



Универзитет у Београду
Математички факултет

ВЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ)

ЈУН 1 - 19. септембар 2025. године

1. Општи чланови X_n и Y_n независних низова независних случајних величина имају униформну $\mathcal{U}(0, 1)$ расподелу. Ако је

$$Z_n = \ln \left(\prod_{k=1}^n \frac{e^{-(p+q)}}{X_k^p (1 - Y_k)^q} \right)^{\frac{1}{\sqrt{n}}},$$

где су p и q позитивни реални бројеви, испитати конвергенцију у расподели низа случајних величина (Z_n) .

2. Нека обележје X има закон расподеле

$$\begin{pmatrix} 0 & 1 & 2 \\ 1 - \pi & \pi(1 - p) & \pi p \end{pmatrix},$$

где су p и π непознати параметри, такви да $p \in (0, 1)$ и $\pi \in (0, 1)$. На основу узорка обима n методом максималне веродостојности одредити оцену за вероватноћу $P\{X = 2\}$ и испитати постојаност те оцене.

3. За густину расподеле обележја X важи да је

$$f(x; \theta) = \frac{\theta}{(1+x)^{\theta+1}}, \quad x > 0, \theta > 0.$$

Изабран је узорак обима 50 и добијено је да је $\sum_{k=1}^{50} \ln(1+x_k) = 27.475$. На основу тог узорка добијени 93.2% интервал поверења за непознати параметар θ је $(1.378, c)$. Израчунати реалан број c .

Решења задатака

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

$$Z_n = \ln \left(\prod_{k=1}^n \frac{e^{-(p+q)}}{X_k^p (1-Y_k)^q} \right)^{\frac{1}{\sqrt{n}}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \left(\ln e^{-(p+q)} - p \ln X_k - q \ln(1-Y_k) \right) = \frac{1}{\sqrt{n}} \sum_{k=1}^n W_k$$

где је $W_k = -(p+q) - p \ln X_k - q \ln(1-Y_k)$. За општи члан X_n низа случајних величина важи да је

$$E(\ln X_k) = \int_0^1 \ln x dx = (x \ln x) \Big|_{x=0}^1 - \int_0^1 x \frac{1}{x} dx = - \int_0^1 dx = -1,$$

$$E(\ln X_k)^2 = \int_0^1 (\ln x)^2 dx = (x(\ln x)^2) \Big|_{x=0}^1 - \int_0^1 x \frac{2 \ln x}{x} dx = -2 \int_0^1 \ln x dx = 2,$$

па је $D(\ln X_k) = E(\ln X_k)^2 - (E(\ln X_k))^2 = 1$. Симетријом за $\ln(1-Y_k)$ добијамо да је $E(\ln(1-Y_k)) = -1$ и $D(\ln(1-Y_k)) = 1$. Сада је, због независности низова случаних величина X_k и Y_k ,

$$EW_k = -(p+q) - pE(\ln X_k) - qE(\ln(1-Y_k)) = -(p+q) - p(-1) - q(-1) = 0,$$

$$DW_k = p^2 D(\ln X_k) + q^2 D(\ln(1-Y_k)) = p^2 + q^2 < +\infty.$$

Како је W_n низ независних случајних величина са истом расподелом и коначним математичким очекивањем и дисперзијом, то на основу централне граничне теореме следи да је

$$\frac{\sum_{k=1}^n W_k - E\left(\sum_{k=1}^n W_k\right)}{\sqrt{D\left(\sum_{k=1}^n W_k\right)}} = \frac{\sum_{k=1}^n W_k}{\sqrt{n}\sqrt{p^2+q^2}} \xrightarrow{D} W^*, \quad W^* \in \mathcal{N}(0,1).$$

Одавде следи да низ случајних величина (Z_n) у расподели конвергира ка случајној величини Z^* , где је $Z^* \in \mathcal{N}(0, p^2 + q^2)$.

