Испит 30.06.2022. МНЛР

Данијел Алексић

Децембар 2022.

Задатак 1. *За густину расподеле обележја X важи да је $f(x;\theta) = \theta e^{1-\theta x}$, $x \geq \frac{1}{\theta}$, $\theta > 0$. На основу узорка обима n, одредити оцену максималне веродостојности параметра $\frac{1}{\theta}$. Испитати непристрасност и постојаност добијене оцене.*

Решење. На основу познате теореме, уколико је $\hat{\theta}$ оцена максималне веродостојности за θ, онда је $1/\hat{\theta}$ ОМВ за $1/\theta$. Стога, наћи ћемо ОМВ за θ, за почетак. Функција веродостојности дата је са

$$\begin{aligned} L(\theta) &= \prod_{k=1}^{n} \theta e^{1-\theta x_k} I\{x_k \geq 1/\theta\} \\ &= \theta^n e^{n-\theta\sum_{k=1}^{n} x_k} \prod_{k=1}^{n} I\{\theta \geq 1/x_k\} \\ &= \theta^n e^{n-\theta\sum_{k=1}^{n} x_k} I\{\theta \geq \max_k (1/x_k)\}. \end{aligned}$$

Кад θ расте, расте и θ^n, али опада $e^{n-\theta\sum_{k=1}^{n} x_k}$. Нама је циљ да максимизујемо, по θ, $L(\theta)$, па ћемо прибећи следећем трику. Нека је $\tilde{\theta}$ оно које максимизује $\theta^n e^{n-\theta\sum_{k=1}^{n} x_k}$. Ако је оно веће од максимума унутар индикатора, оно је оцена. Иначе, максимум је оцена. Отуда, ОМВ за θ је

$$\hat{\theta} = \max\left\{\tilde{\theta}, \max_k (1/X_k)\right\}.$$

Нађимо, $\tilde{\theta}$. Максимизираћемо, уместо $\tilde{L}(\theta) = \theta^n e^{n-\theta\sum_{k=1}^{n} x_k}$, његов логаритам,

$$\ln \tilde{L}(\theta) = n \ln \theta + n - \theta \sum_{k=1}^{n} x_k.$$

Први извод му је

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

што даје

$$\tilde{\theta} = \frac{1}{\overline{X}_n}.$$

Други извод је тривијално негативан. Уочимо да је $\overline{X}_n$, као просечна вредност, сигурно веће од најмање вредности у узорку, то јест важи да је

$$\overline{X}_n \geq \min_k X_k.$$

Како су по услову сви елементи узорка позитивни, узмемо ли реципрочну вредност, добијамо да је

$$\frac{1}{\overline{X}_n} \leq \frac{1}{\min_k X_k} = \max_k \frac{1}{X_k}.$$

Дакле, максимум за ОМВ увек се остварује за $\max_k \frac{1}{X_k}$, па је

$$\hat{\theta} = \max_k \frac{1}{X_k}.$$

Овиме смо нашли оцену максималне веродостојности за параметар θ. ОМВ за $\frac{1}{\theta}$, који ћемо назвати, због лакшег праћења, τ, јесте

$$\hat{\tau} = \frac{1}{\hat{\theta}} = \frac{1}{\max_k \frac{1}{X_k}} = \min_k X_k.$$

Да бисмо му испитали непристрасност, морамо му наћи очекивање, а за то морамо наћи расподелу. Нека је, зато, $x \geq 1/\theta$ (јасно је да је носач за минимум исти као и за сваки од елемената узорка). Имамо да је

$$\begin{aligned}
F_{\hat{\tau}}(x) &= \mathbf{P}\{\hat{\tau} \leq x\} \\
&= \mathbf{P}\{\min_k X_k \leq x\} \\
&= 1 - \mathbf{P}\{\min_k X_k > x\} \\
&= 1 - (\mathbf{P}\{X_1 > x\})^n \\
&= 1 - \left(\int_x^\infty \theta e^{1-\theta t} dt\right)^n \\
&= 1 - e^n \left(\underbrace{\int_x^\infty \theta e^{-\theta t} dt}_{\text{реп екс. рас.}}\right)^n \\
&= 1 - e^n (e^{-\theta x})^n \\
&= 1 - e^{n(1-\theta x)}.
\end{aligned}$$

