Числовые характеристики нормального распределения

- Если X случайная величина, имеющая нормальное распределение с параметрами a и σ, то **математическое ожидание совпадает с параметром** a, **дисперсия с** σ^2.

$$M\,X = a,\ D\,X = \sigma^2,\ \sigma[X] = \sigma.$$

Вычисление вероятности попадания в интервал нормально распределённой с.в.

- Если случайная величина X **имеет нормальное распределение с параметрами** a , σ , то справедливы формулы

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

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

- Доказательство первой формулы следует из равенства

$$F_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(x-a)^2}{2\sigma^2}}dx = F_s\left(\frac{x-a}{\sigma}\right).$$

Пример

- Давление на выходе газопровода можно рассматривать как случайную величину, имеющую нормальное распределение с параметрами

$$a = 50 кгс/см^2 \quad , \sigma = 2 кгс/см^2$$

- Какова вероятность, что давление не превысит $47\, кгс/см^2$?
- Каким с вероятностью 0,995 может быть максимальное давление?

Решение. Обозначим через X давление на выходе газопровода. По условию $X \in N(50, 2^2)$. $P(X \le 47) =$

$$= F_s\left(\frac{47-50}{2}\right) = \Phi\left(\frac{-3}{2}\right) + \frac{1}{2} = -\Phi\left(\frac{3}{2}\right) + \frac{1}{2} =$$

$$= -0,43 + 0,5 = 0,07$$

Решение примера

- Найдём с вероятностью 0,995 максимальное давление. Обозначим через x_m максимальное давление, которое возможно с вероятностью 0,995

$$P(X \le x_m) = 0,995$$

- С другой стороны $P(X \le x_m) = F_s\left(\frac{x_m - 50}{2}\right).$
- Поэтому

$$F_s\left(\frac{x_m - 50}{2}\right) = 0,995 \Rightarrow \Phi\left(\frac{x_m - 50}{2}\right) = 0,495.$$

- По таблицам функции Лапласа $\frac{x_m - 50}{2} = 2,6.$
- Отсюда

$$x_m = 2,6 \cdot 2 + 50 = 55,2.$$

Симметричный интервал.

- Вычислим вероятность попадания в симметричный интервал относительно математического ожидания

$$\boldsymbol{P}(\left|\boldsymbol{X}-\boldsymbol{a}\right|<\delta)=\boldsymbol{P}(\boldsymbol{a}-\delta<\boldsymbol{X}<\boldsymbol{a}+\delta)$$

$$\boldsymbol{P}(\boldsymbol{a}-\delta<\boldsymbol{X}<\boldsymbol{a}+\delta)=\boldsymbol{F_s}(\frac{\boldsymbol{a}+\delta-\boldsymbol{a}}{\sigma})-\boldsymbol{F_s}(\frac{\boldsymbol{a}-\delta-\boldsymbol{a}}{\sigma})=\mathbf{1}-\mathbf{2}\boldsymbol{F_s}(\frac{\delta}{\sigma}).$$

$$\boldsymbol{P}(\boldsymbol{a}-\delta<\boldsymbol{X}<\boldsymbol{a}+\delta)=\boldsymbol{\Phi}(\frac{\boldsymbol{a}+\delta-\boldsymbol{a}}{\sigma})-\boldsymbol{\Phi}(\frac{\boldsymbol{a}-\delta-\boldsymbol{a}}{\sigma})=\mathbf{2}\boldsymbol{\Phi}(\frac{\delta}{\sigma}).$$

Правило 3-х сигм.

- В случае $\delta = 3\sigma$ вероятность отклонения случайной величины на величину больше, чем 3σ мала . Действительно,

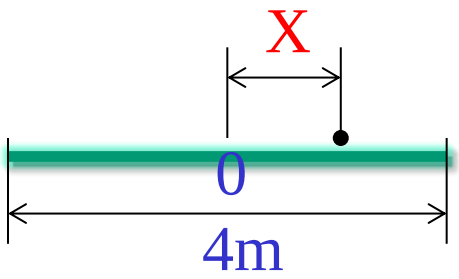
$$P(|X-a|>3\sigma)=1-P(|X-a|<3\sigma)=$$

$$=1-2\Phi(3)\approx 1-2\cdot 0,4987=0,0026.$$

- Таким образом, **значениями** случайной величины, которые отклоняются от математического ожидания **больше, чем на** 3σ, часто **пренебрегают**, **как маловероятными**.

