Универзитет у Београду
Математички факултет

ВЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ, 3Р)

21. септембар 2023. године

---

1. За густину расподеле општег члана низа случајних величина (X_n) важи да је

$$f_{X_n}(x) = \frac{e^{-x}}{(1+e^{-x})^2},\ x \in \mathbb{R}.$$

Ако је $Y_n = \frac{1}{n(1+e^{-X_n})}$, испитати сва четири типа конвергенције низа случајних величина (Y_n).

2. Компанија "SunTech" је развила нову врсту соларног панела и жели да испита просечну ефикасност у претварању сунчеве светлости у електричну енергију. За густину расподеле обележја X, које представља ефикасност соларних панела, важи да је $f(x;\alpha) = \alpha x^{\alpha-1}$, $x \in (0,1)$, где је $\alpha > 0$ непознат параметар. Оснивачи компаније сматрају да је просечна ефикасност соларних панела једнака 0.7. Пре постављања на тржиште, "SunTech" је послао узорак од 10 нових панела у лабораторију на којима је извршена анализа ефикасности и добијени су следећи резултати: 0.991, 0.619, 0.959, 0.374, 0.228, 0.760, 0.628, 0.551, 0.779, 0.409. На основу добијеног узорка, са прагом значајности 0.1 тестирати најмоћнијим тестом да ли су власници у праву или је просечна ефикасност соларних панела мања.

3. Број поена на пријемном испиту за упис на Математички факултет има нормалну расподелу. Изабрано је 30 ученика на случајан начин и регистрован је њихов резултат на пријемном испиту:

број поена	[0,20]	(20,40]	(40,60]	(60,80]	(80,100]
број ученика	3	2	9	11	5

На основу овог узорка, одредити 92.5% интервал поверења за просечан број поена свих ученика на пријемном испиту, такав да вероватноћа да просечан број поена свих ученика на пријемном испиту буде мањи од леве границе интервала поверења буде дупло мања од вероватноће да просечан број поена свих ученика на пријемном испиту буде већи од десне границе интервала поверења.

ВЕРОВАТНОЋА И СТАТИСТИКА Б (4МНЛ, 3Р) - Писмени испит 21. септембар 2023.

Решења задатака

1. За функцију расподеле низа случајних величина X_n имамо да је

$$F_{X_n}(x) = \int_{-\infty}^{x} \frac{e^{-t}}{(1+e^{-t})^2} dt.$$

Ако уведемо смену $u = 1 + e^{-t}$, добијамо да је

$$F_{X_n}(u) = \int_{+\infty}^{w} (-1) \frac{(u-1)}{u^2(u-1)} du = \int_{w}^{+\infty} \frac{du}{u^2} = \frac{1}{w}$$

где је $w = 1 + e^{-x}$, па је

$$F_{X_n}(x) = \frac{1}{1+e^{-x}},\ x \in \mathbb{R}.$$

Приметимо да је сада

$$nY_n = F(X_n),$$

па је

$$F_{nY_n}(y) = P\{nY_n \leq y\} = P\{F(X_n) \leq y\} = P\{X_n \leq F^{-1}(y)\} = F(F^{-1}(y)) = y,\ y \in (0,1).$$

Дакле, добијамо да је $nY_n \in \mathcal{U}(0,1)$, па је

$$P\left\{Y_n \leq y\right\} = P\{nY_n \leq ny\} = ny,\ y \in \left(0, \frac{1}{n}\right),$$

односно следи $Y_n \in \mathcal{U}\left(0, \frac{1}{n}\right)$. Покажимо да ће низ Y_n конвергирати ка 0.
За конвергенцију у средње квадратном важи

$$E\left(Y_n\right)^2 = DY_n + (EY_n)^2 = \frac{1}{n^2}\left(\frac{1}{12} + \frac{1}{4}\right) = \frac{1}{3n^2} \to 0,\ n \to \infty.$$

Према томе, закључујемо да низ Y_n конвергира у средње квадратном ка нули, па следствено томе важи $Y_n \xrightarrow{P} 0$ и $Y_n \xrightarrow{D} 0$ када $n \to \infty$. У циљу испитивања скоро сигурне конвергенције приметимо да је

$$(\forall \varepsilon > 0)\ P\left\{|Y_n - 0| > \varepsilon\right\} \leq \frac{E\left|Y_n\right|^2}{\varepsilon^2} = \frac{1}{3\varepsilon^2 n^2}.$$

На основу претходног следи да ред $\sum_{n=1}^{\infty} P\left\{|Y_n - 0| > \varepsilon\right\}$ конвергира јер конвергира ред $\sum_{n=1}^{\infty} \frac{1}{3\varepsilon^2 n^2}$. Одавде следи да низ (Y_n) конвергира скоро сигурно ка нули.