Дакле, имамо да је

$$F_{\hat{\tau}}(x) = \begin{cases} 0, & x < 1/\theta \\ 1 - e^{n(1-\theta x)}, & x \geq 1/\theta, \end{cases}$$

па је

$$f_{\hat{\tau}}(x) = n\theta e^{n(1-\theta x)},$$

за $x \geq 1/\theta$, а нула иначе. Сада можемо срачунати да је

$$\begin{aligned}
\mathbf{E}(\hat{\tau}) &= \int_{1/\theta}^\infty x f_{\hat{\tau}}(x) dx \\
&= \int_{1/\theta}^\infty x n\theta e^{n(1-\theta x)} dx \\
&= e^n \int_{1/\theta}^\infty x n\theta e^{-n\theta x} dx \\
&= \begin{bmatrix} xn\theta = t & dx = \frac{dt}{n\theta} \\ 1/\theta \mapsto n & \infty \mapsto \infty \end{bmatrix} \\
&= \frac{e^n}{n\theta} \int_n^\infty t e^{-t} dt \\
&= \begin{bmatrix} u = t & dv = e^{-t} dt \\ du = dt & v = -e^{-t} \end{bmatrix} \\
&= \frac{e^n}{n\theta} \left(-te^{-t}\Big|_n^\infty + \int_n^\infty e^{-t} dt\right) \\
&= \frac{e^n}{n\theta} \left(ne^{-n} + e^{-n}\right)
\end{aligned}$$

$$= \frac{n+1}{n}\frac{1}{\theta}.$$

Дакле, оцена није непристрасна, али јесте асимптотски непристрасна. Даље можемо рачунати да је

$$\begin{aligned}
\mathbf{E}(\hat{\tau} - 1/\theta)^2 &= \mathbf{E}(\hat{\tau}^2) - \frac{2}{\theta}\mathbf{E}(\hat{\tau}) + \frac{1}{\theta^2} \\
&= \mathbf{E}(\hat{\tau}^2) - \frac{2}{\theta}\frac{n+1}{n}\frac{1}{\theta} + \frac{1}{\theta^2} \\
&= \mathbf{E}(\hat{\tau}^2) + \frac{1}{\theta^2}\frac{-n-2}{n},
\end{aligned}$$

а затим

$$\begin{aligned}
\mathbf{E}(\hat{\tau^2}) &= \int_{1/\theta}^{\infty} x^2 n\theta e^{n(1-\theta x)} dx \\
&= e^n \int_{1/\theta}^{\infty} x^2 n\theta e^{-n\theta x} dx \\
&= \begin{bmatrix} xn\theta = t & dx = \frac{dt}{n\theta} \\ 1/\theta \mapsto n & \infty \mapsto \infty \end{bmatrix} \\
&= e^n \int_n^{\infty} \frac{t^2}{n^2\theta^2} n\theta e^{-t} \frac{dt}{n\theta} \\
&= \frac{e^n}{n^2\theta^2} \int_n^{\infty} t^2 e^{-t} dt \\
&= \begin{bmatrix} u = t^2 & dv = e^{-t}dt \\ du = 2tdt & v = -e^{-t} \end{bmatrix} \\
&= \frac{e^n}{n^2\theta^2} \left(-t^2 e^{-t}\Big|_n^{\infty} + 2\int_n^{\infty} te^{-t} dt \right) \\
&= \frac{e^n}{n^2\theta^2} \left(n^2 e^{-n} + 2e^{-n}(n+1) \right) \\
&= \frac{1}{\theta^2}\left(1 + \frac{2n+2}{n^2}\right) \\
&= \frac{n^2+2n+2}{\theta^2 n^2}.
\end{aligned}$$

Сада је

$$\begin{aligned}
\mathbf{E}(\hat{\tau} - 1/\theta)^2 &= \mathbf{E}(\hat{\tau}^2) + \frac{1}{\theta^2}\frac{-n-2}{n} \\
&= \frac{n^2+2n+2}{\theta^2 n^2} + \frac{1}{\theta^2}\frac{-n^2-2n}{n^2} \\
&= \frac{1}{\theta^2}\frac{2}{n^2},
\end{aligned}$$

што тежи нули кад $n \to \infty$. Следи да је оцена постојана на основу неједнакости Чебишова:

$$\mathbf{P}\{|\hat{\tau} - 1/\theta| > \varepsilon\} \le \frac{\mathbf{E}(\hat{\tau} - 1/\theta)^2}{\varepsilon^2} \to 0.$$

