Универзитет у Београду
Математички факултет

ВЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ, 3Р)

8. јун 2023. године

---

1. Испитати све четири врсте конвергенције низа независних случајних величина $(X_n)_{n\in\mathbb{N}}$ ако је карактеристична функција φ_n општег члана тог низа дата са
$$\varphi_n(t) = \frac{e^{\frac{it}{n}}}{2024 - 2023 \cdot e^{\frac{it}{n}}}.$$
2. За густину расподеле обележја X важи да је $f(x) = \frac{\theta}{(1+x)^{\theta+1}}$, $x > 0$, $\theta > 0$. На основу узорка обима n, методом максималне веродостојности одредити оцену непознатог параметра $\frac{1}{\theta}$, а затим испитати ефикасност добијене оцене.
3. Дечак баца коцку за игру 120 пута и добија следећи узорак:

горња страна коцке	1	2	3	4	5	6
фреквенција	20	14	23	12	26	25

.

(а) Користећи овај узорак, са прагом значајности 0.05 тестирати хипотезу да је коцка за игру правилна.

(б) Дечак је уместо 120 пута бацио коцку n пута. Испоставило се да су релативне фреквенције у свакој од шест класа остале исте као у горњој табели. Колико најмање може бити n, ако је донет супротан закључак него у случају под (а)?

Универзитет у Београду
Математички факултет

ВЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ, 3Р)

8. јун 2023. године

---

1. Испитати све четири врсте конвергенције низа независних случајних величина $(X_n)_{n\in\mathbb{N}}$ ако је карактеристична функција φ_n општег члана тог низа дата са
$$\varphi_n(t) = \frac{e^{\frac{it}{n}}}{2024 - 2023 \cdot e^{\frac{it}{n}}}.$$
2. За густину расподеле обележја X важи да је $f(x) = \frac{\theta}{(1+x)^{\theta+1}}$, $x > 0$, $\theta > 0$. На основу узорка обима n, методом максималне веродостојности одредити оцену непознатог параметра $\frac{1}{\theta}$, а затим испитати ефикасност добијене оцене.
3. Дечак баца коцку за игру 120 пута и добија следећи узорак:

горња страна коцке	1	2	3	4	5	6
фреквенција	20	14	23	12	26	25

.

(а) Користећи овај узорак, са прагом значајности 0.05 тестирати хипотезу да је коцка за игру правилна.

(б) Дечак је уместо 120 пута бацио коцку n пута. Испоставило се да су релативне фреквенције у свакој од шест класа остале исте као у горњој табели. Колико најмање може бити n, ако је донет супротан закључак него у случају под (а)?

ВЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ, 3Р) - Писмени испит 8. јун 2023.

Решења задатака

1. Приметимо да је

$$\lim_{n\to\infty}\varphi_n(t)=\lim_{n\to\infty}\frac{e^{\frac{it}{n}}}{2024-2023\cdot e^{\frac{it}{n}}}=1,$$

па на основу теореме о непрекидности закључујемо

$$\varphi_n(t)\xrightarrow{n\to\infty}1,$$

односно

$$X_n\xrightarrow{D}0.$$

Како је $X_n\xrightarrow{D}0$, то важи и $X_n\xrightarrow{P}0$. У циљу испитивања средње квадратне конвергенције рачунамо $E(X_n^2)$. Имамо да је редом

$$\varphi_n'(t)=\frac{\frac{i}{n}e^{\frac{it}{n}}\cdot\left(2024-2023\cdot e^{\frac{it}{n}}\right)+e^{\frac{it}{n}}\cdot 2023\frac{i}{n}e^{\frac{it}{n}}}{\left(2024-2023\cdot e^{\frac{it}{n}}\right)^2}=\frac{2024ie^{\frac{it}{n}}}{n\left(2024-2023\cdot e^{\frac{it}{n}}\right)^2}.$$

$$\varphi_n''(t)=\frac{-2024e^{\frac{it}{n}}\left(2024+2023\cdot e^{\frac{it}{n}}\right)}{n^2\left(\left(2024-2023\cdot e^{\frac{it}{n}}\right)\right)^3}.$$

Одавде добијамо да је

$$EX_n^2=\frac{\varphi''(0)}{i^2}=\frac{2024\cdot\left(2024+2023\right)}{n^2}\xrightarrow{n\to\infty}0,$$

па имамо да низ (X_n) конвергира средње квадратно. За скоро скигурно конвергенцију на основу Чебишевљеве неједнакости важи

$$P\{|X_n|>\varepsilon\}\leq\frac{E|X_n|^2}{\varepsilon^2}=\frac{2024\cdot\left(2024+2023\right)}{n^2\varepsilon^2}.$$

Следи

$$\sum_{n=1}^{\infty}P\{|X_n|>\varepsilon\}\leq\sum_{n=1}^{\infty}\frac{2024\cdot\left(2024+2023\right)}{n^2\varepsilon^2},$$

те низ (X_n) конвергира скоро сигурно ка нули.

