



Универзитет у Београду
Математички факултет

БЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ)

8. јун 2024. године

1. Општи члан X_n низа независних случајних величина има униформну $\mathcal{U}(0, 1)$ расподелу. Ако је

$$Y_n = \ln \left(\prod_{k=1}^n \frac{1}{e^{X_k}} \right)^{\frac{1}{\sqrt{n}}},$$

испитати конвергенцију у расподели низа случајних величина (Y_n) .

2. Нека обележје X има закон расподеле

$$\begin{pmatrix} -1 & 0 & 2 \\ \frac{\theta(1-\theta)}{4} & \frac{\theta(1-\theta)}{4} & 1 - \frac{\theta(1-\theta)}{2} \end{pmatrix},$$

где је $\theta \in (0, 1)$. За оцену непознатог параметра $\theta(1 - \theta)$ на основу узорка обима n предложене су две оцене T_1 и T_2 , где је T_1 оцена добијена методом максималне веродостојности и $T_2 = \frac{4}{5}(2 - \bar{X}_n)$. Одредити која од предложених оцена је боља у средње квадратном смислу, а затим испитати да ли је нека од њих ефикасна.

3. Рачунарска лабораторија Математичког факултета тврди да у току испитног рока студенти проведу просечно $2\sqrt{\pi}$ сати дневно на сајту факултета. За густину расподеле обележја X , које представља време које студент проведе дневно на сајту, важи да је

$$f(x; \theta) = \frac{2}{\theta^2} x e^{-\left(\frac{x}{\theta}\right)^2}, \quad x \geq 0,$$

где је $\theta > 0$ непознат параметар. Посматрано је 7 случајно одабраних студената и забележено је колико просечно времена су провели на сајту факултета у току једног дана: 0.71, 4.21, 5.59, 3.17, 3.42, 2.74, 3.48. На основу добијеног узорка, са прагом значајности 0.1 тестирати најмоћнијим тестом да ли је рачунарска лабораторија у праву или је просечно време проведено на сајту факултета у току једног дана веће.

Решења задатака

1. За низ случајних величина (Y_n) важи да је

$$Y_n = \ln \left(\prod_{k=1}^n \frac{1}{e^{X_k}} \right)^{\frac{1}{\sqrt{n}}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \ln \left(\frac{1}{e^{X_k}} \right) = \frac{1}{\sqrt{n}} \sum_{k=1}^n -(1 + \ln X_k) = \frac{1}{\sqrt{n}} \sum_{k=1}^n Z_k,$$

где је $Z_k = -(1 + \ln X_k)$. Функција расподеле случајне величине Z_k дата је са:

$$\begin{aligned} F_{Z_k}(z) &= P\{Z_k \leq z\} = P\{-(1 + \ln X_k) \leq z\} \\ &= P\{X_k > e^{-(z+1)}\} = 1 - P\{X_k \leq e^{-(z+1)}\} \\ &= 1 - e^{-(z+1)}, \quad z \geq -1, \end{aligned}$$

па је одговарајућа густина

$$f_{Z_k}(z) = e^{-(z+1)}, \quad z \geq -1.$$

Сада је

$$\begin{aligned} EZ_k &= \int_{-1}^{+\infty} z e^{-(z+1)} dz = \int_0^{+\infty} (t-1) e^{-t} dt = \Gamma(2) - \Gamma(1) = 0, \\ EZ_k^2 &= \int_{-1}^{+\infty} z^2 e^{-(z+1)} dz = \int_0^{+\infty} (t-1)^2 e^{-t} dt = \Gamma(3) - 2\Gamma(2) + \Gamma(1) = 1, \end{aligned}$$

па је $DZ_k = EZ_k^2 - (EZ_k)^2 = 1$. Како је Z_n низ независних случајних величина са истом расподелом и коначним математичким очекивањем и дисперзијом, то на основу централне граничне теореме следи да је

$$\frac{\sum_{k=1}^n Z_k - E\left(\sum_{k=1}^n Z_k\right)}{\sqrt{D\left(\sum_{k=1}^n Z_k\right)}} = \frac{\sum_{k=1}^n Z_k}{\sqrt{n}} \xrightarrow{D} Z^*, \quad Z^* \in \mathcal{N}(0, 1).$$

Одавде следи да низ случајних величина (Y_n) у расподели конвергира ка случајној величини Z^* , где је $Z^* \in \mathcal{N}(0, 1)$.

