Оглавление

ЧАСТЬ 2. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ.

Глава 3. Одномерные случайные величины.

При рассмотрении схемы Бернулли (для *n* независимых опытов) мы можем вычислить значения вероятностей по формуле Бернулли

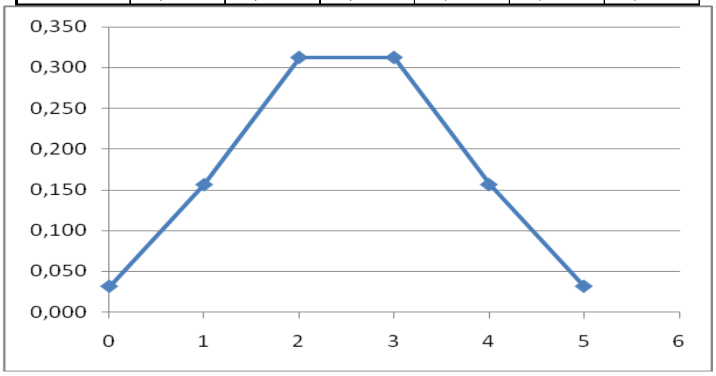
$$P_n(m) = C_n^m p^m q^{n-m}$$

для всех возможных значений числа успехов *m=0,1,…,n*
Наши вычисления можем расположить в виде таблицы

m	**0**	**1**	**…**	**n**
$P_n(m)$	$P_n(0)$	$P_n(1)$	$P_n(m)$	$P_n(n)$

Результаты можем отобразить в виде так называемого многоугольника вероятностей, например, для *n=5,p=0.5*:

m	0	1	2	3	4	5
$P_n(m)$	0,031	0,156	0,313	0,313	0,156	0,031



Можно говорить о величине X, принимающей значения $x_i = 0,1,...,n$ с вероятностями $p_i = P_n(i)$ - то есть о величине X, случайным образом принимающей определенные значения или случайной величине. Можно еще так ввести понятие случайной величины:

случайной величиной называется величина, которая в результате опыта может принять то или иное значение, неизвестно заранее, которое именно.

Случайные события в этом случае: $\{X = x_i\}$ - {случайная величина принимает одно из возможных значений}. Таким образом, задавая распределение вероятностей для случайной величины, мы перебираем все возможные варианты несовместных случайных событий. Как мы знаем, такие случайные события образуют полную группу, а соответствующие вероятности удовлетворяют условию нормировки – в сумме равны 1 (вероятности достоверного события).
Случайные величины мы будем обозначать большими латинскими X,Y,Z… или X_1, X_2, X_3… В некоторых источниках случайные величины обозначают греческими буквами ξ,η,θ… Принимаемые значения будем обозначать x_1 ,x_2,x_2…

Далее введем строгое определение случайной величины.

Пусть (Ω, S, p) – вероятностное пространство (Ω - пространство элементарных событий, S - σ-алгебра событий, p -вероятности событий); $\Re$ - множество вещественных чисел. Будем обозначать X случайную величину, x - принимаемые этой величиной значения.

Определение. ***Случайной величиной*** X называется числовая функция, определённая на пространстве элементарных событий Ω, которая каждому элементарному событию ω ставит в соответствие число $x = X(\omega) \in \Re$, причем функция $X(\omega)$ должна быть такова: $\forall x \; A = \{\omega: \; X(\omega) < x\} \in S$ (для любого x все такие события A принадлежат сигма-алгебре событий) и определена вероятность $p(A) = p\{X < x\}$.

Случайная величина, принимающая конечное или счётное число значений, называется ***дискретной***. Сокращенно будем писать "д.с.в.".
Под ***непрерывной случайной величиной*** будем понимать случайную величину, принимающую любые вещественные значения на некотором интервале или любое вещественное значение на всей числовой оси. Сокращенно будем писать "н.с.в.".

Определение. ***Законом распределения вероятностей случайной величины*** (или ***распределением)*** называется любое правило (таблица, функция, график), позволяющее находить вероятности произвольных событий из алгебры событий.

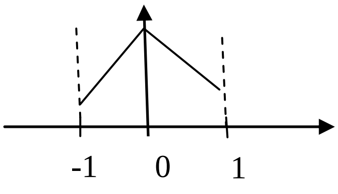
Закон распределения дискретной случайной величины может представлять собой таблицу, в которой значениям, принимаемым случайной величиной, сопоставлены их вероятности, причём, события $\{X = x_i\}, i = 1,2,\ldots,n$ образуют полную группу событий, то есть $\sum_{i=1}^{n} p_i = 1$ (условие нормировки):

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

Такая таблица представляет собой ***ряд распределения***.
Замечание. Случайная величина может принимать отрицательные значения, например, вполне допустимо распределение вида:

X	-1	0	1
P	0,1	0,7	0,2

Закон распределения можно задать графически, например, с помощью ***многоугольника распределения*** – в дискретном случае, в частности для приведенного примера, многоугольник вероятностей будет выглядеть так:



Значение случайной величины, которому соответствует наибольшая вероятность, называется ***модой***.

