Содержание

ГЛАВА 1. СЛУЧАЙНЫЕ СОБЫТИЯ

1. 1. Основные понятия теории вероятностей

Событие. Классификация событий

Теория вероятностей **–** раздел математики, который изучает закономерности, имеющие место в однородных массовых испытаниях.

Испытание **–** комплекс, каких-либо условий, действий. Например: стрелок стреляет по мишени; подбрасывается монета; из колоды карт наугад извлекается карта и т. д.

Массовые однородные испытания – такие испытания, которые теоретически могут быть продолжены до бесконечности.

Теория вероятностей интересуется только одной стороной явления: произошло оно в серии массовых однородных испытании или нет.

Исход испытания – возможный результат испытания.

Исходя из этого возникает основное неопределяемое понятие теории вероятностей – событие.

Событие – абстракция исхода, испытания (произошло явление в массовых однородных испытаниях или нет). Приведем примеры событий:

- стрелок стреляет по мишени, разделенной на четыре области. Выстрел – это испытание. Попадание в определенную область мишени – событие;
- в урне имеются цветные шары. Из урны наудачу берут один шар. Извлечение шара из урны есть испытание. Появление шара определенного цвета – событие.

Обозначаются события большими буквами латинского алфавита: A, B, C и т.д.

В необходимых случаях применяются индексы. Все события можно подразделить на три вида: достоверные, невозможные и случайные.

Достоверным называется событие, которое происходит при любом исходе испытания.

Невозможное – такое событие, которое не происходит ни при каком исходе испытания.

Случайным называется событие, которое при испытании может произойти или не произойти.

Поясним данные определения примерами. Появление 10 очков при однократном подбрасывании игральной кости есть событие невозможное. Выпадение не более шести очков при однократном подбрасывании игральной кости – событие достоверное. Появление 5 очков при одном бросании игральной кости есть событие случайное. Игральной костью называют кубик, сделанный из однородного материала, на гранях которого обозначено число очков от 1 до 6.

Среди случайных событий можно выделить:

- ***равновозможные*** события, это такие события, для которых существует равноправие отдельных исходов испытания;
- ***единственно возможные*** события, это такие события, если при испытании обязательно поступит хотя бы одно из них.

Примером единственновозможных и равновозможных событий можно считать появление того или иного числа очков на брошенной игральной кости.

Классическое определение вероятности

Вероятность – одно из основных понятий теории вероятностей. Оно выражает меру объективной возможности наступления события. Существует несколько определений этого понятия. Приведем определение, которое называют классическим.

Вероятностью события *A* называют отношение числа благоприятствующих этому событию исходов к общему числу всех равновозможных несовместных элементарных исходов, образующих полную группу.

Итак, вероятность события А определяется формулой

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

где m – число элементарных исходов, благоприятствующих A; n – число всех возможных элементарных исходов испытания.

Пример. Из полной колоды в 36 карт наудачу извлекается одна. Какова вероятность, что это туз?

Решение:

1. Испытание: из 36 карт извлекается 1.
2. Событие A: появился туз.
3. n = 36 (общее число возможных элементов исходов).
4. m = 4 (число исходов благоприятствует событию A).
5. $P(A)=\frac{m}{n}=\frac{4}{36}=\frac{1}{9}$.

Из определения вероятности вытекают следующие ее свойства.

С в о й с т в о 1. Вероятность достоверного события равна единице.

Действительно, если событие достоверно, то каждый элементарный исход испытания благоприятствует событию. В этом случае $m=n$, следовательно

$$P(A)=\frac{m}{n}=\frac{n}{n}=1.$$

С в о й с т в о 2. Вероятность невозможного события равна нулю.

Действительно, если событие невозможно, то ни один из элементарных исходов испытания не благоприятствует событию. В этом случае $m=0$, следовательно

$$P(A)=\frac{m}{n}=\frac{0}{n}=0.$$

С в о й с т в о 3. Вероятность случайного события есть положительное число, заключенное между нулем и единицей.

Действительно, случайному событию благоприятствует лишь часть из общего числа элементарных исходов испытания. В этом случае $0<m<n,$ значит, $0<\frac{m}{n}<1$,

$$0<P(A)<1.$$

Итак, вероятность любого события удовлетворяет двойному неравенству

$$0\le P(A)\le 1.$$

Основные формулы комбинаторики

Комбинаторика – наука о комбинациях. Она изучает количество комбинаций, подчиненных определенным условиям, которые можно составить из элементов, безразлично какой природы, заданного конечного множества. При непосредственном вычислении вероятностей часто используют формулы комбинаторики. Приведем наиболее употребительные из них.

Перестановками P_n из n элементов называют соединения, состоящие из одних и тех же n различных элементов и отличающихся только порядком их расположения:

$$P_n=n!,$$

где $n!=1\cdot 2\cdot 3\cdots n.$ Причем по определению $0!=1!=1$.

Пример. Скольким числом способов можно расставить на полке восьмитомник?

Решение. Искомое число перестановок

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

Размещениями A_n^m из n элементов по m называются такие соединения, которые отличаются друг от друга либо составом элементов, либо их порядком.

$$A_n^m = n\cdot(n-1)\cdot(n-2)\cdots[n-(m-1)] = \frac{n!}{(n-m)!}.$$

Пример. Сколько различных двузначных чисел можно составить из множества цифр $\{1; 2; 3; 4\}$, причем так, чтобы цифры числа были различны?

Решение. Искомое число чисел

$$A_4^2 = \frac{4!}{(4-2)!} = \frac{1\cdot 2\cdot 3\cdot 4}{1\cdot 2} = 12.$$

Сочетаниями C_n^m из n элементов по m называются такие соединения, которые отличаются друг от друга только составом элементов, порядок соединения элементов не важен.

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

Пример. Скольким числом способов можно в группе из 30 человек распределить три бесплатные путевки?

Решение. Искомое число способов

$$C_n^m = \frac{30!}{3!(30-3)!} = \frac{30!}{3!27!} = 4\ 060.$$

Следует отметить, что число размещений, перестановок и сочетаний связаны равенством

$$A_n^m = P_m C_n^m.$$

При решении задач комбинаторики используют следующие правила.

П р а в и л о с у м м ы. Если некоторый объект ***а*** может быть выбран из совокупности объектов ***r*** способами, а другой объект ***в*** может быть выбран ***s*** способами, то выбрать либо ***а***, либо ***в*** можно ***r + s*** способами.

П р а в и л о п р о и з в е д е н и я. Если объект ***а*** можно выбрать из совокупности объектов ***r*** способами и после каждого такого выбора объект ***в*** можно выбрать ***s*** способами

бами, то пара объектов **(а, *в*)** в указанном порядке может быть выбрана ***rs*** способами.

Относительная частота. Статистические определения вероятности

Относительная частота наряду с вероятностью принадлежит к основным понятиям теории вероятностей.

Относительной частотой события называют отношение числа испытаний, в которых событие произошло, к общему числу фактически произведенных испытаний. Таким образом, относительная чистота события A определяется формулой

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

где m - число появлений события, n - общее число испытаний.

Сопоставляя определение вероятности и относительной частоты, заключаем: определение вероятности не требует, чтобы испытания производились в действительности; определение же относительности частоты предполагает, что испытания были произведены фактически. Другими словами, вероятность вычисляют до опыта, а относительную частоту – после опыта.

Пример. Отдел технического контроля обнаружил три нестандартных детали в партии из 80 случайно отобранных деталей. Относительная частота появления нестандартных деталей

$$W(A)=\frac{3}{80}.$$

Длительные наблюдения показали, что если в одинаковых условиях производят опыты, в каждом из которых число испытаний достаточно велико, то относительная частота обнаруживает свойства устойчивости. Это свойство состоит в том, что в различных опытах относительная частота изменяется мало (тем меньше, чем больше произведено испытаний), колеблясь около некоторого постоянного числа**.** Оказалось, что это постоянное число есть вероятность появления события.

Статистической вероятностью события A называется число, около которого группируются относительные чистоты этого события, причем при неизменных условиях и неограниченном возрастании числа испытаний относительная частота незначительно отличается от этого числа.

Таким образом, если опытным путем установлена относительная частота, то полученное число можно принять за приближенное значение вероятности.

Пример. По данным государственной статистики РФ, относительная частота рождения девочек за последние десять лет характеризуется следующими числами (числа расположены в порядке следования начиная с 1990 г.): 0,489; 0,490; 0,471; 0,478; 0,482; 0,462; 0,484; 0,485; 0,490; 0,482.

Относительная частота колеблется около числа 0,482, которое можно принять за приближенное значение вероятности рождения девочек.

Заметим, что статистические данные различных стран дают примерно то же значение относительной частоты.

Решение типовых задач

Задача 1. Имеется 100 одинаковых деталей, среди которых 3 бракованных. Найти вероятность того, что взятая наудачу деталь без брака.

Решение. В этой задаче производится испытание – извлекается одна деталь. Число всех исходов испытания равно 100, т. к. может быть взята любая деталь из 100. Эти исходы несовместны, равновозможны, единственно возможны. Таким образом, $n=100$. Событие A- появилась деталь без брака. Всего в партии 97 деталей без брака, следовательно, число исходов, благоприятных появлению события A равно 97 . Итак, $m=97$. Тогда $P(A)=\frac{97}{100}=0{,}97.$

Задача 2. Код банковского сейфа состоит из 6 цифр. Найти вероятность того, что наудачу выбранный код содержит различные цифры?

Решение. Так как на каждом из шести мест в шестизначном шифре может стоять любая из десяти цифр: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, то всех различных шестизначных номеров по правилу произведения будет $n=10\cdot10\cdot10\cdot10\cdot10\cdot10=10^6$. Номера, в которых все цифры различны, - это размещения из 10 элементов (10 цифр) по 6. Поэтому число благоприятствующих исходов $m=A_{10}^6$. Искомая вероятность равна

$$P(A)=\frac{A_{10}^6}{10^6}=\frac{10\cdot9\cdot8\cdot7\cdot6\cdot5}{10^6}=0{,}151.$$

Задача 3. Между шестью фирмами (А, Б, В, Г, Д, Е), занимающимися продажей компьютерной техники, проводится жеребьевка на предмет очередности предъявления своей продукции на выставке потенциальным потребителям. Какова вероятность того, что очередь будет выстроена по порядку, т. е. А, Б, В, Г, Д, Е?

Решение. Исход испытания – случайное расположение фирм в очереди. Число всех возможных исходов равно числу всех перестановок из шести элементов (фирм), т.е. $n=P_6=6!$. Число исходов, благоприятствующих событию A: m=1, если очередь выстроена по порядку. Тогда $P(A)=\frac{1}{6!}=\frac{1}{1\cdot2\cdot3\cdot4\cdot5\cdot6}=0{,}0014.$

Задача 4. В компании 10 акционеров, из них трое имеют привилегированные акции. На собрание акционеров явилось 6 человек. Найти вероятность того, что среди явившихся акционеров:

а) все трое акционеров с привилегированными акциями отсутствуют;

б) двое присутствуют и один не явился.

Решение

а) испытанием является отбор 6 человек из 10 акционеров. Число всех исходов испытания равно числу сочетаний из 10 по 6, т. е.

$$n = C_{10}^{6} = \frac{10!}{6! \cdot 4!} = \frac{7 \cdot 8 \cdot 9 \cdot 10}{1 \cdot 2 \cdot 3 \cdot 4} = 210.$$

Пусть событие A - среди шести человек нет ни одного с привилегированными акциями. Исход, благоприятствующий событию A,- отбор шести человек среди семи акционеров, не имеющих привилегированных акций. Число всех исходов, благоприятствующих событию A, будет $m = C_{7}^{6} = \frac{7!}{6! \cdot 1!} = 7.$

Искомая вероятность

$$P(A) = \frac{m}{n} = \frac{7}{210} = \frac{1}{30};$$

б) пусть событие B- среди шести явившихся акционеров двое с привилегированными акциями, а остальные четыре – с общими акциями. Число всех исходов, $n = C_{10}^{6} = 210.$ Число способов выбора двух человек из необходимых трех $m_1 = C_3^2 = \frac{3!}{2! \cdot 1!} = 3.$ Число способов выбора оставшихся четырех акционеров среди семи с общими акциями $m_2 = \frac{7!}{4! \cdot 3!} = \frac{5 \cdot 6 \cdot 7}{1 \cdot 2 \cdot 3} = 35.$Тогда число всех способов отбора по правилу произведения $m = m_1 \cdot m_2 = 3 \cdot 35 = 105$.

Искомая вероятность равна

$$P(A) = \frac{m}{n} = \frac{105}{210} = 0{,}5.$$

Задачи (1 – 10)

1. В двух из 14 составленных кассиром счетов имеются ошибки. Ревизор решил проверить наудачу 5 счетов. Какова вероятность, что а) ошибки не будут обнаружены; б) будет обнаружена хотя бы одна ошибка.

2. Магазин получает товар партиями по 100 штук. Если пять взятых наудачу образцов соответствуют стандартам, партия товара поступает на реализацию. В очередной партии 8 единиц товара с дефектом. Найти вероятность того, что товар поступит на реализацию.

3. В ящике 20 деталей, 5 из них с дефектом. Наудачу извлекают три детали. Найти вероятность того, что среди извлеченных деталей: а) две дефектных; б) хотя бы одна с дефектом.

4. В цехе работают шесть мужчин и четыре женщины. По табельным номерам наудачу отобраны семь человек. Найти вероятность того, что среди отобранных лиц окажутся три женщины.

5. На курсах повышения квалификации бухгалтеров учат определять правильность оформления накладной. Для проверки преподаватель предлагает проверить 12 накладных, 5 из которых содержат ошибки. Наудачу выбирают три накладных. Найти вероятность того, что а) из трех накладных одна с ошибками; б) хотя бы одна с ошибками.

6. Данное предприятие в среднем выпускает 20 % продукции высшего сорта и 70% первого сорта. Найти вероятность того, что случайно взятое изделие окажется первого или высшего сорта.

7. Совет директоров состоит из 3 бухгалтеров, 3 менеджеров и 2 инженеров. Планируется создать подкомитет из трех его членов. Найти вероятность того, что в подкомитет войдут: а) два бухгалтера и менеджер; б) бухгалтер, менеджер и инженер; в) хотя бы один бухгалтер.

8. В группе 12 студентов, среди которых 8 отличников. По списку наудачу отобраны 9 студентов. Найти вероятность того, что среди отобранных студентов пять отличников.

9. Среди 20 компьютеров, поступивших в ремонт в мастерскую, 12 на гарантийном обслуживании. Мастер наудачу берет 2 компьютера для ремонта. Найти вероятность того, что а) оба компьютера находятся на гарантийном обслуживании; б) хотя бы один на гарантии.

10. В ящике 100 деталей, из них 10 бракованных. Наудачу извлечены 4 детали. Найтивероятность того, что среди извлеченных деталей: а) нет бракованных; б) нет годных.

1.2. Классические теоремы теории вероятностей

Теоремы сложения вероятностей

События $A_1, A_2, \ldots, A_n$ называются ***несовместными***, если появление одного из них исключает появление других событий в одном и том же испытании.

Например, если из колоды карт наудачу извлекается одна, то событие A - появился король; событие B - появилась дама; событие C - появился туз – несовместны.

Два события A и $\bar{A}$ называются ***противоположными***, если они несовместны и одно из них обязательно должно произойти в испытании.

Например, если брошена монета, то событие A - появился «герб» и противоположное ему событие $\bar{A}$ - появилось «число».

События $A_1, A_2, \ldots, A_n$ образуют ***полную группу***, если они являются единственно возможными и несовместными.

Пример. Приобретены два лотерейных билета. 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.

Суммой несовместных событий $A_1, A_2, \ldots, A_n$ называется такое событие, которое состоит в наступлении какого-либо одного из событий $A_1, A_2, \ldots, A_n$.

Пример. Испытание: Подбрасываем игральную кость.
Событие A – выпала «пятерка»
Событие B – выпало четное число
Событие $(A+B)$ –выпала «пятерка» или четное число.

Теорема. Вероятность суммы конечного числа несовместных событий равна сумме вероятностей этих событий

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

Доказательство:

1. Докажем теорему для случая двух событий, т.е. $P(A+B) = P(A) + P(B)$. Введем обозначения: n - общие число возможных элементарных исходов испытания; m_1 - число исходов, благоприятствующих событию A; m_2 - число исходов, благоприятствующих события B. Число элемен-

тарных исходов, благоприятствующих наступлению либо события A, либо события B, равно $m_1 + m_2$. Следовательно,

$$P(A+B) = \frac{(m_1 + m_2)}{n} = \frac{m_1}{n} + \frac{m_2}{n}.$$

Приняв во внимание, что $\frac{m_1}{n} = P(A)$ и $\frac{\mathrm{m}_2}{\mathrm{n}} = P(B)$, окончательно получим

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

2. Методом математической индукции докажем теорему для любого конечного числа событий. Предположим, что для n событий утверждение верно, т.е.

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n),$$

и докажем для $(n+1)$ события. Тогда

$$P(A_1 + A_2 + \ldots + A_n + A_{n+1}) = P\left[\underbrace{(A_1 + A_2 + \ldots + A_n)}_{A} + \underbrace{A_{n+1}}_{B}\right] =$$

$$= P(A_1 + A_2 + \ldots + A_n) + P(A_{n+1}).$$

С учетом нашего предположения, имеем

$$P(A_1 + A_2 + \ldots + A_n + A_{n+1}) = \sum_{i=1}^{n+1} P(A_i).$$

Следствие 1. Сумма вероятностей событий A_1, $A_2, \ldots$, A_n, образующих полную группу, равна единице**:**

$$P(A_1) + P(A_2) + \ldots + P(A_n) = 1.$$

Следствие 2. Сумма вероятностей противоположных событий равна единице:

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

Замечание. Если вероятность одного из двух противоположных событий обозначена через p, то вероятность другого события обозначают через q. Таким образом, следствие 2 примет вид:

$$p + q = 1.$$

Пример. Вероятность того, что день будет дождливым, $p = 0{,}7$. Найти вероятность того, что день будет ясным.

Решение. События «день дождливый» и «день ясный» - противоположные, поэтому искомая вероятность

$$q = 1 - p = 1 - 0{,}7 = 0{,}3.$$

События A_1, A_2, ..., A_n называются ***совместными***, если появление одного из них не исключает появление других событий в одном и том же испытании.

Пример. Испытание: Выполняется метеорологический прогноз.

Событие A – ожидается дождь.

Событие B – ожидается ветер.

События A и B совместны.

Суммой n совместных событий A_1, A_2, ..., A_n называется такое событие, которое состоит в наступлении хотя бы одного из событий A_1, A_2, ..., A_n.

Теорема. Вероятность суммы двух совместных событий равна сумме вероятностей этих событий минус вероятность их совместного появления:

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

Доказательство. Поскольку события A и B, по условию совместны, то событие $(A+B)$ наступит, если наступит одно из следующих трех несовместных событий: $A\overline{B}$, $\overline{A}B$ или AB. По теореме сложения вероятностей несовместных событий,

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

(*)

Событие A произойдет, если наступит одно из двух несовместных событий: $A\overline{B}$ или AB. По теореме сложения вероятностей несовместных событий имеем

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

Отсюда

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

(**)

Аналогично имеем

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

Отсюда

$$P(\bar{A}B)=P(B)-P(AB). \qquad (***)$$

Подставив (**) и (***) в (*), окончательно получим

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

Замечание. При использовании полученной формулы следует иметь в виду, что события A и B могут быть как независимыми, так и независимы.

Для независимых событий

$$P(A+B)=P(A)+P(B)-P(A)P(B);$$

для зависимых событий

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

Следствие. Вероятность суммы конечного числа совместных событий вычисляется по формуле

$$P(A_1+A_2+\ldots+A_n)=\sum_i P(A_i)-\sum_{i,j} P(A_i A_j)+$$

$$+\sum_{i,j,k} A_i A_j A_k-\ldots+(-1)^{n-1}\sum_{(i,n)} P\left(\prod_{i=1}^{n} A_i\right).$$

Теоремы умножения вероятностей

Зависимыми называются такие два события, при появлении одного из которых вероятность появления другого изменяется.

Классическим примером зависимых событий является выборка без возвращения.

Пример. В урне имеется 5 шаров, из которых три белых. Извлекается сначала один шар, а затем не возвращая его в урну, второй. Какова вероятность появления белого шара при первом испытании и при втором?

Испытание: Сначала извлекается один шар, затем не возвращая его в урну второй.

Событие A – первый шар белый.

Событие B – второй шар белый.

$$P(A)=\frac{m}{n}=\frac{3}{5}.$$

После первого этапа испытания в урне осталось четыре шара, причем белых из них только два. Поэтому $P_A(B)=\frac{m}{n}=\frac{2}{4}=0{,}5$.

Вероятность $P_A(B)$ называется ***условной*** – вероятность появления события B, при условии, что событие A уже произошло.

Произведением событий $A_1, A_2, \ldots, A_n$ называется такое событие, которое предполагает совместное появления всех этих событий $A_1, A_2, \ldots, A_n$.

Например, если A, B, C – появление «герба» соответственно в первом, втором и третьем бросаниях монеты, то ABC – выпадение «герба» во всех трех испытаниях.

Теорема. Вероятность произведения двух зависимых событий равна произведению вероятности одного из них на условную вероятность другого, вычисленную в предположении, что первое событие уже наступило:

$$P(AB)=\begin{cases}P(A)P_A(B)\\ P(B)P_B(A)\end{cases}.$$

Доказательство. Пусть событию A из n испытаний благоприятствует m исходов. Тогда при появлении A, событию B будет благоприятствовать $k\le m$ исходов. Поэтому совместному появлению событии A и B будет благоприятствовать только k исходов.

$$P(A\cdot B)=\frac{k}{n}=\frac{k}{n}\cdot\frac{m}{m}=\frac{m}{n}\cdot\frac{k}{m}=P(A)\cdot P_A(B).$$

Вторая часть теоремы доказывается аналогично.

Следствие. Вероятность совместного появления n событий равна произведению вероятности одного из них на условные вероятности всех остальных, причем вероятность каждого последующего события вычисляется в предположении, что все предыдущие события уже появились:

$$P(A_1A_2A_3\ldots A_n)=P(A_1)P_{A_1}(A_2)P_{A_1A_2}(A_3)\ldots P_{A_1A_2\ldots A_{n-1}}(A_n).$$

Независимыми называются такие два события, при появлении одного из которых, вероятность появления другого не меняется.

Примером независимых событий может служить выборка с возвращением.

Несколько событий называют ***независимыми в совокупности*** (или просто независимыми), если независимы каждые два из них и независимы каждое событие и все возможные произведения остальных. Например, если события A_1, A_2, A_3 независимы в совокупности, то независимы события A_1 *и* A_2, A_1 *и* A_3, A_2 *и* A_3; A_1 *и* A_2A_3, A_2 *и* A_1A_3, A_3 *и* A_1A_2.

Теорема. Вероятность произведения двух независимых событий равна произведению вероятностей этих событий:

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

Доказательство. Так как события A и B - независимы, то $P_A(B)=P(B)$ и $P_B(A)=P(A)$. Тогда по теореме о вероятности произведения зависимых событий

$$P(AB)=\begin{cases}P(A)\cdot P_A(B)=P(A)\cdot P(B)\\ P(B)\cdot P_B(A)=P(B)\cdot P(A)\end{cases}.$$

Таким образом, $P(AB)=P(A)\cdot P(B)$.

Следствие. Вероятность совместного появления нескольких событий, независимых в совокупности, равна произведению вероятностей этих событий:

$$P(A_1A_2\ldots A_n)=P(A_1)P(A_2)\ldots P(A_n).$$

Вероятность появления хотя бы одного события

Пусть в результате испытания могут появиться n событий, независимых в совокупности, либо некоторые из них (в частности, только одно или ни одного), причем вероятности появления каждого из событий известны. Как найти вероятность того, что наступит хотя бы одно из этих событий? Например, если в результате испытания могут появиться три события, то появление хотя бы одного из этих событий означает наступление либо одного, либо двух, либо трех событий. Ответ на поставленный вопрос дает следующая теорема.

Теорема. Вероятность появления хотя бы одного из событий $A_1, A_2, \ldots, A_n$, независимых в совокупности, равна разности между единицей и произведением вероятностей противоположных событий $\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_n$:

$$P(A) = 1 - q_1 q_2 \ldots q_n .$$

Доказательство. Обозначим через A событие, состоящее в появлении хотя бы одного из событий $A_1, A_2, \ldots, A_n$. События A и $\bar{A}_1 \bar{A}_2 \ldots \bar{A}_n$ (ни одного из событий не наступило) противоположны, следовательно, сумма их вероятностей равна единице:

$$P(A) + P(\bar{A}_1 \bar{A}_2 \ldots \bar{A}_n) = 1 .$$

Отсюда, пользуясь теоремой умножения, получим

$$P(A) = 1 - P(\bar{A}_1 \bar{A}_2 \ldots \bar{A}_n) = 1 - P(\bar{A}_1) P(\bar{A}_2) \ldots P(\bar{A}_n)$$

или

$$P(A) = 1 - q_1 q_2 \ldots q_n .$$

Следствие. Если события $A_1, A_2, \ldots, A_n$ имеют одинаковую вероятность, равную p, то вероятность появления хотя бы одного из этих событий:

$$P(A) = 1 - q^n .$$

Формулы полной вероятности и Бейеса

Пусть событие A может наступить при условии появления одного из несовместных событий (гипотез) H_1, H_2, ..., H_n, которые образуют полную группу. Пусть известны вероятности этих событий и условные вероятности $P_{H_1}(A), P_{H_2}(A), \ldots, P_{H_n}(A)$ события A. Как найти вероятность события A? Ответ на этот вопрос дает следующая теорема.

Теорема. Вероятность события A, которое может наступить лишь при условии появления одной из несовместных гипотез $H_1, H_2, \ldots, H_n$, образующих полную группу, равна сумме произведений вероятностей каждой из этих гипотез на соответствующую условную вероятность события A:

$$P(A) = P(H_1)P_{H_1}(A) + P(H_2)P_{H_2}(A) + \ldots + P(H_n)P_{H_n}(A),$$

где $\sum\limits_{i=1}^{n} P(H_i) = 1$. Эту формулу называют «формулой полной вероятности».

Доказательство. По условию, событие A может наступить, если наступит одна из несовместных гипотез $H_1, H_2, \ldots, H_n$. Другими словами, появление события A означает осуществление одного, безразлично какого, из несовместных событий $H_1A, H_2A, \ldots, H_nA$. Пользуясь для вычисления вероятности события A теоремой сложения, получим

$$P(A) = P(H_1A) + P(H_2A) + \ldots + P(H_nA). \quad (*)$$

Остается вычислить каждое из слагаемых. По теореме умножения вероятностей зависимых событий имеем

$$P(H_1A) = P(H_1)P_{H_1}(A);\ P(H_2A) = P(H_2)P_{H_2}(A); \ldots;\ P(H_nA) = P(H_n)P_{H_n}(A).$$

Подставим правые части этих равенств в соотношение (*), получим формулу полной вероятности

$$P(A) = P(H_1)P_{H_1}(A) + P(H_2)P_{H_2}(A) + \ldots + P(H_n)P_{H_n}(A).$$

Допустим, что произведено испытание, в результате которого появилось событие A. Поставим своей задачей определить, как изменились (в связи с тем, что событие A уже наступило) вероятности гипотез. Другими словами, будем искать условные вероятности:

$$P_A(H_1),\ P_A(H_2), \ldots, P_A(H_n).$$

Найдем условную вероятность $P_A(H_i)$. По теореме умножения имеем

$$P(AH_i) = P(A)P_A(H_i) = P(H_i)P_{H_i}(A).$$

Отсюда

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

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

$$P_A(H_i) = \frac{P(H_i)P_{H_i}(A)}{P(H_1)P_{H_1}(A) + P(H_2)P_{H_2}(A) + + \ldots + P(H_n)P_{H_n}(A)} =$$

$$= \frac{P(H_i) \cdot P_{H_i}(A)}{\sum\limits_{i=1}^{n} P(H_i)P_{H_i}(A)}.$$

Полученную формулу называют формулой Бейеса. Формула Бейеса позволяет переоценить вероятности гипотез после того, как становится известным результат испытания, в итоге которого появилось событие A.

Решение типовых задач

Задача 1. Стрелок производит два выстрела по мишени. Вероятность попадания при каждом выстреле 0,8. Составить полную группу событий и найти их вероятности.

Решение.

Испытание – Производится два выстрела по мишени.

Событие A – оба раза промахнулся.

Событие B – попал один раз.

Событие C – оба раза попал.

$P(A)=q^2=(0{,}2)^2=0{,}04$.

$P(B)=pq+qp=0{,}8\cdot 0{,}2\cdot 2=0{,}32.$

$P(C)=p^2=(0{,}8)^2=0{,}64.$

Контроль: $P(A)+P(B)+P(C)=1$.

Задача 2. Согласно прогнозу метеорологов Р(дождь)=0,4; Р(ветер)=0,7; Р(дождь и ветер)=0,2. Какова вероятность того, что будет дождь или ветер?

Решение. По теореме сложения вероятностей и в силу совместности предложенных событий имеем:

Р(дождь или ветер или то и другое)=Р(дождь) +Р(ветер) –Р(дождь и ветер)=0,4+0,7-0,2=0,9.

Задача 3. На станции отправления имеется 8 заказов на отправку товара: пять – внутри страны, а три – на экспорт. Какова вероятность того, что два выбранных наугад заказа окажутся предназначенными для потребления внутри страны?

Решение. Событие A – первый взятый наугад заказ – внутри страны. Событие B – второй тоже предназначен для внутреннего потребления. Нам необходимо найти вероятность $P(A\cdot B)$. Тогда по теореме об умножении вероятностей зависимых событий имеем

$$P(A\cdot B)=P(A)\cdot P_A(B)=\frac{5}{8}\cdot\frac{4}{7}=\frac{5}{14}.$$

Задача 4. Из партии изделий товаровед наудачу отбирает изделия высшего сорта. Вероятность того, что выбранная вещь окажется высшего сорта равна, 0,8; первого сорта – 0,7; второго сорта – 0,5. Найти вероятность того, что из трех наудачу отобранных изделий будут:

а) только два высшего сорта;

б) все разные.

Решение. Пусть событие A_1 - изделие высшего сорта; событие A_2- изделие первого сорта; событие A_3- изделие второго сорта.

По условию задачи $P(A_1)=0{,}8$; $P(A_2)=0{,}7$; $P(A_3)=0{,}5.$ События A_1, A_2, A_3- независимы.

а) Событие A – только два изделия высшего сорта будет выглядеть так $A=A_1A_1A_2+A_1A_1A_3$, тогда

$$P(A)=P(A_1A_1A_2+A_1A_1A_3)=P(A_1)\cdot P(A_1)\cdot P(A_2)+P(A_1)\cdot P(A_1)\cdot P(A_3)=$$

$$=(0{,}8)^2\cdot 0{,}7+(0{,}8)^2\cdot 0{,}5=0{,}768.$$

б) Событие B – все три изделия различны – выразим так: $B=A_1\cdot A_2\cdot A_3$, тогда $P(B)=0{,}8\cdot 0{,}7\cdot 0{,}5=0{,}28$.

Задача 5. Вероятности попадания в цель при стрельбе из трех орудий таковы: p_1=0,8; p_2=0,7; p_3=0,9. Найти вероятность хотя бы одного попадания (событие A) при одном залпе из всех орудий.

Решение. Вероятность попадания в цель каждым из орудий не зависит от результатов стрельбы из других орудий, поэтому рассматриваемые события A_1 (попадание первого орудия), A_2 (попадание второго орудия) и A_3 (попадание третьего орудия) независимы в совокупности.

Вероятности событий, противоположных событиям A_1, A_2 и A_3 (т.е. вероятности промахов), соответственно равны:

$q_1=1-p_1=1-0{,}8=0{,}2;$

$q_2=1-p_2=1-0{,}7=0{,}3;$

$q_3=1-p_3=1-0{,}9=0{,}1.$

Искомая вероятность

$$P(A)=1-q_1q_2q_3=1-0{,}2\cdot 0{,}3\cdot 0{,}1=0{,}994.$$

Задача 6. В типографии имеется 4 печатных машины. Для каждой машины вероятность того, что она работает в данный момент, равна 0,9. Найти вероятность того, что в данный момент работает хотя бы одна машина (событие A).

Решение. События «машина работает» и «машина не работает» (в данный момент) – противоположные, поэтому сумма их вероятностей равна единице:

$$p+q=1.$$

Отсюда вероятность того, что машина в данный момент не работает, равна

$$q=1-p=1-0{,}9=0{,}1.$$

Искомая вероятность

$$P(A)=1-q^4=1-0{,}1^4=0{,}999.$$

Задача 7. Фирма имеет три источника поставки комплектующих – фирмы *A,B,C*. На долю фирмы *A* приходится 50 % общего объема поставок, *B* – 30 % и *C* – 20 %. Из практики известно, что 10 % поставляемых фирмой *A* деталей – бракованные, фирмой *B* – 5 % и *C* – 6 %. Найти вероятность того, что наудачу выбранная деталь будет бракованной.

Решение. Производится испытание – извлекается одна деталь.

Событие *A* – появилась бракованная деталь.

Гипотеза H_1 – деталь фирмы *A*.

Гипотеза H_2 – деталь фирмы *B*.

Гипотеза H_3 – деталь фирмы *C*.

Тогда, согласно формуле полной вероятности, искомая вероятность равна:

$$P(A)=P(H_1)\cdot P_{H_1}(A)+P(H_2)\cdot P_{H_2}(A)+\ldots+P(H_n)\cdot P_{H_n}(A),$$

$$P(H_1)=\frac{50}{100}=0{,}5;\ P_{H_1}(A)=\frac{10}{100}=0{,}1.$$

$$P(H_2)=\frac{30}{100}=0{,}3;\ P_{H_2}(A)=\frac{5}{100}=0{,}05.$$

$$P(H_3)=\frac{20}{100}=0{,}2;\ P_{H_3}(A)=\frac{6}{100}=0{,}06.$$

$$P(A)=0{,}5\cdot 0{,}1+0{,}3\cdot 0{,}05+0{,}2\cdot 0{,}06=0{,}077.$$

Задача 8. В центральную бухгалтерию корпорации поступили пачки накладных для проверки и обработки. 90 % пачек были признаны удовлетворительными: они содержали только 1 % неправильно заполненных накладных. Остальные 10 % пачек были признаны неудовлетворительными, так как содержали 5 % неверно оформленных накладных. Взятая наугад из пачки накладная оказалась оформленной неверно. Учитывая это, какова вероятность того, что вся пачка накладных будет признана несоответствующей стандарту?

Решение. Испытание – проверяется пачка накладных.
Событие A – взятая наугад накладная оказалась неверной.
Гипотеза H_1 – пачка не соответствует стандарту.
Гипотеза H_2 – пачка соответствует стандарту.

Необходимо узнать вероятность гипотезы H_1 при условии, что событие A произошло. Согласно формуле Бейеса имеем:

$$P_A(H_1)=\frac{P(H_1)\cdot P_{H_1}(A)}{P(H_1)\cdot P_{H_1}(A)+P(H_2)\cdot P_{H_2}(A)}.$$

$$P(H_1)=\frac{10}{100}=0{,}1;\ P_{H_1}(A)=\frac{5}{100}=0{,}05.$$

$$P(H_2)=\frac{90}{100}=0{,}9;\ P_{H_2}(A)=\frac{1}{100}=0{,}01.$$

$$P_A(H_1)=\frac{0{,}1\cdot 0{,}05}{0{,}1\cdot 0{,}05+0{,}9\cdot 0{,}01}\approx 0{,}257.$$

Задачи (11 – 20)

11. Два бухгалтера обрабатывают равное количество счетов. Вероятность того, что первый бухгалтер допустит ошибку, равна 0,005, для второго эта вероятность равна 0,01. При проверке счетов была найдена ошибка. Найти вероятность того, что ее допустил первый бухгалтер.

12. Вероятности получения прибыли с акций компаний A, B и C для акционера соответственно равны 0,6; 0,7; 0,5. Найти вероятность того, что а) только один вид акций составит прибыль акционера, б) хотя бы один вид акций принесет доход их обладателю.

13. Надежность первого банка – 0,95; для второго – 0,8; для третьего – 0,85. Предприниматель совершил вклад во все три банка. Найти вероятность того, что предпринимателю вернут вклад: а) два банка; б) хотя бы один банк.

14. Станок работает при условии одновременного функционирования узлов *A*, *B*, *C*, которые работают независимо друг от друга. Вероятность поломки этих узлов 0,2; 0,3; 0,1 соответственно. Какова вероятность того, что станок выйдет из строя?

15. Вероятность того, что в определенный день торговой базе потребуется двух- тонная машина, равна 0,9, пятитонная – 0,7. Определить вероятность того, что торговой базе потребуется хотя бы одна автомашина.

16. Вероятность своевременного возвращения кредитов каждым из трех заемщиков банку независимы и соответственно равны: 0,6; 0,9; 0,7. Найти вероятность следующих событий:

а) только два заемщика возвратят кредит своевременно;

б) хотя бы один из заемщиков возвратит кредит своевременно.

17. Заводом послана автомашина за различными материалами на четыре базы.

Вероятность наличия нужного материала на первой базе равна 0,9, на второй –

0,95, на третьей – 0,8, на четвертой – 0,6. Найти вероятность того, что только на

одной базе не окажется нужного материала.

18. Вероятность того, что налоговая инспекция предъявит штраф первому предприятию – 0,2, второму – 0,3, третьему – 0,15. Найти вероятность того, что будут оштрафованы: а) три предприятия; б) два предприятия.

19. Технологический процесс состоит из нескольких операций. Вероятность того, что во время первой операции изделие получит повреждение, равна 0,1, а во время второй операции – 0,05. Какова вероятность того, что после двух операций изделие окажется поврежденным?

20. Банк может выдать кредит одному из трех клиентов с вероятностью $p_1 = 0{,}4,\ p_2 = 0{,}3,\ p_3 = 0{,}3$ соответственно. Найти вероятность того, что кредит получит только один клиент.

1.3. Повторные независимые испытания

Схема и формула Бернулли

Если производится несколько испытаний, причем вероятность события A в каждом испытании не зависит от исходов других испытаний, то такие испытания называют ***независимыми*** относительно события A.

В независимых испытаниях событие A может иметь либо различные вероятности, либо одну и ту же вероятность. Будем далее рассматривать лишь такие независимые испытания, в которых событие A имеет одну и ту же вероятность.

Классическая схема таких испытаний носит название схемы Бернулли, она подразумевает выполнение трех основных требований:

1. Все n испытаний независимы друг от друга.
2. Каждое испытание имеет два исхода (событие A произошло или не произошло).

3. Вероятность события A в каждом испытании постоянна $P(A)=p$, тогда вероятность ненаступления события A в каждом испытании также постоянна и равна $P(\overline{A})=1-p=q$.

Поставим перед собой задачу найти вероятность того, что в n испытаниях, удовлетворяющих схеме Бернулли, событие A произойдет ровно m раз. Искомую вероятность обозначим $P_n(m)$.

Поставленную задачу можно решить с помощью теоремы Бернулли.

Теорема. Если проводится n испытаний, удовлетворяющих схеме Бернулли, то вероятность того, что событие A произойдет в них ровно m раз вычисляется по формуле

$$P_n(m)=C_n^m \cdot p^m \cdot q^{n-m} \quad, m=\left(\overline{0,n}\right).$$

Доказательство. Пусть событие A в n испытаниях произошло m раз, тогда наступило событие: $\underbrace{A\cdot A\cdot\ldots\cdot A}_{m}\cdot\underbrace{\overline{A}\cdot\overline{A}\cdot\ldots\cdot\overline{A}}_{n-m}$.

Вероятность этого события по теореме умножения вероятностей независимых событий равна $P\left(A\ldots A\cdot\overline{A}\ldots\overline{A}\right)=p^m\cdot q^{n-m}$.

Таких сложных событий может быть столько, сколько можно составить сочетаний из n элементов по m элементов, т.е. C_n^m.

Так как эти сложные события несовместны, то по теореме сложения вероятностей несовместных событий искомая вероятность равна сумме вероятностей всех возможных сложных событий. Поскольку же вероятности всех этих событий одинаковы, то искомая вероятность (появления m раз события A в n испытаниях) равна вероятности одного сложного события, умноженной на их число:

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

или

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

Полученную формулу называют ***формулой Бернулли***.

