УНИВЕРЗИТЕТ “СВ. КИРИЛ И МЕТОДИЈ”– СКОПЈЕ
Природно-математички Факултет
Институт за математика

Ирена Стојковска

ОСНОВИ НА СТАТИСТИКА

Задачи

Скопје, 2020

Содржина

1 Елементи од теорија на веројатност **3**

2 Дескриптивна статистика **9**

3 Основни поими на математичката статистика **21**

4 Оценување на параметри **23**

4.1 Непристрасни и конзистентни оценувачи 23

4.2 Најефикасни оценувачи и доволни статистики 28

4.3 Методи за наоѓање на оценувачи 37

4.4 Интевали на доверба 49

5 Тестирање на хипотези **57**

5.1 Тестирање на параметарски хипотези. Нејман-Пирсонов тест . . . 58

5.2 Тестови за параметрите на нормална распределба 65

5.3 Тестирање на непараметарски хипотези 70

6 Регресиона анализа **79**

Прилози **85**

Литература **87**

1

Елементи од теорија на веројатност

Задача 1.1. Ако $X_1, X_2, ..., X_n$ се независни и еднакво распределени случајни променливи со Бернулиеви $0, 1$ распределби со параметар p, $0 < p < 1$, тогаш случајната променлива $Y = X_1 + X_2 + ... + X_n \sim \mathcal{B}(n, p)$. Покажи.

Задача 1.2. Врската меѓу Поасонова и Биномна распределба е следната

$$\lim_{n\to\infty,\ np\to a} \binom{n}{k} p^k (1-p)^{n-k} = \frac{a^k}{k!} e^{-a}.$$

Покажи.

Решение. Имаме дека

$$\begin{aligned}
\binom{n}{k} p^k (1-p)^{n-k} &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \\
&= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \\
&= \frac{(np)^k}{k!} \cdot \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} (1-p)^{n-k} \\
&= \frac{(np)^k}{k!} \big(1 - \frac{1}{n}\big)\big(1 - \frac{2}{n}\big) \cdots \big(1 - \frac{k-1}{n}\big)(1-p)^{n-k}.
\end{aligned}$$

Бидејќи, $(1-p)^{n-k} = \big((1-p)^{-1/p}\big)^{-np+kp} \to e^{-a}$, кога $np \to a$ и $p \to 0$, имаме дека

$$\binom{n}{k} p^k (1-p)^{n-k} \to \frac{a^k}{k!} e^{-a},$$

кога $n \to \infty,\ np \to a,\ p \to 0$. ∎

Задача 1.3. Ако $X_1, X_2, ..., X_n$ се независни случајни променливи, при што $X_i \sim P(a_i)$, $i = 1, 2, ..., n$, тогаш случајната променлива $Y = X_1 + X_2 + ... + X_n \sim \mathcal{P}(a)$, каде $a = a_1 + a_2 + ... + a_n$. Покажи.

Задача 1.4. Ако случајната променлива X има $\mathcal{N}(\mu, \sigma^2)$ распределба, тогаш случајната променлива $Y = \frac{X-\mu}{\sigma}$ има $\mathcal{N}(0, 1)$ распределба. Покажи.

Решение. Нека $x \in \mathbb{R}$ е произволен. Тогаш, за функцијата на распределба за Y имаме

$$F_Y(x) = P\{Y \leq x\} = P\{\tfrac{X-\mu}{\sigma} \leq x\} = P\{X \leq \sigma x + \mu\} = F_X(\sigma x + \mu),$$

од каде густината на распределба на Y е

$$\begin{aligned} p_Y(x) &= F_Y'(x) = F_X'(\sigma x + \mu) \cdot \sigma = p_X(\sigma x + \mu) \cdot \sigma = \\ &= \sigma \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(\sigma x + \mu - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \end{aligned}$$

и заклучуваме дека $Y \sim \mathcal{N}(0, 1)$. ■

Задача 1.5. Покажи дека за случајната променлива $X \sim \mathcal{N}(\mu, \sigma^2)$ важи **правилото на три сигми**, односно

$$P\{\mu - 3\sigma \leq X \leq \mu + 3\sigma\} = 0,9973 = 99,7\%.$$

Решение. Од Својство 1.4 имаме дека $Y = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$. Тогаш,

$$P\{\mu - 3\sigma \leq X \leq \mu + 3\sigma\} = P\{-3\sigma \leq X - \mu \leq 3\sigma\} =$$

$$= P\{-3 \leq \tfrac{X-\mu}{\sigma} \leq 3\} = P\{-3 \leq Y \leq 3\} =$$

$$= \Phi_0(3) - \Phi_0(-3) = 2\Phi_0(3) = 2 \cdot 0,49867 = 0,9973 \approx 99,7\%,$$

каде $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2} du$ е Лапласовиот интеграл чии вредности се читаат од таблица. ■

Задача 1.6. Ако $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, ..., n$ се независни случајни променливи, тогаш $Y = a_1X_1 + ... + a_nX_n \sim \mathcal{N}(\mu, \sigma^2)$ каде $\mu = a_1\mu_1 + ... + a_n\mu_n$ и $\sigma^2 = a_1^2\sigma_1^2 + ... + a_n^2\sigma_n^2$. Покажи.

Решение. Карактеристичните функции на случајните променливи $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, ..., n$ се

$$\varphi_{X_i}(t) = Ee^{itX_i} = e^{it\mu_i - \sigma_i^2 t^2/2}, \quad i = 1, 2, ..., n.$$

Од својството на карактеристична функција на линеарно зависна променлива т.е. $\varphi_{aX+b}(t) = e^{itb} \cdot \varphi_X(at)$, имаме

$$\varphi_{a_iX_i}(t) = \varphi_{X_i}(a_it) = e^{ita_i\mu_i - \sigma_i^2 a_i^2 t^2/2}.$$

Бидејќи X_i, $i = 1, 2, ..., n$ се независни, за карактеристичната функција на $Y = a_1X_1 + ... + a_nX_n$ имаме

$$\begin{aligned} \varphi_Y(t) &= \varphi_{a_1X_1+...+a_nX_n}(t) = \prod_{i=1}^{n} \varphi_{a_iX_i}(t) = \prod_{i=1}^{n} e^{ita_i\mu_i - \sigma_i^2 a_i^2 t^2/2} = \\ &= e^{it\sum_{i=1}^n a_i\mu_i - (\sum_{i=1}^n \sigma_i^2 a_i^2)t^2/2}, \end{aligned}$$

значи $Y \sim \mathcal{N}(\mu, \sigma^2)$ каде $\mu = \sum_{i=1}^n a_i\mu_i$ и $\sigma^2 = \sum_{i=1}^n a_i^2\sigma_i^2$. ∎

Задача 1.7. Ако $X \sim \mathcal{U}(0, \beta)$, тогаш $Y = (-\frac{1}{\beta}\ln(\frac{X}{\beta})) \sim \mathcal{E}(\beta)$. Покажи.

Задача 1.8. Ако X_i, $i = 1, 2, ..., n$ се независни случајни променливи така што $X_i \sim \Gamma(\alpha_i, \beta)$, $i = 1, 2, ..., n$, тогаш $Y = X_1 + ... + X_n \sim \Gamma(\alpha, \beta)$ каде $\alpha = \alpha_1 + ... + \alpha_n$. Покажи.

Решение. Карактеристичните функции на случајните променливи $X_i \sim \Gamma(\alpha_i, \beta)$, $i = 1, 2, ..., n$ се

$$\varphi_{X_i}(t) = (1 - it\beta)^{-\alpha_i}, \quad i = 1, 2, ..., n.$$

Бидејќи X_i, $i = 1, 2, ..., n$ се независни, за карактеристичната функција на $Y = X_1 + ... + X_n$ имаме

$$\varphi_Y(t) = \varphi_{X_1+...+X_n}(t) = \prod_{i=1}^{n}(1 - it\beta)^{-\alpha_i} = (1 - it\beta)^{-\sum_{i=1}^n \alpha_i},$$

од каде заклучуваме дека $Y \sim \Gamma(\alpha, \beta)$ каде $\alpha = \sum_{i=1}^n \alpha_i$. ∎

Задача 1.9. Ако $X \sim \mathcal{N}(0, 1)$, тогаш $Y = X^2 \sim \chi_1^2$. Покажи.

Решение. Нека $x > 0$, тогаш за функцијата на распределба на Y имаме

$$F_Y(x) = P\{Y \le x\} = P\{X^2 \le x\} = P\{|X| \le \sqrt{x}\} = F_X(\sqrt{x}) - F_X(-\sqrt{x}).$$

Густината на распределба на Y е

$$\begin{aligned} p_Y(x) &= F_Y'(x) = F_X'(\sqrt{x}) \cdot \tfrac{1}{2\sqrt{x}} - F_X'(-\sqrt{x}) \cdot (-\tfrac{1}{2\sqrt{x}}) = \\ &= \tfrac{1}{2\sqrt{x}} \cdot (p_X(\sqrt{x}) + p_X(-\sqrt{x})) = \tfrac{1}{2\sqrt{x}} \cdot (\tfrac{1}{\sqrt{2\pi}}e^{-x/2} + \tfrac{1}{\sqrt{2\pi}}e^{-x/2}) = \\ &= \tfrac{1}{2\sqrt{x}} \cdot 2 \cdot \tfrac{1}{\sqrt{2\pi}}e^{-x/2} = \tfrac{1}{\sqrt{2\pi x}}e^{-x/2} = \tfrac{x^{-1/2}e^{-x/2}}{2^{1/2}\Gamma(1/2)}, \end{aligned}$$

затоа што $\Gamma(1/2) = \sqrt{\pi}$. Значи, $Y \sim \chi_1^2$. ∎

Задача 1.10. Ако $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, тогаш $Y = \frac{(X-\mu)^2}{\sigma^2} \sim \chi_1^2$. Покажи.

Задача 1.11. Ако X_i, $i = 1, 2, ..., n$ се независни и еднакво распределени случајни променливи со $\mathcal{N}(0,1)$ распределби, тогаш $Y = X_1^2 + ... + X_n^2 \sim \chi_n^2$. Покажи.

Решение. Од Својство 1.9 имаме дека $X_i^2 \sim \chi_1^2 \equiv \Gamma(\frac{1}{2}, 2)$, $i = 1, 2, ..., n$. Од независноста на X_i, $i = 1, 2, ..., n$ и Својство **??** имаме дека $Y = X_1^2 + ... + X_n^2 \sim \Gamma(\frac{n}{2}, 2) \equiv \chi_n^2$. ■

Задача 1.12. Ако $X_i \sim \chi_{m_i}^2$, $i = 1, 2, ..., n$ се независни случајни променливи, тогаш $Y = X_1 + X_2 + ... + X_n \sim \chi_m^2$, каде $m = m_1 + m_2 + ... + m_n$. Покажи.

Задача 1.13. Ако $X \sim \mathcal{N}(0,1)$ и $Y \sim \chi_n^2$ се независни случајни променливи, тогаш $Z = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n$. Покажи.

Решение. Од независноста на X и Y, за распределбата на случајниот вектор (X, Y) имаме

$$p_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{y^{n/2-1}}{2^{n/2}\Gamma(n/2)} e^{-y/2}, \quad y > 0.$$

Воведуваме нови случајни променливи $U = \frac{X}{\sqrt{Y/n}}$ и $V = Y$, од каде имаме дека $X = U\sqrt{V/n}$ и $Y = V$. Тогаш, за Јакобијанот на трансформацијата имаме

$$J = \begin{vmatrix} \sqrt{\frac{v}{n}} & \frac{u}{\sqrt{n}} \cdot \frac{1}{2\sqrt{v}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{v}{n}}.$$

Густината на распределба на случајниот вектор (U, V) е

$$\begin{aligned} p_{U,V}(u,v) &= p_{X,Y}(x(u,v), y(u,v)) \cdot |J| = \\ &= \frac{1}{\sqrt{2\pi}} e^{-(1/2)u^2(v/n)} \cdot \frac{v^{n/2-1}}{2^{n/2}\Gamma(n/2)} e^{-v/2} \cdot (\tfrac{v}{n})^{1/2}, \quad v > 0. \end{aligned}$$

Бараната маргинална густина е

$$\begin{aligned} p_U(u) &= \int_0^{+\infty} p_{U,V}(u,v)dv = \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2^{n/2}\Gamma(n/2)} \cdot \int_0^{+\infty} e^{-\frac{u^2 v}{2n}} \cdot v^{n/2-1} \cdot e^{-v/2} \cdot (\tfrac{v}{n})^{1/2} dv = \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2^{n/2}\Gamma(n/2)n^{1/2}} \cdot \int_0^{+\infty} e^{-(\frac{u^2}{n}+1)\cdot\frac{v}{2}} \cdot v^{\frac{n+1}{2}-1} dv. \end{aligned}$$

Ставаме смена $t = (\frac{u^2}{n} + 1) \cdot \frac{v}{2}$, и тогаш

$$\begin{aligned} p_U(u) &= \frac{1}{\sqrt{2\pi}} \cdot \frac{2^{\frac{n+1}{2}-1} \cdot 2 \cdot (\frac{u^2}{n}+1)^{-1}}{2^{n/2}\Gamma(n/2)n^{1/2}(\frac{u^2}{n}+1)^{\frac{n+1}{2}-1}} \cdot \int_0^{+\infty} e^{-t} \cdot t^{\frac{n+1}{2}-1} dt = \\ &= \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(n/2)} \cdot (\frac{u^2}{n} + 1)^{-\frac{n+1}{2}} = \frac{(\frac{u^2}{n}+1)^{-\frac{n+1}{2}}}{B(1/2, n/2)\sqrt{n}}, \end{aligned}$$

затоа што $B(1/2, n/2) = \frac{\Gamma(n/2)\sqrt{\pi}}{\Gamma(\frac{n+1}{2})}$. Значи, $U \sim t_n$ т.е. $Z = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n$. ■

Задача 1.14. Ако $X_1 \sim \mathcal{N}(0,1)$ и $X_2 \sim \mathcal{N}(0,1)$ се независни, тогаш $Y = (\frac{X_1}{X_2})$ има Кошева распределба. Покажи.

Задача 1.15. Ако $X_1 \sim \chi^2_{n_1}$ и $X_2 \sim \chi^2_{n_2}$ се независни случајни променливи, тогаш $Y = \frac{X_1/n_1}{X_2/n_2} \sim F_{n_1,n_2}$. Покажи.

Решение. Од независноста на случајните променливи X_1 и X_2, за распределбата на случајниот вектор (X_1, X_2) имаме

$$p_{X_1,X_2}(x_1,x_2) = \frac{x_1^{n_1/2-1}}{2^{n_1/2}\Gamma(n_1/2)}e^{-x_1/2} \cdot \frac{x_2^{n_2/2-1}}{2^{n_2/2}\Gamma(n_2/2)}e^{-x_2/2}, \quad x_1 > 0,\ x_2 > 0.$$

Воведуваме нови случајни променливи $Y_1 = \frac{X_1/n_1}{X_2/n_2}$ и $Y_2 = X_2/n_2$, од каде имаме дека $X_1 = n_1 Y_1 Y_2$ и $X_2 = n_2 Y_2$. Тогаш, Јакобијанот на трансформацијата е

$$J = \begin{vmatrix} n_1 y_2 & n_1 y_1 \\ 0 & n_2 \end{vmatrix} = n_1 n_2 y_2.$$

Густината на распределба на случајниот вектор (Y_1, Y_2) е

$$\begin{aligned} p_{Y_1,Y_2}(y_1,y_2) &= p_{X_1,X_2}(x_1(y_1,y_2), x_2(y_1,y_2)) \cdot |J| = \\ &= \frac{(n_1y_1y_2)^{n_1/2-1}}{2^{n_1/2}\Gamma(n_1/2)}e^{-\frac{n_1y_1y_2}{2}} \cdot \frac{(n_2y_2)^{n_2/2-1}}{2^{n_2/2}\Gamma(n_2/2)}e^{-\frac{n_2y_2}{2}} n_1n_2y_2, \quad y_1 > 0,\ y_2 > 0. \end{aligned}$$

Бараната маргинална густина е

$$\begin{aligned} p_{Y_1}(y_1) &= \int_0^{+\infty} p_{Y_1,Y_2}(y_1,y_2)dy_2 = \\ &= \frac{(n_1y_1)^{n_1/2-1}}{2^{n_1/2}\Gamma(n_1/2)} \cdot \frac{(n_2)^{n_2/2-1}}{2^{n_2/2}\Gamma(n_2/2)} \cdot n_1n_2 \int_0^{+\infty} y_2^{\frac{n_1}{2}-1} e^{-\frac{n_1y_1y_2}{2}} y_2^{\frac{n_2}{2}-1} e^{-\frac{n_2y_2}{2}} y_2\, dy_2 = \\ &= \frac{(n_1y_1)^{n_1/2-1}}{2^{n_1/2}\Gamma(n_1/2)} \cdot \frac{(n_2)^{n_2/2-1}}{2^{n_2/2}\Gamma(n_2/2)} \cdot n_1n_2 \int_0^{+\infty} y_2^{\frac{n_1+n_2}{2}-1} e^{-\frac{(n_1y_1+n_2)y_2}{2}} dy_2. \end{aligned}$$

Ставаме смена $t = \frac{(n_1y_1+n_2)y_2}{2}$ и добиваме

$$\begin{aligned} p_{Y_1}(y_1) &= \frac{(n_1y_1)^{n_1/2-1}}{2^{n_1/2}\Gamma(n_1/2)} \cdot \frac{(n_2)^{n_2/2-1}}{2^{n_2/2}\Gamma(n_2/2)} \cdot n_1n_2 \cdot \frac{2^{(n_1+n_2)/2}}{(n_1y_1+n_2)^{(n_1+n_2)/2}} \int_0^{+\infty} t^{\frac{n_1+n_2}{2}-1} e^{-t} dt = \\ &= \frac{(n_1y_1)^{n_1/2}}{\Gamma(n_1/2)} \cdot \frac{(n_2)^{n_2/2}}{\Gamma(n_2/2)} \cdot \frac{1}{(n_1y_1+n_2)^{(n_1+n_2)/2}} \cdot \Gamma(\tfrac{n_1+n_2}{2}) = \\ &= \frac{(\frac{n_1}{n_2})^{n_1/2} \cdot y_1^{n_1/2-1}}{B(n_1/2, n_2/2) \cdot (1+\frac{n_1y_1}{n_2})^{(n_1+n_2)/2}}, \end{aligned}$$

од каде заклучуваме дека $Y_1 \sim F_{n_1,n_2}$ т.е. $Y = \frac{X_1/n_1}{X_2/n_2} \sim F_{n_1,n_2}$. ■

Задача 1.16. Ако $X \sim F_{n_1,n_2}$, тогаш $Y = (\frac{1}{X}) \sim F(n_2, n_1)$. Покажи.

Задача 1.17. Ако $X \sim t_n$, тогаш $Y = X^2 \sim F(1, n)$. Покажи.

Задача 1.18. Ако $X_1 \sim \mathcal{N}(0,1)$ и $X_2 \sim \mathcal{N}(0,1)$ се независни случајни променливи, тогаш $Y = (\frac{X_1}{X_2})^2 \sim F_{1,1}$. Покажи.

Задача 1.19. Ако $X_1 \sim \mathcal{E}(\beta_1)$ и $X_2 \sim \mathcal{E}(\beta_2)$ се независни случајни променливи, тогаш $Y = \frac{1}{X_1+X_2} \sim Bet(\beta_1, \beta_2)$. Покажи.

Задача 1.20. Ако $X \sim F_{n_1,n_2}$, тогаш $Y = \frac{(n_1/n_2)X}{1+(n_1/n_2)X} \sim Bet(\frac{n_1}{2}, \frac{n_2}{2})$. Покажи.

2

Дескриптивна статистика

Задача 2.1. Нека $x_1, ..., x_n$ и $y_1, ..., y_n$ се две низи од податоци кои одговараат на обележјата X и Y соодветно, и за кои важи $y_i = ax_i + b$, $i = 1, 2, ..., n$, каде $a, b = const$. Покажи дека $\overline{y} = a\overline{x} + b$ и $\overline{s}_y^2 = a^2\overline{s}_x^2$.

Решение. Користејќи ги дефинициите на аритметичка средина и дисперзија на податоците, имаме

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = \frac{1}{n}\sum_{i=1}^{n} ax_i + \frac{1}{n}\sum_{i=1}^{n} b = a\overline{x} + \frac{1}{n}\cdot nb = a\overline{x} + b,$$

$$\overline{s}_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b - (a\overline{x} + b))^2 = \frac{1}{n}\sum_{i=1}^{n} a^2(x_i - \overline{x})^2 = a^2\overline{s}_x^2,$$

што требаше да се покаже. ∎

Задача 2.2. По првите n_1 мерења $x_1', ..., x_{n_1}'$ кои одговараат на статистичкото обележје X добиена е средина $\overline{x}_1$ и варијанса $\overline{s}_1^2$, а од новите n_2 мерења $x_1'', ..., x_{n_2}''$ добиена е средина $\overline{x}_2$ и варијанса $\overline{s}_2^2$. Ако сите $n_1 + n_2 = n$ мерења се сфатат како една низа од статистички податоци, тогаш се добива средина $\overline{x}$ и варијанса $\overline{s}_0^2$. Покажи дека

а) $\overline{x} = \frac{1}{n}(n_1\overline{x}_1 + n_2\overline{x}_2)$,

б) $\overline{s}_0^2 = \frac{1}{n}(n_1(\overline{s}_1^2 + (\overline{x}_1 - \overline{x})^2) + n_2(\overline{s}_2^2 + (\overline{x}_2 - \overline{x})^2))$.

Решение. За првото тврдење имаме

$$\overline{x} = \frac{1}{n}(\sum_{i=1}^{n_1} x_i' + \sum_{i=1}^{n_2} x_i'') = \frac{1}{n}(n_1\overline{x}_1 + n_2\overline{x}_2),$$

додека за второто тврдење имаме

$$\begin{aligned}
\overline{s}_0^2 &= \frac{1}{n}\Big(\sum_{i=1}^{n_1}(x_i' - \overline{x})^2 + \sum_{i=1}^{n_2}(x_i'' - \overline{x})^2\Big) = \\
&= \frac{1}{n}\Big(\sum_{i=1}^{n_1}(x_i' - \overline{x}_1 + \overline{x}_1 - \overline{x})^2 + \sum_{i=1}^{n_2}(x_i'' - \overline{x}_2 + \overline{x}_2 - \overline{x})^2\Big) = \\
&= \frac{1}{n}\Big(\sum_{i=1}^{n_1}((x_i' - \overline{x}_1)^2 - 2(x_i' - \overline{x}_1)(\overline{x}_1 - \overline{x}) + (\overline{x}_1 - \overline{x})^2) + \\
&\quad + \sum_{i=1}^{n_2}((x_i'' - \overline{x}_2)^2 - 2(x_i'' - \overline{x}_2)(\overline{x}_2 - \overline{x}) + (\overline{x}_2 - \overline{x})^2)\Big) = \\
&= \frac{1}{n}(n_1\overline{s}_1^2 + n_1(\overline{x}_1 - \overline{x})^2 + n_2\overline{s}_2^2 + n_2(\overline{x}_2 - \overline{x})^2) = \\
&= \frac{1}{n}(n_1(\overline{s}_1^2 + (\overline{x}_1 - \overline{x})^2) + n_2(\overline{s}_2^2 + (\overline{x}_2 - \overline{x})^2)),
\end{aligned}$$

затоа што $\sum_{i=1}^{n_1}(x_i' - \overline{x}_1) = 0$ и $\sum_{i=1}^{n_2}(x_i'' - \overline{x}_2) = 0$. ■

Задача 2.3. По анкетирањето на 100 возачи во врска со просечната дневна потрошувачка на гориво (во литри) добиените резултати прикажани се во Табела 2.1.

потрошувачка (во литри)	2	4	5	6	8	10	11	12	13	14
број на возачи	5	10	10	12	18	12	8	10	9	6

Табела 2.1: Просечна дневна потрошувачка на гориво.

Обработи ги дадените податоци (одреди ги бројните карактеристики и направи ги соодветните графички прикази).

