С.М. Бородич, А.И. Курильчик, С.А. Шлапаков

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Курс лекций

2011

УДК 519.2(075.8)
ББК 22.17я73
Б83

Авторы: старшие преподаватели кафедры геометрии и математического анализа УО «ВГУ им. П.М. Машерова» **С.М. Бородич, А.И. Курильчик;** заведующий кафедрой геометрии и математического анализа УО «ВГУ им. П.М. Машерова», кандидат физико-математических наук, доцент **С.А. Шлапаков**

Рецензент:
заведующий кафедрой информатики и ИТ УО «ВГУ им. П.М. Машерова», кандидат физико-математических наук, доцент *А.И. Бочкин*

Научный редактор:
доцент кафедры геометрии и математического анализа УО «ВГУ им. П.М. Машерова», кандидат физико-математических наук *Т.Л. Сурин*

Курс лекций подготовлен в соответствии с учебной программой курса «Теория вероятностей и основы математической статистики». В нем содержится лекционный курс дисциплины, который сопровождается демонстрационными примерами, способствующими лучшему пониманию учебного материала.

Издание предназначено для студентов педагогических специализаций, обучающихся по специальностям «Математика. Физика», «Математика. Информатика» как дневной, так и заочной форм обучения. Оно окажется полезным всем тем, кто изучает курс теории вероятностей и математической статистики, и может быть использовано при самообразовании.

УДК 519.2(075.8)
ББК 22.17я73

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

Курс «Теория вероятностей и математическая статистика» является составной частью цикла математических дисциплин, из которых складывается фундамент математического образования современного специалиста.

Настоящее учебное издание основано на материале лекций, читаемых авторами в течение ряда лет для студентов математического и физического факультетов ВГУ им. П.М. Машерова, обучающихся по специальностям «Математика. Информатика», «Математика. Физика» и «Физика. Математика». Это издание призвано помочь студентам овладеть основами теории вероятностей и математической статистики в такой степени, чтобы они могли не только осознанно применять полученные знания в процессе обучения и работы, но и, по мере необходимости, углублять и расширять их путем дальнейшего самообразования. Следует отметить, что методы теории вероятностей и математической статистики находят весьма широкое применение в различных приложениях, с которыми нередко сталкиваются в своей дальнейшей профессиональной деятельности выпускники как естественных, так и гуманитарных факультетов университетов.

Материал лекций разбит на четыре части: 1) случайные события и их вероятности, 2) случайные величины, 3) предельные теоремы теории вероятностей, 4) элементы математической статистики. Изложение лекционного курса сопровождается демонстрационными примерами, способствующими лучшему пониманию и усвоению учебной дисциплины.

Учебное издание методологически подходит для организации самостоятельной работы и будет полезным студентам дневной и, в особенности, заочной форм обучения.

Часть I. СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ ВЕРОЯТНОСТИ

1.1. Случайные события. Пространство элементарных событий

Пусть проводится некоторое испытание (опыт, эксперимент, наблюдение), исход которого заранее предсказать нельзя. Такие эксперименты в теории вероятностей называют *случайными*. При этом рассматриваются только те случайные эксперименты, которые можно повторять, хотя бы теоретически, при неизменном комплексе условий произвольное число раз.

Случайным событием (или просто: *событием*) называется любой факт, который в результате испытания может произойти или не произойти.

События обозначаются, как правило, заглавными буквами латинского алфавита (при необходимости с индексами): A, B_1, C_3 и т.д.

Событие называется *достоверным*, если оно обязательно наступит в данном испытании, и *невозможным*, если оно заведомо в данном испытании не произойдет.

Достоверное событие будем обозначать буквой Ω (ome\'га) греческого алфавита, а невозможное – символом $\varnothing$.

Пример 1.1. Испытание: однократное бросание игральной кости. Событие A – выпадение 4 очков, событие B – выпадение четного числа очков, событие C – выпадение 7 очков, событие D – выпадение целого числа очков. Событие C – невозможное, а событие D – достоверное. ●

Непосредственные (теоретически возможные) исходы случайного эксперимента, из которых в результате эксперимента происходит ровно один, называются *элементарными событиями (элементарными исходами)*. Для любого события A в данном эксперименте можно выделить совокупность элементарных событий, при наступлении которых происходит это событие. Такие элементарные события называются *благоприятствующими событию* A. Таким образом, любое событие, связанное с данным экспериментом, можно представить как совокупность всех благоприятствующих этому событию элементарных исходов.

Пример 1.2. Испытание: один раз бросают игральную кость. Здесь 6 элементарных событий $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6$. Событие ω_i означает, что в результате бросания кости выпало i очков, $i = 1,\ 2,\ 3,\ 4,\ 5,\ 6$. Пусть событие A – это выпадение четного числа очков. Тогда благоприятствующими событию A будут элементарные события $\omega_2, \omega_4, \omega_6$, и значит, $A = \{\omega_2,\ \omega_4,\ \omega_6\}$. ●

Очевидно, что достоверное событие представляется совокупностью всех элементарных событий.

Множество всех элементарных событий данного эксперимента называется *пространством элементарных событий* и обозначается, как и достоверное событие, буквой Ω.

1.2. Отношения между событиями

Два события называются *несовместными*, если появление одного из них исключает появление другого в одном и том же испытании. В противном случае события называются *совместными*.

События $A_1, A_2, A_3, \ldots, A_n$ называются *попарно несовместными* (по-другому: *взаимоисключающими*), если любые два из них несовместны.

Говорят, что событие A *влечет* событие B, и пишут $A \subset B$ или $B \supset A$, если при наступлении события A обязательно происходит и событие B. Это означает, что любой элементарный исход, благоприятствующий событию A, является таковым и по отношению к событию B.

События и отношения между ними можно иллюстрировать с помощью *диаграмм Венна*. При этом пространство элементарных исходов (достоверное событие) Ω изображается прямоугольником, элементарные исходы – точками этого прямоугольника, а случайные события – областями внутри него. На рис. 1 показано отношение $A \subset B$.

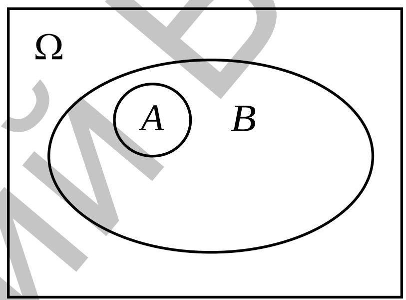

Рис. 1.

Пример 1.3. Испытание: однократное бросание игральной кости. Событие A – выпадение 2 очков, событие B – выпадение четного числа очков. При наступлении события A наступает и событие B, т.е. $A \subset B$. ●

Если $A \subset B$ и $B \subset A$, то события A и B называют *эквивалентными (равносильными, равными)*; записывают: $A = B$.

1.3. Алгебраические операции над событиями

Суммой событий $A_1, A_2, \ldots, A_n$ называется событие A, которое состоит в появлении хотя бы одного из этих событий. Записывают:

$$A = A_1 + A_2 + \cdots + A_n \quad \text{или} \quad A = \sum_{i=1}^{n} A_i .$$

Произведением событий $A_1, A_2, \ldots, A_n$ называется событие A, состоящее в появлении каждого из этих событий. Пишут:

$$A = A_1 A_2 \cdots A_n \quad \text{или} \quad A = \prod_{i=1}^{n} A_i .$$

Разностью двух событий A и B называется событие C, состоящее в том, что событие A произошло, а событие B не произошло. Записывают:

$$C = A - B .$$

Событие $\overline{A} = \Omega - A$ называется *дополнительным (противоположным)* к событию A. Оно заключается в непоявлении события A. Очевидно, что $\overline{\Omega} = \varnothing$, $\overline{\varnothing} = \Omega$, $\overline{\overline{A}} = A$.

Операции над событиями обладают следующими свойствами:

$A + B = B + A$, $\qquad AB = BA$,

$(A+B)+C=A+(B+C)$, $\quad (AB)C=A(BC)$,

$(A+B)C=AC+BC$, $\quad AB+C=(A+C)(B+C)$,

$A+A=A,$ $\quad AA=A,$

$A+\Omega=\Omega,$ $\quad A\Omega=A,$

$A+\varnothing=A,$ $\quad A\varnothing=\varnothing$,

$A+\overline{A}=\Omega,$ $\quad A\overline{A}=\varnothing$,

$\overline{A+B}=\overline{A}\overline{B}$, $\quad \overline{AB}=\overline{A}+\overline{B}$.

В справедливости этих формул можно убедиться с помощью диаграмм Венна.

1.4. Вероятность события. Классическое определение вероятности

Вероятность события – это числовая характеристика степени возможности его появления в рассматриваемом опыте. Вероятность события A обозначается символом $P(A)$.

Математических определений вероятности существует несколько, все они дополняют и обобщают друг друга.

Будем рассматривать вначале так называемую *классическую вероятностную модель*, которая используется для описания опытов с конечным числом равновозможных элементарных исходов.

Несколько событий в данном опыте называются *равновозможными*, если ни одно из них не является объективно более возможным, чем другие.

Равновозможность возникает обычно в силу симметрии в эксперименте (симметричная монета, правильная игральная кость, хорошо перемешанная колода карт, отсутствие оснований предпочесть один результат другому).

Пусть пространство элементарных событий эксперимента состоит из конечного числа элементарных событий, причем все они равновозможны. Пусть A – произвольное событие, связанное с данным экспериментом. *Вероятность (классическая вероятность) события A* определяется формулой

$$P(A)=\frac{|A|}{|\Omega|}, \tag{1.1}$$

где $|\Omega|$ – число всех элементарных событий, $|A|$ – число элементарных событий, благоприятствующих событию A.

Из определения (1.1) вытекают следующие свойства вероятности:

1. $0\le P(A)\le 1$;
2. $P(\Omega)=1$;
3. $P(\varnothing)=0$;
4. Если A и B – несовместные события (т.е. $AB=\varnothing$), то
$$P(A+B)=P(A)+P(B);$$
5. $P(A)+P(\overline{A})=1$.

1.5. Статистическое определение вероятности

Классическое определение вероятности применимо лишь в случаях, когда элементарные исходы испытания равновозможны. Во многих практических задачах это условие не выполняется. Например, при бросании неправильной игральной кости выпадение ее различных граней не равновозможно. В таких случаях применяются другие способы определения вероятностей. В их основе лежит эксперимент.

Пусть произведено n испытаний, при этом некоторое событие A наступило в k случаях. Величина

$$P^*(A) = \frac{k}{n}$$

называется *относительной частотой* события A.

При проведении нескольких серий из n испытаний, когда n сравнительно мало́, относительная частота $P^*(A)$ принимает значения, которые могут сильно отличаться друг от друга. Однако по мере увеличения n относительная частота $P^*(A)$ стабилизируется, т.е. приближается к некоторой постоянной (*свойство статистической устойчивости* относительной частоты).

Статистической вероятностью события A называется число $P(A)$, около которого группируются значения относительной частоты, приближаясь к нему при увеличении числа n.

Таким образом,

$$P(A) \approx P^*(A) = \frac{k}{n}$$

при больших значениях n.

Позже мы покажем, что при неограниченном увеличении числа испытаний относительная частота появления случайного события в определенном смысле (по вероятности) сходится к вероятности этого события (теорема Я. Бернулли).

Относительная частота $P^*(A)$ обладает всеми приведенными в пункте 1.4 свойствами классической вероятности:

1. $0 \le P^*(A) \le 1$;
2. $P^*(\Omega) = 1$;
3. $P^*(\varnothing) = 0$;
4. $P^*(A+B) = P^*(A) + P^*(B)$, если A и B – несовместные события;
5. $P^*(A) + P^*(\overline{A}) = 1$.

Следовательно, этими же свойствами будет обладать и статистическая вероятность события.

1.6. Геометрическое определение вероятности

Геометрическое определение вероятности является обобщением классического определения на случай, когда пространство элементарных событий Ω не является конечным множеством.

Рассматривается некоторый случайный эксперимент с бесконечным числом равновозможных элементарных исходов. Допустим, что каждому элементарному исходу ставится в соответствие некоторая точка на прямой, плоскости или в пространстве (при этом разным исходам соответствуют разные точки). Пусть Ω – множество всех таких точек. Таким образом, рассматриваемый эксперимент можно интерпретировать как случайный выбор точки $X \in \Omega$, а множество Ω – как пространство его элементарных исходов. Случайным событиям данного эксперимента соответствуют различные подмножества множества Ω, при этом подмножество A интерпретируется как случайное событие, заключающееся в том, что $X \in A$.

Будем предполагать, что и для самого множества Ω, и для всех рассматриваемых его подмножеств определена мера μ (длина, площадь, объем), причем предполагаем, что мера $\mu(\Omega)$ конечна.

Геометрической вероятностью события A называется число

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}.$$

Пример 1.4. На линии связи длиной 10 км произошел разрыв. Найти вероятность того, что разрыв произошел не далее, чем в 2-х км от начала линии. Предполагается равная возможность разрыва в любых точках линии связи.

Решение. $\Omega = [0, 10]$, $A = [0, 2]$. По формуле геометрической вероятности находим $P(A) = \frac{2}{10} = 0{,}2$. ●

Геометрическая вероятность обладает всеми перечисленными в пункте 1.4 свойствами классической вероятности. Однако имеются также и отличия в свойствах этих вероятностей. Например, если классическая вероятность события равна нулю, то это событие невозможно, а для геометрической вероятности это, вообще говоря, не так: геометрическая вероятность любого элементарного исхода равна нулю.

1.7. Формулы сложения вероятностей

Как известно (см. пункты 1.4, 1.5 и 1.6), вероятность суммы двух несовместных событий A и B равна сумме вероятностей этих событий:

$$P(A+B) = P(A) + P(B). \qquad (1.2)$$

С помощью метода математической индукции эта формула легко обобщается на случай n попарно несовместных событий $A_1, \ldots, A_n$:

$$P(A_1 + \ldots + A_n) = P(A_1) + \ldots + P(A_n). \qquad (1.3)$$

Теорема 1.1. Для любых событий A и B имеет место формула

$$P(A+B) = P(A) + P(B) - P(AB). \qquad (1.4)$$

Доказательство. $A + B = A + B\overline{A}$, $B = AB + B\overline{A}$. Так как $A(B\overline{A}) = \varnothing$ и $(AB)(B\overline{A}) = \varnothing$, то в силу формулы (1.2) имеем:

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

Подставляя в первое из этих равенств вероятность $P(B\overline{A})$, выраженную из второго равенства, получаем доказываемую формулу. ◄

Теорема 1.2. Для любых событий $A_1, A_2, \ldots, A_n$ имеет место формула

$$P(A_1 + A_2 + \ldots + A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i A_j) +$$
$$+ \sum_{i<j<k} P(A_i A_j A_k) - \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n). \qquad (1.5)$$

Эту формулу можно доказать методом математической индукции по n на основе формулы (1.4).

Замечание 1.1. Формулы (1.2)–(1.5) часто называют *формулами сложения вероятностей*. ●

1.8. Условная вероятность. Формула умножения вероятностей

Пусть A и B – два события, причем $P(B) \neq 0$.

