Прикладная статистика 1. Введение.

Рябенко Евгений
riabenko.e@gmail.com

10 февраля 2014 г.

Выборка

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

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

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

n — **объём выборки**.

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

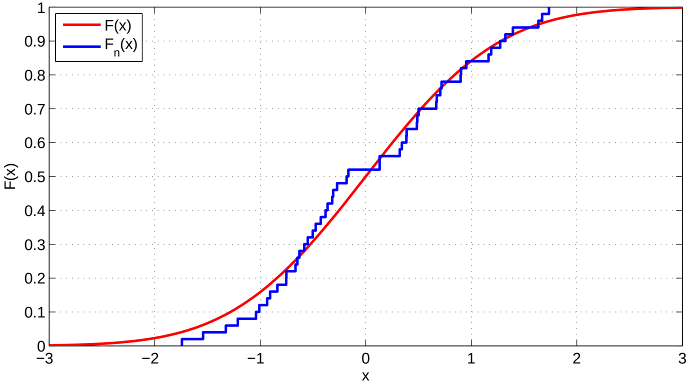
Пусть $F(x)$ — функция распределения элемента простой выборки:

$$F(x) = P\left(X_i \leq x\right).$$

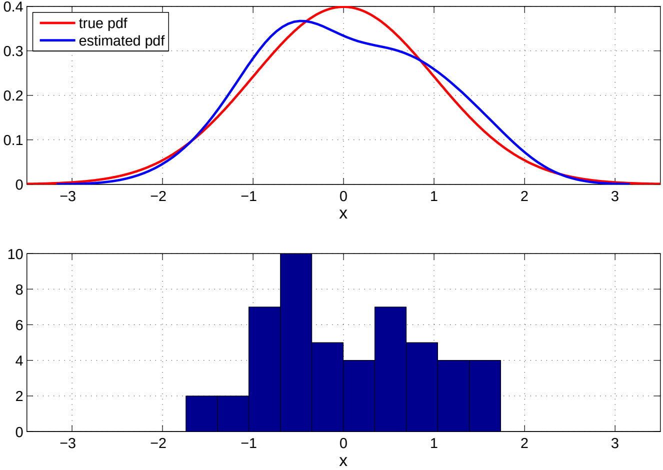
Основная задача статистики — описание $F(x)$ по реализации выборки.

Функция распределения

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



Плотность распределения



Статистика

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

Примеры:

- **выборочное среднее:**

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

- **выборочная дисперсия:**

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

- **несмещённая выборочная дисперсия:**

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

Статистика

Вариационный ряд:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

- k-я **порядковая статистика:** $X_{(k)}$;

Квантиль порядка $\alpha \in (0,1)$ случайной величины X:

$$X_\alpha : \quad P\left(X < X_\alpha\right) \leq \alpha, \ P\left(X \leq X_\alpha\right) \geq \alpha.$$

- **выборочный** α**-квантиль:** $X_{([n\alpha])}$;
- **выборочная медиана:**

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

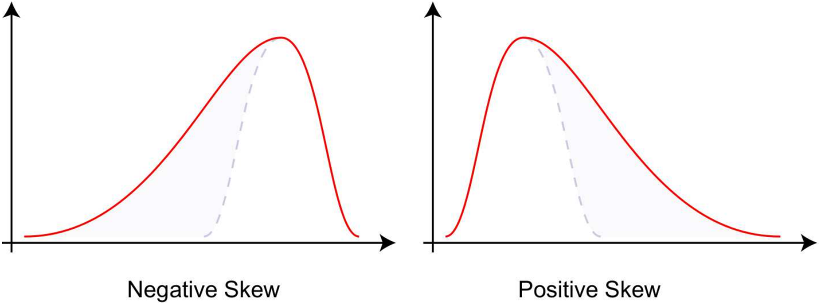
- **выборочный интерквартильный размах:**

$$IQR = X_{([3n/4])} - X_{([n/4])}.$$

Статистика

Коэффициент ассиметрии (skewness):

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



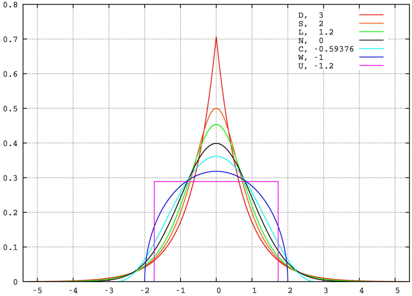
- **выборочный коэффициент ассиметрии:**

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

Статистика

Коэффициент эксцесса (kurtosis):

