Тема 2. Законы распределения случайных величин

1.2.1. Дискретные случайные величины

Во многих прикладных задачах интерес представляет не сам исход эксперимента, а некоторое число, связанное с этим исходом. Например, при подбрасывании двух игральных костей интерес представляет сумма выпавших очков. Тогда каждому исходу (1,1), (1,2), …, (5,1), …, (6,6) будет соответствовать число, равное сумме очков, т. е. 2, 3, …, 6, …, 12. При двух выстрелах по цели интерес представляет число попаданий. Выборочное пространство включает в себя следующие исходы: $E = \{ПП, ПН, НП, НН\}$, где П – попал, Н – не попал.

Каждому исходу можно поставить в соответствие следующие числа: 2, 1, 1, 0.

Определение. *Случайной величиной называется числовая функция* $X(e_i)$, *где* e_i *– исходы эксперимента.*

В литературе также встречается еще одно определение случайной величины как величины, которая в результате опыта может принимать то или иное значение, причем заранее (до опыта) неизвестное.

Случайные величины обозначаются заглавными буквами X, Y, Z, …, а их конкретные значения – маленькими буквами x, y, z…

Пример. Два носка выбираются случайным образом из ящика, в котором находится 5 коричневых и 3 зеленых. Записать элементы выборочного пространства, соответствующие вероятности, и значения случайной величины X, где X – число коричневых носков.

Решение

Пусть событие B состоит в вытаскивании коричневого носка.

Событие G состоит в вытаскивании зеленого носка.

Результаты можно представить в табл. 1.1.

Таблица 1.1

Элементы выборочного пространства	*BB*	*BG*	*GB*	*GG*
Значения случай-	2	1	1	0

ной величины				
Вероятности	$\frac{5}{8}\cdot\frac{4}{7}=\frac{5}{14}=0.35$	$\frac{5}{8}\cdot\frac{3}{7}=\frac{15}{56}=0.27$	$\frac{3}{8}\cdot\frac{5}{7}=\frac{15}{56}=0.27$	$\frac{3}{8}\cdot\frac{2}{7}=\frac{3}{28}=0.11$

Таким образом, для случайной величины, принимающей, например, значение 2, можно написать $P(X=2)=\frac{5}{14}$.

В данном примере рассмотрена случайная величина *дискретного* типа.

Определение. *Дискретной случайной величиной называется величина, которая принимает конечное или бесконечное счетное множество значений.*

Определение. *Для дискретной случайной величины X функция* $f(x)=P(X=x)$, *заданная для каждого значения x, называется законом распределения вероятностей для случайной величины X.*

Из аксиом вероятностей следует, что функция $f(x)$ может быть законом распределения случайной величины X тогда, и только тогда, когда выполняются следующие условия:

1. $f(x)\geq 0$ для каждого возможного значения X.
2. $\sum\limits_{x} f(x)=1$, суммирование для всех возможных значений X.

Закон распределения $f(x)=P(X=x)$, представленный для указанного примера в табл. 1.2, называется еще *рядом распределения*.

Таблица 1.2

Ряд распределения случайной величины

Значение x_i	2	1	0	
Pi	0.35	0.27 + 0.27 = 0.54	0.11	$\sum p_i=1$

Ряд распределения представляется графически в виде *многоугольника распределения*. В этом случае по оси x откладываются все возможные значения X, а по оси y – соответствующие им вероятности. Для примера с выбором носков многоугольник распределения имеет вид (рис. 1.5).

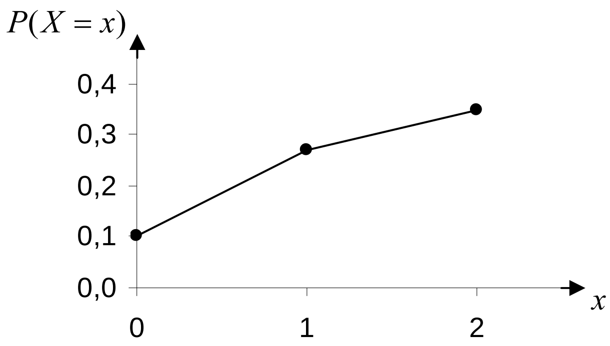

Рис. 1.5. Многоугольник распределения
числа вынутых коричневых носков

Пример. Проверить, является ли функция $f(x) = \dfrac{x+2}{25}$ законом распределения дискретной случайной величины для x, принимающего значения 1, 2, 3, 4, 5.

Решение

Подставляя различные значения x в $f(x)$, получаем $f(1) = \dfrac{3}{25}$; $f(2) = \dfrac{4}{25}$; $f(3) = \dfrac{5}{25}$; $f(4) = \dfrac{4}{25}$; $f(5) = \dfrac{5}{25}$.

Так как все значения функции $f(x)$ неотрицательны и в сумме дают

$$f(1) + f(2) + f(3) + f(4) + f(5) = \frac{3}{25} + \frac{4}{25} + \frac{5}{25} + \frac{6}{25} + \frac{7}{25} = 1,$$

то $f(x)$ может являться законом распределения дискретной случайной величины, принимающей значения $\{1,2,3,4,5\}$.

1.2.2. Функция распределения дискретной случайной величины

Во многих задачах необходимо знать вероятность того, что случайная величина X примет значение меньше (или равное) некоторого значения x.

Запишем вероятность того, что X принимает значение меньше (или равное) x как $F(x) = P(X \le x)$.

Определение. *Функция* $F(x) = P(X \le x)$ *называется функцией распределения случайной величины X.*

Для дискретной случайной величины

$$F(x) = \sum_{x_i < x} P(x_i), \qquad (1.23)$$

где x_i – значения, принимаемые случайной величиной *X*;

$P(x_i)$ – значения вероятностей при $X = x_i$;

x – некоторое фиксированное значение *X*.

Свойства функции распределения вытекают из аксиом вероятностей и их следствий и определяются в виде:

1. $F(-\infty)=0$, $F(+\infty)=1$.

2. *Если* $a<b$, *то* $F(a)\le F(b)$ *для любых значений a и b* (*функция неубывающая*).

При решении практических задач часто возникает необходимость вычислять вероятность того, что случайная величина примет значение в интервале от *a* до *b* или $P(a\le X<b)$. Эта вероятность выражается через $F(x)$ следующим образом.

Пусть событие *A* состоит в том, что $X<b$, событие *B* состоит в том, что $X<a$, событие *C* состоит в том, что $a\le X<b$.

Учитывая, что $A=B+C$, по теореме сложения вероятностей имеем

$P(X<b)=P(X<a)+P(a\le X<b)$ или

$F(b)=F(a)+P(a\le X<b)$,

отсюда

$$P(a\le X<b)=F(b)-F(a). \tag{1.24}$$

Пример. Определить функцию распределения числа гербов при четырех подбрасываниях монеты.

Решение

Объем выборочного пространства в данном случае $n=16$:

$f(0)=P(X=0)=\frac{1}{16}$; $f(1)=P(X=1)=\frac{4}{16}$;

$f(2)=P(X=2)=\frac{6}{16}$;

$f(3)=P(X=3)=\frac{4}{16}$; $f(4)=P(X=4)=\frac{1}{16}$.

Ряд распределения в данном случае представлен в табл. 1.3.

Таблица 1.3

Ряд распределения случайной величины *X*

x	0	1	2	3	4

p	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$	$\sum p_i = 1$

Используя формулу (1.23), определим значения функции распределения $F(x)$:

$$F(0) = f(0) = 1/16\ ;$$

$$F(1) = f(0) + f(1) = 5/16\ ;$$

$$F(2) = f(0) + f(1) + f(2) = 11/16\ ;$$

$$F(3) = f(0) + f(1) + f(2) + f(3) = 15/16\ ;$$

$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1\ .$$

Функция распределения представляется в следующем виде:

$$F(x) = \begin{cases} 0, & x < 0; \\ \dfrac{1}{16}, & 0 \le x < 1; \\ \dfrac{5}{16}, & 1 \le x < 2; \\ \dfrac{11}{16}, & 2 \le x < 3; \\ \dfrac{15}{16}, & 3 \le x < 4; \\ 1, & x \ge 4. \end{cases}$$

Данная функция распределения может быть определена не только для значений, принимаемых случайной величиной, но и для всех вещественных чисел. Например:

$F(1.7) = \dfrac{5}{16}$; $F(100) = 1$; $F(-1) = 0$ и т. д.

