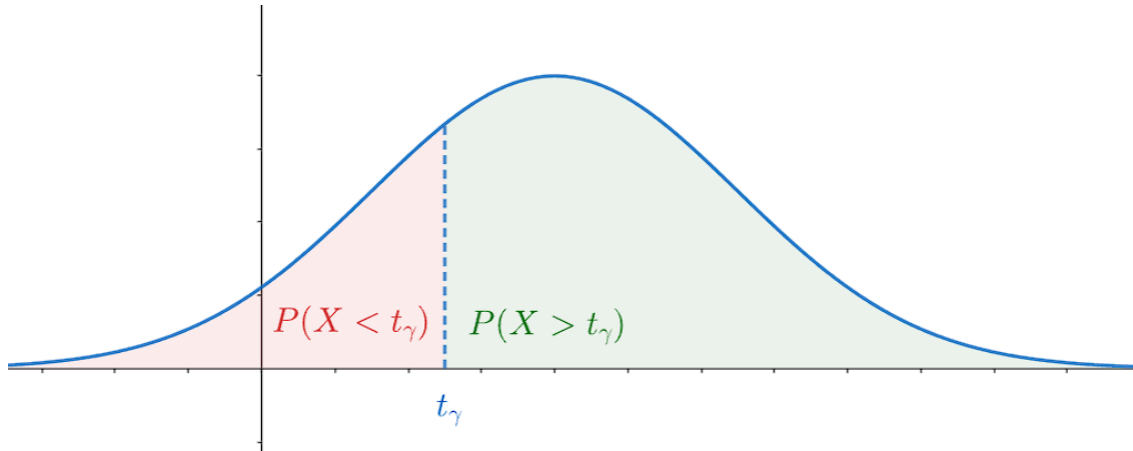


Лекция 5.

Квантиль распределение

Предполагаем, что распределение абсолютно непрерывное, и $F(x)$ - функция распределения

Def. 1. Число t_γ называется квантилем распределения уровня γ , если значения функции распределения $F(t_\gamma) = \gamma$ или $P(X < t_\gamma) = \gamma$ ($t_\gamma = F^{-1}(\gamma)$)



Ex. Медиана - квантиль уровня $\frac{1}{2}$

Def. 2. Число t_α называется квантилем уровня значимости α , если $P(X > t_\alpha) = \alpha$ или $F(t_\alpha) = 1 - \alpha$
Ясно, что $\gamma = 1 - \alpha$

Интервальные оценки

Недостатки точечных оценок - неизвестно насколько они далеки от реального значения параметра и насколько им можно доверять. Особенно это заметно при малых выборках. Поэтому мы указываем интервал, в котором лежит этот параметр с заданной вероятностью (надежностью) γ . Такие оценки называются интервальными (доверительными)

Def. Интервал $(\theta_\gamma^-; \theta_\gamma^+)$ называется доверительным интервалом параметра θ надежности γ , если вероятность $P(\theta_\gamma^- < \theta < \theta_\gamma^+) = \gamma$

Nota. Если имеем дискретную случайную величину, то $P(\theta_\gamma^- < \theta < \theta_\gamma^+) \geq \gamma$

Nota. Так как параметр θ - константа, то бессмысленно говорить о его попадании в интервал.

Правильно: доверительный интервал накрывает параметр θ с вероятностью γ

Nota. 1. $\alpha = 1 - \gamma$ называется уровнем значимости доверительного интервала

Nota. 2. Обычно пытаются строить симметричный доверительный интервал относительно несмещенной оценки θ^*

Nota. 3. Возникает вопрос, какой уровень γ выбрать для исследования. Стандартные уровни надежности γ : 0.9, 0.95, 0.99, 0.999. Самый мейнстримный - 0.95. В малых выборках используют 0.9

Вспомним основную теорему:

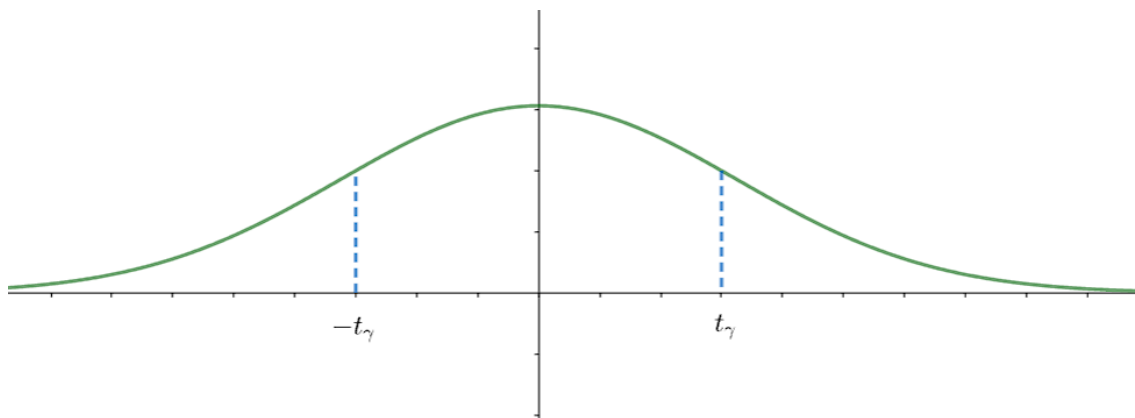
$\square (X_1, \dots, X_n)$ - выборка объема n из $N(\alpha, \sigma^2)$