Пример

- При артиллерийской стрельбе по полосе шириной 4 м, перпендикулярной направлению стрельбы случайное отклонение X от средней линии полосы не превышает 30 м. Найти вероятность попадания в полосу, если отклонение X подчиняется нормальному закону с нулевым средним.



Предельные теоремы

Неравенства Чебышёва

Теорема.

- 1. Если случайная величина $X \geq 0$ **неотрицательна** и имеет **конечное математическое ожидание** MX, то для любого числа $\delta > 0$ справедливо **первое неравенство Чебышёва**

$$P(X \geq \delta) \leq \frac{MX}{\delta}.$$

- 2. Если случайная величина X имеет **конечную дисперсию**, то для любого числа $\delta > 0$ справедливо **второе неравенство Чебышёва**

$$P(|X - MX| \geq \delta) \leq \frac{DX}{\delta^2}.$$

Доказательство

- Обозначим I_{A_δ} индикатор события $A_\delta = \{\omega : X(\omega) \geq \delta\}$ Очевидно, справедливы неравенства $X \geq X\, I_{A_\delta} \geq \delta I_{A_\delta}$.
- Поэтому

$$MX \geq M\delta I_{A_\delta} = \delta M I_{A_\delta} = \delta P(A_\delta) = \delta P\{X \geq \delta\}.$$

- Отсюда

$$P(X \geq \delta) \leq \frac{MX}{\delta}.$$

- Для доказательства второго неравенства применим первое неравенство к случайной величине $(X - MX)^2$

$$P(|X - MX| \geq \delta) = P\{(X - MX)^2 \geq \delta^2\} \leq \frac{M(X - MX)^2}{\delta^2} = \frac{DX}{\delta^2}.$$

Закон больших чисел. Теорема Чебышёва.

- Рассмотрим последовательность случайных величин

$$X_1, X_2, \cdots, X_n, \cdots$$

Обозначим $$\eta_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- **Теорема.** Если случайные величины $X_1, X_2, \cdots, X_n, \cdots$ **независимы** , а их **дисперсии ограничены одним и тем же числом** $DX_n \le c$, то для любого числа $\varepsilon > 0$ справедливо равенство

$$\lim_{n\to\infty} P\left\{\left|\eta_n - M\eta_n\right| \ge \varepsilon\right\} = 0$$

Доказательство

- Применим второе неравенство Чебышёва и свойства дисперсии

$$P\left\{\left|\eta_n - M\eta_n\right| \geq \varepsilon\right\} \leq \frac{D\eta_n}{\varepsilon^2} = \frac{D(\frac{1}{n}\sum_{i=1}^{n} X_i)}{\varepsilon^2} = \frac{\sum_{i=1}^{n} DX_i}{n^2\varepsilon^2} \leq \frac{cn}{n^2\varepsilon^2}.$$

Отсюда $$P\left\{\left|\eta_n - M\eta_n\right| \geq \varepsilon\right\} \to 0$$

при $n \to \infty.$

Следствие

- **Предельное равенство в теореме Чебышёва называется законом больших чисел**.
- В частном случае теоремы, когда случайные величины $X_1, X_2, \cdots, X_n, \cdots$ имеют одно и то же математическое ожидание

$$MX_1 = MX_2 = \cdots M\,X_n = \cdots = m,$$

получаем по теореме Чебышёва $\lim\limits_{n\to\infty} P\left\{\left|\eta_n - m\right| \geq \varepsilon\right\} \to 0.$

- В этом случае будем говорить, что последовательность η_n **арифметических средних сходится по вероятности к математическому ожиданию** m.

Теорема Бернулли

- Если μ_n - число появления события A в n независимых испытаниях, $p = P(A) -$ вероятность появления события A в одном испытании, то для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| \geq \varepsilon\right\} \to 0.$$

- Т.е. **относительная частота события** A равная $\frac{\mu_n}{n}$ **стремится по вероятности** к **вероятности события** A.

Доказательство.

- Обозначим $\xi_i = I_{A_i}$ случайную величину, принимающую значение 1, если событие A появляется в $i-$ ом испытании и 0, если не появляется. Тогда $M\xi_i = p, i = 1, 2, \cdots$. Полагая

$$\mu_n = \xi_1 + \cdots + \xi_n$$

учитывая, что $\eta_n = \dfrac{\mu_n}{n}$,

по следствию теоремы Чебышёва получаем

$$\lim_{n\to\infty} P\left\{\left|\eta_n - p\right| \geq \varepsilon\right\} \to 0.$$

Центральная предельная теорема

- **Общая форма** центральной предельной теоремы утверждает: **Каковы бы ни были распределения случайных величин**

$$X_1, X_2, \cdots, X_n, \cdots$$

при достаточно общих условия функция распределения суммы

$$\zeta_n = X_1 + X_2 + \cdots + X_n$$

стремится при $n \to \infty$ **к функции распределения нормальной случайной величины.**

- Дадим точную формулировку частного случая центральной предельной теоремы.