Рассмотрим построение графика функции распределения для предыдущего примера.

По оси x откладываются значения случайной величины X, по оси y – значения функции $F(x)$. Полученный график представлен на рис. 1.6.

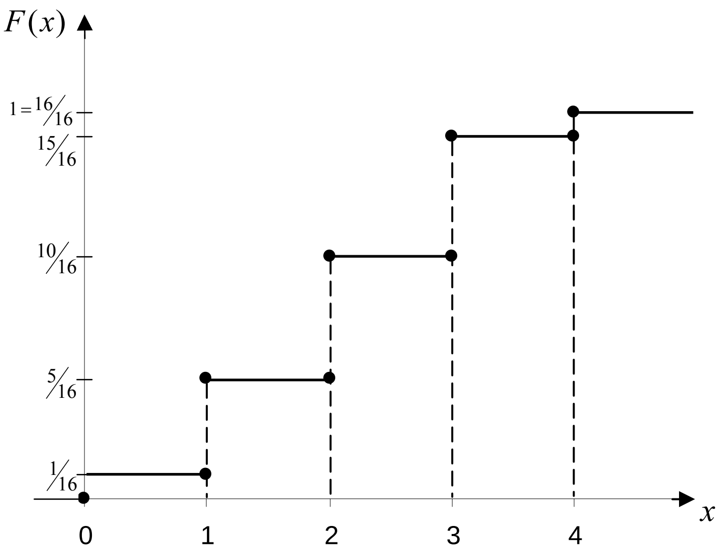

Рис. 1.6. График функции распределения числа выпавших гербов при четырех подбрасываниях монеты

Во всех точках разрыва функция распределения делает скачок, равный вероятности того, что случайная величина X примет данное значение.

Для дискретной случайной величины X, принимающей значения $x_1 < x_2 < \ldots < x_n$, $f(x_1) = F(x_1)$ и $f(x_i) = F(x_i) - F(x_{i-1})$, где $i = 2, 3, \ldots, n$.

Функция распределения $F(x)$ является универсальной характеристикой случайной величины. С помощью функции распределения задается закон распределения как для дискретных, так и для непрерывных случайных величин.

1.2.2. Непрерывные случайные величины

Определение. *Непрерывной случайной величиной называется случайная величина, которая может принять любое значение из заданного числового отрезка.*

Вероятность любого отдельного значения *непрерывной случайной* величины равна нулю. Этот вывод можно получить из соотношения (1.24), согласно которому $P(a \le X < b) = F(b) - F(a)$ для дискретных случайных величин.

Если неограниченно уменьшать отрезок (a,b), полагая $b \to a$, то в пределе получим не вероятность попадания на участок, а вероятность того, что случайная величина $X = a$, т. е.

$$P(X = a) = \lim_{b \to a} P(a \le X < b) = \lim_{b \to a} [F(b) - F(a)].$$

Если функция $F(x)$ непрерывна в точке *a*, то этот предел равен нулю.

Непрерывными случайными величинами называют еще величины, функция распределения которых везде непрерывна. Таким образом, обладать нулевой вероятностью могут не только невозможные (как определялось ранее), а и возможные события. Это появляется при рассмотрении экспериментов, не сводящихся к схеме случаев.

Как указывалось ранее, закон распределения для непрерывной случайной величины может быть задан с помощью функции распределения.

Кроме этого, для задания закона распределения *непрерывной случайной величины* используется функция $f(x) = F'(x)$, которая называется *плотностью вероятности* и которая является производной от функции распределения. Поэтому ее еще называют *дифференциальной* функцией, а функцию распределения называют *интегральной* функцией.

Кривая, изображающая плотность распределения, называется *кривой распределения*; пример кривой распределения представлен на рис. 1.7.

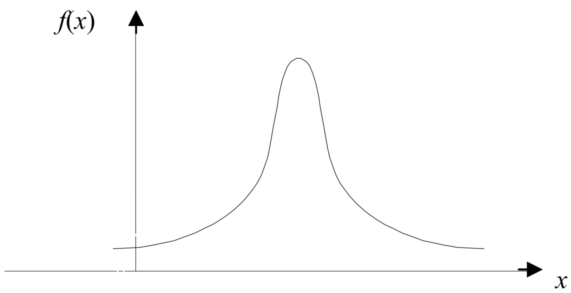

Рис. 1.7. График плотности распределения, или кривая распределения

Вероятность попадания непрерывной случайной величины на отрезок от *a* до *b* определяется в виде

$$P(a < X < b) = \int_a^b f(x)dx. \qquad (1.25)$$

Геометрически вероятность попадания случайной величины X на участок (a,b) равна площади под кривой распределения, опирающейся на этот участок (см. рис. 1.8).

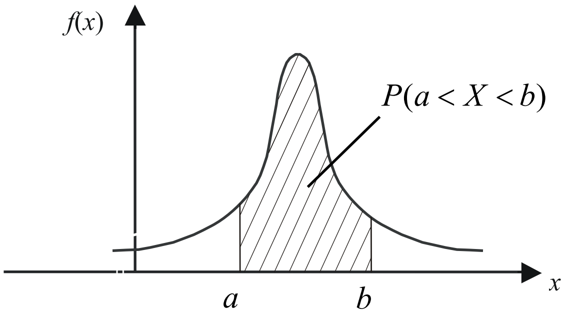

Рис. 1.8. Геометрическая интерпретация вероятности попадания случайной величины на отрезок от a до b

Заметим, что $f(c)$ – величина плотности распределения в точке c – не определяет значение $P(X=c)$, как в случае дискретной случайной величины. Для непрерывных случайных величин вероятность определяется для некоторого интервала и, как уже указывалось, $P(X=c)=0$ для любого значения c.

Из этого следует, что не имеет значения, включаются точки a и b в интервал или нет, т. е.

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

1.2.3. Свойства плотности вероятности

Из аксиом вероятности следует, что плотность вероятности $f(x)$ непрерывной случайной величины X удовлетворяет:

1. $f(x)\ge 0$ для $-\infty<x<+\infty$.
2. $\int\limits_{-\infty}^{+\infty} f(x)dx=1$.

Функция распределения непрерывной случайной величины X с плотностью распределения $f(x)$ определяется в виде

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt \tag{1.26}$$

для $-\infty < x < +\infty$.

Свойства функции распределения для непрерывной случайной величины такие же, как для дискретной случайной величины, т. е.

$F(-\infty) = 0$, $F(+\infty) = 1$;

$F(a) \le F(b)$, если $a < b$;

$P(a \le X \le b) = F(b) - F(a)$.

1.2.4. Числовые характеристики случайных величин

К основным числовым характеристикам относятся характеристики положения и характеристики рассеяния значений случайной величины.

Формулы для определения этих характеристик зависят от того, является ли случайная величина дискретной или непрерывной.

Основной характеристикой положения или расположения случайной величины является *математическое ожидание*, обозначаемое $M(x)$ и определяемое по следующим формулам:

$$M(X) = \sum_{i}^{k} x_i p_i \quad \text{для дискретных случайных величин;} \tag{1.27}$$

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx \quad \text{для непрерывных случайных величин.} \tag{1.28}$$

В формуле (1.27) x_i – возможные значения случайной величины, p_i – соответствующие им вероятности.

В формуле (1.28) $f(x)$ – плотность распределения случайной величины.

Предполагается, что сумма $\sum x_i p_i$ и $\int_{-\infty}^{+\infty} xf(x)dx$ абсолютно сходятся, в противном случае $M(x)$ не существует.