Напомена. У случају да се не уочи приказана веза између случајних величина X_n и Y_n, наводимо одређивање функције расподеле случајне величине Y_n по дефиницији. Заиста, важи да је

$$\begin{aligned}
F_{Y_n}(y) = P\{Y_n \leq y\} &= P\left\{\frac{1}{n(1+e^{-X_n})} \leq y\right\} \\
&= P\left\{e^{-X_n} > 1 - \frac{1}{ny}\right\} = P\left\{X_n \leq -\ln\left(1 - \frac{1}{ny}\right)\right\} \\
&= F_{X_n}\left(-\ln\left(1 - \frac{1}{ny}\right)\right) = \frac{1}{1 + e^{\ln\left(1 - \frac{1}{ny}\right)}} \\
&= ny, \quad y \in \left(0, \frac{1}{n}\right).
\end{aligned}$$

2. Просечна ефикасност соларних панела дата је са
$$EX = \int_0^1 x\alpha x^{\alpha-1}dx = \alpha\int_0^1 x^{\alpha}dx = \frac{\alpha}{\alpha+1}.$$
Одавде следи да треба тестирати $H_0 : \frac{\alpha}{\alpha+1} = 0.7$ против $H_1 : \frac{\alpha}{\alpha+1} < 0.7$, што је еквивалентно са $H_0 : \alpha = \frac{7}{3}$ против $H_1 : \alpha < \frac{7}{3}$. Критичну област ћемо тражити помоћу Нејман-Пирсонове леме. Да би лема могла да се примени обе хипотезе морају бити просте. Нулта хипотеза свакако јесте проста $H_0 : \alpha = \alpha_0$, где је $\alpha_0 = \frac{7}{3}$, а алтернативну хипотезу напишимо као $H_1 : \alpha = \alpha_1,\ \alpha_1 < \frac{7}{3}$, те онда и она јесте проста јер смо параметру доделили тачно једну вредност.

За Нејман-Пирсонову лему је потребно одредити функцију веродостојности која је у овом случају дата са
$$L(x;\alpha) = \prod_{i=1}^{n}\alpha x_i^{\alpha-1} = \alpha^n\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1}.$$
Користећи лему, најбоља критична област је одређена са
$$\begin{aligned}
W &= \left\{\frac{L(x;\alpha_1)}{L(x;\alpha_0)} \geq k\right\} = \left\{\frac{\alpha_1^n\left(\prod_{i=1}^{n}x_i\right)^{\alpha_1-1}}{\alpha_0^n\left(\prod_{i=1}^{n}x_i\right)^{\alpha_0-1}} \geq k\right\}\\
&= \left\{\left(\frac{\alpha_1}{\alpha_0}\right)^n\left(\prod_{i=1}^{n}x_i\right)^{\alpha_1-\alpha_0} \geq k\right\} = \left\{\left(\prod_{i=1}^{n}x_i\right)^{\alpha_1-\alpha_0} \geq k_1\right\}\\
&= \left\{(\alpha_1-\alpha_0)\sum_{i=1}^{n}\ln x_i \geq \ln k_1\right\} = \left\{\sum_{i=1}^{n} -\ln x_i \geq c\right\},
\end{aligned}$$
где је $\alpha_1 - \alpha_0 < 0$ и c нека константа у односу на дати узорак, коју треба одредити да бисмо добили тачан израз за критичну област. На основу датог прага значајности, можемо одредити c тако да важи
$$P_{H_0}\left\{\sum_{i=1}^{n} -\ln x_i \geq c\right\} = 0.1.$$
Како при H_0, X има бета $\beta(\alpha_0, 1)$ расподелу, знамо да $-\ln X$ има експоненцијалну $\mathcal{E}(\alpha_0)$ расподелу, одакле имамо
$$\sum_{i=1}^{n} -\ln X_i \in \gamma(n, \alpha_0),$$
односно
$$2\alpha_0\sum_{i=1}^{n} -\ln X_i \in \gamma\left(n, \frac{1}{2}\right) = \chi^2_{2n}.$$
Сада за конкретне вредности $n = 10$, $\alpha_0 = \frac{7}{3}$, вредност c можемо добити из једначине
$$0.1 = P_{H_0}\left\{2\alpha_0\sum_{i=1}^{10} -\ln X_i \geq 2\alpha_0 c\right\}$$
$$c = \frac{1}{2\alpha_0}\chi^2_{20;0.1} = \frac{3}{14}\cdot 28.412 = 6.088.$$
Дакле, најбоља критична област је
$$W = \left\{\sum_{i=1}^{n} -\ln x_i \geq 6.088\right\}.$$
Коначно, на основу датог узорка добијамо да је
$$\sum_{i=1}^{10} -\ln x_i = 5.472,$$
што не припада критичној области, па не одбацујемо нулту хипотезу.

