Числовые характеристики случайной величины

Лекция 2

Числовые характеристики

1. Характеристики положения случайной величины на числовой оси (мода *Mo*, медиана *Me*, математическое ожидание $M(X)$).

2. Характеристики разброса случайной величины около среднего значения (дисперсия $D(X)$, среднее квадратическое отклонение $\sigma(x)$).

3. Характеристики формы кривой $y = \varphi(x)$ (асимметрия *As*, эксцесс *Ex*).

Математическое ожидание

Математическое ожидание случайной величины *X* указывает некоторое среднее значение, около которого группируются все возможные значения *X*.

Для дискретной случайной величины, которая может принимать лишь конечное число возможных значений, математическим ожиданием называют сумму произведений всех возможных значений случайной величины на вероятность этих значений:

Для непрерывной случайной величины *X*, имеющей заданную плотность распределения φ(*x*) математическим ожиданием называется следующий интеграл:

Свойства математического ожидания

1. $M(C) = C$, где $C = const$;

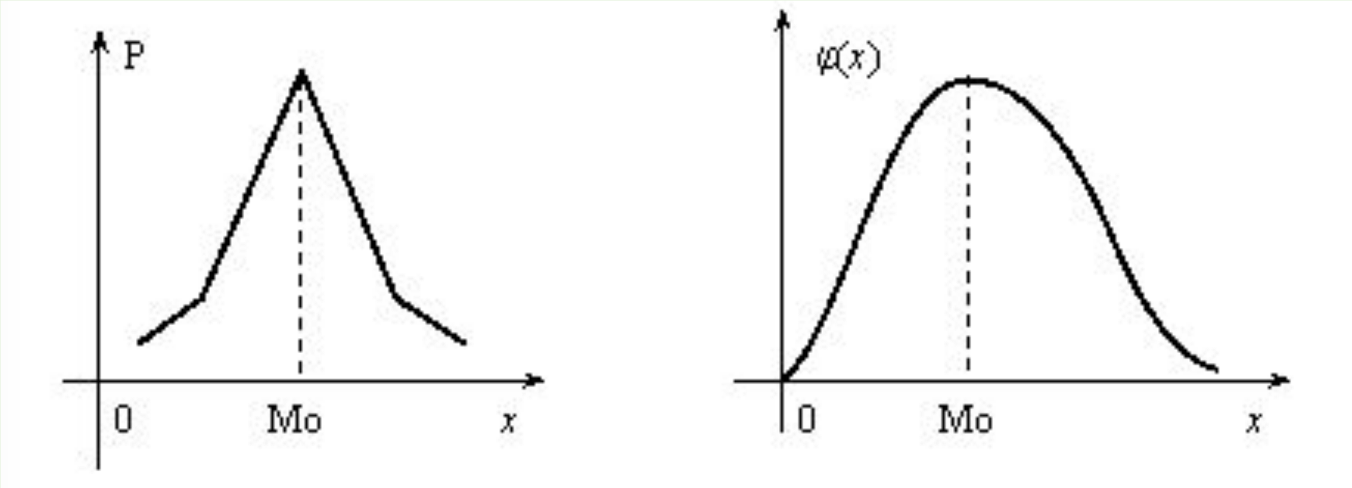
2. $M(C \cdot X) = C \cdot M(X)$;

3. $M(X \pm Y) = M(X) \pm M(Y)$, где X и Y – любые случайные величины;

4. $M(X \cdot Y) = M(X) \cdot M(Y)$, где X и Y – независимые случайные величины.

Мода

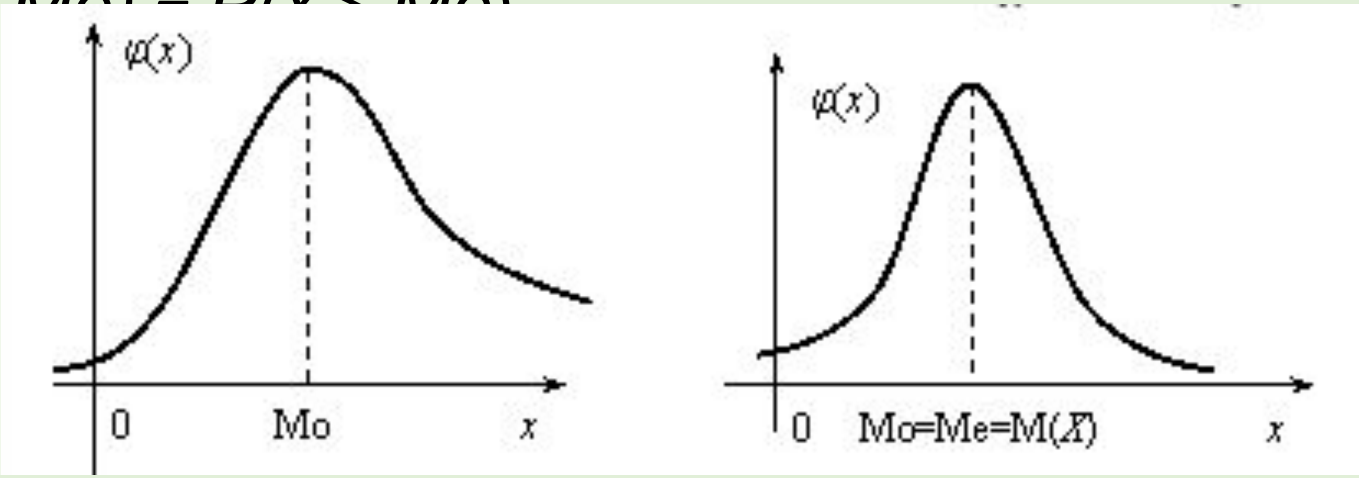
Модой дискретной случайной величины, обозначаемой *Mo*, называется ее наиболее вероятное значение, а модой непрерывной случайной величины – значение, при котором плотность вероятности максимальна.



Медиана

Медианой непрерывной случайной величины *X* называется такое ее значение Me, для которого одинаково вероятно, окажется ли случайная величина меньше или больше *Me*, т. е.

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



Дисперсия

Дисперсией случайной величины называется математическое ожидание квадрата ее отклонения от математического ожидания

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

Дисперсию случайной величины X удобно вычислять по формуле:

а) для дискретной величины

б) для непрерывной случайной величины

Дисперсия обладает следующими свойствами:

1. $D(C) = 0$, где $C = const$;
2. $D(C \times X) = C^2 \cdot D(X)$;
3. $D(X \pm Y) = D(X) + D(Y)$, если X и Y независимые случайные величины.

Среднее квадратическое отклонение

Средним квадратическим отклонением случайной величины *X* называется арифметический корень из дисперсии, т.е.

$\sigma(X) =$ [illegible]

Моменты случайных величин

Начальным моментом k-го порядка α_k случайной величины *X* называется математическое ожидание величины X^k, т.е. $\alpha_k = M(X^k)$.

Начальный момент первого порядка – это математическое ожидание случайной величины.

Центральным моментом k-го порядка μ_k случайной величины *X* называется математическое ожидание величины $(X - M(X))^k$, т.е. $\mu_k = M(X - M(X))^k$.

Центральный момент второго порядка – это дисперсия случайной величины.