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

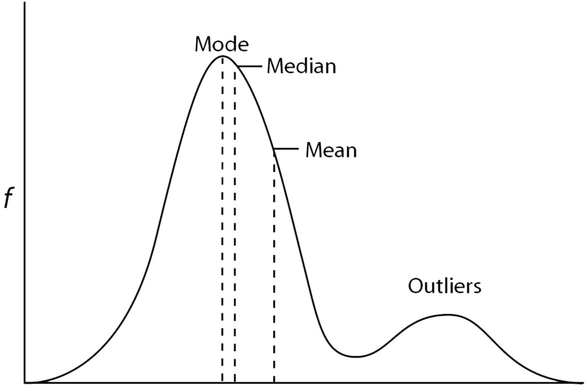


- **выборочный коэффициент эксцесса:**

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

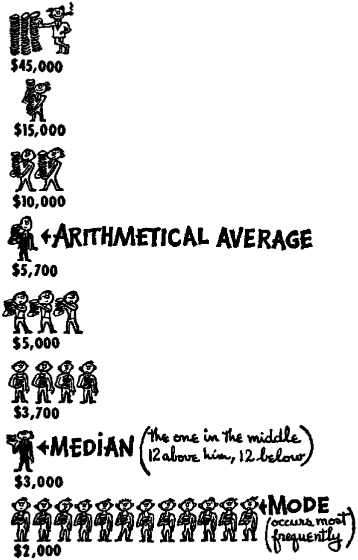
Оценки центральной тенденции

Выборочное среднее — среднее значение в выборке.
Медиана — центральный элемент вариационного ряда.
Мода — самое распространённое значение в выборке.

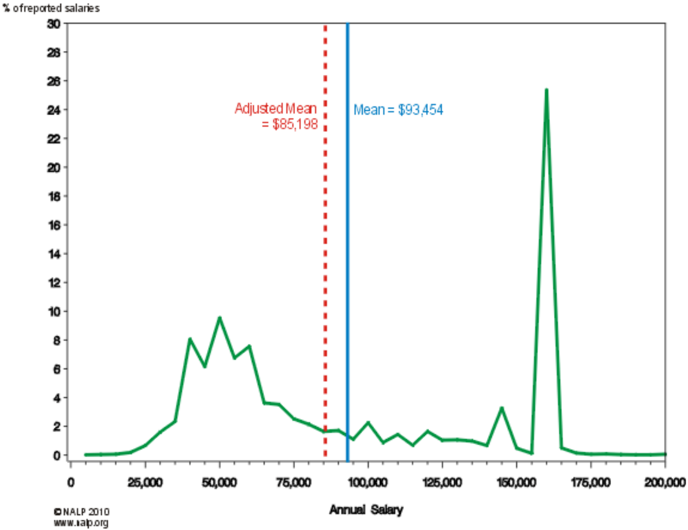


Оценки центральной тенденции

How to lie with statistics (Huff, 1954)



Оценки центральной тенденции



Salary Distribution Curve for the class of 2009 Shows Relatively Few Salaries Were Close to the Mean (NALP data, 2009)

Точечные оценки

Пусть распределение генеральной совокупности параметрическое:

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

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

Состоятельность: $\operatorname*{plim}_{n\to\infty} \hat{\theta} = \theta$.
Несмещённость: $\mathbb{E}\hat{\theta} = \theta$.
Асимптотическая несмещённость: $\lim_{n\to\infty} \mathbb{E}\hat{\theta} = \theta$.
Оптимальность: $\mathbb{D}\hat{\theta}^* = \min_{\hat{\theta}:\ \mathbb{E}\hat{\theta}=\theta} \mathbb{D}\hat{\theta}$.
Робастность: устойчивость $\hat{\theta}$ относительно малых отклонений реального распределения X от $F(X, \theta)$ (в частности, относительно наличия выбросов).

Интервальные оценки

Оценим параметр θ двумя статистиками:

$$P\left(\theta \in [C_L, C_U]\right) \geq 1 - \alpha,$$

α — уровень доверия, C_L, C_U — верхний и нижний 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 параметра.

Интервальные оценки

Пример 1: доверительный интервал для среднего $X \sim N\left(\mu, \sigma^2\right)$ при известной дисперсии σ^2.

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

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \sqrt{n}\frac{\bar{X} - \mu}{\sigma} \sim N(0,1) \Rightarrow$$

$$P\left(\mu \in \left[\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]\right) = 1 - \alpha.$$

$z_{1-\frac{\alpha}{2}}$ — $\left(1 - \frac{\alpha}{2}\right)$-квантиль стандартного нормального распределения; при $\alpha = 0.05$ получаем $z_{0.975} \approx 1.96$.