3. Нека просечан број поена на пријемном испиту има нормалну $\mathcal{N}(m, \sigma^2)$ расподелу. Како је σ^2 непознат параметар, користићемо случајну величину $T_n = \frac{\bar{X}_n - m}{\bar{S}_n}\sqrt{n-1}$, која има Студентову t_{n-1} расподелу.

Одредимо прво реализоване вредности статистика. С обзиром да немамо тачне вредности елемената узорка, већ само број елемената у одговарајућем интервалу, можемо сматрати да су елементи случајно расподељени унутар интервала и да је једнака вероватноћа да се налазе са леве и десне стране средине интервала, па можемо њом апроксимирати вредности.

На основу података из задатка, можемо одредити узорачку средину и узорачку дисперзију:

$$\bar{x}_{30} = \frac{3 \cdot 10 + 2 \cdot 30 + 9 \cdot 50 + 11 \cdot 70 + 5 \cdot 90}{30},$$

$$\bar{s}_{30}^2 = \frac{3 \cdot (10 - 58.67)^2 + 2 \cdot (30 - 58.67)^2 + 9 \cdot (50 - 58.67)^2 + 11 \cdot (70 - 58.67)^2 + 5 \cdot (90 - 58.67)^2}{30},$$

тј. $\bar{x}_{30} = 58.67$ и $\bar{s}_{30}^2 = 542.98$. Одредимо сада 92.5% интервал поверења за m у општем случају. Потребно је одредити статистике U_n и V_n за које је $P\{U_n \le V_n\} = 1$ такве да важи

$$P\{U_n \le m \le V_n\} = 0.95,$$

$$P\{\bar{X}_n - V_n \le \bar{X}_n - m \le \bar{X}_n - U_n\} = 0.925,$$

$$P\left\{\frac{\bar{X}_n - V_n}{\bar{S}_n}\sqrt{n-1} \le \frac{\bar{X}_n - m}{\bar{S}_n}\sqrt{n-1} \le \frac{\bar{X}_n - U_n}{\bar{S}_n}\sqrt{n-1}\right\} = 0.925.$$

Из услова задатка имамо да за овакав интервал поверења треба да важи $P\{m < U_n\} = 0.025$ и $P\{m > V_n\} = 0.05$, односно

$$P\left\{T_n < \frac{\bar{X}_n - V_n}{\bar{S}_n}\sqrt{n-1}\right\} = 0.05, \quad P\left\{T_n > \frac{\bar{X}_n - U_n}{\bar{S}_n}\sqrt{n-1}\right\} = 0.025.$$

Користећи да T_n има Студентову расподелу, добијамо да је

$$\frac{\bar{X}_n - V_n}{\bar{S}_n}\sqrt{n-1} = -t_{n-1;2\cdot 0.05},$$

$$\frac{\bar{X}_n - U_n}{\bar{S}_n}\sqrt{n-1} = t_{n-1;2\cdot 0.025}.$$

Према томе,

$$U_n = \bar{X}_n - \frac{\bar{S}_n}{\sqrt{n-1}}t_{n-1;0.05},$$

$$V_n = \bar{X}_n + \frac{\bar{S}_n}{\sqrt{n-1}}t_{n-1;0.1}.$$

Односно интервал поверења је

$$I_m^{(n)} = \left[\bar{X}_n - \frac{\bar{S}_n}{\sqrt{n-1}}t_{n-1;0.05},\ \bar{X}_n + \frac{\bar{S}_n}{\sqrt{n-1}}t_{n-1;0.1}\right].$$

У конкретном случају, $\beta = 0.0.925$, $t_{29;0.05} = 2.045$ и $t_{29;0.1} = 1.699$. Тражени интервал је

$$I_m^{(30)} = \left[58.67 - 2.045 \cdot \frac{\sqrt{542.98}}{\sqrt{29}}, 58.67 + 1.699 \cdot \frac{\sqrt{542.98}}{\sqrt{29}}\right],$$

$$I_m^{(30)} = [49.821, 66.022].$$