Условной вероятностью события A при условии, что произошло событие B, называется величина

$$P(A \mid B) = \frac{P(AB)}{P(B)}. \qquad (1.6)$$

Из формулы (1.6) следует равенство

$$P(AB) = P(B) P(A \mid B), \qquad (1.7)$$

называемое *формулой (теоремой) умножения вероятностей*.

Формула (1.7) обобщается на случай n событий:

Теорема 1.3. Пусть события $A_1, A_2, \ldots, A_n$ таковы, что $P(A_1 A_2 \ldots A_{n-1}) \neq 0$. Тогда

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

Эта теорема легко доказывается методом математической индукции по n на основе формулы (1.7).

Пример 1.5. В урне находится 5 белых, 4 синих и 3 черных шара. Наудачу последовательно вынимают 3 шара. Какова вероятность того, что 1-й шар будет белым, 2-й – синим, 3-й – черным?

Решение. Введем обозначения событий: A_1 – первый вынутый шар – белый, A_2 – второй – синий, A_3 – третий – черный. Находим $P(A_1) = \frac{5}{12}$, $P(A_2 \mid A_1) = \frac{4}{11}$, $P(A_3 \mid A_1 A_2) = \frac{3}{10}$. В силу теоремы 1.3

$$P(A_1 A_2 A_3) = P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_1 A_2) = \frac{5}{12} \cdot \frac{4}{11} \cdot \frac{3}{10} = \frac{1}{22}. \; ●$$

1.9. Независимые события

События A и B, вероятности которых $P(A) \neq 0$ и $P(B) \neq 0$, называются *независимыми*, если

$$P(A|B)=P(A) \text{ и } P(B|A)=P(B). \qquad (1.8)$$

Замечание 1.2. Если выполняется одно из равенств (1.8), то выполняется и другое (т.е. в определении независимости событий *A* и *B* достаточно потребовать выполнения одного из них). Действительно,

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

$$\iff P(B|A)=P(B). \bullet$$

Теорема 1.4. События *A* и *B*, имеющие ненулевую вероятность, являются независимыми тогда и только тогда, когда

$$P(AB)=P(A)P(B). \qquad (1.9)$$

Доказательство. Утверждение этой теоремы вытекает из следующей ниже цепочки эквивалентных равенств и замечания 1.2.

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

Учитывая теорему 1.4, дадим также следующее определение независимости событий, распространяющееся и на события с нулевой вероятностью.

События *A* и *B* называются *независимыми*, если для них выполнено равенство (1.9).

Несколько событий называют *попарно независимыми*, если каждые два из них независимы.

События $A_1, \ldots, A_n$ называют *независимыми в совокупности* (или просто: *независимыми*), если независимы каждые два из них и независимы каждое событие и всевозможные произведения остальных.

Методом математической индукции формула (1.9) может быть распространена на *n* событий, независимых в совокупности:

$$P(A_1A_2\ldots A_n)=P(A_1)P(A_2)\ldots P(A_n). \qquad (1.10)$$

1.10. Формула полной вероятности

Говорят, что события $H_1, H_2, \ldots, H_n$ образуют *полную группу*, если они попарно несовместны (т.е. $H_iH_j=\varnothing$ при $i\neq j$) и $\sum_{i=1}^{n} H_i=\Omega$.

События $H_1, H_2, \ldots, H_n$, образующие полную группу, называются *гипотезами*.

Теорема 1.5. Пусть события $H_1, H_2, \ldots, H_n$ образуют полную группу. Тогда для любого события A имеет место формула

$$P(A)=\sum_{i=1}^{n} P(H_i)P(A|H_i), \qquad (1.11)$$

называемая *формулой полной вероятности*.

Доказательство. В силу того, что $\sum_{i=1}^{n} H_i=\Omega$, имеем

$$A=A\Omega=A\sum_{i=1}^{n} H_i=\sum_{i=1}^{n} AH_i.$$

Из того, что $H_i H_j = \varnothing$ при $i \neq j$, следует, что $(AH_i)(AH_j) = \varnothing$, т.е. события AH_i и AH_j тоже несовместны. Тогда в силу формулы (1.3) для вероятности суммы попарно несовместных событий

$$P(A) = \sum_{i=1}^{n} P(AH_i).$$

По формуле умножения вероятностей

$$P(AH_i) = P(H_i)P(A \mid H_i).$$

Из двух последних равенств и следует формула (1.11). ◄

Пример 1.6. В первой коробке содержится 20 ламп, из них 18 стандартных; во второй коробке – 10 ламп, из них 9 стандартных. Из второй коробки наудачу взята лампа и переложена в первую. Найти вероятность того, что лампа, наудачу извлеченная из первой коробки, будет стандартной.

Решение. Введем обозначения событий: A – из первой коробки извлечена стандартная лампа, H_1 – лампа, переложенная из второй коробки в первую, была стандартной, H_2 – эта лампа была нестандартной. Тогда

$$P(H_1) = \frac{9}{10},\ \ P(H_2) = \frac{1}{10},\ \ P(A \mid H_1) = \frac{19}{21},\ \ P(A \mid H_2) = \frac{18}{21}.$$

По формуле полной вероятности находим

$$P(A) = P(H_1)P(A \mid H_1) + P(H_2)P(A \mid H_2) = \frac{9}{10} \cdot \frac{19}{21} + \frac{1}{10} \cdot \frac{18}{21} = 0{,}9. \ \bullet$$

1.11. Формула Байеса (формула гипотез)

Теорема 1.6. Пусть события $H_1, H_2, \ldots, H_n$ образуют полную группу. Тогда условная вероятность события H_k $(k = 1, 2, \ldots, n)$ при условии, что в результате эксперимента произошло событие A, может быть вычислена по формуле

$$P(H_k \mid A) = \frac{P(H_k)P(A \mid H_k)}{\sum\limits_{i=1}^{n} P(H_i)P(A \mid H_i)}. \qquad (1.12)$$

Эта формула называется *формулой Байеса* (*формулой гипотез*).

Доказательство. Согласно определению условной вероятности

$$P(H_k \mid A) = \frac{P(AH_k)}{P(A)}.$$

Применяя к числителю этой дроби формулу умножения вероятностей, а к знаменателю – формулу полной вероятности, получаем равенство (1.12). ◄

Пример 1.7. В условиях примера 1.6 лампа, извлеченная наудачу из первой коробки после того, как в нее переложили лампу из второй, оказалась стандартной. Какова вероятность того, что из второй коробки в первую переложили нестандартную лампу?

Решение. Определим вероятность события (гипотезы) H_2 при условии, что событие A (извлеченная из первой коробки лампа стандартна) уже произошло:

$$P(H_2 \mid A) = \frac{P(H_2)P(A|H_2)}{P(H_1)P(A|H_1)+P(H_2)P(A|H_2)} = \frac{\frac{1}{10}\cdot\frac{18}{21}}{0{,}9} = \frac{2}{21}. \bullet$$

1.12. Схема Бернулли. Формула Бернулли

Пусть проводится серия n испытаний, в каждом из которых может произойти или не произойти событие A. При этом предполагается, что вероятность наступления события A в каждом отдельном испытании не зависит от исходов других испытаний и равна одному и тому же числу p $(0 < p < 1)$. Такая последовательность испытаний называется *схемой Бернулли*.

Ставится задача: вычислить вероятность события B_k, состоящего в том, что в n испытаниях, проведенных по схеме Бернулли, событие A произойдет ровно k раз $(0 \le k \le n)$. Искомую вероятность будем обозначать $P_n(k)$.

Пусть событие A_i означает появление события A в i-м испытании, тогда $P(A_i) = p$ для любого i.

Если принять всю последовательность из n испытаний за один опыт, то элементарные исходы этого опыта можно представить как произведения вида

$$A_1\overline{A}_2A_3\ldots\overline{A}_{n-1}\overline{A}_n. \tag{1.13}$$

В силу сделанных выше предположений множители каждого такого произведения являются независимыми в совокупности событиями. Благоприятствующими событию B_k будут лишь те элементарные исходы, которые содержат k множителей типа A_i и $n-k$ множителей типа $\overline{A}_i$. Число таких исходов равно C_n^k, при этом вероятность каждого из них в силу формулы (1.10) равна

$$p^k q^{n-k} \quad (q = 1-p).$$

Поскольку элементарные исходы опыта (произведения вида (1.13)) являются попарно несовместными событиями, то в силу формулы (1.3) получаем

$$P_n(k) = C_n^k p^k q^{n-k}. \tag{1.14}$$

Эта формула называется *формулой Бернулли*.

1.13. Предельные теоремы в схеме Бернулли

Использование формулы Бернулли (1.14) при больших значениях n затруднительно в вычислительном плане. Возникает необходимость в отыскании приближенных формул для вычисления вероятности $P_n(k)$. Такие формулы дают нам предельные теоремы.

Теорема 1.7 (теорема Пуассона). Пусть при неограниченном увеличении числа n испытаний $(n \to \infty)$ вероятность p наступления события A в каждом испытании стремится к нулю $(p \to 0)$, причем $np = \lambda$ – постоянная величина. Тогда

$$\lim_{n\to\infty} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \qquad (1.15)$$

Доказательство. Преобразуем формулу Бернулли (1.14) с учетом того, что $p = \dfrac{\lambda}{n}$:

$$P_n(k) = \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} =$$

$$= \frac{n(n-1)(n-2)\ldots(n-(k-1))}{k!}\cdot\frac{\lambda^k}{n^k}\cdot\left(1-\frac{\lambda}{n}\right)^n\cdot\left(1-\frac{\lambda}{n}\right)^{-k} =$$

$$= \frac{\lambda^k}{k!}\cdot 1\cdot\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\cdot\left(1-\frac{\lambda}{n}\right)^n\cdot\left(1-\frac{\lambda}{n}\right)^{-k}.$$

Переходя к пределу при $n\to\infty$ и учитывая, что

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \lim_{n\to\infty}\left[\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right]^{-\lambda} = e^{-\lambda},$$

получаем формулу (1.15). ◄

Из равенства (1.15) при больших n и малых p вытекает приближенная *формула Пуассона*

$$P_n(k) \approx \frac{\lambda^k}{k!}e^{-\lambda}, \quad \lambda = np, \quad k = 0, 1, 2, \ldots, n.$$

Эту формулу обычно используют, когда $p < 0{,}1$, а $np < 10$.

Пример 1.8. Завод отправил на базу 500 доброкачественных изделий. Вероятность того, что в пути изделие повредится, равна 0,002. Найти вероятность того, что на базу прибудут 3 негодных изделия.

Решение. Так как $n = 500,\ p = 0{,}002,$ то $\lambda = 500\cdot 0{,}002 = 1$. По формуле Пуассона находим

$$P_{500}(3) = \frac{1}{3!}e^{-1} \approx 0{,}06. \ \bullet$$

Сформулируем без доказательства еще две очень важные предельные теоремы, относящиеся к случаю, когда число испытаний n в схеме Бернулли велико, а вероятность p не близка к нулю.

Теорема 1.8 (локальная теорема Муавра–Лапласа). Если вероятность p появления события A в каждом из n испытаний постоянна и отлична от 0 и 1, то

$$P_n(k) \approx \frac{1}{\sqrt{npq}}\cdot\varphi\left(\frac{k-np}{\sqrt{npq}}\right),$$

где

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \text{ (функция Гаусса).}$$

Приближение тем точнее, чем больше n.

Для вычисления значений функции $\varphi(x)$ используется специальная таблица.

Обозначим через $P_n(k_1, k_2)$ вероятность того, что в n испытаниях, проводимых по схеме Бернулли, событие A наступит не менее k_1 и не более k_2 раз.

Теорема 1.9 (интегральная теорема Муавра–Лапласа). Если вероятность p наступления события A в каждом испытании постоянна и отлична от 0 и 1, то

$$P_n(k_1, k_2) \approx \Phi_0\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{k_1 - np}{\sqrt{npq}}\right), \qquad (1.16)$$

где

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \text{ (нормированная функция Лапласа)}. \qquad (1.17)$$

Равенство (1.16) тем точнее, чем больше n.

Значения функции $\Phi_0(x)$ находятся по таблице. Отметим некоторые важные свойства этой функции:

1. $\Phi_0(x)$ – нечетная функция, т.е. $\Phi_0(-x) = -\Phi_0(x)$;
2. $\Phi_0(x)$ монотонно возрастает;
3. $\lim\limits_{x \to +\infty} \Phi_0(x) = 0{,}5$.

Часть II. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

2.1. Понятие случайной величины. Дискретные и непрерывные случайные величины

Переменная величина, которая, в зависимости от исхода опыта, т.е. в зависимости от случая, принимает различные действительные значения, называется *случайной величиной* (сокращенная запись: с.в.).

Случайные величины будем обозначать прописными латинскими буквами (при необходимости с индексами): X, Y_1, Z_i и т.д., а их возможные значения – соответствующими строчными буквами с индексами: $x_1, x_2, \ldots, y_{11}, y_{12}, \ldots, z_{j1}, z_{j2}, \ldots$.

Пример 2.1. Опыт: однократное бросание игральной кости.

1) С.в. X – число выпавших очков. Возможные значения с.в. X: $x_1 = 1,\ x_2 = 2,\ \ldots,\ x_6 = 6.$

2) С.в. Y принимает значение 0, если выпало четное число очков, и значение 1, если число выпавших очков нечетное. Возможные значения с.в. Y: $y_1 = 0,\ y_2 = 1.$ ●

Различают два основных типа случайных величин: *дискретные* и *непрерывные* случайные величины.

Дискретной случайной величиной называют с.в., которая принимает отдельные, изолированные друг от друга возможные значения. Очевидно, что множество возможных значений дискретной с.в. является конечным или счетным (т.е. его элементы могут быть занумерованы натуральными числами).

Непрерывной случайной величиной называют с.в., которая может принимать любые значения из некоторого конечного или бесконечного промежутка. (Более строгое определение непрерывной с.в. будет дано ниже.)

Законом распределения случайной величины называется любое правило (таблица, функция, график), позволяющее находить вероятности произвольных событий, связанных с этой с.в. (например, вероятность того, что она примет какое-то значение или попадет в какой-то интервал).

Замечание 2.1. Вместо термина «закон распределения» часто употребляют более простой термин «распределение». ●

2.2. Закон распределения дискретной случайной величины. Основные дискретные распределения

Пусть X – дискретная с.в., которая принимает значения $x_1, x_2, \ldots$.

Законом распределения (или просто: *распределением*) *дискретной с.в.* X называется соответствие между возможными значениями x_i $(i = 1, 2, \ldots)$ и их вероятностями $p_i = P(X = x_i)$; его можно задать таблично, аналитически (в виде формулы) или графически.

При табличном задании закона распределения дискретной с.в. первая строка содержит возможные значения с.в., а вторая – их вероятности:

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

Такую таблицу обычно называют *рядом распределения*. Так как события $X = x_1$, $X = x_2$, …, $X = x_n$ несовместны и образуют полную группу, то $\sum_{i=1}^{n} p_i = 1$.

