Важнейшие законы распределения

Обычная практика при решении прикладных задач, связанных с изучением тех или иных СВ, – моделирование анализируемых величин и зависимостей с помощью типовых законов распределения.

К настоящему времени сформирован довольно обширный список хорошо изученных законов распределения, которые позволяют получить вероятностные модели большинства практически значимых ситуаций.

Имеют готовую реализацию во всех программных пакетах, поддерживающих методы статистической обработки

В данной презентации рассматриваются самые известные типовые законы. Позднее (по мере необходимости) будут использоваться и некоторые другие типовые распределения.

1. Дискретные законы

Распределение Бернулли

СВ X имеет ***распределение Бернулли***, если у нее два возможных значения

$$x_1 = 0 \quad \text{и} \quad x_2 = 1,$$

вероятности которых равны, соответственно,

$$q = 1 - p \quad \text{и} \quad p.$$

p, $0 < p < 1$ – параметр распределения Бернулли.

X	0	1
p	$1 - p$	p

Распределение Бернулли

Условия возникновения.

Пусть производится некоторое испытание, в результате которого с вероятностью $\boldsymbol{p}$ может наступить событие $\boldsymbol{A}$.

СВ $\boldsymbol{X}$ – число появлений события $\boldsymbol{A}$ в одном испытании (0 или 1) – имеет распределение Бернулли с параметром $\boldsymbol{p}$.

Числовые характеристики.

$$\boldsymbol{M(X) = p, \quad D(X) = pq, \quad \sigma(X) = \sqrt{pq}}.$$

Упражнение: вычислить самостоятельно указанные числовые характеристики

Распределение Бернулли

<u>Примеры</u>.

- СВ ***X*** – число выпадений герба при одном подбрасывании монеты (0 или 1); ***p*** = 1/2.
- СВ ***X*** – число кликов по рекламному объявлению в ГИС при одном показе этого объявления пользователю.

  CTR = ***p*** – *показатель кликабельности* рекламы.

Биномиальный закон распределения

СВ ***X*** имеет ***биномиальное распределение***, если ее возможные значения

$x_0 = 0,\quad x_1 = 1,\quad x_2 = 2,\quad \ldots,\quad x_{n-1} = n - 1,\quad x_n = n,$

а соответствующие им вероятности определяются по формуле Бернулли:

$$p_i = P(X = i) = C_n^i \cdot p^i \cdot q^{n-i}, \quad i = 0, 1\ 2,\ \ldots, n,$$

где $0 < p < 1,\ q = 1 - p.$

n и ***p*** – параметры биномиального распределения.

Биномиальный закон распределения

Условия возникновения.

Пусть производится $\boldsymbol{n}$ независимых испытаний, в каждом из которых событие $\boldsymbol{A}$ появляется с одной и той же вероятностью $\boldsymbol{p}$.

СВ $\boldsymbol{X}$ – число появлений события $\boldsymbol{A}$ в $\boldsymbol{n}$ испытаниях – имеет биномиальное распределение с параметрами $\boldsymbol{n}$ и $\boldsymbol{p}$.

Числовые характеристики.

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

Распределение Пуассона

СВ ***X*** имеет ***распределение Пуассона***, если ее возможные значения

$$x_0 = 0, \quad x_1 = 1, \quad x_2 = 2, \quad \ldots,$$

(бесконечное, но счетное множество значений),

а соответствующие им вероятности определяются формулой

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

где ***a*** > 0 – некоторое число, называемое *параметром закона Пуассона*.

Распределение Пуассона

Числовые характеристики.

$$\boxed{M(X) = D(X) = a,}$$

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

Характеристическое свойство распределения Пуассона

Распределение Пуассона

Условия возникновения.

1. Распределение Пуассона – предельное для биномиального, когда

$$\begin{cases} n \to \infty, \\ p \to 0, \end{cases} \quad \text{причем} \quad \lim_{\substack{n\to\infty \\ p\to 0}} np = a.$$

Другое название закона – *закон редких явлений*.

На практике распределение Пуассона с параметром $a = np$ может приближенно применяться вместо биномиального, когда число опытов n очень велико, а вероятность p наступления события A в каждом опыте очень мала.

Распределение Пуассона

2. Последовательность случайных моментов возникновения некоторых однородных событий называется ***потоком событий***.

Поток событий может обладать следующими свойствами.

1) ***Стационарность***.

