

4. ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

4.1. Введение

В главах 2-3 были рассмотрены эмпирические распределения и методы вычисления их числовых характеристик. Но обработка экспериментальных данных не ограничивается рассмотренными методами. Обычно исследователь, получив данные эксперимента на одной или нескольких группах испытуемых и определив по ним некоторые обобщающие числовые характеристики (среднее, стандартное отклонение и др.), пытается найти ответ на следующие вопросы: насколько точно полученные результаты можно обобщить для более широкой совокупности (например, на всех жителей данного возраста)? Как хорошо его данные согласуются с данными других исследователей? Насколько достоверно различие экспериментальных данных, полученных в разных группах испытуемых или в одной и той же группе, но в разные промежутки времени? Существует ли связь между различными признаками, изучаемыми в проводимом исследовании, и если да, то насколько она сильна?

В ряде случаев исследователь пытается установить такую экспериментальную зависимость между изучаемыми признаками, чтобы по значениям одного из них, легко поддающегося измерению, установить значение другого, измерить который трудно или невозможно.

Конечно, в зависимости от целей конкретного исследования задачи могут быть различными и не ограничиваются приведенным перечнем.

Методы математической статистики, с помощью которых можно получить ответы на поставленные выше вопросы, рассматриваются в следующих главах. Чаще всего эти методы основаны на использовании тех или иных согласующихся с условиями проводимого эксперимента математических моделей, разработанных теорией вероятностей.

В данной главе рассматриваются некоторые ее элементарные, положения в том минимальном объеме, который необходим для дальнейшего изложения.

4.2. Определение вероятности

4.2.1. Испытание, событие, случайная величина

Под *испытанием* (*опытом*) в теории вероятностей принято понимать наблюдение какого-либо явления при соблюдении определенного комплекса условий, который должен каждый раз строго выполняться при повторении данного испытания. Если то же самое явление наблюдается при другом комплексе условий, то это уже другое испытание.

Когда речь идет о соблюдении комплекса условий данного испытания, имеется в виду постоянство значений всех факторов, контролируемых в данном испытании. Но при этом, как правило, имеет место большое число неконтролируемых факторов, которые трудно или невозможно учесть.

Результаты испытаний можно охарактеризовать качественно и количественно.

Качественная характеристика заключается в регистрации какого-либо явления, которое может наблюдаться или не наблюдаться при данном испытании. Любое из этих явлений называется в теории вероятностей *событием*.

События делятся на:

невозможные (в результате опыта никогда не произойдут),	достоверные (в результате опыта происходят всегда),	случайные (в результате опыта событие может произойти или не произойти).
---	---	--

Теория вероятностей рассматривает именно случайные события. При этом предполагается, что испытание может быть повторено неограниченное (по крайней мере, теоретически) число раз. Например, выполнение штрафного броска в баскетболе есть испытание, а попадание в кольцо — событие.

Другим примером события, часто приводимым в учебниках по теории вероятностей, является выпадение определенного числа очков (от 1 до 6) при бросании игральной кости.

События в теории вероятностей принято обозначать начальными прописными латинскими буквами A, B, C, ...

Случайные события называются *несовместными* если появление одного исключает появление другого. В противном случае они называются *совместными*.

Если в результате опыта произойдет хоть одно из некой группы событий, то они образуют *полную группу*. Появление хотя бы одного события из полной группы – достоверное событие.

Если, по условиям испытания нет никаких оснований предполагать, что один из исходов появляется чаще других, то все исходы являются *равновозможными*.

Два события называются *независимыми*, если появление одного из них не изменяет вероятности другого.

Количественная характеристика испытания состоит в определении значений некоторых величин, которыми интересуются при данном испытании (например, число подтягиваний на перекладине или время на беговой дистанции). В силу действия большого числа неконтролируемых факторов эти величины могут принимать различные значения в результате испытания. Причем до испытания невозможно предсказать значение величины, поэтому она называется *случайной величиной*.

4.2.2. Вероятность событий

Вероятность какого либо события – численное выражение возможности его наступления.

В некоторых простейших случаях вероятности событий могут быть легко определены непосредственно исходя из условий испытаний.

Представим себе общую схему таких испытаний.

Пусть испытание имеет n возможных несовместных исходов, т. е. отдельных событий, могущих появиться в результате данного испытания; причем при каждом повторении испытания возможен один и только один из этих исходов. Кроме того, пусть по условиям испытания, нет никаких оснований предполагать, что один из исходов появляется чаще других, т. е. все исходы являются равновозможными.

Допустим теперь, что при n равновозможных несовместных исходах интерес представляет некоторое событие A , появляющееся при каждом из m исходов и не появляющееся при остальных $n-m$ исходах. Тогда принято говорить, что в данном испытании имеется n случаев, из которых m благоприятствуют появлению события A .

Вероятность события A равна отношению числа исходов, благоприятствующих событию A , к общему числу всех равновозможных несовместных исходов опыта:

$$P(A) = \frac{m}{n}. \quad (4.1)$$

Формула (4.1) представляет собой так называемое *классическое определение вероятности* по Лапласу, пришедшее из области азартных игр, где теория вероятностей применялась для определения перспективы выигрыша.

Статистическое определение вероятности.

Будем фиксировать число испытаний, в результате которых появилось некоторое событие A . Пусть было проведено N испытаний, в результате которых событие A появилось ровно n_N раз. Тогда число n_N называется частотой события, а отношение $\frac{n_N}{N}$ — частотой (относительной частотой) события.

Замечательным экспериментальным фактом является то, что частота события при большом числе повторений испытания начинает мало изменяться и стабилизируется около некоторого определенного значения, в то время как при малом числе повторений она принимает различные, совершенно случайные значения. Поэтому интуитивно ясно, что если при неограниченном повторении испытания частота события будет стремиться к вполне определенному числовому значению, то это значение можно принять и качестве объективной характеристики события A . Такое число $P(A)$, связанное с событием A , называется вероятностью события A .

Математически неограниченное число повторений испытания записывается в виде предела (*lim*) при N , стремящемся к бесконечности (∞):

$$P(A) = \lim_{N \rightarrow \infty} \frac{n_N}{N}.$$

Поскольку n_N никогда не может превзойти N , то вероятность оказывается заключенной в интервале $0 \leq P(A) \leq 1$.

Следует отметить, что приведенное определение вероятности является абстрактным, оно не может быть экспериментально проверено, так как на практике нельзя реализовать бесконечно большое число повторений испытания.

Пусть проводятся независимые испытания, при каждом из которых вероятность события A неизменна. Справедливо утверждение, называемое *законом больших чисел* или *теоремой Бернулли*: если N достаточно

велико, то с вероятностью сколь угодно близкой к единице, отличие $\frac{n_N}{N}$ от $P(A)$ меньше любого наперед заданного положительного числа или, в символьной записи, $P\left(\left|\frac{n_N}{N} - P(A)\right| > \varepsilon\right) = 1$. Т.е. много раз бросая монету, мы “почти наверняка” будем получать примерно равные частоты выпадения герба и цифры.

4.3. Действия над событиями

В этом разделе приводятся основные правила операций над событиями с использованием для наглядности их графического изображения в виде диаграмм.

Вначале введем понятие “поле событий” как совокупности всех случайных событий данного испытания, для которых определены вероятности. На рис. 4.1 поле событий изображено в виде заштрихованного прямоугольника.

1. Сумма (объединение) событий (рис. 4.2) представляет собой сложное событие, состоящее в появлении хотя бы одного из событий A и B . Объединение событий обозначается как $A \cup B$, или $A + B$.

2. Произведением (пересечением) событий A и B называется их совместное появление (рис. 4.3). Обозначается произведение событий как $A \cap B$, или $A \bullet B$.

3. Достоверным событием называется событие, которое обязательно происходит в результате данного испытания (рис. 4.4). Оно обозначается обычно как E .

4. Невозможное событие – событие, которое не может произойти в результате данного испытания. Принятое обозначение – $\emptyset$.

5. Несовместными называются события, которые в результате данного испытания не могут произойти вместе (рис. 4.5). Примеры несовместных событий: попадание и промах при выстреле, выпадение двух и трех очков при бросании игральной кости. Рис. 4.5 наглядно показывает, что для несовместных событий $A \bullet B = \emptyset$.

6. Противоположным к A событием называется событие, состоящее в не появлении события A (рис. 4.6). Обозначается противоположное событие символом $\bar{A}$. Примеры противоположных событий: промах и попадание при выстреле, выпадение герба или цифры при одном подбрасывании монеты.



