Испит 10.06.2022. МНЛ

Данијел Алексић

Јануар 2023.

Задатак 1. *Из популације чије је обележје* X *извучен је узорак Нека је* $Y = \ln X$*. Са прагом*

I_k	$(0, 8)$	$[8, 18)$	$[24, 32)$
n_k	42	28	30

значајности 0.1 *тестирати хипотезу да обележје* Y *има нормалну расподелу са једнаким очекивањем и дисперзијом.*

Решење. За почетак допунимо табелу за X интервалима који фале: Сада можемо прећи на

I_k	$(0, 8)$	$[8, 18)$	$[18, 24)$	$[24, 32)$	$[32, +\infty)$
n_k	42	28	0	30	0

табелу за Y: Ми тестирамо нулту хипотезу $H_0 : Y \sim \mathcal{N}(m, m)$; јасно, користимо Хи-квадрат

I_k	$(-\infty, \ln 8)$	$[\ln 8, \ln 18)$	$[\ln 18, \ln 24)$	$[\ln 24, \ln 32)$	$[\ln 32, +\infty)$
n_k	42	28	0	30	0

тест. За то, морамо наћи оцену параметра m методом максималне веродостојности. Функција веродостојности је

$$L(m) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi m}} e^{-\frac{(x_k-m)^2}{2m}} = \left(\frac{1}{2\pi m}\right)^n e^{-\sum_{k=1}^{n} \frac{(x_k-m)^2}{2m}}.$$

Логаритам функције веродостојности је

$$\begin{aligned}
\ln L(m) &= -\frac{n}{2}\ln(2\pi m) - \sum_{k=1}^{n} \frac{(x_k-m)^2}{2m} \\
&= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln m - \frac{1}{2m}\sum_{k=1}^{n}(x_k-m)^2 \\
&= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln m - \frac{1}{2m}\left(\sum x_k^2 - 2m\sum x_k + nm^2\right) \\
&= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln m - \frac{1}{2m}\sum x_k^2 + \sum x_k - \frac{nm}{2},
\end{aligned}$$

па је

$$\frac{\partial}{\partial m}\ln L(m) = -\frac{n}{2m} + \frac{1}{2m^2}\sum x_k^2 - \frac{n}{2},$$

те је једначина веродостојности $\frac{\partial}{\partial m}\ln L(m) = 0$, односно, након множења са $2m^2$,

$$-nm + \sum x_k^2 - nm^2 = 0.$$

Означимо ли $u = \sum_{k=1}^{n} x_k^2$, добијамо да је

$$-m^2 - m + u = 0.$$

Како је m уједно и дисперзија, не може бити негативно, те једно од два решења отпада и остаје

$$\hat{m} = \frac{-1 + \sqrt{1+4u}}{2}.$$

Да се показати да је максимум класичним методама. Како немамо све елементе x_k узорка, већ само припадност интервалу, користићемо класичну „фору” са средином интервала: узимамо да су одговарајући иксеви 4, 13, затим било шта, јер се множи нулом, затим 28 и поново било шта. За ипсилоне њихове логаритме. Добиће се $u = 5.980347$, као и $\hat{m} \approx 2$. Јасно, код нас је $n = 42 + 28 + 30 = 100$.

Следеће што треба срачунати јесте вероватноћа припадности интервалима. У преводу, треба попунити табелу На пример, вероватноћа p_2 рачуна се као

I_k	$(-\infty, \ln 8)$	$[\ln 8, \ln 18)$	$[\ln 18, \ln 24)$	$[\ln 24, \ln 32)$	$[\ln 32, +\infty)$
n_k	42	28	0	30	0
p_k					
np_k					

$$F_{\mathcal{N}(\hat{m},\hat{m})}(\ln 18) - F_{\mathcal{N}(\hat{m},\hat{m})}(\ln 8).$$

За логаритме ћемо лако, било директно на калкулатору, било фором попут $\ln 18 = \ln 2 + 2\ln 3$, па се сви логаритми своде на логаритам двојке и тројке. Већи проблем је то што не користимо функцију расподеле стандардне нормалне расподеле, за коју једино имамо таблицу. Ту је згодно уочити следеће

$$F_{\mathcal{N}(\hat{m},\hat{m})}(x) = \mathbf{P}\{\mathcal{N}(\hat{m},\hat{m}) \leq x\} = \mathbf{P}\left\{\frac{\mathcal{N}(\hat{m},\hat{m}) - \hat{m}}{\sqrt{\hat{m}}} \leq \frac{x - \hat{m}}{\sqrt{\hat{m}}}\right\} = \Phi\left(\frac{x-\hat{m}}{\sqrt{\hat{m}}}\right).$$

Сада се може искористити таблица и попунити да је Свуда је $np_k \geq 5$, па не спајамо категорије.