Среднее число событий λ, появляющихся в единицу времени, постоянно. Это число называется *интенсивностью* (*плотностью*) потока.

2) ***Ординарность***.

Вероятность появления на малом промежутке времени Δt двух и более событий пренебрежимо мала по сравнению с вероятностью появления на нем одного события.

События происходят поодиночке, а не парами, тройками, и т. д.

3) ***Отсутствие последействия***.

Для любых неперекрывающихся промежутков времени число событий, попадающих на один из этих промежутков, не зависит от числа событий, попадающих на другие.

События наступают независимо друг от друга

Распределение Пуассона

Поток событий, обладающий свойствами стационарности, ординарности и отсутствия последействия, называется ***простейшим*** (***стационарным пуассоновским***) потоком.

Пусть имеется простейший поток событий с плотностью $\boldsymbol{\lambda}$.

СВ $\boldsymbol{X}$ – число событий, возникающих на промежутке времени длины $\boldsymbol{\tau}$, имеет распределение Пуассона с параметром $\boldsymbol{a} = \boldsymbol{\lambda\tau}$:

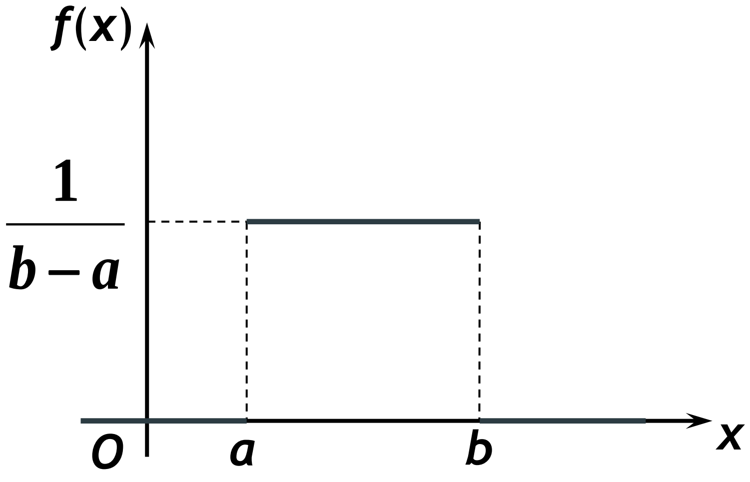
$$\boldsymbol{P(X = k) = \frac{(\lambda\tau)^k}{k!} \cdot e^{-\lambda\tau}, \quad k = 0, 1, 2, \ \ldots}$$

2. Непрерывные законы

Равномерное распределение

СВ *X* имеет ***равномерное на*** (***a***, ***b***) распределение, если плотность ее распределения имеет вид

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

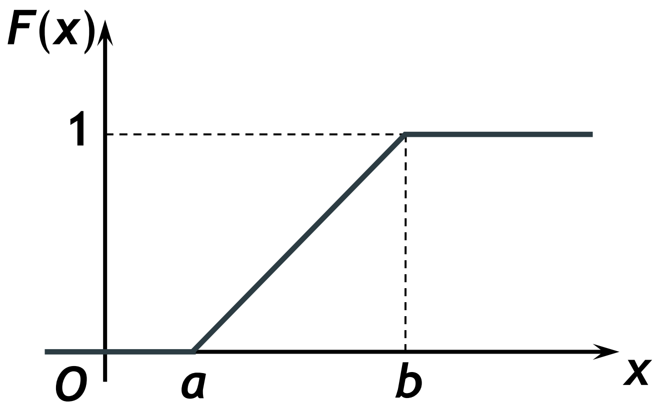


Все возможные значения СВ ***X*** принадлежат интервалу (***a***, ***b***), причем в пределах этого интервала плотность вероятности постоянна

Равномерное распределение

Функция распределения (интегральная функция).

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



Числовые характеристики.