Напомена: Задатак је могуће решити и применом теореме непрекидности. Приметимо да је

$$\varphi_{\ln X_k}(t) = E(e^{it \ln X_k}) = E(X_k^{it}) = \int_0^1 x^{it} dx = \frac{1}{1+it}.$$

Даље, имамо

$$\begin{aligned} \varphi_{Y_n}(t) &= \varphi_{-\sqrt{n} - \frac{1}{\sqrt{n}} \sum_{k=1}^n \ln X_k}(t) = e^{-i\sqrt{n}t} \prod_{k=1}^n \varphi_{\ln X_k} \left(-\frac{t}{\sqrt{n}} \right) \\ &= e^{-i\sqrt{n}t} \frac{1}{\left(1 - \frac{it}{\sqrt{n}}\right)^n} = e^{-i\sqrt{n}t} e^{-n \ln \left(1 - \frac{it}{\sqrt{n}}\right)} \\ &= e^{-i\sqrt{n}t} e^{-n \left(-\frac{it}{\sqrt{n}} + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)} \\ &= e^{-\frac{t^2}{2}}, \quad n \rightarrow \infty, \end{aligned}$$

при чему смо искористили да је $\ln(1-x) = -x - \frac{x^2}{2} + o(x^2)$, $x \rightarrow 0$. Додатно, функција $\varphi_{Y_n}(t)$ је непрекидна у 0, па на основу теореме непрекидности следи да Y_n конвергира у расподели ка случајној величини која има нормалну $\mathcal{N}(0, 1)$ расподелу.

2. Закон расподеле се може записати као

$$\begin{aligned} p(x; \theta) &= \left(\frac{\theta(1-\theta)}{4} \right)^{I\{x=-1\}} \left(\frac{\theta(1-\theta)}{4} \right)^{I\{x=0\}} \left(1 - \frac{\theta(1-\theta)}{2} \right)^{I\{x=2\}} \\ &= \left(\frac{\theta(1-\theta)}{4} \right)^{1-I\{x=2\}} \left(1 - \frac{\theta(1-\theta)}{2} \right)^{I\{x=2\}}, \end{aligned}$$

или

$$p(x; \alpha) = \left(\frac{\alpha}{4} \right)^{1-I\{x=2\}} \left(1 - \frac{\alpha}{2} \right)^{I\{x=2\}},$$

где је $\alpha = \theta(1-\theta)$.

Претпоставимо да имамо реализован узорак $(x_1, x_2, \dots, x_n)$. Функција веродостојности је

$$\begin{aligned} L(\alpha) &= \prod_{k=1}^n p(x_k; \alpha) = \prod_{k=1}^n \left(\frac{\alpha}{4} \right)^{1-I\{x_k=2\}} \left(1 - \frac{\alpha}{2} \right)^{I\{x_k=2\}} \\ &= \left(\frac{\alpha}{4} \right)^{n - \sum_{k=1}^n I\{x_k=2\}} \left(1 - \frac{\alpha}{2} \right)^{\sum_{k=1}^n I\{x_k=2\}}. \end{aligned}$$

па је логаритам веродостојности

$$\ln L(\alpha) = \left(n - \sum_{k=1}^n I\{x_k = 2\} \right) \ln \frac{\alpha}{4} + \sum_{k=1}^n I\{x_k = 2\} \ln \left(1 - \frac{\alpha}{2} \right).$$

Желимо да пронађемо $\alpha > 0$ које максимизује $\ln L(\alpha)$, те стога одређујемо стационарну тачку:

$$\frac{\partial}{\partial \alpha} \ln L(\alpha) = \frac{n - \sum_{k=1}^n I\{x_k = 2\}}{\frac{\alpha}{4}} \cdot \frac{1}{4} + \frac{\sum_{k=1}^n I\{x_k = 2\}}{1 - \frac{\alpha}{2}} \cdot \left(-\frac{1}{2} \right) = 0,$$

одакле је

$$\alpha = \frac{2n - 2 \sum_{k=1}^n I\{x_k = 2\}}{n}.$$

Још треба проверити да ли се у добијеној вредности постиже максимум. Како је

$$\frac{\partial^2}{\partial \alpha^2} \ln L(\alpha) = -\frac{n - \sum_{k=1}^n I\{x_k = 2\}}{\alpha^2} - \frac{\sum_{k=1}^n I\{x_k = 2\}}{(2 - \alpha)^2} < 0,$$

ово заиста јесте тачка у којој функција $\ln L(\alpha)$ достиже максимум. Према томе, на основу овог узорака, оцена методом максималне веродостојности је

$$T_1 = \frac{2n - 2 \sum_{k=1}^n I\{X_k = 2\}}{n}.$$

Очекивања оцена су:

$$ET_1 = E \left(\frac{2n - 2 \sum_{k=1}^n I\{X_k = 2\}}{n} \right) = 2 - \frac{2}{n} \sum_{k=1}^n P\{X_k = 2\} = 2 - 2P\{X_k = 2\} = 2 - 2 \left(1 - \frac{\alpha}{2} \right) = \alpha,$$

$$ET_2 = E \left(\frac{4}{5} (2 - \bar{X}_n) \right) = \frac{8}{5} - \frac{4}{5} E\bar{X}_n = \frac{8}{5} - \frac{4}{5} EX_1 = \frac{8}{5} - \frac{4}{5} \left(2 - \frac{5}{4} \alpha \right) = \alpha,$$

односно, обе оцене су непристрасне оцене параметра α па можемо поредити њихове дисперзије.

