$\sum_{k=1}^{m} p_{k1}$	$\sum_{k=1}^{m} p_{k1}$	…	$\sum_{k=1}^{m} p_{k1}$	$\sum_{i=1}^{n}\sum_{k=1}^{m} p_{ki} = 1$

Система двух случайных величин (X, Y) характеризуется набором следующих моментов:

– начальный момент порядка $l + s$:

$$\alpha_{l,s} = \sum_k \sum_i x_i^l y_k^s p_{ki}\,; \qquad (2.4.3)$$

– центральный момент порядка $l + s$:

$$\mu_{l,s} = \sum_k \sum_i (x_i - m_X)^l (y_k - m_Y)^s p_{ki}\,. \qquad (2.4.4)$$

В частности,

$$m_X = \sum_i \sum_k x_i\, p_{ki} = \sum_i x_i p_i\,,$$

$$m_Y = \sum_k \sum_i y_k p_{ki} = \sum_k y_k p_k\,, \qquad (2.4.5)$$

$$D_X = \sum_i \sum_k (x_i - m_X)^2 p_{ki} = \sum_i (x_i - m_X)^2 p_i = \sum_i x_i^2 p_i - m_X^2\,;$$

$$D_Y = \sum_k \sum_i (y_k - m_Y)^2 p_{ki} = \sum_k (y_k - m_Y)^2 p_k = \sum_k y_k^2 p_k - m_Y^2, \qquad (2.4.6)$$

$$K_{XY} = \sum_i \sum_k (x_i - m_X)\ (y_k - m_Y)\ p_{ki} = \sum_i \sum_k x_i y_k p_{ki} - m_X m_Y\,.$$

В общем случае Y и X связаны вероятностной зависимостью, справедливой лишь в среднем, так как при фиксированном значению $X = x$ зависимая переменная Y имеет случайный разброс (столбец значений) из-за ошибок измерения, влияния неучтенных факторов или других причин. Таким образом, фиксированному значению $X = x_i$ соответствует усредненное значение $Y_{x_i} = M[Y/X = x_i]$ – условное математическое ожидание, вычисляемое по формуле

$$Y_{x_i} = \tilde{y}_i = \frac{1}{p_i} \sum_{k=1}^{m} y_k p_{ki}\,. \qquad (2.4.7)$$

В итоге исходная таблица $\{x_i, y_k\}$ эквивалентна таблице $\{x_i, \tilde{y}_i\}$ (табл. 2).

Таблица 2

x_i	x_1	x_2	…	x_n
$\tilde{y}_i$	$\tilde{y}_1$	$\tilde{y}_2$	…	$\tilde{y}_n$
p_i	p_1	p_2	…	p_n

Условное математическое ожидание $Y_x = M[Y/X = x]$ называется ***регрессией Y*** на ***X***, график зависимости $Y_x(x)$ называется линией регрессии. Аналогично определяется регрессия X на Y.

Рассмотрим модель линейной по параметрам регрессии Y на X, находящей линейную комбинацию $Y_x(x) = f(x) = \sum_{j=1}^{n_\beta} \beta_j f_j(x)$ базисных функций f_j, которая лучше всего, в смысле метода наименьших квадратов, аппроксимирует массив $\{x_i, \tilde{y}_i\}$. В этом случае результаты наблюдений представляются в виде

$$\tilde{y}_i = f(x_i) = \sum_j \beta_j f_j(x_i) + \varepsilon_i,$$

где ε_i – случайные некоррелированные ошибки наблюдений в предположении, что $M[\varepsilon_i] = 0$, $D[\varepsilon_i] = M[\varepsilon_i^2] = \sigma_i^2 = \sigma^2 w_i$, w_i – заданные «веса» дисперсий ошибок наблюдений. Таким образом, при выбранных базисных функциях f_j оценки $\overline{\beta}_j$ коэффициентов β_j определяются из условия

$$\varepsilon(\beta_j) = \sum_i \varepsilon_i^2 \tilde{p}_i = \sum_i [\tilde{y}_i - f(x_i)]^2 \tilde{p}_i = \min, \qquad (2.4.8)$$

где $\tilde{p}_i = \dfrac{p_i}{w_i}$.

Для регрессионной модели $f(x) = \beta_1 f_1(x) + \beta_2 f_2(x) + \beta_3 f_3(x)$ из необходимых условий минимума функции $\varepsilon(\beta_j)$ в (2.4.8) следует, что оценки $\overline{\beta}_j$ коэффициентов β_j являются решениями линейной системы алгебраических уравнений, обозначенной как (2.4.9):

$$\overline{\beta}_1 \sum_{i=1}^{n} f_1^2(x_i)\tilde{p}_i + \overline{\beta}_2 \sum_{i=1}^{n} f_2(x_i) f_1(x_i)\tilde{p}_i + \overline{\beta}_3 \sum_{i=1}^{n} f_3(x_i) f_1(x_i)\tilde{p}_i = \sum_{i=1}^{n} \tilde{y}_i f_1(x_i)\tilde{p}_i;$$

$$\overline{\beta}_1 \sum_{i=1}^{n} f_1(x_i) f_2(x_i)\tilde{p}_i + \overline{\beta}_2 \sum_{i=1}^{n} f_2^2(x_i)\tilde{p}_i + \overline{\beta}_3 \sum_{i=1}^{n} f_3(x_i) f_2(x_i)\tilde{p}_i = \sum_{i=1}^{n} \tilde{y}_i f_2(x_i)\tilde{p}_i;$$

$$\overline{\beta}_1 \sum_{i=1}^{n} f_1(x_i) f_3(x_i)\tilde{p}_i + \overline{\beta}_2 \sum_{i=1}^{n} f_2(x_i) f_3(x_i)\tilde{p}_i + \overline{\beta}_3 \sum_{i=1}^{n} f_3^2(x_i)\tilde{p}_i = \sum_{i=1}^{n} \tilde{y}_i f_3(x_i)\tilde{p}_i. \qquad (2.4.9)$$

В случае $D[\varepsilon_i] = M[\varepsilon_i^2] = \sigma^2 = \text{const}$ полагаем $w_i = 1$ и, таким образом, под $\tilde{p}_i$ подразумеваем просто p_i.

Качество аппроксимации результатов наблюдений регрессивной моделью определяется остаточной дисперсией

$$s^2 = \frac{\varepsilon}{n - n_\beta}, \qquad (2.4.10)$$

где n_β – число оцениваемых параметров β_j.

Корреляционная матрица оценок $\overline{\beta}_j$ равна произведению s^2 на матрицу, обратную по отношению к матрице коэффициентов при $\overline{\beta}_j$ в системе (2.4.9).

2.4.1. Простая линейная регрессия $f(x) = \beta_1 + \beta_2 x$

Рассмотрим простую линейную регрессию, которая считается выполненной $f(x) = \sum_{j=1}^{2} \beta_j x^{j-1} = \beta_1 + \beta_2 x$, если, выбирая базисные функции в виде $f_1 = 1$, $f_2 = x$, найдем оценки коэффициентов β_1 и β_2 из условия минимизации $\sum_i [\tilde{y}_i - \beta_1 - \beta_2 x_i]^2 p_i$, т. е. решения системы (2.4.9) в случае постоянства дисперсии ошибок наблюдения при дополнительном условии $\beta_3 = 0$:

$$\overline{\beta}_1 + \overline{\beta}_2 \sum_{i=1}^{n} x_i\, p_i = \sum_{i=1}^{n} \tilde{y}_i\, p_i;$$

$$\overline{\beta}_1 \sum_{i=1}^{n} x_i\, p_i + \overline{\beta}_2 \sum_{i=1}^{n} x_i^2\, p_i = \sum_{i=1}^{n} \tilde{y}_i x_i\, p_i. \qquad (2.4.9')$$

В этом случае $\overline{\beta}_1$ и $\overline{\beta}_2$ можно выразить через точечные оценки числовых характеристик системы дискретных случайных величин (X,Y):

$$f(x) = \overline{y} + \overline{r}_{xy} \frac{\overline{\sigma}_y}{\overline{\sigma}_x} (x - \overline{x}), \qquad (2.4.11)$$

где, согласно формулам (2.4.5)–(2.4.6),

$\overline{x} = \sum_{i=1}^{n} x_i p_i$ – оценка ***математического ожидания*** по массиву $\{x_i\}$;

$\overline{y} = \sum_{k=1}^{m} y_k p_k$ – оценка ***математического ожидания*** по массиву $\{y_k\}$;

$\overline{\sigma}_x^2 = \sum_i (x_i - \overline{x})^2 p_i = \overline{D}_x$ – оценка ***дисперсии*** D_x по массиву $\{x_i\}$;

$\overline{\sigma}_y^2 = \sum_k (y_k - \overline{y})^2 p_k = \overline{D}_y$ – оценка ***дисперсии*** D_y по массиву $\{y_k\}$;

$\overline{K}_{xy} = \sum_{i,k} (x_i - \overline{x})(y_k - \overline{y}) p_{ki}$ – оценка ***ковариации*** по массиву $\{x_i, y_k\}$;

$\overline{r}_{xy} = \dfrac{\overline{K}_{xy}}{\overline{\sigma}_x \overline{\sigma}_y}$ – ***коэффициент корреляции***, значение которого по модулю равно единице в случае линейной зависимости *Y* и *X*. Таким образом, $|\overline{r}_{xy}|$ характеризует степень тесноты линейной зависимости между *Y* и *X*, проявляющейся в том, что при возрастании одной случайной величины другая проявляет тенденцию также возрастать (в этом случае $\overline{r}_{xy} > 0$) или убывать (в таком случае $\overline{r}_{xy} < 0$). В первом случае говорят, что *Y* и *X* связаны положительной корреляцией, а во втором корреляция отрицательна. При этом зависимость тем ближе к линейному закону, чем $|\overline{r}_{xy}|$ ближе к единице слева. Если $\overline{r}_{xy} = 0$, то это означает только отсутствие линейной связи между *Y* и *X*, любой другой вид связи может при этом присутствовать.

*ПРИМЕР 1. По результатам эксперимента получена таблица наблюдений системы случайных величин (*X*, *Y*):

Таблица 3

Y	*X*					
	10	20	30	40	50	60
5	0,01	0,035	0,025	0,015	0,0	0,0
10	0,0	0,10	0,065	0,125	0,025	0,0
15	0,0	0,0	0,05	0,085	0,135	0,015
20	0,0	0,0	0,015	0,045	0,065	0,08
25	0,0	0,0	0,0	0,025	0,035	0,05

Оценить данную матрицу распределения системы случайных величин (*X*, *Y*) на простую линейную регрессию.

Решение. Данная задача может быть решена следующим образом.

Согласно (2.3.1), можно осуществить контроль вероятностей распределения

$$\sum_{i=1}^{6} \sum_{k=1}^{5} p_{ki} = 1.$$

Отобразим выборку графически в координатах $\{x, y\}$, при этом каждую пару $\{x_i, y_k\}$ изобразим кругом с центром в точке $\{x_i, y_k\}$ и радиусом, пропорциональным соответствующей вероятности p_{ik} (рис. 2.4.1).

