Костанайский государственный педагогический институт
Естественно-математический факультет
Кафедра физико-математических и общетехнических дисциплин

Калжанов М.У.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Костанай, 2018

УДК 519.2(075.8)
ББК 22.171.я 73
К17

Составитель:
Калжанов М. У. – к.ф.-м.н., доцент кафедры физико-математических и общетехнических дисциплин

Рецензенты:
Тастанов М.Г. – кандидат физико-математических наук, доцент кафедры информатики, профессор КГУ им. А.Байтурсынова
Касымова А.Г. – кандидат физико-математических наук, доцент кафедры физико-математических и общетехнических дисциплин

К 17 Калжанов М.У. Теория вероятностей и математическая статистика. Учебное пособие. / Калжанов М.У. – Костанай, 2018. - 141 с.

ISBN 978-601-7934-50-7

В учебное пособие включены основные понятия и утверждения теории вероятностей и математической статистики. В нем, наряду с теоретическим материалом, представлены задачи прикладного характера и множество примеров, которые непосредственно применяются на практических занятиях.

Предназначен для студентов специальностей 5В010900-Математика, 5В011000-Физика, изучающих курс теории вероятностей и математической статистики.

УДК 519.2(075.8)
ББК 22.171.я 73

Печатается по решению Научно-методического Совета Костанайского государственного педагогического института

ISBN 978-601-7934-50-7

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

«Теория вероятностей и математическая статистика» являетсядисциплиной, в которой изложены основныесведения, построение и анализ математических моделей, учитывающих случайные факторы. **Основной задачей** является ознакомление студентов с основами теории вероятностей и математической статистики в рамках конечномерных случайных величин без строгого применения теории меры и функционального анализа.

Вучебном пособииданы основные фундаментальные понятия теории вероятностей, аксиоматическое обоснование теории вероятностей и вытекающие из этого свойства вероятности; свойств случайных величин и их распределении схему Бернулли и связанные с ним предельные теоремы, числовые характеристики случайных величин; цени Маркова, характеристические функции; законы больших чисел и центральную предельную теорему, условные распределения вероятностей и условные математические ожидания; основные понятия математической статистики; точные оценки и методы их получения; классификацию оценок; интервальные оценки; основы теории проверки гипотез; корреляционную теорию случайных процессов; основные понятия Марковского процесса, свойства винеровского и пуассоновского процессов.

Представленытрадиционные математические модели, правильно отражающие те или иные стороны реальных случайных явлений, решениезадач на классическое и геометрическое определения вероятности; нахождениезаконараспределений функций от случайных величин; вычислениемоментовслучайных величин; применениепростейших вариантовцентральных предельных теорем к конкретным модельным задачам; находитьэмпирические функции распределения, выборочные моменты оценки методов наименьших квадратов, максимального правдоподобия; строить доверительные интервалы для неизвестных параметров; биномиальной и нормальной совокупности; вычислить простейшие стохастические интегралы; находить предельные распределения для простейших цепей Маркова; находить распределения простейших функционалов от Марковских процессов и т.д.

1. ПРЕДМЕТ ТЕОРИИ ВЕРОЯТНОСТЕЙ. СЛУЧАЙНЫЕ СОБЫТИЯ. АЛГЕБРА СОБЫТИЙ

В различных разделах науки и техники нередко возникают ситуации, когда результат каждого из многих проводимых опытов заранее предугадать невозможно, однако можно исследовать закономерности, возникающие при проведении серии опытов. Нельзя, например, точно сказать, какая сторона монеты окажется сверху при данном броске: герб или цифра – но при большом количестве бросков число выпадений герба приближается к половине количества бросков; нельзя заранее предсказать результат одного выстрела из данного орудия по данной цели, но при большом числе выстрелов частота попадания приближается к некоторому постоянному числу. Исследование вероятностных закономерностей массовых однородных явлений составляет предмет **теории вероятностей**.

Основным интуитивным понятием классической теории вероятностей является **случайное событие**. События, которые могут произойти в результате опыта, можно подразделить на три вида:

а) **достоверное событие** – событие, которое всегда происходит при проведении опыта;

б) **невозможное событие** – событие, которое в результате опыта произойти не может;

в) **случайное событие** – событие, которое может либо произойти, либо не произойти. Например, при броске игральной кости достоверным событием является выпадение числа очков, не превышающего 6, невозможным – выпадение 10 очков, а случайным – выпадение 3 очков.

Алгебра событий

Определение 1.1. **Суммой** $\boldsymbol{A+B}$двух событий A и Bназывают событие, состоящее в том, что произошло хотя бы одно из событий Aи B. **Суммой нескольких событий**, соответственно, называется событие, заключающееся в том, что произошло хотя бы одно из этих событий.

Пример 1. Два стрелка делают по одному выстрелу по мишени. Если событие A – попадание первого стрелка, а событие B – второго, то сумма $A+B$ – это хотя бы одно попадание при двух выстрелах.

Пример 2. Если при броске игральной кости событием A_iназвать выпадение i очков, то выпадение нечетного числа очков является суммой событий $A+A_2+A_3$.

Назовем все возможные результаты данного опыта его *исходами* и предположим, что множество этих исходов, при которых происходит событие A(исходов, *благоприятных*событию A), можно представить в виде некоторой области на плоскости. Тогда множество исходов, при которых произойдет событие $A+B$, является объединением множеств исходов, благоприятных событиям A или B (см. Рисунок 1.1).

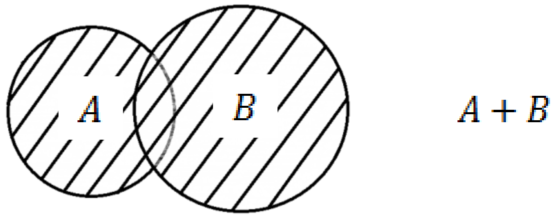

Рисунок 1.1

Определение 1.2. **Произведением $\boldsymbol{AB}$** событий A и B называется событие, состоящее в том, что произошло и событие A, и событие B. Аналогично **произведением нескольких событий** называется событие, заключающееся в том, что произошли все эти события.

Пример 3. В примере 1 (два выстрела по мишени) событием AB будет попадание обоих стрелков.

Пример 4. Если событие A состоит в том, что из колоды карт извлечена карта пиковой масти, а событие B – в том, что из колоды вынута дама, то событием AB будет извлечение из колоды дамы пик.

Геометрической иллюстрацией множества исходов опыта, благоприятных появлению произведения событий A и B, является пересечение областей, соответствующих исходам, благоприятным A и B (см. Рисунок 1.2).

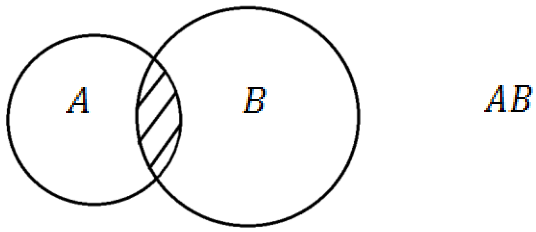

Рисунок 1.2

Определение 1.3. **Разностью $\boldsymbol{A \backslash B}$** событий A и B называется событие, состоящее в том, что A произошло, а B – нет.

*Пример*5. Вернемся к примеру 1, где $A \backslash B$– попадание первого стрелка при промахе второго.

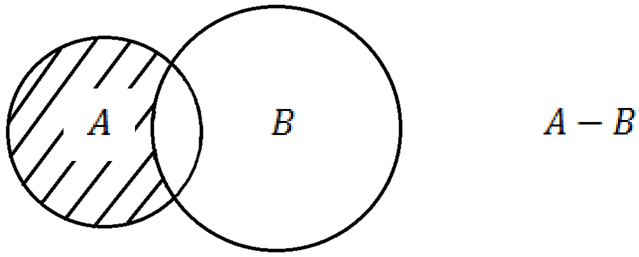

Рисунок 1.3

Пример 6. В примере 4$A\B$– извлечение из колоды любой карты пиковой масти, кроме дамы. Наоборот, $B\backslash A$ – извлечение дамы любой масти, кроме пик (см. Рисунок 1.3).

Введем еще несколько категорий событий.

Определение 1.4. События A и B называются **совместными**, если они могут произойти оба в результате одного опыта. В противном случае (то есть если они не могут произойти одновременно) события называются **несовместными.**

Примеры: совместными событиями являются попадания двух стрелков в примере 1 и появление карты пиковой масти и дамы в примере 4; несовместными – события $A_1 - A_6$в примере 2.

Замечание 1. Если изобразить графически области исходов опыта, благоприятных несовместным событиям, то они не будут иметь общих точек.

Замечание 2. Из определения несовместных событий следует, что их произведение является невозможным событием.

Определение 1.5. Говорят, что события $A_1, A_2, \ldots, A_n$образуют **полную группу**, если в результате опыта обязательно произойдет хотя бы одно из событий этой группы.

Замечание. В частности, если события, образующие полную группу, попарно несовместны, то в результате опыта произойдет *одно и только одно* из них. Такие события называют **элементарными событиями**.

Пример. В примере 2 события $A_1 - A_6$(выпадение одного, двух,…, шести очков при одном броске игральной кости) образуют полную группу несовместных событий.

Определение 1.6. События называются **равновозможными**, если нет оснований считать, что одно из них является более возможным, чем другое.

Примеры: выпадение любого числа очков при броске игральной кости, появление любой карты при случайном извлечении из колоды, выпадение герба или цифры при броске монеты и т.п.

Классическое определение вероятности

При изучении случайных событий возникает необходимость количественно сравнивать возможность их появления в результате опыта. Например, при последовательном извлечении из колоды пяти карт более возможна ситуация, когда появились карты разных мастей, чем появление пяти карт одной масти; при десяти бросках монеты более возможно чередование гербов и цифр, нежели выпадение подряд десяти гербов, и т.д. Поэтому с каждым таким событием связывают по определенному правилу некоторое число, которое тем больше, чем более возможно событие. Это число называется **вероятностью события** и является вторым основным понятием теории вероятностей.

Отметим, что само понятие вероятности, как и понятие случайного события, является аксиоматическим и поэтому не поддается строгому определению. То, что в дальнейшем будет называться различными определениями вероятности, представляет собой способы вычисления этой величины.

Определение 1.7. Если все события, которые могут произойти в результате данного опыта,

а) попарно несовместны;

б) равновозможны;

в) образуют полную группу,

то говорят, что имеет место **схема случаев.**

Можно считать, что случаи представляют собой все множество исходов опыта. Пусть их число равно n (число возможных исходов), а при m из них происходит некоторое событие A (число благоприятных исходов).

Определение 1.8. **Вероятностью события** A называется отношение числа исходов опыта, благоприятных этому событию, к числу возможных исходов:

$$p(A) = \frac{m}{n} \tag{1.1}$$

–*классическое определение вероятности*.

Свойства вероятности.

Из определения 1.8 вытекают следующие свойства вероятности:

Свойство 1. Вероятность достоверного события равна единице.

Доказательство. Так как достоверное событие всегда происходит в результате опыта, то все исходы этого опыта являются для него благоприятными, то есть $m = n$, следовательно,

$$p(A) = 1.$$

Свойство 2. Вероятность невозможного события равна нулю.

Доказательство. Для невозможного события ни один исход опыта не является благоприятным, поэтому $m = 0$ и $p(A) = 0$.

Свойство 3. Вероятность случайного события есть положительное число, заключенное между нулем и единицей.

Доказательство. Случайное событие происходит при некоторых исходах опыта, но не при всех, следовательно, $0 < m < n$, и из (1.1) следует, что $0 < p(A) < 1$.

Пример. Из урны, содержащей 6 белых и 4 черных шара, наудачу вынут шар. Найти вероятность того, что он белый.

Решение. Будем считать элементарными событиями, или исходами опыта, извлечение из урны каждого из имеющихся в ней шаров. Очевидно, что эти события удовлетворяют всем условиям, позволяющим считать их схемой случаев. Следовательно, число возможных исходов равно 10, а число исходов, благоприятных событиюA (появлению белого шара) – 6 (таково количество белых шаров в урне). Значит,

$$p(A) = \frac{m}{n} = \frac{6}{10} = 0{,}6.$$

Относительная частота. Статистическое определение вероятности

Классическое определение вероятности применимо только для очень узкого класса задач, где все возможные исходы опыта можно свести к схеме случаев. В большинстве реальных задач эта схема неприменима. В таких ситуациях требуется определять вероятность события иным образом. Для этого введем вначале понятие **относительной частоты**$\boldsymbol{W(A)}$события A как отношения числа опытов, в которых наблюдалось событие A, к общему количеству проведенных испытаний:

$$W(A) = \frac{M}{N} \qquad (1.2)$$

где N – общее число опытов, M – число появлений событияA.

Большое количество экспериментов показало, что если опыты проводятся в одинаковых условиях, то для большого количества испытаний относительная частота изменяется мало, колеблясь около некоторого постоянного числа. Это число можно считать вероятностью рассматриваемого события.

Определение 1.9. **Статистической вероятностью события** считают его относительную частоту или число, близкое к ней.

Замечание 1. Из формулы (1.2) следует, что свойства вероятности, доказанные для ее классического определения, справедливы и для статистического определения вероятности.

Замечание 2. Для существования статистической вероятности события A требуется:

1) возможность производить неограниченное число испытаний;
2) устойчивость относительных частот появления A в различных сериях достаточно большого числа опытов.

Замечание 3. Недостатком статистического определения является неоднозначность статистической вероятности.

Пример. Если в задаче задается вероятность попадания в мишень для данного стрелка (скажем, $p = 0{,}7$), то эта величина получена в результате изучения статистики большого количества серий выстрелов, в которых этот стрелок попадал в мишень около семидесяти раз из каждой сотни выстрелов.

Основные формулы комбинаторики

При вычислении вероятностей часто приходится использовать некоторые формулы *комбинаторики* – науки, изучающей комбинации, которые можно составить по определенным правилам из элементов некоторого конечного множества. Определим основные такие комбинации.

Определение 1.10. **Перестановки**– это комбинации, составленные из всех n элементов данного множества и отличающиеся только порядком их расположения. Число всех возможных перестановок

$$P_n = n! \tag{1.3}$$

Пример. Сколько различных списков (отличающихся порядком фамилий) можно составить из 7 различных фамилий?

Решение. $P_7 = 7! = 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040.$

Определение 1.11. **Размещения** – комбинации из m элементов множества, содержащего n различных элементов, отличающиеся либо составом элементов, либо их порядком. Число всех возможных размещений

$$A_n^m = n(n-1)(n-2)\ldots(n-m+1) \tag{1.4}$$

Пример. Сколько возможно различных вариантов пьедестала почета (первое, второе, третье места), если в соревнованиях принимают участие 10 человек?

Решение.$A_{10}^3 = 10 \cdot 9 \cdot 8 = 720.$

Определение 1.12. **Сочетания** – неупорядоченные наборы из m элементов множества, содержащего n различных элементов (то есть наборы, отличающиеся только составом элементов). Число сочетаний

$$C_n^m = \frac{n!}{m!(n-m)!}. \tag{1.5}$$

Пример. В отборочных соревнованиях принимают участие 10 человек, из которых в финал выходят трое. Сколько может быть различных троек финалистов?

Решение. В отличие от предыдущего примера, здесь не важен порядок финалистов, следовательно, ищем число сочетаний из 10 по 3:

$$C_{10}^3 = \frac{10!}{3!\,7!} = \frac{8 \cdot 9 \cdot 10}{6} = 120.$$

2. ГЕОМЕТРИЧЕСКИЕ ВЕРОЯТНОСТИ. ТЕОРЕМА СЛОЖЕНИЯ ВЕРОЯТНОСТЕЙ. ПРОТИВОПОЛОЖНЫЕ СОБЫТИЯ

Одним из недостатков классического определения вероятности является то, что оно неприменимо к испытаниям с бесконечным количеством исходов. В таких случаях можно воспользоваться понятием **геометрической вероятности**.

Пусть на отрезок L наудачу брошена точка. Это означает, что точка обязательно попадет на отрезок L и с равной возможностью может совпасть с любой точкой этого отрезка. При этом вероятность попадания точки на любую часть отрезка Lне зависит от расположения этой части на отрезке и пропорциональна его длине. Тогда вероятность того, что брошенная точка попадет на отрезок *l*, являющийся частью отрезка L, вычисляется по формуле:

$$p = \frac{l}{L} \tag{2.1}$$

где l– длина отрезкаl, а L– длина отрезка L.

Можно дать аналогичную постановку задачи для точки, брошенной на плоскую областьS и вероятности того, что она попадет на часть этой области s:

$$p = \frac{s}{S} \qquad (2.1`)$$

где s – площадь части области, аS– площадь всей области.

В трехмерном случае вероятность того, что точка, случайным образом расположенная в теле V, попадет в его частьv, задается формулой:

$$p = \frac{\mathrm{v}}{V} \qquad (2.1``)$$

где v– объем части тела, аV– объем всего тела.

Пример 1. Найти вероятность того, что точка, наудачу брошенная в круг, не попадет в правильный шестиугольник, вписанный в него.

Решение. Пусть радиус круга равен R, тогда сторона шестиугольника тоже равна R. При этом площадь круга$S = \pi R^2$, а площадь шестиугольника$s = \frac{3\sqrt{3}}{2}R^2$. Следовательно,

$$p = \frac{\mathrm{S} - \mathrm{s}}{S} = \frac{\pi R^2 - \frac{3\sqrt{3}}{2}R^2}{\pi R^2} = \frac{\pi - 3\sqrt{3}}{2\pi} \approx 0{,}174.$$

Пример 2. На отрезок ABслучайным образом брошены три точки: C,DиM Найти вероятность того, что из отрезков AC, ADиAM можно построить треугольник.

Решение. Обозначим длины отрезков AC, ADи AMчерезx, yиzи рассмотрим в качестве возможных исходов множество точек трехмерного пространства с координатами (x,y,z). Если принять длину отрезка равной 1, то эти множество возможных исходов представляет собой куб с ребром, равным 1. Тогда множество благоприятных исходов состоит из точек, для координат которых выполнены неравенства треугольника:

$$x+y>z, x+z>y, y+z>x.$$

Это часть куба, отрезанная от него плоскостями

$$x+y=z, x+z=y, y+z=x$$

(одна из них, плоскость $x+y=z$,проведена на рисунке 2.1). Каждая такая плоскость отделяет от куба пирамиду, объем которой равен

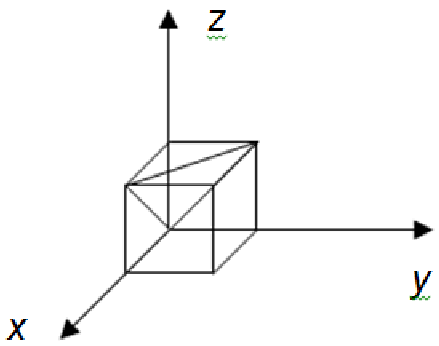

Рисунок 2.1

$$\frac{1}{3}\cdot\frac{1}{2}\cdot 1=\frac{1}{6}.$$

Следовательно, объем оставшейся части

$$\mathrm{v}=1-3\cdot\frac{1}{6}=\frac{1}{2}.$$

Тогда

$$p=\frac{\mathrm{v}}{V}=\frac{1}{2}:1=\frac{1}{2}.$$

Теорема сложения вероятностей

Теорема 2.1 (теорема сложения).Вероятность$P(A+B)$суммы событий A и B равна

$$P(A+B)=p(A)+p(B)-p(AB). \qquad (2.2)$$

Доказательство.

Докажем теорему сложения для схемы случаев. Пусть n – число возможных исходов опыта, m_A– число исходов, благоприятных событию A, m_B– число исходов, благоприятных событию B, а m_{AB}– число исходов опыта, при которых происходят оба события (то есть исходов, благоприятных произведению AB). Тогда число исходов, при которых имеет место событие $A+B$, равно$m_A+m_B-m_{AB}$(так как в сумме $(m_A+m_B)m_{AB}$учтено дважды: как исходы, благоприятные A, и исходы, благоприятныеB). Следовательно, вероятность суммы можно определить по формуле (1.1):

$$P(A + B) = \frac{m_A + m_B - m_{AB}}{n} = \frac{m_A}{n} + \frac{m_B}{n} - \frac{m_{AB}}{n} == p(A) + p(B) - p(AB),$$

что и требовалось доказать.

Следствие 1. Теорему 2.1 можно распространить на случай суммы любого числа событий. Например, для суммы трех событий A, BиC

$$P(A + B + C) = p(A) + p(B) + p(C) - p(AB) - p(AC) - p(BC) + +p(ABC). \quad (2.3)$$

и т.д.

Следствие 2. Если события A, иBнесовместны, то $m_{AB} = 0$, и, следовательно, вероятность суммы несовместных событий равна сумме их вероятностей:

$$P(A + B) = p(A) + p(B). \quad (2.4)$$

*Определение 2.1.***Противоположными событиями** называют два несовместных события, образующих полную группу. Если одно из них назвать A, то второе принято обозначать$\overline{A}$.

Замечание. Таким образом, $\overline{A}$заключается в том, что событие A не произошло.

Теорема 2.2. Сумма вероятностей противоположных событий равна 1:

$$p(A) + p(\overline{A}) = 1. \quad (2.5)$$

Доказательство.

Так как A и $\overline{A}$образуют полную группу, то одно из них обязательно произойдет в результате опыта, то есть событие $A + \overline{A}$является достоверным. Следовательно,$P(A + \overline{A}) = 1$.Но, так как A и$\overline{A}$ несовместны, из (2.4) следует, что $P(A + \overline{A}) = p(A) + p(\overline{A})$. Значит,$p(A) + p(\overline{A}) = 1$, что и требовалось доказать.

Замечание. В ряде задач проще искать не вероятность заданного события, а вероятность события, противоположного ему, а затем найти требуемую вероятность по формуле (2.5).

Пример. Из урны, содержащей 2 белых и 6 черных шаров, случайным образом извлекаются 5 шаров.Найти вероятность того, что вынуты шары разных цветов.

Решение. Событие$\overline{A}$, противоположное заданному, заключается в том, что из урны вынуто 5 шаров одного цвета, а так как белых шаров в ней всего два, то этот цвет может быть только черным. Множество возможных исходов опыта найдем по формуле (1.5):

$$n = C_8^5 = \frac{8!}{5! \cdot 3!} = \frac{6 \cdot 7 \cdot 8}{6} = 56,$$

а множество исходов, благоприятных событию $\overline{A}$– это число возможных наборов по 5 шаров только из шести черных:

$$m_{\overline{A}} = C_6^5 = 6.$$

Тогда

$$P(\overline{A}) = \frac{6}{56} = \frac{3}{28}, \text{ а } P(A) = 1 - \frac{3}{28} = \frac{25}{28}.$$

Теорема умножения вероятностей

Определение 2.2. Назовем **условной вероятностью** $\boldsymbol{p(B/A)}$ **события** $\boldsymbol{B}$ вероятность события B при условии, что событие A произошло.

Замечание. Понятие условной вероятности используется в основном в случаях, когда осуществление события A изменяет вероятность события B.

Примеры:

1) пусть событие A – извлечение из колоды в 32 карты туза, а событие B – то, что и вторая вынутая из колоды карта окажется тузом. Тогда, если после первого раза карта была возвращена в колоду, то вероятность вынуть вторично туз не меняется: $p(B) = p(A) = \frac{4}{32} = \frac{1}{8} = 0{,}125$. Если же первая карта в колоду не возвращается, то осуществление события A приводит к тому, что в колоде осталась 31 карта, из которых только 3 туза. Поэтому $p(B/A) = \frac{3}{31} \approx 0{,}097$.

2) если событие A – попадание в самолет противника при первом выстреле из орудия, а B – при втором, то первое попадание уменьшает маневренность самолета, поэтому $p(B/A)$ увеличится по сравнению с $p(A)$.

Теорема 2.3 (теорема умножения). Вероятность произведения двух событий равна произведению вероятности одного из них на условную вероятность другого при условии, что первое событие произошло:

$$p(AB) = p(A) \cdot p(B/A). \tag{2.6}$$

Доказательство.

Воспользуемся обозначениями теоремы 2.1. Тогда для вычисления $p(B/A)$ множеством возможных исходов нужно считать

m_A(так как Aпроизошло), а множеством благоприятных исходов – те, при которых произошли и A, и B (m_{AB}). Следовательно,

$$p(B/A) = \frac{m_{AB}}{m_A} = \frac{m_{AB}}{n} \cdot \frac{n}{m_A} = p(AB)/p(A),$$

откуда следует утверждение теоремы.

Пример. Для поражения цели необходимо попасть в нее дважды. Вероятность первого попадания равна 0,2, затем она не меняется при промахах, но после первого попадания увеличивается вдвое. Найти вероятность того, что цель будет поражена первыми двумя выстрелами.

Решение. Пусть событие A – попадание при первом выстреле, а событие B – попадание при втором. Тогда $p(A) = 0{,}2, p(B/A) = 0{,}4, p(AB) = 0{,}2 \cdot 0{,}4 = 0{,}8$.

*Следствие.*Если подобным образом вычислить вероятность события BA, совпадающего с событием AB, то получим, что

$$(BA) = p(B) \cdot p(A/B).$$

Следовательно,

$$p(A) \cdot p(B/A) = p(B) \cdot p(A/B). \qquad (2.7)$$

Определение 2.3. Событие B называется **независимым** от события A, если появление события A не изменяет вероятности B, то есть$p(B/A) = p(B)$.

Замечание. Если событие B не зависит отA, то и A не зависит от B. Действительно, из (2.7) следует при этом, что$p(A) \cdot p(B) = p(B) \cdot p(A/B)$,откуда $p(A/B) = p(A)$. Значит, **свойство независимости событий взаимно**.

Теорема умножения для независимых событий имеет вид:

$$p(AB) = p(A) \cdot p(B), \qquad (2.8)$$

то есть вероятность произведения независимых событий равна произведению их вероятностей.

При решении задач теоремы сложения и умножения обычно применяются вместе.

Пример. Два стрелка делают по одному выстрелу по мишени. Вероятности их попадания при одном выстреле равны соответственно 0,6 и 0,7. Найти вероятности следующих событий:

A – хотя бы одно попадание при двух выстрелах;

B– ровно одно попадание при двух выстрелах;

C – два попадания;

D– ни одного попадания.

Решение. Пусть событие H_1– попадание первого стрелка, H_2 – попадание второго. Тогда

$$A = H_1 + H_2, B = H_1 \cdot \overline{H}_2 + \overline{H}_1 \cdot H_2, C = H_1 \cdot H_2, D = \overline{H}_1 \cdot \overline{H}_2.$$

События H_1 и H_2 совместны и независимы, поэтому теорема сложения применяется в общем виде, а теорема умножения – в виде (2.8). Следовательно,

$$p(C) = 0{,}6 \cdot 0{,}7 = 0{,}42, p(A) = 0{,}6 + 0{,}7 - 0{,}42 = 0{,}88,$$
$$p(B) = 0{,}6 \cdot 0{,}3 + 0{,}7 \cdot 0{,}4 = 0{,}46,$$

(так как события $H_1 \cdot \overline{H}_2$и $\overline{H}_1 \cdot H_2$несовместны), $p(D) = 0{,}4 \cdot 0{,}3 = 0{,}12$.Заметим, что события A и D являются противоположными, поэтому

$$p(A) = 1 - p(D).$$

Вероятность появления хотя бы одного события

Теорема 2.4. Вероятность появления хотя бы одного из попарно независимых событий$A_1, A_2, \dots, A_n$ равна

$$p(A) = 1 - q_1 q_2 \dots q_n, \quad (2.9)$$

где q_i– вероятность события $\overline{A}_i$, противоположного событию A_i.

Доказательство.

Если событие A заключается в появлении хотя бы одного события из $A_1, A_2, \dots, A_n$, то события A и $\overline{A}_1, \overline{A}_2, \dots, \overline{A}_n$ противоположны, поэтому по теореме 2.2 сумма их вероятностей равна 1. Кроме того, поскольку $A_1, A_2, \dots, A_n$независимы, то независимы и $\overline{A}_1, \overline{A}_2, \dots, \overline{A}_n$, следовательно,$p(\overline{A}_1 \overline{A}_2 \dots \overline{A}_n) = p(\overline{A}_1) p(\overline{A}_2) \dots p(\overline{A}_n) = q_1 q_2 \dots q_n$. Отсюда следует справедливость формулы (2.9).

Пример. Сколько нужно произвести бросков монеты, чтобы с вероятностью не менее 0,9 выпал хотя бы один герб?

Решение. Вероятность выпадения герба при одном броске равна вероятности противоположного события (выпадения цифры) и равна 0,5. Тогда вероятность выпадения хотя бы одного герба при n выстрелах равна $1 - (0{,}5)^n$. Тогда из решения неравенства $1 - (0{,}5)^n > 0{,}9$следует, что $n > log_2 10 \geq 4$.

3. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ И ФОРМУЛА БАЙЕСА. СХЕМА И ФОРМУЛА БЕРНУЛЛИ

Определение 3.1. Пусть событие A может произойти только совместно с одним из событий $H_1, H_2, \dots, H_n$, образующих полную группу несовместных событий. Такие события $H_1, H_2, \dots, H_n$, называются **гипотезами.**

Теорема 3.1. Вероятность события A, наступающего совместно с гипотезами $H_1, H_2, \dots, H_n$, равна:

$$p(A) = \sum_{i=1}^{n} p(H_i)\, p(A/H_i), \tag{3.1}$$

где $p(H_i)$ – вероятность -й гипотезы, а$p(A/H_i)$– вероятность события A при условии выполнения этой гипотезы. Формула (3.1) носит название **формулы полной вероятности**.

Доказательство.

Можно определить событие A как сумма попарно несовместных событий$AH_1, AH_2, \dots, AH_n$, Тогда из теорем сложения и умножения вытекает, что

$$p(A) = p(AH_1 + AH_2 + \dots + AH_n) =$$
$$= p(AH_1) + p(AH_2) + \dots + p(AH_n) = \sum_{i=1}^{n} p(H_i)\, p(A/H_i),$$

что и требовалось доказать.

Пример. Имеются три одинаковые урны с шарами. В первой из них 3 белых и 4 черных шара, во второй – 2 белых и 5 черных, в третьей – 10 черных шаров. Из случайно выбранной урны наудачу вынут шар. Найти вероятность того, что он белый.

Решение. Будем считать гипотезами H_1, H_2и H_3выбор урны с соответствующим номером. Так как по условию задачи все гипотезы равновозможны, то$p(H_1) = p(H_2) = p(H_3) = \frac{1}{3}$. Найдем условную вероятность A при реализации каждой гипотезы:

$$p(A/H_i) = \frac{3}{7}, p(A/H_2) = \frac{2}{7}, p(A/H_3) = 0.$$

Тогда

$$p(A) = \frac{1}{3} \cdot \frac{3}{7} + \frac{1}{3} \cdot \frac{2}{7} + \frac{1}{3} \cdot 0 = \frac{5}{21} \approx 0{,}238.$$

Формула Байеса (теорема гипотез)

Пусть известен результат опыта, а именно то, что произошло событие A. Этот факт может изменить априорные (то есть известные до опыта) вероятности гипотез. Например, в предыдущем примере извлечение из урны белого шара говорит о том, что этой урной не могла быть третья, в которой нет белых шаров, то есть $p(H_3/A) = 0$.Для переоценки вероятностей гипотез при известном результате опыта используется **формула Байеса**:

$$p(H_i/A) = \frac{p(H_i)p(A/H_i)}{p(A)}. \tag{3.2}$$

Действительно, из (2.7) получим, что$p(A)p(H_i/A) = p(H_i)p(A/H_i)$,откуда следует справедливость формулы (3.2).

Пример. После двух выстрелов двух стрелков, вероятности попаданий которых равны 0,6 и 0,7, в мишени оказалась одна пробоина. Найти вероятность того, что попал первый стрелок.

Решение. Пусть событие A – одно попадание при двух выстрелах, а гипотезы: H_1 – первый попал, а второй промахнулся, H_2 – первый промахнулся, а второй попал, H_3 – оба попали, H_4 – оба промахнулись. Вероятности гипотез: $p(H_1) = 0{,}6 \cdot 0{,}3 = 0{,}18, p(H_2) = 0{,}4 \cdot 0{,}7 = 0{,}28, p(H_3) = 0{,}6 \cdot 0{,}7 = 0{,}42, p(H_4) = 0{,}4 \cdot 0{,}3 = 0{,}12.$

Тогда

$$p(A/H_1) = p(A/H_2) = 1, p(A/H_3) = p(A/H_4) = 0.$$

Следовательно, полная вероятность

$$p(A) = 0{,}18 \cdot 1 + 0{,}28 \cdot 1 + 0{,}42 \cdot 0 + 0{,}12 \cdot 0 = 0{,}46.$$

Применяя формулу Байеса, получим:

$$p(H_1/A) = \frac{0{,}18 \cdot 1}{0{,}46} = \frac{9}{23} \approx 0{,}391.$$

Схема повторения испытаний. Формула Бернулли

Рассмотрим серию из n испытаний, в каждом из которых событие A появляется с одной и той же вероятностью p, причем результат каждого испытания не зависит от результатов остальных. Подобная постановка задачи называется **схемой повторения испытаний**. Найдем вероятность того, что в такой серии событие A произойдет ровно k раз (неважно, в какой последовательности). Интересующее нас событие представляет собой сумму равновероятных несовместных событий, заключающихся в том, что A произошло в некоторых k испытаниях и не произошло в остальных$n - k$ испытаниях. Число таких событий равно числу сочетаний изnпоk, то есть C_n^k, а вероятность каждого из них: $p^k q^{n-k}$, где $q = 1 - p$– вероятность того, что в данном опыте A не произошло. Применяя теорему сложения для несовместных событий, получим **формулу Бернулли**:

$$p_n(k) = C_n^k \cdot p^k \cdot q^{n-k}. \qquad (3.3)$$

Пример. Для получения приза нужно собрать 5 изделий с особым знаком на этикетке. Найти вероятность того, что придется купить 10 изделий, если этикетки с этим знаком имеют 5% изделий.

Решение. Из постановки задачи следует, что последнее купленное изделие имеет особый знак. Следовательно, из предыдущих девяти эти знаки имели 4 изделия. Найдем вероятность этого по формуле Бернулли:

$$p_9(4) = C_9^4 \cdot (0{,}05)^4 \cdot (0{,}95)^5 = 0{,}0006092.$$

Тогда

$$p = 0{,}0006092 \cdot 0{,}05 = 0{,}0000304.$$

Приближение Пуассона для схемы Бернулли

Формула Бернулли требует громоздких расчетов при большом количестве испытаний. Можно получить более удобную для расчетов приближенную формулу, если при большом числе испытаний вероятность появления A в одном опыте мала, а произведение $np = \lambda$сохраняет постоянное значение для разных серий опытов (то есть среднее число появлений события A в разных сериях испытаний остается неизменным). Применим формулу Бернулли:

$$p_n(k) = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} p^k (1-p)^{n-k} =$$
$$= \frac{n(n-1)\dots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Найдем предел полученного выражения при$n \to \infty$:

$$p_n(k) \approx \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(1 \cdot \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\dots\left(1 - \frac{k-1}{n}\right)\left(1 - \frac{\lambda}{n}\right)^{n-k}\right) =$$

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

Таким образом, **формула Пуассона**

$$p_n(k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{3.4}$$

позволяет найти вероятность k появлений события A для массовых (n велико) и редких(p мало) событий.

4. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ. ЗАКОН РАСПРЕДЕЛЕНИЯ И ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ ДИСКРЕТНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Наряду с понятием случайного события в теории вероятности используется и более удобное понятие *случайной величины*.

Определение 4.1. **Случайной величиной** называется величина, принимающая в результате опыта одно из своих возможных значений, причем заранее неизвестно, какое именно.

Будем обозначать случайные величины заглавными буквами латинского алфавита $(X, Y, Z, \ldots)$, а их возможные значения – соответствующими малыми буквами $(x_i, y_i, z_i, \ldots)$.

Примеры: число очков, выпавших при броске игральной кости; число появлений герба при 10 бросках монеты; число выстрелов до первого попадания в цель; расстояние от центра мишени до пробоины при попадании.

