Универзитет у Београду
Математички факултет

ВЕРОВАТНОЋА И СТАТИСТИКА Б (3Р, 4МНЛ)

29. јануар 2024. године

---

1. Општи члан X_n низа независних случајних величина има густину расподеле

$$f(x;m) = \frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{x}{2}}, \quad x \geq 0,\, m \in \mathbb{N},$$

док општи члан Y_n низа независних случајних величина има закон расподеле $P\{Y_n = -2024\} = P\{Y_n = 2024\} = \frac{1}{2}$. Ако су за сваки природан број n случајне величине X_n и Y_n независне, испитати конвергенцију у расподели низа случајних величина (Z_n), где је

$$Z_n = (nm(m+2))^{-\frac{1}{2}} \sum_{k=1}^{n} X_k Y_k.$$

2. За густину расподеле обележја X важи да је $f(x) = x\theta^{-\frac{x^2}{2}} \ln\theta$, $x > 0$, $\theta > 1$. На основу узорка обима n, методом максималне веродостојности одредити оцену непознатог параметра $\frac{1}{\ln\theta}$, а затим испитати ефикасност добијене оцене.

3. Претпоставља се да снага Intel процесора има нормалну $\mathcal{N}(m_1, 16)$ расподелу, док снага AMD процесора има нормалну $\mathcal{N}(m_2, 16)$ расподелу. Тим истраживача жели да испита колико се разликују просечне снаге процесора ова два произвођача. Истраживање је рађено на истом броју Intel и AMD процесора и констатовано је да би, под претпоставком да су просечне снаге процесора једнаке, апсолутна разлика узорачких просечних снага процесора била мања од 0.8 са вероватноћом 0.99. Одредити дужину 97% интервала поверења за разлику просечних снага процесора та два произвођача.

ВЕРОВАТНОЋА И СТАТИСТИКА Б (3Р, 4МНЛ) - Писмени испит 29. јануар 2024.

Решења задатака

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

$$\begin{aligned} EX_n &= \int_0^\infty x \cdot f(x;m)dx = \int_0^\infty x \cdot \frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{x}{2}} dx \\ &= \frac{2^{\frac{m}{2}+1}\Gamma\left(\frac{m}{2}+1\right)}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} \int_0^\infty \frac{1}{2^{\frac{m}{2}+1}\Gamma\left(\frac{m}{2}+1\right)} x^{\frac{m}{2}} e^{-\frac{x}{2}} dx \\ &= 2 \cdot \frac{m}{2} = m, \end{aligned}$$

$$\begin{aligned} EX_n^2 &= \int_0^\infty x^2 \cdot f(x;m)dx = \int_0^\infty x^2 \cdot \frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{x}{2}} dx \\ &= \frac{2^{\frac{m}{2}+2}\Gamma\left(\frac{m}{2}+2\right)}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} \int_0^\infty \frac{1}{2^{\frac{m}{2}+2}\Gamma\left(\frac{m}{2}+2\right)} x^{\frac{m}{2}+1} e^{-\frac{x}{2}} dx \\ &= 2^2 \cdot \left(\frac{m}{2}+1\right) \cdot \frac{m}{2} = m(m+2), \end{aligned}$$

па је $DX_n = EX_n^2 - (EX_n)^2 = m^2 + 2m - m^2 = 2m$, док за низ (Y_n) важи да је $EY_n = -2024 \cdot \frac{1}{2} + 2024 \cdot \frac{1}{2} = 0$ и $DY_n = EY_n^2 = 2024^2$. Даље, због независности случајних величина X_n и Y_n имамо да је

$$E(X_nY_n) = EX_n \cdot EY_n = 0,$$
$$D(X_nY_n) = E(X_n^2Y_n^2) - (E(X_nY_n))^2 = EX_n^2 \cdot EY_n^2 = 2024^2 m(m+2).$$

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

$$\frac{\sum\limits_{k=1}^n X_kY_k - E\left(\sum\limits_{k=1}^n X_kY_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^n X_kY_k\right)}} = \frac{\sum\limits_{k=1}^n X_kY_k}{\sqrt{n \cdot 2024^2 m(m+2)}} = \frac{\sum\limits_{k=1}^n X_kY_k}{2024\sqrt{nm(m+2)}} \xrightarrow{D} Z^*, \quad Z^* \in \mathcal{N}(0,1).$$

Одавде добијамо граничну расподелу низа случајних величина (Z_n):

$$Z_n = (nm(m+2))^{-\frac{1}{2}} \sum_{k=1}^n X_kY_k = 2024 \cdot \frac{\sum\limits_{k=1}^n X_kY_k}{2024\sqrt{nm(m+2)}} \xrightarrow{D} \tilde{Z}, \quad \tilde{Z} \in \mathcal{N}(0, 2024^2).$$

Дакле, низ случајних величина (Z_n) у расподели конвергира ка случајној величини $\tilde{Z}$, где је $\tilde{Z} \in \mathcal{N}(0, 2024^2)$.