Центральная предельная теорема.Частный случай

- Если случайные величины $X_1, X_2, \cdots, X_n, \cdots$ **независимы, одинаково распределены** , то при $n \to \infty$ для любого $x \in R$

$$P\left(\frac{X_1 + X_2 + \cdots + X_n - na}{\sigma\sqrt{n}} < x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = F_s(x),$$

- где $a = MX_i, \sigma^2 = DX_i, i = 1, 2, \cdots.$

Следствие. Случай биномиального распределения

- Если μ_n - число появления события A в n независимых испытаний, $p = P(A) -$ вероятность появления события A в одном испытании $q = 1 - p$, то

$$\lim_{n\to\infty} P\left(\frac{\mu_n - np}{\sqrt{npq}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = F_s(x).$$

- Говорят, что биномиальная случайная величина μ_n имеет **асимптотически нормальное распределение**.

Интегральная теорема Муавра - Лапласа

- Справедлива формула

$$P\left(\alpha \le \mu_n \le \beta\right) = P\left(\frac{\alpha - np}{\sqrt{npq}} \le \frac{\mu_n - np}{\sqrt{npq}} \le \frac{\beta - np}{\sqrt{npq}}\right)$$

- Отсюда по центральной предельной теореме:

$$P\left(\alpha \le \mu_n \le \beta\right) \approx \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right).$$

Пример

- Вероятность выпуска бракованного изделия равна 0,05. Какова вероятность, что количество брака окажется в партии из 1000 изделий будет не более 70.
- **Решение**. По условию

$$\boldsymbol{n = 1000;\ p = 0{,}05;\ np = 50;\ npq = 50 \cdot 0{,}95 = 47{,}5.}$$

- Обозначим через μ число бракованных изделий в партии.

Продолжение решения примера

- **По интегральной теореме Муавра – Лапласа**

имеем

$$P\left(0 \le \mu \le 70\right) = \Phi\left(\frac{70-50}{\sqrt{47,5}}\right) - \Phi\left(\frac{0-50}{\sqrt{47,5}}\right)$$

Локальная теорема Муавра-Лапласа.

- Если μ_n число появления события A в n независимых испытаний, а величина

$$x_m = \frac{m - np}{\sqrt{npq}}, \qquad \infty < a \le x_m \le b < \infty,$$

- ограничена при $n \to \infty$, то справедлива формула

$$P(\mu_n = m) = \frac{\varphi(x_m)}{\sqrt{npq}}(1 + \alpha_n), \quad n \to \infty,$$

- где $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ - плотность стандартного нормального распределения, а $|\alpha_n| \le \frac{C}{\sqrt{n}}, \quad C > 0$

Распределение хи-квадрат (распределение Пирсона)

- Распределение хи-квадрат относится к выборочным распределениям.
- Если случайные величины X_k независимы и нормальны N(0,1), то известно, что величина

$$\chi^2 = \sum_{k=1}^{n} X_k^{\,2}$$

- имеет распределение хи-квадрат с n степенями свободы.

