



Универзитет у Београду
Математички факултет

БЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ)

3. септембар 2024. године

1. Општи чланови X_n и Y_n независних низова независних случајних величина имају карактеристичне функције, редом, дате са

$$\varphi_{X_n}(t) = \frac{1}{5}(3 + \cos t + \cos 2t),$$

$$\varphi_{Y_n}(t) = (2 - e^{it})^{-1}.$$

Ако је

$$Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k Y_k,$$

испитати конвергенцију у расподели низа случајних величина (Z_n) .

2. За густину расподеле обележја X важи да је

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

За оцену непознатог параметра θ на основу узорка обима n предложена је оцена T , где је $T = k\hat{\theta}$ и $\hat{\theta}$ оцена за θ добијена методом максималне веродостојности. Одредити k тако да T буде непристрасна оцена, а затим испитати постојаност предложене оцене.

3. Претпоставља се да број дана до првог следећег дана са растом тржишне вредности акција на берзи $S\&P500$ има геометријску расподелу. Расподела времена између два таква дана дата је у следећој табели

број дана	1	2 и 3	4	5 и 6	8 и више
учесталост	1532	1098	194	107	17

Са прагом значајности 0.05 испитати да ли је та претпоставка тачна.

Решења задатака

1. Приметимо да је

$$\varphi_{X_n}(t) = \frac{3}{5} + \frac{e^{it} + e^{-it} + e^{2it} + e^{-2it}}{10},$$

одакле добијамо да општи члан X_n низа случајних величина има закон расподеле

$$X_n : \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ \frac{1}{10} & \frac{1}{10} & \frac{3}{5} & \frac{1}{10} & \frac{1}{10} \end{pmatrix},$$

па је $EX_n = 0$, $EX_n^2 = 1$ и $DX_n = 1$. Са друге стране, имамо да је

$$\varphi_{Y_n}(t) = \frac{1}{2 - e^{it}} = \frac{1}{2} \frac{1}{1 - \frac{1}{2}e^{it}} = \frac{1}{2} \sum_{k=0}^{\infty} \left(\frac{e^{it}}{2}\right)^k = \sum_{k=0}^{\infty} \frac{1}{2^{k+1}} e^{itk}.$$

Одавде закључујемо да општи члан Y_n има закон расподеле $P\{Y_n = k\} = \frac{1}{2^{k+1}}$, за $k \in \mathbb{N}_0$.
Одредимо сада EY_n и EY_n^2 . Важи

$$EY_n = \sum_{k=0}^{\infty} \frac{k}{2^{k+1}} = 1, \quad EY_n^2 = \sum_{k=0}^{\infty} \frac{k^2}{2^{k+1}} = 3.$$

Претходно добијамо јер из

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n, \quad |z| < 1,$$

диференцирањем добијамо редом

$$\frac{z}{(1-z)^2} = z \sum_{n=1}^{\infty} n z^{n-1} = \sum_{n=1}^{\infty} n z^n, \quad \sum_{n=1}^{\infty} n^2 z^n = \frac{z(1+z)}{(1-z)^3},$$

одакле за $z = \frac{1}{2}$ следи тражено.

Даље, због независности случајних величина X_n и Y_n имамо да је

$$E(X_n Y_n) = EX_n \cdot EY_n = 0,$$

$$D(X_n Y_n) = E(X_n^2 Y_n^2) - (E(X_n Y_n))^2 = EX_n^2 \cdot EY_n^2 = 3.$$

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

$$\frac{\sum_{k=1}^n X_k Y_k - E\left(\sum_{k=1}^n X_k Y_k\right)}{\sqrt{D\left(\sum_{k=1}^n X_k Y_k\right)}} = \frac{\sum_{k=1}^n X_k Y_k}{\sqrt{3n}} = \frac{\sum_{k=1}^n X_k Y_k}{\sqrt{3n}} \xrightarrow{D} Z^*, \quad Z^* \in \mathcal{N}(0, 1).$$

Одавде добијамо граничну расподелу низа случајних величина (Z_n) :

$$Z_n = \frac{\sum_{k=1}^n X_k Y_k}{\sqrt{n}} = \sqrt{3} \cdot \frac{\sum_{k=1}^n X_k Y_k}{\sqrt{3n}} \xrightarrow{D} \tilde{Z}, \quad \tilde{Z} \in \mathcal{N}(0, 3).$$

Дакле, низ случајних величина (Z_n) у расподели конвергира ка случајној величини $\tilde{Z}$, где је $\tilde{Z} \in \mathcal{N}(0, 3)$.