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

$$L(\theta) = L(\theta | x_1, x_2, \ldots, x_n) = \prod_{i=1}^n x_i \theta^{-\frac{x_i^2}{2}} \ln\theta, \quad \theta > 1, \, , x_1, \ldots, x_n > 0,$$

тј.

$$L(\theta) = \left(\prod_{i=1}^n x_i\right) \theta^{-\sum_{i=1}^n \frac{x_i^2}{2}} (\ln\theta)^n, \quad \theta > 1, \, , x_1, \ldots, x_n > 0,$$

па је

$$\ln L(\theta) = n \ln(\ln\theta) - \left(\frac{1}{2}\sum_{i=1}^n x_i^2\right) \ln\theta + \sum_{i=1}^n \ln x_i.$$

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

$$\frac{\partial}{\partial\theta}\ln L(\theta) = -\frac{n}{\theta}\left[\frac{1}{2n}\sum_{i=1}^{n}x_i^2 - \frac{1}{\ln\theta}\right] = \frac{n}{\theta\ln\theta} - \frac{1}{2\theta}\sum_{i=1}^{n}x_i^2 = 0 \quad\Longrightarrow\quad \theta = \exp\frac{2n}{\sum_{i=1}^{n}x_i^2}.$$

Још треба проверити да ли се у добијеној вредности постиже максимум. Како је $\frac{\partial}{\partial\theta}\ln L(\theta)$ позитивна за $\theta < \exp\frac{2n}{\sum_{i=1}^{n}x_i^2}$, а негативна за $\theta > \exp\frac{2n}{\sum_{i=1}^{n}x_i^2}$, па је функција $\frac{\partial}{\partial\theta}\ln L(\theta)$ растућа за $\theta < \exp\frac{2n}{\sum_{i=1}^{n}x_i^2}$, а опадајућа за $\theta > \exp\frac{2n}{\sum_{i=1}^{n}x_i^2}$. Према томе, $\exp\frac{2n}{\sum_{i=1}^{n}x_i^2}$ је тачка у којој $\ln L(\theta)$ достиже максимум. Према томе, на основу овог узорка, оцена методом максималне веродостојности је

$$\hat{\theta} = \exp\frac{2n}{\sum_{i=1}^{n}X_i^2}.$$

Како је оцена методом максималне веродостојности инваријантна у односу на трансформацију Борел-мерљивом функцијом, одавде закључујемо да је оцена непознатог параметра $\frac{1}{\ln\theta}$ дата са

$$T_\theta = \frac{\sum_{i=1}^{n}X_i^2}{2n}$$

Да би имало смисла да испитујемо ефикасност оцене, морамо прво да проверимо да ли је она непристрасна. За почетак, одредимо расподелу случајне величине X_1^2. Имамо да је

$$\begin{aligned}
P\{X_1^2 \le x\} &= P\{X_1 \le \sqrt{x}\} = \int_0^{\sqrt{x}} t\theta^{-\frac{t^2}{2}}\ln\theta dt \\
&= \left|\begin{matrix}\frac{t^2}{2} = w \\ tdt = dw\end{matrix}\right| = \ln\theta\int_0^{\frac{x}{2}}\theta^{-w}dw \\
&= \ln\theta\left(-\frac{\theta^{-w}}{\ln\theta}\right)\Big|_{w=0}^{\frac{x}{2}} = \ln\theta\left(\frac{1}{\ln\theta} - \frac{\theta^{-\frac{x}{2}}}{\ln\theta}\right) \\
&= 1 - \theta^{-\frac{x}{2}} = 1 - e^{-x\cdot\frac{\ln\theta}{2}}, \quad x \ge 0,
\end{aligned}$$

одакле закључујемо да случајна величина X_1^2 има експоненцијалну $\mathcal{E}\left(\frac{\ln\theta}{2}\right)$ расподелу, па следи да $\sum_{i=1}^{n}X_i^2$ има гама $\gamma\left(n, \frac{\ln\theta}{2}\right)$ расподелу.
Сада имамо да је

$$E(T_\theta) = E\left(\frac{\sum_{i=1}^{n}X_i^2}{2n}\right) = \frac{1}{2n}\cdot\frac{2n}{\ln\theta} = \frac{1}{\ln\theta},$$

