Числовые характеристики ДСВ

Закон распределения полностью характеризует случайную величину. Однако, когда невозможно найти закон распределения, или этого не требуется, можно ограничиться нахождением значений, называемых числовыми характеристиками случайной величины. Эти величины определяют некоторое среднее значение, вокруг которого группируются значения случайной величины, и степень их разбросанности вокруг этого среднего значения.

Математическим ожиданием (средним значением, средним) дискретной случайной величины X, имеющей возможные значения $x_1, x_2, \ldots, x_n$ с вероятностями $p_1, p_2, \ldots, p_n$ называется равенство:

$$M(X) = MX = \sum_{i=1}^{n} x_i p_i = m_x$$

Замечание: Если дискретная случайная величина может принимать бесконечное счётное множество значений $x_1, \ldots, x_n, \ldots$ с вероятностями $p_1, \ldots, p_n, \ldots$, то её математическое ожидание определяется равенством: $MX = m_x = \sum_{i=1}^{\infty} x_i p_i$ при условии, что последний ряд абсолютно сходится. В противном случае говорят, что мат. ожидание не существует.

MX не показывает наиболее вероятное значение случайной величины X, а только характеризует ту величину, около которой колеблются возможные значения X, то есть выражает "среднее ожидаемое" значение случайной величины.

Вероятностный смысл математического ожидания: математическое ожидание приблизительно равно (тем точнее, чем больше испытаний) среднему арифметическому наблюдаемых значений случайной величины.

Доказательство: Пусть в n испытаниях случайная величина X приняла значение x_1 m_1 раз, x_2 - m_2 раз,…, $x_к$ – $m_к$ раз, $m_1+m_2+\ldots+m_к=n$. Найдём среднее арифметическое

$$\bar{X} = \frac{m_1x_1 + m_2x_2 + \cdots + m_kx_k}{n} = x_1\frac{m_1}{n} + x_2\frac{m_2}{n} + \cdots + x_k\frac{m_k}{n} =$$
$$= x_1w_1 + x_2w_2 + \cdots + x_kw_k \approx x_1p_1 + x_2p_2 + \cdots + x_kp_k = M(X), \text{ где}$$
$$\frac{m_1}{n} = w_1 \approx p_1, \ldots, \frac{m_k}{n} = w_k \approx \mathrm{p_k}.$$

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

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

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

Пример: Согласно американским статистическим таблицам смертности вероятность, что 25-летний человек проживёт ещё год, равна 0.992 (соответственно, что умрёт – 0.008). Страховая компания предлагает такому человеку застраховать свою жизнь на год на сумму 1000 долларов. Страховой взнос -10 долларов. Найти мат.ожидание прибыли компании. Пусть случайная величина X – прибыль компании. Запишем закон распределения

X	+10	-990
p	0.992	0.008

Тогда $M(X)=10\cdot 0.992-990\cdot 0.008=9.92-7.92=2$.

Итак, ожидаемая средняя прибыль положительна, что даёт страховой компании возможность продолжить своё дело: оставлять резервный капитал для выплаты страховых сумм, получать прибыль и т. д.

Свойства математического ожидания.

1) Математическое ожидание постоянной равно этой же постоянной, то есть если $C\equiv const$, то $MC=C$.

Доказательство: Постоянную величину “C” можно рассматривать как случайную величину, принимающую значение “C” с вероятностью 1. Тогда по формуле

$MC=C$, что и требовалось доказать.

2) $M(CX)=C\cdot MX$, где $C\equiv const$.

Доказательство: Для дискретной случайной величины

$$M(CX)=\sum_{i=1}^{\infty}Cx_i p_i = C\sum_{i=1}^{\infty}x_i p_i = CMX$$

3) Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий слагаемых.

$$M(X+Y)=M(X)+M(Y)$$

Это свойство справедливо для произвольного числа случайных величин.

4) Математическое ожидание произведения двух **независимых** случайных величин равно произведению их математических ожиданий.

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

Это свойство справедливо для произвольного числа случайных величин.

5) Математическое ожидание среднего арифметического значения n одинаково распределенных взаимно независимых случайных величин равно математическому ожиданию каждой из этих величин.

Доказательство: пусть $X_1, X_2, \dots, X_n$ одинаково распределённые случайные величины, математические ожидания каждой из которых одинаковы и равны

$M(X_1)=M(X_2)=\cdots=M(X_1)=a$, тогда $\bar{X}=(X_1+\cdots+X_n)/n$ и $M(\bar{X})=\frac{1}{n}\big(M(X_1)+\cdots+M(X_n)\big)=\frac{na}{n}=a.$

Математическое ожидание дискретной случайной величины может не совпадать ни с одним из ее возможных значений X.

Дисперсия.

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

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

Y	0	100
p	0,5	0,5

Найдем $M(X) = 49\cdot0,1 + 50\cdot0,8 + 51\cdot0,1 = 50$, $M(Y) = 0\cdot0,5 + 100\cdot0,5 = 50$. Как видно, математические ожидания обеих величин равны, но если для X $M(X)$ хорошо описывает поведение случайной величины, являясь ее наиболее вероятным возможным значением (причем остальные значения ненамного отличаются от 50), то значения Y существенно отстоят от $M(Y)$. Следовательно, наряду с математическим ожиданием желательно знать, насколько значения случайной величины отклоняются от него. Для характеристики этого показателя служит дисперсия.

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

$$D(X)=DX= M\,(X - M(X))^2$$

(Если взять просто отклонение, то это ничего не даст: $M\big(X - M(X)\big) = 0$ - отклонения разных знаков компенсируют друг друга).

Если известен закон распределения случайной величины X, то для дискретной случайной величины дисперсию можно вычислить по формуле

$$DX = \sum_{i=1}^{\infty}(x_i - MX)^2 P(X = x_i)\,.$$

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

Пример: Найдем дисперсию случайной величины X (числа стандартных деталей среди отобранных) в примере, приведённом выше. Вычислим значения квадрата отклонения каждого возможного значения от математического ожидания:

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

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

Существует более удобная для расчетов формула для вычисления дисперсии

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

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

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

$$= M(X^2) - 2M^2(X) + M^2(X) = M(X^2) - M^2(X),$$ что и требовалось доказать.

Пример: Вычислим дисперсии случайных величин X и Y, рассмотренных в начале этого раздела. $D(X) = (49^2 \cdot 0,1 + 50^2 \cdot 0,8 + 51^2 \cdot 0,1) - 50^2 = 2500,2 - 2500 = 0,2$.

$D(Y) = (0^2 \cdot 0,5 + 100^2 \cdot 0,5) - 50^2 = 5000 - 2500 = 2500$. Итак, дисперсия второй случайной величины в несколько тысяч раз больше дисперсии первой. Таким образом, даже не зная законов распределения этих величин, по известным значениям дисперсии мы можем утверждать, что X мало отклоняется от своего математического ожидания, в то время как для Y это отклонение весьма существенно.

Свойства дисперсии

1) Дисперсия постоянной величины C равна нулю:

$$D(C) = 0.$$

Доказательство: $D(C) = M((C - M(C))^2) = M((C - C)^2) = M(0) = 0$.

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

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

Доказательство: $D(CX) = M((CX - M(CX))^2) = M((CX - CM(X))^2) = M(C^2(X - M(X))^2) = C^2 D(X)$.

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

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

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

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

Дисперсия суммы постоянной и случайной величин равна дисперсии случайной величины.

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

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

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

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

Доказательство: $D(C+X)=M[(C+X)-M(C+X)]^2=M[C+X-C-MX]^2=$
$=M[X-MX]^2=DX$, что и требовалось доказать.