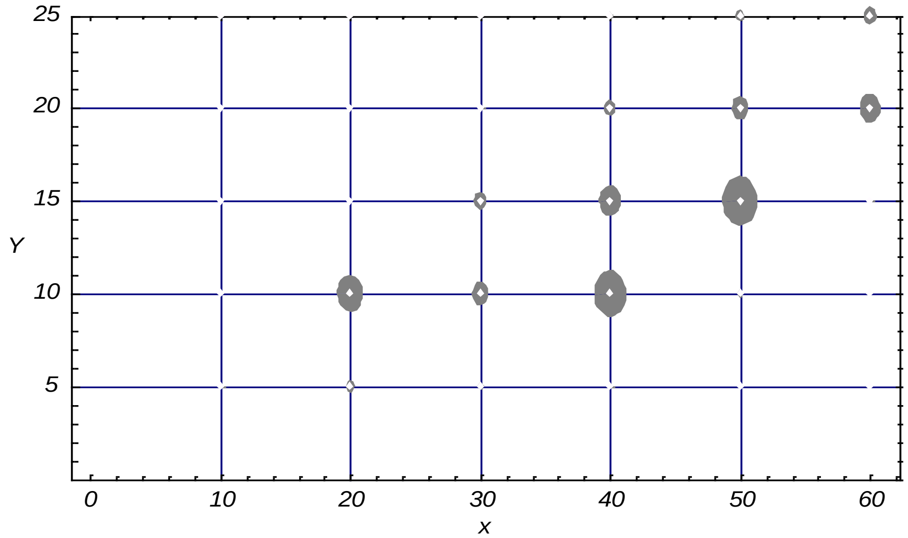


Рис. 2.4.1

Строим график вероятностей распределения (гистограмму системы случайных величин) (рис. 2.4.2).

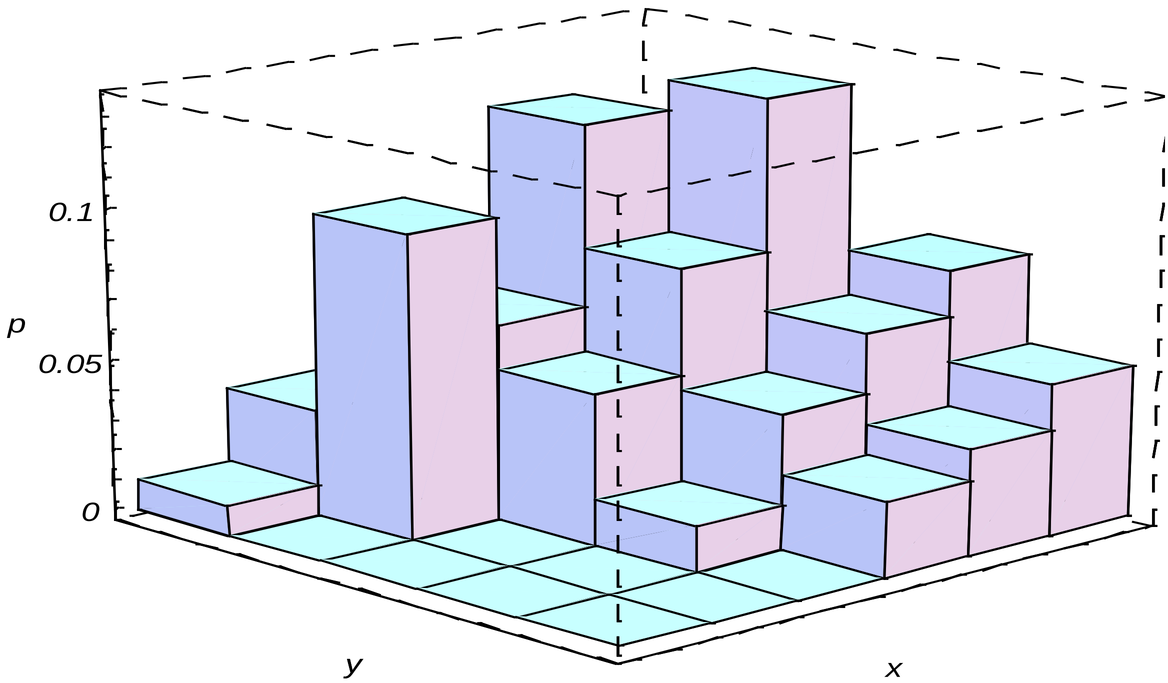


Рис. 2.4.2

Возможен плоский вариант гистограмм (наложение построчных распределений) (см. рис. 2.4.3).

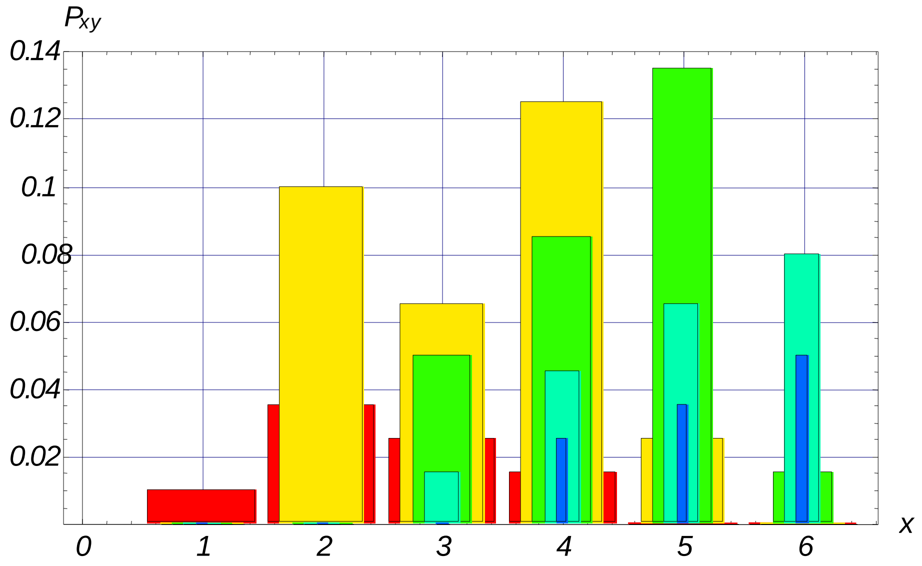


Рис. 2.4.3

Найдем одномерный закон распределения для Y по формуле (2.4.2):

$$p_k = P\{Y = y_k\} = \sum_{i=1}^{6} p_{ki} \,.$$

Таблица 4

y_k	5	10	15	20	25
p_k	0,085	0,315	0,285	0,205	0,11

Построим гистограмму частного распределения для Y (рис. 2.4.4).

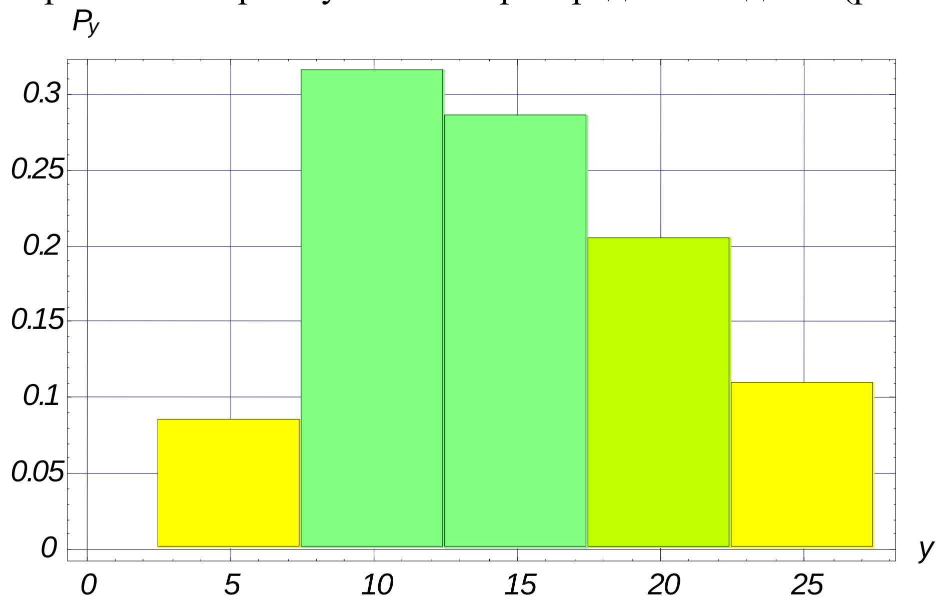


Рис. 2.4.4

Найдем одномерный закон распределения для X по формуле (2.4.2):

$$p_i = P\{X = x_i\} = \sum_{k=1}^{5} p_{ki} .$$

Таблица 5

x_i	10	20	30	40	50	60
p_i	0,01	0,135	0,155	0,295	0,26	0,145

Построим гистограмму частного распределения для X (рис. 2.4.5).

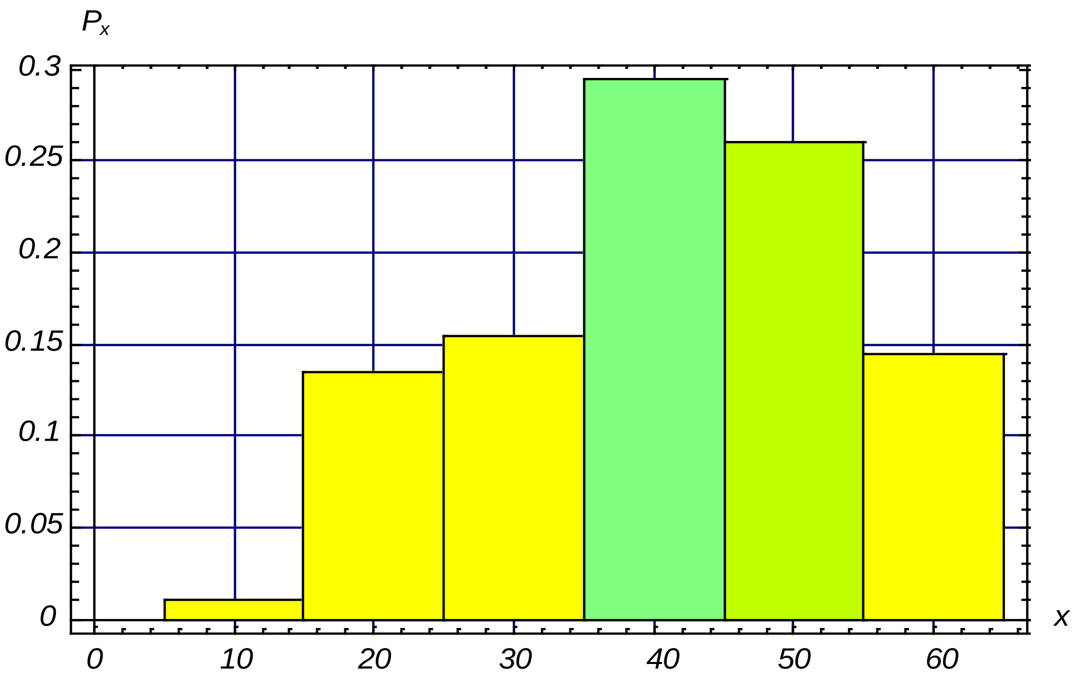


Рис. 2.4.5

После чего можно привести более полный вариант табл. 1, расширенный одномерными законами распределения.

Таблица 6

Y	X						p_Y
	10	20	30	40	50	60	
5	0,01	0,035	0,025	0,015	0,0	0,0	0,085
10	0,0	0,10	0,065	0,125	0,025	0,0	0,315
15	0,0	0,0	0,05	0,085	0,135	0,015	0,285
20	0,0	0,0	0,015	0,045	0,065	0,08	0,205
25	0,0	0,0	0,0	0,025	0,035	0,05	0,11
p_X	0,01	0,135	0,155	0,295	0,26	0,145	1,0

По формуле (2.4.7) можно построить условное математическое ожидание:

$$Y_{x_i} = \tilde{y}_i = \frac{1}{p_i}\sum_{k=1}^{5} y_k p_{ki},$$

например,

$$\tilde{y}_4 = (5\cdot 0{,}015+10\cdot 0{,}125+15\cdot 0{,}085+20\cdot 0{,}045+25\cdot 0{,}035)/0{,}295 \approx 13{,}983.$$

Тогда исходная табл. 6 для $\{x_i,\ y_k\}$ будет эквивалентна таблице $\{x_i,\ \tilde{y}_i\}$ (табл. 7).

