Прикладной статистический анализ данных.
1. Введение: распределения, статистики, оценки, гипотезы.

Юлиан Сердюк @jserdyuk
Ольга Кравцова @oakravts
cs.msu.psad@gmail.com

14.02.2025

Зачем нужен этот курс

- специфические статистические методы для конкретных постановок задач
- границы применимости методов
- статистическое мышление

Описание случайных величин

Дискретная случайная величина X принимает счётное множество значений $A = \{a_1, a_2, \dots\}$ с вероятностями $p_1, p_2, \dots, \sum\limits_i p_1 = 1$.

$f_X(a_i) = \mathbf{P}(X = a_i) = p_i$ — **функция вероятности**.

Непрерывная случайная величина задаётся с помощью **функции распределения**:

$$F_X(x) = \mathbf{P}(X \leqslant x)$$

или **плотности распределения**:

$$f_X(x) : \int\limits_a^b f_X(x)\,dx = \mathbf{P}(a \leqslant X \leqslant b)\,.$$

Характеристики распределений

- **матожидание** — среднее значение X:

$$\mathbb{E}X = \int x\, dF(x) = \int x f(x)\, dx$$

- **дисперсия** — мера разброса X:

$$\mathbb{D}X = \mathbb{E}\left((X - \mathbb{E}X)^2\right)$$

- **квантиль** порядка $\alpha \in (0,1)$:

$$X_\alpha: \quad \mathbf{P}(X \leqslant X_\alpha) \geqslant \alpha, \quad \mathbf{P}(X \geqslant X_\alpha) \geqslant 1-\alpha$$

  эквивалентное определение: $X_\alpha = F^{-1}(\alpha)$
- n-**процентиль** — квантиль порядка $\frac{n}{100}$;
- **медиана** — квантиль порядка 0.5, центральное значение распределения:

$$\operatorname{med} X: \quad \mathbf{P}(X \leqslant \operatorname{med} X) \geqslant 0.5, \quad \mathbf{P}(X \geqslant \operatorname{med} X) \geqslant 0.5$$

- **интерквартильный размах**:

$$IQR = X_{0.75} - X_{0.25}$$

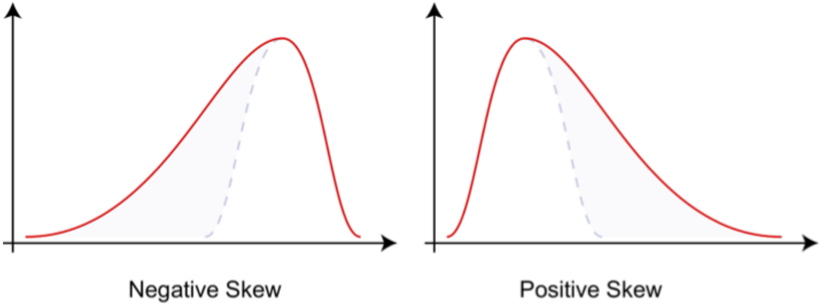
- **мода** — точка максимума функции вероятности или плотности:

$$\operatorname{mode} X = \underset{x}{\operatorname{argmax}}\, f(x)$$

Характеристики распределений

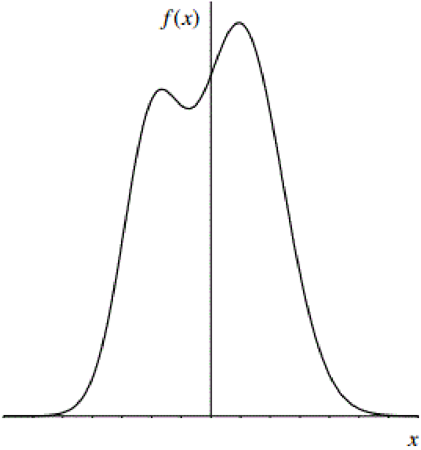
- **коэффициент ассиметрии** (skewness):

$$\gamma_1 = \mathbb{E}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathbb{D}X}}\right)^3$$



Характеристики распределений

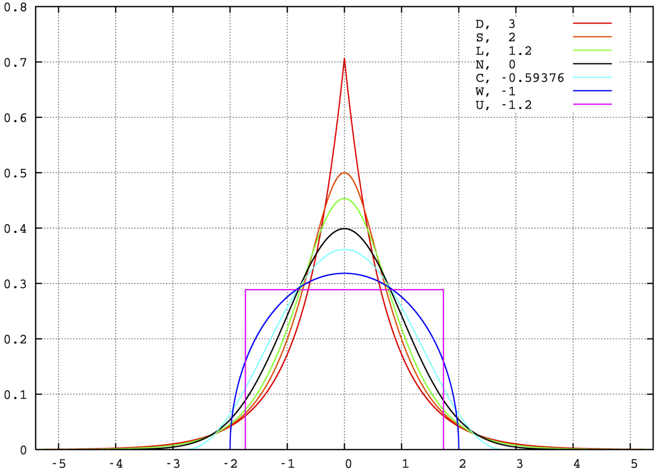
$\gamma_1 = 0$ — необходимое, но не достаточное условие симметричности:



Характеристики распределений

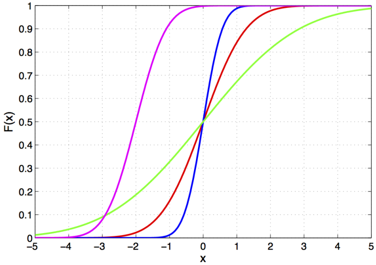
- **коэффициент эксцесса** (excess, без вычитания тройки — kurtosis):

