б) Пошто је X позитивна случајна величина, следи да ће и Z бити позитивна случајна величина. Дакле, за $z \geq 0$ важи

$$\begin{aligned} F_Z(z) &= P\{Z \leq z\} = P\{\min\{X, X^2\} \leq z\} = 1 - P\{\min\{X, X^2\} > z\} \\ &= 1 - P\{X > \max\{z, \sqrt{z}\}\} = P\{X \leq \max\{z, \sqrt{z}\}\} \\ &= F_X(\max\{z, \sqrt{z}\}) = 1 - e^{-\lambda \max\{z, \sqrt{z}\}}. \end{aligned}$$

Према томе,

$$F_Z(z) = \begin{cases} 0, & z < 0, \\ 1 - e^{-\lambda \max\{z, \sqrt{z}\}}, & z \geq 0. \end{cases}$$

7. Нека је X апсолутно непрекидна случајна величина на носачу $[a, b], a, b \in \mathbb{R}$ са функцијом расподеле F_X. Одредити расподелу случајне величине $Y = F_X(X)$.

Решење 7. За $y \in (0, 1)$ важи

$$F_Y(y) = P\{Y \leq y\} = P\{F_X(X) \leq y\} = P\{X \leq F_X^{-1}(y)\} = F_X(F_X^{-1}(y)) = y, \ \ y \in (0, 1).$$

Дакле, како је

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ y, & 0 \leq y < 1 \\ 1, & y \geq 1, \end{cases}$$

следи да Y има униформну $\mathcal{U}(0, 1)$.

Основни појмови у статистици

Уведимо сада појмове које ћемо често користити.

Ω - популација; популација је скуп елемената чије карактеристике желимо да испитамо;

$X : \Omega \to \mathbb{R}$ - обележје популације; обележјем популације представљамо карактеристике популације које нас интересују;

Случајни вектор $(X_1, X_2, \ldots, X_n)$ код кога су $X_1, X_2, \ldots, X_n$ независне и једнако расподељене зовемо прост случајни узорак (ПСУ). Свака реална функција од узорка $T = T(X_1, X_2, \ldots, X_n)$, која не зависи од непознатих параметара, назива се статистика.

npr. $X_i \sim \mathcal{N}(0, \sigma^2)$ nepoznato

$\Rightarrow \sum_{i=1}^{n} X_i + \sigma^2$ nije statistika

Познате статистике

1. узорачка средина: $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$
2. узорачка дисперзија: $\overline{S_n}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_n})^2$
3. варијациони низ: низ $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ сачињен од елемената узорка који је растуће сортиран; статистику $X_{(i)}$ зовемо i-та статистика поретка
   $X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$ - прва статистика поретка
   $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$ - n-та статистика поретка

   $X_{(i)}(\omega) = i$- ti član rastuće sortiranog niza: $X_1(\omega), \ldots, X_n(\omega)$
4. узорачка медијана:

$$Me = \begin{cases} X_{(\frac{n+1}{2})}, & n \text{ непарно}, \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}, & n \text{ парно}. \end{cases}$$

   npr. (4,-3,6) - realizovane vrednosti uzorka
   Me
   (3,8,2,0) => Me = (2+3)/2
5. узорачки распон: $R = X_{(n)} - X_{(1)}$

8. Нека је $(X_1, X_2, \ldots, X_n)$ прост случајан узорак (ПСУ) из експоненцијалне $\mathcal{E}(\beta)$ расподеле. Наћи расподелу узорачке средине.

Решење 8. Узорачка средина је дефинисана као

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Сетимо се да је збир независних експоненцијалних расподела са истим параметром заправо гама расподела и важи следеће

$$Z = \sum_{i=1}^{n} X_i \sim \Gamma(n, \beta).$$

Даље, тражимо функцију расподеле за узорачку средину

$$F_{\overline{X}_n}(t) = P\{\overline{X}_n \le t\} = P\{Z \le n \cdot t\} = F_Z(nt) = \int_0^{nt} \frac{x^{n-1}\beta^n e^{-\beta x}}{(n-1)!}dx, \quad t \in [0, \infty). \tag{1}$$

$= \frac{Z}{n}$

▲

9. Нека је $(X_1, X_2, \ldots, X_n)$ прост случајан узорак из популације чије обележје има функцију расподеле F_X и $t \in \mathbb{R}$ фиксиран број. Одредити очекивање и расподелу од $nF_n(t)$.