Рис. 4.1. Поле событий

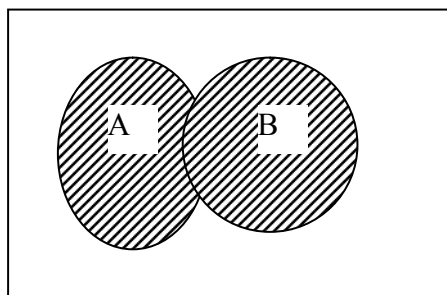


Рис. 4.2. Сумма событий

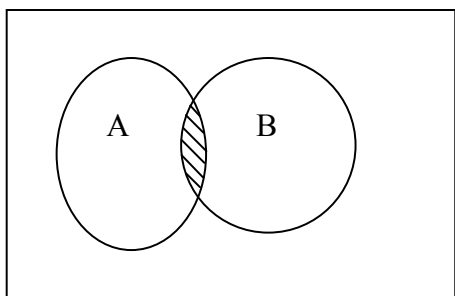


Рис. 4.3. Произведение событий



Рис. 4.4. Достоверное событие

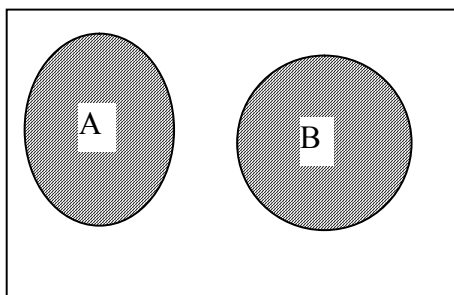


Рис. 4.5. Несовместные события

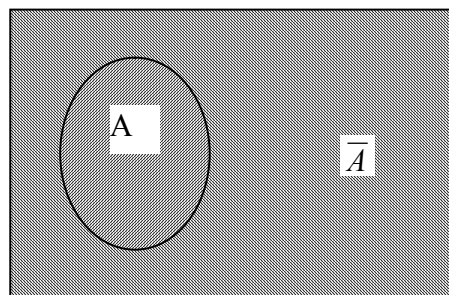


Рис. 4.6. Противоположные события

4.4. Исчисление вероятностей

4.4.1. Примеры непосредственного определения вероятностей

Рассмотрим несколько примеров на вычисление вероятностей по формуле (4.1).

Пример 4.1

Испытание состоит в подбрасывании игральной кости, на каждой из граней которой проставлено число очков (от 1 до 6). Какова вероятность того, что: 1) выпадает 2 очка? 2) выпадает нечетное число очков?

Решение 1: В данном испытании имеется 6 равновероятных случаев (выпадение 1, 2, 3, 4, 5, 6 очков), так как нет оснований предполагать, что появление какого-то определенного числа очков более вероятно (если, конечно, кость симметрична). Поэтому вероятность выпадения любого числа очков, в том числе и 2, при одном подбрасывании равна $1/6$.

Событию A , заключающемуся в появлении нечетного числа очков, благоприятствуют три случая (выпадение 1, 3 и 5), поэтому по формуле (4.1) получаем

$$P(A) = \frac{3}{6} = 0,5.$$

Решение 2: В данном испытании имеется 2 равновероятных исхода (выпадение четного числа очков (т.е. 2, 4, 6) и нечетного), так как кость симметрична, то очевидно, что эти исходы равновероятные.

Событию A , заключающемуся в появлении нечетного числа очков, благоприятствуют 1 случай из двух, поэтому по формуле (4.1) получаем $P(A) = 1/2 = 50\%$.

Отметим, что построенную таким образом пространство элементарных событий непригодно для расчета вероятности того, что выпадает 2 очка, так как этому событию не благоприятствует не один из введенных нами элементарных исходов.

Пример 4.2

В урне 5 белых и 10 черных шаров, не отличающихся по размеру. Шары тщательно перемешивают и затем наугад вынимают 1 шар. Какова вероятность того, что вынутый шар окажется белым?

Решение. В этом примере имеется 15 равновероятных (шары не отличаются по размеру) исходов опыта, причем событию A (появлению белого шара) благоприятствуют 5 из них, поэтому искомая вероятность составит $5/15 = 33.33\%$.

4.4.2. Основные правила вычисления вероятностей сложных событий

Ниже приведены основные правила, позволяющие определить вероятность появления сложного события на основании известных вероятностей составляющих его более простых событий.

1. Вероятность достоверного события равна единице:

$$P(E) = 1. \quad (4.2)$$

2. Вероятность объединения (суммы) несовместных событий равна сумме их вероятностей:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) \quad (4.3)$$

Эти два равенства являются аксиомами теории вероятностей, т. е. принимаются в качестве исходных, но требующих доказательства свойств вероятностей. На их основе строится вся теория вероятностей.

Все остальные, приведенные ниже без доказательств формулы могут быть выведены из принятых аксиом.

3. Вероятность невозможного события равна нулю:

$$P(\emptyset) = 0. \quad (4.4)$$

4. Вероятность события, противоположного событию A , равна

$$P(\bar{A}) = 1 - P(A). \quad (4.5)$$

Формула (4.5) оказывается полезной на практике в тех случаях, когда вычисление вероятности непосредственно события A затруднительно, в то время как вероятность противоположного события находится просто (см. ниже п.9).

5. Теорема сложения вероятностей. Вероятность объединения произвольных событий равна сумме их вероятностей за вычетом вероятности произведения событий:

$$P(A \cup B) = P(A) + P(B) - P(AB). \quad (4.6)$$

Для несовместных событий $P(AB) = 0$ и формула (4.6) переходит в (4.3).

6. Условная вероятность. Если требуется найти вероятность события B при условии, что произошло некоторое другое событие A , то такую ситуацию характеризуют с помощью условной вероятности $P(B|A)$. Условная вероятность равна отношению вероятности произведения событий A и B к вероятности события A :

$$P(B|A) = \frac{P(AB)}{P(A)} \quad (4.7)$$

В тех случаях, когда события A и B несовместны, $P(AB) = 0$ и соответственно $P(B|A) = 0$.

7. Определение условной вероятности в виде (4.7) дает возможность записать следующую формулу для вычисления вероятности произведения событий (*теорема умножения вероятностей*)

$$P(AB) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \quad (4.8)$$

8. Поскольку вероятность события A (или B) для независимых событий по определению не изменяется при появлении другого события, то условная вероятность $P(A|B)$ совпадает с вероятностью события A , а условная вероятность $P(B|A)$ — с $P(B)$. Вероятности $P(A)$ и $P(B)$ в отличие от условных вероятностей называются безусловными.

$$P(A|B) = P(A), \quad P(B|A) = P(B), \quad (4.9)$$

Теорема умножения вероятностей для независимых событий записывается следующим образом:

$$P(A_1 \cdot A_2 \cdot \dots \cdot A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n), \quad (4.10)$$

т. е. вероятность произведения независимых событий равна произведению их вероятностей.

9. Вычислим вероятность появления хотя бы одного события в n испытаниях

A — появление в n испытаниях **хотя бы** один раз интересующего нас события.

$\bar{A}$ — интересующее нас событие не появилось в n испытаниях **ни разу**.

A_1 — интересующее нас событие появилось в первом испытании.

A_2 — интересующее нас событие появилось во втором испытании.

....

A_n — интересующее нас событие появилось в n -ом испытании.

$$P(A) = 1 - P(\bar{A}) = 1 - P(\bar{A}_1 \cdot \bar{A}_2 \cdot \dots \cdot \bar{A}_n) = 1 - P(\bar{A}_1) \cdot P(\bar{A}_2) \cdot \dots \cdot P(\bar{A}_n) \quad (4.11)$$

10. Формула полной вероятности.

Если событие A может произойти только при появлении одного из несовместных событий $H_1, H_2, \dots, H_n$, то

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \dots + P(H_n)P(A|H_n). \quad (4.12)$$

Пример 4.3

В урне 5 белых, 20 красных и 10 черных шаров, не отличающихся по размеру. Шары тщательно перемешивают и затем наугад вынимают 1 шар. Какова вероятность того, что вынутый шар окажется белым или черным?

Решение. Пусть событие A — появление белого или черного шара. Разобьем это событие на более простые. Пусть B_1 — появление белого шара, а B_2 — черного. Тогда, $A = B_1 + B_2$ по определению суммы событий. Следовательно $P(A) = P(B_1 + B_2)$. Так как B_1 и B_2 — несовместные события, то по теореме о вероятности суммы несовместных событий (формула 4.3) $P(B_1 + B_2) = P(B_1) + P(B_2)$.

