Семинар - 8.

Теория:

Доверительные интервалы

Одним из способов поиска доверительного является метод *центральной статистики*. Данный метод применим в случае выборки $\boldsymbol{X}$ из абсолютно непрерывного распределения $\mathbb{P}_\theta$.

Определение (Центральная статистика)**.** Функция от выборки $G(\boldsymbol{X}, \theta)$, которая может зависеть от параметра θ, называется *центральной статистикой*, если

- $G(\boldsymbol{X}, \theta)$ непрерывна и монотонна по θ при любом фиксированном $\boldsymbol{X}$;
- $\mathbb{P}_\theta(G(\boldsymbol{X}, \theta) \le t) = F_G(t)$ - непрерывная функция по t, не зависящая от θ (т.е. распределение $G(\boldsymbol{X}, \theta)$ не зависит от θ).

Таким образом, если $G(\boldsymbol{X}, \theta)$ - центральная статистика, то множество $\{\theta : t_1 < G(\boldsymbol{X}, \theta) < t_2\}$ является интервалом. Формально центральная статистика не является статистикой, так как может зависеть от неизвестного параметра θ.

Теорема 1 (Фишер)**.** *Пусть* $(X_1, \ldots, X_n)$ *– выборка объема* n *из распределения* $\mathcal{N}(\mu, \sigma^2)$. *Тогда* $\overline{X}$ *и* S^2 *независимы и*

$$\sqrt{n} \cdot \frac{\overline{X} - \mu}{\sigma} \sim \mathcal{N}(0, 1), \qquad \frac{nS^2}{\sigma^2} \sim \chi^2(n-1).$$

Определение (Распределение Стьюдента)**.** Распределением Стьюдента (или t-распределением) с n степенями свободы называется распределение случайной величины

$$T = \frac{\xi}{\sqrt{\eta/n}},$$

где ξ и η – независимые случайные величины, причем $\xi \sim \mathcal{N}(0,1)$, $\eta \sim \chi^2(n)$. Обозначение: $T \sim t(n)$, $T \sim t_n$.

Задачи с семинара:

1. $X_1, \ldots, X_n$ – выборка из $\mathcal{N}(\theta, \sigma^2)$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для математического ожидания θ при известной дисперсии σ^2.
2. $X_1, \ldots, X_n$ – выборка из $\mathcal{N}(\mu, \theta^2)$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для дисперсии θ^2 при известном математическом ожидании μ.
3. $X_1, \ldots, X_n$ – выборка из $\mathcal{N}(\theta, \sigma^2)$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для математического ожидания θ при неизвестной дисперсии σ^2.
4. $X_1, \ldots, X_n$ – выборка из $\mathcal{N}(\mu, \theta^2)$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для дисперсии θ^2 при неизвестном математическом ожидании μ.
5. Пусть $X_1, \ldots, X_n$ – выборка из пуассоновского распределения $\mathrm{Pois}(\theta)$ с параметром $\theta > 0$. Постройте асимптотический доверительный интервал уровня доверия $1-\alpha$ для параметра θ.
6. Пусть $X_1, \ldots, X_n$ – выборка из распределения Бернулли $\mathrm{Ber}(\theta)$ с параметром $\theta \in (0,1)$. Постройте асимптотический доверительный интервал уровня доверия $1-\alpha$ для параметра θ.
7. $X_1, \ldots, X_n$ – выборка из равномерного распределения $\mathcal{U}(0, \theta)$, $\theta > 0$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для параметра θ.

2. $X_1, \ldots, X_n$ – выборка из $\mathcal{N}(\mu, \theta^2)$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для дисперсии θ^2 при известном математическом ожидании μ.

$$X_1, X_2, X_3 \ldots X_n \sim \mathcal{N}(\mu, \theta^2)$$

$$\mathbb{P}(T_1(X) < \theta^2 < T_2(X)) = 1-\alpha.$$

a) $\hat{\theta} = ?$

1) $G(X, \theta) = ?$

начинаем с оценки

получаем незав. распред. от. θ.