Правило двух сигм: если $X \sim N\left(\mu, \sigma^2\right)$, то $P\left(|X - \mu| \le 2\sigma\right) \approx 0.954$.

Если X распределена не нормально, то можно утверждать только $P\left(|X - \mathbb{E}X| \le 2\mathbb{D}X\right) \ge 0.75$ (из неравенства Чебышева).

Интервальные оценки

Пример 2: непараметрический доверительный интервал для медианы.

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

$$P\left(\operatorname{med} X_i \in \left[X_{(l)}, X_{(u)}\right]\right) = \frac{1}{2^n} \sum_{i=l}^{u} C_n^i.$$

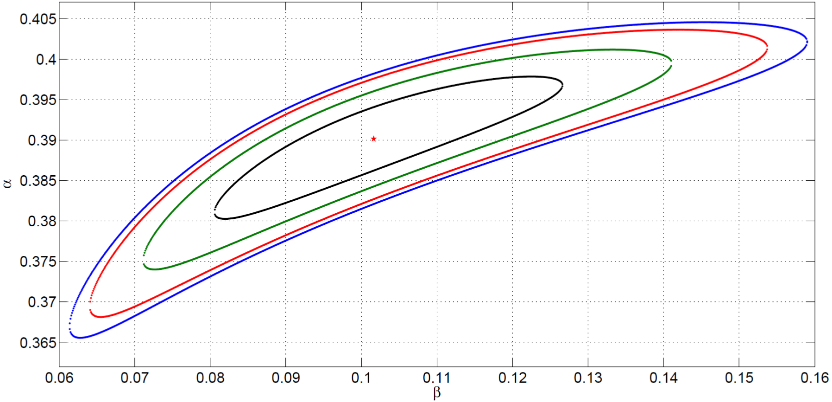
Чтобы построить как можно более узкий доверительный интервал, l и u выбираются так, чтобы соответствующие им слагаемые были как можно больше.

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

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

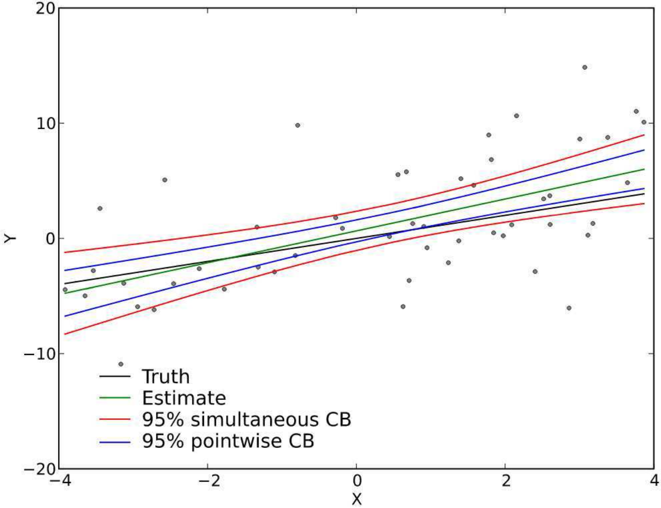
Интервальные оценки

Доверительная область для пары неизвестных параметров (α, β):



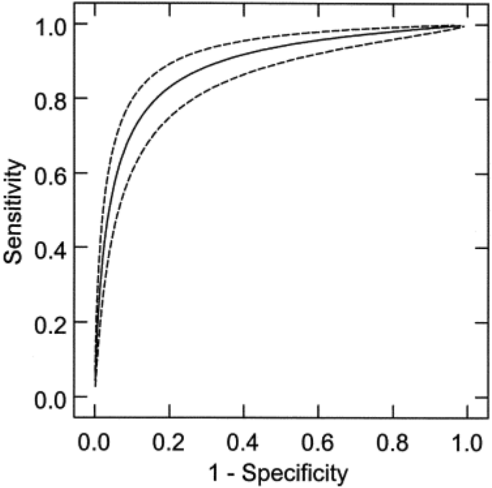
Интервальные оценки

Доверительная лента для функции $Y = \beta_0 + \beta_1 x$:



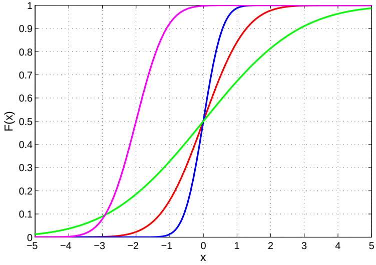
Интервальные оценки

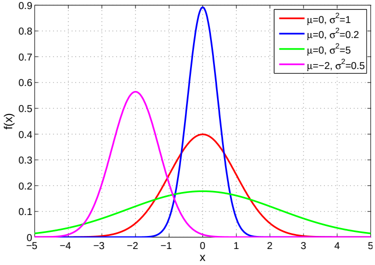
Доверительная лента для ROC-кривой:



Нормальное распределение

$X \in \mathbb{R} \sim N\left(\mu, \sigma^2\right)$, $\sigma^2 > 0$ — предельное распределение суммы слабо взаимозависимых сл. в.