Если множество возможных значений с.в. X счетно, то ряд $\sum_{i=1}^{\infty} p_i$ сходится, причем $\sum_{i=1}^{\infty} p_i = 1$.

Закон распределения дискретной с.в. можно изобразить графически. Для этого в прямоугольной системе координат строят точки (x_i, p_i), а затем соединяют их отрезками. Полученную ломаную называют *многоугольником распределения*.

Рассмотрим некоторые наиболее часто встречающиеся на практике законы распределений дискретных случайных величин.

Биномиальное распределение. Дискретная с.в. X имеет *биномиальное распределение* (или *распределена по биномиальному закону*), если она принимает значения $k = 0, 1, 2, \ldots, n$ с вероятностями

$$p_k = P(X = k) = C_n^k p^k q^{n-k},$$

где $0 < p < 1$, $q = 1 - p$.

Данное определение корректно, так как

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

Как видим, вероятности p_k находятся по формуле Бернулли, полученной в пункте 1.12. Таким образом, случайную величину X, распределенную по биномиальному закону, можно рассматривать как число появлений события A в n испытаниях, проводимых по схеме Бернулли; при этом p – вероятность наступления события A в каждом из испытаний.

Распределение Пуассона. Дискретная с.в. X имеет *распределение Пуассона* (или *распределена по закону Пуассона*), если она принимает значения $k = 0, 1, 2, \dots$ с вероятностями

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

где $\lambda > 0$ – параметр.

Распределение Пуассона определено корректно, так как

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Отметим, что распределение Пуассона возникло в теореме Пуассона (п. 1.13) как предельное распределение для числа появлений события A в n испытаниях схемы Бернулли, когда $n \to \infty$, а $p \to 0$ так, что $np = \lambda$ – постоянная величина.

Геометрическое распределение. Дискретная с.в. X имеет *геометрическое распределение*, если она принимает значения $k = 1, 2, \dots$ с вероятностями

$$p_k = P(X = k) = pq^{k-1},$$

где $0 < p < 1,\ q = 1 - p.$

Данное определение корректно, так как

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} pq^{k-1} = p\sum_{k=1}^{\infty} q^{k-1} = p \cdot \frac{1}{1-q} = \frac{p}{p} = 1.$$

Геометрическое распределение также связано со схемой Бернулли: по этому закону распределена с.в. X, равная числу испытаний, проведенных до первого появления события A.

2.3. Функция распределения случайной величины

Функцией распределения с.в. X называется функция $F(x)$, которая для любого $x \in \mathbf{R}$ равна вероятности события $X < x$:

$$F(x) = P(X < x). \qquad (2.1)$$

Теорема 2.1. Функция распределения $F(x)$ обладает следующими свойствами:

1. $0 \le F(x) \le 1$ для любого $x \in \mathbf{R}$;
2. $F(x)$ – неубывающая функция на $\mathbf{R}$;
3. $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$;
4. $P(x_1 \le X < x_2) = F(x_2) - F(x_1)$, если $x_2 > x_1$;
5. $F(x)$ непрерывна слева в любой точке $x_0 \in \mathbf{R}$: $F(x_0 - 0) = F(x_0)$, где $F(x_0 - 0) = \lim\limits_{x \to x_0 - 0} F(x)$.

Докажем свойства 2–4.

Доказательство. 2. Пусть $x_1 < x_2$. Событие $X < x_2$ есть сумма двух несовместных событий $X < x_1$ и $x_1 \le X < x_2$. Поэтому

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

т.е.

$$F(x_2) = F(x_1) + P(x_1 < X \le x_2). \qquad (2.2)$$

В силу неотрицательности вероятности отсюда следует, что

$$F(x_1) \le F(x_2).$$

3. $\lim\limits_{x\to-\infty} F(x) = \lim\limits_{x\to-\infty} P(X < x) = P(X < -\infty) = 0$,

поскольку событие $X < -\infty$ невозможно.

$$\lim_{x\to+\infty} F(x) = \lim_{x\to+\infty} P(X < x) = P(X < +\infty) = 1,$$

поскольку событие $X < +\infty$ достоверно.

4. Свойство 4 следует из равенства (2.2). ◄

2.4. Непрерывные случайные величины

Выше было дано понятие непрерывной с.в. Приведем теперь более строгое определение.

Случайная величина X называется *непрерывной* с.в., если существует такая неотрицательная, интегрируемая по Риману на интервале $(-\infty, +\infty)$ функция $p(x)$, что при всех $x \in \mathbf{R}$

$$F(x) = \int_{-\infty}^{x} p(t)dt, \qquad (2.3)$$

где $F(x)$ – функция распределения с.в. X. Функция $p(x)$ называется *плотностью распределения вероятностей* (*плотностью распределения* или просто *плотностью*) случайной величины X.

Функция распределения непрерывной с.в. непрерывна на всей числовой оси, так как непрерывен по переменному верхнему пределу интеграл в правой части равенства (2.3).

Теорема 2.2. Плотность распределения вероятностей непрерывной с.в. обладает следующими свойствами:

1. $p(x) \ge 0$ при всех $x \in \mathbf{R}$;
2. $P(a \le X < b) = \int_a^b p(x)dx$ при $a < b$;
3. $\int_{-\infty}^{+\infty} p(x)dx = 1$ (условие нормировки);
4. $F'(x) = p(x)$ в точках непрерывности функции $p(x)$;
5. $P(x \le X < x + \Delta x) = p(x)\Delta x + o(\Delta x)$ в точках непрерывности $p(x)$.

Докажем свойство 2.

Доказательство. Используем свойство 4 функции распределения (теорема 2.1), равенство (2.3) и свойство аддитивности сходящегося несобственного интеграла:

$$P(a \le X < b) = F(b) - F(a) = \int_{-\infty}^{b} p(x)dx - \int_{-\infty}^{a} p(x)dx = \int_a^b p(x)dx. \ ◄$$

Замечание 2.2. Плотность распределения непрерывной с.в. называют также *дифференциальным законом распределения* (или просто *законом распределения*) этой с.в. ●

Теорема 2.3. Если X – непрерывная с.в., то для любого $x_0 \in \mathbf{R}$

$$P(X = x_0) = 0. \tag{2.4}$$

Доказательство. Используя свойство 4 функции распределения (п. 2.3, теорема 2.1), выводим

$$P(X = x_0) = \lim_{x \to x_0 + 0} P(x_0 \le X < x) =$$

$$= \lim_{x \to x_0 + 0} (F(x) - F(x_0)) = F(x_0 + 0) - F(x_0).$$

В силу непрерывности $F(x)$ отсюда получаем равенство (2.4). ◄

Следствие. Для непрерывной с.в. X справедливы равенства

$$P(a \le X < b) = P(a < X < b) = P(a \le X \le b) = P(a < X \le b).$$

2.5. Основные непрерывные распределения

Рассмотрим основные законы распределения непрерывных случайных величин.

Равномерное распределение. Непрерывная с.в. X имеет *равномерное распределение* на отрезке $[a, b]$, если ее плотность распределения задана формулой

$$p(x) = \begin{cases} \dfrac{1}{b-a} & \text{при } x \in [a, b], \\ 0 & \text{при } x \notin [a, b]. \end{cases}$$

Находим функцию распределения с.в. X, распределенной по равномерному закону:

$$F(x) = \int_{-\infty}^{x} p(t)dt = \begin{cases} 0 & \text{при } x < a, \\ \dfrac{x-a}{b-a} & \text{при } a \le x \le b, \\ 1 & \text{при } x > b. \end{cases}$$

Графики плотности $p(x)$ и функции распределения $F(x)$ для равномерного распределения на отрезке $[a, b]$ изображены на рис. 2.

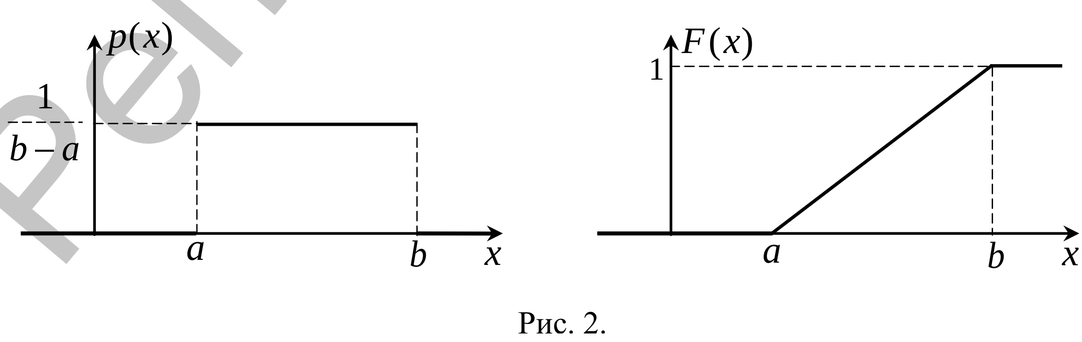

Рис. 2.

Показательное распределение. Непрерывная с.в. X имеет *показательное* (*экспоненциальное*) распределение, если ее плотность имеет вид

$$p(x) = \begin{cases} 0 & \text{при} \quad x < 0, \\ \lambda e^{-\lambda x} & \text{при} \quad x \geq 0, \end{cases}$$

где $\lambda > 0$ – параметр распределения.

Для функции распределения с.в. X, распределенной по показательному закону, нетрудно получить следующее выражение:

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

Графики плотности $p(x)$ и функции распределения $F(x)$ для показательного распределения изображены на рис. 3.

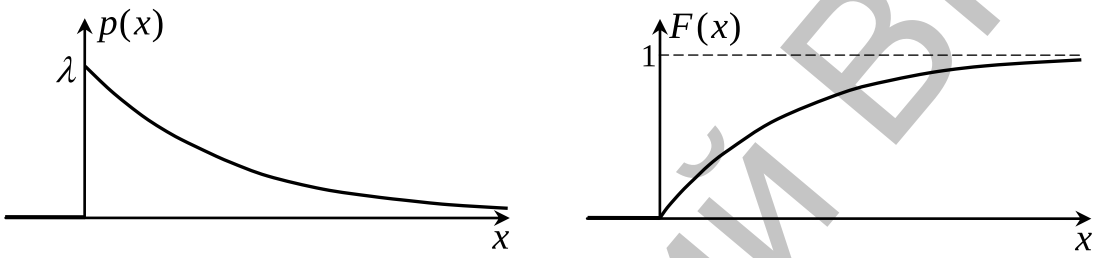

Рис. 3.

Нормальное распределение. Непрерывная с.в. X имеет *нормальное* распределение с параметрами $m \in \mathbf{R}$ и $\sigma > 0$, если ее плотность имеет вид

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbf{R}.$$

Функция распределения с.в. X, распределенной по нормальному закону, выражается формулой

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt = \Phi\left(\frac{x-m}{\sigma}\right), \tag{2.5}$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \ \ (\textit{функция Лапласа}).$$

Замечание 2.3. Функция Лапласа $\Phi(x)$ связана с нормированной функцией Лапласа $\Phi_0(x)$ (п. 1.13, формула 1.17) равенством

$$\Phi(x) = \Phi_0(x) + 0{,}5. \ \bullet$$

Графики плотности и функции распределения для нормального распределения изображены на рис. 4.

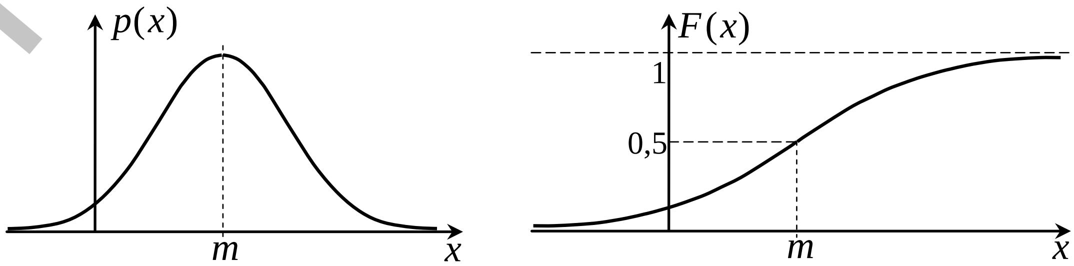

Рис. 4.

Найдем вероятность попадания с.в. X, распределенной по нормальному закону, в интервал (a, b). Учитывая следствие из теоремы 2.3, воспользуемся свойством 4 функции распределения (п. 2.3) и формулой (2.5):

$$P(a < X < b) = F(b) - F(a) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right).$$

В силу замечания 2.4 имеем также

$$P(a < X < b) = \Phi_0\left(\frac{b-m}{\sigma}\right) - \Phi_0\left(\frac{a-m}{\sigma}\right). \qquad (2.6)$$

Пример 2.2. Вычислить вероятность попадания с.в. X, распределенной по нормальному закону с параметрами m и σ, в интервал $(m-3\sigma,\ m+3\sigma)$.

Решение. Полагая в формуле (2.6) $a = m-3\sigma,\ b = m+3\sigma$, получаем

$$P(m-3\sigma < X < m+3\sigma) = \Phi_0\left(\frac{3\sigma}{\sigma}\right) - \Phi_0\left(-\frac{3\sigma}{\sigma}\right) = 2\Phi_0(3)$$

(здесь была учтена нечетность функции $\Phi_0(x)$). По таблице значений функции $\Phi_0(x)$ находим: $\Phi_0(3) = 0{,}49865$. Следовательно,

$$P(m-3\sigma < X < m+3\sigma) = 0{,}9973.$$

Таким образом, практически достоверно, что с.в. X принимает свои значения в промежутке $(m-3\sigma, m+3\sigma)$. Полученный результат называется «правилом трех σ». ●

2.6. Понятие многомерной случайной величины

Пусть в некотором эксперименте одновременно наблюдается n случайных величин $X_1, X_2, \ldots, X_n$.

Упорядоченный набор случайных величин $(X_1, X_2, \ldots, X_n)$ называется *n-мерной (многомерной) случайной величиной*, или *n-мерным случайным вектором*.

Одномерные случайные величины X_i $(i = 1, 2, \ldots, n)$ называются *компонентами (составляющими)* n-мерной с.в. $(X_1, X_2, \ldots, X_n)$.

Многомерные случайные величины могут быть *дискретными*, *непрерывными* и *смешанными* в зависимости от типа их компонент. В первом случае все компоненты многомерной с.в. являются дискретными случайными величинами, во втором – непрерывными. Компонентами смешанной многомерной с.в. являются случайные величины разных типов.

2.7. Закон распределения дискретной двумерной случайной величины

Рассмотрим дискретную двумерную с.в. (X, Y). Пусть с.в. X может принимать только значения $x_1, x_2, \ldots$, а с.в. Y – только значения $y_1, y_2, \ldots$. Тогда множество возможных значений двумерной с.в. (X, Y) составляют пары значений (x_i, y_j), $i = 1, 2, \ldots,\ j = 1, 2, \ldots$.

Через $P(X = x_i, Y = y_j)$ будем обозначать вероятность произведения событий $X = x_i$ и $Y = y_j$.