Таким образом, если к случайной величине X прибавить постоянную C, то MX изменится на величину C: $M(C+X)=C+MX$, но дисперсия при этом не изменится.

6) Дисперсия среднего арифметического значения n одинаково распределенных взаимно независимых случайных величин, дисперсия каждой из которых равна D, равна $D(\overline{X}) = \frac{D}{n}$.

Доказательство: $D(\overline{X}) = D\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{1}{n^2}\left(D(X_1)+\cdots+D(X_n)\right) = \frac{nD}{n^2} = \frac{D}{n}$.

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

Средним квадратическим отклонением σ случайной величины X называется квадратный корень из дисперсии:

$$\sigma(X) = \sqrt{D(X)}\ , \qquad \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}.$$

Пример: В предыдущем примере средние квадратические отклонения X и Y равны соответственно $\sigma_x = \sqrt{0{,}2} \approx 0{,}447; \quad \sigma_y = \sqrt{2500} = 50.$

Пример: Испытывается устройство, состоящее из четырех независимо работающих приборов. Вероятности отказа каждого из приборов равны соответственно $p_1=0{,}3;\ p_2=0{,}4;\ p_3=0{,}5;\ p_4=0{,}6$. Найти математическое ожидание и дисперсию числа отказавших приборов.

Решение: Принимая за случайную величину число отказавших приборов, видим что эта случайная величина может принимать значения 0, 1, 2, 3 или 4.

Для составления закона распределения этой случайной величины необходимо определить соответствующие вероятности. Примем $q_i = 1 - p_i$.

1) Не отказал ни один прибор.

$p(0) = q_1q_2q_3q_4 = 0{,}7 \cdot 0{,}4 \cdot 0{,}5 \cdot 0{,}4 = 0{,}084.$

2) Отказал один из приборов.

$p(1) = p_1q_2q_3q_4 + q_1p_2q_3q_4 + q_1q_2p_3q_4 + q_1q_2q_3p_4 = 0{,}302.$

3) Отказали два прибора.

$p(2) = p_1p_2q_3q_4 + p_1q_2p_3q_4 + p_1q_2q_3p_4 + q_1p_2p_3q_4 + q_1p_2q_3p_4 + q_1q_2p_3p_4 = 0{,}38.$

4) Отказали три прибора.

$p(3) = p_1p_2p_3q_4 + p_1p_2q_3p_4 + p_1q_2p_3p_4 + q_1p_2p_3p_4 = 0{,}198.$

5) Отказали все приборы.

$p(4) = p_1p_2p_3p_4 = 0{,}036.$

Получаем закон распределения:

x	0	1	2	3	4
x^2	0	1	4	9	16
p	0,084	0,302	0,38	0,198	0,036

Математическое ожидание:

$M(X) = 0{,}302 + 2 \cdot 0{,}38 + 3 \cdot 0{,}198 + 4 \cdot 0{,}036 = 1{,}8.$

$M(X^2) = 0{,}302 + 4 \cdot 0{,}38 + 9 \cdot 0{,}198 + 16 \cdot 0{,}036 = 4{,}18.$

Дисперсия:

$D(X) = M(X^2) - [M(X)]^2 = 4{,}18 - 3{,}24 = 0{,}94.$

В финансовом анализе *M(X)* и *D(X)* интерпретируются следующим образом. Пусть известно распределение доходности X некоторого актива (акции). Тогда *M(X)* выражает среднюю (прогнозную) доходность актива, а $D(X)$ (или $\sigma(X)$) – меру отклонения от ожидаемого среднего значения, т.е. **риск** данного актива.

Непрерывные случайные величины

Определение и свойства функции распределения сохраняются и для непрерывной случайной величины, для которой функцию распределения можно считать одним из видов задания закона распределения. Но для непрерывной случайной величины вероятность каждого отдельного ее значения равна 0. Это следует из свойства 4 функции распределения: $P(X = a) = F(a) - F(a) = 0$ (для дискретных СВ из нулевой вероятности следовала невозможность события и наоборот, т.е. вероятность концентрировалась около конкретных значений X, теперь она разлита (размазана) по бесконечному числу точек). Поэтому для такой случайной величины имеет смысл говорить только о вероятности ее попадания в некоторый интервал.

Следствие: вероятность попадания СВ в интервал (a,b) не зависит от того закрытый этот интервал или открытый.

Функция распределения непрерывной случайной величины является ее исчерпывающей вероятностной характеристикой. Но она имеет недостаток, заключающийся в том, что по ней трудно судить о характере распределения случайной величины в небольшой окрестности той или иной точки числовой оси. Более наглядное представление о характере распределения непрерывной случайной величины в окрестностях различных точек дается функцией, которая называется плотностью вероятности или дифференциальным законом распределения случайной величины.

Пусть имеется непрерывная случайная величина X с функцией распределения $F(x)$. Вычислим вероятность попадания этой случайной величины на элементарный участок $[x, x+\Delta x)$. Имеем, $P(x \le X < x+\Delta x) = F(x+\Delta x) - F(x)$.

Деля обе части последнего равенства на Δx имеем

$$\frac{P(x < X < x+\Delta x)}{\Delta x} = \frac{F(x+\Delta x) - F(x)}{\Delta x}.$$

Последнее равенство называется средней вероятностью, которая приходится на единицу длины этого участка.

Считая функцию распределения $F(x)$ дифференцируемой, перейдем в последнем равенстве к пределу при $\Delta x \to 0$. Имеем

$$\lim_{\Delta x \to 0} \frac{P(x < X < x+\Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = F'(x).$$

Плотностью распределения вероятностей (просто **плотностью** или **дифференциальной функцией**) непрерывной случайной величины X в точке x называется производная ее функции распределения в этой точке. Обозначим ее через

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

Плотность распределения $f(x)$, как и функция распределения $F(x)$ является одной из форм закона распределения, но в отличии от функции распределения эта форма не универсальна: она существует только для непрерывных случайных величин.

Дадим теперь более строгое определение непрерывной случайной величины:

случайная величина X называется **непрерывной**, если ее функция распределения $F(x)$ непрерывна на всей оси Ox, а плотность распределения $f(x)$ существует везде, за исключением, быть может, конечного числа точек.

График плотности распределения $f(x)$ называется **кривой распределения**.

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

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

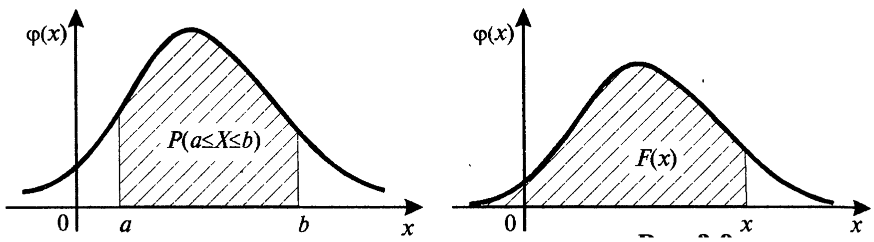
Свойства плотности распределения

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

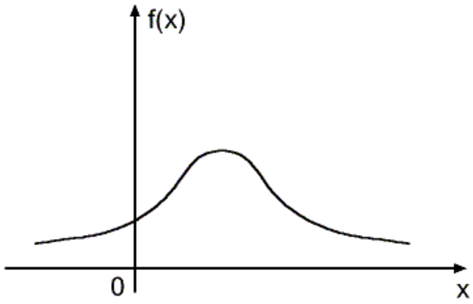
2) $F(x) = \int\limits_{-\infty}^{x} f(t)dt$, что следует из определения плотности распределения.

3) Вероятность попадания случайной величины в интервал (a, b) определяется формулой $p(a < X < b) = \int\limits_{a}^{b} f(x)dx.$

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



Геометрически $F(x)$ и $P(a < X < b)$ равны площади заштрихованных фигур.



