Карактеристична функција

Дефиниција. Нека је X случајна величина са функцијом расподеле F. Карактеристична функција случајне величине X (односно њене функције расподеле) је функција $\varphi_X : \mathbb{R} \to \mathbb{C}$ одређена са

$$\varphi_X(t) = E(e^{itX}) = \int_{-\infty}^{+\infty} e^{itx} dF(x).$$

Особине карактеристичне функције:

1. $|\varphi(t)| \leq \varphi(0) = 1$,
2. $\varphi(-t) = \overline{\varphi(t)}$,
3. φ је равномерно непрекидна,
4. $\varphi_{aX+b}(t) = \varphi_X(at)e^{itb}$.

Теорема 1. *Нека је X случајна величина са карактеристичном функцијом $\varphi(t)$. Ако за неки природан број n важи $E|X|^n < \infty$, онда за свако $k \leq n$ постоји извод $\varphi^{(k)}(t)$ и важи једнакост*

$$\varphi^{(k)}(t) = \int_{-\infty}^{+\infty} (ix)^k e^{itx} dF(x), \ k = 0, 1, ..., n,$$

и тада је $EX^k = \frac{\varphi^{(k)}(0)}{i^k}$ $k-$ти моменат. $\star$

Теорема 2. *(Теорема о производу) Нека су $X_1, ..., X_n$ независне случајне величине. Тада је*

$$\varphi_{X_1+\cdots+X_n}(t) = \varphi_{X_1}(t) \cdots \varphi_{X_n}(t).$$

$\star$

Карактеристична функција јединствено одређује расподелу, односно ако две функције расподела имају исту карактеристичну функцију, онда су оне идентичне.

Задатак 1. Одредити карактеристичну функцију случајне величине:

а) X_1 која има $\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$ расподелу;

б) X_2 која има биномну $\mathcal{B}(n, p)$ расподелу;

в) X_3 која има Пуасонову $\mathcal{P}(\lambda)$ расподелу и доказати да ако X_3 има Пуасонову $\mathcal{P}(\lambda)$ расподелу, а X_4 има Пуасонову $\mathcal{P}(\mu)$ расподелу и независне су, онда њихов збир $X_3 + X_4$ има Пуасонову $\mathcal{P}(\lambda + \mu)$ расподелу;

г) X_5 чија густина расподеле је $f(x) = \frac{1}{2}e^{-|x|}, x \in \mathbb{R}$, а затим израчунати очекивање EX_5 и дисперзију DX_5.

Решење. а) Како је $X_1 : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$, то

$$\varphi_{X_1}(t) = Ee^{itX_1} = (1-p)e^{it\cdot 0} + pe^{it} = 1 - p + pe^{it}.$$

б) Како је $P\{X_2 = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \ k = 0, 1, ..., n$, то

$$\varphi_{X_2}(t) = Ee^{itX_2} = \sum_{k=0}^{n} e^{itk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^{it})^k (1-p)^{n-k} = (pe^{it} + 1 - p)^n.$$

в) Како је $P\{X_3 = k\} = \frac{\lambda^k e^{-\lambda}}{k!},\ k \in \mathbb{N}_0$, то

$$\varphi_{X_3}(t) = Ee^{itX_3} = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{-\lambda(1-e^{it})}.$$

Слично, $\varphi_{X_4}(t) = e^{-\mu(1-e^{it})}$. Тада

$$\varphi_{X_3+X_4}(t) = \varphi_{X_3}(t)\varphi_{X_4}(t) = e^{-\lambda(1-e^{it})} e^{-\mu(1-e^{it})} = e^{-(\lambda+\mu)(1-e^{it})}.$$

Дакле, $X_3 + X_4 \in \mathcal{P}(\lambda + \mu)$.

г) Како је $f(x) = \frac{1}{2} e^{-|x|}, x \in \mathbb{R}$, то

$$\begin{aligned}\varphi_{X_5}(t) = Ee^{itX_5} &= \int_{-\infty}^{+\infty} e^{itx} \frac{1}{2} e^{-|x|} dx = \frac{1}{2} \left(\int_{-\infty}^{0} e^{x(it+1)} dx + \int_{0}^{\infty} e^{-x(1-it)} dx \right) \\ &= \frac{1}{2(it+1)} + \frac{1}{2(1-it)} = \frac{1}{1+t^2}.\end{aligned}$$

Треба још одредити очекивање и дисперзију. Користећи теорему 1 следи

$$\begin{aligned}\varphi'_{X_5}(t) &= -\frac{2t}{(1+t^2)^2} \implies EX_5 = \frac{\varphi'_{X_5}(0)}{i} = 0 \\ \varphi''_{X_5}(t) &= -\frac{2-6t^2}{(1+t^2)^3} \implies EX_5^2 = \frac{\varphi''_{X_5}(0)}{i^2} = 2 \implies DX_5 = EX_5^2 - (EX_5)^2 = 2.\end{aligned}$$

✓

Задатак 2. Одредити расподелу случајне величине за чију карактеристичну функцију $\varphi(t)$ важи:

а) $\varphi(t) = \frac{3+\cos t}{4}$;

б) $\varphi(t) = (1-\beta)(1+\alpha)^{-1}(1+\alpha e^{-it})(1-\beta e^{it})^{-1}, 0 < \alpha < \beta < 1$;

в) $\varphi(t) = \frac{1}{2-e^{it}}$.

Решење. а) Приметимо да

$$\varphi(t) = \frac{3}{4} + \frac{\cos t}{4} = \frac{3}{4} + \frac{1}{4} \frac{e^{it} + e^{-it}}{2} = \frac{1}{8} e^{-it} + \frac{3}{4} e^{it\cdot 0} + \frac{1}{8} e^{it} \implies X : \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \end{pmatrix}$$

б) Приметимо да

$$\begin{aligned}\varphi(t) &= \frac{1-\beta}{1+\alpha} \frac{1+\alpha e^{-it}}{1-\beta e^{it}} = \frac{1-\beta}{1+\alpha} (1+\alpha e^{-it}) \sum_{k=0}^{\infty} (\beta e^{it})^k \\ &= \frac{1-\beta}{1+\alpha} \left(\sum_{k=0}^{\infty} (\beta e^{it})^k + \sum_{k=0}^{\infty} \alpha \beta^k e^{it(k-1)} \right) \\ &= \frac{1-\beta}{1+\alpha} \left(\sum_{k=0}^{\infty} (\beta e^{it})^k + \alpha e^{-it} + \alpha\beta \sum_{k=0}^{\infty} (\beta e^{it})^k \right) \\ &= \frac{1-\beta}{1+\alpha} \alpha e^{-it} + \frac{1-\beta}{1+\alpha} (1+\alpha\beta) \sum_{k=0}^{\infty} (\beta e^{it})^k\end{aligned}$$

$$\Longrightarrow P\{X=-1\} = \alpha\frac{1-\beta}{1+\alpha}$$
$$P\{X=k\} = (1+\alpha\beta)\frac{1-\beta}{1+\alpha}\beta^k,\ k \in \mathbb{N}_0$$

в) Приметимо да

$$\varphi(t) = \frac{1}{2-e^{it}} = \frac{1}{2}\frac{1}{1-\frac{1}{2}e^{it}} = \frac{1}{2}\sum_{k=0}^{\infty}\left(\frac{e^{it}}{2}\right)^k = \sum_{k=0}^{\infty}\frac{1}{2^{k+1}}e^{itk}$$
$$\Longrightarrow P\{X=k\} = \frac{1}{2^{k+1}},\ k \in \mathbb{N}_0$$

✓

Задатак 3. Ако су Y и Z независне случајне величине такве да је $X = Y + Z$, где X има $\mathcal{U}(0, n+1)$, а Y има $\mathcal{U}(0,1)$ расподелу, одредити расподелу случајне величине Z.

Решење. Како $X \in \mathcal{U}(0, n+1)$, то следи да

$$\varphi_X(t) = \int_0^{n+1} e^{itx}\frac{1}{n+1}dx = \frac{1}{n+1}\frac{e^{it(n+1)}-1}{it}.$$

Слично, како $Y \in \mathcal{U}(0,1)$, то следи да

$$\varphi_Y(t) = \int_0^1 e^{itx}dx = \frac{e^{it}-1}{it}.$$

Из независности следи да је $\varphi_X(t) = \varphi_Y(t)\varphi_Z(t)$. Према томе

$$\varphi_Z(t) = \frac{\varphi_X(t)}{\varphi_Y(t)} = \frac{\frac{1}{n+1}\frac{e^{it(n+1)}-1}{it}}{\frac{e^{it}-1}{it}} = \frac{1}{n+1}\frac{e^{it(n+1)}-1}{e^{it}-1} = \frac{1}{n+1}\sum_{k=0}^{n}(e^{it})^k$$
$$\Longrightarrow Z : \begin{pmatrix} 0 & 1 & \cdots & n \\ \frac{1}{n+1} & \frac{1}{n+1} & \cdots & \frac{1}{n+1} \end{pmatrix}$$

✓

Борел-Кантелијева лема

Нека је $(A_n)_{n\in\mathbb{N}}$ низ догађаја. Дефинишемо следеће догађаје:

- $\limsup\limits_{n\to\infty} A_n = \overline{\lim\limits_{n\to\infty}} A_n = \bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty} A_k$,
- $\liminf\limits_{n\to\infty} A_n = \underline{\lim\limits_{n\to\infty}} A_n = \bigcup\limits_{n=1}^{\infty}\bigcap\limits_{k=n}^{\infty} A_k$.

Лема 1. *1. Ако је (A_n) произвољан низ догађаја и $\sum\limits_{n=1}^{\infty} P(A_n) < \infty$, онда важи $P(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty} A_k) = 0$.*

2. Ако је (A_n) низ независних догађаја и $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$, онда важи $P(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty} A_k) = 1$.

⋆

Последица. *Нека је* (A_n) *низ независних догађаја. Тада важи*

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) = 0 \iff \sum_{n=1}^{\infty} P(A_n) < \infty.$$

⋆

Чебишовљева неједнакост: Нека је X случајна величина са коначним моментом $r-$тог реда, $r > 0$. За свако $\varepsilon > 0$ важи

$$P\left\{|X| \geq \varepsilon\right\} \leq \frac{E|X|^r}{\varepsilon^r}.$$

Специјално, $P\left\{|X - EX| \geq \varepsilon\right\} \leq \frac{E|X-EX|^2}{\varepsilon^2} = \frac{DX}{\varepsilon^2}$.

Задатак 4. За низ случајних величина (X_n) важи да је $EX_n = n$, а $DX_n = 1$. Ако је c фиксиран број који припада интервалу $(0, 1)$, израчунати вероватноћу да ће бесконачно много пута бити $X_n < cn$ кад $n \in \mathbb{N}$.

Решење. Нека је $A_n = \{X_n < cn\}$. Треба одредити $P(\overline{\lim_{n\to\infty}} A_n)$. Приметимо да

$$\begin{aligned} P(A_n) &= P\left\{X_n < cn\right\} = P\left\{X_n - EX_n < cn - n\right\} = P\left\{X_n - EX_n < (c-1)n\right\} \\ &\leq P\left\{X_n - EX_n < (c-1)n\right\} + P\left\{X_n - EX_n > (1-c)n\right\} \\ &= P\left\{|X_n - EX_n| > (1-c)n\right\} \\ &\leq \frac{DX_n}{(n(1-c))^2} = \frac{1}{(n(1-c))^2} \end{aligned}$$

Дакле, $\sum\limits_{n=1}^{\infty} P(A_n) \leq \sum\limits_{n=1}^{\infty} \frac{1}{(n(1-c))^2} < \infty$. Из Борел-Кантелијеве леме следи да је

$$P(\overline{\lim_{n\to\infty}} A_n) = 0.$$

✓

Конвергенција низова случајних величина

Дефиниција. Низ случајних величина (X_n) конвергира у вероватноћи ка случајној величини X ако важи

$$(\forall \varepsilon > 0)\ \lim_{n\to\infty} P\left\{|X_n - X| > \varepsilon\right\} = 0.$$

За конвергенцију у вероватноћи користи се ознака $X_n \xrightarrow{P} X, n \to \infty$.

Дефиниција. Низ случајних величина (X_n) конвергира скоро сигурно (са верова тноћом 1) ка случајној величини X ако важи

$$P\left\{\lim_{n\to\infty} X_n = X\right\} = 1.$$

За скоро сигурну конвергенцију користи се ознака $X_n \xrightarrow{\text{С.С.}} X, n \to \infty$.

Дефиниција. Низ случајних величина (X_n) конвергира у средње квадратном ка случајној величини X ако важи

$$\lim_{n\to\infty} E\left(|X_n - X|^2\right) = 0.$$

За средње квадратну конвергенцију користи се ознака $X_n \xrightarrow{\text{С.К.}} X, n \to \infty$.

Дефиниција. Низ случајних величина (X_n) конвергира у расподели (или слабо конвергира) ка случајној величини X ако за сваку ограничену и непрекидну функцију $f : \mathbb{R} \to \mathbb{R}$ важи

$$\lim_{n\to\infty} E(f(X_n)) = E(f(X)).$$

За конвергенцију у расподели користи се ознака $X_n \xrightarrow{D} X, n \to \infty$.

Теорема 3. *Ако за* $(\forall \varepsilon > 0)$ $\sum\limits_{n=1}^{\infty} P\{|X_n - X| > \varepsilon\} < \infty$, *тада* $X_n \xrightarrow{C.C.} X, n \to \infty$. ⋆

Теорема 4. *Ако* $(\exists \varepsilon > 0)$ $\sum\limits_{n=1}^{\infty} P\{|X_n - X| > \varepsilon\} = \infty$ *и* (X_n) *је низ независних случајних величина, тада* $X_n \overset{C.C.}{\nrightarrow} X, n \to \infty$. ⋆

Теорема 5. *Нека је* X_n *случајна величина са функцијом расподеле* F_n *и нека је* X *случајна величина са функцијом расподеле* F. *Тада низ* (X_n) *конвергира у расподели ка случајној величини* X *ако и само ако за сваку тачку* x *која је тачка непрекидности функције* F *важи*

$$\lim_{n\to\infty} F_n(x) = F(x).$$

⋆

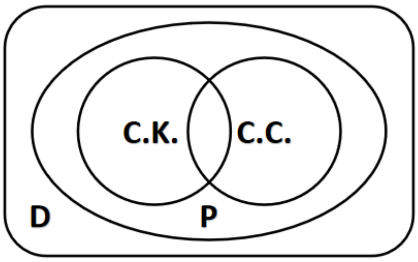
Теорема 6. *(Теорема непрекидности) Нека је* (F_n) *произвољан низ функција расподеле и* (φ_n) *одговарајући низ карактеристичних функција.*

а) Ако $X_n \xrightarrow{D} X, n \to \infty$, *онда за свако* $t \in \mathbb{R}$ *важи* $\varphi(t) = \lim\limits_{n\to\infty} \varphi_n(t)$ *и при томе је* φ *карактеристична функција случајне величине* X.

б) Ако за сваки реалан број t *постоји* $\lim\limits_{n\to\infty} \varphi_n(t)$ *и ако је функција* $\varphi(t) = \lim\limits_{n\to\infty} \varphi_n(t)$ *непрекидна у нули, онда је* φ *карактеристична функција неке случајне величине* X *и важи* $X_n \xrightarrow{D} X, n \to \infty$.

⋆

Односи између разних видова конвергенције



Теорема 7. *Ако* $X_n \xrightarrow{D} c, n \to \infty$, *тада* $X_n \xrightarrow{P} c, n \to \infty$, *где је* c *константа.* ⋆

Задатак 5. Нека је $\Omega = \{\omega_k | k \in N\}$ скуп елементарних исхода неког експеримента и $P\{\omega_k\} = \frac{6}{\pi^2 k^2}$. За сваки природан број n нека је $X_n(\omega_k) = \begin{cases} n, k = n, \\ 0, k \neq n, \end{cases}$, а $Y_n = 1 - \frac{X_n}{n^2}$. Испитати све четири врсте конвергенције низа случајних величина (X_n), односно низа случајних величина (Y_n).

Решење. Из закона расподеле

$$X_n : \begin{pmatrix} 0 & n \\ 1 - \frac{6}{\pi^2 n^2} & \frac{6}{\pi^2 n^2} \end{pmatrix}$$

очекујемо да ће низ случајних величина X_n конвергирати ка 0.

Испитајмо конвергенцију у вероватноћи и расподели низа (X_n):

$$(\forall \varepsilon > 0)\ P\{|X_n - 0| > \varepsilon\} = P\{X_n > \varepsilon\} \leq P\{X_n = n\} = \frac{6}{\pi^2 n^2} \to 0,\ n \to \infty$$
$$\implies X_n \xrightarrow{P} 0, n \to \infty$$
$$\implies X_n \xrightarrow{D} 0, n \to \infty$$

Испитајмо скоро сигурну конвергенцију низа (X_n):

$$(\forall \varepsilon > 0)\ \sum_{n=1}^{\infty} P\{|X_n - 0| > \varepsilon\} \leq \sum_{n=1}^{\infty} P\{X_n = n\} = \sum_{n=1}^{\infty} \frac{6}{\pi^2 n^2} < \infty$$
$$\implies X_n \xrightarrow{\text{С.С.}} 0, n \to \infty$$

Из једнакости $E|X_n - 0|^2 = \frac{6}{\pi^2} \nrightarrow 0, n \to \infty$, видимо да $X_n \overset{\text{С.К.}}{\nrightarrow} 0, n \to \infty$.

Испитајмо сада конвергенцију низа (Y_n). Како је $Y_n = 1 - \frac{X_n}{n^2}$, очекујемо да ће низ Y_n конвергирати ка 1. Приметимо да

$$\begin{aligned} X_n \xrightarrow{\text{С.С.}} 0, n \to \infty &\iff P\{\lim_{n\to\infty} X_n = 0\} = 1 \\ &\implies P\left\{\lim_{n\to\infty} \frac{X_n}{n^2} = 0\right\} = 1 \\ &\iff P\left\{\lim_{n\to\infty} \left(1 - \frac{X_n}{n^2}\right) = 1\right\} = 1 \\ &\iff Y_n \xrightarrow{\text{С.С.}} 1, n \to \infty \end{aligned}$$

Из претходног следи $Y_n \xrightarrow{P} 1, n \to \infty$ и $Y_n \xrightarrow{D} 1, n \to \infty$.

Остаје још да се испита конвергенција у средње квадратном низа (Y_n). Како је $E|Y_n - 1|^2 = E\left|\frac{X_n^2}{n^4}\right| = \frac{6}{\pi^2 n^4} \to 0,\ n \to \infty$, важи и $Y_n \xrightarrow{\text{С.К.}} 1, n \to \infty$. ✓

Задатак 6. За сваки природан број n случајна величина X_n има униформну $U[0, \frac{1}{n}]$ расподелу, а независно од ње случајна величина Y_n има закон расподеле $\begin{pmatrix} 0 & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. Ако је $Z_n = X_n + Y_n$, испитати све четири врсте конвергенције низа случајних величина (Z_n).

Решење. На основу датих расподела случајних величина X_n и Y_n видимо да можемо очекивати да ће оба низа (X_n) и (Y_n) конвергирати ка 0, као и низ (Z_n).

Средње квадратну конвергенцију испитаћемо директним рачунањем:

$$E|Z_n|^2 = EX_n^2 + 2EX_nEY_n + EY_n^2 = \frac{4}{3n^2} \to 0,\ n \to \infty \implies Z_n \xrightarrow{\text{С.К.}} 0, n \to \infty$$

Испитајмо скоро сигурну конвергенцију помоћу теореме 3:

$$(\forall \varepsilon > 0)\ P\{|Z_n - 0| > \varepsilon\} \leq \frac{E|Z_n|^2}{\varepsilon^2} = \frac{4}{3\varepsilon^2 n^2}.$$

Како је $\sum_{n=1}^{\infty} P\{|Z_n - 0| > \varepsilon\} \leq \sum_{n=1}^{\infty} \frac{4}{3\varepsilon^2 n^2} < \infty$, закључујемо да $Z_n \xrightarrow{\text{С.С.}} 0, n \to \infty$, као и $Z_n \xrightarrow{P} 0, n \to \infty$, и $Z_n \xrightarrow{D} 0, n \to \infty$. ✓

Задатак 7. Дат је низ независних случајних величина чији општи члан X_n има униформну $U[0,n]$ расподелу. Ако је $Y_n = \min\{1, X_n\}$, испитати све четири врсте конвергенције низа случајних величина (Y_n).

Решење. Како није очигледно шта ће бити гранична случајна величина низа (Y_n), прво ћемо одредити функцију расподеле случајне величине Y_n, затим применити теорему 5:

$$\begin{aligned} F_{Y_n}(y) &= P\{Y_n \le y\} = P\{\min\{1, X_n\} \le y\} = 1 - P\{\min\{1, X_n\} > y\} \\ &= 1 - P\{1 > y, X_n > y\} = 1 - P\{X_n > y\} = P\{X_n \le y\} = \frac{y}{n},\ 0 \le y < 1 \end{aligned}$$

$$\implies F_{Y_n}(y) = \begin{cases} 0, & y < 0 \\ \frac{y}{n}, & 0 \le y < 1 \\ 1, & y \ge 1 \end{cases} \to \begin{cases} 0, & y < 1 \\ 1, & y \ge 1 \end{cases} = F(y)$$

Функција $F(x)$ је функција расподеле (дегенерисане) случајне величине 1. Из претходног закључујемо да $Y_n \xrightarrow{D} 1, n \to \infty$, као и $Y_n \xrightarrow{P} 1, n \to \infty$.

Испитајмо скоро сигурну конвергенцију. Како је за $\varepsilon > 0$

$$P\{|Y_n - 1| \ge \varepsilon\} = P\{Y_n \le 1 - \varepsilon\} = \begin{cases} 0, & \varepsilon > 1, \\ \frac{1-\varepsilon}{n}, & \varepsilon \in (0,1], \end{cases}$$

видимо да постоји бар једно ε, на пример $\varepsilon = \frac{1}{2}$, за које је $\sum\limits_{n=1}^{\infty} P\{|Y_n - 1| \ge \varepsilon\} = \infty$ и (Y_n) је низ независних случајних величина (јер је (X_n) низ независних случајних величина), применом теореме 4 имамо да $Y_n \overset{\text{С.С.}}{\nrightarrow} 1, n \to \infty$.

Директним рачунањем испитујемо средње квадратну конвергенцију:

$$\begin{aligned} E|Y_n - 1|^2 &= E(\min\{1, X_n\})^2 - 2E(\min\{1, X_n\}) + 1 = \frac{1}{3n} \to 0,\ n \to \infty \\ &\implies Y_n \xrightarrow{\text{С.К.}} 1, n \to \infty. \end{aligned}$$

✓

Теорема 8. *(Коши-Шварцова неједнакост) За случајне величине X и Y такве да је $EX^2 < \infty$ и $EY^2 < \infty$ важи*

$$|EXY| \le \sqrt{EX^2 EY^2}.$$

⋆

Задатак 8. Ако низ случајних величина (X_n) конвергира у средње квадратном смислу ка случајној величини X, онда $EX_n \to EX$ и $DX_n \to DX$ кад $n \to \infty$. Доказати.

Решење. Како низ случајних величина (X_n) конвергира у средње квадратном смислу ка случајној величини X, важи једнакост

$$\lim_{n\to\infty} E|X_n - X|^2 = 0.$$

Како је

$$0 \le |EX_n - EX| = |E(X_n - X)| \le E|X_n - X| = \sqrt{(E|X_n - X|)^2} \le \sqrt{E(X_n - X)^2} \to 0,$$

применом теореме о два полицајца следи да $EX_n \to EX, n \to \infty$.

Даље је

$$EX_n^2 = E(X_n - X + X)^2 = E(X_n - X)^2 + 2E(X(X_n - X)) + EX^2.$$

Видимо да први сабирак тежи 0, јер низ случајних величина (X_n) конвергира у средње квадратном смислу ка случајној величини X. Посматрајмо други сабирак:

$$\begin{aligned}0 \le |E(X(X_n - X))| &\le E|X(X_n - X)| = \sqrt{(E|X(X_n - X)|)^2}\\ &\le \sqrt{E(X^2(X_n - X)^2)} \le \sqrt{EX^2 E(X_n - X)^2} \to 0.\end{aligned}$$

Последња неједнакост је добијена применом Коши-Шварц неједнакости. Како је $EX_n \to EX$ и $EX_n^2 \to EX^2$, добијамо да $DX_n \to DX, n \to \infty$. ✓

Последица. *Ако* $X_n \xrightarrow{C.K.} c, n \to \infty$, *онда* $EX_n \to c,\ DX_n \to 0,\ n \to \infty$ ⋆

Задатак 9. Случајна величина Y има униформну $\mathcal{U}[0,1]$ расподелу. За сваки природан број n нека је $X_n = a_n I\{Y < \frac{1}{n}\} + YI\{Y \ge \frac{1}{n}\}$. Одредитити низ бројева (a_n) тако да за свако $n \in \mathbb{N}$ буде $EX_n = 0$ и за тако изабране a_n испитати све четири врсте конвергенције низа случајних величина (X_n).

Решење. Прво одредимо низ бројева a_n такав да важи $EX_n = 0$:

$$\begin{aligned}EX_n &= a_n E\left(I\left\{Y < \frac{1}{n}\right\}\right) + E\left(YI\left\{Y \ge \frac{1}{n}\right\}\right)\\ &= a_n P\left\{Y < \frac{1}{n}\right\} + \int_{\frac{1}{n}}^{1} y dy = a_n \frac{1}{n} + \frac{1}{2} - \frac{1}{2n^2} = 0\\ &\implies a_n = \frac{1}{2n} - \frac{n}{2}\end{aligned}$$

Нађимо функцију расподеле случајне величине X_n, $F_{X_n}(x) = P\{X_n \le x\}$. Потребно је да разматрамо следеће случајеве

1. за $y < 0 \implies x < \frac{1}{2n} - \frac{n}{2}$: $F_{X_n}(x) = 0$
2. за $0 \le y < \frac{1}{n} \implies \frac{1}{2n} - \frac{n}{2} \le x < \frac{1}{n}$: $F_{X_n}(x) = \frac{1}{n}$
3. за $\frac{1}{n} \le y < 1 \implies \frac{1}{n} \le x < 1$: $F_{X_n}(x) = \frac{1}{n} + P\{\frac{1}{n} \le X_n \le x\}$
$= \frac{1}{n} + P\{\frac{1}{n} \le Y \le x\} = \frac{1}{n} + x - \frac{1}{n} = x$
4. за $y \ge 1 \implies x \ge 1$: $F_{X_n}(x) = 1$

Дакле, функција расподеле случајне величине X_n једнака је

$$F_{X_n}(x) = \begin{cases} 0,\ x < \frac{1}{2n} - \frac{n}{2} \\ \frac{1}{n},\ \frac{1}{2n} - \frac{n}{2} \le x < \frac{1}{n} \\ x,\ \frac{1}{n} \le x < 1 \\ 1,\ x \ge 1 \end{cases} \xrightarrow{n \to \infty} \begin{cases} 0,\ x < 0 \\ x,\ 0 \le x < 1 \\ 1,\ x \ge 1 \end{cases} = F_Y(x),$$

па закључујемо да $X_n \xrightarrow{D} Y, n \to \infty$, где $Y \in \mathcal{U}(0,1)$.