Можно заметить, что множество возможных значений для перечисленных случайных величин имеет разный вид: для первых двух величин оно конечно (соответственно 6 и 11 значений), для третьей величины множество значений бесконечно и представляет собой множество натуральных чисел, а для четвертой – все точки отрезка, длина которого равна радиусу мишени. Таким образом, для первых трех величин множество значений из отдельных (дискретных), изолированных друг от друга значений, а для четвертой оно представляет собой непрерывную область. По этому показателю случайные величины подразделяются на две группы: дискретные и непрерывные.

Определение 4.2. Случайная величина называется **дискретной**, если она принимает отдельные, изолированные возможные значения с определенными вероятностями.

*Определение 4.3.*Случайная величина называется **непрерывной**, если множество ее возможных значений целиком заполняет некоторый конечный или бесконечный промежуток.

Дискретные случайные величины

Для задания дискретной случайной величины нужно знать ее возможные значения и вероятности, с которыми принимаются эти значения. Соответствие между ними называется **законом распределения** случайной величины. Он может иметь вид таблицы, формулы или графика.

Таблица, в которой перечислены возможные значения дискретной случайной величины и соответствующие им вероятности, называется **рядом распределения**:

x_i	x_1	x_2	$\dots$	x_n	$\dots$
p_i	p_1	p_2	$\dots$	p_n	$\dots$

Заметим, что событие, заключающееся в том, что случайная величина примет одно из своих возможных значений, является достоверным, поэтому $\sum_{i=1}^{n(\infty)} = 1$.

Пример. Два стрелка делают по одному выстрелу по мишени. Вероятности их попадания при одном выстреле равны соответственно 0,6 и 0,7. Составить ряд распределения случайной величины X – числа попаданий после двух выстрелов.

Решение. Очевидно, что X может принимать три значения: 0, 1 и 2. Их вероятности найдены в примере, рассмотренном в лекции 3. Следовательно, ряд распределения имеет вид:

x_i	0	1	2
p_i	0,12	0,46	0,42

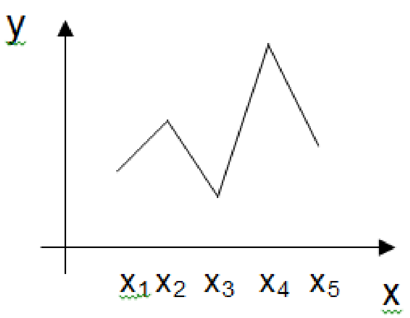

Рисунок 4.1

Графически закон распределения дискретной случайной величины можно представить в виде **многоугольника распределения** – ломаной, соединяющей точки плоскости с координатами (x_i, p_i)(см. Рисунок 4.1).

Функция распределения

Определение 4.4. **Функцией распределения** $F(x)$случайной величины X называется вероятность того, что случайная величина примет значение, меньшее x:

$$F(x) = p(X < x). \qquad (4.1)$$

Свойства функции распределения.

1) $0 \le F(x) \le 1$.Действительно, так как функция распределения представляет собой вероятность, она может принимать только те значения, которые принимает вероятность.

2) Функция распределения является неубывающей функцией, то есть $F(x_2) \ge F(x_1)$при $x_2 > x_1$. Это следует из того, что

$$F(x_2) = p(X < x_2) = p(X < x_1) + p(x_1 \le X < x_2) \ge F(x_1).$$

3) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$. В частности, если все возможные значения X лежат на интервале $[a, b]$, то $F(x) = 0$при$x \le a$и $F(x) = 1$ при $x \ge b$. Действительно, $X < a$– событие невозможное, а $X < b$– достоверное.

4) Вероятность того, что случайная величина примет значение из интервала $[a, b]$, равна разности значений функции распределения на концах интервала:

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

Справедливость этого утверждения следует из определения функции распределения (см. свойство 2).

Для дискретной случайной величины значение $F(x)$ в каждой точке представляет собой сумму вероятностей тех ее возможных значений, которые меньше аргумента функции.

Пример. Найдем $F(x)$ для предыдущего примера:

$$F(x) = \begin{cases} 0, x \le 0 \\ 0{,}12, 0 < x \le 1 \\ 0{,}12 + 0{,}46 = 0{,}58, 1 < x \le 2 \\ 0{,}58 + 0{,}42 = 1, x > 2. \end{cases}$$

Соответственно график функции распределения имеет ступенчатый вид (см. Рисунок 4.2).

Биномиальное распределение

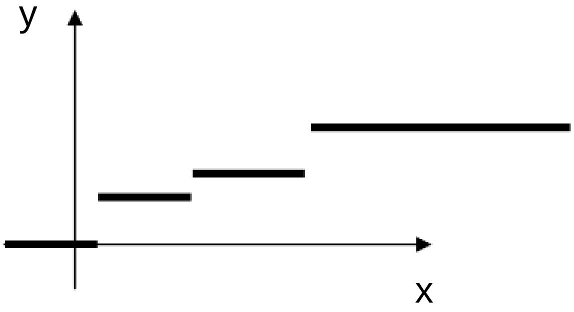

Рисунок 4.2

Вернемся к схеме независимых испытаний и найдем закон распределения случайной величины X – числа появлений события A в серии из n испытаний. Возможные значения A: $0, 1, \ldots, n$. Соответствующие им вероятности можно вычислить по формуле Бернулли:

$$p(X = k) = C_n^k p^k q^{n-k} \qquad (4.2)$$

(p – вероятность появления A в каждом испытании).

Такой закон распределения называют **биномиальным**, поскольку правую часть равенства (4.2) можно рассматривать как общий член разложения бинома Ньютона:

$$(p + q)^n = C_n^n p^n + C_n^{n-1} p^{n-1} q + \cdots + C_n^k p^k q^{n-k} + \cdots + C_n^0 q^n.$$

Пример. Составим ряд распределения случайной величины X – числа попаданий при 5 выстрелах, если вероятность попадания при одном выстреле равна 0,8.

$$p(x = 0) = 1\cdot(0{,}2)^5 = 0{,}00032;\ p(x = 1) = 5\cdot 0{,}8\cdot(0{,}2)^4 = 0{,}0064;$$
$$p(x = 2) = 10\cdot(0{,}8)^2\cdot(0{,}2)^3 = 0{,}0512;\ p(x = 3) = 10\cdot(0{,}8)^3\cdot(0{,}2)^2 = 0{,}2048;$$
$$p(x = 4) = 5\cdot(0{,}8)^4\cdot 0{,}2 = 0{,}4096;\ p(x = 5) = 1\cdot(0{,}8)^5 = 0{,}32768.$$

Таким образом, ряд распределения имеет вид:

x	0	1	2	3	4	5
p	0.00032	0.0064	0.0512	0.2048	0.4096	0.32728

Распределение Пуассона

Рассмотрим дискретную случайную величину X, принимающую только целые неотрицательные значения $(0, 1, 2, \ldots, m, \ldots)$, последовательность которых не ограничена. Такая случайная величина называется распределенной **по закону Пуассона**, если вероятность того, что она примет значение m, выражается формулой:

$$p(X = m) = \frac{a^m}{m!} e^{-a}, \tag{4.3}$$

где a – некоторая положительная величина, называемая *параметром* закона Пуассона.

Покажем, что сумма всех вероятностей равна 1:

$$\sum_{m=0}^{\infty} p(X = m) = e^{-a} \sum_{m=0}^{\infty} \frac{a^m}{m!} = e^{-a} \cdot e^{a} = 1,$$

(использовано разложение в ряд Тейлора функции e^x).

Рассмотрим типичную задачу, приводящую к распределению Пуассона. Пусть на оси абсцисс случайным образом распределяются точки, причем их распределение удовлетворяет следующим условиям:

1) вероятность попадания некоторого количества точек на отрезок длины lзависит только от длины отрезка и не зависит от его расположения на оси (то есть точки распределены с одинаковой средней плотностью);
2) точки распределяются независимо друг от друга (вероятность попадания какого-либо числа точек на данный отрезок не зависит от количества точек, попавший на любой другой отрезок);
3) практическая невозможность совпадения двух или более точек.

Тогда случайная величина X – число точек, попадающих на отрезок длины l – распределена по закону Пуассона, где a – среднее число точек, приходящееся на отрезок длины $l.$

Замечание. В лекции 3 говорилось о том, что формула Пуассона выражает биномиальное распределение при большом числе опытов и малой вероятности события. Поэтому закон Пуассона часто называют *законом редких явлений*.

5. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ ИПЛОТНОСТЬ РАСПРЕДЕЛЕНИЯ НЕПРЕРЫВНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Определение и свойства функции распределения сохраняются и для непрерывной случайной величины, для которой функцию распределения можно считать одним из видов задания закона распределения. Но для непрерывной случайной величины вероятность каждого отдельного ее значения равна 0. Это следует из свойства 4 функции распределения: $p(X = a) = F(a) - F(a) = 0$. Поэтому для такой случайной величины имеет смысл говорить только о вероятности ее попадания в некоторый интервал.

Вторым способом задания закона распределения непрерывной случайной величины является так называемая плотность распределения (плотность вероятности, дифференциальная функция).

Определение 5.1. Функция $f(x)$, называемая **плотностью распределения** непрерывной случайной величины, определяется по формуле:

$$f(x) = F^{'}(x), \tag{5.1}$$

то есть является производной функции распределения.

Свойства плотности распределения.

1) $f(x) \geq 0$, так как функция распределения является неубывающей.

2) $F(x) = \int_{-\infty}^{x} f(t)dt,$ что следует из определения плотности распределения.

3) Вероятность попадания случайной величины в интервал (a, b) определяется формулой

$$p(a < X < b) = \int\limits_{a}^{b} f(x)\,dx.$$

Действительно,

$$p(a < X < b) = F(b) - F(a) = \int\limits_{-\infty}^{b} f(x)\,dx - \int\limits_{-\infty}^{a} f(x)\,dx = \int\limits_{a}^{b} f(x)\,dx.$$

4) $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ (условие нормировки). Его справедливость следует из того, что

$$\int_{-\infty}^{+\infty} f(x)\,dx = F(+\infty), \text{ а } \lim_{x \to +\infty} F\,(x) = 1.$$

5) $\lim\limits_{x \to \pm\infty} f\,(x) = 0,$ так как $F(x) \to const$ при $x \to \pm\infty$.

Таким образом, график плотности распределения представляет собой кривую, расположенную выше оси Ox причем эта ось является ее горизонтальной асимптотой при $x \to \pm\infty$ (последнее справедливо только для случайных величин, множеством возможных значений которых является все множество действительных чисел). Площадь криволинейной трапеции, ограниченной графиком этой функции, равна единице.

Замечание. Если все возможные значения непрерывной случайной величины сосредоточены на интервале $[a,b]$, то все интегралы вычисляются в этих пределах, а вне интервала$[a,b] f(x) \equiv 0$.

Пример 1. Плотность распределения непрерывной случайной величины задана формулой

$$f(x) = \frac{C}{1+x^2}, -\infty < x < +\infty.$$

Найти: а) значение константы C;б) вид функции распределения;в) $p(-1 < x < 1)$.

Решение.

а) значение константы Cнайдем из свойства 4:

$$\int_{-\infty}^{+\infty} \frac{C}{1+x^2} dx = C arctg x \Big|_{-\infty}^{+\infty} = C\pi = 1, \text{ откуда } C = \frac{1}{\pi}.$$

$$\text{б)} F(x) = \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{1+t^2} dt = \frac{1}{\pi} arctg\ t \Big|_{-\infty}^{x} = \frac{1}{\pi}\left(arctg\ x + \frac{\pi}{2}\right) = \frac{1}{\pi} arctg\ x + \frac{1}{2}.$$

$$\text{в) } p(-1 < x < 1) = \frac{1}{\pi}\int_{-1}^{1} \frac{1}{1+x^2} dx = \frac{1}{\pi} arctg\ x \Big|_{-1}^{1} = \frac{1}{\pi}\left(\frac{\pi}{4} + \frac{\pi}{4}\right) = 0{,}5.$$

Пример 2. Функция распределения непрерывной случайной величины имеет вид:

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

Найти плотность распределения.

Решение.

$$f(x) = \begin{cases} 0', x \le 2, \\ \left(\dfrac{x-2}{2}\right)', 2 < x \le 4, \\ 1', x > 4, \end{cases} = \begin{cases} 0, x \le 2, \\ 0{,}5, 2 < x \le 4, \\ 0, x > 4. \end{cases}$$

Равномерный закон распределения

Часто на практике мы имеем дело со случайными величинами, распределенными определенным типовым образом, то есть такими, закон распределения которых имеет некоторую стандартную форму. В прошлой лекции были рассмотрены примеры таких законов распределения для дискретных случайных величин (биномиальный и Пуассона). Для непрерывных случайных величин тоже существуют часто встречающиеся виды закона распределения, и в качестве первого из них рассмотрим равномерный закон.

Определение 5.2. Закон распределения непрерывной случайной величины называется **равномерным**, если на интервале, которому принадлежат все возможные значения случайной величины, плотность распределения сохраняет постоянное значение ($f(x) = const$при $a \leq x \leq b$, $f(x) = 0$ при $x < a,\ x > b.$

Найдем значение, которое принимает $f(x)$при$x \in [a, b]$. Из условия нормировки следует, что $\int_a^b f(x)\,dx = \int_a^b c\,dx = c(b-a) = 1$, откуда$f(x) = c = \frac{1}{b-a}$.

Вероятность попадания равномерно распределенной случайной величины на интервал $[\alpha, \beta]$($a \leq \alpha < \beta \leq b$)равна при этом

$$\int_{\alpha}^{\beta} \frac{1}{b-a}\,dx = \frac{\beta - \alpha}{b - a}.$$

Вид функции распределения для нормального закона:

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

Пример. Автобусы некоторого маршрута идут с интервалом 5 минут. Найти вероятность того, что пришедшему на остановку пассажиру придется ожидать автобуса не более 2 минут.

Решение. Время ожидания является случайной величиной, равномерно распределенной в интервале $[0, 5]$.

Тогда $f(x) = \frac{1}{5}, p(0 \leq x \leq 2) = \frac{2}{5} = 0{,}4$.

6. НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ. НОРМАЛЬНАЯ КРИВАЯ. ФУНКЦИЯ ЛАПЛАСА

Определение 6.1. Непрерывная случайная величина называется распределенной по **нормальному закону**, если ее плотность распределения имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad (6.1)$$

Замечание. Таким образом, нормальное распределение определяется двумя параметрами: a и σ.

График плотности нормального распределения называют **нормальной кривой (кривой Гаусса)**. Выясним, какой вид имеет эта кривая, для чего исследуем функцию (6.1).

1) Область определения этой функции: $(-\infty, +\infty)$.

2) $f(x) > 0$ при любом x (следовательно, весь график расположен выше оси Ox).

3) $\lim\limits_{|x|\to+\infty} f(x) = 0$, то есть ось Ox служит горизонтальной асимптотой графика при $x \to \pm\infty$.

4) $f'(x) = -\frac{x-a}{\sigma^3\sqrt{2\pi}} e^{\frac{(x-a)^2}{2\sigma^2}} = 0$ при $x = a$; $f'(x) > 0$, при $x > a, f'(x) < 0$ при $x < a$. Следовательно, $\left(a, \frac{1}{\sigma\sqrt{2\pi}}\right)$ – точка максимума.

5) $F(x-a) = f(a-x)$, то есть график симметричен относительно прямой $x = a$.

6) $f''(x) = \frac{1}{\sigma^3\sqrt{2\pi}} e^{\frac{(x-a)^2}{2\sigma^2}} \left(1 - \frac{(x-a)^2}{\sigma^2}\right) = 0$, при $x = a \pm \sigma$, то есть точки $\left(a \pm \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$ являются точками перегиба.

Примерный вид кривой Гаусса изображен на рисунке 6.1.

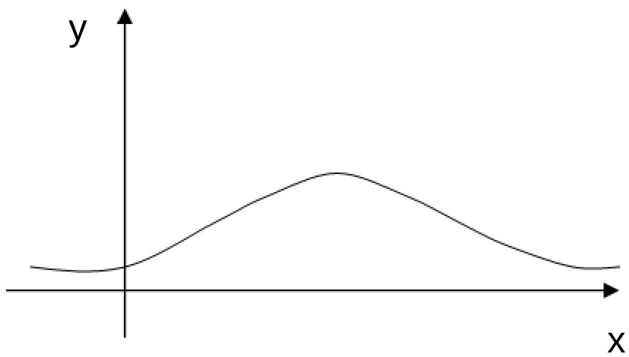

Рисунок 6.1

Найдем вид функции распределения для нормального закона:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt. \qquad (6.2)$$

Перед нами так называемый «неберущийся» интеграл, который невозможно выразить через элементарные функции. Поэтому для вычисления значений $F(x)$приходится пользоваться таблицами. Они составлены для случая, когда $a = 0$, а $\sigma = 1$.

Определение 6.2. Нормальное распределение с параметрами $a = 0, \sigma = 1$ называется **нормированным**, а его функция распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \qquad (6.3)$$

–**функцией Лапласа**.

Замечание. Функцию распределения для произвольных параметров можно выразить через функцию Лапласа, если сделать замену:$t = \frac{x-a}{\sigma}$, тогда

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}} dt.$$

Найдем вероятность попадания нормально распределенной случайной величины на заданный интервал:

$$p(\alpha < x < \beta) = F(\beta) - F(\alpha) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right) \qquad (6.4)$$

Пример. Случайная величина X имеет нормальное распределение с параметрами $a = 3, \sigma = 2$. Найти вероятность того, что она примет значение из интервала (4, 8).

Решение.

$$p(4 < x < 8) = F(8) - F(4) = \Phi\left(\frac{8-3}{2}\right) - \Phi\left(\frac{4-3}{2}\right) =$$
$$= \Phi(2{,}5) - \Phi(0{,}5) = 0{,}9938 - 0{,}6915 = 0{,}3023.$$

Правило «трех сигм»

Найдем вероятность того, что нормально распределенная случайная величина примет значение из интервала $(a - 3\sigma, a + 3\sigma)$:

$$p(\alpha - 3\sigma < x < a + 3\sigma) = \Phi(3) - \Phi(-3) =$$
$$= 0{,}9986 - 0{,}0014 = 0{,}9973.$$

Следовательно, вероятность того, что значение случайной величины окажется *вне* этого интервала, равна 0,0027, то есть составляет 0,27% и может считаться пренебрежимо малой. Таким образом, на практике можно считать, что *все* возможные значения нормально распределенной случайной величины лежат в интервале $(a-3\sigma, a+3\sigma)$.

Полученный результат позволяет сформулировать **правило «трех сигм»**: *если случайная величина распределена нормально, то модуль ее отклонения от* $x=a$ *не превосходит* 3σ.

Показательное распределение

Определение 6.3. **Показательным (экспоненциальным)** называют распределение вероятностей непрерывной случайной величины X, которое описывается плотностью

$$f(x)=\begin{cases}0, x<0,\\ \lambda e^{-\lambda x}, x\geq 0.\end{cases} \tag{6.5}$$

В отличие от нормального распределения, показательный закон определяется только одним параметром λ. В этом его преимущество, так как обычно параметры распределения заранее не известны и их приходится оценивать приближенно. Понятно, что оценить один параметр проще, чем несколько.

Найдем функцию распределения показательного закона:

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{0} 0\cdot dt+\lambda\int_{0}^{x} e^{-\lambda t}\,dt=1-e^{-\lambda x}.$$

Следовательно,

$$F(x)=\begin{cases}0, x<0\\ 1-e^{-\lambda x}, x\geq 0.\end{cases} \tag{6.6}$$

Теперь можно найти вероятность попадания показательно распределенной случайной величины в интервал (a,b):

$$p(a<x<b)=e^{-\lambda a}-e^{-\lambda b}. \tag{6.7}$$

Значения функции e^{-x}можно найти из таблиц.

Функция надежности

Пусть *элемент* (то есть некоторое устройство) начинает работать в момент времени $t_0=0$ и должен проработать в течение периода времени t. Обозначим за T непрерывную случайную величину – время безотказной работы элемента, тогда функция $F(t) = p(T > t)$ определяет вероятность отказа за время t. Следовательно, вероятность безотказной работы за это же время равна

$$R(t) = p(T > t) = 1 - F(t). \tag{6.8}$$

Эта функция называется **функцией надежности**.

Показательный закон надежности

Часто длительность безотказной работы элемента имеет показательное распределение, то есть

$$F(t) = 1 - e^{-\lambda t}.$$

Следовательно, функция надежности в этом случае имеет вид:

$$R(t) = 1 - \left(1 - e^{-\lambda t}\right) = e^{-\lambda t}.$$

Определение 6.4. **Показательным законом надежности** называют функцию надежности, определяемую равенством

$$R(t) = e^{-\lambda t}, \tag{6.9}$$

где λ – интенсивность отказов.

Пример. Пусть время безотказной работы элемента распределено по показательному закону с плотностью распределения $f(t) = 0{,}1e^{-0{,}1t}$. при $t \geq 0$. Найти вероятность того, что элемент проработает безотказно в течение 10 часов.

Решение. Так как $\lambda = 0{,}1$, $R(10) = e^{-0{,}1\cdot 10} = e^{-1} = 0{,}368$.

7. ОСНОВНЫЕ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ДИСКРЕТНЫХ И НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Закон распределения (функция распределения и ряд распределения или плотность вероятности) полностью описывают поведение случайной величины. Но в ряде задач достаточно знать некоторые числовые характеристики исследуемой величины (например, ее среднее значение и возможное отклонение от него), чтобы ответить на поставленный вопрос. Рассмотрим основные числовые характеристики дискретных случайных величин.

Математическое ожидание

Определение 7.1. **Математическим ожиданием** дискретной случайной величины называется сумма произведений ее возможных значений на соответствующие им вероятности:

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n. \qquad (7.1)$$

Если число возможных значений случайной величины бесконечно, то$M(X) = \sum_{i=1}^{\infty} x_i\, p_i$, если полученный ряд сходится абсолютно.

Замечание 1. Математическое ожидание называют иногда *взвешенным средним*, так как оно приближенно равно среднему арифметическому наблюдаемых значений случайной величины при большом числе опытов.

Замечание 2. Из определения математического ожидания следует, что его значение не меньше наименьшего возможного значения случайной величины и не больше наибольшего.

Замечание 3. Математическое ожидание дискретной случайной величины есть *неслучайная* (постоянная) величина. В дальнейшем увидим, что это же справедливо и для непрерывных случайных величин.

Пример 1. Найдем математическое ожидание случайной величины X – числа стандартных деталей среди трех, отобранных из партии в 10 деталей, среди которых 2 бракованных. Составим ряд распределения дляX. Из условия задачи следует, что X может принимать значения 1, 2, 3.

$$p(1) = \frac{C_8^1 \cdot C_2^2}{C_{10}^3} = \frac{1}{15}, p(2) = \frac{C_8^2 \cdot C_2^1}{C_{10}^3} = \frac{17}{15}, p(3) = \frac{C_8^3}{C_{10}^3} = \frac{7}{15}.$$

Тогда

$$M(X) = 1 \cdot \frac{1}{15} + 2 \cdot \frac{7}{15} + 3 \cdot \frac{7}{15} = 2{,}4.$$

Пример 2. Определим математическое ожидание случайной величины X – числа бросков монеты до первого появления герба. Эта величина может принимать бесконечное число значений (множество возможных значений есть множество натуральных чисел). Ряд ее распределения имеет вид:

X	1	2	…	n	…
p	0,5	$(0{,}5)^2$	…	$(0{,}5)^n$	…

Тогда

$$M(X)=\sum_{n=1}^{\infty}\frac{n}{2^n}=\frac{1}{2}+2\cdot\left(\frac{1}{2}\right)^2+3\cdot\left(\frac{1}{2}\right)^3+\cdots+n\cdot\left(\frac{1}{2}\right)^n+\cdots=$$

$$=\sum_{n=1}^{\infty}\frac{1}{2^n}+\frac{1}{2}\sum_{n=1}^{\infty}\frac{1}{2^n}+\cdots+\frac{1}{2^n}\sum_{n=1}^{\infty}\frac{1}{2^n}+\cdots=$$
$$=1\cdot\left(1+\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^n}+\cdots\right)=1\cdot 2=2,$$

(при вычислении дважды использовалась формула суммы бесконечно убывающей геометрической прогрессии:

$$S=\frac{b_1}{1-q}.$$

откуда

$$\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^n}+\cdots=1{,}1+\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^n}+\cdots=2).$$

Свойства математического ожидания

1) Математическое ожидание постоянной равно самой постоянной:

$$M(C)=C. \qquad (7.2)$$

Доказательство. Если рассматривать C как дискретную случайную величину, принимающую только одно значение C с вероятностью $p=1$, то $M(C)=C\cdot 1=C$.

2) Постоянный множитель можно выносит за знак математического ожидания:

$$M(CX)=C\cdot M(X) \qquad (7.3)$$

Доказательство. Если случайная величина X задана рядом распределения

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

то ряд распределения дляCX имеет вид:

Cx_i	Cx_1	Cx_2	…	Cx_n
p_i	p_1	p_2	…	p_n

Тогда

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

Определение 7.2. Две случайные величины называются **независимыми**, если закон распределения одной из них не зависит от того, какие значения приняла другая. В противном случае случайные величины **зависимы**.

Определение 7.3. Назовем **произведением независимых случайных величин X и Y** случайную величину XY, возможные значения которой равны произведениям всех возможных значений X на все возможные значения Y, а соответствующие им вероятности равны произведениям вероятностей сомножителей.

3) Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий:

$$M(XY) = M(X)M(Y) \tag{7.4}$$

Доказательство. Для упрощения вычислений ограничимся случаем, когда X и Yпринимают только по два возможных значения:

x_i	x_1	x_2
p_i	p_1	p_2

y_i	y_1	y_2
q_i	q_1	q_2

Тогда ряд распределения дляXYвыглядит так:

XY	x_1y_1	x_2y_1	x_1y_2	x_2y_2
p	p_1q_1	p_2q_1	p_1q_2	p_2q_2

Следовательно,

$$M(XY) = x_1y_1 \cdot p_1q_1 + x_2y_1 \cdot p_1q_2 + x_1y_2 \cdot p_1q_2 + x_2y_2 \cdot p_2q_2 =$$
$$= y_1q_1(x_1p_1 + x_2p_2) + y_2q_2(x_1p_1 + x_2p_2) =$$
$$= (y_1q_1 + y_2q_2) \cdot (x_1p_1 + x_2p_2) = M(X)M(Y).$$

Замечание 1. Аналогично можно доказать это свойство для большего количества возможных значений сомножителей.

Замечание 2. Свойство 3 справедливо для произведения любого числа независимых случайных величин, что доказываетсяметодом математической индукции.

Определение 7.4. Определим **сумму случайных величин X и Y** как случайную величину $X+Y$, возможные значения которой равны суммам каждого возможного значения X с каждым возможным значением Y; вероятности таких сумм равны произведениям вероятностей слагаемых (для зависимых случайных величин – произведениям вероятности одного слагаемого на условную вероятность второго).

4) Математическое ожидание суммы двух случайных величин (зависимых или независимых) равно сумме математических ожиданий слагаемых:

$$M(X+Y)=M(X)+M(Y). \tag{7.5}$$

Доказательство.

Вновь рассмотрим случайные величины, заданные рядами распределения, приведенными при доказательстве свойства 3. Тогда возможными значениями $X+Y$ являются $x_1+y_1, x_1+y_2, x_2+y_1, x_2+y_2$. Обозначим их вероятности соответственно как p_{11}, p_{12}, p_{21} и p_{22}. Найдем

$$M(X+Y)=(x_1+y_1)p_{11}+\ (x_1+y_2)p_{12}+\ (x_2+y_1)p_{21}+(x_2+y_2)p_{22}=$$
$$=x_1(p_{11}+p_{12})+x_2(p_{21}+p_{22})+y_1(p_{11}+p_{21})+y_2(p_{12}+p_{22}).$$

Докажем, что $p_{11}+\ p_{22}=\ p_1$. Действительно, событие, состоящее в том, что $X+Y$ примет значения x_1+y_1 или x_1+y_2 и вероятность которого равна $p_{11}+\ p_{22}$, совпадает с событием, заключающемся в том, что $X=x_1$ (его вероятность – p_1). Аналогично доказывается, что

$$p_{21}+\ p_{22}=\ p_2, p_{21}+\ p_{22}=\ q_1, p_{12}+\ p_{22}=\ q_2.$$

Значит,

$$M(X+Y)=x_1p_1+x_2p_2+y_1q_1+y_2p_2=M(X)+M(Y).$$

Замечание. Из свойства 4 следует, что сумма любого числа случайных величин равна сумме математических ожиданий слагаемых.

Пример. Найти математическое ожидание суммы числа очков, выпавших при броске пяти игральных костей.

Найдем математическое ожидание числа очков, выпавших при броске одной кости:

$$M(X_1)=(1\ +\ 2\ +\ 3\ +\ 4\ +\ 5\ +\ 6)\cdot\frac{1}{6}=\frac{7}{2}.$$

Тому же числу равно математическое ожидание числа очков, выпавших на любой кости. Следовательно, по свойству 4

$$M(X) = 5 \cdot \frac{1}{6} = \frac{5}{6}.$$

Дисперсия

Для того, чтобы иметь представление о поведении случайной величины, недостаточно знать только ее математическое ожидание. Рассмотрим две случайные величины: X и Y, заданные рядами распределения вида

X	49	50	51
p	0,1	0,8	0,1

Y	0	100
p	0,5	0,5

Найдем $M(X) = 49 \cdot 0{,}1 + 50 \cdot 0{,}8 + 51 \cdot 0{,}1 = 50$, $M(Y) = 0 \cdot 0{,}5 + 100 \cdot 0{,}5 = 50$. Как видно, математические ожидания обеих величин равны, но если для$X M(X)$хорошо описывает поведение случайной величины, являясь ее наиболее вероятным возможным значением (причем остальные значения ненамного отличаются от 50), то значения Yсущественно отстоят от $M(Y)$. Следовательно, наряду с математическим ожиданием желательно знать, насколько значения случайной величины отклоняются от него. Для характеристики этого показателяслужит дисперсия.

Определение 7.5. **Дисперсией (рассеянием)** случайной величины называется математическое ожидание квадрата ее отклонения от ее математического ожидания:

$$D(X) = M\big(X - M(X)\big)^2. \qquad (7.6)$$

*Пример.*Найдем дисперсию случайной величины X (числа стандартных деталей среди отобранных) в примере 1 данной лекции. Вычислим значения квадрата отклонения каждого возможного значения от математического ожидания:

$$(1 - 2{,}4)^2 = 1{,}96;\ (2 - 2{,}4)^2 = 0{,}16;\ (3 - 2{,}4)^2 = 0{,}36.$$

Следовательно,

$$D(X) = 1{,}96 \cdot \frac{1}{15} + 0{,}16 \cdot \frac{7}{15} + 0{,}36 \cdot \frac{7}{15} = \frac{28}{75} \approx 0{,}373.$$

Замечание 1. В определении дисперсии оценивается не само отклонение от среднего, а его квадрат. Это сделано для того, чтобы отклонения разных знаков не компенсировали друг друга.

Замечание 2. Из определения дисперсии следует, что эта величина принимает только неотрицательные значения.

Замечание 3. Существует более удобная для расчетов формула для вычисления дисперсии, справедливость которой доказывается в следующей теореме:

Теорема 7.1.

$$D(X) = M(X^2) - M^2(X). \tag{7.7}$$

Доказательство.

Используя то, что $M(X)$– постоянная величина, и свойства математического ожидания, преобразуем формулу (7.6) к виду:

$$D(X) = M\big(X - M(X)\big)^2 = M\big(X^2 - 2X \cdot M(X) + M^2(X)\big) =$$
$$= M(X^2) - 2M(X) \cdot M(X) + M^2(X) =$$
$$= M(X^2) - 2M^2(X) + M^2(X) = M(X^2) - M^2(X),$$

что и требовалось доказать.

Пример. Вычислим дисперсии случайных величин X и Y, рассмотренных в начале этого раздела.

$$M(X) = (49^2 \cdot 0{,}1 + 50^2 \cdot 0{,}8 + 51^2 \cdot 0{,}1) - 50^2 = 2500{,}2 - 2500 = 0{,}2.$$
$$M(Y) = (0^2 \cdot 0{,}5 + 100^2 \cdot 0{,}5) - 50^2 = 5000 - 2500 = 2500.$$

Итак, дисперсия второй случайной величины в несколько тысяч раз больше дисперсии первой. Таким образом, даже не зная законов распределения этих величин, по известным значениям дисперсии мы можем утверждать, что X мало отклоняется от своего математического ожидания, в то время как дляYэто отклонение весьма существенно.

Свойства дисперсии

1) Дисперсия постоянной величины C равна нулю:

$$D(C) = 0. \tag{7.8}$$

Доказательство.

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

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

$$D(CX) = C^2 D(X). \tag{7.9}$$

Доказательство.

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

3) Дисперсия суммы двух независимых случайных величин равна сумме их дисперсий:

$$D(X + Y) = D(X) + D(Y). \qquad (7.10)$$

Доказательство.

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

Следствие 1. Дисперсия суммы нескольких взаимно независимых случайных величин равна сумме их дисперсий.

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

4) Дисперсия разности двух независимых случайных величин равна сумме их дисперсий:

$$D(X - Y) = D(X) + D(Y). \qquad (7.11)$$

Доказательство.

$$D(X - Y) = D(X) + D(-Y) = D(X) + (-1)^2 D(Y) = D(X) + D(Y).$$

Дисперсия дает среднее значение квадрата отклонения случайной величины от среднего; для оценки самого отклонения служит величина, называемая средним квадратическим отклонением.

Определение 7.6. **Средним квадратическим отклонением**σ случайной величины X называется квадратный корень из дисперсии:

$$\sigma = \sqrt{D(X)}. \qquad (7.12)$$

Пример. В предыдущем примере средние квадратические отклонения X и Y равны соответственно

$$\sigma_x = \sqrt{0{,}2} \approx 0{,}447;\ \sigma_y = \sqrt{2500} = 50.$$

Числовые характеристики непрерывных случайных величин

Распространим определения числовых характеристик случайных величин на непрерывные случайные величины, для которых плотность распределения служит в некотором роде аналогом понятия вероятности.

Определение 7.7. **Математическим ожиданием** непрерывной случайной величины называется

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx. \qquad (7.13)$$

Замечание 1. Общее определение дисперсии сохраняется для непрерывной случайной величины таким же, как и для дискретной (опр. 7.5), а формула для ее вычисления имеет вид:

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - M^2(X). \qquad (7.14)$$

Среднее квадратическое отклонение вычисляется по формуле (7.12).

Замечание 2. Если все возможные значения непрерывной случайной величины не выходят за пределы интервала $[a, b]$, то интегралы в формулах (7.13) и (7.14) вычисляются в этих пределах.

Пример. Плотность распределения случайной величины X имеет вид:

$$f(x) = \begin{cases} 0, x < 2, \\ -\frac{3}{4}(x^2 - 6x + 8), 2 \le x \le 4, \\ 0, x > 4. \end{cases}$$

Найти $M(X)$, $D(X)$, σ.