Перечень возможных значений пар компонент (x_i, y_j) и соответствующих каждой такой паре вероятностей

$$p_{ij} = P(X = x_i, Y = y_j)$$

называется *законом распределения* (или просто *распределением*) *дискретной двумерной с.в.* (X, Y).

Если множества возможных значений случайных величин *X* и *Y* конечны, то распределение дискретной двумерной с.в. (X, Y) можно задать в виде *таблицы совместного распределения*:

$X \backslash Y$	y_1	…	y_j	…	$\boldsymbol{y_m}$
x_1	p_{11}	…	p_{1j}	…	$\boldsymbol{p_{1m}}$
…	…	…	…	…	**…**
x_i	p_{i1}		p_{ij}	…	$\boldsymbol{p_{im}}$
…	…	…	…	…	**…**
$\boldsymbol{x_n}$	$\boldsymbol{p_{n1}}$		$\boldsymbol{p_{nj}}$	**…**	$\boldsymbol{p_{nm}}$

Так как события, означающие одновременное выполнение равенств $X = x_i, Y = y_j$ $(i = 1, 2, \ldots, n; j = 1, 2, \ldots, m)$, являются попарно несовместными и образуют полную группу, то

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1.$$

Одномерные законы распределения отдельных компонент выражаются через вероятности совместных значений по формулам

$$p_{i\bullet} = P(X = x_i) = \sum_{j=1}^{m} p_{ij}, \qquad p_{\bullet j} = P(Y = y_j) = \sum_{i=1}^{n} p_{ij}. \qquad (2.7)$$

Доказательство первой формулы (2.7) (вторая доказывается аналогично). Событие $X = x_i$ можно представить как сумму попарно несовместных событий

$$(X, Y) = (x_i, y_1),\ (X, Y) = (x_i, y_2),\ \ldots,\ (X, Y) = (x_i, y_m).$$

Воспользовавшись формулой сложения вероятностей для попарно несовместных событий (формула (1.3)), получаем первое равенство (2.7). ◄

2.8. Функция распределения многомерной случайной величины

Функцией распределения n-мерной с.в. $(X_1, X_2, \ldots, X_n)$ называется функция $F(x_1, x_2, \ldots, x_n)$, которая для любых действительных $x_1, x_2, \ldots, x_n$ равна вероятности произведения событий $X_1 < x_1,\ X_2 < x_2, \ldots,\ X_n < x_n$:

$$F(x_1, x_2, \ldots, x_n) = P(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n).$$

В частности, для двумерной с.в. (X, Y) функция распределения определяется равенством

$$F(x, y) = P(X < x, Y < y).$$

Геометрически функция распределения $F(x, y)$ означает вероятность попадания случайной точки (X, Y) в бесконечный квадрант с вершиной в точке $M(x, y)$, лежащий левее и ниже ее (заштрихованная область на рис. 5).

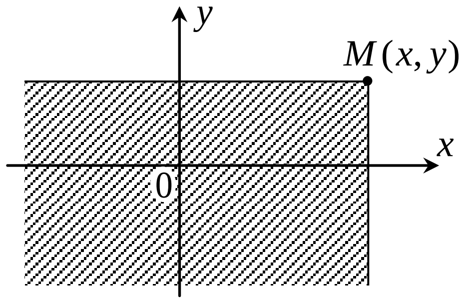

Рис. 5.

Теорема 2.4. Функция распределения $F(x, y)$ обладает следующими свойствами:

1. $0 \le F(x, y) \le 1$ в любой токе $(x, y) \in \mathbf{R}^2$;
2. $F(x, y)$– неубывающая функция по каждому из аргументов;
3. $\lim\limits_{x \to -\infty} F(x, y) = \lim\limits_{y \to -\infty} F(x, y) = \lim\limits_{\substack{x \to -\infty \\ y \to -\infty}} F(x, y) = 0$;
4. $\lim\limits_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1$;
5. $\lim\limits_{y \to +\infty} F(x, y) = F_X(x), \quad \lim\limits_{x \to +\infty} F(x, y) = F_Y(y)$,

где $F_X(x)$ и $F_Y(y)$ – функции распределения случайных величин X и Y, т.е. $F_X(x) = P(X < x),\ F_Y(y) = P(Y < y)$.

6. $F(x, y)$ непрерывна слева в любой точке $(x_0, y_0) \in \mathbf{R}^2$ по каждому из аргументов x и y.

Приведем доказательство первого равенства утверждения 5.

Доказательство. $\lim\limits_{y \to +\infty} F(x, y) = P(X < x, Y < +\infty)$. Поскольку событие $Y < +\infty$ является достоверным, то произведение событий $X < x$ и $Y < +\infty$ равно событию $X < x$ (см. п. 1.3). Таким образом,

$$\lim_{y \to +\infty} F(x, y) = P(X < x) = F_X(x). \blacktriangleleft$$

Геометрическая интерпретация доказанного свойства следующая: при $y \to +\infty$ заштрихованный на рис. 5 квадрант превращается в полуплоскость, лежащую слева от вертикальной прямой, проходящей через точку M; $P(X < x) = F_X(x)$ – вероятность попадания случайной точки (X, Y) в эту полуплоскость.

2.9. Непрерывная двумерная случайная величина

По аналогии с одномерным случаем дадим строгое определение непрерывной двумерной с.в.

Двумерная случайная величина (X,Y) называется *непрерывной*, если ее функция распределения $F(x,y)$ непрерывна в $\mathbf{R}^2$ и существует такая неотрицательная интегрируемая по Риману в бесконечных пределах по каждой из переменных функция $p(x,y)$, что

$$F(x,y) = \int_{-\infty}^{x} ds \int_{-\infty}^{y} p(s,t)dt. \qquad (2.8)$$

Функция $p(x,y)$ называется *плотностью распределения* (или *совместной плотностью*) двумерной с.в. (X,Y).

Теорема 2.5. Пусть (X,Y) – непрерывная двумерная с.в. с плотностью распределения $p(x,y)$. Тогда:

1. $p(x,y) \geq 0$ при всех $(x,y) \in \mathbf{R}^2$.
2. $P((X,Y) \in G) = \iint_G p(x,y)dxdy$, G – область в $\mathbf{R}^2$.
3. $\int_{-\infty}^{+\infty} ds \int_{-\infty}^{+\infty} p(s,t)dt = 1.$
4. $p(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$, если $(x,\ y)$ – точка непрерывности функции $p(x,y)$.
5. $P(x \leq X < x + \Delta x,\ y \leq Y < y + \Delta y) \approx p(x,y)\Delta x \Delta y.$
6. X и Y – непрерывные случайные величины, причем их плотности $p_X(x)$ и $p_Y(y)$ выражаются через совместную плотность $p(x,y)$ следующим образом:

$$p_X(x) = \int_{-\infty}^{+\infty} p(x,y)dy, \quad p_Y(y) = \int_{-\infty}^{+\infty} p(x,y)dx. \qquad (2.9)$$

Замечание 2.4. Функция распределения непрерывной n-мерной с.в. $(X_1,\ X_2,\ \ldots,\ X_n)$ по определению имеет вид

$$F(x_1,\ \ldots,\ x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} p(t_1,\ \ldots,\ t_n)dt_1 \ldots dt_n.$$

Свойства плотности $p(x_1,\ \ldots,\ x_n)$ аналогичны свойствам плотности $p(x,\ y)$. В частности, плотность распределения одномерной с.в. X_1 выражается через плотность $p(x_1,\ \ldots,\ x_n)$ следующим образом:

$$p_{X_1}(x_1) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} p(x_1,\ x_2,\ x_3,\ \ldots,\ x_n)dx_2 dx_3 \ldots dx_n. \ \bullet$$

2.10. Условные законы распределения компонент двумерных случайных величин

Условным законом распределения одной из компонент двумерной с.в. $(X,\ Y)$, называется ее закон распределения, найденный при условии, что другая компонента приняла определенное значение (или попала в какой-то определенный интервал).

1. Пусть (X, Y) – дискретная двумерная с.в., компонента X которой принимает значения $x_1, x_2, \ldots, x_n$, а компонента Y – значения $y_1, y_2, \ldots, y_m$. Как известно (см. п. 2.7), закон распределения двумерной с.в. (X, Y) задается набором вероятностей

$$p_{ij} = P(X = x_i, Y = y_j) \qquad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m).$$

Для краткости записи условную вероятность того, что с.в. X примет значение x_i при условии, что $Y = y_j$, обозначим через $p(x_i \mid y_j)$:

$$p(x_i \mid y_j) = P(X = x_i \mid Y = y_j).$$

Согласно определению условной вероятности (п. 1.8)

$$p(x_i \mid y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\bullet j}}.$$

Совокупность условных вероятностей

$$p(x_1 \mid y_j),\ p(x_2 \mid y_j),\ \ldots,\ p(x_n \mid y_j)$$

представляет собой *условный закон распределения с.в. X при условии, что* $Y = y_j$.

Аналогично определяется *условный закон распределения с.в. Y при условии, что* $X = x_i$:

$$p(y_j \mid x_i) = P(Y = y_j \mid X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{\bullet i}} \quad (j = 1, 2, \ldots, m).$$

Замечание 2.5. Сумма вероятностей условного закона распределения равна единице:

$$\sum_{i=1}^{n} p(x_i \mid y_j) = 1, \qquad \sum_{j=1}^{m} p(y_j \mid x_i) = 1. \tag{2.10}$$

Доказательство первого равенства (2.10) (второе доказывается аналогично). С учетом того, что $\sum_{i=1}^{n} p_{ij} = p_{\bullet j}$ (п. 2.7, вторая формула (2.7)), выводим

$$\sum_{i=1}^{n} p(x_i \mid y_j) = \sum_{i=1}^{n} \frac{p_{ij}}{p_{\bullet j}} = \frac{1}{p_{\bullet j}} \sum_{i=1}^{n} p_{ij} = \frac{p_{\bullet j}}{p_{\bullet j}} = 1. \ \bullet$$

2. Пусть (X, Y) – непрерывная двумерная с.в. с совместной плотностью распределения $p(x, y)$. Через $p_X(x)$ и $p_Y(y)$ обозначаем, как и выше, плотности распределений одномерных случайных величин X и Y.

Условной плотностью с.в. X при условии $Y = y$ называется функция $p_X(x \mid y)$, определяемая соотношением

$$p_X(x \mid y) = \frac{p(x, y)}{p_Y(y)} \qquad (p_Y(y) \neq 0). \tag{2.11}$$

Аналогично определяется *условная плотность с.в. Y при условии* $X = x$:

$$p_Y(y \mid x) = \frac{p(x, y)}{p_X(x)} \qquad (p_X(x) \neq 0). \tag{2.12}$$

В силу равенств (2.9) условные плотности можно выразить через совместную плотность следующим образом:

$$p_X(x\,|\,y) = \frac{p(x,y)}{\int\limits_{-\infty}^{+\infty} p(x,y)dx}, \qquad p_Y(y\,|\,x) = \frac{p(x,y)}{\int\limits_{-\infty}^{+\infty} p(x,y)dy}.$$

Как и любая плотность распределения, условные плотности обладают следующими свойствами:

$$p_X(x\,|\,y) \ge 0, \qquad \int\limits_{-\infty}^{+\infty} p_X(x\,|\,y)dx = 1,$$

$$p_Y(y\,|\,x) \ge 0, \qquad \int\limits_{-\infty}^{+\infty} p_Y(y\,|\,x)dy = 1.$$

2.11. Независимые случайные величины

Случайные величины $X_1, \ldots, X_n$ называются *независимыми* (по-другому: *независимыми в совокупности, взаимно независимыми*), если для любых $x_1, \ldots, x_n \in \mathbf{R}$ являются независимыми в совокупности события $X_1 < x_1, \ldots, X_n < x_n$ (см. п. 1.9).

Теорема 2.6. Для независимости случайных величин $X_1, \ldots, X_n$ необходимо и достаточно, чтобы для любых $x_1, \ldots, x_n \in \mathbf{R}$ выполнялось равенство

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n), \tag{2.13}$$

где $F(x_1, \ldots, x_n)$ – функция распределения n-мерной с.в. $(X_1, \ldots, X_n)$, $F_{X_i}(x_i)$ – функция распределения с.в. X_i, $i = 1, \ldots, n$.

В случае $n = 2$ утверждение теоремы 2.6 очевидно: равенство (2.13) переписывается в виде

$$P(X_1 < x_1,\ X_2 < x_2) = P(X_1 < x_1)P(X_2 < x_2),$$

что и означает независимость событий $X_1 < x_1$ и $X_2 < x_2$. Доказательство теоремы в общем случае читателю рекомендуется провести самостоятельно.

Теорема 2.7. Непрерывные случайные величины $X_1, \ldots, X_n$ независимы тогда и только тогда, когда n-мерная с.в. $(X_1, \ldots, X_n)$ непрерывна и ее плотность $p(x_1, \ldots, x_n)$ представляется в виде

$$p(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdot \ldots \cdot p_{X_n}(x_n),$$

где $p_{X_i}(x_i)$ – плотность с.в. X_i, $i = 1, \ldots, n$.

Доказательство утверждения теоремы для случая $n = 2$ (в общем случае доказательство аналогично). В силу теоремы 2.6 независимость случайных величин X_1 и X_2 равносильна выполнению равенства

$$F(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2). \tag{2.14}$$

Необходимость. С учетом определения непрерывной одномерной с.в. (п. 2.4) из равенства (2.14) выводим

$$F(x_1, x_2) = \int\limits_{-\infty}^{x_1} p_{X_1}(s)ds \cdot \int\limits_{-\infty}^{x_2} p_{X_2}(t)dt = \int\limits_{-\infty}^{x_1} ds \int\limits_{-\infty}^{x_2} p_{X_1}(s)p_{X_2}(t)dt.$$

Отсюда следует, что (X_1, X_2) является непрерывной двумерной с.в. с плотностью

$$p(x_1, x_2) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \qquad (2.15)$$

(см. определение непрерывной двумерной с.в. и ее плотности (п. 2.9)).

Достаточность. Проинтегрировав обе части равенства (2.15) по x_1 и x_2 в промежутках $(-\infty, x_1)$ и $(-\infty, x_2)$ соответственно, получим равенство (2.14). ◄

Для дискретных случайных величин справедлива следующая теорема.