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

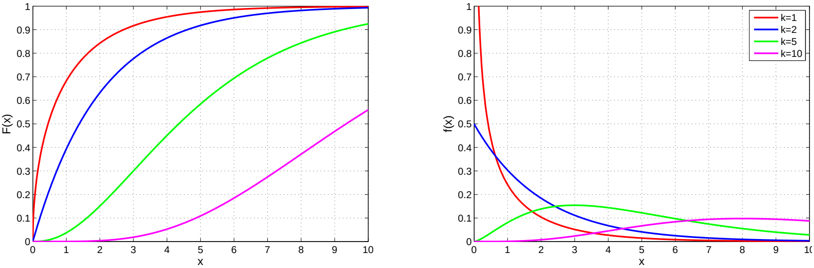
$$f(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right), \qquad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Нормальное распределение

$$\begin{aligned}
\mathbb{E}X &= \mu, \\
\operatorname{med} X &= \mu, \\
\operatorname{mode} X &= \mu, \\
\mathbb{D}X &= \sigma^2, \\
\gamma_1(X) &= 0, \\
\gamma_2(X) &= 0.
\end{aligned}$$

Распределение хи-квадрат

$X \in \mathbb{R}_+ \sim \chi^2_k,\ k \in \mathbb{N}$ — распределение суммы квадратов k независимых стандартных нормальных сл. в.



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

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

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

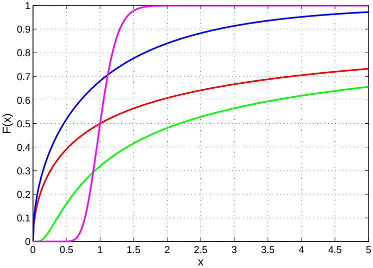
Распределение хи-квадрат

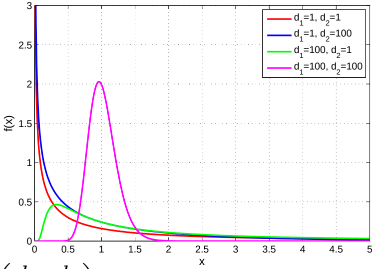
$$
\begin{aligned}
\mathbb{E}X &= k,\\
\operatorname{med} X &\approx k\left(1-\frac{2}{9k}\right)^3,\\
\operatorname{mode} X &= \max\left(k-2,0\right),\\
\mathbb{D}X &= 2k,\\
\gamma_1\left(X\right) &= \sqrt{8/k},\\
\gamma_2\left(X\right) &= 12/k.
\end{aligned}
$$

- Пусть $X^k = X_1,\dots,X_k, \quad X_i \sim N\left(0,1\right)$, тогда
$$\sum_{i=1}^{k} X_i^2 \sim \chi_k^2.$$

Распределение Фишера

$X \in \mathbb{R}_+ \sim F(d_1, d_2)\,,\ d_1, d_2 > 0$ — распределение отношения двух независимых нормированных хи-квадрат сл. в.





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

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

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

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

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

Распределение Фишера

$$\mathbb{E}X = \frac{d_2}{d_2 - 2} \text{ при } d_2 > 2,$$

$$\operatorname{mode} X = \frac{d_1 - 2}{d_1} \frac{d_2}{d_2 + 2} \text{ при } d_1 > 2,$$

$$\mathbb{D}X = \frac{2d_2^2 (d_1 + d_2 - 2)}{d_1 (d_2 - 2)^2 (d_2 - 4)} \text{ при } d_2 > 4,$$

$$\gamma_1 (X) = \frac{(2d_1 + d_2 - 2) \sqrt{8 (d_2 - 4)}}{(d_2 - 6) \sqrt{d_1 (d_1 + d_2 - 2)}} \text{ при } d_2 > 6,$$

$$\gamma_2 (X) = 12 \frac{d_1(5d_2 - 22)(d_1 + d_2 - 2) + (d_2 - 4)(d_2 - 2)^2}{d_1(d_2 - 6)(d_2 - 8)(d_1 + d_2 - 2)} \text{ при } d_2 > 8.$$