Решение.

$$M(X) = -\frac{3}{4}\int_2^4 x(x^2 - 6x + 8)dx = -\frac{3}{4}\left(\frac{x^4}{4} - 2x^3 + 4x^2\right)\Big|_2^4 = 3;$$

$$D(X) = -\frac{3}{4}\int_2^4 x^2(x^2 - 6x + 8)dx - 9 = -\frac{3}{4}\left(\frac{x^5}{5} - \frac{3x^4}{2} + \frac{8x^3}{3}\right)\Big|_2^4 - 9 =$$
$$= 9{,}2 - 9 = 0{,}2;\ \sigma = \sqrt{0{,}2} \approx 0{,}447.$$

Числовые характеристики случайных величин, имеющих некоторыестандартные законы распределения

1. *Биномиальное распределение*

Для дискретной случайной величины X, представляющей собой число появлений события A в серии из n независимых испытаний (см. лекцию 6), $M(X)$можно найти, используя свойство 4 математического ожидания. Пусть X_1 – число появлений A в первом испытании, X_2 – во втором и т.д. При этом каждая из случайных величин X_iзадается рядом распределения вида

X_i	0	1
p_i	q	p

Следовательно, $M(X_i) = p$. Тогда

$$M(X) = \sum_{i=1}^{n} M(X_i) = \sum_{i=1}^{n} p = np.$$

Аналогичным образом вычислим дисперсию:

$$D(X_i) = 0^2 \cdot q + 1^2 \cdot p - p^2 = p - p^2 = p(1-p),$$

откуда по свойству 4 дисперсии

$$D(X) = \sum_{i=1}^{n} D(X_i) = np(1-p) = npq.$$

2. *Закон Пуассона*

Если $p(X = m) = \frac{a^m}{m!} e^{-a}$, то

$$M(X) = \sum_{m=1}^{\infty} m \frac{a^m}{m!} e^{-a} = ae^{-a} \sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} = ae^{-a}e^{a} = a.$$

(использовалось разложение в ряд Тейлора функции e^x).

Для определения дисперсии найдем вначале

$$M(X^2) = \sum_{m=1}^{\infty} m^2 \frac{a^m}{m!} e^{-a} = a \sum_{m=1}^{\infty} m \frac{a^{m-1}}{(m-1)!} e^{-a} =$$

$$= a \sum_{m=1}^{\infty} \big((m-1)+1\big) \frac{a^{m-1}}{(m-1)!} e^{-a} =$$

$$= a \left(\sum_{m=1}^{\infty} (m-1) \frac{a^{m-1}}{(m-1)!} e^{-a} + \sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} e^{-a} \right) = a(a+1).$$

Поэтому $D(X) = a^2 + a - a^2 = a.$

Замечание. Таким образом, обнаружено интересное свойство распределения Пуассона: математическое ожидание равно дисперсии (и равно единственному параметру a, определяющему распределение).

3. *Равномерное распределение*

Для равномерно распределенной на отрезке $[a, b]$непрерывной случайной величины $M(X) = \int_a^b x \frac{1}{b-a} dx = \frac{x^2}{2(b-a)} \Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$, то есть математическое ожидание равномерно распределенной случайной величины равно абсциссе середины отрезка$[a, b]$.

Дисперсия

$$D(X) = \int_a^b x^2 \frac{1}{b-a} dx - \frac{(a+b)^2}{4} = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} =$$

$$= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12}.$$

4. *Нормальное распределение*

Для вычисления математического ожидания нормально распределенной случайной величины воспользуемся тем, что *интеграл Пуассона*$\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi}$.

$$M(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{\frac{(x-a)^2}{2\sigma^2}} dx = \left(z = \frac{x-a}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma z + a) e^{-\frac{z^2}{2}} dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma z e^{-\frac{z^2}{2}} dz + \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = 0 + \frac{a}{\sqrt{2\pi}} \sqrt{2\pi} = a,$$

(первое слагаемое равно 0, так как подынтегральная функция нечетна, а пределы интегрирования симметричны относительно нуля).

$$D(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-a)^2 e^{\frac{(x-a)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \cdot z e^{\frac{z^2}{2}} dz =$$

$$= \left(u = z, dv = z e^{-\frac{z^2}{2}}\right) = \frac{\sigma^2}{\sqrt{2\pi}} \left(-z \cdot e^{-\frac{z^2}{2}} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz \right) =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left(-0 + \sqrt{2\pi}\right) = \sigma^2.$$

Следовательно, параметры нормального распределения (a и σ) равны соответственно математическому ожиданию и среднему квадратическому отклонению исследуемой случайной величины.

8. СЛУЧАЙНЫЕ ВЕКТОРЫ (СИСТЕМЫ НЕСКОЛЬКИХ СЛУЧАЙНЫХ ВЕЛИЧИН). ЗАКОН РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ДИСКРЕТНОЙ ДВУМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Наряду с одномерными случайными величинами, возможные значения которых определяются одним числом, теория вероятностей рассматривает и многомерные случайные величины. Каждое возможное значение такой величины представляет собой упорядоченный набор нескольких чисел. Геометрической иллюстрацией этого понятия служат точки -мерного пространства, каждая координата которых является случайной величиной (дискретной или непрерывной), или n-мерные векторы. Поэтому многомерные случайные величины называют еще случайными векторами.

Двумерные случайные величины

1. *Дискретные двумерные случайные величины*

Закон распределения дискретной двумерной случайной величины (X,Y)имеет вид таблицы с двойным входом, задающей перечень возможных значений каждой компоненты и вероятности $p(x_i,y_i)$,с которыми величина принимает значение (x_i,y_i):

Y	X					
	x_1	x_2	$\dots$	x_i	$\dots$	x_n
y_1	$p(x_1,y_1)$	$p(x_2,y_1)$	$\dots$	$p(x_i,y_1)$	$\dots$	$p(x_n,y_1)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_i	$p(x_1,y_i)$	$p(x_2,y_i)$	$\dots$	$p(x_i,y_i)$	$\dots$	$p(x_n,y_i)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_m	$p(x_1,y_m)$	$p(x_2,y_m)$	$\dots$	$p(x_i,y_m)$	$\dots$	$p(x_n,y_m)$

При этом сумма вероятностей, стоящих во всех клетках таблицы, равна 1.

Зная закон распределения двумерной случайной величины, можно найти законы распределения ее составляющих. Действительно, событие $X=x_1$ представляется собой сумму несовместных событий

$$(X=x_1,Y=y_1),(X=x_1,Y=y_2),\dots,(X=x_1,Y=y_m),$$

поэтому

$$p(X = x_1) = p(x_1,\ y_1) + p(x_1,\ y_2) + \cdots + p(x_1,\ y_m),$$

(в правой части находится сумма вероятностей, стоящих в столбце, соответствующем $X = x_1$). Так же можно найти вероятности остальных возможных значений X. Для определения вероятностей возможных значений Yнужно сложить вероятности, стоящие в строке таблицы, соответствующей $Y = y_i$.

Пример 1. Дан закон распределения двумерной случайной величины:

Y	X		
	-2	3	6
-0,8	0,1	0,3	0,1
-0,5	0,15	0,25	0,1

Найти законы распределения составляющих.

Решение. Складывая стоящие в таблице вероятности «по столбцам», получим ряд распределения дляX:

X	-2	3	6
p	0,25	0,55	0,2

Складывая те же вероятности «по строкам», найдем ряд распределения дляY:

Y	-0,8	-0,5
p	0,5	0,5

2. *Непрерывные двумерные случайные величины*

Определение 8.1. **Функцией распределения** $\boldsymbol{F(x,y)}$двумерной случайной величины (X,Y) называется вероятность того, что $X < x$, а$Y < y$:

$$F(x,y) = p(X < x, Y < y\ ) \qquad (8.1)$$

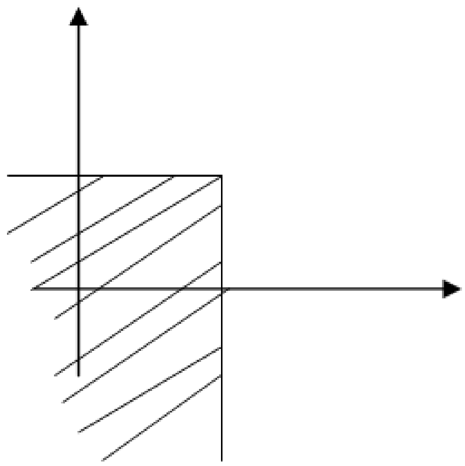

Рисунок 8.1

Это означает, что точка (X,Y) попадет в область, заштрихованную на рисунке 8.1, если вершина прямого угла располагается в точке (x,y).

Замечание. Определение функции распределения справедливо как для непрерывной, так и для дискретной двумерной случайной величины.

Свойства функции распределения

1) $0 \le F(x,y) \le 1$(так как $F(x,y)$ является вероятностью).

2) $F(x,y)$ есть неубывающая функция по каждому аргументу:

$$F(x_2,y) \ge F(x_1,y), \text{ если} x_2 > x_1;$$
$$F(x,y_2) \ge F(x,y_1), \text{ если} y_2 > y_1.$$

Доказательство.

$$F(x_2,y) = p(X < x_2, Y < y) = p(X < x_1, Y < y) +$$
$$+ p(x_1 \le X < x_2, Y < y) \ge p(X < x_1, Y < y) = F(x_1,y).$$

Аналогично доказывается и второе утверждение.

3) Имеют место предельные соотношения:

а) $F(-\infty, y) = 0$;b) $F(x,-\infty) = 0$;c) $F(-\infty,-\infty) = 0$;d) $F(\infty,\infty) = 1$.

Доказательство. События а), b) и с) невозможны (так как невозможно событие $X < -\infty$ или $Y < -\infty$), а событие d) достоверно, откуда следует справедливость приведенных равенств.$x = \infty$.

4) При$y = \infty$функция распределения двумерной случайной величины становится функцией распределения составляющей X:

$$F(x,\infty) = F_1(x).$$

При$x = \infty$ функция распределения двумерной случайной величины становится функцией распределения составляющей Y:

$$F(\infty,y) = F_2(y).$$

Доказательство. Так как событие $Y < \infty$ достоверно, то

$$F(x,\infty) = p(X < x) = F_1(x).$$

Аналогично доказывается второе утверждение.

Определение 8.2. **Плотностью совместного распределения вероятностей (двумерной плотностью вероятности)** непрерывной двумерной случайной величины называется смешанная частная производная 2-го порядка от функции распределения:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}. \qquad (8.2)$$

Замечание. Двумерная плотность вероятности представляет собой предел отношения вероятности попадания случайной точки в прямоугольник со сторонами Δx и Δy к площади этого прямоугольника при $\Delta x \to 0, \Delta y \to 0$.

Свойства двумерной плотности вероятности

1) $f(x, y) \geq 0$ (см. предыдущее замечание: вероятность попадания точки в прямоугольник неотрицательна, площадь этого прямоугольника положительна, следовательно, предел их отношения неотрицателен).

2) $F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y) dx dy$(следует из определения двумерной плотности вероятности).

3) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$(поскольку это вероятность того, что точка попадет на плоскость Oxy, то есть достоверного события).

Вероятность попадания случайной точки в произвольную область.

Пусть в плоскости Oxy задана произвольная область D. Найдем вероятность того, что точка, координаты которой представляют собой систему двух случайных величин (двумерную случайную величину) с плотностью распределения $f(x, y)$, попадет в область D. Разобьем эту область прямыми, параллельными осям координат, на прямоугольники со сторонами Δx и Δy. Вероятность попадания в каждый такой прямоугольник равна$f(\xi_i, \eta_i)\Delta x \Delta y$, где (ξ_i, η_i)– координаты точки, принадлежащей прямоугольнику. Тогда вероятность попадания точки в область Dесть предел интегральной суммы $\sum_{i=1}^{n} f(\xi_i, \eta_i)\Delta x \Delta y$, то есть

$$p\big((X, Y) \subset D\big) = \iint_D f(x, y) dx dy. \qquad (8.3)$$

Отыскание плотностей вероятности составляющихдвумерной случайной величины.

Выше было сказано, как найти функцию распределения каждой составляющей, зная двумерную функцию распределения. Тогда по определению плотности распределения

$$f_1(x) = \frac{dF_1(x)}{dx} = \frac{dF(x, \infty)}{dx} = \frac{d\left(\int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x, y)\right)}{dx} = \int_{-\infty}^{\infty} f(x, y) dy. \qquad (8.4)$$

Аналогично находится

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx. \qquad (8.4')$$

Условные законы распределения составляющих дискретной двумерной случайной величины

Рассмотрим дискретную двумерную случайную величину и найдем закон распределения составляющей X при условии, что Y примет определенное значение (например, $Y = y_1$). Для этого воспользуемся формулой Байеса, считая гипотезами события $X = x_1$, $X = x_2, \ldots$, $X = x_n$, а событием A – событие $Y = y_1$. При такой постановке задачи нам требуется найти условные вероятности гипотез при условии, что A произошло. Следовательно,

$$p(x_i/y_1) = \frac{p(x_i, y_1)}{p(y_1)}.$$

Таким же образом можно найти вероятности возможных значений X при условии, что Yпринимает любое другое свое возможное значение:

$$p\left(x_i/y_{\mathrm{j}}\right) = \frac{p\left(x_i, y_{\mathrm{j}}\right)}{p\left(y_{\mathrm{j}}\right)}. \qquad (8.5)$$

Аналогично находят условные законы распределения составляющей Y:

$$p\left(y_{\mathrm{j}}/x_i\right) = \frac{p\left(x_i, y_{\mathrm{j}}\right)}{p(x_i)}. \qquad (8.5`)$$

Пример. Найдем закон распределения X при условии $Y = -0{,}8$ и закон распределения Y при условии $X = 3$для случайной величины, рассмотренной в примере 1.

$$p(x_1/y_1) = \frac{0{,}1}{0{,}5} = \frac{1}{5} = 0{,}2;\ p(x_2/y_1) = \frac{0{,}3}{0{,}5} = \frac{3}{5} = 0{,}6;$$
$$p(x_3/y_1) = \frac{0{,}1}{0{,}5} = \frac{1}{5} = 0{,}2; p(y_1/x_2) = \frac{0{,}3}{0{,}55} = \frac{6}{11};$$
$$p(y_2/x_2) = \frac{0{,}25}{0{,}55} = \frac{5}{11}.$$

Условные законы распределения составляющих дискретной двумерной случайной величины

Определение 8.3. **Условной плотностью $\boldsymbol{\varphi(x/y)}$**распределения составляющих X при данном значении $Y = y$ называется

$$\varphi(x/y) = \frac{f(x,y)}{f_2(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx}. \qquad (8.6)$$

Аналогично определяется условная плотность вероятности Yпри$X = x$:

$$\psi(y/x) = \frac{f(x,y)}{f_1(x)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dy}. \qquad (8.6`)$$

Равномерное распределение на плоскости

Система двух случайных величин называется **равномерно распределенной на плоскости**, если ее плотность вероятности $f(x,y) = const$ внутри некоторой области и равна 0 вне ее. Пусть данная область – прямоугольник вида $a \leq x \leq b, c \leq y \leq d$.Тогда из свойств$f(x,y)$ следует, что

$$f(x,y) = \begin{cases} \dfrac{1}{S_{\text{пр}}} = \dfrac{1}{(b-a)(d-c)}, \text{внутри прямоугольрика,} \\ 0, \text{вне его.} \end{cases}$$

Найдем двумерную функцию распределения:

$$F(x,y) = \frac{1}{(b-a)(d-c)} \int_{c}^{y} \int_{a}^{x} dxdy = \frac{(x-a)(y-c)}{(b-a)(d-c)},$$

при $a < x < b, c < y < d, F(x,y) = 0$ при $x \leq a$ или$y \leq c, F(x,y) = 1$ при $x \geq b, y \geq d.$

Функции распределения составляющих, вычисленные по формулам, приведенным в свойстве 4 функции распределения, имеют вид:

$$F_1(x) = \frac{(x-a)}{(b-a)}, F_2(y) = \frac{(y-c)}{(d-c)}.$$

9. НЕКОТОРЫЕ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ОДНОМЕРНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Определение 9.1. **Начальным моментом порядка k случайной величины X называется математическое ожидание величины X^k:**

$$\mathrm{v}_k = M(X^k). \tag{9.1}$$

В частности, $\mathrm{v}_1 = M(X), \mathrm{v}_2 = M(X^2)$. Следовательно, дисперсия

$$D(X) = \mathrm{v}_2 - \mathrm{v}_1{}^2.$$

*Определение 9.2.***Центральным моментом порядка k**случайной величины X называется математическое ожидание величины $\big(X - M(X)\big)^k$:

$$\mu_k = M\big(X - M(X)\big)^k. \tag{9.2}$$

В частности, $\mu_1 = M(X - M(X)) = 0, \mu_2 = M\left((X - M(X))^2\right) = D(X)$.

Можно получить соотношения, связывающие начальные и центральные моменты:

$$\mu_2 = \mathrm{v}_2 - \mathrm{v}_1{}^2, \mu_3 = \mathrm{v}_3 - 3\mathrm{v}_2\mathrm{v}_1 + 2\mathrm{v}_1{}^2, \mu_4 = \mathrm{v}_4 - 4\mathrm{v}_3\mathrm{v}_1 + 6\mathrm{v}_2\mathrm{v}_1{}^2 - 3\mathrm{v}_1{}^4.$$

Мода и медиана

Такая характеристика случайной величины, как математическое ожидание, называется иногда *характеристикой положения*, так как она дает представление о положении случайной величины на числовой оси. Другими характеристиками положения являются мода и медиана.

Определение 9.3. **Модой** $\boldsymbol{M}$ дискретной случайной величины называется ее наиболее вероятное значение, **модой** $\boldsymbol{M}$ непрерывной случайной величины – значение, в котором плотность вероятности максимальна.

Пример 1.

Если ряд распределения дискретной случайной величины X имеет вид:

X	1	2	3	4
p	0,1	0,7	0,15	0,05

то $M = 2$.

Пример 2.

Для непрерывной случайной величины, заданной плотностью распределения $f(x) = \frac{1}{\pi(1+x^2)}$, модой является абсцисса точки максимума: $M = 0$.

Замечание 1. Если кривая распределения имеет больше одного максимума, распределение называется **полимодальным**, если эта кривая не имеет максимума, но имеет минимум – **анти-модальным**.

Замечание 2. В общем случае мода и математическое ожидание не совпадают. Но, если распределение является симметричным и модальным (то есть кривая распределения симметрична относительно прямой $x = M$) и имеет математическое ожидание, оно совпадает с модой.

Определение 9.4. **Медианой** M_e непрерывной случайной величины называют такое ее значение, для которого

$$p(X < M_e) = p(X > M_e) \tag{9.3}$$

Графически прямая $x = M_e$делит площадь фигуры, ограниченной кривой распределения, на две равные части.

Замечание. Для симметричного модального распределения медиана совпадает с математическим ожиданием и модой.

Определение 9.5. Для случайной величины X с функцией распределения $F(X)$**квантилью порядка** $\boldsymbol{p}(0 < p < 1)$называется число K_p такое, что $F(K_p) \leq p, F(K_p + 0) \geq p$. В частности, если $F(X)$ строго монотонна, K_p: $F(K_p) = p$.

Асимметрия и эксцесс

Если распределение не является симметричным, можно оценить асимметрию кривой распределения с помощью центрального момента 3-го порядка. Действительно, для симметричного распределения все нечетные центральные моменты равны 0 (как интегралы от нечетных функцийв симметричных пределах), поэтому выбран нечетный момент наименьшего порядка, не тождественно равный 0. Чтобы получить безразмерную характеристику, его делят наσ^3 (так как μ_3 имеет размерность куба случайной величины).

Определение 9.6. **Коэффициентом асимметрии** случайной величины называется

$$S_k = \frac{\mu_3}{\sigma^3}\ . \tag{9.4}$$

В частности, для кривой, изображенной на рисунке 9.1,$S_k > 0$, а на рисунке 9.2 –$S_k < 0$.

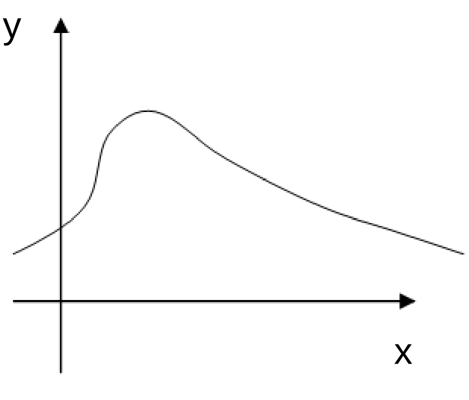

Рисунок 9.1

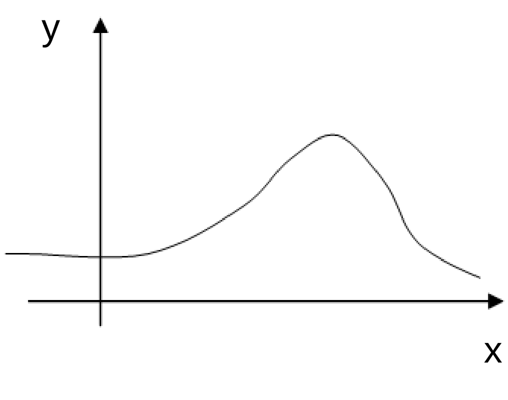

Рисунок 9.2

Для оценки поведения кривой распределения вблизи точки максимума (для определения того, насколько «крутой» будет его вершина) применяется центральный момент 4-го порядка.

Определение 9.7. **Эксцессом** случайной величины называется величина

$$E_x = \frac{\mu_4}{\sigma^4} - 3. \qquad (9.5)$$

Замечание. Можно показать, что для нормального распределения $\frac{\mu_4}{\sigma^4} = 3$, и, соответственно, $E_x = 0$. Для кривых с более острой вершиной $E_x > 0$, в случае более плоской вершины $E_x < 0$.

Числовыехарактеристики двумерных случайных величин

Такие характеристики, как начальные и центральные моменты, можно ввести и для системы двух случайных величин.

Определение 9.8. **Начальным моментом порядка $\boldsymbol{k, s}$** двумерной случайной величины (X, Y) называется математическое ожидание произведения X^k на Y^s:

$$\alpha_{k,s} = M(X^k, Y^s). \qquad (9.6)$$

Для дискретных случайных величин $\alpha_{k,s} = \sum_i \sum_{\mathrm{j}} x_i^k y_{\mathrm{j}}^s\, p_{i\mathrm{j}}$ для непрерывных случайных величин $\alpha_{k,s} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k, y^s f(x, y) dxdy$.

Определение 9.9. **Центральным моментом порядка $\boldsymbol{k, s}$** двумерной случайной величины (X, Y) называется математическое ожидание произведения $\big(X - M(X)\big)^k$ на $\big(Y - M(Y)\big)^s$:

$$\mu_{k,s} = M\big(X - M(X)\big)^k \big(Y - M(Y)\big)^s \qquad (9.7)$$

Для дискретных случайных величин

$$\mu_{k,s} = \sum_i \sum_{\mathrm{j}} \big(x_i - M(X)\big)^k \big(y_i - M(Y)\big)^s p_{\mathrm{ij}},$$

для непрерывных случайных величин

$$\mu_{k,s} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \big(x - M(X)\big)^k \big(y - M(Y)\big)^s f(x, y) dxdy.$$

При этом $M(X) = \alpha_{1,0}, M(Y) = \alpha_{0,1}, D(X) = \mu_{2,0}, D(Y) = \mu_{0,2}$.

Корреляционный момент и коэффициент корреляции

Определение 9.10. **Корреляционным моментом** системы двух случайных величин называется второй смешанный центральный момент:

$$K_{xy} = \mu_{1,1} = M\Big(\big(X - M(X)\big)\big(Y - M(Y)\big)\Big) \qquad (9.8)$$

Для дискретных случайных величин

$$K_{xy} = \sum_{i}\sum_{\mathrm{j}}\big(x_i - M(X)\big)\big(y_i - M(Y)\big)\, p_{i\mathrm{j}}$$

для непрерывных случайных величин

$$K_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\big(x - M(X)\big)\big(y - M(Y)\big)\, f(x,y)dxdy.$$

Безразмерной характеристикой коррелированности двух случайных величин является **коэффициент корреляции**

$$r_{xy} = \frac{K_{xy}}{\sigma_x\sigma_y}. \qquad (9.9)$$

Корреляционный момент описывает связь между составляющими двумерной случайной величины. Действительно, убедимся, что для независимых X и $Y K_{xy} = 0$. В этом случае $f(x,y) = f_1(x)f_2(y)$, тогда

$$K_{xy} = \int_{-\infty}^{\infty}\big(x - M(X)\big)\, f_1(x)dx \int_{-\infty}^{\infty}\big(y - M(Y)\big)\, f_2(y)dy = \mu_1(x)\mu_2(y) = 0.$$

Итак, две независимые случайные величины являются и некоррелированными. Однако понятия коррелированности и зависимости не эквивалентны, а именно, величины могут быть зависимыми, но при этом некоррелированными. Дело в том, что коэффициент корреляции характеризует не всякую зависимость, а только линейную. В частности, если$Y = aX + b$, то $r_{xy} = \pm 1$.Найдем возможные значения коэффициента корреляции.

Теорема 9.1.$|r_{xy}| \le 1$.

Доказательство. Докажем сначала, что $|K_{xy}| \le \sigma_x\sigma_y$ Действительно, если рассмотреть случайную величину $Z_1 = \sigma_y X - \sigma_x Y$ и найти ее дисперсию, то получим: $D(Z_1) = 2\sigma_x^2\sigma_y^2 - 2\sigma_x\sigma_y K_{xy}$. Так как дисперсия всегда неотрицательна, то $2\sigma_x^2\sigma_y^2 - 2\sigma_x\sigma_y K_{xy} \geq 0$ откуда $|K_{xy}| \leq \sigma_x\sigma_y$. Отсюда $\left|\frac{K_{xy}}{\sigma_x\sigma_y}\right| = |r_{xy}| \leq 0$, что и требовалось доказать.

10. ФУНКЦИИ ОТ СЛУЧАЙНЫХ ВЕЛИЧИН. ФУНКЦИЯ ОДНОГО СЛУЧАЙНОГО АРГУМЕНТА, ЕЕ РАСПРЕДЕЛЕНИЕ И МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

В предыдущих лекциях рассматривались некоторые законы распределения случайных величин. При решении задач часто удобно бывает представить исследуемую случайную величину как функцию других случайных величин с известными законами распределения, что помогает установить и закон распределения заданной случайной величины.

Определение 10.1. Если каждому возможному значению случайной величины X соответствует одно возможное значение случайной величины Y, тоYназывают **функцией случайного аргумента X**:$Y = \varphi(X)$.Выясним, как найти закон распределения функции по известному закону распределения аргумента.

1) Пусть аргумент X – дискретная случайная величина, причем различным значениям X соответствуют различные значения Y. Тогда вероятности соответствующих значений X и Yравны.

Пример 1. Ряд распределения дляX имеет вид:

X5678

p0,10,20,30,4

Найдем закон распределения функции Y = 2X^2- 3:

Y476995125

p0,10,20,30,4

(при вычислении значений Y в формулу, задающую функцию, подставляются возможные значения X).

2) Если разным значениям X могут соответствовать одинаковые значения Y, то вероятности значений аргумента, при которых функция принимает одно и то же значение, складываются.

Пример 2. Ряд распределения дляX имеет вид:

X0123

p0,10,20,30,4

Найдем закон распределения функции $Y = X^2 - 2X$:

Y-103

p0,20,40,4

(так как $Y = 0$ при $X = 0$ и $X = 2$, то $p(Y = 0) = p(X = 0) + p(X = 2) = 0{,}1 + 0{,}3 = 0{,}4$).

4) Если X – непрерывная случайная величина, $Y=\varphi(X), \varphi(x)$– монотонная и дифференцируемая функция, а$\psi(y)$– функция, обратная к$\varphi(x)$, то плотность распределения $g(y)$случайной функции Yравна:

$$g(y)=f\big(\psi(y)\big)\big|\psi^{'}(y)\big|. \qquad (10.1)$$

Пример 3. $f(x)=\frac{1}{\pi(1+x^2)}, Y=x^3$.

Тогда $\psi(y)=\sqrt[3]{y},\ g(y)=\frac{1}{\pi\left(1+y^{\frac{2}{3}}\right)}\cdot\left(\frac{1}{3}y^{-\frac{2}{3}}\right)=\frac{1}{3\pi y^{\frac{2}{3}}\left(1+y^{\frac{2}{3}}\right)}$.

Математическое ожидание функции одного случайного аргумента

Пусть $Y=\varphi(X)$– функция случайного аргумента X, и требуется найти ее математическое ожидание, зная закон распределения X.

1) Если X – дискретная случайная величина, то

$$M(Y)=M\big(\varphi(x)\big)=\sum_{i=1}^{n}\varphi(x_i)p_i. \qquad (10.2)$$

Пример 3. Найдем$M(Y)$для примера 1:

$$M(Y)=47\cdot 0{,}1\ +\ 69\cdot 0{,}2\ +\ 95\cdot 0{,}3\ +\ 125\cdot 0{,}4\ =\ 97.$$

2) Если X – непрерывная случайная величина, то $M(Y)$можно искать по-разному. Если известна плотность распределения $g(y)$, то

$$M(Y)=\int_{-\infty}^{\infty}yg(y)\,dy. \qquad (10.3)$$

Если же $g(y)$найти сложно, то можно использовать известную плотность распределения $f(x)$:

$$M(Y)=\int_{-\infty}^{\infty}\varphi(x)f(x)\,dx. \qquad (10.4)$$

В частности, если все значения X принадлежат промежутку (a,b), то

$$M(Y)=\int_{a}^{b}\varphi(x)f(x)\,dx. \qquad (10.4`)$$

Функция двух случайных величин. Распределение суммынезависимых слагаемых

Определение 10.2. Если каждой паре возможных значений случайных величин X и Y соответствует одно возможное значение случайной величины Z, то Zназывают **функцией двух случайных аргументов**XиY:$Z = \varphi(X, Y)$.

Рассмотрим в качестве такой функции сумму $X + Y$. В некоторых случаях можно найти ее закон распределения, зная законы распределения слагаемых.

1) ЕслиXиY– дискретные *независимые* случайные величины, то для определения закона распределения$Z = X + Y$нужно найти все возможные значения Zи соответствующие им вероятности.

Пример 4. Рассмотрим дискретные случайные величины XиY, законы распределения которых имеют вид:

X-213Y012

p0,30,40,3p0,20,50,3

Найдем возможные значенияZ:

-2 + 0 = -2 (p= 0,3·0,2 = 0,06), -2 + 1 = -1 (p = 0,3·0,5 = 0,15),

-2 + 2 = 0 (p= 0,3·0,3 = 0,09),1 + 0 = 1 (p = 0,4·0,2 = 0,08),

1 + 1 = 2 (p = 0,4·0,5 = 0,2), 1 + 2 = 3(p = 0,4·0,3 = 0,12),

3 + 0 = 3 (p = 0,3·0,2 = 0,06), 3 + 1 = 4 (p = 0,3·0,5 = 0,15),

3 + 2 = 5 (p= 0,3·0,3 = 0,09).

Сложив вероятности повторившегося дважды значения Z = 3, составим ряд распределения дляZ:

Z	-2	-1	0	1	2	3	4	5
p	0,06	0,15	0,09	0,08	0,2	0,18	0,15	0,09

2) ЕслиXиY– непрерывные *независимые* случайные величины, то, если плотность вероятности хотя бы одного из аргументов задана на $(-\infty, \infty)$одной формулой, то плотность суммы $g(z)$ можно найти по формулам

$$g(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z - x)\, dx = \int_{-\infty}^{\infty} f_1(z - y) f_2(y)\, dy, \qquad (10.5)$$

где $f_1(x), f_2(y)$– плотности распределения слагаемых. Если возможные значения аргументов неотрицательны, то

$$g(z) = \int_0^z f_1(x) f_2(z - x)\, dx = \int_0^z f_1(z - y) f_2(y)\, dy. \qquad (10.6)$$

Замечание. Плотность распределения суммы двух независимых случайных величин называют **композицией**.

Устойчивость нормального распределения

Определение 10.3. Закон распределения вероятностей называется **устойчивым**, если композиция таких законов есть тот же закон (возможно, отличающийся другими значениями параметров).

В частности, свойством устойчивости обладает нормальный закон распределения: композиция нормальных законов тоже имеет нормальное распределение, причем ее математическое ожидание и дисперсия равны суммам соответствующих характеристик слагаемых.

11. НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ НА ПЛОСКОСТИ. ЛИНЕЙНАЯ РЕГРЕССИЯ. ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ

Определение 11.1. **Нормальным законом распределения на плоскости** называют распределение вероятностей двумерной случайной величины (X, Y), если

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r_{xy}^2}} e^{\frac{1}{2\left(1-r_{xy}^2\right)}\left(\frac{(x-a_1)^2}{\sigma_x^2}+\frac{(y-a_2)^2}{\sigma_y^2}-2r_{xy}\frac{x-a_1}{\sigma_x}\cdot\frac{y-a_2}{\sigma_y}\right)}. \quad (11.1)$$

Таким образом, нормальный закон на плоскости определяется 5 параметрами: $a_1, a_2, \sigma_x, \sigma_y, r_{xy}$,где$a_1, a_2$– математические ожидания, σ_x, σ_y– средние квадратические отклонения, r_{xy}– коэффициент корреляции X и Y.Предположим, что $r_{xy} = 0$, то есть Xи Y некоррелированы. Тогда из (11.1) получим:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-0,5\left(\frac{(x-a_1)^2}{\sigma_x^2}+\frac{(y-a_2)^2}{\sigma_y^2}\right)} =$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}} e^{\frac{(x-a_1)^2}{2\sigma_x^2}} \cdot \frac{1}{\sigma_y\sqrt{2\pi}} e^{\frac{(y-a_1)^2}{2\sigma_y^2}} = f_1(x)f_2(y).$$

Следовательно, из некоррелированности составляющих нормально распределенной двумерной случайной величины следует их независимость, то есть для них понятия независимости и некоррелированности равносильны.

Линейная регрессия

Пусть составляющие X и Y двумерной случайной величины (X, Y)зависимы. Будем считать, что одну из них можно приближенно представить как линейную функцию другой, например

$$Y \approx g(X) = \alpha + \beta X, \quad (11.2)$$

и определим параметры αиβ с помощью метода наименьших квадратов.

Определение 11.2. Функция $g(X) = \alpha + \beta X$называется **наилучшим приближением** Yв смысле метода наименьших квадратов, если математическое ожидание $M\big(Y-g(X)\big)^2$принимает наименьшее возможное значение; функцию $g(X)$называют **среднеквадратической регрессией** YнаX.

Теорема 11.1. Линейная средняя квадратическая регрессия Yна X имеет вид:

$$g(X)=m_y+r\frac{\sigma_y}{\sigma_x}(X-m_x) \quad (11.3)$$

где

$$m_x=M(X), m_y=M(Y), \sigma_x=\sqrt{D(X)}, \sigma_y=\sqrt{D(Y)}, r=\frac{K_{xy}}{\sigma_x\sigma_y}-$$

–коэффициент корреляции X и Y.

Доказательство. Рассмотрим функцию

$$F(\alpha,\beta)=M(Y-\alpha-\beta X)^2 \quad (11.4)$$

и преобразуем ее, учитывая соотношения$M(X-m_x)=M\big(Y-m_y\big)=0$, $M\Big((X-m_x)\big(Y-m_y\big)\Big)=K_{xy}=r\sigma_x\sigma_y$:

$$F(\alpha,\beta)=\sigma_y^2+\beta^2\sigma_x^2-2r\sigma_x\sigma_y\beta+\big(m_y-\alpha-\beta m_x\big)^2.$$

Найдем стационарные точки полученной функции, решив систему

$$\begin{cases}\dfrac{\partial F}{\partial\alpha}=2\big(m_y-\alpha-\beta m_x\big)=0,\\ \dfrac{\partial F}{\partial\beta}=2\beta\sigma_x^2-2r\sigma_x\sigma_y=0.\end{cases}$$

Решением системы будет$\beta=r\frac{\sigma_y}{\sigma_x}, \alpha=m_y-r\frac{\sigma_y}{\sigma_x}m_x$.

Можно проверить, что при этих значениях функция $F(\alpha,\beta)$ имеет минимум, что доказывает утверждение теоремы.

Определение 11.3. Коэффициент $\beta=r\frac{\sigma_y}{\sigma_x}$называется **коэффициентом регрессии** Y**на** X, а прямая

$$y-m_y=r\frac{\sigma_y}{\sigma_x}(x-m_x)- \quad (11.5)$$

–**прямой среднеквадратической регрессии** Y**на**X.

Подставив координаты стационарной точки в равенство (11.4), можно найти минимальное значение функции $F(\alpha,\beta)$, равное $\sigma_y^2(1-r^2)$.