Вычислим вероятности событий B_1 и B_2 . В этом примере имеется 35 равновозможных (шары не отличаются по размеру) исходов опыта, событию B_1 (появлению белого шара) благоприятствуют 5 из них, поэтому $P(B_1) = \frac{5}{35}$. Аналогично, $P(B_2) = \frac{10}{35}$. Следовательно, $P(A) = \frac{5}{35} + \frac{10}{35} = \frac{15}{35}$.

Пример 4.4

Ведутся поиски двух преступников. Каждый из них независимо от другого может быть обнаружен в 2006/2007 учебный год

течение суток с вероятностью 0,5. Какова вероятность того, что в течение суток будет обнаружен хотя бы один преступник?

Решение. Пусть событие A – “обнаружен хотя бы один преступник”. Разобьем это событие на более простые. Пусть B_1 – обнаружен первый преступник, а B_2 – обнаружен второй преступник. Тогда, $A=B_1+B_2$ по определению суммы событий. Следовательно $P(A)=P(B_1+B_2)$. Так как B_1 и B_2 – совместные события, то по теореме о вероятности суммы событий (формула 4.6)

$$P(B_1+B_2) = P(B_1)+P(B_2)-P(B_1 B_2) = 0,5+0,5 - 0,25=0,75.$$

Можно решать и через обратное событие: $P(A)=1-P(\bar{A})=1-P(\bar{B}_1 \cdot \bar{B}_2)=1-\frac{1}{2} \cdot \frac{1}{2}=1-\frac{1}{4}=\frac{3}{4}$.

Пример 4.5 а)

Преступник имеет 3 ключа. В темноте он открывает дверь выбирая ключ случайным образом. На открытие каждой из дверей он тратит 5 сек. Найти вероятность того, что он откроет все двери за 15 сек.

Решение. Пусть событие A – “открыты все двери”. Разобьем это событие на более простые. Пусть B – “открыта 1-я”, C – “открыта 2-я”, а D – “открыта 3-я”. Тогда, $A=BCD$ по определению произведения событий. Следовательно $P(A)=P(BCD)$. По теореме о вероятности произведения независимых событий (формула 4.10) $P(BCD) = P(B)P(C)P(D)$.

Вычислим вероятности событий B , C и D . В этом примере имеется 3 равновозможных (каждый ключ выбираем из трех) исходов опыта. Каждому из событий B , C и D благоприятствует 1 из них, поэтому

$$P(B)=P(C)=P(D)=\frac{1}{3}. \quad P(A)=\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}.$$

Пример 4.5 б)

Изменим задачу: считаем, что преступник – забывчивый человек. Пусть преступник открыв дверь, оставляет ключ в ней. Какова тогда вероятность, что он откроет все двери за 15 сек?

Решение. Событие A – “открыты все двери”. Опять, $A=BCD$ по определению произведения событий. Следовательно $P(A)=P(BCD)$. Но, теперь события B , C и D – зависимы. По теореме о вероятности произведения зависимых событий $P(BCD) = P(B)P(C|B)P(D|BC)$.

Вычислим вероятности: $P(B)=\frac{1}{3}$, $P(C|B)=\frac{1}{2}$ (ключа осталось только два и один из них подходит!),

$$P(D|BC)=\frac{1}{1} \text{ и, значит, } P(A)=\frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{6}.$$

Пример 4.6

Ведутся поиски двух преступников. Каждый из них независимо от другого может быть обнаружен в течение суток с вероятностью 0,5. После поимки одно из них, в связи с увеличением количества сотрудников, занятых в поисках, вероятность найти второго возрастает до 0,7. Какова вероятность того, что в течение суток будут обнаружены оба преступника.

Решение. Пусть событие A – “обнаружены два преступника”. Разобьем это событие на более простые. Пусть B_1 – обнаружен первый преступник, а B_2 – обнаружен второй преступник, после того, как пойман первый. Тогда, $A=B_1B_2$ по определению произведения событий. Следовательно $P(A)=P(B_1B_2)$. Так как B_1 и B_2 – зависимые события, то по теореме о вероятности произведения зависимых событий (формула 4.8) $P(B_1B_2) = P(B_1)P(B_2|B_1) = 0,5 \cdot 0,7=0,35$.

Пример 4.7

Найти вероятность того, что при подбрасывании монеты 10 раз герб выпадет хотя бы 1 раз.

Решение. Пусть событие A – “герб выпадет хотя бы 1 раз”. Рассмотрим обратное событие: $\bar{A}$ – “герб не выпадет ни разу”. Очевидно, что обратное событие легче чем исходное разбить на более простые. Пусть A_1 – герб не выпал при первом броске, A_2 – герб не выпал при втором броске, ... A_{10} – герб не выпал при 10-м броске. Все события $A_1 \dots A_{10}$ независимы, следовательно, (формула 4.11)

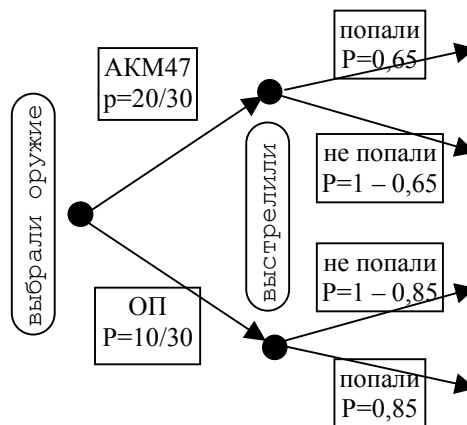
$$P(A)=1-P(\bar{A})=1-P(\bar{A}_1 \cdot \bar{A}_2 \cdot \dots \cdot \bar{A}_{10})=1-P(\bar{A}_1) \cdot P(\bar{A}_2) \cdot \dots \cdot P(\bar{A}_{10})=1-\frac{1}{2} \cdot \frac{1}{2} \cdot \dots \cdot \frac{1}{2}=1-\frac{1}{1024}=\frac{1023}{1024}.$$

Пример 4.8

В проведении операции по освобождению заложников участвуют 2 группы снайперов: 10 человек с винтовкой ОП21 и 20 человек с АКМ47. Вероятность поражения из ОП21 – 0,85, а АКМ47 – 0,65. Найти вероятность того, что при одном выстреле произвольного снайпера преступник будет поражен.

Решение. Пусть событие A – “преступник поражен”. Разобьем это событие на более простые. Преступник может быть поражен либо из ОП21, либо из АКМ47. Вероятность того, что произвольный снайпер вооружен ОП21 (событие H_1) равна 10/30. Вероятность того, что произвольный снайпер вооружен АКМ47 (событие H_2) равна 20/30.

В подобных задачах полезно изобразить “дерево возможностей”:



Вероятность того, что преступник поражен равна (формула 4.12)

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = \frac{10}{30} \cdot 0,85 + \frac{20}{30} \cdot 0,65.$$

4.4.3. Комбинаторика

Правило сложения. Если первое действие можно выполнить n различными способами, а второе — m способами, то выполнить первое **ИЛИ** второе действие можно $m + n$ способами.

Правило умножения. Если первое действие можно выполнить n различными способами, а второе — m способами, то выполнить первое **И** второе действие (в таком порядке) можно nm способами.

Эти правила можно обобщить на случай 3-х, 4-х и более действий.

Пример 4.9. Рекламный плакат мебельной фабрики утверждает, что возможно составить 100000 различных вариантов расстановки производимых ей шкафов если купить хотя бы 5 шкафов. Верно ли это, если выпускается всего 10 различных типов шкафов?

Решение. Вычислим, сколькими способами можно расставить 5 шкафов рядом друг с другом. Первую позицию можно заполнить 10-ю различными способами (принцип сложения), вторую — также 10-ю (никто не мешает купить и второй шкаф той же модели, что и первый), третью — опять 10-ю и т.д. Вообще имеем (принцип умножения) $10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 100000$ различных вариантов расстановки пяти шкафов рядом друг с другом. Если же купить шкафов больше, чем 5, то, очевидно, вариантов расстановки будет еще больше.

Вывод: реклама является добросовестной.

Пример 4.10. Из тщательно перемешанной колоды в 52 карты вытягивают 3 карты. Сколько существует различных вариантов карт на руках у игрока?

Решение. В данном опыте производится 3 действия: вытягивание 1-й карты, 2-й карты и 3-й карты.

Вычислим, сколькими способами можно вытянуть 1-ую карту. Так как всего в колоде 52 карты, то имеем 52 различных способа. (Здесь мы применили принцип сложения: карта может быть двойка пик **ИЛИ** тройка пик **ИЛИ** ... **ИЛИ** туз червей. Значит, всего имеем $1+1+\dots+1=52$ способа.)