4) $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ (условие нормировки). Его справедливость следует из того, что $\int\limits_{-\infty}^{+\infty} f(x)dx = F(+\infty)$, а $\lim\limits_{x \to +\infty} F(x) = 1$.

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

Пример : Функция распределения непрерывной случайной величины имеет вид:

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

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

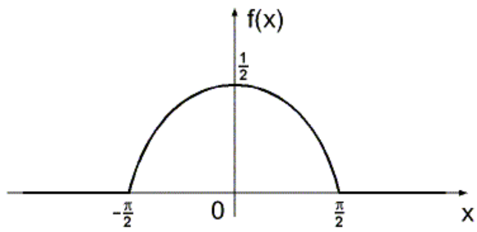
Пример: $$f(x)=\begin{cases} a\cos x, & x\in\left(-\dfrac{\pi}{2};\dfrac{\pi}{2}\right)\\ 0, & x\notin\left(-\dfrac{\pi}{2};\dfrac{\pi}{2}\right)\end{cases}$$

1) Найти коэффициент a;
2) Построить кривую распределения случайной величины X;
3) Найти вероятность попадания случайной величины X на участок от 0 до $\dfrac{\pi}{4}$;
4) Найти и построить функцию распределения *F(x)* случайной величины X.

Решение: 1) Коэффициент a определим используя нормировочное свойство

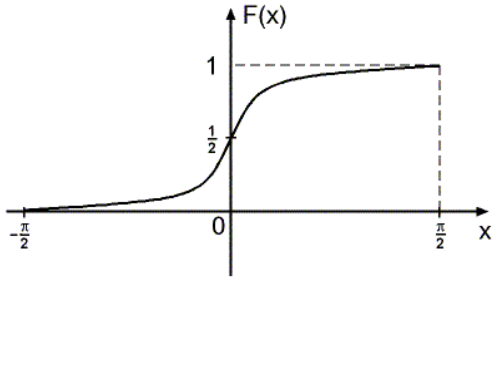
$$\int_{-\infty}^{\infty} f(x)dx=1 \Rightarrow \int_{-\infty}^{\infty} f(x)dx=\underbrace{\int_{-\infty}^{-\frac{\pi}{2}} f(x)dx}_{=0}+\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(x)dx+\underbrace{\int_{\frac{\pi}{2}}^{\infty} f(x)dx}_{=0}=a\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos x\,dx=2a=1 \quad\Rightarrow\quad a=\frac{1}{2}$$

2) $$f(x)=\begin{cases} \dfrac{1}{2}\cos x, & x\in\left(-\dfrac{\pi}{2};\dfrac{\pi}{2}\right)\\ 0, & x\notin\left(-\dfrac{\pi}{2};\dfrac{\pi}{2}\right)\end{cases}$$



3) $$P\left(0<X<\frac{\pi}{4}\right)=\int_{0}^{\frac{\pi}{4}}\frac{\cos x}{2}dx=\frac{\sqrt{2}}{4}.$$

4) Найдем *F(x)*:



1°. $x<-\dfrac{\pi}{2} \quad\Rightarrow\quad F(x)=\underbrace{\int_{-\infty}^{x} f(x)dx}_{=0}=0;$

2°. $-\frac{\pi}{2} < x < \frac{\pi}{2}$

$$\Rightarrow \quad F(x) = \int_{-\infty}^{x} f(x)dx = \underbrace{\int_{-\infty}^{-\frac{\pi}{2}} f(x)dx}_{=0} + \int_{-\frac{\pi}{2}}^{x} f(x)dx = \frac{1}{2}\int_{-\frac{\pi}{2}}^{x} \cos x\, dx = \frac{1}{2}\sin x \Big|_{-\frac{\pi}{2}}^{x} = \frac{1}{2}(\sin x + 1)$$

3°. $x > \frac{\pi}{2}$

$$\Rightarrow \quad F(x) = \int_{-\infty}^{x} f(x)dx = \underbrace{\int_{-\infty}^{-\frac{\pi}{2}} f(x)dx}_{=0} + \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(x)dx + \underbrace{\int_{\frac{\pi}{2}}^{x} f(x)dx}_{=0} =$$

$$= \frac{1}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos x\, dx = \frac{1}{2}\sin x \Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 1$$

Числовые характеристики непрерывных случайных величин

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

Математическим ожиданием непрерывной случайной величины называется

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

При этом, конечно, предполагается, что несобственный интеграл сходится.

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

$$M(X) = \int_{a}^{b} xf(x)dx.$$

Дисперсией непрерывной случайной величины называется математическое ожидание квадрата ее отклонения.

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

По аналогии с дисперсией дискретной случайной величины, для практического вычисления дисперсии используется формула:

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

Пример: Плотность распределения случайной величины X имеет вид:

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

Найти $M(X)$, $D(X)$, σ.

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

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

Пример: $f(x)=\begin{cases}0, x\le 1\\ \frac{A}{x^4}, x>1\end{cases}$

Определить A, $F(x)$, $M(X)$,$D(X)$,$P(2<X<3)$.

Решение: из условия нормировки $\int_1^\infty \frac{A}{x^4}dx=1$ следует $A=3;$

$F(x)=\int_{-\infty}^x f(x)dx$, при $x\le 1, F(x)=0$, при $x>1$ $F(x)=0+\int_1^x \frac{3}{x^4}dx=1-\frac{1}{x^3}$

$$F(x)=\begin{cases}0, x\le 1\\ 1-\frac{1}{x^3}, x>1\end{cases};$$

$$M(X)=\int_1^\infty x\frac{3}{x^4}dx=\int_1^\infty \frac{3}{x^3}dx=3\frac{x^{-2}}{-2}\Big|_1^\infty=3\left(\frac{1}{2}\right)=\frac{3}{2};$$

$$D(X)=\int_1^\infty x^2\frac{3}{x^4}dx-\left(\frac{3}{2}\right)^2=\int_1^\infty \frac{3}{x^2}dx-\left(\frac{3}{2}\right)^2=3\frac{x^{-1}}{-1}\Big|_1^\infty-\frac{9}{4}=\frac{3}{4};$$

$$P(2<X<3)=\int_2^3 \frac{3}{x^4}dx=3\frac{x^{-3}}{-3}\Big|_2^3=3\left(\frac{3^{-3}}{-3}-\frac{2^{-3}}{-3}\right)=\frac{1}{8}-\frac{1}{27}=\frac{19}{216}.$$

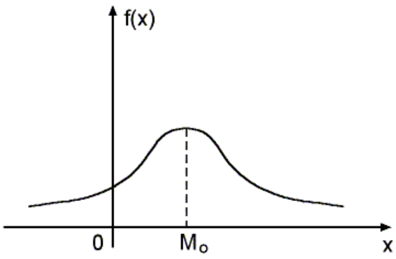
Другие характеристики СВ

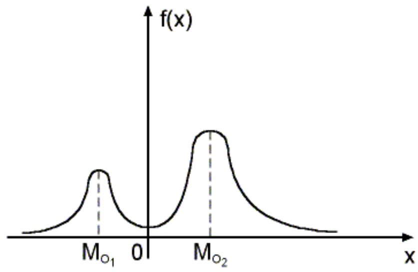
1. **Характеристики положения**

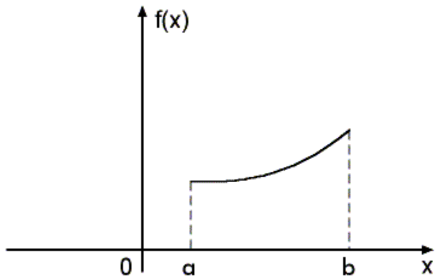
Математическое ожидание, мода, медиана характеризуют положение случайной величины на числовой оси, т.е. указывают некоторое среднее, около которого группируются все возможные значения СВ.