Решение. Бројните карактеристики на дадените податоци се

$$\begin{aligned}
\overline{x} &= \frac{1}{100}(5\cdot 2 + 10\cdot 4 + 10\cdot 5 + 12\cdot 6 + 18\cdot 8 + 12\cdot 10 + \\
&\quad +8\cdot 11 + 10\cdot 12 + 9\cdot 13 + 6\cdot 14) = 8,45,
\end{aligned}$$

$$\begin{aligned}
\overline{s}^2 &= \frac{1}{100}(5\cdot 2^2 + 10\cdot 4^2 + 10\cdot 5^2 + 12\cdot 6^2 + 18\cdot 8^2 + 12\cdot 10^2 + \\
&\quad +8\cdot 11^2 + 10\cdot 12^2 + 9\cdot 13^2 + 6\cdot 14^2) - 8,45^2 = 11,7875,
\end{aligned}$$

$$s^2 = \frac{n}{n-1}\overline{s}^2 = \frac{100}{100-1} \cdot 11,7875 = 11,9066.$$

Модата како вредност која се прима со најголема честота, е вредноста 8 (честотата е 18). Бидејќи $n = 100 = 2 \cdot 50$ е парен број, медијаната m се пресметува според формулата

$$m = \frac{1}{2}(x'_{\frac{n}{2}} + x'_{\frac{n}{2}+1}) = \frac{1}{2}(x'_{50} + x'_{51}) = \frac{1}{2}(8+8) = 8.$$

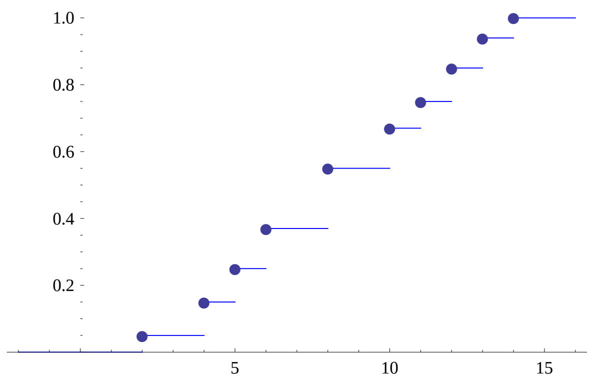
Бидејќи $n = 100 = 4 \cdot 25$ е број од облик $n = 4k$, кварталите Q_1 и Q_3 се пресметуваат според формулите

$$Q_1 = \frac{1}{4}(x'_{\frac{n}{4}} + 3x'_{\frac{n}{4}+1}) = \frac{1}{4}(x'_{25} + 3x'_{26}) = \frac{1}{4}(5 + 3 \cdot 6) = 5,75,$$

$$Q_3 = \frac{1}{4}(3x'_{\frac{3n}{4}} + x'_{\frac{3n}{4}+1}) = \frac{1}{4}(3x'_{75} + x'_{76}) = \frac{1}{4}(3 \cdot 11 + 12) = 11,25.$$

Емпириската функција на распределба е $F_n(x) = \frac{n_x}{n}$, $x \in \mathbb{R}$, каде n_x е бројот на податоци со вредност помала или еднаква на x. Така, за $x < 2$ имаме дека $n_x = 0$, па $F_n(x) = 0$. За $2 \leq x < 4$ имаме дека $n_x = 5$, па $F_n(x) = 5/100 = 0,05$. За $4 \leq x < 5$ имаме дека $n_x = 5 + 10 = 15$, па $F_n(x) = 15/100 = 0,15$. За $5 \leq x < 6$ имаме дека $n_x = 5 + 10 + 10 = 25$, па $F_n(x) = 25/100 = 0,25$ и.т.н. Функцијата $F_n(x)$ заедно со нејзиниот график се прикажани подолу.

$$F_n(x) = \begin{cases} 0 & , x < 2 \\ 0,05 & , 2 \leq x < 4 \\ 0,15 & , 4 \leq x < 5 \\ 0,25 & , 5 \leq x < 6 \\ 0,37 & , 6 \leq x < 8 \\ 0,55 & , 8 \leq x < 10 \\ 0,67 & , 10 \leq x < 11 \\ 0,75 & , 11 \leq x < 12 \\ 0,85 & , 12 \leq x < 13 \\ 0,94 & , 13 \leq x < 14 \\ 1 & , x \geq 14 \end{cases}$$

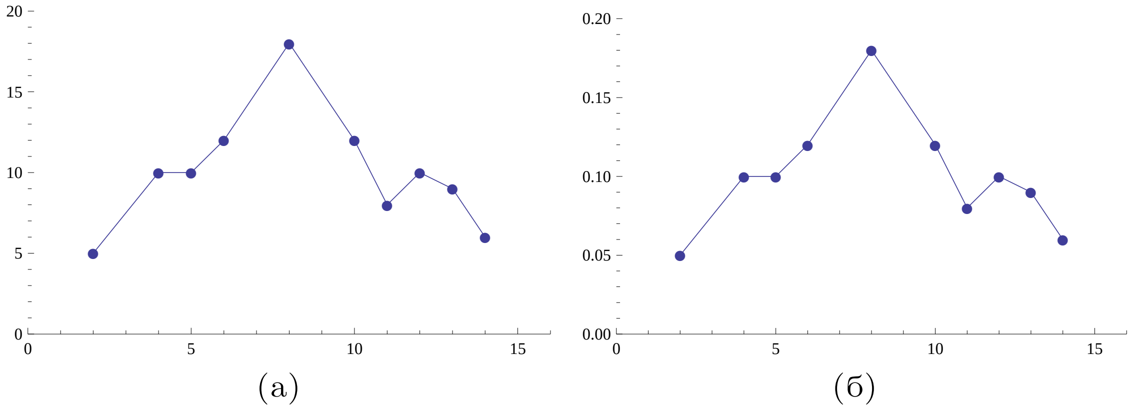


Соодветните релативни честоти на податоците дадени со Табела 2.1 се во Табела 2.2. Полигонот на честоти и полигонот на релативни честоти прикажани се на Слика 2.1. Од овие графички прикази се согледува дека распределбата на податоците е скоро симетрична.

Доколку сакаме да ги прикажеме дадените податоци графички со хистограми, потребно е претходно да ги поделиме во интервали.

потрошувачка (во литри)	2	4	5	6	8	10	11	12	13	14
релативни честоти	0,05	0,10	0,10	0,12	0,18	0,12	0,08	0,10	0,09	0,06

Табела 2.2: Релативни честоти на податоците од Табела 2.1.



Слика 2.1: (а) Полигон на честоти, Табела 2.1, (б) Полигон на релативни честоти, Табела 2.2

Бројот на интервали r може да се одреди на повеќе начини:

$$r \approx \sqrt{n} = \sqrt{100} = 10,$$

$$r \approx 1 + 3,21 \log n 1 + 3,21 \log 100 = 7,42 \approx 7,$$

$$r \approx 5 \log n = 5 \log 100 = 10.$$

Треба да се земе во предвид и препорачливата референца $5-10\%$ од вкупниот број на податоци n и не повеќе од 30% од n, што во нашиот случај ($n = 100$) значи дека препорачливо е r да е број меѓу 5 и 10, но не поголем од 30. Од до сега кажаното, може да се одлучиме за $r = 7$.

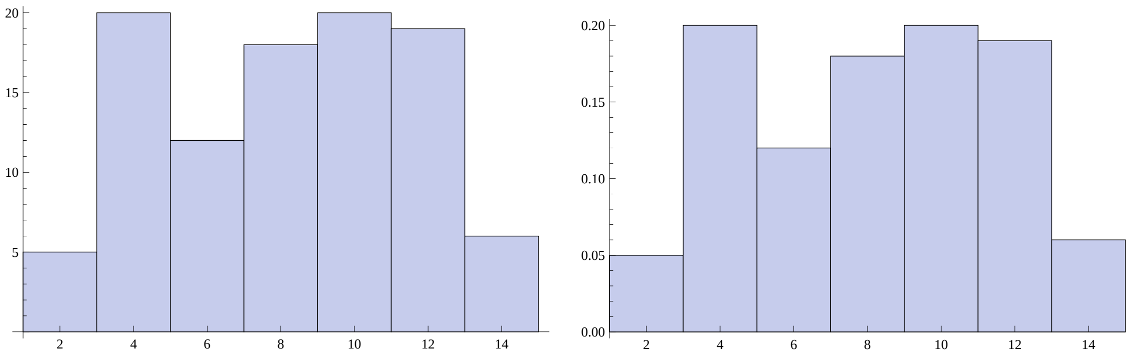
Потоа, го прошируваме интервалот во кој се наоѓаат податоците, тоа е интервалот $[2, 14]$. Едно проширување е $[1, 15]$. За ова проширување, должината на секој од подинтервалите ќе изнесува

$$h = \frac{15-1}{7} = \frac{14}{7} = 2.$$

Со Табела 2.3 дадени се податоците од Табела 2.1 групирани во интервали, заедно со релативните честоти. Соодветните графички прикази со хистограм на честоти и хистограм на релативни честоти прикажани се на Слика 2.2. Добиените хистограми (Слика 2.2) не укажуваат на некоја поголема симетричност на распределбата на податоците.

интервал	[1,3]	(3,5]	(5,7]	(7,9]	(9,11]	(11,13]	(13,15]
честоти	5	20	12	18	20	19	6
релативни честоти	0,05	0,20	0,12	0,18	0,20	0,19	0,06

Табела 2.3: Податоците од Табела 2.1 групирани во интервали.



(а) (б)

Слика 2.2: (а) Хистограм на честоти, Табела 2.3, (б) Хистограм на релативни честоти, Табела 2.3

Друг начин на поделба на интервали ќе даде друга претстава за распределбата на податоците. На пример, ако се одлучиме за $r = 10$ интервали и ако за проширен интервал го земеме интервалот $[1,5 - 14,5]$, тогаш должината на секој од подинтервалите ќе биде

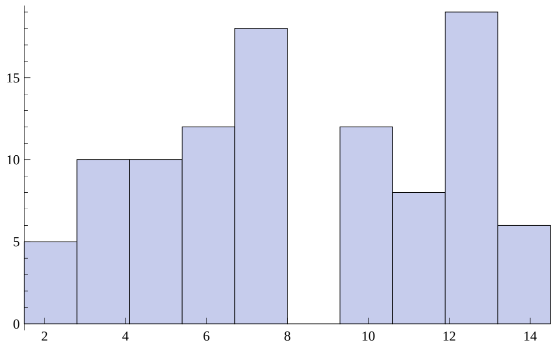
$$h = \frac{14,5 - 1,5}{10} = \frac{13}{10} = 1,3.$$

Групираните податоци во 10 интервали заедно со релативните честоти се прикажани во Табела 2.4. Соодветните графички прикази со хистограм на честоти и хистограм на релативни честоти прикажани се на Слика 2.3. Хистограмите (Слика 2.3) даваат различна претстава за распределбата на истите податоци, но многу поблиска до онаа слика која ја даваат полигоните. ■

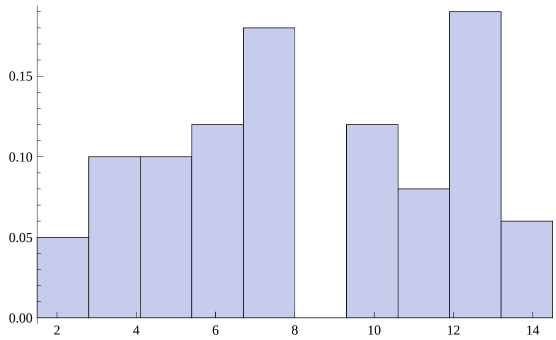
интервал	[1,5-2,8]	(2,8-4,1]	(4,1-5,4]	(5,4-6,7]	(6,7-8]	(8-9,3]
честоти	5	10	10	12	18	0
релативни честоти	0,05	0,10	0,10	0,12	0,18	0

интервал	(9,3-10,6]	(10,6-11,9]	(11,9-13,2]	(13,2-14,5]
честоти	12	8	19	6
релативни честоти	0,12	0,08	0,19	0,06

Табела 2.4: Податоците од Табела 2.1 групирани во интервали.



(а)



(б)

Слика 2.3: (а) Хистограм на честоти, Табела 2.3, (б) Хистограм на релативни честоти, Табела 2.3

Задача 2.4. На случаен начин се избрани 200 стебла на кои им е измерен дијаметарот на напречниот пресек (во cm) и добиените резултати прикажани се во Табела 2.5.

дијаметар (во cm)	40-43	43-46	46-49	49-52	52-55	55-58	58-61
број на стебла	2	7	40	87	58	5	1

Табела 2.5: Дијаметар на напречниот пресек на стеблото.

Обработи ги дадените податоци (одреди ги бројните карактеристики и направи ги соодветните графички прикази).

Решение. Податоците кои се дадени во оваа задача се веќе групирани во интервали, така да пресметвањето на $\overline{x}$, $\overline{s}^2$ е преку средините на интервалите (Табела 2.6).

$$\begin{aligned}\overline{x} &= \frac{1}{200}(2\cdot 41,5+7\cdot 44,5+40\cdot 47,5+87\cdot 50,5+\\ &\quad +58\cdot 53,5+5\cdot 56,5+1\cdot 59,5)=50,665,\end{aligned}$$

$$\begin{aligned}\overline{s}^2 &= \frac{1}{200}(2\cdot 41,5^2+7\cdot 44,5^2+40\cdot 47,5^2+87\cdot 50,5^2+\\ &\quad +58\cdot 53,5^2+5\cdot 56,5^2+1\cdot 59,5^2)-50,665^2=7,76,\end{aligned}$$

$$s^2=\frac{n}{n-1}\overline{s}^2=\frac{200}{200-1}\cdot 7,76=7,799.$$

Модата е средината на оној интервал кој содржи најголем број на податоци, тоа е интервалот $(49,52]$ кој содржи 87 податоци, па модата е 50,5.

Ирена Стојковска

Дадената табела со честоти, Табела 2.5, ја дополнуваме со средините на интервалите, кумулативни честоти, релативни честоти, и кумулативни релативни честоти. Се добива Табела 2.6.

интервал	40-43	43-46	46-49	49-52	52-55	55-58	58-61
средина на инт.	41,5	44,5	47,5	50,5	53,5	56,5	59,5
честоти	2	7	40	87	58	5	1
кумулат. честоти	2	9	49	136	194	199	200
релативни честоти	0,01	0,035	0,2	0,435	0,29	0,025	0,005
кумул. рел. честоти	0,01	0,045	0,245	0,68	0,97	0,995	1

Табела 2.6: Дополнета Табела 2.5.

Медијаната и кварталите Q_1 и Q_3 се пресметуваат на следниот начин. Бидејќи медијаната е $p = 50$-ти процентил, таа се наоѓа на

$$1 + (n-1) \cdot p\% = 1 + (200-1) \cdot 0,50 = 100,5\text{-тото}$$

место, што значи дека се наоѓа во интервалот $(49, 52]$ и изнесува

$$m = 49 + (100,5 - 49) \cdot \frac{52-49}{87} = 50,77586.$$

Слично, кварталот Q_1 се наоѓа на $1 + (200-1) \cdot 0,25 = 50,75$-тото место, значи повторно во интервалот $(49, 52]$, и изнесува

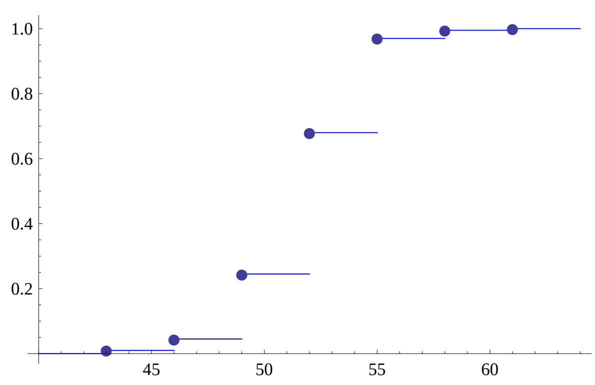
$$Q_1 = 49 + (50,75 - 49) \cdot \frac{52-49}{87} = 49,06034.$$

Кварталот Q_3 се наоѓа на $1 + (200-1) \cdot 0,75 = 150,25$-тото место, значи во интервалот $(52, 55]$, и изнесува

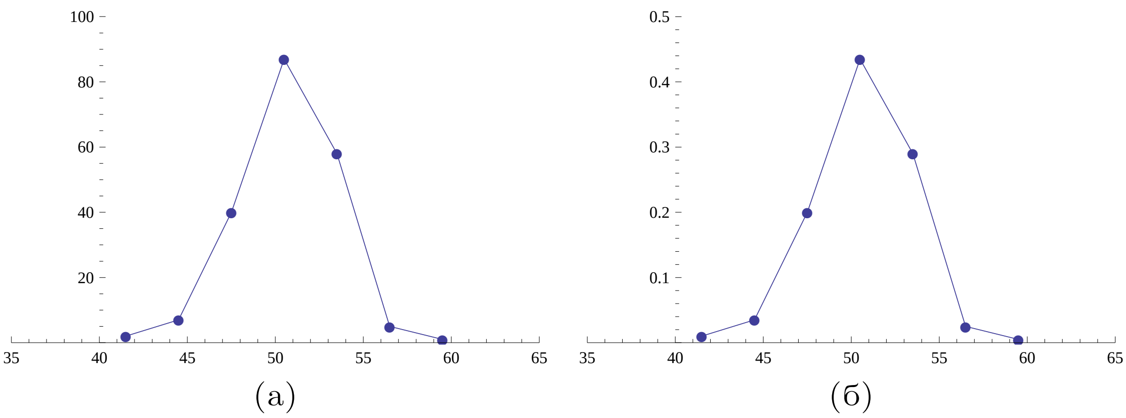
$$Q_3 = 52 + (150,25 - 136) \cdot \frac{55-52}{58} = 52,73707.$$

Емпириската функција на распределба $F_n(x) = \frac{n_x}{n}$, $x \in \mathbb{R}$, каде n_x е бројот на податоци со вредност помала или еднаква на x, заедно со нејзиниот график се прикажани подолу.

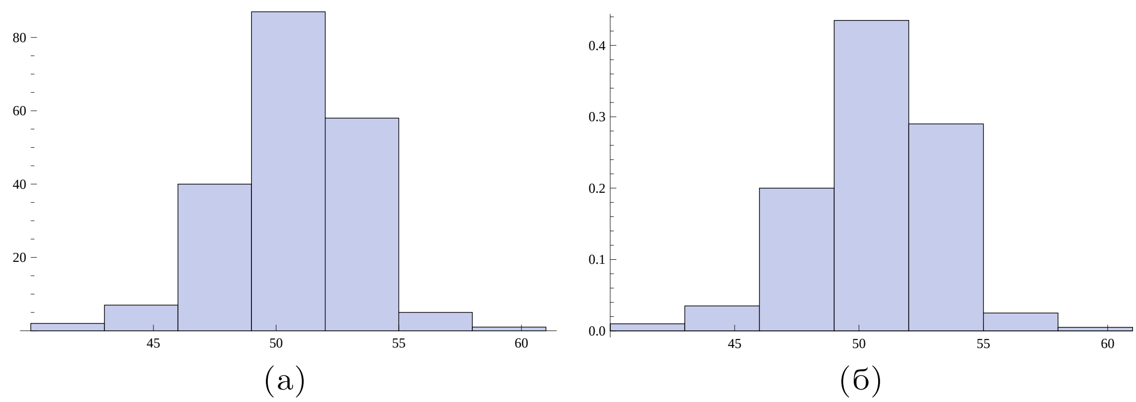
$$F_n(x) = \begin{cases} 0 & , x < 43 \\ 0,01 & , 43 \le x < 46 \\ 0,045 & , 46 \le x < 49 \\ 0,245 & , 49 \le x < 52 \\ 0,68 & , 52 \le x < 55 \\ 0,97 & , 55 \le x < 58 \\ 0,995 & , 58 \le x < 61 \\ 1 & , x \ge 61 \end{cases}$$



Полигонот на честоти и полигонот на релативни честоти дадени се на Слика 2.4. Хистограмот на честоти и хистограмот на релативни честоти дадени се на Слика 2.5. Од овие графички прикази се согледува симетричноста на распределбата на податоците. ■



Слика 2.4: (а) Полигон на честоти, Табела 2.6, (б) Полигон на релативни честоти, Табела 2.6



Слика 2.5: (а) Хистограм на честоти, Табела 2.6, (б) Хистограм на релативни честоти, Табела 2.6

Задача 2.5. Направено е испитување кај 150 студенти за да се увиди врската меѓу оценката по математика во последната година од средното образование (X) и оценката на испитот по математика на факултет (Y). Добиените резултати прикажани се во Табела 2.7.

$X \backslash Y$	5	6	7	8	9	10
2	2	1	0	1	0	0
3	5	35	2	5	0	1
4	3	1	10	15	6	6
5	1	0	0	6	18	32

Табела 2.7: Оценки по предметот математика во последната година од средното образование (X) и оценката на испитот по математика на факултет (Y).

Обработи ги дадените податоци (одреди ги бројните карактеристики и испитај ја зависноста меѓу обележјата).

Решение. За наоѓање на бројните карактеристики, најнапред ги одредуваме маргиналните распределби на фрекфенциите.

X	2	3	4	5
g_i	4	48	41	57

Y	5	6	7	8	9	10
h_j	11	37	12	27	24	39

Табела 2.8: Маргинални фрекфенции.

Тогаш, имаме

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{r} g_i a_i = \frac{1}{150}(4\cdot 2 + 48\cdot 3 + 41\cdot 4 + 57\cdot 5) = 4,006667,$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{s} h_i b_i = \frac{1}{150}(11\cdot 5 + 37\cdot 6 + 12\cdot 7 + 27\cdot 8 + 24\cdot 9 + 39\cdot 10) = 7,886667,$$

$$\overline{s}_x^2 = \frac{1}{n}\sum_{i=1}^{r} g_i a_i^2 - \overline{x}^2 = 0,8066222,$$

$$\overline{s}_y^2 = \frac{1}{n}\sum_{i=1}^{s} h_i b_i^2 - \overline{y}^2 = 2,913822,$$

$$\overline{s}_{xy} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s} a_i b_j f_{ij} - \overline{x}\,\overline{y} = 1,03404,$$

па коефициентот на корелација е

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = 0,6744826,$$

што значи дека постои статистички сигнификантна корелација ($|r_{xy}| \geq 0,5$) меѓу оценките по математика во последната година од средното образование и иститe на испитот по математика на факултет. Понатаму, за да ја испитаме статистичката зависност меѓу обележјата, ги пресметуваме оцтапувањето од статистичката независност

$$f^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{f_{ij}^2}{g_i h_j} - 1 = 1,077948,$$

и степенот на статистичката зависност

$$o = \frac{f^2}{\min\{r,s\} - 1} = \frac{1,077948}{4-1} = 0,359827 \approx 36\%,$$

што значи дека меѓу оценката по математика во последната година од средното образование и оценката на испитот по математика на факултет постои статистичка зависност од 36%.

За одредување на кривите на регресија, ги бараме најнапред условните распределби на фрекфенциите, за да ги најдеме соодветните аритметички средини на податоците од условните распределби.

$X\|Y=5$	2	3	4	5
честота	2	5	3	1

$X\|Y=6$	2	3	4	5
честота	1	35	1	0

$X\|Y=7$	2	3	4	5
честота	0	2	10	0

$X\|Y=8$	2	3	4	5
честота	1	5	15	6

$X\|Y=9$	2	3	4	5
честота	0	0	6	18

$X\|Y=10$	2	3	4	5
честота	0	1	6	32

Табела 2.9: Условни распределби на фрекфенции за обележјето X.

Соодветните аритметички средини на податоците од условните распределби за X се

$$\overline{x}(5) = 3,272727,\ \overline{x}(6) = 3,102564,\ \overline{x}(7) = 3,75,$$

$$\overline{x}(8) = 3,962963,\ \overline{x}(9) = 4,428571,\ \overline{x}(10) = 4,658537.$$

$Y \mid X = 2$	5	6	7	8	9	10
честота	2	1	0	1	0	0

$Y \mid X = 3$	5	6	7	8	9	10
честота	5	35	2	5	0	1

$Y \mid X = 4$	5	6	7	8	9	10
честота	3	1	10	15	6	6

$Y \mid X = 5$	5	6	7	8	9	10
честота	1	0	0	6	18	32

Табела 2.10: Условни распределби на фрекфенции за обележјето Y.

Соодветните аритметички средини на податоците од условните распределби за Y се

$$\overline{y}(2) = 6,\ \overline{y}(3) = 6,34,\ \overline{y}(4) = 7,92683,\ \overline{y}(5) = 9,196721.$$

Графичкиот приказ на кривата на регресија на Y во зависност од X е искршена линија со темиња во $(a_i, \overline{y}(a_i))$, $i = 1, ..., r$, а кривата на регресија на X во зависност од Y е искршена линија со темиња во $(\overline{x}(b_i), b_j)$, $j = 1, ..., s$. Правите на регресија кои ги апроксимираат овие две криви се

$$y = \overline{y} + \frac{s_{xy}}{s_x^2}(x - \overline{x}) = 1,2819x + 2,7504,$$

$$x = \overline{x} + \frac{s_{xy}}{s_y^2}(y - \overline{y}) = 0,3549y + 1,2079,$$

соодветно.

Задача 2.6. При едно испитување направено на пазарот за посетеноста на супермаркетите, во еден супермаркет, во 40 случајно одбрани денови, измерени се просечниот процент на намалувањето на цените (x) и бројот на посетители во минута (y) и добиени се следните дводимензионални податоци:

(0, 12), (0, 13), (0, 13), (1, 14), (1, 14), (2, 14), (2, 14), (2, 16),
(3, 17), (3, 17), (4, 18), (4, 18), (4, 18), (4, 18), (5, 19), (5, 19),
(6, 19), (6, 19), (6, 19), (7, 19), (7, 20), (7, 20), (7, 20), (7, 20),
(7, 20), (7, 20), (8, 21), (9, 21), (10, 21), (11, 21), (11, 22), (13, 22),
(13, 22), (14, 23), (15, 25), (16, 26), (17, 26), (18, 27), (22, 29), (26, 30)

Обработи ги дадените податоци (одреди ги бројните карактеристики и испитај ја зависноста меѓу обележјата).

Задачи за самостојна работа

Задача 2.7. За случајно избрани 60 дипломирани студенти, забележан е просекот од студиите и добиени се следните резултати (подредени):

6.00, 6.00, 6.00, 6.26, 6.26, 6.54, 6.94, 7.00, 7.03, 7.08,
7.17, 7.24, 7.26, 7.28, 7.32, 7.40, 7.42, 7.46, 7.53, 7.65,
7.65, 7.71, 7.83, 7.86, 7.87, 7.87, 7.89, 7.89, 7.90, 7.92,
7.94, 7.94, 7.95, 8.04, 8.08, 8.18, 8.18, 8.22, 8.26, 8.31,
8.58, 8.59, 8.67, 8.68, 8.73, 8.81, 8.83, 8.93, 8.94, 9.09,
9.10, 9.15, 9.25, 9.26, 9.45, 9.49, 9.58, 9.64, 9.81, 10.00

За дадените податоци важи $\sum x_i = 480.91$ и $\sum x_i^2 = 3912.29$.

а) Најди ги бројните карактеристики (средина, дисперзија, мода, обсег, медијана, кваталите Q_1 и Q_3) на дадените податоци.

б) Групирај ги дадените податоци во интервали, и на така зададените интервални податоци пресметај ги бројните карактеристики.

в) Прикажи ги графички распределбата на податоците и бројните карактеристики. Што се воочува од графичките прикази?

Задача 2.8. Стап со должина 2 метра на случаен начин се крши на два дела. Овој експеримент е повторен 45 пати при што се забележани вредностите (подредени по големина) на помалиот дел (во метри):

0.04, 0.04, 0.07, 0.13, 0.14, 0.21, 0.26, 0.27, 0.28,
0.29, 0.29, 0.3, 0.31, 0.31, 0.31, 0.34, 0.34, 0.37,
0.39, 0.4, 0.46, 0.48, 0.53, 0.59, 0.6, 0.62, 0.63,
0.64, 0.65, 0.67, 0.71, 0.74, 0.75, 0.77, 0.78, 0.79,
0.8, 0.84, 0.87, 0.87, 0.92, 0.99, 0.99, 0.99, 1.00.

а) Групирај ги дадените податоци во интервали, и на така зададените интервални податоци пресметај ги бројните карактеристики.

б) Прикажи ги графички распределбата на податоците и бројните карактеристики. Што се воочува од графичките прикази?

Задача 2.9. Зависноста меѓу примањата по глава на жител (во иљади денари) и потрошувачката на електрична енергија (во kWh) е дадена со следната табела

примања	9,0	9,0	11,0	12,5	12,5	15,0
потрошувачка	8	10	12	13	15	18
примања	15,0	18,0	30,0	45,5	50,0	65,0
потрошувачка	20	25	30	42	48	52

Најди ги коефициентот на корелација, степенот на статистичка зависност меѓу примањата по глава на жители потрошувачката на електрична енергија. Најди ги правите на регресија. Коментирај ги корелацијата и зависноста.

Ирена Стојковска

3

Основни поими на математичката статистика

Задача 3.1. Нека $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележјето X со функција на распределба $F(x)$. Најди ги распределбите на подредените статистики $X_{(1)} = \min\{X_1, ..., X_n\}$ и $X_{(n)} = \max\{X_1, ..., X_n\}$.