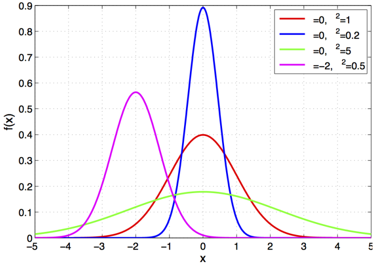
$$\gamma_2 = \frac{\mathbb{E}\left(X - \mathbb{E}X\right)^4}{\left(\mathbb{D}X\right)^2} - 3$$



Нормальное распределение

$$X \in \mathbb{R} \sim N\left(\mu, \sigma^2\right), \ \sigma^2 > 0$$





$$F\left(x\right) = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad \Phi\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$$

$$f\left(x\right) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \qquad \phi\left(x\right) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

Нормальное распределение

- предельное распределение суммы слабо взаимозависимых сл. в.
- $\mathbb{E}X = \operatorname{med} X = \operatorname{mode} X = \mu$, $\mathbb{D}X = \sigma^2$, все моменты более высокого порядка нулевые
- пусть $X_1, \ldots, X_n$ независимы, $X_i \sim N\left(\mu_i, \sigma_i^2\right)$, тогда $\forall a_1, \ldots, a_n$

$$\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

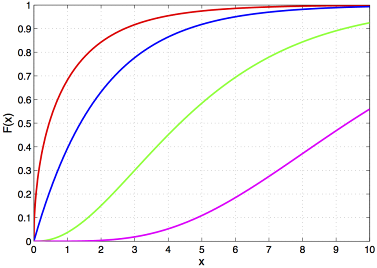
- центральная предельная теорема: пусть $X_1, \ldots, X_n$ i.i.d. с $\mathbb{E}X$ и $\mathbb{D}X < \infty$, тогда

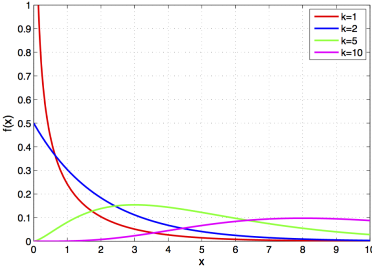
$$\frac{1}{n} \sum_{i=1}^n X_i \sim\approx N\left(\mathbb{E}X, \frac{\mathbb{D}X}{n}\right)$$

- пример: погрешность измерения

Распределение хи-квадрат

$X \in \mathbb{R}_+ \sim \chi^2_k,\ k \in \mathbb{N}$





$$F(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right)}\gamma\left(\frac{k}{2}, \frac{x}{2}\right)$$

$$f(x) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$$

$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ — гамма-функция
$\gamma(a, x) = \int_0^x e^{-t}t^{a-1}dt$ — нижняя неполная гамма-функция

Распределение хи-квадрат

- пусть $X_1, \ldots, X_k$ — i.i.d., $\quad X_i \sim N(0,1)$, тогда

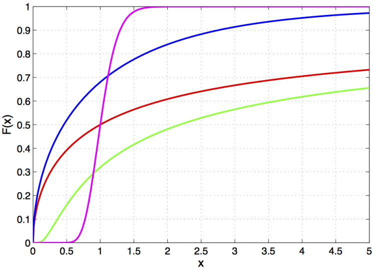
$$\sum_{i=1}^{k} X_i^2 \sim \chi_k^2$$

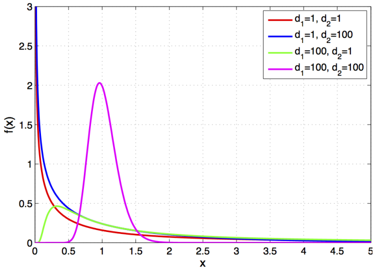
- пример: нормированная выборочная дисперсия:

$$(n-1)\frac{S_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

Распределение Фишера

$$X \in \mathbb{R}_+ \sim F\left(d_1, d_2\right), \; d_1, d_2 > 0$$





$$F\left(x\right) = I_{\frac{d_1 x}{d_1 x + d_2}}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)$$

$$f\left(x\right) = \sqrt{\frac{\left(d_1 x\right)^{d_1} d_2^{d_2}}{\left(d_1 x + d_2\right)^{d_1 + d_2}}} \Bigg/ x B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)$$

$B\left(a, b\right) = \int_0^1 t^{a-1}\left(1-t\right)^{b-1} dt$ — бета-функция

$I_x\left(a, b\right) = \frac{B(x; a, b)}{B(a, b)}$ — регуляризованная неполная бета-функция

$B\left(x; a, b\right) = \int_0^x t^{a-1}\left(1-t\right)^{b-1} dt$ — неполная бета-функция

Распределение Фишера

- пусть $X_1 \sim \chi^2_{d_1}$, $X_2 \sim \chi^2_{d_2}$, X_1 и X_2 независимы, тогда

$$\frac{X_1/d_1}{X_2/d_2} \sim F(d_1, d_2)$$

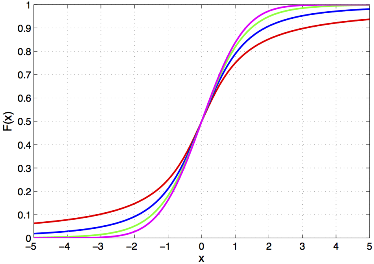
- если $X \sim F(d_1, d_2)$, то

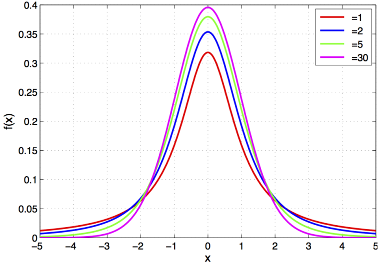
$$Y = \lim_{d_2 \to \infty} d_1 X \sim \chi^2_{d_1}$$

