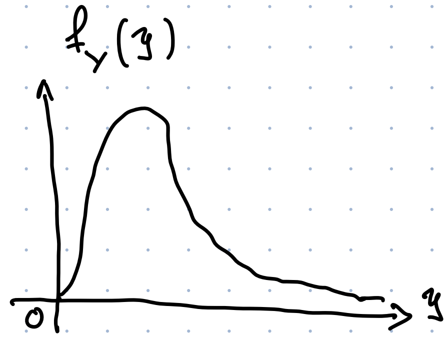
① Распределения

$X_0, X_1, X_2, \dots, X_k \sim iid\ \mathcal{N}(0, 1)$



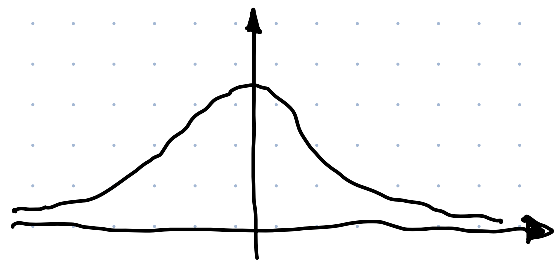
Опр. „Хи-квадрат" χ^2_k

$$Y = X_1^2 + X_2^2 + \dots + X_k^2 \sim \chi^2_k$$

$\mathbb{E}(Y) = k$

$Var(Y) = 2k$

Опр. р.-е. Стьюдента



$$Z = \frac{X_0}{\sqrt{Y/k}} \sim t(k)$$

$\mathbb{E}(Z) = 0 \quad Var(Z) = \frac{k}{k-2},\ k > 2$

$$t(1) = \frac{\mathcal{N}(0,1)}{\mathcal{N}(0,1)} \text{ — распределение Коши}$$

? $\sqrt{\chi^2_1/1}$

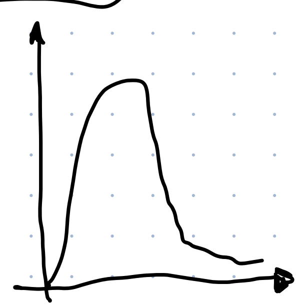
$$\boxed{\begin{array}{l} t(k) \xrightarrow{k\to\infty} \mathcal{N}(0,1) \\[1em] \dfrac{\chi^2_k}{k} \xrightarrow{P} 1 \end{array}}$$

Опр. р.-е. Фишер

$X \sim \chi^2_k$

$Y \sim \chi^2_m$

$$F = \frac{\sqrt{X/k}}{\sqrt{Y/m}} \sim F(k, m)$$



Т. Фишера

$X_1, \dots, X_n \sim iid\ N(\mu, \sigma^2)$

тогда:

1) $\bar{X},\ s^2$ независимы (Д/з.)

2) $\frac{(n-1)\cdot s^2}{\sigma^2} \sim \chi^2_{n-1}$

s^2 – несмещенная оценка дисперсии

σ^2 – теор. дисперсия

? Аргументация

② Точные критерии для одной выборки

Выборка: $X_1, \dots, X_n \sim iid$

Предпосылки:

- большое n → iid
- нет выбросов

Союзник:

- ЦПТ

$CI_\mu:\ \bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot se(\bar{x})$

$se(\bar{x}) = \frac{\hat{\sigma}}{\sqrt{n}}$, $\hat{\sigma} = s$ несмещ.

$z = \frac{\bar{x} - \mu_0}{se(\bar{x})} \underset{H_0}{\overset{asy}{\sim}} N(0;1)$

$z_{1-\frac{\alpha}{2}}$ vs z_{obs}

z-критерий

Предпосылки:

- маленькое n → iid
- конкретное р.-е. выборки

$N(\mu, \sigma^2)$

Союзник:

- Р.-е. Стьюдента
- Т. Фишера

$CI_\mu:\ \bar{x} \pm t_{1-\frac{\alpha}{2}}(n-1) \cdot se(\bar{x})$

$t = \frac{\bar{x} - \mu_0}{se(\bar{x})} \underset{H_0}{\sim} t(n-1)$

$t_{1-\frac{\alpha}{2}}$ vs t_{obs}

t-критерий

$\bar{X}$ – нормальное.