Модой M_0 дискретной случайной величины называется ее наиболее вероятное значение. Для непрерывной случайной величины мода – такое значение случайной величины, при которой плотность распределения имеет максимум.

Если многоугольник распределения для дискретной случайной величины или кривая распределения для непрерывной случайной величины имеет два или несколько максимумов, то такое распределение называется **двухмодальным** или **многомодальным**. Если распределение имеет минимум, но не имеет максимума, то оно называется **антимодальным**.







Медианой Me случайной величины X называется такое ее значение, относительно которого равновероятно получение большего или меньшего значения случайной величины.

$$P(X < Me) = P(X > Me)$$

Геометрически медиана — это абсцисса точки, в которой площадь, ограниченная кривой распределения, делится пополам. Каждая из этих площадей равна 1/2, так как вся площадь, ограниченная кривой распределения, равна единице.

Следовательно, $F(Me)=P(X<Me)=\frac{1}{2}$

Если распределение вероятностей случайной величины X симметрично относительно некоторой прямой $x=a$ и унимодально, то $MX=M_0=Me$.

Для асимметричных распределений это не так.

Пример: Если ряд распределения дискретной случайной величины X имеет вид:

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

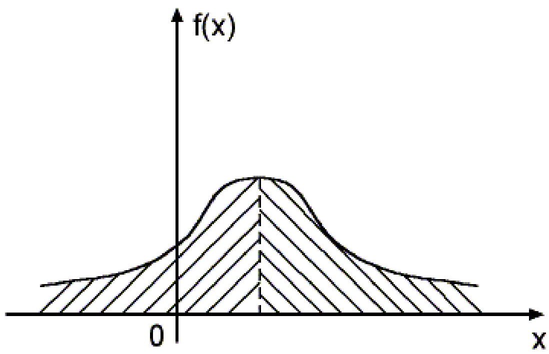
то $M_0= 2$.

Пример: Для непрерывной случайной величины, заданной плотностью распределения $f(x)=\dfrac{1}{\pi(1+x^2)}$, модой является абсцисса точки максимума: $M_0 = 0$.

Пример: Найти a, $M(X)$, M_0, Me для СВ: $f(x)=\begin{cases} 0, x<2 \\ a(x-2)(4-x), 2\le x\le 4 \\ 0, x>4 \end{cases}$.

Ответ: a=0.75, $M_0=Me=M(X)=3$.

2. Моменты



В теории вероятностей оказывается целесообразным рассматривать математическое ожидание целой положительной степени СВ. Например, для СВ, заданной законом распределения

X	1	2	5	100
p	0.6	0.2	0.19	0.01

получим: $M(X)$=2.95 , $M(X^2)$=106.15. Во втором результате мы видим влияние значения СВ при x=100, т.е. с помощью таких операций можно выявлять какие-либо особенности СВ.

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

$$\nu_k = M\,(X^k).$$

В частности, $\nu_1 = M(X)$, $\nu_2 = M(X^2)$. Следовательно, дисперсия $D(X) = \nu_2 - \nu_1^2$.

Для дискретной случайной величины X $\nu_k = \sum_{i=1}^{n} x_i^k p_i$

где x_i — значение случайной величины X, p_i — соответствующие вероятности.

Для непрерывной случайной величины X $\nu_k = \int_{-\infty}^{\infty} x^k f(x)dx$, где $f(x)$ - плотность распределения.

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

$$\mu_k = M((X - M(X))^k)$$

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

Можно доказать, что если распределение вероятностей случайной величины X симметрично относительно MX, то все центральные моменты нечётного порядка равны нулю:

$$\mu_{2к+1}=0\ ,\ \ k=1, 2, \dots$$

Можно получить соотношения, связывающие начальные и центральные моменты: $\mu_2 = \nu_2 - \nu_1^2,\ \ \mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^2,\ \ \mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4.$

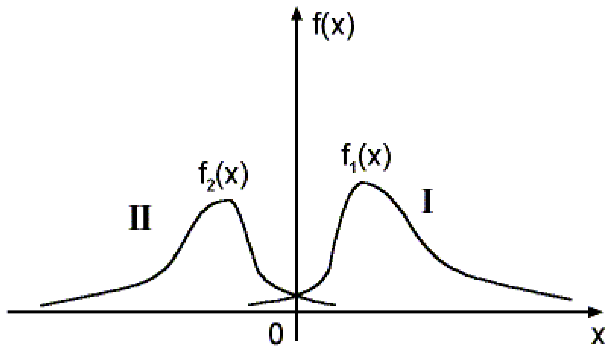
3. Асимметрия и эксцесс

Чем больше моментов случайных величин известны, тем более детальное представление о законе распределения мы имеем. В теории вероятностей и её приложениях используют две числовые характеристики случайных величин, основанные на центральных моментах 3-го и 4-го порядков соответственно, — коэффициент асимметрии и эксцесс. Моменты более высокого порядка на практике обычно не используются. Коэффициент асимметрии и эксцесс дают представление о форме плотности распределения или многоугольнике распределения. Эти характеристики являются безразмерными величинами.

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

Коэффициентом асимметрии (“скошенности”, от английского skew — “косой”) случайной величины X называется число

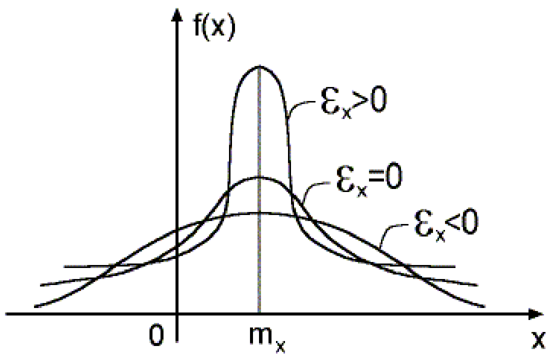
$$\alpha_x (= S_K) = \frac{\mu_3}{\sigma^3}$$



На рисунке изображены две асимметричные кривые распределения $f_1(x)$ и $f_2(x)$; I имеет положительную асимметрию (S_K>0); II - отрицательную (S_K< 0).

Четвёртый центральный момент μ_4 служит для характеристики так называемой “крутости”, то есть островершинности или плосковершинности распределения. Это свойство характеризуется с помощью **эксцесса**

$$\varepsilon_x = \frac{\mu_4}{\sigma^4} - 3.$$



Число 3 вычитается из отношения $\frac{\mu_4}{\sigma^4}$ потому, что для весьма распространённого и часто встречающегося нормального распределения (о нём мы расскажем позже) отношение $\frac{\mu_4}{\sigma^4} = 3$. Для нормального распределения ε_x=0 (см. рисунок)

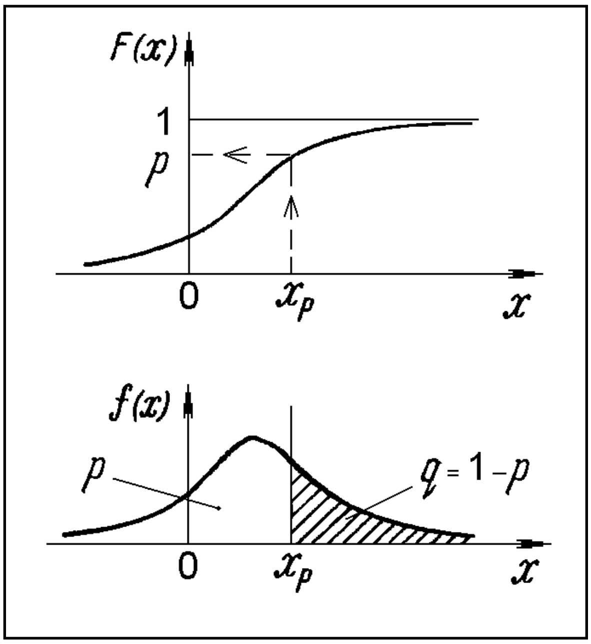
Характеристикой ε_x пользуются главным образом для симметричных распределений.