Следствие 1. Вероятность того, что в n испытаниях, удовлетворяющих схеме Бернулли, событие A пройдет хотя бы один раз, вычисляется по формуле

$$P_n(m \geq 1) = 1 - P_n(0) = 1 - q^n.$$

Следствие 2. Вероятность того, что в n испытаниях, удовлетворяющих схеме Бернулли, событие A произойдет от m_1 до m_2 раз, вычисляется по формуле

$$P_n(m_1 \leq m \leq m_2) = \sum_{m=m_1}^{m_2} P_n(m) = \sum_{m=m_1}^{m_2} C_n^m \cdot p^m \cdot q^{n-m}.$$

Локальная теорема Лапласа

При большом числе испытаний применение формулы Бернулли к нахождению вероятности того, что в n испытаниях событие появится ровно m раз, довольно затруднительно. Поэтому к решению подобных задач применяют ***локальную теорему Лапласа***, которая позволяет приближенно найти искомую вероятность.

Теорема. Если в схеме испытаний Бернулли число испытаний достаточно велико, а вероятность появления события A в каждом испытании фиксирована $(0 < p < 1)$, то

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x),$$

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

Свойства функции $\varphi(x)$

1. Область определения функции: $x \in (-\infty; \infty)$.
2. Область значений функции: $\varphi(x) > 0$.
3. Функция четная, т. е. $\varphi(-x) = \varphi(x)$.

4. График функции имеет горизонтальную асимптоту $y=0$, так как $\lim\limits_{x\to\pm\infty}\varphi(x)=0$.

5. Максимальное значение функции $\varphi_{\max}=\varphi(0)=\dfrac{1}{\sqrt{2\pi}}$.

6. График функции имеет две точки перегиба $\left(-1;\dfrac{1}{\sqrt{2\pi e}}\right)$ и $\left(1;\dfrac{1}{\sqrt{2\pi e}}\right)$.

7. $\int\limits_{-\infty}^{\infty}\varphi(x)dx=1$.

График функции $\varphi(x)$ называется кривой Гаусса (рис. 1.1).

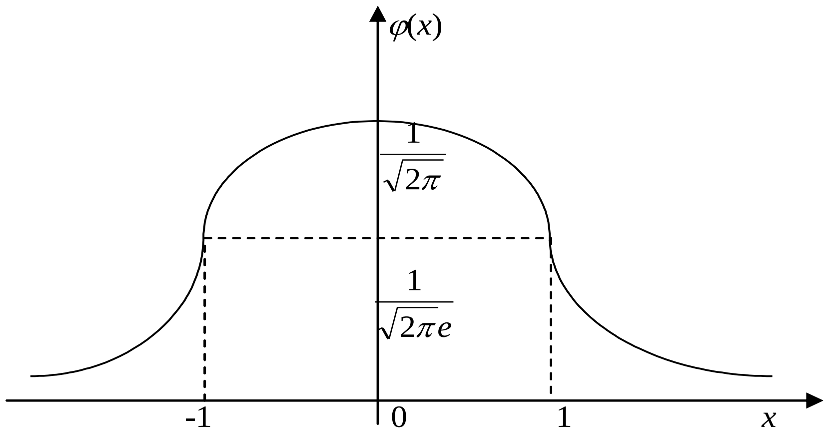


Рис. 1.1.

Функция $\varphi(x)$ табулирована. Таблица ее значений приведена в приложении 1. Для значений аргумента $|x|\geq 4$ функцию принято считать равной нулю. Например значение функции $\varphi(-5{,}34)=\varphi(5{,}34)\approx 0$.

Теорема Пуассона

Если при большом числе испытаний вероятность появления события близка к нулю, то для нахождения вероятности того, что в n испытаниях событие произойдет ровно m раз, используют теорему Пуассона.

Теорема. Если в условиях схемы Бернулли число испытаний неограниченно возрастает, а вероятность появления события A в каждом испытании уменьшается так, что $np=\lambda$ остается постоянной величиной, то

$$P_n(m)\approx\frac{\lambda^m}{m!}e^{-\lambda}.$$

Формула Пуассона тем точнее для расчетов, чем больше n и меньше p. Обычно ею пользуются при $p\le 0{,}03$ или при условии $\lambda=n\cdot p\le 10$.

Следствие 1. В условиях теоремы Пуассона вероятность того, что событие A произойдет хотя бы один раз, вычисляется по формуле

$$P_n(m\ge 1)\approx 1-e^{-\lambda}.$$

Следствие 2. В условиях теоремы Пуассона вероятность того, что событие A произойдет от m_1 до m_2 раз, вычисляется по формуле

$$P_n(m_1\le m\le m_2)=\sum_{m=m_1}^{m_2}\frac{\lambda^m}{m!}e^{-\lambda}.$$

Для облегчения вычислений по формуле Пуассона существуют таблицы для нахождения значений выражения $P_n(m;\lambda)\approx\frac{\lambda^m}{m!}e^{-\lambda}$ в зависимости от параметров m и λ (приложение 2).

Интегральная теорема Лапласа

Вновь предположим, что производится n испытаний, в каждом из которых вероятность появления события A постоянна и равна $p, (0<p<1)$. Предположим, что необходимо найти вероятность того, что в этих испытаниях событие A произойдет не менее m_1 и не более m_2 раз (от m_1 до m_2 раз).

Если число слагаемых на отрезке $[m_1; m_2]$ невелико, то для решения данной задачи можно использовать теорему о вероятности суммы попарно несовместных событий и тогда

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

Каждую из вероятностей $P_n(m)$ в этой формуле можно вычислять по одной из формул (Бернулли, Пуассона, локальная Лапласа), удовлетворяющих условию задачи.

В случае если число испытаний велико, а также число слагаемых на отрезке $[m_1; m_2]$ велико, к решению применяют интегральную теорему Лапласа.

Теорема. Если выполняется схема испытаний Бернулли, то при большом числе испытаний справедлива формула

$$P_n(m_1 \le m \le m_2) \approx \Phi(x_2) - \Phi(x_1),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$; $x_1 = \frac{m_1 - np}{\sqrt{npq}}$; $x_2 = \frac{m_2 - np}{\sqrt{npq}}$.

Свойства функции $\Phi(x)$

1. Область определения функции: $x \in (-\infty; \infty)$.
2. Область значений функции: $\Phi(x) \in (-0{,}5; 0{,}5)$.
3. Функция нечетная, т.е. $\Phi(-x) = -\Phi(x)$.
4. Функция монотонно возрастает на всей области определения.
5. График функции имеет две горизонтальные асимптоты $y = \pm\frac{1}{2}$, так как $\lim\limits_{x \to \pm\infty} \Phi(x) = \pm\frac{1}{2}$

График функции $\Phi(x)$ имеет вид (рис. 1.2).

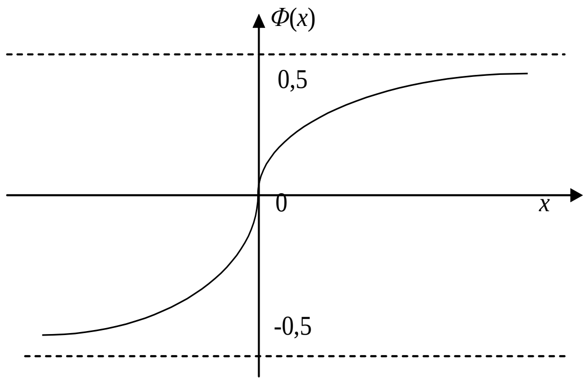


Рис. 1.2.

Функция $\Phi(x)$ табулирована. Таблица ее значений приведена в приложении 3. Для значений аргумента $x>5$ функцию принято считать равной 0,5. Например значение функции $\Phi(-10{,}3)=-\Phi(10{,}3)\approx -0{,}5$.

Следствие 1. Вероятность абсолютной величины отклонения числа появлений события *A* от произведения *np* не более чем на $\varepsilon>0$ находится по формуле

$$P\left(|m-np|\le\varepsilon\right)\approx 2\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right).$$

Доказательство. Заменим неравенство $|m-np|\le\varepsilon$ ему равносильными:

$$-\varepsilon\le m-np\le\varepsilon \text{ или } np-\varepsilon\le m\le np+\varepsilon.$$

По интегральной теореме Лапласа искомая вероятность

$$P\left(|m-np|\le\varepsilon\right)=P\left(np-\varepsilon\le m\le np+\varepsilon\right)=\Phi\left(\frac{np+\varepsilon-np}{\sqrt{npq}}\right)-\Phi\left(\frac{np-\varepsilon-np}{\sqrt{npq}}\right)\approx 2\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right).$$

Следствие 2. Вероятность того, что абсолютная величина отклонения относительной частоты появления события *A* от его вероятности не превысит $\varepsilon>0$, находится по формуле

$$P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)\approx 2\Phi\left(\frac{\varepsilon n}{\sqrt{npq}}\right).$$

Доказательство проводится аналогично.

Наивероятнейшее число появлений события в независимых испытаниях

Число m_0 (наступления события в n испытаниях, удовлетворяющих схеме Бернулли) называют ***наивероятнейшим***, если вероятность того, что событие наступит в этих испытаниях m_0 раз, будет максимальной.

Наивероятнейшее число m_0 определяют из двойного неравенства

$$np - q \le m_0 \le np + p,$$

причем:

1) если число $np - q$ – дробное, то существует одно наивероятнейшее число m_0;
2) если число $np - q$ – целое, то существуют два наивероятнейших числа, а именно
m_0 и m_0+1;
3) если число np – целое, то наивероятнейшее число $m_0=np$.

Задача 1. Вероятность выигрыша по одному любому лотерейному билету равна 0,02. Чему равна вероятность выигрыша: а) по трем билетам; б) не более двух билетов; в) хотя бы по одному билету; для владельца четырех билетов.

Решение: $n=4$; $p=0{,}02$; $q=0{,}98$.

Так как число испытаний мало, применяем формулу Бернулли $P_n(m)=C_n^m \cdot p^m \cdot q^{n-m}$.

а) $P_4(3)=C_4^3\cdot(0{,}02)^3\cdot(0{,}98)^1 \approx 3\cdot 10^{-5}$;

б) $P_4(0\le m\le 2)=P_4(0)+P_4(1)+P_4(2)=C_4^0\cdot(0{,}02)^0\cdot(0{,}98)^4 +$
$+C_4^1\cdot(0{,}02)^1\cdot(0{,}98)^3+C_4^2\cdot(0{,}02)^2\cdot(0{,}98)^1=0{,}099$;

в) $P_4(m\ge 1)=1-P_4(0)=1-(0{,}98)^4=0{,}078$.

Задача 2. Доля изделий высшего сорта на данном предприятии составляет 90 %. Найти вероятность того, что среди наудачу взятых 100 изделий высшего сорта окажется 84 изделия.

Решение: n=100; p=0,9; m=84; np=90.

Так как число испытаний n=100 – велико, а произведение $np=\lambda$=90>10, применяем к решению локальную формулу Лапласа:

$$P_{100}(84)\approx\frac{1}{\sqrt{100\cdot 0{,}9\cdot 0{,}1}}\cdot\varphi\left(\frac{84-90}{\sqrt{100\cdot 0{,}9\cdot 0{,}1}}\right)=\frac{1}{3}\cdot\varphi(-2).$$

Так как функция $\varphi(x)$- четная, имеем: $P_{100}(84)\approx\frac{1}{3}\cdot\varphi(2)$.

По таблице приложения 1 находим $\varphi(2)=0{,}054$.

Искомая вероятность

$$P_{100}(84)\approx\frac{1}{3}\cdot 0{,}054\approx 0{,}018.$$

Задача 3. Магазин получил 1000 бутылок минеральной воды. Вероятность того, что при перевозке бутылка окажется разбитой, равна 0,003. Найти вероятность того, что магазин получил разбитых бутылок: а) ровно две; б) меньше двух; в) больше одной; г) хотя бы одну.

Решение: n=1000; p=0,003; $\lambda=1000\times 0{,}003=3$.

Так как число испытаний n=1000 – велико, p – мало, а $\lambda=n\cdot p=3<10$, применяем формулу Пуассона:

а) $P_{1000}(2)\approx\frac{3^2\cdot e^{-3}}{2!}\approx 0{,}224$;

б) $P_{1000}(m<2)=P_{1000}(0)+P_{1000}(1)\approx\frac{3^0\cdot e^{-3}}{0!}+\frac{3\cdot e^{-3}}{1!}\approx 0{,}049+0{,}149\approx 0{,}198$;

в) $P_{1000}(m>1)=1-P_{1000}(m\le 1)=1-[P_{1000}(0)+P_{1000}(1)]\approx 0{,}8$;

г) $P_{1000}(m\ge 1)=1-e^{-3}=1-(2{,}71)^{-3}\approx 0{,}95$.

Задача 4. Доля изделий высшего сорта продукции составляет 80 %. Найти вероятность того, что в партии из 900 изделий высшего сорта будет: а) заключено между 700 и 750; б) не меньше 750; в) не больше 600.

Решение: n=900, p=0,8, q=0,2, np=720. $\sqrt{npq}=\sqrt{900\cdot 0{,}8\cdot 0{,}2}=12$.

Согласно интегральной теореме Лапласа

а) $P_{900}(700\le m\le 750)\approx \Phi\left(\frac{750-720}{12}\right)-\Phi\left(\frac{700-720}{12}\right)=\Phi\left(\frac{5}{2}\right)-\Phi\left(-\frac{5}{3}\right)$.

Функция $\Phi(x)$ - нечетная, поэтому $P_{900}(700\le m\le 750)=\Phi(2{,}5)+\Phi(1{,}67)$.

По таблице приложения 3 находим $\Phi(2{,}5)=0{,}4938$; $\Phi(1{,}67)=0{,}4521$.

Искомая вероятность $P_{900}(700\le m\le 750)=0{,}4938+0{,}4521=0{,}9459$

б) $P_{900}(m\ge 750)=P_{900}(750\le m\le 900)=\Phi\left(\frac{900-720}{12}\right)-\Phi\left(\frac{750-720}{12}\right)=$

$=\Phi(15)-\Phi(2{,}5)=0{,}5-0{,}4938=0{,}0062$;

в) $P_{900}(m\le 600)=P_{900}(0\le m\le 600)=\Phi\left(\frac{600-720}{12}\right)-\Phi\left(\frac{0-720}{12}\right)=$

$=\Phi(-10)-\Phi(-60)=-\Phi(10)+\Phi(60)=-0{,}5+0{,}5=0$.

Задача 5. Предприятие поставляет свою продукцию 15 магазинам, от каждого из которых может поступить заявка на очередной день с вероятностью 0,6, независимо от заявок других магазинов. Найти наивероятнейшее число заявок.

Решение: n=15; p=0,6; q=0,4. Найдем наивероятнейшее число заявок из двойного неравенства:

$$np-q\le m_0\le np+p.$$

Подставив данные задачи, получим

$$15\cdot 0{,}6-0{,}4\le m_0\le 15\cdot 0{,}6+0{,}6;$$

$$8{,}6\le m_0\le 9{,}6.$$

Так как m_0 – целое число, то наивероятнейшее число m_0=9.

Задачи (21 – 30)

21. Вероятность того, что фирма, проведя рекламную кампанию, продаст единицу своей продукции, составляет 0,8. Найти вероятность того, что из 100 изделий фирма реализует не менее 75.

22. Вероятность своевременной поставки продукции для каждого из пяти поставщиков постоянна и равна 0,7. Найти вероятность того, что своевременно поставят продукцию от двух до четырех поставщиков.

23. На 150 предприятиях края была произведена аудиторская проверка хозяйственой деятельности. Найти вероятность того, что у 50 предприятий были выявлены серьезные нарушения, если вероятность подобных нарушений для каждого объекта составляет 0,3.

24. Вероятность производственной травмы в течение года равна 0,000 5. Какова вероятность того, что из 10 000 рабочих пострадает 3 человека? Каково наивероятнейшее число пострадавших?

25. Вероятность того, что случайно выбранный лицевой счет клиента отделения сбербанка содержит ошибки равна 0,05. Если при выборочной проверке счетов обнаружится, что не менее 6 % отобранных счетов содержат ошибки, то оператор увольняется с работы. Найти вероятность того, что оператор будет уволен, если ревизор проверит 500 счетов.

26. Необходимо перевезти 500 единиц товара грузовым автомобилем, но определенное количество товара может получить в дороге повреждение. Вероятность повреждения единицы товара равна 0,2. Необходимо найти наивероятнейшее число поврежденного товара, чтобы отправить сразу такое количество товара, которое могло бы гарантировать, что необходимую партию товара довезут.

27. В автобусном парке имеется 100 машин. Вероятность выхода автобуса на линию
равна 0,9. Для обеспечения нормальной работы маршрутов необходимо иметь на линиях не менее 90 машин. Определить вероятность нормального функционирования автобусных маршрутов.

28. Фотолаборатория взяла на себя обязательство выполнить 130 заказов для клиента, с вероятностью выполнения одного заказа 0,9. Найти вероятность того, что фотолаборатория выполнит 110 заказов.

29. Вероятность того, что посетитель магазина совершит покупку, равна 0,5. Найти вероятность того, что из 8 посетителей покупку сделает: а) не более двух человек, б) не менее двух человек.

30. 90 % изделий данного предприятия – это продукция высшего сорта. Найти вероятность того, что из 600 приобретенных Вами изделии первого сорта будет от 500 до 550.

Вопросы для самоконтроля

1. Среди 10 деталей 2 бракованные. Наудачу извлекается деталь. Найти вероятности возможных событий.

101. $\frac{2}{10}$ и $\frac{10}{2}$. ***102.*** $\frac{1}{10}$ и $\frac{1}{20}$. ***103.*** $\frac{2}{10}$ и $\frac{8}{10}$. ***104.*** $\frac{2}{10}$ и $\frac{8}{10}$.

2. Вероятность попадания стрелком в цель 0,7. Найти вероятность противоположного события.

201. 0,5. ***202.*** 1,7. ***203.*** $\frac{1}{70}$. ***204.*** 0,3.

3. События А и В несовместны. Какова вероятность Р(А + В) ?

301. $P(A + B) = P(A) + P(B) - P(AB)$. ***302.*** $P(A + B) = P(A) + P(B)$.

303. $P(A + B) = P(A) - P(B)$.   ***304.*** $P(A + B) = P_A(B) + P(A)$.

4. Вероятность того, что студент сдаст первый экзамен, – 0,7; второй – 0,6. Найти вероятность того, что он сдаст один экзамен.

401. $P = 0{,}7 + 0{,}6 - 0{,}7 \cdot 0{,}6 = 0{,}88$.   ***402.*** $P = 0{,}7 \cdot 0{,}3 + 0{,}6 \cdot 0{,}4 = 0{,}45$.

403. $P = 0{,}7 \cdot 0{,}4 + 0{,}3 \cdot 0{,}6 = 0{,}46$.   ***404.*** $P = 0{,}7 \cdot (1 - 0{,}6) = 0{,}28$.

5. Два события называются …, если при появлении одного из них изменяется вероятность появления другого.

501. совместными.   ***502.*** равновозможными.   ***503.*** зависимыми.   ***504.*** случайными.

6. Вероятность появления хотя бы одного из событий A_1, A_2, …, A_n, независимых в совокупности, определяется по формуле

601. $P(A) = p_1 + p_2 + \ldots + p_n$.   ***602.*** $P(A) = 1 - p^n$.

603. $P(A) = p_1 q_1 + \ldots + p_n q_n$.   ***604.*** $P(A) = 1 - q_1 q_2 \ldots q_n$.

7. Вероятность того, что посетитель данного магазина совершит покупку равна 0,6. Найти вероятность того, что из пяти первых посетителей сделают покупку трое.

701. $P_5(3) = C_5^3 \cdot (0{,}6)^3 \cdot (0{,}4)^2$.   ***702.*** $P_5(3) = \dfrac{(5 \cdot 0{,}6)^3}{3!} e^{-3}$.

703. $P_5(3) = \dfrac{1}{\sqrt{5 \cdot 0{,}6 \cdot 0{,}4}} \cdot \varphi\left(\dfrac{3-5}{\sqrt{1{,}2}}\right)$.   ***704.*** $P_5(3) = 1 - P_5(2)$.

8. Локальная функция Лапласа $\varphi(x)$ - …, а интегральная функция Лапласа $\Phi(x)$ -…

801. возрастающая, убывающая.   ***802.*** четная, нечетная.

803. нечетная, возрастающая.   ***804.*** убывающая, четная.

9. Событие, которое предполагает совместное появление событий A_1 , A_2, …, A_n называется … этих событий.

901. произведением. ***902.*** суммой. ***903.*** комбинацией. ***904.*** разностью.

10. Если в условиях схемы Бернулли число испытаний неограниченно возрастает, а вероятность появления события А в каждом испытании уменьшается так, что $np = \lambda \le 10$***, то вероятность того, что в*** n ***испытаниях событие А произойдет ровно*** m ***раз, вычисляют по формуле.***

1001. Бернулли. ***1002.*** Пуассона.

1003. Локальной Лапласа. ***1004.*** Интегральной Лапласа.

ГЛАВА 2. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

2.1. Случайная величина. Закон распределения случайной величины. Функция распределения

Виды случайных величин

Случайной называют величину, которая в результате испытания примет одно и только одно возможное значение, наперед неизвестное и зависящее от случайных причин, которые заранее не могут быть учтены.

Пример 1. Число родившихся мальчиков среди ста новорожденных есть случайная величина, которая имеет следующие возможные значения: 0, 1, 2, …, 100.

Пример 2. Расстояние, которое пролетит снаряд при выстреле из орудия, есть случайная величина. Действительно расстояние зависит не только от установки прицла, но и от многих других причин (силы и направление ветра, температуры и т.д.), которые не могут быть полностью учтены. Возможные значения этой величины принадлежат некоторому промежутку (a, b).

Обозначаются случайные величины прописными буквами X, Y, Z, а их возможные значения – соответствующими строчными буквами x, y, z. Например, если случайная величина X имеет три возможных значения, то они будут обозначены так: x_1, x_2, x_3.

Вернемся к примерам, приведенным выше. В первом из них случайная величина X могла принять одно из следующих возможных значений: 0, 1, 2, …, 100. Эти значения отделены одно от другого промежутками, в которых нет возможных значений X. Таким образом, в этом примере случайная величина принимает отдельные, изолированные возможные значения. Во втором примере случайная величина могла принять любое из значений промежутка (a, b). Здесь нельзя отделить одно возможное значение от другого промежутком, не содержащим возможных значений случайной величины.

В зависимости от множества значений, принимаемых случайной величиной, выделяют дискретные и непрерывные случайные величины.

Дискретной случайной величиной (ДСВ) называется случайная величина, множество значений которой конечно или бесконечно, но счетно.

Непрерывной случайной величиной (НСВ) называется случайная величина, которая может принимать все значения из некоторого конечного или бесконечного промежутка.

Закон распределения случайной величины

На первый взгляд может показаться, что для задания дискретной случайной величины достаточно перечислить все ее возможные значения. В действительности это не так: случайные величины могут иметь одинаковые перечни возможных значений, а вероятности их – различные. Поэтому для задания дискретной случайной величины недостаточно перечислить все возможные ее значения, нужно еще указать их вероятности.

Законом распределения случайной величины называется соответствие между ее возможными значениями и вероятностями, с которыми она их принимает.

Закон распределения можно задавать тремя способами.

1. Табличный.

Табличным способом можно задать только дискретную случайную величину. В этом случае первая строка таблицы содержит возможные значения, а вторая – их вероятности:

X	x_1	x_2	$\dots$	x_n
P	p_1	p_2	$\dots$	p_n

Приняв во внимание, что в одном испытании случайная величина принимает одно и только одно возможное значение, заключаем, что события $X = x_1, X = x_2, \dots, X = x_n$ образуют полную группу; следовательно, сумма вероятностей этих событий, т.е. сумма вероятностей второй строки таблицы, равна единице:

$$\sum_{i=1}^{n} p_i = p_1 + p_2 + \dots + p_n = 1.$$

Если множество возможных значений X бесконечно (счетно), то ряд $p_1 + p_2 + \dots$ сходится и его сумма равна единице.

2. Графический.

Для наглядности закон распределения дискретной случайной величины можно изобразить и графически, для чего в прямоугольной системе координат строят точки (x_i, p_i), а затем соединяют их отрезками. Полученную фигуру называют ***многоугольником распределения*** (рис. 2.1).

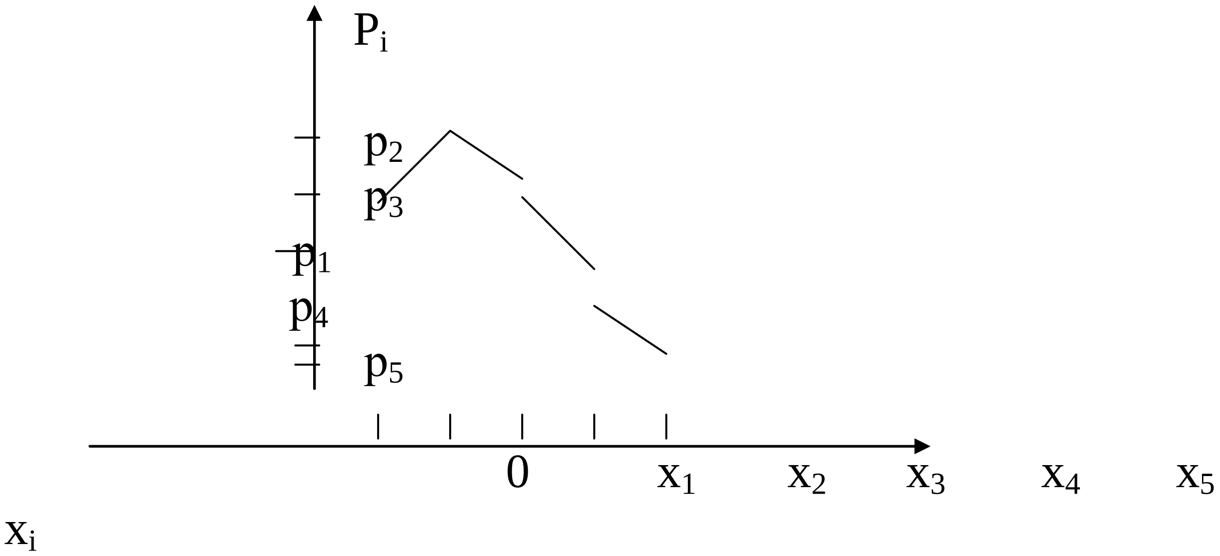


Рис. 2.1.

Для непрерывной случайной величины данный график будет представлять собой ***кривую распределения*** вероятностей.

3. Аналитический.

Закон распределения случайной величины можно задавать также формулой. Приведем пример случайной величины, распределенной по биномиальному закону, тогда вероятность того, что в n испытаниях случайная величина примет значение, равное m, вычисляется по формуле Бернулли:

$$P_n(X = m) = C_n^m \cdot p^m \cdot q^{n-m}, m = \left(\overline{0, n}\right).$$

Основные законы распределения дискретных случайных величин

Равномерный закон распределения.

Дискретная случайная величина X называется распределенной по равномерному закону, если она принимает свои возможные значения с постоянной вероятностью

$$P(X = x_i) = \frac{1}{n}, \quad \left(i = \overline{1, n}\right)$$

Запишем равномерный закон в виде таблицы

X	x_1	x_2	…	x_n
P	$\frac{1}{n}$	$\frac{1}{n}$	…	$\frac{1}{n}$

Биномиальный закон распределения.

Для этого закона дискретная случайная величина X – число появлений события A при проведении испытаний, удовлетворяющих схеме Бернулли.

Вероятности ее возможных значений вычисляются по формуле

$$P_n(X = m) = C_n^m \cdot p^m \cdot q^{n-m}; m = \left(\overline{0, n}\right).$$

Запишем биномиальный закон в виде таблицы

X	0	1	…	m	…	n
P	q^n	npq^{n-1}	…	$C_n^m \cdot p^m \cdot q^{n-m}$	…	p^n

Закон распределения *Пуассона*.

Для этого закона дискретная случайная величина X – число появлений события A при проведении испытаний, удовлетворяющих теореме Пуассона.

Тогда вероятность того, что в n испытаниях случайная величина X примет значение равное m вычисляется по формуле

$$P_n(X = m) = \frac{\lambda^m}{m!} e^{-\lambda}; \quad m = (0, 1, 2 \ldots), \text{ где } \lambda = np.$$

Запишем закон Пуассона в виде таблицы

X	0	1	…	m	…

P	$e^{-\lambda}$	$\lambda e^{-\lambda}$	…	$\frac{\lambda^m}{m!}e^{-\lambda}$	…

Геометрический **закон распределения.**

Для этого закона дискретная случайная величина X – число проведенных испытаний до первого появления события A, если испытания удовлетворяют схеме Бернулли.

Тогда вероятность того, что в n испытаниях случайная величина X примет значение равное m вычисляется по формуле

$$P_n(X=m)=p\cdot q^{m-1}, (m=1,2,3\ldots).$$

Запишем геометрический закон в виде таблицы

X	1	2	3	…	m	…
P	p	$p\cdot q$	$p\cdot q^2$	…	$p\cdot q^{m-1}$	…

Гипергеометрический закон распределения.

Пусть имеется множество из n элементов, из которых s элементов обладают фиксированным свойством. Пусть из этого множества осуществляется выборка из r элементов.

Тогда дискретная случайная величина X – число элементов с фиксированным свойством, оказавшихся в выборке, а вероятности ее возможных значений определяются по формуле

$$P_n(X=m)=\frac{C_s^m\cdot C_{n-s}^{r-m}}{C_n^r}.$$

Функция распределения, ее свойства и график

Пусть x – действительное число. Вероятность события, состоящего в том, что X примет значение меньше x, т.е. вероятность события $X<x$, обозначим через $F(x)$. Если x изменяется, то изменяется и $F(x)$, т.е. $F(x)$ – функция от x.

Функцией распределения называют функцию *F(x)*, определяющую вероятность того, что случайная величина X в результате испытания примет значение меньшее x, т.е.

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

Функция распределения представляет собой общий способ задания любых типов случайных величин.

Свойства функции распределения

С в о й с т в о 1. Значения функции распределения принадлежат отрезку $[0;1]$:

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

Доказательство. Свойство вытекает из определения функции распределения как вероятности: вероятность всегда есть неотрицательное число, не превышающее единицы.

С в о й с т в о 2. $F(x)$ - неубывающая функция, т.е.

$$F(x_2) \ge F(x_1), \text{ если } x_2 > x_1.$$

Доказательство. Пусть $x_2 > x_1$. Событие, состоящее в том, что X примет значение меньшее x_2, можно подразделять на следующие два несовместных события: 1) X примет значение, меньшие x_1, с вероятностью $P(X < x_1)$; 2) X примет значение, удовлетворяющее неравенству $x_1 \le X < x_2$, с вероятностью $P(x_1 \le X < x_2)$. По теореме сложения имеем

$$P(X < x_2) = P(X < x_1) + P(x_1 \le X < x_2).$$

Отсюда

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

или

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

(*)

Так как любая вероятность есть число неотрицательное, то $F(x_2) - F(x_1) \ge 0$, или $F(x_2) \ge F(x_1)$, что и требовалось доказать.

Следствие 1. Вероятность того, что случайная величина примет значение, заключенное в интервале (a, b), равна приращению функции распределения на этом интервале:

$$P(a < X < b) = F(b) - F(a).$$

Это важное следствие вытекает из формулы (*), если положить $x_2 = b$ и $x_1 = a$.

Пример. Случайная величина *X* задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le -1; \\ \frac{x}{4} + \frac{1}{4} & \text{при} \quad -1 < x \le 3; \\ 1 & \text{при} \quad x > 3. \end{cases}$$

Найти вероятность того, что в результате испытания *X* примет значение, принадлежащее интервалу (0, 2):

$$P(0 < X < 2) = F(2) - F(0).$$

Решение. Так как на интервале (0, 2) по условию,

$$F(x) = \frac{x}{4} + \frac{1}{4},$$

то

$$F(2) - F(0) = \left(\frac{2}{4} + \frac{1}{4}\right) - \left(\frac{0}{4} + \frac{1}{4}\right) = \frac{1}{2}.$$

Итак,

$$P(0 < x < 2) = 0{,}5.$$

Следствие 2. Вероятность того, что непрерывная случайная величина *X* примет одно определенное значение, равна нулю.

С в о й с т в о 3. Если возможные значения случайной величины принадлежат интервалу (a, b), то: 1) $F(x) = 0$ при $x \le a$; 2) $F(x) = 1$ при $x \ge b$.

Доказательство. 1. Пусть $x_1 \le a$. Тогда событие $X < x_1$ невозможно (так как значений, меньших x_1, величина *X* по условию не принимает) и, следовательно, вероятность его равна нулю.

2. Пусть $x_2 \ge b$. Тогда событие $X < x_2$ достоверно (так как все возможные значения X меньше x_2) и, следовательно, вероятность его равна единице.

Следствие. Если возможные значения непрерывной случайной величины расположены на всей оси x, то справедливы следующие предельные соотношения:

$$\lim_{x\to-\infty} F(x)=0; \quad \lim_{x\to\infty} F(x)=1.$$

Доказанные свойства позволяют представить, как выглядит график функции распределения непрерывной случайной величины.

График расположен в полосе, ограниченной прямыми $y=0,\ y=1$ (первое свойство).

При возрастании x в интервале (a, b), в котором заключены все возможные значения случайной величины функция возрастает (второе свойство).

При $x \le a$ ординаты графика равны нулю; при $x \ge b$ ординаты графика равны единице (третье свойство).

График функции распределения непрерывной случайной величины изображен на рис. 2.2.

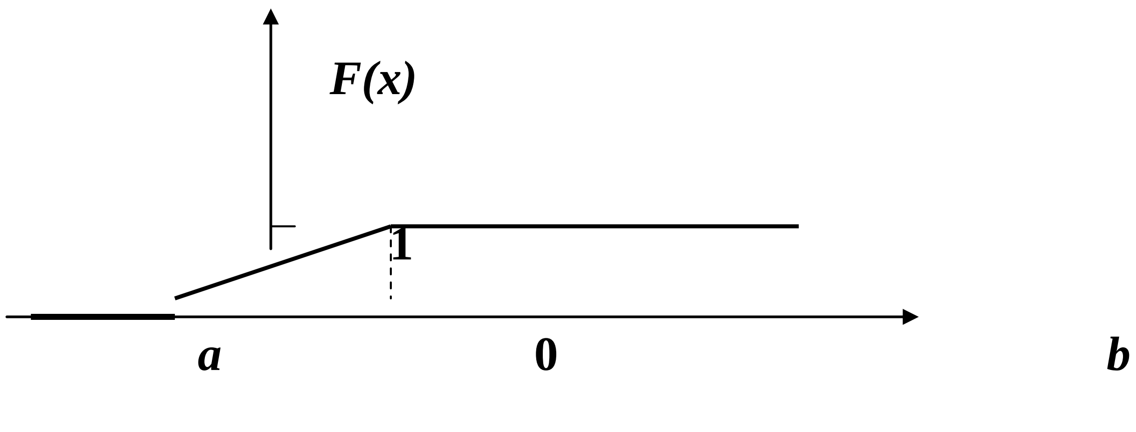


Рис. 2.2.

Замечание. График функции распределения дискретной случайной величины имеет ступенчатый вид. Убедимся в этом на примере.

Пример. Дискретная случайная величина *X* задана законом распределения

X	1	4	8
P	0,3	0,1	0,6

Найти функцию распределения и построить ее график.

Решение. Если $x \le 1$, то $F(x) = 0$(третье свойство).

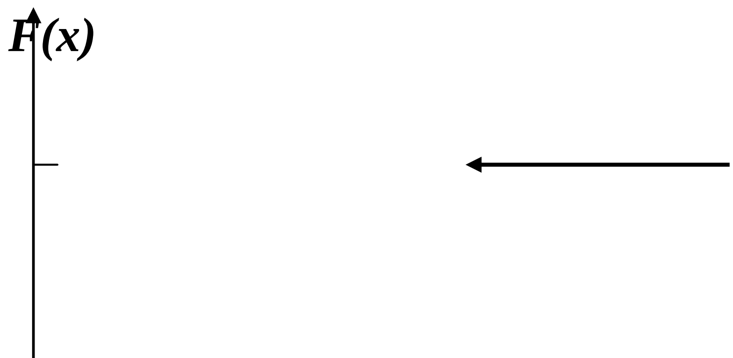
Если $1 < x \le 4$, то $F(x) = 0{,}3$. Действительно, *X* может принять значение 1 с вероятностью 0,3.

Если $4 < x \le 8$, то $F(x) = 0{,}4$. Действительно, если x_1удовлетворяет неравенству

$4 < x_1 \le 8$, то $F(x_1)$ равно вероятности события $X < x_1$, которое может быть осуществлено, когда *X* примет значение 1 (вероятность этого события равна 0,3) или значение 4 (вероятность этого события равна 0,1). Поскольку эти два события несовместны, то по теореме сложения вероятность события $X < x_1$ равна сумме вероятностей 0,3 + 0,1=0,4. Если $x > 8$, то $F(x) = 1$. Действительно, событие $X \le 8$ достоверно, следовательно, его вероятность равна единице. Итак, функция распределения аналитически может быть записана так:

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le 1, \\ 0{,}3 & \text{при} \quad 1 < x \le 4, \\ 0{,}4 & \text{при} \quad 4 < x \le 8, \\ 1 & \text{при} \quad x > 8. \end{cases}$$

График этой функции приведен на рис. 2.3.



1

0,4
0,3

0 1 4 8

x

Рис. 2.3.

Решение типовых задач

Задача 1. В магазине куплено 3 электроприбора: чайник, утюг и пылесос. Вероятность выхода из строя в течение гарантийного срока для каждого из них соответственно равны $p_1 = 0{,}05,\ \ p_2 = 0{,}1,\ \ p_3 = 0{,}2$. Составить закон распределения случайной величины X – числа приборов, вышедших из строя в течение гарантийного срока.

Решение. X – число приборов, вышедших из строя, имеет следующие возможные значения:

$x_1 = 0$ – все три прибора не выйдут из строя в течении гарантийного срока;

$x_2 = 1$ – один прибор выйдет из строя;

$x_3 = 2$ – два прибора выйдут из строя;

$x_4 = 3$ – три прибора выйдут из строя.

Найдем соответствующие этим значениям вероятности. По условию, вероятности выхода из строя приборов равны: $p_1 = 0{,}05;\ p_2 = 0{,}1;\ p_3 = 0{,}2,$ тогда вероятности того, что приборы будут рабочими в течение гарантийного срока равны:

$q_1 = 1 - p_1 = 1 - 0{,}05 = 0{,}95;$

$q_2 = 1 - p_2 = 1 - 0{,}1 = 0{,}9;$

$q_3 = 1 - p_3 = 1 - 0{,}2 = 0{,}8.$

$P(X = 0) = q_1 \cdot q_2 \cdot q_4 = 0{,}95 \cdot 0{,}9 \cdot 0{,}8 = 0{,}684.$

$P(X = 1) = q_1 \cdot q_2 \cdot p_3 + q_1 \cdot p_2 \cdot q_3 + p_1 \cdot q_2 \cdot q_3 = 0{,}95 \cdot 0{,}9 \cdot 0{,}2 + 0{,}95 \cdot 0{,}1 \cdot 0{,}8 +$
$+ 0{,}05 \cdot 0{,}9 \cdot 0{,}8 = 0{,}283.$

$P(X = 2) = p_1 \cdot p_2 \cdot q_3 + p_1 \cdot q_2 \cdot p_3 + q_1 \cdot p_2 \cdot p_3 = 0{,}05 \cdot 0{,}1 \cdot 0{,}8 + 0{,}05 \cdot 0{,}9 \cdot 0{,}2 +$
$+ 0{,}95 \cdot 0{,}1 \cdot 0{,}2 = 0{,}032.$

$P(X = 3) = p_1 \cdot p_2 \cdot p_3 = 0{,}05 \cdot 0{,}1 \cdot 0{,}2 = 0{,}001.$

Закон распределения имеет вид:

X	0	1	2	3
P	0,684	0,283	0,032	0,001

Проверка:

$P = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0{,}684 + 0{,}283 + 0{,}032 + 0{,}001 = 1.$

Задача 2. Предприятие выпускает 90 % изделий высшего сорта. Составить закон распределения случайной величины X – числа изделий высшего сорта из трех, взятых наудачу изделий.