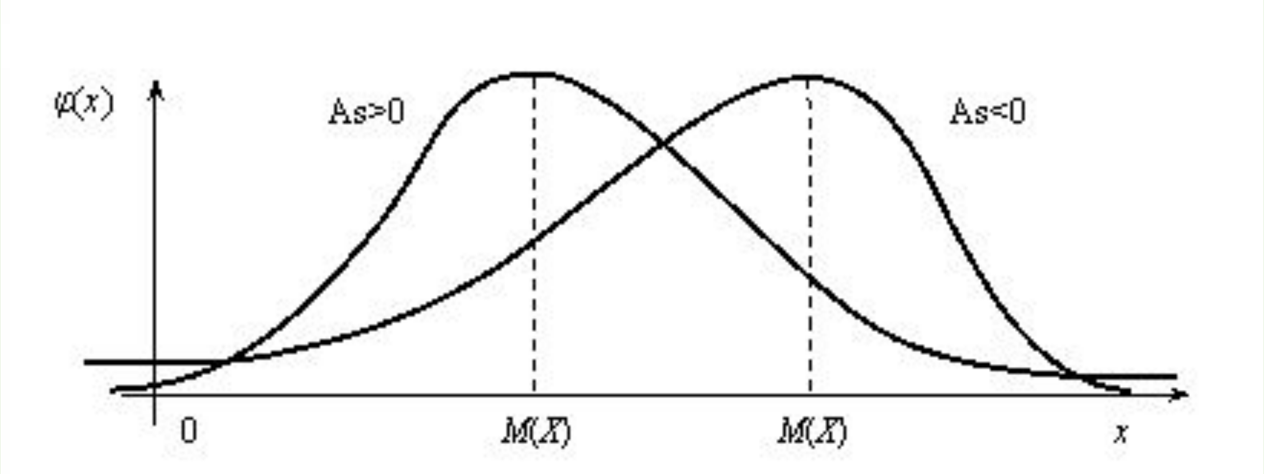
Для дискретной случайной величины начальный момент выражается суммой $\alpha_k =$, а центральный – суммой $\mu_k =$ где $p_i = p(X = x_i)$. Для начального и центрального моментов непрерывной случайной величины можно получить следующие равенства:

$\alpha_k =$, $\mu_k =$,

где φ(*x*) – плотность распределения случайной величины X

Коэффициент асимметрии

Если коэффициент асимметрии отрицательный, то это говорит о большом влиянии на величину m_3 отрицательных отклонений. В этом случае кривая распределения более полога слева от *M*(*X*). Если коэффициент As положительный, то кривая распределения более полога справа.

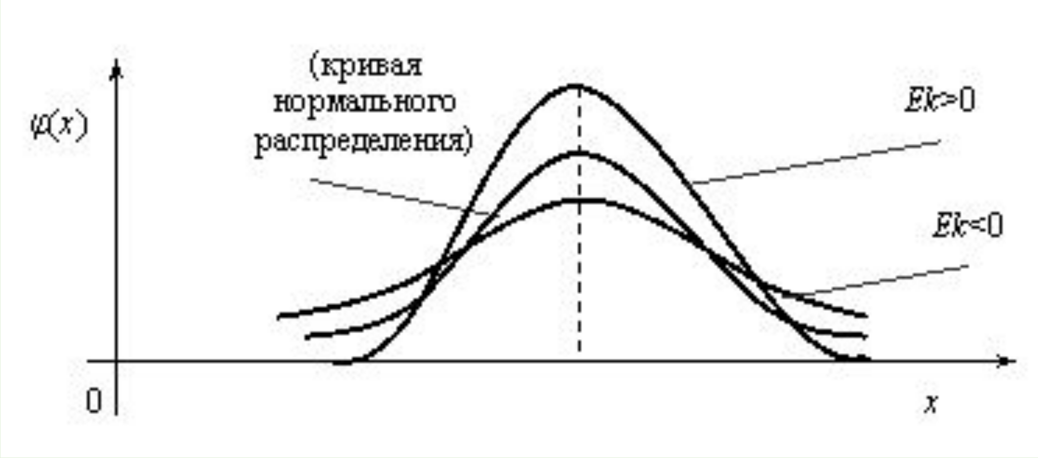


Эксцесс

Эксцессом *Ek* называется величина

$Ek = \mu_4 / \sigma^4 - 3.$

Эксцесс служит для сравнения данного распределения с нормальным, у которого эксцесс равен нулю.



Основные распределения дискретной случайной величины

Биноминальное распределение

Пусть в каждом из n независимых испытаний событие A может произойти с одной и той же вероятностью p(следовательно, вероятность непоявления $q = 1 - p$).

Дискретная случайная величина X – число наступлений события A– имеет распределение, которое называется ***биномиальным***.

X	0	1	…	k	…	n
P	$C_n^0 \cdot p^0 \cdot q^n$	$C_n^1 \cdot p^1 \cdot q^{n-1}$	…	$C_n^k \cdot p^k \cdot q^{n-k}$	…	$C_n^n \cdot p^n \cdot q^0$

Распределение Пуассона

Это распределение представляет собой предельный случай биномиального, когда вероятность *p* очень мала, а число испытаний *n* велико.

Дискретная случайная величина *X*, которая может принимать только целые неотрицательные значения с вероятностями

Распределение Пуассона

Закон Пуассона описывает число событий *k*, происходящих за одинаковые промежутки времени. При этом полагается, что события появляются независимо друг от друга с постоянной средней интенсивностью, которая характеризуется параметром $\lambda = n \cdot p$

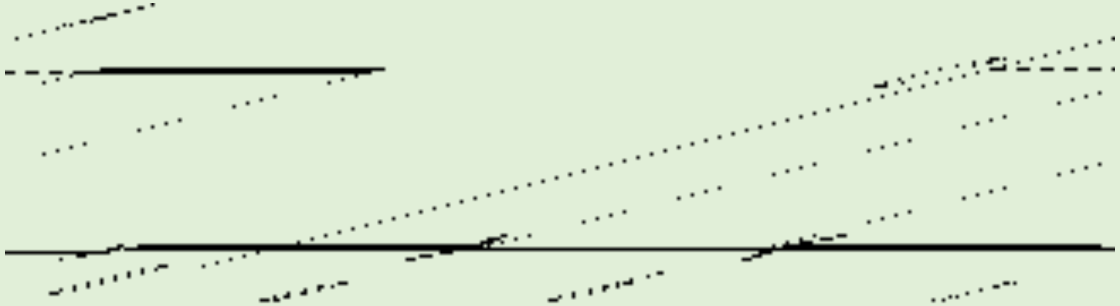
По распределению Пуассона распределено, например число посетителей магазина или банка за определенный промежуток времени, при этом λ – среднее число посетителей за это время.

Предположим, что в среднем в магазин приходит 2,1 покупатель в минуту. Тогда,

Равномерное распределение

Непрерывная случайная величина *X* имеет равномерное распределение на отрезке [*a*, *b*], если ее плотность имеет следующий вид:

График плотности распределения

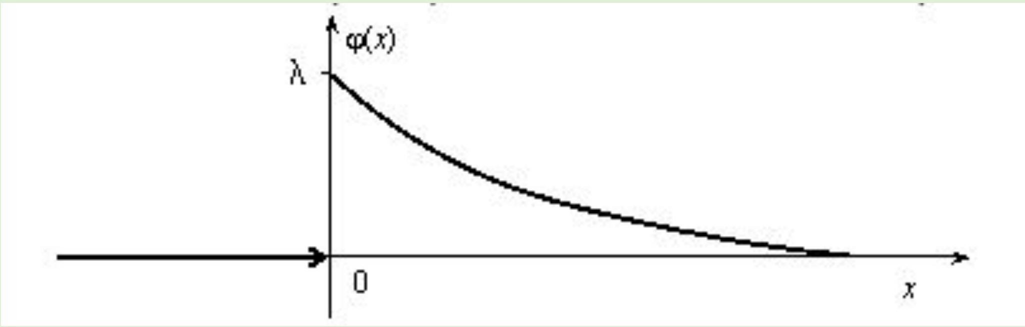


Показательное распределение

Непрерывная случайная величина *X*, функция плотности которой задается выражением

[illegible]

называется случайной величиной, имеющей **показательное**, или экспоненциальное, распределение. Здесь параметр λ постоянная положительная величина.