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

Упражнение: построить самостоятельно функцию распределения равномерного закона;

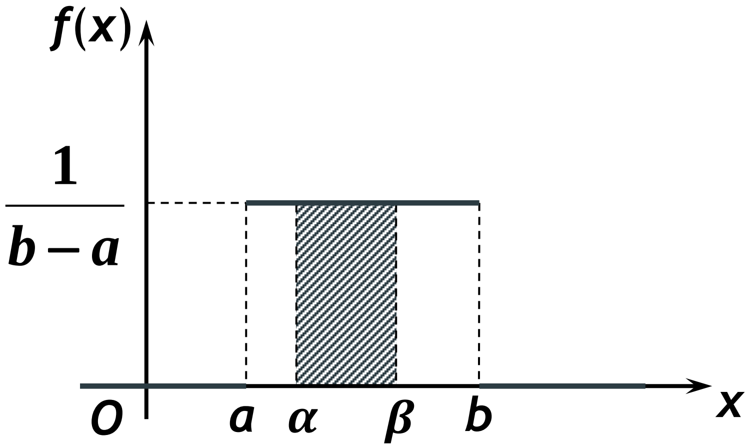
вычислить его числовые характеристики

Равномерное распределение

Вероятность попадания СВ в интервал $(\alpha, \beta) \subset (a, b)$

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha) = \frac{\beta - \alpha}{b - a}.$$

Проверить!



Равномерное распределение

Примеры.

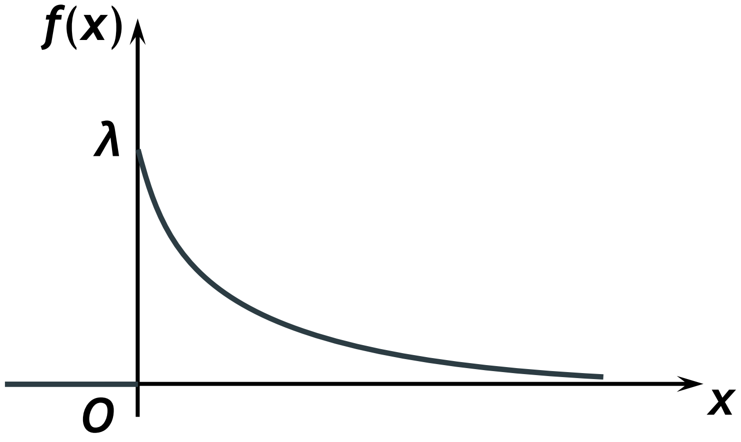
1) Ошибки округления (в том числе, на ЭВМ): равномерное распределение на интервале (-0.5$\cdot\mathbf{10^{-k}}$, 0.5 $\cdot\,\mathbf{10^{-k}}$), где $\boldsymbol{k}$ – число разрядов.

2) Транспорт приходит строго по расписанию с интервалом $\boldsymbol{t_0}$. Пассажир приходит на остановочный пункт в случайный момент времени (не связанный с расписанием транспорта).

СВ $\boldsymbol{T}$ – время ожидания транспорта – имеет равномерное на $(\mathbf{0},\ \boldsymbol{t_0})$ распределение.

Показательное (экспоненциальное) распределение

СВ ***X*** имеет ***показательное распределение***, если плотность распределения имеет вид



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

где $\boldsymbol{\lambda} > 0$ – параметр показательного распределения.

Показательное распределение

Функция распределения (интегральная функция).

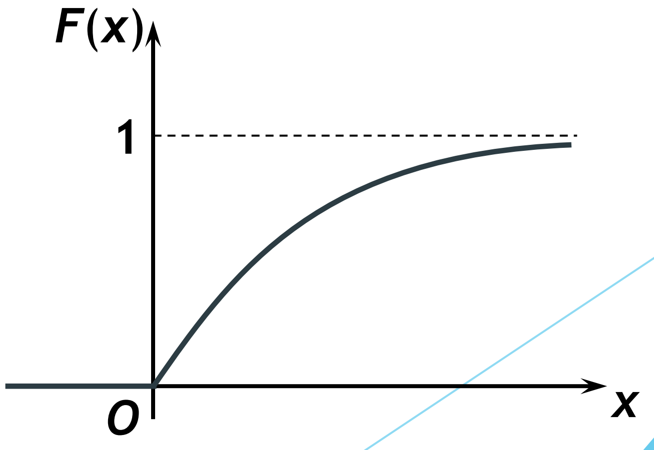
При $x \le 0$ $\quad F(x) = \int\limits_{-\infty}^{x} 0\,dt = 0$,

при $x > 0$

$$F(x) = \int\limits_{-\infty}^{0} 0\,dt + \int\limits_{0}^{x} \lambda \cdot e^{-\lambda t}dt = 0 - e^{-\lambda t}\Big|_0^x = 1 - e^{-\lambda x}.$$

Итог:

$$F(x) = \begin{cases} 0 & \text{при } x \le 0, \\ 1 - e^{-\lambda x} & \text{при } x > 0, \end{cases}$$