- $F(x, d_1, d_2) = F(1/x, d_2, d_1)$
- возникает в дисперсионном и регрессионном анализе

Распределение Стьюдента

$$X \in \mathbb{R} \sim St\left(\nu\right), \ \nu > 0$$





$$F\left(x\right) = \frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right)$$

$$f\left(x\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Распределение Стьюдента

- $\mathbb{E}X = 0$ при $\nu > 1$, $\operatorname{med} X = \operatorname{mode} X = 0$ всегда
- пусть $Z \sim N(0,1)$ и $V \sim \chi^2_\nu$ независимы, тогда

$$T = \frac{Z}{\sqrt{V/\nu}} \sim St(\nu)$$

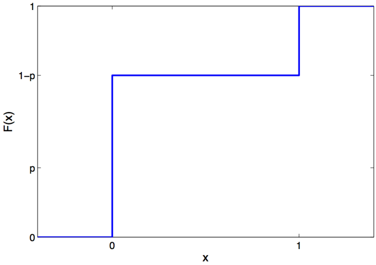
- если $X \sim St(\nu)$, то

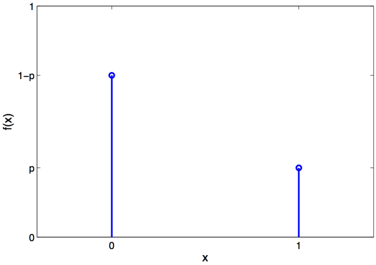
$$Y = \lim_{\nu \to \infty} X \sim N(0,1)$$

- возникает при оценке среднего значения сл. в. с неизвестной дисперсией

Распределение Бернулли

$$X \in \{0, 1\} \sim Ber\left(p\right), \quad p \in (0, 1)$$





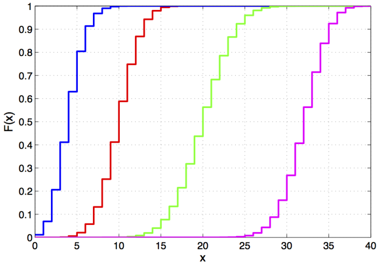
$$F\left(x\right) = \begin{cases} 0, & x < 0, \\ 1 - p, & 0 \leqslant x < 1, \\ 1, & x \geqslant 1. \end{cases}$$

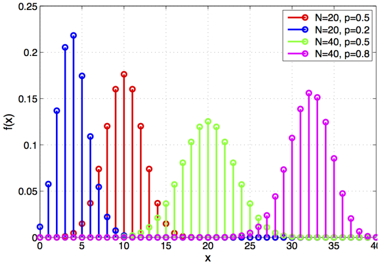
$$f\left(x\right) = \begin{cases} 1 - p, & x = 0, \\ p, & x = 1. \end{cases}$$

- пример: результат подбрасывания монеты

Биномиальное распределение

$$X \in \{0, \ldots, N\} \sim Bin\,(N, p)\,, \;\; N \in \mathbb{N}, \;\; p \in [0, 1]$$





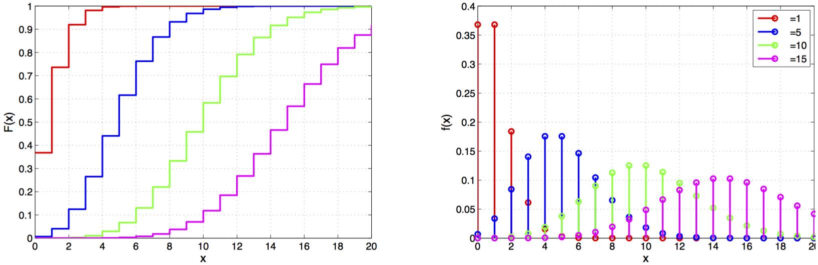
$$F\,(x) = I_{1-p}\,(N - x, 1 + x)$$
$$f\,(x) = C_N^x p^x\,(1-p)^{N-x}$$

Биномиальное распределение

- пусть $X_1, \dots, X_n$ независимы, $X_i \sim Ber(p)$, тогда
$$\sum_{i=1}^{n} X_i \sim Bin(n, p).$$
- $Bin\left(1, p\right) = Ber\left(p\right)$
- если $N > 20$ и p не слишком близко к нулю или единице, то для $X \sim Bin\left(N, p\right)$ справедлива нормальная аппроксимация:
$$F_X\left(x\right) \approx \Phi\left(\frac{x - Np}{\sqrt{Np\left(1 - p\right)}}\right)$$
- пример: число попаданий из N бросков в баскетбольное кольцо

Распределение Пуассона

$$X \in \{0, 1, 2, \dots\} \sim Pois(\lambda), \ \lambda > 0$$



$$F(x) = e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!}$$

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Распределение Пуассона

- распределение числа независимых событий в фиксированном временном или пространственном интервале
- $\mathbb{E}X = \mathbb{D}X = \lambda$
- пусть $X_1, \dots, X_n$ независимы, $\quad X_i \sim Pois\left(\lambda_i\right)$, тогда

$$\sum_{i=1}^{n} X_i \sim Pois\left(\sum_{i=1}^{n} \lambda_i\right)$$

- если $X \sim Pois\left(\lambda\right)$, $Y = \sqrt{X}$, то при больших λ

$$F_Y\left(x\right) \approx \Phi\left(\frac{x - \sqrt{\lambda}}{\sqrt{\lambda}}\right)$$

- пример: количество изюма в булочке с изюмом

Выборка

Генеральная совокупность — множество объектов, свойства которых подлежат изучению в рассматриваемой задаче.