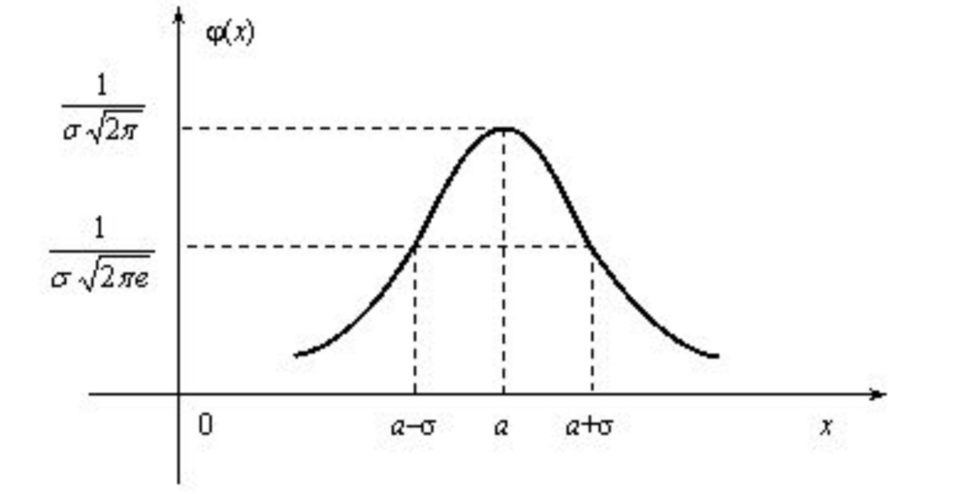


Нормальное распределение

Случайная величина *X* имеет нормальное распределение (или распределение по закону Гаусса), если ее плотность вероятности имеет вид:

где параметры *a* – любое действительное число и σ >0.

График дифференциальной функции нормального распределения называют нормальной кривой (кривой Гаусса).



Примеры решения задач

Решение задачи на классическую вероятность

Абонент забыл последние 2 цифры телефонного номера, но помнит, что они различны и образуют двузначное число, меньшее 30. С учетом этого он набирает наугад 2 цифры. Найти вероятность того, что это будут нужные цифры.

Решение задачи на классическую вероятность

Абонент забыл последние 2 цифры телефонного номера, но помнит, что они различны и образуют двузначное число, меньшее 30. С учетом этого он набирает наугад 2 цифры. Найти вероятность того, что это будут нужные цифры.

Решение: подсчитаем количество всех возможных двузначных чисел с разными цифрами, меньшее 30, которые может набрать абонент:

10	12	13	14	15	16	17	18	19
20	21	23	24	25	26	27	28	29

Ответ: 1/18.

Схема Бернулли

Из n аккумуляторов за год хранения k выходит из строя. Наудачу выбирают m аккумуляторов. Определить вероятность того, что среди них l исправных.

n=100, k=7,m=5, l=3

Схема Бернулли

Из *n* аккумуляторов за год хранения *k* выходит из строя. Наудачу выбирают *m* аккумуляторов. Определить вероятность того, что среди них *l* исправных.

n=100, k=7, m=5, l=3

Решение: Имеем схему Бернулли с параметрами p=7/100=0,07 (вероятность того, что аккумулятор выйдет из строя), n=5 (число испытаний), k=5−3

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

неисправных аккумуляторов).

Получаем

$$P_5(2) = C_5^2 \cdot 0,07^2 \cdot (1-0,07)^{5-2} = \frac{5!}{3!2!} \cdot 0,07^2 \cdot 0,93^3 = 0,0394.$$

Ответ: 0,0394.

Теоремы сложения и умножения вероятностей

Задача: трое учащихся на экзамене независимо друг от друга решают одну и ту же задачу. Вероятности ее решения этими учащимися равны 0,8, 0,7 и 0,6 соответственно. Найдите вероятность того, что хотя бы один учащийся решит задачу.

Теоремы сложения и умножения вероятностей

РЕШЕНИЕ. Введем событие X = (Хотя бы один учащийся решит задачу) и противоположное ему $\overline{X}$ = (Ни один учащийся не решит задачу).

Введем вспомогательные события:
A_1 = (Первый учащийся решил задачу),
A_2 = (Второй учащийся решил задачу),
A_3 = (Третий учащийся решил задачу),
вероятности $P(A_1)=0,8$, $P(A_2)=0,7$, $P(A_2)=0,6$.

Выразим событие $\overline{X}=\overline{A_1}\cdot\overline{A_2}\cdot\overline{A_3}$. Считаем вероятность как вероятность произведения независимых событий:

$$P\left(\overline{X}\right)=P\left(\overline{A_1}\cdot\overline{A_2}\cdot\overline{A_3}\right)=P\left(\overline{A_1}\right)\cdot P\left(\overline{A_2}\right)\cdot P\left(\overline{A_3}\right)=(1-0,8)(1-0,7)(1-0,6)=$$
$$=0,2\cdot 0,3\cdot 0,4=0,024.$$

Тогда вероятность искомого события $P(X)=1-P\left(\overline{X}\right)=1-0,024=0,976$.

ОТВЕТ. 0,976.

Формула полной вероятности

Задача. Из 1000 ламп 380 принадлежат к 1 партии, 270 – ко второй партии, остальные к третьей. В первой партии 4% брака, во второй - 3%, в третьей – 6%. Наудачу выбирается одна лампа. Определить вероятность того, что выбранная лампа – бракованная.

Формула полной вероятности

Решение. Введем полную группу независимых гипотез:
Hi = (Лампа принадлежат i -ой партии), $i = 1, 2, 3$.
Найдем вероятности гипотез по классическому определению вероятностей. Всего ламп 1000, из них 1-ой партии принадлежат 380, то есть $P(H1) = \frac{380}{1000} = 0,38$, 2-ой партии принадлежат 270, то есть $P(H2) = \frac{270}{1000} = 0,27$, остальные 1000-380-270=350 ламп принадлежат 3-ей партии, поэтому $P(H3) = \frac{350}{1000} = 0,35$.

Введем событие A = (Лампа бракованная). По условию даны априорные вероятности:
$P(A \mid H1) = 0,04$, $P(A \mid H2) = 0,03$, $P(A \mid H3) = 0,06$.

Вероятность события A найдем по формуле полной вероятности:
$$P(A) = P(A \mid H1)P(H1) + P(A \mid H2)P(H2) + P(A \mid H3)P(H3) =$$
$$= 0,38 \cdot 0,04 + 0,27 \cdot 0,03 + 0,35 \cdot 0,06 = 0,0443.$$

Ответ. 0,0443 (или 4,43%).

Формула Байеса

Задача. Из 30 стрелков 12 попадает в цель с вероятностью 0.6, 8 - с вероятностью 0.5 и 10 – с вероятностью 0.7. Наудачу выбранный стрелок произвел выстрел, поразив цель. К какой из групп вероятнее всего принадлежал этот стрелок?

