Математическая статистика. ДЗ 1.

Прохоров Юрий, 776

Задача 1

(a) Сгенерировать выборку $X_1, \dots, X_n \sim \mathrm{Be}(\theta)$, где $n = 100,\ \theta = \frac{1}{3}$.

(b) Построить графики теоретической и эмпирической функций распределения.

(c) Найти предельное распределение

$$\xi_n = \sqrt{n} \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right|$$

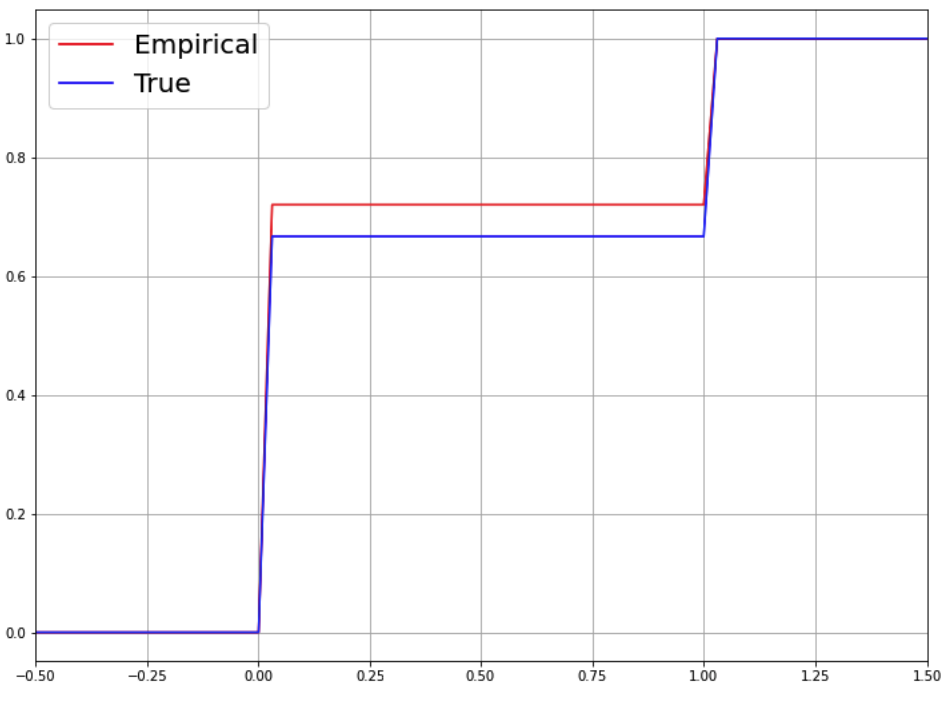
(d) Вычислить $\lim\limits_{n \to \infty} \mathbb{P}\left\{\xi_n < \Delta\right\}$.

(e) Положив значение предела в (d) равным $\gamma = 0.9$, найти соответствующее $\Delta = \Delta(\gamma, \theta)$. Сделать оценку $\Delta(\gamma, \theta) \le \overline{\Delta}(\gamma)$.

(f) Оценить ошибку приближения $\mathbb{P}\left\{\xi_n < \Delta\right\}$ предельным значением по неравенству Берри-Эссеена.

Решение:

(a), (b) Сгенерировал в Python выборку.



(c) Перепишем

$$\xi_n = \sqrt{n} \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right| = \sqrt{n} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\{X_k = 0\} - (1 - \theta) \right| = \left| \frac{\sum_{k=1}^{n} X_k - n\theta}{\sqrt{n}} \right| = |\eta_n|$$

По центральной предельной теореме,

$$\eta_n = \sqrt{\theta(1-\theta)} \cdot \frac{\sum_{k=1}^{n} X_k - n \cdot \mathbb{E}X_1}{\sqrt{n \cdot \mathbb{V}X_1}} \xrightarrow[n \to \infty]{d} \eta, \qquad \eta \sim \mathcal{N}\big(0, \theta(1-\theta)\big)$$