Выборка — конечное множество объектов, отобранных из генеральной совокупности для проведения измерений.

$$X^n = (X_1, \ldots, X_n).$$

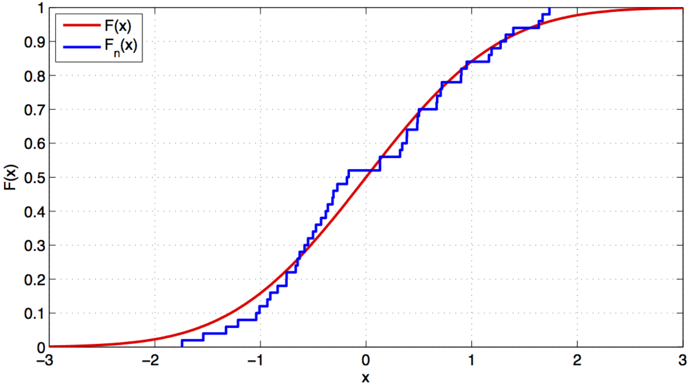
n — **объём выборки**.

X^n — **простая выборка**, если $X_1, \ldots, X_n$ — независимые одинаково распределённые случайные величины (i.i.d.).

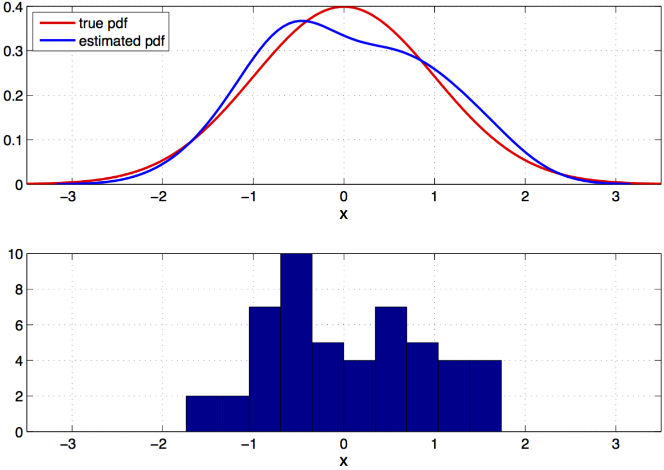
Основная задача статистики — описание $F_X(x)$ по реализации выборки.

Функция распределения

$F_n(x) = \frac{1}{n}\sum\limits_{i=1}^{n}[X_i \leqslant x]$ — **эмпирическая функция распределения.**



Плотность распределения



Статистика

Статистика $T(X^n)$ — любая измеримая функция выборки.

- выборочное среднее:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- выборочная дисперсия:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

вариационный ряд:

$$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$$

ранг элемента выборки X_i:

$$\operatorname{rank}(X_i) = r\colon X_i = X_{(r)}$$

- k-я порядковая статистика: $X_{(k)}$
- выборочный α-квантиль: $X_{([n\alpha])}$
- выборочная медиана:

$$m = \begin{cases} X_{(k+1)}, & \text{если } n = 2k+1, \\ \frac{X_{(k)}+X_{(k+1)}}{2}, & \text{если } n = 2k. \end{cases}$$

Статистика

- выборочный интерквартильный размах:

$$IQR_n = X_{([0.75n])} - X_{([0.25n])}$$

- выборочный коэффициент ассиметрии:

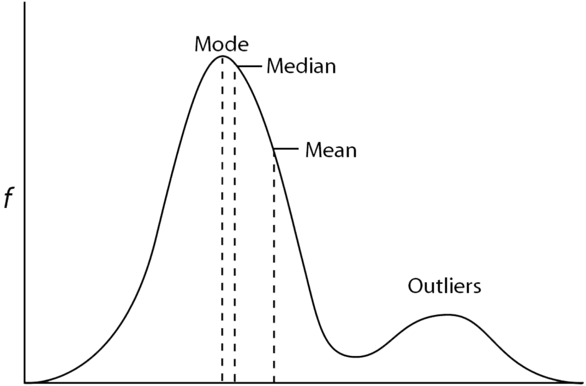
$$g_1 = \frac{\sqrt{n}\sum\limits_{i=1}^{n}\left(X_i - \bar{X}\right)^3}{\left(\sum\limits_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right)^{3/2}}$$

- выборочный коэффициент эксцесса:

$$g_2 = \frac{n\sum\limits_{i=1}^{n}\left(X_i - \bar{X}\right)^4}{\left(\sum\limits_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right)^{2}} - 3$$

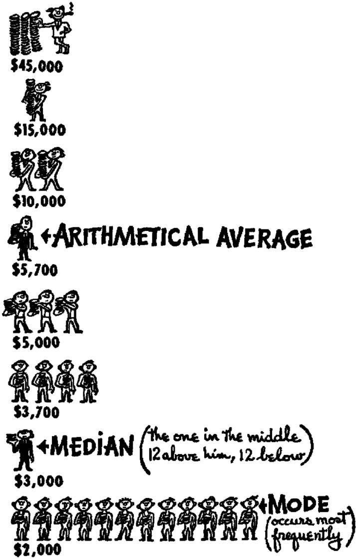
Оценки центральной тенденции

Выборочное среднее — среднее арифметическое по выборке.
Выборочная медиана — центральный элемент вариационного ряда.
Выборочная мода — самое распространённое значение в выборке.



Оценки центральной тенденции

(Huff, 1954):



Точечные оценки

Пусть распределение генеральной совокупности параметрическое:

$$F(x) = F(x, \theta).$$

Статистика $\hat{\theta}_n = \hat{\theta}\left(X^n\right)$ — точечная оценка параметра θ.
Какая оценка лучше?