<u>Пример 1</u>: подбрасываем один раз игральный кубик. Обозначим за X количество выпавших очков. Тогда $\Omega = \{1,2,3,4,5,6\}$. *X равняется одному из элементарных событий, каждое событие происходит с вероятностью 1/6. Получаем распределение:*

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

Обратим теперь внимание на требования второй части определения, а именно, на события $\{X < x_i\}$ $\forall i$ *из алгебры событий.*

Все события $\{X = 1\}, \{X = 2\}, \{X = 3\}, \{X = 4\}, \{X = 5, \{X = 6\}$ *включим в алгебру событий S. По определению, всевозможные суммы и произведения этих событий тоже должны входить в S.*

Рассмотрим случайное событие $\{X < 1\}$ *- это невозможное событие,* $P\{X < 1\} = 0$. *По определению алгебры событий S невозможное событие должно в него входить.*

Событие $\{X < 2\} = \{X = 1\}$ *- событие из S,* $P\{X < 2\} = P\{X = 1\} = 1/6$.

Событие $\{X < 3\} = \{X = 1\} + \{X = 2\}$ *- сумма событий из S, в силу несовместности событий, вероятность такого события есть* $P\{X < 3\} = P\{X = 1\} + P\{X = 2\} = 1/6 + 1/6 = 2/6$.

Аналогично:

$$P\{X < 4\} = P\{X = 1\} + P\{X = 2\} + P\{X = 3\} = \\ = 1/6 + 1/6 + 1/6 = 3/6 .$$

$$P\{X < 5\} = P\{X = 1\} + P\{X = 2\} + P\{X = 3\} + P\{X = 4\} = \\ = 1/6 + 1/6 + 1/6 + 1/6 = 4/6 .$$

$$P\{X < 6\} = P\{X = 1\} + P\{X = 2\} + P\{X = 3\} + P\{X = 4\} + P\{X = 5\} = \\ = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 5/6$$

Наконец, событие $\{X < a\}$, *где* $a > 6$ *есть событие достоверное, его вероятность равна 1. Такое событие тоже должно входить в алгебру событий S.*

Таким образом, мы определили все возможные события $A = \{X < x_i\}$, *показали, что они входят в алгебру событий и вычислили их вероятности.*

Задание вероятностей $P\{X < x\}$ для всех возможных значений *x* играет важную роль в теории вероятностей.

Функция распределения

<u>Определение</u>. <u>Функцией распределения</u> случайной величины называется функция $F(x)$, которая для $\forall x \in \Re$ равна вероятности события $P\{X < x\}$:

$$\boxed{F(x) = P\{X < x\}}$$

<u>Замечание 1</u>. Функцию распределения еще называют ***<u>интегральной функцией распределения</u>***. Английский аналог – ***<u>cumulative distribution function</u>*** (CDF). Это полезно знать, если для решения задач использовать какой-либо математический пакет, например, Matlab.

<u>Замечание 2.</u> Задание функции распределения является самым универсальным способом описания распределения вероятностей случайной величины.

Свойства функции распределения.

$F(x)$ есть неубывающая, непрерывная слева функция, удовлетворяющая свойствам:

1. $0 \le F(x) \le 1$
2. $F(-\infty) = 0, \;\; F(+\infty) = 1$
3. $P\{a \le x \le b\} = F(a) - F(b)$

   (вероятность попадания в интервал).

Замечание 1. Согласно свойству 3, для непрерывной случайной величины вероятность попасть в точку а:

$$P\{X = a\} = P\{a \le x \le a\} = F(a) - F(a) = 0.$$

Поэтому

$$P\{a \le x \le b\} = P\{a < x < b\} = P\{a \le x < b\} = P\{a < x \le b\}.$$

Замечание 2. События $\{X < x\}$ и $\{X \ge x\}$ противоположны, следовательно:

$$P\{X \ge x\} = 1 - P\{X < x\} = 1 - F(x)$$

Функция распределения дискретной случайной величины:

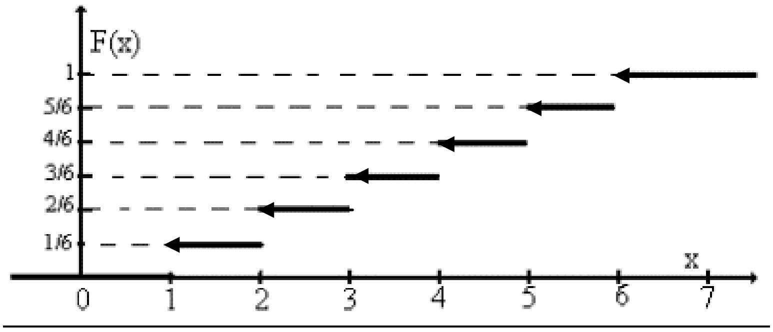
$$\boxed{F(x) = \sum_{x_k < x} p_k}$$

Для построения функции распределения дискретной случайной величины, мы должны расположить все значения x_k, принимаемые случайной величиной X, на оси x и для каждого интервала $x_k < x < x_{k+1}$ найти значение вероятности случайного события $\{X < x_{k+1}\}$, равное сумме вероятностей $\sum_{i=1}^{k} p_i$.