Вычислим, сколькими способами можно вытянуть 2-ую карту. Так как в колоде осталось 51 карта, то, значит, второе действие можно выполнить 51-м способом.

Аналогично рассуждая, находим, что 3-е действие можно осуществить 50-ю способами.

Всего различных вариантов **расположения** карт на руках у игрока будет $52 \cdot 51 \cdot 50 = 132600$ способов. Для ответа осталось разделить это число на $3 \cdot 2 \cdot 1$ — это кол-во способов перетасовать эти 3 розданные карты.

Ответ: 22100.

В общем виде:

Если стоит задача вычислить сколькими способами можно расположить “в ряд” (т.е. важен порядок их следования) вытянутые m предметов из коробки содержащей различных n предметов, то имеем так называемую ситуацию “перестановок”.

Вычислим это количество: первую позицию можно заполнить n способами, вторую — $n - 1$ способом, третью — $n - 2$ способом, и т.д. Искомое количество способов заполнить **все n позиций** равно (по принципу умножения)

$$n(n-1)(n-2)(n-3)\dots(n-m+1)$$

Если стоит задача вычислить сколькими способами можно расположить “в ряд” (т.е. важен порядок их следования) вытянутые m предметов из коробки содержащей различных n предметов, то имеем так называемую ситуацию “перестановок”.

Вычислим это количество: первую позицию можно заполнить n способами, вторую – $n - 1$ способом, третью – $n - 2$ способом, и т.д. Искомое количество способов равно (по принципу умножения)

$$A_n^m = n(n-1)(n-2)(n-3)\dots(n-m+1).$$

Но, поскольку нам не важно какой именно элемент стоит на каком месте, то необходимо A_n^m разделить на количество способов по разному переставлять уже выбранные элементы. А это количество равно $A_n^m = n(n-1)(n-2)(n-3)\dots 3 \cdot 2 \cdot 1 = n!$ (читается n факториал).

Искомое количество способов заполнить **все n позиций** равно $A_n^m / n!$ и обозначается C_n^m .

Пример 4.11. В совбесе ООН 11 членов: 5 постоянных и 6 так называемые “малые нации”. Для принятия решения, надо, чтобы было 7 голосов “ЗА”, причем следующим образом: все постоянные + как минимум 2 временных. Сколько всего вариантов голосования? Сколько всего можно организовать выигрышных коалиций? (Выигрышной коалицией называется такая, когда как бы не голосовали противники решение все равно будет принято.)

Решение. Так, как голосуют 11 делегаций и у них есть 2 выбора (“за”, “против”), то по принципу умножения имеем $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^{11} = 2048$ – вариантов голосования. Так как все постоянные члены должны проголосовать “за”, то выигрышная коалиция определяется только временными членами, а кол-во – количеством способов выбрать 2 или 3 или 4 или 5 или 6 временных членов, голосующих “за”.

Имеем $1 \cdot \left(C_6^2 + C_6^3 + C_6^4 + C_6^5 + C_6^6 \right) = 15 + 20 + 15 + 6 + 1 = 57$ способов, причем 15 – число так называемых

минимальных выигрышных коалиций.

4.4.4. Схема Бернулли

Пусть производится n одинаковых независимых опытов. В каждом испытании некоторое событие A может произойти с вероятностью p (а, значит, не произойти с вероятностью $q = 1 - p$).

Вычислим вероятность того, что событие произойдет **ровно k раз** в проведенных n опытах.

$P_n(0) = \underbrace{q \cdot q \cdot \dots \cdot q}_n = q^n$ – вероятность того, что **во всех** опытах событие не произойдет (см. также пример

4.7)

$P_n(1) = \underbrace{p \cdot q \cdot \dots \cdot q}_n + \underbrace{q \cdot p \cdot \dots \cdot q}_n + \dots + \underbrace{q \cdot q \cdot \dots \cdot p}_n = n \cdot p \cdot q^{n-1}$ – вероятность того, что событие произойдет **ровно**

в одном опыте (из проведенных n опытов).

$P_n(2) = \frac{n(n-1)}{2} p^2 q^{n-2}$ – вероятность того, что событие произойдет **ровно 2 раза** в n опытах.

.....
 $P_n(k) = C_n^k p^k q^{n-k}$ – вероятность того, что событие произойдет **ровно k раз** в n опытах.

.....
 $P_n(n) = \underbrace{p \cdot p \cdot \dots \cdot p}_n = p^n$ вероятность того, что событие произойдет **во всех** опытах.

4.5. Случайные величины

Случайные величины (С.В.) – численное значение появляющееся в результате опыта, и принимающее произвольное значение из заранее определенного множества.

Существует два типа случайных величин: дискретные и непрерывные.

Дискретные случайные величины принимают в результате испытания одно из изолированного дискретного множества значений. Они хорошо подходят для описания результатов измерений, связанных с подсчетом и выражаемых целыми числами.

Примеры дискретных случайных величин: число зарегистрированных правонарушений, совершенных сделок, подтягиваний на перекладине и т. п.

Вероятность принятия дискретной случайной величиной каждого из возможных ее значений больше нуля. Эта вероятность может быть записана как

$$P(\{X = x_i\}) = p_i,$$

где $i = \dots -1, 0, 1 \dots$

Здесь X — обозначение случайной величины; x_i — конкретные числовые значения, принимаемые дискретной случайной величиной; p_i — вероятности этих значений.

Индекс i может в общем случае пробегать значения от $-\infty$ до ∞ .

Функция $P(\{X = x_i\})$, связывающая значения дискретной случайной величины с их вероятностями, называется ее **распределением (законом распределения)**. Обычно закон распределения записывается в виде таблицы вида

X	x_1	x_2	...	x_n	...	
P	p_1	p_2	...	p_n	...	

Пример 4.12. Пусть X — число очков выпавшее на игральной кости при одном броске. Тогда, эта с.в. распределена по закону

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

Непрерывные случайные величины в результате испытания могут принимать любые значения из некоторого интервала.

Примеры непрерывных случайных величин: спортивный результат в беге или прыжках, рост и масса тела человека, сила мышц и др.

Поскольку число возможных значений непрерывной случайной величины бесконечно велико и чаще всего нет оснований предположить, что одни значения появляются существенно чаще других, то вероятность принятия непрерывной случайной величиной каждого отдельного значения оказывается равной нулю. По этой причине нельзя описать распределение непрерывной случайной величины в виде вероятностей ее отдельных значений, как в случае дискретных случайных величин. Здесь необходимы другие подходы, которые рассмотрены в разделах 4.6 и 4.7.

4.6. Функция распределения

Рассмотрим вероятность того, что случайная величина X окажется меньше или равной некоторому заданному **числу** x , т. е.

$$F(x) = P(\{X \leq x\}). \quad (4.11)$$

Эта вероятность, рассматриваемая как функция переменной x , называется функцией распределения случайной величины X . Она используется для записи распределений как дискретных, так и непрерывных случайных величин. Для дискретной случайной величины функция распределения имеет скачки. Функция распределения непрерывной случайной величины будет непрерывной функцией (рис. 4.8).

Как было сказано ранее, вероятность принятия непрерывной случайной величиной какого-либо конкретного значения равна 0. Но для непрерывной случайной величины обычно и надо определить вероятность попадания ее в заданный интервал $[x_1; x_2]$, которая по известной функции распределения находится как

$$P(\{x_1 \leq X < x_2\}) = F(x_2) - F(x_1) \quad (4.12)$$

В этом выражении совершенно не обязательно записывать интервал таким образом. Можно было бы записать $x_1 < X < x_2$, $x_1 \leq X \leq x_2$ или $x_1 < X \leq x_2$, при этом вероятность попадания случайной величины в интервал не изменится. Это связано с тем, что, как уже отмечалось, функция распределения случайной непрерывной величины не имеет скачков ни при каких значениях x .

Свойства функции распределения совпадают со свойствами эмпирической функции распределения (см. 2.3.4)

1. $F(x)$ неубывающая функция.
2. $\lim_{x \rightarrow -\infty} F(x) = 0$
3. $\lim_{x \rightarrow +\infty} F(x) = 1$

График функция распределения представляет собой теоретический аналог полигона накопленных частот, рассмотренного в разделе 2.3.3.

