Математическая статистика. ДЗ 11.

Прохоров Юрий, 776

Задача 1

Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — простая выборка, $X_1 \sim \mathcal{N}(\theta_1, \theta_2)$. Построить γ-доверительную область для оценки параметров $\theta = (\theta_1, \theta_2)$.

Решение:

Требуется построить такую область $\Omega_\gamma \in \mathbb{R}^n$, что $\mathbb{P}_\theta\{\mathbf{X} \in \Omega_\gamma\} \geq \gamma$.

Идея: раньше мы строили центральную статистику G_1 и смотрели на вероятность ее попадания в некоторый промежуток. Теперь же мы построим две статистики $G_1(\mathbf{X}, \theta_1, \theta_2)$ и $G_2(\mathbf{X}, \theta_1, \theta_2)$, то есть такие функции, что

1. распределения G_1 и G_2 не зависят от θ,
2. статистики G_1 и G_2 — независимые случайные величины.

Воспользуемся **леммой Фишера**:

- $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^n X_i \sim \mathcal{N}\left(\theta_1, \dfrac{\theta_2}{n}\right)$;
- $\sum\limits_{i=1}^n \dfrac{\left(X_i - \overline{X}\right)^2}{\theta_2} \sim \chi^2(n-1)$;
- случайные величины $\overline{X}$ и $\sum\limits_{i=1}^n \dfrac{\left(X_i - \overline{X}\right)^2}{\theta_2}$ независимы.

Она подсказывает, что можно взять

$$G_1(\mathbf{X}, \theta) = \sqrt{n}\frac{\overline{X} - \theta_1}{\sqrt{\theta_2}} \sim \mathcal{N}(0,1), \qquad G_2(\mathbf{X}, \theta) = \sum_{i=1}^n \frac{\left(X_i - \overline{X}\right)^2}{\theta_2} \sim \chi^2(n-1)$$

Из независимости этих статистик получаем

$$\mathbb{P}_\theta\Big\{u_1 \leq G_1(\mathbf{X}, \theta) \leq u_2,\ v_1 \leq G_2(\mathbf{X}, \theta) \leq v_2\Big\} = \mathbb{P}_\theta\big\{u_1 \leq G_1(\mathbf{X}, \theta) \leq u_2\big\} \cdot \mathbb{P}_\theta\big\{v_1 \leq G_2(\mathbf{X}, \theta) \leq v_2\big\} =$$
$$= \Big[F_{G_1}(u_2) - F_{G_1}(u_1)\Big] \cdot \Big[F_{G_2}(v_2) - F_{G_2}(v_1)\Big]$$

С другой стороны,

$$\mathbb{P}_\theta\Big\{u_1 \leq G_1(\mathbf{X}, \theta) \leq u_2,\ v_1 \leq G_2(\mathbf{X}, \theta) \leq v_2\Big\} =$$
$$= \mathbb{P}_\theta\left\{\underbrace{u_1 \leq \sqrt{n}\frac{\overline{X} - \theta_1}{\sqrt{\theta_2}} \leq u_2, \quad v_1 \leq \sum_{i=1}^n \frac{\left(X_i - \overline{X}\right)^2}{\theta_2} \leq v_2}_{\Omega_\gamma}\right\} \geq \gamma$$

Будем строить для обоих статистик симметричные интервалы, то есть выберем u_1, u_2, v_1, v_2 такими, что

$$F_{G_1}(u_1) + F_{G_1}(u_2) = 1, \qquad F_{G_2}(v_1) + F_{G_1}(v_2) = 1$$

Имеем с учетом этих равенств:

$$F_{G_1}(u_2) - F_{G_1}(u_1) = \gamma_1 \qquad \Longrightarrow \qquad u_2 = \lambda_{\frac{1-\gamma_1}{2}}, \qquad u_1 = -u_2 = -\lambda_{\frac{1-\gamma_1}{2}}$$

$$F_{G_2}(v_2) - F_{G_2}(v_1) = \gamma_2 \qquad \Longrightarrow \qquad v_2 = \mu_{\frac{1-\gamma_2}{2}}, \qquad u_1 = \mu_{\frac{1+\gamma_2}{2}}$$

Здесь λ_α — α-квантиль распределения $\mathcal{N}(0,1)$, а μ_β — β-квантиль распределения $\chi^2(n-1)$.

Нам нужно, чтобы хотя бы было выполнено равенство

$$\Big[F_{G_1}(u_2) - F_{G_1}(u_1)\Big] \cdot \Big[F_{G_2}(v_2) - F_{G_2}(v_1)\Big] = \gamma_1\gamma_2 = \gamma.$$