Прво нам треба дисперзија од X :

$$DX = EX^2 - (EX)^2 = \frac{\alpha}{4} + 4 - 2\alpha - \left(2 - \frac{5}{4} \alpha \right)^2 = \frac{\alpha(52 - 25\alpha)}{16}.$$

Сада лако рачунамо дисперзије оцена:

$$\begin{aligned} DT_1 &= D \left(2 - \frac{2}{n} \sum_{k=1}^n I\{X_k = 2\} \right) = \frac{4}{n^2} D \left(\sum_{k=1}^n I\{X_k = 2\} \right) \\ &= \frac{4}{n^2} n D(I\{X_k = 2\}) = \frac{4}{n} P\{X_1 = 2\}(1 - P\{X_1 = 2\}) \\ &= \frac{4}{n} \left(1 - \frac{\alpha}{2} \right) \frac{\alpha}{2} = \frac{\alpha(2 - \alpha)}{n}. \end{aligned}$$

$$DT_2 = D \left(\frac{4}{5} (2 - \bar{X}_n) \right) = \frac{16}{25} D\bar{X}_n = \frac{16}{25n} DX_1 = \frac{16}{25n} \frac{\alpha(52 - 25\alpha)}{16} = \frac{\alpha(52 - 25\alpha)}{25n}.$$

Како је за свако $n \in \mathbb{N}$ испуњено $\frac{\alpha(2-\alpha)}{n} < \frac{\alpha(52-25\alpha)}{25n} \iff 50 < 52$, закључујемо да је $DT_1 < DT_2$, па је оцена T_1 боља у средње квадратном смислу од оцене T_2 , па има смисла испитати њену ефикасност. Испитујемо регуларност дате фамилије расподела:

$$\sum_{x \in \{-1, 0, 2\}} \frac{\partial}{\partial \alpha} p(x; \alpha) = \frac{1}{4} + \frac{1}{4} - \frac{1}{2} = 0.$$

Дата фамилија расподела је регуларна, па можемо рачунати Рао-Крамерову доњу границу дисперзија. Имамо да је

$$\ln p(x; \alpha) = (1 - I\{x = 2\}) \ln \frac{\alpha}{4} + I\{x = 2\} \ln \left(1 - \frac{\alpha}{2} \right),$$

$$\frac{\partial}{\partial \alpha} \ln p(x; \alpha) = \frac{1 - I\{x = 2\}}{\alpha} - \frac{I\{x = 2\}}{2 - \alpha},$$

$$\frac{\partial^2}{\partial \alpha^2} \ln p(x; \alpha) = -\frac{1 - I\{x = 2\}}{\alpha^2} - \frac{I\{x = 2\}}{(2 - \alpha)^2},$$

па је

$$\begin{aligned} I(\alpha) &= E \left(-\frac{\partial^2}{\partial \alpha^2} \ln p(X; \alpha) \right) = E \left(\frac{1 - I\{x = 2\}}{\alpha^2} + \frac{I\{x = 2\}}{(2 - \alpha)^2} \right) \\ &= \frac{1 - P\{x = 2\}}{\alpha^2} + \frac{P\{x = 2\}}{(2 - \alpha)^2} = \frac{1 - (1 - \frac{\alpha}{2})}{\alpha^2} + \frac{1 - \frac{\alpha}{2}}{(2 - \alpha)^2} \\ &= \frac{1}{2\alpha} + \frac{1}{2(2 - \alpha)} = \frac{1}{\alpha(2 - \alpha)}. \end{aligned}$$

Одавде добијамо да је доња граница Рао-Крамера једнака

$$G = \frac{1}{nI(\alpha)} = \frac{\alpha(2 - \alpha)}{n}.$$

Према томе, како је $ef(T_1) = \frac{G}{DT_1} = 1$, закључујемо да је ова оцена ефикасна.

