



Универзитет у Београду
Математички факултет

ВЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ)

1. јул 2024. године

1. Општи члан X_n низа независних случајних величина има експоненцијалну $\mathcal{E}(1)$ расподелу. Ако је

$$Y_n = n \cdot \min\{X_1, X_2, \dots, X_{n^2}\},$$

испитати све четири врсте конвергенције низа случајних величина (Y_n) .

2. Из популације чије обележје X има нормалну $\mathcal{N}(\theta, \theta^2)$, $\theta > 0$, расподелу извучен је узорак обима n . За оцену непознатог параметра θ на основу тог узорака предложене су следеће оцене:

$$T_1 = \bar{X}_n, \quad T_2 = k\sqrt{n\bar{S}_n^2}, \quad T = \alpha T_1 + (1 - \alpha)T_2,$$

где је $k \in \mathbb{R}$, $\alpha \in [0, 1]$. Одредити k тако да T_2 буде непристрасна оцена, а затим наћи α за које оцена T има најмању средње квадратну грешку.

3. Број пријављених кандидата за упис на Математички факултет на студијски програм Информатика има нормалну $\mathcal{N}(m_1, \sigma^2)$ расподелу, а на програм Математика нормалну $\mathcal{N}(m_2, 0.45\sigma^2)$ расподелу, где је $\sigma > 1$ и познато. На случајном узорку од 12 година добијено је да је просечан број пријављених кандидата на студијском програму Информатика 294.6, док је на случајном узорку од 10 година добијено да је просечан број пријављених кандидата на студијском програму Математика 191.3. Тестира се хипотеза да је разлика у средњим вредностима броја пријављених кандидата на студијским програмима Информатика и Математика једнака 100, против алтернативе да је разлика већа од 100. Формиран је тест такав да је вероватноћа тачног прихватања нулте хипотезе 7 пута већа од вероватноће погрешног одбацивања нулте хипотезе. Такође, моћ предложеног теста, при алтернативи да је разлика у средњим вредностима броја пријављених кандидата на студијским програмима Информатика и Математика једнака $105 - 0.25\sigma$, износи 0.85. Одредити који закључак треба донети на основу датих узорака.

БЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ) - Писмени испит 1. јул 2024.

Решења задатака

1. Функција расподеле случајне величине Y_n дата је са

$$\begin{aligned} F_{Y_n}(y) &= P\{Y_n \leq y\} = P\{n \min\{X_1, X_2, \dots, X_{n^2}\} \leq y\} = P\left\{\min\{X_1, X_2, \dots, X_{n^2}\} \leq \frac{y}{n}\right\} \\ &= 1 - P\left\{\min\{X_1, X_2, \dots, X_{n^2}\} > \frac{y}{n}\right\} = 1 - \left(P\left\{X_1 > \frac{y}{n}\right\}\right)^{n^2} \\ &= 1 - \left(e^{-\frac{y}{n}}\right)^{n^2} = 1 - e^{-ny}, \quad y \geq 0. \end{aligned}$$

Следи да је

$$F_{Y_n}(z) = \begin{cases} 0, & y < 0 \\ 1 - e^{-\lambda ny}, & y \geq 0 \end{cases} \xrightarrow{n \rightarrow \infty} \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases} = F(z),$$

где је $F(z)$ функција расподеле (дегенерисане) случајне величине 0. Из претходног закључујемо $Y_n \xrightarrow{D} 0$, $n \rightarrow \infty$, као и $Y_n \xrightarrow{P} 0$, $n \rightarrow \infty$. Испитајмо сада скоро сигурну конвергенцију. Како је за свако $\varepsilon > 0$

$$P\{|Y_n| \geq \varepsilon\} = P\{n \min\{X_1, X_2, \dots, X_{n^2}\} \geq \varepsilon\} = e^{-\varepsilon n} = \left(\frac{1}{e^\varepsilon}\right)^n,$$

то је $\sum_{n=1}^{\infty} P\{|Y_n| \geq \varepsilon\} < \infty$, па закључујемо $Z_n \xrightarrow{c.c.} 0$, $n \rightarrow \infty$. У циљу испитивања средње квадратне конвергенције рачунамо $E(Y_n^2)$. Имамо да је редом

$$\begin{aligned} E(Y_n)^2 &= \int_0^{\infty} y^2 f_{Y_n}(y) dy = \int_0^{\infty} y^2 n e^{-ny} dy = \left[\begin{matrix} ny = u \\ ndx = du \end{matrix} \right] \\ &= \frac{1}{n^2} \int_0^{\infty} u^2 e^{-u} du = \frac{1}{n^2} \Gamma(3) \\ &= \frac{2}{n^2}. \end{aligned}$$

Како $E(Y_n)^2 = \frac{2}{n^2} \rightarrow 0$, $n \rightarrow \infty$, следи да низ (Y_n) конвергира средње квадратно.