Таблица 7

x_i	10	20	30	40	50	60
$\tilde{y}_i$	5	8,704	11,774	13,983	17,115	21,207
p_i	0,01	0,135	0,155	0,295	0,26	0,145

Вычислим оценки числовых характеристик системы $(X,\ Y)$: математического ожидания, дисперсии, среднеквадратичного отклонения, ковариации, корреляционной матрицы, коэффициента корреляции и нормированной корреляционной матрицы:

$$\bar{x} = \sum_{i=1}^{6} x_i p_i = 10\cdot 0{,}01+20\cdot 0{,}135+30\cdot 0{,}155+40\cdot 0{,}295+50\cdot 0{,}26+60\cdot 0{,}145=40{,}95;$$

$$\bar{y} = \sum_{k=1}^{5} y_k p_k = 5\cdot 0{,}085+10\cdot 0{,}315+15\cdot 0{,}285+20\cdot 0{,}205+25\cdot 0{,}11 = 14{,}7;$$

аналогично

$$\bar{\sigma}_x = \sqrt{\sum_{i=1}^{6}(x_i-\bar{x})^2 p_i} \approx 12{,}712;$$

$$\bar{\sigma}_y = \sqrt{\sum_{k=1}^{5}(y_k-\bar{y})^2 p_k} \approx 5{,}693;$$

$$\bar{K}_{xy} = \sum_{k=1}^{5}\sum_{i=1}^{6}(x_i-\bar{x})(y_k-\bar{y})p_{ki} = 48{,}785;$$

$$\bar{r}_{xy} = \frac{\bar{K}_{xy}}{\bar{\sigma}_x\bar{\sigma}_y} \approx 0{,}674.$$

По формуле $f(x) = \bar{y} + \bar{r}_{xy}\dfrac{\bar{\sigma}_y}{\bar{\sigma}_x}(x - \bar{x})$ выполним линейную регрессию

$$f(x) = 2{,}3375 + 0{,}3019\, x.$$

Построим график полученной линейной зависимости на фоне графически отображенной выборки (рис. 2.4.6).

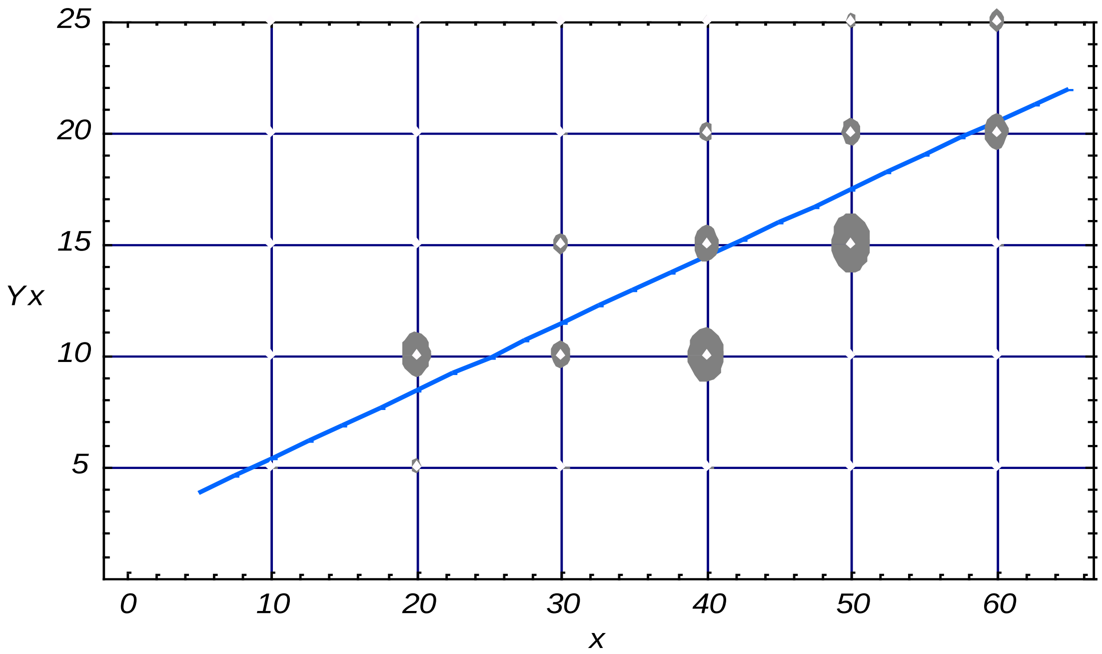


Рис. 2.4.6

Качество аппроксимации результатов наблюдений простой линейной регрессивной моделью определяется остаточной дисперсией по формуле (2.4.10):

$$s^2 = \frac{\varepsilon}{n - n_\beta} \approx \frac{0{,}2017}{6 - 2} \approx 0{,}0504.$$

2.4.2. Линейная полиномиальная регрессия $f(x) = \beta_1 + \beta_2 x + \beta_3 x^2$

Рассмотрим линейную (по параметрам β_j) регрессию, находящую линейную комбинацию $f(x) = \beta_1 + \beta_2 x + \beta_3 x^2$, которая лучше всего аппроксимирует массив $\{x_i,\ \tilde{y}_i\}$ в смысле метода наименьших квадратов, т. е. определяющую коэффициенты β_j из условия

$$\varepsilon = \sum_i [\tilde{y}_i - f(x_i)]^2 \tilde{p}_i = \min,$$

что приводит к частному случаю системы (2.4.9) при $f_1(x) = 1$, $f_2(x) = x$, $f_3(x) = x^2$:

$$\begin{aligned}
&\bar{\beta}_1 \sum_{i=1}^{n} \tilde{p}_i + \bar{\beta}_2 \sum_{i=1}^{n} x_i \tilde{p}_i + \bar{\beta}_3 \sum_{i=1}^{n} x_i^2 \tilde{p}_i = \sum_{i=1}^{n} \tilde{y}_i \tilde{p}_i;\\
&\bar{\beta}_1 \sum_{i=1}^{n} x_i \tilde{p}_i + \bar{\beta}_2 \sum_{i=1}^{n} x_i^2 \tilde{p}_i + \bar{\beta}_3 \sum_{i=1}^{n} x_i^3 \tilde{p}_i = \sum_{i=1}^{n} \tilde{y}_i x_i \tilde{p}_i; \qquad (2.4.9'')\\
&\bar{\beta}_1 \sum_{i=1}^{n} x_i^2 \tilde{p}_i + \bar{\beta}_2 \sum_{i=1}^{n} x_i^3 \tilde{p}_i + \bar{\beta}_3 \sum_{i=1}^{n} x_i^4 \tilde{p}_i = \sum_{i=1}^{n} \tilde{y}_i x_i^2 \tilde{p}_i.
\end{aligned}$$

*ПРИМЕР 2. По табл. 3 матрицы распределения системы случайных величин (X, Y) найти оценки параметров модели линейной полиномиальной регрессии вида

$$f(x) = \beta_1 + \beta_2 x + \beta_3 x^2.$$

Решение. В связи с тем, что в предыдущем примере коэффициент корреляции, характеризующий степень тесноты линейной зависимости между Y и X, $\bar{r}_{xy} \approx 0{,}674$ не очень близок к единице, оценим матрицу распределения системы случайных величин по табл. 3 на линейную полиномиальную регрессию вида

$$f(x) = \beta_1 + \beta_2 x + \beta_3 x^2.$$

В данном случае с учетом постоянства дисперсии ошибок наблюдений полагаем $\tilde{p}_i = p_i$ в системе (2.4.9″). Вычислим коэффициенты системы (2.4.9″), подсчитав предварительно суммы вида:

$$\sum_{i=1}^{6} x_i p_i = \bar{x} = 40{,}95; \qquad \sum_{i=1}^{6} x_i^2 p_i = 1838{,}5;$$

$$\sum_{i=1}^{6} x_i^3 p_i = 87975; \qquad \sum_{i=1}^{6} x_i^4 p_i = 4406650;$$

$$\sum_{i=1}^{6} \tilde{y}_i p_i = \bar{y} = 14{,}7; \qquad \sum_{i=1}^{6} \tilde{y}_i x_i p_i = 650{,}75;$$

$$\sum_{i=1}^{6} \tilde{y}_i x_i^2 p_i = 30912{,}3.$$

Подставляя вычисленные суммы коэффициентами в систему (2.4.9″) и разрешая последнюю относительно $\bar{\beta}_j$, получим

$$f(x) = 4{,}884 + 0{,}1567x + 0{,}00185x^2.$$

Построим графики $4{,}884 + 0{,}1567x + 0{,}00185x^2$ и $2{,}3375 + 0{,}3019\,x$ на фоне графически отображенной выборки $\{x_i,\ \tilde{y}_i\}$, которая изображена кружками с центром в точке $\{x_i,\ \tilde{y}_i\}$ и радиусами, пропорциональными соответствующей вероятности p_i (рис. 2.4.7).

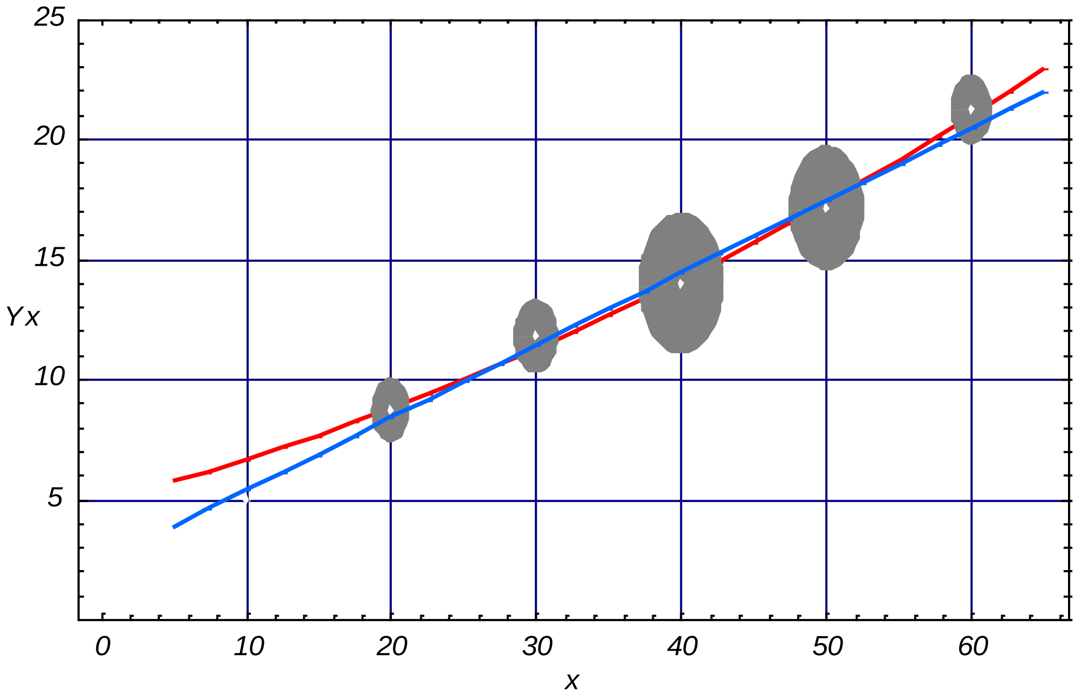


Рис. 2.4.7

Качество аппроксимации результатов наблюдений трехпараметрической линейной регрессивной моделью определяется остаточной дисперсией по формуле (2.4.10):

$$s^2 = \frac{\varepsilon}{n - n_\beta} \approx \frac{0{,}09812}{6-3} \approx 0{,}0327.$$

Из сравнения остаточных дисперсий последних двух регрессивных моделей следует, что последняя модель более адекватна результатам наблюдений.

*ПРИМЕР 3. Найти оценки параметров линейной полиномиальной регрессии вида $f(x) = \beta_1 + \beta_2 x + \beta_3 x^2$ по таблице результатов независимых экспериментов с соответствующими «весами» дисперсий ошибок наблюдений (табл. 8).