Итак, характеристики положения, моменты, асимметрия, эксцесс – наиболее употребляемые характеристики СВ. Часто закон распределения (плотность распределения) – полная характеристика СВ – или не нужна, или не может быть получена. Тогда ограничиваются приближённым описанием СВ с помощью числовых характеристик, каждая из которых выражает какое-либо свойство распределения.

4. Квантиль и процентные точки

При построение интервальных оценок и использовании статистических критериев используются понятия квантилей и процентных точек.

Квантилью уровня p или p– квантилью непрерывной случайной величины X, обладающей непрерывной функцией распределения $F(x)$, называется такое возможное значение x_p этой случайной величины, для которого выполняется $F(x_p)=P(X< x_p)=p$.



Смысл квантиля показан на первом рисунке: по квантилю приписанная вероятность определяется по кривой функции распределения ; квантиль делит площадь под кривой плотности распределения на две части – p слева от квантиля и $q = 1 – p$ — справа , т.е. вероятность нахождения случайной величины слева от квантиля равна p, справа $q = 1 – p$.

Частным случаем является медиана – это 0.5- квантиль. Некоторые квантили получили особое название. Квантили $x_{0,25}$ и $x_{0,75}$ получили название соответственно верхняя и нижняя квантиль. Квантили $x_{0,1}$, $x_{0,2}$,…, $x_{0,9}$ называются децили.

Под Q – процентной точкой ($0< Q < 100$) случайной величины X понимается такое её возможное значение w_Q, для которого выполняется соотношение

$$1 - F(w_Q) = P(X \geq w_Q) = Q/100.$$

Связь между этими величинами $x_p = w_{100(1-p)}$.

Например: для случая стандартного нормального распределения

а) p=0,25 , $u_{0,25}$= -0.675 ; б) Q=5% , $w_{5\%}$=1.65 .

Для наиболее часто встречающихся законов распределения составлены таблицы квантилей и процентных точек. Очевидно, что достаточно иметь только таблицу одной из этих величин, так как, если известна 0,9-квантиль, то известна 10%-точка того же распределения.

Равномерный закон распределения

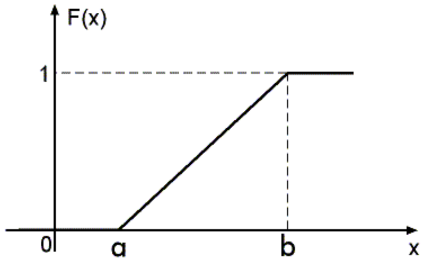
Говорят, что случайная величина X имеет **равномерное распределение** на участке (a, b) (обозначается кратко: $X \sim R(a,b)$), если её плотность на этом участке постоянна, то есть $f(x)=A=const$ при $x \in (a, b)$ и равна нулю вне его. Для определения A найдём

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_{a}^{b} f(x)dx + \int_{b}^{\infty} f(x)dx = A\int_{a}^{b} dx = 1 \quad \Rightarrow A = \frac{1}{b-a}$$

то есть

$$f(x) = \begin{cases} \dfrac{1}{b-a} & , \quad x \in (a,b); \\ 0 & , \quad x \notin (a,b). \end{cases}$$

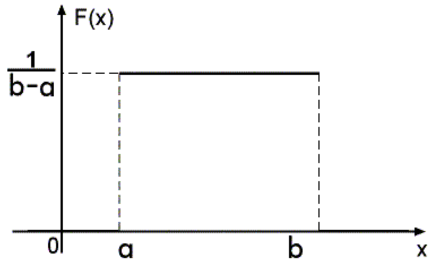
Найдём $F(x) \stackrel{\Delta}{=} \int\limits_{-\infty}^{x} f(x)dx$:



$$1) x \le a \quad F(x) = \int\limits_{-\infty}^{x} f(x)dx = 0;$$

$$2) x \in (a,b] \quad F(x) = \int\limits_{-\infty}^{x} f(x)dx = \int\limits_{-\infty}^{a} f(x)dx + \int\limits_{a}^{x} f(x)dx = \frac{1}{b-a}\int\limits_{a}^{x} dx = \frac{x-a}{b-a};$$

$$3) x > b \quad F(x) = \int\limits_{-\infty}^{x} f(x)dx = \int\limits_{-\infty}^{a} f(x)dx + \int\limits_{a}^{b} f(x)dx + \int\limits_{b}^{x} f(x)dx = \frac{1}{b-a}\int\limits_{a}^{b} dx = 1.$$



$$MX = \int\limits_{-\infty}^{\infty} xf(x)dx = \int\limits_{a}^{b} \frac{x}{b-a}dx = \frac{a+b}{2}: \quad DX = \int\limits_{-\infty}^{\infty}\left(x - \frac{a+b}{2}\right)^2 \frac{dx}{b-a} = \frac{(b-a)^2}{12} \Rightarrow \sigma_X = \frac{b-a}{2\sqrt{3}}.$$

Мода —M_0 — не существует.

Медиана — $M_e = \dfrac{a+b}{2}$ — из соображения симметрии.

Из тех же соображений третий центральный момент случайной величины X $\mu_3=0$ и коэффициент асимметрии равен 0.

Для определения эксцесса найдём четвёртый центральный момент

$$\mu_4 = \frac{1}{b-a}\int\limits_{a}^{b}\left(x - \frac{a+b}{2}\right)^4 dx = \frac{1}{b-a} \cdot \left.\frac{\left(x - \frac{a+b}{2}\right)^5}{5}\right|_a^b =$$

$$= \frac{1}{5(b-a)}\left[\left(b - \frac{a+b}{2}\right)^5 - \left(a - \frac{a+b}{2}\right)^5\right] = \frac{1}{5(b-a)}\left[\frac{(b-a)^5 - (a-b)^5}{2^5}\right] =$$

$$= \frac{2(b-a)^5}{5\cdot 32\cdot(b-a)} = \frac{(b-a)^4}{80} \quad \Rightarrow \quad \varepsilon_x \stackrel{\Delta}{=} \frac{\mu_4}{\sigma_x^4} - 3 = -1{,}2$$

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

Решение: Время ожидания является случайной величиной, равномерно распределенной в интервале [0, 5]. Тогда $f(x) = \dfrac{1}{5}, \; p(0 \le x \le 2) = \dfrac{2}{5} = 0{,}4.$

Показательный закон распределения

Показательным (экспоненциальным) (обозначение: $X \sim E_x(\lambda)$) называется распределение вероятностей непрерывной случайной величины X, которое описывается плотностью

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

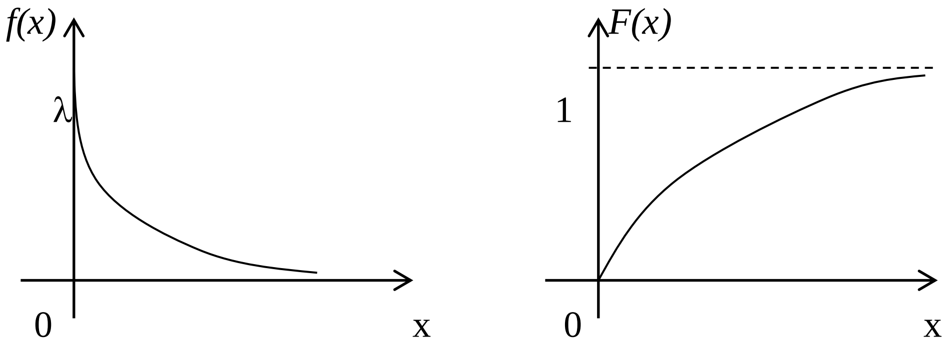
где λ - положительное число.