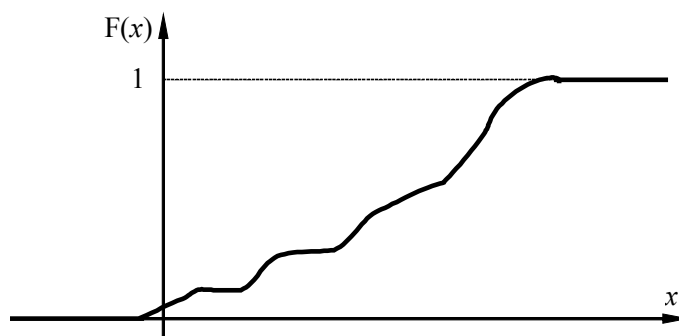


Рис. 4.8. Функция распределения непрерывной случайной величины

4.7. Плотность распределения вероятностей

Для непрерывных случайных величин вводится понятие плотность распределения вероятностей, или “плотность вероятностей”, играющее исключительно важную роль при их описании.

Плотность вероятностей — это производная от функции распределения непрерывной случайной величины, т.е.

$$f(x) = \frac{dF(x)}{dx} \quad (4.14)$$

Более подробно при рассмотрении конкретных непрерывных распределений об этой функции рассказано в разделе 4.9. Типичный вид графика плотности вероятностей показан на рис. 4.9.

Вероятность попадания непрерывной случайной величины в интервал между значениями x_1 и x_2 пропорциональна площади под кривой плотности вероятностей, заключенной между точками x_1 и x_2 . Эта вероятность математически записывается в виде интеграла от $f(x)$ в пределах x_1 и x_2 .

$$P(\{x_1 \leq X < x_2\}) = \int_{x_1}^{x_2} f(x) dx. \quad (4.15)$$

Свойства $f(x)$:

1. $f(x) \geq 0$;
2. $f(x) = 0$ при $x < x_{\min}$;
3. $f(x) = 0$ при $x > x_{\max}$;
4. $\int_{-\infty}^{+\infty} f(x) dx = 1$.

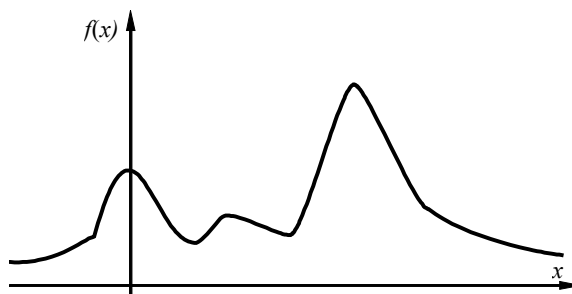


Рис. 4.9. Вид графика плотности распределения вероятностей

График плотности вероятностей является теоретическим аналогом полигона частот и гистограммы, (2.3.2 и 2.3.4).

4.8. Числовые характеристики случайных величин

Распределение случайной величины, заданное в виде функции распределения или плотности вероятностей, полностью ее характеризует. Однако такая исчерпывающая характеристика случайной величины сложна и далеко не всегда необходима. Для решения многих практических задач не нужно знать распределение случайной величины, а достаточно иметь лишь некоторые обобщающие числовые характеристики этого распределения.

4.8.1. Математическое ожидание

Для более наглядного определения математического ожидания (среднего значения) случайной величины рассмотрим подход к этому понятию на конкретном примере.

Пусть имеется дискретная случайная величина X с возможными значениями $x_1, x_2, \dots, x_k$ и вероятностями этих значений $p_1, p_2, \dots, p_k$. В качестве примера X рассмотрим случайную величину — количество правонарушений за сутки. Каждое из значений $x_1, x_2, \dots, x_n$ (отмечено 0; 1; ... нарушений) будет наблюдаться некоторое число раз. Обозначим эти числа через $n_1, n_2, \dots, n_k$. Очевидно, что сумма $n_1 + n_2 + \dots + n_k = n$.

Таким образом, имеем n наблюдений за случайной величиной X , т. е. выборку объема n . Определим выборочное среднее арифметическое: $\bar{x}_{Bn} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \sum_{i=1}^k x_i \frac{n_i}{n}$.

Здесь индекс n при $\bar{x}_B$ обозначает, что среднее арифметическое вычислено по n наблюдениям.

Теперь представим, что испытание, состоящее в регистрации количества правонарушений, повторяется неограниченное число раз. Здесь, абстрагируясь от физической реализуемости такого эксперимента, будем считать, что наблюдению доступна вся теоретически бесконечная генеральная совокупность значений случайной величины X .

Согласно статистическому определению вероятности, данному в разделе 4.2.2, относительные частоты событий стремятся к их вероятностям при неограниченном повторении испытания.

Поэтому в пределе при $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \bar{x}_{Bn} = \lim_{n \rightarrow \infty} \sum_{i=1}^k x_i \frac{n_i}{n} = \sum_{i=1}^k x_i \lim_{n \rightarrow \infty} \frac{n_i}{n} = \sum_{i=1}^k x_i p_i.$$

Таким образом, выборочное среднее арифметическое случайной величины X стремится при неограниченном повторении испытания (при неограниченном увеличении объема выборки) к некоторому постоянному числу, так как в последней сумме x_i и p_i — постоянные числа. Это число носит название математического ожидания (среднего значения) случайной величины.

Математическое ожидание обозначает как $M(X)$ или m_x .

Математическое ожидание дискретной случайной величины равно сумме всех ее возможных значений, умноженных на вероятности этих значений:

$$M(X) = m_x = \sum_i x_i P(X = x_i) = \sum_i x_i p_i. \quad (4.16)$$

В этой записи $\sum_i$ означает, что суммирование производится по всем возможным i .

Только что рассмотренный пример показывает, что математическое ожидание — абстрактное понятие. Оно является теоретическим аналогом выборочного среднего арифметического.

Математическое ожидание равно среднему значению генеральной совокупности.

Для непрерывных случайных величин математическое ожидание определяется с помощью плотности вероятностей по формуле:

$$M(X) = m_x = \int_{-\infty}^{\infty} x f(x) dx. \quad (4.17)$$

Свойства математического ожидания

1. $M(C) = C$,
2. $M(CX) = CM(X)$.
3. $M(X+Y) = M(X) + M(Y)$,
4. $M(X \cdot Y) = M(X)M(Y)$, если X и Y — независимые с.в.

4.8.2. Дисперсия и стандартное отклонение

Точно так же, как математическое ожидание, являющееся теоретическим аналогом среднего арифметического, можно ввести теоретические аналоги всех числовых характеристик выборки, рассмотренных в гл. 3. Для этого нужно в соответствующих формулах для выборочных характеристик

заменить все средние арифметические на математические ожидания.

Дисперсией случайной величины X называется математическое ожидание квадрата отклонений случайной величины от ее математического ожидания (сравните с определением п. 3.4.2). Дисперсия обозначается как $D(X)$, или σ^2 .

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

Для дискретных случайных величин

$$D(X) = \sum_i (x_i - M(X))^2 p_i, \quad (4.18)$$

т. е. дисперсия дискретной случайной величины равна сумме квадратов отклонений отдельных значений случайной величины от ее математического ожидания, умноженных на вероятности этих значений.

Для непрерывных случайных величин

$$D(X) = \int_{-\infty}^{\infty} (x - M(X))^2 f(x) dx. \quad (4.19)$$

Положительный корень квадратный из дисперсии называется средним квадратическим (стандартным) отклонением случайной величины.

Эта величина обозначается, как σ_x

$$\sigma_x = \sqrt{D}. \quad (4.20)$$

Дисперсия и стандартное отклонение характеризуют изменчивость (вариативность) случайной величины. Чем сильнее случайная величина отклоняется от своего математического ожидания, тем больше величины $D(X)$ и σ_x . Последнюю (σ_x) использовать удобнее, так как его размерность совпадает с размерностью случайной величины (Если С.В. X – количество выигранных в лотерею \$, то D измеряется в долларах “в квадрате”, что не имеет никакого прикладного смысла, а σ_x – это доллары!).

Свойства дисперсии

1. $D(C) = 0$,
2. $D(CX) = C^2 \cdot D(X)$.
3. $D(X + Y) = D(X) + D(Y)$, если X и Y – независимые с.в.

Обобщение 4.8.1 и 4.8.2.

Пусть $X_1, \dots, X_n$ – с.в., а $C_1, \dots, C_n$ – константы. Тогда $M(C_1 X_1 + \dots + C_n X_n) = C_1 M(X_1) + \dots + C_n M(X_n)$. В частности $M\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{M(X_1) + \dots + M(X_n)}{n}$. Если более того $X_1, \dots, X_n$ – независимые, то

$$D(C_1 X_1 + \dots + C_n X_n) = C_1^2 D(X_1) + \dots + C_n^2 D(X_n). \text{ В частности } D\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{D(X_1) + \dots + D(X_n)}{n^2}.$$

Пример 4.13.