Напомена: У претходном задатку нисмо морали одредити закон расподеле за низ величина (Y_n) . Наиме, директно методом карактеристичних функција налазимо први и други моменат. Заиста, имамо да је

$$\varphi'_{Y_n}(t) = \frac{ie^{it}}{(2 - e^{it})^2}, \quad \varphi''_{Y_n}(t) = \frac{e^{it}(2 + e^{it})}{(e^{it} - 2)^3}.$$

Сада је

$$EY_n = \frac{\varphi'_{Y_n}(0)}{i} = 1, \quad EY_n^2 = \frac{\varphi''_{Y_n}(0)}{i^2} = 3.$$

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

$$\begin{aligned} L(\theta) &= \prod_{k=1}^n f(x_k; \theta) = \prod_{k=1}^n \frac{\theta}{\sqrt{x_k}} e^{-2\theta\sqrt{x_k}} \\ &= \theta^n \exp \left\{ -2\theta \sum_{k=1}^n \sqrt{x_k} \right\} \left(\prod_{k=1}^n \sqrt{x_k} \right)^{-1}. \end{aligned}$$

па је логаритам веродостојности

$$\ln L(\theta) = n \ln \theta - 2\theta \sum_{k=1}^n \sqrt{x_k} - \ln \left(\prod_{k=1}^n \sqrt{x_k} \right).$$

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

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

одакле је

$$\theta = \frac{n}{2 \sum_{k=1}^n \sqrt{x_k}}.$$

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

$$\frac{\partial^2}{\partial \theta^2} \ln L(\theta) = -\frac{n}{\theta^2} < 0,$$

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

$$\hat{\theta} = \frac{n}{2 \sum_{k=1}^n \sqrt{X_k}}.$$

Да бисмо одредили k , тако да T_1 буде непристрасна оцена, треба да важи

$$E(T_1) = E(k\hat{\theta}) = \theta.$$

Нека је $U = 2 \sum_{k=1}^n \sqrt{X_k}$, а $V_k = 2\sqrt{X_k}$. За почетак, одредимо расподелу случајних величина V_k , $k = 1, 2, \dots, n$. Приметимо да

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

па је

$$F_{V_k}(v) = P\{V_k \leq v\} = P\left\{2\sqrt{X_k} \leq v\right\} = P\left\{X_k \leq \frac{v^2}{4}\right\} = 1 - e^{-2\theta\sqrt{\frac{v^2}{4}}} = 1 - e^{-\theta v}, \quad v \geq 0.$$

Дакле, случајне величине V_k имају експоненцијалну $\mathcal{E}(\theta)$ расподелу и независне су јер су случајне величине X_k независне. Закључујемо да

$$U = 2 \sum_{k=1}^n \sqrt{X_k} = \sum_{k=1}^n V_k \in \gamma(n, \theta).$$

Сада је

$$E(\hat{\theta}) = E\left(\frac{n}{U}\right) = nE\left(\frac{1}{U}\right) = n \int_0^n \frac{1}{u} \frac{u^{n-1} \theta^n e^{-\theta u}}{\Gamma(n)} du = n \frac{\Gamma(n-1)}{\Gamma(n)} \theta \int_0^n \frac{u^{n-2} \theta^{n-1} e^{-\theta u}}{\Gamma(n-1)} du = \frac{n}{n-1} \theta.$$

Дакле, за $k = \frac{n-1}{n}$, оцена T_1 ће бити непристрасна оцена параметра θ . Одредимо још дисперзију оцено T . Важи да је

$$\begin{aligned} D(T) &= E(T^2) - (ET)^2 = E\left(\frac{(n-1)^2}{U^2}\right) - \theta^2 = (n-1)^2 E\left(\frac{1}{U^2}\right) - \theta^2 \\ &= (n-1)^2 \int_0^\infty \frac{1}{u^2} \frac{u^{n-1} \theta^n e^{-\theta u}}{\Gamma(n)} du - \theta^2 = (n-1)^2 \theta^2 \frac{\Gamma(n-2)}{\Gamma(n)} \int_0^\infty \frac{u^{n-3} \theta^{n-2} e^{-\theta u}}{\Gamma(n-2)} du - \theta^2 \\ &= \frac{n-1}{n-2} \theta^2 - \theta^2 = \frac{\theta^2}{n-2} \end{aligned}$$

Остаје још да испитамо постојаност ове оцено. Постојаност подразумева да оцена конвергира у вероватноћи ка параметру θ . Нека је $\varepsilon > 0$ произвољно:

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

при чему је коришћена Чебишовљева неједнакост. Дакле,

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

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

3. Наведене класе не покривају читав узорачки простор, па морамо да направимо нову табелу у којој ћемо да покријемо све:

број дана	1	2 и 3	4	5 и 6	7	8 и више
учесталост	1532	1098	194	107	0	17

Обим узорка је $n = 1532 + 1098 + 194 + 107 + 0 + 17 = 2948$.