Задатак 2. *Општи члан X_n низа независних случајних величина има Бернулијеву $Ber(p)$ расподелу. Ако је $Y_k = X_k X_{k+1}$ и $Z_n = \frac{1}{n}\sum_{k=1}^{n} Y_{2k}$, испитати сва четири типа конвергенције низа случајних величина Z_n.*

Решење. У суми у изразу за Z_n ређају се $Y_2, Y_4, Y_6, \ldots$, односно $X_2X_3, X_4X_5, X_6X_7, \ldots$. Штос је у томе што су ипсилони, са кораком два, независни! Свакако, за свако k, X_kX_{k+1} има исту расподелу (производ две независне $Ber(p)$) па су они и једнако расподељени. Није тешко наћи, због независности, да је $\mathbf{E}Y_k = p^2$, што је коначан број. Следи да су испуњени услови јаког закона великих бројева, па $Z_n \stackrel{c.c.}{\to} p^2$, те му конвергира и у вероватноћи и у расподели.

Остаје још испитати средњеквадратну ковергенцију. Тривијално, $\mathbf{E}Z_n = p^2$, па је онда $\mathbf{E}(Z_n - p^2)^2 = \mathbf{D}(Z_n)$, а имамо да је (због независности)

$$\begin{aligned}
\mathbf{D}(Z_n) &= \frac{1}{n^2}\sum_{k=1}^{n}\mathbf{D}(Y_{2k}) \\
&= \frac{1}{n^2}n\mathbf{D}(Y_2) \\
&= \frac{\mathbf{D}(Y_2)}{n} \\
&= \frac{\mathbf{E}(Y_2^2) - (\mathbf{E}Y_2)^2}{n} \\
&= \frac{\mathbf{E}(X_2^2X_3^2) - p^4}{n} \\
&= \frac{\mathbf{E}(X_2^2)\mathbf{E}(X_3^2) - p^4}{n} \\
&= \frac{(\mathbf{E}(X_2^2))^2 - p^4}{n} \\
&= \frac{(\mathbf{D}X_2 + (\mathbf{E}X_2)^2)^2 - p^4}{n} \\
&= \frac{(p(1-p) + p^2)^2 - p^4}{n} \\
&= \frac{p^2(1-p)^2 + 2p^3(1-p) + p^4 - p^4}{n} \\
&= \frac{p^2(1-p)^2 + 2p^3(1-p)}{n},
\end{aligned}$$

што евидентно тежи нули, па Z_n тежи ка p^2 и у средњеквадратном.

Задатак 3. *За густину расподеле обележја X, које представља животни век одређених уређаја подвргнутих тесту, важи да је $f(x;\alpha) = \frac{\alpha}{x^{\alpha+1}}$, $x \geq 1$, $\alpha > 1$. Истраживачи сматрају да је просечан животни век уређаја 6 сати. Извршено је мерење на узорку од 12 уређаја и добијени су подаци* $2.471, 1.167, 4.329, 2.568, 2.400, 3.112, 1.990, 7.846, 3.504, 2.551, 3.034, 2.223$. *На основу добијеног узорка, са прагом значајности 0.1 тестирати униформно најмоћнијим тестом да ли су истраживачи у праву или је просечан животни век уређаја мањи.*

Решење. За почетак испитујемо везу очекиваног животног века $\mathbf{E}(X)$ и параметра α. Имамо да је

$$\begin{aligned}
\mathbf{E}(X) &= \int_1^\infty x\frac{\alpha}{x^{\alpha+1}}dx \\
&= \alpha\int_1^\infty x^{-\alpha}dx \\
&= \alpha\frac{x^{-\alpha+1}}{-\alpha+1}\Bigg|_1^\infty \\
&= \alpha\frac{1}{(1-\alpha)x^{\alpha-1}}\Bigg|_1^\infty \\
&= \frac{\alpha}{\alpha-1}.
\end{aligned}$$

Јасно, треба тестирати $H_0 : \frac{\alpha}{\alpha-1} = 6$ против $H_1 : \frac{\alpha}{\alpha-1} < 6$, што је еквивалентно са тестирањем $H_0 : \alpha = \frac{6}{5}$ против $H_1 : \alpha > \frac{6}{5}$.