Решење 9. Сетимо се да је емпиријска функција расподеле дефинисана на следећи начин

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \le t\}, \quad \forall t \in \mathbb{R}. \quad = \frac{\text{br. elemenata iz uzorka <=t}}{n} \tag{2}$$

Дакле, $nF_n(t)$ је само сума индикатора. Како је узорак прост и случајан то имамо да збир индикатора заправо биномна расподела, тј. важи

$$nF_n(t) \sim Bin(n, \underbrace{F_X(t)}_{P\{X<=t\}}). \tag{3}$$

nezavisni,
jer su X_i
nezavisne

Сада је

$$\mathbb{E}(nF_n(t)) = nF_X(t).$$

▲

Тачкасте оцене

Обележје популације X је случајна величина чија расподела може зависити од параметра θ. Како можемо оценити тај параметар? Посматрајмо ПСУ $(X_1, X_2, \ldots, X_n)$ у коме свака случајна величина X_i има исту расподелу као X. Тачкаста оцена параметра θ је дата са:

$$\hat{\theta}_n = f(X_1, X_2, \ldots, X_n),$$

где је f нека статистика.

Наведимо неке важне особине које може имати оцена $\hat{\theta}_n$:

- оцена $\hat{\theta}_n$ је непристрасна ако је $E(\hat{\theta}_n) = \theta$; ако $E(\hat{\theta}_n) \to \theta$, кад $n \to +\infty$, оцена је асимптотски непристрасна
- оцена $\hat{\theta}_n$ је постојана ако:

$$(\forall \epsilon > 0)P\{|\hat{\theta}_n - \theta| > \epsilon\} \to 0, \quad n \to +\infty$$

za velike vrednisti n ocena $\hat{\theta}_n$ je bliska θ

Тврђење. Ако важи: $E(\hat{\theta}_n - \theta)^2 \to 0$, кад $n \to +\infty$, тада је оцена $\hat{\theta}_n$ постојана.

želimo da bude malo

Величина $E(\hat{\theta}_n - \theta)^2$ се зове средњеквадратна грешка оцене $\hat{\theta}_n$ и означава се са $MSE(\hat{\theta}_n)$. Приметимо да уколико је оцена $\hat{\theta}_n$ непристрасна, важи $MSE(\hat{\theta}_n) = D(\hat{\theta}_n)$.

Дакле, да бисмо показали постојаност непристрасне оцене, довољно је да покажемо да $D(\hat{\theta}_n) \to 0$, кад $n \to +\infty$.

Од две оцене параметра θ боља је она која има мању средњеквадртну грешку.

10. Применом претходног задатка показати да је $F_n(t)$ непристрасна и постојана оцена за $F_X(t)$.

Решење 10. Како је $\mathbb{E}(nF_n(t)) = n\mathbb{E}(F_n(t))$, видимо из претходног задатка да је $n\mathbb{E}(F_n(t)) = nF_X(t)$, односно $\mathbb{E}(F_n(t)) = F_X(t)$. Дакле, оцена је непристрасна.

Да бисмо испитали постојаност непристрасне оцене, довољно је да покажемо да важи:

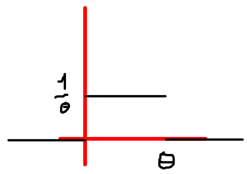
$$\lim_{n\to+\infty} D(F_n(t)) = 0.$$

Како $nF_n(t) \sim Bin(n, F_X(t))$, знамо да је $D(nF_n(t)) = nF_X(t)(1 - F_X(t))$. Користећи познато својство дисперзије $D(nF_n(t)) = n^2 D(F_n(t))$, добијамо:

$$D(F_n(t)) = \frac{1}{n} F_X(t)(1 - F_X(t)),$$

што заиста тежи 0 кад $n \to +\infty$ ▲

11. Нека је $X_1, X_2, \ldots, X_n$ прост случајан узорак из популације чије обележје X има униформну $\mathcal{U}[0, \theta]$ расподелу. За оцену непознатог параметра θ предложене су следеће две оцене:



$$\widehat{\theta_1} = c \cdot X_{(n)},$$

$$\widehat{\theta_2} = 2 \cdot \overline{X_n}.$$

(а) Одредити константу $c \in \mathbb{R}$ тако да прва оцена буде непристрасна оцена параметра θ.

(б) Испитати непристрасност друге оцене.

(в) Испитати да ли су оцене постојане и на основу тога одредити која је боља.

Решење 11.