Состоятельность: $\underset{n\to\infty}{\operatorname{plim}} \hat{\theta}_n = \theta$.
Несмещённость: $\mathbb{E}\hat{\theta}_n = \theta$.
Асимптотическая несмещённость: $\lim\limits_{n\to\infty} \mathbb{E}\hat{\theta}_n = \theta$.
Оптимальность: $\mathbb{D}\hat{\theta}_n = \min\limits_{\hat{\theta}\,:\,\mathbb{E}\hat{\theta}=\theta} \mathbb{D}\hat{\theta}$.
Робастность: устойчивость $\hat{\theta}_n$ относительно

- отклонений истинного распределения X от модельного семейства
- выбросов, содержащихся в выборке

Метод максимума правдоподобия

Популярный метод получения точечных оценок:

$$
\begin{aligned}
X &\sim f(x,\theta),\\
X^n &= (X_1,\dots,X_n),\\
L(X^n,\theta) &= \prod_{i=1}^{n} f(X_i,\theta),\\
\hat{\theta}_{MLE} &\equiv \operatorname*{argmax}_{\theta} L(X^n,\theta).
\end{aligned}
$$

Удобно прологарифмировать:

$$
\begin{aligned}
\log L(X^n,\theta) &= \sum_{i=1}^{n} f(X_i,\theta),\\
\hat{\theta}_{MLE} &\equiv \operatorname*{argmax}_{\theta} \log L(X^n,\theta).
\end{aligned}
$$

Производные функции правдоподобия

Score function:

$$S(\theta) \equiv \frac{\partial}{\partial\theta} \log L(\theta)$$

ОМП — решение score equation:

$$S(\theta) = 0$$

Информация Фишера:

$$I(\theta) \equiv -\frac{\partial^2}{\partial\theta^2} \log L(\theta)$$

Дисперсия ОМП:

$$\mathbb{D}\hat{\theta}_{MLE} \approx I^{-1}\left(\hat{\theta}_{MLE}\right)$$

Свойства ОМП

- состоятельность:
$$\operatorname*{plim}_{n\to\infty} \hat{\theta}_{MLE} = \theta$$
- асимптотическая нормальность: при $n \to \infty$
$$\hat{\theta}_{MLE} \sim N\left(\theta, I^{-1}(\theta)\right)$$
- эффективность: ОМП имеют наименьшую дисперсию среди всех состоятельных оценок
- инвариантность: $g\left(\hat{\theta}_{MLE}\right)$ — ОМП-оценка для $g\left(\theta\right)$

Интервальные оценки

Доверительный интервал:

$$\mathbf{P}(\theta \in [C_L, C_U]) \geqslant 1 - \alpha,$$

$1 - \alpha$ — уровень доверия,
C_L, C_U — нижний и верхний доверительные пределы.

Неверная интерпретация: неизвестный параметр лежит в пределах построенного доверительного интервала с вероятностью $1 - \alpha$.

Верная интерпретация: при бесконечном повторении процедуры построения доверительного интервала на аналогичных выборках в $100(1 - \alpha)\%$ случаев он будет содержать истинное значение θ.

Для нормального распределения

$$X \sim N\left(\mu, \sigma^2\right), \quad X^n = \left(X_1, \dots, X_n\right),$$

$\bar{X}_n$ — оценка $\mathbb{E}X = \mu$,

ЦПТ: $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow$

$$\mathbf{P}\left(\mu - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leqslant \bar{X}_n \leqslant \mu + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \Rightarrow$$

доверительный интервал для μ:

$$\mathbf{P}\left(\bar{X}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha,$$

$z_{1-\frac{\alpha}{2}}$ — квантиль стандартного нормального распределения.

Для ненормальных распределений

ЦПТ: если X^n — выборка из $F(x)$, $F(x)$ не слишком скошено и $n > 30$, то

$$\bar{X}_n \sim\approx N\left(\mathbb{E}X, \frac{\mathbb{D}X}{n}\right) \Rightarrow$$

доверительный интервал для $\mathbb{E}X$:

$$\mathbf{P}\left(\bar{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbb{D}X}{n}} \leqslant \mathbb{E}X \leqslant \bar{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbb{D}X}{n}}\right) \approx 1 - \alpha.$$

Если дисперсия неизвестна:

$$\mathbf{P}\left(\bar{X}_n - t_{n-1,1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}} \leqslant \mathbb{E}X \leqslant \bar{X}_n + t_{n-1,1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}\right) \approx 1 - \alpha,$$

$t_{n-1,1-\frac{\alpha}{2}}$ — квантиль распределения Стьюдента с $n-1$ степенью свободы.

Квантили

Непараметрический доверительный интервал для медианы непрерывного распределения.

$$X^n = (X_1, \dots, X_n), \quad X \sim F(x) \Rightarrow$$

$$\mathbf{P}\left(\operatorname{med} X \in \left[X_{(r)}, X_{(n-r+1)}\right]\right) = \frac{1}{2^n} \sum_{i=r}^{n-r+1} C_n^i.$$

При $n > 10$ применима нормальная аппроксимация:

$$\mathbf{P}\left(\operatorname{med} X \in \left[X_{\left(\left\lfloor \frac{n-\sqrt{n} z_{1-\frac{\alpha}{2}}}{2} \right\rfloor\right)}, X_{\left(\left\lceil \frac{n+\sqrt{n} z_{1-\frac{\alpha}{2}}}{2} \right\rceil\right)}\right]\right) \approx 1 - \alpha.$$

Аналогично строится непараметрический доверительный интервал для любого квантиля $X_\alpha, \alpha \in (0,1)$:

$$\mathbf{P}\left(X_\alpha \in \left[X_{(l)}, X_{(u)}\right]\right) = \sum_{i=l}^{u} C_n^i \alpha^i (1-\alpha)^{n-i}.$$