Пример. В студенческой группе организована лотерея. Разыгрываются две вещи стоимостью по 10 руб. и одна стоимостью 30 руб. Определить математическое ожидание чистого выигрыша для студента, если он приобрел 1 билет стоимостью 1 руб., а всего билетов 50.

Решение. Пусть X – случайная величина, характеризующая сумму чистого выигрыша для студента.

X может принять значение: 1, если студент ничего не выиграет;
9, если его выигрыш – 10 руб.;
29, если его выигрыш – 30 руб.

Чтобы определить математическое ожидание выигрыша, необходимо определить вероятность каждого выигрыша:

$$P(X=-1)=\frac{47}{50}=0.94;\ P(X=9)=\frac{2}{50}=0.04;\ P(X=29)=\frac{1}{50}=0.02.$$

Закон распределения случайной величины X имеет вид

X	-1	9	29
p	0.94	0.04	0.02

$$\sum p_i = 0.94+0.04+0.02=1$$

$$M[X]=\sum_{i-1}^{3} x_i p_i = -1\cdot 0.94 + 9\cdot 0.04 + 29\cdot 0.02 =$$
$$= -0.94 + 0.36 + 0.58 = 0.$$

Пример. Случайная величина X, принимающая значения размеров диаметра болта, имеет плотность распределения

$$f(x)=\begin{cases} \dfrac{4}{\pi(1+x^2)} & \text{для } 0<x<1; \\ 0 & \text{в противном случае.} \end{cases}$$

Определить математическое ожидание случайной величины X.

Решение. Так как случайная величина X непрерывного типа, то

$$M[X]=\int_0^1 x\cdot\frac{4}{\pi(1+x^2)}dx=\frac{4}{\pi}\int_0^1\frac{x}{1+x^2}dx=\frac{\ln 4}{\pi}=0.4413.$$

1.2.5. Свойства математических ожиданий

1. *Математическое ожидание постоянной величины равно этой постоянной.*

Доказательство

Постоянную величину a можно рассматривать как случайную, которая принимает лишь одно значение a с вероятностью 1, поэтому $Ma = a\cdot 1 = a$.

2. *Постоянный множитель можно выносить за знак математического ожидания, т. е.* $M[kX]=kM[X]$.

Доказательство

KX – это случайная величина, которая принимает значение Kx_i и $P(KX = kx_i) = p_i$, $i = 1,2,...,\ n$. Математическое ожидание KX:

$$M[KX] = kx_1p_1 + kx_2p_2 + ... + kx_np_n =$$
$$= k(x_1p_1 + ... + k_np_n) = KM[X].$$

Следующие свойства приводятся без доказательства.

3. *Математическое ожидание суммы конечного числа случайных величин равно сумме их математических ожиданий*:

$$M[X + Y] = M[X] + M[Y].$$

Следствие. Математическое ожидание разности случайных величин равно разности их математических ожиданий:

$$M[X - Y] = M[X] - M[Y].$$

4. *Математическое ожидание произведения конечного числа независимых случайных величин равно произведению их математических ожиданий:*

$$M[XY] = M[X] \cdot M[Y].$$

5. *Если все значения случайной величины X уменьшить* (*увеличить*) *на одно и то же число C, то математическое ожидание её уменьшится* (*увеличится*) *на то же число C.*

Следствие. Математическое ожидание отклонения случайной величины X от её математического ожидания равно нулю.

1.2.6. Дисперсия случайной величины

Математическое ожидание не может в достаточной степени характеризовать случайную величину.

Пусть случайные величины X и Y заданы следующими законами распределения, представленными в нижеприведенных таблицах.

X	-0.1	-0.01	0	0.01	0.1
p	0.1	0.2	0.4	0.2	0.1

Y	-20	-10	0	10	20
p	0.3	0.1	0.2	0.1	0.3

Математические ожидания их одинаковы и равны нулю:

$$M[X] = -0.1 \cdot 0.1 - 0.01 \cdot 0.2 + 0 \cdot 0.4 + 0.01 \cdot 0.2 + 0.1 \cdot 0.1 = 0;$$

$$M[Y] = -20 \cdot 0.3 - 10 \cdot 0.1 + 0 \cdot 0.2 + 10 \cdot 0.1 + 20 \cdot 0.3 = 0.$$

Однако характер распределения их различный. Случайная величина X может принимать значения, мало отличающиеся от математического ожидания.

Случайная величина Y может принимать значения, значительно отклоняющиеся от математического ожидания, и вероятности их не малы.

Так, при одинаковой средней величине осадков, выпадающих в двух местностях за год, нельзя сказать, что климат этих местностей одинаков.

Иными словами, по математическому ожиданию нельзя судить о том, какие отклонения от него возможны. А тем не менее умение дать оценку рассеяния имеет важное значение.

Самой используемой характеристикой разброса значений случайной величины является дисперсия, обозначаемая символами $D[X]$, D_X или σ^2.

Определение. *Дисперсией случайной величины называется математическое ожидание квадрата отклонения её от математического ожидания:*

$$D[X] = M[X - M[X]]^2. \tag{1.29}$$

Для дискретной случайной величины

$$D[X] = (x_1 - m_X)^2 p_1 + \ldots + (x_n - m_X)^2 p_n = \sum_{i=1}^{n} (x_i - m_X)^2 p_i. \tag{1.30}$$

Для непрерывной случайной величины

$$D[X] = \int_{-\infty}^{+\infty} (x - m_X)^2 f(x)dx, \tag{1.31}$$

где m_X – значение математического ожидания.

Пример. Вычислить дисперсии для случайных величин X и Y, законы распределения которых приведены в начале данного параграфа.

Решение

$$D[X] = (-0.1 - 0)^2 \cdot 0.1 + (-0.01 - 0)^2 \cdot 0.2 + (0 - 0)^2 \cdot 0.4 +$$
$$+ (0.01 - 0)^2 \cdot 0.2 + (0.1 - 0)^2 \cdot 0.1 = 0.00204;$$

$$D[Y] = (-20 - 0)^2 \cdot 0.3 + (-10 - 0)^2 \cdot 0.1 + (0 - 0)^2 \cdot 0.2 +$$
$$+ (10 - 0)^2 \cdot 0.1 + (20 - 0)^2 \cdot 0.3 = 260.$$

Таким образом, при одинаковых математических ожиданиях дисперсия величины X очень мала, а случайной величины Y значительная.

В общем случае, если дисперсия случайной величины мала, то малы отклонения от матожидания, а если существуют значения x_i, сильно отклоняющиеся от матожидания, то они маловероятны.

Если же дисперсия велика, то это указывает на существование значений случайной величины, которые сильно отклоняются от её математического ожидания, причем не все они маловероятны.

Кроме дисперсии, характеристикой рассеяния является *среднее квадратическое отклонение* σ, которое является корнем квадратичным из дисперсии: $\sigma = \sqrt{D[X]}$. Среднее квадратическое отклонение имеет размерность значений случайной величины, в то время как дисперсия имеет размерность квадрата размерности значений случайной величины.

Математическое ожидание, дисперсия и среднеквадратическое отклонение – это теоретические величины, и они не являются случайными. Это постоянные величины.

1.2.7. Свойства дисперсии

1. *Дисперсия постоянной величины равна нулю.*

Доказательство

Так как постоянную величины a можно считать случайной величиной, которая принимает только одно значение a с вероятностью 1 $M[a] = a$, поэтому

$$D[a] = (a - a)^2 \cdot 1 = 0.$$

2. *Постоянный множитель можно выносить за знак дисперсии, возводя его при этом в квадрат, т. е.*

$$D[KX] = K^2 D[X],$$

где K – постоянная величина.