Вернемся к примеру 1. Вероятности всех событий $\{X < x\}$ мы уже вычислили и можем записать функцию распределения в этом случае:

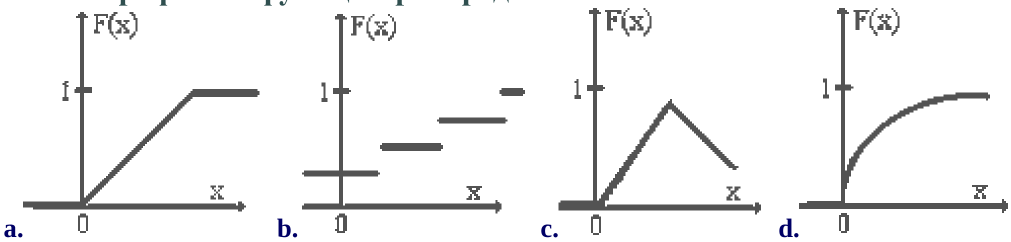
$$F(x) = \begin{cases} 0, \;\; x \le 1 \\ 1/6, \;\; 1 < x \le 2 \\ 2/6, \;\; 2 < x \le 3 \\ 3/6, \;\; 3 < x \le 4 \\ 4/6, \;\; 4 < x \le 5 \\ 5/6, \;\; 5 < x \le 6 \\ 1, \;\; x > 6 \end{cases}$$

График этой функции будет выглядеть так:



Тесты:

Укажите рисунок, на котором изображен график, который не может быть графиком функции распределения



Независимые случайные величины.

Пусть имеются дискретные случайные величины X и Y, распределение вероятностей которых задано таблицами:

X	x_1	x_2	...	x_n
P	p_{x1}	p_{x2}	...	p_{xn}

и

Y	y_1	y_2	...	y_m
P	p_{y1}	p_{y2}	...	p_{ym}

Случайные величины X и Y называются ***независимыми***, если независимы случайные события $\{X = x_i\}$ и $\{Y = y_j\}$ для всех *i* и *j*.

Для независимых случайных величин X и Y можно определить, например, их линейную комбинацию $Z = aX + bY$:

$Z = aX + bY$	$ax_1 + by_1$	$ax_2 + by_1$	...	$ax_n + by_n$
P	$p_{x1} \cdot p_{y1}$	$p_{x2} \cdot p_{y1}$	...	$p_{xn} \cdot p_{yn}$

- вероятности представляют собой вероятности произведения событий $\{X = x_i\} \cdot \{Y = y_j\}$.

В дальнейшем будем использовать такую запись: $\{X = x_i, Y = y_j\}$. В силу независимости случайных величин, вероятность произведения событий равна произведению вероятностей.

Замечание. Для совпадающих значений *Z* соответствующие вероятности складываются.

Пример.

X	*-1*	*0*	*1*
P	*0,3*	*0,4*	*0,3*

и

Y	*0*	*1*	*2*
P	*0,1*	*0,2*	*0,7*

Вычислим все возможные значения и соответствующие вероятности (для каждого значения x_i перебираем все значения y_j):

$Z = 2X - Y$	-2	-3	-4	0	-1	-2	2	1	0
P	0,3*0,1	0,3*0,2	0,3*0,7	0,4*0,1	0,4*0,2	0,4*0,7	0,3*0,1	0,3*0,2	0,3*0,7

$Z = 2X - Y$	-2	-3	-4	0	-1	-2	2	1	0
P	0,03	0,06	0,21	0,04	0,08	0,28	0,03	0,06	0,21

Учитывая одинаковые значения:

$Z = 2X - Y$	-4	-3	-2	-1	0	1	2
P	0,21	0,06	0,03+0,28=0,31	0,08	0,04+0,21=0,25	0,06	0,03

В сумме вероятности для Z равны 1, следовательно, эта таблица задает распределение.