3. Просечно време које студенти проведу на сајту дато је са

$$EX = \int_0^{+\infty} \frac{2}{\theta^2} x^2 e^{-\left(\frac{x}{\theta}\right)^2} dx = \theta \int_0^{+\infty} t^{\frac{1}{2}} e^{-t} dt = \theta \Gamma \left(\frac{3}{2} \right) = \theta \frac{\sqrt{\pi}}{2}.$$

Одавде следи да треба тестирати $H_0 : \theta \frac{\sqrt{\pi}}{2} = 2\sqrt{\pi}$ против $H_1 : \theta \frac{\sqrt{\pi}}{2} > 2\sqrt{\pi}$, што је еквивалентно са $H_0 : \theta = 4$ против $H_1 : \theta > 4$. Критичну област ћемо тражити помоћу Нејман-Пирсонове леме. Да би лема могла да се примени обе хипотезе морају бити просте. Нулта хипотеза

свакако јесте проста, а алтернативну хипотезу напишимо као $H_1 : \theta = \theta_1$, $\theta_1 > 4$, те онда и она јесте проста јер смо параметру доделили тачно једну вредност.

За Нејман-Пирсонову лему је потребно одредити функцију веродостојности која је у овом случају дата са

$$L(x; \theta) = \prod_{i=1}^n \frac{2}{\theta^2} x_i e^{-\left(\frac{x_i}{\theta}\right)^2} = \left(\frac{2}{\theta^2}\right)^n e^{-\frac{1}{\theta^2} \sum_{i=1}^n x_i^2} \prod_{i=1}^n x_i.$$

Користећи лему, најбоља критична област је одређена са

$$\begin{aligned} W &= \left\{ \frac{L(x; \theta_1)}{L(x; 4)} \geq k \right\} = \left\{ \frac{\left(\frac{2}{\theta_1^2}\right)^n e^{-\frac{1}{\theta_1^2} \sum_{i=1}^n x_i^2} \prod_{i=1}^n x_i}{\left(\frac{2}{4^2}\right)^n e^{-\frac{1}{4^2} \sum_{i=1}^n x_i^2} \prod_{i=1}^n x_i} \geq k \right\} \\ &= \left\{ e^{-\left(\frac{1}{\theta_1^2} - \frac{1}{4^2}\right) \sum_{i=1}^n x_i^2} \geq k_1 \right\} = \left\{ -\left(\frac{1}{\theta_1^2} - \frac{1}{4^2}\right) \sum_{i=1}^n x_i^2 \geq k_2 \right\} \\ &= \left\{ \sum_{i=1}^n x_i^2 \geq c \right\}, \end{aligned}$$

јер је $\theta_1 > 4$, па је $-\left(\frac{1}{\theta_1^2} - \frac{1}{4^2}\right) > 0$ и c нека константа у односу на дати узорак, коју треба одредити да бисмо добили тачан израз за критичну област. На основу датог прага значајности, можемо одредити c тако да важи

$$P_{H_0} \left\{ \sum_{i=1}^n x_i^2 \geq c \right\} = 0.1.$$

Приметимо да је

$$F_X(x) = \frac{2}{\theta^2} \int_0^x t e^{-\frac{t^2}{\theta^2}} dt = \int_0^{\frac{x^2}{\theta^2}} e^{-u} du = 1 - e^{-\left(\frac{x}{\theta}\right)^2}, \quad x \geq 0,$$

па је

$$F_{X^2}(x) = P\{X^2 \leq x\} = P\{X \leq \sqrt{x}\} = F_X(\sqrt{x}) = 1 - e^{-\frac{x}{\theta^2}}, \quad x \geq 0,$$

односно случајна величина X^2 има експоненцијалну $\mathcal{E}\left(\frac{1}{\theta^2}\right)$ расподелу.

Како при H_0 , X^2 има експоненцијалну $\mathcal{E}\left(\frac{1}{16}\right)$ расподелу, одакле имамо

$$\sum_{i=1}^n X_i^2 \in \gamma\left(n, \frac{1}{16}\right),$$

односно

$$\frac{1}{8} \sum_{i=1}^n X_i^2 \in \gamma\left(n, \frac{1}{2}\right) = \chi_{2n}^2.$$

Сада за конкретан узорак обима $n = 7$, вредност c можемо добити из једначине

$$0.1 = P_{H_0} \left\{ \frac{1}{8} \sum_{i=1}^7 X_i^2 \geq \frac{c}{8} \right\}$$

$$c = 8\chi_{14;0.1}^2 = 8 \cdot 21.064 = 168.512.$$

Дакле, најбоља критична област је

$$W = \left\{ \sum_{i=1}^n x_i^2 \geq 168.512 \right\}.$$

Коначно, на основу датог узорака добијамо да је

$$\sum_{i=1}^7 x_i^2 = 90.8396,$$

што не припада критичној области, па не одбацујемо нулту хипотезу.