Доказательство

Пусть X – данная случайная величина, тогда KX – новая случайная величина. Используя свойство 2 математического ожидания, имеем

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

3. *Дисперсия случайной величины равна математическому ожиданию квадрата случайной величины без квадрата математического ожидания:*

$$D[X] = M\left[X^2\right] - M^2(X). \qquad (1.32)$$

Доказательство

Обозначим $M(X)$ через a, имеем

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

Используя свойство 2 математического ожидания, $M(2aX)=2aM(X)$, а $M\left(a^2\right)=a^2$, так как a – величина постоянная.

Поэтому

$$D[X]=M\left[X^2\right]-2aM[X]+a^2=M\left[X^2\right]-2a^2+a^2=M\left[X^2\right]-M^2[X].$$

Пример. Найти дисперсию случайной величины X, имеющей следующий закон распределения:

X	1	2	3	4	5
p	0.1	0.2	0.3	0.3	0.1

Решение

$M[X]=1\cdot 0.1+2\cdot 0.2+3\cdot 0.3+4\cdot 0.3+5\cdot 0.1=3.1$;

$M\left[X^2\right]=1^2\cdot 0.1+2^2\cdot 0.2+3^2\cdot 0.3+4^2\cdot 0.3+5^2\cdot 0.1=10.9$;

$D[X]=10.9-3.1^2=1.29$.

Формула (1.32) упрощает вычисление дисперсии.

Следующие свойства приводятся без доказательства.

4. *Дисперсия суммы конечного числа независимых случайных величин равна сумме их дисперсий*:

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

5. *Дисперсия разности независимых случайных величин равна сумме их дисперсий:*

$$D[X-Y]=D[X]+D[Y].$$

1.2.8. Моменты

Математическое ожидание, определение которого было дано в предыдущем параграфе, называется еще *моментом*, а точнее, *начальным моментом первого порядка* случайной величины. Этот момент обозначается μ и называется еще *средним,* или *генеральным средним*.

Определение. *Начальным моментом r-го порядка дискретной случайной величины называется*

$$\mu_r^1 = \sum_x x^r p(x) \tag{1.33}$$

для $r = 0,1,2\ldots$

Определение. *Начальным моментом r-го порядка для непрерывной случайной величины называется*

$$\mu_r^1 = \int_{-\infty}^{+\infty} x^r f(x)dx. \tag{1.34}$$

Термин «момент» пришел из физики. Если $p(x)$ является точкой масс, действующей перпендикулярно оси X на расстоянии x от начала координат, то μ_1^1 является «центром тяжести масс», т. е. первый момент, деленный на $\sum p(x) = 1$. М*атематическое ожидание* μ_1^1 обозначается μ .

Определение. *Центральным моментом порядка r для дискретной случайной величины называется величина*

$$\mu_r = \sum_x (x-\mu)^r p(x), \tag{1.35}$$

где μ – математическое ожидание.

Определение. *Центральным моментом порядка r для непрерывной случайной величины называется величина*

$$\mu_r = \int_{-\infty}^{+\infty} (x-\mu)^r f(x)dx. \tag{1.36}$$

Дисперсия является центральным моментом второго порядка для случайной величины, обозначаемым $D[X]$, σ^2 или $\mathrm{var}(X)$.

Третий центральный момент μ_3 описывает *симметрию* распределения, через него вводится показатель асимметрии в виде

$$A_X = \frac{\mu_3}{\sigma^3}. \tag{1.37}$$

1.2.9. Наиболее часто встречающиеся дискретные распределения

Одним из наиболее широко известных распределений является *биномиальный закон распределения*.

Считается, что случайная величина имеет биномиальный закон распределения, если выполняются следующие предположения или схема Бернулли:

– имеется n независимых испытаний;

– каждое испытание имеет только два исхода, обозначенных как «успех» и «неудача»;

– вероятность «успеха» p и «неудачи» $q = 1 - p$ являются постоянными от испытания к испытанию.

Что считать «успехом» и «неудачей» – зависит от поставленной задачи.

Пример. При анализе качества выпускаемой продукции обнаружение контролером среди n деталей бракованного считается «успехом» и годного изделия – «неудачей».

При биномиальном распределении испытания должны быть независимыми. Это требование выполняется для экспериментов, где независимость вытекает из самого эксперимента (подбрасывание монеты, подбрасывание игральной кости) или в экспериментах с возвращением.

Пример. Предположим, 5 % деталей в корзине – с браком. Вероятность вытаскивания бракованного изделия при первом вытаскивании $p = 0.05$. Если первая деталь возвращается в корзину, то вероятность вытащить бракованную деталь сохраняется и равна $p = 0.05$. В этом случае испытания являются независимыми. Если бракованная деталь не возвращается в корзину, то вероятность вытаскивания бракованной детали изменяется. Испытания не являются независимыми.

Вероятность того, что при n испытаниях случайная величина X примет значение m, равное числу «успехов», определяется по биномиальному закону распределения в виде

$$P(X = m) = P_n(m) = C_n^m p^m q^{n-m} . \qquad (1.38)$$

Формула (1.38) носит название *формулы Бернулли*.

Биномиальное распределение имеет два параметра: n и p.

Среднее $\mu = np$.

Дисперсия $\sigma^2 = npq$.

Пример. Производится 4 независимых выстрела, вероятность попадания $p = 0.25$. Определить закон распределения случайной величины X, равной числу попаданий.

Решение. Случайная величина X распределена по биномиальному закону, которая в данном случае принимает значения $0,1,2,3,4$. Вероятность того, что случайная величина X примет значение i, где $i = 0,1,2,3,4$, определяется по формуле (1.38) в виде

$$P(X=i)=C_4^i p^i q^{4-i}.$$

Закон распределения представляется следующей таблицей (рядом распределения):

X	0	1	2	3	4
p	0.316	0.421	0.211	0.047	0.004

$$\sum_i p_i = 0.999 \approx 1$$

Гипергеометрическое распределение применяется к испытаниям, которые проводятся без возвращения.

Гипергеометрическое распределение, подобно биномиальному, состоит из двух исходов: «успеха» и «неудачи», но в данном случае необходимо знать пропорцию «успехов» и «неудач».

Считается, что случайная величина X имеет гипергеометрическое распределение, если выполняются следующие условия:

– испытания проводятся без возвращения;
– каждый исход эксперимента является «успехом» или «неудачей»;
– число испытаний N конечно;
– число успехов M в испытаниях известно.

В этом случае

$$P(X=k)=C_M^k C_{N-M}^{n-k} \Big/ C_N^n, \tag{1.39}$$

где N, M, n – натуральные числа;

$M \le N$, $n \le N$, $k=0,1,\dots,\min(n,M)$.

Пример. Из партии, содержащей M белых и $N-M$ черных шаров, наудачу извлекается n шаров. Какова вероятность того, что среди выбранных шаров окажется k белых?

Решение. В этом примере случайная величина X, равная числу белых шаров, имеет гипергеометрическое распределение, так как испытания проводятся без возвращения, число испытаний N конечно, результат испытания «успех» – выбор белого шара, «неудача» – выбор черного шара. Число «успехов» или число белых шаров известно и равно M:

$$P(X=k)=C_M^k \cdot C_{N-M}^{n-k} \Big/ C_N^n.$$

Пример. 24 человека, среди которых 8 женщин, подали заявление на работу. Если 5 претендентов выбраны случайно, то какова вероятность, что среди них будет 3 женщины?

Решение. Число испытаний $N = 24$. Выборка из 5-ти человек сделана без возвращения. Известно число «успехов» ($M = 8$ женщин) и «неудач» ($N - M = 24 - 8 = 16$ мужчин).

Вероятность, что $X = 3$, определяется в виде

$$P(X = 3) = C_8^3 \cdot C_{16}^2 / C_{24}^5 = 0.1581.$$

1.2.10. Геометрическое распределение

Для бесконечной последовательности испытаний в схеме Бернулли случайная величина X, равная числу испытаний до первого успеха включительно, имеет геометрическое распределение

$$P(X = k) = (1 - p)^{k-1} p, \qquad (1.40)$$

где p – вероятность успеха.