Таблица 8

X	–3	–2	–1	0	1	2	3
Y	–10	0	4	5	4	2	–2
w	1/2	2/3	1	1	1	2/3	1/2

Найти ковариационную матрицу оценок параметров β_j.

Решение. В связи с отсутствием повторов в таблице результатов, для данного случая имеем $\tilde{y}_i = y_i$, $p_i = 1/n = 1/7$, следовательно: $\tilde{p}_i = \dfrac{p_i}{w_i} = \dfrac{1}{7}\dfrac{1}{w_i}$.

Подсчитаем суммы в системе (2.3.9″):

$$\sum_{i=1}^{7}\tilde{p}_i = 10/7;\quad \sum_{i=1}^{7}x_i\,\tilde{p}_i = 0;\quad \sum_{i=1}^{7}x_i^2\tilde{p}_i = 50/7;\quad \sum_{i=1}^{7}\tilde{y}_i\tilde{p}_i = -8/7;$$

$$\sum_{i=1}^{7}x_i^3\tilde{p}_i = 0;\quad \sum_{i=1}^{7}x_i^4\tilde{p}_i = 374/7;\quad \sum_{i=1}^{7}\tilde{y}_i x_i\,\tilde{p}_i = 54/7;\quad \sum_{i=1}^{7}\tilde{y}_i x_i^2\tilde{p}_i = -196/7.$$

Тогда система (2.4.9″) примет вид:

$$\begin{aligned} &10\,\overline{\beta}_1 + \; + 50\,\overline{\beta}_3 = -8;\\ &50\,\overline{\beta}_2 \; = 54;\\ &50\,\overline{\beta}_1 + \; + 374\,\overline{\beta}_3 = -196. \end{aligned} \qquad (2.4.9''')$$

Разрешая эту систему относительно $\overline{\beta}_j$, получим

$$f(x) = 5{,}49 + 1{,}08x - 1{,}258x^2.$$

Построим график полученной зависимости на фоне графически отображенной выборки (см. рис. 2.4.8).

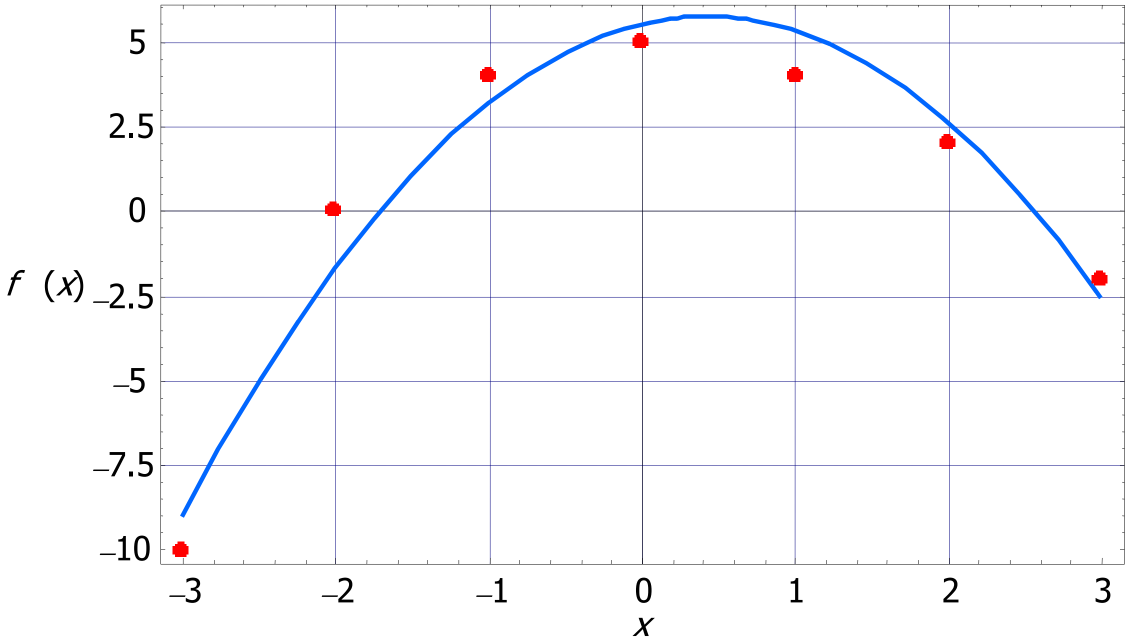


Рис. 2.4.8

Добавим в табл. 8 вычисленные значения *f*(*X*).

Таблица 8′

X	–3	–2	–1	0	1	2	3
Y	–10	0	4	5	4	2	–2
f(*X*)	–9,07	–1,70	3,15	5,49	5,31	2,62	–2,59
w	1/2	2/3	1	1	1	2/3	1/2

Подсчитаем по (2.4.8) величину

$$\varepsilon = \sum_{i=1}^{7} \frac{[y_i - f(x_i)]^2}{w_i} = 10{,}022,$$

что позволяет, вычисляя по формуле (2.4.10) остаточную дисперсию, оценить дисперсию оценки регрессии

$$s^2 = \frac{10{,}022}{7-3} = 2{,}5055.$$

Наконец, находя матрицу, обратную по отношению к матрице коэффициентов при $\overline{\beta}_j$ в системе (2.4.9′′′),

$$\begin{pmatrix} 10 & 0 & 50 \\ 0 & 50 & 0 \\ 50 & 0 & 374 \end{pmatrix}^{-1} = \begin{pmatrix} 0{,}302 & 0 & -0{,}040 \\ 0 & 0{,}02 & 0 \\ -0{,}040 & 0 & 0{,}008 \end{pmatrix}$$

и умножая ее на s^2, получим корреляционную матрицу оценок $\overline{\beta}_j$:

$$\begin{pmatrix} 0{,}756 & 0 & -0{,}101 \\ 0 & 0{,}05 & 0 \\ -0{,}101 & 0 & 0{,}02 \end{pmatrix}.$$

2.4.3. Линейная множественная регрессия $f(x_1, x_2) = \beta_1 + \beta_2 x_1 + \beta_3 x_2$

Предположим, что зависимость между переменными имеет вид:

$$y = \beta_1 + \beta_2 x_1 + \beta_3 x_2,$$

где x_1 и x_2 принимают заданные фиксированные значения, причем между x_1 и x_2 нет линейной зависимости. Результаты наблюдений без повторов $\{x_{1i}, x_{2i}, y_i\}$, $i = 1, 2, \ldots, n$ представляются в виде

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \varepsilon_i.$$

Предположим, что ошибки наблюдений независимы и имеют равные дисперсии, т. е. полагаем $w_i = 1$.

Оценки $\overline{\beta}_j$ параметров β_j данной регрессионной модели могут быть также найдены из условия (2.4.8), что приводит к частному случаю системы (2.4.9) при $f_1 = 1$, $f_2 = x_1$, $f_3 = x_2$, $\tilde{p}_i = 1/n$:

$$\begin{aligned} &\overline{\beta}_1 n + \overline{\beta}_2 \sum_{i=1}^{n} x_{1i} + \overline{\beta}_3 \sum_{i=1}^{n} x_{2i} = \sum_{i=1}^{n} y_i\,; \\ &\overline{\beta}_1 \sum_{i=1}^{n} x_{1i} + \overline{\beta}_2 \sum_{i=1}^{n} x_{1i}^2 + \overline{\beta}_3 \sum_{i=1}^{n} x_{1i} x_{2i} = \sum_{i=1}^{n} y_i x_{1i}\,; \\ &\overline{\beta}_1 \sum_{i=1}^{n} x_{2i} + \overline{\beta}_2 \sum_{i=1}^{n} x_{1i} x_{2i} + \overline{\beta}_3 \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} y_i x_{2i}\,. \end{aligned} \qquad (2.4.12)$$

*ПРИМЕР 4. Температура объекта Y зависит от процентного содержания X_1 компоненты A в теплоносителе и температуры окружающей среды X_2. Ниже приведены результаты 11 замеров этих данных (табл. 9).

Таблица 9

Y	6	8	1	0	5	3	2	–4	10	–3	5
X_1	1	4	9	11	3	8	5	10	2	7	6
X_2	8	2	–8	–10	6	–6	0	–12	4	–2	–4

Используя эту выборку, найти оценки $\overline{\beta}_j$ параметров β_j данной регрессионной модели $y = \beta_1 + \beta_2 x_1 + \beta_3 x_2$ и вычислить оценку дисперсии случайных ошибок наблюдений.

Решение. Предварительно вычислим суммы при $n = 11$ в системе (2.4.12):

$$\sum_{i=1}^{n} x_{1i} = 66; \quad \sum_{i=1}^{n} x_{2i} = -22; \quad \sum_{i=1}^{n} y_i = 33; \quad \sum_{i=1}^{n} x_{1i}^2 = 506;$$

$$\sum_{i=1}^{n} x_{2i}^2 = 484; \quad \sum_{i=1}^{n} x_{1i}x_{2i} = -346; \quad \sum_{i=1}^{n} y_i x_{1i} = 85; \quad \sum_{i=1}^{n} y_i x_{2i} = 142.$$

Подставляя вычисленные суммы коэффициентами в систему (2.4.12) и разрешая последнюю относительно $\overline{\beta}_j$, получим

$$f(x_1, x_2) = 14 - 2\,x_1 - 0{,}5\,x_2.$$

Построим график полученной зависимости (плоскость изображена с ребра) на фоне графически отображенной выборки (табл. 9).

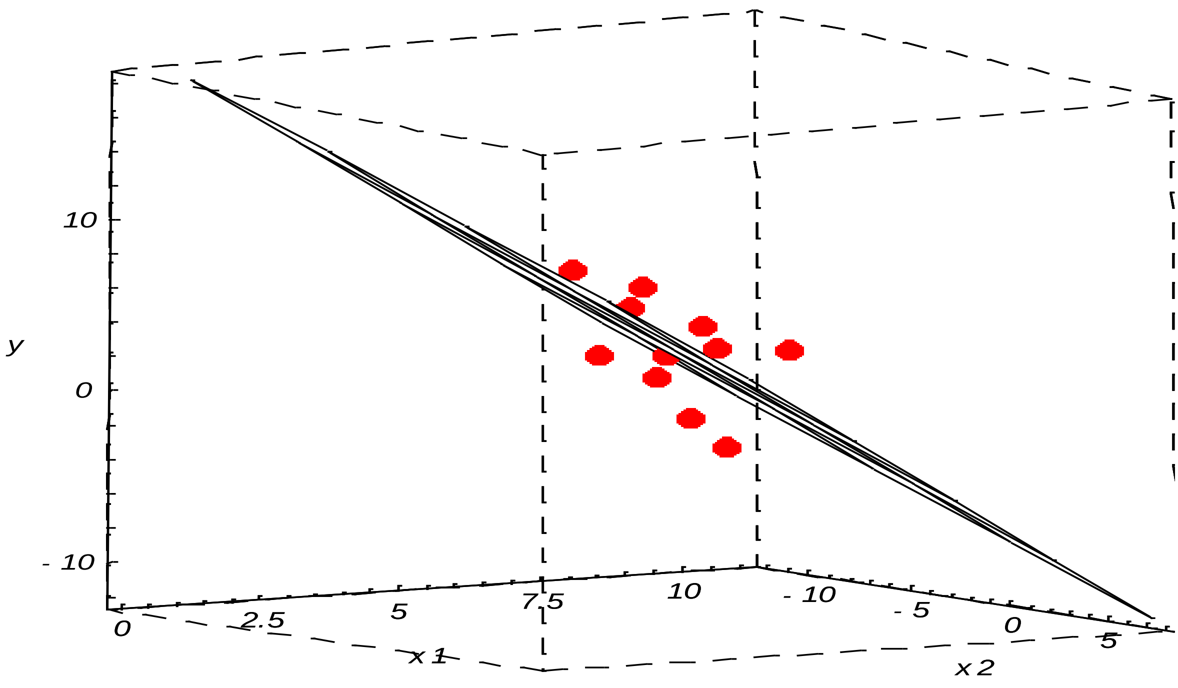


Рис. 2.4.9

Добавим в табл. 9 вычисленные значения $f(x_{1i}, x_{2i})$:

Y	6	8	1	0	5	3	2	–4	10	–3	5
f	8	5	0	–3	5	1	4	0	8	1	4
X_1	1	4	9	11	3	8	5	10	2	7	6
X_2	8	2	–8	–10	6	–6	0	–12	4	–2	–4

Подсчитаем по (2.4.8) величину $\varepsilon = \sum_{i=1}^{11} \frac{[y_i - f(x_{1i}, x_{2i})]^2}{w_i} = 68$, что позволяет, вычисляя по формуле (2.4.10) остаточную дисперсию, оценить дисперсию оценки регрессии $s^2 = \frac{68}{11-3} = 8{,}5$.