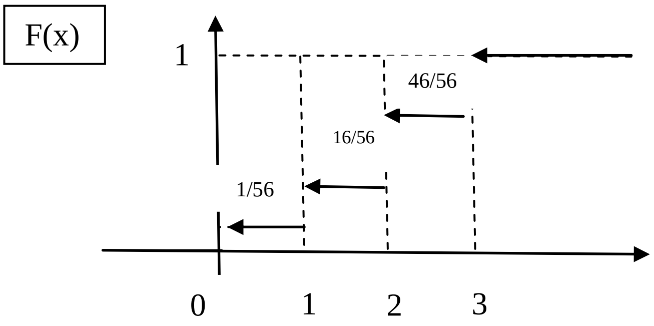
<u>Задача</u>. В урне 5 белых и 3 черных шара. Достали 3 шара. Найти распределение случайной величины, равной числу белых шаров в выборке. Построить функцию распределения.

С.в. X принимает значения от 0 до 3. Соответствующие вероятности вычисляем по формулам $\dfrac{C_5^m \cdot C_3^{3-m}}{C_8^3}$:

X	0	1	2	3
P	1/56	15/56	30/56	10/56

$F(x)$	x<0	0<x<1	1<x<2	2<x<3	x>3
P	0	1/56	16/56	46/56	1

("накопленные вероятности")



Индикатор события.

Случайная величина *I(A)* называется ***индикатором события A***, если она принимает значение 1, когда событие *A* происходит и принимает значение 0, если событие *A* не происходит, вероятность *p* определяется вероятностью события *A*:

X	*0*	*1*
P	*1-p*	*p*

Индикатор события позволяет описывать распределения достаточно сложных событий.

Пример 1. Игральный кубик бросается 1 раз. Найти распределение вероятностей для события $A = \{\text{выпало четное число очков}\}$. *Введем индикатор этого события, тогда:*

I	*0*	*1*
P	*3/6*	*3/6*

Пример 2. Подбрасываем два игральных кубика. Событие $A = \{\text{сумма очков равна } 5\}$.

I	*0*	*1*
P	*32/36*	*4/36*

Непрерывные случайные величины.

<u>Плотностью распределения</u> непрерывной случайной величины называется

$$\boxed{f(x) = F'(x)}$$

<u>Свойства</u>.

1. $f(x) \geq 0$
2. $\int\limits_{-\infty}^{\infty} f(x)dx = 1 \quad - \quad$ *условие нормировки*
3. $F(x) = \int\limits_{-\infty}^{x} f(t)dt$
4. $P\{a \leq x \leq b\} = F(a) - F(b) = \int\limits_{a}^{b} f(x)dx$

<u>Определение непрерывной случайной величины.</u>

С.в. X называется ***непрерывной***, если

$$\exists f(x): \begin{array}{l} 1)\ f(x) \geq 0 \\ 2)\ \forall x\ F(x) = \int\limits_{-\infty}^{x} f(t)dt \ \Rightarrow \ f(x) = F'(x) \end{array}$$

<u>Замечание</u>.

$f(x)$ называют также ***дифференциальной функцией распределения***. Английский термин: **probability density function (pdf)**.

Пример. *Задана функция* $f(x) = \dfrac{a}{(1+x^2)}$. *При каких а эта функция является плотностью распределения?*

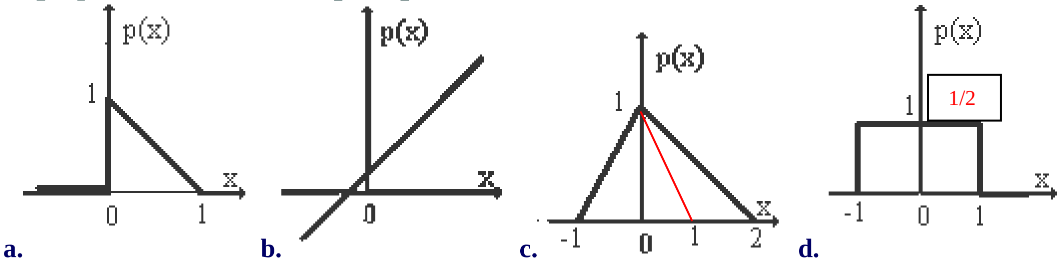
Условие нормировки: $\int\limits_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow$

$$a\int\limits_{-\infty}^{\infty}\frac{dx}{1+x^2} = a\cdot arctgx|_{-\infty}^{\infty} = a(\pi/2+\pi/2) = a\cdot\pi = 1 \Rightarrow a = 1/\pi$$

Распределение с плотностью $f(x) = \dfrac{1}{\pi}\cdot\dfrac{1}{(1+x^2)}$ *называется* *распределением Коши*.

Тесты:

Укажите рисунок, на котором изображен график, который может быть графиком плотности распределения



○	a.	**Рисунок под буквой a**
○	b.	**Рисунок под буквой b**
○	c.	**Рисунок под буквой c (с испралением)**
○	d.	**Рисунок под буквой d (с исправлением)**

5.

Какое свойство не обязательно для функции распределения ?

○	a.	**F(X) не отрицательна**
○	b.	**F(X) не имеет разрывов**
○	c.	**F(X) не убывает с ростом x**
○	d.	**F(X) не более 1**

Основные числовые характеристики.