\bar{x} - выборочное среднее, S^2 - исправленная дисперсия, D^* - выборочная дисперсия

Тогда:

1. $\sqrt{n} \frac{\bar{x} - a}{\sigma} \in N(0, 1)$
2. $\sum_{i=1}^n \left(\frac{X_i - a}{\sigma} \right)^2 = \frac{n\tilde{\sigma}^2}{\sigma^2} \in H_n$, где $n\tilde{\sigma}^2 = \sum_{i=1}^n (X_i - a)^2$
3. $\sum_{i=1}^n \left(\frac{X_i - \bar{x}}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{nD^*}{\sigma^2} \in H_{n-1}$
4. $\sqrt{n} \frac{\bar{x} - a}{S} \in T_{n-1}$
5. \bar{x} и S^2 - независимы

Nota. Если $F(x)$ - функция симметричного относительно $x = 0$ распределения, то $P(|X| < t) = 2F(t) - 1$



Доверительные интервалы для параметров нормального распределения

Пусть $\vec{X} = (X_1, \dots, X_n)$ - выборка объема n из $N(a, \sigma^2)$. Хотим найти интервалы для параметров a и σ^2

I. Доверительный интервал для параметра a при известном значении σ^2

По пункту 1 из теоремы $\sqrt{n} \frac{\bar{x} - a}{\sigma} \in N(0, 1)$

$$P\left(-t_\gamma < \sqrt{n} \frac{\bar{x} - a}{\sigma} < t_\gamma\right) = P\left(\left|\sqrt{n} \frac{\bar{x} - a}{\sigma}\right| < t_\gamma\right) = 2F_0(t_\gamma) - 1 = \gamma$$

$$F_0(t_\gamma) = \frac{1+\gamma}{2} \implies t_\gamma - \text{квантиль уровня } \frac{1+\gamma}{2} \text{ для } N(0, 1), \text{ где } F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$$

Решая неравенство, получаем $-t_\gamma < \sqrt{n} \frac{\bar{x} - a}{\sigma} < t_\gamma$

$$-t_\gamma \frac{\sigma}{\sqrt{n}} < \bar{x} - a < t_\gamma \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - t_\gamma \frac{\sigma}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{\sigma}{\sqrt{n}} - \text{симметричный интервал относительно } \bar{x}$$

Доверительный интервал надежности γ : $\left(\bar{x} - t_\gamma \frac{\sigma}{\sqrt{n}}, \bar{x} + t_\gamma \frac{\sigma}{\sqrt{n}}\right)$, где t_γ - квантиль $N(0, 1)$

уровня $\frac{1+\gamma}{2}$

II. Доверительный интервал для параметра a при неизвестном σ^2

Из пункта 4 из теоремы $\sqrt{n} \frac{\bar{x} - a}{S} \in T_{n-1}$

$$P\left(-t_\gamma < \sqrt{n} \frac{\bar{x} - a}{S} < t_\gamma\right) = P\left(\left|\sqrt{n} \frac{\bar{x} - a}{S}\right| < t_\gamma\right) = 2F_{T_{n-1}}(t_\gamma) = \gamma$$

$$F_{T_{n-1}}(t_\gamma) = \frac{1+\gamma}{2} \implies t_\gamma - \text{квантиль } T_{n-1} \text{ уровня } \frac{1+\gamma}{2}$$

Аналогично с примером выше получаем интервал $\left(\bar{x} - t_\gamma \frac{S}{\sqrt{n}}, \bar{x} + t_\gamma \frac{S}{\sqrt{n}}\right)$, где t_γ - квантиль

T_{n-1} уровня $\frac{1+\gamma}{2}$

III. Доверительный интервал для параметра σ^2 при неизвестном a

По пункту 3 из теоремы $\sum_{i=1}^n \left(\frac{X_i - \bar{x}}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{nD^*}{\sigma^2} \in H_{n-1}$

Пусть χ_1^2 и χ_2^2 - квантили H_{n-1} уровней $\frac{1-\gamma}{2}$ и $\frac{1+\gamma}{2}$

$$\text{Тогда } P\left(\chi_1^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_2^2\right) = F_{H_{n-1}}(\chi_1^2) - F_{H_{n-1}}(\chi_2^2) = \frac{1-\gamma}{2} - \frac{1+\gamma}{2} = \gamma$$

$$\chi_1^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_2^2$$

$$\frac{1}{\chi_2^2} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi_1^2}$$

$$\frac{(n-1)S^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_1^2} \text{ или } \frac{nD^*}{\chi_2^2} < \sigma^2 < \frac{nD^*}{\chi_1^2}$$