Играем в следующую игру: один раз бросаем игральную кость и получаем столько \$, сколько выпало очков. Цена игры: 4\$. Выгодно ли играть?

Пусть с.в. X – количество очков, выпавшее при броске игральной кости. (см. пример 4.12)

Вычислим $M(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3,5$ – именно столько очков (а, значит, и \$) ”в

среднем” мы будем получать если играть достаточно долго. Значит, игра невыгодна для нас. Мы ”в среднем” теряем 0.5\$ в каждой игре. Для вычисления $D(X)$ обычно пользуются формулой

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

С.В. X^2 имеет следующее распределение

X^2	1	$2^2=4$	$3^2=9$	$4^2=16$	$5^2=25$	$6^2=36$
P	1/6	1/6	1/6	1/6	1/6	1/6

Вычислив $M(X^2)=91/6$ находим $D(X) = 91/6 - 441/36 = 105/36 = 2\frac{11}{12}$.

4.8.3. Моменты

Математическое ожидание и дисперсия представляют собой частные случаи общих числовых

характеристик случайной величины, называемых моментами.

Ниже кратко рассматриваются лишь так называемые центральные моменты случайной величины.

S -ым центральным моментом случайной величины X называется математическое ожидание s -й степени отклонения случайной величины от ее математического ожидания:

$$\mu_s = M((X - M(X))^s).$$

В частности, при $s = 2$ второй центральный момент случайной величины есть дисперсия.

На практике часто используются также третий и четвертый центральные моменты, позволяющие судить о симметричности и остроте вершины кривой распределения случайной величины.

Если $\mu_3 = 0$, то распределение симметрично относительно математического ожидания, если $\mu_3 > 0$, то преобладают положительные отклонения от математического ожидания, если $\mu_3 < 0$ — отрицательные. Для удобства применяется так называемый коэффициент асимметрии, который является безразмерной величиной и определяется как

$$\gamma_3 = \frac{\mu_3}{\sigma^3} \quad (4.21)$$

Об остроте вершины кривой распределения судят по коэффициенту эксцесса:

$$\gamma_4 = \frac{\mu_4}{\sigma^4} - 3 \quad (4.22)$$

Если $\gamma_4 > 0$, то распределение имеет острый пик, если $\gamma_4 < 0$ (минимальное значение $\gamma_4 = -2$), то распределение имеет плосковершинную форму по сравнению с рассмотренным ниже нормальным распределением, для которого $\gamma_4 = 0$.

4.9. Биномиальное распределение. Распределение Бернулли и Пуассона.

Пусть проводятся n испытаний по схеме Бернулли (см. 4.4.3.). Событие A может произойти в результате этой серии опытов 0 раз, 1 раз, ... n раз. Рассмотрим случайную величину — число испытаний в которых событие A произошло. Имеем дискретную с.в. с законом распределения

X	0	1	...	k	...	n
P	$P_n(0) = q^n$	$P_n(1) = n \cdot p \cdot q^{n-1}$	...	$P_n(k) = C_n^k p^k q^{n-k}$	...	$P_n(n) = p^n$

Также говорят, что с.в. X распределена по **биномиальному** закону с параметрами n и p и пишут $X \sim B(n; p)$.

Если $n = 1$, то говорят, что с.в. X имеет **распределение Бернулли** параметром p .

Теорема. Пусть $X_1, X_2, \dots, X_n$ — независимые с.в. распределенные по Бернулли с одинаковым параметром p . Пусть $S_n = X_1 + X_2 + \dots + X_n$. Тогда $S_n \sim B(n, p)$.

Числовые характеристики биномиального закона.

$$M(X) = n \cdot p,$$

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

Если n — велико, а p — мало, то вычисления вероятности по формуле $P_n(k) = C_n^k p^k q^{n-k}$ на практике невозможно. При этих условиях используется **формула Пуассона** для вычисления вероятности маловероятных событий в массовых испытаниях: $P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, где $\lambda = np$. Соответствующая с.в. распределена по закону Пуассона.

Пример 4.14.

Вероятность встретить блондина в равна 20%. Опрашивается 1000 человек. Как распределено X — число блондинов в этой выборке? Найти вероятность того, что среди опрошенных 100 человек будет 20 блондинов.

Поскольку имеем 100 испытаний по схеме Бернулли (с вероятностью успеха $p = 20\% = 0,2$), то $X \sim B(1000; 0,2)$. Искомая вероятность вычисляется как $P_{1000}(20) = C_{1000}^{20} 0,2^{20} 0,8^{880}$, и довести это выражение до конкретного числа сложная задача даже с использованием компьютера. Но, поскольку имеем право

использовать формулу Пуассона, то приблизительно имеем $\lambda = 100 \cdot 0.2 = 20$, $P_{100}(20) = \frac{20^{20}}{20!} e^{-20}$ а это значение уже легко вычислить даже в Excel и получить $0.08883531739208530000 = 8,88\%$.

4.10. Нормальное распределение

4.10.1. Определение и значение

Большинство экспериментальных исследований, в том числе и в области правоведения, связано с измерениями, результаты которых могут принимать практически любые значения в заданном интервале и, как уже было отмечено, описываются моделью непрерывных случайных величин. Поэтому в дальнейшем будут рассматриваться в основном непрерывные случайные величины и связанные с ними непрерывные распределения.

Одним из непрерывных распределений, имеющим основополагающую роль в математической статистике, является нормальное, или гауссово*, распределение.

Нормальное распределение является самым важным в статистике. Это объясняется целым рядом причин.

1. Прежде всего, многие экспериментальные наблюдения можно успешно описать с помощью нормального распределения. Следует сразу же отметить, что не существует распределений эмпирических данных, которые были бы в точности нормальными, поскольку (как будет показано ниже) нормально распределенная случайная величина находится в пределах от $-\infty$ до $+\infty$, чего никогда не бывает на практике. Однако нормальное распределение очень часто хорошо подходит как приближение.

Проводятся ли измерения IQ, роста и других физиологических параметров — везде на результаты оказывает влияние очень большое число случайных факторов (естественные причины и ошибки измерения). Причем, как правило, действие каждого из этих, факторов незначительно. Опыт показывает, что результаты именно в таких случаях будут распределены приближенно нормально.

2. Нормальное распределение хорошо подходит в качестве аппроксимации (приближенного описания) других распределений (например, биномиального).

3. Многие распределения, связанные со случайной выборкой, при увеличении объема последней переходят в нормальное.

4. Нормальное распределение обладает рядом благоприятных математических свойств, во многом обеспечивших его широкое применение в статистике.

В то же время следует отметить, что в природе встречается много экспериментальных распределений, для описания которых модель нормального распределения малопримодна. Для этого в математической статистике разработан ряд методов, некоторые из которых приводятся в следующих главах.

Говорят, что с.в. распределена по нормальному закону с параметрами m и σ и записывать $X \rightarrow N(m, \sigma^2)$ если ее плотность вероятностей задается следующим образом

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad -\infty < x < \infty. \quad (4.23)$$

График плотности (нормальная кривая) представлен на рис. 4.10.

Укажем основные свойства нормального распределения $N(m, \sigma^2)$.

1. Нормальная кривая имеет колоколообразную форму, симметричную относительно точки $x = m$, с точками перегиба, абсциссы которых отстоят от m на $\pm \sigma$.

2. Для нормального распределения математическое ожидание $M(X) = m$, дисперсия равна σ^2 и, следовательно, стандартное отклонение равно σ .

3. Как видно из выражения (4.23), нормальное распределение полностью определяется двумя параметрами: m и σ — математическим ожиданием и стандартным отклонением.

График плотности вероятности нормального распределения показывает, что для нормально распределенной случайной величины вероятность отклонения от среднего значения m быстро уменьшается с ростом величины отклонения.

* Гаусс Карл Фридрих (1777—1855) — немецкий ученый в области практической и прикладной математики. Распределение названо по его имени, так как он ввел его одним из первых.

4. Медиана и мода нормального распределения совпадают и равны математическому ожиданию m .

5. Коэффициенты асимметрии и эксцесса нормального распределения равны нулю ($\gamma_3 = 0$, $\gamma_4 = 0$).

Последнее свойство (5) используется для проверки предположения о нормальности распределения генеральной совокупности (гл. 6).