Показательное распределение

Числовые характеристики.

$$M(x)=\int_{0}^{+\infty} x\cdot\lambda e^{-\lambda x}dx = \lim_{b\to+\infty}\int_{0}^{b} x\cdot\lambda e^{-\lambda x}dx =$$

$$=\begin{bmatrix} u=x & du=dx \\ dv=\lambda e^{-\lambda x}dx & v=-e^{-\lambda x}\end{bmatrix} =$$

Интегрирование по частям

$$=\lim_{b\to+\infty}\left(-x\cdot\lambda e^{-\lambda x}\Big|_{0}^{b}+\int_{0}^{b}\lambda e^{-\lambda x}\,dx\right) = \frac{1}{\lambda}.$$

Показательное распределение

$$D(x) = \int_0^{+\infty} x^2 \cdot \lambda e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2 = \lim_{b \to +\infty} \int_0^{b} x^2 \cdot \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} =$$

$$= \begin{bmatrix} u = x^2 & du = 2x \cdot dx \\ dv = \lambda e^{-\lambda x} dx & v = -e^{-\lambda x} \end{bmatrix} = \quad \dots \quad =$$

Дважды интегрирование по частям

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Проверить!

Окончательно:

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

Показательное распределение

<u>Вероятность попадания СВ в интервал</u> $(\boldsymbol{a}, \boldsymbol{b})$, $\boldsymbol{a} > 0$, $\boldsymbol{b} > 0$.

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

<u>Условия возникновения</u>.

Пусть имеется простейший поток событий с плотностью $\boldsymbol{\lambda}$.

СВ $\boldsymbol{T}$ – интервал времени между двумя соседними событиями в потоке – имеет показательное распределение с параметром $\boldsymbol{\lambda}$:

$$f(t) = \lambda \cdot e^{-\lambda t}, \qquad t > 0.$$

Нормальный закон распределения (закон Гаусса)

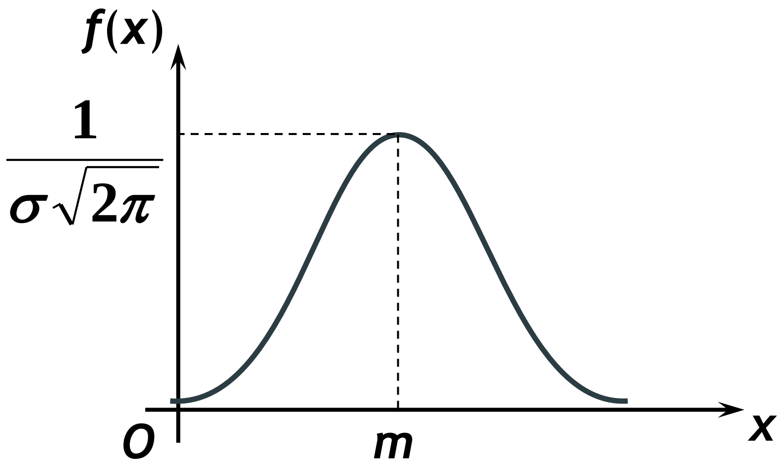
Нормальный закон распределения (закон Гаусса) играет исключительно важную роль в теории вероятностей и занимает среди законов распределения особое место.

Это – наиболее часто встречающийся на практике закон распределения.

Нормальный закон распределения

СВ ***X*** имеет ***нормальное распределение с параметрами m*** и ***σ***, если плотность распределения имеет вид

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



Кривая нормального распределения называется ***нормальной кривой*** или ***кривой Гаусса***

Нормальный закон распределения

Числовые характеристики.