Испитајмо конвергенцију у вероватноћи:

$$(\forall \varepsilon > 0)\ P\{|X_n - Y| > \varepsilon\} = P\{X_n \neq Y\} \le P\left\{Y < \frac{1}{n}\right\} = \frac{1}{n} \to 0, n \to \infty$$

$$\implies X_n \xrightarrow{P} Y, n \to \infty$$

Испитајмо скоро сигурну конвергенцију, односно испитајмо да ли важи

$$P\left\{\omega | \lim_{n\to\infty} X_n(\omega) = Y(\omega)\right\} = 1.$$

Нека је ω_0 такво да $P\{\omega|0 < Y(\omega) < 1\} = 1$ и постоји n_0 такво да је $Y(\omega_0) \geq \frac{1}{n_0} > \frac{1}{n}$ за $\forall n > n_0$. За $n \geq n_0$ имамо да је $X_n = \left(\frac{1}{2n} - \frac{n}{2}\right) I\left\{Y < \frac{1}{n}\right\}\right) + YI\left\{Y \geq \frac{1}{n}\right\} = Y$, односно почевши од неког n_0 низ је константан, тј. $X_n(\omega_0) \to Y(\omega_0),\ n \to \infty$. Тада

$$P\{\omega | \lim_{n\to\infty} X_n(\omega) = Y(\omega)\} \geq P\{\omega|0 < Y(\omega) < 1\} = 1$$
$$\implies X_n \xrightarrow{\text{С.С.}} Y, n \to \infty.$$

За испитивање средње квадратне конвергенције искористићемо претходни задатак. Приметимо да је $EY = \frac{1}{2}$ и да је $EX_n = 0$ (из услова задатка). Очигледно $EX_n \not\to EY,\ n \to \infty$, па закључујемо да $X_n \overset{\text{С.К.}}{\not\to} Y, n \to \infty$. ✓

Задатак 10**.** У зависности од вредности реалног параметра α $(\alpha > 0, \alpha \neq 1)$, испитати све четири врсте конвергенције низа независних случајних величина чији општи члан X_n има униформну $\mathcal{U}\left([0, \frac{1}{n^\alpha}] \cup [1, 1 + \frac{1}{n}]\right)$ расподелу.

Решење. Густина је дата са

$$f_n(x) = \begin{cases} \frac{n^\alpha}{1+n^{\alpha-1}},\ x \in [0, \frac{1}{n^\alpha}] \cup [1, 1 + \frac{1}{n}], \\ 0, \ \text{иначе.} \end{cases}$$

Рачунамо функцију расподеле случајне величине X_n, како бисмо испитали конвергенцију у расподели.

$$F_n(x) = P\left\{X_n \leq x\right\} = (*)$$

1. за $x < 0:\ (*) = 0;$
2. за $0 \leq x < \frac{1}{n^\alpha}:\ (*) = \int\limits_0^x \frac{n^\alpha}{1+n^{\alpha-1}} dt = \frac{n^\alpha}{1+n^{\alpha-1}} x;$
3. за $\frac{1}{n^\alpha} \leq x < 1:\ (*) = P\left\{X_n \in [0, \frac{1}{n^\alpha}]\right\} = \int\limits_0^{\frac{1}{n^\alpha}} \frac{n^\alpha}{1+n^{\alpha-1}} dt = \frac{1}{1+n^{\alpha-1}};$
4. за $1 \leq x < 1 + \frac{1}{n}:\ (*) = \int\limits_0^{\frac{1}{n^\alpha}} \frac{n^\alpha}{1+n^{\alpha-1}} dt + \int\limits_1^x \frac{n^\alpha}{1+n^{\alpha-1}} dt = \frac{1}{1+n^{\alpha-1}} + \frac{n^\alpha}{1+n^{\alpha-1}}(x-1);$
5. за $x \geq 1 + \frac{1}{n}:\ (*) = 1.$

Дакле, функција расподеле случајне величине X_n је

$$F_n(x) = \begin{cases} 0,\ x < 0, \\ \frac{n^\alpha}{1+n^{\alpha-1}} x,\ 0 \leq x < \frac{1}{n^\alpha}, \\ \frac{1}{1+n^{\alpha-1}},\ \frac{1}{n^\alpha} \leq x < 1, \\ \frac{1}{1+n^{\alpha-1}} + \frac{n^\alpha}{1+n^{\alpha-1}}(x-1),\ 1 \leq x < 1 + \frac{1}{n}, \\ 1,\ x \geq 1 + \frac{1}{n}. \end{cases}$$

Ако је $0 < \alpha < 1$:

$$F_n(x) \to \begin{cases} 0, \ x < 0, \\ 1, \ x \geq 0. \end{cases}$$

Према томе $X_n \xrightarrow{D} 0$, па и $X_n \xrightarrow{P} 0$.

Ако је $\alpha > 1$:

$$F_n(x) \to \begin{cases} 0, \ x < 1, \\ 1, \ x \geq 1. \end{cases}$$

Према томе $X_n \xrightarrow{D} 1$, па и $X_n \xrightarrow{P} 1$.

Сада испитујемо средње квадратну конвергенцију.

Ако је $0 < \alpha < 1$:

$$\begin{aligned} E|X_n - 0|^2 = EX_n^2 &= \int_0^{\frac{1}{n^\alpha}} x^2 \frac{n^\alpha}{1+n^{\alpha-1}} dx + \int_1^{1+\frac{1}{n}} x^2 \frac{n^\alpha}{1+n^{\alpha-1}} dx \\ &= \frac{n^\alpha}{1+n^{\alpha-1}} \left(\frac{1}{n^{3\alpha}} \frac{1}{3} + \left(\left(1 + \frac{1}{n}\right)^3 - 1 \right) \frac{1}{3} \right) \\ &= \frac{1}{3n^{2\alpha}(1+n^{\alpha-1})} + \frac{n^{\alpha-1}}{1+n^{\alpha-1}} + \frac{n^{\alpha-2}}{1+n^{\alpha-1}} + \frac{n^{\alpha-3}}{3(1+n^{\alpha-1})} \to 0. \end{aligned}$$

Према томе $X_n \xrightarrow{\text{С.К.}} 0$.

Ако је $\alpha > 1$:

$$\begin{aligned} E|X_n - 1|^2 &= E(X_n^2 - 2X_n + 1) = EX_n^2 - 2EX_n + 1 \\ &= \frac{n^\alpha}{1+n^{\alpha-1}} \left(\frac{1}{3n^{3\alpha}} + \frac{1}{n} + \frac{1}{n^2} + \frac{1}{3n^3} - 2\int_0^{\frac{1}{n^\alpha}} x dx - 2\int_1^{1+\frac{1}{n}} x dx \right) + 1 \\ &= \frac{1}{3n^{2\alpha}(1+n^{\alpha-1})} - \frac{1}{n^\alpha(1+n^{\alpha-1})} + \frac{n^{\alpha-1}}{1+n^{\alpha-1}} \left(\frac{1}{3n^2} - 1 \right) + 1 \to 0. \end{aligned}$$

Према томе $X_n \xrightarrow{\text{С.К.}} 1$.

Остало је још да испитамо скоро сигурну конвергенцију.

Ако је $0 < \alpha < 1$:

Испитујемо да ли за свако $\varepsilon > 0$ важи да је $\sum_{n=1}^{\infty} P\{|X_n - 0| \geq \varepsilon\} < \infty$.

Нека је $\varepsilon > 0$ такво да је, за свако n почев од неког n_0 $\varepsilon > \frac{1}{n^\alpha}$. Тада је

$$\begin{aligned} P\{|X_n - 0| \geq \varepsilon\} = P\{X_n \geq \varepsilon\} &= P\left\{X_n \in \left[1, 1 + \frac{1}{n}\right]\right\} \\ &= \int_1^{1+\frac{1}{n}} \frac{n^\alpha}{1+n^{\alpha-1}} dx = \frac{n^{\alpha-1}}{1+n^{\alpha-1}} \to 1 \neq 0. \end{aligned}$$

Према томе, постоји ε тако да $P\{|X_n - 0| \geq \varepsilon\} \not\to 0$, па $\sum_{n=1}^{\infty} P\{|X_n - 0| \geq \varepsilon\}$ не конвергира. Како су случајне величине независне, закључујемо да $X_n \overset{\text{С.С.}}{\not\to} 0$.

Ако је $\alpha > 1$: Испитујемо да ли за свако $\varepsilon > 0$ важи да је $\sum\limits_{n=1}^{\infty} P\left\{|X_n - 1| \geq \varepsilon\right\} < \infty$.

Приметимо да

$$P\left\{|X_n - 1| \geq \varepsilon\right\} = P\left\{X_n \geq 1 + \varepsilon\right\} + P\left\{X_n \leq 1 - \varepsilon\right\} = (**)$$

а) Ако је $0 < \varepsilon < 1$:

За велико n такво да је $\frac{1}{n} < \varepsilon$ и $\frac{1}{n^\alpha} < 1 - \varepsilon$

$$(**) = P\left\{0 \leq X_n \leq \frac{1}{n^\alpha}\right\} = \frac{1}{1 + n^{\alpha - 1}}.$$

б) Ако је $\varepsilon \geq 1$:

$$(**) = 0.$$

Да би $X_n \stackrel{\text{С.С.}}{\to} 1$, мора бити $\sum\limits_{n=1}^{\infty} P\left\{|X_n - 1| \geq \varepsilon\right\} < \infty$ за свако $\varepsilon > 0$. Ово важи за свако $\varepsilon \geq 1$, па остаје још да покажемо да ли важи и за свако $\varepsilon \in (0, 1)$. Како је за $0 < \varepsilon < 1$, $\sum\limits_{n=1}^{\infty} P\left\{|X_n - 1| \geq \varepsilon\right\} \approx \sum\limits_{n=1}^{\infty} \frac{1}{1+n^{\alpha-1}}$, морамо да раздвојимо случајеве $1 < \alpha \leq 2$ и $\alpha > 2$.

1) Ако је $1 < \alpha \leq 2$:
   $\sum\limits_{n=1}^{\infty} \frac{1}{1+n^{\alpha-1}} = \infty$, па закључујемо да $X_n \stackrel{\text{С.С.}}{\not\to} 1$.

2) Ако је $\alpha > 2$:
   $\sum\limits_{n=1}^{\infty} \frac{1}{1+n^{\alpha-1}} < \infty$, па је и $\sum\limits_{n=1}^{\infty} P\left\{|X_n - 1| \geq \varepsilon\right\} < \infty$ за свако $\varepsilon \in (0, 1)$. За $\varepsilon \geq 1$ ова сума такође конвергира (једнака је 0), па закључујемо да $X_n \stackrel{\text{С.С.}}{\to} 1$.

✓

Задатак 11. Дат је низ независних случајних величина чији општи члан X_n има $\begin{pmatrix} \frac{-1}{2^n} & \frac{1}{2^n} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ расподелу и нека је $S_n = \sum\limits_{k=1}^{n} X_k$. Доказати да низ случајних величина (S_n) конвергира у расподели ка случајној величини S_∞, где S_∞ има униформну $U[-1, 1]$ расподелу.

Решење. Показаћемо да $\varphi_{S_n}(t) \to \varphi_{S_\infty}(t),\ n \to \infty,\ \forall t \in \mathbb{R}$, где су $\varphi_{S_n}(t)$ и $\varphi_{S_\infty}(t)$ карактеристичне функције одговарајућих случајних величина. Ако покажемо додатно да је $\varphi_{S_\infty}(t)$ непрекидна у 0, на основу теореме непрекидности ће следити тражени резултат. Приметимо да је

$$\varphi_{S_\infty}(t) = E(e^{itS_\infty}) = \int_{-1}^{1} e^{itx} \frac{1}{2} dx = \begin{cases} 1,\ t = 0, \\ \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t},\ t \neq 0. \end{cases}$$

Видимо да је $\varphi_{S_\infty}(t)$ непрекидна у 0. Даље,

$$\varphi_{S_n}(t) = \prod_{k=1}^{n} \varphi_{X_k}(t) = \prod_{k=1}^{n} E(e^{itX_k}) = \prod_{k=1}^{n} \left(\frac{1}{2} e^{-it\frac{1}{2^k}} + \frac{1}{2} e^{it\frac{1}{2^k}}\right) = \prod_{k=1}^{n} \cos\left(\frac{t}{2^k}\right).$$

Према томе,

$$\begin{aligned}
\lim_{n\to\infty}\varphi_{S_n}(t) &= \lim_{n\to\infty}\cos\left(\frac{t}{2}\right)\cos\left(\frac{t}{2^2}\right)\cdots\cos\left(\frac{t}{2^n}\right)\\
&= \begin{cases} 1,\ t=0\\ \lim\limits_{n\to\infty}\frac{1}{\sin\left(\frac{t}{2^n}\right)}\cos\left(\frac{t}{2}\right)\cos\left(\frac{t}{2^2}\right)\cdots\sin\left(\frac{t}{2^n}\right)\cos\left(\frac{t}{2^n}\right),\ t\neq 0\end{cases}\\
&= \begin{cases} 1,\ t=0\\ \lim\limits_{n\to\infty}\frac{1}{\sin\left(\frac{t}{2^n}\right)}\frac{1}{2^n}\sin t=\frac{\sin t}{t},\ t\neq 0.\end{cases}
\end{aligned}$$

Дакле, $S_n \overset{D}{\to} S_\infty$, кад $n\to\infty$. ✓

Закон великих бројева

Нека је (X_n) низ случајних величина дефинисаних на истом простору вероватноћа. Нека је $S_n=\sum\limits_{k=1}^{n}X_k$. Ако:

- $\frac{S_n-ES_n}{n}\overset{P}{\to}0,\ n\to\infty$, онда важи слаби закон великих бројева за низ (X_n);
- $\frac{S_n-ES_n}{n}\overset{\text{С.С}}{\to}0,\ n\to\infty$, онда важи јаки закон великих бројева за низ (X_n).

Теорема 9. ***(Колмогоровљева)*** *Ако је (X_n) низ независних случајних величина и ако је $\sum\limits_{n=1}^{\infty}\frac{DX_n}{n^2}<\infty$, онда важи јаки закон великих бројева за низ (X_n).* ⋆

Теорема 10. ***(Колмогоровљева)*** *Ако је (X_n) низ независних случајних величина са истом расподелом и ако је $EX_1<\infty$, онда важи јаки закон великих бројева за низ (X_n).* ⋆

Задатак 12. Испитати да ли закон великих бројева важи за низ независних случајних величина чији општи члан X_n има:

а) закон расподеле $\begin{pmatrix} -\sqrt{\ln n} & \sqrt{\ln n}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$;

б) густину расподеле $f_{X_n}(x)=ne^{-nx}, x\geq 0$;

в) закон расподеле $\begin{pmatrix} -2^n & 2^n\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$.

Решење. а) Како је $EX_n=0$ и $DX_n=EX_n^2=\ln n$,

$$\sum_{n=1}^{\infty}\frac{DX_n}{n^2}=\sum_{n=1}^{\infty}\frac{\ln n}{n^2}<\infty,$$

па важи јаки закон великих бројева, а самим тим и слаби закон великих бројева.

б) Како је

$$EX_n=\int\limits_0^{\infty}xne^{-nx}dx=\frac{1}{n}$$

и

$$DX_n = EX_n^2 - (EX_n)^2 = \int\limits_0^\infty x^2 n e^{-nx} dx - \frac{1}{n^2} = \frac{1}{n^2},$$

важи да је

$$\sum_{n=1}^{\infty} \frac{DX_n}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^4} < \infty,$$

па важи јаки закон великих бројева, а самим тим и слаби закон великих бројева.

в) Показаћемо да не важи слаби закон великих бројева (па самим тим ни јаки закон великих бројева), тако што ћемо показати да постоји $\varepsilon > 0$ такво да $P\left\{|\frac{S_n - ES_n}{n}| > \varepsilon\right\} \not\to 0,\ n \to \infty$. Нека је $\varepsilon = \frac{1}{2}$. Тада

$$\begin{aligned} P\left\{\left|\frac{S_n - ES_n}{n}\right| > \frac{1}{2}\right\} &= P\left\{\left|\frac{S_n}{n}\right| > \frac{1}{2}\right\} \geq P\left\{\frac{S_n}{n} > \frac{1}{2}\right\} \\ &= P\left\{\sum_{k=1}^{n} X_k > \frac{n}{2}\right\} \geq P\left\{\sum_{k=1}^{n} X_k > 2^n\right\} \\ &\geq P\left\{X_{n-1} = 2^{n-1}, X_n = 2^n\right\} = \frac{1}{2}\frac{1}{2} \not\to 0,\ n \to \infty. \end{aligned}$$

Дакле, не важе ни слаби ни јаки закон великих бројева.

✓

Задатак 13. Дат је низ независних случајних величина чији општи члан X_n има густину расподеле $f_{X_n}(x) = \frac{x^n e^{-x}}{n!}, x > 0$. Испитати да ли за овај низ важи закон великих бројева.

Решење. Желимо да испитамо да ли $\frac{S_n - ES_n}{n} \xrightarrow{P} 0,\ n \to \infty$. Како је

$$EX_n = \int\limits_0^\infty x \frac{x^n e^{-x}}{n!} dx = \frac{1}{n!} \int\limits_0^\infty x^{n+1} e^{-x} dx = \frac{\Gamma(n+2)}{n!} = n+1,$$

добијамо да је

$$ES_n = \sum_{k=1}^{n} EX_k = \sum_{k=1}^{n} (k+1) = \frac{n(n+3)}{2}.$$

Пошто конвергенција у вероватноћи повлачи конвергенцију у расподели, ако покажемо да $U_n = \frac{S_n}{n} - \frac{n+3}{2} \overset{D}{\not\to} 0,\ n \to \infty$, онда $\frac{S_n}{n} - \frac{n+3}{2} \overset{P}{\not\to} 0\ n \to \infty$, па не важи закон великих бројева. Показаћемо да $\varphi_{U_n}(t) \not\to \varphi_0(t) = 1$, бар за једно $t \in \mathbb{R}$. Пошто је

$$\varphi_{X_n}(t) = E(e^{itX_n}) = \int\limits_0^\infty e^{itx} \frac{x^n e^{-x}}{n!} dx = \frac{1}{n!} \int\limits_0^\infty x^n e^{-x(1-it)} dx = \frac{1}{(1-it)^{n+1}}$$

$$\Longrightarrow\ \varphi_{\frac{S_n}{n}}(t) = \prod_{k=1}^{n} \varphi_{X_k}\left(\frac{t}{n}\right) = \left(\frac{1}{1 - \frac{it}{n}}\right)^{\sum\limits_{k=1}^{n} k + n} = \left(\frac{1}{1 - \frac{it}{n}}\right)^{\frac{n(n+3)}{2}}$$

$$\Longrightarrow\ \varphi_{U_n}(t) = e^{-\frac{n+3}{2}it}\varphi_{\frac{S_n}{n}}(t) = e^{-\frac{n+3}{2}it}\left(\frac{1}{1-\frac{it}{n}}\right)^{\frac{n(n+3)}{2}} = e^{-\frac{n+3}{2}it}e^{-\frac{n(n+3)}{2}\ln\left(1-\frac{it}{n}\right)}$$
$$= e^{-\frac{n+3}{2}it}e^{-\frac{n(n+3)}{2}\left(-\frac{it}{n}+\frac{t^2}{2n^2}+o(\frac{1}{n^2})\right)} \to e^{-\frac{t^2}{4}},\ n\to\infty.$$

Како је $e^{-\frac{t^2}{4}}$ вредност карактеристичне функције у тачки t случајне величине која има нормалну $\mathcal{N}(0,\frac{1}{2})$ расподелу, закључујемо да $U_n \overset{D}{\not\to} 0$ $n\to\infty$, па ни $U_n \overset{P}{\not\to} 0$ $n\to\infty$, те не важи слаби закон великих бројева. ✓

Задатак 14. Нека је низ случајних величина (X_n) такав да за сваки природан број n важи да је $EX_n = 0$ и $DX_n \le C$ где је C константа која је већа од 0, и било који члан X_n зависи само од претходног X_{n-1} и следећег X_{n+1}, а независан је од осталих чланова низа. Доказати да за овај низ важи слаби закон великих бројева.

Решење. Пошто је $ES_n = \sum\limits_{i=1}^{n} EX_i = 0$, желимо да испитамо да ли $\frac{S_n}{n} \overset{P}{\to} 0,\ n\to\infty$. Нека је $\varepsilon > 0$ произвољно. Тада

$$P\left\{\left|\frac{S_n}{n}-0\right|\ge\varepsilon\right\} = P\left\{|S_n|\ge n\varepsilon\right\} \le \frac{ES_n^2}{n^2\varepsilon^2} = \frac{E\left(\sum\limits_{i=1}^{n}X_i\right)^2}{n^2\varepsilon^2}$$
$$= \frac{E\left(\sum\limits_{i=1}^{n}X_i^2+\sum\limits_{i\ne j}X_iX_j\right)}{n^2\varepsilon^2} = \frac{\sum\limits_{i=1}^{n}EX_i^2+\sum\limits_{i\ne j}EX_iX_j}{n^2\varepsilon^2}$$
$$= \frac{\sum\limits_{i=1}^{n}DX_i+\sum\limits_{|i-j|=1}EX_iX_j+\sum\limits_{|i-j|>1}EX_iX_j}{n^2\varepsilon^2}$$
$$\le \frac{nC+\sum\limits_{|i-j|=1}EX_iX_j+\sum\limits_{|i-j|>1}EX_iEX_j}{n^2\varepsilon^2}$$
$$\le \frac{nC+2(n-1)C}{n^2\varepsilon^2} \to 0,\ n\to\infty,$$

при чему је коришћено да је $EX_iX_j \le E|X_iX_j| \le \sqrt{EX_i^2EX_j^2}$.

Према томе, $\frac{S_n}{n}\overset{P}{\to} 0$ $n\to\infty$, па важи слаби закон великих бројева. ✓

Централна гранична теорема

Теорема 11. *Нека је (X_n) низ независних случајних величина са истом расподелом, математичким очекивањем $EX_1 = m$ и коначном дисперзијом $DX_1 = \sigma^2 > 0$. Ако је $S_n = \sum\limits_{k=1}^{n} X_k$, онда за свако $x\in\mathbb{R}$ при $n\to\infty$ важи*

$$P\left\{\frac{S_n-nm}{\sigma\sqrt{n}}\le x\right\} \to \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}dt,$$

односно $\frac{S_n-ES_n}{\sqrt{DS_n}} \overset{D}{\to} X^\star \in \mathcal{N}(0,1)$. ⋆

Задатак 15. Рачунар у процесу сабирања бројева врши заокруживање на најближи цео број. Претпоставља се да су грешке настале заокруживањем независне и униформно расподељене на сегменту $[-0.5, 0.5]$.

а) Ако се сабира 1500 бројева, израчунати вероватноћу да апсолутна вредност укупне грешке буде већа од 15.

б) Колико се највише бројева може сабрати, па да са вероватноћом 0.9 вредност укупне грешке буде мања од 10?

Решење. а) Нека је $X_k,\ k = 1, ..., n$, грешка при $k-$том заокруживању. Треба одредити

$$P\left\{\left|\sum_{k=1}^{n} X_k\right| > 15\right\}.$$

Из дефиниције случајних величина $X_k,\ k = 1, ..., n$ можемо закључити да $X_k \in \mathcal{U}[-0.5, 0.5]$. Према томе $EX_k = 0$ и $DX_k = \frac{1}{12},\ k = 1, ..., n$. Како је $n = 1500$ довољно велико (значајно веће од 30), може се применити централна гранична теорема, односно можемо закључити да

$$\frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} \xrightarrow{D} X^\star \in \mathcal{N}(0,1).$$

Према томе

$$\begin{aligned}
P\left\{\left|\sum_{k=1}^{n} X_k\right| > 15\right\} &= 1 - P\left\{\left|\sum_{k=1}^{n} X_k\right| \leq 15\right\} \\
&= 1 - P\left\{\left|\frac{\sum\limits_{k=1}^{n} X_k}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}}\right| \leq \frac{15}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}}\right\} \\
&= 1 - P\left\{\left|\frac{\sum\limits_{k=1}^{n} X_k}{\sqrt{\frac{1500}{12}}}\right| \leq \frac{15}{\sqrt{\frac{1500}{12}}}\right\} \\
&\approx 1 - P\left\{|X^\star| \leq \frac{15}{\sqrt{125}}\right\} \\
&= 1 - P\left\{-\frac{15}{\sqrt{125}} \leq X^\star \leq \frac{15}{\sqrt{125}}\right\} \\
&= 1 - \Phi(1.34) + \Phi(-1.34) \\
&= 2 - 2\Phi(1.34) = 0.18.
\end{aligned}$$

б) Треба одредити n, тако да важи

$$P\left\{\sum_{k=1}^{n} X_k < 10\right\} = 0.9.$$

Очекујемо да је n довољно велико, па важи

$$P\left\{\frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} < \frac{10 - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}}\right\} = 0.9$$

$$P\left\{\frac{\sum\limits_{k=1}^{n} X_k}{\sqrt{\frac{n}{12}}} < \frac{10}{\sqrt{\frac{n}{12}}}\right\} = 0.9$$

$$P\left\{X^\star < \sqrt{\frac{1200}{n}}\right\} = 0.9$$

$$\Phi\left(\sqrt{\frac{1200}{n}}\right) = 0.9 \implies \sqrt{\frac{1200}{n}} = \Phi^{-1}(0.9) = 1.28.$$

Дакле, $n = 732.42$. Међутим, број бројева које треба сабрати мора бити целобројна вредност, а пошто је питање колико највише бројева треба сабрати, решење је 732.

✓

Задатак 16**.** Кишне капи облика сфере, чија дужина полупречника (у mm) има униформну $\mathcal{U}[1,2]$ расподелу, падају у кофу. Сваког секунда упадне 20 кишних капи. Ако је вероватноћа да се након двосатног падања кише кофа неће препунити водом једнака 0.965, израчунати запремину (у литрима) те кофе.

Решење. Нека је $X_k,\ k = 1, ..., n$, запремина кишне капи и $R_k,\ k = 1, ..., n$, њен полупречник, $R_k \in \mathcal{U}[1,2]$. Тада је $X_k = \frac{4}{3}\pi R_k^3$. Треба одредити v тако да важи

$$P\left\{\sum_{k=1}^{n} X_k < v\right\} = 0.965.$$

Пошко сваке секунде упадне 20 капи, током 2 сата ће упасти $n = 20 \cdot 7200 = 144000$ капи. Како је n довољно велико, можемо применити централну граничну теорему, односно

$$\frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} \xrightarrow{D} X^\star \in \mathcal{N}(0,1).$$

Према томе

$$P\left\{\frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} < \frac{v - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}}\right\} = 0.965$$

$$\implies \frac{v - nEX_1}{\sqrt{nDX_1}} = \Phi^{-1}(0.965) = 1.81.$$

Како је

$$EX_1 = \frac{4}{3}\pi E R_1^3 = \frac{4}{3}\pi \int\limits_1^2 x^3 dx = 5\pi,$$

$$DX_1 = EX_1^2 - (EX_1)^2 = \frac{16}{9}\pi^2 \int\limits_1^2 x^6 dx - 25\pi^2 = \frac{457}{63}\pi^2,$$

то следи да је

$$v = 1.81\sqrt{144000\frac{457}{63}\pi + 144000 \cdot 5\pi} = 2267758.33472,$$

односно запремина кофе је приближно 2.27 литра. ✓

Задатак 17. Нека је (X_n) низ независних случајних величина које имају исту расподелу и нека је $EX_n = 0$, а $DX_n < +\infty$. Ако је

$$\lim_{n\to\infty} P\left\{\frac{\sum\limits_{k=1}^{n} X_k}{\sqrt{n}} > 1\right\} = \frac{1}{3}$$

Израчунати дисперзију DX_n.

Решење. Из централне граничне теореме важи

$$P\left\{\frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} \le x\right\} \to \Phi(x)$$

$$\iff P\left\{\frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} > x\right\} \to 1 - \Phi(x).$$

Желимо да $1 - \Phi(x) = \frac{1}{3}$, односно $\Phi(x) = \frac{2}{3}$. Према томе $x = \Phi^{-1}(\frac{2}{3}) = 0.43$. Комбинујући са претходним добија се

$$P\left\{\frac{\sum\limits_{k=1}^{n} X_k}{\sqrt{nDX_1}} > 0.4307273\right\} \to \frac{1}{3}$$

$$P\left\{\frac{\sum\limits_{k=1}^{n} X_k}{\sqrt{n}} > 0.43\sqrt{DX_1}\right\} \to \frac{1}{3}.$$

Из услова задатка следи да је $0.433\sqrt{DX_1} = 1$, односно $DX_1 = 5.41$. ✓

Задатак 18. Општи чланови X_n и Y_n независних низова независних случајних величина имају униформну $\mathcal{U}[0,1]$ и експоненцијалну $\mathcal{E}(2)$ расподелу. Ако је $Z_n = \frac{\sum_{k=1}^{n}(X_k - Y_k)}{\sqrt{n}}$, испитати конвергенцију у расподели низа случајних величина (Z_n).