Мы рассмотрели законы описания распределений случайной величины. В частности, *табличный* и графический способ для д.с.в. и функциональный способ задания, – с помощью функции плотности вероятности, - для н.с.в.
Универсальным способом задания является задание функции распределения.
Закон распределения полностью описывает случайную величину.
Однако зачастую нам достаточно знать вместо полного описания распределения случайной величины только некоторые числовые характеристики.
Основной числовой характеристикой является математическое ожидание. С помощью математического ожидания определяются остальные числовые характеристики – начальные и центральные моменты.

Математическое ожидание MX .

Определение. Для дискретной случайной величины X , принимающей значения x_i с соответствующими вероятностями p_i :

$$\boxed{MX = \sum_{i=1}^{n} x_i \cdot p_i}$$

(в случае счетного числа значений $n = \infty$ и ряд предполагается абсолютно сходящимся).

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

$$\boxed{MX = \int_{-\infty}^{\infty} x \cdot f(x) dx}$$

(интеграл предполагается абсолютно сходящимся).
Замечание. Математическое ожидание является средневзвешенным значений случайной величины, в которой в качестве весов принимаются соответствующие вероятности:

$$MX = \frac{\sum_{i=1}^{n} (x_i \cdot p_i)}{\sum_{i=1}^{n} p_i}, \text{ где } \sum_{i=1}^{n} p_i = 1$$

Математическое ожидание является центром рассеивания случайной величины.
Механическим аналогом является центр масс:

$$\frac{\sum_{i=1}^{n} (x_i \cdot m_i)}{\sum_{i=1}^{n} m_i}, \text{ где } \sum_{i=1}^{n} m_i = m$$

(где m_i - массы, сосредоточенные в точках с координатами x_i).

Часто в литературе встречается такое обозначение математического ожидания случайной величины ξ: $E\xi$.

В англоязычных статьях математическое ожидание называется ***mean*** ("среднее").

Математическое ожидание имеет размерность случайной величины: $[MX]=[X]$.

Примеры вычисления математического ожидания.

Д.с.в.

1.

X	*-1*	*0*	*1*
P	*0,5*	*0,3*	*0,2*

$MX=-1\cdot 0.5+0\cdot 0.3+1\cdot 0.2=-0.3$ *(мат.ожидание может быть отрицательным)*

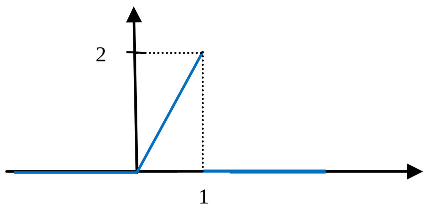
2. *Из 1000 выигрышных билетов на 10 билетов приходится выигрыш 500 рублей, на 50 – 50 рублей, на 100 – 10 рублей, на 150 билетов – 1 рубль. Найти математическое ожидание выигрыша на один билет.*

X	*0*	*1*	*10*	*50*	*500*
P	*690/1000=0,69*	*150/100=0,15*	*100/1000=0,1*	*50/1000=0,05*	*10/1000=0,01*

$MX=1\cdot 0.15+10\cdot 0.1+50\cdot 0.05+500\cdot 0.01=8.65$ *рублей.*

Н.с.в.

1. Пусть плотность н.с.в. задана графиком:



$$f(x)=\begin{cases}0, & x\le 0\\ 2x, & 0<x\le 1,\\ 0, & x>1\end{cases} \qquad MX=\int_0^1 x\cdot 2x\,dx=2x^3/3\big|_0^1=2/3$$

2. Пусть с.в. имеет плотность $f(x)=\dfrac{1}{\pi}\cdot\dfrac{1}{(1+x^2)}$, $x\in\Re$.

Тогда $MX=\displaystyle\int_{-\infty}^{\infty} x\cdot\frac{1}{\pi}\cdot\frac{1}{(1+x^2)}dx=\frac{1}{2\pi}\ln|1+x^2|\,\Big|_{-\infty}^{\infty}$ - не существует!!!

Свойства математического ожидания.

1. Математическое ожидание от константы: $M(c)=c$. В частности, $M(MX)=MX$.
2. Вынесение константы: $M(cX)=cMX$.
3. $M(X\pm Y)=MX\pm MY$.
4. $M(X-MX)=0$.
5. Для <u>независимых</u> с.в. X и Y: $M(XY)=MX\cdot MY$.

<u>Пример</u>. *Пусть* $MX=0.1$, $MY=0.3$. *Найти мат. ожидание* $Z=3X-2Y+1$.
$MZ=M(3X-2Y+1)=3MX-2MY+1=0.7$

Начальный момент порядка k:

$$\boxed{\alpha_k = M(X^k)}$$

В случае дискретной случайной величины:

$$\alpha_k = \sum_{i=1}^{n} x_i^{\ k} p_i$$

В случае непрерывной случайной величины:

$$\alpha_k = \int_{-\infty}^{\infty} x^k f(x)dx;$$

Замечание.

$$\alpha_1 = MX$$

.