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

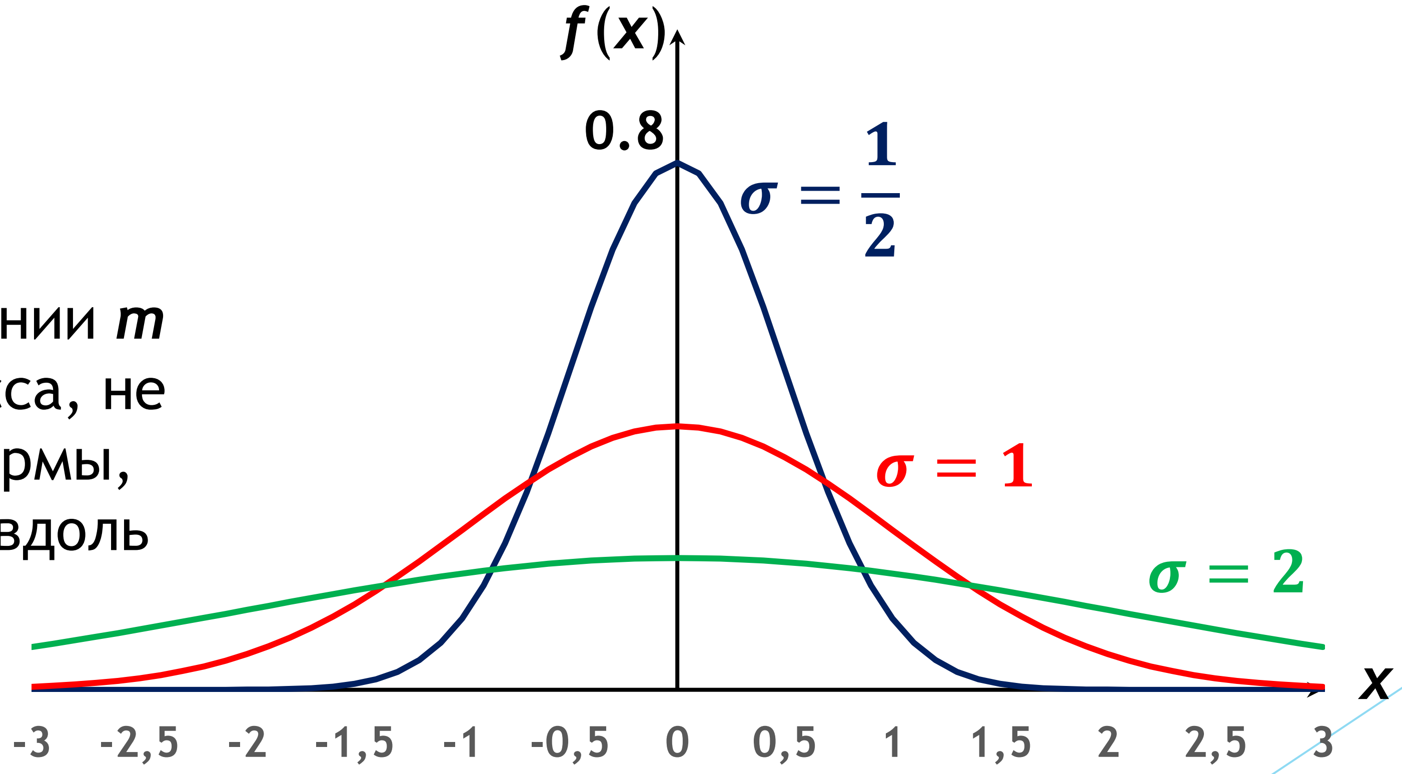
Функция распределения (интегральная функция).

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{t^2}{2}} dt \; .$$

Замена переменной $\left[t = \frac{x-m}{\sigma}\right]$

Нормальный закон распределения

Вид кривой нормального распределения при $m = 0$ и различных значениях σ:



При изменении m кривая Гаусса, не изменяя формы, смещается вдоль оси Ox

Стандартное нормальное распределение

Нормированным (стандартным) нормальным распределением называется нормальное распределение с параметрами $\boldsymbol{m} = 0$ и $\boldsymbol{\sigma} = 1$.

Плотность нормированного распределения имеет вид

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

функция распределения (интегральная функция):

$$F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt .$$

Обе функции табулированы (таблицы в учебниках и справочниках), а также реализованы в программных платформах, поддерживающих статистическую обработку данных

Нормальный закон распределения

Вероятность попадания СВ в интервал (a, b).

$$P(a < X < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} e^{-\frac{t^2}{2}} dt.$$

Может быть выражена через табулированные функции.
Например:

$$P(a < X < b) = F(b) - F(a) = F_0\left(\frac{b-m}{\sigma}\right) - F_0\left(\frac{a-m}{\sigma}\right).$$

Нормальный закон распределения

В частности, вероятность отклонения СВ $\boldsymbol{X}$ от своего среднего значения не более, чем на заданную величину $\boldsymbol{\delta}$:

$$\boldsymbol{P}(|\boldsymbol{X}-\boldsymbol{m}|<\boldsymbol{\delta}) = \boldsymbol{P}(\boldsymbol{m}-\boldsymbol{\delta}<\boldsymbol{X}<\boldsymbol{m}+\boldsymbol{\delta}) = \boldsymbol{2F_0}\left(\frac{\boldsymbol{\delta}}{\boldsymbol{\sigma}}\right)-\boldsymbol{1}.$$

Нормальный закон распределения

Пример.