Формула Байеса

Решение. Введем полную группу гипотез:
$H1$ = (Стрелок принадлежал первой группе),
$H2$ = (Стрелок принадлежал второй группе),
$H3$ = (Стрелок принадлежал третьей группе).

По классическому определению вероятности:

$$P(H1) = \frac{12}{30} = \frac{2}{5},\ P(H2) = \frac{8}{30} = \frac{4}{15},\ P(H3) = \frac{10}{30} = \frac{1}{3}.$$

Введем событие A = (Стрелок попал в мишень). Выпишем условные вероятности:

$$P(A\,|\,H1) = 0,6,\ P(A\,|\,H2) = 0,5,\ P(A\,|\,H3) = 0,7.$$

Найдем сначала вероятность события A по формуле полной вероятности:

$$P(A) = P(A\,|\,H1)P(H1) + P(A\,|\,H2)P(H2) + P(A\,|\,H3)P(H3) =$$

$$= \frac{2}{5}\cdot 0,6 + \frac{4}{15}\cdot 0,5 + \frac{1}{3}\cdot 0,7 \approx 0,607.$$

Формула Байеса

Теперь найдем апостериорные вероятности того, что стрелок принадлежал i-ой группе, если он попал в цель, по формуле Байеса.

$$P(H1 \mid A) = \frac{P(H1)P(A \mid H1)}{P(A)} = \frac{\frac{2}{5} \cdot 0,6}{0,607} \approx 0,395,$$

$$P(H2 \mid A) = \frac{P(H2)P(A \mid H2)}{P(A)} = \frac{\frac{4}{15} \cdot 0,5}{0,607} \approx 0,22,$$

$$P(H3 \mid A) = \frac{P(H3)P(A \mid H3)}{P(A)} = \frac{\frac{1}{3} \cdot 0,7}{0,607} \approx 0,384.$$

Таким образом, вероятнее всего стрелок принадлежал первой группе.

Биноминальный закон распределения

Задача. В городе 4 коммерческих банка. У каждого риск банкротства в течение года составляет 20%. Составьте ряд распределения числа банков, которые могут обанкротиться в течение следующего года.

Биноминальный закон распределения

Решение.

Пусть X – дискретная случайная величина, равная числу банков, которые могут обанкротиться в течение следующего года. Она может принимать значения 0, 1, 2, 3 и 4. X распределена по биномиальному закону с параметрами $n=4$, $p=20\%=0,2$, поэтому найдем соответствующие вероятности по формуле Бернулли: $P(X=k)=P_n(k)=C_n^k p^k (1-p)^{n-k}$. Получаем:

$$P(X=0)=C_4^0\cdot 0,2^0\cdot 0,8^4=0,8^4=0,4096$$

$$P(X=1)=C_4^1\cdot 0,2^1\cdot 0,8^3=4\cdot 0,2\cdot 0,8^3=0,4096$$

$$P(X=2)=C_4^2\cdot 0,2^2\cdot 0,8^2=6\cdot 0,2^2\cdot 0,8^2=0,1536$$

$$P(X=3)=C_4^3\cdot 0,2^3\cdot 0,8^1=4\cdot 0,2^3\cdot 0,8=0,0256$$

$$P(X=4)=C_4^4\cdot 0,2^4\cdot 0,8^0=0,2^4=0,0016$$

Таким образом, закон распределения случайной величины X имеет вид:

x_i	0	1	2	3	4
p_i	0,4096	0,4096	0,1536	0,0256	0,0016

Расчеты произведены правильно, так как сумма $\sum p_i = 1$.

Закон распределения Пуассона

Задача. Среднее число самолетов, взлетающих с полевого аэродрома за одни сутки, равно 10. Найти вероятность того, что за 6 часов взлетят:
А) три самолета,
Б) не менее двух самолетов

Закон распределения Пуассона

Решение. Будем использовать формулу Пуассона:

$P_t(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} = \frac{(10t)^k}{k!} e^{-10t}$ - вероятность того, что за время t суток с полевого аэродрома взлетят ровно k самолетов. Здесь $\lambda = 10$ - интенсивность потока взлетов (10 взлетов в сутки).

Так как 6 часов равны $t = \frac{1}{4}$ суток, получаем вероятности.

А) вероятность того, что за 6 часов взлетят три самолета $P_{1/4}(3) = \frac{(10/4)^3}{3!} e^{-(10/4)} \approx 0,214$.

Б) вероятность того, что за 6 часов взлетят не менее двух самолетов:

$$P_{1/4}(k \geq 2) = 1 - P_{1/4}(k < 2) = 1 - P_{1/4}(0) - P_{1/4}(1) =$$

$$= 1 - \frac{(10/4)^0}{0!} e^{-(10/4)} - \frac{(10/4)^1}{1!} e^{-(10/4)} = 1 - \left(1 + \frac{10}{4}\right) e^{-5/2} \approx 0,713.$$

Ответ: 0,214; 0,713.

Примеры решения задач по математической статистике

Простой вариационный ряд

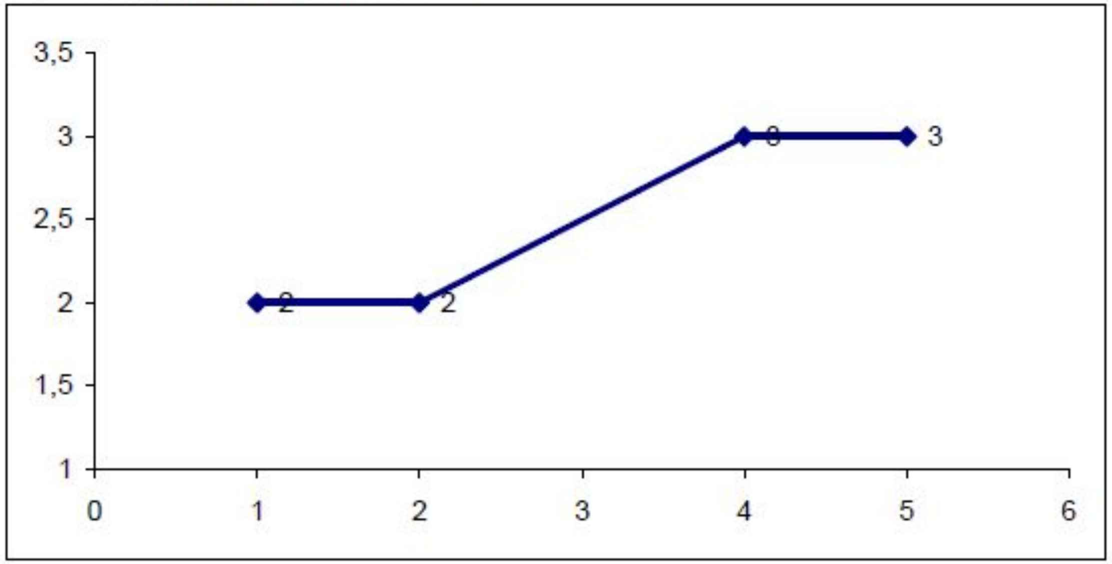
Задача 1. Дан следующий вариационный ряд
1 2 3 4 5 6 7 8 9 10
1 1 2 2 4 4 4 5 5 5
Требуется
1) Построить полигон распределения
2) Вычислить выборочную среднюю, дисперсию, моду, медиану.
3) Построить выборочную функцию распределения
4) Найти несмещенные оценки математического ожидания и дисперсии.