Решение. Случайная величина X – число изделий высшего сорта среди трех отобранных изделий может принимать одно из значений: 0, 1, 2, 3. Вероятности этих значений вычисляются по формуле Бернулли:

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

$n = 3;\ p = 0{,}9;\ q = 0{,}1;\ m = 0{,}1{,}2{,}3.$

$P_3(X = 0) = (0{,}1)^3 = 0{,}001;\ P_3(X = 1) = C_3^1 \cdot 0{,}9^1 \cdot 0{,}1^2 = 0{,}027.$

$P_3(X = 2) = C_3^2 \cdot 0{,}9^2 \cdot 0{,}1 = 0{,}243;\ P_3(X = 3) = 0{,}9^3 = 0{,}729.$

Закон распределения случайной величины X:

X	0	1	2	3
P	0,001	0,027	0,243	0,729

Проверка: $P = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$ = 0,001 + 0,027 + 0,243 +
+ 0,729 = 1.

Задача 3. На сборку поступило 30 деталей, из них 25 стандартных. Сборщик берет наудачу три детали. Составить закон распределения случайной величины X – числа стандартных деталей среди трех взятых.

Решение. Случайная величина X – число стандартных деталей среди отобранных, подчиняется гипергеометрическому закону распределения $\left(n = 30; s = 25; r = 3; m = \left(\overline{0,3}\right)\right)$. Воспользуемся формулой: $P_n(x = m) = \dfrac{C_s^m \cdot C_{n-s}^{r-m}}{C_n^r}$

$$P_{30}(X = 0) = \frac{C_{25}^0 \cdot C_5^3}{C_{30}^3} = \frac{1}{406}; \quad P_{30}(X = 1) = \frac{C_{25}^1 \cdot C_5^2}{C_{30}^3} = \frac{25}{406}$$

$$P_{30}(X = 2) = \frac{C_{25}^2 \cdot C_5^1}{C_{30}^3} = \frac{150}{406}; \quad P_{30}(X = 3) = \frac{C_{25}^3 \cdot C_5^0}{C_{30}^3} = \frac{230}{406}$$

Закон распределения случайной величины X:

X	0	1	2	3
P	$\frac{1}{406}$	$\frac{25}{406}$	$\frac{150}{406}$	$\frac{230}{406}$

Проверка: $P = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \frac{1}{406} + \frac{25}{406} + \frac{150}{406} + \frac{230}{406} = 1.$

Задача 4. Дискретная случайная величина задана законом распределения

X	2	4	5	7
P	0,2	0,1	0,3	0,4

Составить функцию распределения $F(x)$ и построить ее график. Найти вероятность того, что в результате испытания X примет значение, принадлежащие интервалу (3;6).

Решение. 1) По определению $F(x)=P(X<x)$, т.е. $F(x)$ есть вероятность того, что случайная величина X примет значение меньше, чем x.

1. При $x \le 2$ $F(x)=P(X<2)=0.$

2. При $2<x\le 4$ $F(x)=P(X<4)=P(X=2)=0{,}2.$

3. При $4<x\le 5$ $F(x)=P(X<5)=P(X=2)+P(X=4)=0{,}2+0{,}1=0{,}3.$

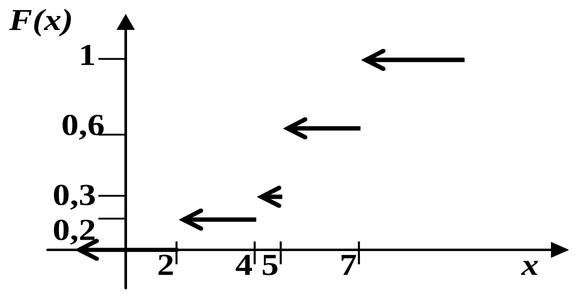
4. При $5<x\le 7$ $F(x)=P(X<7)=P(X=2)+P(X=4)+P(X=5)=$
$=0{,}2+0{,}1+0{,}3=0{,}6.$

5. При $x>7$ $F(x)=P(X=2)+P(X=4)+P(X=5)+P(X=7)=$
$=0{,}2+0{,}1+0{,}3+0{,}4=1.$

Таким образом функция распределения примет вид:

$$F(x)=\begin{cases}0, & x\le 2\\ 0{,}2, & 2<x\le 4\\ 0{,}3, & 4<x\le 5\\ 0{,}6, & 5<x\le 7\\ 1, & x>7\end{cases}$$

Построим график *F(x)*:



2) Найдем вероятность $P(3<X<6)$ по формуле $P(a<X<b)=F(b)-F(a)$, тогда $P(3<X<6)=F(6)-F(3)=0{,}6-0{,}2=0{,}4.$

Задачи (31 – 40)

31. Известно, что 20 % хабаровчан предпочитают добираться на работу личным автотранспортом. Случайно выбраны 4 человека. Составить закон распределения числа людей, предпочитающих добираться на работу личным автотранспортом, среди отобранных. Составить функцию распределения, построить ее график.

32. В некотором цехе брак составляет 5 % всех изделий. Составить закон распределения числа бракованных изделий в случайно отобранной партии из трех изделий. Составить функцию распределения и построить ее график.

33. Вероятность выигрыша по одному лотерейному билету – 0,1. Составить закон распределения случайного числа выигрышных билетов среди пяти купленных. Составить функцию распределения, построить ее график.

34. Торговый агент в среднем контактирует с 6 потенциальными покупателями в день. Из опыта ему известно, что вероятность того, что потенциальный покупатель совершит покупку, равна 0,1. Составить закон распределения ежедневного числа продаж для агента. Составить функцию распределения, построить ее график.

35. Практика показывает, что 7 % накладных, проходящих проверку в бухгалтерии, оказываются неправильно оформленными. Наугад отобраны пять накладных. Составить закон распределения случайного числа накладных, не содержащих ошибки. Составить функцию распределения, построить ее график.

36. Для того чтобы проверить точность своих финансовых счетов, компания регулярно пользуется услугами аудиторов для проверки бухгалтерских проводок счетов. Известно, что служащие компании при обработке входящих счетов допускают 5 % ошибок. Аудитор случайно отбирает 3 входящих документа.

Составить закон распределения числа ошибок, выявленных аудитором. Составить функцию распределения, построить ее график.

37. Предприятие в среднем выпускает 90 % изделий высшего сорта. Составить закон распределения случайного числа изделий высшего сорта из взятых наугад четырех изделий. Составить функцию распределения, построить ее график.

38. Контролер проверяет на соответствие стандарту 5 изделий. Вероятность того, что каждое из изделий будет признано годным, равна 0,9. Составить закон распределения числа стандартных изделий среди проверенных. Составить функцию распределения, построить ее график.

39. В ходе аудиторской проверки строительной компании аудитор случайным образом отбирает 5 счетов. Известно, что 3 % счетов содержат ошибки. Составить закон распределения правильных счетов. Составить функцию распределения, построить ее график.

40. Устройство состоит из трех независимо работающих элементов. Вероятность отказа в одном опыте для каждого элемента равна 0,1. Составить закон распределения случайного числа отказавших элементов в одном опыте. Составить функцию распределения, построить ее график.

2.2. Числовые характеристики дискретных случайных величин

Наиболее полно описывает случайную величину описывает ее закон распределения, который задает все возможные ее значения и соответствующие вероятности. Однако зачастую нет необходимости характеризовать случайную величину так досконально. Бывает достаточно указать несколько параметров случайной величины.

Такие параметры, выражающие наиболее важные особенности случайной величины, называются ***числовыми характеристиками*** случайной величины.

Первой группой таких характеристик являются характеристики положения случайной величины на числовой оси. К ним относятся: математическое ожидание, мода, медиана – это некоторые числа, вокруг которых группируются все возможные значения случайной величины.

К другой группе числовых характеристик относятся характеристики, оценивающие меру рассеяния (разброса) случайной величины. Таковыми прежде всего, являются дисперсия и среднее квадратичное отклонение. Это некоторые числа, которые показывают, сколь больших отклонений значений случайной величины от ее среднего значения можно ожидать.

Математическое ожидание дискретной случайной величины

Математическим ожиданием дискретной случайной величины называют сумму произведений всех ее возможных значений на их вероятности.

Пусть случайная величина X может принимать только значения $x_1, x_2, \ldots, x_n$, вероятности которых соответственно равны $p_1, p_2, \ldots, p_n$. Тогда математическое ожидание *M(X)* случайной величины *X* определяется равенством

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

Выясним вероятностный смысл математического ожидания.

Пусть произведено *n* испытаний, в которых случайная величина *X* приняла m_1 раз значение x_1, m_2 раз значение $x_2, \ldots, m_k$ раз значение x_k, причем $m_1 + m_2 + \ldots + m_k = n$. Тогда сумма всех значений, принятых *X*, равна

$$x_1 m_1 + x_2 m_2 + \ldots x_k m_k \text{ .}$$

Найдем среднее арифметическое $\overline{X}$ всех значений, принятых случайной величиной, для чего разделим найденную сумму на общее число испытаний:

$$\overline{X} = \frac{(x_1 m_1 + x_2 m_2 + \ldots + x_k m_k)}{n}$$

или

$$\overline{X} = x_1\left(\frac{m_1}{n}\right) + x_2\left(\frac{m_2}{n}\right) + \ldots + x_k\left(\frac{m_k}{n}\right).$$

Заметив, что отношение $\frac{m_1}{n}$ - относительная частота W_1 значения x_1, $\frac{m_2}{n}$ - относительная частота W_2 значения x_2 и т.д., тогда

$$\overline{X} = x_1 W_1 + x_2 W_2 + \ldots + x_k W_k \text{ .}$$

Допустим, что число испытаний достаточно велико. Тогда, как известно, относительная частота приближенно равна вероятности появления события.

$$W_1 \approx p_1,\ W_2 \approx p_2, \ldots,\ W_k \approx p_k \text{ .}$$

Заменив относительные частоты соответствующими вероятностями, получим

$$\overline{X} \approx x_1 p_1 + x_2 p_2 + \ldots + x_k p_k \text{ .}$$

Правая часть этого приближенного равенства есть *M(X)*.

Итак,

$$\overline{X} \approx M(X) \text{ .}$$

Вероятностный смысл полученного результата таков: математическое ожидание приближенно равно (тем точнее, чем больше число испытаний) среднему арифметическому наблюдаемых значений случайной величины.

Математические операции над случайными величинами

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

Y	y_1	y_2	…	y_m
P'	p'_1	p'_2	…	p'_m

Пусть случайные величины X и Y заданы следующими законами распределения

Две случайные величины называют ***независимыми***, если закон распределения одной из них не зависит от того, какие возможные значения приняла другая величина. В противном случае случайные величины ***зависимы***. Несколько случайных величин называют ***взаимно независимыми***, если законы распределения любого числа из них не зависят от того, какие возможные значения приняли остальные величины.

Произведением постоянной C на случайную величину X называется новая случайная величина $Z = CX$, которая принимает свои значения $z_k = Cx_i$ с вероятностями $P(Z = z_k) = P(X = x_i)$.

Суммой двух случайных величин X и Y называется случайная величина $Z = X + Y$ для которой

1) возможные значения равны всевозможным суммам возможных значений случайных величин X и Y, т.е.

$$z_k = x_i + y_j; \quad \left(i = (\overline{1,n}); \quad j = (\overline{1,m})\right);$$

2) соответствующие ей вероятности находятся из условия:

$$P(Z=z_k)=\sum_{z_k=x_i+y_j} P(X=x_i)\cdot P(Y=y_j);\ \left(i=(\overline{1,n});\ j=(\overline{1,m})\right)$$

в случае, если случайные величины X и Y независимы и по формуле

$$P(Z=z_k)=\sum_{z_k=x_i+y_j} P(X=x_i)P_{X=x_i}(Y=y_j);\ \left(i=(\overline{1,n});\ j=(\overline{1,m})\right),$$

если случайные величины X и Y зависимы.

Замечание. Аналогично определяется разность случайных величин.

Произведением двух случайных величин X и Y называется случайная величина $Z=XY$, для которой

1) возможные значения равны всевозможным произведениям возможных значений
   случайных величин X и Y, т.е.
   $$z_k=x_i\cdot y_j;\ \left(i=(\overline{1,n});\ j=(\overline{1,m})\right);$$
2) соответствующие ей вероятности находятся из условия
   $$P(Z=z_k)=\sum_{z_k=x_i\cdot y_j} P(X=x_i)\cdot P(Y=y_j);\ \left(i=(\overline{1,n});\ j=(\overline{1,m})\right)$$

в случае, если случайные величины X и Y независимы и по формуле

$$P(Z=z_k)=\sum_{z_k=x_i\cdot y_j} P(X=x_i)\cdot P_{x=x_i}(Y=y_j);\ \left(i=(\overline{1,n});\ j=(\overline{1,m})\right),$$

если случайные величины X и Y зависимы.

Свойства математического ожидания

С в о й с т в о 1. Математическое ожидание постоянной величины равно самой постоянной:

$$M(C)=C.$$

Доказательство. Будем рассматривать постоянную C как дискретную случайную величину, которая имеет одно возможное значение C и принимает его с вероятностью $p=1$. Следовательно,

$$M(C)=C\cdot 1=C.$$

С в о й с т в о 2. Постоянный множитель можно выносить за знак математического ожидания:

$$M(CX)=CM(X).$$

Доказательство. Пусть случайная величина X задана законом распределения вероятностей:

X	x_1	x_2	$\ldots$	x_n
P	p_1	p_2	$\ldots$	p_n

Запишем закон распределения случайной величины CX:

CX	Cx_1	Cx_2	$\ldots$	Cx_n
P	p_1	p_2	$\ldots$	p_n

Математическое ожидание случайной величины CX:

$$M(CX)=Cx_1p_1+Cx_2p_2+\ldots+Cx_np_n=$$
$$=C(x_1p_1+x_2p_2+\ldots+x_np_n)=CM(X).$$

Итак,

$$M(CX)=CM(X).$$

С в о й с т в о 3. Математическое ожидание суммы (разности) двух случайных величин равно сумме (разности) математических ожиданий слагаемых:

$$M(X\pm Y)=M(X)\pm M(Y).$$

Доказательство. 1. Докажем равенство $M(X+Y)=M(X)+M(Y)$. Согласно определению математического ожидания и определению суммы двух случайных величин будем иметь

$$M(X+Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i+y_j\right)p_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i p_{ij}+\sum_{i=1}^{n}\sum_{j=1}^{m}y_j p_{ij}=$$

$$=\sum_{i=1}^{n}x_i\cdot\sum_{j=1}^{m}p_{ij}+\sum_{j=1}^{m}y_j\cdot\sum_{i=1}^{n}p_{ij}=\sum_{i=1}^{n}x_i p_i+\sum_{j=1}^{m}y_j p'_j=M(X)+M(Y).$$

2. Доказательство равенства $M(X-Y)=M(X)-M(Y)$ предлагается провести самостоятельно.

Следствие. Математическое ожидание суммы (разности) нескольких случайных величин равно сумме (разности) математических ожиданий слагаемых.

С в о й с т в о 4. Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий:

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

Доказательство. Из определений математического ожидания и произведения случайных величин имеем

$$M(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_{ij}.$$

Так как по условию случайные величины X и Y независимы, то

$$p_{ij}=P(X=x_i)\cdot P(Y=y_j)=p_i p'_j.$$

Тогда $M(X\cdot Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_i p'_j=\sum_{i=1}^{n}x_i p_i\cdot\sum_{j=1}^{m}y_j p'_j=M(X)\cdot M(Y).$

Следствие. Математическое ожидание произведения нескольких взаимно независимых случайных величин равно произведению их математических ожиданий.

С в о й с т в о 5. Если все значения случайной величины X уменьшить (увеличить) на одно и то же число C, то ее математическое ожидание уменьшится (увеличится) на то же число C:

$$M(X\pm C)=M(X)\pm C.$$

Доказательство. Воспользуемся свойствами1 и 3.

Тогда $M(X\pm C)=M(X)\pm M(C)=M(X)\pm C.$

С в о й с т в о 6. Математическое ожидание отклонения случайной величины X от ее математического ожидания равно нулю:

$$M(X - M(X)) = 0.$$

Доказательство. Так как $M(X)$ – число постоянное, то $M(M(X))=M(X)$.

Тогда $M(X - M(X)) = M(X) - M(M(X)) = M(X) - M(X) = 0$.

Дисперсия дискретной случайной величины

Легко указать такие случайные величины, которые имеют одинаковые математические ожидания, но различные возможные значения. Рассмотрим, например, дискретные случайные величины X и Y, заданные следующими законами распределения:

X	-0,01	0,01
P	0,5	0,5

Y	-100	0,01
P	0,5	0,5

Найдем математические ожидания этих величин:

$$M(X) = -0{,}01 \cdot 0{,}5 + 0{,}01 \cdot 0{,}5 = 0,$$

$$M(Y) = -100 \cdot 0{,}5 + 100 \cdot 0{,}5 = 0.$$

Здесь математические ожидания обеих величин одинаковы, а возможные значения различны, причем X имеет возможные значения, близкие к математическому ожиданию, а Y – далекие от своего математического ожидания. Таким образом, зная лишь математическое ожидание случайной величины, еще нельзя судить ни о том, какие возможные значения она может принимать, ни о том, как они рассеяны вокруг математического ожидания. Другими словами, математическое ожидание полностью случайную величину не характеризует.

По этой причине наряду с математическим ожиданием вводят и другие числовые характеристики. Так, например, для того, чтобы оценить, как рассеяны возможные значения случайной величины вокруг ее математического ожидания, пользуются, в частности, числовой характеристикой, которую называют дисперсией.

На первый взгляд может показаться, что для оценки рассеяния проще всего вычислить все возможные значения отклонения случайной величины и затем найти их среднее значение. Однако такой путь ничего не даст, так как среднее значение отклонения, т.е. $M[X-M(X)]$, для любой случайной величины равно нулю. Это свойство уже было доказано выше и объясняется тем, что одни возможные отклонения положительны, а другие – отрицательны; в результате их взаимного погашения среднее значение отклонения равно нулю. Эти соображения говорят о целесообразности заменить возможные отклонения их абсолютными значениями или их квадратами. Правда, в случае, когда возможные отклонения заменяют их абсолютными значениями, приходится оперировать с абсолютными величинами, что приводит иногда к серьезным затруднениям. Поэтому чаще всего идут по другому пути, т.е. вычисляют среднее значение квадрата отклонения, которое и называют дисперсией.

Дисперсией (рассеянием) дискретной случайной величины называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

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

Пусть случайная величина задана законом распределения

X	x_1	x_2	…	x_n

P	p_1	p_2	…	p_n

Тогда квадрат отклонения имеет следующий закон распределения:

$[X-M(X)]^2$	$[x_1-M(X)]^2$	$[x_2-M(X)]^2$	…	$[x_n-M(X)]^2$
P	p_1	p_2	…	p_n

По определению математического ожидания

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

$$=[x_1-M(X)]^2 p_1+[x_2-M(X)]^2 p_2+\ldots+[x_n-M(X)]^2 p_n.$$

Таким образом, для того чтобы найти дисперсию, достаточно вычислить сумму произведений возможных значений квадрата отклонения на их вероятности. Для вычисления дисперсии часто бывает удобно пользоваться следующей теоремой.

Теорема. Дисперсия равна разности между математическим ожиданием квадрата случайной величины X и квадратом ее математического ожидания

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

Доказательство. Математическое ожидание $M(X)$ есть постоянная величина, следовательно, $2M(X)$ и $M^2(X)$ есть также постоянные величины. Приняв это во внимание и пользуясь свойствами математического ожидания (постоянный множитель можно вынести за знак математического ожидания, математическое ожидание суммы равно сумме математических ожиданий слагаемых), упростим формулу, выражающую определение дисперсии:

$$D(X)=M[X-M(X)]^2=M\left[X^2-2XM(X)+M^2(X)\right]=$$

$$=M(X^2)-2M(X)M(X)+M^2(X)=$$

$$=M(X^2)-2M^2(X)+M^2(X)=M(X^2)-M^2(X).$$

Итак,

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

Свойства дисперсии

Свойство 1. Дисперсия постоянной величины C равна нулю:

$$D(C)=0.$$

Доказательство. По определению дисперсии,

$$D(C)=M\left([C-M(C)]^2\right)$$

Пользуясь первым свойством математического ожидания, получим

$$D(C)=M\left[(C-C)^2\right]=M(0)=0.$$

Итак,

$$D(C)=0.$$

Свойство 2. Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

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

Доказательство. По определению дисперсии имеем

$$D(CX)=M\left([CX-M(CX)]^2\right).$$

Пользуясь вторым свойством математического ожидания (постоянный множитель можно вынести за знак математического ожидания), получим

$$D(CX)=M\left([CX-CM(X)]^2\right)=M\left(C^2[X-M(X)]^2\right)=C^2M\left([X-M(X)]^2\right)=C^2D(X).$$

Итак,

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

Свойство 3. Дисперсия суммы (разности) двух независимых случайных величин равна сумме дисперсии этих величин:

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

Доказательство. 1. Докажем равенство $D(X+Y)=D(X)+D(Y)$. По формуле для вычисления дисперсии имеем

$$D(X+Y)=M\left[(X+Y)^2\right]-[M(X+Y)]^2.$$

Раскрыв скобки и пользуясь свойствами математического ожидания суммы нескольких величин и произведения двух независимых случайных величин, получим

$$D(X+Y)=M\left[X^2+2XY+Y^2\right]-[M(X)+M(Y)]^2=$$

$$=M(X^2)+2M(X)\cdot M(Y)+M(Y^2)-M^2(X)-2M(X)\cdot M(Y)-M^2(Y)=$$

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

Итак,

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

2. Доказательство равенства $D(X-Y)=D(X)+D(Y)$ предлагается провести самостоятельно.

Следствие 1. Дисперсия суммы (разности) нескольких взаимно независимых случайных величин равна сумме дисперсий этих величин.

Следствие 2. Дисперсия суммы постоянной величины и случайной равна дисперсии случайной величины:

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

Доказательство. Величины C и X независимы, поэтому, по третьему свойству

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

В силу первого свойства $D(C)=0$. Следовательно,

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

С в о й с т в о 4. Дисперсия произведения двух независимых случайных величин X и Y вычисляется по формуле

$$D(XY)=D(X)D(Y)+M^2(X)D(Y)+M^2(Y)D(X).$$

Доказательство. Предлагается провести самостоятельно.

Среднее квадратическое отклонение

Для оценки рассеяния возможных значений случайной величины вокруг ее среднего значения, кроме дисперсии служат и некоторые другие характеристики. К их числу относится среднее квадратическое отклонение.

Среднем квадратическим отклонением случайной величины X называют квадратный корень из дисперсии

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

Легко показать, что дисперсия имеет размерность, равную квадрату размерности случайной величины. Так как среднее квадратическое отклонение равно квадратному корню из дисперсии, то размерность $\sigma(X)$ совпадает с размерностью X. Поэтому в тех случаях, когда желательно, чтобы оценка рассеяния имела размерность случайной величины, вычисляют среднее квадратическое отклонение, а не дисперсию. Например, если X выражается в линейных метрах, то $\sigma(X)$ будет выражаться также в линейных метрах, а $D(X)$ – в квадратных метрах.

Числовые характеристики основных дискретных случайных величин

В параграфе 2.1 нами были изучены пять основных законов распределения дискретных случайных величин: равномерный, биноминальный, закон распределения Пуассона, геометрический и гипергеометрический.

Приведем ряд теорем, позволяющих находить числовые характеристики случайных величин, распределенных по одному из этих законов.

Теорема. Математическое ожидание и дисперсия равномерно распределенной дискретной случайной величины вычисляются по формулам

$$M(X) = \frac{\sum\limits_{i=1}^{n} x_i}{n}; \quad D(X) = \frac{\sum\limits_{i=1}^{n} x_i^2}{n} - \left[\frac{\sum\limits_{i=1}^{n} x_i}{n}\right]^2.$$

Доказательство. Пусть случайная величина *X* распределения равномерно, тогда ее закон распределения имеет вид

X	x_1	x_2	…	x_n
P	$\frac{1}{n}$	$\frac{1}{n}$	…	$\frac{1}{n}$

Найдем числовые характеристики этой случайной величины.

$$M(X)=\frac{x_1}{n}+\frac{x_2}{n}+\ldots+\frac{x_n}{n}=\frac{\sum\limits_{i=1}^{n}x_i}{n};$$

$$D(X)=M(X^2)-\left[M(X)\right]^2=\frac{\sum\limits_{i=1}^{n}x_i{}^2}{n}-\left[\frac{\sum\limits_{i=1}^{n}x_i}{n}\right]^2.$$

Теорема. Математическое ожидание и дисперсия дискретной случайной величины, распределения по биномиальному закону вычисляется по формулам

$$M(X)=n\cdot p;\ D(X)=n\cdot p\cdot q.$$

Доказательство. Пусть случайная величина X_i- число появления события *A* в *i*-том испытании, $i=\left(\overline{1,n}\right)$. Тогда случайная величина $X=X_1+X_2+\ldots+X_n$ - число появлений события *A* во всех *n* испытаниях.

Согласно свойствам математического ожидания и дисперсии

$$M(X)=M(X_1+X_2+\ldots+X_n)=M(X_1)+M(X_2)+\ldots+M(X_n);$$
$$D(X)=D(X_1+X_2+\ldots+X_n)=D(X_1)+D(X_2)+\ldots+D(X_n).$$

Каждая из случайных величин X_i имеет один и тот же закон распределения

X_i	0	1
P_i	q	p

Тогда $M(X_i)=0\cdot q+1\cdot p=p$;

$$D(X_i) = M(X_i^2) - [M(X_i)]^2 = 0^2 \cdot q + 1^2 \cdot p - p^2 = p - p^2 = p(1-p) = pq.$$

Таким образом искомые числовые характеристики

$$M(X) = p + p + \ldots + p = np;\ D(X) = pq + pq + \ldots + pq = npq.$$

Теорема. Математическое ожидание и дисперсия дискретной случайной величины, распределенной по закону Пуассона вычисляются по формуле

$$M(X) = D(X) = \lambda\ .$$

Доказательство. Согласно определениям распределения Пуассона, и математического ожидания имеем, что

$$M(X) = \sum_{m=0}^{\infty} m\frac{\lambda^m}{m!}e^{-\lambda} = \sum_{m=0}^{\infty} \frac{\lambda^m}{(m-1)!}e^{-\lambda} =$$

$$= \lambda e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda.$$

Равенство $D(X) = \lambda$ предлагается доказать самостоятельно.

Замечание. Из курса высшей математики известно, что ряд вида

$$\sum_{m=0}^{\infty} \frac{x^m}{m!} = 1 + x + \frac{x^2}{2!} + \ldots = e^x$$

представляет собой сходящийся на всей числовой оси ряд Маклорена.

Теорема. Математическое ожидание и дисперсия дискретной случайной величины, распределенной по геометрическому закону, вычисляются по формулам

$$M(X) = \frac{1}{p};\ D(X) = \frac{q}{p^2}$$

Доказательство. По определению геометрического закона случайная величина имеет следующий закон распределения:

X	1	2	3	…	m
P	p	pq	pq^2	…	pq^{m-1}

Тогда

$$M(X)=\sum_{i=1}^{\infty}x_i p_i = p+2pq+3pq^2+\ldots+mpq^{m-1}=p\sum_{m=1}^{\infty}mq^{m-1};$$

$$D(X)=\sum_{i=1}^{\infty}x_i^2 p_i - M^2(X) = p+2^2pq+3^2pq^2+\ldots$$

$$\ldots+m^2pq^{m-1}-M^2(X)=p\sum_{m=1}^{\infty}m^2q^{m-1}-M^2(X).$$

Из теории рядов известно, что

$$\sum_{m=1}^{\infty} m\cdot q^{m-1}=\frac{1}{(1-q)^2};\quad \sum_{m=1}^{\infty} m^2\cdot q^{m-1}=\frac{1+q}{(1-q)^3}.$$

Тогда

$$M(X)=p\cdot\frac{1}{(1-q)^2}=\frac{1}{p};\ D(X)=p\cdot\frac{1+q}{p^3}-\frac{1}{p^2}=\frac{q}{p^2}.$$

Теорема. Если случайная величина X имеет гипергеометрический закон распределения, то

$$M(X)=r\cdot\frac{s}{n};\ D(X)=r\cdot\frac{s}{n-1}\left(1-\frac{s}{n}\right)\left(1-\frac{r}{n}\right).$$

Примем данную теорему без доказательства.

Решение типовых задач

Задача 1. Дискретная случайная величина задана законом распределения

X	2	4	5	7
P	0,2	0,1	0,3	0,4

Найти: математическое ожидание *M(X)*, дисперсию *D(X)* и среднее квадратическое отклонение $\sigma(X)$.

Решение. По формуле $M(X)=\sum_{i=1}^{n}x_i p_i$ находим математическое ожидание X:

$$M(X)=2\cdot 0{,}2+4\cdot 0{,}1+5\cdot 0{,}3+7\cdot 0{,}4=5{,}1.$$

По формулам $D(X)=M(X^2)-[M(X)]^2$ и $\sigma(X)=\sqrt{D(X)}$ найдем дисперсию и среднее квадратическое отклонение:

$$M(X^2)=\sum_{i=1}^{n} x_i^2 p_i = 2^2\cdot 0{,}2+4^2\cdot 0{,}1+5^2\cdot 0{,}3+7^2\cdot 0{,}4=29{,}5.$$

$$D(X)=29{,}5-(5{,}1)^2=3{,}49;\ \sigma(X)=\sqrt{3{,}49}=1{,}87.$$

X	1	3	4
P	0,1	?	0,6

Y	0	2	3
P	0,2	0,4	?

Задача 2. Независимые случайные величины X и Y заданы законами распределения:

а) найти $P(X=3), P(Y=3)$;

б) составить закон распределения случайной величины $Z=X+Y$. Найти *M(Z)*, *D(Z)* и проверить выполняемость свойств $M(X+Y)=M(X)+M(Y); D(X+Y)=D(X)+D(Y)$; в) составить закон распределения $V=X\cdot Y$. Найти *M(V)* и проверить выполняемость свойства $M(X\cdot Y)=M(X)\cdot M(Y)$.

Решение: а) Так как

$P(X=1)+P(X=3)+P(X=4)=1,\ P(Y=0)+P(Y=2)+P(Y=3)=1,$

то $P(X=3)=1-(0{,}1+0{,}6)=0{,}3,\ P(X=3)=1-(0{,}2+0{,}4)=0{,}4.$

Запишем закон распределения случайных величин X и Y с учетом их вероятности:

X	1	3	4
P	0,1	0,3	0,6

Y	0	2	3
P	0,2	0,4	0,4

б) Суммой случайных величин X и Y называется случайная величина $Z=X+Y$, возможные значения которой равны суммам каждого возможного значения величины X с каждым возможным значением величины Y. Если X и Y независимы, то вероятности возможных значений $Z=X+Y$ равны произведениям вероятностей слагаемых:

$Z=X+Y$	1+0=1	1+2=3	1+3=4	3+0=3

P	$0{,}1\cdot 0{,}2=0{,}02$	$0{,}1\cdot 0{,}4=0{,}04$	$0{,}1\cdot 0{,}4=0{,}04$	$0{,}3\cdot 0{,}2=0{,}06$

3+2=5	3+3=6	4+0=4	4+2=6	4+3=7
$0{,}3\cdot 0{,}4=0{,}12$	$0{,}3\cdot 0{,}4=0{,}12$	$0{,}6\cdot 0{,}2=0{,}12$	$0{,}6\cdot 0{,}4=0{,}24$	$0{,}6\cdot 0{,}4=0{,}24$

Одинаковые значения величины *Z* объединяем, складывая их вероятности. Закон распределения случайной величины *Z* будет иметь вид:

$Z=X+Y$	1	3	4	5	6	7
P	0,02	0,1	0,16	0,12	0,36	0,24

$M(Z)=1\cdot 0{,}02+3\cdot 0{,}1+4\cdot 0{,}16+5\cdot 0{,}12+6\cdot 0{,}36+7\cdot 0{,}24=5{,}4;$

$M(X)=1\cdot 0{,}1+3\cdot 0{,}3+4\cdot 0{,}6=3{,}4;$ $M(Y)=0\cdot 0{,}2+2\cdot 0{,}4+3\cdot 0{,}4=2;$

$M(X)+M(Y)=3{,}4+2=5{,}4.$

Итак, $M(X+Y)=M(X)+M(Y);$

$M(Z^2)=1\cdot 0{,}02+9\cdot 0{,}1+16\cdot 0{,}16+25\cdot 0{,}12+36\cdot 0{,}36+43\cdot 0{,}24=31{,}2;$

$D(Z)=M(Z^2)-[M(Z)]^2=31{,}2-(5{,}4)^2=2{,}04;$

$M(X^2)=1\cdot 0{,}1+9\cdot 0{,}3+16\cdot 0{,}6=12{,}4;$ $\mathrm{D(X)}=12{,}4-(3{,}4)^2=0{,}84;$

$M(Y^2)=0\cdot 0{,}2+4\cdot 0{,}4+9\cdot 0{,}4=5{,}2;$ $D(Y)=5{,}2-2^2=1{,}2;$

$D(X)+D(Y)=0{,}84+1{,}2=2{,}04.$

Итак $: D(X+Y)=D(X)+D(Y).$

в) Составим закон распределения $V=X\cdot Y$. Произведением случайных величин *X* и *Y* называется случайная величина $V=X\cdot Y$, возможные значения которой равны произведениям каждого возможного значения *X* на каждое возможное значение *Y*. Если *X* и *Y* независимы, то вероятности возможных значений $V=X\cdot Y$ равны произведениям вероятностей сомножителей:

$V=X\cdot Y$	$1\cdot 0=0$	$1\cdot 2=2$	$1\cdot 3=3$	$3\cdot 0=0$
P	0,02	0,04	0,04	0,06

$3\cdot 2=6$	$3\cdot 3=9$	$4\cdot 0=0$	$4\cdot 2=8$	$4\cdot 3=12$
0,12	0,12	0,12	0,24	0,24

Одинаковые значения величины $V = X \cdot Y$ объединяем, складывая их вероятности.

Закон распределения $V = X \cdot Y$ записываем так:

$V = X \cdot Y$	0	2	3	6	8	9	12
P	0,2	0,04	0,04	0,12	0,24	0,12	0,24

Найдем

$M(V) = 0 \cdot 0{,}2 + 2 \cdot 0{,}04 + 3 \cdot 0{,}04 + 6 \cdot 0{,}12 + 8 \cdot 0{,}24 + 9 \cdot 0{,}12 + 12 \cdot 0{,}24 = 6{,}8.$

$M(X) \cdot M(Y) = 3{,}4 \cdot 2 = 6{,}8.$

Таким образом, $M(X \cdot Y) = M(X) \cdot M(Y).$

Задачи (41 – 50)

41. Даны законы распределения двух независимых случайных величин X и Y.

1. Составить закон распределения случайной величины Z.

X	3	6	9
P	0,6	0,3	0,1

Y	5	15	25
P	0,9	0,05	0,05

2. Найти числовые характеристики случайной величины Z.

$Z = \frac{1}{3}X - \frac{1}{5}Y$.

42. Даны законы распределения двух независимых случайных величин X и Y .

1. Составить закон распределения случайной величины Z.
2. Найти числовые характеристики случайной величины Z.

X	10	20	30
P	0,1	0,5	0,4

Y	20	25	30
P	0,5	0,4	0,1

$Z = 5X - 4Y$.

43. Даны законы распределения двух независимых случайных величин X и Y.

1. Составить закон распределения случайной величины Z.
2. Найти числовые характеристики случайной величины Z.

X	0	1	2
P	0,1	0,3	0,6

Y	1	2	3
P	0,8	0,1	0,1

$Z = (4X) \cdot Y$.

44. Даны законы распределения двух независимых случайных величин X и Y.

1. Составить закон распределения случайной величины Z.

X	2	5	8
P	0,7	0,1	0,2

Y	2	4	6
P	0,35	0,4	0,25

2. Найти числовые характеристики случайной величины Z.

$Z = X + \frac{1}{2}Y$.

45. Даны законы распределения двух независимых случайных величин X и Y.

1. Составить закон распределения случайной величины Z.
2. Найти числовые характеристики случайной величины Z.

X	–2	0	1
P	0,3	0,2	0,5

Y	–1	1	2
P	0,1	0,7	0,2

$Z = 3X - Y$.

46. Даны законы распределения двух независимых случайных величин X и Y.

1. Составить закон распределения случайной величины Z.
2. Найти числовые характеристики случайной величины Z.

X	2	4	6
P	0,6	0,2	0,2

Y	–1	0	2
P	0,15	0,25	0,6

$Z = \frac{1}{2}X + 2Y$.

47. Даны законы распределения двух независимых случайных величин X и Y.

1. Составить закон распределения случайной величины Z.
2. Найти числовые характеристики случайной величины Z.

X	1	3	5
P	0,3	0,5	0,2

Y	7	8	9
P	0,4	0,3	0,3

$Z = X \cdot Y$.

48. Даны законы распределения двух независимых случайных величин X и Y.

1. Составить закон распределения случайной величины Z.
2. Найти числовые характеристики случайной величины

X	1	2	4
P	0,1	0,6	0,3

Y	0	3	4
P	0,2	0,5	0,3

Z.

$Z = 2X + Y$.

49. Даны законы распределения двух независимых случайных величин X и Y.

1. Составить закон распределения случайной величины Z.
2. Найти числовые характеристики случайной величины Z.

X	1	3	4	6
P	0,1	0,2	0,2	0,5

Y	1	2	5
P	0,15	0,55	0,3

$Z = X^2 + Y$.

50. Даны законы распределения двух независимых случайных величин X и Y.

1. Составить закон распределения случайной величины Z.
2. Найти числовые характеристики случайной величины Z.

X	–1	0	2
P	0,6	0,3	0,1

Y	2	6	10
P	0,5	0,4	0,1

$Z = X \cdot Y$.

2.3. Плотность распределения вероятностей и числовые
характеристики непрерывных случайных величин

Плотность распределения вероятностей, ее свойства

Непрерывную случайную величину наряду с функцией распределения можно также задать, используя другую функцию, которую называют плотностью распределения или плотностью вероятности (иногда ее называют дифференциальной функцией).

Плотностью распределения вероятностей непрерывной случайной величины *X* называют функцию $f(x)$ - первую производную от функции распределения $F(x)$:

$$f(x)=F'(x).$$

Из этого определения следует, что функция распределения является первообразной для плотности распределения.

Пусть $F(x)$ - функция распределения непрерывной случайной величины *X*. По определению плотности распределения $f(x)=F'(x)$, или в иной форме:

$$f(x)=\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}.$$

Как известно, разность $F(x+\Delta x)-F(x)$ определяет вероятность того, что *X* примет значение, принадлежащее интервалу $(x; x+\Delta x)$. Таким образом, предел отношения вероятности того, что непрерывная случайная величина примет значение, принадлежащее интервалу $(x; x+\Delta x)$, к длине этого интервала $(\text{при}\,\Delta x\to 0)$ равен значению плотности распределения в точке *x*.

Замечание. Для описания распределения вероятностей дискретной случайной величины плотность распределения неприменима.

Свойства плотности распределения

С в о й с т в о 1. Плотность распределения – неотрицательная функция:

$$f(x) \geq 0.$$

Доказательство. Функция распределения – неубывающая функция, следовательно, ее производная $F'(x) = f(x)$ - функция неотрицательная.

Геометрически это свойство означает, что точки, принадлежащие графику плотности распределения, расположены либо над осью *Ox*, либо на этой оси.

График плотности распределения называют ***кривой распределения.***

С в о й с т в о 2. Вероятность того, что непрерывная случайная величина *X* примет значение, принадлежащее интервалу (*a*,*b*), равна определенному интегралу от плотности распределения, взятому в пределах от *a* до *b*:

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

Доказательство. Из свойств функции распределения известно, что

$$P(a < X < b) = F(b) - F(a).$$

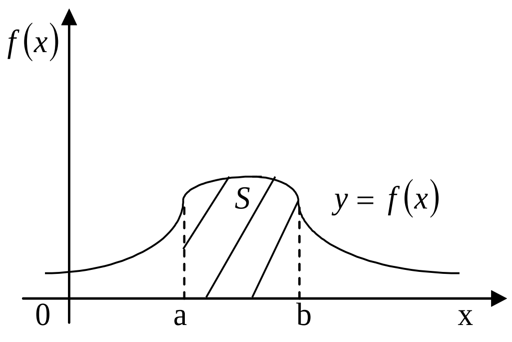
По формуле Ньютона – Лейбница,

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

Таким образом,

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

Геометрически полученный результат можно истолковать так: вероятность того, что непрерывная случайная величина примет значение, принадлежащее интервалу



(a,b), равна площади криволинейной трапеции, ограниченной осью Ox, кривой распределения $f(x)$ и прямыми $x=a$ и $x=b$ (рис.2.4).