Решење. Како X_n има униформну $\mathcal{U}[0,1]$ расподелу, следи да је $E(X_n) = \frac{1}{2}$ и $D(X_n) = \frac{1}{12}$. Слично, пошто Y_n има експоненцијалну $\mathcal{E}(2)$ расподелу, следи да је $E(Y_n) = \frac{1}{2}$ и $D(Y_n) = \frac{1}{4}$. Из централне граничне теореме следи да

$$\frac{\sum\limits_{k=1}^{n} X_k - E\Big(\sum\limits_{k=1}^{n} X_k\Big)}{\sqrt{D\Big(\sum\limits_{k=1}^{n} X_k\Big)}} = \frac{\sum\limits_{k=1}^{n} X_k - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \xrightarrow{D} X^* \in \mathcal{N}(0,1) \tag{1}$$

и

$$\frac{\sum\limits_{k=1}^{n} Y_k - E\Big(\sum\limits_{k=1}^{n} Y_k\Big)}{\sqrt{D\Big(\sum\limits_{k=1}^{n} Y_k\Big)}} = \frac{\sum\limits_{k=1}^{n} Y_k - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \xrightarrow{D} Y^* \in \mathcal{N}(0,1). \tag{2}$$

Сада, из (1) можемо закључити да

$$\frac{\sum\limits_{k=1}^{n} X_k - \frac{n}{2}}{\sqrt{n}} \xrightarrow{D} \frac{1}{\sqrt{12}} X^*.$$

Слично, из (2) закључујемо да

$$\frac{\sum\limits_{k=1}^{n} Y_k - \frac{n}{2}}{\sqrt{n}} \xrightarrow{D} \frac{1}{\sqrt{4}} Y^*.$$

Познато је да збир две случајне величине са нормалном расподелом има нормалну расподелу, а пошто је стандардна нормална расподела симетрична око нуле, Y^* и $-Y^*$ имају исту расподелу. С обиром да су низови случајних величина међусобно независни, морају бити независне и њихове граничне вредности. Одатле следи да $\frac{1}{\sqrt{12}}X^* - \frac{1}{\sqrt{4}}Y^*$ има нормалну расподелу са очекивањем

$$E\Big(\frac{1}{\sqrt{12}}X^* - \frac{1}{\sqrt{4}}Y^*\Big) = E\Big(\frac{1}{\sqrt{12}}X^*\Big) - E\Big(\frac{1}{\sqrt{4}}Y^*\Big) = \frac{1}{\sqrt{12}}E(X^*) - \frac{1}{\sqrt{4}}E(Y^*) = 0,$$

и дисперзијом

$$D\Big(\frac{1}{\sqrt{12}}X^* - \frac{1}{\sqrt{4}}Y^*\Big) = D\Big(\frac{1}{\sqrt{12}}X^*\Big) + D\Big(\frac{1}{\sqrt{4}}Y^*\Big) = \frac{1}{12}D(X^*) + \frac{1}{4}D(Y^*) = \frac{1}{12} + \frac{1}{4} = \frac{1}{3},$$

при чему је за рачунање дисперзије коришћена независност случајних величина. Дакле,

$$Z_n = \frac{\sum\limits_{k=1}^{n}(X_k - Y_k)}{\sqrt{n}} = \frac{\sum\limits_{k=1}^{n} X_k - \frac{n}{2}}{\sqrt{n}} - \frac{\sum\limits_{k=1}^{n} Y_k - \frac{n}{2}}{\sqrt{n}} \xrightarrow{D} \frac{1}{\sqrt{12}}X^* - \frac{1}{\sqrt{4}}Y^* \in \mathcal{N}\Big(0, \frac{1}{3}\Big).$$

✓

Издвојене расподеле

У наставку ће бити приказане неке расподеле случајних величина и нека њихова својства.

◇ Дискретне:

- ***Бернулијева расӣодела:*** $X \in Ber(p),\ p \in [0,1]$

  закон расподеле: $X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$, односно

  $$p(x;p) = P\{X = x\} = p^x(1-p)^{1-x},\ x \in \{0,1\}$$

  очекивање: $E(X) = p$

  дисперзија: $D(X) = p(1-p)$

  ⋆ Ако су $X_1, ..., X_n$ независне случајне величине са Бернулијевом $Ber(p)$ расподелом, онда $\sum\limits_{k=1}^{n} X_k$ има биномну $B(n,p)$ расподелу.

- ***Биномна расӣодела:*** $X \in \mathcal{B}(m,p),\ p \in [0,1],\ m \in \mathbf{N}$

  закон расподеле: $p(x;m,p) = P\{X = x\} = \binom{m}{x} p^x(1-p)^{n-x},\ x \in \{0,...,m\}$

  очекивање: $E(X) = mp$

  дисперзија: $D(X) = mp(1-p)$

  ⋆ Ако су $X_1, ..., X_n$ независне случајне величине са биномном $B(m,p)$ расподелом, онда $\sum\limits_{k=1}^{n} X_k$ има биномну $B(nm,p)$ расподелу. Због ове особине, каже се да је биномна расподела адитивна по првом параметру.

- ***Геомеӣријска расӣодела:*** $X \in \mathcal{G}(p),\ p \in [0,1]$

  закон расподеле: $p(x;p) = P\{X = x\} = (1-p)^{x-1}p,\ x \in \mathbf{N}$

  очекивање: $E(X) = \frac{1}{p}$

  дисперзија: $D(X) = \frac{1-p}{p^2}$

  ⋆ Ако су $X_1, ..., X_n$ независне случајне величине са геометријском $\mathcal{G}(p)$ расподелом, онда $\sum\limits_{k=1}^{n} X_k$ има негативну биномну $\mathcal{NB}(n,p)$ расподелу.

- ***Неīаӣивна биномна расӣодела:*** $X \in \mathcal{NB}(m,p),\ p \in [0,1],\ m \in \mathbf{N}$

  закон расподеле: $p(x;m,p) = P\{X = x\} = \binom{x-1}{m-1} p^m(1-p)^{x-m}, x \in \{m, m+1, ...\}$

  очекивање: $E(X) = \frac{m}{p}$

  дисперзија: $D(X) = \frac{m(1-p)}{p^2}$

  ⋆ Ако су $X_1, ..., X_n$ независне случајне величине са негативном биномном $\mathcal{NB}(m,p)$ расподелом, онда $\sum\limits_{k=1}^{n} X_k$ има негативну биномну $\mathcal{NB}(nm,p)$ расподелу. Негативна биномна расподела је адитивна по првом параметру.

- ***Равномерна (дискреӣна униформна) расӣодела:*** $X \in \mathcal{DU}(m),\ m \in \mathbf{N}$

закон расподеле: $X : \begin{pmatrix} 1 & 2 & \cdots & m \\ \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \end{pmatrix}$, односно

$$p(x; m) = P\{X = x\} = \frac{1}{m} I\{x \leq m\},\ x \in \mathbf{N}$$

очекивање: $E(X) = \frac{m+1}{2}$

дисперзија: $D(X) = \frac{m^2-1}{12}$

- ***Пуасонова расподела:*** $X \in \mathcal{P}(\lambda),\ \lambda > 0$

  закон расподеле: $p(x; \lambda) = P\{X = x\} = \frac{\lambda^x e^{-\lambda}}{x!},\ x \in \mathbf{N}_0$

  очекивање: $E(X) = \lambda$

  дисперзија: $D(X) = \lambda$

  ⋆ Ако су $X_1, ..., X_n$ независне случајне величине са Пуасоновом $\mathcal{P}(\lambda)$ расподелом, онда $\sum\limits_{k=1}^{n} X_k$ има Пуасонову $\mathcal{P}(n\lambda)$ расподелу. Дакле, Пуасонова расподела је адитивна по параметру

◇ Непрекидне:

- ***Униформна расподела:*** $X \in \mathcal{U}(a, b),\ a, b \in \mathbb{R},\ a < b$

  густина расподеле: $f(x; a, b) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{иначе.} \end{cases}$

  фунција расподеле: $F(x; a, b) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b. \end{cases}$

  очекивање: $E(X) = \frac{a+b}{2}$

  дисперзија: $D(X) = \frac{(b-a)^2}{12}$

- ***Експоненцијална расподела:*** $X \in \mathcal{E}(\lambda),\ \lambda > 0$

  густина расподеле: $f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$

  фунција расподеле: $F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$

  очекивање: $E(X) = \frac{1}{\lambda}$

  дисперзија: $D(X) = \frac{1}{\lambda^2}$

  ⋆ Ако су $X_1, ..., X_n$ независне случајне величине са експоненцијалном $\mathcal{E}(\lambda)$ расподелом, онда $\sum\limits_{k=1}^{n} X_k$ има гама $\gamma(n, \lambda)$ расподелу.

- ***Гама расподела:*** $X \in \gamma(a, b),\ a > 0,\ b > 0$

  густина расподеле: $f(x; a, b) = \begin{cases} \frac{x^{a-1} b^a e^{-bx}}{\Gamma(a)}, & x \geq 0, \\ 0, & x < 0, \end{cases}$ где је $\Gamma(a)$ гама функција

  фунција расподеле: $F(x; a, b) = \int\limits_0^x f(t; a, b) dt$

очекивање: $E(X) = \frac{a}{b}$

дисперзија: $D(X) = \frac{a}{b^2}$

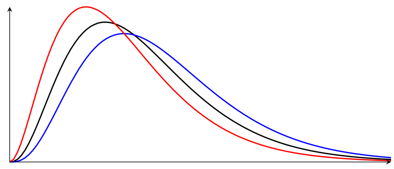
⋆ Ако су $X_1, ..., X_n$ независне случајне величине са гама $\gamma(a,b)$ расподелом, онда $\sum\limits_{k=1}^{n} X_k$ има гама $\gamma(na,b)$ расподелу. Дакле, гама расподела је адитивна по првом параметру

⋆ Ако $X \in \gamma(a,b)$, онда $2bX \in \chi^2_{2a}$.

- ***Хи-квадрат расподела са m степени слободе:*** $X \in \chi^2_m,\ m \in \mathbb{N}$

густина расподеле:

$$f(x;m) = \begin{cases} \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$



фунција расподеле: $F(x;m) = \int\limits_0^x f(t;m)dt$

очекивање: $E(X) = m$

дисперзија: $D(X) = 2m$

⋆ Ако су $X_1, ..., X_n$ независне случајне величине са χ^2_m расподелом, онда $\sum\limits_{k=1}^{n} X_k$ има χ^2_{nm} расподелу. Дакле, χ^2 расподела има особину адитивности.

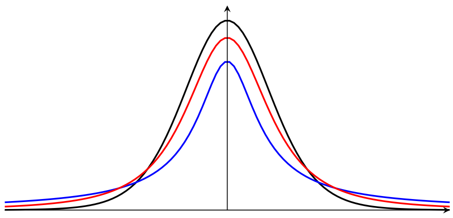
⋆ Ако $Y \in \mathcal{N}(0,1)$, онда $Y^2 \in \chi^2_1$.

⋆ Ако $X \in \chi^2_m$, онда $X \in \gamma(\frac{m}{2}, \frac{1}{2})$.

- ***Студентова расподела расподела са m степени слободе:*** $X \in t_m,\ m \in \mathbb{N}$

⋆ Ако су $X_0, X_1, ..., X_m$ независне случајне величине са $\mathcal{N}(0,1)$ расподелом, онда

$$\frac{X_0}{\sqrt{\frac{\sum\limits_{k=1}^{m} X_k^2}{m}}}$$



има Студентову t_m расподелу.

- ***Бета расподела:*** $X \in \beta(a,b),\ a > 0,\ b > 0$

густина расподеле: $f(x;a,b) = \begin{cases} \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, & x \in (0,1), \\ 0, & \text{иначе}, \end{cases}$ где је $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ бета функција

фунција расподеле: $F(x;a,b) = \int\limits_0^x f(t;a,b)dt$

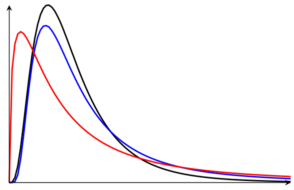
очекивање: $E(X) = \frac{a}{a+b}$

дисперзија: $D(X) = \frac{ab}{(a+b)^2(a+b+1)}$

⋆ Ако $X \in \beta(a,1)$, онда $-\ln X \in \mathcal{E}(a)$.

⋆ Ако $X \in \beta(a,b)$, онда $\frac{2bX}{2a(1-X)} \in \mathcal{F}_{2a,2b}$.

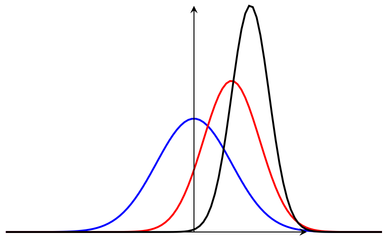
- ***Фишерова расӣодела:*** $X \in \mathcal{F}_{a,b},\ a > 0, b > 0$



$\star$ Ако $X \in \mathcal{F}_{a,b}$, онда $\frac{1}{X} \in \mathcal{F}_{b,a}$.

$\star$ Ако $Y \in \chi^2_a$ и $Z \in \chi^2_b$ и независне су, онда $\frac{\frac{Y}{a}}{\frac{Z}{b}} \in \mathcal{F}_{a,b}$

- ***Нормална расӣодела:*** $X \in \mathcal{N}(m, \sigma^2),\ m \in \mathbb{R},\ \sigma^2 > 0$



густина расподеле: $f(x; m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$

фунција расподеле:

$$\Phi(x; m, \sigma^2) = \int\limits_{-\infty}^{x} f(t; m, \sigma^2) dt$$

очекивање: $E(X) = m$

дисперзија: $D(X) = \sigma^2$

$\star$ Ако $X \in \mathcal{N}(m, \sigma^2)$, онда $\frac{X-m}{\sigma} \in \mathcal{N}(0, 1)$.

$\star$ Ако су $X_1, ..., X_n$ независне случајне величине са $\mathcal{N}(m, \sigma^2)$ расподелом, онда

- $\sum\limits_{k=1}^{n} X_k \in \mathcal{N}(nm, n\sigma^2)$;
- $\overline{X}_n \in \mathcal{N}(m, \frac{\sigma^2}{n})$;
- $\sum\limits_{k=1}^{n} a_k X_k \in \mathcal{N}(m \sum\limits_{k=1}^{n} a_k, \sigma^2 \sum\limits_{k=1}^{n} a_k^2)$, где су $a_1, ..., a_n$ реални бројеви;
- $\sum\limits_{k=1}^{n} \left(\frac{X_k - m}{\sigma}\right)^2 \in \chi^2_n$.

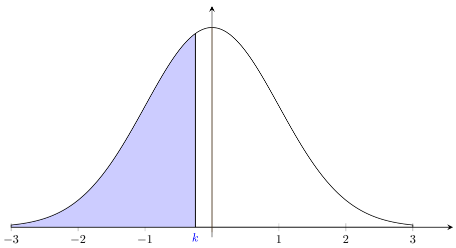
Наӣомена: Са $\Phi_0(x)$ ће у наставку бити означена вероватноћа која је приказана у таблицама, односно

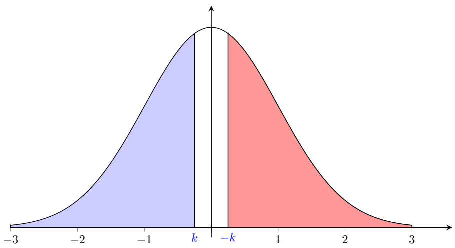
$$\Phi_0(x) = \int_0^x f(t; 0, 1) dt = \Phi(x; 0, 1) - \Phi(0; 0, 1) = \Phi(x; 0, 1) - 0.5.$$

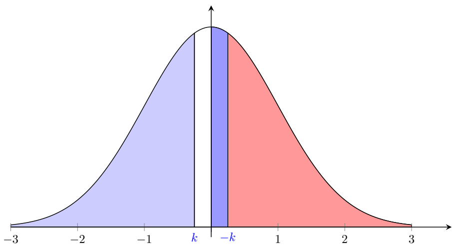
Задатак 19. Ако случајна величина X има нормалну $\mathcal{N}(2, a)$ расподелу, израчунати позитиван реалан број a такав да је $P\{X > 1.5\} = 0.59871$.

Решење.

$$0.59871 = P\{X > 1.5\} = P\left\{\frac{X-2}{\sqrt{a}} > \frac{1.5-2}{\sqrt{a}}\right\}$$
$$= 1 - P\left\{\frac{X-2}{\sqrt{a}} \le \frac{1.5-2}{\sqrt{a}}\right\} = 1 - \Phi\Big(\frac{1.5-2}{\sqrt{a}}\Big)$$
$$\Longrightarrow\ \Phi\Big(-\frac{0.5}{\sqrt{a}}\Big) = 1 - 0.59871 = 0.40129$$







$$\begin{aligned}\Longrightarrow \Phi^{-1}(0.40129) &= \Phi^{-1}(0.5-0.09871)\\ &= -\Phi_0^{-1}(0.09871)\\ &= -0.25\end{aligned}$$

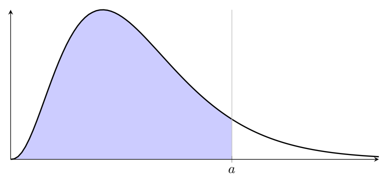
$$\Longrightarrow -\frac{0.5}{\sqrt{a}} = \Phi^{-1}(0.40129) = -0.25$$

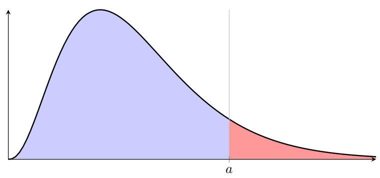
$$\Longrightarrow a = \left(\frac{0.5}{0.25}\right)^2 = 4$$

✓

Задатак 20**.** Ако случајна величина X има χ^2_7 расподелу, израчунати реалан број a такав да је $P\{X \leq a\} = 0.9$.

Решење. $P\{X \leq a\} = 0.9 \implies P\{X > a\} = 1 - 0.9 = 0.1 \implies a = \chi^2_{7;0.1} = 12.017$

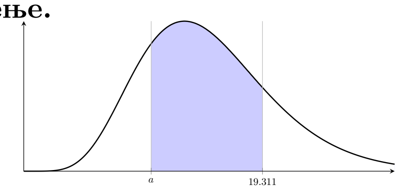




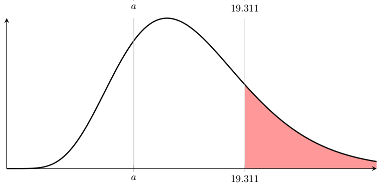
✓

Задатак 21**.** Ако случајна величина X има χ^2_{15} расподелу, израчунати реалан број a такав да је $P\{a < X < 19.311\} = 0.6$.

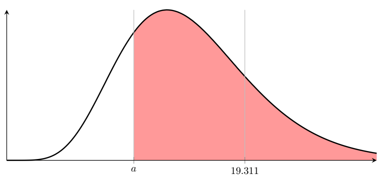
Решење.



$$\begin{aligned}0.6 &= P\{a < X < 19.311\}\\ &= P\{X \geq a\} - P\{X \geq 19.311\}\end{aligned}$$



$$\Longrightarrow P\{X \geq 19.311\} = 0.2$$



$$\begin{aligned}P\{X > a\} &= P\{a < X < 19.311\}\\ &\quad + P\{X \geq 19.311\}\\ &= 0.6 + 0.2 = 0.8\end{aligned}$$

$$\Longrightarrow a = \chi^2_{15;0.8} = 10.307$$

✓

Задатак 22. Ако случајна величина X има χ^2_{100} расподелу, израчунати реалан број a такав да је $P\{X > a\} = 0.7$.

Решење. Како $X \in \chi^2_{100}$, одговарајућу вредност није могуће прочитати у таблици. У том случају, када је n довољно велико, $X \sim \mathcal{N}(100, 200)$. Према томе

$$0.7 = P\{X > a\} = P\left\{\frac{X-100}{\sqrt{200}} > \frac{a-100}{\sqrt{200}}\right\} \approx 1 - \Phi\left(\frac{a-100}{\sqrt{200}}\right)$$

$$\implies \frac{a-100}{\sqrt{200}} = \Phi^{-1}(0.3) = \Phi_0^{-1}(0.5 - 0.2) = -\Phi^{-1}(0.2) \approx -0.525$$

$$\implies a = 100 - \Phi^{-1}(0.3)\sqrt{200} \approx 92.57$$

✓

Задатак 23. Случајна величина X има нормалну $\mathcal{N}(0, 1)$ расподелу, а случајна величина Y има нормалну $\mathcal{N}(0, 4)$ расподелу. Ако су X и Y независне, одредити реалан број a такав да случајна величина $X^2 + aY^2$ има χ^2_2 расподелу.

Решење. $X \in \mathcal{N}(0, 1) \implies X^2 \in \chi^2_1$

$$Y \in \mathcal{N}(0, 4) \implies \frac{Y}{\sqrt{4}} \in \mathcal{N}(0, 1) \implies \frac{Y^2}{4} \in \chi^2_1$$

Из адитивности χ^2 расподеле следи да $X^2 + \frac{Y^2}{4} \in \chi^2_2$, па можемо закључити да је $a = \frac{1}{4}$. ✓

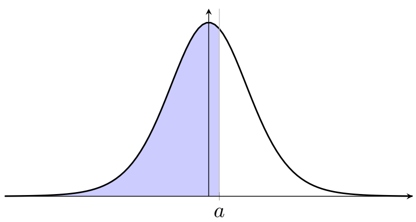
Задатак 24. Ако случајна величина X има Студентову t_{10} расподелу, израчунати реалан број a такав да је:

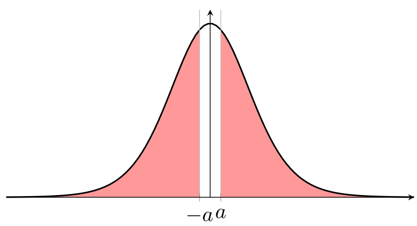
а) $P\{X < a\} = 0.6$;

б) $P\{X > a\} = 0.6$;

в) $P\{|X| < a\} = 0.6$.

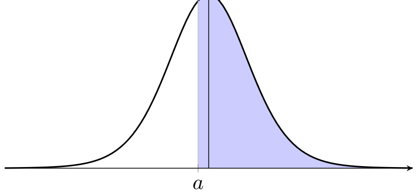
Решење. а) $P\{X < a\} = 0.6 \implies P\{X \geq a\} = 0.4 \implies a = t_{10;2\cdot 0.4} = 0.260$

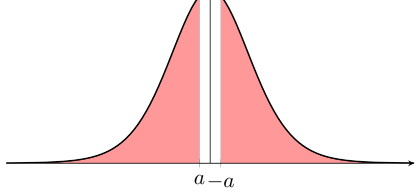




б) $P\{X > a\} = 0.6 \implies P\{X \leq a\} = 0.4$

$$\implies P\{X \geq -a\} = 0.4 \implies a = -t_{10;2\cdot 0.4} = -0.260$$



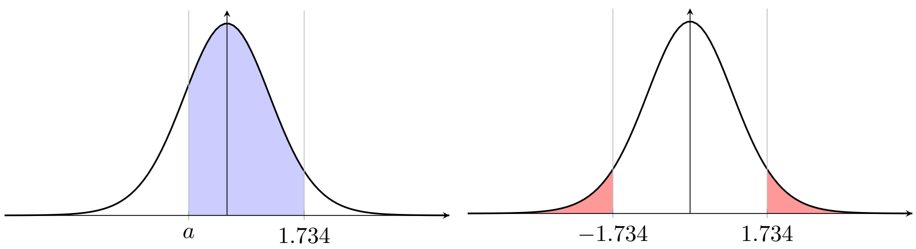


в) $P\{|X| < a\} = 0.6 \implies P\{|X| \geq a\} = 0.4 \implies a = t_{10;0.4} = 0.879$

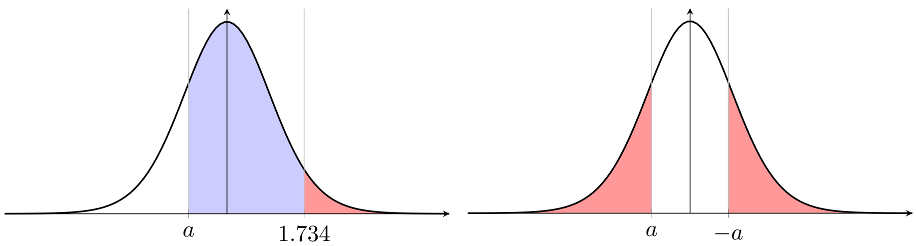
✓

Задатак 25. Ако случајна величина X има Студентову t_{18} расподелу, израчунати реалан број a такав да је $P\{a < X < 1.734\} = 0.75$.

Решење.



Из таблица, слично као у претходном задатку, следи да је $P\{|X| \geq 1.734\} = 0.10$, односно $P\{X \geq 1.734\} = 0.05$.



Из претходног и услова задатке следи да је

$$0.75 = P\{a < X < 1.734\} = P\{X < 1.734\} - P\{X \leq a\} = 0.95 - P\{X \leq a\}$$

$$\Longrightarrow \; P\{X \leq a\} = 0.20$$

$$P\{|X| > -a\} = P\{X \leq a\} + P\{X \geq -a\} = 0.4 \;\Longrightarrow\; a = -t_{18;0.4} = -0.862 \qquad \checkmark$$

Задатак 26. Ако случајна величина X има Студентову t_7 расподелу, израчунати реалан број a такав да је $P\{-1.895 < X < a\} = 0.35$.

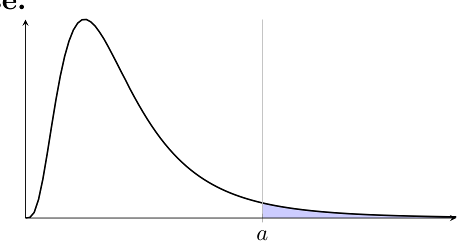
Решење. Из таблица следи да је $P\{|X| > 1.895\} = 0.10$, односно $P\{X < -1.895\} = 0.05$. Из претходног и услова задатке следи да је

$$0.35 = P\{-1.895 < X < a\} = P\{X < a\} - P\{X \leq -1.895\} = P\{X < a\} - 0.05$$

$$\Longrightarrow \; P\{X < a\} = 0.40 \;\Longrightarrow\; P\{X > -a\} = 0.4 \;\Longrightarrow\; a = -t_{7;2\cdot 0.4} = -0.263 \qquad \checkmark$$

Задатак 27. Ако случајна величина X има Фишерову $F_{10,12}$ расподелу, израчунати реалан број а такав да је $P\{X \geq a\} = 0.05$.

Решење.



$$P\{X \geq a\} = 0.05$$
$$\Longrightarrow \; a = F_{10,12;0.05} = 2.75$$

$\checkmark$

Задатак 28**.** Ако случајна величина X има Фишерову $F_{10,12}$ расподелу, израчунати реалан број a такав да је $P\{X \geq a\} = 0.95$.

Решење. $0.95 = P\{X \geq a\} = P\left\{\frac{1}{X} \leq \frac{1}{a}\right\} \implies P\left\{\frac{1}{X} > \frac{1}{a}\right\} = 0.05$

$\implies \frac{1}{a} = F_{12,10;0.05} = 2.91 \implies a = 0.3436$ ✓

Математичка статистика

Статистика се бави прикупљањем података, њиховом анализом и доношењем закључака на основу њих. Циљ је да се пронађе најбољи математички модел, којим би се подаци могли описати и на основу кога истраживач може да донесе закључке.

Популација је скуп елемената чије карактеристике проучавамо. Елементе популације називамо **јединкама популације**. Претпоставимо да желимо да донесемо закључке о некој особини елемената популације (нпр. популацију чине сви студенти на неком факултету, а интересује нас просечна оцена). Такву карактеристику називамо **обележјем** популације.

Математички, обележје је функција која јединкама популације додељује вредности из скупа реалних бројева[1] $X : \Omega \to \mathbb{R}$, где је Ω популација, а $X(\omega) \in \mathbb{R}$ је вредност обележја X за једнинку $\omega \in \Omega$.

У вероватноћи је била позната расподела случајне величине, а треба одредити вероватноћу неког конкретног догађаја. Са друге стране у статистици знамо који догађај се реализовао, али не знамо расподелу обележја. Дакле, циљ нам је да одредимо расподелу обележја X. Означимо је са $F(\cdot; \theta)$, где је θ параметар расподеле. Постоје два типа проблема:

1) Параметарски: Знамо тип расподеле, али не знамо вредност параметра (нпр. знамо да обележје има нормалну расподелу, али не знамо који су њени параметри),

2) Непараметарски: Непозната расподела.