Распределение Фишера

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

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

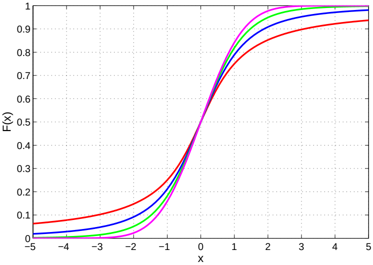
- Если $X \sim F\left(d_1, d_2\right)$, то

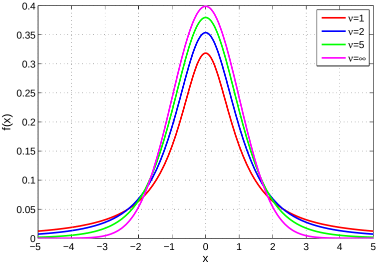
$$Y = \lim_{d_2 \to \infty} d_1 X \sim \chi^2_{d_1}.$$

- $F\left(x, d_1, d_2\right) = F\left(1/x, d_2, d_1\right).$

Распределение Стьюдента

$X \in \mathbb{R} \sim St\left(\nu\right), \; \nu > 0$ —распределение отношения независимых стандартной нормальной сл. в. и корня из нормированной хи-квадрат сл. в.





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

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

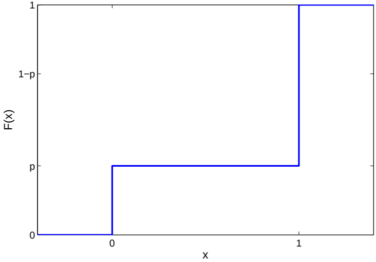
Распределение Стьюдента

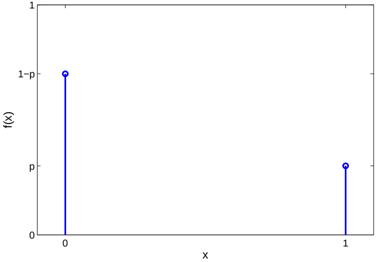
$$\begin{aligned}
\mathbb{E}X &= 0 \text{ при } \nu > 1,\\
\operatorname{med} X &= 0,\\
\operatorname{mode} X &= 0,\\
\mathbb{D}X &= \begin{cases} \frac{\nu}{\nu-2}, & \nu > 2,\\ \infty, & 1 < \nu \le 2, \end{cases},\\
\gamma_1(X) &= 0 \text{ при } \nu > 3,\\
\gamma_2(X) &= \begin{cases} \frac{6}{\nu-4}, & \nu > 4,\\ \infty, & 2 < \nu \le 4. \end{cases}.
\end{aligned}$$

- Пусть $Z \sim N(0,1)$, $\quad V \sim \chi^2_\nu$, тогда
$$T = \frac{Z}{\sqrt{V/\nu}} \sim St(\nu).$$
- Если $X \sim St(\nu)$, то
$$Y = \lim_{\nu\to\infty} X \sim N(0,1).$$

Распределение Бернулли

$X \in \{0,1\} \sim Ber\left(p\right), \;\; p \in (0,1)$ — распределение, моделирующее испытание Бернулли.





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

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

Распределение Бернулли

$$\mathbb{E}X = p,$$

$$\operatorname{med} X = \begin{cases} 0, & 1-p > p, \\ 0.5, & 1-p = p, \\ 1, & 1-p < p, \end{cases}$$

$$\operatorname{mode} X = \begin{cases} 0, & 1-p > p, \\ \{0,1\}, & 1-p = p, \\ 1, & 1-p < p, \end{cases}$$

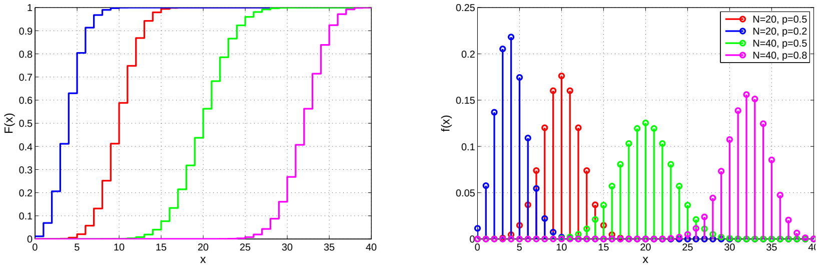
$$\mathbb{D}X = p\left(1-p\right),$$

$$\gamma_1\left(X\right) = \frac{1-2p}{\sqrt{p\left(1-p\right)}},$$

$$\gamma_2\left(X\right) = \frac{1-6p\left(1-p\right)}{p\left(1-p\right)}.$$

Биномиальное распределение

$X \in \{0, \ldots, N\} \sim Bin\,(N, p)\,,\ N \in \mathbb{N},\ \ p \in [0, 1]$ — распределение числа успехов в N независимых испытаниях Бернулли.