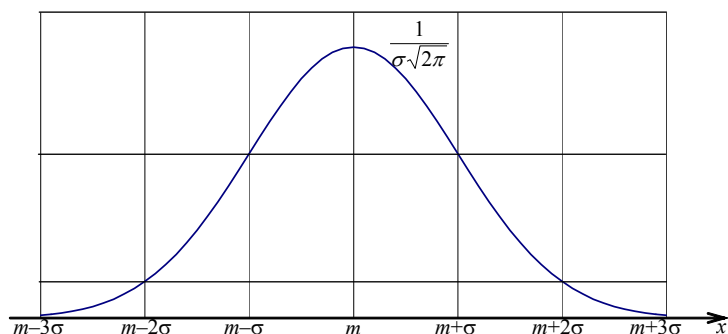


Рис. 4.10. Плотность вероятностей нормального распределения

4.10.2. Нормированное нормальное распределение

Формула (4.23) описывает целое семейство нормальных кривых, зависящих, как было сказано выше, от двух параметров — m и σ , которые могут принимать любые значения, поэтому возможно бесконечно много нормально распределенных совокупностей.

Чтобы избежать неудобств, связанных с расчетами для каждого конкретного случая по достаточно сложной формуле (4.23), используют так называемое нормированное (или стандартное) нормальное распределение $N(0;1)$, для которого составлены подробные таблицы.

Нормированное нормальное распределение имеет параметры $m=0$ и $\sigma=1$. Это распределение получается, если **пронормировать** нормально распределенную величину X по формуле:

$$U = \frac{X - m}{\sigma} \quad (4.24)$$

Плотность распределения вероятностей нормированного нормального распределения записывается в виде:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, -\infty < u < \infty.$$

На кривой нормированного нормального распределения (рис. 4.11) указаны в процентах доли площадей соответствующих отмеченным значениям нормированного отклонения u , по отношению к общей площади под кривой, равной 1 (100%). Эти площади определяют вероятности попадания случайной величины в соответствующие интервалы.

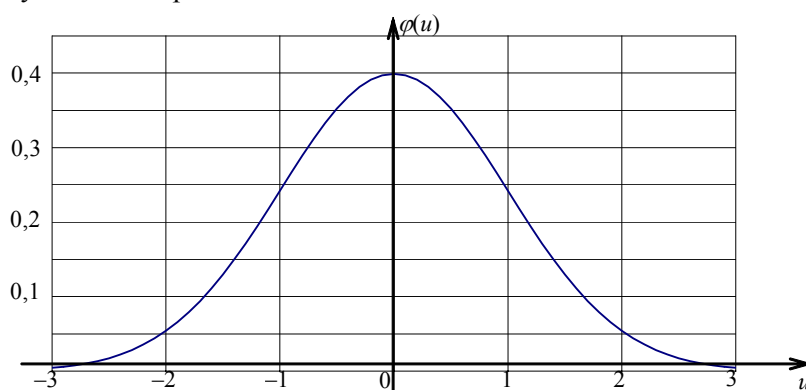


Рис. 4.11. Нормированное нормальное распределение

Таблица значений — ординат нормальной кривой приведена в специальных таблицах. Значения $\varphi(u)$ для некоторых характерных нормированных отклонений представлены в табл. 4.1.

Таблица 4.1
Ординаты нормальной кривой

Нормированное отклонение, u	0	$\pm 0,5$	$\pm 1,0$	$\pm 2,0$	$\pm 3,0$
Ордината нормальной кривой, $\varphi(u)$	0,399	0,352	0,242	0,054	0,004

4.10.3. Вероятность попадания в заданный интервал

Очень часто исследователя интересует вопрос: какова вероятность того, что изучаемый признак генеральной совокупности находится в заданных границах (например, вероятность того, что результат измерения IQ для группы испытуемых окажется в пределах 115 — 125)? Если предполагается нормальное распределение признака в генеральной совокупности, то получить ответ на этот вопрос очень просто.

Как говорилось ранее, вероятность попадания нормально распределенной случайной величины в заданный интервал $[x_1; x_2]$ можно определить по функции распределения: $P(x_1 < X < x_2) = F(x_2) - F(x_1)$ или

с помощью функции плотности вероятностей: $P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x) dx$.

Итак, вероятность попадания с.в. $U \rightarrow N(0;1)$ в заданный интервал:

$$P(a < U < b) = P(U < b) - P(U < a) = \Phi(b) - \Phi(a),$$

где Φ — принятое обозначение для функции нормированного нормального распределения которое имеет следующий вид:

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-t^2/2} dt, \quad (4.25)$$

при этом $\Phi(-u) = 1 - \Phi(u)$.

Часто представляет интерес вероятность попадания с.в. $U \rightarrow N(0;1)$ в симметричный интервал. Тогда

$$P(-a < U < a) = \Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1 \quad (4.25^*)$$

Учитывая свойства функции Лапласа, получаем:

Интеграл, входящий в выражение (4.25), не выражается в элементарных функциях, поэтому для вычисления функции $\Phi(u)$ используют вспомогательную функцию — функцию Лапласа (интеграл вероятностей):

$$\Phi_0(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-t^2/2} dt \quad (4.26)$$

который табулируется. Функция Лапласа является нечетной, т.е. $\Phi_0(-u) = -\Phi_0(u)$.

В книгах по теории вероятности приведена либо таблица значений функции Лапласа $\Phi_0(u)$, либо $\Phi(u)$.

Чтобы найти вероятность попадания нормально распределенной случайной величины $X \rightarrow N(m, \sigma^2)$ в заданный интервал $[x_1; x_2]$ с помощью функции Лапласа, сначала с.в. X нормализуется (см. 4.24), а затем используется следующая формула:

$$\begin{aligned} P(x_1 < X < x_2) &= P(x_1 < m + \sigma U < x_2) = P(x_1 - m < \sigma U < x_2 - m) = P\left(\frac{x_1 - m}{\sigma} < U < \frac{x_2 - m}{\sigma}\right) = \\ &= \Phi_0\left(\frac{x_2 - m}{\sigma}\right) - \Phi_0\left(\frac{x_1 - m}{\sigma}\right) \end{aligned} \quad (4.27)$$

Пример 4.14. Вычислить $P(0,74 < X < 3,26)$ если $X \rightarrow N(2,9)$.

Решение.

$$P(0,74 < X < 3,26) = P(0,74 < 2 + \sqrt{9}U < 3,26) = P(-1,26 < 3U < 1,26) = P(-0,42 < U < 0,42) = 2P(0,42) - 1 = 2 \cdot 0,6628 - 1 = 0,3256$$

4.10.4. Правило трех сигм

В табл. 4.2 приведены полученные по формуле (4.28) вероятности того, что нормально распределенная случайная величина отклонится от своего среднего значения m , не более, чем на $\pm 0,5\sigma$, $\pm\sigma$, $\pm 2\sigma$, $\pm 3\sigma$.

Таблица 4.2

Вероятности попадания нормально распределенной случайной величины в заданный интервал

Границы интервала, $m \pm x$	$m \pm 0.5\sigma$	$m \pm \sigma$	$m \pm 2\sigma$	$m \pm 3\sigma$
Вероятность попадания в интервал	0,3829	0,6827	0,9545	0,9973

Из табл. 4.2 следует, что $P[-3\sigma < (X - m) < 3\sigma] = 0,9973$.

Это выражение известно в статистике как “правило трех сигм”. Оно означает, что с вероятностью 0,9973 (практически с единичной) нормально распределенная случайная величина окажется в пределах $\pm 3\sigma$ от среднего значения. Иначе говоря, отклонения от среднего больше чем на $+3\sigma$ можно ожидать примерно в 1 случае из 370 испытаний.

4.11. Применение нормального распределения.

Теорема Муавра-Лапласа.

Пусть X — с.в. распределенная по биномиальному закону с параметрами n и p (т.е. $X \rightarrow B(n, p)$).

Если $n \geq 30$, $np(1-p) \geq 5$, то можно приблизительно описать закон распределения с.в. с помощью $N(np, np(1-p))$.

Тогда вероятность того, что дискретная с.в. X примет значение k (т.е. событие произойдет ровно k раз в проведенных n опытах) вычисляют так: $P(X = k) = P\left(k - \frac{1}{2} < X < k + \frac{1}{2}\right) = P\left(-\frac{1}{2} < np + \sqrt{npq} \cdot U < k + \frac{1}{2}\right)$, где U — нормированная нормальная с.в.

Пример 4.15. Вычислить вероятность того, что при 144 бросках игральной кости двойка выпадет от 28 до 30 раз.