Најпрецизније је да извршимо **попис** - регистровање вредности обележја на свакој јединки популације, али се то обично не ради (некад нам није ни доступна читава популација, а и ако јесте, најчешће се то не ради због времена, финансија, итд.), већ се вади подскуп популације - **узорак**, на основу кога се доносе закључци. Ако се тај подскуп бира на случајан начин, говоримо о **случајном узорку**. Желимо да узорак буде **репрезентативан**, односно да добро осликава вредности обележја на читавој популацији. Постоје различити начини бирања узорака, а све у циљу добијања најрепрезентативнијег узорка и доношења најпрецизнијих закључака.

Просто случајно узорковање подразумева да су сви могући узорци истог обима једнако вероватни. Може се узорковати без понављања и са понављањем. Ми ћемо увек претпостављати да је одабран **прост случајан узорак са понављањем**.

Прост случајан узорак обима n је n-димензионална случајна величина $\mathbf{X} = (X_1, ..., X_n)$, где су X_i независне случајне величине[2] и све имају исту расподелу као обележје X. Реализовани узорак је n-димензионални вектор $(x_1, ..., x_n) \in \mathbb{R}^n$ (регистроване вредности на одабраном узорку).

[1]Сматрамо да је обележје нумеричка карактеристика, јер ако није, увек је можемо кодирати, нпр. пол кодирамо са 0 и 1.

[2]Случајне величине јер су то вредности обележја на **случајно** одабраним јединкама популације.

Како се не би радило са $n-$димензионим случајним величинама, од узорка се формирају једнодимензионе случајне величине које се називају статистике. **Статистиком** се назива било која Борелова функција од узорка. Оне зависе од узорка и константи, а не смеју зависити од непознатих параметара. Неки од примера које ћемо користити:

Узорачка средина: $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$

Узорачка дисперзија: $\overline{S}_n^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$

Поправљена узорачка дисперзија: $\widetilde{S}_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$

Статистике поретка: $X_{(1)} = \min\{X_1, ..., X_n\}, ..., X_{(n)} = \max\{X_1, ..., X_n\}$

Нека је $\mathbf{X}_n$ узорак из популације чије обележје X има нормалну $\mathcal{N}(m, \sigma^2)$ расподелу. Тада:

- $\frac{\overline{X}_n - m}{\sigma}\sqrt{n} \in \mathcal{N}(0,1)$;
- $\frac{n\overline{S}_*^2}{\sigma^2} \in \chi_n^2$, где је $\overline{S}_*^2 = \frac{1}{n}\sum\limits_{k=1}^{n}(X_k - m)^2$;
- $\frac{n\overline{S}_n^2}{\sigma^2} \in \chi_{n-1}^2$;
- $\frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1} \in t_{n-1}$;

Нека су дати независни узорци $\mathbf{X}_n$ из популације чије обележје X има нормалну $\mathcal{N}(m_1, \sigma_1^2)$ расподелу и $\mathbf{Y}_k$ из популације чије обележје Y има нормалну $\mathcal{N}(m_2, \sigma_2^2)$ расподелу. Тада

- $\frac{n(k-1)\sigma_2^2\overline{S}_n^2(X)}{k(n-1)\sigma_1^2\overline{S}_k^2(Y)} \in \mathcal{F}_{n-1,k-1}$;
- Ако је $\sigma_1^2 = \sigma_2^2 = \sigma^2$: $\frac{(\overline{X}_n - m_1) - (\overline{Y}_k - m_2)}{\sqrt{n\overline{S}_n^2(X) + k\overline{S}_k^2(Y)}}\sqrt{\frac{nk}{n+k}(n+k-2)} \in t_{n+k-2}$.

Оцењивање параметара

Претпоставимо да имамо параметарски проблем, тј. да је статистички модел одређен до на непознате параметре расподеле. Дакле, претпостављамо да $X : F(\cdot;\theta)$, при чему је познат тип расподеле F, али не знамо вредност параметра θ.

Разликујемо два типа оцена:

1) **Тачкасте оцене -** оцењујемо конкретну вредност параметра,

2) **Интервалне оцене -** оцењујемо границе интервала коме припада параметар са великом вероватноћом.

За почетак се бавимо тачкастим оценама.

Оцена параметра θ је нека статистика $\widehat{\theta}_n$ (функција од узорка), која се бира тако да најбоље осликава вредност параметра који оцењује. Постоје различите методе добијања оцена, али за почетак, погледајмо које су то особине које би квалитетна оцена требало да има:

Непристрасност: $\widehat{\theta}_n$ је непристрасна оцена параметра θ, ако је $E(\widehat{\theta}_n) = \theta$.

⋆ Ако оцена $\widehat{\theta}_n$ није непристрасна, али $\lim\limits_{n\to\infty} E(\widehat{\theta}_n) = \theta$, онда је оцена асимптотски непристрасна.

Постојаност: Оцена $\widehat{\theta}_n$ је постојана оцена параметра θ, ако $\widehat{\theta}_n \xrightarrow{P} \theta$, односно, ако за

$$(\forall \varepsilon > 0) \ \lim_{n\to\infty} P\{|\widehat{\theta}_n - \theta| > \varepsilon\} = 0.$$

⋆ Ако је $\widehat{\theta}_n$ непристрасна оцена параметра θ и $D(\widehat{\theta}_n) \to 0, n \to \infty$, онда је $\widehat{\theta}_n$ постојана оцена параметра θ, јер је

$$P\{|\widehat{\theta}_n - \theta| > \varepsilon\} \leq \frac{E(\widehat{\theta}_n - \theta)^2}{\varepsilon^2} = \frac{D(\widehat{\theta}_n)}{\varepsilon^2}$$

Као што смо већ рекли, постоје различите методе добијања оцена, па морамо наћи неку меру квалитета оцена, која ће нам помоћи да изаберемо најбољу. Постоји више начина за то, али често се користи следећи критеријум поређења средње квадратних грешки. Односно

- Ако су $\widehat{\theta}_n$ и $\widetilde{\theta}_n$ две оцене параметра θ, кажемо да је $\widehat{\theta}_n$ боља од $\widetilde{\theta}_n$ у средње квадратном смислу, ако је

$$MSE(\widehat{\theta}_n) = E(\widehat{\theta}_n - \theta)^2 < E(\widetilde{\theta}_n - \theta)^2 = MSE(\widetilde{\theta}_n).$$

- Ако су $\widehat{\theta}_n$ и $\widetilde{\theta}_n$ непристрасне оцене, довољно је да упоређујемо дисперзије, јер је онда

$$MSE(\widehat{\theta}_n) = E(\widehat{\theta}_n - \theta)^2 = E(\widehat{\theta}_n - E(\widehat{\theta}_n))^2 = D(\widehat{\theta}_n).$$

  Дакле, у том случају боља је она оцена која има мању дисперзију.

Задатак 29. Обележје X посматране популације има нормалну $\mathcal{N}(m, m)$ расподелу. За оцену непознатог параметра m предложене су две статистике: узорачка средина $\overline{X}_n$ и поправљена узорачка дисперзија $\widetilde{S}_n^2$. Испитати непристрасност и постојаност тих оцена. Испитати која је оцена боља у средње квадратном смислу.

Решење. Испитујемо непристрасност обе оцене:

$$E(\overline{X}_n) = E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n} n E(X_1) = m,$$

$$E(\widetilde{S}_n^2) = E\left(\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2\right) = E\left(\frac{n}{n-1}\overline{S}_n^2\right) = E\left(\frac{n\overline{S}_n^2}{m}\right)\frac{m}{n-1} = m,$$

при чему је коришћено да случајна величина $\frac{n\overline{S}_n^2}{m} \in \chi^2_{n-1}$, па је њено очекивање једнако $n-1$, а дисперзија $2(n-1)$.

Обе оцене су непристрасне, па су њихове средње квадратне грешке једнаке њиховим дисперзијама. Дакле, довољно је да упоредимо дисперзије ових оцена да бисмо испитали која је оцена боља у средње квадратном смислу. Како је

$$D(\overline{X}_n) = D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2} n D(X_1) = \frac{m}{n},$$

$$D(\widetilde{S}_n^2) = D\left(\frac{n}{n-1}\overline{S}_n^2\right) = D\left(\frac{m}{n-1}\frac{n\overline{S}_n^2}{m}\right) = \frac{m^2}{(n-1)^2} D\left(\frac{n\overline{S}_n^2}{m}\right) = \frac{2(n-1)m^2}{(n-1)^2} = \frac{2m^2}{n-1},$$

то је

$$D(\overline{X}_n) > D(\widetilde{S}_n^2) \iff \frac{m}{n} > \frac{2m^2}{n-1} \iff \frac{n-1}{n} > 2m$$

Када је $\frac{n-1}{n} > 2m$, онда је $\widetilde{S}_n^2$ боља у средње квадратном смислу, а када је $\frac{n-1}{n} < 2m$, онда је боља $\overline{X}_n$. Ако важи једнакост подједнако су добре. С обзиром да не постоји универзално правило, које важи за свако $m > 0$, не можемо закључити да је нека универзално боља.

Остало је још да испитамо постојаност ових оцена. Постојаност подразумева да оцена конвергира у вероватноћи ка параметру m. Нека је $\varepsilon > 0$ произвољно:

$$P\{|\overline{X}_n - m| > \varepsilon\} \leq \frac{E(\overline{X}_n - m)^2}{\varepsilon^2} = \frac{D(\overline{X}_n)}{\varepsilon^2} = \frac{m}{n\varepsilon^2},$$

при чему је коришћена Чебишовљева неједнакост. Дакле,

$$0 \leq P\{|\overline{X}_n - m| > \varepsilon\} \leq \frac{m}{n\varepsilon^2} \to 0, n \to \infty$$

па по теореми о два полицајца $P\{|\overline{X}_n - m| > \varepsilon\} \to 0, n \to \infty$ и то важи за $\forall \varepsilon > 0$. Закључујемо да $\overline{X}_n \xrightarrow{P} m$, односно да је $\overline{X}_n$ постојана оцена за m. Даље,

$$0 \leq P\{|\widetilde{S}_n^2 - m| > \varepsilon\} \leq \frac{D(\widetilde{S}_n^2)}{\varepsilon^2} = \frac{2m^2}{(n-1)\varepsilon^2} \to 0, n \to \infty$$

Теорема о два полицајца да је да $P\{|\widetilde{S}_n^2 - m| > \varepsilon\} \to 0, n \to \infty$, за ε одабрано произвољно. Па ово важи за свако $\varepsilon > 0$, те $\widetilde{S}_n^2 \xrightarrow{P} m$. Дакле, оцена $\widetilde{S}_n^2$ је такође постојана оцена за m. ✓

Задатак 30. Обележје X има нормалну $\mathcal{N}(m, 1), m \in (0, 1)$, расподелу. За оцену непознатог параметра m, на основу узорка обима n, предложене су две статистике: узорачка средина $\overline{X}_n$ и $T_n = \overline{X}_n I\{0 \leq \overline{X}_n \leq 1\} + I\{\overline{X}_n > 1\}$. Испитати која оцена је боља у средње квадратном смислу.

Решење. Прво ћемо испитати непристрасност оцене $T_n = \overline{X}_n I\{0 \leq \overline{X}_n \leq 1\} + I\{\overline{X}_n > 1\}$:

$$E(T_n) = E(\overline{X}_n I\{0 \leq \overline{X}_n \leq 1\} + I\{\overline{X}_n > 1\}) = E(\overline{X}_n I\{0 \leq \overline{X}_n \leq 1\}) + P\{\overline{X}_n > 1\}.$$

Знамо да $\overline{X}_n \in \mathcal{N}(m, \frac{1}{n})$, па $Z_n = \sqrt{n}(\overline{X}_n - m) \in \mathcal{N}(0, 1)$. Дакле, $\overline{X}_n$ можемо записати као $\frac{Z_n}{\sqrt{n}} + m$, па је

$$\begin{aligned}
E(T_n) &= E\left(\left(\frac{Z_n}{\sqrt{n}} + m\right) I\{0 \leq \frac{Z_n}{\sqrt{n}} + m \leq 1\}\right) + P\left\{\frac{Z_n}{\sqrt{n}} + m > 1\right\} \\
&= \frac{1}{\sqrt{n}} E\left(Z_n I\{-m\sqrt{n} \leq Z_n \leq (1-m)\sqrt{n}\}\right) + mP\{-m\sqrt{n} \leq Z_n \leq (1-m)\sqrt{n}\} \\
&\quad + P\{Z_n > (1-m)\sqrt{n}\} \\
&= \frac{1}{\sqrt{n}} \int_{-m\sqrt{n}}^{(1-m)\sqrt{n}} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + m\Phi((1-m)\sqrt{n}) - m\Phi(-m\sqrt{n}) + 1 - \Phi((1-m)\sqrt{n}) \\
&\neq m
\end{aligned}$$

Оцена није непристрасна, међутим, кад $n \to \infty$:

$$\frac{1}{\sqrt{n}} \int_{-m\sqrt{n}}^{(1-m)\sqrt{n}} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \to 0,$$

јер

$$\Phi((1-m)\sqrt{n}) \to \Phi(\infty) = 1$$

и

$$\Phi(-m\sqrt{n}) \to \Phi(-\infty) = 0.$$

па

$$E(T_n) \to 0 + m - 0 + 1 - 1 = m, n \to \infty.$$

Дакле, T_n није непристрасна, али јесте асимптотски непристрасна.

Сада хоћемо да испитамо која од предложених оцена је боља у средње квадратном смислу, односно желимо да упоредимо $MSE(\overline{X}_n) = E(\overline{X}_n - m)^2$ и $MSE(T_n) = E(T_n - m)^2$. Приметимо да

$$T_n = \begin{cases} 0, \text{ ако } \overline{X}_n < 0 \\ \overline{X}_n, \text{ ако } 0 \le \overline{X}_n \le 1 \\ 1, \text{ ако } \overline{X}_n > 1 \end{cases} \implies (T_n - m)^2 = \begin{cases} m^2, \text{ ако } \overline{X}_n < 0 \\ (\overline{X}_n - m)^2, \text{ ако } 0 \le \overline{X}_n \le 1 \\ (1-m)^2, \text{ ако } \overline{X}_n > 1 \end{cases}$$

Дакле,

$$(T_n - m)^2 - (\overline{X}_n - m)^2 = \begin{cases} m^2 - (\overline{X}_n - m)^2, \text{ ако } \overline{X}_n < 0 \\ 0, \text{ ако } 0 \le \overline{X}_n \le 1 \\ (1-m)^2 - (\overline{X}_n - m)^2, \text{ ако } \overline{X}_n > 1 \end{cases}$$

- Ако $\overline{X}_n < 0$: $m \in (0,1)$, па је $(\overline{X}_n - m)^2 > m^2$, тј. $m^2 - (\overline{X}_n - m)^2 < 0$.
- Ако $\overline{X}_n > 1$: $m \in (0,1)$, па је $(\overline{X}_n - m)^2 > (1-m)^2$, тј. $(1-m)^2 - (\overline{X}_n - m)^2 < 0$.

Стога, $(T_n - m)^2 - (\overline{X}_n - m)^2 \le 0$, па је и $E((T_n - m)^2 - (\overline{X}_n - m)^2) \le 0$, односно $E(T_n - m)^2 \le E(\overline{X}_n - m)^2$, па је T_n ефикаснија од $\overline{X}_n$. ✓

Задатак 31. Обележје X има униформну $\mathcal{U}[0,\theta]$, $\theta > 0$, расподелу. За оцену непознатог параметра θ, на основу узорка обима n, предложене су две статистике: $2\overline{X}_n$ и $\frac{n+1}{n}X_{(n)}$. Испитати која оцена је боља у средње квадратном смислу.

Решење. Испитаћемо непристрасност за почетак:

$$E(2\overline{X}_n) = 2E(X_1) = 2 \cdot \frac{\theta}{2} = \theta,$$
$$E\left(\frac{n+1}{n}X_{(n)}\right) = \frac{n+1}{n}E(X_{(n)}).$$

Да бисмо израчунали очекивање случајне величине $X_{(n)}$ одредимо прво њену расподелу:

$$F_{(n)}(x) = P\{X_{(n)} \le x\} = P\{\max\{X_1, ..., X_n\} \le x\} = P\{X_1 \le x, ..., X_n \le x\}$$
$$= (P\{X_1 \le x\})^n = \left(\frac{x}{\theta}\right)^n, \ x \in [0,\theta],$$

односно

$$f_{(n)}(x) = \frac{d}{dx}F_{(n)}(x) = n\frac{x^{n-1}}{\theta^n}.$$

Дакле,

$$E\left(\frac{n+1}{n}X_{(n)}\right) = \frac{n+1}{n}\int_0^\theta xn\frac{x^{n-1}}{\theta^n}dx = \frac{n+1}{\theta^n}\frac{x^{n+1}}{n+1}\Big|_0^\theta = \theta.$$

Обе оцене су непристрасне, па можемо упоређивати дисперзије. Како је

$$\begin{aligned}
D(2\overline{X}_n) &= 4D(\overline{X}_n)=\frac{4}{n^2}nD(X_1) = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n},\\
D\left(\frac{n+1}{n}X_{(n)}\right) &= \frac{(n+1)^2}{n^2}D(X_{(n)})\\
&= \frac{(n+1)^2}{n^2}(E(X_{(n)}^2) - (E(X_{(n)}))^2)\\
&= \frac{(n+1)^2}{n^2}\left(\int_0^\theta x^2 n\frac{x^{n-1}}{\theta^n}dx - \left(\int_0^\theta xn\frac{x^{n-1}}{\theta^n}dx\right)^2\right)\\
&= \frac{(n+1)^2}{n^2}\left(\frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2\right)\\
&= \frac{(n+1)^2}{n^2}\frac{n}{(n+1)^2(n+2)}\theta^2 = \frac{\theta^2}{n(n+2)},
\end{aligned}$$

то је

$$D\left(\frac{n+1}{n}X_{(n)}\right) < D(2\overline{X}_n) \iff \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{3n} \iff 3 < n+2$$

Ово важи кад год је $n > 1$, па закључујемо да је оцена $\frac{n+1}{n}X_{(n)}$ боља у средње квадратном смислу. ✓

Ефикасност (непристрасних) оцена

Фамилија расподела је класа која садржи све расподеле одређеног типа, нпр.

- ⋆ Фамилија $\mathcal{N}(m,\sigma^2)$ расподела, $m \in \mathbb{R}$, $\sigma > 0$, садржи расподеле $\mathcal{N}(0,1)$, $\mathcal{N}(0,2)$, $\mathcal{N}(1,2)$,...
- ⋆ Фамилија $\mathcal{U}(1,\theta)$ расподела, $\theta > 1$ садржи расподеле $\mathcal{U}(1,1.5)$, $\mathcal{U}(1,2)$, $\mathcal{U}(1,10)$,...
- ⋆ Фамилија $\mathcal{P}(\lambda)$ расподела, $\lambda > 0$ садржи расподеле $\mathcal{P}(1)$,$\mathcal{P}(2)$,$\mathcal{P}(10)$,...

Фамилија расподела $\{p(x;\theta)\}$[3] је **регуларна** у смислу Рао-Крамера, ако:

1) Узорачки простор $\mathcal{X} = \{x | p(x;\theta) > 0\}$ (скуп свих вредности које може узети случајна величина са том расподелом) не зависи од параметра θ;
2) За свако $x \in \mathcal{X}$ $p(x;\theta)$ је диференцијабилно по θ;
3) $E\Big(\frac{\partial}{\partial\theta}\ln p(X;\theta)\Big) = 0$.

У суштини, потребно је формално проверити само услов 3) (прва два се лако проверавају). Ако је $p(x;\theta)$ закон расподеле, онда је

$$3) \iff E\Big(\frac{1}{p(X;\theta)}\frac{\partial p(X;\theta)}{\partial\theta}\Big)=0 \iff \sum_{x\in\mathcal{X}}\frac{1}{p(x;\theta)}\frac{\partial p(x;\theta)}{\partial\theta}p(x;\theta)=0 \iff \sum_{x\in\mathcal{X}}\frac{\partial p(x;\theta)}{\partial\theta}=0.$$

Дакле, за дискретне фамилије, довољно је доказати да је $\sum\limits_{x\in\mathcal{X}}\frac{\partial p(x;\theta)}{\partial\theta} = 0$. Ако је у питању фамилија непрекидних расподела, важе аналогни услови с тим што се закон расподеле $p(x;\theta)$ замењује густином $f(x;\theta)$, а сума интегралом, па је услов 3) еквивалентан са $\int\limits_{x\in\mathcal{X}}\frac{\partial f(x;\theta)}{\partial\theta}dx = 0$.

[3] Може бити и дискретног и апсолутно непрекидног типа.

За регуларне фамилије дефинишемо **информациону функцију Фишера**

$$I(\theta) = E\left(\frac{\partial}{\partial\theta}\ln p(X;\theta)\right)^2 = E\left(-\frac{\partial^2}{\partial\theta^2}\ln p(X;\theta)\right)$$

и **Рао-Крамерову доњу границу дисперзије свих непристрасних оцена параметра** θ

$$G = \frac{1}{nI(\theta)}.$$

За сваку непристрасну оцену $\hat{\theta}_n$ важи да је $G \le D(\hat{\theta}_n)$, а ако је $G = D(\hat{\theta}_n)$, онда кажемо да је оцена $\hat{\theta}_n$ **ефикасна**.

У складу са тиме, дефинишемо **ефикасност** оцене $\hat{\theta}_n$

$$ef(\hat{\theta}_n) = \frac{G}{D(\hat{\theta}_n)}\begin{cases} < 1, & \text{оцена није ефикасна} \\ = 1, & \text{оцена јесте ефикасна.}\end{cases}$$

Задатак 32. Обележје X има густину расподеле $f(x;\theta) = \theta x^{\theta-1}, x \in (0,1], \theta > 0$. За оцену непознатог параметра θ предложена је статистика $V = \frac{n-1}{-\sum\limits_{k=1}^{n}\ln X_k}$. Испитати да ли је та оцена ефикасна.

Решење. Почињемо испитивањем регуларности дате фамилије расподела. Узорачки простор је $(0,1]$, па не зависи од непознатог параметра θ. Желимо да испитамо да ли је $\int\limits_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x;\theta)dx = 0$. Како је

$$\frac{\partial}{\partial\theta}f(x;\theta) = \frac{\partial}{\partial\theta}(\theta x^{\theta-1}) = x^{\theta-1} + \theta x^{\theta-1}\ln x,$$

па је

$$\begin{aligned}\int_0^1 \frac{\partial}{\partial\theta}f(x;\theta)dx &= \int_0^1 x^{\theta-1}dx + \int_0^1 \theta x^{\theta-1}\ln x dx = \begin{vmatrix}\ln x = t \implies x = e^t \\ \frac{dx}{x} = dt\end{vmatrix} \\ &= \frac{x^\theta}{\theta}\Big|_0^1 + \theta\int_{-\infty}^0 e^{\theta t}tdt = \begin{vmatrix} u = t \implies du = dt \\ dv = e^{\theta t} \implies v = \frac{e^{\theta t}}{\theta}\end{vmatrix} \\ &= \frac{1}{\theta} + \theta\left(\frac{te^{\theta t}}{\theta}\Big|_{-\infty}^0 - \int_{-\infty}^0 \frac{e^{\theta t}}{\theta}dt\right) = \frac{1}{\theta} - \frac{1}{\theta} = 0.\end{aligned}$$

Фамилија јесте регуларна. Рачунамо информациону функцију Фишера, да бисмо добили доњу границу дисперзије свих непристрасних оцена:

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

Дакле,

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

па је

$$I(\theta) = E\left(\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}.$$

Према томе, доња граница дисперзије свих непристрасних оцена једнака је

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

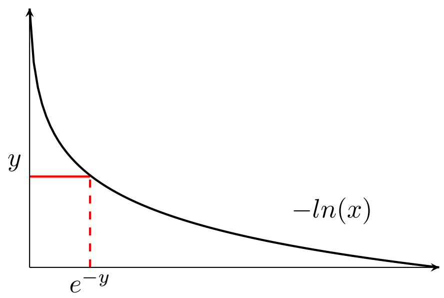
Да би имало смисла да испитујемо ефикасност оцене, морамо прво да проверимо да ли је она непристрасна. Проверавамо да ли је

$$E(V) = E\left(\frac{n-1}{-\sum\limits_{k=1}^{n} \ln X_k}\right) = \theta.$$

Нека је $U = -\sum\limits_{k=1}^{n} \ln X_k$, а $Y_k = -\ln X_k$. Пошто је V реална функција од U, довољно је да одредимо расподелу случајне величине U и онда је $E(V) = E\left(\frac{n-1}{U}\right)$ што ћемо лако израчунати.

За почетак, испитујемо расподелу случајних величина $Y_k = -\ln X_k, k = 1, ..., n$. Приметимо да

$$F_{Y_k}(y) = P\{Y_k \leq y\} = P\{-\ln X_k \leq y\} = (\star)$$



$y < 0$: $(\star) = 0$;

$y \geq 0$:

$$\begin{aligned}(\star) &= P\{e^{-y} \leq X_k \leq 1\}\\ &= \int_{e^{-y}}^{1} \theta x^{\theta-1} dx\\ &= x^\theta \Big|_{e^{-y}}^{1} = 1 - e^{-\theta y}\end{aligned}$$

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

$$U = \sum_{k=1}^{n} Y_k = -\sum_{k=1}^{n} \ln X_k \in \gamma(n, \theta).$$

Враћамо се на очекивање случајне величине V:

$$\begin{aligned}E(V) = E\left(\frac{n-1}{U}\right) = (n-1)E\left(\frac{1}{U}\right) &= (n-1)\int_0^\infty \frac{1}{u}\frac{u^{n-1}\theta^n e^{-\theta u}}{\Gamma(n)} du\\ = (n-1)\int_0^\infty \frac{u^{n-2}\theta^{n-1}e^{-\theta u}}{\Gamma(n-1)} du \cdot \frac{\Gamma(n-1)}{\Gamma(n)}\theta &= \frac{(n-1)(n-2)!}{(n-1)!}\theta = \theta,\end{aligned}$$

Према томе, V јесте непристрасна оцена, па сада рачунамо њену дисперзију, да бисмо је упоредили са доњом границом G и видели да ли је ефикасна. Можемо приметити да

$$\begin{aligned}D(V) = E(V^2) - (E(V))^2 &= E\left(\left(\frac{n-1}{U}\right)^2\right) - \theta^2\\ &= (n-1)^2 E\left(\left(\frac{1}{U}\right)^2\right) - \theta^2\end{aligned}$$

$$
\begin{aligned}
&= (n-1)^2 \int_0^\infty \frac{1}{u^2} \frac{u^{n-1}\theta^n e^{-\theta u}}{\Gamma(n)} du - \theta^2 \\
&= (n-1)^2 \int_0^\infty \frac{u^{n-3}\theta^{n-2} e^{-\theta u}}{\Gamma(n-2)} du \cdot \frac{\Gamma(n-2)}{\Gamma(n)} \theta^2 - \theta^2 \\
&= (n-1)^2 \frac{(n-3)!}{(n-1)!} \theta^2 - \theta^2 = \frac{n-1}{n-2}\theta^2 - \theta^2 = \frac{\theta^2}{n-2}.
\end{aligned}
$$

Дакле,

$$
ef(V) = \frac{G}{DV} = \frac{\frac{\theta^2}{n}}{\frac{\theta^2}{n-2}} = \frac{n-2}{n} < 1,
$$

па ова оцена није ефикасна (али јесте асимптотски ефикасна). ✓

Метод максималне веродостојности

Метод максималне веродостојности представља један начин добијања тачкастих оцена непознатих параметара. Интуитивно, метод максималне веродостојности проналази вредност параметра за коју је највероватније да се добије реализовани узорак који је истраживачу на располагању.