Условие нормировки выполняется: $\int_0^\infty \lambda e^{-\lambda x}dx = \frac{\lambda}{-\lambda}e^{-\lambda x}\Big|_0^\infty = 1$.

Найдем закон распределения:

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{0}0dx+\lambda\int_{0}^{x}e^{-\lambda x}dx=1-e^{-\lambda x}. \qquad F(x)=\begin{cases}0, & при \quad x<0\\ 1-e^{-\lambda x} & при \quad x\ge 0\end{cases}$$

Графики функции распределения и плотности распределения:



Найдем математическое ожидание случайной величины, подчиненной показательному распределению.

$$M(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_{0}^{\infty}x\lambda e^{-\lambda x}dx=\begin{Bmatrix}u=x; & e^{-\lambda x}dx=dv;\\ du=dx; & -\dfrac{e^{-\lambda x}}{\lambda}=v;\end{Bmatrix}=\lambda\left(-\frac{xe^{-\lambda x}}{\lambda}\Big|_0^\infty+\int_0^\infty\frac{e^{-\lambda x}}{\lambda}dx\right)=$$

$$=\int_0^\infty e^{-\lambda x}dx=-\frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty=\frac{1}{\lambda}.$$

Результат получен с использованием того факта, что

$$xe^{-\lambda x}\Big|_0^\infty=\lim_{x\to\infty}\frac{x}{e^{\lambda x}}-0=\begin{Bmatrix}По \quad правилу\\ Лопиталя\end{Bmatrix}=\lim_{x\to\infty}\frac{1}{-\lambda e^{\lambda x}}=0.$$

Для нахождения дисперсии найдем величину $M(X^2)$

$M(X^2)=\int\limits_{-\infty}^{\infty}x^2f(x)dx=\int\limits_{0}^{\infty}\lambda x^2e^{-\lambda x}dx$. Дважды интегрируя по частям, аналогично рассмотренному случаю, получим: $M(X^2)=\frac{2}{\lambda^2}$;

Тогда $D(X)=M(X^2)-[M(X)]^2=\frac{1}{\lambda^2}$. Итого: $M(X)=\frac{1}{\lambda}$; $D(X)=\frac{1}{\lambda^2}$; $\sigma_x=\frac{1}{\lambda}$.

Видно, что в случае показательного распределения математическое ожидание и среднее квадратическое отклонение равны.

Мода M_0=0, коэффициент асимметрия: $\alpha_x = 0$, эксцесс - $\varepsilon_x = 6$.

Также легко определить и вероятность попадания случайной величины, подчиненной показательному закону распределения, в заданный интервал.

$$P(a<x<b)=F(b)-F(a)=e^{-\lambda a}-e^{-\lambda b}.$$

Пример: Среднее время обслуживания покупателя 20 минут. Чему равна вероятность простоя в очереди от 20 до 40 минут?

$\frac{1}{\lambda}=20$, $\quad P(20<x<40)=e^{-1}-e^{-2}=0.23$

Экспоненциальное распределение является одним из основных распределений, используемых в теории надежности.

Например, продолжительность безотказной работы многих технических устройств, а также время задержки вылета самолѐта по вине технических служб аэропорта удовлетворительно описываются соответствующими экспоненциальными распределениями.

Экспоненциальное распределение описывает наработку до отказа объектов, у которых в результате сдаточных испытаний отсутствует период приработки, а назначенный ресурс установлен до окончания периода нормальной эксплуатации.

Экспоненциальный закон характерен для распределения случайных величин, изменение которых обусловлено влиянием доминирующего фактора. Он используется при рассмотрении внезапных отказов деталей в тех случаях, когда явления изнашивания и усталости выражены настолько слабо, что ими можно пренебречь.

Допустим, некоторое устройство начинает работать в момент времени t_0=0, а через какое– то время *t* происходит отказ устройства.

Обозначим *T* непрерывную случайную величину – длительность безотказной работы устройства. Таким образом, функция распределения *F(t) = P(T<t)* определяет вероятность отказа за время длительностью *t*.

Вероятность противоположного события (безотказная работа в течение времени *t*) равна *R(t) = P(T>t) = 1 – F(t)*.

Функцией надежности *R(t)* называют функцию, определяющую вероятность безотказной работы устройства в течение времени *t*.

Часто на практике длительность безотказной работы подчиняется показательному закону распределению. Функция надежности для какого- либо устройства при показательном законе распределения равна: $R(t)=1-F(t)=e^{-\lambda t}$.

Данное соотношение называют **показательным законом надежности**.

Важным свойством, позволяющим значительно упростить решение задач теории надежности, является то, что вероятность безотказной работы устройства на интервале времени t не зависит от времени предшествующей работы до начала рассматриваемого интервала, а зависит только от длительности времени t.

Таким образом, безотказная работа устройства зависит только от интенсивности отказов λ и не зависит от безотказной работы устройства в прошлом.

Так как подобным свойством обладает только показательный закон распределения, то этот факт позволяет определить, является ли закон распределения случайной величины показательным или нет.

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

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

Нормальный закон распределения (обозначается : $X \sim N(m,\sigma)$)

Нормальный закон распределения занимает центральное место в теории вероятностей. Это обусловлено тем, что этот закон проявляется во всех случаях, когда случайная величина является результатом действия большого числа различных факторов. К нормальному закону приближаются все остальные законы распределения.

Говорят, что случайная величина X распределена по **нормальному закону** (закону Гаусса) с параметрами m, σ, если

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

При этом $F(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$.

Замечание: В случае, если $m=0$ и $\sigma=1$, получаем так называемое нормированное или стандартное распределение $X \sim N(0,1)$:

$\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ (из локальной теоремы Лапласа) и связь $f(x) = \frac{1}{\sigma}\varphi(\frac{x-m}{\sigma})$, а также

$F_0(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$. Легко проверить, что $F(x) = F_0\left(\dfrac{x-m}{\sigma}\right)$

Учитывая, что $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$ и в силу симметрии $f(x)$ относительно нуля

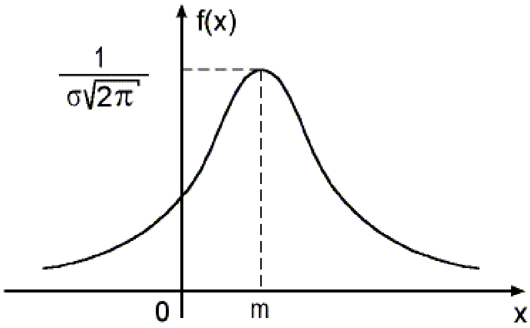
$\displaystyle\int_{-\infty}^{0} f(x)dx = \frac{1}{2} = 0{,}5$, получим $P(-\infty<X<0)=0{,}5$.

Легко получить, что $F_0(x)=P(-\infty<X<x)=P(-\infty<X<0)+P(0<X<x)=0{,}5+\Phi_0(x)$, где

$\Phi_0(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{0}^{x} e^{-\frac{t^2}{2}} dt$ - функция Лапласа, введенная ранее. Напомним,

$\Phi_0(-x) = -\Phi_0(x)$, поэтому $F(-x)=1-F(x)$.