Эта величина называется **остаточной дисперсией** Y относительно X и характеризует величину ошибки, допускаемой при замене Y на $g(X) = \alpha + \beta X$. При $r = \pm 1$ остаточная дисперсия равна 0, то есть равенство (11.2) является не приближенным, а точным. Следовательно, при $r = \pm 1$ Y и X связаны *линейной* функциональной зависимостью. Аналогично можно получить прямую среднеквадратической регрессии X на Y:

$$x - m_x = r\frac{\sigma_x}{\sigma_y}\left(y - m_y\right) \qquad (11.6)$$

и остаточную дисперсию X относительно Y. При $r = \pm 1$ обе прямые регрессии совпадают. Решив систему из уравнений (11.5) и (11.6), можно найти точку пересечения прямых регрессии – точку с координатами $\left(m_x, m_y\right)$, называемую **центром совместного распределения величин X и Y.**

Линейная корреляция

Для двумерной случайной величины (X, Y) можно ввести так называемое **условное математическое ожидание Y** при $X = x$. Для дискретной случайной величины оно определяется как

$$M(Y|X = x) = \sum_{j=1}^{m} y_j p\left(y_j/x\right) \qquad (11.7)$$

для непрерывной случайной величины –

$$M(Y|X = x) = \int_{-\infty}^{\infty} y\psi(y/x)\,dy. \qquad (11.8)$$

Определение 11.4. **Функцией регрессии Y на X** называется условное математическое ожидание

$$M(Y/x) = f(x).$$

Аналогично определяется условное математическое ожидание X и функция регрессии X на Y.

Определение 11.5. Если обе функции регрессии X на Y и Y на X линейны, то говорят, что X и Y связаны **линейной корреляционной зависимостью**.

При этом графики линейных функций регрессии являются прямыми линиями, причем можно доказать, что эти линии совпадают с прямыми среднеквадратической регрессии.

Теорема 11.2. Если двумерная случайная величина (X, Y) распределена нормально, то X и Y связаны линейной корреляционной зависимостью.

Доказательство. Найдем условный закон распределения Y при $X = x$ $\left(\psi(y/x) = \frac{f(x,y)}{f_1(x)}\right)$используя формулу двумерной плотности вероятности нормального распределения (11.1) и формулу плотности вероятности X:

$$f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-a_1)^2}{2\sigma_x^2}}. \tag{11.9}$$

Сделаем замену $u = \frac{x-a_1}{\sigma_x}, v = \frac{y-a_2}{\sigma_y}$.

Тогда

$$\psi(y/x) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(v-ru)^2}{2(1-r^2)}} = \frac{1}{\sigma_y\sqrt{1-r^2}\sqrt{2\pi}} e^{-\frac{\left(y-\left(a_2+r\frac{\sigma_y}{\sigma_x}(x-a_1)\right)\right)^2}{2\sigma_y^2(1-r^2)}}.$$

Полученное распределение является нормальным, а его математическое ожидание$M(Y/x) = a_2 + r\frac{\sigma_y}{\sigma_x}(x - a_1)$ есть функция регрессии Y на X (см. определение 11.4)). Аналогично можно получить функцию регрессии XнаY:

$$M(Y/x) = a_1 + r\frac{\sigma_x}{\sigma_y}(y - a_2).$$

Обе функции регрессии линейны, поэтому корреляция междуX и Y линейна, что и требовалось доказать. При этом уравнения прямых регрессии имеют вид

$$y - a_2 = r\frac{\sigma_y}{\sigma_x}(x - a_1), x - a_1 = r\frac{\sigma_x}{\sigma_y}(y - a_2),$$

то есть совпадают с уравнениями прямых среднеквадратической регрессии (см. формулы (11.5), (11.6)).

12. РАСПРЕДЕЛЕНИЯ «ХИ-КВАДРАТ», СТЬЮДЕНТА И ФИШЕРА

Рассмотрим некоторые распределения, связанные с нормальным и широко применяющиеся в математической статистике.

Распределение «хи-квадрат»

Пусть имеется несколько нормированных нормально распределенных случайных величин: $X_1, X_2, \dots, X_n (a_i = 0, \sigma_i = 1)$. Тогда сумма их квадратов

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \qquad (12.1)$$

является случайной величиной, распределенной по так называемому **закону «хи-квадрат»**с$k = n$степенями свободы; если же слагаемые связаны каким-либо соотношением (например,$\sum X_i = n\bar{X}$), то число степеней свободы $k = n - 1$.

Плотность этого распределения

$$f(x) = \begin{cases} 0, x \le 0, \\ \frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} e^{-\frac{x}{2}} x^{\frac{k}{2} - 1}, x > 0. \end{cases} \qquad (12.2)$$

Здесь $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$– гамма-функция;
в частности, $\Gamma(n + 1) = n!$

Следовательно, распределение «хи-квадрат» определяется одним параметром – числом степеней свободы k.

Замечание 1. С увеличением числа степеней свободы распределение «хи-квадрат» постепенно приближается к нормальному.

Замечание 2. С помощью распределения «хи-квадрат» определяются многие другие распределения, встречающиеся на практике, например, распределение случайной величины$\sqrt{\chi^2}$– длины случайного вектора $(X_1, X_2, \dots, X_n)$координаты которого независимы и распределены по нормальному закону.

Распределение Стьюдента

Рассмотрим две независимые случайные величины: Z, имеющую нормальное распределение и нормированную (то есть $M(Z) = 0, \sigma(Z) = 1$), иV, распределенную по закону «хи-квадрат» сk степенями свободы. Тогда величина

$$T = \frac{Z}{\sqrt{\frac{V}{k}}} \qquad (12.3)$$

имеет распределение, называемое **-распределением или распределением Стьюдента** с k степенями свободы.

С возрастанием числа степеней свободы распределение Стьюдента быстро приближается к нормальному.

Распределение F Фишера-Снедекора

Рассмотрим две независимые случайные величины U и V, распределенные по закону «хи-квадрат» со степенями свободы k_1 и k_2 и образуем из них новую величину

$$F = \frac{U/k_1}{V/k_2} \tag{12.4}$$

Ее распределение называют **распределением F Фишера-Снедекора** со степенями свободы k_1 и k_2. Плотность его распределения имеет вид

$$f(x) = \begin{cases} 0, x \le 0, \\ C_0 = \frac{x^{\frac{k_1-2}{2}}}{(k_2+k_1x)^{\frac{k_1+k_2}{2}}}, x > 0, \end{cases} \tag{12.5}$$

где $C_0 = \frac{\Gamma\left(\frac{k_1+k_2}{2}\right)k_1^{\frac{k_1}{2}}k_2^{\frac{k_2}{2}}}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}$. Таким образом, распределение Фишера определяется двумя параметрами – числами степеней свободы.

13. ЗАКОН БОЛЬШИХ ЧИСЕЛ. НЕРАВЕНСТВО ЧЕБЫШЕВА. ТЕОРЕМЫ ЧЕБЫШЕВА И БЕРНУЛЛИ

Изучение статистических закономерностей позволило установить, что при некоторых условиях суммарное поведение большого количества случайных величин почти утрачивает случайный характер и становится закономерным (иначе говоря, случайные отклонения от некоторого среднего поведения взаимно погашаются). В частности, если влияние на сумму отдельных слагаемых является равномерно малым, закон распределения суммы приближается к нормальному. Математическая формулировка этого утверждения дается в группе теорем, называемой **законом больших чисел**.

Неравенство Чебышева

Неравенство Чебышева, используемое для доказательства дальнейших теорем, справедливо как для непрерывных, так и для дискретных случайных величин. Докажем его для дискретных случайных величин.

Теорема 13.1(неравенство Чебышева).

$$p(|X - M(X)| < \varepsilon) \ge D(X)/\varepsilon^2. \tag{13.1}$$

Доказательство. Пусть X задается рядом распределения

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

$$D(X) = \left(x_1 - M(X)\right)^2 p_1 + \left(x_2 - M(X)\right)^2 p_2 + \dots + \left(x_n - M(X)\right)^2 p_n.$$

Так как события $|X - M(X)| < \varepsilon$ и $|X - M(X)| \geq \varepsilon$ противоположны, то $p(|X - M(X)| < \varepsilon) + p(|X - M(X)| \geq \varepsilon) = 1$, следовательно, $p(|X - M(X)| < \varepsilon) = 1 - p(|X - M(X)| \geq \varepsilon)$. Найдем $p(|X - M(X)| \geq \varepsilon)$.

$$D(X) = \left(x_1 - M(X)\right)^2 p_1 + \left(x_2 - M(X)\right)^2 p_2 + \dots + \left(x_n - M(X)\right)^2 p_n.$$

Исключим из этой суммы те слагаемые, для которых $|X - M(X)| < \varepsilon$. При этом сумма может только уменьшиться, так как все входящие в нее слагаемые неотрицательны. Для определенности будем считать, что отброшены первые k слагаемых. Тогда

$$D(X) \geq \left(x_{k+1} - M(X)\right)^2 p_{k+1} + \left(x_{k+2} - M(X)\right)^2 p_{k+2} + \dots + \\ + \left(x_n - M(X)\right)^2 p_n \geq \varepsilon^2 (p_{k+1} + p_{k+2} + \dots + p_n).$$

Отметим, что $(p_{k+1} + p_{k+2} + \dots + p_n)$ есть вероятность того, что $|X - M(X)| \geq \varepsilon$, так как это сумма вероятностей всех возможных значений *X*, для которых это неравенство справедливо. Следовательно, $D(X) \geq \varepsilon^2 \cdot p(|X - M(X)| \geq \varepsilon)$, или $p(|X - M(X)| \geq \varepsilon) \leq D(X)/\varepsilon^2$. Тогда вероятность противоположного события $p(|X - M(X)| < \varepsilon) \geq D(X)/\varepsilon^2$, что и требовалось доказать.

Теоремы Чебышева и Бернулли

Теорема 13.2 (теорема Чебышева). Если $X_1, X_2, \dots, X_n$ – попарно независимые случайные величины, дисперсии которых равномерно ограничены $(D(X_i) \leq C)$, то для сколь угодно малого числа ε вероятность неравенства

$$\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}\right| < \varepsilon$$

будет сколь угодно близка к 1, если число случайных величин достаточно велико.

Замечание. Иначе говоря, при выполнении этих условий

$$\lim_{n\to\infty} p\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{M(X_1) + M(X_2) + \cdots + M(X_n)}{n}\right| < \varepsilon\right) = 1.$$

Доказательство. Рассмотрим новую случайную величину $\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$ и найдем ее математическое ожидание. Используя свойства математического ожидания, получим, что

$$M\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{M(X_1) + M(X_2) + \cdots + M(X_n)}{n}.$$

Применим к $\bar{X}$ неравенство Чебышева:

$$p\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{M(X_1) + M(X_2) + \cdots + M(X_n)}{n}\right| < \varepsilon\right) \geq$$
$$\geq 1 - \frac{D\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)}{\varepsilon^2}.$$

Так как рассматриваемые случайные величины независимы, то, учитывая условие теоремы, имеем:

$$D\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{D(X_1) + D(X_2) + \cdots + D(X_n)}{n^2} \leq \frac{C_n}{n^2} = \frac{C}{n}.$$

Используя этот результат, представим предыдущее неравенство в виде:

$$p\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{M(X_1) + M(X_2) + \cdots + M(X_n)}{n}\right| < \varepsilon\right) \geq 1 - \frac{C}{n\varepsilon^2}.$$

Перейдем к пределу при $n \to \infty$:

$$\lim_{n\to\infty} p\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{M(X_1) + M(X_2) + \cdots + M(X_n)}{n}\right| < \varepsilon\right) \geq 1.$$

Поскольку вероятность не может быть больше 1, можно утверждать, что

$$\lim_{n\to\infty} p\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{M(X_1) + M(X_2) + \cdots + M(X_n)}{n}\right| < \varepsilon\right) = 1.$$

Теорема доказана.

Следствие. Если $X_1, X_2, \ldots, X_n$ – попарно независимые случайные величины с равномерно ограниченными дисперсиями, имеющие одинаковое математическое ожидание, равное a, то для любого сколь угодно малого $\varepsilon > 0$ вероятность неравенства $\left|\frac{X_1+X_2+\cdots+X_n}{n} - a\right| < \varepsilon$ будет как угодно близка к 1, если число случайных величин достаточно велико. Иначе говоря, $\lim_{n\to\infty} p\left(\left|\frac{X_1+X_2+\cdots+X_n}{n} - a\right| < \varepsilon\right) = 1$.

Вывод: среднее арифметическое достаточно большого числа случайных величин принимает значения, близкие к сумме их математических ожиданий, то есть утрачивает характер случайной величины. Например, если проводится серия измерений какой-либо физической величины, причем: а) результат каждого измерения не зависит от результатов остальных, то есть все результаты представляют собой попарно независимые случайные величины;б) измерения производятся без систематических ошибок (их математические ожидания равны между собой и равны истинному значению a измеряемой величины);в) обеспечена определенная точность измерений, следовательно, дисперсии рассматриваемых случайных величин равномерно ограничены; то при достаточно большом числе измерений их среднее арифметическое окажется сколь угодно близким к истинному значению измеряемой величины.

Теорема Бернулли

Теорема 13.3 (теорема Бернулли). Если в каждом из n независимых опытов вероятность p появления события A постоянна, то при достаточно большом числе испытаний вероятность того, что модуль отклонения относительной частоты появлений A в n опытах от p будет сколь угодно малым, как угодно близка к 1:

$$\lim_{n\to\infty} p\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1. \qquad (13.2)$$

Доказательство. Введем случайные величины $X_1, X_2, \ldots, X_n$, где X_i – число появлений A в-м опыте. При этом X_i могут принимать только два значения: 1(с вероятностью p) и 0 (с вероятностью $q = 1 - p$). Кроме того, рассматриваемые случайные величины попарно независимы и их дисперсии равномерно ограничены (так как $D(X_i) = pq, p + q = 1$, откуда $pq \leq 1/4$). Следовательно, к ним можно применить теорему Чебышева при $M_i = p$:

$$\lim_{n\to\infty} p\left(\left|\frac{X_1+X_2+\cdots+X_n}{n} - p\right| < \varepsilon\right) = 1.$$

Но$\frac{X_1+X_2+\cdots+X_n}{n}=\frac{m}{n}$, так как X_iпринимает значение, равное 1, при появлении A в данном опыте, и значение, равное 0, если A не произошло. Таким образом,

$$\lim_{n\to\infty} p\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=1$$

что и требовалось доказать.

Замечание. Из теоремы Бернулли *не следует*, что $\lim\limits_{n\to\infty}\frac{m}{n}=p$. Речь идет лишь о *вероятности* того, что разность относительной частоты и вероятности по модулю может стать сколь угодно малой. Разница заключается в следующем: при обычной сходимости, рассматриваемой в математическом анализе, для всех n, начиная с некоторого значения, неравенство $\left|\frac{m}{n}-p\right|<\varepsilon$выполняется всегда; в нашем случае могут найтись такие значения n, при которых это неравенство неверно. Этот вид сходимости называют **сходимостью по вероятности**.

14. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА ЛЯПУНОВА. ПРЕДЕЛЬНАЯ ТЕОРЕМА МУАВРА-ЛАПЛАСА

Закон больших чисел не исследует вид предельного закона распределения суммы случайных величин. Этот вопрос рассмотрен в группе теорем, называемых **центральной предельной теоремой.** Они утверждают, что закон распределения суммы случайных величин, каждая из которых может иметь различные распределения, приближается к нормальному при достаточно большом числе слагаемых. Этим объясняется важность нормального закона для практических приложений.

Характеристические функции

Для доказательства центральной предельной теоремы используется метод характеристических функций.

Определение 14.1. **Характеристической функцией**случайной величины Xназывается функция

$$g(t)=M\left(e^{itX}\right) \tag{14.1}$$

Таким образом,$g(t)$представляет собой математическое ожидание некоторой комплексной случайной величины $U=e^{itX}$, связанной с величиной X. В частности, если X – дискретная случайная величина, заданная рядом распределения, то

$$g(t) = \sum_{k\to 1}^{n} e^{itx_k} p_k \qquad (14.2)$$

Для непрерывной случайной величины с плотностью распределения $f(x)$

$$g(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx \qquad (14.3)$$

Пример 1. Пусть X – число выпадений 6 очков при одном броске игральной кости. Тогда по формуле (14.2) $g(t) = e^{it0} \cdot \frac{5}{6} + e^{it1} \cdot \frac{1}{6} = \frac{5+e^{it}}{6}$.

Пример 2. Найдем характеристическую функцию для нормированной непрерывной случайной величины, распределенной по нормальному закону $\left(f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}\right)$. По формуле (14.3)

$$g(t) = \int_{-\infty}^{+\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx - \frac{x^2}{2}} dx = e^{\frac{t^2}{2}}$$

(использовалась формула $\int_{-\infty}^{+\infty} e^{-Ax^2 \pm 2Bx - C} dx = \sqrt{\frac{\pi}{A}} e^{-\frac{AC-B^2}{A}}$ и то, что $i^2 = -1$).

Свойства характеристических функций

1. Функцию $f(x)$ можно найти по известной функции $g(t)$ по формуле

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{itx} g(t) dt \qquad (14.4)$$

(преобразование (14.3) называется *преобразованием Фурье*, а преобразование (14.4) – *обратным преобразованием Фурье*).

2. Если случайные величины X и Y связаны соотношением $Y = aX$, то их характеристические функции связаны соотношением

$$g_y(t) = g_x(at). \qquad (14.5)$$

3. Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых: для $Y = \sum_{k=1}^{n} X_k$.

$$g_y(t) = g_{x_1}(t) \cdot g_{x_2}(t) \cdot \ldots \cdot g_{x_n}(t). \qquad (14.6)$$

Теорема 14.1 (центральная предельная теорема для одинаково распределенных слагаемых). Если $X_1, X_2, \dots, X_n, \dots$– независимые случайные величины с одинаковым законом распределения, математическим ожиданием m и дисперсией σ^2, то при неограниченном увеличении n закон распределения суммы$Y_n = \sum_{k=1}^{n} X_k$неограниченно приближается к нормальному.

Доказательство.

Докажем теорему для непрерывных случайных величин $X_1, X_2, \dots, X_n$ (доказательство для дискретных величин аналогично). Согласно условию теоремы, характеристические функции слагаемых одинаковы: $g_x(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)dx$. Тогда по свойству 3 характеристическая функция суммы Y_n будет$g_{y_n}(t) = g_x^n(t)$. Разложим функцию $g_x(t)$в ряд Маклорена:

$$g_x(t) = g_x(0) + g_x'(0)t + \left(\frac{g_x''(0)}{2} + \alpha(t)\right)t^2,$$

где $\alpha(t) \to 0$ при $t \to 0$.

Найдем

$$g_x(0) = \int_{-\infty}^{+\infty} f(x)dx = 1,$$

$$g_x'(0) = \int_{-\infty}^{+\infty} ixe^{itx} f(x)dx\,\Big|_{t=0} = i\int_{-\infty}^{+\infty} xe^{itx} f(x)dx\,\Big|_{t=0} = i\int_{-\infty}^{+\infty} xf(x)dx = im.$$

Если предположить, что $m = 0$(то есть перенести начало отсчета в точку m), то$g_x'(0) = 0$.

$g_x''(0) = -\int_{-\infty}^{+\infty} x^2 e^{itx} f(x)dx\,\Big|_{t=0} = -\int_{-\infty}^{+\infty} x^2 f(x)dx = -\sigma^2$(так как $m = 0$). Подставив полученные результаты в формулу Маклорена, найдем, что

$$g_x(t) = 1 - \left(\frac{\sigma^2}{2} - \alpha(t)\right)t^2.$$

Рассмотрим новую случайную величину$Z_n = \frac{Y_n}{\sigma\sqrt{n}}$, отличающуюся от Y_nтем, что ее дисперсия при любом n равна 0. Так как Y_nи Z_n связаны линейной зависимостью, достаточно доказать, что Z_nраспределена по нормальному закону, или, что то же самое, что ее характеристическая функция приближается к характеристической функции нормального закона (см. пример 2). По свойству характеристических функций

$$g_{Z_n}(t) = g_{y_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(g_x\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \left(\frac{\sigma^2}{2} - \alpha\left(\frac{t}{\sigma\sqrt{n}}\right)\right)\frac{t^2}{n\sigma^2}\right)^n.$$

Прологарифмируем полученное выражение:

$$\ln g_{Z_n}(t) = n\ln(1-k), \text{где } k = \left(\frac{\sigma^2}{2} - \alpha\left(\frac{t}{\sigma\sqrt{n}}\right)\right)\frac{t^2}{n\sigma^2}, \lim_{n\to\infty} k = 0.$$

Разложим $\ln(1-k)$в ряд при$n \to \infty$, ограничившись двумя членами разложения, тогда$\ln(1-k) \approx -k$. Отсюда

$$\lim_{n\to\infty} \ln g_{Z_n}(t) = \lim_{n\to\infty} n(-k) = \lim_{n\to\infty}\left(-\frac{t^2}{2} + \alpha\left(\frac{t}{\sigma\sqrt{n}}\right)\frac{t^2}{\sigma^2}\right) = \frac{t^2}{2} + \lim_{n\to\infty}\frac{t^2}{\sigma^2}\alpha\left(\frac{t}{\sigma\sqrt{n}}\right),$$

где последний предел равен 0, так как $\alpha(t) \to 0$ при$t \to 0$. Следовательно, $\lim_{n\to\infty} \ln g_{Z_n}(t) = -\frac{t^2}{2}$, то есть $\lim_{n\to\infty} \ln g_{Z_n}(t) = e^{-\frac{t^2}{2}}$– характеристическая функция нормального распределения. Итак, при неограниченном увеличении числа слагаемых характеристическая функция величины Z_n неограниченно приближается к характеристической функции нормального закона; следовательно, закон распределения Z_n(иY_n) неограниченно приближается к нормальному. Теорема доказана.

А.М.Ляпунов доказал центральную предельную теорему для условий более общего вида:

Теорема 14.2 (теорема Ляпунова). Если случайная величина X представляет собой сумму очень большого числа взаимно независимых случайных величин, для которых выполнено условие:

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} b_k}{\left(\sum_{k=1}^{n} D_k\right)^{\frac{3}{2}}} \qquad (14.7)$$

где b_k– третий абсолютный центральный момент величины X_k, аD_k – ее дисперсия, то X имеет распределение, близкое к нормальному (условие Ляпунова означает, что влияние каждого слагаемого на сумму ничтожно мало).

Практически можно использовать центральную предельную теорему при достаточно небольшом количестве слагаемых, так как вероятностные расчеты требуют сравнительно малой точности. Опыт показывает, что для суммы даже десяти и менее слагаемых закон их распределения можно заменить нормальным.

Частным случаем центральной предельной теоремы для дискретных случайных величин является теорема Муавра-Лапласа.

Теорема 14.3 (теорема Муавра-Лапласа). Если производится n независимых опытов, в каждом из которых событие A появляется с вероятностью p то справедливо соотношение:

$$p\left(\alpha < \frac{Y-np}{\sqrt{npq}} < \beta\right) = \Phi(\beta) - \Phi(\alpha), \qquad (14.8)$$

где Y– число появлений события A в n опытах $q = 1 - p$.

Доказательство.

Будем считать, что$Y = \sum_{i=1}^{n} X_i$, где X_i – число появлений события A в -мопыте. Тогда случайную величину $Z = \frac{Y-m_y}{\sigma_y}$ (см. теорему 14.1) можно считать распределенной по нормальному закону и нормированной, следовательно, вероятность ее попадания в интервал (α, β)можно найти по формуле

$$p(\alpha < Z < \beta) = \Phi(\beta) - \Phi(\alpha).$$

Поскольку Y имеет биномиальное распределение,$m_y = np, D_y = npq,\ \sigma_y = \sqrt{npq}$. Тогда$Z = \frac{Y-np}{\sqrt{npq}}$. Подставляя это выражение в предыдущую формулу, получим равенство (14.8).

Следствие.В условиях теоремы Муавра-Лапласа вероятность $p_n(k)$ того, что событие A появится в n опытах ровно k раз, при большом количестве опытов можно найти по формуле:

$$p_n(k) = \frac{1}{\sqrt{npq}} \cdot \varphi(x) \qquad (14.9)$$

где$x = \frac{k-np}{\sqrt{npq}}$ а $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$(значения этой функции приводятся в специальных таблицах).

Пример 3. Найти вероятность того, что при 100 бросках монеты число выпадений герба окажется в пределах от 40 до 60.

Применим формулу (14.8), учитывая, что $n = 0{,}5$.

Тогда $np = 100 \cdot 0{,}5 = 50$, $\sqrt{npq} = \sqrt{100 \cdot 0{,}5 \cdot (1 - 0{,}5)} = 5$.

Тогда, если $40 < Y < 60, -2 < \frac{Y-50}{5} < 2$.

Следовательно,

$$p(40 < Y < 60) = p\left(-2 < \frac{Y-50}{5} < 2\right) =$$
$$= \Phi(2) - \Phi(2) = 0{,}9772 - 0{,}0228 = 0{,}9544.$$

Пример 4. В условиях предыдущего примера найти вероятность того, что выпадет 45 гербов.

Найдем$x = \frac{45-50}{5} = -1$, тогда

$$p_{100}(45) \approx \frac{1}{5} \cdot \varphi(-1) = \frac{1}{5} \cdot \varphi(1) = \frac{1}{5} \cdot 0{,}2420 = 0{,}0484.$$

15. ОСНОВНЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ. ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРКА

Математическая статистика занимается установлением закономерностей, которым подчинены массовые случайные явления, на основе обработки статистических данных, полученных в результате наблюдений. Двумя основными задачами математической статистики являются:

- определение способов сбора и группировки этих статистических данных;
- разработка методов анализа полученных данных в зависимости от целей исследования, к которым относятся:

а) оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения, вид которого известен; оценка зависимости от других случайных величин и т.д.;

б) проверка статистических гипотез о виде неизвестного распределения или о значениях параметров известного распределения.

Для решения этих задач необходимо выбрать из большой совокупности однородных объектов ограниченное количество объектов, по результатам изучения которых можно сделать прогноз относительно исследуемого признака этих объектов.

Определим основные понятия математической статистики.

Генеральная совокупность– все множество имеющихся объектов.

Выборка – набор объектов, случайно отобранных из генеральной совокупности.

Объем генеральной совокупности N и объем выборки n – число объектов в рассматриваемой совокупности.

Виды выборки:

Повторная – каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность;

Бесповторная – отобранный объект в генеральную совокупность не возвращается.

Замечание. Для того, чтобы по исследованию выборки можно было сделать выводы о поведении интересующего нас признака генеральной совокупности, нужно, чтобы выборка правильно представляла пропорции генеральной совокупности, то есть была **репрезентативной** (представительной). Учитывая закон больших чисел, можно утверждать, что это условие выполняется, если каждый объект выбран случайно, причем для любого объекта вероятность попасть в выборку одинакова.

Первичная обработка результатов

Пусть интересующая нас случайная величина X принимает в выборке значение $x_1 - n_1$раз, $x_2 - n_2$раз, …, $x_k - n_k$раз, причем$\sum_{i=1}^{k} n_k = n$,где n – объем выборки. Тогда наблюдаемые значения случайной величины$x_1, x_2, \dots, x_k$называют **вариантами**, а $n_1, n_2, \dots, n_k$– **частотами**. Если разделить каждую частоту на объем выборки, то получим **относительные частоты** $w_i = \frac{n_i}{n}$.Последовательность вариант, записанных в порядке возрастания, называют **вариационным** рядом, а перечень вариант и соответствующих им частот или относительных частот – **статистическим рядом**:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
w_i	w_1	w_2	…	w_k

Пример.При проведении 20 серий из 10 бросков игральной кости число выпадений шести очков оказалось равным 1,1,4,0,1,2,1,2,2,0,5,3,3,1,0,2,2,3,4,1.Составим вариационный ряд: 0,1,2,3,4,5. Статистический ряд для абсолютных и относительных частот имеет вид:

x_i	0	1	2	3	4	5
n_i	3	6	5	3	2	1
w_i	0,15	0,3	0,25	0,15	0,1	0,05

Если исследуется некоторый непрерывный признак, то вариационный ряд может состоять из очень большого количества чисел. В этом случае удобнее использовать **группированную выборку**. Для ее получения интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько равных частичных интервалов длиной h, а затем находят для каждого частичного интервала n_i – сумму частот вариант, попавших в -й интервал. Составленная по этим результатам таблица называется **группированным статистическим рядом**:

Номера интервалов	1	2	…	k
Границы интервалов	$(a, a+h)$	$(a+h, a+2h)$	…	$(b-h, b)$
Сумма частот вариант, попавших в интервал	n_1	n_2	…	n_k

Полигон частот.
Выборочная функция распределения и гистограмма

Для наглядного представления о поведении исследуемой случайной величины в выборке можно строить различные графики. Один из них – **полигон частот**: ломаная, отрезки которой соединяют точки с координатами $(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k)$, где x_iоткладываются на оси абсцисс, аn_i– на оси ординат. Если на оси ординат откладывать не абсолютные (n_i), а относительные w_iчастоты, то получим **полигон относительных частот** (см. Рисунок 15.1)**.**

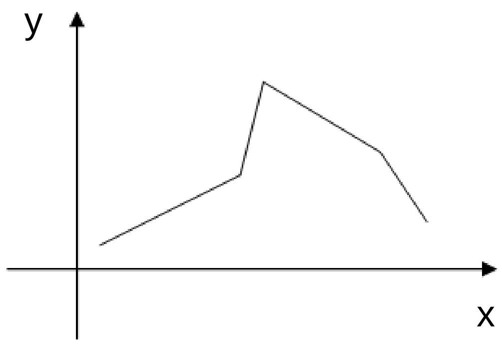

Рисунок 15.1

По аналогии с функцией распределения случайной величины можно задать некоторую функцию, относительную частоту события$X < x$.

Определение 15.1. **Выборочной (эмпирической) функцией распределения** называют функцию$F^*(x)$, определяющую для каждого значения x относительную частоту события$X < x$*.* Таким образом,

$$F^*(x) = \frac{n_x}{n}, \qquad (15.1)$$

где n_x – число вариант, меньших x, n – объем выборки.

Замечание. В отличие от эмпирической функции распределения, найденной опытным путем, функцию распределения $F(x)$генеральной совокупности называют *теоретической функцией распределения.*$F(x)$определяет вероятность события $X < x$, а$F^*(x)$ – его относительную частоту. При достаточно большихn, как следует из теоремы Бернулли, $F^*(x)$стремится по вероятности к $F(x)$.

Из определения эмпирической функции распределения видно, что ее свойства совпадают со свойствами $F(x)$, а именно:

1) $0 \le F^*(x) \le 1$.
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$x > x_k$.

Для непрерывного признака графической иллюстрацией служит **гистограмма**, то есть ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h, а высотами – отрезки длинойn_i/h (гистограмма частот) илиw_i/h(гистограмма относительных

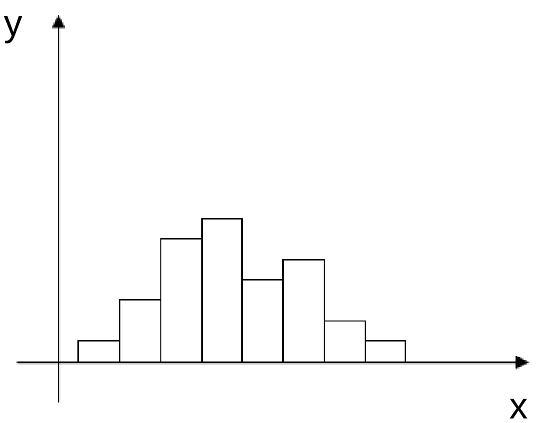

Рисунок 15.2

частот). В первом случае площадь гистограммы равна объему выборки, во втором – единице (см. Рисунок 15.2).

16. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины.

Определение 16.1. **Выборочным средним** называется среднее арифметическое значений случайной величины, принимаемых в выборке:

$$\overline{x}_B = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_2 x_n}{n} = \frac{\sum_{i=1}^{k} n_i x_i}{n}, \qquad (16.1)$$

где x_i – варианты, n_i–частоты.

Замечание. Выборочное среднее служит для оценки математического ожидания исследуемой случайной величины. В дальнейшем будет рассмотрен вопрос, насколько точной является такая оценка.

Определение 16.2. **Выборочной дисперсией** называется

$$D_{\text{в}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_B)^2}{n} = \frac{\sum_{i=1}^{n} n_i (x_i - \overline{x}_B)^2}{n}, \qquad (16.2)$$

а **выборочным средним квадратическим отклонением**–

$$\sigma_B = \sqrt{D_{\text{в}}}. \qquad (16.3)$$

Так же, как в теории случайных величин, можно доказать, что справедлива следующая формула для вычисления выборочной дисперсии:

$$D = \overline{x^2} - (\bar{x})^2. \qquad (16.4)$$

Пример 1. Найдем числовые характеристики выборки, заданной статистическим рядом

x_i	2	5	7	8
n_i	3	8	7	2

$$\bar{x}_{\text{в}} = \frac{2 \cdot 3 + 5 \cdot 8 + 7 \cdot 7 + 8 \cdot 2}{20} = 5{,}55; D_B = \frac{4 \cdot 3 + 25 \cdot 8 + 49 \cdot 7 + 64 \cdot 2}{20}$$
$$= 5{,}55^2 = 3{,}3475;$$

$$\sigma_B = \sqrt{3{,}3475} = 1{,}83.$$

Другими характеристиками вариационного ряда являются:

- **мода** $\boldsymbol{M_0}$ – варианта, имеющая наибольшую частоту (в предыдущем примере$M_0 = 5$).

- **медиана** $\boldsymbol{m_e}$–варианта, которая делит вариационный ряд на две части, равные по числу вариант. Если число вариант нечетно $(n = 2k + 1)$, то $m_e = x_{k+1}$, а при четном $n = 2k m_e = \frac{x_k + x_{k+1}}{2}$. В частности, в примере 1$m_e = \frac{5+7}{2} = 6$.

Оценки начальных и центральных моментов (так называемые эмпирические моменты) определяются аналогично соответствующим теоретическим моментам:

- начальным эмпирическим моментом порядка $\boldsymbol{k}$ называется

$$M_k = \frac{\sum n_i x_i^k}{n}. \tag{16.5}$$

В частности, $M_1 = \frac{\sum n_i x_i}{n} = \bar{x}_B$, то есть начальный эмпирический момент первого порядка равен выборочному среднему.

- **центральным эмпирическим моментом порядка $\boldsymbol{k}$** называется

$$m_k = \frac{\sum n_i (x_i - \bar{x}_{\text{в}})^k}{n}. \tag{16.6}$$

В частности, $m_2 = \frac{\sum n_i (x_i - \bar{x}_B)^2}{n} = D_{\text{в}}$, то есть центральный эмпирический момент второго порядка равен выборочной дисперсии.

Статистическое описание и вычисление характеристик двумерного случайного вектора

При статистическом исследовании двумерных случайных величин основной задачей является обычно выявление связи между составляющими.

Двумерная выборка представляет собой набор значений случайного вектора: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Для нее можно определить выборочные средние составляющих: $\bar{x}_B = \frac{\sum x_i}{n}, \bar{y}_B = \frac{\sum y_i}{n}$и соответствующие выборочные дисперсии и средние квадратические отклонения. Кроме того, можно вычислить **условные средние**: $\bar{y}_x$– среднее арифметическое наблюдавшихся значений Y, соответствующих $X = x$, и $\bar{x}_y$– среднее значение наблюдавшихся значений X, соответствующих $Y = y$.

Если существует зависимость между составляющими двумерной случайной величины, она может иметь разный вид: функциональная зависимость, если каждому возможному значению X соответствует одно значение Y, и статистическая, при которой изменение одной величины приводит к изменению распределения другой. Если при этом в результате изменения одной величины меняется среднее значение другой, то статистическую зависимость между ними называют корреляционной.