Код метода максималне веродостојности, оцену параметра θ тражимо као вредност за коју се достиже максимум функције веродостојности

$$
L(\theta) = \prod_{k=1}^{n} p(x_k; \theta).
$$

Дакле, функција веродостојности је заправо узорачка расподела, посматрана као функција од параметра θ. Оцена максималне веродостојности је:

$$
\hat{\theta} = \underset{\theta}{argmax} L(\theta).
$$

У случају дискретног обележја:

$$
L(\theta) = \prod_{k=1}^{n} p(x_k; \theta),
$$

а у случају апсолутно непрекидног обележја:

$$
L(\theta) = \prod_{k=1}^{n} f(x_k; \theta).
$$

Због монотоности логаритма и чињенице да логаритам функције и сама функција достижу максимум у истој тачки, оцена максималне веродостојности се може тражити као максимум функције $\ln L(\theta)$, што је у већини случајева једноставније. Дакле, оцену ћемо најчешће тражити као

$$
\hat{\theta} = \underset{\theta}{argmax} \ln L(\theta),
$$

који се углавном налази у стационарним тачкама функције $\ln L(\theta)$. Међутим, ако функција веродостојности није диференцијабилна, морамо наћи алтернативни начин за одређивање максимума функције.

Резултат оцењивања је оцена максималне веродостојности $\hat{\theta}$, што је статистика (случајна величина). За реализовани узорак $(x_1, ..., x_n)$, можемо израчунати реализовану вредност статистике $\hat{\theta}$ на том узорку и добити конкретну вредност. На пример, ако је оцена максималне веродостојности $\overline{X}_n$, на основу реализованог узорка $(1, 2, 3, 4, 5)$ непознати параметар ћемо оценити са $\frac{1+2+3+4+5}{5} = 3$.

Задатак 33. Обележје X има Пуасонову $\mathcal{P}(\lambda)$ расподелу. На основу узорка обима n методом максималне веродостојности одредити оцену непознатог параметра λ, а затим испитати непристрасност и ефикасност тако добијене оцене.

Решење. Закон расподеле обележја X је дат са:

$$P\{X = x\} = \frac{\lambda^x e^{-\lambda}}{x!},\ x \in \{0, 1, ...\}.$$

Претпоставимо да имамо реализовани узорак $(x_1, ..., x_n)$. Функција веродостојности је тада

$$L(\lambda) = \prod_{k=1}^{n} P\{X = x_k\} = \prod_{k=1}^{n} \frac{\lambda^{x_k} e^{-\lambda}}{x_k!} = \frac{e^{-n\lambda}\lambda^{\sum\limits_{k=1}^{n} x_k}}{\prod\limits_{k=1}^{n} x_k!},$$

а њен логаритам

$$\ln L(\lambda) = -n\lambda + \sum_{k=1}^{n} x_k \ln \lambda - \ln\left(\prod_{k=1}^{n} x_k!\right).$$

Желимо да пронађемо вредност λ која максимизује ову функцију. Тражимо стационарну тачку

$$\frac{\partial}{\partial \lambda} \ln L(\lambda) = -n + \frac{\sum\limits_{k=1}^{n} x_k}{\lambda} = 0 \implies \lambda = \frac{\sum\limits_{k=1}^{n} x_k}{n} = \overline{x}_n.$$

Треба проверити да ли је добијено решење заиста максимум: $\frac{\partial}{\partial \lambda} \ln L(\lambda)$ је позитивна за $\lambda < \overline{x}_n$, а негативна за $\lambda > \overline{x}_n$, па функција $\ln L(\lambda)$ расте за $\lambda < \overline{x}_n$, а опада за $\lambda > \overline{x}_n$. Према томе, $\overline{x}_n$ је тачка у којој $\ln L(\lambda)$ достиже максимум. Дакле, оцена параметра λ методом максималне веродостојности на основу реализованог узорка $(x_1, ..., x_n)$ је

$$\hat{\lambda} = \overline{x}_n.$$

Ово је број, конкретна вредност која се рачуна на основу вредности изабраних у узорак. Као случајна величина, оцена методом максималне веродостојности је

$$\hat{\lambda} = \overline{X}_n.$$

Испитујемо непристрасност добијене оцене:

$$E(\hat{\lambda}) = E(\overline{X}_n) = E(X_1) = \lambda.$$

Оцена је непристрасна, па има смисла да испитујемо њену ефикасност. Испитујемо регуларност дате фамилије расподела:

$$\sum_{x=0}^{\infty} \frac{\partial}{\partial \lambda} p(x; \lambda) = 0 \ \ (?)$$

Приметимо да

$$\frac{\partial}{\partial \lambda} p(x; \lambda) = \frac{\partial}{\partial \lambda}\left(\frac{\lambda^x e^{-\lambda}}{x!}\right) = \frac{1}{x!}\left(x\lambda^{x-1}e^{-\lambda} - \lambda^x e^{-\lambda}\right).$$

Односно

$$\sum_{x=0}^{\infty}\frac{\partial}{\partial\lambda}p(x;\lambda)=e^{-\lambda}\left(\sum_{x=1}^{\infty}\frac{x\lambda^{x-1}}{x!}-\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}\right)=e^{-\lambda}\left(\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}-\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}\right)=e^{-\lambda}(e^{\lambda}-e^{\lambda})=0.$$

Дакле, фамилија јесте регуларна. Рачунамо информациону функцију Фишера

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

Како је

$$\begin{aligned}\ln p(x;\lambda)&=\ln\left(\frac{\lambda^x e^{-\lambda}}{x!}\right)=x\ln\lambda-\lambda-\ln x!,\\ \frac{\partial}{\partial\lambda}\ln p(x;\lambda)&=\frac{x}{\lambda}-1,\\ \frac{\partial^2}{\partial\lambda^2}\ln p(x;\lambda)&=-\frac{x}{\lambda^2},\end{aligned}$$

то

$$I(\lambda)=E\left(\frac{X}{\lambda^2}\right)=\frac{\lambda}{\lambda^2}=\frac{1}{\lambda},$$

па је

$$G=\frac{1}{nI(\lambda)}=\frac{\lambda}{n}.$$

Рачунамо дисперзију оцене:

$$D(\hat{\lambda})=D(\overline{X}_n)=\frac{D(X_1)}{n}=\frac{\lambda}{n}.$$

Дакле,

$$ef(\lambda)=\frac{G}{D(\lambda)}=\frac{\frac{\lambda}{n}}{\frac{\lambda}{n}}=1,$$

па закључујемо да је ова оцена ефикасна. ✓

Задатак 34. Обележје X има густину расподеле $f(x;\theta)=\theta^2xe^{-\theta x}, x\geq 0,\theta>0$. На основу узорка обима n методом максималне веродостојности одредити оцену непознатог параметра θ.

Решење. Претпоставимо да имамо реализовани узорак $(x_1,...,x_n)$. Функција веродостојности је

$$L(\theta)=\prod_{k=1}^{n}f(x_k;\theta)=\theta^{2n}e^{-\theta\sum\limits_{k=1}^{n}x_k}\prod_{k=1}^{n}x_k,$$

па је

$$\ln L(\theta)=2n\ln\theta-\theta\sum_{k=1}^{n}x_k+\ln\left(\prod_{k=1}^{n}x_k\right).$$

Желимо да пронађемо $\theta>0$ које максимизује $\ln L(\theta)$. Одређујемо стационрну тачку:

$$\frac{\partial}{\partial\theta}\ln L(\theta)=\frac{2n}{\theta}-\sum_{k=1}^{n}x_k=0\implies\frac{2n}{\theta}=\sum_{k=1}^{n}x_k\implies\theta=\frac{2n}{\sum x_k}=\frac{2}{\overline{x}_n}.$$

Још треба проверити да ли се у добијеној вредности постиже максимум. Како је $\frac{\partial}{\partial\theta}\ln L(\theta)$ позитивна за $\theta < \frac{2}{\overline{x}_n}$, а негативна за $\theta > \frac{2}{\overline{x}_n}$, па је функција $\ln L(\theta)$ растућа за $\theta < \frac{2}{\overline{x}_n}$, а опадајућа за $\theta > \frac{2}{\overline{x}_n}$. Према томе, $\frac{2}{\overline{x}_n}$ је тачка у којој $\ln L(\theta)$ достиже максимум. Дакле, оцена параметра θ методом максималне веродостојности на основу реализованог узорка $(x_1, ..., x_n)$ је

$$\hat{\theta} = \frac{2}{\overline{x}_n}.$$

Када оцену запишемо као случајну величину добија се:

$$\hat{\theta} = \frac{2}{\overline{X}_n}.$$

✓

Задатак 35. Вероватноћа да се догађај A оствари при неком експерименту је $p,\ 0 < p < 1$. Експерименти се независно понављају или до прве појаве догађаја A или до $N-$тог покушаја, где је $N \geq 1$ и унапред познат број. Број експеримената до краја серије се бележи. На основу регистрованих n бројева, тј. на основу n серија, методом максималне веродостојности одредити оцену непознатог параметра p.

Решење. Нека обележје X представља број експеримената до краја серије. Највећи могући број експеримената је N (ако се догађај A није реализовао у првих $N-1$ покушаја, без обзира да ли се реализовао или не у N-том покушају) а најмањи могући број експеримената је 1 (ако се догађај A реализовао у првом покушају). Дакле, случајна величина X узима вредности из скупа $\{1, 2, ..., N\}$. Потребно је још да одредимо вероватноће са којима узима те вредности.

Ако $k \in \{1, ..., N-1\}$, онда догађај $\{X = k\}$ значи да се догађај A реализовао први пут у k-том покушају. Другачије речено, догађај A се није реализовао у првих $k-1$ покушаја, догађај A се реализовао у k-том покушају. Према томе, вероватноћа тог догађаја је

$$P\{X = k\} = (1-p)^{k-1}p.$$

Међутим, ако је $k = N$, онда је

$$P\{X = N\} = (1-p)^{N-1}.$$

Дакле, расподела обележја X одређена је следећим законом:

$$p(x;p) = P\{X = x\} = \begin{cases} (1-p)^{x-1}p, & \text{за } x \in \{1, ..., N-1\} \\ (1-p)^{x-1}, & \text{за } x = N. \end{cases}$$

Претпоставимо да имамо реализовани узорак $(x_1, ..., x_n)$, при чему x_k представља број експеримената у k-тој серији. Да бисмо одредили функцију веродостојности, треба нам краћи запис расподеле обележја X. Приметимо да је

$$p(x;p) = (1-p)^{x-1}p^{1-I\{x=N\}},\ \ x \in \{1, 2, ..., N\}.$$

Функција веродостојности је тада

$$L(p) = \prod_{k=1}^{n} p(x_k;p) = (1-p)^{\sum\limits_{k=1}^{n} x_k - n} \cdot p^{\,n - \sum\limits_{k=1}^{n} I\{x_k = N\}},$$

а њен логаритам

$$\ln L(p) = \left(\sum_{k=1}^{n} x_k - n\right)\ln(1-p) + \left(n - \sum_{k=1}^{n} I\{x_k = N\}\right)\ln p.$$

Желимо да пронађемо p које максимизује $\ln L(p)$. Тражимо стационарну тачку из

$$\frac{\partial \ln L(p)}{\partial p} = -\frac{\sum\limits_{k=1}^{n} x_k - n}{1-p} + \frac{n - \sum\limits_{k=1}^{n} I\{x_k = N\}}{p} = 0,$$

одакле је

$$\begin{aligned}
&\frac{n - \sum\limits_{k=1}^{n} I\{x_k = N\}}{p} = \frac{\sum\limits_{k=1}^{n} x_k - n}{1-p}, \\
\iff\ & n - np - \sum_{k=1}^{n} I\{x_k = N\} + p\sum_{k=1}^{n} I\{x_k = N\} = p\sum_{k=1}^{n} x_k - np, \\
\iff\ & n - \sum_{k=1}^{n} I\{x_k = N\} = p\left(\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} I\{x_k = N\}\right), \\
\implies\ & p = \frac{n - \sum\limits_{k=1}^{n} I\{x_k = N\}}{\sum\limits_{k=1}^{n} x_k - \sum\limits_{k=1}^{n} I\{x_k = N\}}.
\end{aligned}$$

Желимо да проверимо да ли је заиста максимум. Како је $\frac{\partial \ln L(p)}{\partial p} > 0$, ако је $p < \frac{n - \sum_{k=1}^{n} I\{x_k=N\}}{\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} I\{x_k=N\}}$, а $\frac{\partial \ln L(p)}{\partial p} < 0$, ако је $p > \frac{n - \sum_{k=1}^{n} I\{x_k=N\}}{\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} I\{x_k=N\}}$, то функција $\ln L(p)$ расте до $\frac{n - \sum_{k=1}^{n} I\{x_k=N\}}{\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} I\{x_k=N\}}$, а након тога опада, па ово јесте тачка у којој функција $\ln L(p)$ достиже максимум.

Према томе, на основу овог узорка, оцена методом максималне веродостојности је

$$\hat{p} = \frac{n - \sum\limits_{k=1}^{n} I\{x_k = N\}}{\sum\limits_{k=1}^{n} x_k - \sum\limits_{k=1}^{n} I\{x_k = N\}},$$

па је оцена, као случајна величина,

$$\hat{p} = \frac{n - \sum\limits_{k=1}^{n} I\{X_k = N\}}{\sum\limits_{k=1}^{n} X_k - \sum\limits_{k=1}^{n} I\{X_k = N\}}.$$

✓

Задатак 36. Методом максималне веродостојности одредити оцену непознатог параметра θ на основу узорка обима n из популације чије обележје X има:

а) униформну $\mathcal{U}[0,\theta], \theta > 0$, расподелу;

б) униформну $\mathcal{U}[-\theta,\theta], \theta > 0$, расподелу;

в) униформну $\mathcal{U}[0,\theta], \theta \geq 1$, расподелу;

г) униформну $\mathcal{U}[0,\theta], \theta \in \mathbb{N}$, расподелу. У овом случају испитати непристрасност и постојаност тако добијене оцене.

Решење. а) Ако обележје X има униформну $\mathcal{U}[0,\theta], \theta > 0$, расподелу, тада је густина расподеле

$$f(x;\theta) = \frac{1}{\theta} I\{0 \leq x \leq \theta\}.$$

Функција веродостојности је тада

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I\{x_i \leq \theta\} = \frac{1}{\theta^n} \prod_{i=1}^{n} I\{x_i \leq \theta\} = \frac{1}{\theta^n} I\{x_{(n)} \leq \theta\}.$$

Приметимо да функција веродостојности није диференцијабилна по θ, па је потребно на други начин одредити θ за које ће функција веродостојности достизати максимум. Да би функција веродостојности имала максимум, потребно је да индикатор узме вредност 1, у супротном функција веродостојности је 0. Како је функција $\frac{1}{\theta^n}$ опадајућа по θ, треба да изаберемо што мање θ, али такво да је $x_{(n)} \leq \theta$ (да би индикатор узео вредност 1). Према томе, оцена добијена методом максималне веродостојности је

$$\hat{\theta} = X_{(n)}.$$

б) Ако обележје X има униформну $\mathcal{U}[-\theta,\theta], \theta > 0$, расподелу, тада је густина расподеле

$$f(x;\theta) = \frac{1}{2\theta} I\{-\theta \leq x \leq \theta\}.$$

Функција веродостојности је тада

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \frac{1}{2\theta} I\{-\theta \leq x_i \leq \theta\} \\ &= \frac{1}{(2\theta)^n} \prod_{i=1}^{n} I\{-\theta \leq x_i \leq \theta\} \\ &= \frac{1}{(2\theta)^n} I\{-\theta \leq x_{(1)}, x_{(n)} \leq \theta\} \\ &= \frac{1}{(2\theta)^n} I\{\theta \geq x_{(n)}, \theta \geq -x_{(1)}\}. \end{aligned}$$

Приметимо да функција веродостојности није диференцијабилна по θ, па је потребно на други начин одредити θ за које ће функција веродостојности достизати максимум. Да би функција веродостојности имала максимум, потребно је да индикатор узме вредност 1, у супротном функција веродостојности је 0. Како је функција $\frac{1}{(2\theta)^n}$ опадајућа по θ, треба да изаберемо што мање θ, али такво да је $\theta \geq x_{(n)}, \theta \geq -x_{(1)}$ (да би индикатор узео вредност 1). Према томе, оцена добијена методом максималне веродостојности је

$$\hat{\theta} = \max\{X_{(n)}, -X_{(1)}\}.$$

в) Ако обележје X има униформну $\mathcal{U}[0,\theta], \theta \geq 1$, расподелу, тада је густина расподеле

$$f(x;\theta) = \frac{1}{\theta} I\{x \leq \theta,\ \theta \geq 1\}.$$

Функција веродостојности је тада

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I\{x_i \leq \theta, \theta \geq 1\} = \frac{1}{\theta^n} \prod_{i=1}^{n} I\{x_i \leq \theta, \theta \geq 1\} = \frac{1}{\theta^n} I\{x_{(n)} \leq \theta, \theta \geq 1\}.$$

Приметимо да функција веродостојности није диференцијабилна по θ, па је потребно на други начин одредити θ за које ће функција веродостојности достизати максимум. Да би функција веродостојности имала максимум, потребно је да индикатор узме вредност 1, у супротном функција веродостојности је 0. Како је функција $\frac{1}{\theta^n}$ опадајућа по θ, треба да изаберемо што мање θ, али такво да је $\theta \geq x_{(n)}, \theta \geq 1$ (да би индикатор узео вредност 1). Према томе, оцена добијена методом максималне веродостојности је

$$\hat{\theta} = \max\{X_{(n)}, 1\}.$$

г) Ако обележје X има униформну $\mathcal{U}[0,\theta], \theta \in \mathbb{N}$, расподелу, тада је густина расподеле

$$f(x;\theta) = \frac{1}{\theta} I\{x \leq \theta,\ \theta \in \mathbb{N}\}.$$

Функција веродостојности је тада

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I\{x_i \leq \theta, \theta \in \mathbb{N}\} = \frac{1}{\theta^n} \prod_{i=1}^{n} I\{x_i \leq \theta, \theta \in \mathbb{N}\} = \frac{1}{\theta^n} I\{x_{(n)} \leq \theta, \theta \in \mathbb{N}\}.$$

Тражимо θ за које ће функција веродостојности бити највећа. Како је функција $\frac{1}{\theta^n}$ опадајућа по θ, треба да изаберемо што мање θ, али такво да је $\theta \geq x_{(n)}, \theta \in \mathbb{N}$ (да би индикатор узео вредност 1). Према томе, оцена добијена методом максималне веродостојности је

$$\hat{\theta} = \begin{cases} X_{(n)}, & \text{ако је } X_{(n)} \in \mathbb{N} \\ \left[X_{(n)}\right] + 1, & \text{ако је } X_{(n)} \notin \mathbb{N}. \end{cases}$$

Одредимо сада расподелу оцене $\hat{\theta}$.

Ако је $X_{(n)} \in \mathbb{N}$, онда је $\hat{\theta} = k$ исто што и $X_{(n)} = k$, а ако $X_{(n)} \notin \mathbb{N}$, онда $[X_{(n)}] + 1 = k$ је исто што и $[X_{(n)}] = k - 1$, па се $X_{(n)}$ мора наћи у интервалу $(k-1, k)$. То се може објединити на следећи начин:

$$\begin{aligned} P\{\hat{\theta} = k\} &= P\{k - 1 < X_{(n)} \leq k\} \\ &= P\{X_{(n)} \leq k\} - P\{X_{(n)} \leq k - 1\} \\ &= P\{\max_{1 \leq i \leq n} X_i \leq k\} - P\{\max_{1 \leq i \leq n} X_i \leq k - 1\} \\ &= P\{X_1 \leq k, \ldots, X_n \leq k\} - P\{X_1 \leq k - 1, \ldots, X_n \leq k - 1\} \\ &= \prod_{i=1}^{n} P\{X_i \leq k\} - \prod_{i=1}^{n} P\{X_i \leq k - 1\} \\ &= \left(\frac{k}{\theta}\right)^n - \left(\frac{k-1}{\theta}\right)^n = \frac{1}{\theta^n}(k^n - (k-1)^n), \quad k = 1, 2, \ldots, \theta. \end{aligned}$$

Испитајмо непристрасност ове оцене:

$$\begin{aligned}
E(\hat{\theta}) &= \sum_{k=1}^{\theta} kP\{\hat{\theta}=k\} \\
&= \sum_{k=1}^{\theta} k\left(\frac{1}{\theta}\right)^n (k^n-(k-1)^n) \\
&= \left(\frac{1}{\theta}\right)^n \left(\sum_{k=1}^{\theta} k^{n+1} - \sum_{k=1}^{\theta} k(k-1)^n\right) \\
&= \left(\frac{1}{\theta}\right)^n \left(\theta^{n+1} + \sum_{k=1}^{\theta-1} k^{n+1} - \sum_{k=1}^{\theta-1} (k+1)k^n\right) \\
&= \left(\frac{1}{\theta}\right)^n \left(\theta^{n+1} + \sum_{k=1}^{\theta-1} k^{n+1} - \sum_{k=1}^{\theta-1} k^{n+1} - \sum_{k=1}^{\theta-1} k^n\right) \\
&= \left(\frac{1}{\theta}\right)^n (\theta^{n+1}-(\theta-1)^n-(\theta-2)^n-\cdots-1^n) \\
&= \theta - \frac{1^n+\ldots(\theta-1)^n}{\theta^n} \\
&\neq \theta.
\end{aligned}$$

Закључујемо да оцена није непристрасна, али јесте асимптотски непристрасна.

Испитајмо још постојаност оцене $\hat{\theta}$:

$$\begin{aligned}
(\forall\epsilon>0) \quad P\{|\hat{\theta}-\theta|\geq\epsilon\} &= P\{\theta-\hat{\theta}\geq\epsilon\} \\
&= P\{\hat{\theta}\leq\theta-\epsilon\} \\
&\leq P\{\hat{\theta}\leq\theta-1\} \\
&= 1-P\{\hat{\theta}=\theta\} \\
&= 1-\left(1-\frac{(\theta-1)^n}{\theta^n}\right) \to 0.
\end{aligned}$$

Добили смо да $(\forall\epsilon>0) \quad P\{|\hat{\theta}-\theta|\geq\epsilon\}\to 0$, па је оцена $\hat{\theta}$ постојана.

✓

Задатак 37. Обележје X има експоненцијалну $\mathcal{E}(\lambda)$ расподелу, где је λ непознати параметар. На основу узорка $(2.3, 3.4, 1.2, 2.5, 0.6)$ методом максималне веродостојности одредити оцену за вероватноћу $P\{X\geq 1\}$.

Решење. Вероватноћа коју треба да оценимо је

$$P\{X\geq 1\} = 1-P\{X<1\} = 1-(1-e^{-\lambda}) = e^{-\lambda}.$$

Приметимо да ако одредимо оцену параметра λ, заменом у овај израз одредићемо оцену тражене вероватноће.

Одредимо оцену непознатог параметра λ методом максималне веродостојности. Претпоставимо да имамо реализовани узорак $(x_1,...,x_n)$. Функција веродостојности једнака је

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda\sum\limits_{i=1}^{n} x_i}.$$

Логаритмовањем добијамо

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

Желимо да нађемо вредност λ која максимизује $\ln L(\lambda)$:

$$\frac{\partial}{\partial \lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0,$$

па је оцена параметра λ добијена методом максималне веростојности

$$\hat{\lambda} = \frac{1}{\overline{X}_n}.$$

Оцена тражене вероватноће једнака је $e^{-\hat{\lambda}} = e^{-\frac{1}{\overline{X}_n}}$.

На основу реализованог узорка из задатка добијамо да је оцена вероватноће $P\{X \geq 1\}$ једнака $e^{-\frac{1}{2}}$. ✓

Задатак 38. Обележје X има густину расподеле $f(x;\theta_1,\theta_2) = \frac{1}{\theta_2} e^{-\frac{x-\theta_1}{\theta_2}}, x \geq \theta_1, \theta_1 > 0, \theta_2 > 0$. На основу узорка обима n методом максималне веродостојности одредити оцене непознатих параметара θ_1 и θ_2.

Решење. Претпоставимо да имамо реализовани узорак $(x_1, ..., x_n)$. Функција веродостојности је тада

$$L(\theta_1,\theta_2) = \prod_{i=1}^{n} \frac{1}{\theta_2} e^{-\frac{x_i-\theta_1}{\theta_2}} I\{x_i \geq \theta_1\} = \frac{1}{(\theta_2)^n} e^{-\frac{\sum_{i=1}^{n} x_i - n\theta_1}{\theta_2}} I\{x_{(1)} \geq \theta_1\}.$$

Приметимо да функција веродостојности није диференцијабилна по θ_1. Индикатор ће узети вредност 1 ако је $x_{(1)} \geq \theta_1$. Функција веродостојности је растућа по θ_1, па закључујемо да је оцена параметра θ_1 добијена методом максималне веродостојности

$$\hat{\theta}_1 = X_{(1)}.$$

Сада, кад смо одредили оцену параметра θ_1, фиксирамо параметар θ_1 у функцији веродостојности и посматрамо је као функцију од θ_2

$$L(\theta_2) = \frac{1}{(\theta_2)^n} e^{\frac{1}{\theta_2}(n\hat{\theta}_1 - \sum_{i=1}^{n} x_i)}.$$

Логаритмовањем добијамо

$$\ln L(\theta_2) = -n \ln \theta_2 + \frac{1}{\theta_2}\left(n\hat{\theta}_1 - \sum_{i=1}^{n} x_i\right).$$

Желимо да нађемо вредност θ_2 која максимизује $\ln L(\theta_2)$:

$$\frac{\partial}{\partial \theta_2} \ln L(\theta_2) = -\frac{n}{\theta_2} - \frac{1}{\theta_2^2}\left(n\hat{\theta}_1 - \sum_{i=1}^{n} x_i\right) = 0,$$

па је оцена паараметра θ_2 добијена методом максималне веродостојности

$$\hat{\theta}_2 = \overline{X}_n - X_{(1)}.$$

✓

Интервали поверења

До сада смо видели како се добијају тачкасте оцене непознатих параметара. Тачкаста оцена као резултат даје једну тачку, међутим није познато колико та вредност одступа од стварне вредности параметра. Са друге стране, може се одредити интервална оцена, односно интервал вредности, која са унапред задатком тачношћу садржи непознати параметар.

Нека је θ непознати параметар у расподели обележја X и нека је $(X_1, ..., X_n)$ прост случајан узорак обима n за посматрано обележје. Нека су U_n и V_n статистике дефинисане на основу узорка такве да је $P\{U_n \leq V_n\} = 1$ за које важи

$$P\{U_n \leq \theta \leq V_n\} = \beta.$$

Интервал $[U_n, V_n]$ се назива $\beta\%$ двострани интервал поверења параметра θ, а β је ниво поверења.

Задатак 39. Из популације чије обележје X има нормалну $\mathcal{N}(m, \sigma^2)$ расподелу извучен је узорак обима 10, чија је узорачка средина $\bar{x}_{10} = 5.5$, а узорачка дисперзија $\bar{s}^2_{10} = 36$.

а) Одредити 90% интервал поверења за непознати параметар m.

б) Одредити 90% једнострани (доњи, горњи) и двострани интервал поверења за непознати параметар σ^2, као и за непознати параметар σ.

Решење.

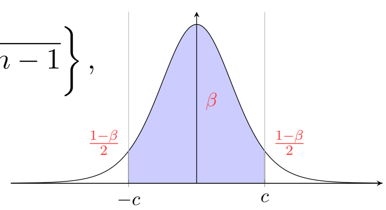
а) Приметимо прво да је σ^2 непознато, па ћемо користити случајну величину $T_n = \frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1}$, која има Студентову t_{n-1} расподелу.

Потребно је да одредимо U_n и V_n такве да важи

$$\begin{aligned} & P\{U_n \leq m \leq V_n\} = \beta \\ \iff & P\left\{\overline{X}_n - V_n \leq \overline{X}_n - m \leq \overline{X}_n - U_n\right\} = \beta \\ \iff & P\left\{\frac{\overline{X}_n - V_n}{\overline{S}_n}\sqrt{n-1} \leq \frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1} \leq \frac{\overline{X}_n - U_n}{\overline{S}_n}\sqrt{n-1}\right\} = \beta \end{aligned}$$

Једино што знамо у овом случају је да је вероватноћа између $\frac{\overline{X}_n - V_n}{\overline{S}_n}\sqrt{n-1}$ и $\frac{\overline{X}_n - U_n}{\overline{S}_n}\sqrt{n-1}$ једнака β, међутим њихов положај може бити произвољан. Због тога се, у случају да није другачије наглашено, увек узимају интервали такви да

$$P\left\{T_n < \frac{\overline{X}_n - V_n}{\overline{S}_n}\sqrt{n-1}\right\} = P\left\{T_n > \frac{\overline{X}_n - U_n}{\overline{S}_n}\sqrt{n-1}\right\},$$



а пошто је Студентова расподела симетрична око 0 то важи да су ове вредности једнаке по апсолутној вредности.