Пример. Вероятность того, что некоторая деталь окажется дефектной, равна p. В случае обнаружения дефекта линию останавливают и делают переналадку. Составить ряд распределения для числа X годных деталей между переналадками.

Решение. Случайная величина X принимает значения $0, 1, 2, \dots, k, \dots$; $X = k$, когда $(k+1)$-я деталь окажется дефектной. Следовательно, $P(X = k) = q^k p$ $(k = 0, 1, 2, \dots)$ и ряд распределения представляется в виде

x_i	0	1	2	…	k	…
p_i	p	qp	q^2p	…	q^kp	…

1.2.11. Распределение Пуассона

Распределение Пуассона и биномиальное распределение имеют некоторое сходство, но имеют и некоторые различия. *Биномиальное распределение* описывает распределение двух возможных исходов: «успеха» и «неудачи» в конечном числе n независимых испытаний. *Распределение Пуассона* сконцентрировано только на числе дискретных исходов на некотором интервале или континууме. Для него неважно число экспериментов n, как для биномиального распределения.

Распределение Пуассона описывает появление редких событий, и его еще называют законом «неправдоподобных» событий.

Распределение Пуассона имеет следующие характеристики:

– дискретное распределение;

– описывает редкие события;

– каждый исход является независимым от другого;

– описывает дискретные исходы на интервале или на континууме;

– исходы на каждом интервале могут быть проранжированы от нуля до бесконечности (случайная величина X может принимать значения $0, 1, 2, ..., m, ...$).

Распределение Пуассона определяется как

$$P(X = k) = \frac{a^k e^{-a}}{k!}. \tag{1.41}$$

Параметр a является средним для данного интервала, значение которого должно сохраняться для всего данного эксперимента.

Значение параметра a для закона Пуассона совпадает с дисперсией, и это используется для подтверждения того, что случайная величина распределена по закону Пуассона.

Распределение Пуассона используется для аппроксимации биномиального закона распределения при $n \geq 25$ и $p \leq 0.1$. В этом случае $a = np$, где n – число независимых экспериментов, p – вероятность «успеха» в одном эксперименте.

Учитывая это, вероятность m «успехов» в n испытаниях определяется в виде

$$P_n(m) = C_n^m p^m q^{n-m} \approx \frac{(np)^m e^{-np}}{m!}. \tag{1.42}$$

Примечание. Существует правило большого пальца, которое состоит в следующем. Если $n > 20$ и $np \leq 7$, то можно использовать распределение Пуассона вместо биномиального.

Пример. 2 % книг имеют дефект в переплете. Определить вероятность того, что 5 из 400 книг имеют дефект в переплете.

Решение. Так как $n > 25$ и $p < 0.1$, то воспользуемся аппроксимацией биномиального закона законом Пуассона. Тогда $a = np = 400 \cdot 0.02 = 8$, $e^{-8} = 0.00034$ (см. табл. V Приложения):

$$P(X = 5) = \frac{8^5 \cdot e^{-8}}{5!} = \frac{32768 \cdot 0.00034}{120} = 0.093.$$

Пример. На ткацком станке нить обрывается в среднем 0.375 раза в течение часа работы станка. Найти вероятность того, что за смену (8 часов) число обрывов нити будет заключено в границах 2 и 4 (не менее двух и не более четырех обрывов).

Решение. Так как задано значение $\lambda = 0.375$ – среднее значение обрыва нити за час, то параметр *a*, характеризующий среднее число обрыва нити за смену (на всем периоде), определяется в виде $a = 0.375 \cdot 8 = 3$. Тогда

$$P(2 \le X \le 4) = P(X = 2) + P(X = 3) + P(X = 4) = 0.616,$$

где

$$P(X = 2) = \frac{3^2 \cdot e^{-3}}{2!} = 0.224; \quad P(X = 3) = \frac{3^3 \cdot e^{-3}}{3!} = 0.224;$$

$$P(X = 4) = \frac{3^4 \cdot e^{-3}}{4!} = 0.168; \quad e^{-3} = 0.0498.$$

Примечание. Если интервалы для параметра a и наблюдений *X* разные, то необходимо, чтобы среднее было определено на том же интервале, что и наблюдения, но не наоборот.

1.2.12. Задание для самостоятельной работы

1. Вероятность поражения вирусным заболеванием куста земляники равна 0.2:

а) составить закон распределения числа кустов земляники, зараженных вирусом, из четырех посаженных кустов;

б) определить математическое ожидание;

в) определить дисперсию и среднее квадратическое отклонение.

Ответ: $m = 0.8$, $\sigma^2 = 0.639$, $\sigma = 0.8$.

2. Стрелок ведет стрельбу по цели с вероятностью попадания при каждом выстреле 0.2. За каждое попадание он получает 5 очков, а в случае промаха очков ему не начисляют.

а) составить закон распределения числа очков, полученных стрелком за три выстрела;

б) определить матожидание, дисперсию и среднее квадратическое отклонение.

Ответ: $m = 3$, $\sigma^2 = 12$.

3. В билете три задачи. Вероятность правильного решения первой

задачи равна 0.9, второй задачи – 0.8, третьей задачи – 0.7. Составить закон распределения числа правильно решенных задач в билете. Определить матожидание и дисперсию.

Ответ: m=2.4, σ^2 =6.22.

4. Из пяти гвоздик – две белого цвета. Составить закон распределения случайной величины, выражающей число белых гвоздик среди двух одновременно взятых. Определить функцию распределения случайной величины числа белых гвоздик.

Ответ: $p1$=0.3, $p2$=0.6, $p3$ = 0.1.

5. Дана функция распределения случайной величины X.

$$F(x)=\begin{cases}0 & \text{при } x\le 0;\\ 0.3 & \text{при } 1<x\le 2;\\ 0.7 & \text{при } 2<x\le 3;\\ 1 & \text{при } x>3.\end{cases}$$

Определить:
а) ряд распределения;
б) $M(X)$, $D(X)$;
в) построить многоугольник распределения;
г) график функции распределения.

Ответ: $M(X)$=2, $D(X)$= 0.6.

6. Абонент забыл последнюю цифру нужного ему номера телефона, однако помнит, что она нечетная. Составить закон распределения числа сделанных им наборов номера телефона до попадания на нужный номер, если последнюю цифру он набирает наудачу, а набранную цифру в дальнейшем не набирает. Определить матожидание и функцию распределения этой случайной величины и построить ее график.

Ответ: $M(X)$=3.

7. При сборке прибора для наиболее точной подгонки основной детали может потребоваться 1,2,3,4 или 5 проб с вероятностями 0.07, 0.21, 0.55, 0.16, 0.01. Сколько деталей надо отпустить сборщику для сборки тридцати приборов?

Ответ: 85.

8. Спортсмен производит ряд попыток забросить мяч в кольцо. При каждой попытке (независимо от других) попадание в кольцо происходит с вероятностью 0.7. Как только мяч попал в кольцо, попытки прекращаются. Случайная величина X – число попыток, которые необходимо произвести. Составить ряд распределения этой случайной величины.

9. Пусть вероятность изготовления нестандартного изделия равна 0.06. Контролер берет из партии изделие и сразу его проверяет. Если изделие оказывается нестандартным, то проверка прекращается и партия бракуется. Если же изделие оказывается стандартным, то контролер берет следующее изделие, но проверяет не более пяти изделий.

Составить закон распределения и функцию распределения этой случайной величины. Построить график функции распределения.

10. Автомобиль должен проехать по улице, на которой установлены три светофора, дающие независимо друг от друга зеленый сигнал в течение 1.5 минут, желтый – в течение 0.3 минут, красный – в течение 1.2 минуты. Составить закон распределения случайной величины, определяющей число остановок автомобиля. Определить $M(X)$, $D(X)$.

Ответ: $M(X)$=1.2, $D(X)$=0.72.