Кривая нормального распределения имеет симметричный, холмообразный вид



Максимальная ордината кривой, равная $\frac{1}{\sigma\sqrt{2\pi}}$, достигается при $x=m$ (то есть мода, а также медиана, равны $M_0=Me=m$).

Можно легко показать, что параметры m и σ, входящие в плотность распределения являются соответственно математическим ожиданием и средним квадратическим отклонением случайной величины X:

$$MX \overset{\Delta}{=} \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{-\frac{(x-m)^2}{2\sigma^2}}dx = \left| \begin{array}{l} t=\dfrac{x-m}{\sigma} \\ x=t\sigma+m \\ dx=\sigma M \end{array} \right| = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(t\sigma+m)e\sigma^{-\frac{t^2}{2}}dt =$$

$$= \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} te^{-\frac{t^2}{2}}dt + \frac{m}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}dt = m \quad .$$

Первый интеграл равен нулю как интеграл от нечётной функции по симметричным пределам. Для вычисления второго интеграла воспользовались тем, что $\int_{-\infty}^{\infty} f(x)dx = 1$. В случае нормального распределения имеем

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}dx = \left| \begin{array}{l} t=\dfrac{x-m}{\sigma} \\ x=t\sigma+m \\ dx=\sigma dt \end{array} \right| = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}dt = 1 \quad \Rightarrow \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}dt = \sqrt{2\pi}$$

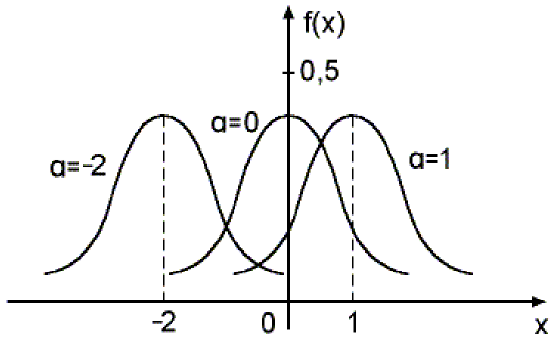
(интеграл Эйлера-Пуссона). Существует и другая форма интеграла Эйлера-Пуассона $\int_{-\infty}^{\infty} e^{-t^2}dt = 2\int_{0}^{\infty} e^{-t^2}dt = \sqrt{\pi}$.

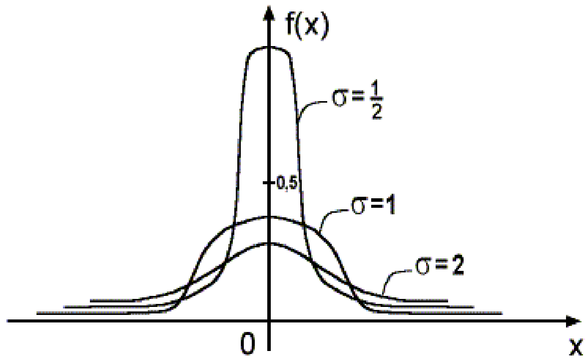
$$DX=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-m)^2e^{-\frac{(x-m)^2}{2\sigma^2}}dx=\left|\begin{array}{c} t=\dfrac{x-m}{\sigma}\\ x=t\sigma+m\\ dx=\sigma dt\end{array}\right|==\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(t\sigma+m-m)^2e^{-\frac{t^2}{2}}\sigma dt=$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2e^{-\frac{t^2}{2}}dt=\begin{bmatrix} u=t\,du=dt\\ dv=te^{-\frac{t^2}{2}}dtv=-e^{-\frac{t^2}{2}}\end{bmatrix}=\frac{2\sigma^2}{\sqrt{2\pi}}\left[-te^{-\frac{t^2}{2}}\Bigg|_{-\infty}^{\infty}+\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt\right]=\frac{\sigma^2}{\sqrt{2\pi}}\sqrt{2\pi}=\sigma^2.$$

(Использовали интегрирование по частям. Первое слагаемое равно нулю, так как e^{-t^2} при $t\to\infty$ убывает быстрее, чем возрастает любая степень t).

При изменении m и σ кривая нормального распределения будет меняться так, как это показано на рисунках (при этом площадь криволинейной трапеции равна 1).





Нетрудно получить, что эксцесс $\varepsilon_x=\frac{\mu_4}{\sigma^4}-3=0$ и коэффициент асимметрия $\alpha_x=0$.

Для вероятности попасть в промежуток мы ранее (интегральная теорема Лапласа) записали

$$P(\alpha<X<\beta)=\int_{\alpha}^{\beta}f(x)dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{\alpha}^{\beta}e^{-\frac{(x-m)^2}{2\sigma^2}}dx=\frac{1}{\sqrt{2\pi}}\int_{\frac{\alpha-m}{\sigma}}^{\frac{\beta-m}{\sigma}}e^{-\frac{t^2}{2}}dt=\Phi_0\left(\frac{\beta-m}{\sigma}\right)-\Phi_0\left(\frac{\alpha-m}{\sigma}\right).$$

Примеры: а) Найти для $X\sim N(0;1^2)$ вероятность $P(0<X<1.56)$:

$$P(0<X<1.56)=\varPhi_0(1.56)-\varPhi_0(0)=\varPhi_0(1.56)=0.44062$$

б) Найти для $X\sim N(0;1^2)$ вероятность $P(X<-2.47)$:

$P(X<-2.47)=P(X>2.47)=0.5-P(0<X<2.47)=0.5-\varPhi_0(2.47)=0.5-0.49324=0.00676$, или $P(X<-2.47)=\varPhi_0(-2.47)-\varPhi_0(-\infty)=0.5+\varPhi_0(-2.47)=0.5-\varPhi_0(2.47)=0.00676$.

в) Найти для $X\sim N(30;10^2)$ вероятность $P(10<X<50)$:

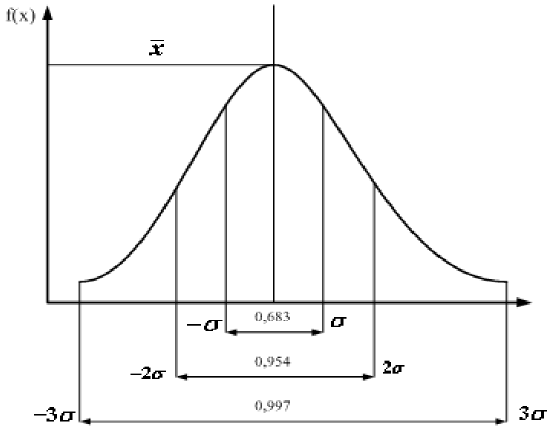
$P(10<X<50)=\varPhi_0\left(\frac{50-30}{10}\right)-\varPhi_0\left(\frac{10-30}{10}\right)=2\varPhi_0(2)=2\cdot0.4772=0.9544$.

Пример: Рост мужчин определённой возрастной группы есть нормально распределённая СВ с m=172 и $\sigma = 6$.

а) Найти долю костюмов 4-го размера (176-182) и 3-го роста (170-176).

Решение: $P(176 < X < 182) = \Phi_0\left(\frac{182-172}{6}\right) - \Phi_0\left(\frac{176-172}{6}\right) = \Phi_0(1.667) - \Phi_0(0.667) = 0.4525 - 0.2486 = 0.204$.

$P(170 < X < 176) = \Phi_0(0.667) - \Phi_0(-0.333) = 0.2486 + 0.1293 = 0.378.$



б) Найти квантиль $x_{0,7}$ и 10% точку этой СВ.

$P(X < t_{0,7}) = 0.7$, $P(X < t_{0,7}) = \Phi_0(t_{0,7}) - \Phi_0(-\infty) = \Phi_0(t_{0,7}) + 0.5 = 0.7$.