Простой вариационный ряд

1) Полигон распределения – это зависимость абсолютной частоты варианта m_i от значения варианта x_i. Эту зависимость можно представить в виде таблицы:

x_i	1	2	4	5
m_i	2	2	3	3

Строим график полигона частот:



Простой вариационный ряд

2) Вычислим выборочную среднюю, дисперсию, моду, медиану.

Выборочная средняя:

$$\overline{x} = \frac{1}{n}\sum x_i m_i = \frac{1}{10}(1\cdot 2 + 2\cdot 2 + 4\cdot 3 + 5\cdot 3) = \frac{33}{10} = 3,3.$$

Выборочная дисперсия

$$D_x = \overline{x^2} - \left(\overline{x}\right)^2 = \frac{1}{n}\sum x_i^2 m_i - 3,3^2 = \frac{1}{10}(1\cdot 2 + 4\cdot 2 + 16\cdot 3 + 25\cdot 3) - 3,3^2 = 2,41.$$

Выборочное среднеквадратичное отклонение:

$$\sigma_x = \sqrt{D_x} = \sqrt{2,41} \approx 1,552.$$

Мода равна варианту, имеющему наибольшую частоту: $x_{Mo} = 4; 5$ (две моды)

Медиана равна среднему варианту выборки: $x_{Me} = 4$.

Простой вариационный ряд

3) Выборочная функция распределения аналогична функции распределения дискретной случайной величины. Для ее нахождения запишем ряд распределения выборки, где

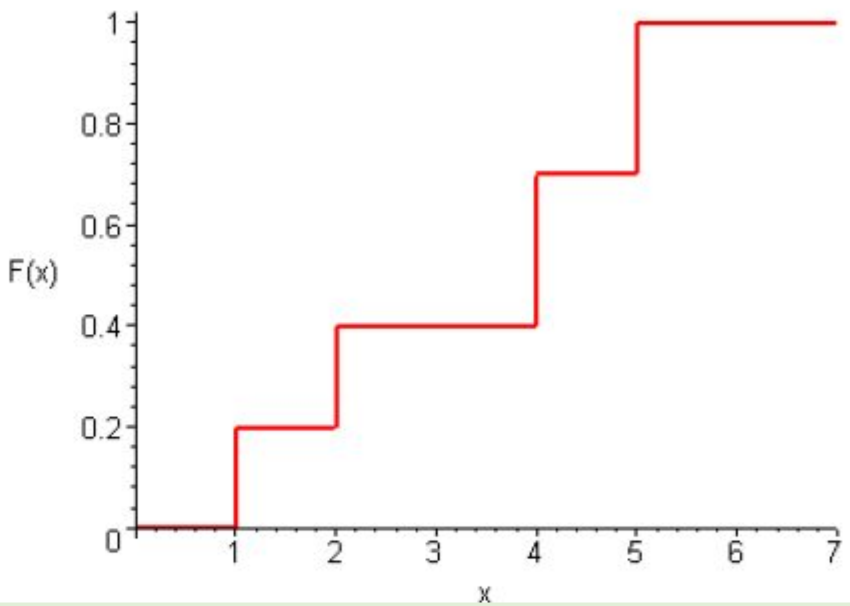
$p_i = \frac{m_i}{n} = \frac{m_i}{10}$ – относительная частота варианта x_i.

x_i	1	2	4	5
p_i	0,2	0,2	0,3	0,3

Тогда

$$F(x) = \begin{cases} 0, \ x \le 1, \\ 0,2, \ 1 < x \le 2, \\ 0,4, \ 2 < x \le 4, \\ 0,7, \ 4 < x \le 5, \\ 1, \ x > 5. \end{cases}$$

Построим график:



Простой вариационный ряд

4) Несмещенная оценка математического ожидания совпадает с выборочной средней: $M(X) = \overline{x} = 3,3$.

Несмещенная оценка дисперсии отличается от выборочной дисперсии в большую сторону: $D(X) = \frac{n}{n-1} D_x = \frac{10}{9} 2,41 \approx 2,678$.

Интервальный ряд

Задача. Проведено выборочное обследование магазинов города. Имеются следующие данные о величине товарооборота для 50 магазинов города (x_i – товарооборот, млн. руб.; n_i – число магазинов).

x_i	25-75	75-125	125-175	175-225	225-275	275-325
n_i	12	15	9	7	4	3

Найти

а) среднее, среднее квадратическое отклонение S и коэффициент V;

б) построить гистограмму и полигон частот.

Интервальный ряд

РЕШЕНИЕ. Перейдем к простому вариационному ряду, выбирая в качестве значений середины интервалов. Получим:

x_i	50	100	150	200	250	300
n_i	12	15	9	7	4	3

Найдем необходимые числовые характеристики.

Выборочная средняя:

$$\overline{X} = \frac{1}{n}\sum x_i n_i = \frac{1}{50}6750 = 135\,.$$

Выборочная дисперсия

$$S^2 = \frac{1}{n}\sum\left(x_i - \overline{X}\right)^2 n_i = \frac{1}{50}271250 = 5425\,.$$

Выборочное среднеквадратичное отклонение:

$$S = \sqrt{S^2} \approx 73,655.$$

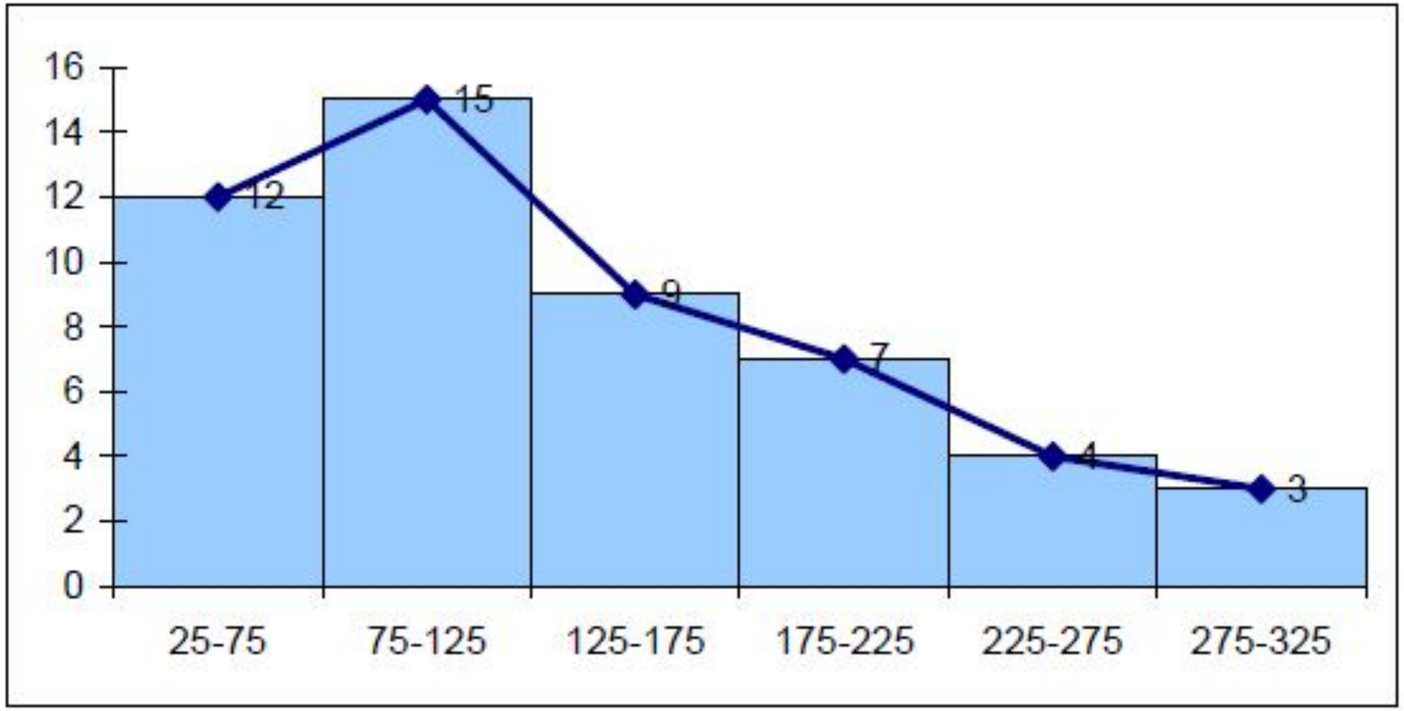
Коэффициент вариации: $V = \frac{S}{\overline{X}} = \frac{73,655}{135}\cdot 100\% \approx 54,56\%$.