2.5. Проверка статистических гипотез

Во многих случаях результаты наблюдений используются для проверки предположений (гипотез) относительно тех или иных свойств распределения случайной величины. В частности, такого рода задачи возникают при сравнении различных технологических процессов или методов обработки по определенным измеряемым признакам, например, по точности, производительности и т. д.

К основным задачам математической статистики относится статистическая проверка гипотез о параметрах распределения и о законах распределения случайной величины. При исследовании различных случайных величин на определённом его этапе появляется возможность выдвинуть ту или иную гипотезу о свойствах изучаемой величины, например, сделать предположение о законе распределения её, или, если закон распределения известен, но не известны его параметры, то сделать предположение об их величине. Наиболее правдоподобную по каким-то соображениям гипотезу называют нулевой (основной) и обозначают H_0. Наряду с основной гипотезой рассматривают другую (альтернативную) гипотезу H_1, противоречащую основной. Выдвинутая нулевая гипотеза нуждается в дальнейшей проверке. При этом могут быть допущены ошибки двух типов:

- ошибка первого рода – отвергнута правильная гипотеза;
- ошибка второго рода – принята неправильная гипотеза.

Вероятность совершить ошибку первого рода (вероятность отвергнуть правильную гипотезу) обычно обозначают α и называют ***уровнем значимости***. Случайную величину Z, служащую для проверки гипотезы, называют ***критерием***. Совокупность значений критерия, при которых нулевую гипотезу отвергают, называют ***критической областью***. Граничные точки критической области z_{kp} называют ***критическими точками***. Различают три вида критической области:

- правосторонняя, определяемая неравенством $Z > z_{kp} > 0$;
- левосторонняя, определяемая неравенством $Z < z_{kp} < 0$;
- двусторонняя, определяемая неравенством $Z < z_1 < z_2 < Z$.

В частности, если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется неравенством $/Z/ > z_{kp} > 0$.

При отыскании критической области задаются уровнем значимости α и ищут критические точки, исходя из требования, чтобы вероятность того, что критерий Z примет значения, лежащие в критической области, была равна принятому уровню значимости. В результате получаем:

- для правосторонней критической области $P(Z > z_{kp}) = \alpha$;
- для левосторонней критической области $P(Z < z_{kp}) = \alpha$;
- для двусторонней симметричной области $P(Z > z_{kp}) = \alpha/2$.

Основной принцип статистической проверки гипотез заключается в следующем: если наблюдаемое значение критерия $Z_{\text{набл}}$, вычисленное по данным выборки, принадлежит критической области, то гипотезу отвергают; если наблюдаемое значение не принадлежит критической области, то нет оснований отвергать гипотезу.

Для многих критериев Z составлены таблицы, позволяющие по α найти критические точки z_{kp} (см. п. 3 прил. 4).

2.5.1. Сравнение двух дисперсий

Рассмотрим гипотезы о параметрах нормального распределения. Пусть имеются две серии опытов, регистрирующие значения некоторой случайной величины и определяющие две выборки объемов n_X и n_Y.

Рассмотрим тестирование гипотезы H_0 о равенстве дисперсий $D_X = D_Y$ при неизвестных математических ожиданиях. Пусть даны две случайные величины X и Y, распределенные по нормальному закону. По данным выборок объёмом n_X и n_Y соответственно подсчитаны исправленные выборочные дисперсии s_x^2 и s_y^2. Требуется при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что $D_X = D_Y$. Такая задача возникает при сравнении точности двух приборов, при сравнении различных методов измерений. Обычно выборочные дисперсии оказываются различными. Возникает вопрос: существенно или нет они различаются? Если различие незначимо, то имеет место нулевая гипотеза, следовательно, приборы имеют одинаковую точность, а различие эмпирических дисперсий объясняется случайными причинами, в частности, случайным отбором объектов выборки.

По данным выборок объёмом n_X и n_Y вычисляют $F_{\text{набл}}$, как отношение большей дисперсии к меньшей:

$$F_{\text{набл}} = \frac{s_{\text{Б}}^2}{s_{\text{М}}^2}.$$

Критическая область строится в зависимости от конкурирующей гипотезы следующим образом: по таблице распределения Фишера по заданному уровню значимости α и вычисленным степеням свободы k_1 и k_2 (см. прил. 6) находят $F_{kp}(\alpha, k_1, k_2)$ для H_1: $D_X > D_Y$ или $F_{kp}(\alpha/2, k_1, k_2)$ для H_1: $D_X \neq D_Y$. Если $F_{\text{набл}} > F_{kp}$, то H_0 отвергают, а при $F_{\text{набл}} < F_{kp}$ нет оснований отвергать H_0.

Величина F удовлетворяет распределению Фишера со степенями свободы k_1, определенной разностью объема выборки с большей дисперсией и единицы, и k_2, определенной разностью объема выборки с меньшей дисперсией и единицы. Поэтому тестирование основано на распределении Фишера. На рис. 2.5.1 приведены сравнительные графики плотностей распределения:

– нормального с $\mu = 0$, $\sigma = 1$ (пунктирная линия);

– Фишера с $k_1 = 1$ и $k_2 = 1$ (тонкая сплошная линия);

– Фишера с $k_1 = 10$ и $k_2 = 10$ (толстая сплошная линия.)

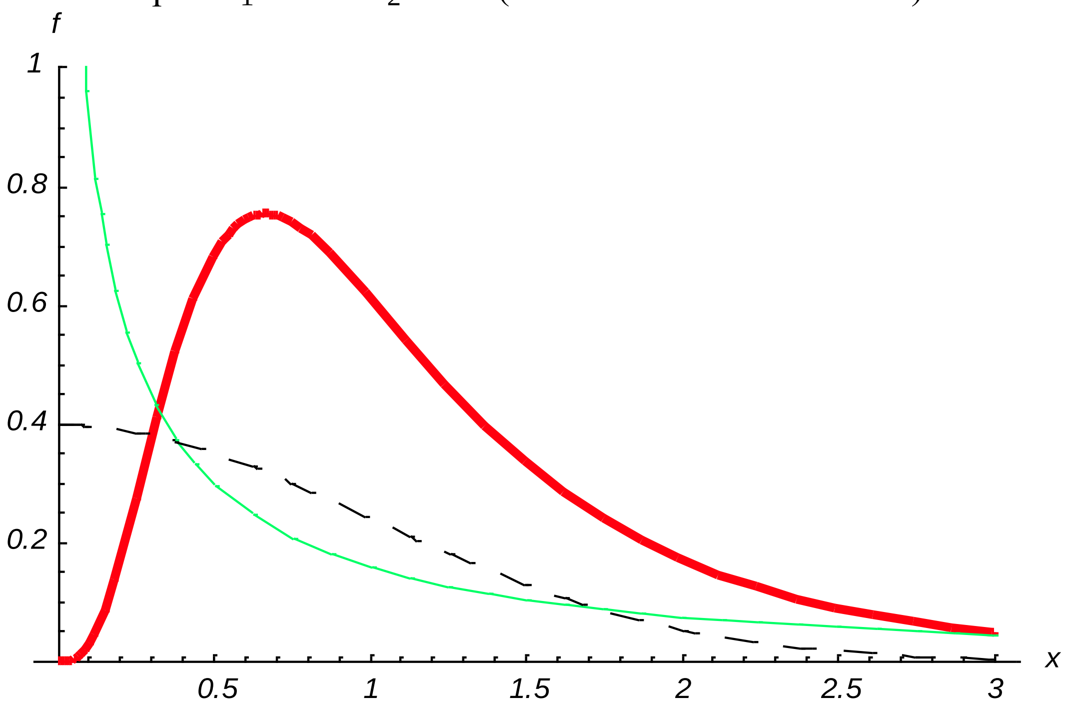


Рис. 2.5.1

*ПРИМЕР 1. По двум малым независимым выборкам объемов $n_X = 11$ и $n_Y = 14$ из нормальных распределений найдены исправленные выборочные дисперсии $s_x^2 = 0{,}76$ и $s_y^2 = 0{,}38$. При уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу H_0: $D_X = D_Y$ о равенстве дисперсий при конкурирующей гипотезе H_1: $D_X > D_Y$.

Решение. Найдем отношение большей исправленной дисперсии к меньшей:

$$F_{\text{набл}} = \frac{s_{\text{Б}}^2}{s_{\text{М}}^2} = 0{,}76/0{,}38 = 2.$$

По условию конкурирующая гипотеза имеет вид H_1: $D_X > D_Y$, поэтому критическая область – правосторонняя. По таблице критических точек распределения Фишера, по уровню значимости $\alpha = 0{,}05$ и числам степеней свободы $k_1 = n_X - 1 = 11 - 1 = 10$ и $k_2 = n_Y - 1 = 14 - 1 = 13$ находим (см. прил. 6) критическую точку:

$$F_{kp}(\alpha, k_1, k_2) = F_{kp}(0{,}05;\ 10;\ 13) = 2{,}67.$$

Так как $F_{\text{набл}} = 2 < F_{kp} = 2{,}67$, то нет оснований отвергать H_0 о равенстве дисперсий. Другими словами, исправленные выборочные дисперсии различаются незначимо.

2.5.2. Сравнение математических ожиданий

Для проверки подобия выборок (соответствия их распределению одной и той же случайной величины) рассмотрим вопрос о значимости расхождения между выборочными значениями математических ожиданий $\bar{x}$ и $\bar{y}$: выдвинем в качестве H_0 равенство математических ожиданий $m_X = m_Y$. Тестирование такой гипотезы основано на нормальном распределении в случае большого объема выборок ($n > 30$), когда дисперсии считаются известными, и на распределении Стьюдента в случае малого объема выборок ($n < 30$), когда дисперсии считаются неизвестными. Сравнительные графики плотностей распределения нормального и Стьюдента приведены на рис. 2.3.1.

Рассмотрим первый случай. Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу H_0: $m_X = m_Y$ о равенстве математических ожиданий двух больших нормальных выборок **с известными дисперсиями D_X и D_Y,** надо вычислить наблюдаемое значение критерия:

$$Z_{\text{набл}} = \frac{\bar{x} - \bar{y}}{\sqrt{D_X / n_X + D_Y / n_Y}}.$$

Далее построить критическую область в зависимости от конкурирующей гипотезы следующим образом:

1. При конкурирующей гипотезе H_1: $m_X \neq m_Y$ по таблице функции Лапласа (см. прил. 3) найти критическую точку z_{kp} из равенства

$$\Phi(z_{kp}) = (1 - \alpha)/2.$$

Если $|Z_{\text{набл}}| < z_{kp}$, то нет оснований отвергать нулевую гипотезу.

Если $|Z_{\text{набл}}| > z_{kp}$, то нулевую гипотезу отвергают.

