Математическая статистика
Лекция 3

Интервальное оценивание

Другой подход к оцениванию параметра, при котором указывают интервал, накрывающий параметр с заданной наперед вероятностью называется **интервальным оцениванием**.

Чем больше уверенность в том, что параметр лежит в интервале, тем шире интервал. Поэтому бессмысленно искать диапазон, внутри которого θ содержится гарантированно — это вся область $\boldsymbol{\Theta}$

Интервальное оценивание

$X_1,\ldots,X_n \sim F(X,\theta)$-семейство распределений с параметром $\theta \in \boldsymbol{\Theta} \subseteq R$

Опр. Пусть $0 < \alpha < 1$. Интервал со случайными концами

$$(T_1, T_2)=(T_1(X_1,\ldots,X_n);\ T_1(X_1,\ldots,X_n))$$

называется *доверительным интервалом* для параметра θ *уровня доверия* $(1-\alpha)$, если для любого $\theta \in \boldsymbol{\Theta}$

$$P(T_1 < \theta < T_2) \geqslant 1-\alpha$$

T_1-нижняя (левая) доверительная граница,

T_2-верхняя (правая) доверительная граница

Лемма Фишера

- Пусть вектор $X=(X_1, . . . , X_n)$ состоит из независимых случайных величин со стандартным нормальным распределением, C — ортогональная матрица, и $Y = C \cdot X$.

Тогда $\sum_{i=1}^{n} X_i^2 - Y_1^2 - \cdots Y_k^2$ для любого $k = 1, . . . , n-1$ не зависит от $Y_1, . . . , Y_k$

и имеет χ^2_{n-k} распределение

(Доказательство см лекцию)

Следствия из леммы Фишера

- Пусть $X_1,\ldots,X_n$ независимы и нормально распределены $N(\mu,\sigma^2)$

Тогда

1) $\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

2) $\bar{X}$ и S^2 независимы

3) $\dfrac{n-1}{\sigma^2} S_0^2 \sim \chi^2_{n-1}$, $\dfrac{n}{\sigma^2} S^2 \sim \chi^2_{n-1}$

4) $\dfrac{\bar{X}-\mu}{S_0/\sqrt{n}} \sim T_{n-1}$

Доказательства см. лекцию

Точный доверительный интервал для математического ожидания

- 1) $X_1,\ldots,X_n \sim N(\mu, \sigma^2)$, где $\mu \in R$ —неизвестный параметр,

а значение $\sigma^2 > 0$ **известно.**

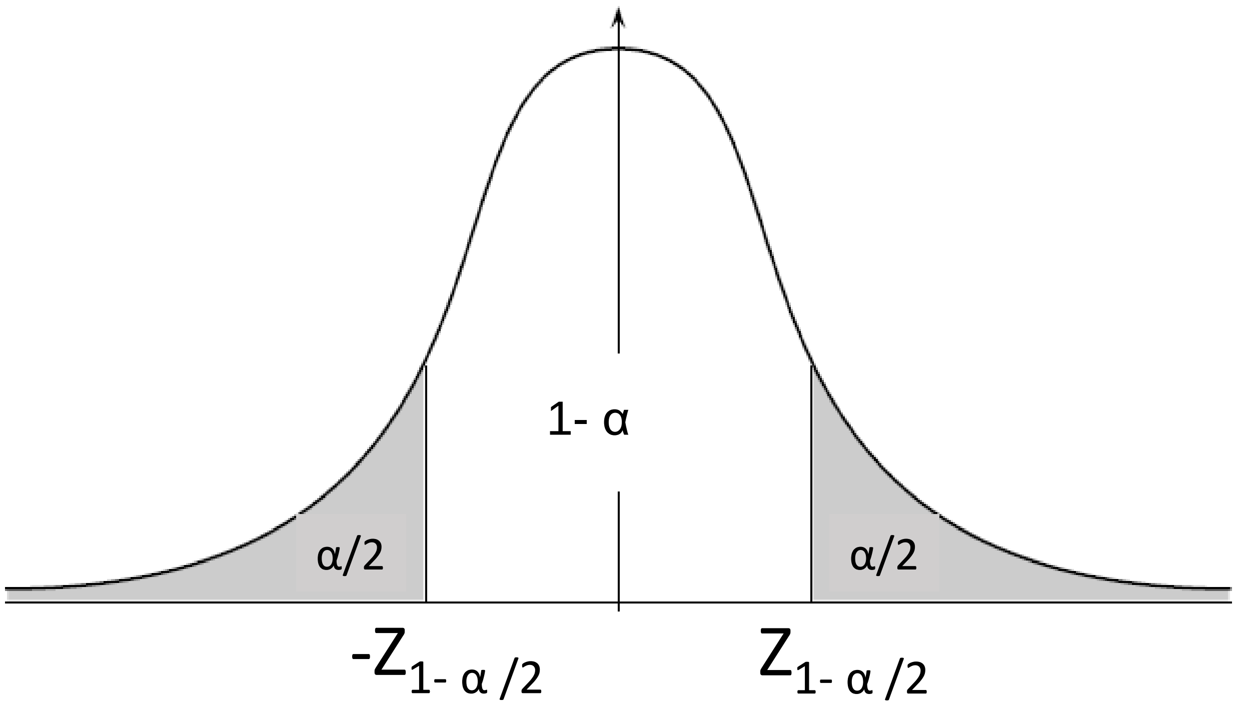
Требуется при произвольном *n* построить доверительный интервал для параметра μ уровня доверия 1- α