Пример.

X	-1	0	1
P	0,5	0,3	0,2

$M(X^2) = (-1)^2 \cdot 0.5 + 0^2 \cdot 0.3 + 1^2 \cdot 0.2 = 0.7$

$M(X^3) = (-1)^3 \cdot 0.5 + 0^3 \cdot 0.3 + 1^3 \cdot 0.2 = -0.3$

Центральный момент порядка k:

$$\boxed{\mu_k = M(X - MX)^k}$$

Для дискретной случайной величины:

$$\mu_k = \sum_{i=1}^{n} (x_i - MX)^k p_i$$

Для непрерывной случайной величины:

$$\mu_k = \int_{-\infty}^{\infty} (x - m_x)^k \cdot f(x)dx$$

Особо выделяют центральный момент второго порядка $\mu_2 = \alpha_2 - \alpha_1^{\ 2}$, который называется дисперсией.

Дисперсия:

$$\boxed{DX = M(X - MX)^2}$$

Дисперсия характеризует рассеяние значений случайной величины относительно центра рассеивания.

Размерность дисперсии равна квадрату размерности случайной величины: $[DX] = [X^2]$.

Часто дисперсию бывает удобнее вычислять по формуле:

$$\boxed{DX = M(X^2) - (MX)^2 = \alpha_2 - \alpha_1^{\ 2}}$$

Английское название – ***variance*** ("отклонение"). Обозначение: $Var(X)$.

Свойства дисперсии.

1. $D(c) = 0$ - дисперсия от константы
2. $D(cX) = c^2 DX$ - вынесение константы из-под знака дисперсии (как оператора)
3. $D(c + X) = DX$ - добавление константы не меняет дисперсию

Для независимых случайных величин X и Y:

4. $D(X \pm Y) = DX + DY$
5. $D(X \cdot Y) = MX^2 \cdot MY^2 - (MX)^2 \cdot (MY)^2$

Среднее квадратическое отклонение (с.к.о.)

$$\boxed{\sigma_x = \sqrt{DX}}$$

Размерность с.к.о. совпадает с размерностью случайной величины.
Замечание. Часто для дисперсии DX используют обозначение $\sigma_x^{\ 2}$.
Английское название – ***standard variance*** ("стандартное отклонение").

Стандартная случайная величина.

Случайная величина, равная отклонению значений случайной величины от математического ожидания, называется ***центрированной*** и обозначается $\overset{0}{X} = X - MX$.
Математическое ожидание центрированной случайной величины равно 0.
Центрированная и нормированная случайная величина $Z = \dfrac{X - MX}{\sigma_x}$ называется ***стандартной***.
Основные числовые характеристики стандартной величины: $MZ = 0$, $DZ = 1$

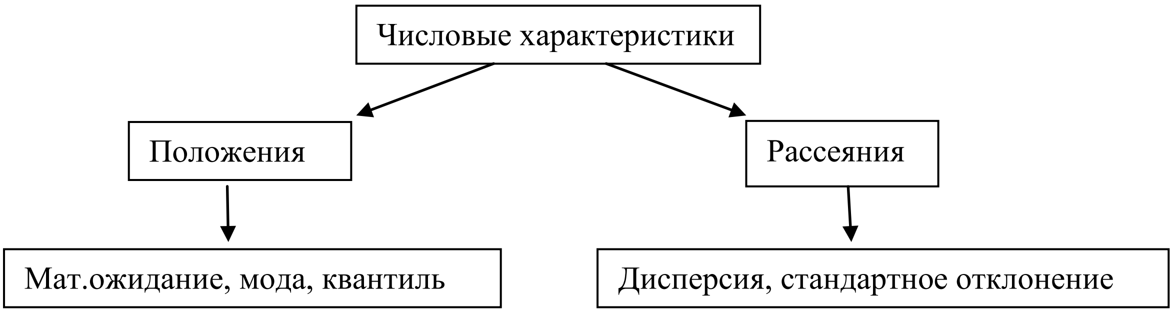
Свойства основных числовых характеристик (сводная таблица).

(X,Y – независимые случайные величины!!!)

Сводная таблица в операторной форме.

Математическое ожидание **MX**	**Дисперсия** **DX**	**С.к.о.** σ_x
Действие на константу:		
$Mc = c$	$Dc = 0$	$\sigma_c = 0$
Вынесение константы:		
$McX = cMX$	$DcX = c^2 DX$	$\sigma_{cX} = \lvert c \rvert \sigma_x$
Действие на сумму и разность:		
$M(X \pm Y) =$ $= MX \pm MY$	$D(X \pm Y) =$ $= DX + DY$	$\sigma(c + X) = \sigma_x$

Классификация числовых характеристик.