Напомена: Приметимо да можемо одредити расподелу низа случајних величина $(X_n)_{n\in\mathbf{N}}$. Заиста, имамо да важи:

$$\begin{aligned}\varphi_n(t)&=\frac{e^{\frac{it}{n}}}{2024-2023\cdot e^{\frac{it}{n}}}=\frac{1}{2024}\cdot e^{\frac{it}{n}}\cdot\sum_{k=0}^{\infty}\left(\frac{2023}{2024}e^{\frac{it}{n}}\right)^k\\&=\sum_{k=0}^{\infty}2023^k\left(\frac{1}{2024}e^{\frac{it}{n}}\right)^{k+1}=\sum_{k=1}^{\infty}2023^{k-1}\left(\frac{1}{2024}e^{\frac{it}{n}}\right)^k,\end{aligned}$$

одакле следи

$$P\left\{X_n=\frac{k}{n}\right\}=\frac{2023^{k-1}}{2024^k},\quad k\in\mathbf{N}.$$

Додатно, уочимо да низ $(nX_n)_{n\in\mathbf{N}}$ има геометријску $\mathcal{G}\left(\frac{1}{2024}\right)$ расподелу.

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

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

па је

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

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

$$\frac{\partial}{\partial \theta} \ln L(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} \ln(1+x_i) = 0 \quad \Longrightarrow \quad \theta = \frac{n}{\sum_{i=1}^{n} \ln(1+x_i)}.$$

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

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

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

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln(1+X_i)}.$$

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

$$T = \frac{1}{n} \sum_{i=1}^{n} \ln(1+X_i).$$

Да би имало смисла да испитујемо ефикасност оцене, морамо прво да проверимо да ли је она непристрасна. Нека је $Y_i = \ln(1+X_i)$. За почетак, одређујемо расподелу случајних величина $Y_i = \ln(1+X_i)$, $i = 1, 2, \ldots, n$. Приметимо да је

$$\begin{aligned} F_{Y_i}(y) &= P\{Y_i \leq y\} = P\{\ln(1+X_i) \leq y\} = P\{X_i \leq e^y - 1\} \\ &= \int_0^{e^y-1} \frac{\theta}{(1+x)^{\theta+1}} dx = 1 - e^{-\theta y}, \quad y \geq 0. \end{aligned}$$

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

$$E(T) = E\left(\frac{1}{n} \sum_{i=1}^{n} \ln(1+X_i)\right) = \frac{1}{n} \sum_{i=1}^{n} E(Y_i) = \frac{1}{\theta},$$

тј. оцена T јесте непристрасна оцена параметра $\frac{1}{\theta}$.

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

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

па је

$$\begin{aligned} \int_0^\infty \frac{\partial}{\partial a} f(x; a) dx &= \int_0^\infty \frac{\ln(1+x)}{a^3 (1+x)^{\frac{1}{a}+1}} dx - \int_0^\infty \frac{1}{a^2 (1+x)^{\frac{1}{a}+1}} dx \\ &= \frac{1}{a^3} \int_0^\infty t e^{-\frac{t}{a}} dt - \frac{1}{a^2} \int_1^\infty \frac{1}{t^{\frac{1}{a}+1}} dt \\ &= \frac{1}{a^3} \cdot a^2 \int_0^\infty u e^{-u} du + \frac{1}{a^2} \cdot \frac{a}{t^{\frac{1}{a}}} \Big|_1^\infty \\ &= \frac{1}{a} \Gamma(2) - \frac{1}{a^2} \cdot a = 0. \end{aligned}$$

Фамилија јесте регуларна. Рачунамо информациону функцију Фишера да бисмо добили доњу границу дисперзије свих непристрасних оцена за параметар $\frac{1}{\theta}$.

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

Дакле, важи

$$\begin{aligned}
\ln f(x;a) &= -\ln a - \left(\frac{1}{a}+1\right)\ln(1+x),\\
\frac{\partial}{\partial a}\ln f(x;\theta) &= -\frac{1}{a}+\frac{1}{a^2}\ln(1+x),\\
\frac{\partial^2}{\partial \theta^2}\ln f(x;\theta) &= \frac{1}{a^2}-\frac{2}{a^3}\ln(1+x),
\end{aligned}$$

па је

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

Према томе, доња граница дисперзије свих непристрасних оцена једнака је

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

Како је T 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

$$D(T) = D\left(\frac{1}{n}\sum_{i=1}^{n}\ln(1+X_1)\right) = \frac{1}{n^2}\sum_{i=1}^{n}D(Y_i) = \frac{1}{n^2}\cdot\frac{n}{\theta^2} = \frac{1}{n\theta^2}.$$

Дакле,

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

па ова оцена јесте ефикасна.

3. Користимо Пирсонов χ^2 тест сагласности са дискретном униформном расподелом. Број категорија је 6, и очекивани број елемената узорка у свакој категорији једнак је 20. Немамо оцењених параметара, те је број степени слободе једнак 5. Како је

$$\chi^2 = \frac{1}{20}\left[(20-20)^2+(14-20)^2+(23-20)^2+(12-20)^2+(26-20)^2+(25-20)^2\right] = \frac{210}{20} = 10.5,$$

те како је $10.5 < \chi^2_{5;0.05} = 11.07$, не одбацујемо нулту хипотезу о правилности коцке.

Супротан закључак био би да се хипотеза одбаци. Вредност тест статистике у овом случају била би

$$\chi^2 = \frac{1}{\frac{n}{6}}\left[\left(\frac{20}{120}n-\frac{n}{6}\right)^2+\left(\frac{14}{120}n-\frac{n}{6}\right)^2+\cdots+\left(\frac{25}{120}n-\frac{n}{6}\right)^2\right] = \frac{17n}{240},$$

те нам је за одбацивање нулте хипотезе неопходно да буде

$$\frac{17n}{240} > 11.07,$$

односно $n > 156.2$. Најмањи обим узорка за који ово важи једнак је 157.