17. ОСНОВНЫЕ СВОЙСТВА СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Получив статистические оценки параметров распределения (выборочное среднее, выборочную дисперсию и т.д.), нужно убедиться, что они в достаточной степени служат приближением соответствующих характеристик генеральной совокупности. Определим требования, которые должны при этом выполняться.

Пусть Θ^*– статистическая оценка неизвестного параметра Θ теоретического распределения. Извлечем из генеральной совокупности несколько выборок одного и того же объема n и вычислим для каждой из них оценку параметра Θ: $\Theta_1^*, \Theta_2^*, \dots, \Theta_k^*$. Тогда оценку Θ^* можно рассматривать как случайную величину, принимающую возможные значения $\Theta_1^*, \Theta_2^*, \dots, \Theta_k^*$. Если математическое ожидание Θ^* не равно оцениваемому параметру, мы будем получать при вычислении оценок систематические ошибки одного знака (с избытком, если $M(\Theta^*) > \Theta$, и с недостатком, если $M(\Theta^*) < \Theta$). Следовательно, необходимым условием отсутствия систематических ошибок является требование $M(\Theta^*) = \Theta$.

Определение 17.2. Статистическая оценка Θ^* называется **несмещенной**, если ее математическое ожидание равно оцениваемому параметруΘ при любом объеме выборки:

$$M(\Theta^*) = \Theta. \qquad (17.1)$$

Смещенной называют оценку, математическое ожидание которой не равно оцениваемому параметру.

Однако несмещенность не является достаточным условием хорошего приближения к истинному значению оцениваемого параметра. Если при этом возможные значения Θ^* могут значительно отклоняться от среднего значения, то есть дисперсия Θ^* велика, то значение, найденное по данным одной выборки, может значительно отличаться от оцениваемого параметра. Следовательно, требуется наложить ограничения на дисперсию.

Определение 17.2. Статистическая оценка называется **эффективной**, если она при заданном объеме выборки n имеет наименьшую возможную дисперсию.

При рассмотрении выборок большого объема к статистическим оценкам предъявляется еще и требование состоятельности.

Определение 17.3. **Состоятельной** называется статистическая оценка, которая при$n \to \infty$стремится по вероятности к оцениваемому параметру (если эта оценка несмещенная, то она будет состоятельной, если при $n \to \infty$ ее дисперсия стремится к 0).

Убедимся, что $\bar{x}_в$ представляет собой несмещенную оценку математического ожидания $M(X)$.

Будем рассматривать $\bar{x}_в$ как случайную величину, а $x_1, x_2, \dots, x_n$, то есть значения исследуемой случайной величины, составляющие выборку,– как независимые, одинаково распределенные случайные величины $X_1, X_2, \dots, X_n$, имеющие математическое ожидание a. Из свойств математического ожидания следует, что

$$M(\bar{X}_в) = M\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = a.$$

Но, поскольку каждая из величин $X_1, X_2, \dots, X_n$имеет такое же распределение, что и генеральная совокупность, $a = M(X)$, то есть $M(\bar{X}_в) = M(X)$, что и требовалось доказать. Выборочное среднее является не только несмещенной, но и состоятельной оценкой математического ожидания. Если предположить, что $X_1, X_2, \dots, X_n$имеют ограниченные дисперсии, то из теоремы Чебышева следует, что их $\bar{X}_в$среднее арифметическое, то есть, при увеличении n стремится по вероятности к математическому ожиданию a каждой их величин, то есть к $M(X)$. Следовательно, выборочное среднее есть состоятельная оценка математического ожидания.

В отличие от выборочного среднего, выборочная дисперсия является смещенной оценкой дисперсии генеральной совокупности. Можно доказать, что

$$M(D_в) = \frac{n-1}{n} D_г, \qquad (17.2)$$

где $D_г$– истинное значение дисперсии генеральной совокупности. Можно предложить другую оценку дисперсии – **исправленную дисперсию** s^2, вычисляемую по формуле

$$s^2 = \frac{n}{n-1} D_в = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_в)^2}{n-1}. \qquad (17.3)$$

Такая оценка будет являться несмещенной. Ей соответствует **исправленное среднее квадратическое отклонение**

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_в)^2}{n-1}} \qquad (17.4)$$

Определение 17.4. Оценка некоторого признака называется **асимптотически несмещенной**, если для выборки $x_1, x_2, \dots, x_n$

$$\lim_{n \to \infty} \frac{x_1 + x_2 + \dots + x_n}{n} = X, \qquad (17.5)$$

где X – истинное значение исследуемой величины.

Способы построения оценок

1. *Метод наибольшего правдоподобия*

Пусть X – дискретная случайная величина, которая в результате n испытаний приняла значения $x_1, x_2, \dots, x_n$. Предположим, что нам известен закон распределения этой величины, определяемый параметром Θ, но неизвестно численное значение этого параметра. Найдем его точечную оценку.

Пусть $p(x_i; \Theta)$ – вероятность того, что в результате испытания величина X примет значение x_i. Назовем **функцией правдоподобия** дискретной случайной величины X функцию аргумента Θ, определяемую по формуле:

$$L(x_1, x_2, \dots, x_n; \Theta) = p(x_1, \Theta) p(x_2, \Theta) \dots p(x_n, \Theta).$$

Тогда в качестве точечной оценки параметра Θ принимают такое его значение $\Theta^* = \Theta(x_1, x_2, \dots, x_n)$, при котором функция правдоподобия достигает максимума. Оценку Θ^* называют **оценкой наибольшего правдоподобия**.

Поскольку функции L и $\ln L$ достигают максимума при одном и том же значении Θ, удобнее искать максимум $\ln L$ – **логарифмической функции правдоподобия**. Для этого нужно:

1) найти производную $\frac{d \ln L}{d\Theta}$;

2) приравнять ее нулю (получим так называемое *уравнение правдоподобия*) и найти критическую точку;

3) найти вторую производную $\frac{d^2 \ln L}{d\Theta^2}$; если она отрицательна в критической точке, то это – точка максимума.

Достоинства метода наибольшего правдоподобия: полученные оценки состоятельны (хотя могут быть смещенными), распределены асимптотически нормально при больших значениях n и имеют наименьшую дисперсию по сравнению с другими асимптотически нормальными оценками; если для оцениваемого параметра θ существует эффективная оценка Θ^*, то уравнение правдоподобия имеет единственное решениеΘ^*; метод наиболее полно использует данные выборки и поэтому особенно полезен в случае малых выборок.

Недостаток метода наибольшего правдоподобия: сложность вычислений.

Для непрерывной случайной величины с известным видом плотности распределения$f(x)$и неизвестным параметром Θ функция правдоподобия имеет вид:

$$L(x_1, x_2, \dots, x_n; \Theta) = f(x_1, \Theta) f(x_2, \Theta) \dots f(x_n, \Theta).$$

Оценка наибольшего правдоподобия неизвестного параметра проводится так же, как для дискретной случайной величины.

2. *Метод моментов*

Метод моментов основан на том, что начальные и центральные эмпирические моменты являются состоятельными оценками соответственно начальных и центральных теоретических моментов, поэтому можно приравнять теоретические моменты соответствующим эмпирическим моментам того же порядка.

Если задан вид плотности распределения$f(x, \Theta)$,определяемой одним неизвестным параметром Θ, то для оценки этого параметра достаточно иметь одно уравнение. Например, можно приравнять начальные моменты первого порядка:

$$\bar{x}_B = M(X) = \int_{-\infty}^{\infty} x f(x; \Theta) dx = \varphi(\Theta)$$

получив тем самым уравнение для определения Θ. Его решение Θ^* будет точечной оценкой параметра, которая является функцией от выборочного среднего и, следовательно, и от вариант выборки:

$$\Theta = \Psi(x_1, x_2, \dots, x_n).$$

Если известный вид плотности распределения $f(x, \Theta_1, \Theta_2)$определяется двумя неизвестными параметрамиΘ_1 и Θ_2, то требуется составить два уравнения, например

$$v_1 = M_1, \mu_2 = m_2.$$

Отсюда$\begin{cases} M(X) = \bar{x}_в \\ D(X) = D_в \end{cases}$– система двух уравнений с двумя неизвестными Θ_1 и Θ_2. Ее решениями будут точечные оценки Θ_1^* и Θ_2^*– функции вариант выборки:

$$\Theta_1 = \Psi_1(x_1, x_2, \ldots, x_n),$$
$$\Theta_2 = \Psi_2(x_1, x_2, \ldots, x_n).$$

3. Метод наименьших квадратов

Если требуется оценить зависимость величин y и x, причем известен вид связывающей их функции, но неизвестны значения входящих в нее коэффициентов, их величины можно оценить по имеющейся выборке с помощью метода наименьших квадратов. Для этого функция $y = \varphi(x)$ выбирается так, чтобы сумма квадратов отклонений наблюдаемых значений $y_1, y_2, \ldots, y_n$ от $\varphi(x_i)$ была минимальной:

$$\sum_{i=1}^{n} (y_i - \varphi(x_i))^2 = min$$

При этом требуется найти стационарную точку функции $\varphi(x; a, b, c, \ldots)$, то есть решить систему:

$$\begin{cases} \sum_{i=1}^{n} \big(y_i - \varphi(x_i; a, b, c, \ldots)\big) \left(\frac{\partial \varphi}{\partial a}\right)_i = 0, \\ \sum_{i=1}^{n} \big(y_i - \varphi(x_i; a, b, c, \ldots)\big) \left(\frac{\partial \varphi}{\partial b}\right)_i = 0, \\ \sum_{i=1}^{n} \big(y_i - \varphi(x_i; a, b, c, \ldots)\big) \left(\frac{\partial \varphi}{\partial c}\right)_i = 0. \\ \ldots \ldots \ldots \end{cases}$$

(решение, конечно, возможно только в случае, когда известен конкретный вид функции φ).

Рассмотрим в качестве примера подбор параметров линейной функции методом наименьших квадратов.

Для того, чтобы оценить параметры a и b в функции $y = ax + b$, найдем

$$\left(\frac{\partial \varphi}{\partial a}\right)_i = x_i; \ \left(\frac{\partial \varphi}{\partial b}\right)_i = 1.$$

Тогда

$$\begin{cases} \sum_{i=1}^{n} \big(y_i - \varphi(ax_i + b)\big)x_i = 0, \\ \sum_{i=1}^{n} \big(y_i - \varphi(ax_i + b)\big) = 0\,. \end{cases}$$

Отсюда

$$\begin{cases} \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i^2 - b\sum_{i=1}^{n} x_i = 0, \\ \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i - bn = 0. \end{cases}$$

Разделив оба полученных уравнения на n и вспомнив определения эмпирических моментов, можно получить выражения дляaи bв виде:

$$a = \frac{(K_{xy})_{\text{в}}}{(D_x)_{\text{в}}}, b = \bar{y}_B - \frac{(K_{xy})_{\text{в}}}{(D_x)_{\text{в}}}\bar{x}_{\text{в}}.$$

Следовательно, связь междуx и y можно задать в виде:

$$y - \bar{y}_{\text{в}} = \frac{\big(K_{xy}\big)_{\text{в}}}{(D_x)_{\text{в}}}(x - \bar{x}_{\text{в}}).$$

4. *Байесовский подход к получению оценок*

Пусть(Y, X) – случайный вектор, для которого известна плотность$p(y|x)$ условного распределения Y при каждом значении $X = x$. Если в результате эксперимента получены лишь значения Y, а соответствующие значения X неизвестны, то для оценки некоторой заданной функции $\varphi(x)$ в качестве ее приближенного значения предлагается искать условное математическое ожидание $M\,(\varphi(x)|Y)$, вычисляемое по формуле:

$$\psi(Y) = \frac{\int \varphi(x)p(Y\ |x)p(x)d\mu(x)}{q(Y)},$$

где $q(y) = \int \varphi(x)p(y|x)p(x)d\mu(x)$, $p(x)$– плотность безусловного распределения $X, q(y)$ – плотность безусловного распределения Y. Задача может быть решена только тогда, когда известна $p(x)$. Иногда, однако, удается построить состоятельную оценку для $q(y)$, зависящую только от полученных в выборке значений Y.

18. ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ. ТОЧНОСТЬ ОЦЕНКИ, ДОВЕРИТЕЛЬНАЯ ВЕРОЯТНОСТЬ (НАДЕЖНОСТЬ), ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ

При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. Поэтому в таком случае лучше пользоваться *интервальными оценками*, то есть указывать интервал, в который с заданной вероятностью попадает истинное значение оцениваемого параметра. Разумеется, чем меньше длина этого интервала, тем точнее оценка параметра. Поэтому, если для оценки Θ^* некоторого параметра Θ справедливо неравенство $|\Theta^* - \Theta| < \delta$, число $\delta > 0$характеризует **точность оценки** (чем меньше δ, тем точнее оценка). Но статистические методы позволяют говорить только о том, что это неравенство выполняется с некоторой вероятностью.

Определение 18.1. **Надежностью (доверительной вероятностью)**оценки Θ^* параметра Θ называется вероятность γ того, что выполняется неравенство$|\Theta^* - \Theta| < \delta$. Если заменить это неравенство двойным неравенством $-\delta < \Theta^* - \Theta < \delta$, то получим:

$$p(\Theta^* - \delta < \Theta < \Theta^* + \delta) = \gamma.$$

Таким образом, γ есть вероятность того, что Θ попадает в интервал$(\Theta^* - \delta, \Theta^* + \delta)$.

Определение 18.2. **Доверительным** называется интервал, в который попадает неизвестный параметр с заданной надежностью γ.

Построение доверительных интервалов

1. *Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии*

Пусть исследуемая случайная величина X распределена по нормальному закону с известным средним квадратическимσ, и требуется по значению выборочного среднего $\bar{x}_B$оценить ее математическое ожидание a. Будем рассматривать выборочное среднее $\bar{x}_B$ как случайную величину $\bar{X}$, а значения вариант выборки $x_1, x_2, \dots, x_n$как одинаково распределенные независимые случайные величины $X_1, X_2, \dots, X_n$, каждая из которых имеет математическое ожидание a и среднее квадратическое отклонение σ.При этом $M(\bar{X}) = a$,$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$(используем свойства математического ожидания и дисперсии суммы независимых случайных величин). Оценим вероятность выполнения неравенства$|\bar{X} - a| < \delta$. Применим формулу для вероятности попадания нормально распределенной случайной величины в заданный интервал:

$$p(|X-a|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Тогда, с учетом того, что

$$\sigma(\bar{X})=\frac{\sigma}{\sqrt{n}}, p(|\bar{X}-a|<\delta)=2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)=2\Phi(t).$$

где

$$t=\frac{\delta\sqrt{n}}{\sigma}.$$

Отсюда

$$\delta=\frac{t\sigma}{\sqrt{n}},$$

и предыдущее равенство можно переписать так:

$$p\left(\bar{x}_{\text{в}}-\frac{t\sigma}{\sqrt{n}}<a<\bar{x}_{\text{в}}+\frac{t\sigma}{\sqrt{n}}\right)=2\Phi(t)=\gamma. \qquad (18.1)$$

Итак, значение математического ожидания a с вероятностью (надежностью) γ попадает в интервал$\left(\bar{x}_{\text{в}}-\frac{t\sigma}{\sqrt{n}};\ \bar{x}_{\text{в}}+\frac{t\sigma}{\sqrt{n}}\right)$, где значение t определяется из таблиц для функции Лапласа так, чтобы выполнялось равенство $2\Phi(t)=\gamma$.

Пример. Найдем доверительный интервал для математического ожидания нормально распределенной случайной величины, если объем выборки $n = 49$, $\bar{x}_{\text{в}}=2{,}8, \sigma = 1{,}4$, а доверительная вероятность $\gamma = 0{,}9$.

Определим t, при котором $\Phi(t) = 0{,}9{:}\,2 = 0{,}45{:}\ t = 1{,}645$. Тогда $2{,}8-\frac{1{,}645\cdot 1{,}4}{\sqrt{49}}<a< 2{,}8+\frac{1{,}645\cdot 1{,}4}{\sqrt{14}}$,или $2{,}471 <a< 3{,}129$. Найден доверительный интервал, в который попадает a с надежностью 0,9.

2. *Доверительный интервал для оценки математического ожидания нормального распределения при неизвестной дисперсии*

Если известно, что исследуемая случайная величина X распределена по нормальному закону с неизвестным средним квадратическим отклонением, то для поиска доверительного интервала для ее математического ожидания построим новую случайную величину

$$T=\frac{\bar{x}_{\text{в}}-a}{\frac{s}{\sqrt{n}}}, \qquad (18.2)$$

где $\bar{x}_{\text{в}}$– выборочное среднее, s – исправленная дисперсия, n – объем выборки. Эта случайная величина, возможные значения которой будем обозначать t, имеет распределение Стьюдента (см. лекцию 12) с $k = n-1$ степенями свободы.

Поскольку плотность распределения Стьюдента $s(t,n) = B_n\left(1+\frac{t^2}{n-1}\right)^{\frac{n}{2}}$, где $B_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)}$ явным образом не зависит от a и σ, можно задать вероятность ее попадания в некоторый интервал $(-t_\gamma, t_\gamma)$, учитывая четность плотности распределения, следующим образом:

$$p\left(\left|\frac{\bar{x}_{\text{в}}-a}{\frac{s}{\sqrt{n}}}\right| < t_\gamma\right) = 2\int_0^{t_\gamma} s(t,n)dt = \gamma.$$

Отсюда получаем:

$$p\left(\bar{x}_{\text{в}} - \frac{t_\gamma s}{\sqrt{n}} < a < \bar{x}_{\text{в}} + \frac{t_\gamma s}{\sqrt{n}}\right) = \gamma \qquad (18.3)$$

Таким образом, получен доверительный интервал для a, где t_γ можно найти по соответствующей таблице при заданных n и γ.

Пример. Пусть объем выборки $n = 25$, $\bar{x}_{\text{в}} = 3$, $s = 1{,}5$. Найдем доверительный интервал для a при $\gamma = 0{,}99$. Из таблицы находим, что $t_\gamma\,(n = 25, \gamma = 0{,}99) = 2{,}797$. Тогда $3 - \frac{2{,}797\cdot 1{,}5}{\sqrt{25}} < a < 3 + \frac{2{,}797\cdot 1{,}5}{\sqrt{25}}$, или $2{,}161 < a < 3{,}839$ – доверительный интервал, в который попадает a с вероятностью 0,99.

3. *Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения*

Будем искать для среднего квадратического отклонения нормально распределенной случайной величины доверительный интервал вида $(s-\delta, s+\delta)$, где s – исправленное выборочное среднее квадратическое отклонение, а для δ выполняется условие: $p\,(|\sigma - s| < \delta) = \gamma$.

Запишем это неравенство в виде: $s\left(1-\frac{\delta}{s}\right) < \sigma < s\left(1+\frac{\delta}{s}\right)$ или, обозначив $q = \frac{\delta}{s}$,

$$s(1-q) < \sigma < s(1+q). \qquad (18.4)$$

Рассмотрим случайную величину χ, определяемую по формуле

$$\chi = \frac{s}{\sigma}\sqrt{n-1},$$

которая распределена по закону «хи-квадрат» с $n-1$ степенями свободы (см. лекцию 12). Плотность ее распределения

$$R(\chi,n)=\frac{\chi^{n-2}e^{\frac{\chi^2}{2}}}{2^{\frac{n-1}{2}}\cdot\Gamma\left(\frac{n-1}{2}\right)}$$

не зависит от оцениваемого параметра σ, а зависит только от объема выборки n. Преобразуем неравенство (18.4) так, чтобы оно приняло вид $\chi_1<\chi<\chi_2$. Вероятность выполнения этого неравенства равна доверительной вероятности γ, следовательно, $\int_{\chi_2}^{\chi_1}R(\chi,n)d\,\chi=\gamma$. Предположим, что $q<1$, тогда неравенство (18.4) можно записать так:

$$\frac{1}{s(1+q)}<\frac{1}{\sigma}<\frac{1}{s(1-q)},$$

или, после умножения на $s\sqrt{n-1}$, $\frac{\sqrt{n-1}}{1+q}<\frac{s\sqrt{n-1}}{\sigma}<\frac{\sqrt{n-1}}{1-q}$. Следовательно, $\frac{\sqrt{n-1}}{1+q}<\chi<\frac{\sqrt{n-1}}{1-q}$. Тогда $\int_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}}R(\chi,n)d\,\chi=\gamma$. Существуют таблицы для распределения «хи-квадрат», из которых можно найти q по заданным n и γ, не решая этого уравнения. Таким образом, вычислив по выборке значение s и определив по таблице значение q, можно найти доверительный интервал (18.4), в который значение σ попадает с заданной вероятностью γ.

Замечание. Если $q>1$, то с учетом условия $\sigma>0$ доверительный интервал для σ будет иметь границы

$$0<\sigma<s(1+q). \qquad (18.5)$$

Пример. Пусть $n=20$, $s=1{,}3$. Найдем доверительный интервал для σ при заданной надежности $\gamma=0{,}95$. Из соответствующей таблицы находим $q(n=20,\gamma=0{,}95)=0{,}37$. Следовательно, границы доверительного интервала: $1{,}3(1-0{,}37)=0{,}819$ и $1{,}3(1+0{,}37)=1{,}781$. Итак, $0{,}819<\sigma<1{,}781$ с вероятностью $0{,}95$.

19. СТАТИСТИЧЕСКАЯ ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ. ОБЩИЕ ПРИНЦИПЫ ПРОВЕРКИ ГИПОТЕЗ

Определение 19.1. **Статистической гипотезой** называют гипотезу о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

Определение 19.2. **Нулевой (основной)** называют выдвинутую гипотезу H_0. **Конкурирующей (альтернативной)** называют гипотезу H_1, которая противоречит нулевой.

Пример. Пусть H_0 заключается в том, что математическое ожидание генеральной совокупности $a = 3$. Тогда возможные варианты H_1: а) $a \neq 3$;б) $a > 3$; в) $a < 3$.

*Определение 19.3.***Простой** называют гипотезу, содержащую только одно предположение, **сложной** – гипотезу, состоящую из конечного или бесконечного числа простых гипотез.

Пример. Для показательного распределения гипотеза H_0: $\lambda = 2$ – простая, H_0:$\lambda > 2$ – сложная, состоящая из бесконечного числа простых (вида $\lambda = c$, где c – любое число, большее 2).

В результате проверки правильности выдвинутой нулевой гипотезы (такая проверка называется **статистической**, так как производится с применением методов математической статистики) возможны ошибки двух видов: **ошибка первого рода**, состоящая в том, что будет отвергнута правильная нулевая гипотеза, и **ошибка второго рода**, заключающаяся в том, что будет принята неверная гипотеза.

Замечание. Какая из ошибок является на практике более опасной, зависит от конкретной задачи. Например, если проверяется правильность выбора метода лечения больного, то ошибка первого рода означает отказ от правильной методики, что может замедлить лечение, а ошибка второго рода (применение неправильной методики) чревата ухудшением состояния больного и является более опасной.

Определение 19.4. Вероятность ошибки первого рода называется **уровнем значимости α.**

Основной прием проверки статистических гипотез заключается в том, что по имеющейся выборке вычисляется значение некоторой случайной величины, имеющей известный закон распределения.

Определение 19.5. **Статистическим критерием** называется случайная величина K с известным законом распределения, служащая для проверки нулевой гипотезы.

*Определение 19.6.***Критической областью** называют область значений критерия, при которых нулевую гипотезу отвергают, **областью принятия гипотезы** – область значений критерия, при которых гипотезу принимают.

Итак, процесс проверки гипотезы состоит из следующих этапов:

1) выбирается статистический критерий K;
2) вычисляется его наблюдаемое значение $K_{набл}$по имеющейся выборке;
3) поскольку закон распределения K известен, определяется (по известному уровню значимости α) **критическое значение** $k_{кр}$, разделяющее критическую область и область принятия гипотезы (например, если $p(K > k_{кр}) = \alpha$, то справа от $k_{кр}$ располагается критическая область, а слева – область принятия гипотезы);

4) если вычисленное значение $K_{\text{набл}}$ попадает в область принятия гипотезы, то нулевая гипотеза принимается, если в критическую область – нулевая гипотеза отвергается.

Различают разные виды критических областей:

- **правостороннюю** критическую область, определяемую неравенством $K > k_{\text{кр}}(k_{\text{кр}} > 0)$;
- **левостороннюю** критическую область, определяемую неравенством $K < k_{\text{кр}}(k_{\text{кр}} < 0)$;
- **двустороннюю** критическую область, определяемую неравенствами $K < k_1, K > k_2$ $(k_1 > k_2)$.

Определение 19.7. **Мощностью критерия** называют вероятность попадания критерия в критическую область при условии, что верна конкурирующая гипотеза.

Если обозначить вероятность ошибки второго рода (принятия неправильной нулевой гипотезы) β, то мощность критерия равна $1 - \beta$. Следовательно, чем больше мощность критерия, тем меньше вероятность совершить ошибку второго рода. Поэтому после выбора уровня значимости следует строить критическую область так, чтобы мощность критерия была максимальной.

Критерий для проверки гипотезы о вероятности события

Пусть проведено n независимых испытаний (n – достаточно большое число), в каждом из которых некоторое событие A появляется с одной и той же, но неизвестной вероятностью p, и найдена относительная частота $\frac{m}{n}$ появлений A в этой серии испытаний. Проверим при заданном уровне значимости α нулевую гипотезу H_0, состоящую в том, что вероятность p равна некоторому значению p_0.

Примем в качестве статистического критерия случайную величину

$$U = \frac{\left(\frac{M}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}}, \tag{19.1}$$

имеющую нормальное распределение с параметрами $M(U) = 0, \sigma(U) = 1$ (то есть нормированную). Здесь $q_0 = 1 - p_0$. Вывод о нормальном распределении критерия следует из теоремы Лапласа (при достаточно большом n относительную частоту можно приближенно считать нормально распределенной с математическим ожиданием p и средним квадратическим отклонением $\sqrt{\frac{pq}{n}}$).

Критическая область строится в зависимости от вида конкурирующей гипотезы.

1) Если H_0: $p = p_0$, а H_1: $p \neq p_0$, то критическую область нужно построить так, чтобы вероятность попадания критерия в эту область равнялась заданному уровню значимости α. При этом наибольшая мощность критерия достигается тогда, когда критическая область состоит из двух интервалов, вероятность попадания в каждый из которых равна $\frac{\alpha}{2}$. Поскольку U симметрична относительно оси Oy, вероятность ее попадания в интервалы$(-\infty; 0)$ и $(0; +\infty)$ равна 0,5, следовательно, критическая область тоже должна быть симметрична относительно Oy.Поэтому$u_{\text{кр}}$ определяется по таблице значений функции Лапласа из условия$\Phi\left(u_{\text{кр}}\right)\frac{1-\alpha}{2}$, а критическая область имеет вид$\left(-\infty; -u_{\text{кр}}\right) \cup \left(u_{\text{кр}}; +\infty\right)$.

Замечание. Предполагается, что используется таблица значений функции Лапласа, заданной в виде$\Phi(x) = \int_0^x e^{\frac{t^2}{2}} dt$, где нижний предел интегрирования равен 0, а не $-\infty$. Функция Лапласа, заданная таким образом, является нечетной, а ее значения на 0,5 меньше, чем значения стандартной функции$\Phi(x)$ (см. лекцию 6).

Далее нужно вычислить наблюдаемое значение критерия:

$$U_{\text{набл}} = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}}. \tag{19.2}$$

Если $|U_{\text{набл}}| < u_{\text{кр}}$, то нулевая гипотеза принимается.

Если$|U_{\text{набл}}| > u_{\text{кр}}$, то нулевая гипотеза отвергается.

2) Если конкурирующая гипотеза H_1: $p > p_0$, то критическая область определяется неравенством $U > u_{\text{кр}}$, то есть является правосторонней, причем$p(U > u_{\text{кр}}) = \alpha$. Тогда$p\left(0 < U < u_{\text{кр}}\right) == \frac{1}{2} - \alpha = \left(\frac{1-2\alpha}{2}\right)$.Следовательно, $u_{\text{кр}}$ можно найти по таблице значений функции Лапласа из условия, что $\Phi\left(u_{\text{кр}}\right) = \frac{1-2\alpha}{2}$. Вычислим наблюдаемое значение критерия по формуле (19.2).

Если $U_{\text{набл}} < u_{\text{кр}}$, то нулевая гипотеза принимается.

Если$U_{\text{набл}} > u_{\text{кр}}$, то нулевая гипотеза отвергается.

3) Для конкурирующей гипотезы H_1: $p < p_0$ критическая область является левосторонней и задается неравенством $U_{\text{набл}} < -u_{\text{кр}}$, где $u_{\text{кр}}$ вычисляется так же, как в предыдущем случае.

Если $U_{\text{набл}} > -u_{\text{кр}}$, то нулевая гипотеза принимается.

Если $U_{\text{набл}} < -u_{\text{кр}}$, то нулевая гипотеза отвергается.

Пример. Пусть проведено 50 независимых испытаний, и относительная частота появления события A оказалась равной 0,12.

Проверим при уровне значимости $\alpha = 0{,}01$ нулевую гипотезу H_0: $p = 0{,}1$ при конкурирующей гипотезе H_1: $p > 0{,}1$.

Найдем $U_{\text{набл}} = \frac{(0{,}12-0{,}1)\sqrt{50}}{\sqrt{0{,}1\cdot 0{,}9}} = 0{,}471$. Критическая область является правосторонней, а $u_{\text{кр}}$ находим из равенства $\Phi(u_{\text{кр}}) = \frac{1-2\cdot 0{,}01}{2} = 0{,}49$. Из таблицы значений функции Лапласа определяем $u_{\text{кр}} = 2{,}33$. Итак, $U_{\text{набл}} < u_{\text{кр}}$, и гипотеза о том, что $p = 0{,}1$, принимается.

Критерий для проверки гипотезы о математическом ожидании

Пусть генеральная совокупность X имеет нормальное распределение, и требуется проверить предположение о том, что ее математическое ожидание равно некоторому числу a_0. Рассмотрим две возможности.

1) Известна дисперсия σ^2 генеральной совокупности. Тогда по выборке объема n найдем выборочное среднее $\bar{x}_B$ и проверим нулевую гипотезу H_0: $M(X) = a_0$.

Учитывая, что выборочное среднее $\bar{X}$ является несмещенной оценкой$M(X)$, то есть $M(\bar{X}) = M(X)$, можно записать нулевую гипотезу так: $M(\bar{X}) = a_0$. Для ее проверки выберем критерий

$$U = \frac{\bar{X}-a_0}{\sigma(\bar{X})} = \frac{(\bar{X}-a_0)\sqrt{n}}{\sigma}. \tag{19.3}$$

Это случайная величина, имеющая нормальное распределение, причем, если нулевая гипотеза справедлива, то $M(U) = 0, \sigma(U) = 1$.

Выберем критическую область в зависимости от вида конкурирующей гипотезы:

- если H_1: $M(\bar{X}) \neq a_0$, то $u_{\text{кр}}: \Phi\left(u_{\text{кр}}\right) = \frac{1-2\alpha}{2}$, критическая область двусторонняя, $U_{\text{набл}} = \frac{(\bar{x}-a_0)\sqrt{n}}{\sigma}$, и, если $|U_{\text{набл}}| < u_{\text{кр}}$, то нулевая гипотеза принимается; если $|U_{\text{набл}}| > u_{\text{кр}}$, то нулевая гипотеза отвергается.

- если H_1: $M(\bar{X}) > a_0$, то $u_{\text{кр}}: \Phi\left(u_{\text{кр}}\right) = \frac{1-2\alpha}{2}$, критическая область правосторонняя, и, если $U_{\text{набл}} < u_{\text{кр}}$, то нулевая гипотеза принимается; если $U_{\text{набл}} > u_{\text{кр}}$, то нулевая гипотеза отвергается.

- если H_1: $M(\bar{X}) < a_0$, то $u_{\text{кр}}: \Phi\left(u_{\text{кр}}\right) = \frac{1-2\alpha}{2}$, критическая область левосторонняя, и, если $U_{\text{набл}} > -u_{\text{кр}}$, то нулевая гипотеза принимается; если $U_{\text{набл}} < -u_{\text{кр}}$, то нулевая гипотеза отвергается.

2) Дисперсия генеральной совокупности неизвестна.

В этом случае выберем в качестве критерия случайную величину

$$T = \frac{(\bar{X} - a_0)\sqrt{n}}{S}, \qquad (19.4)$$

где S – исправленное среднее квадратическое отклонение. Такая случайная величина имеет распределение Стьюдента с $k = n - 1$ степенями свободы. Рассмотрим те же, что и в предыдущем случае, конкурирующие гипотезы и соответствующие им критические области. Предварительно вычислим наблюдаемое значение критерия:

$$T_{\text{набл}} = \frac{(\bar{x} - a_0)\sqrt{n}}{S}. \qquad (19.5)$$

- если H_1: $M(\bar{X}) \neq a_0$, то критическая точка $t_{\text{двуст.кр.}}$ находится по таблице критических точек распределения Стьюдента по известнымα и $k = n - 1$.

Если $|T_{\text{набл}}| < t_{\text{двуст.кр.}}$, то нулевая гипотеза принимается.

Если $|T_{\text{набл}}| > t_{\text{двуст.кр.}}$, то нулевая гипотеза отвергается.

- если H_1: $M(\bar{X}) > a_0$, то по соответствующей таблице находят $t_{\text{двуст.кр.}}(\alpha, k)$ – критическую точку правосторонней критической области. Нулевая гипотеза принимается, если

$$T_{\text{набл}} < t_{\text{двуст.кр.}}$$

- при конкурирующей гипотезе H_1: $M(\bar{X}) < a_0$ критическая область является левосторонней, и нулевая гипотеза принимается при условии $T_{\text{набл}} > -t_{\text{двуст.кр.}}$. Если$T_{\text{набл}} < -t_{\text{двуст.кр.}}$, нулевую гипотезу отвергают.

Критерий для проверки гипотезы о сравнении двух дисперсий

Пусть имеются две нормально распределенные генеральные совокупности X и Y. Из них извлечены независимые выборки объемов соответственно n_1 и n_2, по которым вычислены исправленные выборочные дисперсии s_x^2 иs_y^2. Требуется при заданном уровне значимостиα проверить нулевую гипотезу H_0: $D(X) = D(Y)$ о равенстве дисперсий рассматриваемых генеральных совокупностей. Учитывая несмещенность исправленных выборочных дисперсий, можно записать нулевую гипотезу так:

$$H_0: M\,(s_x^2) = M\,(s_y^2). \qquad (19.6)$$

Замечание. Конечно, исправленные дисперсии, вычисленные по выборкам, обычно оказываются различными. При проверке гипотезы выясняется, является ли это различие незначимым и обусловленным случайными причинами (в случае принятия нулевой гипотезы) или оно является следствием того, что сами генеральные дисперсии различны.

В качестве критерия примем случайную величину

$$F = \frac{S_\sigma^2}{S_м^2} \qquad (19.6)$$

– отношение большей выборочной дисперсии к меньшей. Она имеет распределение Фишера-Снедекора со степенями свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$, где n_1 – объем выборки, по которой вычислена большая исправленная дисперсия, а n_2 – объем второй выборки. Рассмотрим два вида конкурирующих гипотез:

- пусть H_1: $D(X) > D(Y)$. Наблюдаемым значением критерия будет отношение большей из исправленных дисперсий к меньшей: $F_{набл} = \frac{S_\sigma^2}{S_м^2}$. По таблице критических точек распределения Фишера-Снедекора можно найти критическую точку $F_{набл}(\alpha;\ k_1;\ k_2)$. При $F_{набл} < F_{кр.}$ нулевая гипотеза принимается, при $F_{набл} > F_{кр.}$ отвергается.

- если H_1: $D(X) \neq D(Y)$, то критическая область является двусторонней и определяется неравенствами $F < F_1,\ F > F_2$, где $p(F < F_1) = p(F > F_2) = \alpha/2$. При этом достаточно найти правую критическую точку $F_2 = F_{кр.}\left(\frac{\alpha}{2}, k_1, k_2\right)$. Тогда при $F_{набл} < F_{кр.}$ нулевая гипотеза принимается, при $F_{набл} > F_{кр.}$ отвергается.