2. При конкурирующей гипотезе H_1: $m_X > m_Y$ по таблице функции Лапласа (см. прил. 3) найти критическую точку z_{kp} из равенства

$$\Phi(z_{kp}) = (1 - 2\alpha)/2.$$

Если $Z_{\text{набл}} < z_{kp}$, то нет оснований отвергать нулевую гипотезу.

Если $Z_{\text{набл}} > z_{kp}$, то нулевую гипотезу отвергают.

3. При конкурирующей гипотезе H_1: $m_X < m_Y$ по таблице функции Лапласа найти «вспомогательную критическую точку» z_{kp} из равенства

$$\Phi(z_{kp}) = (1 - 2\alpha)/2.$$

Если $Z_{\text{набл}} > -z_{kp}$, то нет оснований отвергать нулевую гипотезу.

Если $Z_{\text{набл}} < -z_{kp}$, то нулевую гипотезу отвергают.

*ПРИМЕР 2. По двум большим независимым выборкам объемов $n_X = 40$ и $n_Y = 50$ нормальных распределений найдены выборочные значениями математических ожиданий $\bar{x} = 130$ и $\bar{y} = 140$. Дисперсии известны $D_X = 80$ и $D_Y = 100$. При уровне значимости $\alpha = 0{,}01$ проверить нулевую гипотезу H_0: $m_X = m_Y$ о равенстве математических ожиданий при конкурирующей гипотезе H_1: $m_X \neq m_Y$.

Решение. Найдем наблюдаемое значение критерия:

$$Z_{\text{набл}} = \frac{\bar{x} - \bar{y}}{\sqrt{D_X / n_X + D_Y / n_Y}} = \frac{130 - 140}{\sqrt{80/40 + 100/50}} = -5.$$

По условию конкурирующая гипотеза имеет вид H_1: $m_X \neq m_Y$, поэтому критическая область – двусторонняя. Найдем критическую точку z_{kp} из равенства

$$\Phi(z_{kp}) = (1 - \alpha)/2 = (1 - 0{,}01)/2 = 0{,}495.$$

По таблице функции Лапласа (см. прил. 3) находим $z_{kp} = 2{,}58$.

Так как $|Z_{\text{набл}}| = |-5| = 5 > z_{kp} = 2{,}58$, то нулевую гипотезу отвергают. Другими словами, выборочными значениями математических ожиданий различаются значимо.

Рассмотрим второй случай. Пусть имеются две выборки объемов n_X и n_Y, на основании которых подсчитаны выборочными значениями математических ожиданий $\bar{x}$ и $\bar{y}$ и исправленные выборочные дисперсии s_x^2 и s_y^2. Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу H_0: $m_X = m_Y$ о равенстве математических ожиданий двух малых нормальных выборок **с неизвестными дисперсиями D_X и D_Y,** надо предварительно проверить гипотезу о равенстве дисперсий (см. п. 2.5.1) по подсчитанным исправленным выборочным дисперсиям s_x^2 и s_y^2. Если не будет оснований отвергать гипотезу о равенстве дисперсий, т. е. дисперсии хотя и неизвестны, но предполагаются одинаковыми, то далее надо вычислить наблюдаемое значение критерия:

$$T_{\text{набл}} = \frac{\bar{x} - \bar{y}}{\sqrt{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}} \sqrt{\frac{n_X n_Y (n_X + n_Y - 2)}{n_X + n_Y}}.$$

Затем построить критическую область в зависимости от конкурирующей гипотезы следующим образом:

1. При конкурирующей гипотезе H_1: $m_X \neq m_Y$ по таблице критических точек распределения Стьюдента (см. прил. 4), по заданному уровню значимости α, помещенному в верхней строке таблицы, и числу степеней свободы $k = n_X + n_Y - 2$ найти критическую точку t_{kp} двусторонней критической области.

Если $|T_{\text{набл}}| < t_{kp}$, то нет оснований отвергать нулевую гипотезу.

Если $|T_{\text{набл}}| > t_{kp}$, то нулевую гипотезу отвергают.

2. При конкурирующей гипотезе H_1: $m_X > m_Y$ по таблице критических точек распределения Стьюдента (см. прил. 4), по заданному уровню значимости α, помещенному в нижней строке таблицы, и числу степеней свободы $k = n_X + n_Y - 2$ найти критическую точку t_{kp} односторонней критической области.

Если $T_{\text{набл}} < t_{kp}$, то нет оснований отвергать нулевую гипотезу.

Если $T_{\text{набл}} > t_{kp}$, то нулевую гипотезу отвергают.

3. При конкурирующей гипотезе H_1: $m_X < m_Y$ по таблице критических точек распределения Стьюдента (см. прил. 4), по заданному уровню значимости α, помещенному в нижней строке таблицы, и числу степеней свободы $k = n_X + n_Y - 2$ найти «вспомогательную критическую точку» t_{kp} односторонней критической области.

Если $T_{набл} > -t_{kp}$, то нет оснований отвергать нулевую гипотезу.

Если $T_{набл} < -t_{kp}$, то нулевую гипотезу отвергают.

*ПРИМЕР 3. По двум малым независимым выборкам объемов $n_X = 12$ и $n_Y = 18$ нормальных распределений найдены выборочные значениями математических ожиданий $\bar{x} = 31{,}2$ и $\bar{y} = 29{,}2$ и исправленные выборочные дисперсии $s_x^2 = 0{,}84$ и $s_y^2 = 0{,}40$. При уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу H_0: $m_X = m_Y$ при конкурирующей гипотезе H_1: $m_X \neq m_Y$.

Решение. Исправленные выборочные дисперсии различны, поэтому проверим предварительно гипотезу о равенстве дисперсий, используя критерий Фишера (см. пример 1).

Найдем отношение большей исправленной дисперсии к меньшей:

$$F_{набл} = \frac{s_{Б}^2}{s_{М}^2} = 0{,}84/0{,}4 = 2{,}1.$$

Дисперсия s_x^2 значительно больше дисперсии s_y^2, поэтому в качестве конкурирующей гипотезы примем гипотезу H_1: $D_X > D_Y$. В этом случае критическая область – правосторонняя. По таблице критических точек распределения Фишера, по уровню значимости $\alpha = 0{,}05$ и числам степеней свободы $k_1 = n_X - 1 = 12 - 1 = 11$ и $k_2 = n_Y - 1 = 18 - 1 = 17$ находим (см. прил. 6) критическую точку:

$$F_{kp}(\alpha, k_1, k_2) = F_{kp}(0{,}05;\ 11;\ 17) = 2{,}41.$$

Так как $F_{набл} = 2{,}1 < F_{kp} = 2{,}41$, то нет оснований отвергать H_0 о равенстве дисперсий. Предположение о равенстве дисперсий не отвергается, поэтому далее проверим гипотезу о равенстве математических ожиданий.

Вычислим наблюдаемое значение критерия Стьюдента:

$$T_{набл} = \frac{\bar{x} - \bar{y}}{\sqrt{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}} \sqrt{\frac{n_X n_Y (n_X + n_Y - 2)}{n_X + n_Y}}.$$

Подставляя числовые значения входящих в эту формулу величин, получим $T_{набл} = 7{,}1$.

По условию конкурирующая гипотеза имеет вид H_1: $m_X \neq m_Y$, поэтому критическая область – двусторонняя. По таблице критических точек распределения Стьюдента, по уровню значимости $\alpha = 0{,}05$, помещенному в верхней строке таблицы, и числу степеней свободы $k = n_X + n_Y - 2 = 12 + 18 - 2 = 28$ находим (см. прил. 4) критическую точку:

$$t_{kp}(\alpha, k) = t_{kp}(0{,}05; 28) = 2{,}05.$$

Так как $|T_{набл}| = 7{,}1 > t_{kp} = 2{,}05$, то нулевую гипотезу о равенстве математических ожиданий отвергаем. Другими словами, математические ожидания различаются значимо.

Вернемся ко второму случаю и рассмотрим далее второй вариант, когда гипотеза о равенстве дисперсий отвергается. Пусть имеются две выборки объемов n_X и n_Y, на основании которых подсчитаны выборочными значениями математических ожиданий $\bar{x}$ и $\bar{y}$ и исправленные выборочные дисперсии s_x^2 и s_y^2. Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу H_0: $m_X = m_Y$ о равенстве математических ожиданий двух малых нормальных выборок **с неизвестными дисперсиями D_X** и **D_Y**, надо предварительно проверить гипотезу о равенстве дисперсий (см. п. 2.5.1) по подсчитанным исправленным выборочным дисперсиям s_x^2 и s_y^2. Пусть гипотеза о равенстве дисперсий отвергается, т. е. дисперсии хотя и не известны, но предполагаются разными. Тестирование такой гипотезы H_0: $m_X = m_Y$ основано на распределении Стьюдента с числом степеней свободы k:

$$k = \frac{\left(s_x^2/n_x + s_y^2/n_y\right)^2}{\dfrac{\left(s_x^2/n_x\right)^2}{n_x - 1} + \dfrac{\left(s_y^2/n_y\right)^2}{n_y - 1}}$$

Далее вычисляют наблюдаемое значение критерия по формуле

$$T_{набл} = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_X + s_y^2/n_Y}}.$$

Затем строят критическую область в зависимости от конкурирующей гипотезы следующим образом:

1. При конкурирующей гипотезе H_1: $m_X \neq m_Y$ по таблице критических точек распределения Стьюдента (см. прил. 4), по заданному уровню значимости α, помещенному в верхней строке таблицы, и числу степеней свободы k найти критическую точку t_{kp} двусторонней критической области.

Если $|T_{\text{набл}}| < t_{kp}$, то нет оснований отвергать нулевую гипотезу.

Если $|T_{\text{набл}}| > t_{kp}$, то нулевую гипотезу отвергают.

2. При конкурирующей гипотезе H_1: $m_X > m_Y$ по таблице критических точек распределения Стьюдента (см. прил. 4), по заданному уровню значимости α, помещенному в нижней строке таблицы, и числу степеней свободы k найти критическую точку t_{kp} односторонней критической области.

Если $T_{\text{набл}} < t_{kp}$, то нет оснований отвергать нулевую гипотезу.

Если $T_{\text{набл}} > t_{kp}$, то нулевую гипотезу отвергают.

3. При конкурирующей гипотезе H_1: $m_X < m_Y$ по таблице критических точек распределения Стьюдента (см. прил. 4), по заданному уровню значимости α, помещенному в нижней строке таблицы, и числу степеней свободы k найти « вспомогательную критическую точку» t_{kp} односторонней критической области.

Если $T_{\text{набл}} > -t_{kp}$, то нет оснований отвергать нулевую гипотезу.

Если $T_{\text{набл}} < -t_{kp}$, то нулевую гипотезу отвергают.

*ПРИМЕР 4. По двум малым независимым выборкам объемов $n_X = 5$ и $n_Y = 5$ нормальных распределений найдены выборочные значениями математических ожиданий $\bar{x} = 13{,}32$ и $\bar{y} = 13{,}80$ и исправленные выборочные дисперсии $s_x^2 = 3{,}37$ и $s_y^2 = 0{,}46$. При уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу H_0: $m_X = m_Y$ при конкурирующей гипотезе H_1: $m_X \neq m_Y$.

Решение. Исправленные выборочные дисперсии различны, поэтому проверим предварительно гипотезу о равенстве дисперсий, используя критерий Фишера (см. пример 1).

Найдем отношение большей исправленной дисперсии к меньшей

$$F_{\text{набл}} = \frac{s_{\text{Б}}^2}{s_{\text{М}}^2} = 3{,}37/0{,}46 \approx 7{,}33.$$

Дисперсия s_x^2 значительно больше дисперсии s_y^2, поэтому в качестве конкурирующей гипотезы примем гипотезу H_1: $D_X > D_Y$. В этом случае критическая область – правосторонняя. По таблице критических точек распределения Фишера, по уровню значимости $\alpha = 0{,}05$ и числам степеней свободы $k_1 = n_X - 1 = 5 - 1 = 4$ и $k_2 = n_Y - 1 = 5 - 1 = 4$ находим (см. прил. 6) критическую точку:

$$F_{kp}(\alpha, k_1, k_2) = F_{kp}(0{,}05; 4; 4) = 6{,}39.$$

Так как $F_{набл} = 7{,}33 > F_{kp} = 6{,}39$, то гипотеза о равенстве дисперсий отклоняется.