а) Достаточно ли этого для z-теста

$\bar{X}$ $\hat{\sigma}$

$$\frac{\bar{X}-\mu_0}{\hat{\sigma}/\sqrt{n}} \overset{asy}{\sim} N(0;1)$$

НЕТ

- А что с размером выборки
- А что с выборкой? наблюдения iid

$\bar{X}$ $\hat{\sigma}$

б) А для t-теста

- iid – ?
- $\bar{X}$ и s^2 незав.

НЕТ

$$\boxed{\hat{\sigma}^2 = s^2}$$

$$t = \frac{\bar{X}-\mu_0}{\sqrt{\hat{\sigma}^2/n}} = \underbrace{\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}}_{N(0,1)} \cdot \frac{\sqrt{\sigma^2/n-1}}{\sqrt{\hat{\sigma}^2/n-1}} = \frac{N(0,1)}{\sqrt{\chi^2_{n-1}/n-1}} = t(n-1)$$

$\hat{\sigma}^2 \leftarrow \sigma^2$

$$\frac{\sqrt{\sigma^2/n-1}}{\sqrt{\hat{\sigma}^2/n-1}} = \frac{1}{\sqrt{\underbrace{\frac{(n-1)\,\hat{\sigma}^2}{\sigma^2}}_{\chi^2_{n-1}}/(n-1)}}$$

Если развивать мысль дальше

- μ неизв. σ^2 изв. $\Rightarrow$ $t \sim N(0,1)$
- σ^2 неизв. μ изв. $\Rightarrow$ Критерии для дисперсий
- σ^2 неизв. μ не изв. $\Rightarrow$

неприменимо на практике

А зачем они?

см. видосы в ютубе

• $\frac{\hat{\sigma}_X}{\hat{\sigma}_Y} \sim F_{\dots}$ $\boxed{?}$

③ Другие точные критерии

Давайте наложим на выборку другие предп.

Упражнение

$X_1, \dots, X_n \sim iid\ Bern(p)$

$H_0: p = 1/2$

$H_A: p \neq 1/2$

→ большое n

→ нет выбросов

ЦПТ

$$Z = \frac{\hat{p} - 1/2}{se(\hat{p})} \overset{asy}{\underset{H_0}{\sim}} N(0,1)$$

$$Z_{obs} \ \text{vs} \ Z_{1-\frac{\alpha}{2}}$$

$$se(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

→ маленькое n

$$n\hat{p} = \sum X_i \sim Bin(n, p)$$

$$S \underset{H_0}{\sim} Bin(n, 1/2)$$

$$S_{obs} \ \text{vs} \ S_{1-\frac{\alpha}{2}}$$

Упражнение

Маша 20 лет 2 года - плохие эпидемии/войны

80 лет

$[0; R]$ - $CI_{0.95}$

геометрическое распределение

Ещё пример: равномерная выборка на золотиве

Упражнение

$X_1, \dots, X_n \sim iid\ Exp(\lambda)$

$\mu = \frac{1}{\lambda}$ $H_0: \mu = \mu_0$

$H_A: \mu \neq \mu_0$

$\to a > 0$ $V \sim Exp(\lambda)$

$a \cdot V \sim Exp\left(\frac{\lambda}{a}\right)$

$\to V \sim Exp(0.5) = \chi^2_2$

$$T = 2\lambda \cdot \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$$

$\frac{1}{\lambda}$ Точный CI:

$$T^L_{cr} \leq T \leq T^R_{cr}$$

$$\dots \leq \frac{1}{\lambda} \leq \dots$$

④ Тест Уэлча (Welch)

$X_1, \dots, X_{n_x} \sim iid$ μ_x σ_x^2

$Y_1, \dots, Y_{n_y} \sim iid$ μ_y σ_y^2

выборки нез.

CI: $\mu_x - \mu_y$

$H_0: \mu_x = \mu_y$

$H_A: \mu_x \neq \mu_y$

Асимптотика:

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{se(\bar{X} - \bar{Y})} \overset{asy}{\sim} N(0, 1)$$

Точное примерное:

$Z_1 \sim t(d)$ Z_{cr} vs Z_{obs}

А что с точным тестом?