Користећи да T_n има Студентову расподелу добијамо да је

$$\frac{\overline{X}_n - V_n}{\overline{S}_n}\sqrt{n-1} = -t_{n-1,1-\beta},$$

$$\frac{\overline{X}_n - U_n}{\overline{S}_n}\sqrt{n-1} = t_{n-1,1-\beta}.$$

Према томе,

$$U_n = \overline{X}_n - \frac{\overline{S}_n}{\sqrt{n-1}}t_{n-1,1-\beta},$$
$$V_n = \overline{X}_n + \frac{\overline{S}_n}{\sqrt{n-1}}t_{n-1,1-\beta}.$$

Односно интервал поверања је

$$I_m = \left[\overline{X}_n - \frac{\overline{S}_n}{\sqrt{n-1}}t_{n-1,1-\beta}, \overline{X}_n + \frac{\overline{S}_n}{\sqrt{n-1}}t_{n-1,1-\beta}\right].$$

У конкретном случају, $\beta = 0.9$ и $t_{9,0.1} = 1.833$. Тражени интервал је

$$I_m = \left[5.5 - 1.833 \cdot \frac{6}{3}, 5.5 + 1.833 \cdot \frac{6}{3}\right] = [1.834, 9.166].$$

б) Прво приметимо да је m непознато, па ћемо користити случајну величину $\frac{n\overline{S}_n^2}{\sigma^2}$, која има χ^2_{n-1} расподелу.

Једнострани доњи интервал $(0, V_n]$: Потребно је да одредимо V_n такво да важи

$$P\{\sigma^2 \leq V_n\} = \beta \iff P\left\{\frac{1}{V_n} \leq \frac{1}{\sigma^2}\right\} = \beta$$
$$\iff P\left\{\frac{n\overline{S}_n^2}{V_n} \leq \frac{n\overline{S}_n^2}{\sigma^2}\right\} = \beta.$$

Из таблица за χ^2 расподелу добијамо да је $\frac{n\overline{S}_n^2}{V_n} = \chi^2_{n-1;\beta}$. Према томе $V_n = \frac{n\overline{S}_n^2}{\chi^2_{n-1;\beta}}$. Интервал поверења непознатог параметра σ^2 је

$$I_{\sigma^2} = \left(0, \frac{n\overline{S}_n^2}{\chi^2_{n-1;\beta}}\right].$$

У конкретном случају, $\beta = 0.9$ и $\chi^2_{9;0.9} = 4.168$, па је тражени интервал $I_{\sigma^2} = (0, 86.3724]$. Интервал поверења непознатог параметра σ се добија узимањем корена из леве и десне границе претходног интервала па се добија $I_\sigma = (0, 9.2937]$.

Једнострани горњи интервал $[U_n, +\infty)$: Потребно је да одредимо U_n такво да важи $P\{\sigma^2 \geq U_n\} = \beta$, односно

$$P\left\{\frac{n\overline{S}_n^2}{U_n} \geq \frac{n\overline{S}_n^2}{\sigma^2}\right\} = \beta.$$

Из таблица за χ^2 расподелу добијамо да је $\frac{n\overline{S}_n^2}{U_n} = \chi^2_{n-1;1-\beta}$. Према томе $U_n = \frac{n\overline{S}_n^2}{\chi^2_{n-1;1-\beta}}$. Интервал поверења непознатог параметра σ^2 је

$$I_{\sigma^2} = \left[\frac{n\overline{S}_n^2}{\chi^2_{n-1;1-\beta}}, \infty\right).$$

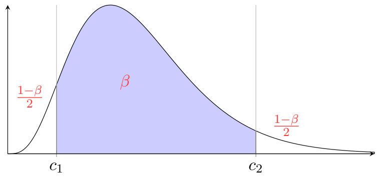
У конкретном случају, $\beta = 0.9$ и $\chi^2_{n-1;1-\beta} = 14.684$, па је тражени интервал поверења $I_{\sigma^2} = [24.5165, +\infty)$. Интервал поверења непознатог параметра σ је $I_\sigma = [4.9514, +\infty)$.

Двострани интервал $[U'_n, V'_n]$: Потребно је да одредимо U'_n и V'_n такве да важи $P\{U'_n \leq \sigma^2 \leq V'_n\} = \beta$, односно

$$P\left\{\frac{n\overline{S}_n^2}{V'_n} \leq \frac{n\overline{S}_n^2}{\sigma^2} \leq \frac{n\overline{S}_n^2}{U'_n}\right\} = \beta.$$

Као и у делу а), интервал ћемо бирати тако да

$$P\left\{\frac{n\overline{S}_n^2}{\sigma^2} > \frac{n\overline{S}_n^2}{U'_n}\right\} = P\left\{\frac{n\overline{S}_n^2}{\sigma^2} < \frac{n\overline{S}_n^2}{V'_n}\right\} = \frac{1-\beta}{2}.$$



Из таблица за χ^2 расподелу добијамо да је $\frac{n\overline{S}_n^2}{U'_n} = \chi^2_{n-1;\frac{1-\beta}{2}}$ и $\frac{n\overline{S}_n^2}{V'_n} = \chi^2_{n-1;\frac{1+\beta}{2}}$. Према томе $U'_n = \frac{n\overline{S}_n^2}{\chi^2_{n-1;\frac{1-\beta}{2}}}$ и $V'_n = \frac{n\overline{S}_n^2}{\chi^2_{n-1;\frac{1+\beta}{2}}}$. Интервал поверења непознатог параметра σ^2 је

$$I_{\sigma^2} = \left[\frac{n\overline{S}_n^2}{\chi^2_{n-1;\frac{1-\beta}{2}}}, \frac{n\overline{S}_n^2}{\chi^2_{n-1;\frac{1+\beta}{2}}}\right].$$

У конкретном случају, $\beta = 0.9$ и $\chi^2_{9;0.05} = 16.919$ и $\chi^2_{9;0.95} = 3.325$. Интервал поверења је $I_{\sigma^2} = [21.2778, 108.2707]$, а интервал поверења непознатог параметра σ је $I_\sigma = [4.6128, 10.4053]$.

✓

Задатак 40**.** Из популације чије обележје X има нормалну $\mathcal{N}(m, \sigma^2)$ расподелу извучен је узорак:

(-5, -3]	(-3, -1]	(-1, -0.5]	(-0.5, 0.5]	(0.5, 1.5]	(1.5, 3.5)
3	13	56	100	60	8

На основу тог узорка добијено је да је $(c, 0.13756)$ 95.8%-ни интервал поверења за непознати параметар m. Израчунати реалан број c.

Решење. Како је σ^2 непознато, користићемо случајну величину $\frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1}$, 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 распоређени 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:

$$\overline{x}_{240} = \frac{3 \cdot (-4) + 13 \cdot (-2) + 56 \cdot (-0.75) + 100 \cdot 0 + 60 \cdot 1 + 8 \cdot 2.5}{240} = 0,$$

$$\overline{s}_{240}^2 = \frac{3 \cdot (-4)^2 + 13 \cdot (-2)^2 + 56 \cdot 0.75^2 + 60 \cdot 1^2 + 8 \cdot 2.5^2}{240} = 1.00625.$$

Одредимо $\beta\%-$ни интервал поверења за m у општем случају. Потребно је да одредимо U_n и V_n такве да важи

$$\begin{aligned}
& P\{U_n \leq m \leq V_n\} = \beta \\
\Longleftrightarrow\ & P\left\{\overline{X}_n - V_n \leq \overline{X}_n - m \leq \overline{X}_n - U_n\right\} = \beta \\
\Longleftrightarrow\ & P\left\{\frac{\overline{X}_n - V_n}{\overline{S}_n}\sqrt{n-1} \leq \frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1} \leq \frac{\overline{X}_n - U_n}{\overline{S}_n}\sqrt{n-1}\right\} = \beta \\
\Longleftrightarrow\ & F_{t_{n-1}}\left(\frac{\overline{X}_n - U_n}{\overline{S}_n}\sqrt{n-1}\right) - F_{t_{n-1}}\left(\frac{\overline{X}_n - V_n}{\overline{S}_n}\sqrt{n-1}\right) = \beta.
\end{aligned}$$

Како је n довољно велико ($n > 30$) одговарајуће вредности се не могу прочитати из таблица Студентове расподеле. У том случају, Студентову расподелу можемо апроксимирати стандардном нормалном расподелом. Односно, горњи израз је еквивалентан са

$$\Phi\left(\frac{\overline{X}_n - U_n}{\overline{S}_n}\sqrt{n-1}\right) - \Phi\left(\frac{\overline{X}_n - V_n}{\overline{S}_n}\sqrt{n-1}\right) = \beta.$$

Како је дата лева граница полазног интеграла, не можемо претпоставити да је у питању симетричан интервал. Према томе, остале прорачуне морамо радити за конкретан узорак.

Из таблица за нормалну расподелу и за реализовану вредност узорка, имамо да је $\Phi(\frac{-0.13756}{\sqrt{1.00625}}\sqrt{239}) = \Phi(-2.12) = 0.017$. Према томе

$$\Phi\left(\frac{\overline{x}_n - c}{\overline{s}_n}\sqrt{n-1}\right) \approx 0.958 + 0.017 = 0.975.$$

Даље из таблица за нормалну расподелу добијамо да је

$$\frac{\overline{x}_n - c}{\overline{s}_n}\sqrt{n-1} = \Phi^{-1}(0.975) = 1.96,$$

па добијамо да је $c = -0.12718$.

✓

Задатак 41**.** Из популације чије обележје X има нормалну $\mathcal{N}(m, 16)$ расподелу извучен је узорак обима 64, чија је узорачка средина $\bar{x}_{64} = 5$. Константовано је да је $\beta\%$ интервал поверења за m једнак (4,6). Израчунати β.

Решење. Пошто је дисперзија расподеле позната, можемо користити случајну величину $X^\star = \frac{\overline{X}_n - m}{\sigma}\sqrt{n}$ која има стандардну нормалну расподелу. На основу тога добија се

$$\begin{aligned}
\beta &= P\{U_n < m < V_n\} \\
\beta &= P\left\{\overline{X}_n - V_n < \overline{X}_n - m < \overline{X}_n - U_n\right\} \\
\beta &= P\left\{\frac{\overline{X}_n - V_n}{\sigma}\sqrt{n} < \frac{\overline{X}_n - m}{\sigma}\sqrt{n} < \frac{\overline{X}_n - U_n}{\sigma}\sqrt{n}\right\}.
\end{aligned}$$

За реализовани узорак, следи

$$\beta = P\left\{\frac{5-6}{4}\sqrt{64} < X^\star < \frac{5-4}{4}\sqrt{64}\right\} = \Phi(2) - \Phi(-2)$$
$$= \Phi(2) - (1-\Phi(2)) = 2\Phi(2) - 1 = 2(0.5 + 0.47725) - 1 = 0.9545.$$

Дакле, $\beta = 0.9545$, односно 95.45%. ✓

Задатак 42**.** Обележје X има униформну $\mathcal{U}[0, 1+\theta], \theta > 0$, расподелу, где је θ непознати параметар. Узет је узорак обима 200 и константовано је да је у узорку 150 елемената који су мањи од 1.

а) Одредити 95% интервал поверења за вероватноћу p, где је $p = P\{X < 1\}$.

б) На основу резултата под (а) одредити 95% интервал поверења за θ.

Решење. а) Уведимо случајну величину Y (индикатор) која узима вредност 1 ако случајна величина X узме вредност мању од један, односно

$$Y : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}.$$

Нас интересује укупан број елемената који је задовољио наведени услов, односно сума претходних индикатора који баш говоре да ли је услов задовољен или не.

Како је узет узорак обима 200 (> 30) можемо да искористимо централну граничну теорему:

$$\frac{\sum\limits_{i=1}^{n} Y_i - E\left(\sum\limits_{i=1}^{n} Y_i\right)}{\sqrt{D\left(\sum\limits_{i=1}^{n} Y_i\right)}} \xrightarrow{D} Y^*, \quad Y^* \in \mathcal{N}(0,1).$$

С обзиром да је

$$E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E(Y_i) = np$$

и

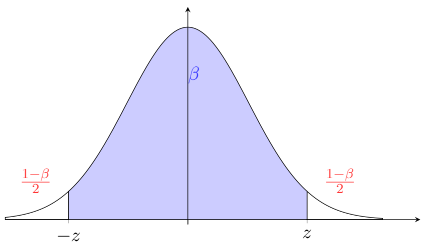
$$D\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} D(Y_i) = np(1-p),$$

интервал поверења за вероватноћу p добијамо из:

$$P\left\{\left|\frac{\sum\limits_{i=1}^{n} Y_i - 200p}{\sqrt{200p(1-p))}}\right| < z\right\} = 0.95.$$

$$P\left\{-z < \frac{\sum\limits_{i=1}^{n} Y_i - 200p}{\sqrt{200p(1-p))}} < z\right\} = 0.95.$$

Из задатка је познато да је $\sum\limits_{i=1}^{200} y_i = 150$.



Из таблица за нормалну расподелу имамо да је $z = \Phi^{-1}(0.95 + \frac{1-0.95}{2}) = 1.96$, па решавањем неједначине

$$\frac{(150-200p)^2}{200p(1-p)} < 1.96^2,$$

добијамо да је $I_p = (0.686, 0.805)$.

б) Користећи део а) и из једнакости $p = P\{X < 1\} = \frac{1}{1+\theta}$ имамо да је

$$\begin{aligned} &0.686 < p < 0.805 \\ &0.686 < \frac{1}{1+\theta} < 0.805 \\ &0.242 < \theta < 0.458 \end{aligned}$$

тј. $I_\theta = (0.242, 0.458)$.

✓

Задатак 43. Из популације чије обележје X има експоненцијалну $\mathcal{E}(\lambda)$ расподелу извучен је узорак:

I_k	[0,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	$[6,\infty)$
n_k	493	378	298	211	171	45	4

Одредити 98% интервал поверења за непознати параметар λ.

Решење. Узет је узорак обима $n = \sum n_k = 1600$ (>30), па можемо да искористимо централну граничну теорему:

$$\frac{\sum\limits_{i=1}^{n} X_i - E\left(\sum\limits_{i=1}^{n} X_i\right)}{\sqrt{D\left(\sum\limits_{i=1}^{n} X_i\right)}} \xrightarrow{D} X^*, \quad X^* \in \mathcal{N}(0,1).$$

Користећи податке из задатка можемо одредити збир вредности из узорака

$$\sum_{k=1}^{n} x_k = 493 \cdot 0.5 + 378 \cdot 1.5 + 298 \cdot 2.5 + 211 \cdot 3.5 + 171 \cdot 4.5 + 45 \cdot 5.5 + 4 \cdot 6.5 = 3340.$$

С обзиром да је $EX_1 = \frac{1}{\lambda}$ и $DX_1 = \frac{1}{\lambda^2}$, интервал поверења за параметар λ добијамо из једнакости:

$$P\left\{\left|\frac{\sum\limits_{i=1}^{1600} X_i - \frac{1600}{\lambda}}{\sqrt{\frac{1600}{\lambda^2}}}\right| < z\right\} = 0.98.$$

Из таблица за нормалну расподелу имамо да је $z = \Phi^{-1}(0.99) = 2.33$. Решавањем неједначине

$$-2.33 < \frac{3340 - \frac{1600}{\lambda}}{\frac{40}{\lambda}} < 2.33,$$

добијамо да је $I_\lambda = (0.451, 0.507)$.

✓

Тестирање статистичких хипотеза

Тестирање статистичких хипотеза је поступак којим се, применом статистичких метода, утврђује да ли се на основу узорка може, и са којом вероватноћом, прихватити претпоставка о конкретној вредности неког параметра или о расподели обележја. На основу тога да ли је тестирање везано за параметар расподеле или за саму расподелу обележја разликујемо параметарске и непараметарске хипотезе, редом. Ми ћемо се углавном бавити параметарским хипотезама.

Нека је расподела обележја X одређена параметром $\theta \in \Theta = \Theta_0 \cup \Theta_1$. На основу узорка $\mathbf{X}_n = (X_1, ..., X_n)$ хоћемо да одредимо у којем од скупова Θ_0 или Θ_1 се налази непознати параметар θ. Хипотеза да $\theta \in \Theta_0$ назива се **нултом хипотезом** и означава са H_0, док се хипотеза да $\theta \in \Theta_1$ назива **алтернативном хипотезом** и означава са H_1. За хипотезу се каже да је **проста** ако се односи на једну вредност параметра којом је расподела обележја одређена, а **сложена** ако садржи читав скуп вредности.

Тест је у потпуности одређен критичном облашћу. **Критична област** теста (или област одбацивања нулте хипотезе) W је део узорачког простора који указује на Θ_1. Ако реализовани узорак $(x_1, ..., x_n) \in W$, онда се одбацује нулта хипотеза, у супротном се не одбацује. Сама критична област од узорка зависи кроз статистику која се назива **тест статистика**.

Постоје две врсте грешака које можемо направити приликом закључивања:

- Грешка I врсте - када је $\theta \in \Theta_0$, а $(x_1, ..., x_n) \in W$, односно када одбацимо нулту хипотезу ($(x_1, ..., x_n) \in W$), а тачна је (јер $\theta \in \Theta_0$).
- Грешка II врсте - када је $\theta \in \Theta_1$, а $(x_1, ..., X_n) \notin W$, односно када не одбацимо нулту хипотезу ($(x_1, ..., x_n) \notin W$), а нетачна је (јер $\theta \in \Theta_1$).

За сваку од ових врста грешака можемо рачунати вероватноћу

- Вероватноћа грешке I врсте:

$$\alpha = P_{H_0}\{\mathbf{X}_n \in W\}.$$

- Вероватноћа грешке II врсте:

$$\beta = P_{H_1}\{\mathbf{X}_n \notin W\}.$$

Најбоље би било да и грешка I врсте и грешка II врсте буду што мање могуће. Међутим, пошто то није могуће, уобичајен приступ је да се приликом тестирања контролише вероватноћа грешке I врсте тако што се захтева да

$$P_{H_0}\{\mathbf{X}_n \in W\} \leq \alpha_0, \tag{3}$$

за неки фиксиран **праг значајности** $\alpha_0 > 0$, обично реда величине $\alpha_0 \in \{0.05, 0.01, 0.001\}$. Праг (ниво) значајности је, дакле, највећа дозвољена вероватноћа одбацивања тачне нулте хипотезе. Циљ тестирања је затим да нађемо критичну област такву да, за дати праг значајности α_0, задовољава услов (3), а да

$$P_{H_1}\{\mathbf{X}_n \in W\} = 1 - \beta$$

буде што веће. Другим речима, вероватноћа грешке I врсте не сме прећи дати ниво значајности, а вероватноћа грешке II врсте треба да буде што мања могућа.

У вези са тим, за дату критичну област W дефинишемо **функцију моћи теста**

$$M(\theta) = P_\theta\{\mathbf{X}_n \in W\}.$$

Очекујемо, дакле, да буде $M(\theta) \leq \alpha_0$ за $\theta \in \Theta_0$ и $M(\theta)$ што веће за $\theta \in \Theta_1$. Вероватноћа одређена функцијом моћи теста $M(\theta)$ за $\theta \in \Theta_1$, односно

$$\gamma = P_{\theta_1}\{\mathbf{X}_n \in W\},\ \theta_1 \in \Theta_1$$

назива се **моћ теста.**

Супремум $\sup\limits_{\theta\in\Theta_0} M(\theta)$ називамо мером теста и она представља највећу вероватноћу грешке I врсте. Увек је $\sup\limits_{\theta\in\Theta_0} M(\theta) \leq \alpha_0$, а најчешће је $\sup\limits_{\theta\in\Theta_0} M(\theta) = \alpha_0$, односно мера и праг значајности теста се поклапају.

Задатак 44**.** У кутији је 10 куглица, црвених и белих. Тестира се (нулта) хипотеза H_0(у кутији су 2 црвене и 8 белих куглица) против (алтернативне) хипотезе H_1(у кутији су више од 2 црвене куглице), тако што се из кутије извлаче две куглице једна за другом, без враћања, па ако су обе извучене куглице црвене, хипотеза H_0 се одбацује, иначе се не одбацује. Израчунати вероватноћу грешке прве врсте и функцију моћи тог теста.

Решење. Ако са θ означимо број црвених куглица у кутији, онда се нулта и алтернативна хипотеза могу написати као

$$H_0(\theta = 2) \qquad H_1(\theta > 2).$$

Пошто је критична област у ствари област одбацивања нулте хипотезе, из текста задатка можемо видети да се H_0 одбацује ако су обе извучене куглице црвене. Односно, критична област W је скуп свих узорака у којима су извучене две црвене куглице. Вероватноћа грешке прве врсте је према томе

$$\alpha = P_{H_0}\{\mathbf{X}_n \in W\} = \frac{2}{10}\frac{1}{9} = \frac{1}{45},$$

при чему је коришћено да је у питању извлачење две куглице једне за другом без враћања. Функција моћи теста је

$$M(\theta) = P_\theta\{\mathbf{X}_n \in W\} = \frac{\theta}{10}\frac{\theta-1}{9} = \frac{\theta(\theta-1)}{90}.$$

✓

Задатак 45**.** Нека је хипотеза H_0(обележје X има густину расподеле $f_0(x) = 1,\ 0 \leq x \leq 1$), а хипотеза H_1(обележје X има густину расподеле $f_1(x) = 2x,\ 0 \leq x \leq 1$). На основу узорка (X_1, X_2) треба се определити за једну од ове две хипотезе. Предложена су два теста чије су критичне области $W_1 = \{(x_1, x_2)|x_1 \geq k_1, x_2 \geq k_1\}$, односно $W_2 = \{(x_1, x_2)|x_1 + x_2 \geq k_2\}$, оба са истим прагом значајности α, где је $\alpha = \frac{1}{8}$. Испитати који је тест бољи. Подразумева се да је познато да обележје X узима вредности из сегмента $[0, 1]$.

Решење. Наведене хипотезе, иако су дате као непараметарске, могу се преформулисати у параметарске на следећи начин: претпоставимо да обележје има фамилију густина расподеле $f(x, \theta) = (2x)^\theta,\ x \in [0, 1],\ \theta \in \{0, 1\}$, и да желимо да тестирамо хипотезу $H_0(\theta = 0)$ против алтернативе $H_1(\theta = 1)$. У наставку можемо користити све што је наведено за параметарске хипотезе.

- Хоћемо да праг значајности првог теста буде $\frac{1}{8}$. Дакле треба да буде

$$\begin{aligned} &P_{H_0}\{(X_1, X_2) \in W_1\} = \frac{1}{8} \\ \Longleftrightarrow\ &P_{H_0}\{X_1 \geq k_1, X_2 \geq k_1\} = \frac{1}{8} \\ \Longleftrightarrow\ &\Big(P_{H_0}\{X_1 \geq k_1\}\Big)^2 = \frac{1}{8} \end{aligned}$$

Како је

$$P_{H_0}\{X_1 \geq k_1\} = \int_{k_1}^{1} (2x)^0 dx = 1 - k_1,$$

треба да буде

$$(1 - k_1)^2 = \frac{1}{8},$$

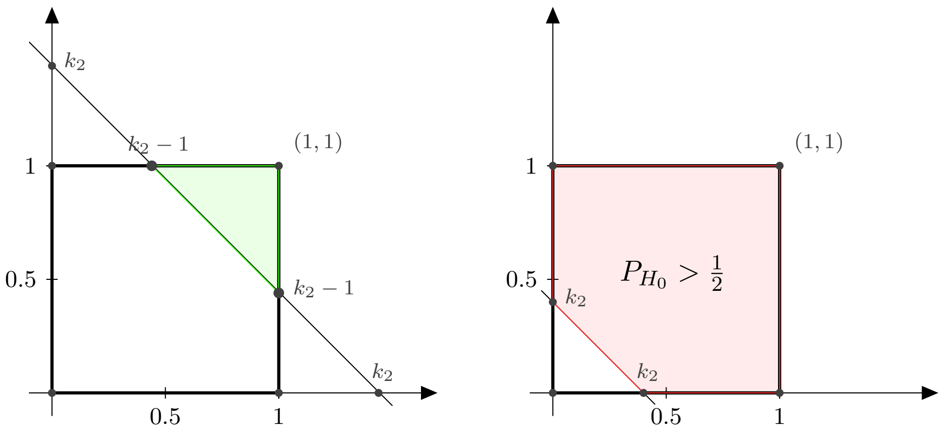
односно

$$k_1 = \frac{4 - \sqrt{2}}{4} \approx 0.65.$$

- Да би праг значајности другог теста био $\frac{1}{8}$ треба да је

$$P_{H_0}\{X_1 + X_2 \geq k_2\} = P_{H_0}\{X_1 \geq k_2 - X_2\} = \frac{1}{8}.$$

Разликујемо две области интеграције за рачунање ове вероватноће - за $k_2 > 1$ и $k_2 < 1$, илустроване на слици. Примећујемо да је за црвену област вероватноћа под нултом хипотезом (униформна расподела) већа од $\frac{1}{2}$ за све $k_2 < 1$, па у том случају не може праг значајности да буде $\frac{1}{8}$, те тај случај одбацујемо.



(а) Може

(б) Не може

Слика 1: Области интеграције за $k_2 > 1$ и $k_2 < 1$

Остаје нам да се фокусирамо само на случај када је $k_2 > 1$ и израчунамо вероватноћу

$$\begin{aligned} P_{H_0}\{X_1 \geq k_2 - X_2\} &= \int_{k_2-1}^{1} \int_{k_2-x_2}^{1} 1 \cdot 1 dx_1 dx_2 = \int_{k_2-1}^{1} (1 - k_2 + x_2) dx_2 \\ &= (1 - k_2)(1 - k_2 + 1) + \frac{1}{2}(1 - (k_2 - 1)^2) = \frac{1}{2}(k_2^2 - 4k_2 + 4). \end{aligned}$$

Изједначавањем овог израза са $\frac{1}{8}$ добијамо

$$\begin{aligned}(k_2-2)^2 &= \frac{1}{4}\\ 2-k_2 &= \frac{1}{2} \quad (\text{јер } k_2 \le 2)\\ k_2 &= \frac{3}{2}.\end{aligned}$$

Дакле, критичне области тестова су

$$W_1 = \{x_1 \ge 0.65, x_2 \ge 0.65\} \text{ и } W_2 = \{x_1 + x_2 \ge 1.5\}.$$

Моћи ових тестова су

$$\begin{aligned}\gamma_1 &= P_{H_1}\{X_1 \ge 0.65, X_2 \ge 0.65\} = (P_{H_1}\{X_1 \ge 0.65\})^2 = \left(\int_{0.65}^{1} 2x dx\right)^2\\ &= (1-0.65^2)^2 = 0.3335\end{aligned}$$

и

$$\begin{aligned}\gamma_2 &= P_{H_1}\{X_1 + X_2 \ge 1.5\} = \int_{0.5}^{1}\int_{1.5-x_2}^{1} 2x_1 \cdot 2x_2 dx_1 dx_2 = \int_{0.5}^{1} 2x_2 \left(\int_{1.5-x_2}^{1} 2x_1 dx_1\right) dx_2\\ &= \int_{0.5}^{1} 2x_2(-x_2^2 + 3x_2 - 1.25)dx_2 = 2\Big(-\frac{x_2^4}{4} + 3\frac{x_2^3}{3} - 0.625x_2^2\Big)\Big|_{0.5}^{1} = 0.34375\end{aligned}$$

Како је $\gamma_2 > \gamma_1$, закључујемо да је други тест бољи, односно моћнији.

✓

Задатак 46. Нека је хипотеза $H_0\left(X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10\\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{pmatrix}\right)$, а хипотеза $H_1\left(X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10\\ 0.01 & 0.42 & 0.01 & 0.01 & 0.5 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \end{pmatrix}\right)$. Са прагом значајности 0.03, одредити најбољу критичну област за тестирање хипотезе H_0 против хипотезе H_1 на основу узорка обима 2. Израчунати вероватноћу грешке друге врсте тако одабраног теста.