Вычислим наблюдаемое значение критерия Стьюдента:

$$T_{набл} = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_X + s_y^2/n_Y}} \approx -0{,}55.$$

Число степеней свободы

$$k = \frac{(s_x^2/n_x + s_y^2/n_y)^2}{\dfrac{(s_x^2/n_x)^2}{n_x - 1} + \dfrac{(s_y^2/n_y)^2}{n_y - 1}} = \frac{(3{,}37/5 + 0{,}46/5)^2}{\dfrac{(3{,}37/5)^2}{5-1} + \dfrac{(0{,}46/5)^2}{5-1}} \approx 5.$$

По условию конкурирующая гипотеза имеет вид H_1: $m_X \neq m_Y$, поэтому критическая область – двусторонняя. По таблице критических точек распределения Стьюдента, по уровню значимости $\alpha = 0{,}05$, помещенному в верхней строке таблицы, и числу степеней свободы $k = n_X + n_Y - 2 = 5 + 5 - 2 = 8$ находим (см. прил. 4) критическую точку:

$$t_{kp}(\alpha, k) = t_{kp}(0{,}05; 8) = 2{,}31.$$

Так как $|T_{набл}| = 0{,}55 < t_{kp} = 2{,}31$, то нет оснований отвергать нулевую гипотезу о равенстве математических ожиданий. Другими словами, математические ожидания различаются незначимо.

2.5.3. Проверка гипотезы о законе распределения

Пусть дана выборка наблюдений случайной величины X: $\{x_1, x_2, \ldots, x_n\}$. Проверяется гипотеза H_0, утверждающая, что X имеет функцию распределения $F(x)$ или плотность распределения $f(x)$. По выборке наблюдений находят оценки неизвестных параметров (если таковые есть) предполагаемого закона распределения случайной величины X. Далее интервал Ω возможных значений случайной величины X разбивается на r непересекающихся подынтервалов $\Omega_k = (a_k, b_k)$, $k = 1, 2, \ldots, r$. Пусть n_k – число элементов выборки, принадлежащих подынтервалу Ω_k. Очевидно, что $\sum_{k=1}^{r} n_k = n$. Используя предполагаемый закон распределения случайной величины X, находят вероятности p_k того, что значение X принадлежит подынтервалу Ω_k:

$$p_k = P(X \in \Omega_k) = \int_{\Omega_k} f(x)dx = F(b_k) - F(a_k), \ \sum_{k=1}^{r} p_k = 1.$$

Далее вычисляют статистическое значение критерия по формуле

$$Z = \sum_{k=1}^{r} \frac{(n_k - np_k)^2}{np_k}.$$

По теореме Пирсона величина Z должна быть распределена по закону χ^2 с $r - l - 1$ степенями свободы, где l – число неизвестных параметров распределения, оцениваемых по выборке (см. рис. 2.3.6).

При заданном уровне значимости α гипотеза о распределении X по закону $F(x)$ отвергается, если $Z > z_{kp}$, и не отвергается, если $Z < z_{kp}$, где z_{kp} определяется по таблице критических точек распределения χ^2 с $r - l - 1$ степенями свободы так, чтобы $P(\chi^2 > z_{kp}) = \alpha$ (см. прил. 5).

*ПРИМЕР 5. По критерию Пирсона при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о распределении случайной величины X по нормальному закону, если задано n_k попаданий выборочных значений случайной величины X в подынтервал $\Omega_k = (a_k, b_k)$:

Ω_k	12÷14	14÷16	16÷18	18÷20	20÷22
n_k	6	8	12	16	13

Решение. Определим объем выборки $\sum_{k=1}^{r} n_k = n$ =55. Найдем по выборке оценки математического ожидания и дисперсии (таким образом $l = 2$) предполагаемого нормального распределения случайной величины X, принимая за x_k середины подынтервалов:

x_k	13	15	17	19	21
n_k	6	8	12	16	13

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{5} x_k n_k = 17{,}8;$$

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{5}(x_k - \bar{x})^2 n_k \approx 6{,}75556; \quad s \approx 2{,}6.$$

Подсчитаем вероятности p_k для предполагаемого нормального распределения случайной величины X по формуле

$$p_k = P(X \in \Omega_k) = \Phi(\frac{b_k - \bar{x}}{s}) - \Phi(\frac{a_k - \bar{x}}{s}),\ k = 1, 2, 3, 4, 5,$$

где a_k и b_k – соответственно нижняя и верхняя границы подинтервалов Ω_k, причем крайние границы расширены до бесконечности ($a_1 = -\infty$, $b_5 = \infty$), а значения функции Лапласа $\Phi(x)$ вычисляются по таблице значений функции Лапласа (см. прил. 3). Добавим также строку плотностей частоты f_k, деля p_k на длину подынтервала: $f_k = p_k/2$. Таким образом, расширенную таблицу выборочного распределения можно представить в виде:

Ω_k	$-\infty \div 14$	$14 \div 16$	$16 \div 18$	$18 \div 20$	$20 \div \infty$
n_k	6	8	12	16	13
p_k	0,07	0,17	0,29	0,27	0,20
f_k	0,035	0,085	0,145	0,135	0,10

Далее вычисляем статистическое значение критерия по формуле

$$Z = \sum_{k=1}^{5}\frac{(n_k - np_k)^2}{np_k} \approx 2{,}83.$$

Затем, при уровне значимости $\alpha = 0{,}05$, определяем z_{kp} по таблице критических точек распределения χ^2 с $r - l - 1 = 5 - 2 - 1 = 2$ степенями свободы (см. прил. 5): $z_{kp}(\alpha, r - l - 1) = z_{kp}(0{,}05{,}2) = 6.0$.

Наконец, так как $Z = 2{,}83 < z_{kp} = 6{,}0$, то делаем вывод о том, что гипотеза о распределении случайной величины X по нормальному закону не отвергается.

Изобразим на фоне точек массива $\{x_k, f_k\}$ график кривой плотности нормального распределения $f(x)$, беря в качестве математического ожидания и дисперсии их подсчитанные оценки: $\bar{x} = 17{,}8$; $s \approx 2{,}6$ (рис. 2.5.2).

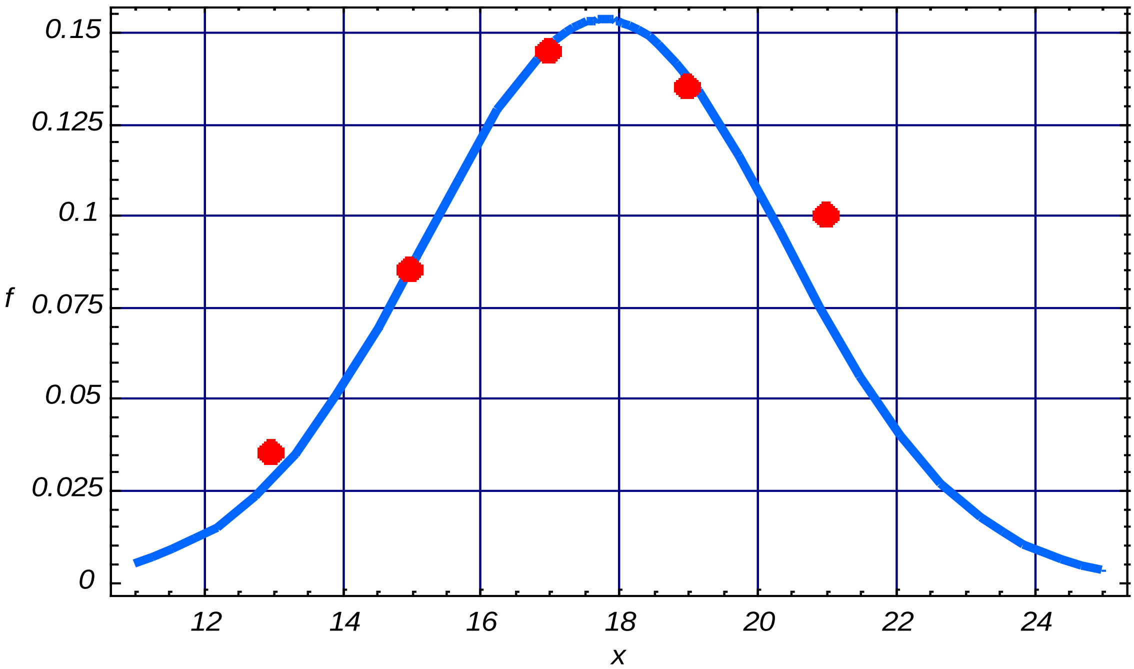


Рис. 2.5.2

Часть 3

ПРИЛОЖЕНИЯ

3.1. Приложение 1

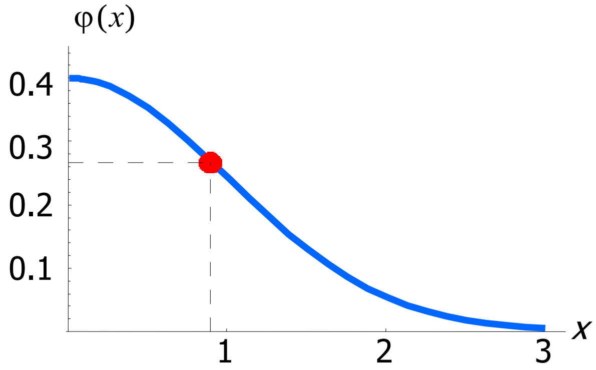
Таблица важнейших дискретных и непрерывных распределений случайной величины *X*

Распределение		Параметры	$M(X)$	$D(X)$
Биномиальное	$P\{X=m\}=C_n^m p^m(1-p)^{n-m}$ $m = 0, 1, ..., n$	$n=1, 2, ...$ $0\le p \le 1$	np	npq
Пуассона	$P\{X = m\} = \dfrac{a^m e^{-a}}{m!}$, $m = 0, 1, 2, ...$	$a > 0$	a	a
Геометрическое	$P\{X=m\}=p(1-p)^m$ $m = 0, 1, ...$	$0\le p\le 1$	$\dfrac{1-p}{p}$	$\dfrac{1-p}{p^2}$
Гипергеометрическое	$P\{X = m\} = \dfrac{C_M^m C_{N-M}^{n-m}}{C_N^n}$, $m = 0, 1, 2, \ldots, \min(M,n)$	$N = 2, 3, \ldots$ $M=1, 2, \ldots<N$ $n=1, 2, \ldots<N$	$\dfrac{nM}{N}$	$\dfrac{nM(N-M)(N-n)}{(N-1)N^2}$
Равномерное	$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in (a,b); \\ 0, & x \notin (a,b) \end{cases}$	(a, b) – любой интервал на оси Ox	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Показательное	$f(x) = \begin{cases} 0, & x < 0; \\ \lambda e^{-\lambda x}, & x \ge 0 \end{cases}$	$\lambda > 0$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
Нормальное (Гаусса)	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$	$-\infty < a < \infty$ $\sigma > 0$	a	σ^2

3.2. Приложение 2

Таблица значений функции

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

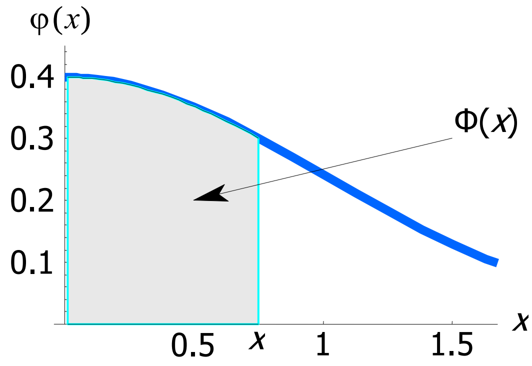


x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	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	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	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
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
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