Решение. За функцијата на распределба на $X_{(1)}$ имаме

$$\begin{aligned} F_{X_{(1)}}(x) &= P\{X_{(1)} \leq x\} = P\{\min\{X_1, ..., X_n\} \leq x\} = \\ &= 1 - P\{\min\{X_1, ..., X_n\} > x\} = 1 - P\{X_1 > x, ..., X_n > x\} = \\ &= 1 - (1 - P\{X_1 \leq x\}) \cdots (1 - P\{X_n \leq x\}) = 1 - (1 - F(x))^n, \end{aligned}$$

додека за функцијата на распределба на $X_{(n)}$ имаме

$$\begin{aligned} F_{X_{(n)}}(x) &= P\{X_{(n)} \leq x\} = P\{\max\{X_1, ..., X_n\} \leq x\} = \\ &= P\{X_1 \leq x, ..., X_n \leq x\} = P\{X_1 \leq x\} \cdots P\{X_n \leq x\} = (F(x))^n. \end{aligned}$$

Задача 3.2. Нека $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележје $X \sim \mathcal{E}(\beta)$. Покажи дека статистиката $T_n = \frac{2}{\beta} \sum_{i=1}^n X_i \sim \chi^2_{2n}$.

Решение. Дефинираме случајна променлива $Y = \frac{2}{\beta} X$, каде $X \sim \mathcal{E}(\beta)$. Тогаш, за распределбата на Y имаме

$$F_Y(y) = P\{Y \leq y\} = P\{\frac{2}{\beta} X \leq y\} = P\{X \leq \frac{\beta y}{2}\} = F_X(\frac{\beta y}{2}),$$

$$p_Y(y) = F'_Y(y) = F'_X(\frac{\beta y}{2}) \cdot \frac{\beta}{2} = p_X(\frac{\beta y}{2}) \cdot \frac{\beta}{2} = \frac{1}{\beta}\, e^{-\frac{1}{\beta} \cdot \frac{\beta y}{2}} \cdot \frac{\beta}{2} = \frac{1}{2}\, e^{-y/2},\ y > 0,$$

односно $Y \sim \mathcal{E}(2) \equiv \Gamma(1, 2)$. Бидејќи $(X_1, X_2, ..., X_n)$ е примерок кој одговара на обележјето $X \sim \mathcal{E}(\beta)$, следи дека $X_i \sim \mathcal{E}(\beta)$, $i = 1, 2, ..., n$, од каде $Y_i = \frac{2}{\beta} X_i \sim \mathcal{E}(2) \equiv \Gamma(1, 2)$, $i = 1, 2, ..., n$. Од Задача 1.8, следи дека $T_n = \frac{2}{\beta} \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i \sim \Gamma(n, 2) \equiv \chi^2_{2n}$, што требаше да се покаже.

Задачи за самостојна работа

Задача 3.3. Нека се дадени два независни примерока $X = (X_1, ..., X_{n_1})$ и $Y = (Y_1, ..., Y_{n_2})$ кои одговараат на обележја со еднаква дисперзија σ^2. Нека $Z = (X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2})$ е примерок добиен со спојување на примероците X и Y. Нека $\overline{X}$, $\overline{Y}$ и $\overline{Z}$ се средините на примероците X, Y и Z соодветно. Покажи дека $D(\overline{Z} - \overline{X}) = \frac{n_2\sigma^2}{n_1(n_1+n_2)}$.

Задача 3.4. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележје $X \sim \mathcal{N}(m_1, \sigma_1^2)$, $(Y_1, ..., Y_m)$ е примерок кој одговара на обележје $Y \sim \mathcal{N}(m_2, \sigma_2^2)$ и нека овие два примерока се независни еден од друг. Покажи дека

а) статистиката $U = \overline{X}_n - \overline{Y}_m = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{j=1}^{m} Y_j \sim \mathcal{N}(m_1 - m_2, \frac{\sigma_1^2}{n} - \frac{\sigma_2^2}{m})$,

б) статистиката $V = \frac{1}{\sigma_1^2}\sum_{i=1}^{n} X_i + \frac{1}{\sigma_2^2}\sum_{j=1}^{m} Y_j \sim \mathcal{N}(\frac{nm_1}{\sigma_1^2} - \frac{nm_2}{\sigma_2^2}, \frac{n}{\sigma_1^2} - \frac{m}{\sigma_2^2})$.

Задача 3.5. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележје X со $EX = m$, $DX = \sigma^2$, $EX^k = m_k$ и $E(X-m)^k = \mu_k$. Покажи дека $D(\overline{S}_n^2) = \frac{(n-1)^2}{n^3}(\mu_4 - \frac{n-3}{n-1}\mu_2^2)$, каде $\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ е дисперзијата на примерокот, а $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ е средината на примерокот.

Задача 3.6. а) Покажи дека за дисперзијата на примерокот $\overline{S}_n^2$ важи $\overline{S}_n^2 \xrightarrow{\text{с.с.}} \sigma^2$, $n \to \infty$.

б) Покажи дека дисперзијата на примерокот $\overline{S}_n^2$ има асимптотска $\mathcal{N}(\sigma^2, \frac{1}{n}(\mu_4 - \sigma^2))$ распределба.

4

Оценување на параметри

4.1 Непристрасни и конзистентни оценувачи

Задача 4.1. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето $X \sim \mathcal{N}(0, \theta)$, каде $0 < \theta < +\infty$ е непознат параметар. Покажи дека $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ е непристрасен оценувач за θ. Најди ја дисперзијата на $\hat{\theta}$. Дали $\hat{\theta}$ е конзистентен оценувач за θ?

Решение. Од $X \sim \mathcal{N}(0, \theta)$ следи дека бројните карактеристики на обележјето X се $EX = 0$ и $DX = \theta$, од каде и за секоја случајна променлива од примерокот важи $EX_i = 0$ и $DX_i = \theta$, $i = 1, 2, ..., n$, од каде $E(X_i^2) = DX_i + (EX_i)^2 = \theta + 0^2 = \theta$. Тогаш,

$$E(\hat{\theta}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) = \frac{1}{n}\sum_{i=1}^{n} \theta = \frac{1}{n} \cdot n\theta = \theta,$$

од каде следи дека $\hat{\theta}$ е непристрасен оценувач за θ.
Од независноста на $X_1, ..., X_n$, следи дека се независни и $X_1^2, ..., X_n^2$, па за дисперзијата на оценувачот $\hat{\theta}$ имаме

$$D(\hat{\theta}) = D(\frac{1}{n}\sum_{i=1}^{n} X_i^2) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i^2).$$

Сега, бидејќи густината на распределба на X е $p(x) = \frac{1}{\sqrt{2\pi\theta}}\, e^{-\frac{x^2}{2\theta}}$, имаме

$$\begin{aligned} D(X_i^2) &= E(X_i^4) - (E(X_i^2))^2 = \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2\theta}}\, dx \; - \theta^2 = \\ &= \frac{1}{\sqrt{2\pi\theta}} \Big(-\theta x^3 e^{-\frac{x^2}{2\theta}} \Big|_{-\infty}^{+\infty} + 3\theta \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2\theta}}\, dx \Big) - \theta^2 = \\ &= 3\theta \cdot E(X_i^2) - \theta^2 = 3\theta^2 - \theta^2 = 2\theta^2. \end{aligned}$$

Па, бараната дисперзија е

$$D(\hat{\theta}) = \frac{1}{n^2}\sum_{i=1}^{n} 2\theta^2 = \frac{1}{n^2}\cdot n \cdot 2\theta^2 = \frac{2\theta^2}{n}.$$

Бидејќи $E(\hat{\theta}) = \theta$ и $D(\hat{\theta}) = \frac{2\theta^2}{n} \to 0$, кога $n \to \infty$, следи дека $\hat{\theta}$ е конзистентен оценувач за θ.

Забелешка. Конзистентноста на $\hat{\theta}$ може да ја покажеме и по дефиниција. Нека $\varepsilon > 0$ е произволен. Кога оценувачот е непристрасен може директно да се примени неравенството на Чебишев, односно

$$0 \le P\{|\hat{\theta} - \theta| \ge \varepsilon\} = P\{|\hat{\theta} - E(\hat{\theta})| \ge \varepsilon\} \le \frac{D(\hat{\theta})}{\varepsilon^2} = \frac{2\theta^2}{n\varepsilon^2} \to 0,\ n \to \infty,$$

од каде $\forall \varepsilon > 0,\ \lim\limits_{n\to\infty} P\{|\hat{\theta} - \theta| \ge \varepsilon\} = 0$, односно $\forall \varepsilon > 0,\ \lim\limits_{n\to\infty} P\{|\hat{\theta} - \theta| < \varepsilon\} = 1$, па $\hat{\theta}$ е конзистентен оценувач за θ.

Задача 4.2. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X кое има густина на распределба $p(x)$ за која важи $p(\theta - x) = p(\theta + x)$, $\forall x \in \mathbb{R}$. Нека $\hat{\theta}$ е оценувач за параметарот θ за кој важи $\hat{\theta}(x_1 + h, ..., x_n + h) = \hat{\theta}(x_1, ..., x_n) + h$, $\forall h \in \mathbb{R}$ и $\hat{\theta}(-x_1, ..., -x_n) = -\hat{\theta}(x_1, ..., x_n)$. Докажи дека, ако постои $E(\hat{\theta})$, тогаш $\hat{\theta}$ е непристрасен оценувач за θ.

Решение. Нека постои $E(\hat{\theta})$. Ако покажеме дека $E(\hat{\theta}) - \theta = 0$, тогаш $\hat{\theta}$ е непристрасен оценувач за θ. Од независниста на случајните променливи во примерокот и условите на задачата, имаме

$$\begin{aligned} E(\hat{\theta}) - \theta = E(\hat{\theta} - \theta) &= \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} (\hat{\theta}(x_1, ..., x_n) - \theta) p(x_1, ..., x_n) dx_1 ... dx_n = \\ = \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} (\hat{\theta}(x_1, ..., x_n) - \theta) p(x_1) ... p(x_n) dx_1 ... dx_n = I, \end{aligned}$$

каде заради упростување на ознаките $p(x_1, ..., x_n)$ означува густина на распределба на примерокот $(X_1, ..., X_n)$. Ако ја ставиме смената $x_k = y_k + \theta$, $k = 1, ..., n$, а потоа и смената $y_k = -z_k$, $k = 1, ..., n$, од условите на задачата за интегралот I имаме

$$\begin{aligned} I &= \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} (\hat{\theta}(y_1 + \theta, ..., y_n + \theta) - \theta) p(y_1 + \theta) ... p(y_n + \theta) dy_1 ... dy_n = \\ &= \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \hat{\theta}(y_1, ..., y_n) p(\theta - y_1) ... p(\theta - y_n) dy_1 ... dy_n = \\ &= (-1)^n \int_{+\infty}^{-\infty} ... \int_{+\infty}^{-\infty} \hat{\theta}(-z_1, ..., -z_n) p(\theta + z_1) ... p(\theta + z_n) dz_1 ... dz_n = \\ &= \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \hat{\theta}(-z_1, ..., -z_n) p(\theta + z_1) ... p(\theta + z_n) dz_1 ... dz_n = \end{aligned}$$

$$= -\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \hat{\theta}(z_1, ..., z_n) p(\theta + z_1)...p(\theta + z_n) dz_1...dz_n =$$
$$= -\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} (\hat{\theta}(z_1 + \theta, ..., z_n + \theta) - \theta) p(\theta + z_1)...p(\theta + z_n) dz_1...dz_n = -I.$$

Односно $I = -I$, од каде $I = 0$, па $\hat{\theta}$ е непристрасен оценувач за θ.

Задача 4.3. Нека $\hat{\theta}_1$ и $\hat{\theta}_2$ се непристрасни оценувачи за параметарот θ добиени од две независни серии на набљудувања. Ако $D(\hat{\theta}_1) = 2D(\hat{\theta}_2)$, одреди ги константите k_1 и k_2 така што $\hat{\theta}_3 = k_1\hat{\theta}_1 + k_2\hat{\theta}_2$ е непристрасен оценувач за θ со минимална дисперзија.

Решение. Од $\hat{\theta}_1$ и $\hat{\theta}_2$ непристрасни оценувачи за θ имаме дека $E(\hat{\theta}_1) = \theta$ и $E(\hat{\theta}_2) = \theta$. И $\hat{\theta}_3$ е непристрасен оценувач за θ, па затоа

$$\theta = E(\hat{\theta}_3) = E(k_1\hat{\theta}_1 + k_2\hat{\theta}_2) = k_1E(\hat{\theta}_1) + k_2E(\hat{\theta}_1) = k_1\theta + k_2\theta = (k_1 + k_2)\theta,$$

од каде $k_1 + k_2 = 1$. Од независноста на $\hat{\theta}_1$ и $\hat{\theta}_2$ имаме

$$\begin{aligned} D(\hat{\theta}_3) &= D(k_1\hat{\theta}_1 + k_2\hat{\theta}_2) = k_1^2 D(\hat{\theta}_1) + k_2^2 D(\hat{\theta}_2) = (2k_1^2 + k_2^2) D(\hat{\theta}_2) = \\ &= (2k_1^2 + (1 - k_1)^2) D(\hat{\theta}_2) = (3k_1^2 - 2k_1 + 1) D(\hat{\theta}_2), \end{aligned}$$

од каде $D(\hat{\theta}_3)$ достигнува минимум за $k_1 = \frac{1}{3}$, од каде следи дека $k_2 = 1 - k_1 = \frac{2}{3}$.

Задача 4.4. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето $X \sim \mathcal{U}(0, a)$, каде $a > 0$ е непознат параметар. Покажи дека $\hat{a}_1 = 2\overline{X}_n$, каде $\overline{X}_n$ е средината на примерокот и $\hat{a}_2 = \max\{X_1, ..., X_n\}$ се конзистентени оценувачи за параметарот a.

Решение. Од $X \sim \mathcal{U}(0, a)$ имаме дека густината на распределба на X е $p(x) = \frac{1}{a}$, $x \in (0, a)$, па за бројните карактеристики на X имаме дека се $EX = \frac{a}{2}$ и $DX = \frac{a^2}{12}$, од каде и засекоја случајна променлива од примерокот имаме $EX_i = \frac{a}{2}$ и $DX_i = \frac{a^2}{12}$, $i = 1, ..., n$. Тогаш, за оценувачот $\hat{a}_1 = 2\overline{X}_n = \frac{2}{n}\sum_{i=1}^{n} X_i$ имаме

$$E(\hat{a}_1) = E\Big(\frac{2}{n}\sum_{i=1}^{n} X_i\Big) = \frac{2}{n}\sum_{i=1}^{n} E(X_i) = \frac{2}{n}\sum_{i=1}^{n} \frac{a}{2} = \frac{2}{n} \cdot n \cdot \frac{a}{2} = a,$$

а заради независноста на случајните променливи во примерокот, имаме

$$D(\hat{a}_1) = D\Big(\frac{2}{n}\sum_{i=1}^{n} X_i\Big) = \frac{4}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{4}{n^2}\sum_{i=1}^{n} \frac{a^2}{12} = \frac{4}{n^2} \cdot n \cdot \frac{a^2}{12} = \frac{a^2}{3n}.$$

Сега, од $E(\hat{a}_1) = a$ и $D(\hat{a}_1) = \frac{a^2}{3n} \to 0$, кога $n \to \infty$, следи дека $\hat{a}_1$ е конзистентен оценувач за a. Да е и забележиме дека $\hat{a}_1$ е и непристрасен оценувач за a.

Ирена Стојковска

За да ги најдеме бројните карактеристики на оценувачот $\hat{a}_2 = \max\{X_1, ..., X_n\}$, прво ја одредуваме неговата распределба. За функцијата на распределба на обележјето X имаме дека е $F(x) = 0$ за $x \leq 0$, $F(x) = \frac{x}{a}$ за $x \in (0, a)$ и $F(x) = 1$ за $x \geq a$. Па, за функцијата на распределба на $\hat{a}_2$ имаме

$$\begin{aligned}
F_{\hat{a}_2}(x) &= P\{\hat{a}_2 \leq x\} = P\{\max\{X_1, ..., X_2\} \leq x\} = P\{X_1 \leq x, ..., X_n \leq x\} = \\
&= P\{X_1 \leq x\} \cdot ... \cdot P\{X_n \leq x\} = \underbrace{F(x) \cdot ... \cdot F(x)}_{n} = (F(x))^n = \frac{x^n}{a^n} \text{ за } x \in (0, a),
\end{aligned}$$

од каде густината на распределба на $\hat{a}_2$ е

$$p_{\hat{a}_2}(x) = F'_{\hat{a}_2}(x) = \left(\frac{x^n}{a^n}\right)' = \frac{nx^{n-1}}{a^n} \text{ за } x \in (0, a).$$

Тогаш, за математичкото очекување и дисперзијата на $\hat{a}_2$ имаме

$$E(\hat{a}_2) = \int_{-\infty}^{+\infty} x\, p_{\hat{a}_2}(x)\, dx = \int_0^a x \cdot \frac{nx^{n-1}}{a^n}\, dx = \frac{n}{a^n} \int_0^a x^n\, dx = \frac{na}{n+1} \xrightarrow[n \to \infty]{} a,$$

$$E(\hat{a}_2^2) = \int_{-\infty}^{+\infty} x^2\, p_{\hat{a}_2}(x)\, dx = \int_0^a x^2 \cdot \frac{nx^{n-1}}{a^n}\, dx = \frac{n}{a^n} \int_0^a x^{n+1}\, dx = \frac{na^2}{n+2},$$

$$D(\hat{a}_2) = E(\hat{a}_2^2) - (E(\hat{a}_2))^2 = \frac{na^2}{n+2} - \left(\frac{na}{n+1}\right) = \frac{na^2}{(n+2)(n+1)} \xrightarrow[n \to \infty]{} 0,$$

од каде следи дека $\hat{a}_2$ е конзистентен оценувач за a, но не е непристрасен оценувач за a.

Забелешка. Конзистентноста на $\hat{a}_2$ може да ја покажеме и по дефиниција, иако $\hat{a}_2$ не е непристрасен оценувач за a. Нека $0 < \varepsilon < a$ е произволен. Тогаш,

$$\begin{aligned}
1 \ &\geq \ P\{|\hat{a}_2 - a| < \varepsilon\} = P\{a - \varepsilon < \hat{a}_2 < a + \varepsilon\} = \int_{a-\varepsilon}^{a+\varepsilon} p_{\hat{a}_2}(x)\, dx = \\
&= \ \int_{a-\varepsilon}^{a} \frac{nx^{n-1}}{a^n}\, dx = \frac{n}{a^n} \cdot \frac{1}{n} \left(a^n - (a-\varepsilon)^n\right) = 1 - \left(\frac{a-\varepsilon}{a}\right)^n \xrightarrow[n \to \infty]{} 1.
\end{aligned}$$

За $\varepsilon > a$, тривијално се покажува дека $P\{|\hat{a}_2 - a| < \varepsilon\} = 1$. Значи, $\forall \varepsilon > 0$, $\lim\limits_{n \to \infty} P\{|\hat{a}_2 - a| < \varepsilon\} = 1$, па $\hat{a}_2$ е конзистентен оценувач за a.

Задача 4.5. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X кое има Бета распределба со параметри 1 и θ, каде $\theta > 0$ е непознат параметар. Покажи дека $\hat{\theta} = \frac{1}{\overline{X}_n} - 1$, каде $\overline{X}_n$ е средината на примерокот, е конзистентен оценувач за параметарот θ.

Ирена Стојковска

Решение. Случајната променлива X има Бета распределба со параметри $\alpha > 0$ и $\beta > 0$, ако има густина на распределба $p(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$, $x \in (0,1)$, каде $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$ е Бета функција за која важи $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Математичко очекување и дисперзија на X се $EX = \frac{\alpha}{\alpha+\beta}$ и $DX = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$ (покажи!).
Тогаш, за обележјето X кое има Бета распределба со параметри 1 и θ имаме дека $EX = \frac{1}{1+\theta}$ и $DX = \frac{\theta}{(2+\theta)(1+\theta)^2}$.
Бидејќи $X_1, ..., X_n$ се независни и еднакво распределени случајни променливи со $EX_i = \frac{1}{1+\theta} < +\infty$, $i = 1, ..., n$, од Теоремата на Хинчин, за низата $\{X_i\}$ важи слабиот закон на големите броеви, односно

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \frac{1}{1+\theta} \implies \frac{1}{\overline{X}_n} \xrightarrow{P} 1+\theta \implies \frac{1}{\overline{X}_n} - 1 \xrightarrow{P} \theta \implies \hat{\theta} \xrightarrow{P} \theta,$$

што значи дека $\hat{\theta}$ е конзистентен оценувач за θ.

Задачи за самостојна работа

Задача 4.6. Обележјето X има $\mathcal{U}(\theta, 2\theta)$ распределба, каде θ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. Покажи дека $T_n = \frac{n+1}{5n+4}(X_{(1)} + 2X_{(n)})$ е непристрасен оценувач за параметарот θ, каде $X_{(1)} = \min\{X_1, ..., X_n\}$ и $X_{(n)} = \max\{X_1, ..., X_n\}$.

Задача 4.7. Нека $X_1, ..., X_n$ се независни случајни променливи со $E(X_i) = \beta t_i$ и $D(X_i) = \sigma^2$, $i = 1, 2, ..., n$, каде $t_1, t_2, ..., t_n$ се познати константи, а β и σ^2 се непознати параметри. Покажи дека $\hat{\beta} = \sum_{i=1}^{n} t_i X_i \big/ \sum_{i=1}^{n} t_i^2$ е непристрасен оценувач за β. Определи доволен услов за да $\hat{\beta}$ е конзистентен оценувач за β.

Задача 4.8. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X, пришто $EX = a$ и $DX < +\infty$, каде a е непознат параметар. Покажи дека $\hat{a} = \frac{2}{n(n+1)}\sum_{k=1}^{n} kX_k$ е конзистентен оценувач за параметарот a.

Задача 4.9. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X, за кое $EX^{2k} < +\infty$. Покажи дека $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^k$ е конзистентен оценувач за $\mu_k = E(X - EX)^k$.

4.2 Најефикасни оценувачи и доволни статистики

Задача 4.10. Нека обележјето X има Поасонова распределба со закон на распределба
$$P(x,\lambda)=\frac{\lambda^x}{x!}\,e^{-\lambda},\ x=0,1,2,...,$$
каде $\lambda>0$ е непознат параметар и нека $(X_1,..,X_n)$ е примерок кој одговара на обележјето X.

а) Покажи дека $\overline{X}_n=\frac{1}{n}\sum_{i=1}^n X_i$ е непристрасен оценувач за параметарот λ и одреди ја неговата диспезија.

б) Покажи дека $\overline{X}_n$ е регуларен оценувач за λ.

в) Испитај ја ефикасноста на $\overline{X}_n$ како оценувач за λ.

Решение. а) Од $EX=\lambda$ и $DX=\lambda$ следи дека и за случајните променливи од примерокот важи $EX_i=\lambda$ и $DX_i=\lambda$, $i=1,...,n$. Тогаш,
$$E(\overline{X}_n)=E\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)=\frac{1}{n}\sum_{i=1}^n E(X_i)=\frac{1}{n}\sum_{i=1}^n \lambda=\frac{1}{n}\cdot n\lambda=\lambda,$$
односно $\overline{X}_n$ е непристрасен оценувач за параметарот λ. За неговата дисперзија имаме
$$D(\overline{X}_n)=D\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)=\frac{1}{n^2}\sum_{i=1}^n D(X_i)=\frac{1}{n^2}\sum_{i=1}^n \lambda=\frac{1}{n^2}\cdot n\lambda=\frac{\lambda}{n},$$
заради независноста на случајните променливи во примерокот.

б) Треба да ги провериме условите за регуларност за оценувачот $\overline{X}_n$ како оценувач за λ. Функцијата на подобност е
$$L(x,\lambda)=\prod_{i=1}^n P(x_i,\lambda)=\prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}\,e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\,e^{-n\lambda},\ x=(x_1,...,x_n)\in\{0,1,2,...\}^n.$$

(i) Множеството $A=\{x|L(x,\lambda)>0\}=\mathbb{R}^n$ не зависи од λ.

(ii) Првиот извод на $L(x,\lambda)$ по λ е
$$\frac{\partial L(x,\lambda)}{\partial\lambda}=\frac{\sum_{i=1}^n x_i\cdot\lambda^{\sum_{i=1}^n x_i-1}-n\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\,e^{-n\lambda}.$$

(iii) Ќе покажеме дека може да се диференцира по λ под знакот за сумата $\sum_{x\in\mathbb{R}^n}L(x,\lambda)$, односно дека $\frac{\partial}{\partial\lambda}\Big(\sum_{x\in\mathbb{R}^n}L(x,\lambda)\Big)=\sum_{x\in\mathbb{R}^n}\frac{\partial L(x,\lambda)}{\partial\lambda}$.

Бидејќи $\sum_{x\in\mathbb{R}^n} L(x,\lambda) = 1$ имаме дека

$$\frac{\partial}{\partial\lambda}\Big(\sum_{x\in\mathbb{R}^n} L(x,\lambda)\Big) = 0.$$

Од друга страна,

$$\begin{aligned}
\sum_{x\in\mathbb{R}^n} \frac{\partial L(x,\lambda)}{\partial\lambda} &= \sum_{x\in\mathbb{R}^n} \frac{\sum_{i=1}^n x_i \cdot \lambda^{\sum_{i=1}^n x_i - 1} - n\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\, e^{-n\lambda} = \\
&= \frac{n}{\lambda}\sum_{x\in\mathbb{R}^n} \frac{\frac{1}{n}\sum_{i=1}^n x_i \cdot \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\, e^{-n\lambda} - n\sum_{x\in\mathbb{R}^n}\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\, e^{-n\lambda} = \\
&= \frac{n}{\lambda}\sum_{x\in\mathbb{R}^n}\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big)\, L(x,\lambda) - n\sum_{x\in\mathbb{R}^n} L(x,\lambda) = \\
&= \frac{n}{\lambda}\, E(\overline{X}_n) - n\cdot 1 = \frac{n}{\lambda}\cdot\lambda - n = 0,
\end{aligned}$$

од каде заклучуваме дека важи $\frac{\partial}{\partial\lambda}\Big(\sum_{x\in\mathbb{R}^n} L(x,\lambda)\Big) = \sum_{x\in\mathbb{R}^n}\frac{\partial L(x,\lambda)}{\partial\lambda}$.