Расчеты в таблице

x_i	50	100	150	200	250	300	**Сумма**
n_i	12	15	9	7	4	3	**50**
$x_i n_i$	600	1500	1350	1400	1000	900	**6750**
$\left(x_i - \overline{X}\right)^2 n_i$	86700	18375	2025	29575	52900	81675	**271250**

Интервальный ряд

Строим гистограмму и полигон частот:



Задачи на построение доверительных интервалов

Строительная компания хочет оценить среднюю стоимость ремонтных работ, выполняемых для клиентов. Каким должен быть объем выборки среди 1200 клиентов строительной фирмы, если среднее квадратическое отклонение по результатам пробного обследования составило 850 у.е., а предельная ошибка выборки не должна превышать 200 у.е. с вероятностью 0,95?

Задачи на построение доверительных интервалов

РЕШЕНИЕ. Найдем объём бесповторной выборки, при предельная ошибка выборки не должна превышать 200 у.е. с вероятностью 0,95.

Предельная ошибка по условию $\Delta_x = t \cdot \mu_x = 200$ у.е., $N = 1200$ клиентов, $S = 850$ у.е., а $t = \Phi^{-1}(\gamma/2) = \Phi^{-1}(0,95/2) = \Phi^{-1}(0,475) = 1,96$. Объем выборки тогда можно вычислить по формуле:

$$n = \frac{t^2 N S^2}{\Delta_x^2 N + t^2 S^2} = \frac{1,96^2 \cdot 1200 \cdot 850^2}{200^2 \cdot 1200 + 1,96^2 \cdot 850^2} \approx 66.$$

ОТВЕТ. 66 клиентов.

Задачи на построение доверительных интервалов

С целью размещения рекламы опрошено 420 телезрителей, из которых данную передачу смотрят 170 человек. С доверительной вероятностью 0,91 найти долю телезрителей, охваченных рекламой в лучшем случае

Задачи на построение доверительных интервалов

С целью размещения рекламы опрошено 420 телезрителей, из которых данную передачу смотрят 170 человек. С доверительной вероятностью 0,91 найти долю телезрителей, охваченных рекламой в лучшем случае

РЕШЕНИЕ. Для оценки неизвестной доли телезрителей используем формулу:

$$w-t\sqrt{\frac{w(1-w)}{n}}<p<w+t\sqrt{\frac{w(1-w)}{n}},$$

где

$n=420$ - объем выборки,

$w=\frac{170}{420}\approx 0,4048$ - относительная частота,

$t=\Phi^{-1}(\gamma/2)=\Phi^{-1}(0,91/2)=\Phi^{-1}(0,5=455)=1,69$ (из таблицы функции Лапласа).

Подставим данные:

$$0,4048-1,69\sqrt{\frac{0,4048(1-0,4048)}{420}}<p<0,4048+1,69\sqrt{\frac{0,4048(1-0,4048)}{420}},$$

$0,364<p<0,445$.

Лучший вариант охвата телезрителей – 44,5% телезрителей.

Ответ. 44,5%.

Основные формулы

Основные формулы комбинаторики

Число перестановок

$$P_n = n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n$$

Число размещений

$$A_m^n = n \cdot (n-1) \cdot \ldots \cdot (n-m+1)$$

Число сочетаний

$$C_n^m = \frac{A_n^m}{P_m} = \frac{n!}{m! \cdot (n-m)!}$$

Число перестановок с повторениями

$$P_n(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}.$$

Число размещений с повторениями

$$\overline{A}_n^k = n \cdot n \cdot \ldots \cdot n = n^k.$$

Число сочетаний с повторениями

$$\overline{C}_n^k = C_{k+n-1}^k = \frac{(k+n-1)!}{(n-1)! \cdot k!}$$

Классическое определение вероятности

$$P(A) = \frac{m}{n},$$

где m - число благоприятствующих событию A исходов, n - число всех элементарных равновозможных исходов в испытании.

Вероятность суммы событий

Теорема сложения вероятностей несовместных событий:

$$P(A+B) = P(A) + P(B)$$

Теорема сложения вероятностей совместных событий:

$$P(A+B) = P(A) + P(B) - P(AB)$$

Вероятность произведения событий

Теорема умножения вероятностей независимых событий:

$$P(A \cdot B) = P(A) \cdot P(B)$$

Теорема умножения вероятностей зависимых событий:

$$P(A \cdot B) = P(A) \cdot P(B|A),$$
$$P(A \cdot B) = P(B) \cdot P(A|B).$$

$P(A|B)$ - условная вероятность события A при условии, что произошло событие B,

$P(B|A)$ - условная вероятность события B при условии, что произошло событие A.

Формула полной вероятности

$$P(A) = \sum_{k=1}^{n} P(H_k) \cdot P(A|H_k),$$

где $H_1, H_2, \ldots, H_n$ - полная группа гипотез.

Формула Бернулли

$$P_n(k) = C_n^k \cdot p^k \cdot (1-p)^{n-k} = \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k}$$

вероятность появления события ровно k раз в n независимых испытаниях, p - вероятность появления события при одном испытании.

Наивероятнейшее число наступления события

Наивероятнейшее число k_0 появления события при n независимых испытаниях (где p - вероятность появления события при одном испытании):

$$np-(1-p) \leq k_0 \leq np+p.$$

Приближенная формула Пуассона

Если число испытаний n велико, и при этом вероятность p наступления события в каждом испытании крайне мала, так что выполняется условие $np<10$, можно применять формулу Пуассона:

$$P_n(k)=\frac{\lambda^k}{k!}\cdot e^{-\lambda}.$$

Здесь $\lambda=n\cdot p$ обозначает среднее число появлений события.

Локальная формула Лапласа

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

вероятность появления события ровно k раз при n независимых испытаниях, p - вероятность появления события при одном испытании, $q=1-p$.
Значения функции $\varphi(x)$ берутся из таблицы.

Интегральная формула Лапласа

$$P_n(m_1,m_2)=\Phi\left(\frac{m_2-np}{\sqrt{npq}}\right)-\Phi\left(\frac{m_1-np}{\sqrt{npq}}\right)$$

вероятность появления события не менее m_1 и не более m_2 раз при n независимых испытаниях, p - вероятность появления события при одном испытании, $q=1-p$.
Значения функции $\Phi(x)$ берутся из таблицы.

Оценка отклонения относительной частоты от постоянной вероятности

$$P\left(\left|\frac{m}{n}-p\right|\leq\varepsilon\right)=2\Phi\left(\varepsilon\cdot\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

ε - величина отклонения, p - вероятность появления события.