Тогда (несложно показать по определению)

$$\xi_n = |\eta_n| \xrightarrow[n\to\infty]{d} |\eta|, \qquad \eta \sim \mathcal{N}\big(0, \theta(1-\theta)\big)$$

(d) Обозначим функцию распределения $\mathcal{N}(0,1)$ через $\Phi(x)$. Тогда

$$\eta \sim \mathcal{N}\big(0, \theta(1-\theta)\big) \qquad \Longrightarrow \qquad F_\eta(x) = \Phi\left(\frac{x}{\sqrt{\theta(1-\theta)}}\right)$$

Тогда имеем

$$\mathbb{P}\{\xi_n < \Delta\} = \mathbb{P}\{-\Delta < \eta_n < \Delta\} = F_{\eta_n}(\Delta) - F_{\eta_n}(-\Delta) = \Big/ \eta_n \xrightarrow[n\to\infty]{d} \eta \Big/ \xrightarrow[n\to\infty]{}$$

$$\xrightarrow[n\to\infty]{} F_\eta(\Delta) - F_\eta(-\Delta) = \Phi\left(\frac{\Delta}{\sqrt{\theta(1-\theta)}}\right) - \Phi\left(\frac{-\Delta}{\sqrt{\theta(1-\theta)}}\right)$$

(e) Пусть этот предел равен $\gamma = 0.9$.

Так как для любого $h \geq 0$ выполнено $\Phi(h) + \Phi(-h) = 1$, то получаем

$$2\Phi\left(\frac{\Delta}{\sqrt{\theta(1-\theta)}}\right) = \gamma + 1 \qquad \Longrightarrow \qquad \Delta(\gamma, \theta) = \sqrt{\theta(1+\theta)}\, \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$$

Так как на самом деле мы не знаем истинного θ, то надо сделать оценку, не зависящую от θ:

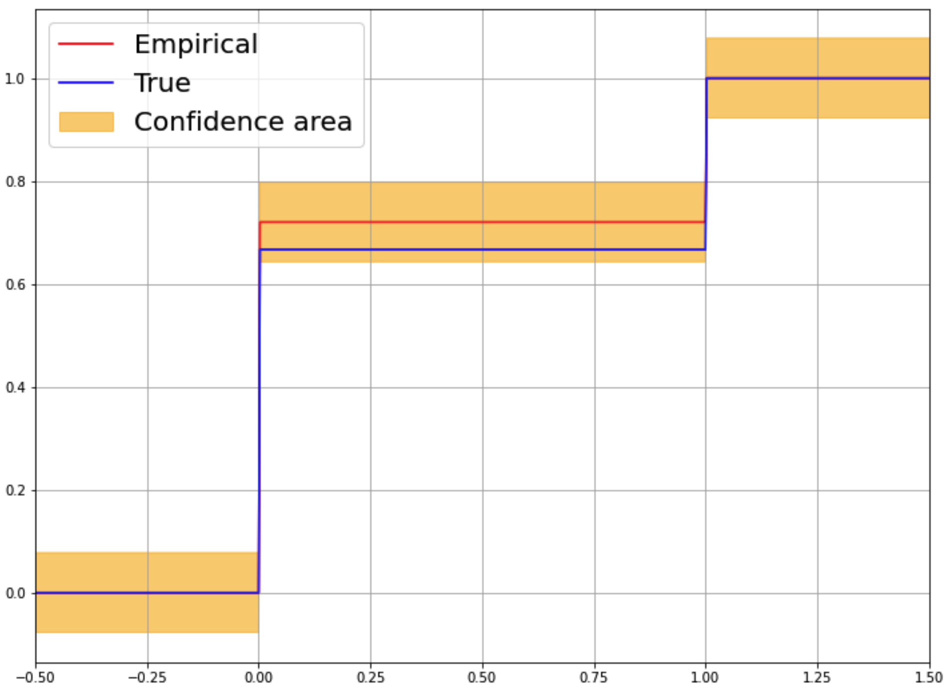
$$\Delta(\gamma, \theta) = \sqrt{\theta(1+\theta)}\, \Phi^{-1}\left(\frac{1+\gamma}{2}\right) \leq \frac{1}{2}\Phi^{-1}\left(\frac{1+\gamma}{2}\right) = \overline{\Delta}(\gamma)$$

Для конкретных чисел $\gamma = 0.9,\ \theta = \frac{1}{3}$, получается, что $\Delta \approx 0.775,\ \overline{\Delta} \approx 0.822$.