1.2.13. Наиболее часто встречающиеся непрерывные распределения

Нормальный закон распределения

Нормальное распределение является краеугольным камнем современной статистической теории. Оно было исследовано впервые в XVIII в., когда ученые наблюдали удивительную регулярность в ошибках измерений. Они обнаружили, что распределение, которое они наблюдали, могло бы быть аппроксимировано непрерывной кривой, которую назвали «нормальной кривой ошибок». Математические свойства нормальных кривых впервые были изучены известными математиками, такими, как Муавр, Лаплас и Гаусс. Иногда нормальную кривую называют *Гауссовой кривой*, а нормальный закон – *Гауссовым*.

Нормальный закон распределения проявляется в тех случаях, когда случайная величина X является результатом действия большого числа различных факторов. Каждый фактор в отдельности на величину X влияет незначительно, и нельзя указать, какой именно в большей степени, чем остальные.

Примерами случайных величин, имеющих нормальное распределение, могут служить отклонения действительных размеров деталей, обработанных на станке, от номинальных размеров; отклонения при стрельбе и др.

Непрерывная случайная величина X имеет нормальное распределение, если её плотность распределения имеет вид

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty<x<\infty \qquad , \tag{1.43}$$

где μ (матожидание) и σ (среднеквадратическое отклонение) являются параметрами нормального распределения. Факт, что случайная величина *X* имеет нормальное распределение, обозначается $N(\mu,\sigma)$.

Функция распределения случайной величины *X*, имеющей нормальное распределение, имеет вид

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx\,. \tag{1.44}$$

График плотности нормального распределения (кривая распределения или кривая Гаусса) имеет вид, представленный на рис. 1.9.

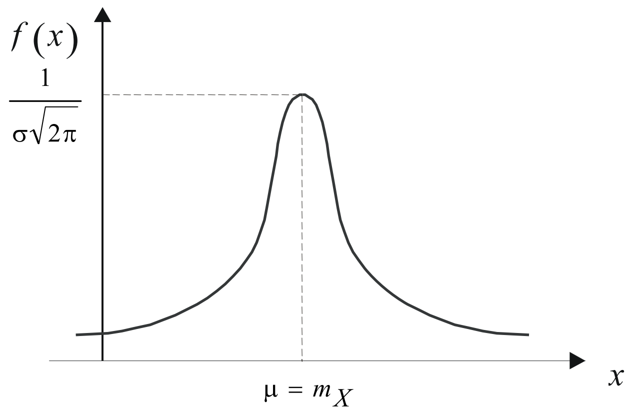

Рис. 1.9. Кривая нормального распределения

Кривая нормального распределения имеет следующие свойства:

– кривая распределения симметрична относительно ординаты, проходящей через точку $\mu=m_x$;

– кривая имеет один максимум при $x=m_x$, равный $\frac{1}{\sigma\sqrt{2\pi}}$;

– при $|x|\to\infty$ ветви кривой асимптотически приближаются к оси 0_X;

– в точках $x=m_x\pm\sigma$ кривая имеет перегиб.

Каждая пара параметров m_X,σ дает свое нормальное распределение.

На рис. 1.10 представлены графики кривых нормальных распределений с параметрами:

$m_x=50, \quad \sigma=5;$

$m_x=80, \quad \sigma=5;$

$m_x = 50, \qquad \sigma = 1.$

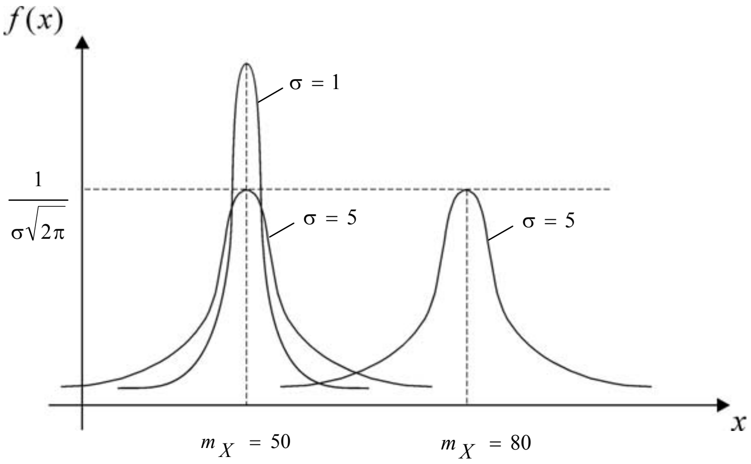

Рис. 1.10. Кривые нормального распределения

Чтобы избежать трудностей при работе с нормальным распределением, нормальное распределение с параметрами m_x, σ преобразуется в единственное распределение с параметрами $m_x = 0$ и $\sigma = 1$, т. е. в $N(0,1)$, которое называется *стандартизованным нормальным распределением*, или Z *-распределением*. Формула преобразования для случайной величины X, имеющей нормальное распределение $N(m_x, \sigma)$ имеет вид

$$Z = \frac{X - m_x}{\sigma}. \tag{1.45}$$

Величина Z представляет отклонение случайной величины X от математического ожидания $\mu = m_x$ в числе стандартных отклонений.

Пример. Определить Z -величину случайной величины X, распределенной по нормальному закону с параметрами $\mu = 50$ и $\sigma = 10$, если $X = 70$.

Решение

$$Z = \frac{70 - 50}{10} = +\frac{20}{10} = +2.$$

Этот результат говорит о том, что случайная величина X отклоняется от математического ожидания на 2 стандартных отклонения вправо.

Кривая нормального распределения с параметрами $\mu = 0$ и $\sigma = 1$ имеет вид, представленный на рис. 1.11.

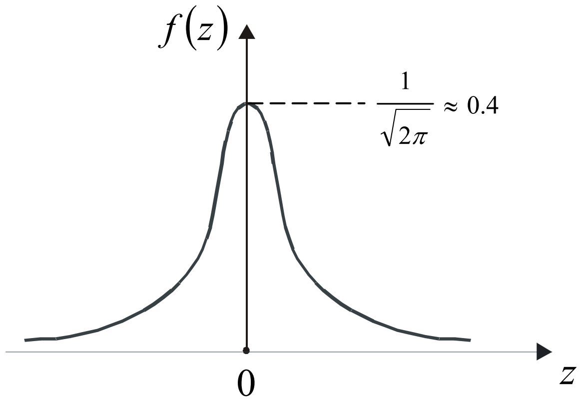

Рис. 1.11. Кривая распределения $N(0,1)$

Кривая распределения $N(0,1)$ имеет следующие свойства:

– она симметрична относительно оси ординат;

– в точке $x=0$ имеет максимум, равный $\frac{1}{\sqrt{2\pi}} \approx 0.4$;

– имеет точки перегиба $x = \pm 1$;

– при $x \to \pm\infty$ кривая распределения приближается к оси абсцисс.

Так как интеграл от плотности распределения $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ не выражается через элементарные функции, то для расчета вероятностей составлены таблицы специальной функции

$$\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz, \tag{1.46}$$

которая называется *интегралом вероятностей,* или *функцией Лапласа* (см. табл. I Приложения).

Как уже указывалось, функция $\Phi(x)$ является нечетной (см. с. 33):

$$\Phi(-x) = \frac{2}{\sqrt{2\pi}} \int_0^{-x} e^{-\frac{z^2}{2}} dz = -\Phi(x), \tag{1.47}$$

поэтому в таблице приведены значения $\Phi(x)$ только для положительных *x*. График функции Лапласа приведен на рис. 1.12.