I_k	$(-\infty, \ln 8)$	$[\ln 8, \ln 18)$	$[\ln 18, \ln 24)$	$[\ln 24, \ln 32)$	$[\ln 32, +\infty)$
n_k	42	28	0	30	0
p_k	0.5235292	0.2131014	0.06196472	0.0522897	0.149115
np_k	52.352920	21.310142	6.196472	5.228970	14.911496

Јасно, код нас је $r = 5$ и број оцењених параметара је $s = 1$. Тест статистика коју користимо је

$$\chi_0^2 = \sum_{k=1}^{r} \frac{(n_k - np_k)^2}{np_k}$$

и она под нултом хипотезом има χ^2_{r-s-1} расподелу, што је код нас χ^2_3. Директан рачун покаже да је $\chi_0^2 = 142.6024$. Критична област је облика $\mathcal{W} = \{\chi_0^2 \geq c\}$, па c добијамо из услова

$$\mathbf{P}_{H_0}\{\chi_0^2 \geq c\} = \alpha = 0.1,$$

па је код нас $1 - \mathbf{P}_{H_0}\{\chi_0^2 \leq c\} = 0.1$

$$c = F^{-1}_{\chi^2_3}(0.9) = 6.251389.$$

Критична област је сада облика $\mathcal{W} = \{\chi_0^2 \geq 6.251389\}$, па тест статистика очигледно упада у критичну област. Одбацујемо нулту хипотезу.

Задатак 2. *Обележје X има густину расподеле*

$$f(x;\sigma^2) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, & x \geq 0 \\ 0, & \textit{иначе.} \end{cases}$$

Како би се одредио 88% интервал поверења за непознати параметар σ, $\sigma > 0$, изабран је узорак обима 100 и добијено је да је $\sum_{i=1}^{100} x_i^2 = 423.1024$. Ако је познато да је доња граница тог интервала поверења 1.327753, одредити горњу границу.

Решење. Наиме, лако је показати да је

$$F_{X^2}(x) = \mathbf{P}\{X^2 \le x\} = \mathbf{P}\{X \le \sqrt{x}\} = \int_0^{\sqrt{x}} \frac{t}{\sigma^2} e^{-\frac{t^2}{2\sigma^2}} dt,$$

па је, диференцирајући по горњој граници

$$f_{X^2}(x) = \frac{\sqrt{x}}{\sigma^2} e^{-\frac{x}{2\sigma^2}} \frac{1}{2\sqrt{x}} = \frac{1}{2\sigma^2} e^{-\frac{1}{2\sigma^2}x},$$

што је густина баш експоненцијалне $\mathcal{E}(1/2\sigma^2)$ расподеле. Дакле, $X^2 \sim \mathcal{E}(1/2\sigma^2)$. Сада следи „намештање". Имамо да је

$$\sum_{k=1}^{100} X_k^2 \sim \gamma\left(100, \frac{1}{2\sigma^2}\right),$$

па је

$$\frac{1}{\sigma^2} \sum_{k=1}^{100} X_k^2 \sim \gamma\left(\frac{200}{2}, \frac{1}{2}\right) = \chi^2_{200}.$$

Како интервали за σ и за σ^2 једнозначно одређују један другог (корен-квадрат крајева), то можемо сматрати да радимо са σ^2. Интервал тражимо у облику $[L_{100}, U_{100}]$, тако да буде

$$\mathbf{P}\{L_{100} \le \sigma^2 \le U_{100}\} = 0.88.$$

Након мало препакивања, добијамо да је то еквивалентно са

$$\mathbf{P}\left\{\frac{\sum X_k^2}{U_{100}} \le \frac{\sum X_k^2}{\sigma^2} \le \frac{\sum X_k^2}{L_{100}}\right\} = 0.88.$$

Нама је дато да је $L_{100} = 1.327753$, па је $\frac{\sum X_k^2}{L_{100}} = 318.6605$. Сада можемо поставити једначину

$$\mathbf{P}\{\chi^2_{200} > 318.6605\} = \alpha_1,$$

те из таблице добити да је $\alpha_1 = 1.87172 \cdot 10^{-7}$, што је приближно нула. Дакле, један реп је нула, „средина" је 0.88, па онда мора бити да је