За почетак треба да израчунамо np_k , $k = 1, 2, \dots, 6$. Вероватноће p_k рачунамо по формули

$$p_k = P_{H_0}\{X \in S_k\} = (1-p)^{k-1}p, \quad k \in \mathbb{N}.$$

Параметар p је непознат, па не можемо израчунати тачне вероватноће, али можемо оценити p методом максималне веродостојности па добити процене вероватноћа. Оцењујемо p методом максималне веродостојности:

$$\begin{aligned} L(x; p) &= p^n (1-p)^{\sum_{i=1}^n x_i - n}, \\ \ln L(x; p) &= n \ln p + \left(\sum_{i=1}^n x_i - n\right) \ln(1-p), \\ \frac{\partial}{\partial p} \ln L(x; p) &= \frac{n}{p} - \frac{\sum_{i=1}^n x_i - n}{1-p} = 0, \\ p &= \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}_n}. \end{aligned}$$

Још треба да проверимо да ли је максимум. Пошто је $\frac{\partial^2}{\partial p^2} \ln L(x; p) < 0$, закључујемо да је оцена максималне веродостојности:

$$\hat{p} = \frac{1}{\bar{X}_n}.$$

Желимо да добијемо оцену p на реализованом узорку. Узорачку средину можемо оценити са:

$$\bar{x}_n = \frac{1 \cdot 1532 + 2.5 \cdot 1098 + 4 \cdot 194 + 5.5 \cdot 107 + 7 \cdot 0 + 8 \cdot 17}{2948} = 1.9598,$$

па на основу овог узорка добијамо да је

$$\hat{p} = \frac{1}{1.9598} = 0.51.$$

Сада можемо да израчунамо вероватноће p_k :

$$\begin{aligned} p_1 &= P\{X = 1\} = 0.51 \cdot (1 - 0.51)^0 = 0.51, \\ p_2 &= P\{X = 2\} + P\{X = 3\} = 0.51 \cdot (1 - 0.51)^1 + 0.51 \cdot (1 - 0.51)^2 = 0.3724, \\ p_3 &= P\{X = 4\} = 0.51 \cdot (1 - 0.51)^3 = 0.06, \\ p_4 &= P\{X = 5\} + P\{X = 6\} = 0.51 \cdot (1 - 0.51)^4 + 0.51 \cdot (1 - 0.51)^5 = 0.0438, \\ p_5 &= P\{X = 7\} = 0.51 \cdot (1 - 0.51)^6 = 0.0071, \\ p_6 &= P\{X \geq 8\} = 1 - P\{X \leq 7\} = 1 - (p_1 + p_2 + p_3 + p_4 + p_5) = 0.0067. \end{aligned}$$

Дакле, имамо табелу

број дана (S_k)	1	2 и 3	4	5 и 6	7	8 и више
учесталост (M_k)	1532	1098	194	107	0	17
p_k	0.51	0.3724	0.06	0.0438	0.0071	0.0067
np_k	1503.48	1097.8352	176.88	129.1224	20.9308	19.7516

Видимо да за све категорије важи $np_k > 5$, те нећемо имати спајање. Критична област је облика

$$W = \{\chi_0^2 \geq c\},$$

где c рачунамо из једнакости

$$P_{H_0}\{\chi_0^2 \geq c\} = 0.05.$$

При H_0 , пошто имамо 6 класа и оценили смо један параметар методом максималне веродостојности, $\chi_0^2 \in \chi_{6-1-1}^2$. Према томе,

$$c = \chi_{4;0.05}^2 = 9.488,$$

па је критична област

$$W = \{\chi_0^2 \geq 9.488\}.$$

Рачунамо вредност тест статистике на добијеном узорку:

$$\begin{aligned} \chi_0^2 &= \sum_{i=1}^6 \frac{(M_k - np_k)^2}{np_k} \\ &= \frac{(1532 - 1503.48)^2}{1503.48} + \frac{(1098 - 1097.8352)^2}{1097.8352} + \frac{(194 - 176.88)^2}{176.88} \\ &\quad + \frac{(107 - 129.1224)^2}{129.1224} + \frac{(0 - 20.9308)^2}{20.9308} + \frac{(17 - 19.7516)^2}{19.7516} \\ &= 27.3024. \end{aligned}$$

Пошто је

$$27.3024 > 9.488,$$

узорак упада у критичну област, па одбацујемо нулту хипотезу.