Математическая статистика. ДЗ 10.

Прохоров Юрий, 776

Задача 1

Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — простая выборка, $X_1 \sim \mathcal{U}[0, \theta]$. Построить γ-д.и. для оценки параметра θ (можно асимптотический д.и.).
д.и. — доверительный интервал

Решение:

Нам нужно построить две измеримые функции $T_1(\mathbf{X})$ и $T_2(\mathbf{X})$, такие что

$$\mathbb{P}_\theta\big\{T_1(\mathbf{X}) \le \theta \le T_2(\mathbf{X})\big\} \ge \gamma$$

Будем строить *центральную статистику* — функцию $G(\mathbf{X}; \theta)$, обладающую свойствами:

- для любого $\mathbf{X}$ функция $G(\mathbf{X}; \theta)$ строго монотонна и непрерывна по θ;
- распределение $G(\mathbf{X}; \theta)$ не зависит от θ.
  Или предельное распределение $G(\mathbf{X}; \theta)$ при $n \to \infty$ не зависит от θ (если строится асимптотический д.и.).

Тогда можно рассмотреть вероятность

$$\mathbb{P}_\theta\big\{z_1 \le G(\mathbf{X}; \theta) \le z_2\big\} = \Big/ \begin{array}{c}\text{распределение}\\ \text{не зависит от } \theta\end{array} \Big/ = \int\limits_{z_1}^{z_2} f_G(x)\, dx$$

С другой стороны

$$\mathbb{P}_\theta\big\{z_1 \le G(\mathbf{X}; \theta) \le z_2\big\} = \Big/ \begin{array}{c}\text{строго}\\ \text{монотонна по } \theta\\ \text{(пусть возрастает)}\end{array} \Big/ = \mathbb{P}_\theta\big\{ \underbrace{G^{-1}(z_1)}_{T_1(\mathbf{X})} \le \theta \le \underbrace{G^{-1}(z_2)}_{T_2(\mathbf{X})} \big\}$$

В последнем выражении можно положить

$$T_1(\mathbf{X}) = G^{-1}(z_1), \qquad T_2(\mathbf{X}) = G^{-1}(z_2),$$

а константы z_1, z_2 найти из условия

$$\mathbb{P}_\theta\big\{T_1(\mathbf{X}) \le \theta \le T_2(\mathbf{X})\big\} = \int\limits_{z_1}^{z_2} f_G(x)\, dx = F_G(z_2) - F_G(z_1) \ge \gamma$$

Часто удобно искать *симметричный* доверительный интервал, то есть дополнительно накладывать на z_1, z_2 ограничение

$$\int_{-\infty}^{z_1} f_G(x)\, dx = \int_{z_2}^{+\infty} f_G(x)\, dx, \qquad \Longleftrightarrow \qquad F_G(z_1) = 1 - F_G(z_2)$$

то есть, что z_1 и z_2 отсекают равные квантили слева и справа.
Такой выбор, однако, не во всех случаях дает оптимальный д.и. (то есть д.и. наименьшей длины).

Рассмотрим несколько способов решения этой задачи.

(a) Применим ЦПТ:

$$\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E} X_i}{\sqrt{\sum_{i=1}^n \mathbb{V} X_i}} = \frac{\overline{X}_n - \frac{\theta}{2}}{\frac{1}{n}\sqrt{n\frac{\theta^2}{12}}} = \underbrace{\sqrt{3n}\left(\frac{2\overline{X}_n}{\theta} - 1\right)}_{G(\mathbf{X};\theta)} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, 1),$$

где $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ — выборочное среднее.

Легко видеть, что выбранная функция $G(\mathbf{X};\theta)$ удовлетворяет определению центральной статистики.

$$G(\mathbf{X};\theta) = \sqrt{3n}\left(\frac{2\overline{X}_n}{\theta} - 1\right) \qquad \Longrightarrow \qquad G^{-1}(z) = \frac{2\overline{X}_n}{1+\frac{z}{\sqrt{3n}}} \approx 2\overline{X}_n\left(1 - \frac{z}{\sqrt{3n}}\right)$$

У нас $G(\mathbf{X};\theta)$ строго убывает по θ, поэтому

$$\mathbb{P}_\theta\big\{z_1 \le G(\mathbf{X};\theta) \le z_2\big\} = \mathbb{P}_\theta\big\{G^{-1}(z_2) \le \theta \le G^{-1}(z_1)\big\} \xrightarrow{n\to\infty} \int\limits_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt = \Phi(z_2) - \Phi(z_1)$$

Будем искать симметричный асимптотический д.и., поэтому положим $z_2 = -z_1 > 0$, и найдем их из условия

$$\Phi(z_2) - \Phi(-z_2) \ge \gamma \qquad \overset{\text{равенство}}{\Longrightarrow} \qquad \Phi(z_2) = \frac{1+\gamma}{2} \qquad \Longrightarrow \qquad z_2 = \lambda_{\frac{1-\gamma}{2}}$$

Итого, асимптотический γ-д.и.:

$$\left[2\overline{X}_n\left(1 + \frac{\lambda_{\frac{1-\gamma}{2}}}{\sqrt{3n}}\right)^{-1},\ \ 2\overline{X}_n\left(1 - \frac{\lambda_{\frac{1-\gamma}{2}}}{\sqrt{3n}}\right)^{-1}\right]$$

Его длина:

$$T_2 - T_1 \approx 2\overline{X}_n\left(1 + \frac{\lambda_{\frac{1-\gamma}{2}}}{\sqrt{3n}}\right) - 2\overline{X}_n\left(1 - \frac{\lambda_{\frac{1-\gamma}{2}}}{\sqrt{3n}}\right) = \frac{4\overline{X}_n\lambda_{\frac{1-\gamma}{2}}}{\sqrt{3n}} = O\left(\frac{1}{\sqrt{n}}\right)$$

(b) Сначала докажем вспомогательное утверждение.