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

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \to \frac{X_i-\mu}{\theta} \sim \mathcal{N}(0,1) \Rightarrow$$

$$G(X, \theta) =$$

$$G(X, \theta^2) = \frac{1}{\theta^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$$

Если μ - тоже неизв. {как в N4}, то вместо μ исп $\bar{X}$ + т. Фишера.

$$\mathbb{P}\left(x_{\gamma_1} < \frac{1}{\theta^2}\sum(X_i-\mu)^2 < x_{\gamma_2}\right) = 1-\alpha$$

$$\parallel$$

$$\mathbb{P}(T_1(X) < \theta^2 < T_2(X)) = 1-\alpha.$$

$$\parallel$$

$$\mathbb{P}\left(\frac{1}{x_{\gamma_2}}\sum(X_i-\mu)^2 < \theta^2 < \frac{1}{x_{\gamma_1}}\sum(X_i-\mu)^2\right)$$

Будем искать центральный дов. интервал:

Находим квантили:

$\gamma_1, \gamma_2 : \gamma_2 - \gamma_1 = 1-\alpha.$

$$\begin{cases} x_{\gamma_1} : F_{\chi^2}(x_{\gamma_1}) = \frac{\alpha}{2} \Rightarrow x_{\gamma_1} = x_{\frac{\alpha}{2}} \\ x_{\gamma_2} : F_{\chi^2}(x_{\gamma_2}) = 1-\frac{\alpha}{2} \Rightarrow x_{\gamma_2} = x_{1-\frac{\alpha}{2}} \end{cases}$$

Наш интервал:

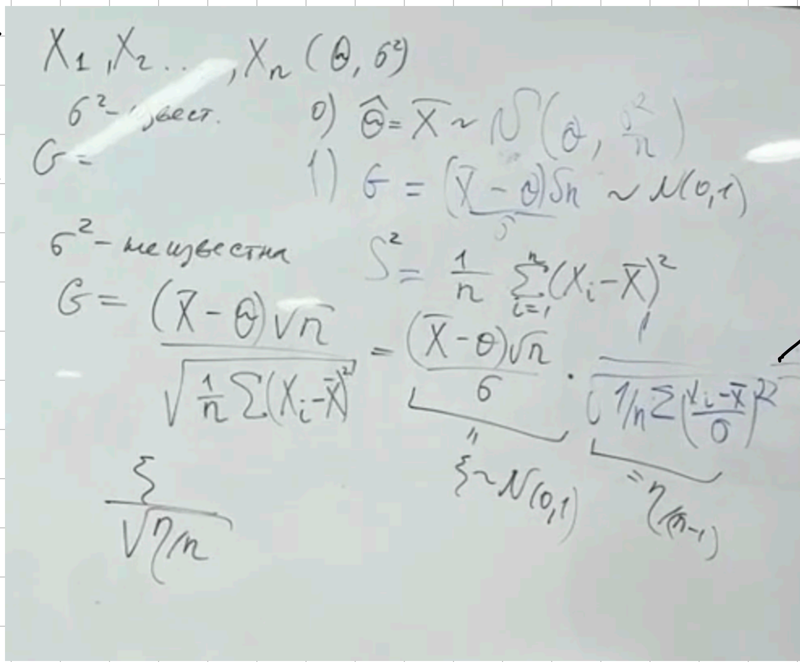
$$\mathbb{P}\left(\frac{1}{x_{1-\frac{\alpha}{2}}}\sum(X_i-\mu)^2 < \theta^2 < \frac{1}{x_{\alpha/2}}\sum(X_i-\mu)^2\right)$$

1. $X_1, \dots, X_n$ – выборка из $\mathcal{N}(\theta, \sigma^2)$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для математического ожидания θ при известной дисперсии σ^2.

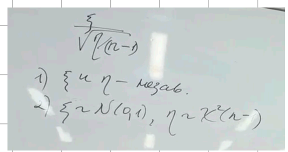
3. $X_1, \dots, X_n$ – выборка из $\mathcal{N}(\theta, \sigma^2)$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для математического ожидания θ при неизвестной дисперсии σ^2.