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

Биномиальное распределение

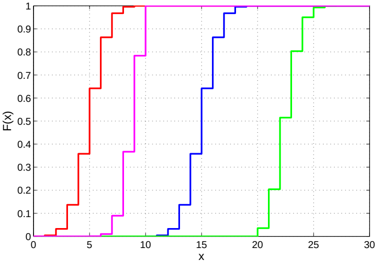
$$\begin{aligned}
\mathbb{E}X &= Np,\\
\operatorname{med} X &= \lfloor Np \rfloor \text{ или } \lceil Np \rceil,\\
\operatorname{mode} X &= \lfloor (N+1)\,p \rfloor \text{ или } \lceil (N+1)\,p \rceil - 1,\\
\mathbb{D}X &= Np\,(1-p)\,,\\
\gamma_1\,(X) &= \frac{1-2p}{\sqrt{Np\,(1-p)}},\\
\gamma_2\,(X) &= \frac{1-Np\,(1-p)}{Np\,(1-p)}.
\end{aligned}$$

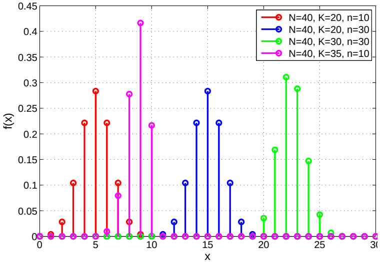
- $X \sim Bin\,(1,p) \Leftrightarrow X \sim Ber\,(p)\,.$
- Если $N > 20$ и p не слишком близко к нулю или единице, то для $X \sim Bin\,(N,p)$ справедливо приближение

$$F_X\,(x) \approx \Phi\left(\frac{x-Np}{\sqrt{Np\,(1-p)}}\right).$$

Гипергеометрическое распределение

$X \sim Hyp(K, N, n)$, $N \in \mathbb{N}_0$, $K, n \in \{0, \dots, N\}$, $X \in \{\max(0, n+K-N), \dots, \min(K, n)\}$ — распределение числа успехов в выборке без возвращения размера n из популяции N с общим числом успехов K.





$$F(x) = \sum_{i=1}^{x} \frac{C_K^i C_{N-K}^{n-i}}{C_N^n},$$

$$f(x) = \frac{C_K^x C_{N-K}^{n-x}}{C_N^n}.$$

Гипергеометрическое распределение

$$\mathbb{E}X = \frac{nK}{N},$$

$$\operatorname{mode} X = \left\lfloor \frac{(n+1)(N+1)}{N+2} \right\rfloor,$$

$$\mathbb{D}X = \frac{nK(N-K)(N-n)}{N^2(N-1)},$$

$$\gamma_1(X) = \frac{(N-2K)(N-2n)\sqrt{N-1}}{(N-2)\sqrt{nK(N-K)(N-n)}},$$

$$\gamma_2(X) = \frac{((N-1)N^2(N(N+1)-6K(N-K)-6n(N-n))+}{nK(N-K)(N-n)(N-2)(N-3)}$$
$$\underline{+6nK(N-K)(N-n)(5N-6))}.$$

Гипергеометрическое распределение

- $X \sim Hyp\,(K, N, 1) \Leftrightarrow X \sim Ber\left(\frac{K}{N}\right)$.
- Пусть $X \sim Hyp\,(K, N, n)$, $Y \sim Bin\left(n, \frac{K}{N}\right)$; если $\frac{K}{N}$ не близко к нулю или единице, а N и K велики по сравнению с n и $\frac{K}{N}$, то

$$F_X\,(x) \approx F_Y\,(x) \approx \Phi\left(\frac{x - \frac{nK}{N}}{\sqrt{\frac{nK}{N}\left(1 - \frac{K}{N}\right)}}\right).$$

Проверка гипотез

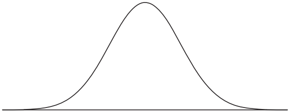
выборка: $X^n = (X_1, \dots, X_n) \sim P \in \Omega;$

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

альтернатива: $H_1 \colon P \notin \omega;$

статистика: $T(X^n), \quad T(X^n) \sim F(x)$ при $P \in \omega;$

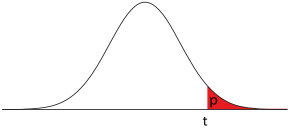
$T(X^n) \not\sim F(x)$ при $P \notin \omega;$



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

реализация статистики: $t = T(x^n);$

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



$$p(x^n) = P(T \geq t \,|H_0)$$

Гипотеза отвергается при $p(x^n) \leq \alpha, \quad \alpha$— уровень значимости.