$$P(T_1 < \mu < T_2) = 1-\alpha$$

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$P\left(-Z_{1-\alpha/2} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{1-\alpha/2}\right) = 1-\alpha$$

$$\boxed{P\left(\overline{X}-Z_{1-\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}+Z_{1-\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha}$$



Точный доверительный интервал для математического ожидания

- 2) $X_1,\dots,X_n \sim N(\mu,\sigma^2)$, где $\mu \in R$ —неизвестный параметр,

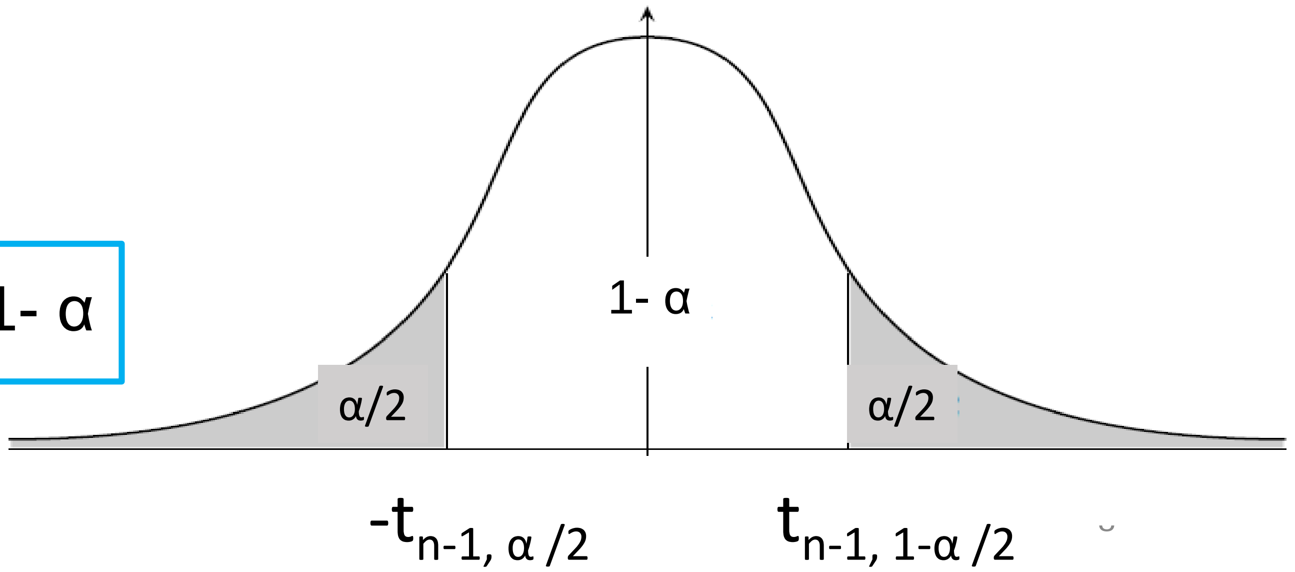
а значение $\sigma^2 > 0$ **неизвестно.**