3.3. Приложение 3

Таблица значений функции

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



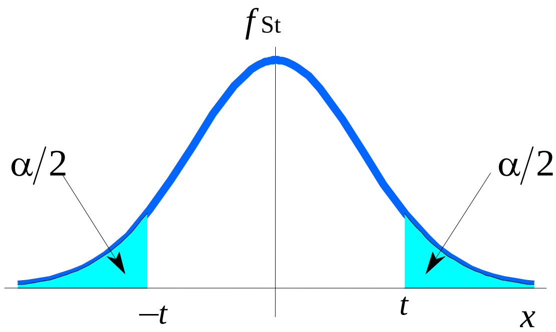
x	Φ(*x*)	*x*	Φ(*x*)	*x*	Φ(*x*)	*x*	Φ(*x*)
0,00	0,0000	0,30	0,1179	0,60	0,2257	0,90	0,3159
0,01	0,0040	0,31	0,1217	0,61	0,2291	0,91	0,3186
0,02	0,0080	0,32	0,1255	0,62	0,2324	0,92	0,3212
0,03	0,0120	0,33	0,1293	0,63	0,2357	0,93	0,3238
0,04	0,0160	0,34	0,1331	0,64	0,2389	0,94	0,3264
0,05	0,0199	0,35	0,1368	0,65	0,2422	0,95	0,3289
0,06	0,0239	0,36	0,1406	0,66	0,2454	0,96	0,3315
0,07	0,0279	0,37	0,1443	0,67	0,2486	0,97	0,3340
0,08	0,0319	0,38	0,1480	0,68	0,2517	0,98	0,3365
0,09	0,0359	0,39	0,1517	0,69	0,2549	0,99	0,3389
0,10	0,0398	0,40	0,1554	0,70	0,2580	1,00	0,3413
0,11	0,0438	0,41	0,1591	0,71	0,2611	1,01	0,3438
0,12	0,0478	0,42	0,1628	0,72	0,2642	1,02	0,3461
0,13	0,0517	0,43	0,1664	0,73	0,2673	1,03	0,3485
0,14	0,0557	0,44	0,1700	0,74	0,2703	1,04	0,3508
0,15	0,0596	0,45	0,1736	0,75	0,2734	1,05	0,3531
0,16	0,0636	0,46	0,1772	0,76	0,2764	1,06	0,3554
0,17	0,0675	0,47	0,1808	0,77	0,2794	1,07	0,3577
0,18	0,0714	0,48	0,1844	0,78	0,2823	1,08	0,3599
0,19	0,0753	0,49	0,1879	0,79	0,2852	1,09	0,3621
0,20	0,0793	0,50	0,1915	0,80	0,2881	1,10	0,3643
0,21	0,0832	0,51	0,1950	0,81	0,2910	1,11	0,3665
0,22	0,0871	0,52	0,1985	0,82	0,2939	1,12	0,3686
0,23	0,0910	0,53	0,2019	0,83	0,2967	1,13	0,3708
0,24	0,0948	0,54	0,2054	0,84	0,2995	1,14	0,3729
0,25	0,0987	0,55	0,2088	0,85	0,3023	1,15	0,3749
0,26	0,1026	0,56	0,2123	0,86	0,3051	1,16	0,3770
0,27	0,1064	0,57	0,2157	0,87	0,3078	1,17	0,3790
0,28	0,1103	0,58	0,2190	0,88	0,3106	1,18	0,3810
0,29	0,1141	0,59	0,2224	0,89	0,3133	1,19	0,3830

Окончание прил. 3

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,20	0,3849	1,60	0,4452	2,00	0,4772	2,60	0,4953
1,21	0,3869	1,61	0,4463	2,02	0,4783	2,62	0,4956
1,22	0,3883	1,62	0,4474	2,04	0,4793	2,64	0,4959
1,23	0,3907	1,63	0,4484	2,06	0,4803	2,66	0,4961
1,24	0,3925	1,64	0,4495	2,08	0,4812	2,68	0,4963
1,25	0,3944	1,65	0,4505	2,10	0,4821	2,70	0,4965
1,26	0,3962	1,66	0,4515	2,12	0,4830	2,72	0,4967
1,27	0,3980	1,67	0,4525	2,14	0,4838	2,74	0,4969
1,28	0,3997	1,68	0,4535	2,16	0,4846	2,76	0,4971
1,29	0,4015	1,69	0,4545	2,18	0,4854	2,78	0,4973
1,30	0,4032	1,70	0,4554	2,20	0,4861	2,80	0,4974
1,31	0,4049	1,71	0,4564	2,22	0,4868	2,82	0,4976
1,32	0,4066	1,72	0,4573	2,24	0,4875	2,84	0,4977
1,33	0,4082	1,73	0,4582	2,26	0,4881	2,86	0,4979
1,34	0,4099	1,74	0,4591	2,28	0,4887	2,88	0,4980
1,35	0,4115	1,75	0,4599	2,30	0,4893	2,90	0,4981
1,36	0,4131	1,76	0,4608	2,32	0,4898	2,92	0,4982
1,37	0,4147	1,77	0,4616	2,34	0,4904	2,94	0,4984
1,38	0,4162	1,78	0,4625	2,36	0,4909	2,96	0,4985
1,39	0,4177	1,79	0,4633	2,38	0,4913	2,98	0,4986
1,40	0,4192	1,80	0,4641	2,40	0,4918	3,00	0,49865
1,41	0,4207	1,81	0,4649	2,42	0,4922	3,20	0,49931
1,42	0,4222	1,82	0,4656	2,44	0,4927	3,40	0,49966
1,43	0,4236	1,83	0,4664	2,46	0,4931	3,60	0,499841
1,44	0,4251	1,84	0,4671	2,48	0,4934	3,80	0,499928
1,45	0,4265	1,85	0,4678	2,50	0,4938	4,00	0,499968
1,46	0,4279	1,86	0,4686	2,52	0,4941		
1,47	0,4292	1,87	0,4693	2,54	0,4945	4,50	0,499997
1,48	0,4306	1,88	0,4699	2,56	0,4948		
1,49	0,4319	1,89	0,4706	2,58	0,4951	5,00	0,499997
1,50	0,4332	1,90	0,4713				
1,51	0,4345	1,91	0,4719				
1,52	0,4357	1,92	0,4726				
1,53	0,4370	1,93	0,4732				
1,54	0,4382	1,94	0,4738				
1,55	0,4394	1,95	0,4744				
1,56	0,4406	1,96	0,4750				
1,57	0,4418	1,97	0,4756				
1,58	0,4429	1,98	0,4761				
1,59	0,4441	1,99	0,4767				

3.4. Приложение 4

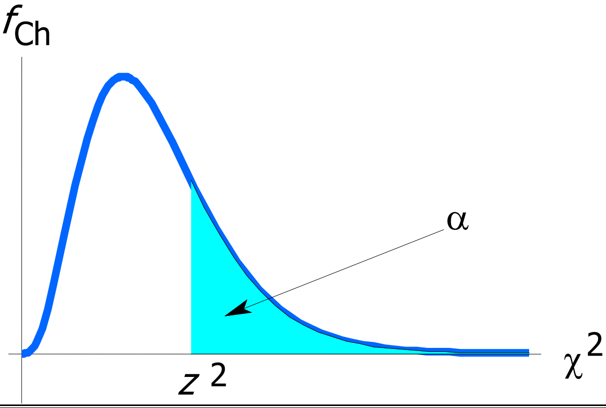
Критические точки
***t*-распределения Стьюдента**



Число степеней свободы k	Уровень значимости α (**двусторонняя** крит. область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,70	31,82	63,70	318,30	637,0
2	2,92	4,30	6,97	9,92	22,33	31 ,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
22	1,72	2,07	2,51	2,82	3,51	3,79
24	1,71	2,06	2,49	2,80	3,47	3,74
26	1,71	2,06	2,48	2,78	3,44	3,71
28	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (**односторонняя** крит. область)					

3.5. Приложение 5

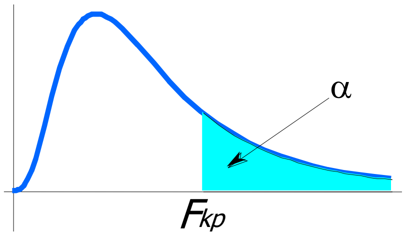
Критические точки z^2 распределения χ^2



Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21 ,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31 ,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

3.6. Приложение 6

Критические точки F_{kp} распределения Фишера



Уровень значимости α = 0,01												
k_2	k_1 – число степеней свободы большей дисперсии											
	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45
Уровень значимости α = 0,05												
k_2	k_1 – число степеней свободы большей дисперсии											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,502	2,45	2,41	2,38

ЛИТЕРАТУРА

1. Вентцель Е.С. Теория вероятностей. – М., 1964. – 576 с.
2. Вентцель Е.С., Овчаров Л.А. Теория вероятностей. – М., 1973. – 364 с.
3. Гмурман В.Е. Теория вероятностей и математическая статистика. – М., 1998. – 480 с.
4. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М., 1979. – 400 с.
5. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. – М., 1997. – 302 с.
6. Магазинников Л.И. Курс лекций по теории вероятностей. – Томск, 1989. – 211 с.
7. Ефимов А.В. Сборник задач по математике. Специальные курсы. Ч.3. – М., 1984. – 607 с.
8. Агапов Г.И. Задачник по теории вероятностей. – М., 1986. – 80 с.
9. Мешалкин Л.Д. Сборник задач по теории вероятностей. – М., 1963. – 155 с.
10. Лихолетов И.И., Мацкевич И.П. Руководство к решению задач по высшей математике, теории вероятностей и математической статистике. – Минск, 1976. – 452 с.
11. Румшинский Л.З. Элементы теории вероятностей. – М., 1976. – 239 с.
12. Данко П.Е., Попов А.Г., Кожевникова Т.Я. Высшая математика в упражнениях и задачах. Ч.2. – М., 1986. – 416 с.
13. Ежов И.И., Скороход А.В., Ядренко М.И. Элементы комбинаторики. – М., 1977. – 80 с.
14. Слободской М.И. Непосредственный подсчет вероятности. – Томск, 1976. – 23 с.
15. Гнеденко Б.В. Курс теории вероятностей. – М., 1965.,1969. – 400 с.
16. Пестова Н.Ф., Самойлова М.В. Практические занятия по теории вероятностей. – Томск, 1975. – 222 с.
17. Гурский Е. Н. Теория вероятностей с элементами математической статистики: учеб. пособие для втузов. – М., 1971. – 328 с.
18. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. – М., 1988. – 480 с.
19. Вентцель Е.С., Овчаров Л.А. Прикладные задачи теории вероятностей. – М., 1983. – 416 с.

ОГЛАВЛЕНИЕ

Учебное издание

ЛАЗАРЕВА Любовь Ивановна
МИХАЛЬЧУК Александр Александрович

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Редактор *Н.Т. Синельникова*

Верстка *Л.А. Егорова*

Отпечатано в Издательстве ТПУ в полном соответствии с качеством предоставленного оригинал-макета

Подписано к печати 07.05.2010 г. Формат 60×84/16.
Бумага «Снегурочка». Печать Xerox.
Усл. печ. л. 8,14. Уч.-изд. л. 7,37.
Заказ . Тираж 300 экз.

Национальный исследовательский
Томский политехнический университет
Система менеджмента качества
Томского политехнического университета сертифицирована
NATIONAL QUALITY ASSURANCE по стандарту ISO 9001:2000

ИЗДАТЕЛЬСТВО ТПУ. 634050, г. Томск, пр. Ленина, 30.
Тел./факс: 8(3822)56-35-35, www.tpu.ru