$\Phi_0(t_{0,7}) = 0.2$, по таблицам $t_{0,7} = 0.525$, $t_{0,7} = \frac{x_{0,7}-172}{6} = 0.525$, $x_{0,7} = 175.15$;

10% точка – это квантиль $x_{0,9}$, находится аналогично: $\Phi_0\left(\frac{x_{0,9}-172}{6}\right) = 0.9 - 0.5 = 0.4$, отсюда $\frac{x_{0,9}-172}{6} = 1.28$ и $x_{0,9} = 179.7$, т.е. 10% мужчин имеют рост не менее 180 см.

При рассмотрении нормального закона распределения выделяется важный частный случай, известный как **правило трех сигм**.

Часто требуется вычислить вероятность того, что отклонение нормально распределенной случайной величины по абсолютной величине меньше заданного положительного числа δ.

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

Если будем брать в качестве $\delta = \sigma, 2\sigma, 3\sigma$, то

$$P(|X - m| < \sigma) = 2\Phi_0(1) = 0.6826 ,$$
$$P(|X - m| < 2\sigma) = 2\Phi_0(2) = 0.9544 ,$$
$$P(|X - m| < 3\sigma) = 2\Phi_0(3) = 0.9973 .$$

То есть, вероятность того, что случайная величина отклонится от своего математического ожидание на величину, большую чем утроенное среднее квадратичное отклонение, практически равна нулю. Это правило называется **правилом трех сигм**.

Не практике считается, что если для какой – либо случайной величины выполняется правило трех сигм, то эта случайная величина имеет нормальное распределение. Это позволяет проверить, является ли изучаемая СВ распределённой нормально или приближённо определить σ: из полученных данных выбирают наибольшее и наименьшее и их разность делят на 6.

Примечание: эта формула применима и к частоте m (число совершения событий в n независимых испытаний) поскольку закон распределения (в данном случае биноминальный) при достаточно большом числе испытаний стремится к нормальному. Вспомним интегральную теорему Лапласа

$P_n(m_1 \le m \le m_2) = \Phi_0\left(\frac{m_2-np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{m_1-np}{\sqrt{npq}}\right)$ и сравним с

$$P(\alpha < X < \beta) = \Phi_0\left(\frac{\beta - m}{\sigma}\right) - \Phi_0\left(\frac{\alpha - m}{\sigma}\right)$$

Пусть X=m, тогда $M(m)$=np и $\sigma(m) = \sqrt{npq}$ и $P(|m - np| < \delta) \approx 2\Phi_0(\frac{\delta}{\sqrt{npq}})$.

Эта формула может быть применима и для относительной частоты m/n

$$P\left(\left|\frac{m}{n} - p\right| < \delta\right) = 2\Phi\left(\delta\sqrt{\frac{n}{pq}}\right).$$

Это вероятность отклонения относительной частоты от вероятности не более чем на заданное число.

Функция одного случайного аргумента и её распределение

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

Определение :Если каждому возможному значению случайной величины X соответствует одно возможное значение случайной величины Y, то Y называют **функцией случайного аргумента** $\boldsymbol{X}$: $Y = \varphi(X)$.

Выясним, как найти закон распределения функции по известному закону распределения аргумента.

1) Пусть аргумент X – дискретная случайная величина

а) Пусть различным значениям X соответствуют различные значения Y. Тогда вероятности соответствующих значений X и Y равны.

(*) Пример: Ряд распределения для X имеет вид:

X	5	6	7	8
p	0,1	0,2	0,3	0,4

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

Y	47	69	95	125
p	0,1	0,2	0,3	0,4

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

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

Пример: Ряд распределения для X имеет вид:

X	0	1	2	3
p	0,1	0,2	0,3	0,4

Найдем закон распределения функции $Y = X^2 - 2X$: Y -1 0 3

p 0,2 0,4 0,4

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

2) Если X – непрерывная случайная величина, $Y = \varphi(X)$, $\varphi(x)$ – монотонная и дифференцируемая функция, а $\psi(y)$ – функция, обратная к $\varphi(x)$, то плотность распределения $g(y)$ случайно функции Y равна: $g(y) = f(\psi(y))\,|\psi'(y)|$.

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

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

Математическое ожидание функции одного случайного аргумента.

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

Если X – дискретная случайная величина, то

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

Пример: Найдем $M(Y)$ для примера **(*)**:

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

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

$$M(Y) = \int_{-\infty}^{\infty} y g(y) dy.$$

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

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

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

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

Пример: $M(X)=3$ и $D(X)=1$, найти мат.ожидание и дисперсию для $Y=2X+3$ **.**

Решение: используя свойства мат.ожидания, получим $M(2X+3)=2M(X)+3=9$, $D(X)=D(2X)+D(3)=4D(X)+0=4$.

Приведём несколько примеров часто встречающихся случаев нахождения закона распределения Y.

1) Линейная функция от нормально распределённой СВ

$$Y = \gamma X + b \ , \ f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$X = \psi(Y) = (Y-b)/\gamma, \ X'(y) = 1/\gamma \ , g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{Y-b}{\gamma}-m\right)^2}{2\sigma^2}} \cdot \frac{1}{\gamma} .$$

Пример : $f(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x-1)^2}{18}}$, т.е. $M(X)$=1 , $D(X)$=9.

Найти $g(y)$ для $Y = -5X + 2$.

Решение: $g(y) = f\left(\frac{2-y}{5}\right) \cdot \left|\left(\frac{2-y}{5}\right)'\right| = \frac{1}{15\sqrt{2\pi}} e^{-\frac{(y+3)^2}{2\cdot 225}}$, т.е. если X – нормально распределённая СВ, то и Y тоже с

$$M(Y) = M(-5x + 2) = M(-5x) + M(2) = -5M(x) + 2 = -5 \cdot 1 + 2 = -3,$$

$$D(Y) = D(-5x + 2) = D(-5x) + D(2) = 25D(X) = 25 \cdot 9 = 225, \sigma = 15 .$$

2) Линейная функция от показательно распределенной случайной величины

$$f(x) = \begin{cases} 0, & x < 0 \\ 5e^{-5x}, & x \ge 0 \end{cases}, \quad Y = 2X + 3$$

Как в предыдущем пункте

$$X = \frac{Y-3}{2}, \; g(y) = f\left(\frac{y-3}{2}\right) \cdot \left|\left(\frac{y-3}{2}\right)'\right| = \begin{cases} 0, & y < 3 \\ \frac{5}{2} e^{-\frac{5}{2}y + \frac{15}{2}}, & y \ge 3 \end{cases} .$$

Можно привести такие соотношения:

если $X \sim N(0; 1)$, то $Y = \sigma X + a \sim N(a; \sigma^2)$;

если $X \sim N(a; \sigma^2)$, то $Y = (X - a)/\sigma \sim N(0; 1)$;

если $X \sim E_x(\boldsymbol{\lambda})$**,** то $Y = \lambda X \sim E_x(1)$**.**

3) X- СВ, равномерно распределённая на $[0; 1]$, найти $g(y)$ для $Y = -lnX$.

$f(x) = \begin{cases} 1, & x \in [0; 1] \\ 0, & x \notin [0; 1] \end{cases}$. Тогда $g(y) = 1 \cdot |-e^{-y}|$. $\quad g(y) = \begin{cases} 0, & y < 0 \\ e^{-y}, & y > 0 \end{cases}$.

Если мы включим датчик случайных чисел, равномерно распределённых на $[0; 1]$ и $x_1, x_2, \ldots, x_n$ – последовательность равномерно распределённых чисел, то соответствующая последовательность $y_1 = -\ln x_1$, $y_2 = -\ln x_2, \ldots, y_n = -\ln x_n$ является последовательностью чисел, распределённых по показательному закону.