Теорема 2.8. Дискретные случайные величины $X_1, \ldots, X_n$ независимы тогда и только тогда, когда для любых $x_1, \ldots, x_n \in \mathbf{R}$ выполняется равенство

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n).$$

Замечание 2.6. Из теорем 2.7 и 2.8, в частности, следует:

1. Если компоненты непрерывной с.в. (X, Y) являются независимыми случайными величинами, то их условные плотности равны безусловным:

$$p_X(x \mid y) = p_X(x), \quad p_Y(y \mid x) = p_Y(y).$$

2. Если компоненты дискретной с.в. (X, Y) независимы, то их условные законы распределения совпадают с безусловными:

$$p(x_i \mid y_j) = p_{i\bullet}, \quad p(y_j \mid x_i) = p_{\bullet j}. \; \bullet$$

2.12. Функции случайных величин. Закон распределения функции одной случайной величины

Пусть $X_1, \ldots, X_n$ – случайные величины, связанные с некоторым опытом, и $f(x_1, \ldots, x_n)$ – действительно значная функция n действительных переменных, область определения которой содержит все возможные значения n-мерной с.в. $(X_1, \ldots, X_n)$. Тогда можно определить с.в. Y, которая принимает свои значения в зависимости от того, какие значения принимают случайные величины $X_1, \ldots, X_n$, а именно: если в результате опыта случайные величины $X_1, \ldots, X_n$ приняли значения $x_1, \ldots, x_n$, то с.в. Y принимает значение $y = f(x_1, \ldots, x_n)$. При этом Y называют *функцией случайных величин* $X_1, \ldots, X_n$ и записывают: $Y = f(X_1, \ldots, X_n)$.

Выясним, как найти закон распределения функции $Y = f(X)$, если известен закон распределения аргумента X.

1. Пусть X – дискретная с.в., принимающая значения $x_1, x_2, \ldots, x_n$ с соответствующими вероятностями $p_1, p_2, \ldots, p_n$. Множество возможных значений с.в. Y составляют все различные числа среди чисел $f(x_1), f(x_2), \ldots, f(x_n)$. Пусть y_j – одно из возможных значений с.в. Y. Событие $Y = y_j$ эквивалентно сумме тех событий $X = x_i$, для которых $f(x_i) = y_j$. Поскольку все слагаемые в этой сумме – события несовместные, то

$$P(Y = y_j) = \sum_{i: f(x_i) = y_j} P(X = x_i) = \sum_{i: f(x_i) = y_j} p_i$$

(суммирование ведется по всем i, для которых $f(x_i) = y_j$).

2. Пусть X – непрерывная с.в. с плотностью распределения $p_X(x)$. Пред-

положим, что все возможные значения с.в. X составляют интервал $(a,\, b)$ (в частности, $a=-\infty$, $b=+\infty$), а функция $f(x)$ строго монотонна и дифференцируема на интервале $(a,\, b)$. В этом случае множеством всех значений с.в. $Y=f(X)$ также является интервал; обозначим его $(c,\, d)$. Можно доказать, что плотность с.в. Y имеет следующий вид:

$$p_Y(y)=\begin{cases} p_X(f^{-1}(y))\cdot|(f^{-1}(y))'| & \text{при } y\in(c,\, d), \\ 0 & \text{при } y\notin(c,\, d), \end{cases} \tag{2.16}$$

где $f^{-1}(y)$ – функция, обратная функции $f(x)$.

Пример 2.3. Пусть с.в. X имеет нормальное распределение с параметрами m и σ (см. п. 2.5). Найти плотность распределения с.в. $Y=aX+b$, где $a,b\in\mathbf{R}$, $a\neq 0$.

Решение. В данном случае имеем: $f(x)=ax+b$ – строго возрастающая (при $a>0$) или строго убывающая (при $a<0$) на всей числовой оси функция. Находим

$$f^{-1}(y)=\frac{y-b}{a}, \qquad (f^{-1}(y))'=\frac{1}{a}.$$

Как легко видеть, множеством возможных значений с.в. Y является интервал $(-\infty,\, +\infty)$. Согласно формуле (2.16) получаем

$$p_Y(y)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\left(\frac{y-b}{a}-m\right)^2}{2\sigma^2}}\cdot\frac{1}{|a|}=\frac{1}{|a|\sigma\sqrt{2\pi}}e^{-\frac{(y-(am+b))^2}{2(|a|\sigma)^2}}.$$

Таким образом, с.в. $Y=aX+b$ также распределена по нормальному закону с параметрами, равными $am+b$ и $|a|\sigma$. ●

Замечание 2.7. Из примера 2.3, в частности, следует: если с.в. X имеет нормальное распределение с параметрами m и σ, то с.в. $Y=\dfrac{X-m}{\sigma}$ также распределена нормально с параметрами 0 и 1. Действительно, в этом случае $a=\dfrac{1}{\sigma}$, $b=-\dfrac{m}{\sigma}$ и значит,

$$am+b=\frac{1}{\sigma}\cdot m-\frac{m}{\sigma}=0, \quad |a|\sigma=\frac{1}{\sigma}\cdot\sigma=1. \text{ ●}$$

2.13. Числовые характеристики случайных величин

Случайные величины, помимо закона распределения, могут описываться числовыми характеристиками, среди которых различают *характеристики положения* (основные: математическое ожидание, мода, медиана) и *характеристики рассеивания* (основные: дисперсия, среднее квадратическое отклонение).

Математическое ожидание. *Математическим ожиданием* (или *средним значением*) *дискретной с.в. X,* имеющей закон распределения $P(X=x_i)=p_i$, $i=1,\, 2,\, \ldots$, называется число $M(X)$, определяемое формулой

$$M(X)=\sum_i x_i p_i\,. \tag{2.17}$$

При этом, если множество возможных значений с.в. X счетно, то условием существования математического ожидания является абсолютная сходимость ряда в правой части равенства (2.17); в противном случае считают, что дискретная с.в. X математического ожидания не имеет.

Математическим ожиданием непрерывной с.в. X с плотностью распределения вероятностей $p(x)$ называется число

$$M(X)=\int_{-\infty}^{+\infty} x\,p(x)dx \tag{2.18}$$

при условии, что интеграл в правой части равенства (2.18) сходится абсолютно.

В дальнейшем будем предполагать, что все рассматриваемые математические ожидания существуют.

Теорема 2.9. Математическое ожидание обладает следующими свойствами:

1. $M(C)=C$, если C – константа (дискретная с.в., принимающая значение C с вероятностью 1);
2. $M(kX)=kM(X)$, если k – константа;
3. $M(X_1+X_2+\ldots+X_n)=M(X_1)+M(X_2)+\ldots+M(X_n)$;
4. $M(X+C)=M(X)+C$, если C – константа;
5. $M(X_1\cdot X_2\cdot\ldots\cdot X_n)=M(X_1)\cdot M(X_2)\cdot\ldots\cdot M(X_n)$ для независимых (в совокупности) случайных величин $X_1,\ X_2,\ \ldots,\ X_n$;
6. Если $X\geq 0$ (т.е. с.в. X принимает только неотрицательные значения), то $M(X)\geq 0$.

Дисперсия. *Дисперсией* с.в. X называется число $D(X)$, определяемое формулой

$$D(X)=M[(X-M(X))^2]=\begin{cases}\sum_i (x_i-M(X))^2 p_i, & \text{если с.в.} X \text{ дискретная},\\ \int_{-\infty}^{+\infty}(x-M(X))^2 p(x)dx, & \text{если с.в.} X \text{ непрерывная}.\end{cases}$$

Теорема 2.10. Дисперсия обладает следующими свойствами:

1. $D(X)\geq 0$;
2. $D(C)=0$, если C – константа;
3. $D(kX)=k^2D(X)$, если k – константа;
4. $D(X_1+X_2+\ldots+X_n)=D(X_1)+D(X_2)+\ldots+D(X_n)$ для независимых (в совокупности) случайных величин $X_1,\ X_2,\ \ldots,\ X_n$;
5. $D(X-Y)=D(X)+D(Y)$ для независимых случайных величин X и Y;
6. $D(X+C)=D(X)$, если C – константа;
7. $D(X)=M(X^2)-[M(X)]^2$.

Среднее квадратическое отклонение. *Средним квадратическим отклонением* с.в. X называется число

$$\sigma_X=\sqrt{D(X)}\,.$$

Мода и медиана. *Модой дискретной с.в.* X, принимающей значения $x_1<x_2<x_3<\ldots x_n$, называется наиболее вероятное ее значение по сравнению с соседними, т.е. такое значение x_m, что

$$P(X = x_m) = \max_{m-1 \le i \le m+1} P(x = x_i).$$

Модой непрерывной с.в. X называется точка локального максимума плотности распределения.

Моду с.в. X обозначают $M_o(X)$.

Медианой непрерывной с.в. X называется число $M_e(X)$, удовлетворяющее условию

$$P(X < M_e(X)) = P(X \ge M_e(X)) = \frac{1}{2}.$$

2.14. Числовые характеристики основных распределений

Найдем математические ожидания и дисперсии дискретных и непрерывных распределений, рассмотренных в пунктах 2.2 и 2.5.

Биномиальное распределение. Как известно (п. 2.2), с.в. X, имеющую биномиальное распределение, можно рассматривать как число появлений события A в n испытаниях, проводимых по схеме Бернулли (p – вероятность наступления события A в каждом из испытаний). Пусть с.в. X_i – число появлений события A в i-м испытании. Тогда

$$X = X_1 + \ldots + X_n.$$

Каждая с.в. X_i $(i = 1, \ldots, n)$ может принимать только два значения: $x_1 = 1$ с вероятностью p и $x_2 = 0$ с вероятностью $q = 1 - p$. Поэтому

$$M(X_i) = 1 \cdot p + 0 \cdot q = p \qquad (i = 1, \ldots, n),$$

$$M(X) = M(X_1) + \ldots + M(X_n) = np.$$

В силу независимости исходов испытаний случайные величины $X_1, \ldots, X_n$ независимы. Поэтому, согласно свойству 4 дисперсии (теорема 2.10),

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

Дисперсию каждой с.в. X_i найдем, используя свойство 7 дисперсии:

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

$$D(X_i) = M(X_i^2) - [M(X_i)]^2 = p - p^2 = p(1-p) = pq.$$

Таким образом,

$$D(X) = npq.$$

Распределение Пуассона. Последовательно находим:

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

$$M(X^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} (k-1+1) \frac{\lambda^{k-1}}{(k-1)!} =$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} =$$

$$= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda = \lambda^2 + \lambda,$$

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

Геометрическое распределение. Вычисляем:

$$M(X) = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1} = p\sum_{k=1}^{\infty} \frac{d(q^k)}{dq} = p\frac{d}{dq}\sum_{k=0}^{\infty} q^k =$$

$$= p\frac{d}{dq}\left(\frac{1}{1-q}\right) = p\frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p};$$

$$M(X^2) = \sum_{k=1}^{\infty} k^2 pq^{k-1} = \sum_{k=1}^{\infty} (k(k-1)+k)pq^{k-1} =$$

$$= pq\sum_{k=2}^{\infty} k(k-1)q^{k-2} + \sum_{k=1}^{\infty} kpq^{k-1} = pq\sum_{k=0}^{\infty} \frac{d^2(q^k)}{dq^2} + M(X) =$$

$$= pq\frac{d^2}{dq^2}\sum_{k=0}^{\infty} q^k + \frac{1}{p} = pq\frac{d^2}{dq^2}\left(\frac{1}{1-q}\right) + \frac{1}{p} = pq\frac{2}{(1-q)^3} + \frac{1}{p} =$$

$$= \frac{2pq}{p^3} + \frac{1}{p} = \frac{2q+p}{p^2} = \frac{q+1}{p^2},$$

$$D(X) = M(X^2) - [M(X)]^2 = \frac{q+1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

Равномерное распределение. Для с.в. X, имеющей равномерное распределение на отрезке $[a, b]$, находим:

$$M(X) = \int_a^b x \cdot \frac{1}{b-a}dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2};$$

$$M(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a}dx = \left.\frac{x^3}{3(b-a)}\right|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3},$$

$$D(X) = M(X^2) - [M(X)]^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

Показательное распределение. При вычислении интегралов используем метод интегрирования по частям:

$$M(X) = \int_0^{+\infty} x\lambda e^{-\lambda x}dx = -\int_0^{+\infty} xde^{-\lambda x} = -xe^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}dx = 0 - \frac{1}{\lambda}e^{-\lambda x}\Big|_0^{+\infty} = \frac{1}{\lambda};$$

$$M(X^2) = \int_0^{+\infty} x^2\lambda e^{-\lambda x}dx = -\int_0^{+\infty} x^2 de^{-\lambda x} =$$

$$= -x^2 e^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} 2xe^{-\lambda x}dx = 0 + \frac{2}{\lambda}\int_0^{+\infty} x\lambda e^{-\lambda x}dx = \frac{2}{\lambda}\cdot\frac{1}{\lambda} = \frac{2}{\lambda^2},$$

$$D(X) = M(X^2) - [M(X)]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Нормальное распределение. Пусть с.в. X имеет нормальный закон распределения с параметрами m и σ. Как известно (замечание 2.7), с.в.

$Y=\dfrac{X-m}{\sigma}$ также распределена нормально с параметрами 0 и 1. Найдем сначала математическое ожидание и дисперсию с.в. *Y*:

$$M(Y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} ye^{-\frac{y^2}{2}}dy=0,$$

так как интеграл сходится, а подынтегральная функция нечетная;

$$M(Y^2)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} y^2e^{-\frac{y^2}{2}}dy=-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} y\,de^{-\frac{y^2}{2}}=$$

$$=-\frac{1}{\sqrt{2\pi}}ye^{-\frac{y^2}{2}}\Bigg|_{-\infty}^{+\infty}+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}dy=0+1=1,$$

$$D(Y)=M(Y^2)-[M(Y)]^2=1-0=1.$$

Теперь с помощью соответствующих свойств математического ожидания и дисперсии (п. 2.13) находим математическое ожидание и дисперсию с.в. $X=\sigma Y+m$:

$$M(X)=\sigma M(Y)+m=m, \qquad D(X)=\sigma^2 D(Y)=\sigma^2.$$

Таким образом, выяснен теоретико-вероятностный смысл параметров нормального распределения: m – математическое ожидание, $\sigma=\sqrt{D(X)}$ – среднее квадратическое отклонение.

Часть III. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

3.1. Неравенства Маркова и Чебышева

Теорема 3.1 (неравенство Маркова). Пусть *X* – неотрицательная с.в., имеющая математическое ожидание. Тогда для любого $\varepsilon>0$ справедливо неравенство

$$P(X\ge\varepsilon)\le\frac{M(X)}{\varepsilon}. \tag{3.1}$$

Доказательство. Докажем неравенство (3.1) для непрерывной с.в. *X* с плотностью распределения $p(x)$.

Учитывая неотрицательность с.в. *X*, получаем

$$M(X)=\int_0^{+\infty} xp(x)dx=\int_0^{\varepsilon} xp(x)dx+\int_{\varepsilon}^{+\infty} xp(x)dx\ge\int_{\varepsilon}^{+\infty} xp(x)dx\ge\varepsilon\int_{\varepsilon}^{+\infty} p(x)dx=\varepsilon P(X\ge\varepsilon).$$

Отсюда следует неравенство (3.1). ◄

Отметим, что неравенство Маркова можно записать в другой форме:

$$P(X<\varepsilon)\ge 1-\frac{M(X)}{\varepsilon}.$$

Теорема 3.2 (неравенство Чебышева). Пусть *X* – с.в., имеющая математическое ожидание и дисперсию. Тогда для любого $\varepsilon>0$ справедливо неравенство

$$P(|X-M(X)|\ge\varepsilon)\le\frac{D(X)}{\varepsilon^2}.$$

Доказательство. Неравенство $|X - M(X)| \geq \varepsilon$ равносильно неравенству $(X - M(X))^2 > \varepsilon^2$. Применяя к неотрицательной с.в.