Рис. 2.4.

Пример. Задана плотность вероятности случайной величины X

$$f(x)=\begin{cases}0, & \text{при } x\le 0,\\ 2x, & \text{при } 0<x\le 1,\\ 0, & \text{при } x>1.\end{cases}$$

Найти вероятность того, что в результате испытания X примет значение, принадлежащее интервалу (0,5; 1).

Решение. Искомая вероятность

$$P(0{,}5<X<1)=2\int\limits_{0{,}5}^{1} x\,dx = x^2\Big|_{0{,}5}^{1}=1-0{,}25=0{,}75.$$

С в о й с т в о 3. Несобственный интеграл от плотности распределения в пределах от $-\infty$ до ∞ равен единице:

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

Доказательство. Несобственный интеграл $\int\limits_{-\infty}^{\infty} f(x)dx$ выражает вероятность события, состоящего в том, что случайная величина примет значение, принадлежащее интервалу $(-\infty;\infty)$. Очевидно, такое событие достоверно, следовательно, вероятность его равна единице.

Геометрически это означает, что вся площадь криволинейной трапеции, ограниченной осью Ox и кривой распределения, равна единице.

В частности, если все возможные значения случайной величины принадлежат интервалу (α,β), то

$$\int\limits_{\alpha}^{\beta} f(x)dx=1.$$

Теорема. Функция распределения непрерывной случайной величины связана с плотностью распределения следующим равенством:

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

Доказательство. Действительно мы обозначили через $F(x)$ вероятность того, что случайная величина примет значение, меньшее x, т.е.

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

Очевидно, неравенство $X<x$ можно записать в виде двойного неравенства $-\infty<X<x$, следовательно,

$$F(x)=P(-\infty<X<x).$$

Полагая в данной формуле $a=-\infty,\ b=x$, имеем согласно свойству 2

$$P(-\infty<X<x)=\int\limits_{-\infty}^{x} f(x)dx.$$

Наконец, заменив $P(-\infty<X<x)$ на $F(x)$, окончательно получим

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

Таким образом, зная плотность распределения, можно найти функцию распределения. Разумеется, по известной функции распределения может быть найдена плотность распределения, а именно:

$$f(x)=F'(x).$$

Числовые характеристики непрерывных случайных величин

Распространим определения числовых характеристик дискретных величин на величины непрерывные. Начнем с математического ожидания.

Пусть непрерывная случайная величина X задана плотностью распределения $f(x)$. Допустим, что все возможные значения X принадлежат отрезку $[a,b]$. Разобьем этот отрезок на n частичных отрезков длиной $\Delta x_1, \Delta x_2, \dots, \Delta x_n$ и выберем в каждом из них произвольную точку x_i $(i = 1, 2, \dots, n)$. Нам надо определить математическое ожидание непрерывной величины по аналогии с дискретной; составим сумму произведения возможных значений x_i на вероятности попадания их в интервал Δx_i (напомним, что произведение $f(x)\Delta x$ приближенно равно вероятности попадания X в интервал Δx):

$$\sum x_i f(x_i)\Delta x_i .$$

Перейдя к пределу при стремлении к нулю длины наибольшего из частных отрезков, получим определенный интеграл $\int_a^b xf(x)dx$.

Математическим ожиданием непрерывной случайной величины X, возможные значения которой принадлежат отрезку $[a,b]$, называют определенный интеграл

$$M(X) = \int_a^b xf(x)dx .$$

Если возможные значения принадлежат всей оси Ox, то

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx .$$

По аналогии с дисперсией дискретной величины определяется и дисперсия непрерывной величины.

Дисперсией непрерывной случайной величины называют математическое ожидание квадрата ее отклонения.

Если возможные значения X принадлежат отрезку $[a,b]$, то

$$D(X) = \int_a^b [x - M(X)]^2 f(x)dx ;$$

если возможные значения принадлежат всей оси x, то

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

Среднее квадратическое отклонение непрерывной случайной величины определяется, как и для величины дискретной, равенством

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

Замечание 1. Можно доказать, что свойства математического ожидания и дисперсии дискретных величин сохраняются и для непрерывных величин.

Замечание 2. Легко получить для вычисления дисперсии более удобные формулы:

$$D(X) = \int_a^b x^2 f(x) dx - [M(X)]^2; \; D(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - [M(X)]^2.$$

Решение типовых задач

Задача 1. Непрерывная случайная величина задана функцией распределения

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

Требуется:

а) найти функцию плотности распределения $f(x)$;

б) найти математическое ожидание $M(X)$, дисперсию $D(X)$ и среднее квадратическое отклонение $\sigma(X)$;

в) построить графики функций $f(x)$ и $F(x)$;

г) найти $P(-1 < X < 1)$.

Решение:

а) по определению функции плотности вероятности $f(x) = F'(x)$

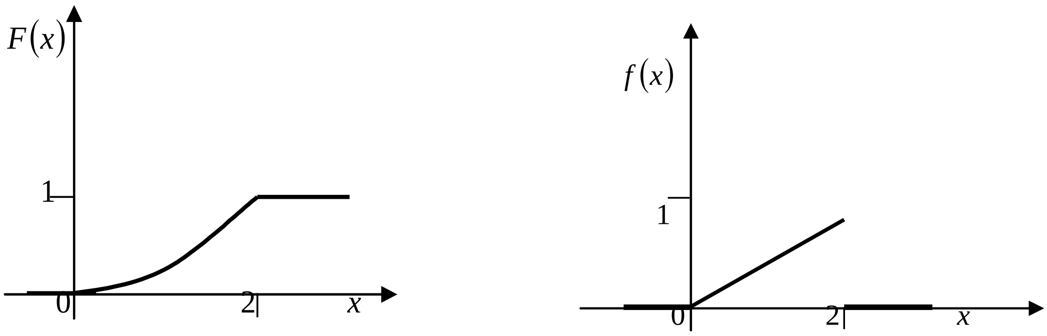
$$f(x)=\begin{cases}0, & x\le 0,\\ \dfrac{x}{2}, & 0<x\le 2,\\ 0, & x>2.\end{cases}$$

б) Для непрерывной случайной величины

$$M(X)=\int\limits_a^b xf(x)dx=\int\limits_0^2\frac{x}{2}\cdot xdx=\frac{1}{6}x^3\Big|_0^2=\frac{4}{3};$$

$$M(X^2)=\int\limits_a^b x^2f(x)dx=\int\limits_0^2 x^2\frac{x}{2}\cdot dx=\frac{1}{8}x^4\Big|_0^2=2;$$

$$D(X)=M(X^2)-[M(X)]^2=2-\left(\frac{4}{3}\right)^2=\frac{2}{9};\ \ \sigma(X)=\sqrt{D(X)}=\sqrt{\frac{2}{9}}\approx 0{,}47.$$



в)

г) для вычисления вероятности попадания непрерывной случайной величины в интервал (α,β) можно применить одну из формул:

$$P(\alpha<X<\beta)=F(\beta)-F(\alpha)\ \text{ или }\ P(\alpha<X<\beta)=\int\limits_\alpha^\beta f(x)dx.$$

Применим первую формулу

$$P(-1<X<1)=F(1)-F(-1)=\frac{1^2}{4}-0=\frac{1}{4}.$$

Задача 2. Случайная величина задана плотностью распределения:

$$f(x)=\begin{cases}0, & x\le 1,\\ \dfrac{c}{8}, & 1<x\le 5,\\ 0, & x>5.\end{cases}$$

Требуется:
а) найти коэффициент *C*;
б) функцию распределения $F(x)$;
в) построить графики функций $F(x)$ и $f(x)$.

Решение:

а) Плотность распределения $f(x)$ должна удовлетворять условиям:

$f(x)\geq 0$; $\int\limits_{-\infty}^{+\infty} f(x)dx=1$, тогда

$$\int\limits_{-\infty}^{\infty} f(x)dx=\int\limits_{-\infty}^{1}0dx+\int\limits_{1}^{5}\frac{c}{8}dx+\int\limits_{5}^{+\infty}0dx=\frac{c}{8}\int\limits_{1}^{5}dx=\frac{c}{8}x\Big|_{1}^{5}=\frac{c}{8}(5-1)=\frac{4c}{8}=\frac{1}{2}C.$$

Так как $\int\limits_{-\infty}^{+\infty} f(x)dx=1$, то $\frac{1}{2}C=1\Rightarrow C=2.$

Таким образом,

$$f(x)=\begin{cases}0, & x\leq 1,\\ \frac{1}{4}, & 1<x\leq 5,\\ 0, & x>5.\end{cases}$$

б) для нахождения функции распределения $F(x)$ воспользуемся формулой

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

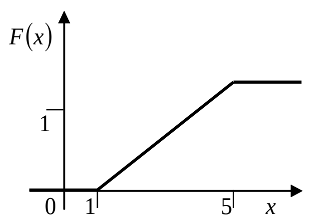
При $x\leq 1$, $f(x)=0 \Rightarrow F(x)=\int\limits_{-\infty}^{x}0dx=0$.

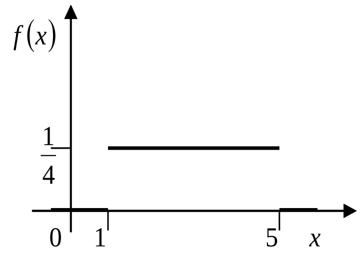
При $1<x\leq 5$, $F(x)=\int\limits_{-\infty}^{1}0dx+\int\limits_{1}^{x}\frac{1}{4}dx=\frac{x}{4}\Big|_{1}^{x}=\frac{1}{4}(x-1).$

При $x>1$, $F(x)=\int\limits_{-\infty}^{1}0dx+\int\limits_{1}^{5}\frac{1}{4}dx+\int\limits_{5}^{x}0dx=1.$

Итак,

$$F(x)=\begin{cases}0, & x\leq 1,\\ \frac{x-1}{4}, & 1<x\leq 5,\\ 1, & x>5.\end{cases}$$





в)

Задачи (51 – 60)

51. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.
2. Найти $M(X)$.
3. Найти вероятность $P(\alpha < X < \beta)$.
4. Построить графики $f(x)$ и $F(x)$.

$$F(x) = \begin{cases} 0, & x \le 1, \\ \dfrac{x-1}{2}, & 1 < x \le 3, \\ 1, & x > 3. \end{cases}$$

$\alpha = 1, \qquad \beta = 2.$

52. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.
2. Найти $M(X)$.
3. Найти вероятность $P(\alpha < X < \beta)$.
4. Построить графики $f(x)$ и $F(x)$.

$$F(x)=\begin{cases}0, & x<2,\\ (x-2)^3, & 2<x\le 3,\\ 1, & x>3.\end{cases}$$

$\alpha=2,\qquad \beta=2{,}5$.

53. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.
2. Найти $M(X)$.
3. Найти вероятность $P(\alpha<X<\beta)$.
4. Построить графики $f(x)$ и $F(x)$.

$$F(x)=\begin{cases}0, & x\le -1,\\ \frac{1}{3}x+\frac{1}{3}, & -1<x\le 2,\\ 1, & x>2.\end{cases}$$

$\alpha=-0{,}5,\qquad \beta=1{,}5.$

54. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.
2. Найти $M(X)$.
3. Найти вероятность $P(\alpha<X<\beta)$.
4. Построить графики $f(x)$ и $F(x)$.

$$F(x)=\begin{cases}0, & x<3,\\ \frac{(x-3)^2}{4}, & 3\le x\le 5,\\ 1, & x>5.\end{cases}$$

$\alpha=3,\qquad \beta=4.$

55. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.

2. Найти $M(X)$.

3. Найти вероятность $P(\alpha < X < \beta)$.

4. Построить графики $f(x)$ и $F(x)$.

$$F(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x^3}{27}, & 0 \le x \le 3, \\ 1, & x > 3. \end{cases}$$

$\alpha = 1, \quad \beta = 2.$

56. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.

2. Найти $M(X)$.

3. Найти вероятность $P(\alpha < X < \beta)$.

4. Построить графики $f(x)$ и $F(x)$.

$$F(x) = \begin{cases} 0, & x \le 1, \\ \dfrac{1}{2}\left(x^2 - x\right), & 1 < x \le 2, \\ 1, & x > 2. \end{cases}$$

$\alpha = 0{,}5, \quad \beta = 1{,}5.$

57. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.

2. Найти $M(X)$.

3. Найти вероятность $P(\alpha < X < \beta)$.

4. Построить графики $f(x)$ и $F(x)$.

$$F(x) = \begin{cases} 0, & x \le 2, \\ \sqrt{x-2}, & 2 < x \le 3, \\ 1, & x > 3. \end{cases}$$

$\alpha = 2{,}25, \quad \beta = 4.$

58. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.
2. Найти $M(X)$.
3. Найти вероятность $P(\alpha < X < \beta)$.
4. Построить графики $f(x)$ и $F(x)$.

$$F(x)=\begin{cases}0, & x<0,\\ x^3, & 0\le x\le 1,\\ 1, & x>1.\end{cases}$$

$\alpha=-2, \qquad \beta=\frac{1}{2}.$

59. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.
2. Найти $M(X)$.
3. Найти вероятность $P(\alpha < X < \beta)$.
4. Построить графики $f(x)$ и $F(x)$.

$$F(x)=\begin{cases}0, & x\le 0,\\ \dfrac{x^2}{4}, & 0<x<2,\\ 1, & x\ge 2.\end{cases}$$

$\alpha=1, \qquad \beta=6.$

60. Случайная величина X задана функцией распределения вероятностей $F(x)$.

Требуется:

1. Найти функцию плотности распределения $f(x)$.
2. Найти $M(X)$.
3. Найти вероятность $P(\alpha < X < \beta)$.
4. Построить графики $f(x)$ и $F(x)$.

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

$\alpha = 2, \qquad \beta = 4{,}5.$

2. 4. Основные законы распределения непрерывных случайных величин

Равномерное распределение

Непрерывная случайная величина X называется ***равномерно распределенной*** на отрезке $[a; b]$, если плотность вероятности постоянна на этом отрезке и равна нулю вне этого отрезка

$$f(x) = \begin{cases} 0, & \text{при } x < a, \\ c, & \text{при } a \le x \le b, \\ 0, & \text{при } x > b. \end{cases}$$

Из свойств функции плотности известно, что $\int\limits_{\alpha}^{\beta} f(x)dx = 1$.

Тогда $\int\limits_{a}^{b} c dx = 1$.

Следовательно, $c(b-a) = 1$ и $c = \dfrac{1}{b-a}$.

Таким образом, плотность $f(x)$ равномерного распределения имеет вид

$$f(x) = \begin{cases} 0, & \text{при } x < a, \\ \dfrac{1}{b-a}, & \text{при } a \le x \le b, \\ 0, & \text{при } x > b. \end{cases}$$

Восстановим функцию распределения $F(x)$ равномерно распределенной случайной величины. Для этого воспользуемся формулой

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

Тогда:

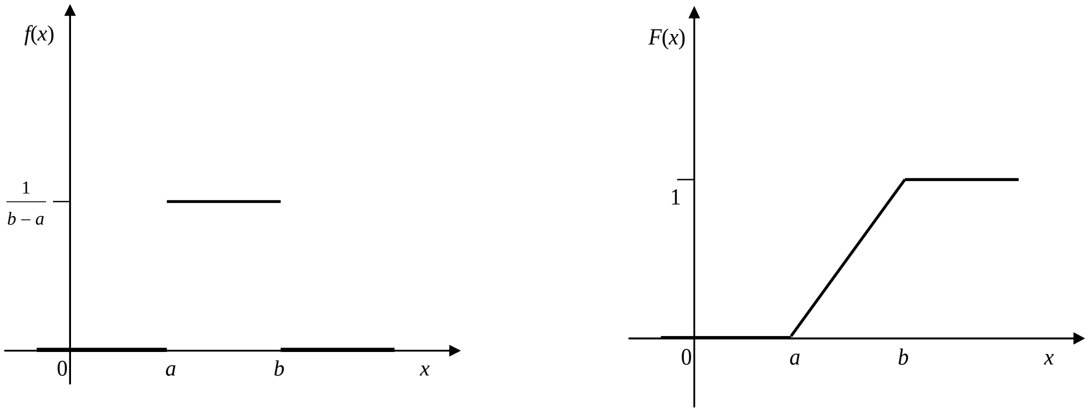
– при $x < a$, $F(x) = \int\limits_{-\infty}^{x} 0dx = 0$;

– при $a \le x \le b$, $F(x)=\int\limits_{-\infty}^{a} 0dx+\int\limits_{a}^{x}\frac{1}{b-a}dx=\frac{x-a}{b-a}$;

– при $x > b$, $F(x)=\int\limits_{-\infty}^{a} 0dx+\int\limits_{a}^{b}\frac{1}{b-a}dx+\int\limits_{b}^{x} 0dx=\frac{b-a}{b-a}=1$.

Таким образом, функция распределения примет вид:

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



Изобразим графики обеих функций (рис.2.5).

Рис. 2.5.

Теорема. Для вычисления *M(X)* и *D(X)* равномерно распределенной случайной величины имеют место равенства

$$M(X)=\frac{a+b}{2};\ D(X)=\frac{(b-a)^2}{12}.$$

Доказательство. Согласно определению $M(X)=\int\limits_{-\infty}^{\infty} xf(x)dx$. Тогда

$$M(X)=\int\limits_{a}^{b} x\cdot\frac{1}{b-a}dx=\left.\frac{x^2}{2(b-a)}\right|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{a+b}{2}.$$

Согласно определению $D(X)=M(X^2)-[M(X)]^2$. Тогда

$$D(X)=\int_a^b x^2\cdot\frac{1}{b-a}dx-\frac{(a+b)^2}{4}=\frac{x^3}{3(b-a)}\bigg|_a^b-\frac{(a+b)^2}{4}=$$

$$=\frac{b^3-a^3}{3(b-a)}-\frac{(a+b)^2}{4}=\frac{b^2+ab+a^2}{3}-\frac{(a+b)^2}{4}=\frac{b^2-2ab+a^2}{12}=\frac{(b-a)^2}{12}.$$

Очевидно, что $\sigma(X)=\frac{b-a}{2\sqrt{3}}$.

Теорема. Пусть непрерывная случайная величина распределена равномерно на отрезке $[a;b]$, и отрезок $[c;d]\subset[a;b]$, тогда

$$P(c<X<d)=\frac{d-c}{b-a}.$$

Доказательство. Согласно свойствам функции распределения

$$P(c<X<d)=F(d)-F(c)=\frac{d-a}{b-a}-\frac{c-a}{b-a}=\frac{d-c}{b-a}.$$

Таким образом, для того, чтобы полностью описать непрерывную случайную величину, имеющую равномерное распределение, достаточно знать концы отрезка, которому принадлежат все возможные значения этой случайной величины.

Показательное распределение

Показательным (экспоненциальным) называют распределение вероятностей случайной величины X, которое описывается плотностью

$$f(x)=\begin{cases}0 & \text{при } x<0;\\ \lambda e^{-\lambda x} & \text{при } x\geq 0,\end{cases}$$

где λ - постоянная положительная величина.

Показательное распределение определяется одним параметром λ. Эта особенность показательного распределения указывает на его преимущество по сравнению с распределениями, зависящими от большего числа параметров. Обычно параметры неизвестны и приходится находить их оценки (приближенные значения); разумеется, проще оценить один параметр, чем два или три и т.д.

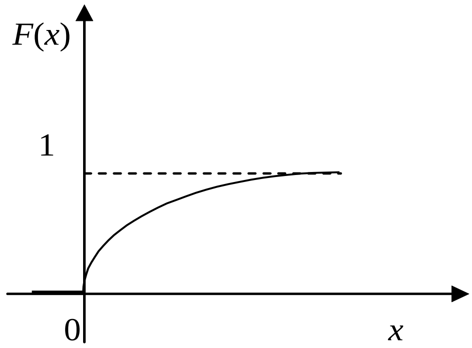
Найдем функцию распределения показательного закона. Воспользуемся формулой $F(x)=\int\limits_{-\infty}^{x} f(x)dx$. Тогда:

- при $x<0,\ F(x)=\int\limits_{-\infty}^{x} 0dx=0$;
- при $x\geq 0,\ F(x)=\int\limits_{-\infty}^{0} 0dx+\int\limits_{0}^{x} \lambda e^{-\lambda x}dx=1-e^{-\lambda x}$.

Итак,

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

Графики плотности и функции распределения показа-



тельного закона изображены на рис. 2.6.

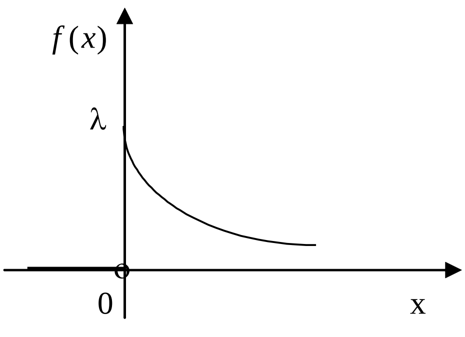


Рис. 2.6.

Теорема. Для *M(X)* и *D(X)* показательного распределения справедливы равенства

$$M(X)=\frac{1}{\lambda};\ D(X)=\frac{1}{\lambda^2}.$$

Доказательство. Согласно определению $M(X)=\int\limits_{-\infty}^{\infty}xf(x)dx$.

Тогда,

$$M(X)=\int\limits_0^{\infty}x\cdot\lambda e^{-\lambda x}dx=\lambda\int\limits_0^{\infty}x\cdot e^{-\lambda x}dx=\begin{vmatrix}u=x;\ du=dx\\ dv=e^{-\lambda x};\ V=-\frac{1}{\lambda}e^{-\lambda e}\end{vmatrix}=$$

$$=\lambda\left(-\frac{x}{\lambda}e^{-\lambda x}\Big|_0^{\infty}+\frac{1}{\lambda}\int\limits_0^{\infty}e^{-\lambda x}dx\right)=\lambda\left(-\frac{1}{\lambda^2}\cdot e^{-\lambda x}\right)\Bigg|_0^{\infty}=-\frac{1}{\lambda}(0-1)=\frac{1}{\lambda}.$$

Соотношение $D(X)=\frac{1}{\lambda^2}$ предлагается доказать самостоятельно.

Очевидно, что $\sigma(X)=\frac{1}{\lambda}$, т.е. математическое ожидание и среднее квадратическое отклонение показательного распределения равны между собой.

Теорема. Вероятность попадания в интервал $(a;b)$ показательно распределенной случайной величины вычисляется по формуле

$$P(a<X<b)=e^{-\lambda a}-e^{-\lambda b}.$$

Доказательство. Используем формулу

$$P(a<X<b)=F(b)-F(a).$$

Учитывая, что $F(a)=1-e^{-\lambda a}$, $F(b)=1-e^{-\lambda b}$, получим

$$P(a<X<b)=e^{-\lambda a}-e^{-\lambda b}.$$

Нормальное распределение

Нормальным называют распределение вероятностей непрерывной случайной величины, которое описывается плотностью

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

Данное распределение определяется двумя параметрами: a и σ, достаточно знать эти параметры, чтобы задать нормальное распределение.

График плотности нормального распределения называют ***нормальной кривой (кривой Гаусса)***.

Исследуем и построим график функции

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

1. Очевидно, функция определена на всей оси абсцисс.
2. При всех значениях x функция принимает положительные значения, т.е. нор- мальная кривая расположена над осью Ox.
3. Предел функции при неограниченном возрастании x (по абсолютной величине) равен нулю: $\lim\limits_{|x|\to\infty} f(x)=0$, т.е. ось Ox служит горизонтальной асимптотой графика.
4. Исследуем функцию на экстремум. Найдем первую производную:

$$f'(x)=-\frac{x-a}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Очевидно, что:

– при $x=a,\ f'(x)=0;$

– при $x<a,\ f'(x)>0;$

– при $x>a,\ f'(x)<0.$

Следовательно:

– при $x\in(-\infty;a),\ f(x)$ - возрастает;

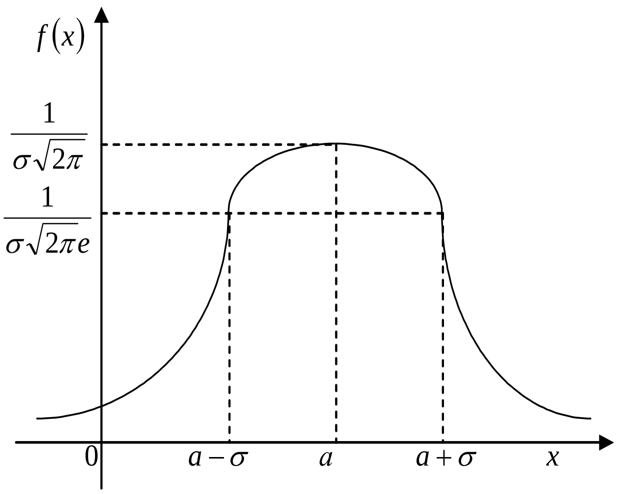
– при $x \in (a;\infty)$, $f(x)$ - убывает;

– при $(x = a)$ функция имеет максимум $f(a) = \frac{1}{\sigma\sqrt{2\pi}}.$

5. Разность $(x - a)$ содержится в аналитическом выражении функции в квадрате, т.е. график функции симметричен относительно прямой $x = a.$
6. Исследуем функцию на точки перегиба. Найдем вторую производную:

$$y'' = -\frac{1}{\sigma^3\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \left[1 - \frac{(x-a)^2}{\sigma^2}\right].$$

Легко видеть, что при $x = a + \sigma$ и $x = a - \sigma$ вторая производная равна нулю, а при переходе через эти точки она меняет знак (в обеих этих точках значение функции равно $\frac{1}{\sigma\sqrt{2\pi e}}$).



Таким образом, точки графика $\left(a - \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$ и $\left(a + \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$ являются точками перегиба. На рис. 2.7 изображена нормальная кривая

Рис. 2.7.

Влияние параметров нормального распределения на форму нормальной

кривой

Выясним, как влияют на форму и расположение нормальной кривой значения параметров a и σ.

Известно, что графики функций $f(x)$ и $f(x-a)$ имеют одинаковую форму; сдвинув график $f(x)$ в положительном направлении оси x на a единиц масштаба при $a>0$ или в отрицательном направлении при $a<0$, получим график $f(x-a)$. Отсюда следует, что изменение величины параметра a не изменяет формы нормальной кривой, а приводит лишь к ее сдвигу вдоль оси Ox: вправо, если a возрастает, и влево, если a убывает.

Рассмотрим форму кривой при изменении параметра σ. Как было указано выше, максимум нормальной кривой равен $\frac{1}{\sigma\sqrt{2\pi}}$. Отсюда следует, что с возрастанием σ максимальная ордината нормальной кривой убывает, а сама кривая становится более пологой, т.е. сжимается к оси Ox; при убывании σ нормальная кривая становится более «островершинной» и растягивается в положительном направлении оси Oy.

Заметим, что при любых значениях параметров a и σ площадь, ограниченная нормальной кривой и осью x, остается равной единице (третье свойство плотности распределения).

На рис. 2.8 изображены нормальные кривые при различных значениях σ и $a=0$. Чертеж наглядно иллюстрирует, как изменение параметра σ сказывается на форме нормальной кривой.

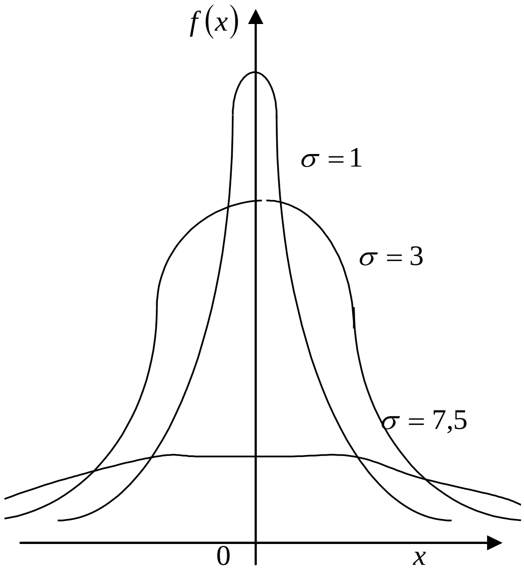


Рис. 2.8.

Замечание 1. ***Общим*** называют нормальное распределение с произвольными параметрами a и $\sigma(\sigma > 0)$.

Нормированным (стандартным) называют нормальное распределение с параметрами $a = 0$ и $\sigma = 1$.

Плотность нормированного распределения

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

Эта функция табулирована (приложение 1).

Замечание 2. Функция *F(x)* общего нормального распределения определяется равенством

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt,$$

а функция нормированного распределения

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

Легко проверить, что $F(x) = F_0\left(\frac{x-a}{\sigma}\right)$.

Замечание 3. Существует связь между функциями общего и нормированного распределения *F(x), F₀(x)* и функцией Лапласа *Ф(x)*:

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

$$F_0(x) = 0{,}5 + \Phi(x).$$

Данные равенства позволяют вычислять значения $F(x)$ и $F_0(x)$ по таблице значений *Ф(x)* (приложение 3).

Вероятностный смысл параметров *a* и σ

Теорема. Для нормально распределенной случайной величины справедливы равенства:

$$M(X) = a;\ D(X) = \sigma^2$$

Доказательство. 1. По определению математического ожидания непрерывной случайной величины,

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{(x-a)^2}{2\sigma^2}} dx = \begin{vmatrix} x = \sigma z + a & \\ z = \dfrac{x-a}{\sigma} & dx = \sigma dz \end{vmatrix} =$$

$$= \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + a)e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma z e^{-\frac{z^2}{2}} dz + \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz.$$

Первое из слагаемых равно нулю (под знаком интеграла нечетная функция; пределы интегрирования симметричны относительно начала координат). Второе из слагаемых равно a $\left(\text{интеграл Пуассона } \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi}\right)$.

Итак, $M(X) = a,$ т.е. математическое ожидание нормального распределения равно параметру *a*.

2. По определению дисперсии непрерывной случайной величины, учитывая, что $M(X) = a,$ имеем

$$D(X)=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}(x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}}dx=$$

$$=\left|\begin{matrix} x-a=\sigma z & \\ z=\dfrac{x-a}{\sigma} & dx=\sigma dz \end{matrix}\right|=\frac{\sigma^2}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} z\cdot z e^{-\frac{z^2}{2}}dz.$$

Интегрируя по частям, положив $u=z,\ d\upsilon=ze^{-\frac{z^2}{2}}dz,$ найдем

$$D(X)=\sigma^2.$$

Следовательно,

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\sigma^2}=\sigma.$$

Итак, среднее квадратическое отклонение нормального распределения равно параметру σ.

Вероятность попадания в заданный интервал нормальной случайной величины

Теорема. Вероятность того, что непрерывная случайная величина, распределенная по нормальному закону, примет значение на интервале (α,β), равна

$$P(\alpha<X<\beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right),$$

где *Ф(х)* – функция Лапласа.

Доказательство. Из свойств функции распределения известно, что $P(\alpha<X<\beta)=F(\beta)-F(\alpha)$. Так как случайная величина распределена нормально, то ее функция распределения связана с функцией Лапласа равенством $F(x)=0{,}5+\Phi\left(\frac{x-a}{\sigma}\right)$ (замечание 3).

Таким образом, искомая вероятность

$$P(\alpha<X<\beta)=0{,}5+\Phi\left(\frac{\beta-a}{\sigma}\right)-0{,}5-\Phi\left(\frac{\alpha-a}{\sigma}\right)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right).$$

Вычисление вероятности заданного отклонения

Часто требуется вычислить вероятность того, что отклонение нормально распределенной случайной величины

X по абсолютной величине меньше заданного положительного числа δ, т.е. требуется найти вероятность осуществления неравенства $|X-a|<\delta$.

Теорема. Вероятность того, что отклонение нормально распределенной случайной величины по абсолютной величине меньше положительного числа δ, находится из соотношения

$$P(|X-a|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Доказательство. Заменим неравенство $|X-a|<\delta$ равносильным ему двойным неравенством

$$-\delta<X-a<\delta, \text{ или } a-\delta<X<a+\delta.$$

Пользуясь предыдущей теоремой

$$P(|X-a|<\delta)=P(a-\delta<X<a+\delta)=$$

$$=\Phi\left[\frac{(a+\delta)-a}{\sigma}\right]-\Phi\left[\frac{(a-\delta)-a}{\sigma}\right]=\Phi\left(\frac{\delta}{\sigma}\right)-\Phi\left(-\frac{\delta}{\sigma}\right),$$

приняв во внимание равенство

$$\Phi\left(-\frac{\delta}{\sigma}\right)=-\Phi\left(\frac{\delta}{\sigma}\right)$$

(функция Лапласа – нечетная), окончательно имеем

$$P(|X-a|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Правило трех сигм

Преобразуем формулу

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

положив $\delta=\sigma t$. В итоге получим

$$P(|X-a|<\sigma t)=2\Phi(t).$$

Если t=3 и , следовательно, $\sigma t=3\sigma,$ то

$$P(|X-a|<3\sigma)=2\Phi(3)\approx 2\cdot 0{,}499=0{,}998,$$

т.е. вероятность того, что отклонение по абсолютной величине будет меньше утроенного среднего квадратического отклонения, равна 0,998.

Другими словами, вероятность того, что абсолютная величина отклонения превысит утроенное среднее квадратическое отклонение, очень мала, а именно равна 0,002 7. Это означает, что лишь в 0,27 % случаев так может произойти. Такие события исходя из принципа невозможности маловероятных событий можно считать практически невозможными. В этом и состоит сущность правила трех сигм: если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратического отклонения.

На практике правило трех сигм применяют так: если распределение изучаемой случайной величины неизвестно, но условие, указанное в приведенном правиле, выполняется, то есть основание предполагать, что изучаемая величина распределена нормально; в противном случае она не распределена нормально.

Решение типовых задач

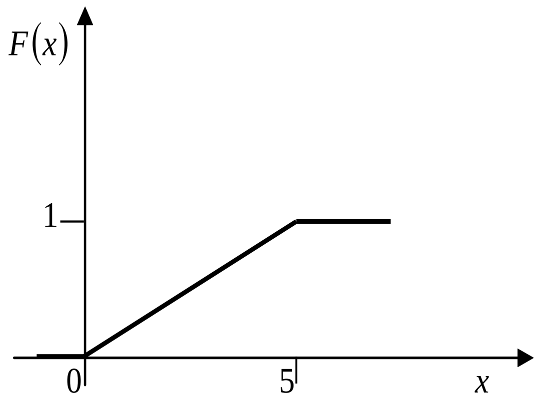
Задача 1. Поезда метро идут строго по расписанию. Интервал движения – 5 минут. Составить *f(x)* и *F(x)* случайной величины *X* – времени ожидания очередного поезда и построить их графики. Найти *M(X), D(X).*

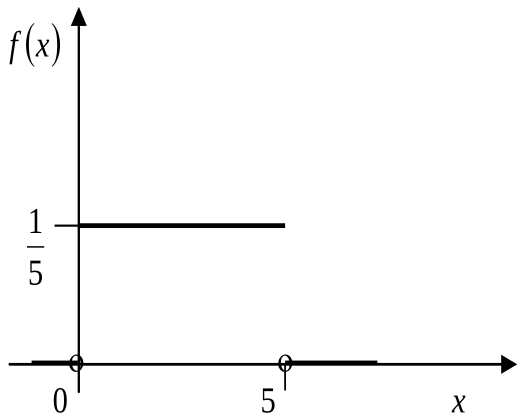
Решение. Случайная величина *X* – время ожидания очередного поезда. Величина *X* распределена равномерно на отрезке [0,5], поэтому воспользуемся формулами

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

Тогда имеем

$$F(x)=\begin{cases}0, & x<0;\\ \dfrac{x}{5}, & 0\le x\le 5;\\ 1, & x>5.\end{cases}\quad f(x)=\begin{cases}0, & x<0;\\ \dfrac{1}{5}, & 0\le x\le 5;\\ 0, & x>5.\end{cases}$$





Математическое ожидание и дисперсия вычисляются по формулам:

$$M(X)=\frac{b+a}{2};\ D(X)=\frac{(b-a)^2}{12}.$$

Тогда

$$M(X)=\frac{5+0}{2}=2{,}5;\ D(X)=\frac{(5-0)^2}{12}=\frac{25}{12}\approx 2{,}08.$$

Задача 2. Непрерывная случайная величина распределена по показательному закону с параметром $\lambda=2$. Составить функцию распределения, функцию плотности этой случайной величины. Найти числовые характеристики и вероятность того, что случайная величина попадет в интервал (0,3;1).

Решение. Очевидно, искомая плотность распределения

$$f(x)=\begin{cases}0, & \text{при} \quad x<0;\\ 2e^{-2x}, & \text{при} \quad x\ge 0.\end{cases}$$

Искомая функция распределения

$$F(x)=\begin{cases}0, & \text{при} \quad x<0;\\ 1-e^{-2x}, & \text{при} \quad x\ge 0.\end{cases}$$

По условию $\lambda=2$. Следовательно,

$M(X)=\sigma(x)=\dfrac{1}{\lambda}=\dfrac{1}{2}=0{,}5;\ D(X)=\dfrac{1}{\lambda^2}=\dfrac{1}{2^2}=0{,}25.$

Для нахождения вероятности $P(0{,}3<X<1)$ воспользуемся формулой $P(a<X<b)=e^{-\lambda a}-e^{-\lambda b}$.

Тогда, $P(0{,}3<X<1)=e^{-(2\cdot 0{,}3)}-e^{-(2\cdot 1)}=e^{-0{,}6}-e^{-2}\approx 0{,}549-0{,}135=0{,}414.$

Задача 3. Детали, выпускаемые цехом, по размеру диаметра распределены по нормальному закону. Стандартная длина диаметра детали равна a=35, среднее квадратическое отклонение $\sigma=4$. Требуется:

а) составить функцию плотности вероятностей;

б) найти вероятность того, что диаметр наудачу взятой детали будет больше $\alpha=34$ и меньше $\beta=40$;

в) найти вероятность того, что диаметр детали отклонится от стандартной длины не больше чем на $\delta=2$.

Решение. 1. Так как непрерывная случайная величина X распределена по нормальному закону, есть ее плотность распределения вероятностей выражается формулой

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

Следовательно,

$$f(x)=\frac{1}{4\sqrt{2\pi}}\cdot e^{-\frac{(x-35)^2}{32}}.$$

2. Для нормально распределенной случайной величины

$$P(\alpha<X<\beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right).$$

Тогда

$$P(34<X<40)=\Phi\left(\frac{40-35}{4}\right)-\Phi\left(\frac{34-35}{4}\right)=\Phi(1{,}25)+\Phi(0{,}25)=0{,}394\ 4+0{,}098\ 7=0{,}493\ 1.$$

3. Последнее задание решаем по формуле

$$P(|X-a|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Таким образом,

$$P(|X-35|<2)=2\Phi\left(\frac{2}{4}\right)=2\Phi(0{,}5)=0{,}3829,$$

где *Φ(x)* – интегральная функция Лапласа (приложение 3).

Задачи (61 – 70)

В задачах 61 – 70. Случайная величина *X* имеет нормальное распределение с параметрами $M(X)$ и $\sigma(X)$.
Требуется:

1. Составить функцию плотности распределения и построить ее график.
2. Найти вероятность того, что случайная величина в результате испытания примет значение, принадлежащее интервалу $(\alpha;\beta)$.
3. Найти вероятность того, что абсолютная величина отклонения значений случайной величины от ее математического ожидания не превысит δ.

61. $M(X)=375;\ \sigma(X)=25;\ \alpha=300;\ \beta=425;\ \delta=0{,}1.$

62. $M(X)=10;\ \sigma(X)=2;\ \alpha=5;\ \beta=12;\ \delta=5.$

63. $M(X)=164;\ \sigma(X)=5{,}5;\ \alpha=153;\ \beta=170;\ \delta=0{,}1.$

64. $M(X)=5;\ \sigma(X)=0{,}81;\ \alpha=4;\ \beta=7;\ \delta=2.$

65. $M(X)=20;\ \sigma(X)=0{,}5;\ \alpha=19;\ \beta=25;\ \delta=1{,}5.$

66. $M(X)=10;\ \sigma(X)=4;\ \alpha=12;\ \beta=14;\ \delta=0{,}1.$

67. $M(X)=25;\ \sigma(X)=4;\ \alpha=13;\ \beta=30;\ \delta=0{,}1.$

68. $M(X)=4{,}5;\ \sigma(X)=0{,}05;\ \alpha=3{,}5;\ \beta=4{,}35;\ \delta=0{,}1.$

69. $M(X)=16;\ \sigma(X)=100;\ \alpha=15{,}75;\ \beta=16{,}3;\ \delta=16{,}25.$

70. $M(X)=12;\ \sigma(X)=4;\ \alpha=10;\ \beta=14;\ \delta=5.$

2.5. Закон больших чисел

Закон больших чисел **–** это совокупность утверждений, что с вероятностью, как угодно близкой к единице (нулю), отклонение средней арифметической достаточно большого числа случайных величин от постоянной – средней арифметической их математических ожиданий не превзойдет (будет больше) сколь угодно малого числа $\varepsilon>0$.