Проверка гипотез

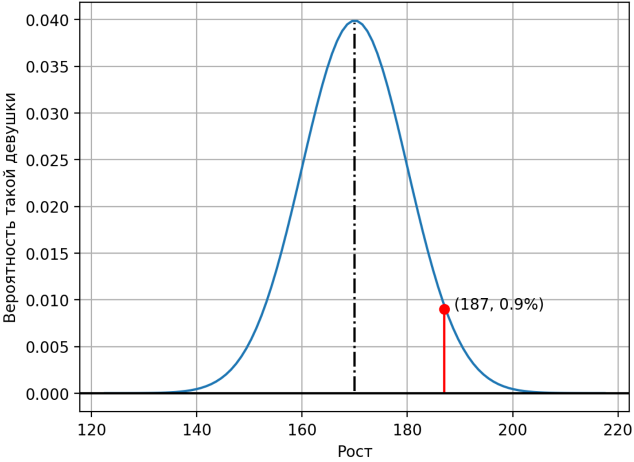
Основной инструмент прикладной статистики – проверка статистических гипотез. Для этого мы делаем следующее:

1. Обозначаем нашу генеральную совокупность.
2. Выдвигаем нулевую и альтернативную гипотезы.
3. Генерируем выборку.
4. Проверяем нашу гипотезу, используя тот или иной статистический метод.
5. Отвергаем нулевую гипотезу в пользу альтернативной, либо НЕ отвергаем нулевую гипотезу (как вариант – придерживаемся нулевой гипотезы).

В большинстве случаев альтернативной гипотезой выбирается та, К КОТОРОЙ мы хотим прийти в конечном итоге.
Пример:
H_0 - мы не смогли научить студентов ПСАД чему-то новому
H_1 - наш курс был полезен для студентов

Проверка гипотез

Гипотеза: рост девушек факультета ВМК описывается нормальным распределением со средним значением 170см и дисперсией 10см.



Проверка гипотез

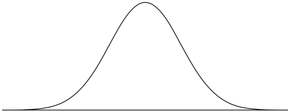
выборка: $X^n = (X_1, \ldots, X_n)\,, \; X \sim \mathbf{P} \in \Omega$

нулевая гипотеза: $H_0\colon \; \mathbf{P} \in \omega, \quad \omega \in \Omega$

альтернатива: $H_1\colon \; \mathbf{P} \notin \omega$

статистика: $T\left(X^n\right), \quad T\left(X^n\right) \sim F(x)$ при $\mathbf{P} \in \omega$

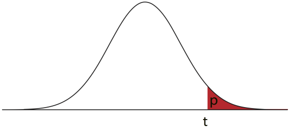
$T\left(X^n\right) \not\sim F(x)$ при $\mathbf{P} \notin \omega$



реализация выборки: $x^n = (x_1, \ldots, x_n)$

реализация статистики: $t = T\left(x^n\right)$

достигаемый уровень значимости: $p\left(x^n\right)$ — вероятность при H_0 получить $T\left(X^n\right) = t$ или ещё более экстремальное



$$p\left(x^n\right) = \mathbf{P}(T \geqslant t \,|\, H_0)$$

Гипотеза отвергается при $p\left(x^n\right) \leqslant \alpha, \quad \alpha$ — уровень значимости

Эквивалентность

Часто можно найти следующее определение p-value:
Для правосторонней альтернативы:

$$p_r\left(x^n\right) = \mathbf{P}(T \geqslant t\,|H_0\,)$$

Для левосторонней альтернативы:

$$p_l\left(x^n\right) = \mathbf{P}(T \leqslant t\,|H_0\,)$$

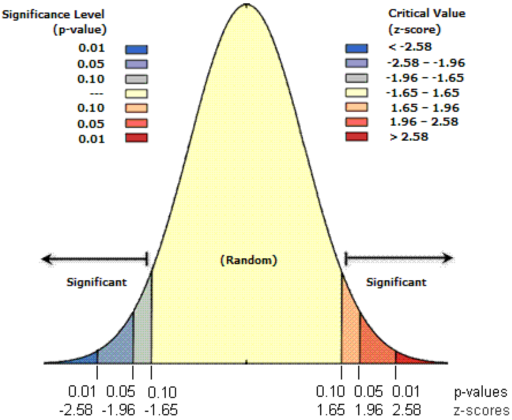
Для двусторонней альтернативы:

$$p\left(x^n\right) = 2 * \min\{p_l, p_r\}$$

Таким образом, p-value – это минимальный уровень доверия, при котором мы бы опровергли нулевую гипотезу.
Нельзя рассматривать p-value как вероятность того, что наша нулевая гипотеза верна или нет!

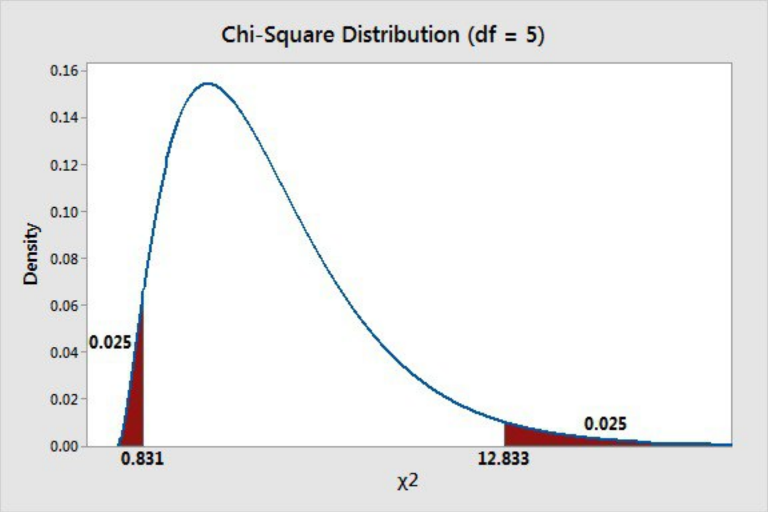
Эквивалентность

Проверка гипотез через p-value и критические значения, по сути, эквивалентны.