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

неравенство Маркова, в котором ε заменено на ε^2, получаем

$$P(|X - M(X)| \geq \varepsilon) = P[(X - M(X))^2 > \varepsilon^2] = P(Y > \varepsilon^2) \leq \frac{M(Y)}{\varepsilon^2} = \frac{D(X)}{\varepsilon^2}. \blacktriangleleft$$

Другая форма неравенства Чебышева:

$$P(|X - M(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}.$$

3.2. Теорема Чебышева

Будем говорить, что последовательность случайных величин $X_1, X_2, \ldots$ является *последовательностью независимых случайных величин*, если при любом натуральном $n \geq 2$ случайные величины $X_1, \ldots, X_n$ независимы.

Теорема 3.3 (теорема Чебышева). Пусть $X_1, X_2, \ldots$ – последовательность независимых случайных величин, обладающих математическими ожиданиями $M(X_i)$ и дисперсиями $D(X_i)$, причем дисперсии ограничены в совокупности (т.е. существует такая константа $C > 0$, что $D(X_i) \leq C$, $i = 1, 2, \ldots$). Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left| \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i) \right| \geq \varepsilon \right) = 0. \qquad (3.2)$$

Доказательство. При произвольном $n \geq 2$ рассмотрим с.в. $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Легко видеть, что $M(Y_n) = \frac{1}{n}\sum_{i=1}^{n} M(X_i)$, $D(Y_n) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i)$ (при получении второго равенства учитывалась независимость случайных величин $X_1, \ldots, X_n$). Применяя к с.в. Y_n неравенство Чебышева и учитывая, что $D(X_i) \leq C$, $i = 1, 2, \ldots$, выводим

$$P(|Y_n - M(Y_n)| \geq \varepsilon) \leq \frac{D(Y_n)}{\varepsilon^2} = \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n} D(X_i) \leq \frac{nC}{n^2\varepsilon^2} = \frac{C}{n\varepsilon^2}, \text{ т.е.}$$

$$P\left(\left| \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i) \right| \geq \varepsilon \right) \leq \frac{C}{n\varepsilon^2}.$$

Переходя в последнем неравенстве к пределу при $n \to \infty$, получаем формулу (3.2). ◄

Следствие. Пусть $X_1, X_2, \ldots$– последовательность независимых одинаково распределенных случайных величин, $M(X_i) = m$, $D(X_i) = \sigma^2$, $i = 1, 2, \ldots$. Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left| \frac{1}{n}\sum_{i=1}^{n} X_i - m \right| \geq \varepsilon \right) = 0. \qquad (3.3)$$

Последовательность случайных величин $Y_1, Y_2, \ldots$ называется *сходящейся по вероятности* к числу b, если для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|Y_n - b\right| \ge \varepsilon\right) = 0$$

(или $\lim_{n\to\infty} P\left(\left|Y_n - b\right| < \varepsilon\right) = 1$); символическая запись:

$$Y_n \xrightarrow{P} b.$$

Замечание 3.1. Соотношение (3.3), выражающее частный случай теоремы Чебышева, можно записать в виде

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} m. \qquad \bullet$$

3.3. Теорема Бернулли

Теорема 3.4 (теорема Я. Бернулли). Пусть k – число наступлений события A в n испытаниях, проводимых по схеме Бернулли. Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{k}{n} - p\right| \ge \varepsilon\right) = 0, \qquad (3.4)$$

где p – вероятность появления события A в каждом из испытаний.

Доказательство. Пусть с.в. X_i – число появлений события A в i-м испытании. Тогда

$$k = \sum_{i=1}^{n} X_i. \qquad (3.5)$$

В силу определения схемы Бернулли (п. 1.12) $X_1, X_2, \ldots$– последовательность независимых одинаково распределенных случайных величин. Ранее (п. 2.14, биномиальное распределение) мы уже установили, что при каждом i

$$M(X_i) = p, \;\; D(X_i) = pq \quad (q = 1 - p).$$

Таким образом, последовательность случайных величин $X_1, X_2, \ldots$ удовлетворяет всем условиям следствия теоремы Чебышева значит

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| \ge \varepsilon\right) = 0.$$

С учетом равенства (3.5) отсюда получаем утверждение (3.4). ◄

Соотношение (3.4) можно заменить на эквивалентное ему соотношение

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

Замечание 3.2. Теорема Бернулли утверждает, что относительная частота $P^*(A) = \dfrac{k}{n}$ сходится по вероятности к вероятности $P(A) = p$:

$$P^*(A) \xrightarrow{P} P(A).$$

Таким образом, эта теорема объясняет смысл свойства устойчивости относительной частоты, которое мы ранее приняли как экспериментальный факт (п. 1.5). ●

3.4. Центральная предельная теорема

Приведем без доказательства вариант центральной предельной теоремы для независимых одинаково распределенных случайных величин.

Теорема 3.5. Пусть $X_1, X_2, \ldots$ – последовательность независимых одинаково распределенных случайных величин, имеющих математическое ожидание $M(X_i)=m$ и дисперсию $D(X_i)=\sigma^2>0$, $i=1, 2, \ldots$. Тогда для любого $x\in\mathbf{R}$

$$P\left(\frac{X_1+\ldots+X_n-nm}{\sigma\sqrt{n}}<x\right)\xrightarrow[n\to\infty]{}\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{t^2}{2}}dt. \qquad (3.6)$$

Замечание 3.3. Соотношение (3.6) означает, что при $n\to\infty$ закон распределения с.в.

$$Z_n=\frac{X_1+\ldots+X_n-nm}{\sigma\sqrt{n}}$$

неограниченно приближается к нормальному закону распределения с параметрами 0 и 1 (п. 2.5, нормальное распределение). Отсюда следует, что и закон распределения с.в. $S_n=\sum\limits_{i=1}^{n}X_i$ при $n\to\infty$ неограниченно приближается к нормальному, но уже с параметрами nm и $\sigma\sqrt{n}$. Действительно,

$$F_{S_n}(x)=P(S_n<x)=P\left(\frac{S_n-nm}{\sigma\sqrt{n}}<\frac{x-nm}{\sigma\sqrt{n}}\right)\xrightarrow[n\to\infty]{}\Phi\left(\frac{x-nm}{\sigma\sqrt{n}}\right),$$

т.е. функция распределения с.в. S_n при любом фиксированном *x* стремится при $n\to\infty$ к функции распределения с.в., распределенной по нормальному закону с параметрами nm и $\sigma\sqrt{n}$. ●

Часть IV. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

4.1. Генеральная совокупность. Выборка. Статистическое распределение выборки и ее графическое представление

Будем предполагать, что каждый исход некоторого случайного эксперимента характеризуется одним из возможных значений с.в. *X*. В таком случае говорят, что с.в. *X* *наблюдается* в данном эксперименте.

В математической статистике множество всех возможных значений наблюдаемой с.в. *X* называют *генеральной совокупностью*. При этом закон распределения с.в. *X* называют *законом распределения (распределением) генеральной совокупности*.

Замечание 4.1. Иногда генеральной совокупностью называют саму с.в. *X*. ●

Повторив *n* раз эксперимент в одинаковых условиях, получим последовательность из *n* наблюденных значений с.в. *X*: $x_1, x_2, \ldots, x_n$. Набор чисел $x_1, x_2, \ldots, x_n$ называется *выборкой*.

Числа x_i $(i=1, 2, \ldots, n)$ называются *элементами выборки*, а их количество n – *объемом выборки*.

Выборку можно упорядочить, расположив ее элементы в неубывающем порядке: $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$.
Полученная таким образом последовательность чисел $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ называется *вариационным рядом.*

Различные значения с.в. X, представленные в выборке, называются *вариантами*.

Если в выборке объема n варианта x_i встретилась n_i раз, то число n_i называется *частотой,* а число $w_i = \dfrac{n_i}{n}$ – *относительной частотой (частостью)* этой варианты. Очевидно, что сумма всех частот равна объему выборки n, а сумма всех относительных частот равна единице.

Статистическим рядом распределения выборки называют перечень вариант и соответствующих им частот или относительных частот. Обычно статистический ряд записывается в виде таблицы, первая строка которой содержит варианты x_i, а вторая – их частоты n_i (или относительные частоты w_i). При этом варианты располагаются в порядке возрастания.

Как правило, статистический ряд распределения составляется в случае, когда наблюдаемая с.в. X является дискретной, причем количество вариант в вариационном ряду не очень велико.

Графически статистический ряд распределения выборки изображается следующим образом. В прямоугольной декартовой системе координат строят точки с координатами (x_i, n_i). Последовательно соединив их отрезками, получают ломаную линию, называемую *полигоном частот*. Аналогично строится *полигон относительных частот*.

Пример 4.1. Имеется выборка значений с.в. X объема $n=15$:

$$0,\ 3,\ -5,\ -3,\ 1,\ 0,\ 1,\ 3,\ 0,\ 0,\ -3,\ 1,\ -1,\ 0,\ -1.$$

Вариационным рядом для нее будет последовательность

$$-5,\ -3,\ -3,\ -1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 3,\ 3.$$

Статистический ряд распределения данной выборки имеет вид

x_i	-5	-3	-1	0	1	3
n_i	1	2	2	5	3	2

или

x_i	-5	-3	-1	0	1	3
w_i	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{5}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

Полигон частот изо-

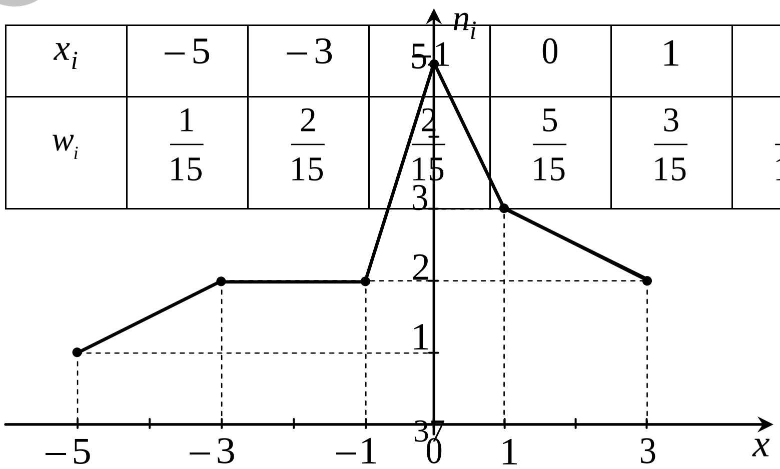

Рис. 6.

бражен на рис. 6.

В случае, когда X является непрерывной с.в. или же вариационный ряд имеет большое количество вариант, элементы выборки объединяют в группы (разряды), представляя результаты опытов в виде *интервального (группированного) статистического ряда распределения*. Для этого интервал, содержащий все значения выборки, разбивают на k непересекающихся интервалов

$$[a_0,\, a_1),\ [a_1,\, a_2),\ \ldots,\ [a_{k-1},\, a_k].$$

Затем для каждого интервала разбиения определяют *частоту* – количество элементов выборки, попавших в этот интервал: n_i – частота, соответствующая i-му интервалу разбиения $(i = 1,\, 2,\, \ldots,\, k)$. Наряду с частотами находят также *относительные частоты* $w_i = \dfrac{n_i}{n}$ (n – объем выборки). Интервальный статистический ряд записывают в виде таблицы, в первой строке которой содержатся интервалы группировки, а во второй – соответствующие частоты n_i или относительные частоты w_i.

Интервальный статистический ряд распределения изображают с помощью *гистограммы частот*. Для этого на оси абсцисс откладывают интервалы группировки и на них, как на основании, строят прямоугольники с высотами $h_i = \dfrac{n_i}{a_i - a_{i-1}}$ $(i = 1,\, 2,\, \ldots,\, k)$. Аналогично строится *гистограмма относительных частот*.

Пример 4.2. Выборка 100 значений наблюдаемой непрерывной с.в. X представлена в виде следующего интервального статистического ряда распределения:

Интервал	[22, 24]	[24, 26]	[26, 28]	[28, 30]	[30,32]	[32, 34]
Частота	2	14	34	40	8	2

Построить гистограмму частот.

Решение. В данном случае $a_i - a_{i-1} = 2$ $(i = 1, 2, 3, 4, 5, 6)$. Находим высоты прямоугольников:

$h_1 = \frac{2}{2} = 1$, $h_2 = \frac{14}{2} = 7$, $h_3 = \frac{34}{2} = 17$, $h_4 = \frac{40}{2} = 20$, $h_5 = \frac{8}{2} = 4$, $h_6 = \frac{2}{2} = 1$.

Гистограмма частот изображена на рис. 7.

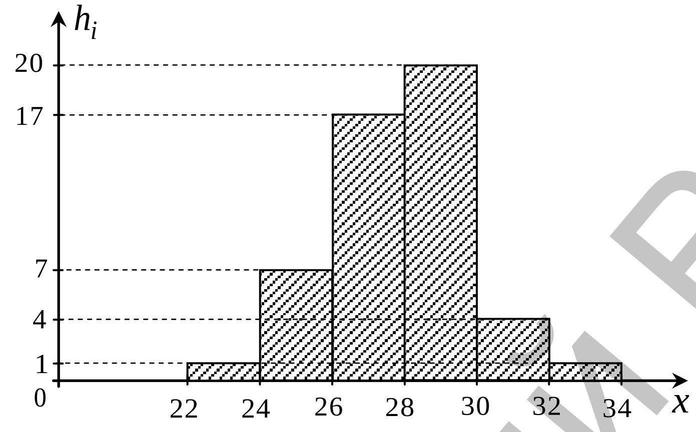

Рис. 7.

4.2. Эмпирическая функция распределения

Пусть в результате наблюдений за с.в. X получена выборка объема n.

Эмпирической (выборочной) функцией распределения с.в. X называют функцию $F_n^*(x)$, которая при каждом $x \in \mathbf{R}$ равна относительной частоте события $X < x$, т.е.

$$F_n^*(x) = \frac{n_x}{n},$$

где n_x – число элементов выборки, меньших x.

Эмпирическая функция распределения обладает следующими свойствами:

1. $0 \le F_n^*(x) \le 1$ для любого $x \in \mathbf{R}$;

2. $F_n^*(x)$ не убывает на $\mathbf{R}$;

3. $F_n^*(x) = 0$ при $x \le x_{\min}$, $F_n^*(x) = 1$ при $x > x_{\max}$, где $x_{\min}$ и $x_{\max}$ – соответственно наименьший и наибольший элементы выборки;

4. $F_n^*(x)$ непрерывна слева в любой точке $x_0 \in \mathbf{R}$.

График эмпирической функции распределения имеет ступенчатый вид. В промежутках между соседними вариантами выборки $F_n^*(x)$ сохраняет постоянное значение. В точках оси Ox, равных вариантам выборки, $F_n^*(x)$ претерпевает скачки; величина скачка в точке $x = x_i$ равна относительной частоте варианты x_i.