20. КРИТЕРИЙ ПИРСОНА ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ О ВИДЕ ЗАКОНА РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

В предыдущей лекции рассматривались гипотезы, в которых закон распределения генеральной совокупности предполагался известным. Теперь займемся проверкой гипотез о предполагаемом законе неизвестного распределения, то есть будем проверять нулевую гипотезу о том, что генеральная совокупность распределена по некоторому известному закону. Обычно статистические критерии для проверки таких гипотез называются **критериями согласия**.

Критерий Пирсона

Достоинством критерия Пирсона является его универсальность: с его помощью можно проверять гипотезы о различных законах распределения.

1. *Проверка гипотезы о нормальном распределении*

Пусть получена выборка достаточно большого объема n с большим количеством различных значений вариант. Для удобства ее обработки разделим интервал от наименьшего до наибольшего из значений вариант на s равных частей и будем считать, что значения вариант, попавших в каждый интервал, приближенно равны числу, задающему середину интервала. Подсчитав число вариант, попавших в каждый интервал, составим так называемую сгруппированную выборку:

варианты...........$x_1 x_2$ ... x_s

частоты............$n_1 n_2$ ... n_s,

где x_i – значения середин интервалов, а n_i – число вариант, попавших в -й интервал (эмпирические частоты).

По полученным данным можно вычислить выборочное среднее $\bar{x}_в$ и выборочное среднее квадратическое отклонение $\sigma_в$. Проверим предположение, что генеральная совокупность распределена по нормальному закону с параметрами $M(X) = \bar{x}_в, D(X) = \sigma_в^2$. Тогда можно найти количество чисел из выборки объема n, которое должно оказаться в каждом интервале при этом предположении (то есть теоретические частоты). Для этого по таблице значений функции Лапласа найдем вероятность попадания в -й интервал:

$$p_i = \Phi\left(\frac{b_i - \bar{x}_в}{\sigma_в}\right) - \Phi\left(\frac{a_i - \bar{x}_в}{\sigma_в}\right),$$

где a_i и b_i– границы-го интервала. Умножив полученные вероятности на объем выборки n, найдем теоретические частоты: $n_i = np_i$.Наша цель – сравнить эмпирические и теоретические частоты, которые, конечно, отличаются друг от друга, и выяснить, являются ли эти различия несущественными, не опровергающими гипотезу о нормальном распределении исследуемой случайной величины, или они настолько велики, что противоречат этой гипотезе. Для этого используется критерий в виде случайной величины

$$\chi^2 = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}. \qquad (20.1)$$

Смысл ее очевиден: суммируются части, которые квадраты отклонений эмпирических частот от теоретических составляют от соответствующих теоретических частот. Можно доказать, что вне зависимости от реального закона распределения генеральной совокупности закон распределения случайной величины (20.1) при$n \to \infty$ стремится к закону распределения χ^2(см. лекцию 12) с числом степеней свободы $k = s - 1 - r$, гдеr– число параметров предполагаемого распределения, оцененных по данным выборки. Нормальное распределение характеризуется двумя параметрами, поэтому $k = s - 3$. Для выбранного критерия строится правосторонняя критическая область, определяемая условием

$$p\left(\chi^2 > \chi^2_{кр}(\alpha, k)\right) = \alpha, \qquad (20.2)$$

где α– уровень значимости. Следовательно, критическая область задается неравенством $\chi^2 > \chi^2_{\text{кр}}(\alpha, k)$ а область принятия гипотезы –$\chi^2 < \chi^2_{\text{кр}}(\alpha, k)$.

Итак, для проверки нулевой гипотезы H_0: генеральная совокупность распределена нормально – нужно вычислить по выборке наблюдаемое значение критерия:

$$\chi^2_{\text{набл}} = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}\,, \tag{20.1`}$$

а по таблице критических точек распределения χ^2 найти критическую точку$\chi^2_{\text{кр}}(\alpha, k)$, используя известные значения α и $k = s - 3$. Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$ - нулевую гипотезу принимают, при $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$ ее отвергают.

2. *Проверка гипотезы о равномерном распределении*

При использовании критерия Пирсона для проверки гипотезы о равномерном распределении генеральной совокупности с предполагаемой плотностью вероятности

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

необходимо, вычислив по имеющейся выборке значение$\bar{x}_{\text{в}}$, оценить параметры a и b по формулам:

$$a^* = \bar{x}_B - \sqrt{3\sigma_{\text{в}}}, b^* = \bar{x}_B + \sqrt{3\sigma_{\text{в}}}, \tag{20.3}$$

где a^* и b^*– оценки a и b. Действительно, для равномерного распределения $M(X) = \frac{a-b}{2}$,$\sigma(x) = \sqrt{D(X)} = \sqrt{\frac{(a-b)^2}{12}} = \frac{a-b}{2\sqrt{3}}$, откуда можно получить систему для определения a^* и b^*:$\begin{cases} \frac{b^*+a^*}{2} = \bar{x}_{\text{в}}, \\ \frac{b^*-a^*}{2\sqrt{3}} = \sigma_{\text{в.}} \end{cases}$ решением которой являются выражения (20.3).

Затем, предполагая, что $f(x) = \frac{1}{b^* - a^*}$, можно найти теоретические частоты по формулам

$$n_1' = np_1 = nf(x)(x_1 - a^*) = n \cdot \frac{1}{b^* - a^*}(x_1 - a^*);$$
$$n_2' = n_3' = \cdots = n_{s-1}' = n \cdot \frac{1}{b^* - a^*}(x_i - x_{i-1}), i = 1,2, \ldots, s-1;$$

$$n'_s = n \cdot \frac{1}{b^* - a^*}(b^* - x_{s-1}).$$

Здесь s – число интервалов, на которые разбита выборка.

Наблюдаемое значение критерия Пирсона вычисляется по формуле (20.1`), а критическое – по таблице с учетом того, что число степеней свободы $k = s - 3$. После этого границы критической области определяются так же, как и для проверки гипотезы о нормальном распределении.

3. *Проверка гипотезы о показательном распределении*

В этом случае, разбив имеющуюся выборку на равные по длине интервалы, рассмотрим последовательность вариант $x_i^* = \frac{x_i - x_{i+1}}{2}$, равноотстоящих друг от друга (считаем, что все варианты, попавшие в - й интервал, принимают значение, совпадающее с его серединой), и соответствующих им частот n_i(число вариант выборки, попавших в i-й интервал). Вычислим по этим данным $\bar{x}_B$ и примем в качестве оценки параметра λ величину $\lambda^* = \frac{1}{\bar{x}_B}$. Тогда теоретические частоты вычисляются по формуле

$$n'_i = n_i p_i = n_i p(x_i < X < x_{i+1}) = n_i\left(e^{-\lambda x_i} - e^{-\lambda x_{i+1}}\right).$$

Затем сравниваются наблюдаемое и критическое значение критерия Пирсона с учетом того, что число степеней свободы $k = s - 2$.

Критерий Колмогорова

Этот критерий применяется для проверки простой гипотезы H_0 о том, что независимые одинаково распределенные случайные величины $X_1, X_2, \dots, X_n$имеют заданную непрерывную функцию распределения$F(x)$.

Найдем функцию эмпирического распределения$F_n(x)$и будем искать границы двусторонней критической области, определяемой условием

$$D_n = \sup_{|x|<\infty} |F_n(x) - F(x)| > \lambda_n. \qquad (20.3)$$

А.Н.Колмогоров доказал, что в случае справедливости гипотезы H_0 распределение статистики D_n не зависит от функции $F(x)$, и при$n \to \infty$

$$p\left(\sqrt{n} D_n < \lambda\right) \to K(\lambda), \lambda > 0,$$

где

$$K(\lambda) = \sum_{m=-\infty}^{\infty} (-1)^m e^{-2m^2\lambda^2} - \qquad (20.4)$$

– критерий Колмогорова, значения которого можно найти в соответствующих таблицах. Критическое значение критерия$\lambda_n(\alpha)$вычисляется по заданному уровню значимости α как корень уравнения$p(D_n \geq \lambda) = \alpha$.

Можно показать, что приближенное значение вычисляется по формуле

$$\lambda_n(\alpha) \approx \sqrt{\frac{z}{2n}} - \frac{1}{6n},$$

где z – корень уравнения $1 - K\left(\sqrt{\frac{\lambda}{2}}\right) = \alpha$.

На практике для вычисления значения статистики D_n используется то, что

$$D_n = \max(D_n^+, D_n^-),$$

где$D_n^+ = \max\limits_{1 \leq m \leq n}\left(\frac{m}{n} - F\left(X_{(m)}\right)\right), D_n^- = \max\limits_{1 \leq m \leq n}\left(F\left(X_{(m)}\right) - \frac{m-1}{n}\right),$
а $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$– вариационный ряд, построенный по выборке $X_1, X_2, \dots, X_n$.

Можно дать следующее геометрическое истолкование критерия Колмогорова: если изобразить на плоскости Oxy графики функций$F_n(x)$, $F_n(x) \pm \lambda_n(\alpha)$(см. Рисунок 20.1), то гипотеза H_0 верна, если график функции $F(x)$ не выходит за пределы области, лежащей между графиками функций $F_n(x) - \lambda_n(\alpha)$ и $F_n(x) + \lambda_n(\alpha)$.

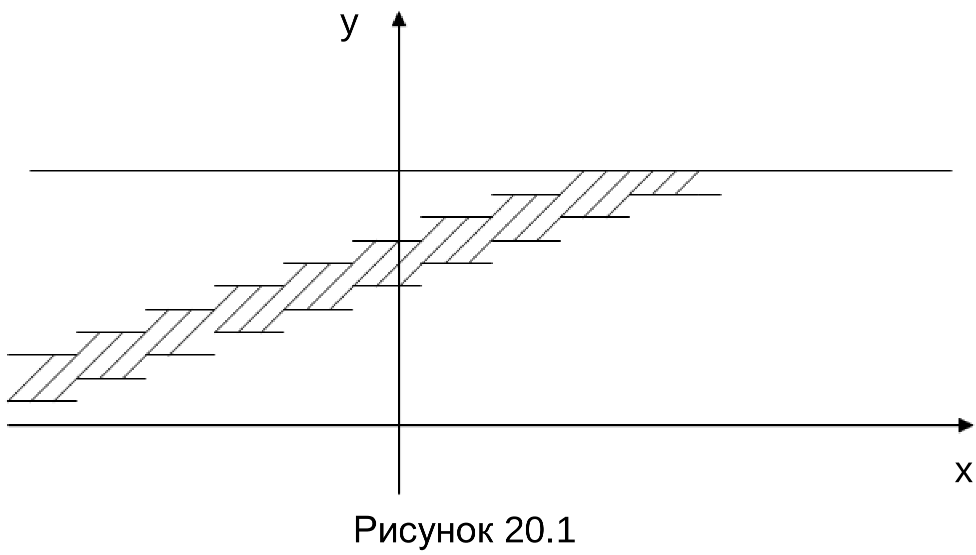

Рисунок 20.1

Приближенный метод проверки нормальности распределения,связанный с оценками коэффициентов асимметрии и эксцесса

Определим по аналогии с соответствующими понятиями для теоретического распределения асимметрию и эксцессэмпирического распределения.

Определение 20.1. **Асимметрия эмпирического распределения** определяется равенством

$$a_s = \frac{m_3}{\sigma_{\text{в}}^3}, \tag{20.5}$$

где m_3 – центральный эмпирический момент третьего порядка.

Эксцесс эмпирического распределения определяется равенством

$$e_k = \frac{m_4}{\sigma_{\text{в}}^4} - 3, \tag{20.6}$$

где m_4 – центральный эмпирический момент четвертого порядка.

Как известно, для нормально распределенной случайной величины асимметрия и эксцесс равны 0. Поэтому, если соответствующие эмпирические величины достаточно малы, можно предположить, что генеральная совокупность распределена по нормальному закону.

21. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Проверка гипотезы о значимости выборочного коэффициента корреляции

Рассмотрим выборку объема n, извлеченную из нормально распределенной двумерной генеральной совокупности(X, Y). Вычислим выборочный коэффициент корреляции r_B. Пусть он оказался не равным нулю. Это еще не означает, что и коэффициент корреляции генеральной совокупности не равен нулю. Поэтому при заданном уровне значимости α возникает необходимость проверки нулевой гипотезыH_0: $r_{\text{г}} = 0$ о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезеH_1:$r_{\text{г}} \neq 0$. Таким образом, при принятии нулевой гипотезы X и Y некоррелированы, то есть не связаны линейной зависимостью, а при отклонении H_0 они коррелированы.

В качестве критерия примем случайную величину

$$T = \frac{r_{\text{в}}\sqrt{n-2}}{\sqrt{1-r_{\text{в}}^2}}, \tag{21.1}$$

которая, при справедливости нулевой гипотезы, имеет распределение Стьюдента (см. лекцию 12) с $k = n - 2$ степенями свободы. Из вида конкурирующей гипотезы следует, что критическая область двусторонняя с границами $\pm t_{\text{кр}}$, где значение $t_{\text{кр}}(\alpha, k)$ находится из таблиц для двусторонней критической области.

Вычислив наблюдаемое значение критерия

$$T_{\text{набл}} = \frac{r_B\sqrt{n-2}}{\sqrt{1-r_{\text{в}}^2}}$$

и сравнив его с $t_{\text{кр}}$, делаем вывод:

- если $|T_{\text{набл}}| < t_{\text{кр}}$ – нулевая гипотеза принимается (корреляции нет);

- если $|T_{\text{набл}}| > t_{\text{кр}}$ – нулевая гипотеза отвергается (корреляция есть).

Ранговая корреляция

Пусть объекты генеральной совокупности обладают двумя качественными признаками (то есть признаками, которые невозможно измерить точно, но которые позволяют сравнивать объекты между собой и располагать их в порядке убывания или возрастания качества). Договоримся для определенности располагать объекты в порядке ухудшения качества.

Пусть выборка объема n содержит независимые объекты, обладающие двумя качественными признаками: A и B. Требуется выяснить степень их связи между собой, то есть установить наличие или отсутствие **ранговой корреляции**.

Расположим объекты выборки в порядке ухудшения качества по признаку A, предполагая, что все они имеют различное качество по обоим признакам. Назовем место, занимаемое в этом ряду некоторым объектом, его **рангом** $\boldsymbol{x_i}$: $x_1 = 1, x_2 = 2, \dots, x_n = n$.

Теперь расположим объекты в порядке ухудшения качества по признаку B, присвоив им ранги y_i, где номер i равен порядковому номеру объекта по признаку A, а само значение ранга равно порядковому номеру объекта по признаку B. Таким образом, получены две последовательности рангов:

по признаку A … $x_1, x_2, \dots, x_n$

по признаку B … $y_1, y_2, \dots, y_n$.

При этом, если, например, $y_3 = 6$, то это означает, что данный объект занимает в ряду по признаку A третье место, а в ряду по признаку B – шестое.

Сравним полученные последовательности рангов.

1. Если $x_i = y_i$ при всех значениях i, то ухудшение качества по признаку A влечет за собой ухудшение качества по признаку B, то есть имеется «полная ранговая зависимость».

2. Если ранги противоположны, то есть $x_1 = 1, y_1 = n$; $x_2 = 2, y_2 = n - 1; \ldots, x_n = n, y_n = 1$, то признаки тоже связаны: ухудшение качества по одному из них приводит к улучшению качества по другому («противоположная зависимость»).

3. На практике чаще всего встречается промежуточный случай, когда ряд y_i не монотонен. Для оценки связи между признаками будем считать ранги $x_1, x_2, \ldots, x_n$ возможными значениями случайной величины X, а $y_1, y_2, \ldots, y_n$ – возможными значениями случайной величины Y. Теперь можно исследовать связь междуX и Y, вычислив для них выборочный коэффициент корреляции

$$r_{\text{в}} = \frac{\sum n_{uv}uv - n\bar{u}\bar{v}}{n\sigma_u\sigma_v}, \tag{21.2}$$

где $u_i = x_i - \bar{x}, v_i = y_i - \bar{y}$(условные варианты). Поскольку каждому рангу x_iсоответствует только одно значениеy_i, то частота любой пары условных вариант с одинаковыми индексами равна 1, а с разными индексами – нулю. Кроме того, из выбора условных вариант следует, что $\bar{u} = \bar{v} = 0$, поэтому формула (21.2) приобретает более простой вид:

$$r_{\text{в}} = \frac{\sum u_i v_i}{n\sigma_u\sigma_v}. \tag{21.3}$$

Итак, требуется найти $\sum u_i\, v_i, \sigma_u$иσ_v.

Можно показать, что$\sum u_i^2 = \sum v_i^2 = \frac{n^3-n}{12}$. Учитывая, что$\bar{x} = \bar{y}$, можно выразить $\sum u_i v_i$через разности рангов $d_i = x_i - y_i = u_i - v_i$. После преобразований получим: $\sum u_i v_i = \frac{n^3-n}{12} - \sum \frac{d_i^2}{2}$, $\sigma_u = \sigma_v = \sqrt{\frac{n^2-1}{12}}$, откуда $n\sigma_u\sigma_v = \frac{n^3-n}{12}$. Подставив эти результаты в (21.3), получим **выборочный коэффициент ранговой корреляции Спирмена**:

$$\rho_{\text{в}} = 1 - \frac{6\sum d_i^2}{n^3-n}. \tag{21.4}$$

Свойства выборочного коэффициента корреляции Спирмена

1. Если междуA и B имеется «полная прямая зависимость», то есть ранги совпадают при всех i, то $\rho_{\text{в}} = 1$. Действительно, при этом $d_i = 0$, и из формулы (21.4) следует справедливость свойства 1.

$2.$ Если междуA и Bимеется «противоположная зависимость», то $\rho_B = -1$. В этом случае, преобразуя $d_i = (2i - 1) - n$, найдем, что $\sum d_i^2 = \frac{n^3-n}{3}$, тогда из (21.4) $\rho_\text{в} = 1 - \frac{6(n^3-n)}{3(n^3-n)} = 1 - 2 = -1$.

$3.$ В остальных случаях $-1 < \rho_\text{в} < 1$, причем зависимость между A и Bтем меньше, чем ближе $|\rho_\text{в}|$ к нулю.

Итак, требуется при заданном уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента ранговой корреляции Спирмена $\rho_\text{г}$ при конкурирующей гипотезе H_1: $\rho_\text{г} \neq 0$. Для этого найдем критическую точку:

$$T_\text{кр} = t_\text{кр}\,(\alpha, k)\sqrt{\frac{1-r_\text{в}^2}{n-2}}, \tag{21.5}$$

где n – объем выборки, $\rho_\text{в}$ – выборочный коэффициент ранговой корреляции Спирмена,$t_\text{кр}\,(\alpha, k)$ – критическая точка двусторонней критической области, найденная по таблице критических точек распределения Стьюдента, число степеней свободы $k = n - 2$.

Тогда, если $|\,\rho_\text{в}\,| < T_\text{кр}$, то нулевая гипотеза принимается, то есть ранговая корреляционная связь между признаками незначима.

Если $|\,\rho_\text{в}\,| > T_\text{кр}$, то нулевая гипотеза отвергается, и между признаками существует значимая ранговая корреляционная связь.

Можно использовать и другой коэффициент – коэффициент ранговой корреляции Кендалла. Рассмотрим ряд рангов $y_1, y_2, \ldots, y_n$, введенный так же, как и ранее, и зададим величины R_i следующим образом: пусть правее y_1 имеется R_1 рангов, больших y_1; правее y_2 – R_2 рангов, больших y_2 и т.д. Тогда, если обозначить $R = R_1 + R_2 + \cdots + R_{n-1}$, то **выборочный коэффициент ранговой корреляции Кендалла** определяется формулой

$$\tau_\text{в} = \frac{4R}{n(n-1)} - 1, \tag{21.6}$$

где n – объем выборки.

Замечание. Легко убедиться, что коэффициент Кендалла обладает теми же свойствами, что и коэффициент Спирмена.

Для проверки нулевой гипотезы H_0: $\tau_\text{г} = 0$ (генеральный коэффициент ранговой корреляции Кендалла равен нулю) при альтернативной гипотезе H_1: $\tau_\text{г} \neq 0$ необходимо найти критическую точку:

$$T_\text{кр} = z_\text{кр}\sqrt{\frac{2(2n+5)}{9n(n-1)}}, \tag{21.7}$$

где n – объем выборки, а $z_{\text{кр}}$ – критическая точка двусторонней критической области, определяемая из условия $\Phi(z_{\text{кр}}) = \frac{1-\alpha}{2}$ по таблицам для функции Лапласа.

Если $|\tau_B| < T_{\text{кр}}$, то нулевая гипотеза принимается (ранговая корреляционная связь между признаками незначима).

Если $|\tau_B| > T_{\text{кр}}$, то нулевая гипотеза отвергается (между признаками существует значимая ранговая корреляционная связь).

22. РЕГРЕССИОННЫЙ АНАЛИЗ

Рассмотрим выборку двумерной случайной величины (X, Y). Примем в качестве оценок условных математических ожиданий компонент их условные средние значения, а именно: **условным средним** $\bar{y}_x$назовем среднее арифметическое наблюдавшихся значений Y, соответствующих $X = x$. Аналогично **условное среднее**$\bar{x}_y$–среднее арифметическое наблюдавшихся значений X, соответствующих $Y = y$. В лекции 11 были выведены уравнения регрессии YнаXи XнаY:

$$M(Y/x) = f(x), M(X/y) = \varphi(y).$$

Условные средние $\bar{y}_x$и $\bar{x}_y$являются оценками условных математических ожиданий и, следовательно, тоже функциями от *x* и *y*, то есть

$$\bar{y}_x = f^*(x) \text{–} \qquad (22.1)$$

–**выборочное уравнение регрессии $\boldsymbol{Y}$на$\boldsymbol{X}$,**

$$\bar{x}_y = \varphi^*(xy) \text{–} \qquad (22.2)$$

–**выборочное уравнение регрессии $\boldsymbol{X}$на$\boldsymbol{Y}$**.

Соответственно функции $f^*(x)$ и $\varphi^*(xy)$ называются **выборочной регрессией $\boldsymbol{Y}$на$\boldsymbol{X}$ и $\boldsymbol{X}$на$\boldsymbol{Y}$**, а их графики – **выборочными линиями регрессии**. Выясним, как определять параметры выборочных уравнений регрессии, если сам вид этих уравнений известен.

Пусть изучается двумерная случайная величина (X, Y), и получена выборка из n пар чисел $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Будем искать параметры прямой линии среднеквадратической регрессии YнаX вида

$$Y = \rho_{yx} x + b, \qquad (22.3)$$

Подбирая параметры ρ_{yx} и b так, чтобы точки на плоскости с координатами $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ лежали как можно ближе к прямой (22.3). Используем для этого метод наименьших квадратов и найдем минимум функции

$$F(\rho, b) = \sum_{i=1}^{n}(Y_i - y_i)^2 = \sum_{i=1}^{n}(\rho x_i + b - y_i)^2. \quad (22.4)$$

Приравняем нулю соответствующие частные производные:

$$\frac{\partial F}{\partial \rho} = 2\sum_{i=1}^{n}(\rho x_i + b - y_i)x_i = 0.$$
$$\frac{\partial F}{\partial b} = 2\sum_{i=1}^{n}(\rho x_i + b - y_i) = 0.$$

В результате получим систему двух линейных уравнений относительно ρ и b:

$$\begin{cases} (\sum x^2)\rho + (\sum x)b = \sum xy, \\ (\sum x)\rho + nb = \sum y. \end{cases} \quad (22.5)$$

Ее решение позволяет найти искомые параметры в виде:

$$\rho_{xy} = \frac{n\sum xy - \sum x \cdot \sum y}{n\sum x^2 - (\sum x)^2}; b = \frac{\sum x^2 \cdot \sum y - \sum x \cdot \sum xy}{n\sum x^2 - (\sum x)^2}. \quad (22.6)$$

При этом предполагалось, что все значения X и Y наблюдались по одному разу.

Теперь рассмотрим случай, когда имеется достаточно большая выборка (не менее 50 значений), и данные сгруппированы в виде *корреляционной таблицы*:

Y	X				
	x_1	x_2	…	x_k	n_y
y_1	n_{11}	n_{21}	…	n_{k1}	$n_{11} + n_{21} + \cdots + n_{k1}$
y_2	n_{12}	n_{22}	…	n_{k2}	$n_{12} + n_{22} + \cdots + n_{k2}$
…	…	…	…	…	…………..
y_m	n_{1m}	n_{2m}	…	n_{km}	$n_{1m} + n_{2m} + \cdots + n_{km}$
n_x	$n_{11} + n_{12} + \cdots + n_{1m}$	$n_{21} + n_{22} + \cdots + n_{2m}$	…	$n_{k1} + n_{k2} + \cdots + n_{km}$	$n = \sum n_x = \sum n_y$

Здесь n_{ij} – число появлений в выборке пары чисел (x_i, y_j).

Поскольку $\bar{x} = \frac{\sum x}{n}, \bar{y} = \frac{\sum y}{n}, \overline{x^2} = \frac{\sum x^2}{n}$, заменим в системе (22.5) $\sum x = n\bar{x}$.

$$\sum y = n\bar{y}, \sum x^2 = n\overline{x^2}, \sum xy = n_{xy}xy,$$

где n_{xy} – число появлений пары чисел (x, y). Тогда система (22.5) примет вид:

$$\begin{cases} (n\overline{x^2})\rho_{xy} + (n\bar{x})b = \sum n_{xy}xy \\ (\bar{x})\rho_{xy} + b = \bar{y} \end{cases}. \qquad (22.7)$$

Можно решить эту систему и найти параметры ρ_{xy} и b, определяющие выборочное уравнение прямой линии регрессии:

$$\bar{y}_x = \rho_{yx}\bar{x} + b.$$

Но чаще уравнение регрессии записывают в ином виде, вводя **выборочный коэффициент корреляции**. Выразим b из второго уравнения системы (22.7):

$$b = \bar{y} - \rho_{yx}\bar{x}.$$

Подставим это выражение в уравнение регрессии: $\bar{y}_x - \bar{y} = \rho_{yx}(x - \bar{x})$. Из (22.7)

$$\rho_{yx} = \frac{\sum n_{xy}xy - n\bar{x}\bar{y}}{n\left(\overline{x^2} - (\bar{x})^2\right)} = \frac{\sum n_{xy}xy - n\bar{x}\bar{y}}{n\tilde{\sigma}_x^2},, \qquad (22.8)$$

где $\tilde{\sigma}_x^2 = \overline{x^2} - (\bar{x})^2$. Введем понятие **выборочного коэффициента корреляции**

$$r_B = \frac{\sum n_{xy}xy - n\bar{x}\bar{y}}{n\tilde{\sigma}_x\tilde{\sigma}_y},$$

и умножим равенство (22.8) на $\frac{\tilde{\sigma}_x}{\tilde{\sigma}_y}$: $\rho_{yx}\frac{\tilde{\sigma}_x}{\tilde{\sigma}_y} = r_B$, откуда $\rho_{yx} = r_{\text{В}}\frac{\tilde{\sigma}_y}{\tilde{\sigma}_x}$. Используя это соотношение, получим выборочное уравнение прямой линии регрессии Y на X вида

$$\bar{y}_x - \bar{y} = r_B\frac{\tilde{\sigma}_y}{\tilde{\sigma}_x}(x - \bar{x}). \qquad (22.9)$$

23. ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Пусть генеральные совокупности $X_1, X_2, \dots, X_p$ распределены нормально и имеют одинаковую дисперсию, значение которой неизвестно. Найдем выборочные средние по выборкам из этих генеральных совокупностей и проверим при заданном уровне значимости нулевую гипотезу H_0: $M(X_1) = M(X_2) = \dots = M(X_p)$ о равенстве всех математических ожиданий. Для решения этой задачи применяется метод, основанный на сравнении дисперсий и названный поэтому **дисперсионным анализом**.

Будем считать, что на случайную величину X воздействует некоторый качественный фактор F, имеющий p уровней: $F_1, F_2, \dots, F_p$.Требуется сравнить «факторную дисперсию», то есть рассеяние, порождаемое изменением уровня фактора, и «остаточную дисперсию», обусловленную случайными причинами. Если их различие значимо, то фактор существенно влияет наX и при изменении его уровня групповые средние различаются значимо.

Будем считать, что количество наблюдений на каждом уровне фактора одинаково и равно q. Оформим результаты наблюдений в виде таблицы:

Номер испытания	Уровни фактора F_{j}			
	F_1	F_2	…	F_p
1 2 … q	x_{11} x_{21} … x_{q1}	x_{12} x_{22} … x_{q2}	… … … …	x_{1p} x_{2p} … x_{qp}
Групповое среднее	$\bar{x}_{\text{гр}1}$	$\bar{x}_{\text{гр}2}$	…	$\bar{x}_{\text{гр}p}$

Определим общую, факторную и остаточную суммы квадратов отклонений от среднего:

$$S_{\text{общ}} = \sum_{\mathrm{j}=1}^{p} \sum_{i=1}^{q} \left(x_{i\mathrm{j}} - \bar{x}\right)^2 - \qquad (23.1)$$

– общая сумма квадратов отклонений наблюдаемых значений от общего среднего$\bar{x}$;

$$S_{\text{факт}} = q \sum_{\mathrm{j}=1}^{p} \left(\bar{x}_{\text{гр}\mathrm{j}} - \bar{x}\right)^2 - \qquad (23.2)$$

– факторная сумма отклонений групповых средних от общей средней, характеризующая рассеяние между группами;

$$S_{\text{ост}} = \sum_{i=1}^{q}\left(x_{i1} - \bar{x}_{\text{гр}1}\right)^2 + \sum_{i=1}^{q}\left(x_{i2} - \bar{x}_{\text{гр}2}\right)^2 + \ldots + \sum_{i=1}^{q}\left(x_{ip} - \bar{x}_{\text{гр}p}\right)^2 \quad (23.3)$$

– остаточная сумма квадратов отклонений наблюдаемых значений группы от своего группового среднего, характеризующая рассеяние внутри групп.

Замечание. Остаточную сумму можно найти из равенства

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}}.$$

Вводя обозначения$R_{\text{j}} = \sum_{i=1}^{q} x_{i\text{j}}, P_{\text{j}} = \sum_{i=1}^{q} x_{i\text{j}}^2$, получим формулы, более удобные для расчетов:

$$S_{\text{общ}} = \sum_{\text{j}=1}^{p} R_{\text{j}} - \frac{\left(\sum_{\text{j}=1}^{p} R_{\text{j}}\right)^2}{pq}, \quad (23.1`)$$

$$S_{\text{факт}} = \sum_{\text{j}=1}^{p} R_{\text{j}}^2 - \frac{\left(\sum_{\text{j}=1}^{p} R_{\text{j}}\right)^2}{pq}. \quad (23.2`)$$

Разделив суммы квадратов на соответствующее число степеней свободы, получим общую, факторную и остаточную дисперсии:

$$s_{\text{общ}}^2 = \frac{S_{\text{общ}}}{pq-1}, s_{\text{факт}}^2 = \frac{S_{\text{факт}}}{p-1}, s_{\text{ост}}^2 = \frac{S_{\text{ост}}}{p(q-1)}, \quad (23.4)$$

Если справедлива гипотеза H_0, то все эти дисперсии являются несмещенными оценками генеральной дисперсии. Покажем, что проверка нулевой гипотезы сводится к сравнению факторной и остаточной дисперсии по критерию Фишера-Снедекора (см. лекцию 12).

1. Пусть гипотеза H_0 правильна. Тогда факторная и остаточная дисперсии являются несмещенными оценками неизвестной генеральной дисперсии и, следовательно, различаются незначимо. Поэтому результат оценки по критерию Фишера-СнедекораF покажет, что нулевая гипотеза принимается. Таким образом, если верна гипотеза о равенстве математических ожиданийгенеральных совокупностей, то верна и гипотеза о равенстве факторной и остаточной дисперсий.

2. Если нулевая гипотеза неверна, то с возрастанием расхождения между математическими ожиданиями увеличивается и факторная дисперсия, а вместе с ней и отношение$F_{\text{набл}} = \frac{s_{\text{факт}}^2}{s_{\text{ост}}^2}$. Поэтому в результате $F_{\text{набл}}$ окажется больше $F_{\text{кр}}$, и гипотеза о равенстве дисперсий будет отвергнута. Следовательно, если гипотеза о равенстве математических ожиданийгенеральных совокупностей ложна, то ложна и гипотеза о равенстве факторной и остаточной дисперсий.

Итак, метод дисперсионного анализа состоит в *проверке по критерию F нулевой гипотезы о равенстве факторной и остаточной дисперсий.*

Замечание. Если факторная дисперсия окажется меньше остаточной, то гипотеза о равенстве математических ожиданий генеральных совокупностей верна. При этом нет необходимости использовать критерий F.

Если число испытаний на разных уровнях различно (q_1испытаний на уровне F_1, q_2– на уровне $F_2, \ldots, q_p$–на уровнеF_p), то

$$S_{\text{общ}} = \left(P_1 + P_2 + \cdots + P_p\right) - \left(R_1 + R_2 + \cdots + R_p\right),$$

где $P_{\text{j}} = \sum_{\text{j}=1}^{q_{\text{j}}} x_{i\text{j}}^2$ – сумма квадратов наблюдавшихся значений признака на уровне F_j;$R_{\text{j}} = \sum_{i=1}^{q_{\text{j}}} x_{i\text{j}}$ – сумма наблюдавшихся значений признака на уровне F_j. При этом объем выборки, или общее число испытаний, равен$n = q_1 + q_2 + \cdots + q_p$.

Факторная сумма квадратов отклонений вычисляется по формуле

$$S_{\text{факт}} = \left(\frac{R_1^2}{q_1} + \frac{R_2^2}{q_2} + \cdots + \frac{R_p^2}{q_p}\right) - \frac{\left(R_1 + R_2 + \cdots + R_p\right)^2}{n}.$$

Остальные вычисления проводятся так же, как в случае одинакового числа испытаний:

$$S_{\text{ост}} = S_{\text{общ}} - S_{\text{факт}}, s_{\text{факт}}^2 = \frac{S_{\text{факт}}}{p-1}, s_{\text{ост}}^2 = \frac{S_{\text{ост}}}{n-p}.$$

24. МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН МЕТОДОМ МОНТЕ-КАРЛО (СТАТИСТИЧЕСКИХ ИСПЫТАНИЙ)

Задачу, для решения которой применяется метод Монте-Карло, можно сформулировать так: требуется найти значение a изучаемой случайной величины. Для его определения выбирается случайная величина X, математическое ожидание которой равно a, и для выборки из n значений X, полученных в n испытаниях, вычисляется выборочное среднее:

$$\overline{x} = \frac{\sum x_i}{n},$$

которое принимается в качестве оценки искомого числа a:

$$a \approx a^* = \overline{x}.$$

Этот метод требует проведения большого числа испытаний, поэтому его иначе называют **методом статистических испытаний**. Теория метода Монте-Карло исследует, как наиболее целесообразно выбрать случайную величину X, как найти ее возможные значения, как уменьшить дисперсию используемых случайных величин, чтобы погрешность при замене a на a^*была возможно меньшей.

Поиск возможных значений X называют **разыгрываниемслучайной величины**. Рассмотрим некоторые способы разыгрывания случайных величин и выясним, как оценить допускаемую при этом ошибку.

Оценка погрешности метода Монте-Карло

Если поставить задачу определения верхней границы допускаемой ошибки с заданной доверительной вероятностью γ, то есть поиска числа δ, для которого

$$p(|\bar{X} - a| \leq \delta) = \gamma,$$

то получим известную задачу определения доверительного интервала для математического ожидания генеральной совокупности (см. лекцию 18). Воспользуемся результатами решения этой задачи для следующих случаев:

1) случайная величины X распределена нормально и известно ее среднее квадратическое отклонение. Тогда из формулы (18.1) получаем: $\delta = \frac{t\delta}{\sqrt{n}}$, где n – число испытаний, δ– известное среднее квадратическое отклонение, а t – аргумент функции Лапласа, при котором $\Phi(t) = \gamma/2$.