Усилим предпосылки!

$$\left.\begin{array}{l} Y_i \sim N(\mu_Y, \sigma_Y^2) \\ X_i \sim N(\mu_X, \sigma_X^2) \\ ?\ \sigma_X^2 = \sigma_Y^2 \end{array}\right] \Rightarrow \cancel{Z \sim t(n_x + n_y - 2)}$$

это бульшит

это неприменимо на практике

! Если $\sigma_X^2 \neq \sigma_Y^2$, то точного р-я для Z не $\exists$.

Проблема Беренца-Фишера $\frac{||}{\smile}$

$\Rightarrow$ Используйте бутстрэп или тест Уэлча.

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\hat{\sigma}_X^2}{n_x} + \frac{\hat{\sigma}_Y^2}{n_y}}} \overset{\text{approx}}{\sim} t_d$$

можно найти d чтобы было похоже

$$t_d = \frac{N(0,1)}{\sqrt{\chi_d^2 / d}}$$

$N(0,1)$

$$Z = \frac{(\bar{x} - \bar{y} - (\mu_x - \mu_y)) / \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_Y^2}{n_y}}}{\sqrt{\left(\frac{\hat{\sigma}_X^2}{n_x} + \frac{\hat{\sigma}_Y^2}{n_y}\right) / \left(\frac{\sigma_X^2}{n_x} + \frac{\sigma_Y^2}{n_y}\right)}} = \sqrt{Q}$$

$$Q \overset{\text{approx}}{\sim} \frac{\chi^2_d}{d}$$

<u>Идея:</u> давайте приравняем все мат. ож. и дисп.

$$\mathbb{E}(\chi^2_d) = d \quad \mathbb{E}\left(\frac{\chi^2_d}{d}\right) = 1 = \mathbb{E}(Q)$$

$$Var(\chi^2_d) = 2d \quad Var\left(\frac{\chi^2_d}{d}\right) = \frac{2}{d} = Var(Q)$$

$\mathbb{E}(\hat{\sigma}_i^2) = \sigma_i^2$ Все оценки дисп. несмещ.

$$\mathbb{E}(Q) = 1$$

$$d = \frac{2}{Var(Q)} = \frac{2}{\frac{1}{\left(\frac{\hat{\sigma}_x^2}{n_x} + \frac{\hat{\sigma}_y^2}{n_y}\right)^2} \cdot Var\left(\frac{\hat{\sigma}_x^2}{n_x} + \frac{\hat{\sigma}_y^2}{n_y}\right)} =$$

$$= \frac{2\left(\frac{\hat{\sigma}_x^2}{n_x} + \frac{\hat{\sigma}_y^2}{n_y}\right)^2}{\frac{1}{n_x^2} \cdot Var(\hat{\sigma}_x^2) + \frac{1}{n_y^2} \cdot Var(\hat{\sigma}_y^2)}$$

$$X_i \sim N(\mu_x, \sigma_x^2)$$

Т. Фишера

$$\frac{(n-1) \cdot s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum (x_i - \bar{x})^2}{\sigma^2}$$

$$Var\left(\frac{1}{\sigma^2} \cdot \sum (x_i - \bar{x})^2\right) = 2 \cdot (n-1)$$

$$Var\left(\sum (x_i - \bar{x})^2\right) = 2(n-1)\sigma^4$$

$$\operatorname{Var}\left((n-1)\cdot\hat{\sigma}^2\right) = 2(n-1)\sigma^4$$

$$\operatorname{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{(n-1)}$$

$$d = \frac{2\cdot\left(\frac{\hat{\sigma}_x^2}{n_x} + \frac{\hat{\sigma}_y^2}{n_y}\right)^2}{\frac{2\hat{\sigma}_x^4}{n_x^2\cdot(n_x-1)} + \frac{2\hat{\sigma}_y^4}{n_y^2\cdot(n_y-1)}} \qquad t(d)$$