Пример 4.3. В условиях примера 4.1 найти эмпирическую функцию распределения с.в. X и построить ее график.

Решение. В данном случае $n = 15$. По статистическому ряду распределения находим

$$F_{15}^{*}(x) = \begin{cases} 0 & \text{при} \quad x \le -5, \\ \dfrac{1}{15} & \text{при} \quad -5 < x \le -3, \\ \dfrac{3}{15} & \text{при} \quad -3 < x \le -1, \\ \dfrac{5}{15} & \text{при} \quad -1 < x \le 0, \\ \dfrac{10}{15} & \text{при} \quad 0 < x \le 1, \\ \dfrac{13}{15} & \text{при} \quad 1 < x \le 3, \\ 1 & \text{при} \quad x > 3. \end{cases}$$

График $F_{15}^{*}(x)$ изображен на рис. 8.

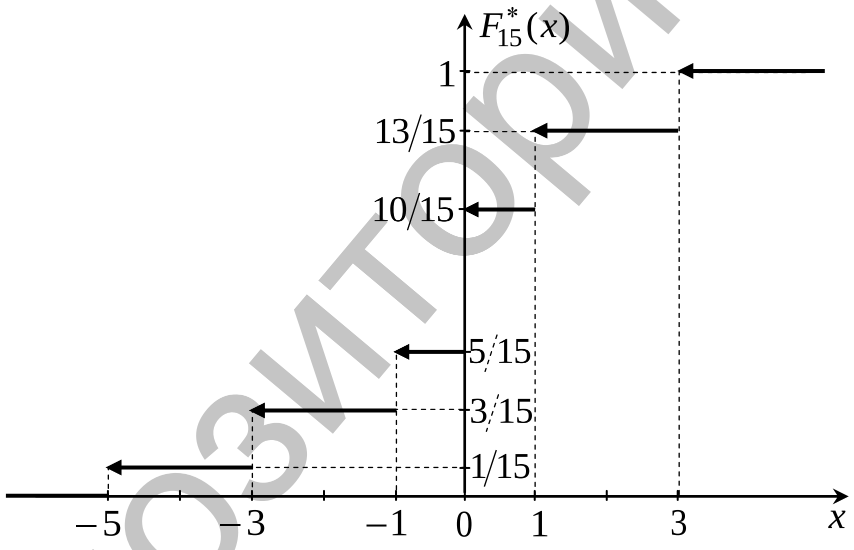

Рис. 8.

4.3. Числовые характеристики выборки

Пусть $x_1, x_2, \ldots, x_n$ – выборка объема n, полученная в результате наблюдений за с.в. X.

Выборочным средним $\bar{x}$ называется среднее арифметическое всех значений выборки:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Выборочной дисперсией $D_в$ называют среднее арифметическое квадратов отклонений значений выборки от выборочного среднего $\bar{x}$:

$$D_в = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

С помощью простых арифметических преобразований можно получить следующую формулу для вычисления выборочной дисперсии:

$$D_в = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - (\bar{x})^2.$$

Замечание 4.2. Если выборка представлена статистическим рядом распределения, то значения $\bar{x}$ и $D_в$ находят по формулам:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k}x_i n_i, \tag{4.1}$$

$$D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = \frac{1}{n}\sum_{i=1}^{k}x_i^2 n_i - (\bar{x})^2, \tag{4.2}$$

где x_i – варианты, n_i – соответствующие им частоты. ●

Замечание 4.3. Для интервального статистического ряда выборочное среднее и выборочную дисперсию также вычисляют по формулам (4.1) и (4.2). В качестве x_i берут середины интервалов ряда, а в качестве n_i – частоты соответствующих интервалов. ●

Выборочное среднее квадратическое отклонение $\sigma_в$ определяется формулой

$$\sigma_в = \sqrt{D_в}.$$

Размахом вариации называется число

$$R = x_{\max} - x_{\min},$$

где $x_{\min}$ – наименьший, а $x_{\max}$ – наибольший элементы выборки.

Модой M_o^* выборки называют варианту, имеющую наибольшую частоту.

Медианой M_e^* выборки называется число, которое делит вариационный ряд

$$x_{(1)},\ x_{(2)},\ \ldots,\ x_{(n)}$$

на две части, содержащие равное число элементов, а именно: если $n = 2k$ $(k \in \mathbf{N})$, то $M_e^* = \dfrac{x_{(k)} + x_{(k+1)}}{2}$; если $n = 2k+1$, то $M_e^* = x_{(k+1)}$.

4.4. Точечные оценки неизвестных параметров распределения

Пусть $F(x, \theta)$ – функция распределения с.в. X, θ – неизвестный параметр. Предполагается, что общий вид функции $F(x, \theta)$ задан. Таким образом, если бы параметр θ был известен, то распределение с.в. X было бы определено полностью.

Пусть в результате n наблюдений за с.в. X получена выборка

$$x_1,\ x_2,\ \ldots,\ x_n. \tag{4.3}$$

По выборке требуется найти приближенное значение параметра θ (или, как говорят, оценить параметр θ).

Элементы выборки (4.3) можно рассматривать последовательно как частные значения n независимых случайных величин

$$X_1, X_2, \ldots, X_n,$$

каждая из которых имеет тот же закон распределения, что и сама с.в. *X*.

Любая функция случайных величин $X_1, X_2, \ldots, X_n$ называется *статистикой*.

Важными примерами статистик являются *выборочное среднее*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

и *выборочная дисперсия*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Пусть $g = g(X_1, X_2, \ldots, X_n)$ – некоторая статистика. Если случайные величины $X_1, X_2, \ldots, X_n$ принимают значения, равные соответствующим элементам выборки (4.3), то статистика g принимает значение

$$g(x_1, x_2, \ldots, x_n),$$

которое называется ее *выборочным значением* (его называют также *реализацией* статистики g).

Отметим, что определенные в пункте 4.3 числовые характеристики выборки $\overline{x}$ (выборочное среднее) и $D_в$ (выборочная дисперсия) являются выборочными значениями одноименных статистик $\overline{X}$ и S^2.

Статистика $\tilde{\theta} = \tilde{\theta}(X_1, X_2, \ldots, X_n)$, выборочное значение (реализация) которой принимается за приближенное значение неизвестного параметра θ, называется его *точечной оценкой* или просто *оценкой*.

Основными свойствами оценок, характеризующими их качество, являются *несмещенность, состоятельность* и *эффективность*.

Оценку $\tilde{\theta}$ параметра θ называют *несмещенной*, если

$$M(\tilde{\theta}) = \theta.$$

В противном случае оценка называется *смещенной*.

Оценку $\tilde{\theta}_n = \tilde{\theta}_n(X_1, X_2, \ldots, X_n)$ параметра θ называют *состоятельной*, если она сходится по вероятности к θ при $n \to \infty$:

$$\tilde{\theta}_n \xrightarrow{P} \theta,$$

т.е. для любого $\varepsilon > 0$ выполняется соотношение

$$\lim_{n\to\infty} P(|\tilde{\theta}_n - \theta| \geq \varepsilon) = 0.$$

Несмещенная оценка $\tilde{\theta}$ параметра θ называется *эффективной*, если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра θ.

Теорема 4.1. Пусть наблюдаемая с.в. *X* имеет математическое ожидание и дисперсию. Тогда выборочное среднее $\overline{X}$ является несмещенной и состоятельной оценкой математического ожидания *m* с.в. *X*.

Доказательство. Поскольку случайные величины X_i распределены так же, как и сама наблюдаемая с.в. *X*, то $M(X_i) = M(X) = m$, $i = 1, 2, \ldots, n$. В силу свойств математического ожидания имеем

$$M(\overline{X}) = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(X_i) = \frac{1}{n} nm = m,$$

и значит, $\overline{X}$ – несмещенная оценка m.

Так как последовательность $X_1, X_2, \ldots, X_n$ состоит из независимых одинаково распределенных случайных величин с конечным математическим ожиданием и дисперсией, то в силу следствия теоремы Чебышева (п. 3.2)

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} m.$$

Следовательно, $\overline{X}$ – состоятельная оценка m. ◄

Замечание 4.4. Примем без доказательства следующий весьма важный для практики результат. Если наблюдаемая с.в. X имеет нормальное распределение с параметрами m и σ, то несмещенная оценка $\overline{X}$ является эффективной оценкой параметра m (математического ожидания с.в. X). При этом $D(\overline{X}) = \dfrac{\sigma^2}{n}$. ●

Теорема 4.2. Пусть наблюдаемая с.в. X имеет математическое ожидание и дисперсию. Тогда выборочная дисперсия S^2 является смещенной и состоятельной оценкой дисперсии σ^2 с.в. X.

Ограничимся лишь доказательством смещенности оценки S_X^2.

Доказательство. Преобразуем формулу выборочной дисперсии:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}[(X_i - M(X)) - (\overline{X} - M(X))]^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - M(X))^2 - 2\cdot\frac{1}{n}(\overline{X} - M(X))\sum_{i=1}^{n}(X_i - M(X)) +$$

$$+ \frac{1}{n}\cdot n(\overline{X} - M(X))^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - M(X))^2 - 2(\overline{X} - M(X))^2 + (\overline{X} - M(X))^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - M(X))^2 - (\overline{X} - M(X))^2.$$

Таким образом,

$$M(S^2) = \frac{1}{n}\sum_{i=1}^{n} M[(X_i - M(X))^2] - M[(\overline{X} - M(X))^2] =$$

$$= \frac{1}{n}\cdot nD(X) - D(\overline{X}) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2.$$

Поскольку $M(S^2) \neq \sigma^2$, то S^2 – смещенная оценка дисперсии σ^2 с.в. X. ◄

Замечание 4.5. Из теоремы 4.2 следует, что статистика

$$\widetilde{S}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{4.4}$$

является несмещенной оценкой дисперсии σ^2. Действительно,

$$S_0^2 = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{n}{n-1} S^2,$$

и значит,

$$M(S_0^2) = \frac{n}{n-1} M(S^2) = \frac{n}{n-1} \cdot \frac{n-1}{n} \sigma^2 = \sigma^2.$$

Можно доказать также, что S_0^2 является состоятельной оценкой σ^2. ●

Статистика S_0^2, определенная формулой (4.4), называется *исправленной выборочной дисперсией*.

Отметим, что при больших значениях n выборочные значения статистик S^2 и S_0^2 отличаются мало. Поэтому на практике оценку S_0^2 используют для оценки дисперсии в основном лишь при малых объемах выборки (обычно при $n \le 30$).

Замечание 4.6. Оценки S^2 и S_0^2 не являются эффективными. В случае, когда известно математическое ожидание m наблюдаемой с.в. X, несмещенной, состоятельной и эффективной оценкой σ^2 является оценка

$$S_1^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - m)^2. \text{ ●}$$

4.5. Метод максимального правдоподобия

Рассмотрим один из наиболее распространенных методов получения точечных оценок неизвестных параметров распределения – метод максимального правдоподобия.

Предположим, что известен вид закона распределения с.в. X, но неизвестен параметр θ, которым определяется этот закон (параметр θ может быть и векторным: $\theta = (\theta_1, \theta_2, \dots, \theta_r)$). Пусть $x_1, x_2, \dots, x_n$ – выборка наблюдений с.в. X, по которой требуется оценить параметр θ.

Функцией правдоподобия для оценки параметра θ называется функция

$$L(\theta) = p(x_1, \theta)\, p(x_2, \theta) \dots p(x_n, \theta),$$

где $p(x, \theta)$ – плотность распределения с.в. X, если X – непрерывная с.в., и $p(x, \theta) = P(X = x)$, если с.в. X дискретная.

Метод максимального правдоподобия состоит в том, что в качестве оценки параметра θ принимается статистика $\tilde{\theta}$ (векторная статистика $\tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_r)$, если $\theta = (\theta_1, \dots, \theta_r)$), выборочное значение (реализация) которой $\tilde{\theta}(x_1, x_2, \dots, x_n)$ является точкой максимума функции правдоподобия. Такую оценку называют *оценкой максимального правдоподобия*.

Так как максимум функций $L(\theta)$ и $\ln L(\theta)$ достигается при одном и том же значении θ, то часто при нахождении оценки максимального правдоподобия используют функцию $\ln L(\theta)$: эта функция более удобна для вычислений.

Пример 4.4. Найти оценку максимального правдоподобия параметра λ распределения Пуассона.

Решение. В данном случае X – дискретная с.в.,

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \quad k=0,1,2,\ldots \quad (\lambda>0).$$

Составляем функцию правдоподобия:

$$L(\lambda)=\frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\,\frac{\lambda^{x_2}}{x_2!}e^{-\lambda}\ldots\frac{\lambda^{x_n}}{x_n!}e^{-\lambda}=e^{-\lambda n}\cdot\lambda^{\sum_{i=1}^{n}x_i}\cdot\frac{1}{x_1!x_2!\ldots x_n!}.$$

Следовательно,

$$\ln L(\lambda)=-\lambda n+\ln\lambda\cdot\left(\sum_{i=1}^{n}x_i\right)-\ln(x_1!x_2!\ldots x_n!).$$

Ищем точку максимума функции $\ln L(\lambda)$:

$$\frac{d\ln L(\lambda)}{d\lambda}=-n+\frac{1}{\lambda}\sum_{i=1}^{n}x_i=0.$$

Отсюда получаем

$$\lambda=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$$

Легко видеть, что $\lambda=\bar{x}$ является точкой максимума функции $\ln L(\lambda)$.

Таким образом, $\tilde{\lambda}=\overline{X}$ – искомая оценка. ●

4.6. Распределения χ^2 (хи-квадрат) и Стьюдента

Рассмотрим два широко распространенных в математической статистике распределения – распределение χ^2 и распределение Стьюдента. Приведем также используемые в дальнейшем теоремы о статистиках, имеющих эти распределения.

Распределение χ^2. Пусть случайные величины $X_1, X_2, \ldots, X_n$ независимы и распределены нормально с параметрами $m=0$ и $\sigma=1$. *Распределением χ^2 (или χ^2-распределением) с n степенями свободы* называется закон распределения с.в.

$$\chi_n^2=\sum_{i=1}^{n}X_i^2.$$

С возрастанием числа степеней свободы *n* распределение χ^2 приближается к нормальному (при $n>30$ распределение χ^2 практически не отличается от нормального).

Для $n\le 30$ составлена таблица, содержащая такие значения $\chi_n^2=\chi_{\alpha,n}^2$, для которых

$$P(\chi_n^2>\chi_{\alpha,n}^2)=\int_{\chi_{\alpha,n}^2}^{+\infty}p(x)dx=\alpha$$

(α – заданный уровень вероятности). Значение $\chi_{\alpha,n}^2$ называется *квантилью распределения* χ^2, отвечающей заданному уровню вероятности α и заданному уровню степеней свободы *n*.