Ряд распределения дискретной случайной величины

Табличный вид:

X_i	x_1	x_2	$\dots$	x_n
p_i	p_1	p_2	$\dots$	p_n

Сумма вероятностей всегда равна 1 (условие нормировки):

$$\sum_{i=1}^{n} p_i = 1$$

Функция распределения (интегральная функция распределения)

Функция распределения случайной величины X определяется по формуле $F(x) = P(X < x)$. Это неубывающая функция, принимающая значения от 0 до 1. Если задана плотность распределения $f(x)$, то функция распределения выражается как интеграл от плотности:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

Плотность распределения (дифференциальная функция распределения)

Плотность распределения случайной величины X определяется по формуле $f(x) = F'(x)$. Существует только для непрерывной случайной величины. Для нее выполняется условие нормировки (площадь под кривой вероятности равна 1):

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1.$$

Вероятность попадания случайной величины в заданный интервал

Может быть вычислена двумя способами:

1) через функцию распределения

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha).$$

2) через плотность распределения

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)\,dx.$$

Математическое ожидание случайной величины

1) Для дискретной случайной величины X, заданной рядом распределения:

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

2) Для непрерывной случайной величины X, заданной плотностью распределения:

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

Дисперсия случайной величины

По определению дисперсия – это второй центральный момент:

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

1) Для дискретной случайной величины X:

$$D(X)=\sum_{i=1}^{n} x_i^2\cdot p_i-(M(X))^2.$$

2) Для непрерывной случайной величины X:

$$M(X)=\int_{-\infty}^{+\infty} f(x)\cdot x^2\,dx-(M(X))^2.$$

Среднее квадратическое отклонение случайной величины

$$\sigma(X)=\sqrt{D(X)}.$$

Коэффициент вариации случайной величины

$$V(X)=\frac{\sigma(X)}{M(X)}.$$

Начальный момент r–го порядка случайной величины

определяется по формуле:

$$\nu_r=M(X^r)$$

В частности, первый начальный момент – это математическое ожидание: $\nu_1=M(X^1)=M(X)$.

Центральный момент r – го порядка случайной величины

определяется по формуле:

$$\mu_r=M\left[(X-M(X))^r\right]$$

В частности, второй центральный момент – это дисперсия:

$$\mu_2=M\left[(X-M(X))^2\right]=D(X).$$

Асимметрия

$$A_s=\frac{\mu_3}{\sigma^3}.$$

Коэффициент асимметрии положителен, если правый хвост распределения длиннее левого (правая часть кривой более пологая), и отрицателен в противном случае. Если распределение симметрично относительно математического ожидания, то его коэффициент асимметрии равен нулю.

Эксцесс

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

Коэффициент эксцесса нормального распределения равен нулю. Он положителен, если пик распределения около математического ожидания острый, и отрицателен, если пик гладкий.

Биномиальное распределение ДСВ

Пусть дискретная случайная величина X - количество "успехов" в последовательности из n независимых случайных экспериментов, таких что вероятность "успеха" в каждом из них равна p ("неуспеха" - $q = 1 - p$).

Закон распределения X имеет вид:

x_k	0	1	...	k	...	n
p_k	q^n	$n \cdot p \cdot q^{n-1}$		$C_n^k \cdot p^k \cdot q^{n-k}$		p^n

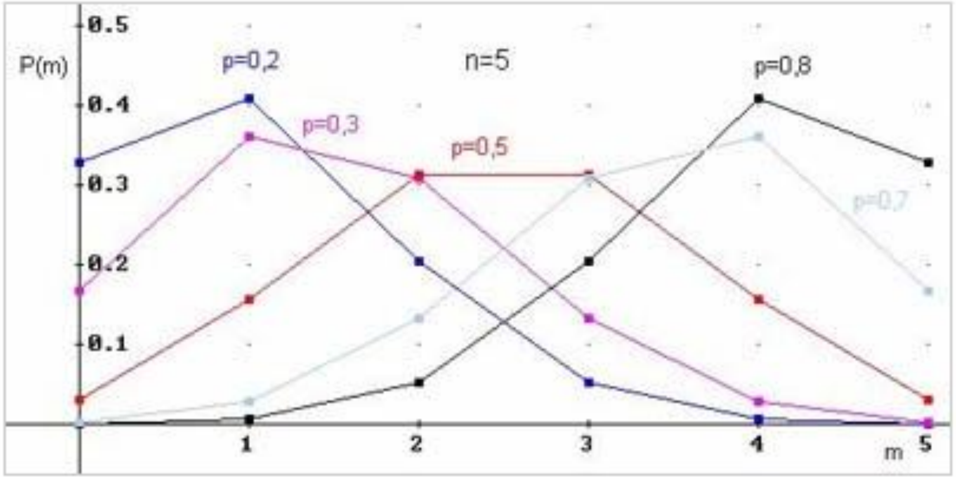
Здесь вероятности находятся по формуле Бернулли:

$$P(X = k) = C_n^k \cdot p^k \cdot (1-p)^{n-k} = C_n^k \cdot p^k \cdot q^{n-k}, k = 0, 1, 2, \dots, n.$$

Числовые характеристики биномиального распределения:

$$M(X) = np, \quad D(X) = npq, \sigma(X) = \sqrt{npq}.$$

Примеры многоугольников распределения для $n = 5$ и различных вероятностей:



Пуассоновское распределение ДСВ

Распределение Пуассона моделирует случайную величину, представляющую собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью и независимо друг от друга.

При условии $p \to 0$, $n \to \infty$, $np \to \lambda = const$ закон распределения Пуассона является предельным случаем биномиального закона. Так как при этом вероятность p события A в каждом испытании мала, то закон распределения Пуассона называют часто законом редких явлений.

Ряд распределения по закону Пуассона имеет вид:

x_k	0	1	...	k	...
p_k	$e^{-\lambda}$	$\lambda e^{-\lambda}$	...	$\frac{\lambda^k}{k!} \cdot e^{-\lambda}$	...

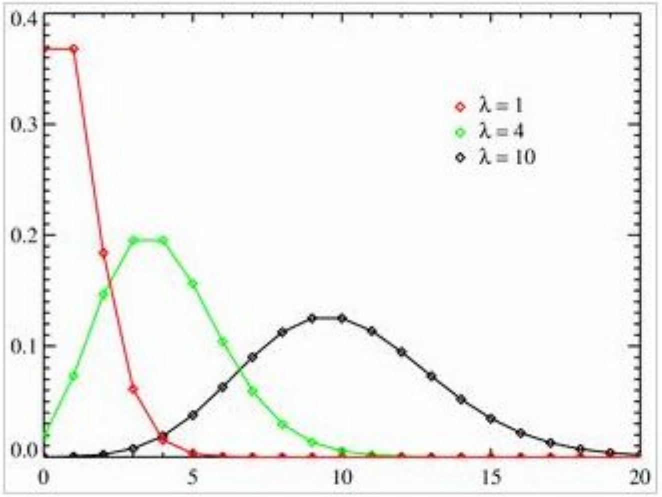
Вероятности вычисляются по формуле Пуассона:

$$P(X=k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, k = 0, 1, 2, \dots$$

Числовые характеристики для распределения Пуассона:

$$M(X) = \lambda, \quad D(X) = \lambda, \sigma(X) = \sqrt{\lambda}.$$

Разные многоугольники распределения при $\lambda = 1; 4; 10$.



Геометрическое распределение ДСВ

Пусть происходит серия независимых испытаний, в каждом из которых событие может появится с одной и той же вероятностью p. Тогда случайная величина X - количество испытаний до первого появления события, имеет геометрическое распределение вероятностей.