Проверка гипотез

Ошибки I и II рода

Задача проверки гипотез несимметрична относительно пары (H_0, H_1).

	H_0 верна	H_0 неверна
H_0 принимается	H_0 верно принята	Ошибка второго рода
H_0 отвергается	Ошибка первого рода	H_0 верно отвергнута

Вероятность ошибки первого рода ограничивается малой величиной α.
Вероятность ошибки второго рода минимизируется путём выбора критерия.

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

$$P\left(p\left(T_1\right) \leq \alpha \,|H_1\right) \geq P\left(p\left(T_2\right) \leq \alpha \,|H_1\right) \quad \forall H_1 \neq H_0,$$
$$P\left(p\left(T_1\right) \leq \alpha \,|H_0\right) = P\left(p\left(T_2\right) \leq \alpha \,|H_0\right),$$

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.

Другие особенности

- По мере увеличения n нулевая гипотеза может сначала приниматься, но потом выявятся более тонкие несоответствия выборки гипотезе H_0, и она будет отвергнута (Statistical vs. clinical significance: http://youtu.be/oqDZO-mfN4Q).
- 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 $100\,(1-\alpha)\%$ доверительный интервал для θ, то H_0 отвергается на уровне значимости α.

Shaken, not stirred

Джеймс Бонд говорит, что предпочитает мартини смешанным, но не взболтанным. Проведём слепой тест: n раз предложим ему пару напитков и выясним, какой из двух он предпочитает.

Выборка: бинарный вектор длины n, 1 — Джеймс Бонд предпочёт смешанный, 0 — взболтанный.

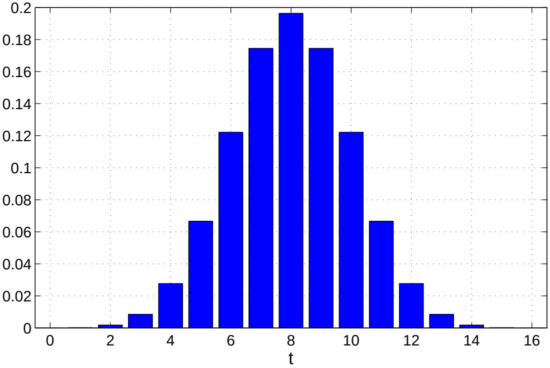
Нулевая гипотеза: Джеймс Бонд не различает два вида мартини, т. е., выбирает наугад.

Статистика t — число единиц в выборке.

Нулевое распределение

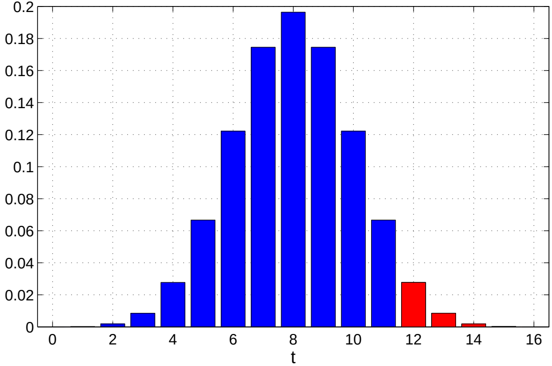
Если нулевая гипотеза справедлива и Джеймс Бонд не различает два вида мартини, то равновероятны все выборки длины n из нулей и единиц.

Пусть $n = 16$, тогда существует $2^{16} = 65536$ равновероятных варианта. Статистика t принимает значения от 0 до 16:



Односторонняя альтернатива

H_1: Джеймс Бонд предпочитает смешанный мартини.
При справедливости такой альтернативы более вероятны большие значения t (т.е., большие t свидетельствуют против H_0 в пользу H_1).
Вероятность того, что Джеймс Бонд предпочтёт смешанный мартини в 12 или более случаях из 16 при справедливости H_0, равна $\frac{2517}{65536} \approx 0.0384$.

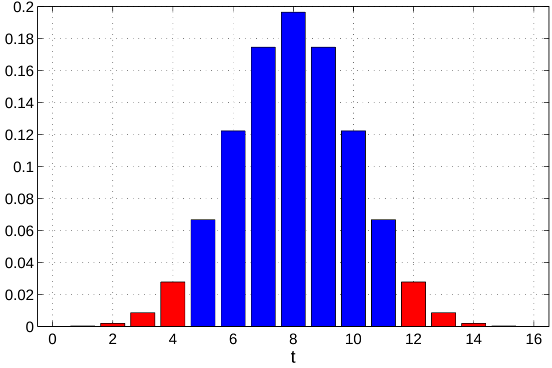


0.0384 — достигаемый уровень значимости при реализации $t = 12$.