Мода

Значение с.в., принимаемое с наибольшей вероятностью. Обозначение M_0X
Локальный минимум функции плотности. Один локальный минимум - ***унимодальное*** распределение. Несколько – ***полимодальное*** распределение.

Квантиль уровня p.

Значение случайной величины x_p, являющееся решением уравнения:

$$F(x_p) = p.$$

$F(x_p) = \int\limits_{-\infty}^{x_p} f(x)dx$ - площадь фигуры, ограниченной графиком плотности $f(x)$, осью x и прямой $x = x_p$.

- $x_{0.25}$ - 0.25-квантиль называется ***первой (или нижней) квартилью***;
- $x_{0.5} = M_eX$ - 0.5-квантиль называется ***медианой*** или ***второй квартилью***;
- $x_{0.75}$ - 0.75-квантиль называется ***третьей (или верхней) квартилью***.

Площади A,B,C,D равны по ¼.

Интерквартильным размахом (англ. *Interquartile range*) называется разность между третьей и первой квартилями, то есть $x_{0.75} - x_{0.25}$. Интерквартильный размах является характеристикой разброса распределения величины и является робастным аналогом дисперсии. Вместе, медиана и ***интерквартильный*** размах могут быть использованы вместо *математического ожидания* и *дисперсии* в случае распределений с большими выбросами, либо при невозможности вычисления последних.

Дециль

Дециль характеризует распределение величин совокупности, при которой девять значений дециля делят ее на десять равных частей. Любая из этих десяти частей составляет 1/10 всей совокупности. Так, первый дециль отделяет 10% наименьших величин, лежащих ниже дециля от 90% наибольших величин, лежащих выше дециля.

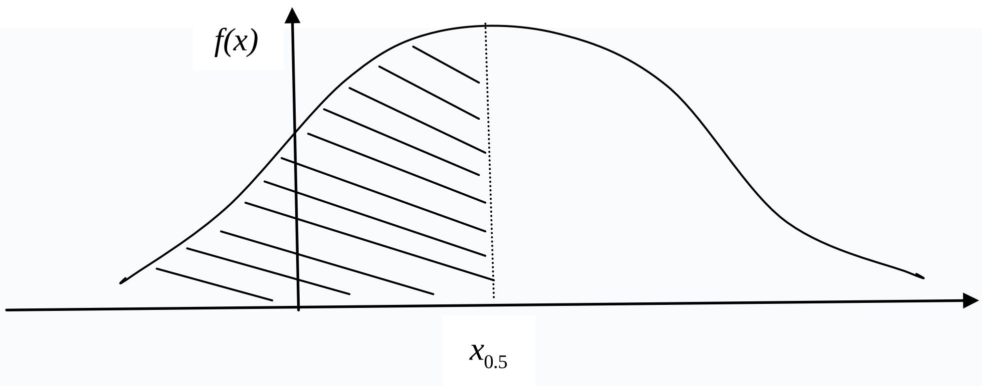
Перцентиль

p-ой называют квантиль уровня α = *p* / 100. При этом обычно рассматривают перцентили для целых *p*, хотя данное требование не обязательно. Соответственно, медиана является 50-ой перцентилью, а первая и третья квартиль — 25-ой и 75-ой перцентилями. В целом, понятия квантиль и перцентиль взаимозаменяемы, также, как и шкалы исчисления вероятностей — абсолютная и процентная. Перцентили также называются ***процентилями*** или ***центилями***.

Медиана. $M_e X = x_{0.5}$

$$F(x_{0.5}) = P\{X < x_{0.5}\} = 1 - F(x_{0.5}) = P\{X > x_{0.5}\} = 0.5.$$

Прямая $x = x_{0.5} = M_e X$ делит площадь под графиком плотности ровно пополам:



Медиана является важной характеристикой распределения случайной величины и так же, как математическое ожидание, может быть использована для центрирования распределения. Однако, медиана более робастна и поэтому может быть более предпочтительной для распределений с так называемыми *тяжёлыми хвостами*.

Медиана определяется для широкого класса распределений (например, для всех непрерывных), а в случае неопределённости, естественным образом доопределяется, в то время как математическое ожидание может быть не определено (например, у распределения Коши).

Пример использования

Предположим, что в одной комнате оказалось 19 бедняков и один миллиардер. Каждый кладет на стол деньги из своего кармана. По пять долларов кладет каждый бедняк, а миллиардер — $1 млрд (10*^{9}). В сумме получается $1 000 000 095. Если мы разделим деньги равными долями на 20 человек, то получим $50 000 004,75. Это будет среднее арифметическое значение суммы наличных, которая была у всех 20 человек в этой комнате.*

Медиана в этом случае будет равна $5 (полусумма десятого и одиннадцатого, срединных значений ранжированного ряда). Можно интерпретировать это следующим образом.