(а) Да би оцена $\widehat{\theta_1}$ била непристрасна, потребно је да важи:

$$E(\widehat{\theta_1}) = c \cdot E(X_{(n)}) = \theta.$$

Дакле, потребно је одредити $E(X_{(n)})$. Израчунајмо прво функцију расподеле од $X_{(n)}$.

$$F_{X_{(n)}}(x) = P\{X_{(n)} \leq x\} = P\{X_1 \leq x, \ldots, X_n \leq x\} = P\{X_1 \leq x\} \cdot \ldots \cdot P\{X_n \leq x\} = (F(x))^n,$$

где је $F(x) = \frac{x}{\theta}$, $\forall x \in [0, \theta]$. Следи:

$$f_{X_{(n)}}(x) = F'_{X_{(n)}}(x) = n\frac{x^{n-1}}{\theta^n}, \quad \forall x \in [0, \theta].$$

Сада можемо израчунати очекивање:

$$E(X_{(n)}) = \int_0^\theta xn\frac{x^{n-1}}{\theta^n}dx = \frac{n}{n+1}\theta.$$

Коначно добијамо да је

$$c = \frac{\theta}{E(X_{(n)})} = \frac{n+1}{n}.$$

(б)

$$E(\widehat{\theta_2}) = E(2\overline{X_n}) = 2E(\overline{X_n}) = 2E(\frac{\sum_{i=1}^n X_i}{n}) = \frac{2}{n}\sum_{i=1}^n E(X_i) = n \cdot \frac{2}{n} \cdot \frac{\theta}{2} = \theta.$$

Дакле, оцена $\widehat{\theta_2}$ јесте непристрасна.

(в) Како су обе оцене непристрасне, боља оцена је она која има мању дисперзију. Израчунајмо зато дисперзије добијених оцена:

$$D(\widehat{\theta_1}) = D\left(\frac{n+1}{n}X_{(n)}\right) = \frac{(n+1)^2}{n^2}D(X_{(n)}) = \frac{(n+1)^2}{n^2}(E(X_{(n)}^2) - (E(X_{(n)}))^2)$$

Дакле, сада је потребно израчунати $E(X_{(n)}^2)$:

$$E(X_{(n)}^2) = \int_0^\theta x^2 n\frac{x^{n-1}}{\theta^n}dx = \theta^2\frac{n}{n+2}$$

Следи:

$$D(\widehat{\theta}_1) = \frac{(n+1)^2}{n^2}\left(\theta^2\frac{n}{n+2} - \theta^2\frac{n^2}{(n+1)^2}\right) = \frac{\theta^2}{n(n+2)}.$$

Рачунамо сада дисперзију друге оцене:

$$D(\widehat{\theta}_2) = D\left(2\cdot\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{4}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{4}{n^2}\cdot n\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

Како $D(\widehat{\theta}_1), D(\widehat{\theta}_2) \to 0$ кад $n \to +\infty$, обе оцене су постојане.
За све $n > 1$ важи $n(n+2) > 3n$, па важи:

$$\frac{\theta^2}{n(n+2)} < \frac{\theta^2}{3n},$$

односно:

$$D(\widehat{\theta}_1) < D(\widehat{\theta}_2).$$

Дакле, боља оцена је $\widehat{\theta}_1$.

▲

Метод момената

Један начин на који можемо добити оцене параметара јесте метод момената. Под њиме подразумевамо изједначавање теоријских и узорачких момената:

motivacija: zakon velikih brojeva

za n veliko: $\overline{X}_n \approx EX_1$

ovo je samo zapis!

$$EX = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n;$$

$$E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2;$$

$$\vdots$$

ispravnije je da stoji $\approx$
umesto =...

$$E(X^k) = \frac{1}{n}\sum_{i=1}^{n} X_i^k;$$

$$D(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \overline{S}_n^2.$$

Погледајмо примере:

12. Методом момената одредити оцене непознатих параметара:

(а) p за $X \sim Bin(N, p)$, где је N познато,

(б) θ за $X \sim \mathcal{U}[0, \theta]$,

(в) p за

$$X : \begin{pmatrix} -1 & 0 & 1 \\ p & p & 1-2p \end{pmatrix},$$

(г) α и β за $X \sim \Gamma(\alpha, \beta)$,

(д) a и b за $X \sim \mathcal{U}[a, b]$

parametar na jednu stranu,
ostalo na drugu

Решење 12.

(а) Непознат параметар p се одређује из $E(X) = \overline{X}_n$. Како је $E(X) = np$, из $np = \overline{X}_n$ добија се $\hat{p} = \frac{1}{n}\overline{X}_n$ као оцена непознатог параметра p.

(б) Како је $E(X) = \frac{\theta}{2}$, из $\frac{\theta}{2} = \overline{X}_n$ се добија $\hat{\theta} = 2\overline{X}_n$.