Утверждение

> Пусть $X_1,\dots,X_n \sim \mathcal{U}[0,\theta]$ — i.i.d., $X_{(n)} = \max_{i=\overline{1,n}} X_i$. Тогда
>
> $$\frac{n(\theta - X_{(n)})}{\theta} \xrightarrow[n\to\infty]{d} \mathrm{Exp}(1)$$

Доказательство.

$$\mathbb{P}\left\{\frac{n(\theta - X_{(n)})}{\theta} < x\right\} = \mathbb{P}\left\{X_{(n)} > \theta - \frac{\theta x}{n}\right\} = 1 - \prod_{i=1}^n \mathbb{P}\left\{X_i < \theta - \frac{\theta x}{n}\right\} =$$
$$= 1 - \left(1 - \frac{x}{n}\right)^n \xrightarrow{n\to\infty} 1 - e^{-x} = F_{\mathrm{Exp}(1)}(x)$$

□

Видно, что функция

$$G(\mathbf{X};\theta) = n\left(1 - \frac{X_n}{\theta}\right), \qquad G^{-1}(z) = \frac{X_{(n)}}{1 - \frac{z}{n}} \approx X_{(n)}\left(1 + \frac{z}{n}\right)$$

подходит под определение центральной статистики. Она строго возрастает по θ, поэтому

$$\mathbb{P}_\theta\big\{z_1 \le G(\mathbf{X};\theta) \le z_2\big\} = \mathbb{P}_\theta\big\{G^{-1}(z_1) \le \theta \le G^{-1}(z_2)\big\} \xrightarrow{n\to\infty} \int\limits_{z_1}^{z_2} e^{-t}\,dt = F(z_2) - F(z_1),$$

где $F(x) = 1 - e^{-x}$ — функция распределения $\mathrm{Exp}(1)$.

Симметричный д.и. здесь искать не будем, потому что тогда его левая граница будет лежать левее $X_{(n)}$, а мы точно знаем, что $\theta \ge X_{(n)}$. Поэтому положим $z_1 = 0$ и найдем z_2 из условия

$$F(z_2) - F(z_1) \ge \gamma \qquad \overset{\text{равенство}}{\Longrightarrow} \qquad z_2 = \lambda_{1-\gamma} = \ln\frac{1}{1-\gamma}, \qquad z_1 = 0$$

Итого, асимптотический γ-д.и.:

$$\left[X_{(n)},\quad X_{(n)}\left(1-\frac{\ln\frac{1}{1-\gamma}}{n}\right)^{-1}\right]$$

Его длина:

$$T_2-T_1\approx X_{(n)}\left(1+\frac{\ln\frac{1}{1-\gamma}}{n}\right)-X_{(n)}=\frac{X_{(n)}\ln\frac{1}{1-\gamma}}{n}=O\left(\frac{1}{n}\right)$$

(c) Построим неасимптотический доверительный интервал. Попробуем использовать величину $X_{(n)}$ и построить интервал вокруг нее. Возьмем в качестве центральной статистики:

$$G(\mathbf{X};\theta)=\frac{X_{(n)}}{\theta}=\max_{i=\overline{1,n}}\frac{X_i}{\theta}\quad - \quad \text{не зависит от } \theta, \text{ т.к. } \frac{X_i}{\theta}\sim\mathcal{U}[0,1]$$

$$G^{-1}(z)=\frac{X_{(n)}}{z}$$

Функция $G(\mathbf{X};\theta)$ строго убывает по θ, поэтому

$$\mathbb{P}_\theta\big\{z_1\le G(\mathbf{X};\theta)\le z_2\big\}=\mathbb{P}_\theta\big\{G^{-1}(z_2)\le\theta\le G^{-1}(z_1)\big\}\xrightarrow{n\to\infty}F_G(z_2)-F_G(z_1)$$

Аналогично пункту (b), не будем строить симметричный критерий. Положим нижнюю границу доверительного интервала равной $X_{(n)}$, т.е. $z_2=1$. Число z_1 найдем из условия

$$F_G(z_2)-F_G(z_1)\ge\gamma\quad\overset{\text{равенство}}{\Longrightarrow}\quad 1-\mathbb{P}\left\{\frac{X_{(n)}}{\theta}<z_1\right\}=1-(z_1)^n=\gamma\quad\Longrightarrow\quad z_1=(1-\gamma)^{1/n}$$

Итого, γ-д.и.:

$$\Big[X_{(n)},\quad X_{(n)}(1-\gamma)^{-1/n}\Big]$$

Его длина:

$$T_2-T_1=X_{(n)}\Big((1-\gamma)^{-1/n}-1\Big)=X_{(n)}\left[\exp\left(\frac{1}{n}\ln\frac{1}{1-\gamma}\right)-1\right]\approx X_{(n)}\frac{\ln\frac{1}{1-\gamma}}{n}=O\left(\frac{1}{n}\right)$$

Стоит отметить, что длина этого доверительного интервала вообще равна длине интервала в пункте (b) при $n\to\infty$.