Формула для вероятностей:

$$P(X=k)=q^k\cdot p, k=0,1,2,\ldots,n,\ldots$$

Ряд распределения геометрического закона:

x_k	0	1	2	...	k	...
p_k	p	$q\cdot p$	$q^2\cdot p$	...	$q^k\cdot p$	...

Числовые характеристики:

$$M(X)=\frac{q}{p},\quad D(X)=\frac{q}{p^2}.$$

Гипергеометрическое распределение ДСВ

Из урны, в которой находятся N шаров (K белых и $N-K$ чёрных шаров), наудачу и без возвращения вынимают n шаров ($n\le N$). Найти закон распределения случайной величины X - равной числу белых шаров среди выбранных.

Случайная величина X может принимать целые значения от 0 до K (если $n<K$, то до n). Вероятности вычисляются по формуле:

$$P(X=k)=\frac{C_K^k\cdot C_{N-K}^{n-k}}{C_N^n},\quad 0\le k\le K.$$

Числовые характеристики:

$$M(X)=\frac{K}{N}\cdot n,\quad D(X)=\frac{K}{N}\cdot n\cdot\frac{N-n}{N}\cdot\frac{N-K}{N-1}.$$

Показательное распределение НСВ

Экспоненциальное или показательное распределение — абсолютно непрерывное распределение, моделирующее время между двумя последовательными свершениями одного и того же события.

Плотность распределения величины X(везде $\lambda > 0$):

$$f(x) = \begin{cases} 0, \ x < 0 \\ \lambda e^{-\lambda x}, \ x \geq 0 \end{cases}$$

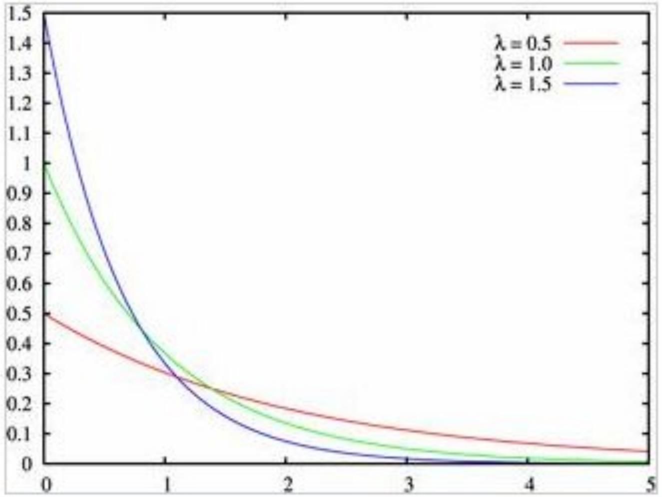
Функция распределения величины X:

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

Числовые характеристики можно найти по формулам:

$$M(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}, \quad \sigma = \frac{1}{\lambda}.$$

Плотность распределения при различных значениях $\lambda > 0$:



Равномерное распределение НСВ

Равномерный закон распределения используется при анализе ошибок округления при проведении числовых расчётов (например, ошибка округления числа до целого распределена равномерно на отрезке), в ряде задач массового обслуживания, при статистическом моделировании наблюдений, подчинённых заданному распределению.

Плотность распределения на отрезке $(a;b)$:

$$f(x)=\begin{cases}0,\ x\le a\\ \frac{1}{b-a},\ a<x\le b,\\ 0,\ x>b,\end{cases}$$

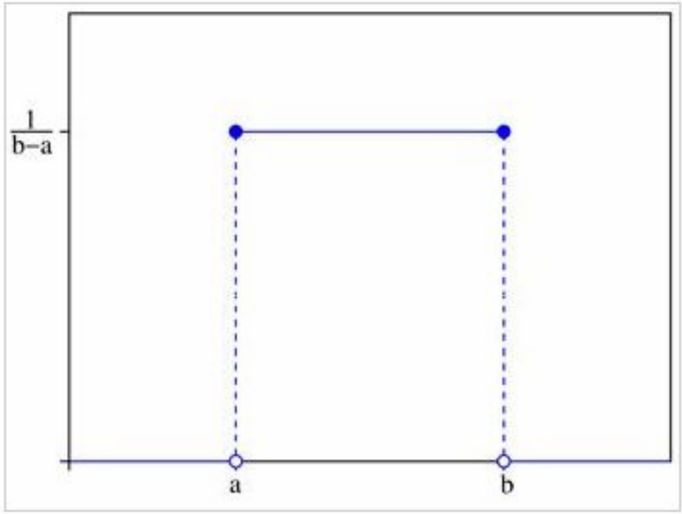
Функция распределения:

$$F(x)=\begin{cases}0,\ x\le a\\ \frac{x-a}{b-a},\ a<x\le b,\\ 1,\ x>b,\end{cases}$$

Числовые характеристики равномерно распределенной случайной величины:

$$M(X)=\frac{a+b}{2},\quad D(X)=\frac{(b-a)^2}{12},\quad \sigma=\frac{b-a}{2\sqrt{3}}.$$

График плотности вероятностей:



Нормальное распределение или распределение Гаусса НСВ

Нормальное распределение, также называемое распределением Гаусса, – распределение вероятностей, которое играет важнейшую роль во многих областях знаний, особенно в физике.

Физическая величина подчиняется нормальному распределению, когда она подвержена влиянию огромного числа случайных помех. Ясно, что такая ситуация крайне распространена, поэтому можно сказать, что из всех распределений в природе чаще всего встречается именно нормальное распределение — отсюда и произошло одно из его названий.

Плотность распределения нормальной случайной величины X имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

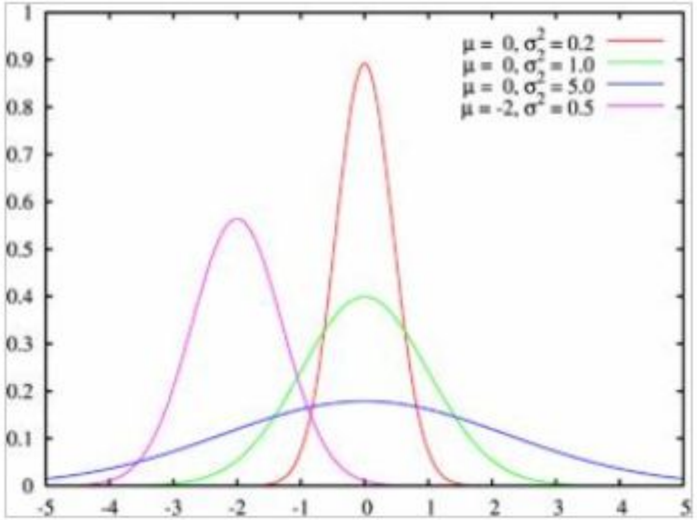
При $a = 0$ и $\sigma = 1$ эта функция принимает вид:

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

Числовые характеристики для нормального распределения:

$$M(X) = a, \quad D(X) = \sigma^2.$$

Пример графика плотности распределения для различных значений среднего и СКО:



Нормальный закон распределения случайной величины с параметрами $a = 0$ и $\sigma = 1$ называется стандартным или нормированным, а соответствующая нормальная кривая - стандартной или нормированной.

Функция Лапласа определяется как:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$$

Вероятность попадания нормально распределенной случайной величины X в заданный интервал (α, β):

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Вероятность отклонения нормально распределенной случайной величины X на величину δ от математического ожидания (по модулю).

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$