Решење. Праг значајности је $\alpha = 0.03$, имамо узорак $\mathbf{X}_2 = (X_1, X_2)$ и тражимо најбољу критичну област W. Због прага значајности, потребно је да важи

$$P_{H_0}\{(X_1, X_2) \in W\} \le 0.03. \tag{4}$$

За произвољне $k_1, k_2 \in \{1, \dots, 10\}$ је

$$P_{H_0}\{(X_1, X_2) = (k_1, k_2)\} = 0.1^2 = 0.01.$$

Критична област W ће бити неки скуп облика $\{(k_{i1}, k_{i2}) \mid i\}$, па ће важити

$$P_{H_0}\{(X_1, X_2) \in W\} = |W| \cdot 0.01.$$

Дакле, вероватноћа под нултом хипотезом да $\mathbf{X}_2$ упадне у критичну област W је сразмерна броју парова у W. Да бисмо имали што већу моћ, желимо да W има што више чланова, али због услова (4), највише можемо имати 3 пара у W.

Критична област ће стога бити облика $W = \{(k_{11}, k_{12}), (k_{21}, k_{22}), (k_{31}, k_{32})\}$.

Моћ теста је

$$P_{H_1}\{(X_1, X_2) \in W\} = P_{H_1}\{X_1 = k_{11}, X_2 = k_{12}\} + P_{H_1}\{X_1 = k_{21}, X_2 = k_{22}\} \\ + P_{H_1}\{X_1 = k_{31}, X_2 = k_{32}\},$$

а овај збир је највећи када је критична област баш

$$W = \{(5,5), (5,2), (2,5)\},$$

јер је под алтернативном хипотезом вероватноћа за вредности 2 и 5 највећа.

Дакле, моћ овог теста је

$$\gamma = P_{H_1}\{(X_1, X_2) \in W\} = 0.5^2 + 2 \cdot 0.42 \cdot 0.5 = 0.67,$$

па је вероватноћа грешке II врсте $\beta = 1 - \gamma = 0.33$. ✓

Нејман–Пирсонова лема и униформно најмоћнији тест

Уколико није унапред позната критична област, потребан нам је начин за одређивање W. Један могући начин је коришћењем Нејман-Пирсонове леме која, при одређеним условима, даје најбољу критичну област, а самим тим и најбољи тест за тестирање одговарајућих хипотеза.

Лема 2 (Нејман-Пирсон). *Ако се тестира хипотеза $H_0(\theta = \theta_0)$ против алтернативе $H_1(\theta = \theta_1)$ тада постоји најмоћнији тест са критичном облашћу W која је мере α, односно*

$$P_{H_0}\{\mathbf{X}_n \in W\} = \alpha,$$

за коју важи да

$$W = \{L(\theta_1) \geq cL(\theta_0)\},$$

за одговарајуће $c > 0$, где је $L(\theta)$ функција веродостојности. ⋆

Ако је $L(\theta) > 0$ за свако θ, онда се критична област може написати као

$$W = \left\{ \frac{L(\theta_1)}{L(\theta_0)} \geq c \right\}$$

Тест је **униформно најмоћнији** (УНМ) за тестирање $H_0(\theta \in \Theta_0)$ против $H_1(\theta \in \Theta_1)$ ако је најмоћнији за тестирање свих хипотеза облика $H_0'(\theta = \theta_0)$ против $H_1'(\theta = \theta_1)$ за сваки избор $\theta_0 \in \Theta_0$ и $\theta_1 \in \Theta_1$.

Када се тестирају хипотезе облика $H_0(\theta = \theta_0)$ против $H_1(\theta \in \Theta_1)$, најбољи тест за тестирање $H_0(\theta = \theta_0)$ против $H_1'(\theta = \theta_1)$, $\theta_1 \in \Theta_1$ (добијен из леме Нејман–Пирсона), је уједно и униформно најмоћнији тест уколико његова критична област не зависи од вредности параметра θ_1 из алтернативе.

Задатак 47. Обележје X има нормалну $\mathcal{N}(0, \sigma^2)$ расподелу, где је σ^2 непознати параметар. Са прагом значајности 0.05, одредити најбољу критичну област за тестирање хипотезе $H_0(\sigma^2 = 1)$ против хипотезе $H_1(\sigma^2 = 2)$ на основу узорка обима 10.

Решење. Пошто треба тестирати просту нулту хипотезу против просте алтернативне хипотезе можемо за одређивање најбоље критичне области, а самим тим и најбољег теста, користити Нејман-Пирсонову лему. За одређивање критичне области, на основу узорка обима $n = 10$, потребна нам је функција веродостојности која је у овом случају

$$L(\sigma^2) = \prod_{i=1}^{n} f(x_i;\sigma^2) = \Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big)^n e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n} x_i^2}.$$

Користећи лему и то да је $L(\theta) > 0$ за свако θ, критична област је

$$W = \left\{\frac{L(2)}{L(1)} \geq k\right\} = \left\{\frac{\Big(\frac{1}{\sqrt{2\pi\cdot 2}}\Big)^n e^{-\frac{1}{2\cdot 2}\sum\limits_{i=1}^{n} x_i^2}}{\Big(\frac{1}{\sqrt{2\pi}}\Big)^n e^{-\frac{1}{2}\sum\limits_{i=1}^{n} x_i^2}} \geq k\right\} = \left\{\frac{1}{2^n} e^{\frac{1}{4}\sum\limits_{i=1}^{n} x_i^2} \geq k\right\}$$

$$= \left\{e^{\frac{1}{4}\sum\limits_{i=1}^{n} x_i^2} \geq k_1\right\} = \left\{\frac{1}{4}\sum_{i=1}^{n} x_i^2 \geq k_2\right\} = \left\{\sum_{i=1}^{n} x_i^2 \geq c\right\}.$$

Дакле, добијена критична област је одређена реализованом вредношћу статистике $\sum\limits_{i=1}^{n} X_i^2$. На основу датог прага значајности, можемо одредити c тако да важи

$$P_{H_0}\left\{\sum_{i=1}^{n} X_i^2 \geq c\right\} = 0.05. \tag{5}$$

Како при H_0, $X_i \in \mathcal{N}(0,1)$, познато је да $X_i^2 \in \chi_1^2$. Пошто је χ^2 раподела адитивна, следи да $\sum\limits_{i=1}^{n} X_i^2 \in \chi_n^2$. Користећи таблице χ^2 расподеле, $n = 10$ и (5), добија се да је

$$c = \chi^2_{10;0.05} = 18.307.$$

Дакле, најбоља критична област за тестирање хипотезе $H_0(\sigma^2 = 1)$ против хипотезе $H_1(\sigma^2 = 2)$ на основу узорка обима 10 је

$$W = \left\{\sum_{i=1}^{n} x_i^2 \geq 18.307\right\}.$$

✓

Задатак 48**.** Из популације чије обележје X има експоненцијалну $\mathcal{E}(\frac{1}{\lambda})$ расподелу извучен је узорак обима n.

а) Одредити најбољу критичну област за тестирање хипотезе $H_0(\lambda = 1)$ против хипотезе $H_1(\lambda \neq 1)$ $(H_1(\lambda = \lambda_1), \lambda_1 \neq 1)$.

б) Испитати да ли је тај тест униформно најмоћнији за тестирање хипотезе $H_0(\lambda = 1)$ против хипотезе $H_1(\lambda \neq 1)$.

в) Израчунати вероватноћу грешке друге врсте тог теста ако је обим узорка 100, алтернативна хипотеза $H_1(\lambda = 2)$, а праг значајности 0.05.

Решење. а) Критичну област ћемо тражити помоћу леме Нејман-Пирсона. Да би лема могла да се примени обе хипотезе морају бити просте. Нулта хипотеза свакако јесте проста, међутим иницијално наведена алтернативна хипотеза $H_1(\lambda \neq 1)$ није. Ако алтернативну хипотезу напишемо као $H_1(\lambda = \lambda_1)$, $\lambda_1 \neq 1$, онда она јесте проста хипотеза јер смо параметру доделили тачно једну вредност

(λ_1). За Нејман-Пирсонову лему је потребно одредити функцију веродостојности која је у овом случају

$$L(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \frac{1}{\lambda^n} e^{-\frac{1}{\lambda}\sum\limits_{i=1}^{n} x_i}.$$

Користећи лему и то да је $L(\lambda) > 0$ за свако λ, најбоља критична област је одређена са

$$W = \left\{ \frac{L(\lambda_1)}{L(1)} \geq k_1 \right\} = \left\{ \frac{\frac{1}{\lambda_1^n} e^{-\frac{1}{\lambda_1}\sum\limits_{i=1}^{n} x_i}}{e^{-\sum\limits_{i=1}^{n} x_i}} \geq k_1 \right\} = \left\{ e^{-(\frac{1}{\lambda_1}-1)\sum\limits_{i=1}^{n} x_i} \geq k_2 \right\}$$

$$= \left\{ -\Big(\frac{1}{\lambda_1} - 1\Big) \sum_{i=1}^{n} x_i \geq k_3 \right\} = \begin{cases} \left\{ \sum\limits_{i=1}^{n} x_i \geq c \right\}, & \lambda_1 > 1 \\ \left\{ \sum\limits_{i=1}^{n} x_i \leq c \right\}, & \lambda_1 < 1 \end{cases}.$$

Дакле, добијена критична област је одређена реализованом вредношћу статистике $\sum\limits_{i=1}^{n} X_i$. На основу датог прага значајности, у сваком од претходна два случаја, можемо одредити c тако да важи

$$P_{H_0}\left\{ \sum_{i=1}^{n} X_i \geq c \right\} = \alpha,$$

у првом случају, односно

$$P_{H_0}\left\{ \sum_{i=1}^{n} X_i \leq c \right\} = \alpha,$$

у другом случају. Како $X_i \in \mathcal{E}(\frac{1}{\lambda})$, познато је да $\sum\limits_{i=1}^{n} X_i \in \gamma(n, \frac{1}{\lambda})$. Односно, користећи везу између гама и χ^2 расподеле, важи $\frac{2}{\lambda}\sum\limits_{i=1}^{n} X_i \in \chi^2_{2n}$. Дакле, при $H_0(\lambda = 1)$, $2\sum\limits_{i=1}^{n} X_i \in \chi^2_{2n}$. Према томе, добија се да је

$$P_{H_0}\left\{ 2\sum_{i=1}^{n} X_i \geq 2c \right\} = \alpha,$$

и $c = \frac{1}{2}\chi^2_{2n;\alpha}$ у првом случају, односно

$$P_{H_0}\left\{ 2\sum_{i=1}^{n} X_i \leq 2c \right\} = \alpha,$$

и $c = \frac{1}{2}\chi^2_{2n;1-\alpha}$ у другом случају.

Дакле, најбоља критична област за тестирање хипотезе $H_0(\lambda = 1)$ против хипотезе $H_1(\lambda = \lambda_1)$, $\lambda_1 \neq 1$, на основу узорка обима n је

$$W = \begin{cases} \left\{ \sum\limits_{i=1}^{n} x_i \geq \frac{1}{2}\chi^2_{2n;\alpha} \right\}, & \lambda_1 > 1 \\ \left\{ \sum\limits_{i=1}^{n} x_i \leq \frac{1}{2}\chi^2_{2n;1-\alpha} \right\}, & \lambda_1 < 1 \end{cases}.$$

б) Видимо да критична област у овом случају зависи од вредности λ_1 из алтернативе, па овај тест није униформно најмоћнији.

в) За $\lambda_1 = 2$, $\alpha = 0.05$ и $n = 100$ вредност константе c и моћ теста можемо тражити користећи централну граничну теорему, будући да је n довољно велико.

Како при H_0 знамо да $\sum\limits_{i=1}^{n} X_i \in \gamma(n, 1)$, закључујемо да ће при H_0 бити $E\Big(\sum\limits_{i=1}^{n} X_i\Big) = n = 100$ и $D\Big(\sum\limits_{i=1}^{n} X_i\Big) = n = 100$.

Критична област за ову алтернативу је $W = \Big\{\sum\limits_{i=1}^{n} x_i \geq c\Big\}$, јер је $\lambda_1 = 2 > 1$.

Прво ћемо одредити вредност c решавањем једначине

$$P_{H_0}\left\{\sum_{i=1}^{n} X_i \geq c\right\} = 0.05,$$

односно

$$0.95 = P_{H_0}\left\{\frac{\sum\limits_{i=1}^{n} X_i - 100}{10} < \frac{c-100}{10}\right\} \approx \Phi\left(\frac{c-100}{10}\right),$$

одакле добијамо

$$\frac{c-100}{10} = \Phi^{-1}(0.95) \approx 1.64,$$

односно $c = 116.4$. Критична област теста нам је дакле $W = \Big\{\sum\limits_{i=1}^{n} x_i \geq 116.4\Big\}$.

При $H_1(\lambda = 2)$ знамо да $\sum\limits_{i=1}^{n} X_i \in \gamma(n, \frac{1}{2})$, па је $E\Big(\sum\limits_{i=1}^{n} X_i\Big) = 200$ и $D\Big(\sum\limits_{i=1}^{n} X_i\Big) = 400$.

Вероватноћу грешке друге врсте рачунамо по формули

$$\beta = P_{H_1}\left\{\sum_{i=1}^{n} X_i < 116.4\right\} = P_{H_1}\left\{\frac{\sum\limits_{i=1}^{n} X_i - 200}{20} < \frac{116.4-200}{20}\right\} = \Phi(-4.1775) \approx 0.$$

✓

Задатак 49**.** Обележје X има густину расподеле $f(x;\theta) = \frac{1-\theta}{x^\theta}, x \in (0,1)$, где је θ непознати параметар такав да је $\theta \in [0,1)$.

а) Одредити најбољу критичну област за тестирање хипотезе $H_0(\theta = 0)$ против хипотезе $H_1(\theta = \theta_0), \theta_0 > 0$, на основу узорка обима n.

б) Испитати да ли је тај тест униформно најмоћнији тест за тестирање хипотезе $H_0(\theta = 0)$ против хипотезе $H_1(\theta > 0)$.

в) Одредити функцију моћи $M(\theta)$ тог теста ако је обим узорка 2, а праг значајности α.

Решење. а) Критичну област ћемо тражити помоћу леме Нејман-Пирсона. Да би лема могла да се примени обе хипотезе морају бити просте, што овде и јесте случај. За Нејман-Пирсонову лему је потребно одредити функцију веродостојности која је у овом случају:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \frac{(1-\theta)^n}{\prod\limits_{i=1}^{n} x_i^{\theta}}.$$

Користећи лему, најбоља критична област је одређена са:

$$W = \left\{ \frac{L(\theta_0)}{L(0)} \geq k \right\} = \left\{ \frac{(1-\theta_0)^n}{\prod\limits_{i=1}^{n} x_i^{\theta_0}} \geq k \right\} = \left\{ \left(\prod_{i=1}^{n} x_i \right)^{\theta_0} \leq k_1 \right\} = \left\{ \prod_{i=1}^{n} x_i \leq c \right\}.$$

Дакле, добијена критична област је одређена реализованом вредношћу статистике $\prod\limits_{i=1}^{n} X_i$. На основу прага значајности α, требало би одредити c тако да важи

$$P_{H_0}\left\{ \prod_{i=1}^{n} X_i \geq c \right\} = \alpha.$$

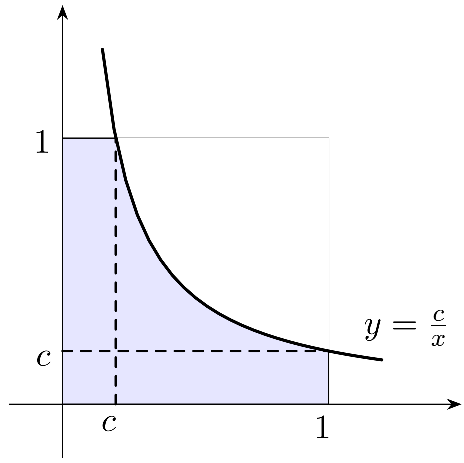
С обзиром да, при H_0, $X_i \in \mathcal{U}(0,1)$, није позната расподела статистика $\prod\limits_{i=1}^{n} X_i$ у општем случају. Према томе, константа c би могла да се одреди неким нумеричким методама за конкретно дате n и α.

б) Видимо да критична област у овом случају не зависи од вредности θ_0 из алтернативе хипотезе, па је овај тест униформно најмоћнији.

в) За $n = 2$ и α, критична област је $W = \{x_1 x_2 \leq c\}$, а функција моћи теста је облика $M(\theta) = P_\theta\{X_1 X_2 \leq c\}$. Остаје још да одредимо c за дато фиксирано α.

Како је α праг значајности, важи једнакост:

$$\begin{aligned}
\alpha &= P_{H_0}\{X_1 X_2 \leq c\} \\
&= \int\int_{xy \leq c} f(x;0) f(y;0)\, dx dy \\
&= 1 - \int_c^1 \int_{\frac{c}{x}}^1 dy dx \\
&= c(1 - \ln c)
\end{aligned}$$



Праг значајности α можемо посматрати као функцију по c: $\alpha(c) = c(1 - \ln c)$. Како је

- α непрекидна функција по c,
- $\alpha'(c) = -\ln c > 0$ (јер је $c \in (0,1)$), тј. функција $\alpha(c)$ је растућа на интервалу $(0,1)$,
- $\alpha(c) \to 0, c \to 0$ и $\alpha(c) \to 1, c \to 1$,

можемо закључити да за свако $\alpha \in (0,1)$ постоји јединствено c такво да важи једнакост $c(1-\ln c) = \alpha$, која се може решити нумерички.

Функција моћи теста је:

$$
\begin{aligned}
M(\theta) &= P_\theta\{X_1X_2 \le c\} \\
&= \int\int_{xy\le c} \frac{(1-\theta)^2}{x^\theta y^\theta} dxdy \\
&= 1 - \int\int_{xy>c} \frac{(1-\theta)^2}{x^\theta y^\theta} dxdy \\
&= 1 - \int_c^1 \int_{\frac{c}{y}}^1 \frac{(1-\theta)^2}{x^\theta y^\theta} dxdy \\
&= 1 - \int_c^1 \left(\frac{1-\theta}{y^\theta} - \frac{1-\theta}{y} c^{1-\theta}\right) dy \\
&= c^{1-\theta}(1-(1-\theta)\ln c).
\end{aligned}
$$

✓

Задатак 50. На основу узорка обима n тестирати хипотезу H_0(обележје X има нормалну $\mathcal{N}(0,1)$ расподелу) против хипотезе H_1(обележје X има густину расподеле $f(x) = \frac{1}{2}e^{-|x|}$). Одредити најбољу критичну област ако је обим узорка 100, а праг значајности 0.05.

Решење. Хипотезе H_0 и H_1 су непараметарске, па ћемо од њих да направимо параметарске увођењем фиктивног параметра. Означимо $f(x;\theta_0) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, а $f(x;\theta_1) = \frac{1}{2}e^{-|x|}$. Сада можемо тестирати $H_0(\theta=\theta_0)$ против $H_1(\theta=\theta_1)$.

Овако дефинисане хипотезе су параметарске и просте, па ћемо критичну област одредити помоћу леме Нејман-Пирсона:

$$
\begin{aligned}
W = \left\{\frac{L(\theta_1)}{L(\theta_0)} \ge k\right\} &= \left\{\frac{\frac{1}{2^n}e^{-\sum\limits_{i=1}^n |x_i|}}{\frac{1}{(\sqrt{2\pi})^n}e^{-\frac{1}{2}\sum\limits_{i=1}^n x_i^2}} \ge k\right\} \\
&= \left\{e^{\frac{1}{2}\sum\limits_{i=1}^n x_i^2 - \sum\limits_{i=1}^n |x_i|} \ge k_1\right\} = \left\{\sum_{i=1}^n (|x_i|-1)^2 \ge c\right\}.
\end{aligned}
$$

Праг значајности је 0.05, па важи једнакост:

$$
P_{H_0}\left\{\sum_{i=1}^{100}(|X_i|-1)^2 \ge c\right\} = 0.05
$$

Уведимо ознаку $Y_i = (|X_i|-1)^2$. Расподела случајних величина Y_i нам није позната, али не морамо ни да је одређујемо, јер је $n = 100$, па можемо применити централну граничну теорему:

$$
\frac{\sum\limits_{i=1}^{100} Y_i - E\left(\sum\limits_{i=1}^{100} Y_i\right)}{\sqrt{D\left(\sum\limits_{i=1}^{100} Y_i\right)}} \xrightarrow{D} Y^\star \in \mathcal{N}(0,1).
$$

Одредимо очекивање и дисперзију случајне величине Y_i при H_0.

$$\begin{aligned}
EY_i &= E(|X_i|-1)^2 = EX_i^2 - 2E|X_i| + 1 = 1 - 2\int_{-\infty}^{+\infty}|x|\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx + 1\\
&= 2 - \frac{4}{\sqrt{2\pi}}\int_0^{+\infty} xe^{-\frac{x^2}{2}}dx = \left|\tfrac{x^2}{2}=t,\ xdx=dt\right|\\
&= 2 - \frac{4}{\sqrt{2\pi}}\int_0^{+\infty} e^{-t}dt = 2 - \frac{4}{\sqrt{2\pi}},\\
EY_i^2 &= EX_i^4 - 4E|X_i|^3 + 6E|X_i|^2 - 4E|X_i| + 1\\
&= \int_{-\infty}^{+\infty} x^4\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx - 8\int_0^{+\infty} x^3\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx + 6 - 4\frac{2}{\sqrt{2\pi}} + 1\\
&= \frac{2}{\sqrt{2\pi}}\int_0^{+\infty}(\sqrt{2t})^4e^{-t}dt - \frac{1}{\sqrt{2\pi}}\int_0^{+\infty}2te^{-t}dt + 7 - 4\frac{2}{\sqrt{2\pi}}\\
&= \frac{2}{\sqrt{2\pi}}2\sqrt{2}\Gamma\left(\frac{5}{2}\right) - \frac{16}{\sqrt{2\pi}}\Gamma(2) + 7 - \frac{8}{\sqrt{2\pi}}\\
&= \frac{4}{\sqrt{\pi}}\frac{3}{4}\sqrt{\pi} - \frac{16}{\sqrt{2\pi}} + 7 - \frac{8}{\sqrt{2\pi}} = 10 - \frac{24}{\sqrt{2\pi}},
\end{aligned}$$

при чему је коришћено да је $\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{4}\Gamma\left(\frac{1}{2}\right) = \frac{3}{4}\sqrt{\pi}$. Дакле,

$$DY_i = E(|X_i|-1)^4 - (E(|X_i|-1))^2 = 6 - \frac{8}{\sqrt{2\pi}} - \frac{8}{\pi}$$

Уврстимо вредности за очекивање и дисперзију и добијамо да је:

$$P_{H_0}\left\{\frac{\sum\limits_{i=1}^{100}(|X_i|-1)^2 - E\left(\sum\limits_{i=1}^{100}(|X_i|-1)^2\right)}{D\left(\sum\limits_{i=1}^{100}(|X_i|-1)^2\right)} \geq \frac{c - 100\left(2-\frac{4}{\sqrt{2\pi}}\right)}{\sqrt{100\left(6-\frac{8}{\sqrt{2\pi}}-\frac{8}{\pi}\right)}}\right\} = 0.05.$$

Из таблица за нормалну расподелу добијамо да је $\frac{c-40.42}{5.12} = \Phi^{-1}(0.95) = 1.645$, па је $c = 48.8424$.

Дакле, тражена критична област је

$$W = \left\{\sum_{i=1}^{100}(|x_i|-1)^2 \geq 48.8424\right\}.$$

✓

Пошто за обележје из нормалне расподеле постоје статистике које се могу повезати са параметром тако да се добије случајна величина за коју знамо расподелу, често (као што је био случај и код интервала поверења) користимо њихова својства да бисмо формирали тестове. У наредним задацима користићемо критичне области које се добијају на основу тих случајних величина чије расподеле су познате када узорак потиче из нормалне расподеле и тада важи $W=\{T_n<c\}$ или $W=\{T_n>c\}$ или $W=\{|T_n|>c\}$, а знак у критичној области зависи од алтернативне хипотезе и слагаће се са знаком у њој.

Задатак 51**.** Из популације чије обележје X има нормалну $\mathcal{N}(m,\sigma^2)$ расподелу извучен је узорак обима 20 и констатовано је да је узорачка средина $\bar{x}_{20} = 53.5$, а узорачка дисперзија $\bar{s}_{20}^2 = 45.85$. Са прагом значајности 0.05 тестирати:

а) хипотезу $H_0(m = 60)$;

б) хипотезу $H_0(\sigma^2 = 50)$ против хипотезе $H_1(\sigma^2 < 50)$.

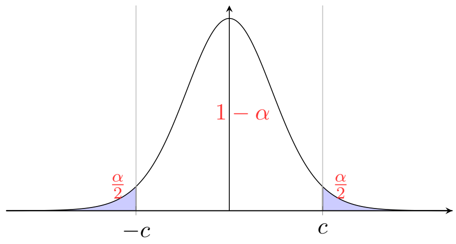
Решење. а) Тестирамо хипотезу $H_0(m = 60)$ против алтернативне $H_1(m \neq 60)$. У задатку није наведена алтернативна хипотеза, па се подразумева она која обухвата све оне вредности параметарског простора које се не налазе у нултој хипотези. За овако дефинисану алтернативну хипотезу, с обзиром да дисперзија није позната, критична област је облика

$$W = \left\{ \left| \frac{\overline{x}_n - m_0}{\overline{s}_n} \sqrt{n-1} \right| \geq c \right\},$$

где је m_0 вредност из нулте хипотезе, а константу c одређујемо тако да важи $P_{H_0}\{\mathbf{X}_n \in W\} = \alpha$.

Односно,

$$P_{H_0} \left\{ \left| \frac{\overline{X}_n - m_0}{\overline{S}_n} \sqrt{n-1} \right| \geq c \right\} = \alpha.$$



Како, при нултој хипотези, $\frac{\overline{X}_n - m_0}{\overline{S}_n} \sqrt{n-1}$ има Студентову t_{n-1} расподелу, из таблица те расподеле добијамо да је

$$c = t_{n-1;\alpha} = t_{19;0.05} = 2.093.$$

Дакле, критична област је

$$W = \left\{ \left| \frac{\overline{x}_n - 60}{\overline{s}_n} \sqrt{n-1} \right| \geq 2.093 \right\}.$$

На основу реализованог узорка имамо да је $\frac{\overline{x}_n - 60}{\overline{s}_n} \sqrt{n-1} = 4.18428$. Приметимо да је $4.18428 \geq 2.093$, што значи да се реализовани узорак налази у критичној области, па одбацујемо H_0.

Напомена:

Уколико је алтернативна хиотеза $H_1(m > 60)$, критична област је облика

$$W = \left\{ \frac{\overline{x}_n - m_0}{\overline{s}_n} \sqrt{n-1} > c \right\}.$$

Уколико је алтернативна хипотеза $H_1(m < 60)$, критична област је облика

$$W = \left\{ \frac{\overline{x}_n - m_0}{\overline{s}_n} \sqrt{n-1} < c \right\}.$$

б) Тестирамо хипотезу $H_0(\sigma^2 = 50)$ против алтернативне $H_1(\sigma^2 < 50)$. За овако дефинисану алтернативну хипотезу критична област је облика

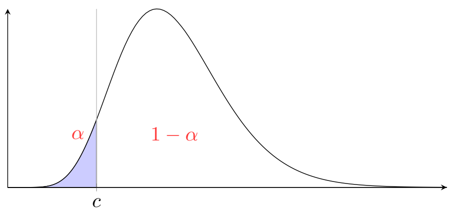
$$W = \left\{ \frac{n\overline{s}_n^2}{\sigma_0^2} \leq c \right\},$$

где је σ_0^2 вредност из нулте хипотезе, а константа c се одређује тако да

$$P_{H_0}\{\mathbf{X}_n \in W\} = \alpha.$$

Односно

$$P_{H_0}\left\{\frac{n\overline{S}_n^2}{\sigma^2}\le c\right\}=\alpha.$$



Познато је да $\frac{n\overline{S}_n^2}{\sigma^2}$ има χ^2_{n-1} расподелу, па из таблица за χ^2 расподелу добијамо да је $c=\chi^2_{n-1;1-\alpha}=\chi^2_{19;0.95}=10.12$.