2) Случайная величина X распределена нормально с неизвестным σ. Воспользуемся формулой (18.3), из которой следует, что $\delta = \frac{t_\gamma s}{\sqrt{n}}$, где s – исправленное выборочное среднее квадратическое отклонение, а t_γопределяется по соответствующей таблице.

3) Если случайная величина распределена по иному закону, то при достаточно большом количестве испытаний $(n > 30)$ можно использовать для оценки δ предыдущие формулы, так как при$n\to\infty$ распределение Стьюдента стремится к нормальному, и границы интервалов, полученные по формулам (18.1) и (18.3), различаются незначительно.

Разыгрывание случайных величин

Определение 24.1. **Случайными числами** называют возможные значения r непрерывной случайной величины R, распределенной равномерно в интервале (0; 1).

1. Разыгрывание дискретной случайной величины

Пусть требуется разыграть дискретную случайную величину X, то есть получить последовательность ее возможных значений, зная закон распределения X:

$$X\ x_1 x_2\ \dots\ x_n$$
$$p\ p_1 p_2\ \dots\ p_n.$$

Рассмотрим равномерно распределенную в $(0,1)$ случайную величину R и разобьем интервал $(0,1)$ точками с координатами$p_1, p_1 + p_2, \dots, p_1 + p_2 + \dots + p_{n-1}$на$n$частичных интервалов $\Delta_1, \Delta_2, \dots, \Delta_n$, длины которых равны вероятностям с теми же индексами.

Теорема 24.1. Если каждому случайному числу $r_{\mathrm{j}}(0 \le r_{\mathrm{j}} < 1)$, которое попало в интервалΔ_i, ставить в соответствие возможное значениеx_i, то разыгрываемая величина будет иметь заданный закон распределения:

$$X\ x_1 x_2\ \dots\ x_n$$
$$p\ p_1 p_2\ \dots\ p_n.$$

Доказательство.

Возможные значения полученной случайной величины совпадают с множеством$x_1, x_2, \dots, x_n$, так как число интервалов равно n, а при попадании r_{j}в интервал Δ_i случайная величина может принимать только одно из значений $x_1, x_2, \dots, x_n$.

Так как R распределена равномерно, то вероятность ее попадания в каждый интервал равна его длине, откуда следует, что каждому значению x_i соответствует вероятность p_i. Таким образом, разыгрываемая случайная величина имеет заданный закон распределения.

Пример. Разыграть 10 значений дискретной случайной величины X, закон распределения которой имеет вид:

X2368

p0,10,30,50,1

Решение. Разобьем интервал (0, 1) на частичные интервалы: Δ_1- (0; 0,1), Δ_2 – (0,1; 0,4), Δ_3 -(0,4; 0,9), Δ_4 – (0,9; 1). Выпишем из таблицы случайных чисел 10 чисел: 0,09; 0,73; 0,25; 0,33; 0,76; 0,52; 0,01; 0,35; 0,86; 0,34. Первое и седьмое числа лежат на интервале Δ_1, следовательно, в этих случаях разыгрываемая случайная величина приняла значение x_1 = 2; третье, четвертое, восьмое и десятое числа попали в интервал Δ_2, что соответствует x_2 = 3; второе, пятое, шестое и девятое числа оказались в интервале Δ_3 – при этом X = x_3= 6; на последний интервал не попало ни одного числа. Итак, разыгранные возможные значения Xтаковы: 2, 6, 3, 3, 6, 6, 2, 3, 6, 3.

2. *Разыгрывание противоположных событий*

Пусть требуется разыграть испытания, в каждом из которых событие A появляется с известной вероятностью p. Рассмотрим дискретную случайную величину X, принимающую значения 1 (в случае, если событие A произошло) с вероятностью p и 0 (если A не произошло) с вероятностью $q = 1 - p$. Затем разыграем эту случайную величину так, как было предложено в предыдущем пункте.

Пример. Разыграть 10 испытаний, в каждом из которых событие A появляется с вероятностью 0,3.

Решение. Для случайной величины X с законом распределения

X10

p0,30,7

получим интервалыΔ_1 – (0; 0,3) и Δ_2 – (0,3; 1). Используем ту же выборку случайных чисел, что и в предыдущем примере, для которой в интервал Δ_1 попадают числа №№1,3 и 7, а остальные – в интервал Δ_2. Следовательно, можно считать, что событие A произошло в первом, третьем и седьмом испытаниях, а в остальных – не произошло.

3. *Разыгрывание полной группы событий*

Если события $A_1, A_2, \dots, A_n$, вероятности которых равны $p_1, p_2, \dots, p_n$, образуют полную группу, то дляиз разыгрывания (то есть моделирования последовательности их появлений в серии испытаний) можно разыграть дискретную случайную величину Xс законом распределенияX $1\ 2\ \dots\ n$, сделав это так же, как в пункте 1. При этом считаем, что

$$p\ p_1 p_2 \dots p_n,$$

если X принимает значение $x_i = i$, то в данном испытании произошло событие A_i.

4. *Разыгрывание непрерывной случайной величины*

а) Метод обратных функций.

Пусть требуется разыграть непрерывную случайную величину X, то есть получить последовательность ее возможных значений $x_i (i = 1, 2, \dots, n)$, зная функцию распределения$F(x)$.

Теорема 24.2. Если r_i– случайное число, то возможное значение x_i разыгрываемой непрерывной случайной величины X с заданной функцией распределения$F(x)$, соответствующее r_i, является корнем уравнения

$$F(x_i) = r_i. \tag{24.1}$$

Доказательство.

Так как $F(x)$ монотонно возрастает в интервале от 0 до 1, то найдется (причем единственное) значение аргумента x_i, при котором функция распределения примет значение r_i. Значит, уравнение (24.1) имеет единственное решение: $x_i = F^{-1}(r_i)$, где F^{-1} – функция, обратная к F. Докажем, что корень уравнения (24.1) является возможным значением рассматриваемой случайной величины X. Предположим вначале, что x_i– возможное значение некоторой случайной величины ξ, и докажем, что вероятность попадания ξ в интервал(c, d)равна$F(d) - F(c)$. Действительно, $c < x_i < d \Leftrightarrow F(c) < r_i < F(d)$в силу монотонности $F(x)$ и того, что $F(x_i) = r_i$. Тогда

$$c < x_i < d \Leftrightarrow F(c) < R < F(d),$$

следовательно,

$$p(c < \xi < d) = p(F(c) < R < F(d)) = F(d) - F(c).$$

Значит, вероятность попадания ξ в интервал(c, d)равна приращению функции распределения $F(x)$ на этом интервале, следовательно, $\xi = X$.

*Пример.*Разыграть 3 возможных значения непрерывной случайной величины X, распределенной равномерно в интервале (5; 8).

Решение.

$F(x) = \frac{x-5}{3}$, то есть требуется решить уравнение$\frac{x_i-5}{3} = r_i, x_i = 3r_i + 5$. Выберем 3 случайных числа: 0,23; 0,09 и 0,56 и подставим их в это уравнение. Получим соответствующие возможные значения X:$x_1 = 5{,}69; x_2 = 5{,}27; x_3 = 6{,}68$.

б) Метод суперпозиции.

Если функция распределения разыгрываемой случайной величины может быть представлена в виде линейной комбинации двух функций распределения:

$$F(x) = C_1F_1(x) + C_2F_2(x)\big(C_{1,2}\big) > 0, \qquad (24.2)$$

то$C_1 + C_2 = 1$, так как при $x \to \infty$ $F(x) \to 1$.

Введем вспомогательную дискретную случайную величину Z с законом распределения

Z12. Выберем 2 независимых случайных числа r_1и r_2 и разыграем возможное

p C_1 C_2

значение Zпо числуr_1 (см. пункт 1). Если $Z = 1$, то ищем искомое возможное значение Xиз уравнения$F_1(x) = r_2$, а если $Z = 2$, то решаем уравнение$F_2(x) = r_2$.

Можно доказать, что при этом функция распределения разыгрываемой случайной величины равна заданной функции распределения.

в) Приближенное разыгрывание нормальной случайной величины.

Так как для R, равномерно распределенной в (0, 1),$\boldsymbol{M(R)} = \frac{\mathbf{1}}{\mathbf{2}}, \boldsymbol{D(R)} = \frac{\mathbf{1}}{\mathbf{12}}$, то для суммы n независимых, равномерно распределенных в интервале (0,1) случайных величин$\sum_{\mathrm{j}=1}^{n} R_{\mathrm{j}}$ $M\left(\sum_{\mathrm{j}=1}^{n} R_{\mathrm{j}}\right) = \frac{n}{2}, D\left(\sum_{\mathrm{j}=1}^{n} R_{\mathrm{j}}\right) = \frac{n}{12}, \sigma = \sqrt{\frac{n}{12}}$. Тогда в силу центральной предельной теоремы нормированная случайная величина $\frac{\sum_{\mathrm{j}=1}^{n} R_{\mathrm{j}} - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$при $n \to \infty$ будет иметь распределение, близкое к нормальному, с параметрами $a = 0$ и $\sigma = 1$. В частности, достаточно хорошее приближение получается при $n = 12$:

$$\sum_{\mathrm{j}=1}^{12} R_{\mathrm{j}} - 6.$$

Итак, чтобы разыграть возможное значение нормированной нормальной случайной величины x, надо сложить 12 независимых случайных чисел и из суммы вычесть 6.

ЗАДАНИЯ ДЛЯ СРС ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

1. Из пяти карточек с буквамиА, Б, В, Г, Днаугад одна за другой выбираются три и располагаются в ряд в порядке появления. Какова вероятность, что получится слово «два»?

2. а) Три одноклассника Иванов, Петров и Сидоров решили подать документы на экономический факультет одного из четырех вузов: БГУ, БНТУ, БГАТУ и БГУИР, причем каждый выбирал себе вуз случайно и независимо от других. Найти вероятности следующих событий:

1) всех одноклассники окажутся в разных вузах;
2) все подадут документы в один и тот же вуз;
3) все подадут документы в БГУ.

б) Студенты из общежития закупят партию из 10 арбузов в том случае, если при нарезке двух из них, выбранных случайным образом, оба окажутся зрелыми. Какова вероятность того, что студенты купят арбузы, среди которых будет 4 незрелых?

в) На одной полкенаугадрасставляютсяn различных книг. Найти вероятность того, что две определенные книги окажутся поставленными рядом (в любом порядке). Задачу решить в общем виде и вычислить конкретный ответ для$n = 2, n = 3, n = 8$.

3. Имеется пять отрезков, длины которых равны соответственно 1, 3, 5, 7 и 9 единицам. Определить вероятность того, что с помощью взятых наудачу трех отрезков из данных пяти можно построить треугольник.

4. Из восьми магазинов с номерами 1,2, …, 8 для проверки выбирают три. Какова вероятность того, что будут проверяться магазины № 5 и № 6?

5. Имеется 6 карточек с буквами А, А, Т, Т, Л, Н. Карточки перемешиваюти затем наугад достают по очереди и располагают в ряд в порядке появления. Какова вероятность, что получится слово «АТЛАНТ»?

6. На пяти карточках написаны цифры от 1 до 5. Опыт состоит в случайном выборе трех карточек и раскладывании их в порядке поступления в ряд слева направо. Найти вероятность следующих событий: A = {появится число 123}; B = {появится число, не содержащее цифры 2}; C = {появится число, состоящее из последовательных цифр}.

7. Десять человек входят в комнату, где имеется всего 7 стульев, и рассаживаются случайным образом, но так, что все стулья оказываются занятыми. Какова вероятность того, что а) два определенных лица окажутся без места? б)4 определенных лица будут сидеть?

8. Фирмы A_1, A_2, A_3, A_4, A_5 предлагают свои условия по выполнению 3различных контрактов C_1, C_2, C_3. Любая фирма может получить только один контракт. Если предположить равновозможность заключения контрактов, чему равна вероятность того, что фирма A_3 получит контракт? Чему равна вероятность того, что фирмы A_1 и A_2 получат контракт?

9. 8 вариантов контрольной работы, написанные каждый на отдельной карточке, перемешиваются и распределяются случайным образом среди шести студентов, сидящих в одном ряду, причем каждый получает по одному варианту. Найти вероятность следующих событий: A = «варианты с номерами 1 и 2 останутся неиспользованными»; B = «варианты 1 и 2 достанутся рядом сидящим студентам»; C = «будут распределены последовательные номера вариантов».

10.A и Bи еще 8 человек стоят в очереди. Определить вероятность того, что A и Bотделены друг от друга тремя лицами.

11. Значения a и b равновозможны в квадрате $|a| \leq 1, |b| \leq 1.$ Найти $x^2 + 2ax + b$вероятности следующих событий: A = «корни квадратного трехчленадействительны», B = «корни квадратного трехчлена положительны».

12. Из отрезка$[-1;2]$ наудачу взяты два числа. Какова вероятность того, что их сумма больше единицы, а произведение меньше единицы?

13. На перекрестке установлен автоматический светофор, в котором одну минуту горит зеленый свет и полминуты красный, затем снова одну минуту – зеленыйи полминуты красный и т.д. В случайный момент времени к перекрестку подъезжает автомобиль. Какова вероятность того, что он проедет перекресток без остановки?

14. К автобусной остановке через каждые четыре минуты подходит автобус линии A и через каждые шесть минут – автобус линии B. Интервал времени между моментами прихода автобуса линии A и ближайшего следующего автобуса линии B равновозможен в пределах от нуля до четырех минут. Определить вероятность того, что: а) первый пришедший автобус окажется автобусом линии A; б) автобус какой-либо линии подойдет в течение двух минут.

15. Два парохода должны подойти к одному и тому же причалу. Время прихода обоих пароходов независимо и равновозможно в течение данных суток. Определить вероятность того, что одному из пароходов придется ожидать освобождение причала, если время стоянки первого парохода один час, а второго – два часа.

16. Иван и Петр договорились о встрече в определенном месте между одиннадцатью и двенадцатью часами. Каждый приходит в случайный момент указанного промежутка и ждет появления другого до истечения часа, но не более 15 мин, после чего уходит. Наблюдаемый результат – пара чисел(x,y), где x– время прихода Петра, y– время прихода Ивана. Определить вероятности следующих событий: A = «встреча состоялась», B = «Петр ждал Ивана все обусловленное время и не дождался», C = «Ивану не пришлось ждать Петра», D =«встреча состоялась после 11 ч 30 мин», E = «Иван опоздал на встречу», F = «встреча состоялась, когда до истечения часа оставалось меньше пяти минут».

17. Какова вероятность не целясь попасть бесконечномалой пулей в квадратную решетку, если толщина прутьев равна a, а расстояние между их средними линиями равно?

18. В течение года фирмы A, B, C, независимо друг от друга, могут обанкротиться с вероятностями 0,06; 0,09 и 0,05 соответственно. Найти вероятности того, что к концу года: 1) все три фирмы будут функционировать; 2) все три фирмы обанкротятся; 3) только одна фирма обанкротится; 4) только две фирмы обанкротятся; 5) хотя бы одна фирма обанкротится.

19. Пусть вероятность того, что в секции магазина по продаже мужской обуви очередной будет продана пара обуви 44-го размера, равна 0,12, 45-го – 0,04, 46-го или большего – 0,01. Найти вероятность того, что очередной будет продана пара мужской обуви не менее 44-го размера.

20. Студент выучил к зачету 15 вопросов из 20. Ему по одному предлагают три вопроса. Найти вероятность того, что только на третий из них он не дает ответа.

21. Для рабочего из маршрутов трамвая № 1, 2, 4, 7 попутными являются маршруты № 1 и 4. Вычислить вероятность того, что к остановке первым подойдет трамвай маршрута попутного для него номера, если по линиям маршрутов № 1, 2, 4, 7 курсируют соответственно 12, 4, 10, 14 поездов.

22. Два охотника стреляют в волка. Для первого охотника вероятность попадания в цель 0,7, для второго – 0,8. Определить вероятность попадания в волка, если каждый охотник: 1) делает по одному выстрелу; 2) делает по два выстрела?

23. Для сигнализации об аварии установлены два независимо работающих сигнализатора. Вероятность того, что при аварии сигнализатор сработает, равна 0,95 для первого сигнализатора и 0,9 для второго. Найти вероятность того, что при аварии сработает: а) только один сигнализатор; б) хотя бы один сигнализатор.

24. Для компании, занимающейся строительством терминалов для аэропортов, вероятность получить контракт в стране A равна 0,4, вероятность выиграть его в стране B равна 0,3. Вероятность того, что контракты будут заключены и в стране A, и в стране B, равна 0,12. Какова вероятность того, что компания получит контракт: а) хотя бы в одной стране; б) только в одной стране?

25. Обследовалась группа из 10000 человек в возрасте свыше 60 лет. Оказалось, что 4000 из них постоянно курит. У 1800 человек из курящих обнаружились серьезные изменения в легких. Среди некурящих серьезные изменения в легких имели 1500 человек. Являются ли курение и наличие серьезных изменений в легких независимыми событиями? (Ответ дать, проверив выполнение равенства $P(AB) = P(A)P\cdot(B)$, где событие A – человек курит, событие B – человек имеет серьезные изменения в легких.

26. В продажу поступают телевизоры трех заводов. Продукция первого завода содержит 20 % телевизоров со скрытым дефектом, второго – 10 % и третьего – 5 %. Какова вероятность приобрести исправный телевизор, если в магазин поступило 30 % телевизоров с первого завода, 20 % – со второго и 50 %– с третьего?

27. Два цеха штампуют однотипные детали. Первый цех дает α % брака, второй –β %. Для контроля отобрано n_1 деталей из первого цеха и n_2 из второго. Эти $n_1 + n_2$ деталей смешаны в одну партию, и из нее наугад извлекают одну деталь. Какова вероятность того, что она бракованная?

28. Курс доллара повышается в течение квартала с вероятностью 0,9 и понижается с вероятностью 0,1. При повышении курса доллара фирма рассчитывает получить прибыль с вероятностью 0,85; при понижении – с вероятностью 0,5. Найти вероятность того, что фирма получит прибыль.

29. Из 10 студентов, пришедших сдавать экзамен по теории вероятностей и взявших билеты, Иванов и Петров знают по 20 билетов из 30, Сидоров успел повторить только 15 билетов, остальные студенты знают все 30 билетов. Экзаменатор наудачу вызывает отвечать одного из студентов. Какова вероятность того, что вызванный сдал экзамен, если знание билета гарантирует сдачу экзамена с вероятностью 0,85, а при незнании билета можно сдать экзамен лишь с вероятностью 0,1?

30. В ящике лежит 20 теннисных мячей, в том числе 15 новых и 5 игранных. Для игры наудачу выбираются два мяча, и после игры возвращаются обратно. Затем для второй игры также наудачу извлекаются еще два мяча. Какова вероятность того, что вторая игра будет проводиться новыми мячами?

31. Два автомата производят одинаковые детали, которые поступают на общий конвейер. Производительность 1-го автомата вдвое больше производительности 2-го. 1-й автомат производит в среднем 60 % деталей отличного качества, а второй – 84 %. Наудачу взятая с конвейера деталь оказалась отличного качества. Найти вероятность того, что эта деталь изготовлена: а) 1-м автоматом; б) 2-м автоматом.

32. Исследованиями психологов установлено, что мужчины и женщины по-разному реагируют на некоторые жизненные обстоятельства. Результаты исследований показали, что 70 % женщин позитивно реагируют на изучаемый круг ситуаций, в то время как 40 % мужчин реагируют на них негативно. 15женщин и 5 мужчин заполнили анкету, в которой отразили свое отношение к предлагаемым ситуациям. Случайно извлеченная анкета содержит негативную реакцию. Чему равна вероятность того, что ее заполнял мужчина?

33. Число грузовых машин, проезжающих мимо бензоколонки, относится к числу легковых машин как 3:2. Вероятность того, что грузовая машина будет заправляться, равна 0,1, легковая – 0,2. Найти вероятность того, что заправляющаяся у бензоколонки машина – грузовая.

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

34. В банк поступило 6 заявлений от физических лиц на получение кредита. Вероятность получить первый кредит для каждого равна $\frac{3}{4}$. Найти вероятности следующих событий:

1) будет выдано ровно 3 кредита;

2) будет выдано не менее двух кредитов.

35. Два баскетболиста делают по три броска в корзину. Вероятность попадания мяча при каждом броске равна соответственно 0,6 и 0,7.

Найти вероятность того, что у обоих будет равное количество попаданий.

36. Игральная кость подбрасывается 5 раз. Найти вероятность того, что два раза появится число очков, кратное трем.

37. Экзаменационный билет состоит из пяти вопросов в виде теста с тремя возможными ответами на каждый из пяти вопросов, из которых нужно выбрать один правильный. Какова вероятность сдать экзамен методом простого угадывания, если достаточно ответить хотя бы на 4 вопроса?

38. Три охотника одновременно выстрелили по волку. Вероятности попадания каждым из охотников одинаковы и равны 0,4. Определить вероятность того, что волк будет убит, если известно, что при одном попадании охотники убивают волка с вероятностью 0,2, при двух – с вероятностью 0,5 и при трех – с вероятностью 0,8.

39.Радиоаппаратура состоит из 2000 элементов. Вероятность отказа одного элемента в течение года равна 0,001. Какова вероятность отказа двух элементов за год? Какова вероятность отказа не менее двух элементов за год?

40.Завод отправил в магазин 5000 лампочек. Вероятность того, что лампочка разобьется при транспортировке равна 0,0002. Найти вероятность того, что в магазин привезли не более трех разбитых лампочек.

41.Среди семян пшеницы 0,6 % семян сорняков. Какова вероятность при случайном отборе 1000 семян обнаружить а)не менее 3 семян сорняков; б)не более 16 семян сорняков; в) ровно 6 семян сорняков?

42.Вероятность изготовления стандартной детали на автомате равна 0,95. Изготовлена партия в 200 деталей. Найти наиболее вероятное число нестандартных деталей в этой партии. Найти вероятность этого количества нестандартных деталей.

43.В камере хранения ручного багажа 80 % всей клади составляют чемоданы, которые вперемешку с другими вещами хранятся на стеллажах. Через окно выдачи были получены все вещи с одного из стеллажей в количестве 50 мест. Найти вероятность того, что среди выданных вещей было 38 чемоданов.

44. На факультете 730 студентов. Вероятность рождения каждого студента в данный день равна $\frac{1}{365}$. Найти вероятность того, что найдутся три студента с одним и тем же днем рождения.

45.Вероятность поражения мишени при одном выстреле равна 0,8. Найти вероятность того, что при 100 выстрелах мишень будет поражена ровно 75 раз.

46.Вероятность рождения мальчика равна 0,51. Найти вероятность того, что среди 100 новорожденных окажется 50 мальчиков?

47.Известно, что в среднем 70% продукции завода является продукцией первого сорта. Какова вероятность того, что в партии из 200 изделий имеется 120 изделий 1-го сорта?

48.Для поступления в колледж необходимо успешно сдать вступительные экзамены. В среднем их успешно сдают 65% абитуриентов. В приемную комиссию поступило 700 заявлений. Какова вероятность того, что хотя бы 500 поступят в колледж?

49.При установившемся технологическом процессе цех выпускает в среднем 80 % продукции первого сорта. Какова вероятность того, что в партии из 125 изделий будет не менее 100 изделий первого сорта?

50. В страховом обществе застраховано 10000 лиц одного возраста и одной социальной группы. Вероятность смерти в течение года для каждого лица равна 0,006. Каждый застрахованный вносит 1 января 12 у.е. страховых, и в случае смерти его родственники получают от общества 1000 у.е. Найти вероятность того, что: а) общество потерпит убыток; б) общество получит прибыль, не меньшую 40000, 60000, 80000 у.е.

51.Вероятность рождения мальчика p = 0,512. Вычислить вероятность событий: A = «среди 100 рожденных будет больше мальчиков, чем девочек», B= «разница между количеством мальчиков и количеством девочек из 100новорожденных не превысит 10».

52. Вероятность выхода из строя за время T одного конденсатора равна 0,2. Определить вероятность того, что за время T из 100 конденсаторов выйдут из строя: а) не менее 20; б) менее 28; в) от 14 до 26 конденсаторов.

53. При социологических опросах граждан каждый человек независимо от других может дать неискренний ответ с вероятностью 0,2. Найти вероятность того, что из 22 500 опросов число неискренних ответов будет не более 4620.

54. В билете три задачи. Вероятность правильного решения первой задачи равна 0,9, второй – 0,8, третьей – 0,7. Составить закон распределения числа правильно решенных задач в билете и вычислить математическое ожидание и дисперсию.

55. Поступающий в институт должен сдать 3 экзамена. Вероятность сдачи первого экзамена 0,9, второго – 0,8, третьего – 0,7. Следующий экзамен поступающий сдает только в случае успешной сдачи предыдущего. Составить закон распределения числа приходов на экзамен для лица, поступающего в институт. Найти математическое ожидание случайной величины.

56. В городе 4 коммерческих банка. У каждого риск банкротства в течение года составляет 10%. Составить закон распределения числа банков, которые могут обанкротиться в течение следующего года и найти числовые характеристики этого распределения.

57. Вероятность поражения земляники вирусным заболеванием равна 0,2. Составить закон распределения числа кустов земляники, зараженных вирусом, из четырех посаженных кустов. Найти математическое ожидание и дисперсию этой случайной величины.

58. Дана функция

$$F(x)=\begin{cases} 0, \text{если } x\le \frac{\pi}{2}, \\ \cos x, \text{если } -\frac{\pi}{2}<x\le 0, \\ 1, \text{если } x>0. \end{cases}$$

Показать, что данная функция является функцией распределения некоторой случайной величины X. Найти вероятность того, что эта случайная величина принимает значения из интервала $\left(-\frac{\pi}{3};0\right)$.

59. Дана функция

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

Является ли она функцией распределения некоторой случайной величины?

60. Является ли функцией распределения некоторой случайной величины функция

$$F(x)=\frac{1}{1+x^2}\,(-\infty<x<+\infty)?$$

61. Является ли функцией распределения некоторой случайной величины каждая из следующих функций:

$$F(x)=\begin{cases} e^x, \text{при } x\le 0, \\ 1, \text{при } x>0. \end{cases}$$

62. Дана функция распределения случайной величины X:

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

Найти плотность вероятности, а также вероятности $P(X=1), P(X<1), P(1\le X<2)$,

63. Случайная величина X, сосредоточенная на интервале $[-1;3]$ задана функцией распределения $F(x)=\frac{1}{4}x+\frac{1}{4}$. Найти вероятность попадания случайной величины X в интервал $[0; 2]$. Построить график функции $F(x)$.

64. Случайная величина X, сосредоточенная на интервале $[2; 6]$ задана функцией распределения $F(x) = \frac{1}{16}(x^2 - 4x + 4)$. Найти вероятность того, что случайная величина X примет значения: а) меньше 4; б) меньше 6; в)не меньше 3; г) не меньше 6.

65. Случайная величина X, сосредоточенная на интервале $(1; 4)$, задана квадратичной функцией $F(x) = ax^2 + bx + c$, имеющей максимум при $x = 4$. Найти параметры a, b, c и вычислить вероятность попадания случайной величины Xв интервал $[2; 3]$.

66. Функция распределения случайной величины X имеет вид

$$F(x) = \begin{cases} 0, \text{при } x < -1, \\ a + b\, arcsin\, x, \text{при } -1 \le x \le 1, \\ 1, \text{при } x > 1. \end{cases}$$

Определить постоянные a и b. Найти плотность вероятности случайной величины X и построить ее график.

67. Плотность распределения вероятностей случайной величины X определяется функцией

$$p(x) = ax^2 e^{-kx} (k > 0, 0 \le x < +\infty).$$

Найти значение коэффициента a. Найти функцию распределения$F(x)$величины X.

68. Случайная величина X имеет плотность

$$p(x) = \begin{cases} \frac{6}{11}(x^2 + x + 1), \text{при } 0 < x \le 1, \\ 0, \text{при } x \le 0 \text{ и при } x > 1. \end{cases}$$

Найти математическое ожидание и дисперсию X.

69. Случайная величина X имеет плотность

$$p(x) = \begin{cases} \frac{2}{\pi} cos^2 x, \text{при } |x| \le \frac{\pi}{2}, \\ 0, \text{при } |x| > \frac{\pi}{2}. \end{cases}$$

Найти математическое ожидание и дисперсию X.

$$M(X) = 0; D(X) = \frac{\pi^2}{12} - \frac{1}{2}.$$

70. Случайная величина X задана плотностью распределения

$$p(x) = \begin{cases} \sin 2x\,, \text{при } x \in \left(0; \frac{\pi}{2}\right), \\ 0, \text{при } x \notin \left(0; \frac{\pi}{2}\right). \end{cases}$$

Найти математическое ожидание функции $Y = X^2$(не находя предварительно плотности распределенияY).

71. Плотность случайной величины X имеет вид

$$p(x) = \begin{cases} ae^{-x}, \text{при } x \geq 0, \\ 0, \text{при } x < 0. \end{cases}$$

Найти коэффициент a. Вычислить моду, медиану, математическое ожидание, дисперсию, начальные и центральные моменты первого, второго и третьего порядков случайной величины X.

72. Случайная величина X задана плотностью распределения

$$p(x) = \begin{cases} 0, \text{при } x \leq 1, \\ \frac{6}{x^7}, \text{при } x > 1. \end{cases}$$

Найти начальные моменты случайной величины X.

73. Плотность вероятности случайной величины X имеет вид

$$p(x) = \begin{cases} \frac{1}{2}\cos x\,, \text{при } -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}, \\ 0, \text{при } x < -\frac{\pi}{2} \text{ и при } x > \frac{\pi}{2}. \end{cases}$$

Найти математическое ожидание и дисперсию случайной величины$Y = \sin 2X$.

74. Случайная величина X имеет функцию распределения

$$F(x) = \begin{cases} 0, \text{при } x \leq 0, \\ x^4, \text{при } 0 < x \leq 1, \\ 1, \text{при } x > 1. \end{cases}$$

Найти математическое ожидание случайной величины $Y = \frac{1}{X+1}$.

75. По данным задачи 8.9 (при $a = \frac{1}{2}, b = \frac{1}{\pi}$) найти моду и медиану распределения; вероятность того, что случайная величина X окажется в промежутке $\left(-\frac{1}{2}, \frac{1}{2}\right)$, математическое ожидание и дисперсию X.

76. Найти математическое ожидание и дисперсию случайной величины, плотность вероятности которой имеет вид

$$p(x) = \frac{1}{2}e^{-|x|} \text{ (распределение Лапласа).}$$

77. Цена деления шкалы измерительного прибора равна 0,1. Показания прибора округляют до ближайшего целого деления. Полагая, что при отсчете ошибка округления распределена по равномерному закону, найти: 1) математическое ожидание, дисперсию и среднее квадратическое отклонение этой случайной величины; 2) вероятность того, что ошибка округления: а) меньше 0,01; б) больше 0,03.

78. Автобусы некоторого маршрута идут строго по расписанию. Интервал движения 4 мин. Найти вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус менее 2 мин.

79. Минутная стрелка часов перемещается скачком в конце каждой минуты. Найти вероятность того, что в данное мгновение часы покажут время, которое отличается от истинного не более чем на 10 с.

80. Случайные величиныX и Yнезависимы и распределены равномерно: X – в интервале $(a; b)$, Y –$(c; d)$. Найти математическое ожидание и дисперсию произведения XY

81. Диаметр круга x измерен приближенно, причем $5 \leq x \leq 6$. Рассматривая диаметр как случайную величину X, распределенную равномерно в интервале $(5; 6)$, найти математическое ожидание и дисперсию площади круга.

82. Ребро куба x измерено приближенно, причем $2 \leq x \leq 3$. Рассматривая длину ребра куба как случайную величину X, распределенную равномерно в интервале $(2; 3)$, найти математическое ожидание и дисперсию объема куба.

83. Пусть случайные величины X_1и X_2независимы и равномерно распределены на отрезке $[-1; 1]$. Найти вероятность того, что$\min\limits_{i=1,2}|x_i| > \frac{1}{2}$.

84. Случайная величина X распределена по показательному закону

$$p(x) = \begin{cases} 0, \text{при } x < 0, \\ 7e^{-7x}, \text{при } x \geq 0. \end{cases}$$

Найти математическое ожидание, дисперсию, среднее квадратическое отклонение и функцию распределения этой случайной величины. Найти вероятность попадания значений случайной величины X в интервал $(0{,}2; 1{,}1)$.

85. Среднее время безотказной работы прибора равно 85ч. Полагая, что время безотказной работы прибора имеет показательный закон распределения, найти: а) выражение его плотности вероятности и функции распределения; б)вероятность того, что в течение 100 ч прибор не выйдет из строя.

а) $F(x)=\left\{0,\text{при } x<0; 1-e^{-\frac{1}{85}x},\text{при } x\geq 0\right\}$;
б) $P(X\geq 100)=1-F(100)=0{,}31$.

86. Найти эксцесс показательного распределения.

87. Производится испытание трех элементов, работающих независимо один от другого. Длительность времени безотказной работы элементов распределена по показательному закону: для первого элемента $p_1(t)=0{,}1e^{-0{,}1t}$; для второго $-p_2(t)=0{,}2e^{-0{,}2t}$ для третьего элемента $p_3(t)=0{,}3e^{-0{,}3t}$. Найти вероятности того, что в интервале времени $(0;10)$часов откажут: а) только один элемент; б) только два элемента; в) хотя бы один элемент; г) все три элемента; д)не менее двух элементов.

88.P%-м ресурсом элемента называется такое число t, что за время t элемент не выходит из строя с вероятностью P. Считается, что время t непрерывной работы электрической лампочки распределено по показательному закону. Найти вероятность того, что лампочка будет гореть в течение 2 лет, если ее 90 %-й ресурс составляет 6 мес.

89. Срок службы жесткого диска компьютера – случайная величина, подчиняющаяся экспоненциальному распределению со средней в 12000 часов. Найти долю жестких дисков, срок службы которых превысит 20000 часов.

$P(T>20000)=0{,}1882$.

90. Срок службы батареек для слуховых аппаратов приблизительно подчиняется экспоненциальному закону с $\lambda=1/12$. Какова доля батареек со сроком службы больше чем 9 дней?

ЗАДАНИЯ ДЛЯ СРС
ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

91.Дана таблица результатов наблюдений над величинами X и Y для шести фермерских хозяйств, где X–количество финансовых вложений на 1 гектар пашни за год, а Y – урожайность этого гектара пашни за год (в некоторых условных единицах)

X	2	4	6	8	10	12
Y	3,5	6,0	7,0	6,0	7,5	8,5

Построить график, отражающий связь X и Y. Рассматривая результаты наблюдений, как выборочные наблюдения случайных величин X и Y, на основе коэффициента корреляции Пирсона оценить их влияние друг на друга. Сделать выводы.

Дополнительно: к количественным данным применить ранговый подход и вычислить коэффициент корреляции Спирмена. Сравнить его с ранее вычисленным коэффициентом Пирсона. Сделать выводы на основе этого сравнения.

92.Для предприятия имеются данные по количеству сделок по продаже товара и затратам на мониторинг рынка (в тыс. долл.) в течение 5 месяцев.

Месяцы	1	2	3	4	5
Количество сделок	6	3	5	9	4
Затраты на мониторинг	3	5	4	4	3

Построить график, отражающий связь этих параметров. Рассматривая данные, как выборочные наблюдения случайных величин, Найдите коэффициент корреляции между ними. Какие рекомендации вы бы дали руководству предприятия?

93.Имеются данные по шести предприятиям, показывающие количество проданных видеомагнитофонов и той продажной ценой (в тыс. руб.) за видеомагнитофон, которая была установлена для каждого предприятия:

№ предприятия	1	2	3	4	5	6
Количество продаж	20	13	15	10	10	17
Цена	4,5	5	5.5	6	5	4.8

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими параметрами. Сделайте вывод о наличии или отсутствии влияния параметров друг на друга.