Двусторонняя альтернатива

H_1: Джеймс Бонд предпочитает какой-то определённый вид мартини. При справедливости такой альтернативы и очень большие, и очень маленькие значения t свидетельствуют против H_0 в пользу H_1).

Вероятность того, что Джеймс Бонд предпочтёт смешанный мартини в 12 или более случаях из 16 при справедливости H_0, равна $\frac{5034}{65536} \approx 0.0768$.



0.0768 — достигаемый уровень значимости при реализации $t = 12$.

Достигаемый уровень значимости

Чем ниже достигаемый уровень значимости, тем сильнее данные свидетельствуют против нулевой гипотезы в пользу альтернативы.

0.0384 — вероятность реализации $t \geq 12$ при условии, что нулевая гипотеза справедлива, т. е. Джеймс Бонд выбирает мартини наугад.

Достигаемый уровень значимости нельзя интерпретировать как вероятность справедливости нулевой гипотезы!

Достигаемый уровень значимости

Пример: утверждается, что осьминог предсказывает результаты матчей чемпионата мира по футболу с участием сборной Германии, выбирая кормушку с флагом страны-победителя. По результатам 13 испытаний ему удаётся верно угадать результаты 11 матчей. Аналогичный предыдущему критерий даёт достигаемый уровень значимости $p \approx 0.0112$.

0.0112 — не вероятность того, что осьминог выбирает кормушку наугад! Эта вероятность равна единице.

$$p = P\left(T \geq t \,|\, H_0\right) \neq P\left(H_0 \,|\, T \geq t\right).$$

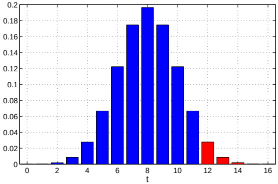
Достигаемый уровень значимости

Пример: пусть Джеймс Бонд выбирает смешанный мартини в 51% случаев (ненаблюдаемая вероятность).

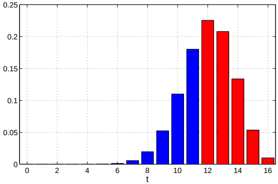
Пусть по итогам 100 испытаний смешанный мартини был выбран 49 раз. Достигаемый уровень значимости против односторонней альтернативы — $p \approx 0.6178$. Нулевая гипотеза не отвергается, при этом сказать, что она верна, было бы ошибкой — Джеймс Бонд выбирает смешанный и взболтанный мартини не с одинаковыми вероятностями!

Мощность

Проверяя нулевую гипотезу против односторонней альтернативы, мы отвергаем H_0 при $t \geq 12$, что обеспечивает достигаемый уровень значимости $p \leq \alpha = 0.05$.



Пусть Джеймс Бонд выбирает смешанный мартини в 75% случаев.



$\mathrm{pow} \approx 0.6302$, т. е., при многократном повторении эксперимента гипотеза будет отклонена только в 63% случаев.

Мощность

Мощность критерия зависит от следующих факторов:

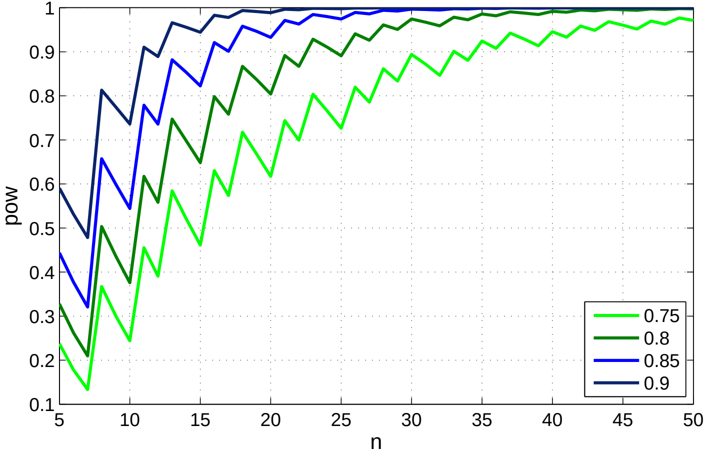
- размер выборки;
- размер отклонения от нулевой гипотезы;
- чувствительность статистики критерия;
- тип альтернативы.

Размер выборки

Особенности прикладной задачи: 1 порция мартини содержит 55 мл джина и 15 мл вермута — суммарно около 25 мл спирта. Смертельная доза алкоголя при массе тела 80 кг составляет от 320 до 960 мл спирта в зависимости от толерантности (от 13 до 38 мартини).

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

Шокирующий график



Падение мощности: объяснение