Пусть X — кол-во выпадений двойки в этих 144 бросках. Очевидно, что X — с.в. распределенная по биномиальному закону $B\left(144, \frac{1}{6}\right)$. Вычислить $P(27 < X < 31) = P_{144}(28) + P_{144}(29) + P_{144}(30)$ по формуле

Бернулли не представляется возможным. Но т.к. $n = 144 \geq 30$ и $np(1-p) = 144 \cdot \frac{1}{6} \cdot \frac{5}{6} = 120 \geq 5$, то применим

теорему Муавра-Лапласа: $X \rightarrow B\left(144, \frac{1}{6}\right) \approx N\left(144 \cdot \frac{1}{6}, 144 \cdot \frac{1}{6} \cdot \frac{5}{6}\right) = N(24, 20)$, и, следовательно

$$P(27 < X < 31) = P\left(27 < 24 + \sqrt{24} \cdot U < 31\right) = P\left(27 - 24 < \sqrt{24} \cdot U < 31 - 24\right) = P\left(\frac{3}{\sqrt{24}} < U < \frac{7}{\sqrt{24}}\right) =$$

$$= P(0,61 < U < 1,43) = P(U < 1,43) - P(U < 0,61) = 0,9236 - 0,7291 = 0,1945$$

Теорема (центральная предельная в упрощенной формулировке).

Если случайные величины $X_1, X_2, \dots, X_n$ — независимы, одинаково распределены (все равно по какому закону!) и имеют конечные мат.ожидания $M[X]$ и дисперсии $D[X]$, то при $n \geq 30$ их среднее арифметическое $\bar{X}_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$ распределено близко к $N\left(M[X], \frac{D[X]}{n}\right)$ или, что то же самое, их сумма S_n распределена близко к $N(n \cdot M[X], n \cdot D[X])$.

Пример 4.16. Игровая кость, используемая мошенниками, подверглась испытаниям в результате которых были установлены вероятности выпадения граней (т.е. построен закон распределения с.в. X — кол-во выпавших очков):

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

Вычислить вероятность того, что при 50 бросках среднее кол-во очков будет между 3 и 4 (т.е. вычислить $P\left(3 < \frac{X_1 + X_2 + \dots + X_{50}}{50} < 4\right)$, где X_i — число полученное при i -ом броске).

Решение.

Очевидно, что $X_1, X_2, \dots, X_{50}$ — независимы, одинаково распределены, следовательно центральная предельная теорема можно аппроксимировать закон распределения с.в. $\bar{X}_{50} = \frac{X_1 + X_2 + \dots + X_{50}}{50}$ с помощью $N\left(M[X], \frac{D[X]}{50}\right)$.

Вычисли мат. ожидание и дисперсию для с.в. X : $M[X]=1\cdot 0,1+2\cdot 0,1+3\cdot 0,1+4\cdot 0,1+5\cdot 0,2+6\cdot 0,4=4,4$ очка, $D[X]=3,04$.

Итак, с.в. $\bar{X}_{50}$ распределена приблизительно по закону $N\left(4,4; \frac{3,04}{50}\right)$. Вычислим теперь

$$P(3 < \bar{X}_{50} < 4) = P\left(3 < 4,4 + \sqrt{\frac{3,04}{50}} \cdot U < 4\right) = P(-1,6 < 0,25 \cdot U < -0,4) = P(-1,62) - P(-6,48) = (1 - 0,94738) - 0 = 0,05262$$

Итак, получить в среднем между 3 и 4 очками маловероятно (вероятность лишь около 5% в то время как для стандартной игральной кости эта вероятность будет более 96%. Получите этот результат самостоятельно).

4.12. Некоторые специальные непрерывные распределения

Нормальное распределение широко применяется как математическая модель для описания экспериментальных данных. В этом разделе будут рассмотрены три распределения, которые играют очень важную роль при обработке результатов, связанных со случайной выборкой объема n , и составляют основу применения критериев значимости и проверки статистических гипотез. Примеры использования этих распределений приводятся в гл. 6, посвященной указанным статистическим методам.

4.12.1. χ^2 -распределение

Если $U_1, U_2, \dots, U_v$ независимые случайные величины, каждая из которых имеет нормированное нормальное распределение с параметрами $m=0$ и $\sigma=1$, то сумма квадратов этих величин

$$\chi^2 = U_1^2 + \dots + U_v^2$$

имеет так называемое χ^2 (хи-квадрат)-распределение. Его плотность вероятностей представлена на рис. 4.12 и зависит от единственного параметра — числа степеней свободы v .

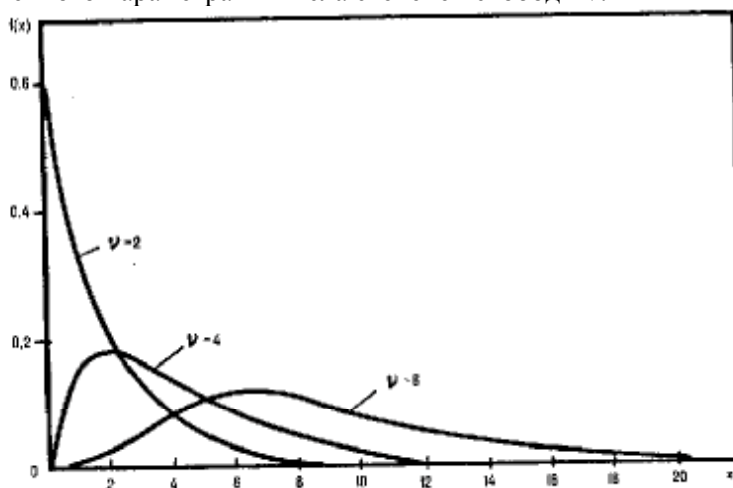


Рис. 4.12. χ^2 -распределение

Кривая χ^2 -распределения имеет положительную асимметрию. С ростом числа степеней свободы v она становится все более симметричной и при $v \geq 30$ переходит в нормальное.

Таблицы χ^2 -распределения приводятся в приложениях к учебникам по ТВ. В этих таблицах обычно содержатся значения x , соответствующие вероятностям $P = 1 - \alpha$, при α , равном 0,05; 0,01 и 0,001 для различного числа степеней свободы v .

4.12.2. t-распределение Стьюдента

Вторым из широко используемых специальных распределений является t-распределение Стьюдента, или просто t-распределение. Это распределение случайной величины:

$$T = \frac{U}{\sqrt{V/v}},$$

где U — случайная величина, имеющая нормированное нормальное распределение; V — случайная величина с распределением χ^2 с ν степенями свободы, t -распределение применяется при проверке статистических гипотез при малом объеме выборки. Эти вопросы рассмотрены в гл. 6. Форма t -распределения полностью определяется одним параметром — числом степеней свободы ν .

Вид кривой плотности t -распределения показан на рис. 4.13. t -распределение симметрично при любом ν и при $\nu \geq 30$ переходит в нормальное с параметрами $\mu = 0$ и $\sigma = \sqrt{\nu/(\nu-2)}$.

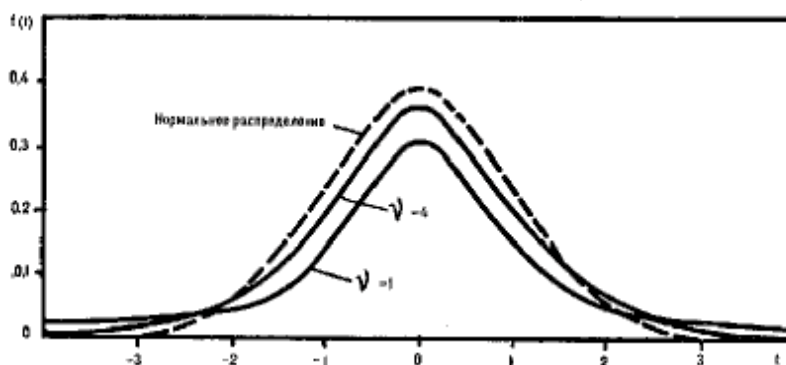


Рис. 4.13. t -распределение Стьюдента

Таблицы t -распределения обычно приводятся в приложениях к учебникам по ТВ.

4.12.3. F-распределение

Если случайные величины U и V независимы и каждая из них распределена как χ^2 с ν_1 и ν_2 степенями свободы соответственно, то величина

$$F = \frac{U / \nu_1}{V / \nu_2}$$

подчиняется так называемому F-распределению, которое зависит от двух параметров — ν_1 и ν_2 , называемых числами степеней свободы, F-распределение применяется в основном в задачах, связанных с дисперсиями. Эти задачи также рассмотрены в гл. 6. Таблицы χ^2 -распределения приводятся в приложениях к учебникам по ТВ.