Неравенства Маркова и Чебышева

Теорема. Пусть случайная величина *X* принимает только неотрицательные значения и имеет конечное математическое ожидание M(X). Тогда для любого $\delta>0$ имеют место неравенства Маркова:

$$P(X\le\delta)\ge 1-\frac{M(X)}{\delta},\ \ P(X>\delta)\le\frac{M(X)}{\delta}.$$

Доказательство. Пусть закон распределения дискретной случайной величины *X* задан таблицей

X	x_1	x_2	$\dots$	x_k	x_{k+1}	$\dots$	x_n
P	p_1	p_2	$\dots$	p_k	p_{k+1}	$\dots$	p_n

Выберем и зафиксируем $\delta>0$. Предположим, что x_1, x_2, …, x_k не превосходят δ, т.е. $x_i\le\delta$ для $i=(\overline{1;k})$, а все x_{k+1}, …, $x_n>\delta$, т.е. $x_i>\delta$, для $i=(\overline{k+1;n})$.

Тогда

$$M(X)=x_1p_1+x_2p_2+\dots+x_kp_k+x_{k+1}p_{k+1}+\dots+x_np_n.$$

Сумма $x_{k+1}p_{k+1}+\dots+x_np_n\le M(X)$. Это неравенство лишь усилится, если заменить каждое из значений x_{k+1}, …, x_n меньшей величиной δ, следовательно,

$$\delta(p_{k+1}+\dots+p_n)\le M(X);$$

$$p_{k+1}+\dots+p_n\le\frac{M(X)}{\delta}.$$

Сумма в левой части неравенства по теореме о сумме несовместных событий дает вероятность того, что $X>\delta$, т.е. $P(X>\delta)=P(X=x_{k+1})+\dots+P(X=x_n)=p_{k+1}+\dots+p_n$ следовательно $P(X>\delta)\le\frac{M(X)}{\delta}$.

Так как сумма противоположных событий равна единице, то

$P(X>\delta)$ $+P(X \le \delta)$ $=1$, следовательно, $P(X \le \delta)$ $=$ $1-$ $P(X>\delta)$, $P(X \le \delta) \ge 1 - \dfrac{M(X)}{\delta}$.

Пример. Вероятность того, что у отдельного вкладчика некоторого сберегательного банка сумма вклада не больше 3 млн. руб., превышает 0,8. Банк обслуживает 1 000 вкладчиков. Какова общая сумма вкладов этого сберегательного банка?

Решение. $P(X \le 3) \ge 0{,}8 = 1 - \dfrac{M(X)}{3}$. Тогда $\dfrac{M(X)}{3} = 0{,}2 \Rightarrow M(X) = 0{,}6$. Общая сумма 600 000 000 руб.

Теорема. Для любого $\varepsilon>0$ и любой случайной величины X, дисперсия которой конечна, выполняются неравенства Чебышева:

$$P(|X - M(X)| \ge \varepsilon) \le \frac{D(X)}{\varepsilon^2}, \quad P(|X - M(X)| < \varepsilon) \ge 1 - \frac{D(X)}{\varepsilon^2}.$$

Доказательство. Неравенство $|X - M(X)| < \varepsilon$ равносильно неравенству $(X - M(X))^2 < \varepsilon^2$. Тогда $P((X - M(X))^2 < \varepsilon^2) \ge 1 - \dfrac{M(X - M(X))^2}{\varepsilon^2}$ (согласно неравенству Маркова и в силу неотрицательности случайной величины $(X - M(X))^2$). Следовательно, $P(|X - M(X)| < \varepsilon) \ge 1 - \dfrac{D(X)}{\varepsilon^2}$.

Так как события противоположны, то $P(|X - M(X)| < \varepsilon) + P(|X - M(X)| \ge \varepsilon) = 1$, и $P(|X - M(X)| \ge \varepsilon) \le \dfrac{D(X)}{\varepsilon^2}$.

Замечание. Для нормально распределенной случайной величины существует точная формула $P(|X - a| < \delta) = 2\Phi\left(\dfrac{\delta}{\sigma}\right)$.

Пример. Вероятность того, что студент учебного заведения в период работы читального зала посетит его, равна 0,3. Оценить вероятность того, что среди 900 студентов читального зала посетят от 240 до 300 человек.

Решение. $M(X) = np = 900 \cdot 0{,}3 = 270$; $D(X) = npq = 189$.

Величина отклонения от *M(X)* равна $|240-270|=|300-270|=30=\varepsilon$.

Тогда $P(|X-270|<30)\ge 1-\frac{189}{900}=0{,}79$.

Закон больших чисел в форме Чебышева

Теорема (непредельная форма). Пусть X_1, X_2, …, X_n – последовательность попарно независимых, однородных случайных величин, имеющих конечные дисперсии D(X_1), D(X_2), …, D(X_n), ограниченные сверху числом *C* ($D(X_i)\le C(i=\overline{1,n})$). Тогда для любого сколь угодно малого числа ε>0, имеют место неравенства

$$P(|\overline{X}-M(\overline{X})|<\varepsilon)\ge 1-\frac{C}{n\varepsilon^2},\ P(|\overline{X}-M(\overline{X})|\ge\varepsilon)\le\frac{C}{n\varepsilon^2}.$$

Доказательство. Вычислим $D(\overline{X})$

$$D(\overline{X})=D\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)=\frac{D(X_1)+D(X_2)+\ldots+D(X_n)}{n^2};$$

так как $D(X_i)\le C$, $(i=\overline{1,n})$, то

$$D(\overline{X})\le\frac{C+C+\ldots+C}{n^2}=\frac{Cn}{n^2}=\frac{C}{n}.$$

Из неравенства Чебышева дальнейшее доказательство теоремы очевидно.

Теорема (предельная форма). Пусть X_1, X_2, …, X_n – последовательность независимых, однородных случайных величин, имеющих конечные *D(X)*, которые ограничены сверху постоянной *C*. Тогда при $n\to\infty$ и любого сколь угодно малого числа $\varepsilon>0$ имеют место предельные равенства:

$$\lim_{n\to\infty}P\left(|\overline{X}-M(\overline{X})|<\varepsilon\right)=1;\ \lim_{n\to\infty}P\left(|\overline{X}-M(\overline{X})|\ge\varepsilon\right)=0.$$

Для доказательства достаточно совершить предельный переход при $n\to\infty$ в неравенствах предыдущей теоремы.

Сущность доказанной теоремы заключается в следующем: хотя отдельные случайные величины могут принимать значения, далекие от своих математических ожиданий, среднее арифметическое достаточно большого числа случайных величин с большой вероятностью принимает значения, близкие к определенному постоянному числу, а именно к числу $\frac{M(X_1)+M(X_2)+...+M(X_n)}{n}=a$. Коротко теорему Чебышева записывают так: $\frac{X_1+X_2+...+X_n}{n} \underset{n\to\infty}{\overset{p}{\to}} a$.

Таким образом, при достаточно большом числе n, среднее арифметическое случайных величин теряет свой случайный характер и ведет себя как постоянная – математическое ожидание. На этом выводе основывается выборочный метод в статистике.

Закон больших чисел в форме Бернулли

Теорема (непредельная форма). Если проводится n повторных независимых испытаний, в каждом из которых вероятность появления событий A постоянна и равна p, то при $n\to\infty$, и для любого сколь угодно малого числа $\varepsilon>0$, имеют место неравенства:

$$P(|W-p|<\varepsilon)\geq 1-\frac{pq}{n\varepsilon^2};\quad P(|W-p|\geq\varepsilon)\leq\frac{pq}{n\varepsilon^2},$$

где $W=\frac{m}{n}$ – относительная частота появления события A.

Доказательство следует из неравенства Чебышева, если в качестве случайной величины X рассмотреть W и воспользоваться тем, что $M(W) = p$, $D(W)=\frac{pq}{n}$.

Теорема (предельная форма). Если в условиях теоремы Бернулли $n\to\infty$, то для любого сколь угодно малого числа $\varepsilon>0$, справедливы предельные соотношения

$$\lim_{n\to\infty} P(|W-p|<\varepsilon)=1;\quad \lim_{n\to\infty} P(|W-p|\geq\varepsilon)=0.$$

Доказательство осуществляется путем предельного перехода в неравенствах предыдущей теоремы.

Итак, теорема Бернулли утверждает, что при $n \to \infty$ относительная частота стремится по вероятности к P. Коротко теорему Бернулли записывают так: $\frac{m}{n} \underset{n\to\infty}{\overset{p}{\to}} P$.

Таким образом, при достаточно больших n относительная частота $W(A)$ теряет свой случайный характер и ведет себя как постоянная вероятность.

Центральная предельная теорема в форме Ляпунова

Теорема. Пусть X_1, X_2, …, X_n – независимые случайные величины, имеющие один и тот же закон распределения с $M(X_i) = a$ и $D(X_i) = \sigma^2$. Тогда при $n \to \infty$, закон распределения суммы $X = X_1 + X_2 + \ldots + X_n$ неограниченно приближается и нормальному.

Примем теорему без доказательства.

Сущность данной теоремы заключается в следующем: если отдельные слагаемые X_i имеют различные законы распределения, то влияние каждого из них на величину суммы случайных величин равномерно мало и закон распределения суммы случайных величин стремится к нормальному.

Вопросы для самоконтроля

1. Закон распределения дискретной случайной величины X имеет вид

X	***2***	***5***	***8***
P	***0,2***	***0,4***	***?***

Найти P(X=8).

101. 0,4. ***102.*** 0,8. ***103.*** 0,2. ***104.*** 0,5.

2. Как найти $P(a<X<b)$ с помощью функции распределения?

201. $\int_a^b F(x)dx$. ***202.*** $0 \le F(x) \le 1$. ***203.*** $F(b)-F(a)$. ***204.*** $F(a<X<b)=0$.

3. Найти математическое ожидание случайной величины, зная ее закон распределения

X	***6***	***3***	***1***
P	***0,2***	***0,3***	***0,5***

301. 3,4. ***302.*** 2,6. ***303.*** 1. ***304.*** 0

4. Две случайные величины называются …, если закон распределения одной из них не зависит от того, какие возможные значения приняла случайная величина

400) несовместными. ***402)*** непрерывными.
403) нестандартными. ***404)*** независимыми.

5. Вероятностный смысл математического ожидания состоит в том, что

501. $M(x) \approx \overline{X}$. ***502.*** $M(x) \approx P(x)$. ***503.*** $M(x) \approx D(x)$. ***504.*** $M(x) \approx W(x)$.

6. Как определяется дисперсия непрерывной случайной величины, если ее возможные значения принадлежат всей числовой оси?

601. $D(x) = \int_a^b x^2 f(x)dx - [M(x)]^2$. ***602.*** $D(x) = \int_{-\infty}^{\infty} x^2 f(x)dx$.

603. $D(x) = \int_{-\infty}^{\infty} [x - M(x)]^2 f(x)dx$. ***604.*** $D(x) = \int_{-\infty}^{x} f(x)dx$.

7. Как найти числовые характеристики случайной величины, распределенной по геометрическому закону?

701. $M(x)=D(x)=\lambda$. ***702.*** $M(x)=\frac{a+b}{2}$; $D(x)=\frac{(b-a)^2}{12}$.

703. *M(x)=np; D(x)=npq*. ***704.*** $M(x)=\frac{1}{p}$; $D(x)=\frac{q}{p^2}$.

8. При каких значениях параметров нормальное распределение называется нормированным?

801. a=0; σ=0. ***802.*** a=0; σ=1. ***803.*** a=1; σ=0. ***804.*** a=1; σ=1.

9. Как найти функцию распределения, если плотность распределения вероятностей известна и интегрируема?

901. $f(x)=F'(x)$. ***902.*** $F(x)=\int\limits_{b}^{a} f(x)dx$.

903. $F(x)=\int\limits_{-\infty}^{x} f(x)dx$. ***904.*** $f(x)=\lim\limits_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}$.

10. Задана нормальная случайная величина функцией плотности

$$f(x)=\frac{1}{2\sqrt{2\pi}}e^{-\frac{(x-5)^2}{8}}.$$

Найти M(x) и D(x)

1001. *M(x)= 2; D(x)=5.* ***1002.*** *M(x)= 5; D(x)=2.*

1003. *M(x)= 2; D(x)=8.* ***1004.*** *M(x)= 5; D(x)=4.*

ГЛАВА 3. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

3.1. Статистические оценки параметров распределения

Основные задачи математической статистики

Математическая статистика изучает закономерности, которые имеют место в массовых совокупностях однородных объектов.

Основные **задачи** математической статистики:

1. Разработка методов сбора и группировки статистических сведений, полученных в результате наблюдений или в результате специально поставленных экспериментов.

2. Разработка методов анализа статистических данных в зависимости от целей исследования. Сюда относятся:

а) оценка неизвестной вероятности события; оценка неизвестной функции
распределения; оценка параметров распределения, вид которого известен; оценка зависимости случайной величины от одной или нескольких случайных величин и др.;

б) проверка статистических гипотез о виде неизвестного распределения или о
величине параметров распределения, вид которого известен.

Современная математическая статистика разрабатывает способы определения числа необходимых испытаний до начала исследования (планирование эксперимента), в ходе исследования (последовательный анализ) и решает многие другие задачи. Современную математическую статистику определяют как науку о принятии решений в условиях неопределенности.

Итак, задача математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Генеральная и выборная совокупности. Виды выборки. Способы отбора

Пусть требуется изучить совокупность однородных объектов относительно некоторого качественного или количественного признака, характеризующего эти объекты.

Иногда проводят сплошное обследование, т.е. обследуют каждый из объектов совокупности относительно признака, которым интересуются. На практике, однако, сплошное обследование применяют сравнительно редко. Чаще случайно отбирают из всей совокупности ограниченное число объектов и подвергают их изучению.

Генеральной совокупностью называют совокупность однородных объектов, подлежащих изучению.

Выборочной совокупностью (выборкой) называется совокупность объектов, отобранных из генеральной совокупности.

Объемом совокупности (выборной или генеральной) называют число объектов этой совокупности. Например, если из 1 000 деталей отобрано для обследования 100 деталей, то объем генеральной совокупности N=1 000, а объем выборки n=100.

При составлении выборки можно поступать двумя способами: после того как объект отобран и над ним произведено наблюдение, он может быть возвращен либо не возвращен в генеральную совокупность. В соответствии со сказанным, выборки подразделяют на повторные и безповторные.

Повторной называют выборку, при которой отобранный объект (перед отбором следующего) возвращается в генеральную совокупность.

Бесповторной называют выборку, при которой отобранный объект в генеральную совокупность не возвращается.

На практике обычно пользуются бесповторным случайным отбором.

Для того, чтобы по данным выборки можно было достаточно уверенно судить об интересующем признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно его представляли. Другими словами, выборка должна правильно представлять пропорции генеральной совокупности. Это требование формулируют так: выборка должна быть ***репрезентативной (представительной).***

На практике применяют различные ***способы отбора.*** Принципиально эти способы можно подразделить на два вида:

1. Отбор, не требующий расчленения генеральной совокупность на части. Сюда

относятся: а) простой случайный бесповторный отбор; б) простой случайный повторный отбор.

2. Отбор, при котором генеральная совокупность разбивается на части. Сюда отно-

сятся: а) типический отбор; б) механический отбор; в) серийный отбор.

Простым случайным называют такой отбор, при котором объекты извлекают по одному из всей генеральной совокупности. Осуществить простой отбор можно с помощью повторной и бесповторной выборки.

Типическим называют отбор, при котором объекты отбираются не из всей генеральной совокупности, а из каждой ее «типической» части. Например, если детали изготовляют на нескольких станках, то отбор производят не из всей совокупности деталей, произведенных всеми станками, а из продукции каждого станка в отдельности. Типическим отбором пользуются тогда, когда обследуемый признак заметно колеблется в различных типических частях генеральной совокупности. Например, если продукция изготовляется на нескольких машинах, среди которых есть более или менее изношенные, то здесь типический отбор целесообразен.

Механическим называют отбор, при котором генеральную совокупность «механически» делят на столько групп, сколько объектов должно войти в выборку, а из каждой группы отбирают один объект. Например, если нужно отобрать 20 % изготовленных станком деталей, то отбирают каждую пятую деталь; если требуется отобрать 5 % деталей, то отбирают каждую двадцатую деталь, и т.д.

Серийным называют отбор, при котором объекты отбирают из генеральной совокупности не по одному, а «сериями», которые подвергаются сплошному обследованию. Например, если изделия изготовляются большой группой станков-автоматов, то подвергают сплошному обследованию продукцию только нескольких станков. Серийным отбором пользуются тогда, когда обследуемый признак колеблется в различных сериях незначительно.

Подчеркнем, что на практике часто применяется комбинированный отбор, при котором сочетаются указанные выше способы. Например, иногда разбивают генеральную совокупность на серии одинакового объема, затем простым случайным отбором выбирают несколько серий и, наконец, из каждой серии простым случайным отбором извлекают отдельные объекты.

Статистическое распределение выборки

Пусть из генеральной совокупности извлечена выборка, причем x_1наблюдалось n_1раз, $x_2 - n_2$ раз, $x_k - n_k$ раз и $\sum n_i = n$ –

объем выборки. Наблюдаемые значения x_i называют ***вариантами***, а последовательность вариант, записанных в возрастающем порядке, – ***вариационным рядом***. Числа наблюдений называют ***частотами***, а их отношения к объему выборки $\frac{n_i}{n} = W_i$ – ***относительными частотами***.

Статистическим распределением выборки называют перечень вариант и соответствующих им частот или относительных частот.

Различают дискретные и интервальные статистические распределения.

Статистическое распределение называется ***дискретным***, если значения признака отличаются друг от друга не менее чем на некоторую постоянную величину

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
W_i	W_1	W_2	…	W_k

$$\sum_{i=1}^{k} n_i = n; \quad \sum_{i=1}^{k} W_i = 1.$$

Для графического представления дискретного распределения используют полигон частот (полигон относительных частот).

Полигоном частот называют ломаную, отрезки которой соединяют точки $(x_1;n_1),(x_2;n_2),\ldots,(x_k;n_k)$. Для построения полигона на оси абсцисс откладывают варианты x_i, а на оси ординат – соответствующие им частоты n_i. Точки $(x_i;n_i)$ соединяют отрезками прямых и получают полигон частот (рис. 3.1).

Полигоном относительных частот называют, ломаную отрезки которой соединяют точки $(x_1;W_1),(x_2;W_2),\ldots,(x_k;W_k)$. Для построения полигона относительных частот на оси абсцисс откладывают варианты x_i, а на оси ординат – соответствующие им относи-

тельные частоты w_i. Точки $(x_i;w_i)$ соединяют отрезками прямых и получают полигон относительных частот.

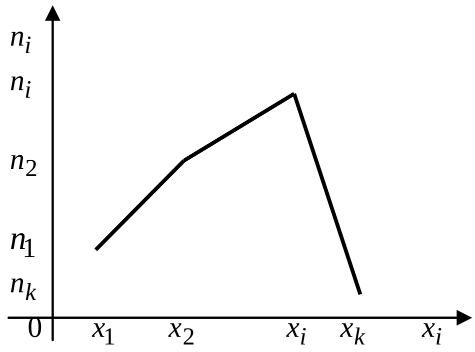


Рис. 3.1.

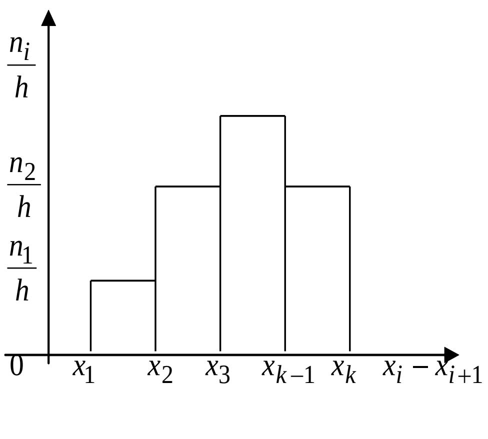
В случае непрерывных случайных величин рассматривают ***интервальное*** статистическое распределение выборки. Оно оформляется в виде следующей таблицы:

$(x_i;x_{i+1})$	$(x_1;x_2)$	$(x_2;x_3)$	…	$(x_{k-1};x_k)$
n_i	n_1	n_2	…	n_{k-1}
W_i	W_1	W_2	…	W_{k-1}

$$\sum_{i=1}^{k} n_i = n; \quad \sum_{i=1}^{k} W_i = 1.$$

Разница между двумя соседними вариантами называется ***шагом*** интервала $h = x_i - x_{i+1}$. От интервального распределения можно перейти к дискретному, взяв на каждом интервале $(x_i;x_{i+1})$ за отдельное значение x_i^* величину $x_i^* = \frac{x_i + x_{i+1}}{2}$, являющуюся серединой этого интервала.

Графической характеристикой интервальных распреде-



лений является гистограмма частот (гистограмма относительных частот).

Рис. 3.2.

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению $\frac{n_i}{h}$ (плотность частоты) (рис. 3.2).

Для построения гистограммы частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс, на расстоянии $\frac{n_i}{h}$.

Площадь i-го частичного прямоугольника равна $\frac{hn_i}{h} = n_i$ - сумме частот вариант i-го интервала; следовательно, ***площадь гистограммы частот равна сумме всех частот, т.е. объему выборки.***

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению $\frac{W_i}{h}$ (плотность относительной частоты).

Для построения гистограммы относительных частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс на расстоянии $\frac{W_i}{h}$.

Площадь i-го частичного прямоугольника равна $\frac{hW_i}{h} = W_i$ - относительной частоте вариант, попавших в i-й интервал. Следовательно, ***площадь гистограммы относительных частот равна сумме всех относительных частот, т.е. единице.***

По виду гистограмм можно предположить, какому теоретическому закону подчинен изучаемый признак генеральной совокупности. Форма гистограммы относительных частот дает представление о форме графика функции плотности $f(x)$ случайной величины.

Эмпирическая функция распределения

Пусть известно статистическое распределение частот количественного признака X. Введем обозначения: n_x - число наблюдений, при которых наблюдалось значение признака, меньшее x; n - общее число наблюдений (объем выборки). Ясно, что относительная частота события $X<x$ равна $\frac{n_x}{n}$. Если x изменяется, то, вообще говоря, изменяется и относительная частота, т.е. относительная частота $\frac{n_x}{n}$ есть функция от x. Так как эта функция находится эмпирическим (опытным) путем, то ее называют эмпирической.

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X<x$.

Итак, по определению,

$$F^*(x)=\frac{n_x}{n},$$

где n_x - сумма частот вариант, меньших x; n – объем выборки.

Из определения функции $F^*(x)$ вытекают следующие ее свойства:

1. Значения эмпирической функции принадлежат отрезку $[0;1]$.
2. $F^*(x)$ - неубывающая функция.

3. Если x_1 - наименьшая варианта, то $F^*(x)=0$ при $x \le x_1$; если x_k - наибольшая варианта, то $F^*(x)=1$ при $x > x_k$.

Эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

Различие между эмпирической и теоретической функциями состоит в том, что теоретическая функция $F(x)$ определяет вероятность события $X<x$, а эмпирическая функция $F^*(x)$ определяет относительную частоту этого же события.

Статистические оценки параметров распределения

Пусть требуется изучить количественный признак генеральной совокупности. Допустим, что из теоретических соображений удалось установить, какое именно распределение имеет признак. Естественно возникает задача оценки параметров, которыми определяется это распределение. Например, если наперед известно, что изучаемый признак распределен в генеральной совокупности нормально, то необходимо оценить (приближенно найти) математическое ожидание и среднее квадратическое отклонение, так как эти два параметра полностью определяют нормальное распределение; если же есть основание считать, что признак имеет, например, распределение Пуассона, то необходимо оценить параметр λ, которым это распределение определяется.

Обычно в распоряжении исследователя имеются лишь данные выборки, например, значения количественного признака $x_1, x_2, \ldots, x_n$, полученные в результате n наблюдений. Через эти данные и выражают оцениваемый параметр. Рассматривая $x_1, x_2, \ldots, x_n$ как независимые случайные величины $X_1, X_2, \ldots, X_n$, можно найти статистическую оценку независимого параметра теоретического распределения.

Пусть одна из характеристик случайной величины X найдена приближенно, путем произведенных независимых опытов (испытаний), обозначим ее θ^*. Тогда случайная величина θ^* - ***статистическая оценка*** неизвестного параметра θ теоретического распределения количественного признака генеральной совокупности.

Статистическая оценка должна удовлетворять трем основным требованиям: несмещенности, эффективности и состоятельности.

Пусть произведено k опытов, в каждом из которых оценка θ^*приняла значения $\theta_1^*, \theta_2^*, \ldots, \theta_k^*$. Если оценка θ^*дает приближенное значение θ с избытком; тогда каждое найденное по данным выборок число $\theta_i^*(i = 1, 2, \ldots, k)$ больше истинного значения θ. Ясно, что в этом случае и математическое ожидание (среднее значение) случайной величины θ^* больше, чем θ, т.е. $M(\theta^*) > \theta$.

Таким образом, использование статистической оценки, математическое ожидание которой не равно оцениваемому параметру, привело бы к систематическим ошибкам. Требование $M(\theta^*) = \theta$ гарантирует избавление от этих ошибок.

Несмещенной называют статистическую оценку θ^*, математическое ожидание которой равно оцениваемому параметру θ при любом объеме выборки, т.е.

$$M(\theta^*) = \theta.$$

Смещенной называют оценку, математическое ожидание которой не равно оцениваемому параметру.

Однако было бы ошибочным считать, что несмещенная оценка всегда дает хорошее приближение оцениваемого параметра. Действительно, возможные значения θ^* могут быть сильно рассеяны вокруг своего среднего значения, т.е. дисперсия $D(\theta^*)$ может быть значительной. В этом случае найденная по данным одной выборки оценка, например θ_1^*, может оказаться весьма удаленной от среднего значения $\overline{\theta^*}$, а значит, и от самого оцениваемого параметра θ; приняв θ_1^* в качестве приближенного значения θ, мы допустили бы большую ошибку. Если же потребовать, чтобы дисперсия θ^* была малой, то возможность допустить большую ошибку будет исключена. По этой причине к статистической оценке предъявляется требование эффективности.

Эффективной называют статистическую оценку, которая (при заданном объеме выборки n) имеет наименьшую возможную дисперсию.

При рассмотрении выборок большего объема (n велико!) к статистическим оценкам предъявляется требование состоятельности.

Состоятельной называют статистическую оценку, которая при $n \to \infty$ стремится, по вероятности, к оцениваемому параметру, т.е. при увеличении количества опытов оценка θ^* параметра должна стремиться (сходиться) к истинному значению этого параметра.

Генеральная и выборочная средняя

Пусть изучается дискретная генеральная совокупность относительно количественного признака X.

Генеральной средней $\bar{x}_Г$ называют среднее арифметическое значений признака генеральной совокупности.

Если все значения $x_1, x_2, \ldots, x_N$ признака генеральной совокупности объема N различны, то

$$\bar{x}_Г = \frac{(x_1 + x_2 + \ldots + x_N)}{N}.$$

Если же значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $N_1, N_2, \ldots, N_k$ причем $N_1 + N_2 + \ldots + N_k = N,$ то

$$\bar{x}_Г = \frac{(x_1 N_1 + x_2 N_2 + \ldots + x_k N_k)}{N},$$

т.е. генеральная средняя есть средняя взвешенная значений признака с весами, равными соответствующим частотам.

Замечание. Пусть генеральная совокупность объема *N* содержит объекты с различными значениями признака *X*, равными $x_1, x_2, \ldots, x_N$. Представим себе, что из этой совокупности наудачу извлекается один объект. Вероятность того, что будет извлечен объект со значением признака, например $x_1,$ очевидно, равна $\frac{1}{N}$. С этой же вероятностью может быть извлечен и любой другой объект. Таким образом, величину признака *X* можно рассматривать как случайную величину, возможные значения которой $x_1, x_2, \ldots, x_n$ имеют одинаковые вероятности, равные $\frac{1}{N}$. Найдем математическое ожидание *M(X)*:

$$M(X) = x_1 \cdot \frac{1}{N} + x_2 \cdot \frac{1}{N} + \ldots + x_N \cdot \frac{1}{N} = \frac{(x_1 + x_2 + \ldots + x_N)}{N} = \bar{x}_Г.$$

Итак, если рассматривать обследуемый признак *X* генеральной совокупности как случайную величину, то математическое ожидание признака равно генеральной средней этого признака:

$$M(X) = \bar{x}_Г.$$

Пусть для изучения генеральной совокупности относительно количественного признака *X* извлечена выборка объема *n*.

Выборочной средней $\bar{x}_В$ называют среднее арифметическое значение признака выборной совокупности.

Если все значения $x_1, x_2, \ldots, x_n$ признака выборки объема *n* различны, то

$$\bar{x}_B = \frac{(x_1 + x_2 + \ldots + x_n)}{n}.$$

Если же значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $n_1, n_2, \ldots, n_k$, причем $n_1 + n_2 + \ldots + n_k = n,$ то

$$\bar{x}_B = \frac{(x_1 n_1 + x_2 n_2 + \ldots + x_k n_k)}{n}$$

или

$$\bar{x}_B = \frac{\sum\limits_{i=1}^{k} n_i x_i}{n},$$

т.е. выборочная средняя есть средняя взвешенная значений признака с весами, равными соответствующим частотам.

Оценка генеральной средней по выборочной средней

Пусть из генеральной совокупности извлечена повторная выборка объема *n* со значениями признака $x_1, x_2, \ldots, x_n$. Пусть генеральная средняя $\bar{x}_Г$ неизвестна и требуется оценить ее по данным выборки. В качестве оценки генеральной средней принимают выборочную среднюю

$$\bar{x}_B = \frac{(x_1 + x_2 + \ldots + x_n)}{n}.$$

Данная оценка удовлетворяет всем трем требованиям.

Докажем несмещенность, т.е.

$$M(\bar{x}_B) = \bar{x}_Г.$$

Будем рассматривать $\bar{x}_B$ как случайную величину и $x_1, x_2, \ldots, x_n$ как независимые, одинаково распределенные случайные величины $X_1, X_2, \ldots, X_n$. Поскольку эти величины одинаково распределены, то они имеют одинаковые числовые характеристики, в частности, одинаковое математическое ожидание, $M(x_i) = a$. Тогда

$$M(\bar{x}_B) = M\,\frac{(x_1 + x_2 + \ldots + x_n)}{n} = a.$$

С другой стороны, $M(X) = \bar{x}_Г = a.$ В результате имеем $M(\bar{x}_B) = \bar{x}_Г$

Эффективность и состоятельность данной оценки предлагается доказать самостоятельно.

Генеральная и выборочная дисперсия

Для того чтобы охарактеризовать рассеяние значений количественного признака X генеральной совокупности вокруг своего среднего значения, вводят сводную характеристику – генеральную дисперсию.

Генеральной дисперсией $D_Г$ называют среднее арифметическое квадратов отклонений значений признака генеральной совокупности от их среднего значения $\bar{x}_Г$.

Если все значения $x_1, x_2, \ldots, x_N$ признака генеральной совокупности объема N различны, то

$$D_Г = \frac{\sum\limits_{i=1}^{N}\left(x_i - \bar{x}_Г\right)^2}{N}.$$

Если же значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $N_1, N_2, \ldots, N_k$ причем $N_1 + N_2 + \ldots + N_k = N,$ то

$$D_Г = \frac{\sum\limits_{i=1}^{k} N_i\left(x_i - \bar{x}_Г\right)^2}{N},$$

т.е. генеральная дисперсия есть средняя взвешенная квадратов отклонений с весами, равными соответствующим частотам.

Выборочной дисперсией D_B называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\bar{x}_B$.

Если все значения $x_1, x_2, \ldots, x_n$ признака выборки объема n различны, то

$$D_B = \frac{\sum\limits_{i=1}^{n}\left(x_i - \bar{x}_B\right)^2}{n}.$$

Если же значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $n_1, n_2, \ldots, n_k$ причем $n_1 + n_2 + \ldots + n_k = n,$ то

$$D_B = \frac{\sum_{i=1}^{k} n_i \left(x_i - \bar{x}_B\right)^2}{n},$$

т.е. выборочная дисперсия есть средняя взвешенная квадратов отклонений с весами, равными соответствующим частотам.

Формула для вычисления дисперсии

Вычисление дисперсии можно упростить, используя следующую теорему.

Теорема. Дисперсия равна среднему квадратов значений признака минус квадрат общей средней:

$$D = \overline{x^2} - \left(\bar{x}\right)^2.$$

Доказательство. Справедливость теоремы вытекает из преобразований:

$$D = \frac{\sum n_i \left(x_i - \bar{x}\right)^2}{n} = \frac{\sum n_i \left(x_i^2 - 2x_i\bar{x} + \left(\bar{x}\right)^2\right)}{n} =$$

$$= \frac{\sum n_i x_i^2}{n} - 2\bar{x}\frac{\sum n_i x_i}{n} + \left(\bar{x}\right)^2 \frac{\sum n_i}{n} = \overline{x^2} - 2\bar{x}\cdot\bar{x} + \left(\bar{x}\right)^2 = \overline{x^2} - \left(\bar{x}\right)^2.$$

Итак, $D = \overline{x^2} - \left(\bar{x}\right)^2$, где $\bar{x} = \frac{\sum n_i x_i}{n}$, $\overline{x^2} = \frac{\sum n_i x_i^2}{n}$.

Оценка генеральной дисперсии по исправленной выборочной

Пусть из генеральной совокупности в результате n независимых наблюдений над количественным признаком X извлечена повторная выборка объема n

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

При этом $n_1 + n_2 + \ldots + n_k = n$.

Требуется по данным выборки оценить неизвестную генеральную дисперсию $D_Г$. Если в качестве оценки генеральной дисперсии принять выборочную дисперсию, то эта оценка будет приводить к систематическим ошибкам, давая заниженное значение генеральной дисперсии. Объясняется это тем, что, как можно доказать, выборочная дисперсия является смешенной оценкой $D_Г$, другими словами, математическое ожидание выборочной дисперсии не равно оцениваемой генеральной дисперсии, а равно

$$M[D_B] = \frac{n-1}{n} D_Г.$$

Легко «исправить» выборочную дисперсию так, чтобы ее математическое ожидание было равно генеральной дисперсии. Достаточно для этого умножить D_B на дробь $\frac{n}{(n-1)}$. Сделав это, получим ***исправленную дисперсию***, которую обычно обозначают через s^2:

$$s^2 = \frac{n}{n-1} D_B = \frac{n}{n-1} \frac{\sum_{i=1}^{k} n_i \left(x_i - \bar{x}_B\right)^2}{n} = \frac{\sum_{i=1}^{k} n_i \left(x_i - \bar{x}_B\right)^2}{n-1}.$$

Исправленная дисперсия является, конечно, несмещенной оценкой генеральной дисперсии. Действительно,

$$M\left[s^2\right] = M\left[\frac{n}{n-1} D_B\right] = \frac{n}{n-1} M[D_B] = \frac{n}{n-1} \cdot \frac{n-1}{n} D_Г = D_Г.$$

Итак, в качестве оценки генеральной дисперсии принимают исправленную
дисперсию

$$s^2 = \frac{\sum_{i=1}^{k} n_i \left(x_i - \bar{x}_B\right)^2}{(n-1)}.$$

Для оценки же среднего квадратического отклонения генеральной совокупности используют «исправленное» среднее квадратическое отклонение, которое равно квадратному корню из исправленной дисперсии:

$$s=\sqrt{\frac{\sum_{i=1}^{k} n_i\left(x_i-\bar{x}_B\right)^2}{(n-1)}}\,.$$

Другие характеристики вариационного ряда

Кроме выборочной средней и выборочной дисперсии применяются и другие характеристики вариационного ряда. Укажем главные из них.

Модой M_0 **называют варианту, которая имеет наибольшую частоту. Например, для ряда**

варианта	1	4	7	9
частота	5	1	20	6

мода равна 7.

Медианой m_e **называют варианту, которая делит вариационный ряд на две части, равные по числу вариант. Если число вариант нечетно, т.е.** $n=2k+1$, **то** $m_e=x_{k+1}$; **при четном** $n=2k$ **медиана**

$$m_e=\frac{\left(x_k+x_{k+1}\right)}{2}.$$

Например, для ряда 2 3 5 6 7 медиана равна 5; для ряда 2 3 5 6 7 9 медиана равна $\frac{(5+6)}{2}=5{,}5.$

Размахом варьирования R называют разность между наибольшей и наименьшей вариантами:

$$R=x_{\max}-x_{\min}\,.$$

Например, для ряда 1 3 4 5 6 10 размах равен 10-1=9.

Размах является простейшей характеристикой рассеяния вариационного ряда.

Средним абсолютным отклонением θ называют среднее арифметическое абсолютных отклонений:

$$\theta=\frac{\sum n_i\left|x_i-\bar{x}_B\right|}{\sum n_i}\,.$$

Например, для ряда

x_i	1	3	6	16
n_i	4	10	5	1

Имеем

$$\bar{x}_B = \frac{4\cdot 1 + 10\cdot 3 + 5\cdot 6 + 1\cdot 16}{4+10+5+1} = \frac{80}{20} = 4;$$

$$\theta = \frac{4\cdot|1-4| + 10\cdot|3-4| + 5\cdot|6-4| + 1\cdot|16-4|}{20} = 2{,}2.$$

Среднее абсолютное отклонение служит для характеристики рассеяния вариационного ряда.

Коэффициентом вариации V называют выраженное в процентах отношение выборочного среднего квадратического отклонения к выборочной средней:

$$V = \frac{\sigma_B}{\bar{x}_B}\cdot 100\,\%.$$

Коэффициент вариации служит для сравнения величин рассеяния по отношению к выборочной средней двух вариационных рядов: тот из рядов имеет большее рассеяние по отношению к выборочной средней, у которого коэффициент вариации больше. Коэффициент вариации – безразмерная величина, поэтому он пригоден для сравнения рассеяний вариационных рядов, варианты которых имеют различную размерность.

Условные варианты

Предположим, что варианты выборки расположены в возрастающем порядке, т.е. в виде вариационного ряда.

Равноотстоящими называют варианты, которые образуют арифметическую прогрессию с разностью h.

Условными называют варианты, определяемые равенством

$$u_i = \frac{(x_i - C)}{h},$$

где C – ложный нуль; h – шаг, т.е. разность между любыми двумя соседними первоначальными вариантами.

Условные варианты используют для упрощенного вычисления числовых характеристик выборки.

Замечание. В качестве ложного нуля можно принять любую варианту. Максимальная простота вычислений достигается, если выбрать в качестве ложного нуля варианту, которая расположена примерно в середине вариационного ряда (часто такая варианта имеет наибольшую частоту). Варианте, которая принята в качестве ложного нуля, соответствует условная варианта, равная нулю.

Пример. Найти условные варианты статистического распределения:

варианта 23,6 28,6 33,6 38,6 43,6

частота 5 20 50 15 10

Решение. Выберем в качестве ложного нуля варианту 33,6 (эта варианта расположена в середине вариационного ряда).

Найдем шаг: $h = 28{,}6 - 23{,}6 = 5.$

Найдем условную варианту:

$$u_1 = \frac{(x_1 - C)}{h} = \frac{(23{,}6 - 33{,}6)}{5} = -2.$$

Аналогично получим: $u_2 = -1,\ u_3 = 0,\ u_4 = 1,\ u_5 = 2.$ Мы видим, что условные варианты – небольшие целые числа. Разумеется, оперировать с ними проще, чем с первоначальными вариантами.