2. Закон расподеле се може записати као

$$p(x; \pi, p) = (1-\pi)^{I\{x=0\}} (\pi(1-p))^{I\{x=1\}} (\pi p)^{I\{x=2\}}, \quad x \in \{0, 1, 2\}.$$

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

$$\begin{aligned} L(\pi, p) &= \prod_{k=1}^n p(x_k; \pi p) = \prod_{k=1}^n (1-\pi)^{I\{x_k=0\}} (\pi(1-p))^{I\{x_k=1\}} (\pi p)^{I\{x_k=2\}} \\ &= (1-\pi)^{\sum_{k=1}^n I\{x_k=0\}} (\pi(1-p))^{\sum_{k=1}^n I\{x_k=1\}} (\pi p)^{\sum_{k=1}^n I\{x_k=2\}} \\ &= (1-\pi)^{n_0} (\pi - \pi p)^{n_1} (\pi p)^{n_2}, \end{aligned}$$

где је $n_j = \sum_{k=1}^n I\{x_k = j\}$, за $j \in \{0, 1, 2\}$, па је логаритам веродостојности

$$\ln L(\pi, p) = n_0 \ln(1-\pi) + n_1 \ln(\pi - \pi p) + n_2 \ln(\pi p).$$

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

$$\frac{\partial}{\partial \pi} \ln L(\pi, p) = -\frac{n_0}{1-\pi} + \frac{n_1 + n_2}{\pi} = 0,$$

$$\frac{\partial}{\partial p} \ln L(\pi, p) = -\frac{n_1}{1-p} + \frac{n_2}{p} = 0,$$

одакле је

$$\pi = \frac{n_1 + n_2}{n}, \quad p = \frac{n_2}{n_1 + n_2}.$$

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

$$\frac{\partial^2}{\partial \pi^2} \ln L(\pi, p) = -\frac{n_1 + n_2}{\pi^2} - \frac{n_0}{(1-\pi)^2} < 0,$$

$$\frac{\partial^2}{\partial p^2} \ln L(\pi, p) = -\frac{n_2}{p^2} - \frac{n_1}{(1-p)^2} < 0,$$

$$\frac{\partial^2}{\partial \pi \partial p} \ln L(\pi, p) = 0,$$

ово заиста јесте тачка у којој функција $\ln L(\pi, p)$ достиже максимум. Према томе, на основу овог узорка, оцена методом максималне веродостојности за вероватноћу $P\{X = 2\}$ је

$$T = \hat{\pi} \cdot \hat{p} = \frac{1}{n} \sum_{k=1}^n I\{X_k = 2\}.$$

Проверимо прво да ли је оцена непристрасна. Очекивање оцене је

$$ET = E\left(\frac{1}{n} \sum_{k=1}^n I\{X_k = 2\}\right) = \frac{1}{n} \sum_{k=1}^n P\{X_k = 2\} = P\{X_k = 2\} = \pi p,$$

па је оцена T непристрасна оцена параметра πp . Сада је дисперзија оцене једнака

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

Постојаност подразумева да оцена конвергира у вероватноћи ка параметру πp . Нека је $\varepsilon > 0$ произвољно:

$$P\{|T - \pi p| > \varepsilon\} \leq \frac{E(T - \pi p)^2}{\varepsilon^2} = \frac{D(T)}{\varepsilon^2} = \frac{\pi p(1 - \pi p)}{n\varepsilon^2},$$

при чему је коришћена Чебишовљева неједнакост. Дакле,

$$0 \leq P\{|T - \pi p| > \varepsilon\} \leq \frac{\pi p(1 - \pi p)}{n\varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty,$$

па следи да $P\{|T - \pi p| > \varepsilon\} \rightarrow 0$, када $n \rightarrow \infty$ и то важи за свако $\varepsilon > 0$. Закључујемо да је T постојана оцена за параметар πp .

3. Нека је $Y_k = \ln(1 + X_k)$, $k = 1, 2, \dots, n$. За почетак, одређујемо расподелу случајних величина $Y_k = \ln(1 + X_k)$. Приметимо да је

$$\begin{aligned} F_{Y_k}(y) &= P\{Y_k \leq y\} = P\{\ln(1 + X_k) \leq y\} = P\{X_k \leq e^y - 1\} \\ &= \int_0^{e^y - 1} \frac{\theta}{(1+x)^{\theta+1}} dx = \left[\frac{1+x}{\theta} \right]_{dx=dt}^t = \int_1^{e^y} \frac{\theta}{t^{\theta+1}} dt \\ &= 1 - e^{-\theta y}, \quad y \geq 0. \end{aligned}$$

Дакле, случајне величине Y_k , $k = 1, 2, \dots, n$ имају $\mathcal{E}(\theta)$ расподелу и независне су јер су случајне величине $X_1, \dots, X_n$ независне. Закључујемо да

$$\sum_{n=1}^{50} \ln(1 + X_n) \in \gamma(50, \theta) \Rightarrow T_{50} = 2\theta \sum_{n=1}^{50} \ln(1 + X_n) \in \chi_{100}^2.$$