Рассматривается производство шариков для подшипников. Номинальный диаметр шариков равен 10 мм; фактический диаметр случаен и распределен по нормальному закону с математическим ожиданием 10 мм и с. к. о. 0.4 мм.

При контроле бракуются все шарики, не проходящие через круглое отверстие диаметром 10.7 мм, и все, проходящие через круглое отверстие диаметром 9.3 мм.

Найти средний процент бракованных шариков при описанных условиях.

Нормальный закон распределения

<u>Пример</u> (продолжение).

Вероятность того, что случайно выбранный шарик будет забракован, равна

$$\boldsymbol{P(|X-m| \geq 0.7) = 1 - P(|X-m| < 0.7)}\,.$$

$$\boldsymbol{P(|X-m| < 0.7) = 2F_0\left(\frac{0.7}{0.4}\right) - 1 \approx 2 \cdot 0.96 - 1 = 0.92},$$

поэтому $\boldsymbol{P(|X-m| \geq 0.7) \approx 0.08}$.

Используя понятие статистической вероятности, можем заключить: в среднем будет браковаться 8% шариков.

Правило «трех сигма»

Определим вероятность того, что СВ $\boldsymbol{X}$, распределенная нормально с параметрами $\boldsymbol{m}$ и $\boldsymbol{\sigma}$, отклонится от своего математического ожидания не более, чем на $\boldsymbol{3\sigma}$:

$$\boldsymbol{P(|X-m|<3\sigma) \;=\; P(m-3\sigma<X<m+3\sigma) \;=}$$
$$\boldsymbol{=\;\; 2F_0(3)-1 \;\;=\;\; 2\cdot 0,9987-1 \;\;=\;\; 0,9973}.$$

С вероятностью, близкой к 1, нормально распределенная СВ будет принимать значения, принадлежащие промежутку **(*m* – 3*σ*, *m* +3*σ*).**

Правило «трех сигма»

При проведении практических расчетов принимают, что практически все возможные значения нормально распределенной СВ принадлежат промежутку **(m – 3σ, m +3σ)**.

Правило «трех сигма»

Пример применения правила «трех сигма».

Результаты IQ-теста формируются таким образом, чтобы они подчинялись нормальному закону распределения с параметрами 100 и 14.

Нормальный закон распределения: условия возникновения

Установление этих условий – в различных формах ***центральной предельной теоремы*** (ЦПТ).

Нормальный закон распределения возникает, когда рассматриваемая СВ может быть представлена в виде суммы достаточно большого числа независимых (или слабо зависимых) элементарных слагаемых, каждое из которых в отдельности сравнительно мало влияет на сумму.

Такая ситуация часто встречается на практике

↓

широкая распространенность нормального закона.

ЦПТ для одинаково распределенных слагаемых

Теорема (ЦПТ для одинаково распределенных слагаемых).

Пусть $X_1,\ X_2,\ \ldots,\ X_n$ - независимые СВ, имеющие одно и то же распределение с математическим ожиданием $\boldsymbol{m}$ и дисперсией $\boldsymbol{\sigma}^2$.

Тогда СВ

$$\boldsymbol{Y} = \sum_{k=1}^{n} \boldsymbol{X}_k$$

имеет асимптотически нормальное распределение с параметрами $\boldsymbol{m_Y = n \cdot m}$ и $\boldsymbol{\sigma_Y = \sigma\sqrt{n}}$.

ЦПТ для одинаково распределенных слагаемых

Замечание 1.

Закон распределения СВ $X_1, X_2, \dots, X_n$ может быть любым (непрерывным или дискретным).

При этом СВ

$$Y = \sum_{k=1}^{n} X_k$$

при достаточно большом n будет иметь распределение, близкое к нормальному.

ЦПТ для одинаково распределенных слагаемых

Замечание 2.

Заключение теоремы может быть сформулировано иначе:

СВ

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

имеет асимптотически нормальное распределение

с параметрами $m_{\overline{X}} = m$ и $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$.

ЦПТ для одинаково распределенных слагаемых

Пример 1.

СВ X_1, X_2, … , X_n имеют показательный закон распределения с параметром $\lambda = 2$.