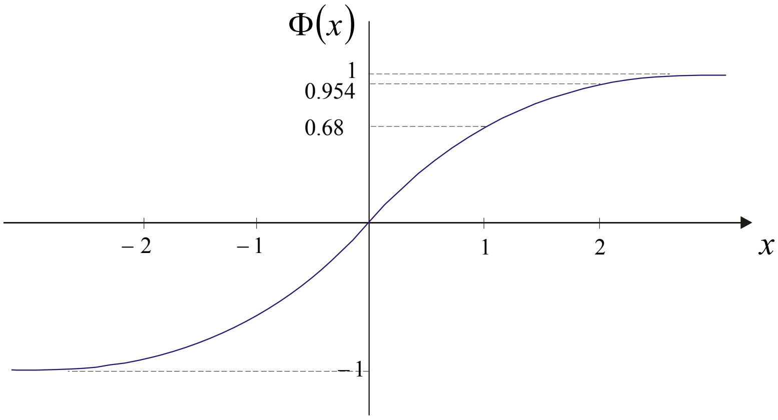

Рис. 1.12. График функции Лапласа

С помощью функции Лапласа можно вычислить вероятность попадания случайной величины в интервал от x_1 до x_2, т. е. $P(x_1 < X < x_2)$:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx = \\ = \frac{1}{\sqrt{2\pi}}\int_0^{x_2} e^{-z^2/2}dz - \frac{1}{\sqrt{2\pi}}\int_0^{x_1} e^{-z^2/2}dz = \Phi_1(x_2) - \Phi_1(x_1). \tag{1.48}$$

Для симметричного интервала

$$P(-t < X < +t) = \frac{1}{2}\Phi(t) - \frac{1}{2}\Phi(-t) = \Phi(t). \tag{1.49}$$

Интеграл вероятностей (или функция Лапласа), для которого составлены таблицы, может определяться еще в виде

$$\Phi_1(x) = \frac{1}{\sqrt{2\pi}}\int_0^{x} e^{-z^2/2}dz, \tag{1.50}$$

где $\Phi_1(x) = \frac{1}{2}\Phi(x)$.

Как известно, вероятность попадания случайной величины на интервал от x_1 до x_2 может определяться и с использованием функции распределения в виде

$$P(x_1 < X < x_2) = F(x_2) - F(x_1). \tag{1.51}$$

Функция распределения стандартизованного нормального распределения $N(0,1)$ имеет вид

$$\Phi_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2} dz. \tag{1.52}$$

Функция распределения случайной величины X, распределенной по нормальному закону с параметрами m_X и σ, выражается через функцию распределения стандартизованного нормального распределения в виде

$$F(x) = \Phi_{N(0,1)}\left(\frac{x - m_X}{\sigma}\right). \tag{1.53}$$

С учетом этого вероятность попадания случайной величины X на участок от x_1 до x_2

$$P(x_1 < X < x_2) = \Phi_{N(0,1)}\left(\frac{x_2 - m_X}{\sigma}\right) - \Phi_{N(0,1)}\left(\frac{x_1 - m_X}{\sigma}\right). \tag{1.54}$$

Функция распределения стандартизованного нормального распределения имеет следующие свойства:

- $\Phi_{N(0,1)}(-\infty) = 0$, $\Phi_{N(0,1)}(+\infty) = 1$;
- $\Phi_{N(0,1)}$ – неубывающая функция;
- из симметричности нормального распределения с параметрами $m_x = 0$, $\sigma = 1$ относительно начала координат следует, что

$$\Phi_{N(0,1)}(-x) = 1 - \Phi_{N(0,1)}(x).$$

Функция стандартизованного нормального распределения $\Phi_{N(0,1)}$ связана с функцией Лапласа соотношениями вида:

$$\Phi_{N(0,1)}(x) = \frac{1}{2}(\Phi(x) + 1); \tag{1.55}$$

$$\Phi_{N(0,1)}(x) = \frac{1}{2}(2\Phi_1(x) + 1). \tag{1.56}$$

Значения функции $\Phi_{N(0,1)}$ протабулированы (см. табл. II Приложения).

Пример. Пусть вес пойманной рыбы подчиняется нормальному закону распределения с $m_x = 375$г и $\sigma = 25$ г. Найти вероятность того, что вес пойманной рыбы будет от 300 до 425 г.

Решение

По формуле (1.54) при $x_1 = 300$, $x_2 = 425$ $m_x = 375$, $\sigma = 25$; с использованием табл. II Приложения

$$P(300 < X < 425) = \Phi_{N(0,1)}\left(\frac{425-375}{25}\right) - \Phi_{N(0,1)}\left(\frac{300-375}{25}\right) =$$
$$= \Phi_{N(0,1)}(2) - \Phi_{N(0,1)}(-3) = \Phi_{N(0,1)}(2) - 1 + \Phi_{N(0,1)}(3) =$$
$$= 0.97725 - 1 + 0.99865 = 0.975.$$

По формуле (1.49), с учетом соотношений (1.50) с использованием табл. I Приложения

$$P(300 < X < 425) = {}^1\!/_2\left(2\Phi_1\left(\frac{425-375}{25}\right) - 2\Phi_1\left(\frac{300-375}{25}\right)\right) =$$
$$= {}^1\!/_2(2\Phi_1(3) + 2\Phi_1(2)) = {}^1\!/_2(2\cdot 0.498 + 2\cdot 0.47725) =$$
$$= {}^1\!/_2(0.996 + 0.9545) = 0.975.$$

На практике часто возникает необходимость определения значения z_p из уравнения $\Phi(z_p) = p$.

В данном случае z_p называется *квантилем уровня* p, где p – заданная вероятность.

Если рассматривается функция стандартизованного нормального распределения $\Phi_{N(0,1)}$, то следует обратить внимание, что значения $\Phi_{N(0,1)}$ в таблице (см. табл. II Приложения) задаются, начиная со значения, равного 0.5.

Пример. Пусть необходимо определить z_p из условия $\Phi_{N(0,1)} = 0.05$.

Решение. По определению функции распределения нужно найти z_p из условия $P(Z < z_p) = 0.05$. Так как значения 0.05 нет в таблице, то, используя свойство функции стандартизованного нормального распределения,

$$\Phi_{N(0,1)}(-z) = 1 - \Phi_{N(0,1)}(z),$$

переходим к $P(Z > z_p) = 1 - P(Z < z_p) = 1 - P(Z < z_p) = 1 - 0.05 = 0.95$, т. е.

$$1 - P(Z < z_p) = \Phi_{N(0,1)}(-z_p) = 0.95;$$

из табл. II Приложения $z_p = -1.64$.

Квантили нормального распределения связаны следующим соотношением:

$$z_p = -z_{1-p}. \tag{1.57}$$

Вероятность того, что абсолютная величина отклонения случайной величины X от своего математического ожидания m_x меньше $\varepsilon > 0$, определяется формулой

$$P(|X-m_x|<\varepsilon)=2\Phi_{N(0,1)}\left(\varepsilon/_{\sigma}\right)-1 \qquad (1.58)$$

или с использованием функции Лапласа $\Phi(x)$ – в виде

$$P(|X-m_x|<\varepsilon)=\Phi\left(\frac{\varepsilon}{\sigma}\right). \qquad (1.59)$$

Эмпирическое правило для случайной величины X, распределенной по нормальному закону распределения, утверждает процент значений случайной величины, находящихся от m_X на расстоянии $\pm 1\sigma$, $\pm 2\sigma$, $\pm 3\sigma$ (табл. 1.4).

Таблица 1.4

Эмпирическое правило

Расстояние от среднего	Процент значений X внутри данного интервала
$m_x \pm 1\sigma$	68
$m_x \pm 2\sigma$	95
$m_x \pm 3\sigma$	99.7

Это правило часто называют *правилом* «3-*х сигм*». Это правило позволяет ориентировочно указать интервал практически возможных значений для случайных величин, подчиненных нормальному закону распределения при известных m_x и σ.

1.2.14. Задание для самостоятельной работы

1. Для нормально распределенной случайной величины определить:
   а) $P\ (12<X<14)$;
   б) $P\ (8<X<12)$, если $m_x = 10$, $\sigma = 4$.

Ответ: 0.14988, 0.38292.

2. При взвешивании получается ошибка, подчиненная нормальному закону с $\sigma = 20$. Найти вероятность того, что взвешивание будет произведено с ошибкой, не превосходящей 10.