94.Дана таблица данных по затратам предприятия на рекламу своей продукции X и объемами продаж этой продукции Y (в условныхденежных единицах) в разные месяцы:

X	3	4	5	3	6
Y	3,5	4,0	5,0	4,0	6,5

Построить график, отражающий связь X и Y. Рассматривая данные, как выборочные наблюдения случайных величин, на основе коэффициента корреляции Пирсона оцените их влияние друг на друга. Сделайте выводы.

Дополнительно: к количественным данным применить ранговый подход и вычислить коэффициент корреляции Спирмена. Сравнить его с ранее вычисленным коэффициентом Пирсона. Сделать выводы на основе этого сравнения.

95.Для цеха имеются данные по себестоимости продукции (в стоимостном выражении) и количеству работников цеха***:***

Себестоимость	2	2.5	3	3.5	4	4.2
Количество работников	4	5	6	7	8	9

Постройте график, отражающий связь величин. Рассматривая данные, как выборочные наблюдения случайных величин, оцените их влияние друг на друга.

96.Для предприятия имеются данные по количеству заключенных сделок на продажу товараи затратам на мониторинг рынка (в тыс. дол.) в течение шести месяцев.

Месяцы	1	2	3	4	5	6
Количество сделок	4	3	5	3	6	5
Затраты на мониторинг	4	5	3	4	3	2

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими параметрами. Сделайте вывод о наличии или отсутствии влиянии этих параметров друг на друга. Что бы вы порекомендовали руководству предприятия?

97.Имеются данные для торгового центра «Нью-Васюки» по затратам на рекламу (тыс. дол.) и количеством проданных шахматных досок (в сотнях штук) в течение последних шести месяцев:

Месяцы	1	2	3	4	5	6
Затратына рекламу	5	4	3	2	6	7
Продано досок	4	4.5	5.2	5.5	6	7

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между ними.

98.Имеются данные для шести предприятий отрасли по производительности труда (X, штук) и средней заработной плате работников предприятия (Y, сотни долларов):

№ предприятия	1	2	3	4	5	6
X	32	30	36	40	41	47
Y	2.0	2.4	2.8	3.0	3.1	3.3

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими параметрами. Сделайте вывод.

99.Для отрасли имеются данные по 6 предприятиям: время эксплуатации оборудования (в годах) и стоимость на обслуживание этого оборудования в течение года(в тыс. долл.):

№ предприятия	1	2	3	4	5	6
Время	1	2	3	4	5	6
Стоимость	4	4.5	5.2	5.5	6	7

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими параметрами. Сделайте вывод.

100.Изучается зависимость себестоимости единицы продукции (Y, тыс. руб.) от величины выпуска продукции (X, тыс. штук) по разным предприятиям отрасли. Результат изучения пяти предприятий таков:

№ предприятия	1	2	3	4	5
X	2	3	4	5	6
Y	1,9	1,7	1,8	1,6	1,4

Построить график, отражающий связь X и Y. Рассматривая данные, как выборочные наблюдения случайных величин, на основе коэффициента корреляции Пирсона выявить тесноту связи между показателями X и Y.

Сделайте вывод.

101.Дана таблица результатов наблюдений над величинами X и Y:

X	-2	-1	1	3	4	6
Y	2.5	3.5	4.0	6.0	5.5	8.5

Постройте график, отражающий связь величин. Рассматривая данные, как выборочные наблюдения случайных величин, оцените влияние случайных величин друг на друга. Сделайте выводы.

102.Затраты Xна усовершенствование колес за год и величина годовой прибыли Yкомпании «Невидимые трамваи» в течение последних 5 лет представлены следующей таблицей:

X	4	1	2	4	3
Y	5	4	3	7	5

Построить график, отражающий связь двух величин. Рассматривая данные, как выборочные наблюдения случайных величин, найдите коэффициент корреляции и сделайте вывод о тесноте связи между этими затратами и успехами компании. (Задача предложена Аносовой О.Д.)

103.Получены данные по затратам предприятия на рекламу своей продукцииX и объемам продаж этой продукции Y (в стоимостном выражении) за пять месяцев:

X	2	3	5	3	5
Y	4	5	3	4	3

Рассматривая данные, как выборочные наблюдения случайных величин, найдите дляXточечные несмещенные оценки математического ожидания и дисперсии. По наблюдениям постройте график, отражающий связь двух параметров. Найдите выборочныйкоэффициент корреляции; сделайте вывод о тесноте связи между этими параметрами. Сформулируйте рекомендации предприятию по коррекции рекламной деятельности.

104.Для одного предприятий имеются данные по количеству уволенных сотрудников мужского и женского пола в течение шести месяцев:

Месяцы	1	2	3	4	5	6
Женщины	2	3	3	2	6	5
Мужчины	4	5	3	4	3	3

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими параметрами. Сделайте вывод о наличии или отсутствии у руководства данного предприятия политики дискриминации по половому признаку.

105.Имеются данные для шести предприятий отрасли по производительности труда (Y, штук) и средней заработной плате работников предприятия (X, у.е.):

№ предприятия	1	2	3	4	5	6
X	32	30	36	40	41	47
Y	2,0	2,4	2,8	3,0	3,1	3,3

Построить график, отражающий связь X и Y. Рассматривая данные, как выборочные наблюдения случайных величин, на основе коэффициента корреляции Пирсона оценить тесноту связи этих двух величин друг на друга. Сделать выводы.

106.Дана таблица результатов наблюдений над величинами X иY:

X	2	4	6	8	10	12
Y	3,5	6,0	7,0	6,0	7,5	8,5

Построить график, отражающий связь X иY. Рассматривая результаты наблюдений, как выборочные наблюдения случайных величин X иY, на основе коэффициента корреляции Пирсона оценить влияние этих величин друг на друга.

107.Известны данные об импорте и потреблении нефти в США за шесть лет (в млн.баррелей в сутки):

Годы	1997	1998	1999	2000	2001	2002
Импорт	8.2	9	10.1	11.3	12	11.5
Потребление	18.3	18.9	19.5	19.7	20	21

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими параметрами. Сделайте вывод о наличии или отсутствии зависимости между потреблением нефти и ее импортом.

108.Имеются данные для шести предприятий отрасли по производительности труда и средней заработной плате работников предприятия:

№ предприятия	1	2	3	4	5	6
Производительност	3	6	9	5	2	5
Зарплата	1	3	4	2	1	3

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите коэффициент корреляции и сделайте вывод о тесноте связи между этими параметрами.

109.Дана таблица результатов наблюдений над величинами X иY:

X	2	4	6	8	10	12
Y	3.5	6.0	7.0	5.0	7.5	8.5

Постройте график, отражающий связь между величинами. Рассматривая данные, как выборочные наблюдения случайных величин, оцените влияние случайных величин друг на друга. Сделайте вывод.

110.Затраты предприятия «Воробышек» на реконструкцию офисных помещений и количество обслуживаемых ею клиентов (в условных единицах) в течение 5 лет представленыследующей таблицей:

X	3	2	4	5	7
Y	3	2	5	6	8

Предполагается линейная зависимость между затратами xи количеством клиентов Yвида $Y = \alpha_0 + \alpha_1 + \varepsilon$ с величиной случайного влияния ε. Получить уравнение линейной регрессии. Выполнить на основе полученного уравнения прогноз о количестве обслуживаемых клиентов, если в последующем году затраты на реконструкцию составят 8 условных единиц.

111.Для одного предприятия имеются данные по количеству сделок на продажу товара и затратам на мониторинг рынка (в у.е.) в течение 5 месяцев:

Месяцы	1	2	3	4	5
Количество сделок	7	4	6	10	10
Затраты на мониторинг	2	6	3	2	3

Построить график, отражающий связь этих параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найти коэффициент корреляции между ними. Какие рекомендации вы бы дали руководству предприятия?

112.Для отрасли имеются данные по 6 предприятиям: время эксплуатации оборудования (в годах) и стоимость на обслуживание этого оборудования в течение года(в у.е.):

Предприятия	1	2	3	4	5	6
Время	1	2	3	4	5	6
Стоимость	4	4.5	5.2	5.5	6	7

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими параметрами. Сделайте вывод.

113.Затраченное на подготовку к экзамену времяи оценка за экзамен для нескольких студентовпредставлены в следующей таблице:

x	2	4	5	8	13
Y	1	3	4	7	8

Предполагается линейная зависимость между временем подготовки xи оценкой Yвида $Y = \alpha_0 + \alpha_1 + \varepsilon$ с величиной случайного влияния ε. Получить уравнение линейной регрессии.

114.Для одного предприятия имеются данные по количеству прогулов работников мужского и женского пола в течение шести месяцев:

Месяцы	1	2	3	4	5	6
Женщины	5	10	3	7	12	5
Мужчины	7	5	6	10	3	15

Постройте график, отражающий связь двух величин. Рассматривая данные, как выборочные наблюдения случайных величин X и Y, найдите тесноту связи между этими параметрами. Сделайте вывод.

115.Для одного предприятия имеются данные по количеству заключенных сделок на продажу товара и затратам на мониторинг рынка (в у.е.) в течение шести месяцев:

Месяцы	1	2	3	4	5	6
Количество сделок	4	3	5	3	6	5
Затраты на мониторинг	4	5	3	4	3	2

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими параметрами. Сделайте вывод о наличии или отсутствии влиянии этих параметров друг на друга. Что бы вы порекомендовали руководству предприятия?

116.Ежегодно устанавливаемые цены X на билеты фиксированного рейса компании «Тише едешь – дальшебудешь» и количество пассажиров Y (в условном измерении), пользующихся услугами этой компании на этот рейс на протяжении пяти лет, представлены таблицей:

X	1	3	5	4	5
Y	3	2	3	1	1

Построить график, отражающий связь этих величин. Рассматривая данные, как выборочные наблюдения случайных величин, найдите коэффициент корреляции между ними и сделайте выводы о степени влияния цены на объемы перевозок.

117.Изучается зависимость себестоимости единицы продукции (X, тыс.руб.) от величины выпуска продукции (Y, тыс.штук) по разным предприятиям отрасли. Результат изучения пяти предприятий таков:

Предприятия	1	2	3	4	5
X	2	3	4	5	6
Y	1.9	1.7	1.8	1.6	1.4

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими показателями.

118.Имеются данные для шести предприятий отрасли по производительности труда и средней заработной плате работников предприятия:

Предприятия	1	2	3	4	5	6
Производительность	3	3	8	6	1	2
Зарплата	2	2	5	2	5	5

Постройте график, отражающий связь двух параметров. Рассматривая данные, как выборочные наблюдения случайных величин, найдите коэффициент корреляции и сделайте вывод о тесноте связи между этими параметрами.

119.Для одного предприятия имеются данные по количеству уволенных сотрудников мужского $\boldsymbol{x}$ и женского $\boldsymbol{Y}$ пола в течение шести месяцев:

Месяцы	1	2	3	4	5	6
Женщины	2	3	3	2	6	5
Мужчины	4	5	3	4	3	3

Предполагается линейная зависимость между $\boldsymbol{x}$ и $\boldsymbol{Y}$ вида $Y = \alpha_0 + \alpha_1 + \varepsilon$ с величиной случайного влияния ε. Постройте уравнение линейной регрессии и сделайте вывод о наличии или отсутствии у руководства данного предприятия политики дискриминации по половому признаку.

120.Имеются данные по шести предприятиям, показывающие количество проданных видеомагнитофонов и их продажной ценой (в тыс. руб. за единицу), которая была установлена каждым предприятием:

№	1	2	3	4	5	6
Количество	20	13	15	10	106	17
Цена	4,5	5	5.5	6	5	4.8

Постройте график, отражающий связь двух показателей. Рассматривая данные, как выборочные наблюдения случайных величин, найдите тесноту связи между этими показателями. Сделайте вывод о наличии или отсутствии их влияния друг на друга.

121.Затраченное на подготовку к экзамену времяи оценка за экзамен для нескольких студентов представлены в следующей таблице:

x	2	4	5	8	13
Y	1	3	4	7	8

Предполагается линейная зависимость между временем подготовки xи оценкой Yвида $Y = \alpha_0 + \alpha_1 + \varepsilon$ с величиной случайного влияния ε. Постройте уравнение линейной регрессии двумя способами, сравнив результаты вычислений между собой.

122.За семь месяцев предприятие получало ежемесячную прибыль (в у.е.): 2, 3, 2, 4, 3, 5, 4. Найдите выборочное среднее и выборочную дисперсию.

123.По данным предыдущей задачи, рассматривая данные, как выборочные наблюдения случайных величин, постройте эмпирическую функцию распределения прибыли, оценку гистограммы плотности распределения с шагом 0.5 и оцените.по графику медиану.

124.За 6 месяцев работы сборочной автоматизированной линии получены такие данные по количеству аварий за каждый месяц работы: 2, 2, 2, 1, 4, Найдите наиболее доброкачественную точечную оценку числа ежемесячных аварий. Какими свойствами должна обладать эта оценка? Найдите вероятность того, что за седьмой месяц произойдет 3 аварии.

125.За десять месяцев работы малое предприятие «Воробышек» получало ежемесячную прибыль (в у.е.): 2, 4, 7, 4, 5, 4, 5, 7, 3, 3. Рассматривая данные, как выборочные наблюдения случайной величины, найдите выборочное среднее и показатели колеблемости / вариации для прибыли по выборке. Является ли выборка однородной?

126.По данным предыдущей задачи постройте эмпирическую функцию распределения прибыли (кумуляту) и оценку гистограммы плотности распределения с шагом 1.25. Найдите с помощью соответствующего графика медиану.

127. Ежедневный доход казино «Версаль» составил за 7 дней ряд значений:2, 3, 4, 1, 5, 6, 2 (в условных единицах). Рассматривая данные, как выборочные наблюдения случайной величины, найдите выборочную среднюю и выборочную дисперсию дохода казино.

128. По данным предыдущей задачи построить эмпирическую функцию распределения дохода казино и оценку гистограммы плотности распределения с шагом 0.8. Найдите на графике моду.

129. За десять днейработы малое предприятие «Дюймовочка» получало дневную прибыль (в у.е.): 4, 5, 8, 5, 9, 3, 4, 3, 3, 3. Рассматривая данные, как выборочные наблюдения случайной величины, найдите выборочное среднее и все меры колеблемости / вариации выборки. Является ли выборка однородной?

130. По данным предыдущей задачи постройте эмпирическую функцию распределения прибыли (кумуляту) и оценку гистограммы плотности распределения с шагом 1.1. Найдите с помощью соответствующего графика медиану.

131. За семь месяцев предприятие получало ежемесячную прибыль (в у.е.): 3, 3, 4, 4, 6, 5, 6. Рассматривая данные, как выборочные наблюдения случайной величины, найдите выборочное среднее и выборочную дисперсию прибыли.

132. По данным предыдущей задачи построить эмпирическую функцию распределения прибыли и оценку гистограммы плотности распределения с шагом 0.6. Найдите на графике медиану.

133. Ежедневный доход казино «Версаль» составил за 7 дней ряд: 2, 3, 4, 1, 5, 6, 2 (в условных единицах). Рассматривая данные, как выборочные наблюдения случайной величины, найдите выборочную среднюю и выборочную дисперсию дохода.

134. По данным предыдущей задачи построить эмпирическую функцию распределения дохода казино и оценку гистограммы плотности распределения с шагом 0.6. Найдите на графике моду.

135. За семь месяцев предприятие получало ежемесячную прибыль (в у.е.): 3, 3, 4, 4, 6, 5, 6. Рассматривая данные, как выборочные наблюдения случайной величины, найдите выборочное среднее и выборочную дисперсию прибыли.

136. По данным предыдущей задачи построить эмпирическую функцию распределения прибыли и оценку гистограммы плотности распределения с шагом 0.6. Найдите на графике медиану.

137. За десять месяцев предприятие получало ежемесячную прибыль (в у.е.): 3; 5; 2; 3; 3; 5; 4; 6; 4; 6. Рассматривая данные, как выборочные наблюдения случайной величины, найдите выборочное среднее и все меры колеблемости / вариации выборки. Является ли выборка однородной?

138. По данным предыдущей задачи построить эмпирическую функцию распределения прибыли (кумуляту) и оценку гистограммы плотности распределения с шагом 0.8. Найдите с помощью соответствующего графика медиану.

139. За десять месяцев работы малое предприятие «Семь гномов» получало ежемесячнуюприбыль (в у.е.): 3, 3, 5, 6, 7, 3, 4, 4, 6, 6. Рассматривая данные, как выборочные наблюдения случайной величины, найдите выборочное среднее и все меры колеблемости / вариации выборки. Является ли выборка однородной?

140. По данным предыдущей задачи построить эмпирическую функцию распределения дохода казино (кумуляту) и оценку гистограммы плотности распределения с шагом 1,1. Найдите с помощью соответствующего графика медиану.

141. Новый официант ресторана «У дядюшки Сэма» Иван Перепелкин решил выяснить, сколько в среднем чаевых за день получают официанты этого ресторана. Для этого он опросил 7 официантов, что составило малую часть работающих в ресторане официантов. На основе опроса выяснилось, что в среднем каждый из опрошенных получает 35 у.е. за день; выборочное среднее квадратическое отклонение оказалось равным 15 у.е. Определить доверительный интервал, который с уровнем доверия (надежностью) 98% накроет истинное значение средней суммы чаевых, получаемых официантом этого ресторана за день. Какими способами можно уменьшить получившийся доверительный интервал?

142. При анализе точности фасовочного автомата было проведено 12 независимых контрольных взвешиваний пачек кофе. Известно, что фасовочный аппарат отрегулирован без смещения, так что его ошибка подчиняется нормальному закону распределения$N(0, \sigma^2)$, но значение параметра σ^2 неизвестно. По результатам контрольных взвешиваний была рассчитана выборочная дисперсия $S^2 = 0{,}7\ (\text{г}^2)$. Получить интервальную оценку для среднего квадратического отклонения ошибки взвешивания с уровнем доверия 0,95.

143. Аналитик рынка ценных бумаг оценивает среднюю доходность определенного вида акций. Случайная выборка из 16 дней показала, что средняя доходность по акциям данного типа составляет 8% с выборочным средним квадратическим отклонением в 4%. Предполагая, что доходность акции подчиняется нормальному закону распределения, определите 99%-ый доверительный интервал для средней доходности интересующего аналитика вида акций.

144. Понятие интервальной оценки неизвестного параметра. Как интервальная оценка связана с точечной оценкой параметра? Покажите эту связь графически. Найдите интервальную оценку математического ожидания нормального распределения с доверительной вероятностью 0,95, если известна дисперсия распределения (равна 4), а среднее арифметическое выборки равно 31,2. Объем выборки равен 25.

145.Опрос 545 случайно отобранных жителей города показал, что 39% из них довольны деятельностью вновь избранного мэра. Построить 98%-ный доверительный интервал для генеральной доли жителей всего города, которые довольны деятельностью мэра. Сколько следует опросить жителей города, чтобы доверительный интервал уменьшился в четыре раза?

146.Изучение роста десятилетних мальчиков одной московской школы на основе случайной выборки объемом 23 мальчика показало, что их средний рост по выборке составляет 118 см с выборочным средним квадратическим отклонением 6см. Найдите 98%-ный доверительный интервал для среднего квадратического отклонения, который характеризует рост всех десятилетних мальчиков московских школ.

147.Для изучения спроса на цветы в городе Нью-Васюки проведен опрос 40 мужчин на выходе из цветочных магазинов. Им задали вопрос: какое количество денег мужчина готов потратить на букет цветов? По результату опроса были вычислены величины – в среднем мужчина готов потратить 350 рублей при среднем квадратическом отклонении 100 рублей. Есть основания полагать, что случайная величина затрат при покупке цветочного букета подчиняется нормальному закону распределения. Найдите интервальную оценку (с доверительной вероятностью 90%) для средней суммы денег, которые готов потратить на букет любой мужчина этого города, склонный к такому поступку.

148. В условиях предшествующей задачи найдите доверительный интервал с тем же уровнем доверия (надежности) для среднего квадратического отклонения по всем тем мужчинам города Нью-Васюки, которые покупают цветы.

149.Случайная выборка 345 людей, обратившихся в брачное агентство, показала, что 210 из них нашли себе пару с его помощью. Построить 95% доверительный интервал для доли всех людей, обратившихся в это агентство инашедших себе супруга с его помощью. Найдите минимальный объем выборки, при котором предельная ошибка выборки для доли всех людей, обратившихся в это агентство и нашедших супруга, не превысит 0,015.

150.На контрольных испытаниях партии радиоприемников, работающих на батарейках, было выбрано 12 радиоприемников. Средняя продолжительность их работы оказалась равной 100 часов, а выборочная дисперсия 625 часов^2 Известно, что продолжительность работы приемника является нормально распределенной случайной величиной. С доверительной вероятностью 99% найдите интервальную оценку для среднего времени работы приемников данной партии.

151.В условиях предыдущей задачи найдите интервальную оценку с той же надежностью для истинного среднего квадратического отклонения времени работы приемников этой партии.

152.Понятие интервальной оценки неизвестного параметра. Как интервальная оценка связана с точечной оценкой параметра? Покажите эту связь графически. Найдите интервальную оценку математического ожидания нормального распределения с доверительной вероятностью 0,95, если известна дисперсия распределения (равна 4), а среднее арифметическое выборки равно 31,2. Объем выборки равен 25.

153.Приведены оценки (по 10-ти бальной шкале) студента Васечкина по пяти различным дисциплинам при первом и втором рубежном контроле:

КР1	7	5	3	8	9
КР2	5	4	7	7	8

На основе рангового коэффициента Спирмена найдите тесноту связи между результатами первого и второго рубежного контроля учебы Васечкина, рассматривая данные, как выборочные наблюдения случайных величин. Сделать выводы.

154.Студенты Петров и Васечкин ответили на пять тестов. Результаты их тестирования таковы:

Тесты	1	2	3	4	5
Петров	8	5	4	7	9
Васечкин	4	6	4	8	5

На основе рангового коэффициента Спирмена найдите значимость различий в результатах тестирования студентов, рассматривая данные, как выборочные наблюдения случайных величин. Сделать выводы.

155.Шесть сотрудников предприятия проходят аттестацию по результатам теста и на основе количества заказов, полученных каждым из них за определенный промежуток времени. Результаты их аттестации таковы:

Сотрудники	1	2	3	4	5	6
Результаты теста	3	2	6	4	1	6
Количество заказов	1	5	5	2	5	6

На основе рангового коэффициента Спирмена найдите тесноту связи между двумя показателями деятельности сотрудников, рассматривая данные, как выборочные наблюдения случайных величин. Сделать выводы.

156.Мужчины и женщины по-разному оценивают положительные человеческие качества. Предложили мужчинам и женщинам на основе десятибалльной шкалы (10 баллов – это максимум) оценить важность следующих пяти качеств в представителях противоположного пола:

Качества	Ум	Доброта	Красота	Юмор	Работоспособность
Мужчины	7	8	8	5	7
Женщины	10	5	3	8	10

Найдите тесноту связи между этими данными, рассматривая данные, как выборочные наблюдения случайных величин. Сделайте вывод о том, насколько близки или далеки мужчины и женщины в оценках качеств партнеров.

157.Приведены оценки студентки Таничкиной по шести различным дисциплинам при первом и втором рубежном контроле:

КР1	7	6	4	8	9	8
КР2	6	4	5	7	9	8

На основе рангового коэффициента Спирмена найдите тесноту связи между результатами учебы Васечкиной при первом и втором рубежном контроле, рассматривая данные, как выборочные наблюдения случайных величин. Сделайте вывод. Приведите конкретные примеры, в одном из которых коэффициент Спирмена был бы равен 1, а в другом был бы равен 0.

158.Рейтинг шести банков оценен двумя экспертами:

№ банка	1	2	3	4	5	6
1эксперт	3	2	1	3	5	6
2эксперт	2	3	1	4	7	9

На основе рангового коэффициента Спирмена найдите тесноту связи между оценками двух экспертов, рассматривая данные, как выборочные наблюдения случайных величин. Сделать выводы.

159.Рейтинги шести банков оценены двумя экспертами:

№ банка	1	2	3	4	5	6
1 эксперт	7	8	9	5	6	7
2 эксперт	3	8	8	6	3	9

Найдите тесноту связи между этими данными, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод осогласованности оценок экспертов.

160.Семь преподавателей школы проходят аттестацию по результатам теста и на основе оценки, поставленной им директором школы. Результаты их аттестации таковы:

Преподаватели	1	2	3	4	5	6	7
Результатытеста	7	2	5	4	3	8	9
Оценкадиректора	6	3	5	7	9	10	8

На основе рангового коэффициента Спирмена найдите тесноту связи между двумя показателями деятельности преподавателей, рассматривая данные, как выборочные наблюдения случайных величин. Сделайте вывод. Приведите конкретные примеры, в одном из которых коэффициент Спирмена был бы равен 1, а в другом примере он был бы равен 0.

161.Приведены результаты тестирования студента Васечкина по двум тестам. Получены следующие оценки на семи вопросам каждого теста:

Вопросы	1	2	3	4	5	6	7
Тест 1	8	5	4	8	3	2	9
Тест 2	7	5	6	9	4	6	8

На основе рангового коэффициента Спирмена найдите тесноту связи между результатами тестирования Васечкина по двум тестам, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод. Приведите конкретные примеры, в одном из которых коэффициент Спирмена был бы равен 1, а в другом примере он был бы равен 0.

162.Пять сотрудников предприятий проходят аттестацию по результатам теста и на основе количества заказов, полученных каждым из них за определенный промежуток времени. Результаты их аттестации таковы:

Сотрудники	1	2	3	4	5
Результаты теста	8	6	4	9	10
Количество заказов	5	6	7	4	8

Найдите тесноту связи между этими данными, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод.

163.Студент Васечкин имеет при первом и втором рубежном контроле обучения по семи дисциплинам следующие оценки:

Дисциплина	1	2	3	4	5	6	7
КР1	9	5	3	8	7	6	7
КР2	10	6	5	9	6	7	8

На основе рангового коэффициента Спирмена найдите тесноту связи между результатами обучения при первом и втором рубежном контроле, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод. Приведите конкретные примеры, в одном из которых коэффициент Спирмена был бы равен 1, а в другом примере он был бы равен 0.

164.Имеются оценки студента Васечкина по шести различным дисциплинам при первом и втором рубежном контроле обучения:

КР1	7	5	3	8	9
КР2	5	4	6	7	8

На основе рангового коэффициента Спирмена найдите тесноту связи учебы Васечкина при первом и втором рубежном контроле, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод. Приведите конкретные примеры, в одном из которых коэффициент Спирмена был бы равен 1, а в другом примере он был бы равен 0.

165.Имеются оценки студента Васечкина за первый курс обучения по пяти дисциплинам, и его ответы на вопрос, сколько часов он потратил на подготовку к каждому экзамену:

Дисциплина	1	2	3	4	5
Оценки	8	5	4	9	7
Количествочасов	10	6	6	15	12

На основе рангового коэффициента Спирмена найдите тесноту связи между результатом экзамена и количеством потраченных на подготовку к экзамену часов, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод. Приведите конкретные примеры, в одном из которых коэффициент Спирмена был бы равен 1, а в другом примере он был бы равен 0.

166.Семь преподавателей школы проходят аттестацию по результатам теста и на основе оценки, поставленной им директором школы. Результаты их аттестации таковы:

Преподаватели	1	2	3	4	5	6	7
Результатытеста	7	2	5	4	3	8	9
Оценкадиректора	6	3	5	7	9	10	8

На основе рангового коэффициента Спирмена найдите тесноту связи между двумя показателями деятельности преподавателей, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод. Приведите конкретные примеры, в одном из которых коэффициент Спирмена был бы равен 1, а в другом примере он был бы равен 0.

167.Студент Васечкин имеет при первом и втором рубежном контроле обучения по семи дисциплинам следующие оценки:

Дисциплина	1	2	3	4	5	6	7
КР1	9	5	3	8	7	6	7
КР2	10	6	5	9	6	7	8

На основе рангового коэффициента Спирмена найдите тесноту связи между результатами обучения при первом и втором рубежном контроле, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод. Приведите конкретные примеры, в одном из которых коэффициент Спирмена был бы равен 1, а в другом 0.

168.Вам известны данные о пяти магазинах, причем вы знаете количество обслуживающего персонала в этих магазинахX оценку качества обслуживания Y (быстрота, приветливость, знание ассортимента и т.д.) по двадцатибальной шкале (20-это максимум):

№ магазина	1	2	3	4	5
X	10	8	13	6	7
Y	15	17	10	7	20

Найдите тесноту связи между этими данными, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод о том, насколько качество обслуживания связано с количеством обслуживающих.

169.Студентка Лисичкина имеет следующие экзаменационные оценки по пяти дисциплинам. На вопрос, сколько часов она потратила на подготовку к каждому экзамену, были получены ответы:

Дисциплина	1	2	3	4	5
Оценки	7	5	6	9	7
Количество часов	10	7	8	12	6

На основе ранговых коэффициентов Спирмена и Кендалла найдите тесноту связи между результатом экзамена и количеством потраченных на подготовку к экзамену часов, рассматривая их, как выборочные наблюдения случайных величин. Показать согласованность в значениях ранговых коэффициентов.

Найдите тесноту связи между этими данными, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод о том, насколько близки или далеки мужчины и женщины в оценках качеств партнеров.

Найдите тесноту связи между этими данными, рассматривая их, как выборочные наблюдения случайных величин. Сделайте вывод о том, насколько затраты на создание фильма влияют на его успех в прокате.

170. Методом наибольшего правдоподобия найдите оценку параметра λ показательного закона распределения времени между авариями, если известно, за 5 месяцев работы сборочной автоматизированной линии получены такие данные по количеству аварий за каждый месяц работы: 3, 4, 1, 0, 2. Найдите вероятность того, что за шестой месяц произойдет 5 аварий.

171. Методом наибольшего правдоподобия найдите оценку вероятности наступления события в биномиальном законе распределения, если известно, что в девяти независимых испытаниях событие наступило ровно 4 раза. Найдите вероятность того, что событие произойдет менее двух раз.

172. Методом наибольшего правдоподобия найдите оценку параметра λ показательного закона распределения времени между авариями, если известно, за 6 месяцев работы сборочной автоматизированной линии получены такие данные по количеству аварий за каждый месяц работы: 2, 3, 2, 1, 4, 1. Найдите вероятность того, что за седьмой месяц произойдет 6 аварий.

173. Методом наибольшего правдоподобия найдите оценку вероятности наступления события в биномиальном законе распределения, если известно, что в девяти независимых испытаниях событие наступило ровно 4 раза. Найдите вероятность того, что событие произойдет менее двух раз.

174. Методом наибольшего правдоподобия найдите оценку параметра λ показательного закона распределения времени между авариями, если известно, за 6 месяцев работы сборочной автоматизированной линии получены такие данные по количеству аварий за каждый месяц работы: 2, 3, 2, 1, 4, 1. Найдите вероятность того, что за седьмой месяц произойдет 6 аварий.

175. Исследуется количество заявок на товар, поступивших на предприятие в течение недели. Предполагается, что количество поступивших заявок починено закону Пуассона. Известно, что число полученных за предыдущие три недели заявок составило, соответственно, 5, 4, 5. Методом наибольшего правдоподобия вывести формулу для оценки параметра λ закона Пуассона на основе результатов выборки. Применив полученную оценку, вычислить вероятность того, что наследующей неделе на предприятие поступит 3 заявки.

176. Трижды проводится эксперимент по одновременному подбрасыванию трех одинаковых монеток (три серии, состоящие из трех бросков). «Орел» выпал три раза в первом эксперименте, два раза во втором эксперименте и два раза в третьем эксперименте. Методом наибольшего правдоподобия найдите оценку вероятности выпадения «Орла» при однократном бросании одной монеты. (задача предложена Аносовой О.Д).

177. Трижды проводится эксперимент по бросанию трех одинаковых монеток (три серии, состоящие из трех бросков). «Орел» не выпал ни разу в первом эксперименте, три раза во втором эксперименте и два раза в третьем. Методом наибольшего правдоподобия оценить вероятность выпадения «Орла» при однократном бросании одной монетки. Как еще можно оценить эту вероятность? Сравните результаты.

178. Методом наибольшего правдоподобия найдите оценку вероятности наступления события в биномиальном законе распределения, если известно, что в девяти независимых испытаниях событие наступило ровно 4 раза. Найдите вероятность того, что событие произойдет менее двух раз.

179. По исследованиям компании Мобильные Телефоны Средиземья среднее число проданных телефонов в первый день в каждом из офисов компании равнялось 15 (участвовали в опросе 10 торговых точек), а во второй день – 17 (участвовали 19 торговых точек). Определить наиболее доброкачественную точечную оценку числа ежедневных продаж телефонов. Какими свойствами должна обладать доброкачественная оценка? (Задача предоставлена Аносовой О.Д.)

180. По исследованиям компании «Тише едешь – дальшебудешь» среднее число проданных билетов за месяц в двенадцати офисах компании было таким: десять в пяти офисах и восемь в оставшихся офисах. Определить наиболее доброкачественную точечную оценку среднего числа ежемесячно продаваемых билетов в одном офисе. Какими свойствами должна обладать доброкачественная оценка?

ЗАКЛЮЧЕНИЕ

В настоящее время возникают много задач из теории и практики носящие стохастический (вероятностный) характер. Для успешного использования необходимо знание основ дисциплин, в которых с помощью математичексих моделей можно описать вероятностное пространство событий и подсчитать вероятности событий в дискретном и непрерывном случае.

Большой класс задач, рассмотреных в данном учебном пособии даст возможность студентам расширить свой научный кругозор. Поднять на более высокий уровень степень восприятия и взаимосязь математических дисцплин, в частности математичексого анализа, линейной алгебры и других.

Представленное учебное пособие поможет студентам освоить основные фундаментальные понятия и законы дисциплины «Теория вероятностей и математическая статистика».

На основе вышесказанного в учебном пособии были представлены основные математические модели дисциплины «Теория вероятностей и математическая статистика. Согласно содержанию и в логической связи между темами показана тесная вязь и обоснование фундаментальных понятий теории вероятностей и математической статистики.

В ходе изложения материала учитывалась последовательность изложения тем, логическая преемственность и завершенность. Для закрепления и усвоения тем были даны задачи с подробными решениями, а также по двум основным разделам – теории вероятностей и математической статистике был предложен большой массив задач для самостоятельной работы.

Необходимо отметить, что в средней школе вводятся элементы теории вероятностей и математической статистики. Именно по этой причине представленное учебное пособие послужит материалом для успешной сдачи контрольных и промежуточных знаний.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

Основная литература

1. Боровков А.А. Теория вероятностей.– М.: Наука, 1976г.
2. Вентцель Е.С. Теория вероятностей. – М.: Наука, 1964 г.
3. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2001 г.
4. Гнеденко Б.В. Курс теории вероятностей.– М.: Наука, 1971 г.
5. Крамер Г. Математические методы статистики. – М.: Мир, 1975 г.
6. Севастьянов Б.А. Курс теории вероятностей и математической статистики.– М.: Наука, 1982 г.

Дополнительная литература

1. Виленкин Н.Я., Потапов В.Г. Задачник-практикум по теории вероятностей с элементами комбинаторики и математической статистики. – М.: Просвещение, 1979 г.
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 1979 г.
3. Гусак А.А. Высшая математика. Часть 2. – Минск: Тетра Системе, 2003 г.
4. Коваленко И.Н., Филиппова А.А. Теория вероятностей и математическая статистика. – М.: Наука, 1982 г.
5. Козлов М.В. Элементы теории вероятностей в примерах и задачах. – М.: Высшая школа, 1990 г.
6. Королев В.С., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике. – М.: Наука, 1985 г.
7. Пугачёв В.С. Теория вероятностей и математическая статистика. – М.: Наука, 1979 г.
8. Сборник задач по математике. Специальные курсы. Под ред. Ефимова А.В. – М.: Наука, 1984 г.
9. Сборник задач по теории вероятностей и математической статистике. Под ред. Свешникова А.А. – М.: Наука, 1965 г.
10. Чистяков В.П. Курс теории вероятностей. – М.: Наука, 1982 г.
11. Чудесенко В.Ф. Сборник заданий по специальным курсам, высшей математики. – М.: Высшая школа, 1983 г.