Теперь можно построить доверительную область:

$$\left(\widehat{F}_n(x) - \frac{\overline{\Delta}}{\sqrt{n}},\ \widehat{F}_n(x) + \frac{\overline{\Delta}}{\sqrt{n}}\right)$$

По построению, при достаточно больших n, истинная функция распределения *полностью* лежит в этой области с вероятностью $\gamma = 0.9$. При $n = 100$ получается такая область:



(f) Обозначим предельное значение:

$$G(\Delta) = \lim_{n\to\infty} \mathbb{P}\left\{\xi_n < \Delta\right\} = \Phi\left(\frac{\Delta}{\sqrt{\theta(1-\theta)}}\right) - \Phi\left(\frac{-\Delta}{\sqrt{\theta(1-\theta)}}\right)$$

Используем неравенство Берри-Эссеена:

> **Неравенство Берри-Эссеена**
>
> Пусть $\{\xi_n\}_{n=1}^{\infty}$ — i.i.d., $\mathbb{E}\xi_1 = 0,\ \mathbb{V}\xi_1 = \sigma^2,\ \mathbb{E}|\xi_1|^3 \le \rho$. Тогда если $\zeta_n = \dfrac{1}{\sigma\sqrt{n}}\sum\limits_{k=1}^{n}\xi_k$, то
>
> $$\sup_{x\in\mathbb{R}} \left|F_{\zeta_n}(x) - \Phi(x)\right| \le \frac{C\rho}{\sigma^{3/2}\sqrt{n}},$$
>
> где $C < 0.48$ — константа (точное значение неизвестно), а $\Phi(x)$ — функция распределения $\mathcal{N}(0,1)$.

У нас рассматриваются случайные величины $(X_n - \theta)$:

$$\mathbb{E}[X_1 - \theta] = 0, \qquad \mathbb{V}[X_1 - \theta] = \theta(1-\theta), \qquad \mathbb{E}|X_1 - \theta|^3 = \theta(1-\theta)(1 - 2\theta + 2\theta^2) \le \frac{1}{8}$$

$$\frac{1}{\sqrt{\theta(1-\theta)}}\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - \theta) = \frac{\eta_n}{\sqrt{\theta(1-\theta)}}$$

Тогда по неравенству Берри-Эссеена:

$$\sup_x \left|F_{\frac{\eta_n}{\sqrt{\theta(1-\theta)}}}(x) - \Phi(x)\right| \le \frac{C\cdot\frac{1}{8}}{[\theta(1-\theta)]^{\frac{3}{4}}\sqrt{n}} \le \frac{1}{16\big[\theta(1-\theta)\big]^{3/4}}\frac{1}{\sqrt{n}}$$

Тогда имеем

$$\begin{aligned}
\Big|\mathbb{P}\{\xi_n < \Delta\} - G(\Delta)\Big| &= \Big|\big(F_{\eta_n}(\Delta) - F_{\eta_n}(-\Delta)\big) - \big(F_{\eta}(\Delta) - F_{\eta}(-\Delta)\big)\Big| \le \\
&\le \big|F_{\eta_n}(\Delta) - F_{\eta}(\Delta)\big| + \big|F_{\eta_n}(-\Delta) - F_{\eta}(-\Delta)\big| \le \\
&\le \left|F_{\frac{\eta_n}{\sqrt{\theta(1-\theta)}}}(\ldots) - \Phi(\ldots)\right| + \left|F_{\frac{\eta_n}{\sqrt{\theta(1-\theta)}}}(\ldots) - \Phi(\ldots)\right| \le \\
&\le \frac{1}{8\big[\theta(1-\theta)\big]^{3/4}}\frac{1}{\sqrt{n}}
\end{aligned}$$