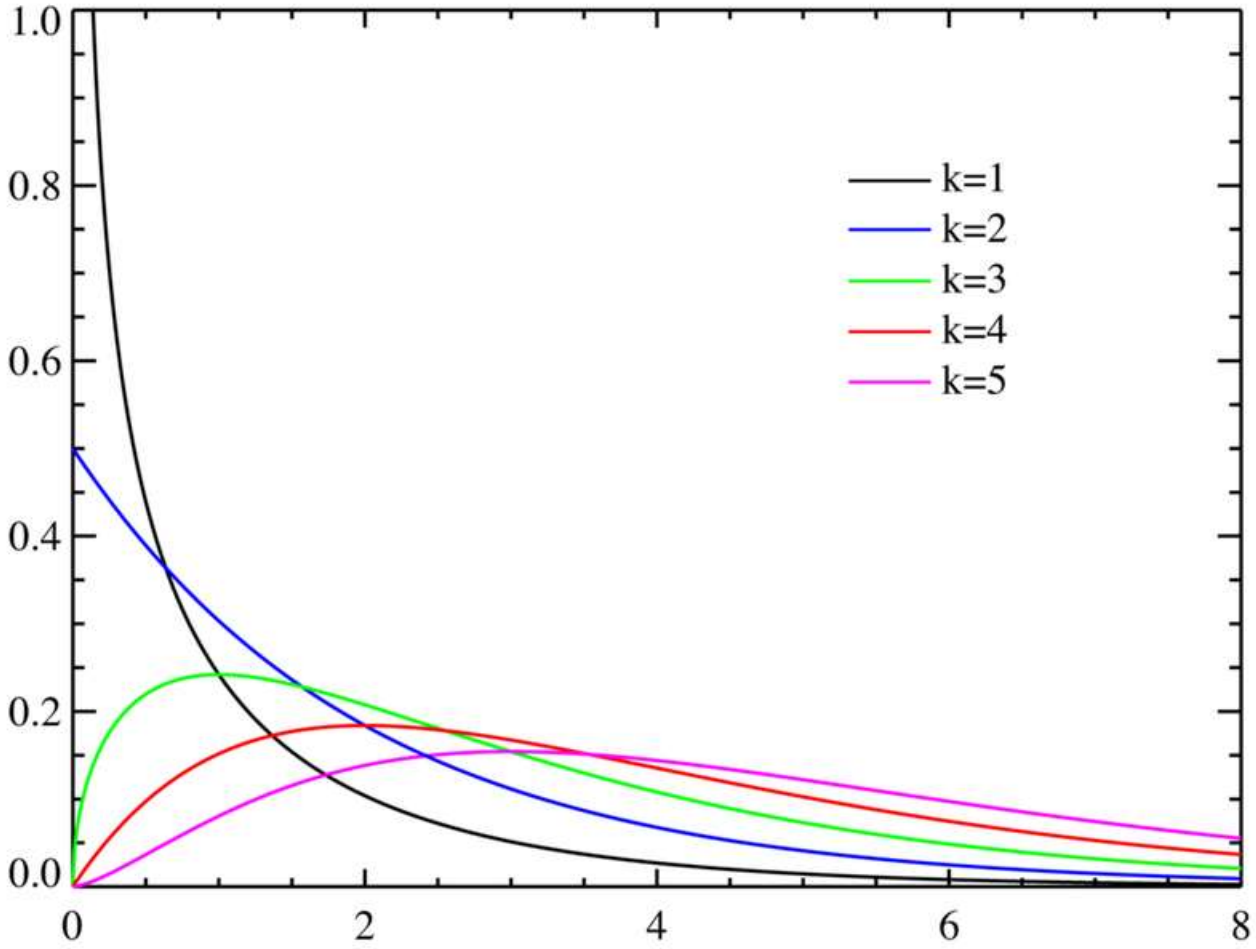
Плотность распределения хи-квадрат с n степенями свободы имеет вид

$$f_n(x)=\begin{cases}\dfrac{e^{-x/2}x^{\frac{n}{2}-1}}{2^{n/2}\Gamma(n/2)}, & x>0\\ 0, & x\le 0\end{cases}$$

- где $\Gamma(x)=\int\limits_0^{+\infty} t^{x-1}e^{-x}dx$ - гамма-функция Эйлера.
- **Замечание**. При больших n распределение Пирсона может быть приближено нормальным $N(n, 2n)$.

Случай линейно-связных величин

- В случае, когда случайные величины $X_1, X_2, \ldots, X_n$ связаны r соотношениями. Их сумма квадратов имеет распределение хи-квадрат с $k = n - r$ степенями свободы


1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
k=1
k=2
k=3
k=4
k=5

Распределение Стьюдента (t-распределение)

- Пусть случайные величины $X, X_1, X_2, \ldots, X_n$ независимы и имеют нормальное распределение $N(0,1)$. Случайная величина

$$T = \frac{X}{\sqrt{\frac{1}{n}\sum_{k=1}^{n} X_k^2}} = \frac{X}{\sqrt{\chi^2/n}}$$

- имеет распределение Стьюдента с n степенями свободы, с плотностью

$$f_n(x) = c_n\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}, c_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi k}\,\Gamma\left(\frac{n}{2}\right)}.$$

Замечание. При больших n распределение Стьюдента приближается нормальным $N(0,\frac{n}{n-2})$.

- Синий график – плотность нормального распределения
- Зелёные и красный – плотность t-распределения (n=1,2,3,4,5)