(iv) Ќе покажеме дека може да се диференцира по λ под знакот за сумата $\sum_{x\in\mathbb{R}^n} U(x)L(x,\lambda)$, односно дека $\frac{\partial}{\partial\lambda}\Big(\sum_{x\in\mathbb{R}^n} U(x)L(x,\lambda)\Big) = \sum_{x\in\mathbb{R}^n} U(x)\frac{\partial L(x,\lambda)}{\partial\lambda}$, каде $U(x) = \frac{1}{n}\sum_{i=1}^n x_i$.

Бидејќи $\sum_{x\in\mathbb{R}^n} U(x)L(x,\lambda) = E(U) = E(\overline{X}_n) = \lambda$ имаме дека

$$\frac{\partial}{\partial\lambda}\Big(\sum_{x\in\mathbb{R}^n} U(x)L(x,\lambda)\Big) = 1.$$

Од друга страна,

$$\begin{aligned}
\sum_{x\in\mathbb{R}^n} U(x)\frac{\partial L(x,\lambda)}{\partial\lambda} &= \sum_{x\in\mathbb{R}^n}\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big)\frac{\sum_{i=1}^n x_i\cdot\lambda^{\sum_{i=1}^n x_i - 1} - n\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\, e^{-n\lambda} = \\
&= \frac{n}{\lambda}\sum_{x\in\mathbb{R}^n}\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big)^2 L(x,\lambda) - n\sum_{x\in\mathbb{R}^n}\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big)L(x,\lambda) = \\
&= \frac{n}{\lambda}\, E((\overline{X}_n)^2) - n\, E(\overline{X}_n) = \frac{n}{\lambda}\cdot(\frac{\lambda}{n} + \lambda^2) - n\lambda = 1,
\end{aligned}$$

затоа што $E(\overline{X}_n)^2 = D(\overline{X}_n) + (E(\overline{X}_n))^2 = \frac{\lambda}{n} + \lambda^2$. Заклучуваме дека важи $\frac{\partial}{\partial\lambda}\Big(\sum_{x\in\mathbb{R}^n} U(x)L(x,\lambda)\Big) = \sum_{x\in\mathbb{R}^n} U(x)\frac{\partial L(x,\lambda)}{\partial\lambda}$, каде $U(x) = \frac{1}{n}\sum_{i=1}^n x_i$.

Ирена Стојковска

(v) Функцијата $f:\Theta\to\mathbb{R}$ (од Дефиниција 4.10) е $f(\lambda)=\lambda$.

Од (i)-(v) следи дека $\overline{X}_n$ е регуларен оценувач за λ.

в) Ја пресметуваме долната граница D_0 од Теоремата на Рао-Крамер (Теорема 4.3). За неа ни треба

$$\ln L(x,\lambda)=\Big(\sum_{i=1}^n x_i\Big)\ln\lambda-\ln\Big(\prod_{i=1}^n x_i!\Big)-n\lambda,$$

од каде

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

па затоа

$$\begin{aligned}
E\Big(\frac{\partial\ln L(\mathbf{X},\lambda)}{\partial\lambda}\Big)^2 &= E\Big(\Big(\sum_{i=1}^n X_i\Big)\,\frac{1}{\lambda}-n\Big)^2=E\Big(\frac{1}{\lambda^2}\Big(\sum_{i=1}^n X_i\Big)^2-\frac{2n}{\lambda}\Big(\sum_{i=1}^n X_i\Big)+n^2\Big)=\\
&= n^2\,E\Big(\frac{1}{\lambda^2}\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)^2-\frac{2}{\lambda}\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)+1\Big)=\\
&= n^2\,E\Big(\frac{1}{\lambda^2}\big(\overline{X}_n\big)^2-\frac{2}{\lambda}\big(\overline{X}_n\big)+1\Big)=\\
&= n^2\,\Big(\frac{1}{\lambda^2}E\big(\overline{X}_n\big)^2-\frac{2}{\lambda}E\big(\overline{X}_n\big)+1\Big)=\\
&= n^2\,\Big(\frac{1}{\lambda^2}(\frac{\lambda}{n}+\lambda^2)-\frac{2}{\lambda}\cdot\lambda+1\Big)=\frac{n}{\lambda}.
\end{aligned}$$

Бидејќи $f'(\lambda)=1$, за D_0 имаме

$$D_0=\frac{(f'(\lambda))^2}{E\Big(\frac{\partial\ln L(\mathbf{X},\lambda)}{\partial\lambda}\Big)^2}=\frac{1}{\frac{n}{\lambda}}=\frac{\lambda}{n}=D(\overline{X}_n),$$

од каде следи дека $\overline{X}_n$ е најефикасен оценувач за λ.

Забелешка. Изразот $E\Big(\frac{\partial\ln L(\mathbf{X},\lambda)}{\partial\lambda}\Big)^2$ може да се пресмета и со помош на информацијата на Фишер $I(\lambda)$, користејќи ја распределбата на обележјето X т.е. $P(x,\lambda)$. Така, имаме

$$\begin{aligned}
E\Big(\frac{\partial\ln L(\mathbf{X},\lambda)}{\partial\lambda}\Big)^2 &= n\,I(\lambda)=n\,E\Big(\frac{\partial\ln P(X,\lambda)}{\partial\lambda}\Big)^2=n\,E\Big(\frac{\partial}{\partial\lambda}\Big(X\ln\lambda-\ln X!-\lambda\Big)\Big)^2=\\
&= n\,E\Big(\frac{X}{\lambda}-1\Big)^2=n\,E\Big(\frac{X^2}{\lambda^2}-2\frac{X}{\lambda}+1\Big)=\frac{n}{\lambda^2}E(X^2)-\frac{2n}{\lambda}EX+n=\\
&= \frac{n}{\lambda^2}(\lambda+\lambda^2)-\frac{2n}{\lambda}\,\lambda+n=\frac{n}{\lambda}.
\end{aligned}$$

Задача 4.11. Нека обележјето X има експоненцијална распределба со густина на распределба

$$p(x,\theta) = \frac{1}{\theta}\, e^{-\frac{x}{\theta}},\ x > 0,$$

каде $\theta > 0$ е непознат параметар и нека $(X_1, .., X_n)$ е примерок кој одговара на обележјето X. Покажи дека $\hat{\theta} = n \cdot \min\{X_1, ..., X_n\}$ е непристрасен оценувач за θ. Испитај ја ефикасноста на $\hat{\theta}$ како оценувач за θ.

Решение. Нека $(X_1, ..., X_n)$ е примерок кој одговаа на обележејето X кое има експоненцијална распределба со функција на распределба

$$F(x,\theta) = \int_{-\infty}^{x} p(u,\theta)\, du = \int_0^x \frac{1}{\theta}\, e^{-\frac{u}{\theta}}\, du = 1 - e^{-\frac{x}{\theta}},\ x > 0.$$

Тогаш, за распределбата на оценувачот $\hat{\theta}$ имаме

$$\begin{aligned}
F_{\hat{\theta}}(x) &= P\{\hat{\theta} \le x\} = P\{n \cdot \min\{X_1, ..., X_n\} \le x\} = P\{\min\{X_1, ..., X_n\} \le \frac{x}{n}\} = \\
&= 1 - P\{\min\{X_1, ..., X_n\} > \frac{x}{n}\} = 1 - P\{X_1 > \frac{x}{n}, ..., X_n > \frac{x}{n}\} = \\
&= 1 - P\{X_1 > \frac{x}{n}\} \cdot ... \cdot P\{X_n > \frac{x}{n}\} = 1 - (1 - P\{X_1 \le \frac{x}{n}\}) \cdot ... \cdot (1 - P\{X_n \le \frac{x}{n}\}) = \\
&= 1 - (1 - F(\frac{x}{n}, \theta))^n = 1 - (e^{-\frac{x}{n\theta}})^n = 1 - e^{-\frac{x}{\theta}},\ x > 0,
\end{aligned}$$

односно и $\hat{\theta}$ има експоненцијална распределба со параметар θ, од каде $E(\hat{\theta}) = \theta$ и $D(\hat{\theta}) = \theta^2$, па $\hat{\theta}$ е непристрасен оценувач за θ.

Ќе покажеме дека за $\hat{\theta}$ важат условите за регуларност. Функцијата на подобност е

$$L(x,\theta) = \prod_{i=1}^{n} p(x_i,\theta) = \prod_{i=1}^{n} \frac{1}{\theta}\, e^{-\frac{x_i}{\theta}} = \frac{1}{\theta^n}\, e^{-\frac{1}{\theta}\sum_{i=1}^n x_i},\ x = (x_1, ..., x_n) \in R_+^n.$$

(i) Множеството $A = \{x | L(x,\theta) > 0\} = \mathbb{R}^n$ не зависи од θ.

(ii) Првиот извод на $L(x,\theta)$ по θ е

$$\frac{\partial L(x,\theta)}{\partial \theta} = \frac{\sum_{i=1}^n x_i - n\theta}{\theta^{n+2}}\, e^{-\frac{1}{\theta}\sum_{i=1}^n x_i}.$$

(iii) Ќе покажеме дека може да се диференцира по θ под знакот за интеграл $\int_{x\in\mathbb{R}^n} L(x,\theta)\, dx$, односно дека $\frac{\partial}{\partial\theta}\left(\int_{x\in\mathbb{R}^n} L(x,\theta)\, dx\right) = \int_{x\in\mathbb{R}^n} \frac{\partial L(x,\theta)}{\partial\theta}\, dx$.

Бидејќи $\int_{x\in\mathbb{R}^n} L(x,\theta)\,dx = 1$ имаме дека

$$\frac{\partial}{\partial\theta}\Big(\int_{x\in\mathbb{R}^n} L(x,\lambda)\,dx\Big) = 0.$$

Од друга страна,

$$\begin{aligned}
\int_{x\in\mathbb{R}^n} \frac{\partial L(x,\theta)}{\partial\theta}\,dx &= \int_{x\in\mathbb{R}^n} \frac{\sum_{i=1}^n x_i - n\theta}{\theta^{n+2}}\; e^{-\frac{1}{\theta}\sum_{i=1}^n x_i}\,dx = \\
&= \frac{n}{\theta^2}\int_{x\in\mathbb{R}^n} \frac{\frac{1}{n}\sum_{i=1}^n x_i}{\theta^n}\; e^{-\frac{1}{\theta}\sum_{i=1}^n x_i} - \frac{n}{\theta}\int_{x\in\mathbb{R}^n} \frac{1}{\theta^n}\; e^{-\frac{1}{\theta}\sum_{i=1}^n x_i} = \\
&= \frac{n}{\theta^2}\int_{x\in\mathbb{R}^n} \Big(\frac{1}{n}\sum_{i=1}^n x_i\Big)\; L(x,\theta) - \frac{n}{\theta}\int_{x\in\mathbb{R}^n} L(x,\theta) = \\
&= \frac{n}{\theta^2}\; E(\overline{X}_n) - \frac{n}{\theta}\cdot 1 = \frac{n}{\theta^2}\cdot\theta - \frac{n}{\theta} = 0,
\end{aligned}$$

затоа што $E(\overline{X}_n) = EX = \theta$, каде X е обележјето, па заклучуваме дека важи $\frac{\partial}{\partial\theta}\Big(\int_{x\in\mathbb{R}^n} L(x,\theta)\,dx\Big) = \int_{x\in\mathbb{R}^n} \frac{\partial L(x,\theta)}{\partial\theta}\,dx$.

(iv) Ќе покажеме дека може да се диференцира по θ под знакот за интеграл $\int_{x\in\mathbb{R}^n} U(x)L(x,\theta)$, односно дека $\frac{\partial}{\partial\theta}\Big(\int_{x\in\mathbb{R}^n} U(x)L(x,\theta)\Big) = \int_{x\in\mathbb{R}^n} U(x)\frac{\partial L(x,\theta)}{\partial\theta}$, каде $U(x) = n\cdot\min\{x_1,...,x_n\}$.

Бидејќи $\int_{x\in\mathbb{R}^n} U(x)L(x,\theta) = E(U) = E(\hat{\theta}) = \theta$ имаме дека

$$\frac{\partial}{\partial\theta}\Big(\int_{x\in\mathbb{R}^n} U(x)L(x,\theta)\Big) = 1.$$

Од друга страна,

$$\begin{aligned}
\int_{x\in\mathbb{R}^n} U(x)\frac{\partial L(x,\theta)}{\partial\theta} &= \int_{x\in\mathbb{R}^n} \big(n\cdot\min\{x_1,...,x_n\}\big)\frac{\sum_{i=1}^n x_i - n\theta}{\theta^{n+2}}\; e^{-\frac{1}{\theta}\sum_{i=1}^n x_i} = \\
&= \frac{n}{\theta^2}\int_{x\in\mathbb{R}^n} \big(n\cdot\min\{x_1,...,x_n\}\big)\;\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big)\; L(x,\theta) - \\
&\quad -\frac{n}{\theta}\int_{x\in\mathbb{R}^n} \big(n\cdot\min\{x_1,...,x_n\}\big)L(x,\theta) = \\
&= \frac{n}{\theta^2}\; E(\hat{\theta}\,\overline{X}_n) - \frac{n}{\theta}\; E(\hat{\theta}) = 1,
\end{aligned}$$

при што последното равенство важи за $n=1$, затоа што тогаш $E(\hat{\theta}\,\overline{X}_n) = E(X_1^2) = 2\theta^2$ и од непристрасноста на $\hat{\theta}$ т.е. $E(\hat{\theta}) = \theta$ се добива равенството. (За останатите вредности на n треба да се провери!). Заклучуваме дека за $n=1$ важи $\frac{\partial}{\partial\theta}\Big(\int_{x\in\mathbb{R}^n} U(x)L(x,\theta)\Big) = \int_{x\in\mathbb{R}^n} U(x)\frac{\partial L(x,\theta)}{\partial\theta}$, каде $U(x) = n\cdot\min\{x_1,...,x_n\}$.

Ирена Стојковска

(v) Функцијата $f : \Theta \to \mathbb{R}$ (од Дефиниција 4.10) е $f(\theta) = \theta$.

Од (i)-(v) следи дека за $n = 1$, оценувачот $\hat{\theta}$ е регуларен оценувач за θ.

Следно, ја пресметуваме долната граница D_0 од Теоремата на Рао-Крамер (Теорема 4.3). За неа ни треба

$$\ln L(x,\theta) = -n\ \ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i,$$

од каде

$$\frac{\partial \ln L(x,\theta)}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i,$$

па затоа

$$\begin{aligned}
E\Big(\frac{\partial \ln L(\mathbf{X},\theta)}{\partial\theta}\Big)^2 &= E\Big(-\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} X_i\Big)^2 = E\Big(\frac{n^2}{\theta^2} - \frac{2n}{\theta^3}\sum_{i=1}^{n} X_i + \frac{1}{\theta^4}\Big(\sum_{i=1}^{n} X_i\Big)^2\Big) = \\
&= n^2\, E\Big(\frac{1}{\theta^2} - \frac{2}{\theta^3}\cdot\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{\theta^4}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big)^2\Big) = \\
&= n^2\, E\Big(\frac{1}{\theta^2} - \frac{2}{\theta^3}\cdot\overline{X}_n + \frac{1}{\theta^4}\big(\overline{X}_n\big)^2\Big) = \\
&= n^2\, \Big(\frac{1}{\theta^2} - \frac{2}{\theta^3}\cdot E\big(\overline{X}_n\big) + \frac{1}{\theta^4}E\big(\overline{X}_n\big)^2\Big) = \\
&= n^2\, \Big(\frac{1}{\theta^2} - \frac{2}{\theta^3}\cdot\theta + \frac{1}{\theta^4}\big(\frac{\theta^2}{n} + \theta^2\big)\Big) = \frac{n}{\theta^2}.
\end{aligned}$$

Бидејќи $f'(\theta) = 1$, за D_0 имаме

$$D_0 = \frac{(f'(\theta))^2}{E\Big(\frac{\partial \ln L(\mathbf{X},\theta)}{\partial\theta}\Big)^2} = \frac{1}{\frac{n}{\theta^2}} = \frac{\theta^2}{n},$$

од каде за $n = 1$ имаме дека $D_0 = \theta^2 = D(\hat{\theta})$, односно непристрасниот регуларен оценувач $\hat{\theta}$ е најефикасен оценувач за θ.

Задача 4.12. Обележјето X има густина на распределба

$$p(x,\theta) = e^{-(x-\theta)},\ \theta < x < +\infty,$$

каде θ е непознат параметар. Покажи дека $U = \min\{X_1, ..., X_n\}$ е доволна статистика за параметарот θ, каде $(X_1, .., X_n)$ е примерок кој одговара на обележјето X.

Ирена Стојковска

Решение. Функцијата на распределба на обележјето X е

$$F(x,\theta) = \int_{-\infty}^{x} p(u,\theta)\, du = \int_{\theta}^{x} e^{-(u-\theta)}\, du = 1 - e^{-(x-\theta)},\ x > \theta.$$

Тогаш, за функцијата на распределба на оценувачот U имаме

$$\begin{aligned} F_U(x) &= P\{U \le x\} = P\{\min\{X_1,...,X_n\} \le x\} = \\ &= 1 - P\{\min\{X_1,...,X_n\} > x\} = 1 - P\{X_1 > x,...,X_n > x\} = \\ &= 1 - P\{X_1 > x\} \cdot ... \cdot P\{X_n > x\} = 1 - (1 - P\{X_1 \le x\}) \cdot ... \cdot (1 - P\{X_n \le x\}) = \\ &= 1 - (1 - F(x,\theta))^n = 1 - (e^{-(x-\theta)})^n = 1 - e^{-n(x-\theta)},\ x > \theta, \end{aligned}$$

од каде густината на распределба на U е

$$p_U(x,\theta) = F_U'(x,\theta) = ne^{-n(x-\theta)},\ x > \theta.$$

Следно, ја изведуваме условната распределба на примерокот $(X_1,..,X_n)$ при услов $U = t$. Нека $(x_1,...,x_n)$ е произволен елемент од доменот $x_i > \theta, i = 1,...,n$, и $t > \theta$. Тогаш,

$$L(x,\theta|U=t) = \frac{L(x,\theta)}{p_U(t)} = \frac{\prod_{i=1}^{n} p(x_i,\theta)}{p_U(t)} = \frac{\prod_{i=1}^{n} e^{-(x_i-\theta)}}{ne^{-n(t-\theta)}} = \frac{e^{-\sum_{i=1}^{n}(x_i-\theta)}}{ne^{-n(t-\theta)}} = \frac{e^{-\sum_{i=1}^{n} x_i}}{ne^{-nt}},$$

што не зависи од оценуваниот праметар θ. Кога $t \le \theta$, условната распределба е 0, па тривијално не зависи од θ. Според тоа, заклучуваме дека оценувачот $U = \min\{X_1,...,X_n\}$ е доволна статистика за параметарот θ.

Задача 4.13. Обележјето X има Поасонова $\mathcal{P}(\lambda)$ распределба, каде $\lambda > 0$ е непознат параметар. Врз основа на примерок $(X_1,..,X_n)$ кој одговара на обележјето X, определи една доволна статистика за параметарот λ.

Решение. Распределбата на обележјето X е

$$P(x,\lambda) = \frac{\lambda^x}{x!}\, e^{-\lambda},\ x = 0,1,2,...,$$

од каде за функцијата на подобност т.е. распределбата на примерокот $(X_1,...,X_n)$ имаме

$$\begin{aligned} L(x,\lambda) &= \prod_{i=1}^{n} P(x_i,\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}\, e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\, e^{-n\lambda} = \\ &= e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i} \cdot \frac{1}{\prod_{i=1}^{n} x_i!} = g(U(x),\theta) \cdot h(x), \end{aligned}$$

за $x = (x_1,...,x_n) \in \{0,1,2,...\}^n$, каде $U(x) = \sum_{i=1}^{n} x_i$, $g(U(x),\theta) = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}$ и $h(x) = \frac{1}{\prod_{i=1}^{n} x_i!}$. Според теоремата за факторизација следи дека $U = \sum_{i=1}^{n} X_i$ е доволна статистика за λ.

Ирена Стојковска

Задача 4.14. Обележјето X има Гаусова $\mathcal{N}(a, \sigma^2)$ распределба, каде a и $\sigma^2 > 0$ се непознати параметри. Нека $(X_1, .., X_n)$ е примерок кој одговара на обележјето X. Покажи дека $(\overline{X}_n, \overline{S}_n^2)$ е доволна статистика за (a, σ^2), каде $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ и $\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ се средината и дисперзијата на примерокот соодветно.

Решение. Густината на распределба на обележјето X е

$$p(x, a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-a)^2}{2\sigma^2}},\ x \in \mathbb{R},$$

од каде за распределбата на примерокот $(X_1, ..., X_n)$ имаме

$$\begin{aligned}
L(x, a, \sigma^2) &= \prod_{i=1}^{n} p(x_i, a, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x_i-a)^2}{2\sigma^2}} = \\
&= \frac{1}{(\sqrt{2\pi\sigma^2})^n}\, e^{-\sum_{i=1}^{n}\frac{(x_i-a)^2}{2\sigma^2}} = \frac{1}{(\sqrt{2\pi\sigma^2})^n}\, e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-a)^2} = \\
&= \frac{1}{(\sqrt{2\pi\sigma^2})^n}\, \exp\{-\frac{1}{2\sigma^2}\Big(\sum_{i=1}^{n}(x_i-\overline{x})^2 + 2(\overline{x}-a)\sum_{i=1}^{n}(x_i-\overline{x}) + n(\overline{x}-a)^2\Big)\} = \\
&= \frac{1}{(\sqrt{2\pi\sigma^2})^n}\, \exp\{-\frac{1}{2\sigma^2}(\overline{s}^2 + n(\overline{x}-a)^2)\} = g(\overline{x}, \overline{s}^2, a, \sigma^2)\cdot h(x),
\end{aligned}$$

за $x = (x_1, ..., x_n) \in \mathbb{R}^n$, каде $g(\overline{x}, \overline{s}^2, a, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n}\ \exp\{-\frac{1}{2\sigma^2}(\overline{s}^2 + n(\overline{x}-a)^2)\}$ и $h(x) = 1$, па според теоремата за факторизација следи дека $(\overline{X}_n, \overline{S}_n^2)$ е доволна статистика за (a, σ^2).

Задачи за самостојна работа

Задача 4.15. Нека обележјето X има биномна $\mathcal{B}(m, p)$ распределба, каде m е познат параметар, а $0 < p < 1$ е непознат параметат и нека $(X_1, .., X_n)$ е примерок кој одгвара на обележјето X. Покажи дека $U_n = \frac{1}{m}\overline{X}_n = \frac{1}{mn}\sum_{i=1}^{n} X_i$ е најефикасен оценувач за параметарот p.

Задача 4.16. Нека обележјето X има рамномерна $\mathcal{U}(\theta_1, \theta_2)$ распределба, каде $\theta_1 < \theta_2$ се непознати параметри и нека $(X_1, .., X_n)$ е примерок кој одгвара на обележјето X.

а) Покажи дека $U_n = \frac{X_{(1)}+X_{(n)}}{2}$ е најефикасен оценувач за медијаната $\frac{\theta_1+\theta_2}{2}$.

б) Покажи дека $V_n = \frac{n+1}{n-1}(X_{(n)} - X_{(1)})$ е најефикасен оценувач за опсегот $\theta_2 - \theta_1$.

Задача 4.17. Нека обележјето X има Гама распределба со параметри p и $\frac{1}{\theta}$ т.е има густина на распределба

$$p(x,\theta) = \frac{\theta^p x^{p-1}}{\Gamma(p)}\, e^{-\theta x},\ x > 0,$$

каде $\theta > 0$ е непознат параметар и нека $(X_1, .., X_n)$ е примерок кој одгвара на обележјето X. Нека $\hat{\theta} = \frac{p}{\overline{X}_n}$ е оценувач за θ, каде $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_n$ е средината на примерокот. Провери дали $\hat{\theta}$ е непристрасен оценувач за θ, ако не е, корегирај го до непристрасност. Испитај ја ефикасноста на корегираниот оценувач како оценувач за θ.
Упатство. Оценувачот $\hat{\theta}_1 = \frac{np-1}{np}\hat{\theta}$ е корекцијата до непристрасност и тој е асимптотски најефикасен оценувач за θ.

Задача 4.18. Обележјето X има Гаусова $\mathcal{N}(0, \sigma^2)$ распределба, каде $\sigma^2 > 0$ е непознат параметар. Врз основа на примерок $(X_1, .., X_n)$ кој одговара на обележјето X, определи една доволна статистика за параметарот σ^2.

Задача 4.19. Обележјето X има густина на распределба

$$p(x,\theta) = \theta x^{(\theta-1)},\ 0 < x < 1,$$

каде $\theta > 0$ е непознат параметар. Покажи дека $U = X_1 \cdot ... \cdot X_n$ е доволна статистика за параметарот θ, каде $(X_1, .., X_n)$ е примерок кој одговара на обележјето X.

Задача 4.20. Нека обележјето X прима вредности $x_1, ..., x_k$ со веројатности $p_1, .., p_k$ соодветно, при што $\sum_{i=1}^{k} p_i = 1$ и нека $(X_1, .., X_n)$ е примерок кој одговара на обележјето X. Нека U_{nj} е број на членови од примерокот $(X_1, .., X_n)$ кои примаат вредност x_j. Покажи дека $(U_{n1}, ..., U_{nk})$ е доволна статистика за $(p_1, ..., p_k)$.

4.3 Методи за наоѓање на оценувачи

Задача 4.21. Нека обележјето X има геометриска распределба $Geo(p)$, каде $0 < p < 1$ е непознат параметар. Со метод на моменти, врз основа на примерок $(X_1, ..., X_n)$ кој одговара на обележјето X, најди оценувач за параметарот p.

Решение. Распределбата на обележјето X е

$$P(k, p) = p(1-p)^k,\ k = 0, 1, 2, ...,$$

па првиот момент на X е

$$E(X) = \frac{1-p}{p}.$$

Сега, според методот на моменти, првиот момент го заменуваме со неговиот оценувач, односно првиот момент на примерокот $Z_{n,1} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, а параметарот p го заменуваме со неговиот оценувач $\hat{p}$. Така, добиваме

$$\overline{X}_n = \frac{1-\hat{p}}{\hat{p}},$$

од каде

$$\hat{p} = \frac{1}{1+\overline{X}_n}$$

е оценувач на параметарот p најден со методот на моменти.

Задача 4.22. Нека обележјето X има рамномерна распределба $\mathcal{U}(\theta_1, \theta_1 + \theta_2)$, каде θ_1 и $\theta_2 > 0$ се непознати параметри. Со метод на моменти, врз основа на примерок $(X_1, ..., X_n)$ кој одговара на обележјето X, најди оценувачи за параметрите θ_1 и θ_2.