Эквивалентность

Для несимметричных непрерывних распределений критические области также равны областям размера $\alpha/2$.



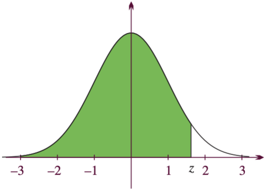
p-значение для дискретных распределений

Для дискретных распределений значение p-value для двусторонней альтернативы часто считается по иной формуле!
Пусть X – множество всех возможных значений статистики T.
Наблюдаемое значение статиктики равно t. Пусть
$\mathcal{I} = \{x \in X : \mathbf{P}(x|H_0) \leqslant \mathbf{P}(t|H_0)\}$.
Тогда p-value для двусторонней альтернативы считается как

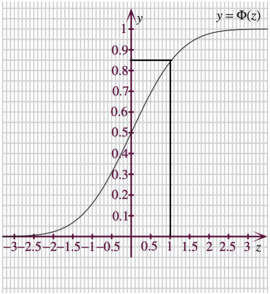
$$p(x^n) = \sum_{i \in \mathcal{I}} \mathbf{P}(i|H_0).$$

Задача

$$\Phi(z) = \int_{-\infty}^{z} \phi(t)\, dt.$$



This is the graph of the standard normal probability density function $\phi(z)$.



This is the graph of the standard normal cumulative distribution function $\Phi(z)$.

Какая из альтернатив рассматривается на графике слева: левосторонняя, правосторонняя и двусторонняя?

Пусть значение статистики равно 1. Чему равно p-value при каждой из альтернатив?

Таблица

Студенты писали тест по курсу глубокого обучения. В тесте было 15 вопросов, каждый вопрос содержал 3 варианта ответа, один из которых был верным.

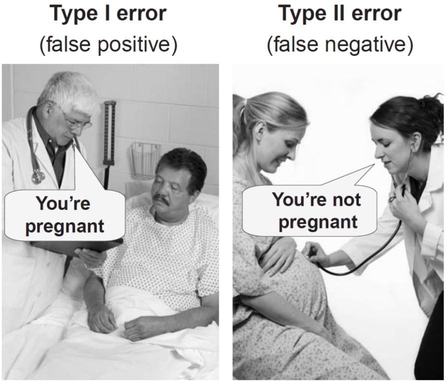
1. Студент A правильно ответил на 1 вопрос из 15. Посчитайте p-value для данного числа при нулевой гипотезе о том, что он ставил ответы случайно, против альтернативы о неслучайности его ответов.
2. Студент B правильно ответил на 8 вопросов из 15. Посчитайте p-value для данного числа при нулевой гипотезе о том, что он ставил ответы случайно против альтернативы о том, что он хорошо прослушал курс и знал ответы на вопросы.

Таблица

Число успехов	Вероятность	Число успехов	Вероятность	Число успехов	Вероятность
0	0.0023	*6*	0.1786	*12*	0.0003
1	0.0171	*7*	0.1148	*13*	0.0000
2	0.0599	*8*	0.0574	*14*	0.0000
3	0.1299	*9*	0.0223	*15*	0.0000
4	0.1948	*10*	0.0067		
5	0.2143	*11*	0.0015		

Ошибки I и II рода

	H_0 верна	H_0 неверна
H_0 принимается	H_0 верно принята	Ошибка второго рода (False negative)
H_0 отвергается	Ошибка первого рода (False positive)	H_0 верно отвергнута



Ошибки I и II рода

Задача проверки гипотез несимметрична относительно пары (H_0, H_1): вероятность ошибки первого рода ограничивается сверху величиной α, а второго рода — минимизируется путём выбора критерия.

Корректный критерий: $\mathbf{P}(p\,(T) \leqslant \alpha \,|H_0\,) \leqslant \alpha \forall\, \mathbf{P} \in \Omega$.
Мощность: $\mathrm{pow} = \mathbf{P}(p\,(T) \leqslant \alpha \,|H_1\,)$.
Состоятельный критерий: $\mathrm{pow} \to 1$ для всех альтернатив H_1 при $n \to \infty$.
T_1 — **равномерно наиболее мощный** критерий, если $\forall T_2$

$$\mathbf{P}(p\,(T_1) \leqslant \alpha \,|H_1\,) \geqslant \mathbf{P}(p\,(T_2) \leqslant \alpha \,|H_1\,) \quad \forall H_1 \neq H_0,$$
$$\mathbf{P}(p\,(T_1) \leqslant \alpha \,|H_0\,) = \mathbf{P}(p\,(T_2) \leqslant \alpha \,|H_0\,)\,,$$

причём хотя бы для одной H_1 неравенство строгое.

Интерпретация результата

Если величина p достаточно мала, то данные свидетельствуют против нулевой гипотезы в пользу альтернативы.
Если величина p недостаточно мала, то данные не свидетельствуют против нулевой гипотезы в пользу альтернативы.

При помощи инструмента проверки гипотез нельзя доказать справедливость нулевой гипотезы!
Absence of evidence $\not\Rightarrow$ evidence of absence.

Статистическая и практическая значимость

Вероятность отвергнуть нулевую гипотезу зависит не только от того, насколько она отличается от истины, но и от размера выборки.
По мере увеличения n нулевая гипотеза может сначала приниматься, но потом выявятся более тонкие несоответствия выборки гипотезе H_0, и она будет отвергнута.

При любой проверке гипотез нужно оценивать **размер эффекта** — степень отличия нулевой гипотезы отличается от истины, и оценивать его практическую значимость.