Требуется при произвольном *n* построить доверительный интервал для параметра μ уровня доверия 1- α

$$\frac{\overline{X}-\mu}{S_0/\sqrt{n}} \sim T_{n-1}$$

$$P\left(-t_{n-1,\,1-\alpha/2} < \frac{\overline{X}-\mu}{S_0/\sqrt{n}} < t_{n-1,\,1-\alpha/2}\right) = 1-\alpha$$

$$\boxed{P\left(\overline{X}-t_{n-1,\,1-\alpha/2}\cdot\frac{S_0}{\sqrt{n}} < \mu < \overline{X}+t_{n-1,\,1-\alpha/2}\cdot\frac{S_0}{\sqrt{n}}\right) = 1-\alpha}$$



Свойства доверительных интервалов

- Длина интервала $$\Delta_1 = z_{1-\alpha/2} \cdot \frac{2\sigma}{\sqrt{n}}$$

$$\Delta_2 = t_{n-1,1-\alpha/2} \cdot \frac{2S_0}{\sqrt{n}}$$

Как изменяется длина интервала с ростом объема выборки, дисперсии, уровня доверия?

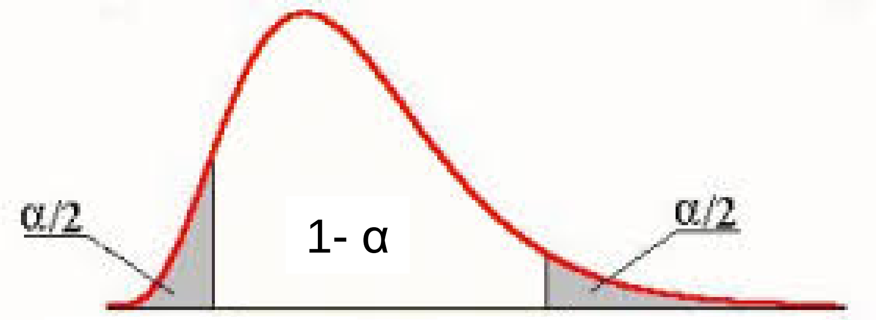
Точный доверительный интервал для дисперсии

- X$_1$,…,Xn$\sim N(\mu,\sigma^2)$, μ- **неизвестно**

P(T_1<σ^2<T_2)= 1- α

- По теореме Фишера $\frac{n-1}{\sigma^2}S_0^2 \sim \chi^2_{n-1}$

P($c_{n-1,\alpha/2}$<$\frac{n-1}{\sigma^2}S_0^2$<$c_{n-1,1-\alpha/2}$)= 1- α



$c_{n-1,\alpha/2}$ $c_{n-1,1-\alpha/2}$

$$\boxed{\mathrm{P}\left(\frac{S_0^2(n-1)}{c_{n-1,1-\alpha/2}} < \sigma^2 < \frac{S_0^2(n-1)}{c_{n-1,\alpha/2}}\right) = 1-\alpha}$$

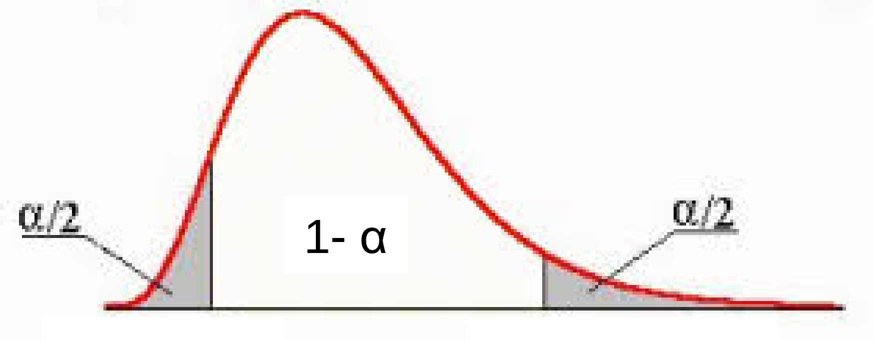
Точный доверительный интервал для дисперсии