Решение. Од $X \sim \mathcal{U}(\theta_1, \theta_1 + \theta_2)$ имаме дека

$$E(X) = \frac{\theta_1 + (\theta_1 + \theta_2)}{2} = \frac{2\theta_1 + \theta_2}{2} \text{ и } D(X) = \frac{(\theta_1 + \theta_2 - \theta_1)^2}{12} = \frac{\theta_2^2}{12},$$

од каде за вториот момент имаме

$$E(X^2) = D(X) + (E(X))^2 = \frac{\theta_2^2}{12} + \left(\frac{2\theta_1 + \theta_2}{2}\right)^2 = \frac{3\theta_1^2 + 3\theta_1\theta_2 + \theta_2^2}{3}.$$

Според методот на моменти, за да ги најдеме оценувачите $\hat{\theta}_1$ и $\hat{\theta}_2$ на θ_1 и θ_2 соодветно, го решаваме следниот систем

$$\begin{cases} Z_{n,1} = \dfrac{2\hat{\theta}_1 + \hat{\theta}_2}{2} \\ Z_{n,2} = \dfrac{3\hat{\theta}_1^2 + 3\hat{\theta}_1\hat{\theta}_2 + \hat{\theta}_2^2}{3} \end{cases},$$

каде $Z_{n,1} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ и $Z_{n,2} = \frac{1}{n}\sum_{i=1}^n X_i^2$ се соодветно првиот и вториот момент на примерокот. Земајќи во предвид дека $\theta_2 > 0$, решенија на последниот систем кои се од наш интерес се

$$\begin{cases} \hat{\theta}_1 = Z_{n,1} - \sqrt{3(Z_{n,2} - Z_{n,1}^2)} \\ \hat{\theta}_2 = 2\sqrt{3(Z_{n,2} - Z_{n,1}^2)} \end{cases},$$

односно оценувачи со метод на моменти за θ_1 и θ_2 се

$$\begin{cases} \hat{\theta}_1 = \overline{X}_n - \sqrt{3\Big(\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2\Big)} = \overline{X}_n - \sqrt{3\overline{S}_n^2} \\ \hat{\theta}_2 = 2\sqrt{3\Big(\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2\Big)} = 2\sqrt{3\overline{S}_n^2} \end{cases}.$$

Задача 4.23. Нека обележјето X има χ_m^2 распределба т.е. неговата густина на распределба е

$$p(x,m) = \frac{x^{\frac{m}{2}-1}}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})}\, e^{-\frac{x}{2}},\ x > 0,$$

каде m е непознат параметар. Со метод на моменти, врз основа на примерок $(X_1, ..., X_n)$ кој одговара на обележјето X, најди два оценувачи за параметарот m. Испитај ја нивната непристрасност и конзистентност.

Решение. За случајната променлива X со χ_m^2 распределба имаме дека $E(X) = m$ и $D(X) = 2m$, од каде $E(X^2) = DX + (E(X))^2 = 2m + m^2$.

Според методот на моменти, бидејќи $E(X) = m$, следи дека еден оценувач за параметарот m е оценувачот $\hat{m} = Z_{n,1} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Додека пак, од $E(X^2) = 2m + m^2$, вториот оценувач $\tilde{m}$ може да го добиеме од $Z_{n,2} = 2\tilde{m} + \hat{m}^2$, од каде $\tilde{m} = \frac{1}{2}(Z_{n,2} - \hat{m}^2) = \frac{1}{2}\overline{S}_n^2$.

Да ја испитаме непристрасноста на оценувачите $\hat{m}$ и $\tilde{m}$. Од

$$E(\hat{m}) = E(\overline{X}_n) = E\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big) = \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n}\cdot n \cdot m = m,$$

следи дека $\hat{m}$ е непристрасен оценувач за m. Потоа, од

$$\begin{aligned} E(\tilde{m}) &= E\Big(\frac{1}{2}\overline{S}_n^2\Big) = E\Big(\frac{1}{2}\Big(\frac{1}{n}\sum_{i=1}^n X_i^2 - \Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)^2\Big)\Big) = \\ &= \frac{1}{2}\Big(\frac{1}{n}\sum_{i=1}^n E(X_i^2) - \frac{1}{n^2}E\Big(\sum_{i=1}^n X_i\Big)^2\Big) = \\ &= \frac{1}{2}\Big(\frac{1}{n}\cdot n \cdot (2m + m^2) - \frac{1}{n^2}\cdot(2nm + n^2m^2)\Big) = \frac{(n-1)m}{n} \to m,\ n \to \infty, \end{aligned}$$

од каде заклучуваме дека $\tilde{m}$ е асимптотски непристрасен оценувач за m. Притоа, искористено е дека $X_1, .., X_n$ се независни и еднакво распределени со χ^2_m распределби, од каде следи дека $\sum_{i=1}^n X_i$ има χ^2_{nm} распределба, па затоа $E(\sum_{i=1}^n X_i) = nm$ и $D(\sum_{i=1}^n X_i) = 2nm$, од каде

$$E(\sum_{i=1}^n X_i)^2 = D(\sum_{i=1}^n X_i) + (E(\sum_{i=1}^n X_i))^2 = 2nm + n^2m^2.$$

Оценувачите најдени со метод на моменти се конзистентни. Конзистентноста ќе ја покажеме и со испитување на нивната дисперзија. Имено од

$$D(\hat{m}) = D(\overline{X}_n) = \frac{DX}{n} = \frac{2m}{n} \to 0,\ n \to \infty,$$

и бидејќи $E(\hat{m}) = m$, следи дека $\hat{m}$ е конзистентен оценувач за m. За диспезијата на оценувачот $\tilde{m}$ имаме

$$D(\tilde{m}) = D(\frac{1}{2}\overline{S}^2_n) = \frac{1}{4}D(\overline{S}^2_n) = \frac{1}{4} \cdot \frac{(n-1)^2}{n^3}(\mu_4 - \frac{n-3}{n-1}\mu_2^2) \to 0,\ n \to \infty,$$

и од $E(\tilde{m}) \to m,\ n \to \infty$, заклучуваме дека $\tilde{m}$ е конзистентен оценувач за m. Притоа, $\mu_2 = E(X - EX)^2$ и $\mu_4 = E(X - EX)^4$ се соодветно вториот и четвртиот централен момент на обележјето X кои не зависат од големината на примерокот n.

Задача 4.24. Нека обележјето X има густина на распределба

$$p(x, \theta) = \theta\, x^{\theta-1},\ 0 < x < 1,$$

каде $\theta > 0$ е непознат параметар. Со метод на моменти, врз основа на податоците од табелата, кои одговараат на обележјето X, најди оценка за параметарот θ.

x_k	0,1	0,2	0,3	0,4	0,5	0,6
f_k	5	12	35	28	17	3

Решение. За математичкото очекување на X имаме

$$E(X) = \int_{-\infty}^{+\infty} x \cdot p(x, \theta)\, dx = \int_0^1 x \cdot \theta\, x^{\theta-1}\, dx = \theta \int_0^1 x^\theta\, dx = \frac{\theta}{\theta + 1}.$$

Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. Според методот на моменти, за оценувачот $\hat{\theta}$ важи

$$Z_{n,1} = \frac{\hat{\theta}}{\hat{\theta} + 1},$$

од каде

$$\hat{\theta} = \frac{Z_{n,1}}{1 - Z_{n,1}} = \frac{\overline{X}_n}{1 - \overline{X}_n}.$$

Ја пресметуваме аритметичката средина на дадените податоци, за кои $n = 5 + 12 + 35 + 28 + 17 + 3 = 100$,

$$\overline{x} = \frac{1}{100}(5 \cdot 0,1 + 12 \cdot 0,2 + 35 \cdot 0,3 + 28 \cdot 0,4 + 17 \cdot 0,5 + 3 \cdot 0,6) = 0,349,$$

и добиваме дека оценка за параметарот θ е

$$\theta = \frac{\overline{x}}{1 - \overline{x}} = \frac{0,349}{1 - 0,349} \approx 0,536098.$$

Задача 4.25. Обележјето X има густина на распределба

$$p(x, \theta) = e^{\theta - x},\ x \geq \theta,$$

каде θ е непознат параметар. Врз основа на примерокот $(X_1, ..., X_n)$, кој одговара на обележјето X, најди максимално подобен оценувач за θ, а потоа испитај ги непристрасноста и конистентноста на најдениот оценувач.

Решение. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. Тогаш, за $x = (x_1, ..., x_n)$, функцијата на подобност е

$$L(x, \theta) = \prod_{i=1}^{n} p(x_n, \theta) = \prod_{i=1}^{n} e^{\theta - x_i} = e^{n\theta - \sum_{i=1}^{n} x_i},\ x_i \geq \theta, i = 1, ..., n.$$

Бидејќи доменот на функцијата на подобност зависи од оценуваниот параметар, максимално подобниот оценувач го бараме со директна максимизација на функцијата на подобност. Така, од

$$\frac{\partial L(x, \theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left(e^{n\theta - \sum_{i=1}^{n} x_i}\right) = n e^{n\theta - \sum_{i=1}^{n} x_i} > 0,\ x_i \geq \theta, i = 1, ..., n,$$

следи дека за $x_i \geq \theta, i = 1, ..., n$, што е еквивалентно со $\min\{x_1, ..., x_n\} \geq \theta$, функцијата $L(x, \theta)$ е растечка по θ, па таа го достигнува својот максимум за најголемата можна вредност за θ, односно за $\theta = \min\{x_1, ..., x_n\}$. Значи, максимално подобен оценувач за θ е $\hat{\theta} = \min\{X_1, ..., X_n\}$.

За функцијата на распределба на $\hat{\theta}$ имаме

$$\begin{aligned} F_{\hat{\theta}}(x) &= P\{\hat{\theta} \leq x\} = P\{\min\{X_1, ..., X_n\} \leq x\} = 1 - P\{\min\{X_1, ..., X_n\} > x\} = \\ &= 1 - P\{X_1 > x, ..., X_n > x\} = 1 - P\{X_1 > x\} \cdot ... \cdot P\{X_n > x\} = \\ &= 1 - (1 - P\{X_1 \leq x\}) \cdot ... \cdot (1 - P\{X_n \leq x\}) = 1 - (1 - F(x))^n, \end{aligned}$$

каде $F(x)$ е функцијата на распределба на облежјето X. Имаме

$$F(x) = \int_{-\infty}^{x} p(u,\theta)\, du = \int_{\theta}^{x} e^{\theta-u}\, du = 1 - e^{\theta-x},\ x \geq \theta,$$

од каде

$$F_{\hat{\theta}}(x) = 1 - (1-F(x))^n = 1 - e^{n(\theta-x)},\ x \geq \theta,$$

па густината на распределба на $\hat{\theta}$ е

$$p_{\hat{\theta}}(x) = F'_{\hat{\theta}}(x) = ne^{n(\theta-x)},\ x \geq \theta.$$

Тогаш, математичкото очекување на $\hat{\theta}$ е

$$\begin{aligned}
E(\hat{\theta}) &= \int_{-\infty}^{+\infty} x \cdot p_{\hat{\theta}}(x)\, dx = \int_{\theta}^{+\infty} x \cdot ne^{n(\theta-x)}\, dx = \\
&= n\Big(-\frac{1}{n}\, x\, e^{n(\theta-x)}\Big|_{\theta}^{+\infty} + \frac{1}{n}\int_{\theta}^{+\infty} e^{n(\theta-x)}\, dx\Big) = \\
&= n\Big(\frac{\theta}{n} - \frac{1}{n^2}e^{n(\theta-x)}\Big|_{\theta}^{\infty}\Big) = n\Big(\frac{\theta}{n} + \frac{1}{n^2}\Big) = \theta + \frac{1}{n} \to \theta,\ n \to \infty,
\end{aligned}$$

односно $\hat{\theta}$ е асимптотски непристрасен оценувач за θ. Да забележиме дека коригираниот оценувач $\tilde{\theta} = \hat{\theta} - \frac{1}{n}$ е непристрасен оценувач за θ.

Да ја испитаме конзистентнста на $\hat{\theta}$. Нека $\varepsilon > 0$ е произволен. Тогаш,

$$\begin{aligned}
P\{|\hat{\theta}-\theta| \geq \varepsilon\} &= 1 - P\{|\hat{\theta}-\theta| < \varepsilon\} = 1 - P\{-\varepsilon < \hat{\theta}-\theta < \varepsilon\} = \\
&= 1 - P\{\theta-\varepsilon < \hat{\theta} < \theta+\varepsilon\} = 1 - \int_{\theta-\varepsilon}^{\theta+\varepsilon} p_{\hat{\theta}}(x)\, dx = \\
&= 1 - \int_{\theta}^{\theta+\varepsilon} ne^{n(\theta-x)}\, dx = 1 + n \cdot \frac{1}{n}\, e^{n(\theta-x)}\Big|_{\theta}^{\theta+\varepsilon} = e^{-n\varepsilon} \to 0,\ n \to \infty,
\end{aligned}$$

од каде следи дека $\hat{\theta}$ е конзистентен оценувач за θ.

Задача 4.26. Обележјето X има густина на распределба

$$p(x,\theta) = \frac{x}{\theta^2}\, e^{-\frac{x}{\theta}},\ x > 0,$$

каде $\theta > 0$ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X.

а) Врз основа на примерокот $(X_1, ..., X_n)$, најди максимално подобен оценувач за θ, а потоа испитај ги непристрасноста и ефикасноста на најдениот оценувач.

б) Покажи дека оценувачот $\frac{1}{3n}\sum_{i=1}^{n} X_i^2$ е непристрасен оценувач за DX.

Ирена Стојковска

Решение. а) Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. Тогаш, функцијата на подобност за $x = (x_1, ..., x_n)$ е

$$L(x,\theta) = \prod_{i=1}^{n} p(x_n,\theta) = \prod_{i=1}^{n} \frac{x_i}{\theta^2}\, e^{-\frac{x_i}{\theta}} = \frac{\prod_{i=1}^{n} x_i}{\theta^{2n}}\, e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i},\ x_i > 0, i = 1, ..., n,$$

од каде

$$\ln L(x,\theta) = \ln \prod_{i=1}^{n} x_i - 2n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i.$$

За да го најдеме максимално подобниот оценувач за θ $(\theta > 0)$, ја решаваме равенката на подобност

$$0 = \frac{\partial \ln L(x,\theta)}{\partial\theta} = -\frac{2n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i,$$

чие решение е $\theta = \frac{1}{2n}\sum_{i=1}^{n} x_i$, од каде максимално подобен оценувач за θ е $\hat{\theta} = \frac{1}{2n}\sum_{i=1}^{n} X_i = \frac{1}{2}\overline{X}_n$, каде $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ е средината на примерокот.

Да ја испитаме непристрассноста на $\hat{\theta}$. Математичкото очекување на X е

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x\cdot p(x,\theta)\,dx = \int_0^{+\infty} x\cdot\frac{x}{\theta^2}\,e^{-\frac{x}{\theta}}\,dx = \\ &= \frac{1}{\theta^2}\Big(-\theta x^2 e^{-\frac{x}{\theta}}\Big|_0^{+\infty} + 2\theta\int_0^{+\infty} x e^{-\frac{x}{\theta}}dx\Big) = \frac{2}{\theta}\int_0^{+\infty} x e^{-\frac{x}{\theta}}dx = \\ &= \frac{2}{\theta}\Big(-\theta x e^{-\frac{x}{\theta}}\Big|_0^{+\infty} + \theta\int_0^{+\infty} e^{-\frac{x}{\theta}}dx\Big) = 2\int_0^{+\infty} e^{-\frac{x}{\theta}}dx = \\ &= -2\theta e^{-\frac{x}{\theta}}\Big|_0^{+\infty} = 2\theta, \end{aligned}$$

од каде

$$E(\hat{\theta}) = E\Big(\frac{1}{2n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{2n}\sum_{i=1}^{n} E(X_i) = \frac{1}{2n}\cdot n\cdot 2\theta = \theta,$$

односно $\hat{\theta}$ е непристрасен оценувач за θ.

Оценувачот $\hat{\theta}$ е регуларен оценувач за θ (покажи!).

За да ја испитаме ефикасниста на $\hat{\theta}$, прво ги наоѓаме вториот момент и дисперзијата на X. За вториот момент на X имаме

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2\cdot p(x,\theta)\,dx = \int_0^{+\infty} x^2\cdot\frac{x}{\theta^2}\,e^{-\frac{x}{\theta}}\,dx = \\ &= \frac{1}{\theta^2}\Big(-\theta x^3 e^{-\frac{x}{\theta}}\Big|_0^{+\infty} + 3\theta\int_0^{+\infty} x^2 e^{-\frac{x}{\theta}}dx\Big) = \frac{3}{\theta}\int_0^{+\infty} x^2 e^{-\frac{x}{\theta}}dx = \\ &= 3\theta\int_0^{+\infty}\frac{x^2}{\theta^2}e^{-\frac{x}{\theta}}dx = 3\theta\cdot E(X) = 3\theta\cdot 2\theta = 6\theta^2, \end{aligned}$$

од каде дисперзијата на X е

$$D(X) = E(X^2) - (E(X))^2 = 6\theta^2 - (2\theta)^2 = 2\theta^2.$$

Сега, изразот $E\Big(\frac{\partial \ln L(\mathbf{X},\theta)}{\partial\theta}\Big)^2$ го пресметуваме со помош на информацијата на Фишер $I(\theta)$, користејќи ја распределбата на обележјето X т.е. $p(x,\theta)$. Па, имаме

$$\begin{aligned}
E\Big(\frac{\partial \ln L(\mathbf{X},\theta)}{\partial\theta}\Big)^2 &= n\, I(\theta) = n\, E\Big(\frac{\partial \ln p(X,\theta)}{\partial\theta}\Big)^2 = n\, E\Big(\frac{\partial}{\partial\theta}\Big(\ln X - 2\ln\theta - \frac{X}{\theta}\Big)\Big)^2 = \\
&= n\, E\Big(-\frac{2}{\theta} + \frac{X}{\theta^2}\Big)^2 = \frac{n}{\theta^4}\, E(-2\theta + X)^2 = \frac{n}{\theta^4}\, E(4\theta^2 - 4\theta X + X^2) = \\
&= \frac{n}{\theta^4}\,(4\theta^2 - 4\theta E(X) + E(X^2)) = \frac{n}{\theta^4}\,(4\theta^2 - 4\theta\cdot 2\theta + 6\theta^2) = \frac{2n}{\theta^2}.
\end{aligned}$$

Од друга страна, заради независносата на случајните променливи во примерокот, за дисперзијата на оценувачот $\hat{\theta}$ имаме

$$D(\hat{\theta}) = D\Big(\frac{1}{2n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{4n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{4n^2}\cdot n\cdot 2\theta^2 = \frac{\theta^2}{2n}.$$

Бидејќи $f(\theta) = \theta$, следни дека $f'(\theta) = 1$, па за D_0 од Теоремата на Рао-Крамер имаме

$$D_0 = \frac{(f'(\theta))^2}{E\Big(\frac{\partial \ln L(\mathbf{X},\theta)}{\partial\theta}\Big)^2} = \frac{1}{\frac{2n}{\theta^2}} = \frac{\theta^2}{2n} = D(\hat{\theta}),$$

од каде заклучуваме дека непристрасниот регуларен оценувач $\hat{\theta}$ е најефикасен оценувач за θ.

б) Да ја испитаме непристрасноста на оценувачот $U = \frac{1}{3n}\sum_{i=1}^{n} X_i^2$ како оценувач за DX. Имаме,

$$E(U) = E\Big(\frac{1}{3n}\sum_{i=1}^{n} X_i^2\Big) = \frac{1}{3n}\sum_{i=1}^{n} E(X_i^2) = \frac{1}{3n}\cdot n\cdot 6\theta^2 = 2\theta^2 = DX,$$

од каде заклучуваме дека U е непристрасен оценувач за DX.

Задача 4.27. Обележјето X има густина на распределба

$$p(x,\theta) = \frac{e^{-|x|}}{2(1-e^{-\theta})},\ |x| \le \theta,$$

каде $\theta > 0$ е непознат параметар. Врз основа на примерокот $(X_1,...,X_n)$, кој одговара на обележјето X, најди максимално подобен оценувач за θ.

Решение. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. За $x = (x_1, ..., x_n)$, функцијата на подобност е

$$L(x,\theta) = \prod_{i=1}^{n} p(x_n,\theta) = \prod_{i=1}^{n} \frac{e^{-|x_i|}}{2(1-e^{-\theta})} = \frac{e^{-\sum_{i=1}^{n}|x_i|}}{2^n(1-e^{-\theta})^n},\ |x_i| \leq \theta, i = 1, ..., n.$$

Бидејќи

$$\frac{\partial L(x,\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\Big(\frac{e^{-\sum_{i=1}^{n}|x_i|}}{2^n(1-e^{-\theta})^n}\Big) = \frac{e^{-\sum_{i=1}^{n}|x_i|}}{2^n} \cdot \frac{-ne^{-\theta}}{(1-e^{-\theta})^{n+1}} < 0,\ |x_i| \leq \theta, i = 1, ..., n,$$

функцијата $L(x,\theta)$ е опаѓачка функција по θ за $|x_i| \leq \theta, i = 1, ..., n$, што е еквивалентно со $\max\{|x_1|, ..., |x_n|\} \leq \theta$, па таа го достигнува својот максимум за минималната можна вредност за θ, односно за $\theta = \max\{|x_1|, ..., |x_n|\}$. Значи, максимално подобен оценувач за θ е $\hat{\theta} = \max\{|X_1|, ..., |X_n|\}$.

Задача 4.28. Еден стрелец ја погодува целта со непозната веројатност p $(0 < p < 1)$. За да ја оцени веројатноста p, стрелецот зема 20 куршуми со кои ја гаѓа целта и при тоа го забележува бројот на погодоци. Тој изведува 5 серии од такви гаѓања (секоја серија од по 20 куршуми) и притоа ги забележал резултатите дадени во табелата. Врз основа на дадените резултати, најди максимално подобна оценка за параметарот p.

серија	1	2	3	4	5
број на погодоци	14	10	12	9	15

Решение. Нека X е број на погодоци на целта при 20 независни гаѓања. Тогаш, X има $\mathcal{B}(20,p)$ распределба, каде p е непознат параметар, односно законот на распределба на X е

$$P(x,p) = \binom{20}{x} p^x (1-p)^{20-x},\ x = 0, 1, ..., 20.$$

Тогаш, реализацијата на функцијата на подобност за дадените податоци е

$$\begin{aligned} L(x,p) &= \prod_{i=1}^{5} P(x_i,p) = P(14,p)\cdot P(10,p)\cdot P(12,p)\cdot P(9,p)\cdot P(15,p) = \\ &= \binom{20}{14} p^{14}(1-p)^{20-14} \cdot \binom{20}{10} p^{10}(1-p)^{20-10} \cdot \binom{20}{12} p^{12}(1-p)^{20-12} \cdot \\ &\quad \cdot \binom{20}{9} p^{9}(1-p)^{20-9} \cdot \binom{20}{15} p^{15}(1-p)^{20-15} = C\cdot p^{60}(1-p)^{40}, \end{aligned}$$

каде $C = \binom{20}{14}\binom{20}{10}\binom{20}{12}\binom{20}{9}\binom{20}{15}$.

Ирена Стојковска

За да ја одредиме вредноста на параметарот p за кој функвијата $L(x, p)$ достигнува максимум, ја решаваме равенката на подобност

$$0 = \frac{\partial \ln L(x,p)}{\partial p} = \frac{\partial}{\partial p}(\ln C + 60 \ln p + 40 \ln(1-p)) = \frac{60}{p} - \frac{40}{1-p},$$

чие решение е $p = 0,6$, што прецтавува максимално подобна оценка за параметарот p најдена врз основа на дадените податоци.

Задача 4.29. Нека обележјето X има закон на распределба

$$P(-2,\theta) = \frac{\theta}{5},\ P(0,\theta) = \frac{\theta}{5},\ P(7,\theta) = 1 - \frac{2\theta}{5},$$

каде θ $(0 < \theta < \frac{5}{2})$ е непознат параметар.

а) Врз основа на податоците: 0, -2, 7, -2, кои одговараат на обележјето X, најди максимално подобна оценка за θ.

б) Врз основа на примерок $(X_1, ..., X_n)$, кој одговара на обележјето X, најди максимално подобен оценувач за θ.

Решение. а) Реализацијата на функцијата на подобност врз основа на дадените податоци е

$$L(x,\theta) = \prod_{i=1}^{4} P(x_i,\theta) = P(0,\theta)\cdot P(-2,\theta)\cdot P(7,\theta)\cdot P(-2,\theta) = \Big(\frac{\theta}{5}\Big)^3\Big(1-\frac{2\theta}{5}\Big).$$

За да ја најдеме максимално подобната оценка за θ $(0 < \theta < \frac{5}{2})$ ја решаваме равенката на подобност

$$0 = \frac{\partial L(x,\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\Big(\frac{\theta^3}{5^3} - \frac{2\theta^4}{5^4}\Big) = \frac{\theta^2}{5^3}\Big(3 - \frac{8\theta}{5}\Big),$$

чие решение е $\theta = \frac{15}{8}$, што прецтавува максимално подобна оценка за θ врз основа на дадените податоци.

б) Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. Нека m_k е број на појавувања на вредноста k, $k = -2, 0, 7$ во реализацијата $(x_1, ..., x_n)$ на примерокот $(X_1, ..., X_n)$, тогаш $m_{-2} + m_0 + m_7 = n$. Функцијата на подобност е

$$L(x,\theta) = \prod_{i=1}^{n} P(x_i,\theta) = (P(-2,\theta))^{m_{-2}}\cdot(P(0,\theta))^{m_0}\cdot(P(7,\theta))^{m_7} = \Big(\frac{\theta}{5}\Big)^{m_{-2}+m_0}\cdot\Big(1-\frac{2\theta}{5}\Big)^{m_7}.$$

Ја решаваме равенката на подобност

$$\begin{aligned} 0 = \frac{\partial L(x,\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta}\Big(\Big(\frac{\theta}{5}\Big)^{m_{-2}+m_0}\cdot\Big(1-\frac{2\theta}{5}\Big)^{m_7}\Big) = \\ &= \frac{1}{5}\Big(\frac{\theta}{5}\Big)^{m_{-2}+m_0-1}\Big(1-\frac{2\theta}{5}\Big)^{m_7-1}\Big((m_{-2}+m_0)\cdot\Big(1-\frac{2\theta}{5}\Big) - \frac{2\theta}{5}\cdot m_7\Big), \end{aligned}$$

од каде следува дека

$$(m_{-2}+m_0)\cdot\left(1-\frac{2\theta}{5}\right)-\frac{2\theta}{5}\cdot m_7=0.$$

Ако искористиме дека $m_{-2}+m_0+m_7=n$, добиваме

$$(n-m_7)\cdot\left(1-\frac{2\theta}{5}\right)-\frac{2\theta}{5}\cdot m_7=0 \implies n\cdot\left(1-\frac{2\theta}{5}\right)-m_7=0,$$

од каде

$$\theta=\frac{5(n-m_7)}{2n}.$$

Значи, $\hat{\theta}=\frac{5(n-M_7)}{2n}$, каде $M_7=\sum_{i=1}^n I\{X_i=7\}$ е максимално подобен оценувач за θ.