Како нам се тражи униформно најмоћнији тест, надамо се да ће лема Нејман-Пирсона да нам помогне. За то, ставимо $\alpha_0 = 6/5$ и $\alpha_1 > 6/5$. Сада критичну област тражимо као

$$\begin{aligned}
\mathcal{W} &= \left\{ \frac{L(\alpha_1)}{L(\alpha_0)} \geq c \right\} \\
&= \left\{ \frac{\prod_{k=1}^n \frac{\alpha_1}{x_k^{\alpha_1+1}}}{\prod_{k=1}^n \frac{\alpha_0}{x_k^{\alpha_0+1}}} \geq c \right\} \\
&= \left\{ \prod_{k=1}^n \frac{\alpha_1}{\alpha_0} x_k^{\alpha_0+1-\alpha_1-1} \geq c \right\} \\
&= \left\{ \frac{\alpha_1^n}{\alpha_0^n} \left(\prod_{k=1}^n x_k \right)^{\alpha_0-\alpha_1} \geq c \right\} \\
&= \left\{ \left(\prod_{k=1}^n x_k \right)^{\alpha_0-\alpha_1} \geq c_1 \right\} \quad \alpha_0 - \alpha_1 < 0 \\
&= \left\{ \prod_{k=1}^n x_k \leq c_2 \right\}.
\end{aligned}$$

Расподела производа унутар критичне области није позната, па се треба снаћи. Углавном смо производе претварали у суме. Производ у суму можемо претворити једино користећи логаритам, па ћемо то и урадити. Све величине су позитивне, па немамо проблем.

$$\mathcal{W} = \left\{ \sum_{k=1}^n \ln x_k \leq c_3 \right\}.$$

Сада је згодно срачунати расподелу логаритма. Није тешко срачунати интеграљењем да је $F_X(x) = 1 - \left(\frac{1}{x}\right)^\alpha$. Онда можемо добити да је

$$\begin{aligned}
\mathbf{P}\{\ln X \leq x\} &= \mathbf{P}\{X \leq e^x\} \\
&= \cdots = 1 - \left(\frac{1}{e^x}\right)^\alpha \\
&= 1 - e^{-\alpha x},
\end{aligned}$$

што је функција расподеле експоненцијалне расподеле с параметром α. Иначе, расподела из задатка зове се Паретова расподела, па смо закључили да логаритам Паретове расподеле има експоненцијалну расподелу. За њу знамо очекивање и дисперзију.

Сада, под нултом хипотезом ($\alpha = \alpha_0 = 6/5$), тражимо константу c_3 као

$$\mathbf{P}_{H_0}\left\{ \sum_k \ln X_k \leq c_3 \right\} = 0.1.$$

Сума унутар витичастих заграда је сума независних и једнако расподељених експоненцијалних случајних величина, па има $\gamma(n, \alpha_0) = \gamma(12, 6/5)$ расподелу. За њу немамо таблицу, али знамо да је $\chi_n^2 = \gamma(n/2, 1/2)$, па некако треба „наместити”. Знамо да је $\ln X \sim \mathcal{E}(6/5)$ при H_0, па је онда $\frac{6}{5}\ln X \sim \mathcal{E}(1)$, па је онда $2\frac{6}{5}\ln X \sim \mathcal{E}(1/2)$ (просто проћи очекивањем и дисперзијом). Зато настављамо као

$$\mathbf{P}_{H_0}\left\{ \sum_{k=1}^{12} \frac{12}{5} \ln X_k \leq \frac{12}{5} c_3 \right\} = 0.1.$$

Сада сума унутар витичасте заграде има $\gamma(12, 1/2) = \gamma(24/2, 1/2)$ расподелу, што је једнако χ_{24}^2 расподели, па је

$$\frac{12}{5} c_3 = F_{\chi_{24}^2}^{-1}(0.1) \approx 15.65868,$$

па је

$$c_3 \approx 6.524452$$

, односно критична област је

$$\mathcal{W} = \left\{ \sum_{k=1}^{n} \ln x_k \leq 6.524452 \right\}.$$

Видимо да критична област никако не зависи од α_1, па овај тест јесте униформно најмоћнији. На крају рачунамо за конкретан узорак[1]

$$\sum_{k=1}^{12} \ln x_k = 12.32553,$$

што није мање од 6.524452, па не одбацујемо нулту хипотезу.

[1]Згодно је вратити на логаритам производа, јер се онда логаритам рачуна ручно само једном, а не 12 пута. Ја сам радио на рачунару па ми свеједно. Кажем за испит.