тј. оцена $\hat{\theta}$ јесте непристрасна оцена параметра $\frac{1}{\ln\theta}$.
Испитајмо сада регуларност фамилије расподела $f\left(x;\frac{1}{\ln\theta}\right)$ у смислу Рао-Крамера и одређујемо $I\left(\frac{1}{\ln\theta}\right)$. Због једноставности формула у наставку, уведимо смену $a = \frac{1}{\ln\theta}$, $\theta > 1$. Узорачки простор је $\mathbb{R}^+$, па не зависи од непознатог параметра a. Желимо да испитамо да ли је $\int_0^\infty \frac{\partial}{\partial a}f(x;a)dx = 0$. Како је

$$\frac{\partial}{\partial a}f(x;a) = \frac{\partial}{\partial a}\left(x(e^{\frac{1}{a}})^{-\frac{x^2}{2}}\frac{1}{a}\right) = e^{-\frac{x^2}{2a}}\left(\frac{x^3}{2a^3} - \frac{x}{a^2}\right),$$

па је

$$\begin{aligned}
\int_0^\infty \frac{\partial}{\partial a}f(x;a)dx &= \frac{1}{2a^3}\int_0^\infty x^3e^{-\frac{x^2}{2a}}dx - \frac{1}{a^2}\int_0^\infty xe^{-\frac{x^2}{2a}}dx \\
&= \frac{1}{2a^2}\int_0^\infty 2a\cdot ue^{-u}du - \frac{1}{a}\int_0^\infty e^{-u}du \\
&= \frac{1}{a}\Gamma(2) - \frac{1}{a} = 0.
\end{aligned}$$

Фамилија јесте регуларна. Рачунамо $I(a)$ да бисмо добили доњу границу дисперзије свих непристрасних оцена за параметар a.

$$I(a) = E\left(-\frac{\partial^2}{\partial a^2}\ln f(X;a)\right).$$

Дакле, важи

$$\ln f(x;a) = -\ln a + \ln x - \frac{x^2}{2a},$$
$$\frac{\partial}{\partial a}\ln f(x;a) = -\frac{1}{a} + \frac{x^2}{2a^2},$$
$$\frac{\partial^2}{\partial a^2}\ln f(x;a) = \frac{1}{a^2} - \frac{x^2}{a^3},$$

па је

$$I(a) = E\left(\frac{X^2}{a^3} - \frac{1}{a^2}\right) = \frac{EX^2}{a^3} - \frac{1}{a^2} = \frac{2a}{a^3} - \frac{1}{a^2} = \frac{1}{a^2}.$$

Према томе, доња граница дисперзије свих непристрасних оцена једнака је

$$G = \frac{1}{nI(a)} = \frac{a^2}{n} = \frac{1}{n\ln^2\theta}.$$

Како је T_θ непристрасна оцена, сада рачунамо њену дисперзију, да бисмо је упоредили са доњом границом G и видели да ли је ефикасна. Како су случајне величине $X_1, \dots, X_n$ независне, имамо да је

$$D(T_\theta) = D\left(\frac{\sum_{i=1}^n X_i^2}{2n}\right) = \frac{1}{4n^2}\sum_{i=1}^n D(X_i)^2 = \frac{1}{4n^2}\cdot\frac{4n}{\ln^2\theta} = \frac{1}{n\ln^2\theta}.$$

Дакле,

$$ef(\hat{\theta}) = \frac{G}{D(T_\theta)} = \frac{\frac{1}{n\ln^2\theta}}{\frac{1}{n\ln^2\theta}} = 1,$$

па ова оцена јесте ефикасна.