Задача 4.30. Обележјето X има густина на распределба

$$p(x,\theta_1,\theta_2)=\frac{1}{\theta_2}\,e^{-\frac{x-\theta_1}{\theta_2}},\ x>\theta_1,$$

каде θ_1 и $\theta_2>0$ се непознати параметри. Со метод на максимална подобност, врз основа на примерок $(X_1,...,X_n)$ кој одговара на обележјето X, најди оценувачи за параметрите θ_1 и θ_2.

Решение. Нека $(X_1,...,X_n)$ е примерок кој одговара на обележјето X. Функцијата на подобност е

$$L(x,\theta_1,\theta_2)=\prod_{i=1}^n p(x,\theta_1,\theta_2)=\prod_{i=1}^n\frac{1}{\theta_2}\,e^{-\frac{x_i-\theta_1}{\theta_2}}=\frac{1}{\theta_2^n}\,e^{-\frac{1}{\theta_2}\sum_{i=1}^n(x_i-\theta_1)},\ x_1>\theta_1,...,x_n>\theta_1,$$

од каде

$$\ln L(x,\theta_1,\theta_2)=-n\ln\theta_2-\frac{1}{\theta_2}\sum_{i=1}^n(x_i-\theta_1)=-n\ln\theta_2+\frac{n\theta_1}{\theta_2}-\frac{1}{\theta_2}\sum_{i=1}^n x_i.$$

Бидејќи

$$\frac{\partial\ln L(x,\theta_1,\theta_2)}{\partial\theta_1}=\frac{n}{\theta_2}>0,$$

за $x_1>\theta_1,...,x_n>\theta_1$, следи дека функцијата $\ln L(x,\theta_1,\theta_2)$ е растечка по θ_1, па го достигнува својот максимум за најголемата можна вредност на θ_1, односно за $\theta_1=\min\{x_1,...,x_n\}$. За истата таа вредност и $L(x,\theta_1,\theta_2)$ достигнува максимум по θ_1, од каде добиваме дека $\hat{\theta_1}=\min\{X_1,...,X_n\}=X_{(1)}$ е максимално подобен оценувач за θ_1.

Равенкита на подобност за $\theta_2>0$ е

$$0=\frac{\partial\ln L(x,\theta_1,\theta_2)}{\partial\theta_2}=-\frac{n}{\theta_2}+\frac{1}{\theta_2^2}\sum_{i=1}^n(x_i-\theta_1),$$

а нејзино решение е $\theta_2=\frac{1}{n}\sum_{i=1}^n(x_1-\theta_1)=\overline{x}-\theta_1$, од каде максимално подобен оценувач за θ_2 е $\hat{\theta_2}=\overline{X}_n-\hat{\theta_1}=\overline{X}_n-X_{(1)}$.

Ирена Стојковска

Задачи за самостојна работа

Задача 4.31. Нека обележјето X има функција на распределба

$$F(x,\delta) = \frac{2\delta x}{\delta^2 + x^2},\ 0 < x < \delta,$$

каде δ е непознат параметар. Со метод на моменти, врз основа на примерок $(X_1, ..., X_n)$ кој одговара на обележјето X, најди оценувач за параметарот δ. Испитај ги непристрасноста и конзистентноста на најдениот оценувач.

Задача 4.32. Обележјето X има "двоен" Пуасонов закон на распределба

$$P(k,\lambda_1,\lambda_2) = \frac{1}{2}\,\frac{\lambda_1^k}{k!}\,e^{-\lambda_1} + \frac{1}{2}\,\frac{\lambda_2^k}{k!}\,e^{-\lambda_2},\ k = 0, 1, 2, ...$$

каде λ_1 и λ_2 $(\lambda_1 < \lambda_2)$ се непознати параметри. Со метод на моменти, врз основа на податоците од табелата, кои одговараат на обележјето X, најди оценки за параметрите λ_1 и λ_2.

x_k	0	1	2	3	4	5	6	7	8	9	10
f_k	13	15	19	16	15	8	6	4	2	1	1

Задача 4.33. Нека обележјето X има закон на распределба

$$P(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{(1 - e^{-\lambda})k!},\ k = 1, 2, ...,$$

каде $\lambda > 0$ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X и нека $\overline{X}_n$ е средината на примерокот. Покажи дека со метод на моменти не може да се најде оценувач $\hat{\lambda}$ за параметарот λ за кој важи $\hat{\lambda} \geq \overline{X}_n$.

Задача 4.34. Нека обележјето X има густина на распределба

$$p(x,\theta_1,\theta_2) = \frac{1}{\theta_2}\, e^{-\frac{x-\theta_1}{\theta_2}},\ x > \theta_1,$$

каде θ_1 и $\theta_2 > 0$ се непознати параметри. Со метод на моменти, врз основа на примерок $(X_1, ..., X_n)$ кој одговара на обележјето X, најди оценувачи за параметрите θ_1 и θ_2.

Задача 4.35. Обележјето X има рамномерна $\mathcal{U}(\theta, 1)$ распределба каде $0 < \theta < 1$ е непознат параметар. Врз основа на примерокот $(X_1, ..., X_n)$, кој одговара на обележјето X, најди максимално подобен оценувач за θ, а потоа испитај ги непристрасноста и конистентноста на најдениот оценувач. Дали најдениот оценувач е доволен оценувач?

Ирена Стојковска

Задача 4.36. Нека X е број на независни испитувања се до појавување на настанот A по прв пат, кој во секое од независните испитувања се појавува со иста веројатнист p $(0 < p < 1)$. Се изведуваат пет серии од независни испитувања и забележан е бројот на испитувања се до појавувањето на настанот A по прв пат: 1, 2, 3, 1, 0, соодветно. Врз основа на овие податоци најди максимално подобна оценка за параметарот p.

Задача 4.37. Нека обележјето X има закон на распределба

$$P\{X = 0,\ p\} = 1 - p,\ P\{X = 1,\ p\} = p,$$

каде p $(0 < p < 1)$ е непознат параметар. Врз основа на примерок $(X_1, ..., X_n)$, кој одговара на обележјето X, најди максимално подобен оценувач за p.

Задача 4.38. Обележјето X има логоритамска нормална распределба со густина на распределба

$$p(x, a, \sigma^2) = \frac{x^{-1}}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(\ln x - a)^2},\ x > 0,$$

каде a и $\sigma^2 > 0$ се непознати параметри. Со метод на максимална подобност, врз основа на примерок $(X_1, ..., X_n)$, кој одговара на обележјето X, најди оценувачи за параметрите a и $\sigma^2 > 0$.

Задача 4.39. Нека обележјето X има густина на распределба која за $0 \leq x \leq 1$ е пропорционална со $x^{\alpha}(1 - x)$, каде $\alpha > -1$ е непознат параметар (за сите други вредности на x, густината на распределба е 0). Врз основа на примерок $(X_1, ..., X_n)$, кој одговара на обележјето X, најди максимално подобен оценувач за параметарот α.

Задача 4.40. Времето T на расипување на една машина има померена експоненцјална распределба со густина на распределба

$$p(t, t_0, \lambda) = \lambda\, e^{-\lambda(t - t_0)}\ t > t_0,$$

каде $\lambda, t_0 > 0$ се параметри. Врз основа на примерок $(T_1, ..., T_n)$, кој одговара на обележјето T, најди максимално подобен оценувач за

а) параметарот λ, под претпоставка дека t_0 е познато,

б) параметарот t_0, под претпоставка дека λ е познато,

в) параметрите λ и t_0, ако и двата параметри се непознати.

4.4 Интевали на доверба

Задача 4.41. При 100 независни гаѓања стрелецот ја погодил целта 50 пати. Најди 99% интервал на доверба за непознатата веројатност за погодок на целта при едно гаѓање во метата.

Решение. Обележјето $X = I_A$, каде настанот A е целта е погодена при едно гаѓање на метата, има $\mathcal{B}(1,p)$ распределба, каде $0 < p < 1$ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X, тогаш интервалот на доверба за p со веројатност на доверба $1-\alpha$ е

$$I_p = \Big(\frac{\mu}{n} - u_{(1-\alpha)/2}\sqrt{\frac{\mu}{n^2}\Big(1-\frac{\mu}{n}\Big)}, \frac{\mu}{n} + u_{(1-\alpha)/2}\sqrt{\frac{\mu}{n^2}\Big(1-\frac{\mu}{n}\Big)}\Big),$$

каде $\mu = \sum_{i=1}^n X_i$. За дадениот примерок важи $n = 100$, $\mu = 50$ и $1-\alpha = 0,99$, од каде $u_{(1-\alpha)/2} = u_{0,495}$ е такво да $\Phi_0(u_{0,495}) = 0,495$, па од таблица имаме $u_{0,495} = 2,575$. Тогаш, бараниот 99% интевал на доверба за p е $I_p = (0,37125; 0,62875)$.

Задача 4.42. Обележјето X на една популација има $\mathcal{U}(0, 1+\theta)$ распределба, каде $\theta > 0$ е непознат параметар. Во примерок со големина $n = 200$ кој одговара на обележејот X има 150 елементи кои се помали од 1. Врз основа на тој примерок, најди 95% интервал на доверба за веројатнста $P\{X < 1\}$, а потоа и 95% интервал на доверба за θ.

Решение. Непознатата веројатност ја означуваме со $p = P\{X < 1\}$. Тогаш, $Y = I_{\{X<1\}}$ има $\mathcal{B}(1,p)$ распределба. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X, тогаш $(Y_1, ..., Y_n)$, каде $Y_i = I_{\{X_i<1\}}$ е примерок кој одговара на обележјето Y. Дадено е дека $n = 200$ и дека 150 елементи од примерокот $(X_1, ..., X_n)$ се помали од 1, што значи дека 150 елементи од примерокот $(Y_1, ..., Y_n)$ се 1-ци, а останатите се 0-ли т.е. $\mu = \sum_{i=1}^{200} Y_i = 150$. Тогаш, интервалот на доверба за p со веројатност на доверба $1-\alpha = 0,95$ е

$$I_p = \Big(\frac{\mu}{n} - u_{(1-\alpha)/2}\sqrt{\frac{\mu}{n^2}\Big(1-\frac{\mu}{n}\Big)}, \frac{\mu}{n} + u_{(1-\alpha)/2}\sqrt{\frac{\mu}{n^2}\Big(1-\frac{\mu}{n}\Big)}\Big),$$

каде $\mu = \sum_{i=1}^n Y_i$. За $n = 200$, $\mu = 150$ и $u_{(1-\alpha)/2} = u_{0,475}$ за кое $\Phi_0(u_{0,475}) = 0,475$, па затоа $u_{0,475} = 1,96$, од каде имаме дека 95% интервал на доверба за p е $I_p = (0,689988; 0,810012)$.

Сега, од

$$p = P\{X < 1\} = \int_{-\infty}^1 p_X(x,\theta)\,dx = \int_0^1 \frac{1}{1+\theta}\,dx = \frac{1}{1+\theta},$$

и од тоа што $I_p = (0,689988; 0,810012)$ е 95% интервал на доверба за p, имаме

$$\begin{aligned} 0,95 &= P\{0,689988 < p < 0,810012\} = P\{0,689988 < \frac{1}{1+\theta} < 0,810012\} = \\ &= P\{1,23455 < 1+\theta < 1,449301\} = P\{0,23455 < \theta < 0,449301\}, \end{aligned}$$

односно $I_\theta = (0,23455; 0,449301)$ е бараниот 95% интервал на доверба за θ.

Задача 4.43. Измерени се висините на 10 студенти на математика и добиени се следните вредности во cm: 165, 182, 173, 154, 157, 162, 168, 176, 168, 170. Под претпоставка дека X-висината на еден студент на математика има нормална распределба, најди 90% интервал на доверба за очекуваната висина на студентите на математика.

Решение. Обележјето X-висината на еден студент на математика има $\mathcal{N}(m,\sigma^2)$ распределба, при што m и σ^2 се непознати параметри. Нека $(X_1,...,X_n)$ е примерок кој одговара на обележјето X. Се бара интервал на доверба за m, кога σ^2 не е позната. Во тој случај имаме дека, интервалот на доверба за m со веројатност на доверба $1-\alpha$ е

$$I_m = \Big(\overline{X}_n - t_{n-1,\alpha}\,\frac{\overline{S}_n}{\sqrt{n-1}}, \overline{X}_n + t_{n-1,\alpha}\,\frac{\overline{S}_n}{\sqrt{n-1}}\Big),$$

каде $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2$ и $t_{n-1,\alpha}$ е таков да $P\{|T_{n-1}| > t_{n-1,\alpha}\} = \alpha$, каде T_{n-1} е случјана променлива со студентова распределба со $(n-1)$ степени на слобода. За дадените податоци имаме $n = 10$, потоа

$$\overline{X}_n = \frac{1}{10}(165+182+173+154+157+162+168+176+168+170) = 167,5,$$

$$\overline{S}_n^2 = \frac{1}{10}(165^2+182^2+173^2+154^2+157^2+162^2+168^2+176^2+168^2+170^2) - 167,5^2 = 64,85,$$

од каде $\overline{S}_n = 8,05295$. Веројатноста на доверба на бараниот интервал е $1-\alpha = 0,9$, од каде $\alpha = 0,1$, па од таблицата за студентова распределба наоѓаме $t_{n-1,\alpha} = t_{9;0,1} = 1,833$. Така, добиваме дека врз основа на дадените податоци, 90% интервал на доверба за очекуваната висина m е $I_m = (162,58; 172,42)$.

Задача 4.44. Обележјето X на една популација има $\mathcal{N}(m, 0,1^2)$ распределба. Одреди ја најмалата големина n на примерокот $(X_1,...,X_n)$ за да должината на 95% интервал на доверба за m да не е поголема од 0,05.

Решение. Обележјето X има $\mathcal{N}(m, 0,1^2)$ распределба, каде m е непознат параметар. Нека $(X_1,...,X_n)$ е примерок кој одговара на обележјето X. Интервал на доверба за m кога $\sigma^2 = 0,1^2$ е позната, со веројатност на доверба $1-\alpha$ е

$$I_m = \Big(\overline{X}_n - u_{(1-\alpha)/2}\,\frac{\sigma}{\sqrt{n}}, \overline{X}_n + u_{(1-\alpha)/2}\,\frac{\sigma}{\sqrt{n}}\Big),$$

каде $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ и $u_{(1-\alpha)/2}$ е таков да $\Phi_0(u_{(1-\alpha)/2}) = (1-\alpha)/2$. Веројатноста на доверба е $1-\alpha = 0,95$, од каде $u_{(1-\alpha)/2} = u_{0,475}$ е таков да $\Phi_0(u_{0,475}) = 0,475$, т.е. $u_{0,475} = 1,96$. Тогаш, за должината на интервалот имаме

$$dI_m = 2\cdot u_{(1-\alpha)/2}\,\frac{\sigma}{\sqrt{n}} = 2\cdot 1,96\cdot\frac{0,1}{\sqrt{n}} = \frac{0,392}{\sqrt{n}}.$$

Се бара најмалата вредност за n за која $dI_m \le 0,05$, односно $\frac{0,392}{\sqrt{n}} \le 0,05$. Решение на последната неравенка е $n \ge 61,4656$ и бидејќи n е природен број, следи дека најмалата вредност на примерокот е $n = 62$.

Задача 4.45. За време на еден студенцки натпревар во трчање на 100 метри, случајно избрани 30 натпреварувачи ги постигнале следните резултати во секунди:

$$15, 12, 14, 15, 18, 16, 12, 17, 15, 23, 25, 11, 16, 14, 10$$
$$17, 16, 13, 16, 19, 15, 22, 12, 18, 15, 20, 14, 19, 16, 13.$$

Под претпоставка дека времето во секунди за кое еден студент - натпреварувач истрчува 100 метри, има нормална $\mathcal{N}(16,\sigma^2)$ распределба, врз основа на дадените податоци, најди 90% интервал на доверба за дисперзијата σ^2.

Решение. Обележјето X има $\mathcal{N}(16,\sigma^2)$ распределба, каде σ^2 е непознат параметар. Нека $(X_1,...,X_n)$ е примерок кој одговара на обележјето X. Тогаш, интервал на доверба за дисперзијата σ^2 со веројатност на доверба $1-\alpha$ е

$$I_{\sigma^2} = \Big(\frac{\sum_{i=1}^n (X_i-m)^2}{\chi^2_{n,\alpha/2}}, \frac{\sum_{i=1}^n (X_i-m)^2}{\chi^2_{n,1-\alpha/2}}\Big),$$

каде $\chi^2_{n,\alpha/2}$ и $\chi^2_{n,1-\alpha/2}$ се такви да $P\{\chi^2_n > \chi^2_{n,\alpha/2}\} = \alpha/2$ и $P\{\chi^2_n > \chi^2_{n,1-\alpha/2}\} = 1-\alpha/2$ соодветно и χ^2_n е случајна променлива со хи-квадрат распределба со n степени на слобода. За дадените податоци имаме дека $n = 30$, а за веројатноста на доверба $1-\alpha = 0,9$, од каде $\alpha = 0,1$, па $\chi^2_{n,\alpha/2} = \chi^2_{30;0,05}$ и $\chi^2_{n,1-\alpha/2} = \chi^2_{30;0,95}$ се такви да $P\{\chi^2_{30} > \chi^2_{30;0,05}\} = 0,05$ и $P\{\chi^2_{30} > \chi^2_{30;0,95}\} = 0,95$ соодветно, од каде $\chi^2_{30;0,05} = 43,773$ и $\chi^2_{30;0,95} = 18,493$. За дадените податоци имаме $\sum_{i=1}^{30} X_i = 478$ и $\sum_{i=1}^{30} X_i^2 = 7970$, па за сумата во интервалот на доверба имаме

$$\sum_{i=1}^{30}(X_i-m)^2 = \sum_{i=1}^{30} X_i^2 - 2m\sum_{i=1}^{30} X_i + 30\cdot m^2 = 7970 - 2\cdot 16\cdot 478 + 30\cdot 16^2 = 354,$$

од каде бараниот 90% интервал на доверба за σ^2 е $I_{\sigma^2} = (8,08718; 19,1424)$.

Ирена Стојковска

Задача 4.46. Нека обележјето X на една популација има нормална $\mathcal{N}(m, \sigma^2)$ распределба. За податоците $x_1, x_2, ..., x_{15}$ кои одговараат на тоа обележје важи $\sum_{i=1}^{15} x_i = 15$ и $\sum_{i=1}^{15} x_i^2 = 27,3$. Врз основа на овие податоци најди 90% интервал на доверба за σ^2.

Решение. Обележјето X има $\mathcal{N}(m, \sigma^2)$ распределба, каде m и σ^2 се непознати параметри. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. Тогаш, интервал на доверба за дисперзијата σ^2 кога m не е познато, со веројатност на доверба $1-\alpha$ е

$$I_{\sigma^2} = \Big(\frac{n\overline{S}_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{n\overline{S}_n^2}{\chi^2_{n-1,1-\alpha/2}}\Big),$$

каде $\overline{S}_n^2$ е дисперзијата на примерокот, а $\chi^2_{n-1,\alpha/2}$ и $\chi^2_{n-1,1-\alpha/2}$ се такви да $P\{\chi^2_{n-1} > \chi^2_{n-1,\alpha/2}\} = \alpha/2$ и $P\{\chi^2_{n-1} > \chi^2_{n-1,1-\alpha/2}\} = 1-\alpha/2$ соодветно и χ^2_{n-1} е случајна променлива со хи-квадрат распределба со $(n-1)$ степени на слобода. За дадените податоци $n = 15$, а за веројатноста на дверба $1-\alpha = 0,9$, од каде $\alpha = 0,1$ па $\chi^2_{n-1,\alpha/2} = \chi^2_{14;0,05}$ и $\chi^2_{n-1,1-\alpha/2} = \chi^2_{14;0,95}$ се такви да $P\{\chi^2_{14} > \chi^2_{14;0,05}\} = 0,05$ и $P\{\chi^2_{14} > \chi^2_{14;0,95}\} = 0,95$ соодветно, од каде $\chi^2_{14;0,05} = 23,685$ и $\chi^2_{14;0,95} = 6,571$. За броителот во интервалот на доверба имаме

$$\begin{aligned} n\overline{S}_n^2 &= 15 \cdot \Big(\frac{1}{15}\sum_{i=1}^{15} X_i^2 - \Big(\frac{1}{15}\sum_{i=1}^{15} X_i\Big)^2\Big) = \sum_{i=1}^{15} X_i^2 - \frac{1}{15}\Big(\sum_{i=1}^{15} X_i\Big)^2 = \\ &= 27,3 - \frac{1}{15}\cdot 15^2 = 27,3 - 15 = 12,3, \end{aligned}$$

од каде бараниот 90% интервал на доверба за σ^2 е $I_{\sigma^2} = (0,519316; 1,87186)$.

Задача 4.47. При 7 расипувања на една машина измерени се следните броеви на непрекинато работење на машината во часови: 53, 48, 50, 54, 51, 50 и 51. Под претпоставка дека X - број на часови на непрекината работа на машината има експоненцијална распределба, најди 95% интервал на доверба за очекуваниот број на часови на непрекината работа на машината.

Решение. Обележјето X - број на часови на непрекината работа на машината има експоненцијална $\mathcal{E}(\beta)$ распределба, каде $\beta > 0$ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X, тогаш статистиката $T = \frac{2}{\beta}\sum_{i=1}^{n} X_i \sim \chi^2_{2n}$ (претходно покажано) може да јаземеме за централна статистика при наоѓање на интервал на доверба за очекуваниот број часови на непрекината работа на машината $E(X) = \beta$. За да го одредиме $(1-\alpha)100\%$ интервал на доверба за β, бараме $a, b \in \mathbb{R}$ такви што $P\{a < T < b\} = 1-\alpha$, односно $P\{T \leq a\} + P\{T \geq b\} = \alpha$. Бараме интервал со што е можно помала должина, па земаме $P\{T \le a\} = \alpha/2$ и $P\{T \ge b\} = \alpha/2$, од каде бидејќи $T \sim \chi^2_{2n}$ имаме дека $a = \chi^2_{2n,1-\alpha/2}$ и $b = \chi^2_{2n,\alpha/2}$. Сега,

$$\begin{aligned} 1-\alpha &= P\{a < T < b\} = P\{a < \frac{2}{\beta}\sum_{i=1}^{n} X_i < b\} = P\{\frac{1}{b} < \frac{\beta}{2\sum_{i=1}^{n} X_i} < \frac{1}{a}\} = \\ &= P\{\frac{2\sum_{i=1}^{n} X_i}{b} < \beta < \frac{2\sum_{i=1}^{n} X_i}{a}\} = P\{\frac{2n\overline{X}_n}{b} < \beta < \frac{2n\overline{X}_n}{a}\}, \end{aligned}$$

што значи $(1-\alpha)100\%$ интервал на доверба за β е интервалот

$$I_\beta = \Big(\frac{2n\overline{X}_n}{b}, \frac{2n\overline{X}_n}{a}\Big),$$

каде $a = \chi^2_{2n,1-\alpha/2}$ и $b = \chi^2_{2n,\alpha/2}$. Од $1-\alpha = 0,95$, следи $\alpha = 0,05$, потоа за дадените податоци имаме дека $n = 7$, од каде $a = \chi^2_{2n,1-\alpha/2} = \chi^2_{14;0,975} = 7,3578$ и $b = \chi^2_{2n,\alpha/2} = \chi^2_{14;0,025} = 26,3417$, а за броеителот во интервалот на доверба имаме

$$2n\overline{X}_n = 2\sum_{i=1}^{n} X_i = 2\cdot(53+48+50+54+51+50+51) = 714.$$

Значи, бараниот 95% интербвал на доверба за очекуваниот број часови на непрекината работа на машината е $I_\beta = (27,1053; 97,0399)$.

Задача 4.48. Врз основа на независни примероци $(X_1, ..., X_{n_1})$ и $(Y_1, ..., Y_{n_2})$ кои одговараат на обележјата $X \sim \mathcal{N}(m_1, \sigma_1^2)$ и $Y \sim \mathcal{N}(m_2, \sigma_2^2)$ соодветно, најди $100(1-\alpha)\%$ интервал на доверба за разликата $m_1 - m_2$.

Решение. Нека $X \sim \mathcal{N}(m_1, \sigma_1^2)$ и $Y \sim \mathcal{N}(m_2, \sigma_2^2)$, а $(X_1, ..., X_{n_1})$ и $(Y_1, ..., Y_{n_2})$ се примероци кои одговараат на обележјата X и Y соодветно. Тогаш,

$$T = \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - (m_1 - m_2)}{\sqrt{n_1\overline{S}^2_X + n_2\overline{S}^2_Y}}\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)} \sim t_{n_1+n_2-2},$$

(претходно покажано), каде $\overline{X}_{n_1}$ и $\overline{S}^2_X$ се средината и дисперзијата на примерокот $(X_1, ..., X_{n_1})$, а $\overline{Y}_{n_2}$ и $\overline{S}^2_Y$ се средината и дисперзијата на примерокот $(Y_1, ..., Y_{n_2})$, па статистиката T може да ја земеме за централна статистика за разликата $m_1 - m_2$. За да го одредиме $(1-\alpha)100\%$ интервал на доверба за $m_1 - m_2$, бараме $a, b \in \mathbb{R}$ такви што $P\{a < T < b\} = 1-\alpha$. Статистиката T има симетрична студентова $t_{n_1+n_2-2}$ распределба, па интервал со минимална должина се добива за $a = -b$, од каде последното равенство преминува во

$P\{|T| < b\} = 1-\alpha$, односно $P\{|T| \geq b\} = \alpha$, од каде $b = t_{n_1+n_2-2,\alpha}$. Сега,

$$
\begin{aligned}
& 1-\alpha = P\{a < T < b\} = P\{-b < T < b\} = \\
= \;& P\{-b < \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - (m_1 - m_2)}{\sqrt{n_1\overline{S}_X^2 + n_2\overline{S}_Y^2}}\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)} < b\} = \\
= \;& P\{\frac{-b\sqrt{n_1\overline{S}_X^2 + n_2\overline{S}_Y^2}}{\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)}} < \overline{X}_{n_1} - \overline{Y}_{n_2} - (m_1 - m_2) < \frac{b\sqrt{n_1\overline{S}_X^2 + n_2\overline{S}_Y^2}}{\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)}}\} \\
= \;& P\{\overline{X}_{n_1} - \overline{Y}_{n_2} - \frac{b\sqrt{n_1\overline{S}_X^2 + n_2\overline{S}_Y^2}}{\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)}} < m_1 - m_2 < \overline{X}_{n_1} - \overline{Y}_{n_2} + \frac{b\sqrt{n_1\overline{S}_X^2 + n_2\overline{S}_Y^2}}{\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)}}\},
\end{aligned}
$$

односно $(1-\alpha)100\%$ интервал на доверба за $m_1 - m_2$ е интервалот

$$
I = \Big(\overline{X}_{n_1} - \overline{Y}_{n_2} - \frac{b\sqrt{n_1\overline{S}_X^2 + n_2\overline{S}_Y^2}}{\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)}}, \overline{X}_{n_1} - \overline{Y}_{n_2} + \frac{b\sqrt{n_1\overline{S}_X^2 + n_2\overline{S}_Y^2}}{\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)}}\Big),
$$

каде $b = t_{n_1+n_2-2,\alpha}$.