Ответ: 0.38292.

3. Диаметр детали, изготовленной в цехе, является случайной величиной, распределенной по нормальному закону. Дисперсия равна 0.0001 и математическое ожидание равно 2.5. Найти границы, в которых с вероятностью 0.9973 заключен диаметр наудачу взятой детали.

Ответ: 2.47, 2.53.

4. Полагая, что рост мужчин определенной возрастной группы есть нормально распределенная случайная величина X с параметрами

m_x = 73, σ^2 = 36, определить: а) выражение плотности вероятности и функции распределения случайной величины X; в) квантиль $x_{0.7}$ и 10 %-ю точку случайной величины X; с) сформулировать правило трех сигм для случайной величины X.

Ответ: в) 177, 180; с) 155, 191.

1.2.15. Нормальное распределение вместо биномиального

Для определенных типов биномиальных проблем нормальное распределение может быть хорошим приближением.

Если объем выборки возрастает, то биномиальное распределение приближается по форме к нормальному, несмотря на значение вероятности p. Если объем выборки небольшой, то биномиальное распределение приближается к нормальному, если $p \approx 0.5$. При больших значениях n определение вероятности с использованием биномиальной формулы становится очень громоздким.

Для того чтобы использовать нормальный закон распределения вместо биномиального, необходимо выполнить следующее:

– преобразовать параметры биномиального распределения n и p в параметры нормального распределения μ и σ по формулам

$$\mu = np \quad \text{и} \quad \sigma = \sqrt{npq}\ ; \tag{1.60}$$

– необходимо определить, является ли нормальное распределение хорошей аппроксимацией для биномиального распределения. Для этого нужно установить, лежит ли интервал $\mu \pm 3\sigma$ между нулем и n, т. е. все возможные значения X должны быть между 0 и n (в соответствии с эмпирическим правилом).

Другое правило для определения, в каком случае нормальная аппроксимация будет достаточно хорошей – это выполнение соотношений

$$np > 5 \text{ и } nq > 5.$$

Пример. Какова вероятность, что в $n = 60$ испытаниях успех наступит 25 раз, если вероятность успеха в одном испытании $p = 0.3$.

Решение. $\mu = np = 60 \cdot 0.3 = 18$, $\sigma = \sqrt{npq} = \sqrt{60 \cdot 0.3 \cdot 0.7} = 3.55$. Интервал $\mu \pm 3\sigma = 18 \pm 3(3.55) = 18 \pm 10.65$, т. е. $0 < 7.35 \le \mu \le 28.65 < 60$, $0 < X < 60$. Кроме того, $np > 5$ и $nq > 5$. Следовательно, аппроксимация будет хорошей.

1.2.16. Поправка на непрерывность

Перевод дискретного распределения в непрерывное не происходит напрямую, для этого необходимо сделать поправку на непрерывность. Эта поправка должна быть сделана для того, чтобы расхождение было как можно меньше и аппроксимация была бы более точной.

Например. Имеем биномиальное распределение с $n=60$ и $p=0.3$. Определить вероятность $P(X \geq 25)$.

Решение. Для определения $P(X \geq 25)$ необходимо определить вероятности для $X = 25, 26, 27, \ldots, 60$ и просуммировать.

Нормальное распределение – непрерывное распределение, и значения X непрерывно располагаются по оси X. Поправка на непрерывность может быть сделана с учетом табл. 1.5.

Таблица 1.5

Поправки на непрерывность

Величины, вероятность которых необходимо определить	Поправки
$X >$	$+ 0.5$
$X \geq$	$- 0.5$
$X <$	$- 0.5$
$X \leq$	$+ 0.5$
$\leq X \leq$	$- 0.5$ и $+ 0.5$
$< X <$	$+ 0.5$ и $- 0.5$
$X =$	$- 0.5$ и $+ 0.5$

Решение. В соответствии с табл. 1.5 в нашем случае необходимо определить $P(X \geq 24.5)$, где $\mu = np = 18$, $\sigma = \sqrt{npq} = 3.55$ (рис. 1.13).

Для этого определяем $Z = \dfrac{X - \mu}{\sigma} = \dfrac{24.5 - 18}{3.55} = 1.83$ (рис. 1.14);

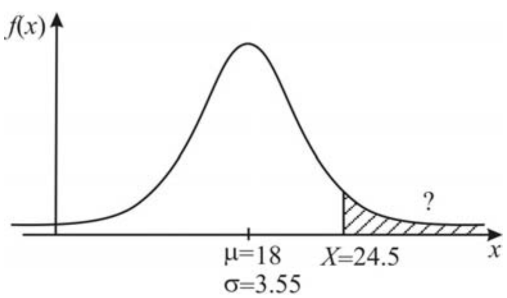

Рис. 1.13

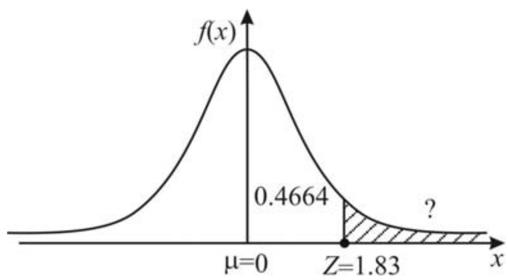

Рис. 1.14

$P(Z \geq 1.83) = 0.5 - 0.4664 = 0.0336$.

1.2.17. Задание для самостоятельной работы

1. Определить вероятность появления 12-ти успехов в 25-ти испытаниях для случайной величины X с использованием нормального распределения, если вероятность успеха в одном испытании $p = 0.4$.

Ответ: $P(X = 15) = 0.117$.

2. Определить $P(X \geq 15)$ для случайной величины X, распределенной по биномиальному закону с $n = 60$ и $p = 0.3$.

Ответ: $P(X \geq 25) = 0.0336$.

3. Согласно статистическим данным 87 % всех студентов колледжа работают. Если эта тенденция сохранится, то какова вероятность, что 100 из 120-ти случайно выбранных студентов работают?

Ответ: 0.0918.

4. Определить $P(13 \leq X \leq 18)$, если вероятность успеха $p = 0.117$ и объем выборки $n = 30$.

Ответ: 0.5863.

1.2.18. Центральная предельная теорема

Центральная предельная теорема представляет собой группу теорем, посвященных установлению условий, при которых возникает нормальный закон распределения. Среди этих теорем важнейшее место принадлежит теореме Ляпунова, которая имеет несколько формулировок. Одна из них приводится ниже.

Теорема Ляпунова. *Распределение суммы независимых случайных величин* X_i $(i = \overline{1, n})$ *приближается к нормальному закону распределения при неограниченном увеличении n, если выполняются следующие условия:*

– все случайные величины X_i *имеют конечные математические ожидания и дисперсии;*

– ни одна из величин X_i *по своему значению резко не отличается от всех остальных, т.е. оказывает ничтожное влияние на их сумму.*

Теорема Ляпунова справедлива не только для непрерывных, но и для дискретных случайных величин. На практике установлено, что распределение суммы независимых случайных величин, у которых дисперсии не отличаются резко друг от друга, довольно быстро приближается к нормальному. Уже при числе слагаемых, большем 10-ти, распределение суммы можно заменить нормальным.

При решении многих практических задач, связанных со случайной величиной $\overline{X}$, являющейся средним арифметическим наблюдаемых значений случайной величины *X*, применяется теорема Ляпунова в следующей формулировке.

Теорема. *Если случайная величина X имеет конечные математическое ожидание* $M[X]$ *и дисперсию* $D[X]$, *то распределение среднего арифметического* $\overline{X}$, *вычисленного по экспериментальным значениям случайной величины X в n независимых испытаниях, проведенных в одинаковых условиях, при* $n \to \infty$ *приближается к нормальному с математическим ожиданием* $M[X]$ *и дисперсией* $D[X]/n$ *или* $\overline{X}$ *имеет* $N\left(M[X], \sqrt{D[X]/n}\right)$.