- $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, μ- **известно**

$P(T_1 < \sigma^2 < T_2) = 1 - \alpha$

Тогда $\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$

$P\left(c_{n,\alpha/2} < \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} < c_{n,1-\alpha/2}\right) = 1 - \alpha$



$c_{n,\alpha/2}$ $c_{n,1-\alpha/2}$

$$\boxed{P\left(\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{c_{n,1-\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{c_{n,\alpha/2}}\right) = 1 - \alpha}$$

Доверительный интервал для разности мат. ожиданий. Связанные пары

- Пусть (X_i, Y_i) – случайная выборка из двумерного нормального распределения, $1 \leq i \leq n$, причем $X_i \sim N(\mu_X, \sigma_X^2), Y_i \sim N(\mu_Y, \sigma_Y^2)$, тогда доверительный интервал для разности математических ожиданий

$$P\left(\overline{X} - \overline{Y} - t_{n-1,\,1-\alpha/2} \cdot \frac{S_0(\Delta)}{\sqrt{n}} < \mu_X - \mu_Y < \overline{X} - \overline{Y} + t_{n-1,\,1-\alpha/2} \cdot \frac{S_0(\Delta)}{\sqrt{n}}\right) = 1 - \alpha$$

Где $t_{n-1,\,1-\alpha/2}$- квантиль уровня 1-α /2 распределения Стьюдента с (n-1) степенями свободы

$$S_0^2(\Delta) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - Y_i - \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)\right)^2$$

Доверительные интервалы для разности математических ожиданий. Независимые выборки, известные дисперсии.

- Даны две независимых выборки из нормального распределения: $X_1, \dots X_{n_X} \sim N(\mu_X, \sigma_X^2), Y_1, \dots, Y_{n_Y} \sim N(\mu_Y, \sigma_Y^2)$,

Доверительный интервал для разницы математических ожиданий

$$\boxed{P\left(\overline{X} - \overline{Y} - z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} < \mu_X - \mu_Y < \overline{X} - \overline{Y} + z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right) = 1 - \alpha}$$

Доверительные интервалы для разности математических ожиданий. Независимые выборки, неизвестные, но равные дисперсии

- Даны две независимых выборки из нормального распределения: $X_1, \dots X_{n_X} \sim N(\mu_X, \sigma^2), Y_1, \dots, Y_{n_Y} \sim N(\mu_Y, \sigma^2)$,
- Доверительный интервал для разницы математических ожиданий

$$\boxed{P\left(\bar{X} - \bar{Y} - t_{n_X+n_Y-2,1-\alpha/2} \cdot \hat{\sigma} \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} < \mu_X - \mu_Y < \bar{X} - \bar{Y} + t_{n_X+n_Y-2,1-\alpha/2} \cdot \hat{\sigma} \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}\right) = 1 - \alpha}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n_X}(X_i - \bar{X})^2 + \sum_{i=1}^{n_Y}(Y_i - \bar{Y})^2}{n_X + n_Y - 2}}$$

Асимптотический доверительный интервал для математического ожидания

- Пусть $X_1,\ldots,Xn$– выборка из некоторого распределения с конечной дисперсией. Тогда для больших выборок

$$P\left(\overline{X} - Z_{1-\alpha/2} \cdot \frac{S_0}{\sqrt{n}} < \mu < \overline{X} + Z_{1-\alpha/2} \cdot \frac{S_0}{\sqrt{n}}\right) \approx 1 - \alpha$$

Асимптотический доверительный интервал для вероятности

- Пусть – $X_1,\ldots,Xn$ выборка из распределения Бернулли B(p) .

Тогда для больших выборок

$$\boxed{P\left(\hat{p}-Z_{1-\alpha/2}\cdot\sqrt{\frac{(\hat{p}-1)\hat{p}}{n}}<p<\hat{p}+Z_{1-\alpha/2}\cdot\sqrt{\frac{(\hat{p}-1)\hat{p}}{n}}\right)\approx 1-\alpha}$$

$$\hat{p}=\bar{X}$$