(в) Из $E(X) = 1 - 2p - p = \overline{X}_n$, добија се $\hat{p} = \frac{1}{3}(1 - \overline{X}_n)$.

obično ovim
redosledom:

(г) Из система једначина $E(X) = \frac{\alpha}{\beta} = \overline{X}_n$, $D(X) = \frac{\alpha}{\beta^2} = \overline{S}_n^2$, добија се $\hat{\alpha} = \frac{\overline{X}_n^{\,2}}{\overline{S}_n^2}$ и $\hat{\beta} = \frac{\overline{X}_n}{\overline{S}_n^2}$.

1. EX
2. DX ili EX^2

kada imamo 2 parametra, treba nam više jednačina, pa biramo one najlakše (д) Оцене $\hat{a}$ и $\hat{b}$ се добијају као решења система једначина $E(X) = \frac{a+b}{2} = \overline{X}_n$, $D(X) = \frac{(b-a)^2}{12} = \overline{S}_n^2$. ▲

13. Методом момената одредити оцену непознатог параметра a за обележје X из експоненцијалне $\mathcal{E}(a)$ расподеле. Испитати непристрасност добијене оцене.

Решење 13. Како је $EX = \frac{1}{a}$, оцена методом момената добија се из једнакости $\frac{1}{a} = \overline{X}_n$, одакле је

$$\hat{a} = \frac{1}{\overline{X}_n} = \frac{n}{\sum_{i=1}^{n} X_i}.$$

Пошто су $X_1, X_2, \dots, X_n$ независне са расподелом $\mathcal{E}(a)$, следи да је случајна величина $Y = X_1 + X_2 + \cdots + X_n \sim \Gamma(n, a)$. Зато је:

$$E(\hat{a}) = E(\frac{n}{Y}) = nE(\frac{1}{Y}) = n\int_0^{+\infty} \frac{1}{y} f_Y(y)dy = n\int_0^{+\infty} \frac{1}{y}\frac{y^{n-1}e^{-ay}a^n}{(n-1)!}dy = na\int_0^{+\infty} \frac{y^{n-1-1}e^{-ay}a^{n-1}}{(n-1)(n-2)!}dy$$

$$= \frac{n}{n-1}a \neq a$$

Дакле, оцена добијена методом момената није непристрасна, али јесте асимптотски непристрасна, јер $E(\hat{a}) = \frac{n}{n-1}a \to \cancel{+\infty}\ a$, кад $n \to +\infty$. ▲

14. Методом момената наћи оцену непознатих параметрара μ и a за X са функцијом густине

$$f(x) = ae^{-a(x-\mu)}, \quad x \geq \mu, \quad \mu \in \mathbb{R}.$$

ima istu raspodelu kao sl. vel. $\underset{\sim \mathcal{E}(a)}{Y} + \mu$

(za vežbu)

Решење 14. Израчунајмо прво $E(X)$:

$$E(X) = \int_{\mu}^{+\infty} xae^{-a(x-\mu)}dx = \int_0^{+\infty} (t+\mu)ae^{-at}dt = \int_0^{+\infty} tae^{-at}dt + \mu\int_0^{+\infty} ae^{-at}dt = \frac{1}{a} + \mu$$

Поставимо прву једначину:

$$\overline{X}_n = \frac{1}{\hat{a}} + \hat{\mu}.$$

Како је потребно оценити два параметра, потребна нам је још једна једначина. Одредимо зато $D(X)$:

$$D(X) = E(X - E(X))^2 = \int_{\mu}^{+\infty} (x - \frac{1}{a} - \mu)^2 ae^{-a(x-\mu)}dx = \int_0^{+\infty} (t - \frac{1}{a})^2 ae^{-at}dt = \frac{1}{a^2}$$

Последња једнакост важи јер је у питању дисперзија случајне величине из $\mathcal{E}(a)$ расподеле. Сада можемо поставити и другу једначину:

$$\overline{S}_n^2 = \frac{1}{\hat{a}^2},$$

одакле добијамо оцену $\hat{a}$. Повратком у прву једначину добијамо и оцену $\hat{\mu}$. ▲

15. Испитати непристрасност оцена $\overline{X}_n$ и $\overline{S}_n^2$ за параметре m и σ^2 за X из нормалне $\mathcal{N}(m, \sigma^2)$ расподеле.

Решење 15. Рађено на часу. ▲

16. Методом максималне веродостојности наћи оцене непознатих параметара:

(а) a за Пуасонову расподелу $\mathcal{P}(a)$,

(б) p за $X \sim Ber(p)$,

(в) p за $X \sim G(p)$,

(г) p за

$$X : \begin{pmatrix} -1 & 0 & 1 \\ p & p & 1-2p \end{pmatrix},$$