Задачи за самостојна работа

Задача 4.49. При мерење на прирастот на масата (во грамови) кај примерок од 12 пилиња во текот на 24 саата, при примена на храна богата со протеини, добиени се следните резултати: 134, 146, 104, 119, 124, 161, 107, 83, 113, 129, 97, 123. Под претпоставка дека X - прираст на масата во грамови на едно пиле има нормална $\mathcal{N}(a, 15^2)$ распределба, најди 95% интервал на доверба за очекуваниот прираст на масата на едно пиле.

Задача 4.50. За податоците $x_1, x_2, ..., x_{10}$ кои одговараат на обележјето X со нормална распределба важи $\sum_{i=1}^{10} x_i = 34$ и $\sum_{i=1}^{10} x_i^2 = 138$. За истото обележје извлечени се нови податоци $y_1, y_2, ..., y_{15}$ за кои $\sum_{i=1}^{15} y_i = 51$ и $\sum_{i=1}^{15} y_i^2 = 207$. За колку се променила должината на 90% интервал на доверба за математичкото очекување на обележјето X?

Задача 4.51. Обележјето X има $\mathcal{U}(0, 2\theta)$ распределба, каде $\theta > 0$ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. Со помош на статистиката $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ и со примена на централната гранична теорема, најди 90% интервал на доверба за θ, врз основа на податоците:

x_k	0,2	0,3	0,6	0,8	1,1	1,2	1,4
f_k	6	7	8	7	8	6	8

Потоа, врз основа на истите податоци, одреди точен интервал на доверба за θ од обликот $\left(\frac{X_{(n)}}{2}, \frac{X_{(n)}}{A}\right)$, каде $X_{(n)} = \max\{X_1, ..., X_n\}$ и спореди ги должините на добиените интервали на доверба.

Задача 4.52. Врз основа на независни примероци $(X_1, ..., X_{n_1})$ и $(Y_1, ..., Y_{n_2})$ кои одговараат на обележјата $X \sim \mathcal{N}(m_1, \sigma_1^2)$ и $Y \sim \mathcal{N}(m_2, \sigma_2^2)$ соодветно, најди $100(1-\alpha)\%$ интервал на доверба за количникот σ_1^2/σ_2^2.
Упатство: За централна статистика земи ја

$$T = \frac{n_1(n_2-1)\sigma_2^2 \overline{S}_X^2}{n_2(n_1-1)\sigma_1^2 \overline{S}_Y^2} \sim F_{n_1-1, n_2-1}.$$

5

Тестирање на хипотези

Задача 5.1. Во една кутија се наоѓаат или 3 црвени и 7 бели топчиња, или 7 црвени и 3 бели топчиња. На случаен начин од кутијата се извлекуваат три топчиња. Ако не се сите црвени, се прифаќа хипотезата:

H_0 : Во кутијата има 3 црвени и 7 бели топчиња,

Ако сите три извлечени топчиња се црвени, се отфрла хипотезата H_0 и се прифаќа алтернативната хипотеза:

H_1 : Во кутијата има 7 црвени и 3 бели топчиња,

Најди ги веројатностите на грешките од прв и втор вид при овој статистички тест.

Решение. Грешката од прв вид настанува при отфрлање на хипотезата H_0, ако таа е точна и нејзината веројатност ја означуваме со α. Отфрлањето на хипотезата H_0, според дадениот статистичи тест, се случува кога сите три извлечени топчиња се црвени. Значи,

$$\alpha = P\{\text{извлечени се три црвени топчиња}|H_0\} = \frac{C_3^3}{C_{10}^3} = \frac{1}{120}.$$

Грешката од втор вид настанува при прифаќање на H_0, ако H_1 е точна, и таа веројатност ја означуваме со β. Според статистичкиот тест, H_0 се прифаќа, ако не сите три извлечени топчиња се црвени, т.е.

$$\begin{aligned}
\beta &= P\{\text{не сите три извлечени топчиња се црвени}|H_1\} = \\
&= 1 - P\{\text{сите три извлечени топчиња се црвени}|H_1\} = \\
&= 1 - \frac{C_7^3}{C_{10}^3} = 1 - \frac{35}{120} = \frac{85}{120}.
\end{aligned}$$

Од тука, моќта на дадениот статистички тест е $p = 1 - \beta = \frac{35}{120} \approx 30\%$, што прецтавува слаба моќ на статистичкиот тест.

5.1 Тестирање на параметарски хипотези. Нејман-Пирсонов тест

Задача 5.2. Монета се фрла 10 пати, од кои ”пара” паднала 6 пати. Со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека монетата е исправна.

Решение. Означуваме со $X = I_A$, каде A е настанот - падната е ”пара” при едно фрлање на монетата. Тогаш, $X \sim \mathcal{B}(1,p)$, каде $p = P(A)$ е непознат параметар $(0 < p < 1)$. Нека $(X_1,..,X_n)$ е примерок кој одговара на обележјето X, тогаш, статистиката $T_n = \sum_{i=1}^n X_i$ го означува бројот на паднати ”пари” при n фрлања на монетата. Треба да креираме статистичко тест за проверка на хипотезата $H_0 : p = \frac{1}{2}$, против алтернативната хипотеза $H_1 : p \neq \frac{1}{2}$.

Ако H_0 е точна, очекуваната вредност на T_n ќе биде $E(T_n) = \frac{n}{2}$, односно критичната област за отфрлање на нултата хипотеза, ќе го има обликот $C = \{(X_1,...,X_n) \mid 0 \leq T_n \leq c_1$ или $c_2 \leq T_n \leq n\}$, за некои $c_1, c_2 \in \{0,1,2,...,n\}$, $c_1 \leq c_2$. Нивото на значајност на тестот е α, што значи $P\{(X_1,...,X_n) \in C|H_0\} = \alpha$. Кога H_0 е точна, распределбата на $T_n \sim \mathcal{B}(n, \frac{1}{2})$ е симетрична, па c_1 и c_2 ги одредуваме така да $P\{0 \leq T_n \leq c_1|H_0\} = \frac{\alpha}{2}$ и $P\{c_2 \leq T_n \leq n|H_0\} = \frac{\alpha}{2}$ (еднакви или најмногу).

За $n = 10$ и $\alpha = 0,05$ добиваме

$$P\{0 \leq T_{10} \leq c_1|H_0\} = \frac{0,05}{2} \text{ и } P\{c_2 \leq T_{10} \leq 10|H_0\} = \frac{0,05}{2},$$

пришто $T_n \sim \mathcal{B}(10, \frac{1}{2})$, ако H_0 е точна, односно

$$\sum_{k=0}^{c_1} \binom{10}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{10-k} = 0,025 \text{ и } \sum_{k=c_2}^{10} \binom{10}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{10-k} = 0,025$$

т.е.

$$\sum_{k=0}^{c_1} \binom{10}{k} = 26,5 \text{ и } \sum_{k=c_2}^{10} \binom{10}{k} = 26,5$$

(или најмногу), од каде добиваме дека $c_1 = 1$ и $c_2 = 9$, па критичната област за $n = 10$ и $\alpha = 0,05$ е $C = \{(X_1,...,X_{10}) \mid 0 \leq T_{10} \leq 1$ или $9 \leq T_{10} \leq 10\}$. За дадениоте податоци $T_{10} = 6$, значи $(X_1,...,X_n) \notin C$, па ја прифаќаме нултата хипотеза H_0 дека монетата е исправна.

Задача 5.3. Вообичаено, бројот на печатни грешки на една страница има Пуасонова распределба со параметар $\lambda = 0,4$. Од една книга случајно се избрани 20 страници и утврдено е дека има вкупно 14 печатни грешки. Врз основа на дадените податоци, со ниво на значајност $\alpha = 0,05$, да се провери хипотезата дека бројот на печатни грешки во книгата е во рамките на стандардот.

Решение. Нека X е број на печатни грешки на една страница од книгата и нека $X \sim \mathcal{P}(\lambda)$. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. Тогаш, статистиката $T_n = \sum_{i=1}^{n} X_i$ е вкупен број на печатни грешки на случајно избрани n страници и $T_n \sim \mathcal{P}(n\lambda)$. Треба да ја тестираме хипотезата $H_0 : \lambda = 0,4$, против алтернативната $H_1 : \lambda \neq 0,4$.

Ако H_0 е точна, тогаш $E(T_n) = 0,4n$, па критичната област ќе го има обликот $C = \{(X_1, ..., X_n) \mid |T_n - 0,4n| \geq c\}$, каде $c \in \{0, 1, 2, ...\}$ е критичната вредност, која ја одредуваме така да $P\{(X_1, ..., X_n) \in C|H_0\} = \alpha$, односно $P\{|T_n - 0,4n| \geq c\} = \alpha$, каде α е нивото на значајност на тестот.

За $n = 20$ и $\alpha = 0,05$, критичната вредност $c \in \{0, 1, 2, ...\}$ ја избираме така да $P\{|T_{20} - 0,4 \cdot 20| \geq c\} = 0,05$, односно $P\{|T_{20} - 8| \geq c\} = 0,05$ (или најмногу), каде $T_{20} \sim \mathcal{P}(8)$. Бидејќи $P\{|T_{20}-8| \geq 6\} = 0,047936$ и $P\{|T_{20}-8| \geq 7\} = 0,020277$, за критичната вредност земаме $c = 6$. За дадените податоци $T_{20} = 14$, од каде $|14 - 8| = 6 \geq 6$, значи $(X_1, ..., X_n) \in C$, па нултата хипотеза $H_0 : \lambda = 0,4$ се отфрла и се прифаќа алтернативната хипотеза $H_1 : \lambda \neq 0,4$.

Задача 5.4. Обележјето X има густина на распределба

$$p(x, \theta) = \frac{1}{\theta}\, e^{-\frac{x}{\theta}},\ x > 0,$$

каде $\theta > 0$ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X.

а) Со користење на теоремата на Нејман-Пирсон, одреди го оптималниот ктитичен домен за проверка на хипотезата $H_0 : \theta = 1$, против алтернативната хипотеза $H_1 : \theta > 1$.

б) За ниво на значајнист $\alpha = 0,05$, при големина на примерокот $n = 100$ и алтернативна хипотеза од облик $H_1 : \theta = 2$, најди ја веројатноста за грешка од втор вид β, на конструираниот статистички тест.

Решение. а) Ќе конструираме тест за проверка на $H_0 : \theta = 1$, против алтернативната хипотеза $H_1 : \theta = \theta_1$, за $\theta_1 > 1$. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X. Функцијата на подобност е

$$L(x, \theta) = \prod_{i=1}^{n} p(x_i, \theta) = \prod_{i=1}^{n} \frac{1}{\theta}\, e^{-\frac{x_i}{\theta}} = \frac{1}{\theta^n}\, e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i},\ x_1 > 0, ..., x_n > 0.$$

Според теоремата на Нејман-Пирсон, постои $c \geq 0$ така што оптималниот критичен домен го има обликот $C = \{x \mid \frac{L(x,\theta_1)}{L(x,\theta_0)} \geq c\}$, каде $\theta_0 = 1$ и $\theta_1 > 1$. За количникот на функциите на подобност имаме

$$\frac{L(x, \theta_1)}{L(x, \theta_0)} = \frac{\frac{1}{\theta_1^n}\, e^{-\frac{1}{\theta_1}\sum_{i=1}^{n} x_i}}{\frac{1}{\theta_0^n}\, e^{-\frac{1}{\theta_0}\sum_{i=1}^{n} x_i}} = \frac{\frac{1}{\theta_1^n}\, e^{-\frac{1}{\theta_1}\sum_{i=1}^{n} x_i}}{e^{-\sum_{i=1}^{n} x_i}} = \frac{1}{\theta_1^n}\, e^{(1-\frac{1}{\theta_1})\sum_{i=1}^{n} x_i},$$

па неравенството $\frac{L(x,\theta_1)}{L(x,\theta_0)} \geq c$ преминува во

$$\frac{1}{\theta_1^n}\, e^{(1-\frac{1}{\theta_1})\sum_{i=1}^n x_i} \geq c \Leftrightarrow e^{(1-\frac{1}{\theta_1})\sum_{i=1}^n x_i} \geq c\,\theta_1^n \Leftrightarrow (1-\frac{1}{\theta_1})\sum_{i=1}^n x_i \geq \ln c\,\theta_1^n,$$

и бидејќи $\theta_1 > 1$, имаме дека $1-\frac{1}{\theta_1} > 0$, па последното неравенство е еквивалентно со

$$\sum_{i=1}^n x_i \geq \frac{\ln c\,\theta_1^n}{1-\frac{1}{\theta_1}} = c_1,$$

па критичниот домен има облик $C = \{x \mid \sum_{i=1}^n x_i \geq c_1\}$, каде c_1 е критична вредност таква да $P\{\sum_{i=1}^n X_i \geq c_1 \mid H_0\} = \alpha$ за дадено ниво на значајност α.

б) Ако $H_0 : \theta = 1$ е точна, тогаш X има густина на распределба $p(x) = e^{-x}$, $x > 0$, односно $X \sim \mathcal{E}(1)$, од каде $E(X) = 1$ и $D(X) = 1$. Случајните променливи $X_1, .., X_n$ се независни и еднакво распределени како обележјетео X, па за нив ја користиме централната гранична теорема, и за $\alpha = 0,05$ и $n = 100$ добиваме

$$\begin{aligned}
0,05 &= P\{\sum_{i=1}^{100} X_i \geq c_1 \mid H_0\} = 1 - P\{\sum_{i=1}^{100} X_i < c_1 \mid H_0\} = \\
&= 1 - P\{\frac{\sum_{i=1}^{100} X_i - 100\cdot 1}{\sqrt{1\cdot 100}} < \frac{c_1 - 100\cdot 1}{\sqrt{1\cdot 100}} \mid H_0\} \approx \\
&\approx 1 - (\frac{1}{2} + \Phi_0(\frac{c_1 - 100\cdot 1}{\sqrt{1\cdot 100}})) = \frac{1}{2} - \Phi_0(\frac{c_1-100}{10}),
\end{aligned}$$

од каде $\Phi_0(\frac{c_1-100}{10}) = 0,45$, па $\frac{c_1-100}{10} = 1,645$, од каде $c_1 = 116,45$, па критичниот домен е $C = \{x \mid \sum_{i=1}^{100} x_i \geq 116,45\}$.

Ако $H_1 : \theta = 2$ е точна, тогаш $X \sim \mathcal{E}(2)$, од каде $E(X) = 2$ и $D(X) = 4$. За да ја најдеме веројатноста за грешка од втор вид, поторно ја применуваме централната гранична теорема за случајните променливи $X_1, ..., X_n$ т.е.

$$\begin{aligned}
\beta &= P\{\sum_{i=1}^{100} X_i < 116,45 \mid H_1\} = P\{\frac{\sum_{i=1}^{100} X_i - 100\cdot 2}{\sqrt{4\cdot 100}} < \frac{116,45 - 100\cdot 2}{\sqrt{4\cdot 100}} \mid H_1\} \approx \\
&\approx \frac{1}{2} + \Phi_0(\frac{116,45 - 100\cdot 2}{\sqrt{4\cdot 100}}) = 0,5 + \Phi_0(-4,18) = 0,5 - \Phi_0(4,18) = \\
&= 0,5 - 0,499979 = 0,000021.
\end{aligned}$$

Задача 5.5. Нека времето на траење (во изминати километри) на една автомобилска гума е случајна променлива со $\mathcal{N}(\theta, 5000^2)$ распределба, каде θ е непознат параметар. Најди ги големината на примерокот и оптималниот критичен домен за тестирање на хипотезата $H_0 : \theta = 30000$, против алтернативната $H_1 : \theta = 35000$, ако веројатностите за грешка од прв и втор вид се $\alpha = 0,01$ и $\beta = 0,02$ соодветно.

Ирена Стојковска

Решение. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X-време на траење (во изминати километри) на една автомобилска гума кое има $\mathcal{N}(\theta, 5000^2)$ распределба со густина на распределба

$$p(x,\theta) = \frac{1}{\sqrt{2\pi 5000^2}}\ e^{-\frac{(x-\theta)^2}{2\cdot 5000^2}},\ x \in \mathbb{R},$$

тогаш функцијата на подобност е

$$L(x,\theta) = \prod_{i=1}^{n} p(x_i,\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi 5000^2}}\ e^{-\frac{(x_i-\theta)^2}{2\cdot 5000^2}} = \frac{1}{(\sqrt{2\pi 5000^2})^n}\ e^{-\frac{1}{2\cdot 5000^2}\sum_{i=1}^{n}(x_i-\theta)^2},$$

за $x \in \mathbb{R}^n$. Според теоремата на Нејман-Пирсон при тестирање на хипотезата $H_0 : \theta = 30000$, против $H_1 : \theta = 35000$, оптималниот критичен домен има облик $C = \{x \mid \frac{L(x,\theta_1)}{L(x,\theta_0)} \geq c\}$, каде $\theta_0 = 30000$, $\theta_1 = 35000$ и $c \geq 0$ е критична вредност. Така, неравенството $\frac{L(x,\theta_1)}{L(x,\theta_0)} \geq c$ преминува во

$$\begin{aligned}
& \frac{\frac{1}{(\sqrt{2\pi 5000^2})^n}\ e^{-\frac{1}{2\cdot 5000^2}\sum_{i=1}^{n}(x_i-35000)^2}}{\frac{1}{(\sqrt{2\pi 5000^2})^n}\ e^{-\frac{1}{2\cdot 5000^2}\sum_{i=1}^{n}(x_i-30000)^2}} \geq c \\
\Leftrightarrow\quad & e^{-\frac{1}{2\cdot 5000^2}(\sum_{i=1}^{n}(x_i-35000)^2-\sum_{i=1}^{n}(x_i-30000)^2)} \geq c \\
\Leftrightarrow\quad & -\frac{1}{2\cdot 5000^2}(\sum_{i=1}^{n}(x_i-35000)^2 - \sum_{i=1}^{n}(x_i-30000)^2) \geq \ln c \\
\Leftrightarrow\quad & \frac{1}{5000}\sum_{i=1}^{n} x_i - \frac{n}{2\cdot 5000^2}(35000^2 - 30000^2) \geq \ln c \\
\Leftrightarrow\quad & \sum_{i=1}^{n} x_i \geq 5000\cdot(\ln c + \frac{15n}{2}) = c_1,
\end{aligned}$$

од каде критичната област има облик $C = \{x \mid \sum_{i=1}^{n} x_i \geq c_1\}$, каде c_1 е критична вредност. Притоа $P\{\sum_{i=1}^{n} X_i \geq c_1 \mid H_0\} = \alpha$ и $P\{\sum_{i=1}^{n} X_i < c_1 \mid H_1\} = \beta$, каде α и β веројатнисти за грешка од прв, односно втор вид.

Ако $H_0 : \theta = 30000$ е точна, тогаш $E(X) = 30000$ и $D(X) = 5000^2$. Ако пак $H_1 : \theta = 35000$ е точна, имаме дека $E(X) = 35000$ и $D(X) = 5000^2$. Сега, за $\alpha = 0,01$ и $\beta = 0,02$, ја применуваме централната гранична теорема на последните две равенства и добиваме

$$\begin{aligned}
& P\{\sum_{i=1}^{n} X_i \geq c_1 \mid H_0\} = 0,01 \text{ и } P\{\sum_{i=1}^{n} X_i < c_1 \mid H_1\} = 0,02 \\
\Leftrightarrow\quad & P\{\frac{\sum_{i=1}^{n} X_i - 30000n}{\sqrt{5000^2 n}} \geq \frac{c_1 - 30000n}{\sqrt{5000^2 n}} \mid H_0\} = 0,01 \text{ и} \\
& P\{\frac{\sum_{i=1}^{n} X_i - 35000n}{\sqrt{5000^2 n}} < \frac{c_1 - 35000n}{\sqrt{5000^2 n}} \mid H_1\} = 0,02
\end{aligned}$$

Ирена Стојковска

$$\Leftrightarrow \quad 1-(\frac{1}{2}+\Phi_0(\frac{c_1-30000n}{\sqrt{5000^2n}}))=0,01 \text{ и } \frac{1}{2}+\Phi_0(\frac{c_1-35000n}{\sqrt{5000^2n}})=0,02$$
$$\Leftrightarrow \quad \Phi_0(\frac{c_1-30000n}{\sqrt{5000^2n}})=0,49 \text{ и } \Phi_0(\frac{c_1-35000n}{\sqrt{5000^2n}})=-0,48$$
$$\Leftrightarrow \quad \Phi_0(\frac{c_1-30000n}{\sqrt{5000^2n}})=0,49 \text{ и } \Phi_0(-\frac{c_1-35000n}{\sqrt{5000^2n}})=0,48$$
$$\Leftrightarrow \quad \frac{c_1-30000n}{\sqrt{5000^2n}}=2,325 \text{ и } -\frac{c_1-35000n}{\sqrt{5000^2n}}=2,055.$$

Со решавање на последниот систем равенки по c_1 и n се добиваат проближните решенија $c_1 \approx 626450$ и $n \approx 19$. Па, за дадените веројатности за грешки од прв и втор вид, оптималната големина на примерокот е $n = 19$, а оптималниот критичен домен е $C = \{x \mid \sum_{i=1}^{19} x_i \geq 626450\}$.

Задача 5.6. Обележјето X има закон на распределба

$$P(x,p)=p(1-p)^x,\ x=0,1,2,...,$$

каде $0 < p < 1$ е непознат параметар. Нека $(X_1,...,X_n)$ е примерок кој одговара на обележјето X.

а) Со користење на теоремата на Нејман-Пирсон, одреди го оптималниот критичен домен за проверка на хипотезата $H_0 : p = \frac{1}{2}$, против алтернативната хипотеза $H_1 : p = \frac{2}{3}$.

б) За ниво на значајнист $\alpha = 0,02$ и големина на примерокот $n = 100$, најди ја веројатноста за грешка од втор вид β, на конструираниот статистички тест.

Решение. а) Ќе конструираме тест за проверка на хипотезата $H_0 : p = \frac{1}{2}$, против алтернативната хипотеза $H_1 : p = \frac{2}{3}$. Нека $(X_1,...,X_n)$ е примерок кој одговара на обележјето X. Функцијата на подобност е

$$L(x,p)=\prod_{i=1}^{n}P(x_i,p)=\prod_{i=1}^{n}p(1-p)^{x_i}=p^n(1-p)^{\sum_{i=1}^n x_i},\ x_i\in\{0,1,2,...\}.$$

Според теоремата на Нејман-Пирсон, постои $c \geq 0$ така што оптималниот критичен домен го има обликот $C = \{x \mid \frac{L(x,p_1)}{L(x,p_0)} \geq c\}$, каде $p_0 = \frac{1}{2}$ и $p_1 = \frac{2}{3}$. За количникот на функциите на подобност имаме

$$\frac{L(x,p_1)}{L(x,p_0)}=\frac{(\frac{2}{3})^n(1-\frac{2}{3})^{\sum_{i=1}^n x_i}}{(\frac{1}{2})^n(1-\frac{1}{2})^{\sum_{i=1}^n x_i}}=(\frac{4}{3})^n(\frac{2}{3})^{\sum_{i=1}^n x_i},$$

па неравенството $\frac{L(x,p_1)}{L(x,p_0)} \geq c$ преминува во

$$(\frac{4}{3})^n(\frac{2}{3})^{\sum_{i=1}^n x_i}\geq c \Leftrightarrow (\frac{2}{3})^{\sum_{i=1}^n x_i}\geq c\,(\frac{3}{4})^n \Leftrightarrow \sum_{i=1}^{n}x_i \leq \log_{\frac{2}{3}}(c\,(\frac{3}{4})^n)=c_1,$$

па критичниот домен има облик $C = \{x \mid \sum_{i=1}^{n} x_i \leq c_1\}$, каде c_1 е критична вредност таква да $P\{\sum_{i=1}^{n} X_i \leq c_1 \mid H_0\} = \alpha$ за дадено ниво на значајност α.

б) Ако $H_0 : p = \frac{1}{2}$ е точна, тогаш X има закон на распределба $P(X = x) = \frac{1}{2}(1-\frac{1}{2})^x$, $x = 0, 1, 2, ...,$, односно $X \sim Geo(\frac{1}{2})$, од каде $E(X) = \frac{1-\frac{1}{2}}{\frac{1}{2}} = 1$ и $D(X) = \frac{1-\frac{1}{2}}{(\frac{1}{2})^2} = 2$. Случајните променливи $X_1, .., X_n$ се независни и еднакво распределени како обележјетео X, па за нив ја користиме централната гранична теорема, и за $\alpha = 0,02$ и $n = 100$ добиваме

$$\begin{aligned} 0,02 &= P\{\sum_{i=1}^{100} X_i \leq c_1 \mid H_0\} = P\{\frac{\sum_{i=1}^{100} X_i - 100 \cdot 1}{\sqrt{2 \cdot 100}} \leq \frac{c_1 - 100 \cdot 1}{\sqrt{2 \cdot 100}} \mid H_0\} \approx \\ &\approx \frac{1}{2} + \Phi_0(\frac{c_1 - 100 \cdot 1}{\sqrt{2 \cdot 100}}) = \frac{1}{2} + \Phi_0(\frac{c_1 - 100}{10\sqrt{2}}), \end{aligned}$$

од каде $\Phi_0(\frac{c_1-100}{10\sqrt{2}}) = -0,48$, односно $\Phi_0(-\frac{c_1-100}{10\sqrt{2}}) = 0,48$, па $-\frac{c_1-100}{10\sqrt{2}} = 2,055$, од каде $c_1 = 70,9379$, па критичниот домен е $C = \{x \mid \sum_{i=1}^{100} x_i \leq 70,9379\}$.