Дакле критична област је

$$W=\left\{\frac{n\overline{s}_n^2}{\sigma_0^2}\le 10.12\right\}.$$

На основу података из задатка имамо да је $\frac{n\overline{s}_n^2}{\sigma_0^2}=18.34>10.12$, па не одбацујемо H_0, јер се реализовани узорак не налази у критичној области.

✓

Задатак 52. Из две популације чија су обележја X које има нормалну $\mathcal{N}(m_1,6^2)$ расподелу и Y које има нормалну $\mathcal{N}(m_2,5^2)$ расподелу извучени су независни узорци обима 12 и 10 и констатовано је да су узорачке средине $\bar{x}_{12}=178$ и $\bar{y}_{10}=176.6$. Са прагом значајности 0.04 тестирати хипотезу $H_0(m_1{=}m_2)$ против хипотезе $H_1(m_1{>}m_2)$.

Решење. Како је потребно да тестирамо једнакост очекивања два обележја $H_0(m_1=m_2)$ против алтернативне $H_1(m_1>m_2)$, посматраћемо критичну област облика

$$W=\{\overline{x}_{12}-\overline{y}_{10}\ge c\}.$$

Дакле, желимо да одбацимо нулту хипотезу о једнакости средњих вредности обележја X и Y, ако се узорачке средине разликују за више од неке унапред одређене вредности.

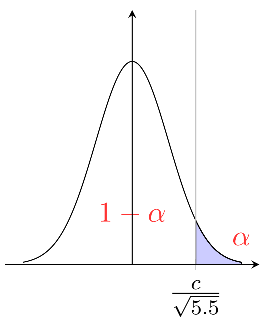
Како су случајне величине $\overline{X}_{12}$ и $\overline{Y}_{10}$ независне и $\overline{X}_{12}\in\mathcal{N}(m_1,\frac{6^2}{12})$, $\overline{Y}_{10}\in\mathcal{N}(m_2,\frac{5^2}{10})$, то $\overline{X}_{12}-\overline{Y}_{10}\in\mathcal{N}(m_1-m_2,5.5)$.

Формирамо критичну област тако да важи $P_{H_0}\{\mathbf{X}_n\in W\}=\alpha$. Праг значајности је 0.04, па важи једнакост

$$0.04=P_{H_0}\left\{\overline{X}_{12}-\overline{Y}_{10}\ge c\right\}=P_{H_0}\left\{\frac{\overline{X}_{12}-\overline{Y}_{10}}{\sqrt{5.5}}\ge\frac{c}{\sqrt{5.5}}\right\}.$$

С обзиром да статистика на левој страни неједнакости има стандардну нормалну расподелу, из таблица за нормалну расподелу добијамо да је

$$\frac{c}{\sqrt{5.5}}=\Phi^{-1}(0.96)=\Phi_0^{-1}(0.46)=1.75,$$



односно $c=4.104$. Критична област је

$$W=\{\overline{x}_{12}-\overline{y}_{10}\ge 4.104\}.$$

На основу података из задатка имамо да је $\overline{x}_{12}-\overline{y}_{10}=1.4<4.104$, па реализовани узорци нису у критичној области и не одбацујемо нулту хипотезу. ✓

Задатак 53. Мерен је систолни (горњи) притисак на узорку од 12 мушкараца и добијено је 130, 148, 122, 140, 132, 142, 124, 150, 170, 136, 146, 140, а на узорку од 13 жена добијене су следеће вредности 140, 150, 130, 132, 150, 138, 123, 124, 160, 138, 170, 144, 108. Сматра се да систолни притисак и код мушкараца и код жена има нормалну расподелу. Ако се претпостави да су дисперзије једнаке, са прагом значајности $\alpha = 0.1$, тестирати хипотезу да су средње вредности притисака мушкараца и жена једнаке против алтернативе да се разликују.

Решење. Нека је X обележје које представља систолни притисак код мушкараца, а Y обележје које представља систолни притисак код жена. Претпоставља се да обележја X и Y имају нормалну расподелу са једнаким дисперзијама. Дакле, претпостављамо да

$$X \in \mathcal{N}(m_1, \sigma^2) \qquad \text{и} \qquad Y \in \mathcal{N}(m_2, \sigma^2).$$

На основу добијених (независних) узорака обима $n_1 = 12$ и $n_2 = 13$, желимо да тестирамо хипотезу да је средња вредност обележја X једнака средњој вредности обележја Y, против алтернативе да се разликују. Односно, тестирамо $H_0(m_1 = m_2)$ против $H_1(m_1 \neq m_2)$.

Није нам позната вредност дисперзије σ^2, али знамо да:

$$\frac{(\overline{X}_{n_1} - m_1) - (\overline{Y}_{n_2} - m_2)}{\sqrt{n_1 \overline{S}^2_{n_1}(X) + n_2 \overline{S}^2_{n_2}(Y)}} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}(n_1 + n_2 - 2)} \in t_{n_1+n_2-2},$$

па можемо да користимо тест статистику:

$$T = \frac{\overline{X}_{12} - \overline{Y}_{13}}{\sqrt{12\overline{S}^2_{12}(X) + 13\overline{S}^2_{13}(Y)}} \sqrt{\frac{12 \cdot 13}{12 + 13}(12 + 13 - 2)},$$

која, при H_0, има Студентову t_{23} расподелу.

Пошто је расподела статистике T симетрична око нуле, формирамо критичну област на следећи начин:

$$W = \{|T| \geq c\}.$$

Вредност константе c можемо да добијемо из задатог прага значајности и познате расподеле тест статистике при H_0 тако да важи

$$P_{H_0}\{|T| \geq c\} = 0.1.$$

Одатле је

$$c = t_{23;0.1} = 1.714.$$

Дакле, критична област је облика:

$$W = \{|T| \geq 1.714\}.$$

Када смо формирали критичну област, рачунамо реализовану вредност тест статистике за добијене узорке да бисмо одлучили да ли ћемо прихватити или одбацити хипотезу о једнакости очекивања:

$$\overline{x}_{12} = \frac{1}{12}\sum_{i=1}^{12} x_i = 140,$$

$$\overline{s}^2_{12}(x) = \frac{1}{12}\sum_{i=1}^{12}(x_i - 140)^2 = 155.34,$$

$$\overline{y}_{13} = \frac{1}{13}\sum_{i=1}^{13} y_i = 139,$$

$$\overline{s}_{13}^2(y) = \frac{1}{13}\sum_{i=1}^{13}(y_i - 139)^2 = 249.54.$$

Дакле, реализована вредност тест статистике је

$$\frac{140-139}{\sqrt{12\cdot 155.34 + 12\cdot 249.54}}\sqrt{\frac{12\cdot 13}{25}23} = \frac{11.98}{11.47} = 0.168.$$

Проверавамо да ли реализована вредност тест статистике упада у критичну област,

$$|0.168| = 0.168 \notin W.$$

Закључујемо да реализована вредност тест статистике не упада у критичну област, па на основу добијених узорака не одбацујемо хипотезу H_0 о једнакости очекивања.

✓

Задатак 54. Из две популације чија су обележја X које има нормалну $\mathcal{N}(m_1, \sigma_1^2)$ расподелу и Y које има нормалну $\mathcal{N}(m_2, \sigma_2^2)$ расподелу извучени су независни узорци обима 8 и 10 и констатовано је да су узорачке дисперзије $\overline{s}_8^2(X) = 46$ и $\overline{s}_{10}^2(Y) = 50$. Са прагом значајности 0.1 тестирати хипотезу $H_0(\sigma_1^2 = \sigma_2^2)$.

Решење. Није наведена алтернативна хипотеза, па претпостављамо најопштији случај, односно тестирамо $H_0(\sigma_1^2 = \sigma_2^2)$ против $H_1(\sigma_1^2 \neq \sigma_2^2)$.

Добра оцена за дисперзију обележја на популацији је узорачка дисперзија, па можемо покушати да формирамо критичну област на следећи начин:

$$W = \{|\overline{s}_{n_1}^2(\mathbf{x}) - \overline{s}_{n_2}^2(\mathbf{y})| \geq c\}$$

Дакле, желимо да одбацимо нулту хипотезу о једнакости дисперзија обележја X и Y, ако се узорачке дисперзије разликују више од неке унапред одређене вредности c. За овакву критичну област, вредност c бисмо рачунали из следеће једнакости

$$P_{H_0}\{|\overline{S}_{n_1}^2(X) - \overline{S}_{n_2}^2(Y)| \geq c\} = 0.1.$$

Међутим, пошто не знамо расподелу статистике $|\overline{S}_{n_1}^2(X) - \overline{S}_{n_2}^2(Y)|$, нити било које њене трансформације која би нам омогућила лак рачун, не можемо одредити c.

Покушаћемо да формирамо критичну област на другачији начин. Познато је да

$$\frac{n_1\overline{S}_{n_1}^2(X)\sigma_2^2(n_2-1)}{n_2\overline{S}_{n_2}^2(Y)\sigma_1^2(n_1-1)} \in \mathcal{F}_{n_1-1,n_2-1},$$

па при H_0

$$\frac{n_1\overline{S}_{n_1}^2(X)(n_2-1)}{n_2\overline{S}_{n_2}^2(Y)(n_1-1)} = \frac{8\cdot\overline{S}_8^2(X)\cdot 9}{10\cdot\overline{S}_{10}^2(Y)\cdot 7} = \frac{72\overline{S}_8^2(X)}{70\overline{S}_{10}^2(Y)} \in \mathcal{F}_{7,9}.$$

Искористићемо ову чињеницу да формирамо критичну област. Желимо да одбацимо нулту хипотезу о једнакости дисперзија, ако однос узорачких дисперзија није близак 1, односно ако је мањи од неке вредности c_1 или већи од неке вредности c_2. Формирамо критичну област

$$W = \left\{ \frac{\overline{s}_8^2(\mathbf{x})}{\overline{s}_{10}^2(\mathbf{y})} \leq c_1 \right\} \cup \left\{ \frac{\overline{s}_8^2(\mathbf{x})}{\overline{s}_{10}^2(\mathbf{y})} \geq c_2 \right\}.$$

Вредности константи c_1 и c_2 можемо одредити из једнакости:

$$P_{H_0}\left\{ \left\{ \frac{\overline{S}_8^2(X)}{\overline{S}_{10}^2(Y)} \leq c_1 \right\} \cup \left\{ \frac{\overline{S}_8^2(X)}{\overline{S}_{10}^2(Y)} \geq c_2 \right\} \right\} = 0.1.$$

Ако није другачије наглашено, стандардно је да се претпостави да су вероватноће скупова у унији једнаке, тј.

$$P_{H_0}\left\{ \frac{\overline{S}_8^2(X)}{\overline{S}_{10}^2(Y)} \leq c_1 \right\} = 0.05 \quad \text{и} \quad P_{H_0}\left\{ \frac{\overline{S}_8^2(X)}{\overline{S}_{10}^2(Y)} \geq c_2 \right\} = 0.05.$$

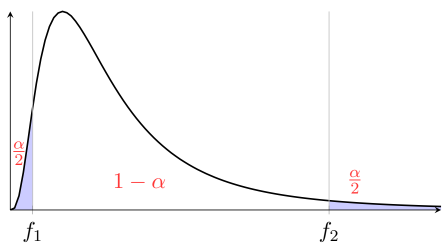
Користећи чињеницу да

$$\frac{72\overline{S}_8^2(X)}{70\overline{S}_{10}^2(Y)} \in \mathcal{F}_{7,9},$$

па

$$\frac{70\overline{S}_{10}^2(Y)}{72\overline{S}_8^2(X)} \in \mathcal{F}_{9,7},$$

коначно можемо израчунати вредности константи c_1 и c_2:



$$\begin{aligned}
&P_{H_0}\left\{ \frac{\overline{S}_8^2(X)}{\overline{S}_{10}^2(Y)} \leq c_1 \right\} = P_{H_0}\left\{ \frac{\overline{S}_{10}^2(Y)}{\overline{S}_8^2(X)} \geq \frac{1}{c_1} \right\} = P_{H_0}\left\{ \frac{70\overline{S}_{10}^2(Y)}{72\overline{S}_8^2(X)} \geq \frac{70}{72}\frac{1}{c_1} \right\} = 0.05 \\
&\quad \Longrightarrow \frac{70}{72}\frac{1}{c_1} = F^{-1}_{\mathcal{F}_{9,7}}(0.95) = 3.7 \\
&\quad \Longrightarrow c_1 = \frac{70}{72}\frac{1}{3.7} = 0.26, \\
&P_{H_0}\left\{ \frac{\overline{S}_8^2(X)}{\overline{S}_{10}^2(Y)} \geq c_2 \right\} = P_{H_0}\left\{ \frac{72\overline{S}_8^2(X)}{70\overline{S}_{10}^2(Y)} \geq \frac{72}{70}c_2 \right\} = 0.05 \\
&\quad \Longrightarrow \frac{72}{70}c_2 = F^{-1}_{\mathcal{F}_{7,9}}(0.95) = 3.3 \\
&\quad \Longrightarrow c_2 = \frac{70}{72}3.3 = 3.21.
\end{aligned}$$

Дакле, критична област је

$$W = \left\{ \frac{\overline{s}_8^2(\mathbf{x})}{\overline{s}_{10}^2(\mathbf{y})} \leq 0.26 \right\} \cup \left\{ \frac{\overline{s}_8^2(\mathbf{x})}{\overline{s}_{10}^2(\mathbf{y})} \geq 3.21 \right\}.$$

Сада можемо да проверимо да ли добијени узорци упадају у критичну област и да одлучимо да ли ћемо одбацити нулту хипотезу. На добијеним узорцима

$$\frac{\overline{s}_8^2(\mathbf{x})}{\overline{s}_{10}^2(\mathbf{y})} = \frac{46}{50} = 0.92,$$

па вредност не упада у критичну област, те не одбацујемо хипотезу о једнакости дисперзија. ✓

Задатак 55. Један лек помаже у лечењу неке болести у 80% случајева. Нови лек за ту болест је помогао у 250 од 300 случајева. Са прагом значајности 0.03 тестирати хипотезу да је нови лек ефикаснији од старог.

Решење. Нека је обележје X такво да има вредност 1 на јединкама којима је помогао нови лек, а 0 на јединкама којима није помогао. Расподела обележја X је тада:

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

где је p вероватноћа да нови лек помогне у лечењу.

Добијен је узорак обима 300 такав да је за 250 јединки одабраних у узорак вредност обележја X једнака 1. Дакле, ако је реализовани узорак $(x_1, ..., x_{300})$, важи да:

$$\sum_{k=1}^{300} x_k = 250, \;\; x_k \in \{0,1\}, \;\; k = 1, ..., 300.$$

На основу добијеног узорка, желимо да закључимо да ли је нови лек ефикаснији од старог. Сматрамо да је нови лек ефикаснији ако је $p > 0.8$, односно ако је вероватноћа да ће пацијенту помоћи нови лек већа него код старог лека. Можемо да тестирамо $H_0(p = 0.8)$ против $H_1(p > 0.8)$. Ако на основу узорка добијемо да треба одбацити нулту хипотезу, закључићемо да је нови лек ефикаснији.

Критичну област формирамо на следећи начин:

$$W = \left\{ \sum_{k=1}^{300} x_k \geq c \right\}.$$

Из задатог прага значајности можемо одредити c тако да важи

$$P_{H_0} \left\{ \sum_{k=1}^{300} X_k \geq c \right\} = 0.03.$$

При H_0:

$$X_k : \begin{pmatrix} 0 & 1 \\ 0.2 & 0.8 \end{pmatrix},$$

према томе $E(X_k) = 0.8$ и $D(X_k) = 0.8(1-0.8) = 0.16, \;\; k = 1, ..., 300.$

Обим узорка је $n = 300 > 30$, случајне величине $X_1, ..., X_n$ су независне и једнако расподељене са коначним очекивањем и дисперзијом, па можемо искористити централну граничну теорему да добијемо расподелу тест статистике при H_0

$$\frac{\sum\limits_{k=1}^{300} X_k - E(\sum\limits_{k=1}^{300} X_k)}{\sqrt{D(\sum\limits_{k=1}^{300} X_k)}} \xrightarrow{D} X^* \in \mathcal{N}(0,1), \;\; n \to \infty.$$

Сада можемо да израчунамо c. При H_0:

$$E\left(\sum_{k=1}^{300} X_k \right) = 300 \cdot 0.8 = 240,$$

$$D\left(\sum_{k=1}^{300} X_k\right) = 300 \cdot 0.16 = 48,$$

па:

$$P_{H_0}\left\{\frac{\sum\limits_{k=1}^{300} X_k - E\left(\sum\limits_{k=1}^{300} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{300} X_k\right)}} \geq \frac{c-240}{\sqrt{48}}\right\} = 0.03.$$

Дакле,

$$\frac{c-240}{4\sqrt{3}} = \Phi^{-1}(0.97) = 1.88$$
$$\implies c = 4\sqrt{3} \cdot 1.88 + 240 = 253.03.$$

Критична област је:

$$W = \left\{\sum_{k=1}^{300} x_k \geq 253.03\right\}.$$

Проверавамо да ли узорак упада у критичну област, односно видимо да је $250 < 253.03$, па узорак не упада у критичну област. Дакле, на основу овог узорка нећемо одбацити H_0, па закључујемо да нови лек није ефикаснији од старог. ✓

Непараметарски тестови

До сада смо се бавили параметарским проблемима, где смо претпостављали да је расподела обележја од интереса одређена до на непознати параметар. Дакле, ако је обележје од интереса X и X има расподелу $F(\cdot, \theta)$, претпостављали смо да знамо тип расподеле F, али не знамо вредност параметра θ (који може бити и вектор). Код непараметарских проблема, не знамо ни тип расподеле F. На основу узорка, код непараметарског тестирања хипотеза, тестирамо да ли је расподела обележја X једнака некој конкретној расподели, тј. тестирамо

$$H_0(F = F_0) \quad \text{против} \quad H_1(F \neq F_0).$$

Пирсонов χ^2 тест

Пирсонов χ^2 тест је један од тестова за тестирање непараметарских хипотеза. Тестирање се спроводи на следећи начин:

- Скуп вредности које узима случајна величина X (узорачки простор) се дели на r дисјунктних подскупова - $S_1, ..., S_r$.
- Ако постоје непознати параметри, оцењују се методом максималне веродостојности.
- Нека је M_k број елемената из узорка обима n који су упали у S_k. Тада је $n = \sum\limits_{k=1}^{r} M_k$.
- Нека је $p_k = P_{H_0}\{X \in S_k\}$. Тада, при H_0, $M_k \in \mathcal{B}(n, p_k)$ расподелу, $k = 1, ..., r$.
- Тест статистика која се користи је $\chi_0^2 = \sum\limits_{k=1}^{r} \frac{(M_k - np_k)^2}{np_k}$ и при H_0 важи да $\chi_0^2 \in \chi^2_{r-s-1}$, где је s број непознатих параметара, оцењених методом максималне веродостојности.

- Критична област је облика $W = \{\chi_0^2 \geq c\}$, где c рачунамо из задатог нивоа значајности теста α и познате расподеле тест статистике при H_0:

$$P_{H_0}\{\chi_0^2 \geq c\} = \alpha.$$

Напомена: Увек прво проверити да ли узорак који је доступан у задатку покрива целу област дефинисаности расподеле која се тестира. Уколико не покрива, треба додати недостајуће вредности и доделити им фреквенције појављивања елемената из узорка једнаке 0. Такође, мора бити $np_k \geq 5$, за свако $k = 1, ..., r$. Ако ово није случај, спајамо класе са најмањим np_k, све док не добијемо да овај услов важи.

Задатак 56. Из популације чије је обележје X извучен је узорак:

X_k	1	2	3	4	5	≥ 6
M_k	45	30	15	6	2	2

Са прагом значајности 0.05 тестирати хипотезу да обележје X има закон расподеле $P\{X = k\} = \frac{1}{2^k},\ k \in \mathbb{N}$.

Решење. Имамо 6 класа, којима су покривене све вредности које може да узме случајна величина X. За почетак, рачунамо вредности np_k, $k = 1, ..., 6$, да видимо да ли ћемо морати да спојимо неке класе (ако је негде $np_k < 5$).

Обим узорка је $n = 45 + 30 + 15 + 6 + 2 + 2 = 100$.

Рачунамо вероватноће $p_k = P_{H_0}\{X \in S_k\}$, где S_k означава k-ту класу:

$$p_k = P_{H_0}\{X = k\} = \frac{1}{2^k},\ \ k = 1, 2, 3, 4, 5,$$

$$p_6 = P_{H_0}\{X \geq 6\} = 1 - P_{H_0}\{X \leq 5\} = 1 - \sum_{k=1}^{5} p_k.$$

Добијају се следеће вредности:

X_k	1	2	3	4	5	≥ 6
M_k	45	30	15	6	2	2
p_k	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
np_k	50	25	12.5	6.25	3.125	3.125

Пошто је $np_5 < 5$ и $np_6 < 5$, спојићемо пету и шесту класу:

X_k	1	2	3	4	≥ 5
M_k	45	30	15	6	4
np_k	50	25	12.5	6.25	6.25

Сада можемо да пређемо на тестирање. Тест статистика $\chi_0^2 = \sum\limits_{k=1}^{5} \frac{(M_k - np_k)^2}{np_k}$ при H_0 има χ^2_{5-0-1} расподелу, јер имамо пет класа и нисмо вршили оцењивање параметара. Критична област је облика:

$$W = \{\chi_0^2 \geq c\}.$$

Вредност константе c налазимо из задатог нивоа значајности и познате расподеле тест статистике при H_0, односно тако да важи

$$P_{H_0}\{\chi_0^2 \geq c\} = 0.05.$$

Дакле,

$$c = \chi^2_{4;0.05} = 9.488.$$

Критична област је облика

$$W = \{\chi^2_0 \geq 9.488\}.$$

Рачунамо вредност тест статистике на добијеном узорку:

$$\chi^2_0 = \frac{(45-50)^2}{50} + \frac{(30-25)^2}{25} + \frac{(15-12.5)^2}{12.5} + \frac{(6-6.25)^2}{6.25} + \frac{(4-6.25)^2}{6.25}$$
$$= 0.5 + 1 + 0.5 + 0.01 + 0.81 = 2.82.$$

Проверавамо да ли узорак упада у критичну област:

$$2.82 < 9.488,$$

па закључујемо да узорак не упада у критичну област, те не одбацујемо H_0. ✓

Задатак 57. Из популације чије је обележје X извучен је узорак:

I_k	[0,1]	[1.5,2.5]	(2.5,3.5]	(3.5,5]
M_k	52	35	9	4

Са прагом значајности 0.02 тестирати хипотезу да обележје X има експоненцијалну расподелу.

Решење. Наведене класе не покривају читав узорачки простор, па морамо да направимо нову табелу у којој ћемо да покријемо све:

I_k	[0,1]	(1,1.5)	[1.5,2.5]	(2.5,3.5]	(3.5,5]	(5,∞)
M_k	52	0	35	9	4	0

Обим узорка је $n = 52 + 35 + 9 + 4 = 100$.

За почетак треба да израчунамо np_k, $k = 1, ..., 6$. Рачунамо p_k по формули

$$p_k = P_{H_0}\{X \in I_k\} = \int_{I_k} \lambda e^{-\lambda x} dx.$$

Параметар λ је непознат, па не можемо израчунати тачне вероватноће, али можемо оценити λ методом максималне веродостојности па добити процене вероватноћа. Оцењујемо λ методом максималне веродостојности:

$$f(x;\lambda) = \lambda e^{-\lambda x}, x \geq 0$$

$$L(\lambda) = \prod_{k=1}^{n} f(x_k, \lambda) = \lambda^n e^{-\lambda \sum\limits_{k=1}^{n} x_k}$$

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{k=1}^{n} x_k$$

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{k=1}^{n} x_k = 0$$

$$\frac{n}{\lambda} = \sum_{k=1}^{n} x_k \implies \lambda = \frac{n}{\sum\limits_{k=1}^{n} x_k} = \frac{1}{\overline{x}_n}$$

Још треба да проверимо да ли је максимум. Пошто је $\frac{\partial \ln L(\lambda)}{\partial\lambda} > 0$, ако је $\lambda < \frac{1}{\overline{x}_n}$, а $\frac{\partial \ln L(\lambda)}{d\partial\lambda} < 0$, ако је $\lambda > \frac{1}{\overline{x}_n}$. Дакле, функција $\log L(\lambda)$ расте до $\frac{1}{\overline{x}_n}$, а после опада. Закључујемо да је оцена максималне веродостојности:

$$\widehat{\lambda} = \frac{1}{\overline{X}_n}.$$

Желимо да добијемо оцену λ на реализованом узорку. Узорачку средину можемо оценити са:

$$\overline{x}_n = \frac{0.5 \cdot 52 + 2 \cdot 35 + 3 \cdot 9 + 4.25 \cdot 4}{100} = 1.4,$$

па на основу овог узорка добијамо да је

$$\widehat{\lambda} = \frac{1}{1.4} = 0.71.$$

Сада можемо да израчунамо вероватноће p_k:

$$\begin{aligned}
p_1 &= P\{X \in I_1\} = \int_0^1 0.71 e^{-0.71x} dx = 0.71 \frac{e^{-0.71x}}{-0.71}\Big|_0^1 = 1 - e^{-0.71} = 0.508, \\
p_2 &= P\{X \in I_2\} = \int_1^{1.5} 0.71 e^{-0.71x} dx = e^{-0.71} - e^{-0.71 \cdot 1.5} = 0.147, \\
p_3 &= P\{X \in I_3\} = \int_{1.5}^{2.5} 0.71 e^{-0.71x} dx = e^{-0.71 \cdot 1.5} - e^{-0.71 \cdot 2.5} = 0.175, \\
p_4 &= P\{X \in I_4\} = \int_{2.5}^{3.5} 0.71 e^{-0.71x} dx = e^{-0.71 \cdot 2.5} - e^{-0.71 \cdot 3.5} = 0.086, \\
p_5 &= P\{X \in I_5\} = \int_{3.5}^{5} 0.71 e^{-0.71x} dx = e^{-0.71 \cdot 3.5} - e^{-0.71 \cdot 5} = 0.055, \\
p_6 &= 1 - p_1 - p_2 - p_3 - p_4 - p_5 = 0.029.
\end{aligned}$$

Последњу вероватноћу коју рачунамо би увек требало рачунати одузимањем претходно израчунатих вероватноћа од 1, због заокруживања, као што је и урађено.

Дакле,

I_k	[0,1]	(1,1.5)	[1.5,2.5]	(2.5,3.5]	(3.5,5]	(5,∞)
M_k	52	0	35	9	4	0
p_k	0.508	0.147	0.175	0.086	0.055	0.029
np_k	50.8	14.7	17.5	8.6	5.5	2.9

Видимо да је $np_6 < 5$, па ћемо шесту класу спојити са класом која има најмање np_k, односно петом класом. Добијамо нову табелу:

I_k	[0,1]	(1,1.5)	[1.5,2.5]	(2.5,3.5]	(3.5,∞)
M_k	52	0	35	9	4
np_k	50.8	14.7	17.5	8.6	8.4

Критична област је облика:

$$W = \{\chi_0^2 \geq c\},$$

где c рачунамо из једнакости

$$P_{H_0}\{\chi_0^2 \geq c\} = 0.02.$$

При H_0, пошто имамо пет класа после срећивања и оценили смо један параметар методом максималне веродостојности, $\chi_0^2 \in \chi^2_{5-1-1}$. Према томе,

$$c = \chi^2_{3;0.02} = 9.837,$$

па је критична област

$$W = \{\chi_0^2 \geq 9.837\}.$$

Рачунамо вредност тест статистике на добијеном узорку:

$$\begin{aligned}
\chi_0^2 &= \sum_{k=1}^{5} \frac{(M_k - np_k)^2}{np_k} \\
&= \frac{(52-50.8)^2}{50.8} + \frac{(0-14.7)^2}{14.7} + \frac{(35-17.5)^2}{17.5} + \frac{(9-8.6)^2}{8.6} + \frac{(4-8.4)^2}{8.4} \\
&= 0.028 + 32.2 + 17.5 + 0.019 + 2.305 = 34.552.
\end{aligned}$$

Пошто је

$$34.552 > 9.837,$$

узорак упада у критичну област, па одбацујемо нулту хипотезу.

✓