задача 1 —

задача 2 —

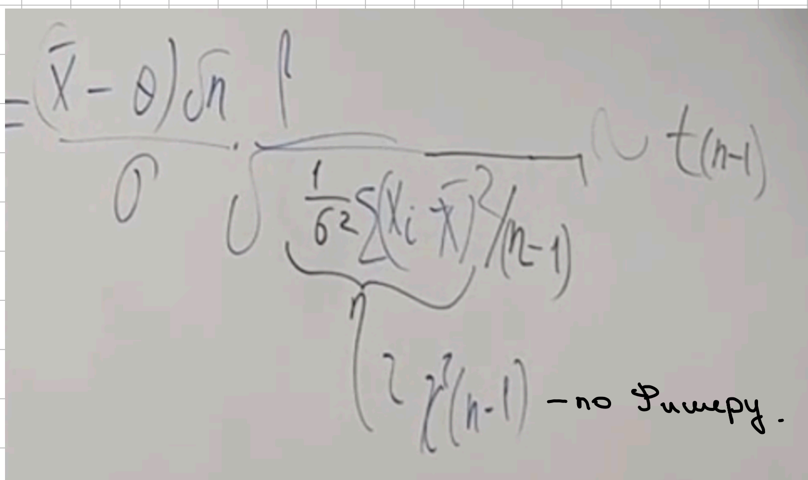


сводим к распред. Стьюдента



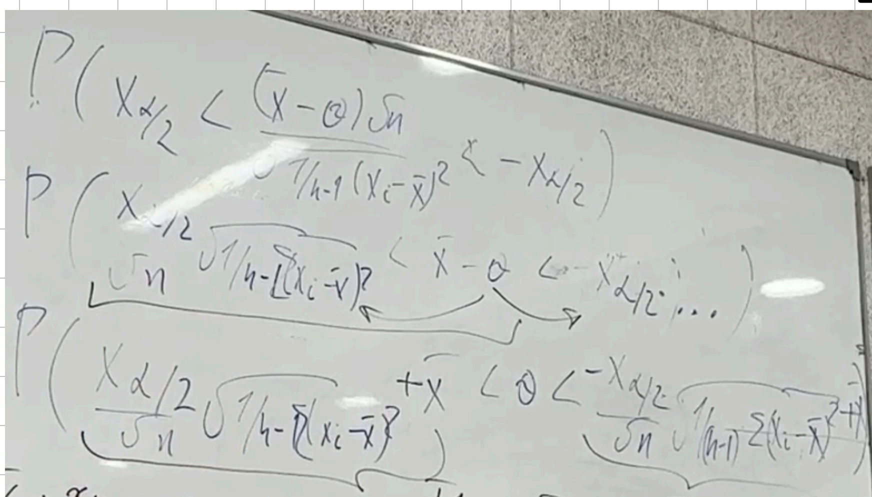
По итогу наша центр. статистика:

$$G(X, \theta) = \frac{(\bar X - \theta)\sqrt n}{\sqrt{\frac1n \sum (x_i - \bar X)^2}} =$$



— по Фишеру.

Строим интервал. Стьюдент симметрич $\Rightarrow x_{1-\frac{\alpha}{2}} = -x_{\alpha/2}$



4. $X_1, \dots, X_n$ – выборка из $\mathcal{N}(\mu, \theta^2)$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для дисперсии θ^2 при неизвестном математическом ожидании μ.

$$X_1 \dots X_n \sim \mathcal{N}(\mu, \theta^2) \quad \leftarrow \text{неизв.}$$

$$\hat{\theta} = \frac{1}{n}\sum (X_i - \mu)^2 \qquad \mu \to \overline{X}.$$

$$\hat{\mu} = \overline{X} \Rightarrow \hat{\theta} = \frac{1}{n}\sum (X_i - \overline{X})^2$$

$$G(X, \theta) = \frac{1}{\sigma^2}\sum (X_i - \overline{X})^2 \sim \chi^2(n-1) \leftarrow \text{Фишер.}$$

Асимптотические доверит. интервалы:

$$\hat{\theta}_{\text{ОМП}} : \frac{\sqrt{n}}{\sigma(\theta)}\left(\hat{\theta}_{\text{ОМП}} - \theta\right) \xrightarrow{n\to\infty} \mathcal{N}(0,1)$$

$$\hat{\theta}_{\text{ОМП}} = \overline{X}, \quad \boxed{\sigma(\theta) = \frac{1}{i_1(\theta)}} \leftarrow \text{ассимпт. дисперсия}$$

Замечание 10. Если модель удовлетворяет условиям регулярности, то в качестве асимптотически нормальной оценки можно взять оценку максимального правдоподобия $\hat{\theta}_{MLE}$. Тогда асимптотическая дисперсия может быть вычислена как $\sigma^2(\theta) = \frac{1}{i_1(\theta)}$ (см. теорему 7.3). Более того, так как $\hat{\theta}_{MLE}$ обладает наименьшей асимптотической дисперсией, то и доверительный интервал, построенный с ее использованием, будет наикратчайший.