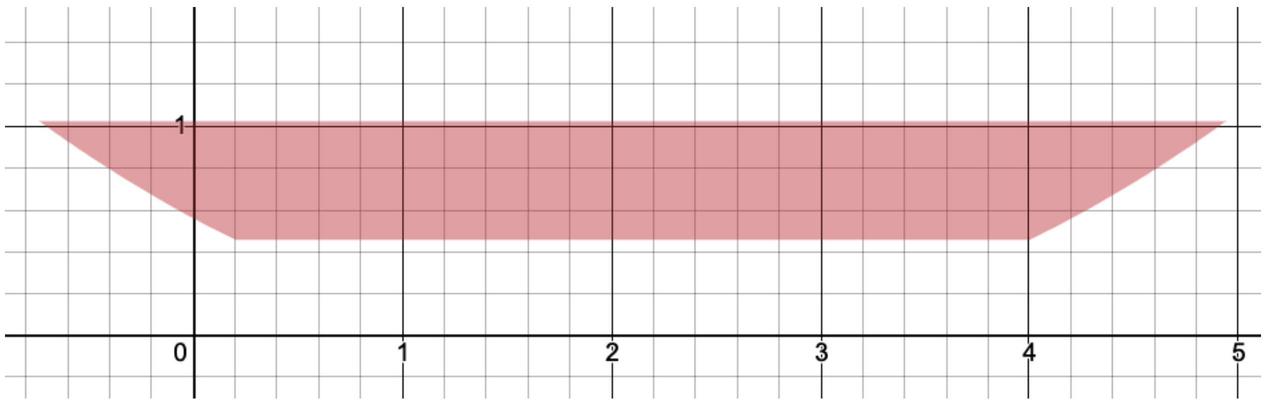
Выбирая разные γ_1 и γ_2, можно строить различные доверительные области. Для определенности возьмем

$$\gamma_1 = \gamma_2 = \sqrt{\gamma}$$

Итого, имеем γ-доверительную область:

$$\Omega_\gamma = \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; -\lambda_{\frac{1-\sqrt{\gamma}}{2}} \le \sqrt{n}\frac{\overline{\mathbf{x}} - \theta_1}{\sqrt{\theta_2}} \le \lambda_{\frac{1-\sqrt{\gamma}}{2}}, \quad \mu_{\frac{1+\sqrt{\gamma}}{2}} \le \sum_{i=1}^{n} \frac{\left(X_i - \overline{\mathbf{x}}\right)^2}{\theta_2} \le v_2 \le \mu_{\frac{1-\sqrt{\gamma}}{2}} \right\}$$

Я сгенерировал выборку из $n = 200$ элементов из распределения $\mathcal{N}(2,1)$. Для $\gamma = 0.99$ доверительная область выглядит вот так:



Задача 2

Доказать, что если оптимальная в среднем квадратичном оценка существует, то она единственна.

Если модель регулярна, то несмещенная оценка $\widehat{\theta}(\mathbf{X})$ называется *эффективной*, если на ней неравенство Рао-Крамера

$$\mathbb{V}_\theta\big[\widehat{\theta}(\mathbf{X})\big] \ge I_n^{-1}$$

обращается в равенство.

Несмещенная оценка $\widehat{\theta}(\mathbf{X})$ называется *оптимальной в среднем квадратичном*, если она имеет минимальную возможную дисперсию.

Решение:

Пусть $\widehat{\theta}_1(\mathbf{X})$ и $\widehat{\theta}_2(\mathbf{X})$ — две несмещенные оптимальные в среднем квадратичном оценки. Докажем, что они равны почти наверное.

Для этого необходимо и достаточно, чтобы

$$\mathbb{V}_\theta\big[\widehat{\theta}_1 - \widehat{\theta}_2\big] = 0$$

В силу несмещенности и оптимальности:

$$\mathbb{E}_\theta\widehat{\theta}_1 = \mathbb{E}_\theta\widehat{\theta}_2 = \theta, \qquad \mathbb{V}_\theta\widehat{\theta}_1 = \mathbb{V}_\theta\widehat{\theta}_2 = \sigma^2$$

Рассмотрим еще одну оценку

$$\widehat{\theta}(\mathbf{X}) = \frac{\widehat{\theta}_1(\mathbf{X}) + \widehat{\theta}_2(\mathbf{X})}{2}, \qquad \mathbb{E}_\theta \widehat{\theta} = \theta$$

Ее дисперсия, с учетом неравенства Коши-Буняковского:

$$\mathbb{V}_\theta \widehat{\theta} = \frac{\mathbb{V}_\theta \widehat{\theta}_1 + \mathbb{V}_\theta \widehat{\theta}_2 + 2\operatorname{cov}\big(\widehat{\theta}_1, \widehat{\theta}_2\big)}{4} \le \frac{2\sigma^2 + 2\sqrt{\mathbb{V}_\theta \widehat{\theta}_1 \cdot \mathbb{V}_\theta \widehat{\theta}_2}}{4} = \sigma^2$$

С другой стороны, из оптимальности оценок $\widehat{\theta}_1$ и $\widehat{\theta}_2$ следует, что дисперсия оценки $\widehat{\theta}$ не может быть меньше σ^2. Значит, $\mathbb{V}_\theta \widehat{\theta} = \sigma^2$.

Из теории вероятностей известно, что неравенство Коши-Буняковского обращается в равенство тогда и только тогда, когда случайные величины линейно зависимы, значит,

$$\widehat{\theta}_1 \overset{\text{п.н.}}{=} \alpha \cdot \widehat{\theta}_2 \qquad \Longrightarrow \qquad \mathbb{E}_\theta \widehat{\theta}_1 = \alpha \cdot \mathbb{E}_\theta \widehat{\theta}_2 \qquad \Longrightarrow \qquad \theta = \alpha\theta, \ \ \forall\theta \qquad \Longrightarrow \qquad \alpha = 1$$