5 Бонус-трек — картинки

Хи-квадрат

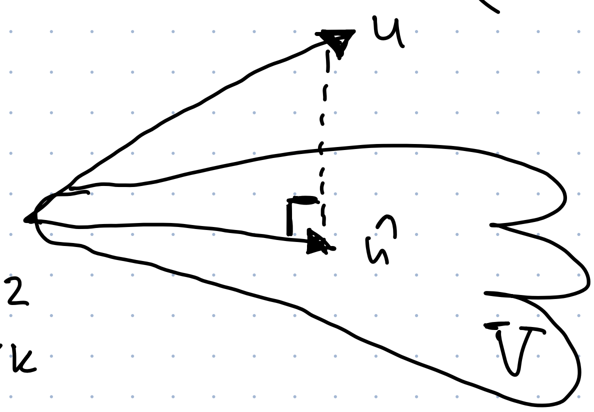
$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}\right)$$

$u \sim N(0, I_n)$

$u \in \mathbb{R}^n$

$V \quad \dim V = k$

$n > k$



$$\|\hat{u}\|^2 = u_1^2 + \ldots + u_k^2 \sim \chi_k^2$$

Квадрат длины проекции

t-распределение

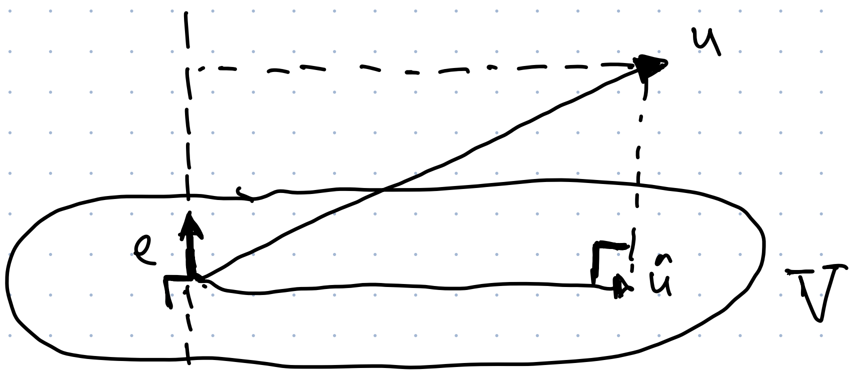
$X_0 \perp\!\!\!\perp \chi^2_n$

$u \sim \mathcal{N}(0, \cancel{\sigma^2} \cdot I_n)$

$Y = \|\hat{u}\|^2 \sim \chi^2_k$

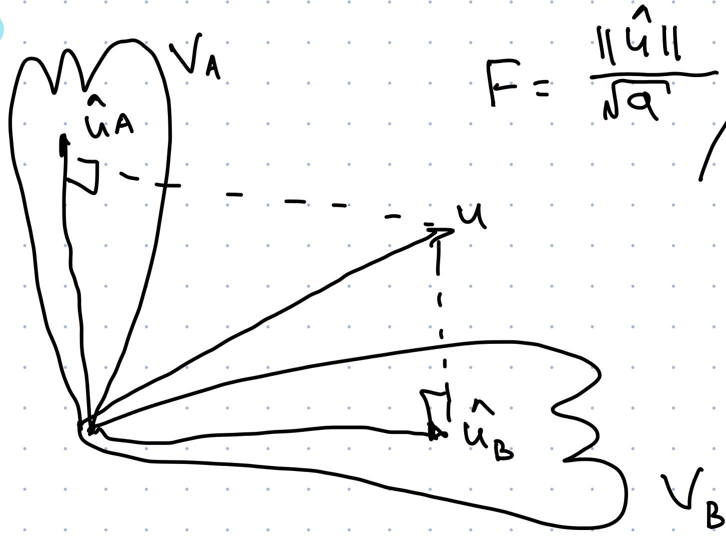
$\hat{u}$ - проекция u на V

$$t = \frac{\mathcal{N}(0,1)}{\sqrt{\chi^2_k / k}} = \frac{\langle u, e \rangle}{\|\hat{u}\| / \sqrt{k}}$$



$e \perp V \qquad \|e\| = 1$

F-распр.



$$F = \frac{\|\hat{u}\|}{\sqrt{a}} \Big/ \frac{\|\hat{u}\|}{\sqrt{b}} \sim F_{a,b}$$