Рассмотрим СВ

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

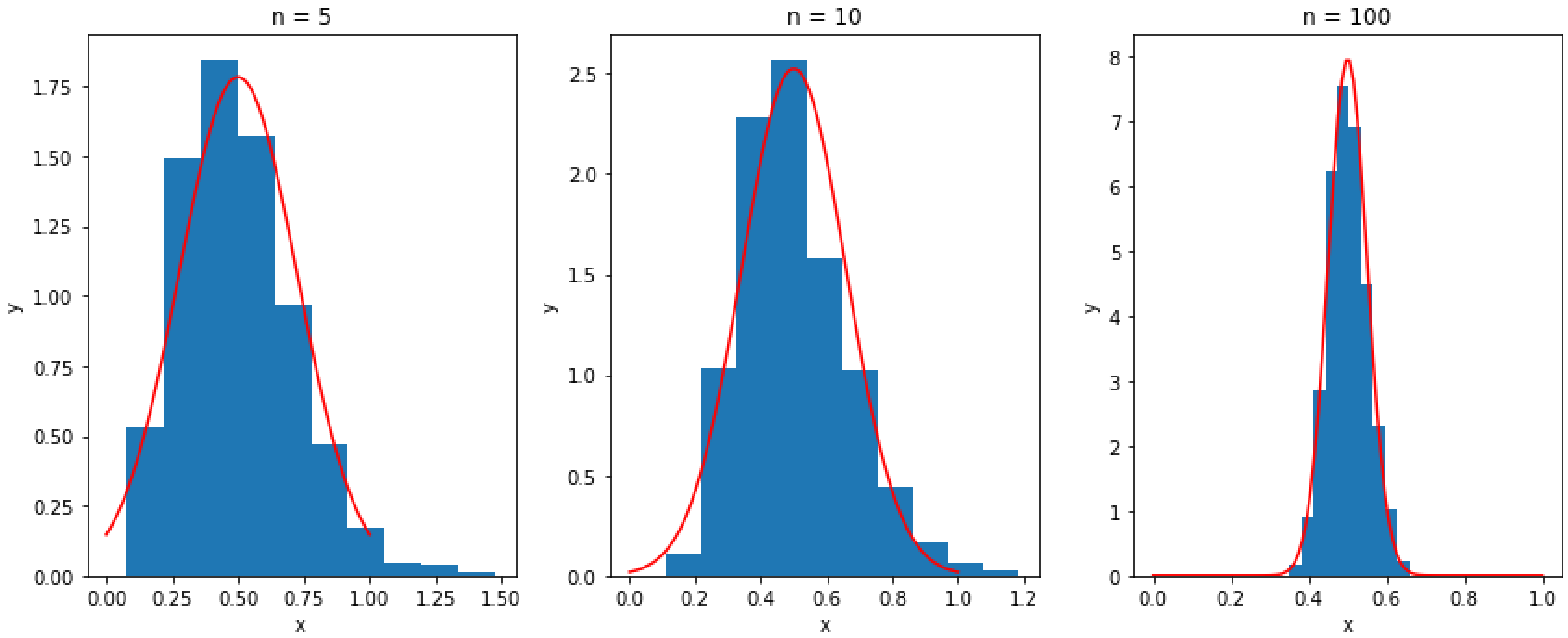
при $n = 5$, $n = 10$, и $n = 100$.

С помощью генератора псевдослучайных чисел сгенерировано по 1000 значений СВ $\bar{X}$ для каждого значения n.

ЦПТ для одинаково распределенных слагаемых

<u>Пример 1</u> (продолжение).

Гистограммы полученных распределений и кривая плотности предельного нормального распределения:



ЦПТ для одинаково распределенных слагаемых

Пример 2 (машина Гальтона).

Машина Гальтона (*доска Гальтона, Galton board*) – первый экземпляр сконструирован в 1873 г. для демонстрации закономерности, описываемой ЦПТ.

Представляет собой ящик, в заднюю стенку которого вбиты штыри в шахматном порядке. Сверху через воронку, расположенную ровно посередине между боковыми стенками, в ящик засыпается большое количество шариков.

При столкновении со штырьками шарики с одинаковой вероятностью могут падать справа или слева от каждого штырька.

ЦПТ для одинаково распределенных слагаемых

Пример 2 (машина Гальтона - продолжение).

Нижняя часть ящика разделена перегородками по числу штырьков в нижнем ряду.