ЦПТ: $$\frac{\sqrt{n}(\overline{X} - \theta)}{\sqrt{\theta}} \xrightarrow{d} \mathcal{N}(0,1)$$

$\overline{X}$ – состоят. для $\theta \Rightarrow$

$$\varphi(\overline{X}) \xrightarrow{\mathbb{P}} \varphi(\theta)$$

$$\hat{\sigma}(\theta) \xrightarrow{\mathbb{P}} \sigma(\theta)$$

$$\hat{\sigma}(\theta) = \sqrt{\overline{X}}, \quad \sigma(\theta) = \sqrt{\theta}$$

$$G = \frac{\sqrt{n}}{\sqrt{\overline{X}}} \cdot (\overline{X} - \theta) = \underbrace{\frac{\sqrt{n}}{\sqrt{\theta}}(\overline{X} - \theta)}_{\mathcal{N}(0,1)\ \text{ЦПТ}} \cdot \frac{\sqrt{\theta}}{\sqrt{\overline{X}}} \xrightarrow{d} \mathcal{N}(0,1)$$

$$\frac{\sqrt{\theta}}{\sqrt{\overline{X}}} \xrightarrow{\mathbb{P}} 1$$

$$\mathbb{P}(\ ) \xrightarrow{n\to\infty} 1-\alpha$$

6. Пусть $X_1, \dots, X_n$ – выборка из распределения Бернулли $\mathrm{Ber}(\theta)$ с параметром $\theta \in (0,1)$. Постройте асимптотический доверительный интервал уровня доверия $1-\alpha$ для параметра θ.

Пример 12. Пусть дана выборка $(X_1, \dots, X_n)$ из распределения с Бернулли $\mathcal{B}\mathrm{er}(\theta)$, $\theta \in (0,1)$. Оценкой максимального правдоподобия $\widehat{\theta}_{MLE} = \overline{X}$. В данном примере можно не искать асимптотическую дисперсию через информацию Фишера, так как она будет совпадать с дисперсией в распределении $\mathcal{B}\mathrm{er}(\theta)$ по ЦПТ 21

$$\frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\theta(1-\theta)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Теперь рассмотрим

$$\frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\overline{X}(1-\overline{X})}} = \frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\theta(1-\theta)}} \cdot \frac{\sqrt{\theta(1-\theta)}}{\sqrt{\overline{X}(1-\overline{X})}}$$

Пользуясь теоремой о наследовании сходимости 22 и состоятельностью оценки $\widehat{\theta}_{MLE} = \overline{X}$, получаем

$$\sqrt{\overline{X}(1-\overline{X})} \xrightarrow{\mathbb{P}} \sqrt{\theta(1-\theta)}.$$

Тогда по лемме Слуцкого 14 произведение сойдется к $\mathcal{N}(0,1)$. Значит

$$\mathbb{P}_\theta\left(x_{\alpha/2} < \frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\overline{X}(1-\overline{X})}} < -x_{\alpha/2}\right) \xrightarrow{n\to\infty} 1-\alpha,$$

где $x_{\alpha/2}$ есть $\alpha/2$-квантиль распределения $\mathcal{N}(0,1)$. Далее осталось выразить θ и получить асимптотически доверительный интервал для θ

$$\left(\overline{X} + \sqrt{\frac{\overline{X}(1-\overline{X})}{n}}x_{\alpha/2},\ \overline{X} - \sqrt{\frac{\overline{X}(1-\overline{X})}{n}}x_{\alpha/2}\right)$$

7. $X_1, \dots, X_n$ – выборка из равномерного распределения $\mathcal{U}(0,\theta)$, $\theta > 0$. Постройте точный доверительный интервал уровня доверия $1-\alpha$ для параметра θ.