Потребно је да одредимо статистике U_{50} и V_{50} дефинисане на основу извученог узорка обима 50 такве да је $P\{U_{50} \leq V_{50}\} = 1$ за које важи

$$P\{U_{50} \leq \theta \leq V_{50}\} = 0.932,$$

односно

$$P\left\{2U_{50} \sum_{n=1}^{50} \ln(1 + X_n) \leq T_{50} \leq 2V_{50} \sum_{n=1}^{50} \ln(1 + X_n)\right\} = 0.932.$$

Користећи да T_n има χ_{100}^2 расподелу, добијамо да је

$$F_{\chi_{100}^2}\left(2V_{50} \sum_{n=1}^{50} \ln(1 + X_n)\right) - F_{\chi_{100}^2}\left(2U_{50} \sum_{n=1}^{50} \ln(1 + X_n)\right) = 0.932.$$

Како је n довољно велико ($n > 30$) одговарајуће вредности се не могу прочитати из таблица χ^2 расподеле. У том случају, χ_{100}^2 расподелу можемо апроксимирати нормалном $\mathcal{N}(100, 200)$ расподелом. Односно горњи израз је еквивалентан са

$$\Phi\left(\frac{2V_{50} \sum_{n=1}^{50} \ln(1 + X_n) - 100}{\sqrt{200}}\right) - \Phi\left(\frac{2U_{50} \sum_{n=1}^{50} \ln(1 + X_n) - 100}{\sqrt{200}}\right) = 0.932.$$

Из таблица за нормалну расподелу и за реализовану вредност узорка, имамо да је

$$\Phi\left(\frac{2 \cdot 1.378 \cdot 27.475 - 100}{\sqrt{200}}\right) = \Phi(-1.717) = 0.043.$$

Према томе,

$$\Phi\left(\frac{2c \sum_{n=1}^{50} \ln(1 + x_n) - 100}{\sqrt{200}}\right) \approx 0.932 + 0.043 = 0.975.$$

Даље, из таблица за нормалну расподелу добијамо да је

$$\frac{2c \sum_{n=1}^{50} \ln(1 + x_n) - 100}{\sqrt{200}} = \Phi^{-1}(0.975) = 1.96,$$

па следи да је $c = 2.324$.

Напомена: Тражени интервал поверења можемо одредити директно применом централне граничне теореме на низ $Y_k = \ln(1 + X_k)$. Наиме, имамо да је

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

односно, с обзиром да је $EY_k = \frac{1}{\theta}$ и $DY_k = \frac{1}{\theta^2}$ важи

$$\frac{\theta \sum_{k=1}^n Y_k - 50}{\sqrt{50}} \xrightarrow{D} Y^*, \quad Y^* \in \mathcal{N}(0, 1).$$

Потребно је да одредимо статистике U_{50} и V_{50} дефинисане на основу извученог узорка обима 50 такве да је $P\{U_{50} \leq V_{50}\} = 1$ за које важи

$$P\{U_{50} \leq \theta \leq V_{50}\} = 0.932,$$

односно

$$P\left\{\frac{U_{50} \sum_{k=1}^n Y_k - 50}{\sqrt{50}} \leq \frac{\theta \sum_{k=1}^n Y_k - 50}{\sqrt{50}} \leq \frac{V_{50} \sum_{k=1}^n Y_k - 50}{\sqrt{50}}\right\} = 0.932.$$

Како је $n > 30$, применом централне граничне теореме добијамо да је

$$\Phi\left(\frac{V_{50} \sum_{k=1}^n Y_k - 50}{\sqrt{50}}\right) - \Phi\left(\frac{U_{50} \sum_{k=1}^n Y_k - 50}{\sqrt{50}}\right) = 0.932.$$

Из таблица за нормалну расподелу и за реализовану вредност узорка, имамо да је

$$\Phi\left(\frac{1.378 \cdot 27.475 - 50}{\sqrt{50}}\right) = \Phi(-1.717) = 0.043.$$

Према томе,

$$\Phi\left(\frac{c \sum_{n=1}^{50} y_k - 50}{\sqrt{50}}\right) \approx 0.932 + 0.043 = 0.975.$$

Даље, из таблица за нормалну расподелу добијамо да је

$$\frac{c \sum_{n=1}^{50} y_k - 50}{\sqrt{50}} = \Phi^{-1}(0.975) = 1.96,$$

па следи да је $c = 2.324$.