Шарики, скатываясь на дно ящика, образуют столбики, высота которых тем больше, чем ближе столбик к середине доски.

При достаточно большом количестве шариков столбики показывают картину, близкую к кривой нормального распределения.

Видео-демонстрация.

Логнормальное распределение

СВ X имеет ***логнормальное распределение***, если СВ **ln**X имеет нормальное распределение.

Обратно:

если СВ Y имеет нормальное распределение, то СВ

$$X = e^{Y}$$

имеет логнормальное распределение.

СВ, имеющая логнормальное распределение, может принимать только положительные значения (в отличие от нормально распределенной СВ)

Логнормальное распределение

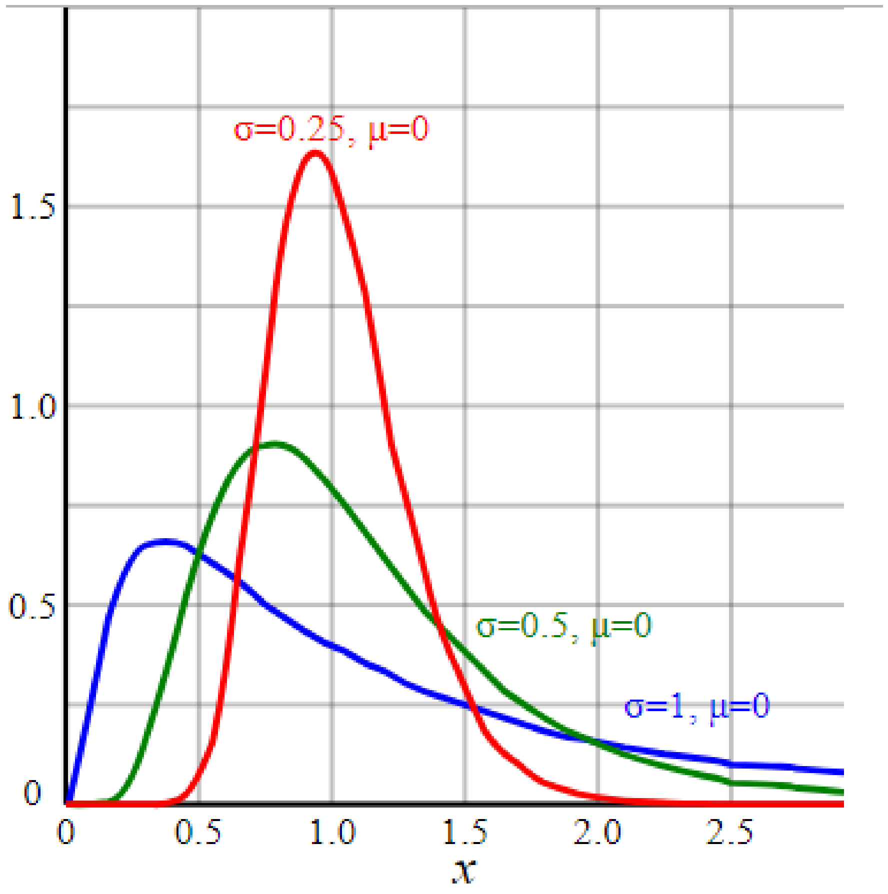
Плотность распределения логнормальной СВ $\boldsymbol{X}$ имеет вид

$$f(x) = \frac{1}{x\cdot\sigma\sqrt{2\pi}}e^{-\frac{(\ln x-m)^2}{2\sigma^2}}, \quad x > 0.$$

$\boldsymbol{m}$ и $\boldsymbol{\sigma}$ – параметры логнормального распределения.

Логнормальное распределение

Графики плотности логнормального распределения при различных значениях $\boldsymbol{\sigma}$ ($\boldsymbol{m}$ = 0):



Логнормальное распределение

Числовые характеристики.

$$M(X) = e^{m+\frac{\sigma^2}{2}}, \quad D(X) = \left(e^{\sigma^2} - 1\right) \cdot e^{2m+\sigma^2}.$$

Функция распределения (интегральная функция).

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\ln x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt = F_0\left(\frac{\ln x - m}{\sigma}\right).$$

Логнормальное распределение

При неограниченном увеличении ***x*** логнормальное распределение приближается к нормальному.

Практика статистических исследований:

при работе с непрерывной СВ, которая по своей природе не может принимать отрицательных значений, эту СВ логарифмируют и проверяют полученные данные на нормальность распределения.