Для $n>30$ и $\alpha\le 0{,}5$ применяют формулу

$$\chi^2_{\alpha,n} = 0{,}5(u_\alpha + \sqrt{2n-1})^2,$$

где u_α (*квантиль уровня* α *нормального распределения с параметрами* $m=0$, $\sigma=1$) находится по таблице функции Лапласа из условия

$$\Phi(u_\alpha) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u_\alpha} e^{-\frac{t^2}{2}}dt = 1-\alpha.$$

Теорема 4.3. Пусть наблюдаемая с.в. *X* распределена нормально с параметрами m и σ. Тогда статистика

$$\frac{(n-1)S_0^2}{\sigma^2}$$

имеет распределение χ^2 с $n-1$ степенями свободы. Здесь *n* – объем выборки, S_0^2 – исправленная выборочная дисперсия (п. 4.4).

Распределение Стьюдента. Пусть $X_0, X_1, \ldots, X_n$ – независимые случайные величины, распределенные нормально с параметрами $m=0$ и $\sigma=1$. *Распределением Стьюдента с n степенями свободы* называется закон распределения с.в.

$$T_n = \frac{X_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}}.$$

При $n \to \infty$ распределение Стьюдента приближается к нормальному с параметрами $m=0$, $\sigma=1$. Практически уже при $n>30$ распределение Стьюдента можно считать приближенно нормальным.

Составлена таблица *квантилей* $t_{\alpha/2,n}$ *распределения Стьюдента*, отвечающих числу степеней свободы *n* и заданному уровню вероятности α. Значение квантили $t_{\alpha/2,n}$ определяется из условия

$$P(|T_n| > t_{\alpha/2,n}) = 2\int_{t_{\alpha/2,n}}^{+\infty} p(x)dx = \alpha.$$

Теорема 4.4. Пусть наблюдаемая с.в. *X* распределена нормально с параметрами m и σ. Тогда статистика

$$\frac{(\overline{X}-m)\sqrt{n}}{S_0}$$

имеет распределение Стьюдента с $n-1$ степенями свободы (*n* – объем выборки, $\overline{X}$ – выборочное среднее, $S_0 = \sqrt{S_0^2}$ – *исправленное выборочное среднее квадратическое отклонение*).

4.7. Интервальное оценивание. Доверительные интервалы для параметров нормального распределения

Пусть θ – неизвестный параметр распределения наблюдаемой с.в. *X*.

Пусть статистики $\tilde{\theta}_1 = \tilde{\theta}_1(X_1, \ldots, X_n)$ и $\tilde{\theta}_2 = \tilde{\theta}_2(X_1, \ldots, X_n)$ таковы, что интервал $(\tilde{\theta}_1, \tilde{\theta}_2)$ содержит (накрывает) истинное значение параметра θ с заданной вероятностью $p = 1-\alpha$:

$$P(\tilde{\theta}_1 < \theta < \tilde{\theta}_2) = 1 - \alpha .$$

Тогда интервал $(\tilde{\theta}_1, \tilde{\theta}_2)$ называется *доверительным интервалом* для параметра θ, вероятность $p = 1 - \alpha$ – *доверительной вероятностью (надежностью)*, а число α – *уровнем значимости*.

Статистики $\tilde{\theta}_1$ и $\tilde{\theta}_2$ называются соответственно *нижней* и *верхней границами* доверительного интервала.

Доверительный интервал для математического ожидания нормально распределенной с.в. Пусть с.в. X распределена нормально с параметрами m и σ. Построим доверительный интервал с заданной доверительной вероятностью $1-\alpha$ для параметра m (математического ожидания с.в. X). Этот доверительный интервал будем искать в виде

$$(\overline{X} - \delta,\ \overline{X} + \delta),$$

где $\overline{X}$ – выборочное среднее (несмещенная оценка параметра m (см. теорему 4.1)). Таким образом, необходимо найти такое $\delta > 0$, чтобы выполнялось соотношение

$$P(\overline{X} - \delta < m < \overline{X} + \delta) = 1 - \alpha;$$

перепишем его в следующем виде:

$$P(|\,\overline{X} - m\,| < \delta) = 1 - \alpha. \tag{4.5}$$

Будем рассматривать два случая: когда параметр σ (среднее квадратическое отклонение с.в. X) известен и когда этот параметр неизвестен.

Случай 1 (параметр σ известен). Известно (мы примем этот факт без доказательства), что с.в.

$$Z_n = \frac{(\overline{X} - m)\sqrt{n}}{\sigma}$$

распределена нормально с параметрами 0 и 1. Поэтому

$$P(|\,\overline{X} - m\,| < \delta) = P\left(|\,Z_n\,| < \frac{\delta\sqrt{n}}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\frac{\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}} e^{-\frac{x^2}{2}} dx = 2\Phi_0\left(\frac{\delta\sqrt{n}}{\sigma}\right),$$

где $\Phi_0(x)$ – нормированная функция Лапласа (п. 1.13). В силу условия (4.5) должно выполняться равенство

$$\Phi_0\left(\frac{\delta\sqrt{n}}{\sigma}\right) = \frac{1-\alpha}{2}.$$

По таблице функции $\Phi_0(x)$ находим значение $u_{\alpha/2}$, для которого

$$\Phi_0(u_{\alpha/2}) = \frac{1-\alpha}{2}$$

Из равенства $\dfrac{\delta\sqrt{n}}{\sigma} = u_{\alpha/2}$ находим $\delta = u_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$. Таким образом, $\left(\overline{X} - u_{\alpha/2}\dfrac{\sigma}{\sqrt{n}},\ \overline{X} + u_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)$ – искомый доверительный интервал.

Случай 2 (параметр σ неизвестен). Рассмотрим с.в.

$$T = \frac{(\overline{X} - m)\sqrt{n}}{S_0},$$

где $S_0 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ (исправленное выборочное среднее квадратическое отклонение). Согласно теореме 4.4 с.в. T распределена по закону Стьюдента с $n-1$ степенями свободы.

Поскольку $|\overline{X} - m| < \delta \quad \Leftrightarrow \quad |T| < \frac{\delta\sqrt{n}}{S_0}$, то δ следует искать таким, чтобы выполнялось соотношение $P\left(|T| < \frac{\delta\sqrt{n}}{S_0}\right) = 1-\alpha$.

По таблице квантилей распределения Стьюдента находим значение $t_{\alpha/2,n-1}$, для которого

$$P\left(|T| < t_{\alpha/2,n-1}\right) = 1-\alpha,$$

что равносильно

$$P\left(|T| > t_{\alpha/2,n-1}\right) = \alpha.$$

(значение $t_{\alpha/2,n-1}$ ищется в зависимости от уровня значимости α и числа степеней свободы $n-1$). Из равенства $\frac{\delta\sqrt{n}}{S_0} = t_{\alpha/2,n-1}$ получаем

$$\delta = t_{\alpha/2,n-1}\frac{S_0}{\sqrt{n}},$$

и, следовательно, искомый доверительный интервал имеет вид

$$\left(\overline{X} - t_{\alpha/2,n-1}\frac{S_0}{\sqrt{n}},\ \overline{X} + t_{\alpha/2,n-1}\frac{S_0}{\sqrt{n}}\right).$$

Доверительный интервал для среднего квадратического отклонения нормально распределенной с.в. Найдем доверительный интервал с заданной доверительной вероятностью $1-\alpha$ для среднего квадратического отклонения σ нормально распределенной с.в. X.

В силу теоремы 4.3 с.в.

$$V^2 = \frac{(n-1)S_0^2}{\sigma^2}$$

имеет распределение χ^2 с $n-1$ степенями свободы. Выберем такие c_1 и c_2, для которых

$$P(V^2 < c_1) = \frac{\alpha}{2}, \qquad P(V^2 > c_2) = \frac{\alpha}{2}.$$

Искомые значения c_1 и c_2 находятся по таблице квантилей распределения χ^2:

$$c_1 = \chi^2_{1-\alpha/2,n-1}, \quad c_2 = \chi^2_{\alpha/2,n-1}.$$

Заметим, что

$$P(c_1 < V^2 < c_2) = 1 - P(V^2 < c_1) - P(V^2 > c_2) = 1-\alpha,$$

т.е.

$$P\left(c_1 < \frac{(n-1)S_0^2}{\sigma^2} < c_2\right) = 1-\alpha .$$

Отсюда следует, что $P\left(S_0\sqrt{\frac{n-1}{c_2}} < \sigma < S_0\sqrt{\frac{n-1}{c_1}}\right) = 1-\alpha$. Таким образом, $\left(S_0\sqrt{\frac{n-1}{\chi^2_{\alpha/2,n-1}}},\ S_0\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2,n-1}}}\right)$ – искомый доверительный интервал.

4.8. Проверка гипотез о виде закона распределения. Критерий согласия χ^2 Пирсона

Статистической гипотезой или просто *гипотезой* называется любое предположение о распределении наблюдаемой с.в., проверяемое по выборке.

Статистические гипотезы о виде неизвестного закона распределения называют *непараметрическими* (в отличие от *параметрических* гипотез, являющихся предположениями о параметрах распределения известного вида).

Пусть необходимо проверить непараметрическую гипотезу, состоящую в том, что наблюдаемая в эксперименте с.в. X распределена по некоторому известному закону (нормальному, биномиальному, Пуассона и т.д.). Проверяемая гипотеза называется *нулевой* и обозначается H_0.

Для проверки гипотезы H_0 производится выборка значений с.в. X. Требуется сделать заключение: согласуются ли данные выборки с высказанным предположением. Для этого используют специально подобранную статистику $Z(X_1, \ldots, X_n)$, по выборочному значению которой судят о справедливости гипотезы H_0.

Правило, согласно которому проверяемая гипотеза H_0 принимается или отвергается, называется *критерием согласия*, а статистика $Z(X_1, \ldots, X_n)$, с помощью которой это правило определяется, называется *статистикой критерия*.

Поскольку проверка статистической гипотезы осуществляется на основании выборочных данных, носящих случайный характер, то всегда присутствует возможность ошибочно отвергнуть гипотезу H_0, когда на самом деле она верна. Поэтому заранее задается малое число α – вероятность, с которой мы можем позволить себе отвергнуть верную гипотезу H_0. Это число называют *уровнем значимости критерия*. Обычно для α используются стандартные значения: $\alpha = 0{,}001;\ 0{,}005;\ 0{,}01;\ 0{,}05;\ 0{,}1$.

Критериев согласия существует много. Мы рассмотрим наиболее часто используемый на практике критерий согласия – *критерий* χ^2 *Пирсона*.

Предположим, что сформулирована гипотеза H_0, состоящая в том, что с.в. X имеет закон распределения известного вида, зависящий от параметров $\theta_1, \ldots, \theta_r$ (например, нормальный закон с параметрами m и σ или закон Пуассона с параметром λ).

Пусть в результате наблюдений за с.в. X получена выборка объема n: $x_1, x_2, \ldots, x_n$.

Для проверки гипотезы H_0 с помощью критерия χ^2 Пирсона поступают следующим образом.

1. Выбирают уровень значимости α.

2. По выборке методом максимального правдоподобия (п. 4.5) оценивают параметры предполагаемого закона распределения.

3. Область возможных значений с.в. X разбивают на k непересекающихся множеств $S_1, S_2, \ldots S_k$. Эти множества представляют собой интервалы в случае, когда X – непрерывная с.в., либо группы отдельных значений, если с.в. X дискретная.

4. Для каждого множества S_i, $i = 1, 2, \ldots, k$ подсчитывают число n_i элементов выборки, попавших в это множество (т.е. находят *эмпирические частоты*).

5. Используя предполагаемый закон распределения с.в. X, вычисляют гипотетические вероятности p_i попадания с.в. X в множества S_i:

$$p_i = P(X \in S_i), \quad i = 1, 2, \ldots, k.$$

6. Находят *теоретические частоты* n_i' попадания значений с.в. X в множества S_i: $$n_i' = np_i.$$

7. Вычисляют выборочное значение $\chi_в^2$ статистики критерия χ^2 Пирсона – случайной величины

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n_i')^2}{n_i'}. \qquad (4.6)$$

Применение статистики (4.6) для проверки гипотезы H_0 основано на следующей теореме.

Теорема 4.5. Если гипотеза H_0 верна, то статистика (4.6) критерия χ^2 асимптотически при $n \to \infty$ распределена по закону χ^2 с $\nu = k - r - 1$ степенями свободы, где k – число множеств S_i, r – число параметров предполагаемого распределения, оцененных по выборке.

8. По таблице квантилей χ^2 – распределения по заданному уровню значимости α и числу степеней свободы $\nu = k - r - 1$ находят значение $\chi^2_{\alpha,\nu}$, удовлетворяющее условию $P(\chi_\nu^2 > \chi^2_{\alpha,\nu}) = \alpha$.

9. Сравнивают значения $\chi_в^2$ и $\chi^2_{\alpha,\nu}$. В соответствии с критерием χ^2 Пирсона гипотеза H_0 принимается, если $\chi_в^2 \le \chi^2_{\alpha,\nu}$ (в этом случае говорят также, что гипотеза H_0 согласуется с результатами наблюдений). Если $\chi_в^2 > \chi^2_{\alpha,\nu}$, то гипотеза H_0 отклоняется.

Замечание 4.7. Необходимым условием применения критерия χ^2 Пирсона является выполнение неравенства $n_i' \ge 5$ для всех множеств S_i. Если для некоторых множествах S_i это условие не выполняется, то их следует объединить с соседними. ●

ЛИТЕРАТУРА

1. Баврин И.И. Теория вероятностей и математическая статистика: учебник / И.И. Баврин. – М.: Высш. шк., 2005. – 160 с.: ил.

2. Боровков А.А. Теория вероятностей / А.А. Боровков. – 4-е изд. – М.: УРСС, 2003.

3. Бочаров П.П., Печинкин А.В. Теория вероятностей. Математическая статистика / П.П. Бочаров, А.В. Печинкин. – 2-е изд. – М.: ФИЗМАТЛИТ, 2005. – 296 с.

4. Герасимович А.И., Матвеева Я.И. Математическая статистика / А.И. Герасимович, Я.И. Матвеева. – Мн.: Выш. шк., 1978. – 200 с.; ил.

5. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В.Е. Гмурман. – 9-е изд., стер. – М.: Высш. шк., 2003. – 479 с.: ил.

6. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для студентов вузов / В.Е. Гмурман. – 9-е изд., стер. – М.: Высш. шк., 2004. – 404 с.: ил.

7. Гнеденко Б.В. Курс теории вероятностей: учебник / Б.В. Гнеденко. – 8-е изд., испр. и доп. – М.: Едиториал УРСС, 2005. – 448 с.

8. Кремер Н.Ш. Теория вероятностей и математическая статистика: учебник для вузов / Н.Ш. Кремер. – 2-е изд., перераб. и доп. – М.: ЮНИТИ-ДАНА, 2004. – 573 с.

9. Теория вероятностей и математическая статистика в задачах: учеб. пособие для вузов / В.А. Ватутин, Г.И. Ивченко, Ю.И. Медведев и др. – 2-е изд., испр. – М.: Дрофа, 2003. – 328 с.: ил.

10. Чистяков В.П. Курс теории вероятностей / В.П. Чистяков. – М.: Агар, 2000. – 255 с.