Статистическая и практическая значимость

- (Lee et al, 2010): за три года женщины, упражнявшиеся не меньше часа в день, набрали значимо меньше веса, чем женщины, упражнявшиеся меньше 20 минут в день ($p < 0.001$). Разница в набранном весе составила 150 г. Практическая значимость такого эффекта сомнительна. Подробности: `http://youtu.be/oqDZ0-mfN4Q`.
- (Ellis, 2010, гл. 2): в 2002 году клинические испытания гормонального препарата Премарин, облегчающего симптомы менопаузы, были досрочно прерваны. Было обнаружено, что его приём ведёт к значимому увеличению риска развития рака груди на 0.08%, риска инсульта на 0.08% и инфаркта на 0.07%. Формально эффект крайне мал, но с учётом численности населения он превращается в тысячи дополнительных смертей.
- (Kirk, 1996): если при испытании гипотетического лекарства, позволяющего замедлить прогресс ослабления интеллекта больных Альцгеймером, оказывается, что разница в IQ контрольной и тестовой групп составляет 13 пунктов, возможно, изучение лекарства стоит продолжить, даже если эта разница статистически незначима.

Другие особенности

- Выбранная статистика может отражать не всю информацию, содержащуюся в выборке. Пример:

$$H_0\colon X \sim N\left(\mu, \sigma^2\right), \quad H_1\colon H_0 \text{ неверна};$$

$$T\left(X^n\right) = g_1.$$

Все симметричные распределения будут признаны нормальными!

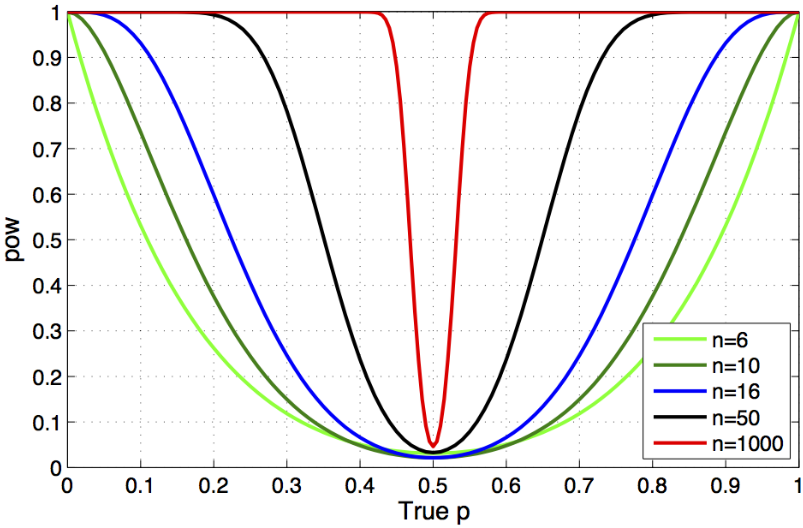
- Гипотезы вида $H_0\colon \theta = \theta_0$ можно проверять при помощи доверительных интервалов для θ:
  - если θ_0 не попадает в $100\left(1-\alpha\right)\%$ доверительный интервал для θ, то H_0 отвергается на уровне значимости α;
  - p-value — максимальное α, при котором θ_0 попадает в соответствующий доверительный интервал.

Мощность

Мощность критерия зависит от следующих факторов:

- размер выборки
- размер отклонения от нулевой гипотезы
- чувствительность статистики критерия
- тип альтернативы

Мощность



Размер выборки

Обеспечение требуемой мощности: размеры выборки подбирается так, чтобы при размере отклонения от нулевой гипотезы не меньше заданного (например, вероятность выбора взболтанного мартини не меньше 0.75) мощность была не меньше заданной.

Литература

Справочники по статистике:

- Кобзарь А.И. *Прикладная математическая статистика*, 2006.
- Kanji G.K. *100 statistical tests*, 2006.

Вводные учебники по статистике:

- Good P.I., Hardin J.W. *Common Errors in Statistics (and How to Avoid Them)*, 2003.
- Reinhart A. *Statistics Done Wrong. The woefully complete guide*, `http://www.statisticsdonewrong.com/`

Литература

Правдоподобие:

- Chihara L., Hesterberg T. *Mathematical Statistics with Resampling and R*, 2011, глава 6.
- Pawitan, Y. *In all likelihood: statistical modelling and inference using likelihood*, 2001.

Бутстреп:

- Hesterberg T., Monaghan S., Moore D.S., Clipson A., Epstein R. *Bootstrap methods and permutation tests*. In Introduction to the Practice of Statistics, 2005. `http://statweb.stanford.edu/~tibs/stat315a/Supplements/bootstrap.pdf`
- Efron B., Tibshirani R. *An Introduction to the Bootstrap*, 1993.

Проверка гипотез:

- Good P.I., Hardin J.W. *Common Errors in Statistics (and How to Avoid Them)*, 2003, глава 2.

Литература

Anscombe F.J. (1973). *Graphs in Statistical Analysis*. American Statistician, 27(1): 17–21.

Begg I.M., Anas A., Farinacci S. (1992). *Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth*. Journal of Experimental Psychology: General, 121(4), 446–458.

Ellis P.D. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*, 2010.

Huff D. *How To Lie With Statistics*, 1954.

Kirk R.E. (1996). *Practical Significance: A Concept Whose Time Has Come*. Educational and Psychological Measurement, 56(5), 746–759.

Lee I.-M., Djoussè L., Sesso H.D., Wang L., Buring J.E. (2010). *Physical Activity and Weight Gain Prevention*. JAMA: the Journal of the American Medical Association, 303(12), 1173–1179.

Marriott, F. H. C. *The Interpretation of Multiple Observations*, 1974.