Сведение первоначальных вариант к равноотстоящим

На практике, как правило, данные наблюдений не являются равноотстоящими числами. Для того чтобы свести выборку наблюдаемых значений признака к случаю равностоящих вариант существует следующий алгоритм:

1. Интервал, в котором заключены все наблюдаемые значения признака (первоначальные варианты), делят на несколько равных частных интервалов $(x_1;x_i),(x_i;x_j),\ldots,(x_k;x_m)$.
2. Находят середины частичных интервалов, которые и образуют последовательность равностоящих вариант

$$y_1=\frac{x_1+x_i}{2};\ y_2=\frac{x_i+x_j}{2};\ldots y_r=\frac{x_k+x_m}{2}.$$

В качестве частоты каждой «новой» варианты принимают общее число первоначальных вариант, попавших в соответствующий частичный интервал.

$$n_1'=n_1+n_2+\ldots+n_{i-1}+\frac{n_i}{2};$$

$$n_2'=\frac{n_i}{2}+n_{i+1}+\ldots+n_{j-1}+\frac{n_j}{2};$$

$$n_r'=\frac{n_k}{2}+n_{k+1}+\ldots+n_{m-1}+\frac{n_m}{2}.$$

Обычные, начальные, центральные, условные эмпирические моменты

Для вычисления сводных характеристик удобно пользоваться эмпирическими моментами, которые вычисляют по данным наблюдений.

Обычным эмпирическим моментом порядка k называют среднее значение *k*-х степеней разностей x_i-C:

$$M_k'=\frac{\sum n_i(x_i-C)^k}{n},$$

где x_i - наблюдаемая варианта; n_i - частота варианты; $n=\sum n_i$ - объем выборки; *C* – ложный нуль.

Начальным эмпирическим моментом порядка k называют обычный момент порядка *k* при *C*=0

$$M_k=\frac{\sum n_i x_i^k}{n}.$$

В частности,

$$M_1 = \frac{\sum n_i x_i}{n} = \bar{x}_B,$$

т.е. начальный эмпирический момент первого порядка равен выборочной средней.

Центральным эмпирическим моментом порядка k называют обычный момент порядка k при $C = \bar{x}_B$

$$m_k = \frac{\sum n_i \left(x_i - \bar{x}_B\right)^k}{n}.$$

В частности,

$$m_2 = \frac{\sum n_i \left(x_i - \bar{x}_B\right)^2}{n} = D_B,$$

т.е. центральный эмпирический момент второго порядка равен выборочной дисперсии.

Вычисление центральных моментов требует довольно громоздких вычислений. Чтобы упростить расчеты, заменяют первоначальные варианты условными.

Условным эмпирическим моментом порядка k называют начальный момент порядка k, вычисленный для условных вариант:

$$M_k^* = \frac{\sum n_i u_i^k}{n} = \frac{\sum n_i \left(\frac{x_i - C}{h}\right)^k}{n}.$$

В частности,

$$M_1^* = \frac{\sum n_i \left(\frac{x_i - C}{h}\right)}{n} = \frac{1}{h}\left[\frac{\sum n_i x_i}{n} - C\frac{\sum n_i}{n}\right] = \frac{1}{h}\left(\bar{x}_B - C\right).$$

Отсюда

$$\bar{x}_B = M_1^* h + C.$$

Таким образом, для того чтобы найти выборочную среднюю, достаточно вычислить условный момент первого порядка, умножить его на h и к результату прибавить ложный нуль C.

Для вычисления выборочной дисперсии можно воспользоваться формулой

$$D_B=\left[M_2^*-\left(M_1^*\right)^2\right]\cdot h^2.$$

Действительно:

$$\left[M_2^*-\left(M_1^*\right)^2\right]\cdot h^2=\left[\frac{\sum n_i\left(\frac{x_i-C}{h}\right)^2}{n}-\left(\frac{\sum n_i\left(\frac{x_i-C}{h}\right)}{n}\right)^2\right]\cdot h^2=$$

$$=\frac{\sum n_i\overline{x_i^2}}{n}-2C\frac{\sum n_i x_i}{n}+C^2\frac{\sum n_i}{n}-\left(\frac{\sum n_i x_i}{n}\right)^2+2C\frac{\sum n_i x_i}{n}-C^2\frac{\sum n_i}{n}=$$

$$=\overline{x^2}-\left(\overline{x}\right)^2=D_B.$$

Метод произведений для вычисления выборочных средней и дисперсии

Цель метода заключается в нахождении условных эмпирических моментов и с их помощью D_B и $\overline{x}_B$.

Алгоритм метода

1. Составляется таблица, в первый столбец которой записывают выборочные вари-
анты, располагая их в возрастающем порядке.
2. Во второй столбец записывают частоты вариант; и их сумму (объем выборки n)
помещают в нижнюю клетку столбца.
3. В третий столбец записывают условные варианты $u_i=\frac{(x_i-C)}{h}$, причем в каче-
стве ложного нуля C выбирают варианту, которая расположена примерно в середине вариационного ряда.
4. Умножают частоты на условные варианты и записывают их произведения n_iu_i в
четвертый столбец; сложив все полученные числа, их сумму $\sum n_iu_i$ помещают в нижнюю клетку столбца.

5. Умножают частоты на квадраты условных вариант и записывают их произведе-

ния $n_i u_i^2$ в пятый столбец; сложив все полученные числа, их сумму $\sum n_i u_i^2$ помещают в нижнюю клетку столбца.

6. Умножают частоты на квадраты условных вариант, увеличенных каждая на еди-

ницу, и записывают произведения $n_i(u_i+1)^2$ в шестой контрольный столбец; сложив все полученные числа, их сумму $\sum n_i(u_i+1)^2$ помещают в нижнюю клетку столбца.

7. На основе данных таблицы вычисляют условные моменты: первого и второго

порядка:

$$M_1^* = \frac{\sum n_i u_i}{n}, \quad M_2^* = \frac{\sum n_i u_i^2}{n}.$$

8. Вычисляют выборочную среднюю и дисперсию по формулам

$$\bar{x}_B = M_1^* h + C, \quad D_B = \left[M_2^* - \left(M_1^* \right)^2 \right] \cdot h^2.$$

Замечание. Шестой столбец служит для контроля вычислений: если сумма $\sum n_i(u_i+1)^2$ окажется равной сумме $\sum n_i u_i^2 + 2\sum n_i u_i + n$ (как и должно быть в соответствии с тождеством $\sum n_i(u_i+1)^2 = \sum n_i u_i^2 + 2\sum n_i u_i + n$), то вычисления проведены правильно.

Точность оценки, доверительная вероятность. Доверительный интервал

Точечной называют оценку, которая определяется одним числом. Все оценки, рассмотренные выше, точечные. При выборки малого объема точечная оценка может значительно отличаться от оцениваемого параметра, т.е. приводить к грубым ошибкам. По этой причине при наибольшем объеме выборки следует пользоваться интервальными оценками.

Интервальной называют оценку, которая определяется двумя числами – концами интервала. Интервальные оценки позволяют установить точность и надежность оценок.

Пусть найденная по данным выборки статистическая характеристика θ^* служит оценкой неизвестного параметра θ. Ясно, что θ^* тем точнее определяет параметр θ, чем больше абсолютная величина разности $\left|\theta-\theta^*\right|$. Другими словами, если $\delta>0$ и $\left|\theta-\theta^*\right|<\delta$, то чем меньше δ, тем оценка точнее. Таким образом, положительное число δ характеризует ***точность оценки***.

Однако статистические методы не позволяют категорически утверждать, что оценка θ^* удовлетворяет неравенству $\left|\theta-\theta^*\right|<\delta$; можно лишь говорить о вероятности γ, с которой это неравенство осуществляется.

Надежностью (доверительной вероятностью) оценки θ по θ^* называют вероятность γ, с которой осуществляется неравенство $\left|\theta-\theta^*\right|<\delta$.

Обычно надежность оценки задается наперед, причем в качестве γ берут число, близкое к единице. Наиболее часто задают надежность, равную 0,95; 0,99 и 0,999.

Пусть вероятность того, что $\left|\theta-\theta^*\right|<\delta$, равна γ

$$P\left(\left|\theta-\theta^*\right|<\delta\right)=\gamma.$$

Заменив неравенство $\left|\theta-\theta^*\right|<\delta$ равносильным ему двойным неравенством $-\delta<\theta-\theta^*<\delta$ или $\theta^*-\delta<\theta<\theta^*+\delta,$ имеем

$$P\left(\theta^*-\delta<\theta<\theta^*+\delta\right)=\gamma.$$

Это соотношение следует понимать так: вероятность того, что интервал $\left(\theta^* - \delta; \theta^* + \delta\right)$ заключает в себе (покрывает) неизвестный параметр θ, равно γ.

Доверительным называют интервал $\left(\theta^* - \delta; \theta^* + \delta\right)$, который покрывает неизвестный параметр с заданной надежностью γ.

Доверительный интервал для оценки математического ожидания нормального распределения при известном σ

Пусть количественный признак X генеральной совокупности распределен нормально, причем среднее квадратическое отклонение σ этого распределения известно. Требуется оценить неизвестное математическое ожидание a по выборочной средней $\bar{x}_B$. Найдем доверительные интервалы, покрывающие параметр a с надежностью γ.

Пусть дана выборочная совокупность объема n: $x_1, x_2, \ldots, x_n$. Так как количественный признак X генеральной совокупности распределен нормально, то элементы выборочной совокупности можно рассматривать как независимые случайные величины: $X_1, X_2, \ldots, X_n$ с одним и тем же законом распределения – нормальным, с параметрами $M(X_i) = a, \sigma(X_i) = \sigma$.

Воспользуемся доказательством следующей теоремы.

Теорема. Числовые характеристики среднего арифметического одинаково распределенных взаимно независимых случайных величин вычисляются по формулам

$$M(\overline{X}) = a;\ D(\overline{X}) = \frac{D}{n};\ \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}.$$

Доказательство

$$M(\overline{X})=M\,\frac{(X_1+X_2+\ldots+X_n)}{n}=\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}=\frac{na}{n}=a.$$
$$D(\overline{X})=D\,\frac{(X_1+X_2+\ldots+X_n)}{n}=\frac{D(X_1)+D(X_2)+\ldots+D(X_n)}{n^2}=\frac{nD}{n^2}=\frac{D}{n}.$$

Очевидно, что $\sigma(\overline{X})=\frac{\sigma}{\sqrt{n}}$.

Потребуем, чтобы выполнялось соотношение
$$P\left(\left|\overline{X}-a\right|<\delta\right)=\gamma,$$

где γ - заданная надежность.

Для нормально распределенной случайной величины имеет место теорема
$$P\left(\left|X-a\right|<\delta\right)=2\Phi\left(\frac{\delta}{\sigma}\right),$$

заменив X на $\overline{X}$ и σ на $\sigma(\overline{X})=\frac{\sigma}{\sqrt{n}}$, получим
$$P\left(\left|\overline{X}-a\right|<\delta\right)=2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)=2\Phi(t),$$

где $t=\frac{\delta\sqrt{n}}{\sigma}$.

Найдя из последнего равенства $\delta=\frac{t\sigma}{\sqrt{n}}$, можно записать
$$P\left(\left|\overline{X}-a\right|<\frac{t\sigma}{\sqrt{n}}\right)=2\Phi(t).$$

Приняв во внимание, что вероятность P задана и равна γ, окончательно имеем
$$P\left(\bar{x}-\frac{t\sigma}{\sqrt{n}}<a<\bar{x}+\frac{t\sigma}{\sqrt{n}}\right)=2\Phi(t)=\gamma.$$

Таким образом, с надежностью γ можно утверждать, что доверительный интервал $\left(\bar{x}-\frac{t\sigma}{\sqrt{n}};\bar{x}+\frac{t\sigma}{\sqrt{n}}\right)$ покрывает неизвестный параметр a; точность оценки $t=\frac{\delta\sqrt{n}}{\sigma}$.

Замечание. Число t определяется из равенства $2\Phi(t)=\gamma$ или $\Phi(t)=\frac{\gamma}{2}$; по таблице функций Лапласа (см. приложение 3) находят аргумент t, которому соответствует значение функции Лапласа, равное $\frac{\gamma}{2}$.

Проведем анализ параметров δ, γ и n.

1. При возрастании объема выборки n число δ убывает и, следовательно, точность
оценки увеличивается.

2. Увеличение надежности оценки $\gamma = 2\Phi(t)$ приводит к увеличению t ($\Phi(t)$ возрас-
тающая функция), следовательно, и к возрастанию δ; другими словами, увеличение надежности классической оценки влечет за собой уменьшение ее точности.

3. Если требуется оценить математическое ожидание с наперед заданной точно-
стью δ и надежность γ, то минимальный объем выборки, который обеспечит эту точность, находят по формуле

$$n = \frac{t^2\sigma^2}{\delta^2}$$

(следствие равенства $\delta = \frac{t\sigma}{\sqrt{n}}$).

Доверительный интервал для оценки математического ожидания нормального распределения при неизвестном σ

Пусть количественный признак X генеральной совокупности распределен нормально, причем среднее квадратическое отклонение σ неизвестно. Требуется оценить неизвестное математическое ожидание a с помощью доверительных интервалов.

Оказывается, что по данным выборки можно построить случайную величину (ее возможные значения будем обозначать через t):

$$T = \frac{\overline{X} - a}{s/\sqrt{n}},$$

которая имеет распределение Стьюдента с $k = n - 1$ степенями свободы (см. пояснение в конце параграфа); здесь $\overline{X}$ –

выборочная средняя, s – «исправленное» среднее квадратическое отклонение, n – объем выборки.

Введем новую точность оценки, обозначив ее t_γ (параметр находится по заданным n и γ, приложение 4), тогда

$$P\left(\left|\frac{\overline{X}-a}{s/\sqrt{n}}\right| < t_\gamma\right) = \gamma.$$

Заменив неравенство в круглых скобках равносильным ему двойным неравенством, получим

$$P\left(\overline{x}-t_\gamma\frac{s}{\sqrt{n}} < a < \overline{x}+t_\gamma\frac{s}{\sqrt{n}}\right) = \gamma.$$

Итак, пользуясь распределением Стьюдента, мы нашли доверительный интервал $\left(\overline{x}-t_\gamma\frac{s}{\sqrt{n}};\overline{x}+t_\gamma\frac{s}{\sqrt{n}}\right)$, покрывающий неизвестный параметр a с надежностью γ.

Пояснение. Если Z – нормальная величина, причем $M(Z)=0$, $\sigma(Z)=1$, а V – независимая от Z величина, распределенная по закону χ^2 с k степенями свободы, то величина

$$T = \frac{Z}{\sqrt{V/k}}$$

распределена по закону Стьюдента с k степенями свободы.

Положим случайная величина

$$Z = \frac{\overline{x}_B - a}{\sigma/\sqrt{n}},$$

также имеет нормальное распределение как линейная функция нормального аргумента $\overline{x}_B$ причем $M(Z)=0$, $\sigma(Z)=1$.

Известно, что случайные величины Z и $V = \frac{(n-1)S^2}{\sigma^2}$ независимы и что величина V распределена по закону χ^2 с $k=n-1$ степенями свободы.

Следовательно, выразив величину T, получим

$$T = \frac{\left(\overline{x}_B - a\right)\sqrt{n}}{S},$$

которая распределена по закону Стьюдента с k=n-1 степенями свободы.

Доверительные интервалы для оценки среднего квадратического отклонения σ нормального распределения

Пусть количественный признак X генеральной совокупности распределен нормально. Требуется оценить неизвестное генеральное среднее квадратическое отклонение σ по «исправленному» выборочному среднему квадратическому отклонению s. Поставим перед собой задачу найти доверительные интервалы, покрывающие параметр σ с заданной надежностью γ.

Потребуем, чтобы выполнялось соотношение $P(|\sigma - s| < \delta) = \gamma$ или $P(s - \delta < \sigma < s + \delta) = \gamma.$

Для того чтобы можно было пользоваться готовой таблицей, преобразуем двойное неравенство

$$s - \delta < \sigma < s + \delta$$

в равносильное неравенство

$$s\left(1 - \frac{\delta}{s}\right) < \sigma < s\left(1 + \frac{\delta}{s}\right).$$

Положив $\frac{\delta}{s} = q$, получим

$$s(1 - q) < \sigma < s(1 + q).$$

Параметр q определяется соответствующими n и γ по приложению 5.

Пример. Количественный признак X генеральной совокупности распределен нормально. По выборке объема n=25 найдено «исправленное» среднее квадратическое отклонение s=0,8. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью 0,95.

Решение. По таблице приложения 5 по данным $\gamma = 0{,}95$ и $n = 25$ найдем $q = 0{,}32$.

Искомый доверительный интервал таков:

$$0{,}8(1-0{,}32)<\sigma<0{,}8(1+0{,}32) \text{ или } 0{,}544<\sigma<1{,}056.$$

Замечание. Выше предполагалось, что $q<1$. Если $q>1$, то неравенство примет вид (учитывая, что $\sigma>0$)

$$0<\sigma<s(1+q).$$

Решение типовых задач

Задача 1. Из большой группы предприятий одной из отраслей промышленности случайным образом отобрано 30, по которым получены показатели основных фондов в млн. руб.: 2; 3; 2; 4; 5; 2; 3; 3; 6; 4; 5; 4; 6; 5; 3; 4; 2; 4; 3; 3; 5; 4; 6; 4; 5; 3; 4; 3; 2; 4.

1. Составить дискретное статистическое распределение выборки.
2. Найти объем выборки.
3. Составить распределение относительных частот.
4. Построить полигон частот.
5. Составить эмпирическую функцию распределения и построить ее график.
6. Найти несмещенные оценки числовых характеристик случайной величины.

Решение

1. Расположим различные значения признака в порядке их возрастания и под каж-

дым из них запишем их частоты. Получим дискретное статистическое распределение выборки:

x_i	2	3	4	5	6
n_i	5	8	9	5	3

где x_i - варианты, n_i - частоты вариант x_i.

2. Сумма частот всех вариант должна быть равной объему выборки.

В данном примере объем выборки равен: n=5 + 8 + 9 + 5 + 3=30.

3. Найдем относительные частоты:

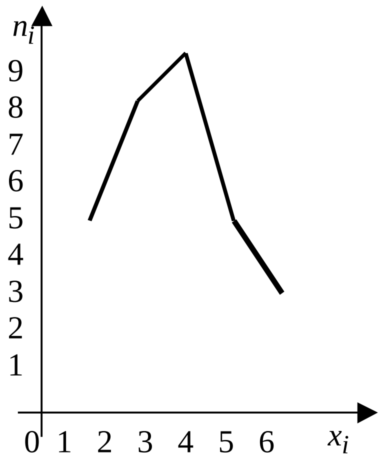
$$W_1 = \frac{5}{30} = \frac{1}{6};\ W_2 = \frac{8}{30} = \frac{4}{15};\ W_3 = \frac{9}{30} = \frac{3}{10};\ W_4 = \frac{5}{30} = \frac{1}{6};\ W_5 = \frac{3}{30} = \frac{1}{10}.$$

Запишем искомое распределение относительных частот

x_i	2	3	4	5	6
W_i	$\frac{1}{6}$	$\frac{4}{15}$	$\frac{3}{10}$	$\frac{1}{6}$	$\frac{1}{10}$

Контроль: $\frac{1}{6} + \frac{4}{15} + \frac{3}{10} + \frac{1}{6} + \frac{1}{10} = 1.$

4. Строим точки с координатами (x_i, n_i) и соединяем их



последовательно отрезками. Полученная ломаная линия называется полигоном частот:

5. Согласно определению эмпирической функцией распределения называется

функция вида

$$F^*(x) = \frac{n_x}{n},$$

где n – объем выборки; n_x - сумма частот вариант, меньших x.

Эмпирическая функция является оценкой функции распределения генеральной совокупности. Наименьшая варианта равна 2, поэтому при $x \le 2, n_x = 0$ и $F^*(x) = 0.$ Значение X<3, а именно, $X = x_1 = 2$ наблюдалось 5 раз. Тогда для $2 < x \le 3$ $n_x = 5$ и $F^*(x) = \frac{5}{30}.$ Значение X<4, а именно, X=2, X=3, наблюдалось 5 + 8 =13 раз. Поэтому для $3 < x \le 4$ $n_x = 13$ и $F^*(x) = \frac{13}{30}.$ Аналогично рассуждая, получаем: для

$4 < x \le 5$ $n_x = 5 + 8 + 9 = 22$ и $F^*(x) = \frac{22}{30}$, для

$5 < x \le 6$ $n_x = 5 + 8 + 9 + 5 = 27$ и $F^*(x) = \frac{27}{30}$

и при $x > 6$ $n_x = 5 + 8 + 9 + 5 + 3 = 30$ и $F^*(x) = \frac{30}{30} = 1.$

Таким образом,

$$F^*(x) = \begin{cases} 0 \text{ при } & x \le 2; \\ \frac{5}{30} \text{ при } & 2 < x \le 3; \\ \frac{13}{30} \text{ при } & 3 < x \le 4; \\ \frac{22}{30} \text{ при } & 4 < x \le 5; \\ \frac{27}{30} \text{ при } & 5 < x \le 6; \\ 1 \text{ при } & x > 6. \end{cases}$$

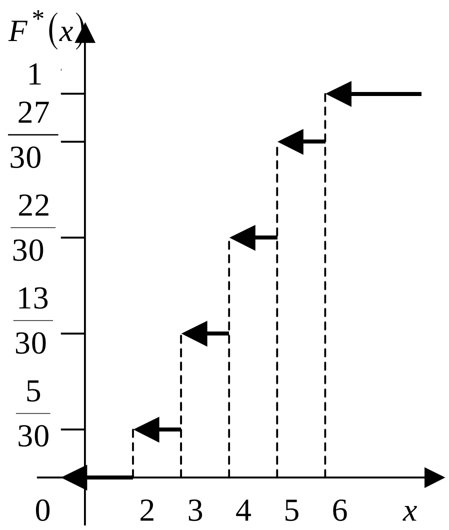


График эмпирической функции имеет вид:

6. Несмещенной оценкой математического ожидания является средняя выборочная:

$$\bar{x}_B = \frac{\sum\limits_{i=1}^{k} x_i n_i}{n} = \frac{2\cdot 5 + 3\cdot 8 + 4\cdot 9 + 5\cdot 5 + 6\cdot 3}{30} = \frac{113}{30} \approx 3{,}77.$$

Несмещенная оценка дисперсии – исправления выборочная дисперсия:

$$s^2 = \frac{n}{n-1} D_B.$$

$$D_B = \overline{x_B^2} - \left(\bar{x}_B\right)^2 = \frac{2^2\cdot 5 + 3^2\cdot 8 + 4^2\cdot 9 + 5^2\cdot 5 + 6^2\cdot 3}{30} - \left(\frac{113}{30}\right)^2 \approx 1{,}42.$$

$$s^2 = \frac{30}{29}\cdot 1{,}42 \approx 1{,}47.$$

Задача 2. Выборочно обследование 30 предприятий машиностроительной промышленности по валовой продукции и получены следующие данные, в млн. руб.:

18,0; 12,0; 11,9; 1,9; 5,5; 14,6; 4,8; 5,6; 4,8; 10,9; 9,7; 7,2; 12,4; 7,6;

9,7; 11,2; 4,2; 4,9; 9,6; 3,2; 8,6; 4,6; 6,7; 8,4; 6,8; 6,9; 17,9; 9,6;

14,8; 15,8.

Составить интервальное распределение выборки с началом $x_0 = 1$ и длиной частичного интервала $h = 3$. Построить гистограмму частот.

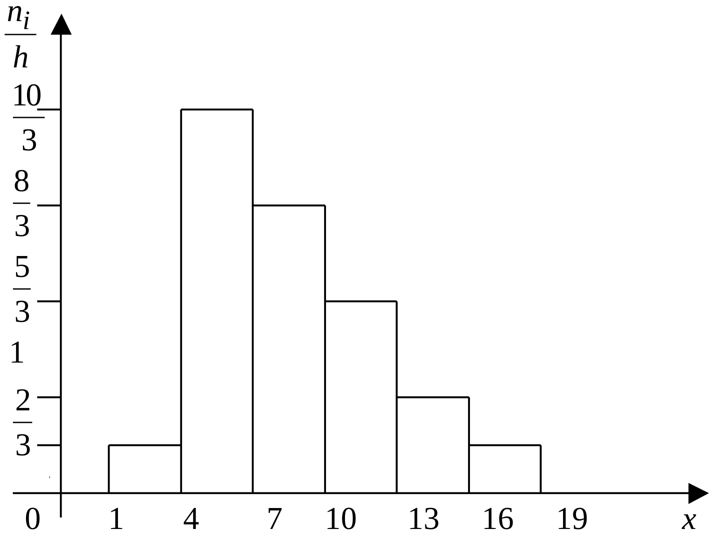
Решение. Для составления интервального распределения составим таблицу, в первой строке которой расположим в порядке возрастания интервалы, длина каждого из которых $h = 3$. Во второй строке запишем количество значений признака в выборке, попавших в этот интервал (т.е. сумму частот вариант, попавших в соответствующий интервал):

$(x_i; x_{i+1})$	1-4	4-7	7-10	10-13	13-16	16-19

n_i	2	10	8	5	3	2

Объем выборки $n = 2 + 10 + 8 + 5 + 3 + 2 = 30$.

Для построения гистограммы частот на оси абсцисс откладываем частичные интервалы, на каждом из них строим прямоугольники высотой $\frac{n_i}{h}$, где n_i – частота i-го частичного интервала, h – шаг (длина интервала), таким образом, гистограмма примет вид:



Указание. Для построения эмпирической функции распределения и нахождения точечных оценок ряда необходимо преобразовать его к дискретному виду по формуле

$$x_i^* = \frac{x_i + x_{i+1}}{2}.$$

Получим

x_i^*	2,5	5,5	8,5	11,5	14,5	17,5
n_i	2	10	8	5	3	2

Задача 3. Из большой партии электроламп случайным образом отобрано 100. Средняя продолжительность горе-

ния ламп в выборке оказалась равной 1000 ч. Найти с надежностью $\gamma = 0{,}95$ доверительный интервал для средней продолжительности a горения ламп во всей партии, если известно, что среднее квадратическое отклонение продолжительности горения лампы $\sigma = 40$ ч и продолжительность горения ламп распределена по нормальному закону.

Решение. По условию $\bar{x}_B = 1000$, $\gamma = 0{,}95$, $\sigma = 40$. Для решения воспользуемся формулой

$$\bar{x}_B - \frac{\sigma t}{\sqrt{n}} < a < \bar{x}_B + \frac{\sigma t}{\sqrt{n}}$$

По приложению 3 находим t из условия:

$$\Phi(t) = \frac{\gamma}{2} = \frac{0{,}95}{2} = 0{,}475; \Rightarrow t = 1{,}96.$$

Тогда доверительный интервал:

$$1000 - \frac{1{,}96 \cdot 40}{\sqrt{100}} < a < 1000 + \frac{1{,}96 \cdot 40}{\sqrt{100}}$$

$$992{,}16 < a < 1007{,}84.$$

Задачи (71 – 80)

В задачах 71 – 80 выборочные совокупности заданы из соответствующих генеральных совокупностей. Требуется:

1. Составить интервальное распределения выборки с шагом h, взяв за начало первого интервала x_0.
2. Построить гистограмму частот.
3. Найти $\bar{x}_B; D_B; \sigma_B; S.$
4. Найти с надежностью γ доверительный интервал для оценки неизвестного математического ожидания признака X генеральной совокупности, если признак X распределен по нормальному закону и его среднее квадратическое отклонение равно $\sigma_Г$.

71. Произведено выборочное обследование 25 магазинов по величине товарооборота.

Получены следующие результаты (в тыс. руб.):

42,5 60,0 63,5 70,5 82,0 83,5 92,0 95,5 100,0
101,0 105,0 108,5 110,0 115,5 120,0 120,5 122,0
130,0 138,5 140,0 142,0 150,5 160,0 162,1 180,5

$\gamma = 0{,}96;\ \sigma_{\Gamma} = 31;\ h = 20;\ x_0 = 42{,}5.$

72. Объем промышленного производства Российской Федерации за период с 1999-2000

годы составил (млрд. руб.) в месяц:

187,6 189,8 223,0 223,2 213,2 228,6 242,3 252,7
271,2 293,7 311,8 358,1
331,7 350,8 387,5 359,2 361,1 384,5 391,6 407,7
417,6 442,7 451,9 476,2

$\gamma = 0{,}98;\ \sigma_{\Gamma} = 86{,}63;\ h = 50;\ x_0 = 185.$

73. Темп роста курса акций 25 фирм по сравнению с предыдущим месяцем составил(%)

104 103,1 102 98 99 94 119 114,8 109,5
103,1 92 97,1 95,2 91,7
104 104,5 92,8 95,8 104,9 77,5 93,1 94,9 99,5
99,7 103

$\gamma = 0{,}95;\ \sigma_{\Gamma} = 8{,}05;\ h = 10;\ x_0 = 75.$

74. Получены результаты выборочного обследования по выполнению плана выработки

на одного рабочего (в %):

90,0 96,0 98,0 98,0 98,5 99,0 101,5 102 102,0
102,5 103,0 103,5 104,0 104,0 104 104,5 105,5
106,0 108,0 108,2 108,7 109,0 112 113,5

$\gamma = 0{,}98;\ \sigma_{\Gamma} = 4{,}7\%;\ h = 5;\ x_0 = 90.$

75. Были испытаны 25 ламп на продолжительность горения и получены следующие

результаты (в часах):

773 792 815 827 843 854 861 869 877 886 889 892
885 901 903 905 911 918 919 923 929 937 941 955
981

$\gamma = 0{,}92;\ \sigma_{\Gamma} = 50;\ h = 40;\ x_0 = 760$.

76. Для прогнозирования спроса на свою продукцию предприятие проводит исследова-
ние, в результате которого получены данные о размере реализованной продукции за некоторый период времени (в тыс. руб.):
42,5 60,0 63,5 70,5 82,0 83,5 92,0 95,5 100,0
101,0 105,0 108,5 110,0 115,5 120,0 130,0 138,5 140,0
142,0 150,5 160,0 162,1 180,5

$\gamma = 0{,}96;\ \sigma_{\Gamma} = 31;\ h = 30;\ x_0 = 40$.

77. В течение 25 лет наблюдался подъем уровня воды в реке во время паводков. Полу-
чены следующие значения (в см.):
266 278 315 336 347 354 368 369 391 408 411 416
427 444 448 457 462 481 483 495 512 536 576

$\gamma = 0{,}96;\ \sigma_{\Gamma} = 65;\ h = 50;\ x_0 = 20$.

78. На предприятии было произведено выборочное обследование заработной платы
рабочих и получены следующие результаты (в руб.)
1360 1550 1600 1690 1750 1750 1800 1880 1890 1920
1950 2000 2020 2050 2050 2050 20801 2120 2150 2200
2250 2340 2420 2450 2600

$\gamma = 0{,}95;\ \sigma_{\Gamma} = 300$ руб; $h = 200$; $x_0 = 1300$.

79. Для определения себестоимости строительно-монтажных работ было произведено
выборочное обследование 25 стороительно-монтажных управлений и получены следующие результаты (тыс. руб.)
1250 1450 1550 1700 1760 1820 1880 1960 2100
2175 2190 2200 2220 2275 2280 2310 2400 2550
2580 2600 2670 2800 2950 3000 3075

$\gamma = 0{,}94;\ \sigma_{\Gamma} = 446$ руб; $h = 400$; $x_0 = 1100$.

80. В районной сберегательной кассе проведено выборочное обследование 25 вкладов,

которое дало следующие результаты (в руб.):

750	2100	3500	3500	4000	5200	5400	5600	5900
6800	7000	7000	7200	7500	7800	7900	8100	8500
8750	8900	9000	10000	11000	12000	12500		

$\gamma = 0{,}95$; $\sigma_{\Gamma} = 2800$ руб; $h = 2000$; $x_0 = 500$.

3.2. Элементы теории корреляции

Функциональная, статистическая и корреляционная зависимости

Зависимость между переменными X и Y называется ***функциональной***, если существует функция $y=f(x)$, по которой каждому значению $x \in X$ ставится в соответствии единственное значение $y \in Y$.

Однако не всякую зависимость между X и Y можно представить в виде функции. Иногда одному фиксированному значению X соответствует множество значений Y.

Статистической называют зависимость, при которой изменение одной из величин влечет изменение распределения другой. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой; в этом случае статистическую зависимость называют ***корреляционной***.

Другими словами, корреляционной зависимостью признака Y от X называют функциональную зависимость ***условного среднего*** $\bar{y}_x$ от X , т.е. $\bar{y}_x = f(x)$, где $\bar{y}_x$ – среднее арифметическое наблюдавшихся значений Y, соответствующих $X = x$. Например, если при $x_1 = 2$ величина Y приняла значения $y_1 = 5, \quad y_2 = 6, \quad y_3 = 10,$ то условное среднее $\bar{y}_{x_1} = \frac{(5+6+1)}{3} = 4.$

Аналогично определяется условное среднее $\bar{x}_y$.

Условным средним $\bar{x}_y$ называется среднее арифметическое наблюдавшихся значений *X*, соответствующих *Y* = *y*.

Уравнение $\bar{y}_x = f(x),$ называется ***выборочным уравнением регрессии*** *Y* на *X*; функцию *f(x)* называют ***выборочной регрессией*** *Y* на *X* , а ее график – ***выборочной линией регрессии*** *Y* на *X*. Аналогично уравнение $\bar{x}_y = \varphi(y)$ называется ***выборочным уравнением регрессии*** *X* на *Y*; функцию $\varphi(x)$ называют ***выборочной регрессией*** *X* на *Y* , а ее график – ***выборочной линией регрессии*** *X* на *Y*.

Основные задачи теории корреляции

1. Установить форму корреляционной связи, т.е. вид функции регрессии (линейная, квадратическая , показательная и пр.).

В случае если обе функции *f(x)* и $\varphi(x)$ линейны, то корреляцию называют линейной, в противном случае нелинейной.

2. Оценить тесноту (силу) корреляционной связи.

Теснота корреляционной связи между *X* и *Y* оценивается по величине рассеяния значений *Y* вокруг условного среднего $\bar{y}_x$. Большое рассеяние свидетельствует о слабой зависимости *Y* от *X* либо об отсутствии этой зависимости. Малое рассеяние указывает на наличие достаточно сильной зависимости; возможно даже, что *X* и *Y* связаны функционально.

Аналогично (по величине рассеяния значений *X* вокруг среднего $\bar{x}_y$) оценивают тесноту корреляционной связи *Y* от *X*.

Корреляционная таблица

Будем считать, что объекты выборочной совокупности характеризуются парой признаков X и Y ,т.е. каждому объекту соответствует пара чисел $(x; y)$. Статистическую совокупность будем обозначать Ω.

Пусть каждому объекту с признаками $(x; y)$ из совокупности Ω поставлена в соответствие точка плоскости XOY с координатами $(x; y)$. Полученное множество точек называется ***диаграммой рассеяния*** статистической совокупности Ω.

Диаграмма рассеяния, как и график функции, дает наглядное представление о статистической совокупности, о виде зависимости между факторами, о тесноте связи.

Если Ω - конечная совокупность объема n, то ее можно описать корреляционной таблицей

$X \backslash Y$	y_1	…	y_j	…	y_l	n_x
x_1	n_{11}	…	n_{1j}	…	n_{1l}	n_{x1}
⋮	⋮	⋮		⋮		⋮
x_i	n_{i1}	…	n_{ij}	…	n_{il}	n_{xi}
⋮	⋮	⋮		⋮		⋮
x_k	n_{k1}	…	n_{kj}	…	n_{kl}	n_{xk}
n_y	n_{y1}	…	n_{yj}	…	n_{yl}	n

где $x_i\left(i = \left(\overline{1,k}\right)\right)$, $y_j\left(j = \left(\overline{1,l}\right)\right)$ – соответственно значения признаков X и Y; n_{xi}, n_{yj} - соответствующие им частоты; n_{ij} – частота, с которой встречается пара (x_i, y_j). По определению,

$n_{xi} = \sum\limits_{j=1}^{l} n_{ij}, n_{yj} = \sum\limits_{i=1}^{k} n_{ij}$. Из таблицы вытекают следующие равенства для объема выборки n: $n = \sum\limits_{i=1}^{k} n_{xi} = \sum\limits_{j=1}^{l} n_{yj} = \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{l} n_{ij}$.

По опытным данным, приведенным в корреляционной таблице, можно судить о форме корреляционной связи между признаками X и Y. С этой целью находятся условные средние $\bar{y}_{xi}$, соответствующие значениям $x_i = \left(i = \overline{1,k}\right)$, и $\bar{x}_{yj}$, соответствующие значениям $y_j = \left(j = \overline{1,l}\right)$ по формулам

$$\bar{y}_{xi} = \frac{\sum\limits_{j=1}^{l} y_j n_{ij}}{n_{xi}}; \quad \bar{x}_{yj} = \frac{\sum\limits_{i=1}^{k} x_i n_{ij}}{n_{yi}}.$$

Эмпирической линией регрессии Y на X (X на Y) называют ломаную линию, соединяющую отрезками точки с координатами $M_i^*\left(x_i; \bar{y}_{xi}\right)$ $\left(M_j^*\left(\bar{x}_{yj}; y_j\right)\right)$.

Форма полученной таким образом эмпирической ломаной является прообразом формы теоретической зависимости.

Уравнение прямой линии регрессии

Пусть требуется найти теоретическое уравнение $\bar{Y}_x = f(x,\ a_1,\ \ldots, a_n)$
регрессии Y на X. Параметры $a_1, \ldots, a_n$ этого уравнения находятся методом "наименьших квадратов". При предполагаемом законе функциональной зависимости f, коэффициенты $a_1, \ldots, a_n$ выбирают «наилучшим» образом так, чтобы величина $\left|\bar{Y}_{xi} - \bar{y}_{xi}\right|$ была наименьшей. Данная величина определяет расстояние от точек $M_i\ \left(x_i; \bar{Y}_{xi}\right)$, лежащих на предполагаемой теоретической кривой до угловых точек $M_i^*\left(x_i; \bar{y}_{xi}\right)$ эмпирической кривой.

Допустим, что количественные признаки X и Y связаны линейной корреляционной зависимостью. В этом случае обе линии регрессии будут прямыми. Тогда $f(x, a_1, \ldots, a_n) = a_1 x + a_2$, а теоретической кривой Y на X будет прямая

$$\overline{Y}_x = a_1 x + a_2 \quad . \tag{*}$$

Найдем параметры a_1 и a_2 так, чтобы точки M_i^*, построенные по данным наблюдений на плоскости XOY как можно ближе лежали к теоретической прямой.

Рассмотрим разность $Y_i - y_i \left(i = \left(\overline{1, n} \right) \right)$, где Y_i - ордината, соответствующая x_i, вычисляется по равенству (*), y_i - наблюдаемая ордината, соответствующая значению x_i.

Составим функцию, зависящую от параметров a_1 и a_2:

$$F(a_1; a_2) = \sum_{i=1}^{n} (Y_i - y_i)^2 = \sum_{i=1}^{n} \left(a_1 x_i + a_2 + y_i \right)^2 .$$

Для отыскания минимума данной функции приравняем к нулю частные производные:

$$F'_{a_1} = 2 \sum_{i=1}^{n} (a_1 x_i + a_2 - y_i) x_i = 0;$$

$$F'_{a_2} = 2 \sum_{i=1}^{n} (a_1 x_i + a_2 - y_i) = 0.$$

Получим систему уравнений

$$\begin{cases} a_1 \sum\limits_{i=1}^{n} x_i^2 + a_2 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i; \\ a_1 \sum\limits_{i=1}^{n} x_i + a_2 n = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

Воспользуемся следующими тождествами:

$$\overline{x} = \frac{\sum x_i}{n} \Rightarrow \sum x_i = \overline{x} \cdot n; \quad \overline{y} = \frac{\sum y_i}{n} \Rightarrow \sum y_i = \overline{y} \cdot n;$$

$$\overline{x^2} = \frac{\sum x_i^2}{n} \Rightarrow \sum x_i^2 = \overline{x^2} \cdot n; \quad \overline{xy} = \frac{\sum x_i y_i}{n} \Rightarrow \sum x_i y_i = \overline{xy} \cdot n .$$

Подставим полученные выражения в систему и разделим каждое уравнение на n

$$\begin{cases} a_1 \overline{x^2} + a_2 \overline{x} = \overline{xy}; \\ a_1 \overline{x} + a_2 = \overline{y}. \end{cases}$$

Решением системы являются коэффициенты:

$$a_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x^2} \text{ и } a_2 = \overline{y} - a_1 \overline{x}.$$

Угловой коэффициент a_1 прямой линии регрессии *Y* на *X* называется ***коэффициентом регрессии*** *Y* на *X* и обычно обозначается $\rho_{yx} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x^2}$. Тогда уравнение (∗) можно записать следующим образом:

$$\overline{Y}_x - \overline{y} = \rho_{yx}\left(x - \overline{x}\right).$$

Аналогично теоретическое уравнение $\overline{X}_y = b_1 y + b_2$ линейной регрессии *X* на *Y* с помощью коэффициента $\rho_{xy} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_y^2}$ приводится к виду

$$\overline{X}_y - \overline{x} = \rho_{xy}\left(y - \overline{y}\right).$$

Сравнивая коэффициенты регрессии *Y* на *X* и *X* на *Y*, можно отметить, что они имеют одинаковые знаки (в силу совпадения числителей и положительности знаменателей).