Разделив нашу компанию на две равные группы по 10 человек, мы можем утверждать, что в первой группе каждый положил на стол не больше $5, во второй же не меньше $5. В общем случае можно сказать, что медиана это то, сколько принес с собой средний человек. Наоборот, среднее арифметическое же совершенно неподходящая характеристика в нашем случае, поскольку выходит, что каждый, будь то бедняк или миллиардер, имел приблизительно $50 000 004,75.

В случае, когда имеется чётное количество случаев и два средних значения различаются, *медианой*, по определению, может служить любое число между ними (например, в выборке {1, 2, 3, 4} медианой, по определению, может служить любое число из интервала (2,3)). На практике в этом случае чаще всего используют среднее арифметическое двух средних значений.

Числовые характеристики сравнения графика плотности с "нормальным".

Коэффициент асимметрии (скошенности).

$$\boxed{A = \frac{\mu^3}{\sigma^3}}$$

Для нормального распределения, с которым происходит сравнение, этот коэффициент равен нулю: $A_{норм.распр.} = \frac{\mu^3}{\sigma^3} = 0$.

Коэффициенты асимметрии других распределений сравниваются с коэффициентом асимметрии нормального распределения. Получаются либо отрицательными (*negative skew*), либо положительными (*positive skew*):

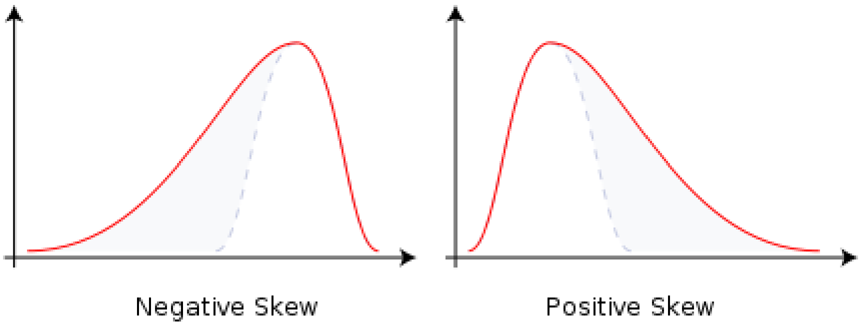


Рисунок. Отрицательная и положительная асимметрия.

Коэффициент эксцесса (островершинности).

$$\boxed{E = \frac{\mu^4}{\sigma^4} - 3}$$

Для нормального распределения значение $\frac{\mu^4}{\sigma^4} = 3$.

<u>Коэффициент эксцесса</u> $\boxed{E = \frac{\mu^4}{\sigma^4} - 3}$ («островершинности») для нормального распределения:

$$E_{норм.распр.} = \frac{\mu^4}{\sigma^4} - 3 = 0.$$

Коэффициенты эксцесса (*kurtosis*) остальных распределений сравниваются с коэффициентом эксцесса нормального распределения:

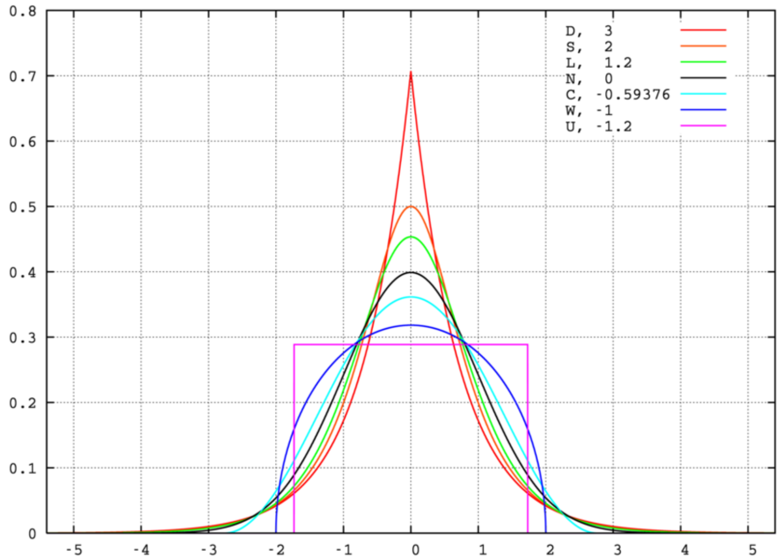


Рисунок. Черным цветом – плотность нормального распределения.
D: Laplace distribution, S: hyperbolic secant distribution, L: logistic distribution,
***N: normal distribution**, C: raised cosine distribution, W: Wigner semicircle distribution,*
U: uniform distribution.

***Замечание**.*

Для расчета коэффициентов асимметрии и эксцесса в Mathcad имеются две встроенные функции:

kurt (x) — коэффициент эксцесса (kurtosis) выборки случайных данных x;

skew(x) — коэффициент асимметрии (skewness) выборки случайных данных X.