Напомена: Није неопходно одредити расподелу случајне величине X_1^2. Директним рачуном налазимо њено очекивање. Заиста, важи да је

$$\begin{aligned}
E(X_1^2) &= \int_0^\infty x^2\cdot f(x)dx = \int_0^\infty x^3\theta^{-\frac{x^2}{2}}\ln\theta dx = \begin{vmatrix}\frac{x^2}{2} = t\\ xdx = dt\end{vmatrix}\\
&= 2\ln\theta\int_0^\infty t\theta^{-t}dt = \begin{vmatrix}u = t, & dv = \theta^{-t}\\ du = dt, & v = -\frac{1}{\theta^t\ln\theta}\end{vmatrix}\\
&= 2\ln\theta\left[\left(-\frac{t}{\theta^t\ln\theta}\right)\Bigg|_{t=0}^{\infty} + \frac{1}{\ln\theta}\int_0^\infty\theta^{-t}dt\right]\\
&= 2\ln\theta\cdot\frac{1}{\ln^2\theta} = \frac{2}{\ln\theta}
\end{aligned}$$

3. Нека обележје X представља снагу Intel процесора, а обележје Y снагу AMD процесора. Како су случајне величине $\bar{X}_n$ и $\bar{Y}_n$ независне и $\bar{X}_n \in \mathcal{N}(m_1, 16)$, $\bar{Y}_n \in \mathcal{N}(m_2, 16)$, то $\bar{X}_n - \bar{Y}_n \in \mathcal{N}\left(m_1 - m_2, \frac{32}{n}\right)$, где смо са n означили обим сваког узорка.

   Одредимо 97% интервала поверења за разлику средњих вредности снага на стандардан начин. Потребно је да одредимо U_n и V_n такве да важи

$$P\{U_n \le m_1 - m_2 \le V_n\} = 0.97,$$

$$P\{\bar{X}_n - \bar{Y}_n + V_n \le \bar{X}_n - \bar{Y}_n - (m_1 - m_2) \le \bar{X}_n - \bar{Y}_n + U_n\} = 0.97,$$

$$P\left\{\frac{\bar{X}_n - \bar{Y}_n + V_n}{\sqrt{\frac{32}{n}}} \le \frac{\bar{X}_n - \bar{Y}_n - (m_1 - m_2)}{\sqrt{\frac{32}{n}}} \le \frac{\bar{X}_n - \bar{Y}_n + U_n}{\sqrt{\frac{32}{n}}}\right\} = 0.97,$$

при чему формирамо такав интервала поверења да важи

$$P\left\{T_n < \frac{\bar{X}_n - \bar{Y}_n + V_n}{\sqrt{\frac{32}{n}}}\right\} = P\left\{T_n > \frac{\bar{X}_n - \bar{Y}_n + U_n}{\sqrt{\frac{32}{n}}}\right\},$$

где је $T_n = \frac{\bar{X}_n - \bar{Y}_n - (m_1 - m_2)}{\sqrt{\frac{32}{n}}} \in \mathcal{N}(0,1)$ тест статистика. Одавде следи да је

$$U_n = \bar{X}_n - \bar{Y}_n - \Phi^{-1}(0.015)\frac{4\sqrt{2}}{\sqrt{n}}, \quad V_n = \bar{X}_n - \bar{Y}_n + \Phi^{-1}(0.985)\frac{4\sqrt{2}}{\sqrt{n}}.$$

Сада имамо да је дужина интервала

$$V_n - U_n = \frac{4\sqrt{2}}{\sqrt{n}}\left(\Phi^{-1}(0.985) - \Phi^{-1}(0.015)\right).$$

Даље, из текста задатка је дато да под претпоставком да је $m_1 - m_2 = 0$, важи да је

$$P\left\{|\bar{X}_n - \bar{Y}_n| < 0.8\right\} = 0.99,$$

односно

$$P\left\{-0.8 \cdot \frac{\sqrt{n}}{4\sqrt{2}} < \frac{\bar{X}_n - \bar{Y}_n - (m_1 - m_2)}{\frac{4\sqrt{2}}{\sqrt{n}}} < 0.8 \cdot \frac{\sqrt{n}}{4\sqrt{2}}\right\} = 0.99,$$

$$P\left\{-\frac{\sqrt{n}}{5\sqrt{2}} < \underbrace{\frac{\bar{X}_n - \bar{Y}_n - 0}{\frac{4\sqrt{2}}{\sqrt{n}}}}_{\sim\mathcal{N}(0,1)} < \frac{\sqrt{n}}{5\sqrt{2}}\right\} = 0.99.$$

Због симетричности нормалне расподеле,

$$-\frac{\sqrt{n}}{5\sqrt{2}} = \Phi^{-1}(0.005),$$

тј.

$$n = (5\sqrt{2}\Phi^{-1}(0.005))^2 \approx 332.$$

Уврстимо ли у разлику граница интервала, имамо

$$V_n - U_n = \frac{4\sqrt{2}}{\sqrt{332}}\left(\Phi^{-1}(0.985) - \Phi^{-1}(0.015)\right) \approx 1.36.$$