Получаем интервал $\left(\frac{(n-1)S^2}{\chi_2^2}, \frac{(n-1)S^2}{\chi_1^2}\right)$, где χ_1^2 и χ_2^2 - квантили H_{n-1} уровней $\frac{1-\gamma}{2}$ и

$\frac{1+\gamma}{2}$

Nota. Данный интервал не симметричен относительно неизвестного параметра σ^2

IV. Доверительный интервал для параметра σ^2 при известном a

По пункту 2 из теоремы $\sum_{i=1}^n \left(\frac{X_i - a}{\sigma} \right)^2 = \frac{n\tilde{\sigma}^2}{\sigma^2} \in H_{n-1}$

Пусть χ_1^2 и χ_2^2 - квантили H_n уровней $\frac{1-\gamma}{2}$ и $\frac{1+\gamma}{2}$

Тогда $P\left(\chi_1^2 < \frac{n\tilde{\sigma}^2}{\sigma^2} < \chi_2^2\right) = F_{H_n}(\chi_1^2) - F_{H_n}(\chi_2^2) = \frac{1-\gamma}{2} - \frac{1+\gamma}{2} = \gamma$

Аналогично получаем интервал $\left(\frac{n\tilde{\sigma}^2}{\chi_2^2}, \frac{n\tilde{\sigma}^2}{\chi_1^2}\right)$, где χ_1^2 и χ_2^2 - квантили H_n уровней $\frac{1-\gamma}{2}$ и

$\frac{1+\gamma}{2}$, $n\tilde{\sigma}^2 = \sum_{i=1}^n (X_i - a)^2$

Nota. $\tilde{\sigma}^2 - D^* = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2aX_i + a^2 - X_i^2 + 2\bar{x}X_i - \bar{x}^2) = \frac{1}{n} (na^2 -$

$2an\bar{x} + 2\bar{x} \cdot n\bar{x} - n\bar{x}^2) = a^2 - 2a\bar{x} + \bar{x}^2 = (a - \bar{x})^2 \implies \tilde{\sigma}^2 = D^* + (a - \bar{x})^2$

Получаем $\left(\frac{n(D^* + (a - \bar{x})^2)}{\chi_2^2}, \frac{n(D^* + (a - \bar{x})^2)}{\chi_1^2}\right)$

Асимптотические доверительные интервалы

Def. Интервал $(\theta_\gamma^-, \theta_\gamma^+)$ называется асимптотическим доверительным интервалом параметра θ уровня γ , если $P(\theta_\gamma^- < \theta < \theta_\gamma^+) \xrightarrow{n \rightarrow \infty} \gamma$

Ex. Доверительный интервал вероятности события A

Пусть $p = P(A)$, $q = 1 - p$, n - число испытаний или объем выборки (X_1, \dots, X_n) , где $X_i \in \{0, 1\}$

$p^* = \frac{n_A}{n} = \bar{x}$ - оценка p

Согласно Центральной предельной теореме $\sqrt{n} \frac{p^* - p}{\sqrt{pq}} = \sqrt{n} \frac{p^* - p}{DX_1} \Rightarrow N(0, 1)$

Так как $p^* \xrightarrow{p} p$, то $\sqrt{n} \frac{p^* - p}{\sqrt{p^*(1-p^*)}} = \underbrace{\sqrt{n} \frac{p^* - p}{\sqrt{p(1-p)}}}_{\Rightarrow N(0,1)} \underbrace{\frac{\sqrt{p(1-p)}}{\sqrt{p^*(1-p^*)}}}_{\xrightarrow{p} 1} \Rightarrow N(0, 1)$

$P\left(\left|\sqrt{n} \frac{p^* - p}{\sqrt{p^*(1-p^*)}}\right| < t_\gamma\right) \xrightarrow{n \rightarrow \infty} 2F_0(t_\gamma) - 1 = \gamma$

$F_0(t_\gamma) = \frac{1+\gamma}{2}$, t_γ - квантиль $N(0, 1)$ уровня $\frac{1+\gamma}{2}$

Получаем $\left|\sqrt{n} \frac{p^* - p}{\sqrt{p^*(1-p^*)}}\right| < t_\gamma$

$|p^* - p| < t_\gamma \frac{\sqrt{p^*(1-p^*)}}{\sqrt{n}}$

Итак, $\left(-t_\gamma \frac{\sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n}}, t_\gamma \frac{\sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n}}\right)$, где t_γ - квантиль $N(0, 1)$ уровня $\frac{1+\gamma}{2}$