$$\mathbf{P}\left\{\chi^2_{200} < \sum X_k^2 / U_{100}\right\} = 0.12,$$

односно

$$U_{100} = \frac{\sum X_k^2}{F^{-1}_{\chi^2_{200}}(0.12)},$$

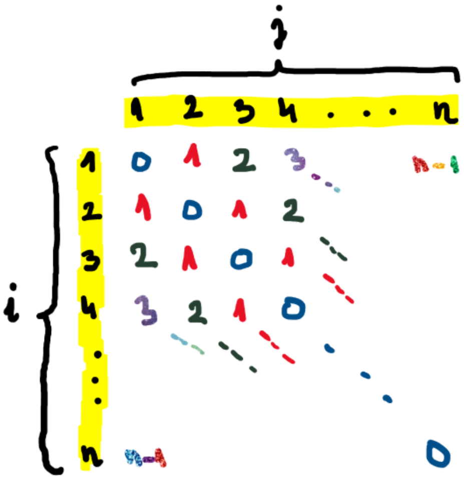
што за наше податке испада 2.393239, што је тражена горња граница.

Задатак 3. *Нека је (X_n) низ случајних величина такав да, за свако i и j, важи $\mathbf{Cov}(X_i, X_j) \le c^{|i-j|}$, где је $|c| \le 1$. Испитати да ли за овај низ важи слаби закон великих бројева.*

Решење. Знамо да за низ (X_n) важи слаби закон великих бројева уколико, за $S_n = X_1 + \cdots + X_n$ важи да $\frac{S_n - \mathbf{E}S_n}{n} \xrightarrow{P} 0$. Прво уочимо шта све може бити разлика $|i - j|$ када се оба индекса „прошетају" кроз скуп $\{1, 2, \ldots, n\}$. То видимо на доњој слици.

Сада рачунамо позивајући се на неједнакост Чебишова и на суму геометријског низа. Нека је $\varepsilon > 0$ произвољно.

$$\begin{aligned}\mathbf{P}\left\{\left|\frac{S_n - \mathbf{E}(S_n)}{n}\right| > \varepsilon\right\} &= \mathbf{P}\{|S_n - \mathbf{E}S_n| > n\varepsilon\} \\ &\le \frac{1}{n^2\varepsilon^2}\mathbf{E}(S_n - \mathbf{E}S_n)^2\end{aligned}$$



$$
\begin{aligned}
&= \frac{1}{n^2\varepsilon^2}\mathbf{E}\left(\sum_{i=1}^{n} X_i - \mathbf{E}\sum_{i=1}^{n} X_i\right)^2 \\
&= \frac{1}{n^2\varepsilon^2}\mathbf{E}\left(\sum_{i=1}^{n} (X_i - \mathbf{E}X_i)\right)^2 \\
&= \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{E}(X_i - \mathbf{E}X_i)(X_j - \mathbf{E}X_j) \\
&= \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{Cov}(X_i, X_j) \\
&\le \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}\sum_{j=1}^{n} c^{|i-j|} \\
&= \frac{1}{n^2\varepsilon^2}\left(nc^0 + 2(n-1)c^1 + 2(n-2)c^2 + \cdots + 2\cdot 1 \cdot c^{n-1}\right) \\
&\le \frac{1}{n^2\varepsilon^2}\left(2nc^0 + 2nc^1 + 2nc^2 + \cdots 2nc^{n-1}\right) \\
&= \frac{2}{n\varepsilon^2}(c^0 + c^1 + \cdots c^{n-1}) \\
&= \frac{2}{n\varepsilon^2}\frac{1-c^n}{1-c} \\
&\to 0 \cdot \frac{1}{1-c} \\
&= 0.
\end{aligned}
$$

Ово је би било довољно када $|c|$ не би могло да буде један. Иначе имамо дељење нулом. Тај се случај мора посебно испитати. Прво уочимо да се у последњем низу неједнакости може ставити свуда $|c|$ уместо c, јер тада само добијамо грубље ограничење (где год се одузимало сада се додаје). Тако је, заправо, суштински остало да се испита $c = 1$.

Нека је X случајна величина за коју је $\mathbf{E}X = 0, \mathbf{D}X = 1$. Формирајмо низ X_n такав да је за свако n испуњено $X_n = X$. Тада је $\mathbf{Cov}(X_i, X_j) = \mathbf{Cov}(X, X) = \mathbf{D}X = 1 \le 1 = c$, па су за овај низ испуњени услови задатка. Даље је $\frac{S_n - \mathbf{E}S_n}{n} = \frac{nX - n\mathbf{E}X}{n} = X$, па је $\mathbf{P}\{|(S_n - \mathbf{E}S_n)/n| > \varepsilon\} = \mathbf{P}\{|X| > \varepsilon\}$ константа, која не тежи нули осим ако је баш нула. Узмемо ли, на пример, $X \sim \mathcal{N}(0,1)$, добијамо да је та константа строго позитивна. Дакле, **за $c = 1$ не мора важити слаби закон великих бројева за низ (X_n), па је одговор на питање из задатка: НЕ.**