2. Прво ћемо одредити k тако да оцена T_2 буде непристрасна. Најпре имамо да је

$$\begin{aligned} E\left(\sqrt{n\bar{S}_n^2}\right) &= \theta E\left(\sqrt{\frac{n\bar{S}_n^2}{\theta^2}}\right) = \theta \int_0^{\infty} x^{\frac{1}{2}} \frac{x^{\frac{n-1}{2}-1} e^{-\frac{x}{\theta^2}}}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}}} dx \\ &= \theta \frac{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}}} \int_0^{\infty} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{\theta^2}}}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} dx \\ &= \theta \sqrt{2} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}, \end{aligned}$$

при чему смо искористили да случајна величина $\frac{n\bar{S}_n^2}{\theta^2}$ има χ_{n-1}^2 расподелу, па директно следи да је за $k = \frac{1}{\sqrt{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$ оцена T_2 непристрасна. Даље, и оцена T_1 је такође непристрасна. Заиста, важи да је

$$E(T_1) = E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} n E X_1 = \theta.$$

Одредимо сада дисперзије ових оцена. Имамо да је

$$D(T_1) = D(\bar{X}_n) = D\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} n D X_1 = \frac{\theta^2}{n},$$

$$\begin{aligned}
D(T_2) &= D\left(k\sqrt{n\bar{S}_n^2}\right) = E\left(k^2 n \bar{S}_n^2\right) - \left(E\left(k\sqrt{n\bar{S}_n^2}\right)\right)^2 \\
&= k^2 \theta^2 E\left(\frac{n\bar{S}_n^2}{\theta^2}\right) - \theta^2 = k^2 \theta^2 (n-1) - \theta^2 \\
&= (k^2(n-1) - 1)\theta^2,
\end{aligned}$$

Како су и T_1 и T_2 непристрасне оцене параметра θ , то је

$$E(T) = E(\alpha T_1 + (1-\alpha)T_2) = \alpha E T_1 + (1-\alpha)E T_2 = \alpha \theta + (1-\alpha)\theta = \theta,$$

а из независности случајних величина $\bar{X}_n$ и $\bar{S}_n^2$, следи независност оцена T_1 и T_2 , па је

$$\begin{aligned}
MSE(T) &= D(T) + (E(T) - \theta)^2 = D(T) \\
&= D(\alpha T_1 + (1-\alpha)T_2) = \alpha^2 D(T_1) + (1-\alpha)^2 D(T_2) \\
&= \alpha^2 (DT_1 + DT_2) - 2\alpha DT_2 + DT_2.
\end{aligned}$$

Из

$$\frac{d}{d\alpha} MSE_T(\alpha) = \frac{d}{d\alpha} (\alpha^2 (DT_1 + DT_2) - 2\alpha DT_2 + DT_2) = 2\alpha (DT_1 + DT_2) - 2DT_2 = 0,$$

следи да је тражено α дато са

$$\alpha = \frac{DT_2}{DT_1 + DT_2} = \frac{(k^2(n-1) - 1)\theta^2}{\left(\frac{1}{n} + k^2(n-1) - 1\right)\theta^2} = \frac{k^2(n-1) - 1}{\frac{1}{n} + k^2(n-1) - 1}.$$

3. Нека су X и Y обележја које представља број пријављених кандидата на студијске програме Информатика и Математика, редом. Тада је $X \in \mathcal{N}(m_1, \sigma^2)$ и $Y \in \mathcal{N}(m_2, 0.45\sigma^2)$, те је на основу датог узорка $\bar{x}_{12} = 294.6$ и $\bar{y}_{10} = 191.3$. Потребно је да тестирамо $H_0(m_1 - m_2 = 100)$ против $H_1(m_1 - m_2 > 100)$, па посматрамо критичну област

$$W = \{\bar{x}_{12} - \bar{y}_{10} \geq c\}.$$

Како су случајне величине $\bar{X}_{12}$ и $\bar{Y}_{10}$ независне и $\bar{X}_{12} \in \mathcal{N}\left(m_1, \frac{\sigma^2}{12}\right)$ и $\bar{Y}_{10} \in \mathcal{N}\left(m_2, \frac{0.45\sigma^2}{10}\right)$, то је $\bar{X}_{12} - \bar{Y}_{10} \in \mathcal{N}(m_1 - m_2, 0.128\sigma^2)$. Са једне стране имамо да је

$$P_{H_0}\{X \notin W\} = 7P_{H_0}\{X \in W\} \implies 1 - P_{H_0}\{X \in W\} = 7P_{H_0}\{X \in W\}$$

па је

$$\begin{aligned}
P_{H_0}\left\{\frac{\bar{X}_{12} - \bar{Y}_{10} - 100}{\sqrt{0.128\sigma^2}} \geq \frac{c - 100}{\sqrt{0.128\sigma^2}}\right\} &= 0.125, \\
\frac{c - 100}{\sqrt{0.128\sigma^2}} &= \Phi^{-1}(0.875) = 1.15, \\
c &= 100 + 0.412\sigma.
\end{aligned} \tag{1}$$

Са друге стране, при $H_1(m_1 - m_2 = 105 - 0.25\sigma)$, моћ теста је

$$\begin{aligned}
P_{H_1}\{X \in W\} &= 0.85, \\
P_{H_1}\left\{\frac{\bar{X}_{12} - \bar{Y}_{10} - 105 + 0.25\sigma}{\sqrt{0.128\sigma^2}} \geq \frac{c - 105 + 0.25\sigma}{\sqrt{0.128\sigma^2}}\right\} &= 0.125, \\
\frac{c - 105 + 0.25\sigma}{\sqrt{0.128\sigma^2}} &= \Phi^{-1}(0.15) = -1.04, \\
c &= 105 - 0.622\sigma.
\end{aligned} \tag{2}$$

Из (1) и (2) следи да је

$$100 + 0.412\sigma = 105 - 0.622\sigma$$

тј.

$$\sigma = 4.836.$$

Сада је $c = 100 + 0.412\sigma = 100 + 0.412 \cdot 4.836 = 101.992$, па је критична област

$$W = \{\bar{x}_{12} - \bar{y}_{10} \geq 101.992\}.$$

На основу података из задатка имамо да је $\bar{x}_{12} - \bar{y}_{10} = 103.3$, па реализовани узорци упадају у критичну област и одбацујемо нулту хипотезу.