Выборочный коэффициент корреляции

Выборочным коэффициентом корреляции r_B признаков *X* и *Y* называется число, равное среднему геометрическому коэффициентов регрессии и имеющее их знак:

$$r_B = \pm\sqrt{\rho_{xy}\rho_{yx}} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y}.$$

Уравнения регрессии с помощью коэффициента корреляции примут вид:

$$\overline{Y}_x - \overline{y} = r_B \frac{\sigma_y}{\sigma_x}\left(x - \overline{x}\right);$$

$$\overline{X}_y - \overline{x} = r_B \frac{\sigma_x}{\sigma_y}\left(y - \overline{y}\right).$$

Прямые регрессии пересекаются в точке $(\overline{x};\overline{y})$, которая называется средней точкой корреляционного графика.

Коэффициент корреляции имеет важное самостоятельное значение. С его помощью оценивается теснота (сила) корреляционной связи между признаками. Коэффициент корреляции r_B обладает следующими свойствами:

1. $|r_B| \le 1$ или $-1 \le r_B \le 1$.
2. Условие $|r_B| = 1$ $(r_B = \pm 1)$ является необходимым и достаточным условием существования линейной функциональной зависимости.
3. При $r_B = 0$ линейной корреляционной связи между признаками не существует (при этом может быть нелинейная корреляционная связь и даже нелинейная функциональная зависимость).

Таким образом, чем ближе по модулю коэффициент линейной корреляции к единице, чем теснее линейная зависимость между *X* и *Y*, чем ближе коэффициент корреляции к нулю, тем слабее линейная зависимость.

О тесноте связи можно судить по значению коэффициента корреляции, используя шкалу Чеддока:

Показания тесноты связи	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99
Характеристика силы связи	слабая	умеренная	заметная	высокая	весьма высокая

Если выборка имеет достаточно большой объем $n \geq 50$ и является репрезентативной, то заключение о тесноте связи признаков X и Y может быть распространено на всю генеральную совокупность.

Так, для оценки коэффициента корреляции $r_Г$ нормально распределенной совокупности можно использовать формулу

$$r_B - 3\frac{1-r_B^2}{\sqrt{n}} \leq r_Г \leq r_B + 3\frac{1+r_B^2}{\sqrt{n}}.$$

Уравнения регрессии в случае равноотстоящих значений признаков

В случае, если значения хотя бы одного из признаков являются равностоящими, полезно использовать условные варианты.

Пусть для определенности значения признака X являются равноотстоящими. Тогда расчет основных параметров уравнения регрессии производится по алгоритму:

1. Рассчитываем условные варианты

$$u_i = \frac{x_i - C}{h},$$

где C – ложный нуль, h – шаг.

2. Находим условные эмпирические моменты первого и второго порядка:

$$M_1^* = \frac{\sum u_i n_i}{n} = \overline{u}; \quad M_2^* = \frac{\sum u_i^2 n_i}{n} = \overline{u^2}.$$

3. Находим

$$\overline{x} = M_1^* h + C = \overline{u}h + C;$$

$$\sigma_x = \sqrt{\overline{u^2} - \left(\overline{u}\right)^2} \cdot h = \sigma_u \cdot h.$$

4. Вычисляем $r_B = \dfrac{\overline{uy} - \overline{u} \cdot \overline{y}}{\sigma_u \cdot \sigma_y}$.

Пусть значения обоих признаков *X* и *Y* являются равноотстоящими соответственно с шагом h_1 и h_2. Тогда целесообразно воспользоваться следующим алгоритмом:

1. Переходим к условным вариантам

$$u_i = \frac{x_i - C_1}{h_1}; \; v_i = \frac{y_j - C_2}{h_2}.$$

2. Находим условные эмпирические моменты

$$\overline{u} \text{ и } \overline{v}\;;\; \overline{u^2} \text{ и } \overline{v^2}.$$

3. Находим

$$\overline{x} = \overline{u}h_1 + C_1;\; \sigma_x = \sigma_u \cdot h_1;$$
$$\overline{y} = \overline{v}h_2 + C_2;\; \sigma_y = \sigma_v \cdot h_2$$

4. Коэффициент корреляции определяется по формуле

$$r_B = \frac{\overline{uv} - \overline{u} \cdot \overline{v}}{\sigma_u \cdot \sigma_v}.$$

5. Составляем уравнения регрессии.

Криволинейная корреляция

Между признаками *X* и *Y* могут существовать и нелинейные корреляционные зависимости (параболическая, гиперболическая, показательная и пр.).

Рассмотрим подробнее случаи параболической и гиперболической зависимости. Предположим между признаками *X* и *Y* – параболическая корреляционная связь. Тогда уравнения регрессии имеют вид:

$$\overline{Y}_x = a_1 x^2 + a_2 x + a_3;$$
$$\overline{X}_y = b_1 y^2 + b_2 y + b_3.$$

Основываясь на выше описанном методе «наименьших квадратов», получим следующую систему линейных уравнений для нахождения параметров:

$$\begin{cases} a_1\overline{x^4} + a_2\overline{x^3} + a_3\overline{x^2} = \overline{x^2 y}; \\ a_1\overline{x^3} + a_2\overline{x^2} + a_3\overline{x} = \overline{xy}; \\ a_1\overline{x^2} + a_2\overline{x} + a_3 = \overline{y}. \end{cases}$$

(*)

Решением системы (*) являются «наилучшие» параметры искомой параболы. Для нахождения параметров b_1, b_2, b_3 необходимо составить идентичную систему уравнений.

В случае гиперболической корреляционной зависимости *Y* от *X* уравнения регрессии имеют вид:

$$\overline{Y}_x = \frac{a_1}{x} + a_2; \ \overline{X}_y = \frac{b_1}{y} + b_2.$$

Метод "наименьших квадратов" приводит процесс составления уравнения регрессии к решению следующей системы:

$$\begin{cases} a_1\left(\overline{\frac{1}{x^2}}\right) + a_2\left(\overline{\frac{1}{x}}\right) = \left(\overline{\frac{1}{x}y}\right); \\ a_1\left(\overline{\frac{1}{x}}\right) + a_2 = \overline{y}. \end{cases}$$

Аналогично составляется и решается система уравнений относительно параметров b_1 и b_2.

Для оценки тесноты нелинейной корреляционной связи используют ***выборочные корреляционные отношения***: η_{yx} – выборочное корреляционное отношение *Y* к *X*; η_{xy} – выборочное корреляционное отношение *X* к *Y*.

Выборочным корреляционным отношением *Y* к *X* называют отношение межгруппового среднего квадратического отклонения к общему среднему квадратическому отклонению признака *Y*:

$$\eta_{yx} = \frac{\sigma_{\text{межгр}}}{\sigma_{\text{общ}}}$$

или в других обозначениях

$$\eta_{yx} = \frac{\sigma_{\bar{y}_x}}{\bar{\sigma}_y}.$$

Здесь

$$\sigma_{\bar{y}_x} = \sqrt{\frac{\sum n_x \left(\bar{y}_x - \bar{y}\right)^2}{n}}; \quad \bar{\sigma}_y = \sqrt{\frac{\sum n_y \left(y - \bar{y}\right)^2}{n}},$$

где n – объем выборки (сумма всех частот); n_x - частота значения x признака X; n_y - частота значения y признака Y; $\bar{y}$ - общая средняя признака X; $\bar{y}_x$ - условная средняя признака Y.

Аналогично определяется выборочное корреляционное отношение X к Y:

$$\eta_{xy} = \frac{\sigma_{\bar{x}_y}}{\bar{\sigma}_x}.$$

Свойства выборочного корреляционного отношения

Поскольку η_{xy} обладает тем же свойством, что и η_{yx}, перечислим свойства только выборочного корреляционного отношения η_{yx}, которое далее для упрощения записи будем обозначать через η и для простоты называть «корреляционным отношением».

С в о й с т в о 1. Корреляционное отношение удовлетворяет двойному неравенству

$$0 \le \eta \le 1.$$

С в о й с т в о 2. Если $\eta = 0$, то признак Y с признаком X корреляционной зависимостью не связан.

С в о й с т в о 3. Если $\eta = 1$, то признак Y связан с признаком X функциональной зависимостью.

С в о й с т в о 4. Выборочное корреляционное отношение не меньше абсолютной величины выборочного коэффициента корреляции: $\eta \ge |r_B|$.

С в о й с т в о 5. Если выборочное корреляционное отношение равно абсолютной величине выборочного коэффициента корреляции, то имеет место точная линейная корреляционная зависимость.

Другими словами, если $\eta = |r_B|$, то точки $(x_1; y_1)$, $(x_2; y_2)$, …, $(x_n; y_n)$ лежат на прямой линии регрессии, найденной способом "наименьших квадратов".

Понятие множественной корреляции

Множественная корреляция **–** это исследование связи между несколькими признаками.

Пусть Z линейно зависит от X и Y, тогда уравнение линейной множественной регрессии имеет вид:

$$z = a_1 x + a_2 y + a_0. \quad (*)$$

Коэффициенты множественной регрессии a_1, a_2, и a_0 находятся методом "наименьших квадратов", т.е. так, чтобы функция $F(a_1, a_2, a_0) = \sum_i (a_1 x_i + a_2 y_i + a_0 - z_i)^2 n_i$ имела минимум.

Раскрывая знак суммы и группируя слагаемые, приводим уравнение (*) к виду:

$$z - \bar{z} = a_1(x - \bar{x}) + a_2(y - \bar{y}),$$

причем коэффициенты регрессии определяются равенствами:

$$a_1 = \frac{r_{xz} - r_{yz} \cdot r_{xy}}{1 - r_{xy}^2} \cdot \frac{\sigma_z}{\sigma_x}; \quad a_2 = \frac{r_{yz} - r_{xz} \cdot r_{xy}}{1 - r_{xy}^2} \cdot \frac{\sigma_z}{\sigma_y},$$

где r_{xz}; r_{yz} и r_{xy} – коэффициенты корреляции соответственно между признаками X и Z; Y и Z; X и Y.

Теснота линейной корреляционной связи признака Z с X и Y оценивается с помощью ***выборочного совокупного коэффициента корреляции:***

$$R = \sqrt{\frac{r_{xz}^2 - 2r_{xy} \cdot r_{xz} \cdot r_{yz} + r_{yz}^2}{1 - r_{xy}^2}}.$$

При этом $0 \le R \le 1$ и при приближении R к единице теснота линейной связи Z с X и Y увеличивается.

Следующей задачей множественной корреляции является задача оценить влияние на Z отдельно признака X и отдельно признака Y. Это осуществляется при помощи ***выборочных частных коэффициентов корреляции:***

$$r_{xz(y)} = \frac{r_{xz} - r_{xy} \cdot z_{yz}}{\sqrt{\left(1-r_{xy}^2\right)\left(1-r_{yz}^2\right)}}; \quad r_{yz(x)} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{\left(1-r_{xy}^2\right)\left(1-r_{xz}^2\right)}}.$$

Первый коэффициент оценивает тесноту линейной корреляционной связи между Z и X, когда Y остается постоянным. Теснота связи между Z и Y (при постоянной X) оценивается вторым коэффициентом корреляции $r_{yz(x)}$.

Эти коэффициенты имеют те же свойства, что и обыкновенный выборочный коэффициент корреляции.

Решение типовых задач

Задача 1. Выборочно обследовано 100 заводов по величине основных производственных фондов X (млн. руб.) и объему готовой продукции Y (млн. руб.). Результаты представлены в корреляционной таблице (табл. 1).

Таблица 1

Y	X					n_y
	5	15	25	35	45	
30	7	1				8
32	2	7	1			10
34	1	5	4	1		11
36		1	15	10	8	34
38			3	12	15	30
40				1	6	7
n_x	10	14	23	24	29	n=100

По данным исследования требуется:

1) в прямоугольной системе координат построить эмпирические ломаные регрессии *Y* на *X* и *X* на *Y*;

2) оценить тесноту линейной корреляционной связи;

3) составить линейные уравнения регрессии *Y* на *X* и *X* на *Y* и построить их графики в одной системе координат.

Решение. 1. Так как при $x = 5$ признак *Y* имеет распределение

Y	30	32	34
n_i	7	2	1

то условное среднее $\bar{y}_{x=5} = \dfrac{30\cdot 7 + 32\cdot 2 + 34\cdot 1}{10} = 30{,}8.$

При x=15 признак *Y* имеет распределение

Y	30	32	34	36
n_i	1	7	5	1

Следовательно $\bar{y}_{x=15} = \dfrac{30\cdot 1 + 32\cdot 7 + 34\cdot 5 + 36\cdot 1}{14} = 32{,}86.$

Аналогично вычисляются все условные средние $\bar{y}_x$. В результате получим таблицу, выражающую корреляционную зависимость $\bar{y}$ от *X* (табл. 2).

Таблица 2

X	5	15	25	35	45
$\bar{y}_x$	30,8	32,86	35,74	37,08	37,86

Так как при y=30 признак *X* имеет распределение

X	5	15
n_j	7	1

то условное среднее $\bar{x}_{y=30} = \frac{5 \cdot 7 + 15 \cdot 1}{8} = 6{,}25.$

При $y = 32$ признак X имеет распределение

X	5	15	25
n_j	2	7	1

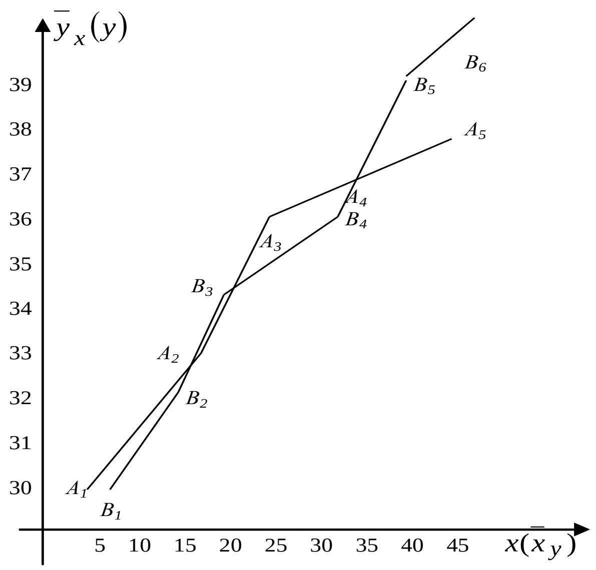
Следовательно $\bar{x}_{y=32} = \frac{5 \cdot 2 + 15 \cdot 7 + 25 \cdot 1}{10} = 14.$

Аналогично вычисляются все $\bar{x}_y$. В результате получим табл. 3.

Таблица 3

Y	30	32	34	36	38	40
$\bar{x}_y$	6,25	14	19,54	32,35	39	43,57

В прямоугольной системе координат построим точки $A_i\left(x_i; \bar{y}_{x_i}\right)$, соединим их отрезками прямых, получим эмпи-



рическую линию регрессии Y на X. Аналогично строятся точки $B_j\left(\bar{x}_y; y_j\right)$ и эмпирическая линия регрессии X на Y.

2. Выдвинув гипотезу о линейной корреляционной зависимости, оценим тесноту связи. Вычислим выборочный коэффициент корреляции

$$r_B = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y},$$

$$\overline{x} = \frac{\sum x_i \cdot n_i}{n}, \quad \overline{y} = \frac{\sum y_j \cdot n_j}{n}, \quad \overline{x^2} = \frac{\sum x_i^2 \cdot n_i}{n}, \quad \overline{y^2} = \frac{\sum y_j^2 \cdot n_j}{n},$$

$$\overline{xy} = \frac{\sum x_i y_j \cdot n_{ij}}{n}, \quad \sigma_x = \sqrt{\overline{x^2} - (\overline{x})^2}, \quad \sigma_y = \sqrt{\overline{y^2} - (\overline{y})^2}$$

$$\overline{x} = \frac{5 \cdot 10 + 15 \cdot 14 + 25 \cdot 23 + 35 \cdot 24 + 45 \cdot 29}{100} = 29{,}8;$$

$$\overline{y} = \frac{30 \cdot 8 + 32 \cdot 10 + 34 \cdot 11 + 36 \cdot 34 + 38 \cdot 30 + 40 \cdot 7}{100} = 35{,}78;$$

$$\overline{x^2} = \frac{5^2 \cdot 10 + 15^2 \cdot 14 + 25^2 \cdot 23 + 35^2 \cdot 24 + 45^2 \cdot 29}{100} = 1059;$$

$$\overline{y^2} = \frac{30^2 \cdot 8 + 32^2 \cdot 10 + 34^2 \cdot 11 + 36^2 \cdot 34 + 45^2 \cdot 30 + 40^2 \cdot 7}{100} = 1287{,}4$$

$$\overline{xy} = \frac{30 \cdot 5 \cdot 7 + 30 \cdot 15 \cdot 1 + 32 \cdot 5 \cdot 2 + 32 \cdot 15 \cdot 7 + 32 \cdot 25 \cdot 1 + 34 \cdot 5 \cdot 1 + 34 \cdot 15 \cdot 5}{100} +$$

$$+ \frac{34 \cdot 25 \cdot 4 + 34 \cdot 35 \cdot 1 + 36 \cdot 15 \cdot 1 + 36 \cdot 25 \cdot 15 + 36 \cdot 35 \cdot 10 + 36 \cdot 45 \cdot 8 + 38 \cdot 25 \cdot 3}{100} +$$

$$+ \frac{38 \cdot 35 \cdot 12 + 38 \cdot 45 \cdot 15 + 40 \cdot 35 + 40 \cdot 45 \cdot 6}{100} = 1095{,}5$$

$$\sigma_x = \sqrt{1059 - (29{,}8)^2} = 13{,}08; \quad \sigma_y = \sqrt{1287{,}4 - (35{,}78)^2} = 2{,}68;$$

$$r_B = \frac{1095{,}5 - 29{,}8 \cdot 35{,}78}{13{,}08 \cdot 2{,}68} = 0{,}83.$$

Так как r_B близок к единице, то между Y и X имеется достаточно тесная корреляционная связь.

3. Подставляя найденные величины в уравнения

$$\overline{Y}_x - \overline{y} = r_B \frac{\sigma_y}{\sigma_x}(x - \overline{x}), \quad \overline{X}_y - \overline{x} = r_B \frac{\sigma_x}{\sigma_y}(y - \overline{y}),$$

получаем искомые уравнения регрессии:

1) уравнение регрессии Y на X

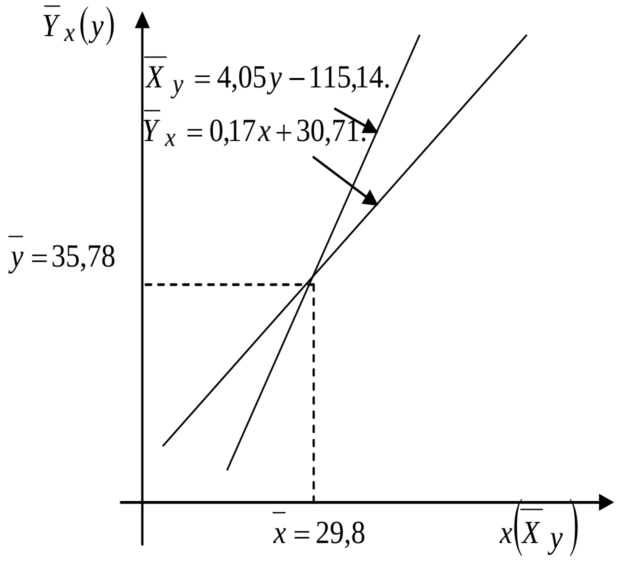
$$\overline{Y}_x - 35{,}78 = 0{,}83 \frac{2{,}68}{13{,}08}(x - 29{,}8), \quad \overline{Y}_x = 0{,}17x + 30{,}71.$$

2) уравнение регрессии X на Y

$$\overline{X}_y - 29{,}8 = 0{,}83 \frac{13{,}08}{2{,}68}(y - 35{,}78), \quad \overline{X}_y = 4{,}05y - 115{,}14.$$

Замечание. Если в корреляционной таблице даны интервальные распределения,
то за значения вариант нужно брать середины частичных интервалов.

Изобразим графики прямых линий регрессии на чертеже.



Так как значения признаков *X* и *Y* являются равноотстоящими, то можно данную задачу решить с помощью условных вариант.

Так, в данном примере

$$C_1 = 25,\ h_1 = 10,\ u_i = \frac{x_i - 25}{10};$$

$$C_2 = 36,\ h_2 = 2,\ v_j = \frac{y_j - 36}{2}.$$

v	u					n_y
	-2	-1	0	1	2	
-3	7	1				8
-2	2	7	1			10

-1	1	5	4	1		11
0		1	15	10	8	34
1			3	12	15	30
2				1	6	7
n_x	10	14	23	24	29	n=100

$$\bar{u} = \frac{-2\cdot 10\cdot 14 + 0\cdot 23 + 1\cdot 24 + 2\cdot 29}{100} = 0{,}48;$$

$$\bar{v} = \frac{-3\cdot 8 - 2\cdot 10 - 1\cdot 11 + 1\cdot 30 + 2\cdot 7}{100} = -0{,}11;$$

$$\overline{u^2} = \frac{4\cdot 10 + 1\cdot 14 + 1\cdot 24 + 4\cdot 29}{100} = 1{,}94; \quad \overline{v^2} = \frac{9\cdot 8 + 4\cdot 10 + 1\cdot 11 + 1\cdot 30 + 4\cdot 7}{100} = 1{,}81;$$

$$\overline{uv} = \frac{(-3)(-2)\cdot 7 + (-3)(-1)\cdot 1 + (-2)(-2)\cdot 2 + (-2)(-1)\cdot 7 + (-1)(-2)\cdot 1 + (-1)(-1)\cdot 5 + (-1)\cdot 1\cdot 1}{100} +$$

$$+\frac{1\cdot 1\cdot 12 + 1\cdot 2\cdot 15 + 2\cdot 1\cdot 1 + 2\cdot 2\cdot 6}{100} = 1{,}4;$$

$$\sigma_u = \sqrt{\overline{u^2} - (\bar{u})^2} = \sqrt{1{,}94 - 0{,}2304} = 1{,}308, \quad \sigma_v = \sqrt{1{,}81 - 0{,}012} = 1{,}34;$$

$$r_B = \frac{1{,}4 - 0{,}48\cdot(-0{,}11)}{1{,}31\cdot 1{,}34} = 0{,}83;$$

$$\bar{x} = \bar{u}h_1 + C_1 = 0{,}48\cdot 10 + 25 = 29{,}8, \quad \bar{y} = \bar{v}h_2 + C_2 = -0{,}11\cdot 2 + 36 = 35{,}78;$$

$$\sigma_x = \sigma_u \cdot h_1 = 1{,}308\cdot 10 = 13{,}08, \quad \sigma_y = \sigma_v \cdot h_2 = 1{,}34\cdot 2 = 2{,}68.$$

Подставляя полученные данные в уравнение регрессии, получим

$$\overline{Y}_x = 0{,}17x + 30{,}71; \quad \overline{X}_y = 4{,}05y - 115{,}14.$$

Задачи (81 – 90)

В задачах 81 – 90 по корреляционной таблице требуется:

1. В прямоугольной системе координат построить эмпирические ломаные регрессии *Y* на *X* и *X* на *Y*, сделать предположение о виде корреляционной связи.
2. Оценить тесноту линейной корреляционной связи.
3. Составить линейные уравнения регрессии *Y* на *X* и *X* на *Y*, построить их графики.

81. В таблице дано распределение объема производственных фондов X (млн руб.) и

объема выпуска готовой продукции однотипных предприятий Y (млн руб.).

Y	X						n_y
	12	17	22	27	32	37	
25	2	4					6
35		6	3				9
45			6	35	4		45
55			2	8	6		16
65				14	7	3	24
n_x	2	10	11	57	17	3	n=100

82. В таблице дано распределение 55 компаний по возрасту сотрудников X и заработ-

ной плате Y (усл. ден. ед.).

Y	X					n_y
	20-30	30-40	40-50	50-60	70-80	
50-80	5	4				9
80-110		12	8	1		21
110-140			5	5		10
140-170			4	7		11
170-200				2	1	3
200-230					1	1
n_x	5	16	17	15	2	n=55

83. В таблице распределение 50 предприятий оптовой торговли по размерам торговой площади X (кв. км.) и объемам реализации Y (млн руб.).

Y	X					n_y
	1-1,5	1,5-2	2-2,5	2,5-3	3-3,5	
5-10	2	1				3
10-15	3	4	3	1		11
15-20		5	10	8		23
20-25			1	6	1	8
25-30				1	4	5
n_x	5	10	14	16	5	n=50

84. В таблице дано распределение 100 однотипных предприятий по основным фондам

X (млн руб.) и себестоимости единицы продукции Y (руб.).

Y	X					n_y
	20	30	40	50	60	
1	8	2				10
3	12	20	8			40
5			10	1		11
7			9	6	2	17
9			10	4	8	22
n_x	20	22	37	11	10	n=100

85. В таблице дано распределение 65 заводов по производству продукции X (тыс. ед.) и

уровню механизации труда Y (%).

Y	X					n_y
	320-370	370-420	420-470	470-520	520-570	
5-20	2	3				5
20-35	1	6	7	1		15
35-50		3	10	9	2	24

50-65			5	4	6	15
65-80			2	3	1	6
n_x	3	12	24	17	9	n=65

86. В таблице дано распределение 50 предприятий по объему выпуска продукции X

(млн руб.) и численности занятых на предприятии Y (чел.).

Y	X					n_y
	2-4	4-6	6-8	8-10	10-12	
30-70	3	4				7
70-110		9	8	1		18
110-150			5	4	1	10
150-190			4	7	2	13
190-230				1	1	2
n_x	3	13	17	13	4	n=50

87. По совокупности 100 предприятий торговли изучается зависимость между ценой

товара X (тыс. руб.) и прибылью торгового предприятия Y (млн руб.).

Y	X						n_y
	5	10	15	20	25	30	
45	2	4					6
55		3	5				8
65			5	35	5		45

75			2	8	17		27
85				4	7	3	14
n_x	2	7	12	47	29	3	n=100

88. В таблице дано распределение 100 предприятий, производящих однородную продукцию, по объему производства X (млн руб.) и себестоимости единицы продукции Y (тыс. руб.).

Y	X					n_y
	0,4-1,4	1,4-2,4	2,4-3,4	3,4-4,4	4,4-5,4	
4-6				2	6	8
6-8			4	7	4	15
8-10	1	1	7	5		14
10-12	2	4	1			7
12-14	3	3				6
n_x	6	8	12	4	10	n=100

89. В таблице дано распределение 50 предприятий по потреблению материалов X (т.) и объему произведенной продукции Y (тыс. ед.).

Y	X					n_y
	9	11	13	15	17	
8	2	6				8
9		4	7	4		15
10		5	7	1	1	14
11			2	4	1	7
12				3	3	6
n_x	2	15	16	12	5	n=50

90. В таблице дано распределение 60 предприятий по стоимости основных производст-

венных фондов *X* (млн руб.) и объему выпуска продукции *Y* (млн руб.).

Y	*X*					n_y
	0-2	2-4	4-6	6-8	8-10	
0-0,2	2	2				4
0,2-0,4	2	7	10			19
0,4-0,6		2	17	7		26
0,6-0,8			4	3	2	9
0,8-1,0					2	2
n_x	4	11	31	10	4	n=60

3.3. Проверка статистических гипотез

Статистическая гипотеза. Статистический критерий

Любую задачу, связанную с анализом статистических данных на языке принятия решений можно представить в виде следующего алгоритма:

1. Сбор статистического материала (выборка).
2. Анализ полученных данных.
3. Выдвижение статистической гипотезы.
4. Проверка выдвинутой гипотезы.
5. Принятие решения.

Такая схема действий является универсальной не только в области статистических исследований.

Решения, принятые на основе анализа статистических данных, называются ***статистическими решениями***. Очевидно, они носят вероятностный характер, поскольку сама выборка является случайной, поэтому принятие статистических решений связано с определенным риском.

Рассмотрим подробнее каждый этап в предложенном алгоритме.

1. Сбор данных

Организация выборки и проведение ее исследования подробно разобраны в курсе общей теории статистики. В целом этот этап зависит от характера произведенной выборки (серийная, повторная и т.д.), ее объема, системы единиц для измеряемого признака.

2. Анализ полученных данных

Первичная статистическая информация представляет собой набор значений признака, т.е. некое числовое множество. Основная задача второго этапа – представить это множество значений в форме, приемлемой для дальнейшего выдвижения гипотезы. Для этой цели служат группировка данных в вариационные ряды и построение полигона и гистограммы относительных частот.

Приведем примеры некоторых, наиболее распространенных видов гистограмм
(рис 3.3).

Из закона больших чисел в форме Бернулли известно, что при увеличении объема выборки и одновременном измельчении интервалов контур гистограммы приближается к функции плотности. На этом факте и основывается следующий этап - выдвижение гипотезы.

4. Выдвижение гипотезы

Часто по эмпирическому распределению выборки можно выдвинуть предположение о теоретическом распределении всей генеральной совокупности. Если же закон распределения известен, а его параметры нет, то можно предположить их величину.

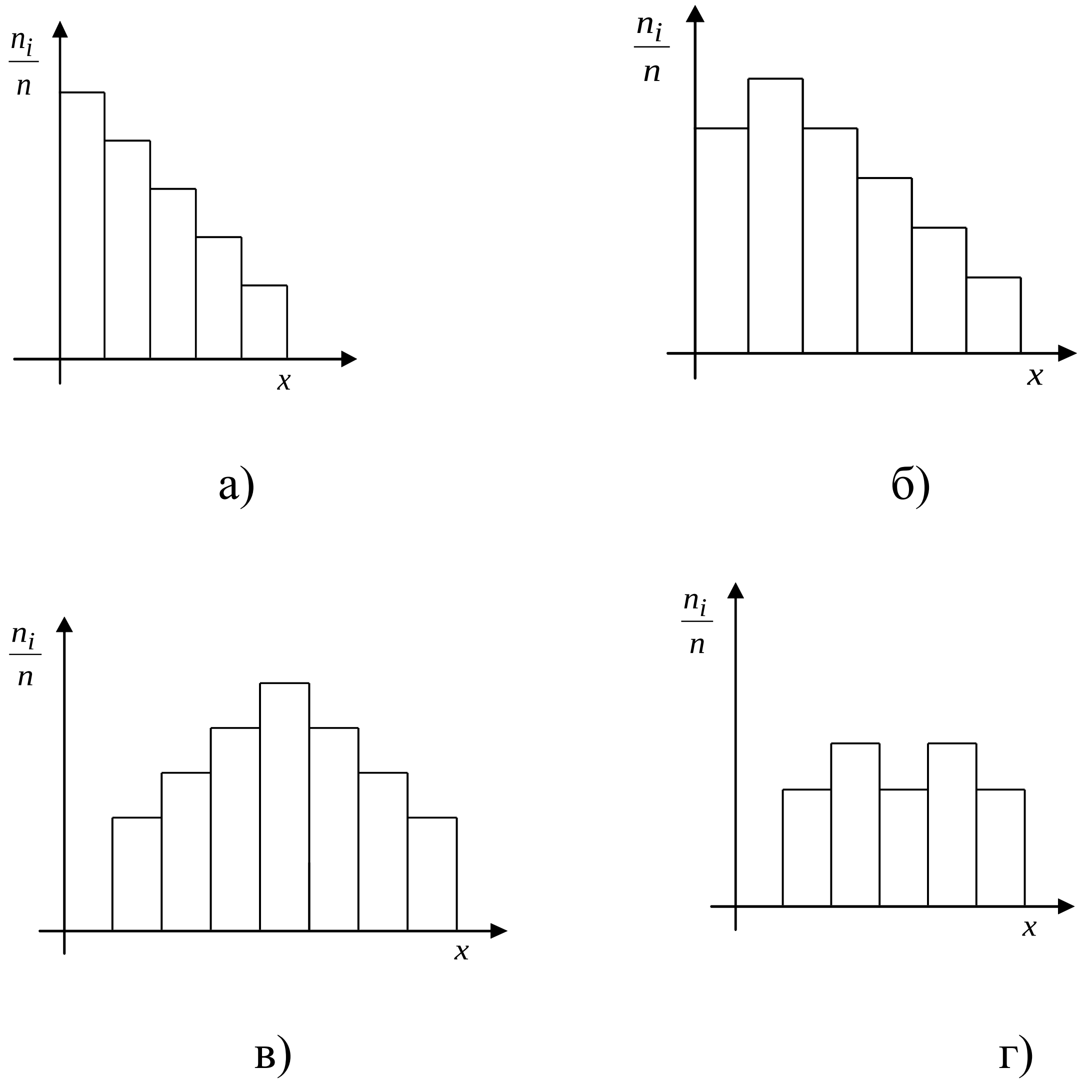


Рис. 3.3.

Статистической называют гипотезу о виде неизвестного распределения, или о параметрах известных распределений.

Например, статистическими являются гипотезы:

1) генеральная совокупность распределена по закону Пуассона;

2) дисперсии двух нормальных совокупностей равны между собой.

В первой гипотезе сделано предположение о виде неизвестного распределения, во второй – о параметрах двух известных распределений.

Наряду с выдвинутой гипотезой рассматривают и противоречащую ей гипотезу. Если выдвинутая гипотеза будет отвергнута, то имеет место противоречащая гипотеза. По этой причине эти гипотезы целесообразно различать.

Нулевой (основной) называют выдвинутую гипотезу H_0.

Конкурирующей (альтернативной) называют гипотезу H_1, которая противоречит нулевой.

Например, если нулевая гипотеза состоит в предположении, что математическое ожидание a нормального распределения равно 10, то конкурирующая гипотеза, в частности, может состоять в предположении, что $a \neq 10$. Коротко это записывают так: H_0: a=10; H_1: $a \neq 10$.

Различают гипотезы, которые содержат только одно и более одного предположений.

Простой называют гипотезу, содержащую только одно предположение.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез.

Для того чтобы выдвинуть гипотезу о том или ином виде теоретического распределения, напомним некоторые из них, изученные в курсе теории вероятностей.

1. Распределение Пуассона (рис. 3.4).

$$P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda},\ m = \left(\overline{0, n}\right);\quad M(X) = D(X) = \lambda.$$

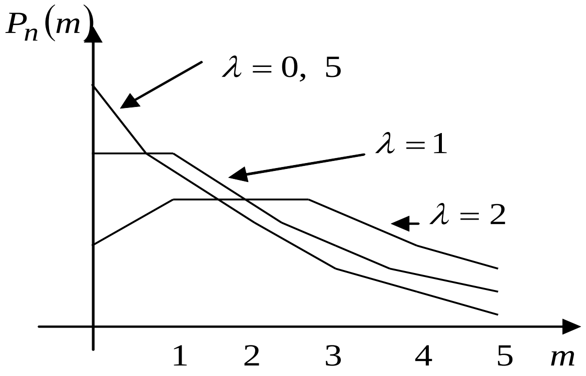


Рис. 3.4.

1. Равномерное распределение (рис. 3.5)

$$f(x) = \begin{cases} 0, & x \notin (a;b) \\ \dfrac{1}{b-a}, & x \in [a;b] \end{cases};\ M(X) = \frac{a+b}{2};\ D(X) = \frac{(b-a)^2}{12}.$$

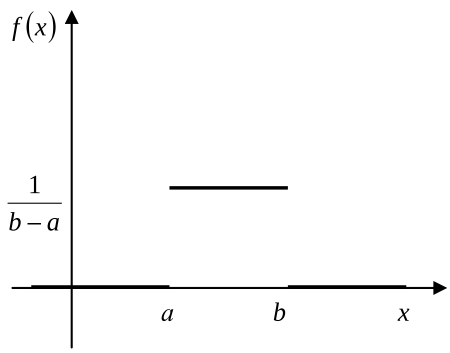


Рис. 3.5.

3. Показательное распределение (рис. 3.6)

$$f(x)=\begin{cases}0, & x<0\\ \lambda e^{-\lambda x}, & x\geq 0\end{cases};\ M(X)\frac{1}{\lambda};D(X)=\frac{1}{\lambda^2}.$$

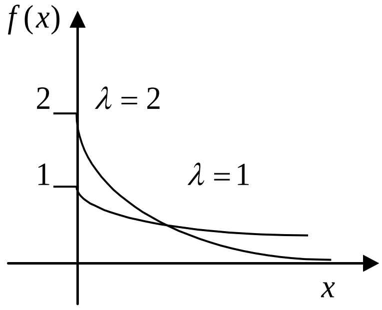


Рис. 3.6.

4. Нормальное распределение (рис. 3.7)

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-a)^2}{2\sigma^2}};\ M(X)=a;D(X)=\sigma^2.$$

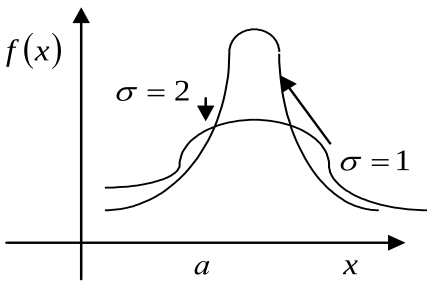


Рис. 3.7.

Сравнивая вид гистограмм приведенных на рис. 3.3, с графиками основных теоретических распределений можно выдвинуть гипотезу о виде распределения всей генеральной совокупности.

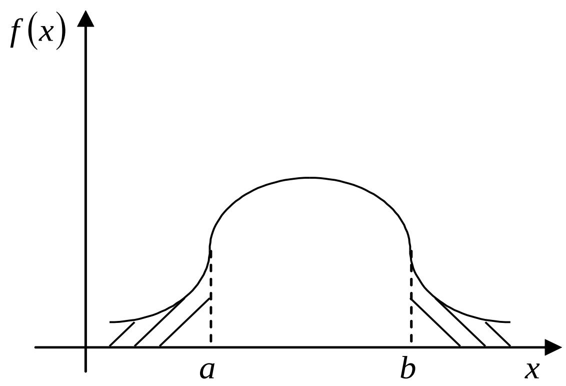
Естественно предположить, что на рис. 3.3 (а) – показательное распределение, на рис. 3.3 (б) – распределение Пуассона, на рис.3.3 (в) – нормальное и наконец на рис. 3.3 (г) – равномерное. Итак, по полученной гистограмме выбирается подходящие теоретическое распределение. Числовые характеристики оцениваются соответствующими выборочными характеристиками.

5. Проверка гипотезы

Проверка гипотезы состоит в том, чтобы установить: можно ли считать расхождение между предполагаемым теоретическим и эмпирическим распределениями несущественным или же существуют коренные (принципиальные) различия? Ответ на этот вопрос дает статистический критерий.

Статистическим критерием называют случайную величину K, которая служит для проверки нулевой гипотезы.

Пусть выдвинута гипотеза о том, что генеральная совокупность распределена нормально и функция плотности



имеет вид (рис. 3.8).

Рис. 3.8.

Тогда можно с уверенностью утверждать, что попадание выборки в заштрихованную область маловероятно, и, напротив, попадание в интервал $(a;b)$ имеет большую вероятность.

Таким образом, если ***наблюдаемое значение критерия*** $\boldsymbol{K_{набл.}}$ (вычисленное по выборке) попадает в интервал $(a;b)$, то это не противоречит гипотезе, если же оно попадает в заштрихованную область, то это ставит гипотезу под сомнение. Следовательно, множество всех возможных значений критерия можно разделить на два непересекающихся множества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, а другая – при которых она принимается.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

Областью принятия гипотезы (областью допустимых значений) называют совокупность значений критерия, при которых нулевую гипотезу принимают.

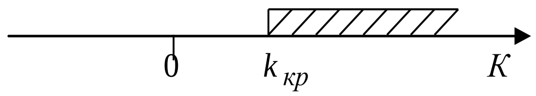
Основной принцип проверки статистических гипотез можно сформулировать так: если наблюдаемое значение критерия принадлежит критической области – гипотезу отвергают, если наблюдаемое значение критерия принадлежит области принятия гипотезы – гипотезу принимают.

Поскольку критерий K – одномерная случайная величина, все ее возможные значения принадлежат некоторому интервалу. Поэтому критическая область и область принятия гипотезы также являются интервалами и, следовательно, существуют точки, которые их разделяют.

Критическими точками (границами) $k_{кр}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Различают одностороннюю (правостороннюю или левостороннюю) и двустороннюю критические области.

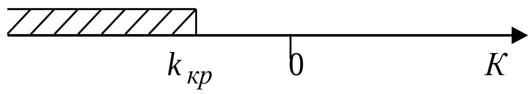
Правосторонней называют критическую область,



определяемую неравенством

$$K > k_{кр}, \quad \text{где } k_{кр} > 0.$$

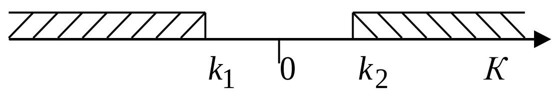
Левосторонней называют критическую область, определяемую неравенством



$$K > k_{кр}, \quad \text{где } k_{кр} < 0.$$

Односторонней называют правостороннюю или левостороннюю критическую область.

Двусторонней называют критическую область, определяемую неравенствами



$$K < k_1, K > k_2, \text{ где } k_2 > k_1.$$

При проверке выдвинутой гипотезы можно допустить два вида ошибок.

1. Если $K_{набл.}$ попало в критическую область и выдвинутая гипотеза H_0 отклоняется, даже если она верна.

Ошибка первого рода состоит в том, что будет отвергнута правильная гипотеза H_0. Вероятность совершить ошибку первого рода принято обозначать через α; ее назы-
вают ***уровнем значимости***.

Замечание. Часто уровень значимости принимают равным 0,05 или 0,01. Если, например, принят уровень значимости, равный 0,05, то это означает, что в пяти случаях из ста имеется риск допустить ошибку первого рода (отвергнуть правильную гипотезу).

2. Возможна и другая ошибка – принять гипотезу H_0, когда она неверна.

Ошибка второго рода – будет принята неправильная гипотеза H_0.

Вероятность совершить ошибку второго рода принято обозначать β; ее называют риск два.

Таблица случаев

Решение по критерию	Истина	
	H_0 верна	H_0 неверна
отклоняется	ошибка первого рода	решение верно
принимается	решение верно	ошибка второго рода

Мощностью критерия называется вероятность отклонить нулевую гипотезу, когда верна конкурирующая гипотеза

$$P_{H_1}\left(\overline{H_0}\right)=1-\beta .$$

Пусть мощность (1-β) возрастает; следовательно, уменьшается вероятность β совершить ошибку второго рода. Таким образом, чем мощность больше, тем вероятность ошибки второго рода меньше.

Итак, при заданном уровне значимости критическую область следует строить так, чтобы мощность критерия была максимальной. Выполнение этого требования должно обеспечить минимальную ошибку второго рода.

Отыскание любой из критических областей (правосторонней, левосторонней и двусторонней) сводится к нахождению критических точек $k_{кр}$.

С этой целью задаются достаточно малым уровнем значимости (обычно 0,05; 0,01). Затем ищут критическую точку исходя из условий:

– $P(K > k_{кр}) = \alpha$ (для правосторонней критической области);

– $P(K < k_{кр}) = \alpha$ (для левосторонней критической области);

– $P(K < k_1) + P(K > k_2) = \alpha$ (для двусторонней критической области).

Для каждого из критериев имеются соответствующие таблицы, по которым находят критические точки $k_{кр}$, удовлетворяющие данным требованиям.

Эмпирические и выравнивающие (теоретические) частоты

Пусть произведено n испытаний, в которых величина X приняла n_1 раз значение x_1, n_2 раз значение x_2, …, n_k раз значение x_k, причем $\sum n_i = n$.

Эмпирическими частотами называют фактические наблюдаемые частоты n_i.

Пусть имеются основания предположить, что изучаемая величина X распределена по некоторому определенному закону. Чтобы проверить, согласуется ли это предположение с данными наблюдений, вычисляют частоты наблюдаемых значений, т.е. находят теоретические частоты n'_i

каждого из наблюдаемых значений в предположении, что величина X распределена по предполагаемому закону.

Выравнивающими (теоретическими) в отличие от фактических наблюдаемых эмпирических частот называют частоты n'_i, найденные теоретически (вычислением).

Опишем способ нахождения теоретических частот.

В случае **дискретного распределения** признака X генеральной совокупности выравнивающие частоты находят с помощью равенства

$$n'_i = nP_i = n \cdot P(X = x_i),$$

где n – число испытаний; P_i - вероятность наблюдаемого значения x_i, вычисленная при допущении, что X имеет предполагаемое распределение.

Итак, выравнивающая частота наблюдаемого значения x_i дискретного распределения равна произведению числа испытаний на вероятность этого наблюдаемого значения.

Важнейшим дискретным распределением является распределение Пуассона. Тогда x_i принимает значения m: 0, 1, 2 …. Вероятности P_i вычисляются по формуле Пуассона:

$$P_i = P(X = m) = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Известно, что параметр λ, которым определяется распределение Пуассона, равен математическому ожиданию этого распределения. Поскольку в качестве оценки математического ожидания принимают выборочную среднюю, то и в качестве оценки λ можно принять выборочную среднюю $\bar{x}_B$.

В случае **непрерывного распределения** вероятности отдельных возможных значений равны нулю. Поэтому весь интервал возможных значений делят на k непересекающихся интервалов и вычисляют вероятности p_i попадания X в i-й частичный интервал, а затем, как и для дискретного распределения, умножают число испытаний на эти вероятности.

Итак, выравнивающие частоты непрерывного распределения находят по равенству

$$n'_i = nP_i = n \cdot P(x_i < X < x_{i+1}),$$

где n – число испытаний; P_i - вероятность попадания X в i-й частичный интервал, вычисленная при допущении, что X имеет предполагаемое распределение.

Выразив P_i через функцию распределения и плотность вероятностей, получим

$$P_i = P(x_i < X < x_{i+1}) = F(x_{i+1}) - F(x_i) = \int_{x_i}^{x_{i+1}} f(x)dx.$$

Согласно теореме о среднем

$$\int_{x_i}^{x_{i+1}} f(x)dx = (x_{i+1} - x_i) \cdot f\left(x_i^*\right),$$

где $(x_{i+1} - x_i)$ длина интервала; x_i^* – любая точка интервала $(x_i; x_{i+1})$.

В качестве точки x_i^* обычно принимают значение середины интервала $(x_i; x_{i+1})$, т.е. $x_i^* = \frac{x_i + x_{i+1}}{2}$.

Таким образом, формула для вычисления теоретических частот примет вид: $n'_i = n \cdot (x_{i+1} - x_i) \cdot f\left(x_i^*\right)$.

Методика вычисления теоретических частот нормального распределения

1) Эмпирическое распределение задано в виде последовательности интервалов одинаковой длины и соответствующих им частот

$(x_i; x_{i+1})$	$(x_1; x_2)$	$(x_2; x_3)$	...	$(x_k; x_{k+1})$

n_i	n_1	n_2	…	n_k

Как найти теоретические частоты, если предполагается, что генеральная совокупность распределена нормально?

Воспользуемся определением выравнивающих частот непрерывного распределения:

$$n'_i = nP_i = n \cdot P(x_i < X < x_{i+1}).$$

Для нормального закона распределения

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Полагая вместо $M(X) = a$, $\bar{x}_B$ несмещенную оценку *M(X)*, а вместо $\sigma_Г - \sigma_B$,
имеем

$$P(x_i < X < x_{i+1}) = \Phi\left(\frac{x_{i+1} - \bar{x}_B}{\sigma_B}\right) - \Phi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right).$$

Алгоритм вычисления теоретических частот

1. Вычислить $\bar{x}_B$ и σ_B, причем в качестве вариант принять среднее арифметическое концов интервалов $x_i^* = \dfrac{x_i + x_{i+1}}{2}$.

2. Пронормировать случайную величину *X*, т.е. перейти к величине $Z = \dfrac{(x_i - \bar{x}_B)}{\sigma_B}$, вычислив концы интервалов $(z_i; z_{i+1})$:

$$z_i = \frac{(x_i - \bar{x}_B)}{\sigma_B},\quad z_{i+1} = \frac{(x_{i+1} - \bar{x}_B)}{\sigma_B},$$

причем наименьшее значение *Z*, т.е. z_1, полагают равным $-\infty$, а наибольшее, т.е. z_k, полагают равным ∞.

3. Вычислить теоретические вероятности P_i попадания *X* в интервалы $(x_i; x_{i+1})$ по равенству (*Ф(z)* – функция Лапласа)

$$P_i = \Phi(z_{i+1}) - \Phi(z_i)$$

и, наконец, найти искомые теоретические частоты $n'_i = nP_i$.

Все вычисления целесообразно внести в таблицу.

i	x_i	x_{i+1}	n_i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i=\Phi(z_{i+1})-\Phi(z_i)$	$n'_i = nP_i$

2) Эмпирическое распределение задано в виде последовательности равноотстоящих вариант и соответствующих им частот. Тогда для нахождения теоретических частот используют формулу

$$n'_i = n\cdot(x_{i+1}-x_i)\cdot f\left(x_i^*\right)=n\cdot h\cdot f\left(x_i^*\right), \qquad (*)$$

где h – длина частичного интервала.

Запишем плотность общего нормального распределения:

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

При a = 0 и σ = 1 получим плотность нормированного распределения:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

или, изменив обозначение аргумента,

$$\varphi(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}.$$

Положив $u=\frac{(x-a)}{\sigma}$, имеем

$$\varphi(u)=\frac{1}{\sqrt{2\pi}}e^{\frac{-(x-a)^2}{2\sigma^2}}.$$

Сравнивая $\varphi(u)$ и $f(x)$, заключаем, что

$$f(x)=\frac{1}{\sigma}\varphi(u).$$

Если математическое ожидание a и среднее квадратическое отклонение σ неизвестны, то в качестве оценок этих параметров принимают соответственно выборочную среднюю $\bar{x}_B$ и выборочное среднее квадратическое отклонение σ_B.

Тогда

$$f(x)=\frac{1}{\sigma_B}\varphi(u), \qquad \text{где } u=\frac{\left(x-\bar{x}_B\right)}{\sigma_B}.$$

Вернемся к формуле (*) $n'_i = n\cdot h\cdot f\left(x_i^*\right)=n\cdot h\cdot\frac{1}{\sigma_B}\cdot\varphi(u_i)$.

Таким образом, $n'_i=\frac{n\cdot h}{\sigma_B}\cdot\varphi(u_i)$, где $u_i=\frac{\left(x_i-\bar{x}_B\right)}{\sigma_B}$.

Алгоритм нахождения теоретических частот

1. Вычислить, например методом произведений выборочную среднюю $\bar{x}_B$ и выборочное среднее квадратическое отклонение σ_B.

2. Перейти к условным вариантам $u_i=\frac{\left(x_i-\bar{x}_B\right)}{\sigma_B}$.

3. Найти значения функции Лапласа $\varphi(u_i)$.

4. Вычислить теоретические частоты по формуле $n'_i=\frac{n\cdot h}{\sigma_B}\cdot\varphi(u_i)$.

Вычисления целесообразно вносить в таблицу

i	x_i	n_i	u_i	$\varphi(u_i)$	$n'_i=\frac{n\cdot h}{\sigma_B}\cdot\varphi(u_i)$

Проверка гипотезы о нормальном распределении генеральной совокупности. Критерий согласия Пирсона

Проверка гипотезы о предлагаемом законе неизвестного распределения производится при помощи специально подобранной случайной величины – критерия согласия.

Критерием согласия называют критерий проверки гипотезы о предполагаемом законе неизвестного распределения.

Имеется несколько критериев согласия: χ^2 («хи-квадрат») Пирсона, Колмогорова, Смирнова и др. Остановимся подробнее на описании применения критерия Пирсона к проверке гипотезы о нормальном распределении генеральной совокупности (критерий аналогично применяется и для других распределений, в этом состоит его достоинство).

Критерий Пирсона служит для сравнения эмпирических и теоретических частот и отвечает на вопрос: случайно ли расхождение этих частот или оно значимо? Но критерий Пирсона, как и любой другой критерий, не доказывает справедливость гипотезы, а лишь устанавливает на принятом уровне значимости ее согласие или несогласие с данными наблюдений.

Итак, пусть по выборке объема n получено эмпирическое распределение

Варианты .... x_i x_1 x_2 ... x_s

Эмпирические частоты ...... n_i n_1 n_2 ... n_s

Допустим, что в предположении нормального распределения генеральной совокупности вычислены теоретические частоты n'_i. При уровне значимости α требуется проверить нулевую гипотезу H_0: генеральная совокупность распределена нормально.

В качестве критерия проверки нулевой гипотезы примем случайную величину

$$\chi^2 = \sum_{i=1}^{s} \frac{\left(n_i - n_i'\right)^2}{n_i'}.$$

Эта величина случайная, так как в различных опытах она принимает различные, заранее неизвестные значения. Очевидно, что чем меньше различаются эмпирические и теоретические частоты, тем меньше величина критерия χ^2 и, следовательно, разница между эмпирическим и теоретическим распределениями несущественна. Критерий согласия χ^2 характеризуется двумя параметрами: уровнем значимости α и числом степеней свободы k.

Число степеней свободы находят по равенству $k = s - 1 - r$, где s – число групп (частичных интервалов) выборки; r – число параметров предполагаемого распределения, которые оценены по данным выборки.

В частности, если предполагаемое распределение – нормальное, то оценивают два параметра (математическое ожидание и среднее квадратическое отклонение), поэтому $r = 2$ и число степеней свободы $k = s - 1 - r = s - 1 - 2 = s - 3$.

Если, например, предполагают, что генеральная совокупность распределена по закону Пуассона, то оценивают один параметр λ, поэтому $r = 1$ и $k = s - 2$.

Поскольку односторонний критерий более «жестко» отвергает нулевую гипотезу, чем двусторонний, построим правостороннюю критическую область исходя из требования, чтобы вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы была равна принятому уровню значимости α:

$$P\left[\chi^2 > \chi^2_{кр}(\alpha; k)\right] = \alpha .$$

Таким образом, правосторонняя критическая область определяется неравенством $\chi^2 > \chi^2_{кр}(\alpha; k)$. Функция плотности данного распределения будет иметь вид (рис. 3.9).

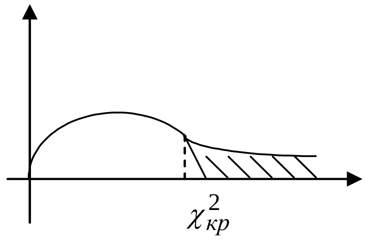


Рис. 3.9.

Существуют специальные таблицы, по которым для заданных k и α находятся соответствующие критические значения критерия $\chi^2_{кр}$ (приложение 6).

Обозначим значение критерия, вычисленное по данным наблюдений, через $\chi^2_{набл}$ и сформулируем ***алгоритм проверки нулевой гипотезы.***

1. По предполагаемому теоретическому распределению находим выравнивающие частоты n'_i.

2. Вычисляем наблюдаемое значение критерия:

$$\chi^2_{набл} = \sum_{i=1}^{s} \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}}.$$

3. Находим число степеней свободы по формуле $k = s -$ 3.

4. По данному значению уровня значимости α и числу степеней свободы k находим критическое значение критерия $\chi^2_{кр}$ (α; k).

5. Сравниваем $\chi^2_{набл}$ и $\chi^2_{кр}$. Если $\chi^2_{набл} < \chi^2_{кр}$, нет оснований отвергнуть нулевую гипотезу. Если $\chi^2_{набл} > \chi^2_{кр}$ - нулевую гипотезу отвергают.

Замечание 1. Объем выборки должен быть достаточно велик, во всяком случае, не менее 50. Каждая группа должна содержать не менее 5 – 8 вариант; малочисленные группы следует объединять в одну, суммируя частоты.

Замечание 2. Для контроля вычислений применяют формулу

$$\chi^2_{набл} = \left[\frac{\sum n_i^2}{n_i'}\right] - n.$$

Решение типовых задач

Задача 1. Используя критерий Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема $n = 200$.

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Решение. 1. Вычислим $\bar{x}_B = \frac{\sum_{i=1}^{k} x_i n_i}{n} = 12{,}63$ и выборочное среднее квадратическое отклонение $\sigma_B = \sqrt{\overline{x_B^2} - \left(\bar{x}_B\right)^2} = 4{,}695$.

2. Вычислим теоретические частоты учитывая, что n = 200, h = 2, σ_B = 4,695, по формуле

$$n'_i = \frac{n \cdot h}{\sigma_B} \cdot \varphi(u_i) = \frac{200 \cdot 2}{4{,}695} \cdot \varphi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right) = 85{,}2 \cdot \varphi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right).$$

Составим расчетную таблицу (значения функции $\varphi(x)$ приведены в приложении 1).

i	x_i	$u_i = \frac{\left(x_i - \bar{x}_B\right)}{\sigma_B}$	$\varphi(u_i)$	$n'_i = \frac{n \cdot h}{\sigma_B} \cdot \varphi(u_i)$
1	5	-1,62	0,1074	9,1
2	7	-1,20	0,1942	16,5
3	9	-0,77	0,2966	25,3
4	11	-0,35	0,3752	32,0

5	13	0,08	0,3977	33,9
6	15	0,51	0,3503	29,8
7	17	0,93	0,2589	22,0
8	19	1,36	0,1582	13,5
9	21	1,78	0,0818	7,0

3. Сравним эмпирические и теоретические частоты. Составим расчетную таблицу, из которой найдем наблюдаемое значение критерия $\chi^2_{набл} = \sum\limits_{i=1}^{s} \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}}$:

i	n_i	n'_i	$\left\lvert n_i - n_i^{'} \right\rvert$	$\left(n_i - n_i^{'}\right)^2$	$\frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}}$
1	15	9,1	5,9	34,81	3,8
2	26	16,5	9,5	90,25	5,5
3	25	25,3	0,3	0,09	0,0
4	30	32,0	2,0	4,0	0,1
5	26	33,9	7,9	62,41	1,8
6	21	29,8	8,8	77,44	2,6
7	24	22,0	2,0	4,0	0,2
8	20	13,5	6,5	42,25	3,1
9	13	7,0	6,0	36,0	5,1
$\sum$	200				$\chi^2_{набл}$=22,2

По таблице критических точек распределения χ^2 (приложение 6), по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = s - 3 = 9 - 3 = 6$ находим критическую точку правосторонней критической области $\chi^2_{кр}$ (0,05; 6) = 12,6.

Так как $\chi^2_{набл}$=22,2 > $\chi^2_{кр}$ = 12,6, гипотезу о нормальном распределении генеральной совокупности отвергаем. Другими словами, эмпирические и теоретические частоты различаются значимо.

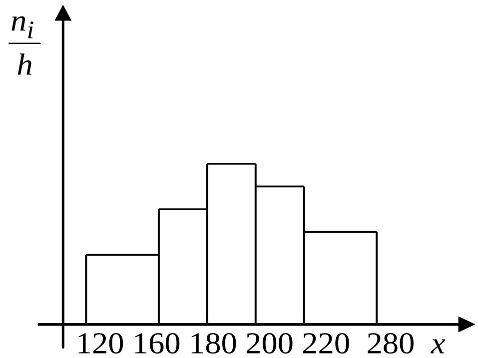
Задача 2. Распределение 50 промышленных предприятий по средней численности работников характеризуются следующими данными:

Численность работников	120-140	140-160	160-180	180-200	200-220	220-240	240-260	260-280
Число предприятий	1	4	10	14	12	6	2	1

Проверить на уровне значимости $\alpha = 0{,}01$ гипотезу о нормальном распределении при помощи критерия Пирсона.

Решение. 1. Ввиду малочисленности частот объединяем первые два и последние три интервала. Получается таблица

$(x_i; x_{i+1})$	120-160	160-180	180-200	200-220	220-280
n_i	5	10	14	12	9



Строим гистограмму:

По виду гистограммы можно предположить, что данная случайная величина подчиняется нормальному закону распределения. Выдвинем и проверим гипотезу – H_0: исследуемая случайная величина имеет нормальный закон распределения.

2. Для вычисления теоретических частот находим $\bar{x}_B$, σ_B, n.

$$\bar{x}_B = 195{,}2;\ \sigma_B = 28{,}5;\ n = 50.$$

3. Найдем теоретические частоты $n'_i = nP_i$, где $P_i = P(x_i < X < x_{i+1})$ - вероятность того, что случайная величина попадет в интервал $(x_i; x_{i+1})$.

Так как предполагаемый закон распределения нормальный, то

$$P_i = \Phi\left(\frac{x_{i+1} - \bar{x}_B}{\sigma_B}\right) - \Phi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right),$$

где $\Phi(x)$ – функция Лапласа (приложение 3). Вычисления приведем в таблице:

i	x_i	x_{i+1}	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n'_i = nP_i$
1	-∞	160	-∞	-1,24	-0,5	-0,3925	0,1075	5,375
2	160	180	-1,24	0,53	-0,3925	-0,2019	0,1906	9,53
3	180	200	-0,53	0,17	-0,2019	0,0675	0,2694	13,47
4	200	220	0,17	0,87	0,0675	0,3079	0,2404	12,02

5	22 0	$+\infty$	0,87	$+\infty$	0,307 9	0,5	0,1921	9,605
$\sum$							1	50

4. Сравним эмпирические и теоретические частоты, используя критерий Пирсона.

Вычислим наблюдаемое значение критерия Пирсона. Для этого составим расчетную таблицу

i	n_i	n'_i	$\left\|n_i - n_i'\right\|$	$\left(n_i - n_i'\right)^2$	$\dfrac{\left(n_i - n_i'\right)^2}{n_i'}$
1	5	5,375	0,375	0,1406	0,0262
2	10	9,53	0,47	0,2209	0,0223
3	14	13,47	0,53	0,2809	0,0209
4	12	12,02	0,02	0,0004	0,0
5	9	9,605	0,605	0,366	0,0381
$\sum$	50	50			$\chi^2_{набл}$=0,10 75

По таблице критических точек распределения χ^2 (приложение 6), по уровню значимости α = 0,01 и числу степеней свободы $k = s - 3 = 5 - 3 = 2$ (s – число интервалов) находим критическую точку правосторонней критической области $\chi^2_{кр}$ (0,01; 2) = 9,2.

Сравним $\chi^2_{набл}$ и $\chi^2_{кр}$. Так как $\chi^2_{набл}$=0,1075 < $\chi^2_{кр}$ = 9,2, нет оснований отклонить гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначимо.

Задачи (91 – 100)

В задачах 91 – 100 даны эмпирические значения случайной величины X. Требуется:

1. Выдвинуть гипотезу о виде распределения.
2. Проверить гипотезу с помощью критерия Пирсона при заданном уровне значимости α.

За значения параметров a и σ принять среднюю выборочную и среднее выборочное квадратическое отклонение, вычисленные по эмпирическим данным.

91. В таблице дано распределение среднегодовой стоимости основных фондов (млн руб.) по 50 предприятиям отрасли.

x_i	100,5	111,2	121,5	132,0	142,5	153,0	163,5
n_i	4	9	18	8	5	4	2

$\alpha = 0{,}05$.

92. По данным, полученным от 50 фермерских хозяйств одного из регионов, составлено распределение численности работников.

x_i	1-5	5-9	9-13	13-17	17-21	21-25
n_i	6	10	17	12	4	1

$\alpha = 0{,}01$.

93. Распределение 60 магазинов по величине товарооборота (млн руб.) характеризуется следующими данными:

x_i	1,0-1,5	1,5-2,0	2,0-2,5	2,5-3,0	3,0-3,5
n_i	5	11	23	13	8

$\alpha = 0{,}05$.

94. Результаты анализа темпов роста стоимости акций 50 компаний (%) имеют сле-
дующее распределение:

x_i	102-104	104-106	106-108	108-110	110-112
n_i	5	10	15	12	8

$\alpha = 0{,}05$.

95. Распределение 50 туристических фирм по средней численности работников харак-
теризуется следующими данными:

x_i	0-5	5-10	10-15	15-20	20-25	25-30
n_i	4	7	11	21	5	2

$\alpha = 0{,}025$.

96. В таблице дано распределение величины дохода торговых предприятий за год
(тыс. руб.):

x_i	20-24	24-28	28-32	32-36	36-40
n_i	10	21	30	17	12

$\alpha = 0{,}01$.

97. Распределение стоимости покупок (руб.) 50 случайно выбранных покупателей ха-
рактеризуется следующими данными:

x_i	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	100-110
n_i	1	3	4	6	11	10	7	5	2	1

$\alpha = 0{,}01$.

98. Результаты анализа 100 промышленных предприятий по возрастной структуре про-

изводственного оборудования характеризуются следующими данными:

x_i	0-5	5-10	10-15	15-20	20-25	25-30	30-35
n_i	4	15	20	26	19	14	2

$\alpha = 0{,}025$.

99. Распределение 50 промышленных предприятий по уровню механизации труда (%)

характеризуется следующими данными:

x_i	5-15	15-25	25-35	35-45	45-55	55-65
n_i	6	5	10	13	9	7

$\alpha = 0{,}01$.

100. В таблице дано распределение расходов на рекламу у предприятий в долях от до-

хода:

x_i	0,2-0,4	0,4-0,6	0,6-0,8	0,8-1	1-1,2	1,2-1,4
n_i	5	17	23	16	7	2

$\alpha = 0{,}01$.

Вопросы для самоконтроля

1. Несмещенной оценкой $\overline{X}_Г$ ***является***

101. $M(\overline{X})$. ***102.*** $\overline{X}$ ***103.*** $\overline{X}_B$. ***104*** $\overline{X^2}$.

2. Дано распределение выборки

x_i	**10**	**15**	**20**
n_i	**2**	**5**	**3**

Найти $\overline{x}_B$

201. 15,5. ***202.*** 1,55. ***203.*** 15. ***204.*** 155.

3. Выборочный коэффициент корреляции r_B ***= 0,85. Оценить тесноту связи.***

301. умеренная; ***302.*** высокая; ***303.*** заметная; ***304.*** весьма высокая.

4. Статистическая оценка не должна удовлетворять требованию

401. несмещенность. ***402.*** эффективность. ***403.*** состоятельность. ***404.*** неотрицательность.

5. Вероятность, с которой осуществляется неравенство $\left|\theta - \theta^*\right| < \delta$***, называют***

501. точностью; ***502.*** оценкой; ***503.*** надежностью; ***504.*** эффектом.

6. Известно, что уровень значимости α ***= 0,05, число степеней свободы k = 13. Каково критическое значение критерия*** $\chi^2_{кр}$***?***

601. 23,7. ***602.*** 22,4. ***603.*** 26,1. ***604.*** 24,7.

7. Согласно статистическому критерию Пирсона, нулевую гипотезу отвергают, если

701. $\chi^2_{набл} > \chi^2_{кр}$. ***702.*** $\chi^2_{набл} \neq \chi^2_{кр}$. ***703.*** $\chi^2_{набл} = \chi^2_{кр}$. ***704.*** $\chi^2_{набл} < \chi^2_{кр}$.

8. Зависимость, при которой изменение одной из величин влечет изменение среднего значения другой величины называется

801. статистической; ***802.*** функциональной;
803. корреляционной; ***804.*** многомерной.

9. В качестве критерия согласия проверки гипотезы о нормальном распределении генеральной совокупности применяется случайная величина $\chi^2 = \sum_{i=1}^{s} \frac{\left(n_i - n_i'\right)^2}{n_i'}$***, где*** n'_i ***- …***

901. стандартизированные частоты; ***902.*** эмпирические частоты;
903. нормированные частоты; ***904.*** теоретические частоты.

10. Найти параметр t в доверительном интервале $\overline{X} - t\frac{\sigma}{\sqrt{n}} < a < \overline{X} + t\frac{\sigma}{\sqrt{n}}$***, если надежность γ = 0,95.***

1001. 0,47. ***1002.*** 2,04. ***1003.*** 1,96. ***1004.*** 0,52.

ПРИЛОЖЕНИЕ 1

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$

	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	332	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	2807	2874	2850	2827	2800	2780	2756	2732	2700	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	251	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	1738
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1529	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	0731	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0450	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	035	034	033	033	032	031	031	030	029	029

	5	7	9	2	5	7	0	3	7	0
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	0056	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

Продолжение приложения 1

	0	1	2	3	4	5	6	7	8	9
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004

3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

ПРИЛОЖЕНИЕ 2

Таблица значений функции $p(m;\lambda)=\frac{\lambda^m}{m!}e^{-\lambda}$

m / λ	0	1	2	3	4	5	6	7	8	9	10
0,1	0,9048	0905	0045	0002	0000						
0,2	8187	1637	0164	0011	0001						
0,3	7408	2222	0333	0033	0003	0000					
0,4	6703	2681	0536	0072	0007	001					
0,5	6065	3033	0758	0126	0016	0002					
0,6	5488	3293	0988	0198	0030	0004	0000				
0,7	4966	3476	1217	0284	0050	0007	0001				

0,8	4493	3595	1438	0383	0077	0012	0002				
0,9	4066	3659	1647	0494	0111	0020	0003	0000			
1	3679	3679	1839	0613	0153	0031	0005	0001	0000	0000	
2	1353	2707	2707	1805	0902	0361	0120	0034	0009	0002	0000
3	0498	1494	2240	2240	1680	1008	0504	0216	0081	0027	0008
4	0183	0733	1465	1954	1954	1563	1042	0595	0298	0132	0052
5	0067	0337	0842	1404	1755	1755	1462	1044	0653	0363	0181
6	0025	0149	0146	0892	1339	1606	1606	1337	1033	0688	0413
7	0009	0064	0223	0521	0921	1277	1490	1490	1304	1014	0710
8	0003	0027	0107	0286	0573	0916	1221	1396	1396	1241	0993
9	0001	0011	0050	0150	0337	0607	0911	1171	1318	1318	1186
10	0000	0004	0023	0076	0189	0378	0631	0901	1126	1251	1251

ПРИЛОЖЕНИЕ 3

Таблица значений функций $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{z^2}{2}} dz$

X	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,43	0,1664	0,86	0,3051	1,29	0,4015
0,01	0,0040	0,44	0,1700	0,87	0,3078	1,30	0,4032
0,02	0,0080	0,45	0,1736	0,88	0,3106	1,31	0,4049
0,03	0,0120	0,46	0,1772	0,89	0,3133	1,32	0,4066
0,04	0,0160	0,47	0,1808	0,90	0,3159	1,33	0,4082
0,05	0,0199	0,48	0,1844	0,91	0,3186	1,34	0,4099
0,06	0,0239	0,49	0,1879	0,92	0,3212	1,35	0,4115
0,07	0,0279	0,50	0,1915	0,93	0,3238	1,36	0,4131
0,08	0,0319	0,51	0,1950	0,94	0,3264	1,37	0,4147
0,09	0,0359	0,52	0,1985	0,95	0,3289	1,38	0,4162
0,10	0,0398	0,53	0,2019	0,96	0,3315	1,39	0,4177
0,11	0,0438	0,54	0,2054	0,97	0,3340	1,40	0,4192
0,12	0,0478	0,55	0,2088	0,98	0,3365	1,41	0,4207
0,13	0,0517	0,56	0,2123	0,99	0,3389	1,42	0,4222

0,14	0,0557	0,57	0,2157	1,00	0,3413	1,43	0,4236
0,15	0,0596	0,58	0,2190	1,01	0,3438	1,44	0,4251
0,16	0,0636	0,59	0,2224	1,02	0,3461	1,45	0,4265
0,17	0,0675	0,60	0,2257	1,03	0,3485	1,46	0,4279
0,18	0,0714	0,61	0,2291	1,04	0,3508	1,47	0,4292
0,19	0,0753	0,62	0,2324	1,05	0,3531	1,48	0,4306
0,20	0,0793	0,63	0,2357	1,06	0,3554	1,49	0,4319
0,21	0,0832	0,64	0,2389	1,07	0,3577	1,50	0,4332
0,22	0,0871	0,65	0,2422	1,08	0,3599	1,51	0,4345
0,23	0,0910	0,66	0,2454	1,09	0,3621	1,52	0,4357
0,24	0,0948	0,67	0,2486	1,10	0,3643	1,53	0,4370
0,25	0,0987	0,68	0,2517	1,11	0,365	1,54	0,4382
0,26	0,1026	0,69	0,2549	1,12	0,3686	1,55	0,4394
0,27	0,1064	0,70	0,2580	1,13	0,3708	1,56	0,4406
0,28	0,1103	0,71	0,2611	1,14	0,3729	1,57	0,4418
0,29	0,1141	0,72	0,2642	1,15	0,3749	1,58	0,4429
0,30	0,117	0,73	0,267	1,16	0,377	1,59	0,444

	9		3		0		1
0,31	0,1217	0,74	0,2703	1,17	0,3790	1,60	0,4452
0,32	0,1255	0,75	0,2734	1,18	0,3810	1,61	0,4463
0,33	0,1293	0,76	0,2764	1,19	0,3830	1,62	0,4474
0,34	0,1331	0,77	0,2794	1,20	0,3849	1,63	0,4484
0,35	0,1368	0,78	0,2823	1,21	0,3869	1,64	0,4495
0,36	0,1406	0,79	0,2852	1,22	0,3883	1,65	0,4505
0,37	0,1443	0,80	0,2881	1,23	0,3907	1,66	0,4515
0,38	0,1480	0,81	0,2910	1,24	0,3925	1,67	0,4525
0,39	0,1517	0,82	0,2939	1,25	0,3944	1,68	0,4535
0,40	0,1554	0,83	0,2967	1,26	0,3962	1,69	0,4545
0,41	0,1591	0,84	0,2995	1,27	0,3980	1,70	0,4554
0,42	0,1628	0,85	0,3023	1,28	0,3997	1,71	0,4564

Продолжение приложения 3

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,72	0,4573	1,94	0,4738	2,32	0,4898	2,76	0,4971
1,73	0,4582	1,95	0,4744	2,34	0,4904	2,78	0,4973
1,74	0,4591	1,96	0,4750	2,36	0,4909	2,80	0,4974

1,75	0,4599	1,97	0,4756	2,38	0,4913	2,82	0,4976
1,76	0,4608	1,98	0,4761	2,40	0,4918	2,84	0,4977
1,77	0,4616	1,99	0,4767	2,42	0,4922	2,86	0,4979
1,78	0,4625	2,00	0,4772	2,44	0,4927	2,88	0,4980
1,79	0,4633	2,02	0,4783	2,46	0,4931	2,90	0,4981
1,80	0,4641	2,04	0,4793	2,48	0,4934	2,92	0,4982
1,81	0,4649	2,06	0,4803	2,50	0,4938	2,94	0,4984
1,82	0,4656	2,08	0,4812	2,52	0,4941	2,96	0,4985
1,83	0,4664	2,10	0,4821	2,54	0,4945	2,98	0,4986
1,84	0,4671	2,12	0,4830	2,56	0,4948	3,00	0,49865
1,85	0,4678	2,14	0,4838	2,58	0,4951	3,20	0,49931
1,86	0,4686	2,16	0,4846	2,60	0,4953	3,40	0,49966
1,87	0,4693	2,18	0,4854	2,62	0,4956	3,60	0,499841
1,88	0,4699	2,20	0,4861	2,64	0,4959	3,80	0,499928
1,89	0,4706	2,22	0,4868	2,66	0,4961	4,00	0,499968
1,90	0,4713	2,24	0,4875	2,68	0,4963	4,50	0,499997
1,91	0,471	2,26	0,488	2,70	0,496	5,00	0,499

	9		1		5		997
1,92	0,4726	2,28	0,4887	2,72	0,4967		
1,93	0,4732	2,30	0,4893	2,74	0,4969		

ПРИЛОЖЕНИЕ 4

Таблица значений $t_\gamma = t(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	2,78	460	861	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,555
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,646	3,439
14	2,16	3,01	4,22	80	1,001	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

ПРИЛОЖЕНИЕ 5

Таблица значений $q = q(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43

12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,211
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

ПРИЛОЖЕНИЕ 6

Критические точки распределения χ^2

Число степеней свободы0	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,039	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	0,31	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	4,26	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

ПРИЛОЖЕНИЕ 7

Список основных формул

1. Случайные события

Классическое определение вероятности: $P(A)=\dfrac{m}{n}$.

Формулы сложения вероятностей:

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

$$P(A+B)=P(A)+P(B)-P(A\cdot B).$$

Формулы умножения вероятностей:

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

$$P(A\cdot B)=P(A)\cdot P_A(B)=P(B)\cdot P_B(A).$$

Формула полной вероятности:

$$P(A)=P(H_1)\cdot P_{H_1}(A)+P(H_2)\cdot P_{H_2}(A)+\ldots+P(H_n)\cdot P_{H_n}(A).$$

Формула Бейеса:

$$P_A(H_i)=\frac{P(H_i)\cdot P_{H_i}(A)}{P(A)}.$$

Формула Бернулли:

$$P_n(m)=C_n^m\cdot p^m\cdot q^{n-m};\ (m=\overline{0,n}).$$

Формула Пуассона:

$$P_n(m)\approx\frac{\lambda^m}{m!}e^{-\lambda}.$$

Локальная формула Лапласа:

$$P_n(m)\approx\frac{1}{\sqrt{npq}}\cdot\varphi\left(\frac{m-np}{\sqrt{npq}}\right).$$

Интегральная формула Лапласа:

$$P_n(m_1\le m\le m_2)\approx\Phi\left(\frac{m_2-np}{\sqrt{npq}}\right)-\Phi\left(\frac{m_1-np}{\sqrt{npq}}\right).$$

2. Случайные величины

Биномиальное распределение:

$$P_n(X=m)=C_n^m\cdot p^m\cdot q^{n-m};\ (m=\overline{0,n}).$$

Пуассоновское распределение:

$$P_n(X=m)\approx\frac{\lambda^m}{m!}e^{-\lambda},\quad \lambda=n\cdot p.$$

Геометрическое распределение:

$$P_n(X=m)=p\cdot q^{m-1}.$$

Гипергеометрическое распределение:

$$P_n(X=m)=\frac{C_s^m\cdot C_{n-s}^{r-m}}{C_n^r}.$$

Числовые характеристики дискретных случайных величин:

$$M(X)=\sum_{i=1}^{n}x_i\cdot p_i;\ D(X)=M(X^2)-(M(X))^2;\ \sigma(X)=\sqrt{D(X)}.$$

Числовые характеристики непрерывных случайных величин:

$$M(X)=\int_{-\infty}^{+\infty}x\cdot f(x)dx;\ D(X)=M(X^2)-(M(X))^2;$$

$$M(X^2)=\int_{-\infty}^{+\infty}x^2\cdot f(x)dx;\ \sigma(X)=\sqrt{D(X)}.$$

Нормальное распределение:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{\frac{-(x-a)^2}{2\sigma^2}};$$

$$P(\alpha<X<\beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right);\ P(|X-a|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Закон больших чисел:

Неравенство Маркова $\quad P(X\le\varepsilon)\ge1-\frac{a}{\varepsilon}.$

Неравенство Чебышева $\quad P(|X-M(X)|\le\varepsilon)\ge1-\frac{D(X)}{\varepsilon^2}.$

Теорема Чебышева $\quad P(|\overline{X}-M(\overline{X})|\le\varepsilon)\ge1-\frac{c}{n\cdot\varepsilon^2}.$

3. Математическая статистика

Точечные оценки параметров распределения:

$$\overline{X}_в = \frac{\sum\limits_{i=1}^{k} x_i \cdot n_i}{n}; D_в = \overline{X^2}_в - (\overline{X}_в)^2; S^2 = \frac{n}{n-1} \cdot D_в; \sigma_в = \sqrt{D_в}.$$

Интервальные оценки параметров распределения:

$$\overline{X}_в - t\frac{\sigma}{\sqrt{n}} < a < \overline{X}_в + t\frac{\sigma}{\sqrt{n}}; \qquad \overline{X}_в - t_\gamma \frac{s}{\sqrt{n}} < a < \overline{X}_в + t_\gamma \frac{s}{\sqrt{n}};$$

$$s(1-q) < \sigma < s(1+q); \quad 0 < \sigma < s(1+q);$$

Выборочные уравнения регрессии:

$$\overline{X}_y - \overline{x} = r_в \frac{\sigma_x}{\sigma_y}(y - \overline{y}); \qquad \overline{Y}_x - \overline{y} = r_в \frac{\sigma_y}{\sigma_x}(x - \overline{x}).$$

Наблюдаемое значение критерия Пирсона: $\chi^2_{набл.} = \sum \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}}$.