Ако $H_1 : p = \frac{2}{3}$ е точна, тогаш $X \sim Geo(\frac{2}{3})$, од каде $E(X) = \frac{1-\frac{2}{3}}{\frac{2}{3}} = \frac{1}{2}$ и $D(X) = \frac{1-\frac{2}{3}}{(\frac{2}{3})^2} = \frac{3}{4}$. За да ја најдеме веројатноста за грешка од втор вид, повторно ја применуваме централната гранична теорема за случајните променливи $X_1, ..., X_n$ т.е.

$$\begin{aligned} \beta &= P\{\sum_{i=1}^{100} X_i > 70,9379 \mid H_1\} = 1 - P\{\frac{\sum_{i=1}^{100} X_i - 100 \cdot \frac{1}{2}}{\sqrt{\frac{3}{4} \cdot 100}} \leq \frac{70,9379 - 100 \cdot \frac{1}{2}}{\sqrt{\frac{3}{4} \cdot 100}} \mid H_1\} \approx \\ &\approx 1 - (\frac{1}{2} + \Phi_0(\frac{70,9379 - 100 \cdot \frac{1}{2}}{\sqrt{\frac{3}{4} \cdot 100}})) = 0,5 - \Phi_0(2,42) = 0,5 - 0,49224 = 0,00776. \end{aligned}$$

Задачи за самостојна работа

Задача 5.7. Една коцка за играње е фрлена 50 пати и при тоа забележано е колку пати се паднала секоја од страните:

број на точки на страната	1	2	3	4	5	6
честота	11	10	8	5	11	5

Со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека коцката е фер.

Задача 5.8. Обележјето X има густина на распределба

$$p(x, \theta) = \frac{2}{\theta^2}(\theta - x),\ 0 < x < \theta,$$

каде $\theta > 0$ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X.

а) Со користење на теоремата на Нејман-Пирсон, одреди го оптималниот критичен домен за проверка на хипотезата $H_0 : \theta = 1$, против алтернативната хипотеза $H_1 : \theta = 2$.

б) За ниво на значајнист $\alpha = 0,05$ и при големина на примерокот $n = 100$, најди ја веројатноста за грешка од втор вид β, на конструираниот статистички тест.

Задача 5.9. Обележјето X има густина на распределба

$$p(x, \theta) = \theta \cdot \ln 2 \cdot 2^{-\theta\, x},\ x > 0,$$

каде $\theta > 0$ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X.

а) Со користење на теоремата на Нејман-Пирсон, одреди го оптималниот критичен домен за проверка на хипотезата $H_0 : \theta = 1$, против алтернативната хипотеза $H_1 : \theta = 2$.

б) За ниво на значајнист $\alpha = 0,05$ и при големина на примерокот $n = 100$, најди ја веројатноста за грешка од втор вид β, на конструираниот статистички тест.

Задача 5.10. Обележјето X има закон на распределба

$$P(x, \theta) = \binom{5}{x} \frac{\theta^x}{(a+\theta)^5},\ x = 0, 1, 2, 3, 4, 5,$$

каде $\theta > 0$ е непознат параметар. Нека $(X_1, ..., X_n)$ е примерок кој одговара на обележјето X.

а) Со користење на теоремата на Нејман-Пирсон, одреди го оптималниот критичен домен за проверка на хипотезата $H_0 : \theta = \frac{1}{2}$, против алтернативната хипотеза $H_1 : \theta = \frac{1}{3}$.

б) За ниво на значајнист $\alpha = 0,05$ и големина на примерокот $n = 100$, најди ја веројатноста за грешка од втор вид β, на конструираниот статистички тест.

5.2 Тестови за параметрите на нормална распределба

Задача 5.11. Една откупна станица го мери процентот на масленост во млекото за кој се претпоставува дека има нормална распределба $\mathcal{N}(m,1)$. Во текот на еден ден собрано е млеко од 25 производители и добиени се следните резултати:

процент на масленост	3,45 - 3,75	3,75 - 4,05	4,05 - 4,35
број на прозводители	5	16	4

Со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека средната вредност на процентот на масленост на млекото е 4.

Решение. Нека X - процент на масленост во млекото, притоа $X \sim \mathcal{N}(m,1)$, каде m е непознат параметар. Врз основа на дадените податоци, треба со ниво на значајност $\alpha = 0,05$, да се тестира хипотезата $H_0 : m = 4$, против $H_1 : m \neq 4$. При тестирање на оваа хипотеза, критичната област го има обликот

$$C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \Big|\frac{\overline{x} - m_0}{\sigma}\sqrt{n}\Big| \geq c_0\},$$

каде $m_0 = 4$, $\sigma = 1$, $n = 25$, аритметичката средина на дадените податоци е

$$\overline{x} = \frac{1}{25}(5 \cdot 3,6 + 16 \cdot 3,9 + 4 \cdot 4,2) = 3,888,$$

а критичната вредност c_0 ја наоѓаме од $c_0 = \Phi_0^{-1}(\frac{1-\alpha}{2}) = \Phi_0^{-1}(\frac{1-0,05}{2}) = \Phi_0^{-1}(0,475) = 1,96$, што се чита од таблицата за нормална распределба. Така, за дадените податоци имаме

$$\Big|\frac{\overline{x} - m_0}{\sigma}\sqrt{n}\Big| = \Big|\frac{3,888 - 4}{1}\sqrt{25}\Big| = 0,56 < 1,96 = c_0,$$

од каде за дадените податоци важи $x \notin C$, значи тие не ѝ противречат на нултата хипотеза $H_0 : m = 4$, па таа се прифаќа.

Задача 5.12. За да се испитаат временските услови во едно туристичко место, за 15 случајно избрани години забележан е бројот на сончеви денови во текот на годината и добиени се следните резултати:

$$220, 218, 180, 230, 245, 253, 227, 182, 194, 228, 231, 192, 260, 268, 251.$$

Со ниво на значајност $\alpha = 0,05$ и под претпоставка дека X-бројот на сончеви денови во текот на една година има нормална распределба, провери ја хипотезата дека средниот број на сончеви денови во текот на годината е 220, против алтернативната хипотеза дека е поголем од 220.

Решение. За обележјето X-број на сончеви денови во текот на една година имаме дека е нормално распределено т.е. $X \sim \mathcal{N}(m, \sigma^2)$, каде m и σ^2 се непознати параметри. Врз основа на дадените податоци, треба со ниво на значајност $\alpha = 0,05$, да се тестира хипотезата $H_0 : m = 220$, против $H_1 : m > 220$. При тестирање на оваа хипотеза, критичната област го има обликот

$$C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \frac{\overline{x} - m_0}{s}\sqrt{n-1} \geq c_0\},$$

каде $m_0 = 220$, $n = 15$, аритметичката средина на дадените податоци е $\overline{x} = 225,3$, дисперзијата е $s^2 = 733,7$, од каде $s = 27,1$, а критичната вредност $c_0 = t_{n-1,2\alpha} = t_{14;0,1} = 1,761$ се чита од таблицата за студентова распределба. Тогаш, за дадените податоци имаме

$$\frac{\overline{x} - m_0}{s}\sqrt{n-1} = \frac{225,3 - 220}{27,1}\sqrt{14} = 0,732 < 1,761 = c_0,$$

значи $x \notin C$, па нултата хипотеза $H_0 : m = 220$ се прифаќа.

Задача 5.13. При мерењето на масата (во kg) на една група ученици, добиени се следните резултати:

маса (во kg)	40 - 50	50 - 60	60- 70	70 - 80	80 - 90
број на ученици	6	7	10	4	3

Со ниво на значајност $\alpha = 0,10$, под претпоставка дека X - масата на еден ученик има нормална распределба, провери ја хипотезата дека средното квадратно отстапување на масата на учениците е 13 kg.

Решение. Дадено е дека обележјето X-маса на еден ученик има $\mathcal{N}(m, \sigma^2)$ распределба, каде m и σ^2 се непознати параметри. Врз основа на дадените податоци, треба со ниво на значајност $\alpha = 0,10$, да се тестира хипотезата $H_0 : \sigma = 13$, против $H_1 : \sigma \neq 13$. При тестирање на оваа хипотеза, критичната област го има обликот

$$C = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : \frac{ns^2}{\sigma_0^2} \geq b \text{ или } \frac{ns^2}{\sigma_0^2} \leq a\},$$

каде $n = 30$, $\sigma_0^2 = 13^2$, $a = \chi^2_{n-1,1-\alpha/2} = \chi^2_{29;0,95} = 17,708$ и $b = \chi^2_{n-1,\alpha/2} = \chi^2_{29;0,05} = 42,557$ се читаат од таблицата за хи-квадрат распределба, а дисперзијата на дадените податоци е $s^2 = 147,67$. Тогаш, за дадените податоци имаме

$$\frac{ns^2}{\sigma_0^2} = \frac{30 \cdot 147,67}{13^2} = 2,016 < 17,708 = a,$$

значи $x \in C$, па нултата хипотеза $H_0 : \sigma = 13$ се отфрла и се прифаќа алтернативната $H_1 : \sigma \neq 13$.

Ирена Стојковска

Задача 5.14. Од два града на случаен начин се избрани 10, односно 12 жители на кои им е измерена висината во *cm* и добиени се следните резултати:

$$\text{град А}:168, 174, 178, 182, 178, 186, 179, 180, 169, 172,$$
$$\text{град Б}:173, 182, 180, 175, 180, 178, 184, 188, 179, 181, 165, 173.$$

Ако висината на случајно избран жител има нормална распределба со средно квадратно отстапување од очекуваната вредност од 5 *cm*, со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека средните висини на жителите од градовите А и Б се еднакви, против алтернативната хипотеза дека средната висина на жителите од градот Б е поголема.

Решение. Нека X е висината на жител од градот А, а Y е висината на жител од градот Б, тогаш $X \sim \mathcal{N}(m_1, 5^2)$ и $Y \sim \mathcal{N}(m_2, 5^2)$, каде m_1 и m_2 се непознати параметри. Врз основа на дадените податоци, треба со ниво на значајност $\alpha = 0,05$, да се тестира хипотезата $H_0 : m_1 = m_2$, против $H_1 : m_1 < m_2$. При тестирање на оваа хипотеза, критичната област го има обликот

$$C = \{(x, y) \in \mathbb{R}^{n_1+n_2} : \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq -c\},$$

каде $n_1 = 10$, $n_2 = 12$, $\sigma_1^2 = \sigma_2^2 = 5^2$, $c = \Phi_0^{-1}(\frac{1-2\alpha}{2}) = \Phi_0^{-1}(\frac{1-2\cdot 0,05}{2}) = \Phi_0^{-1}(0,45) = 1,645$, а аритметичките средини на дадените податоци се $\overline{x} = 176,6$ и $\overline{y} = 178,2$. Тогаш, за дадените податоци имаме

$$\frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{176,6 - 178,2}{\sqrt{\frac{5^2}{10} + \frac{5^2}{12}}} = -0,747 > -1,645 = -c,$$

значи $(x, y) \notin C$, па нултата хипотеза $H_0 : m_1 = m_2$ се прифаќа, односно се прифаќа хипотезата дека средните висини на жителите од градовите А и Б се еднакви.

Задача 5.15. Една група од 16 студенти на еден испит ги покажала следните резултати (во број на поени од можни 100):

студент	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
писмен дел	90	90	80	90	92	88	90	63	70	54	78	86	99	84	56	85
усмен дел	84	84	82	94	90	85	89	62	65	52	72	90	98	89	58	85

Под претпоставка дека бројот на поени на еден студент добиени на писмениот, односно устниот дел од испитот имаат нормални распределби со еднакви дисперзии, врз основа на дадените податоци, со 5% ниво на значајност, провери ја хипотезата дека средните вредности на поените на писмениот и устниот дел се еднакви.

Решение. Нека X е бројот на поени добиен на писмениот испит, а Y е бројот на поени добиен на устниот испит, тогаш $X \sim \mathcal{N}(m_1, \sigma^2)$ и $Y \sim \mathcal{N}(m_2, \sigma^2)$, каде m_1, m_2 и σ^2 се непознати параметри. Врз основа на дадените податоци, треба со ниво на значајност $\alpha = 5\% = 0,05$, да се тестира хипотезата $H_0 : m_1 = m_2$, против $H_1 : m_1 \neq m_2$. При тестирање на оваа хипотеза, критичната област го има обликот

$$C = \{(x,y) \in \mathbb{R}^{n_1+n_2} : \Big|\frac{\overline{x}-\overline{y}}{\sqrt{n_1 s_x^2 + n_2 s_y^2}}\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)}\Big| \geq c\},$$

каде $n_1 = 16$, $n_2 = 16$, $c = t_{n_1+n_2-2,\alpha} = t_{16+16-2;0,05} = t_{30;0,05} = 2,042$, а аритметичките средини и дисперзиите на дадените податоци се $\overline{x} = 80,9375$, $\overline{y} = 79,9375$, $s_x^2 = 167,31$ и $s_y^2 = 178,06$. Тогаш, за дадените податоци имаме

$$\begin{aligned} &\Big|\frac{\overline{x}-\overline{y}}{\sqrt{n_1 s_x^2 + n_2 s_y^2}}\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)}\Big| = \\ = \ &\Big|\frac{80,9375-79,9375}{\sqrt{16\cdot 167,31 + 16\cdot 178,06}}\sqrt{\frac{16\cdot 16}{16+16}(16+16-2)}\Big| = 0,215 < 2,042 = c, \end{aligned}$$

значи $(x,y) \notin C$, па нултата хипотеза $H_0 : m_1 = m_2$ се прифаќа, односно се прифаќа хипотезата дека средните вредности на поените на писмениот и устниот дел од испитот се еднакви.

Задача 5.16. За податоците $x_1, ..., x_{n_1}$ и $y_1, ..., y_{n_2}$ кои одговараат на обележјата X и Y соодветно, познато е дека $n_1 = 10$, $n_2 = 17$, $s_x^2 = 3,2^2$ и $s_y^2 = 3^2$. Врз основа на овие податоци, под претпоставка дека обележјата X и Y имаат нормални распределби $\mathcal{N}(a_1, \sigma_1^2)$ и $\mathcal{N}(a_2, \sigma_2^2)$ соодветно, со ниво на значајност $\alpha = 0,10$, провери ја хипотезата $H_0 : \sigma_1^2 = \sigma_2^2$, против алтернативната $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Решение. Дадено е дека обележјата $X \sim \mathcal{N}(a_1, \sigma_1^2)$ и $Y \sim \mathcal{N}(a_2, \sigma_2^2)$, каде a_1, a_2, σ_1^2 и σ_2^2 се непознати параметри. При тестирање на хипотезата $H_0 : \sigma_1^2 = \sigma_2^2$, против алтернативната $H_1 : \sigma_1^2 \neq \sigma_2^2$, со ниво на значајност $\alpha = 0,10$, врз основа на дадените податоци за кои $s_x^2 > s_y^2$, критичната област има облик

$$C = \{(x,y) \in \mathbb{R}^{n_1+n_2} : \frac{n_1(n_2-1)s_x^2}{n_2(n_1-1)s_y^2} \geq c\},,$$

каде $c = F_{n_1-1,\, n_2-1,\, \alpha/2} = F_{10-1,\, 17-1,\, 0,10/2} = F_{9,\, 16,\, 0,05} = 2,54$ е вредност од таблицата за Фишеровата распределба. Сега, за дадените податоци имаме

$$\frac{n_1(n_2-1)s_x^2}{n_2(n_1-1)s_y^2} = \frac{10\cdot(17-1)\cdot 3,2^2}{17\cdot(10-1)\cdot 3^2} = 1,1898 < 2,54 = c,$$

значи $(x,y) \notin C$, па нултата хипотеза $H_0 : \sigma_1^2 = \sigma_2^2$ се прифаќа.

Ирена Стојковска

Задачи за самостојна работа

Задача 5.17. Измерена е должината на стеблото (во cm) на 30 семиња после една недела од почетокот на ртењето и добиени се следните резултати:

$$4.7, 4.9, 5.2, 5.2, 5.4, 5.5, 5.5, 5.5, 5.7, 5.7, 5.8, 6.0, 6.0, 6.0, 6.0,$$
$$6.1, 6.1, 6.2, 6.4, 6.5, 6.5, 6.7, 6.9, 7.2, 7.5, 7.8, 8.0, 8.5, 8.5, 9.0.$$

Под претпоставка дека X-должината на стеблото по една недела ртење има нормална $\mathcal{N}(m, 2^2)$ распределба, врз основа на дадените податоци, со 5% ниво на значајност, тестирај ја хипотезата $H_0 : m = 6$, против алтернативната $H_1 : m > 6$.

Задача 5.18. Една физичка величина е мерена 26 пати при исти услови. Притоа средно квадратното отстапување на податоците од нивната аритметичка средина е 2,20. Под претпоставка дека физичката величина има нормална распределба, со 5% ниво на значајност, тестирај ја хипотезата дека дисперзијата на физичката величина е 5.

Задача 5.19. Извршено е мерење на масата (во g) на 15 јаболка од сортата "Златен делишес" и 25 јаболка од сортата "Црвен делишес". Добиено е дека средната вредност на масата на јаболката "Златен делишес" е 110,5 g со средно квадратно отстапувањето од 5,1 g, додека средната вредност на масата на јаболката "Црвен делишес" е 115,2 g со средно квадратно отстапувањето од 3,5 g. Под претпоставка дека масатата на двете сорти јаболка има нормална распределба, врз основа на дадените податоци, со 5% ниво на значајност, тестирај ја хипотезата за еднаквост на масите на јаболката од двете сорти.

5.3 Тестирање на непараметарски хипотези

Задача 5.20. Со Пирсоновиот χ^2 критериум, со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека податоците

x_k	(1; 1,2)	(1,2; 1,4)	(1,4; 1,6)	(1,6; 1,8)	(1,8; 2,0)
m_k	33	23	20	15	9

одговараат на обележје X со функција на распределба

$$F(x) = \begin{cases} 0 & , x < 1 \\ \sqrt{x-1} & , 1 \leq x < 2 \\ 1 & , x \geq 2 \end{cases}.$$

Решение. Ја тестираме хипотезата $H_0 : X$ има функција на распределба $F(x)$, против $H_1 : X$ нема функција на распределба $F(x)$. Врз основа на дадените податоци, ја пресметуваме вредноста на Пирсоновата χ^2 тест статистика со пополнување на следната табела:

x_k	(1; 1,2)	(1,2; 1,4)	(1,4; 1,6)	(1,6; 1,8)	(1,8; 2,0)
m_k	33	23	20	15	9
p_k	0,447213	0,185242	0,142141	0,119831	0,105573
np_k	44,7213	18,5242	14,2141	11,9831	10,5573
$(m_k - np_k)^2$	137,389	20,0328	33,4766	9,10169	2,42518
$\frac{(m_k-np_k)^2}{np_k}$	3,07212	1,08144	2,35517	0,759544	0,229716

каде $n = 100$, а веројатностите p_k, $k = 1, ..., 5$ се пресметуваат откако $\mathbb{R}$ ќе се разбие на подинтервали соодветни на дадените податоци т.е. $(-\infty; 1,2]$, $(1,2; 1,4]$, $(1,4; 1,6]$, $(1,6; 1,8]$, $(1,8; +\infty)$, па веројатноста p_k е веројатност да X прима вредности од k-тиот подинтервал, под претпоставка дека H_0 е точна, односно:

$$p_1 = P\{X \in (-\infty; 1,2] | H_0\} = F(1,2) - F(-\infty) = \sqrt{1,2-1} - 0 = 0,447213,$$
$$p_2 = P\{X \in (1,2; 1,4] | H_0\} = F(1,4) - F(1,2) = \sqrt{1,4-1} - \sqrt{1,2-1} = 0,185242,$$

и.т.н. Тогаш, Пирсоновата χ^2 статистика (збирот на броевите од последната редица од табелата) е

$$\overline{\chi}^2 = \sum_{k=1}^{5} \frac{(m_k - np_k)^2}{np_k} = 7,49799.$$

Ирена Стојковска

Критичната област го има обликот

$$C = \{x \mid \overline{\chi}^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i} > \chi^2_{r-1,\alpha}\},$$

каде $r = 5$ е бројот на подинтервали, $\alpha = 0,05$, па од таблица наоѓаме $\chi^2_{r-1,\alpha} = \chi^2_{5-1;0,05} = \chi^2_{4;0,05} = 9,488$. Бидејќи, $\overline{\chi}^2 = 7,49799 < 9,488 = \chi^2_{r-1,\alpha}$ следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Задача 5.21. Во текот на една година забележан е бројот на итни повици неделно за медицинска интервенција и добиени се следните резултати:

број на итни повици	0	1	2	3	4	5
број на недели	6	10	20	10	6	0

Со Пирсоновиот χ^2 критериум, со ниво на значајност $\alpha = 5\%$, провери ја хипотезата дека бројот на итни повици неделно има Поасонова распределба.

Решение. Нека X е број на итни повици неделно за медицинска интервенција. Треба да се тестира хипотезата $H_0 : X$ има $\mathcal{P}(\lambda)$ распределба, против $H_1 : X$ нема $\mathcal{P}(\lambda)$ распределба, каде $\lambda > 0$ е непознат параметар кој треба прво да се оцени. Максимално подобен оценувач за λ е средината на примерокот, па за оценувач за λ ја земаме аритметичката средина на податоците т.е. $\lambda = \overline{x} = 2$. Тогаш, H_0 преминува во $H_0 : X \sim \mathcal{P}(2)$. Врз основа на дадените податоци, ја пресметуваме Пирсоновата χ^2 тест статистика со пополнување на следната табела:

x_k	0	1	2	3	4	5
m_k	6	10	20	10	6	0
p_k	0,135335	0,270671	0,270671	0,180447	0,090223	0,052653
np_k	7,03742	14,0749	14,0749	9,38324	4,6916	2,73796
$(m_k - np_k)^2$	1,07624	16,6047	35,1069	0,380388	1,71192	7,4964
$\frac{(m_k-np_k)^2}{np_k}$	0,152931	1,17974	2,49429	0,0405391	0,364891	2,73796

пришто $n = 52$, а за да ги пресметаме веројатностите p_k, $k = 1, ..., 6$, просторот $\mathbb{R}$ го разбиваме на подинтервали соодветни на дадените податоци т.е. $(\infty, 0]$, $(0, 1]$, $(1, 2]$, $(2, 3]$, $(3, 4]$, $(4, +\infty)$, па затоа:

$$p_1 = P\{X \in (-\infty; 0]|H_0\} = P\{X = 0|H_0\} = \frac{2^0}{0!}\, e^{-2} = 0,135335,$$

$$p_2 = P\{X \in (0, 1]|H_0\} = P\{X = 1|H_0\} = \frac{2^1}{1!}\, e^{-2} = 0,270671,$$

и.т.н. последната веројатност е

$$p_6 = P\{X \in (4, +\infty)|H_0\} = 1 - P\{X \in \{0,1,2,3,4\}|H_0\} = 1 - 0,947347 = 0,052653.$$

Тогаш, Пирсоновата χ^2 статистика е

$$\overline{\chi}^2 = \sum_{k=1}^{6} \frac{(m_k - np_k)^2}{np_k} = 6,97035.$$

Критичната област го има обликот

$$C = \{x \mid \overline{\chi}^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i} > \chi^2_{r-j-1,\alpha}\},$$

каде $r = 6$ е бројот на подинтервали, $j = 1$ е бројот на параметри кои требаше да се оценат, $\alpha = 5\% = 0,05$, па од таблица наоѓаме $\chi^2_{r-j-1,\alpha} = \chi^2_{6-1-1;0,05} = \chi^2_{4;0,05} = 9,488$. Бидејќи, $\overline{\chi}^2 = 6,97035 < 9,488 = \chi^2_{r-j-1,\alpha}$ следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Задача 5.22. Врз основа на податоците од следната табела:

x_k	(0, 5)	(5, 10)	(10, 15)	(15, 20)	(20, 25)
m_k	15	75	100	50	20

со Пирсоновиот χ^2 критериум, со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека податоците одговараат на обележје со нормална распределба.

Решение. Врз основа на дадените податоци за обележјето X, треба да се тестира хипотезата дека $H_0 : X$ има $\mathcal{N}(a, \sigma^2)$ распределба, против $H_1 : X$ нема $\mathcal{N}(a, \sigma^2)$ распределба, каде $a \in \mathbb{R}$ и $\sigma^2 > 0$ се непознати параметри кои треба прво да се оценат. Максимално подобни оценки за a и σ^2 се аритметичката средина и дисперзијата на податоцоте т.е. $a = \overline{x} = 12,21$ и $\sigma^2 = \overline{s}^2 = 25,4351$, односно $\sigma = 5,04$. Тогаш, H_0 преминува во $H_0 : \mathcal{N}(12,21; 5,04^2)$. Врз основа на дадените податоци, ја пресметуваме вредноста на Пирсоновата χ^2 тест статистика со пополнување на следната табела:

x_k	(0, 5)	(5, 10)	(10, 15)	(15, 20)	(10, 25)
m_k	15	75	100	50	20
p_k	0,07636	0,25361	0,37887	0,23059	0,06057
np_k	19,8536	65,9386	98,5062	59,9534	15,7482
$(m_k - np_k)^2$	23,5574	82,109	2,23144	99,0702	18,0778
$\frac{(m_k-np_k)^2}{np_k}$	1,18656	1,24523	0,0226528	1,65245	1,14793

каде $n = 260$, а веројатностите p_k, $k = 1, ..., 5$ се пресметуваат откако $\mathbb{R}$ ќе се разбие на подинтервали соодветни на дадените податоци т.е. $(-\infty, 5]$, $(5, 10]$, $(10, 15]$, $(15, 20]$, $(20, +\infty)$, па веројатноста p_k е веројатност да X прима вредности од k-тиот подинтервал, под претпоставка дека H_0 е точна, односно:

$$\begin{aligned}
p_1 &= P\{X \in (-\infty; 5] | H_0\} = P\{X \leq 5 | H_0\} = P\{\frac{X - 12,21}{5,04} \leq \frac{5 - 12,21}{5,04} | H_0\}, \\
&= 0,5 + \Phi_0(\frac{5 - 12,21}{5,04}) = 0,5 + \Phi_0(-1,43) = 0,5 - \Phi_0(1,43) = \\
&= 0,5 - 0,42364 = 0,07636, \\
p_2 &= P\{X \in (5; 10] | H_0\} = P\{5 < X \leq 10 | H_0\} = \\
&= P\{\frac{5 - 12,21}{5,04} < \frac{X - 12,21}{5,04} \leq \frac{10 - 12,21}{5,04} | H_0\} =, \\
&= \Phi_0(\frac{10 - 12,21}{5,04}) - \Phi_0(\frac{5 - 12,21}{5,04}) = \Phi_0(-0,44) - \Phi_0(-1,43) = \\
&= -\Phi_0(0,44) + \Phi_0(1,43) = -0,17003 + 0,42364 = 0,25361,
\end{aligned}$$

и.т.н. затоа што, кога H_0 е точна имаме дека $\frac{X-12,21}{5,04} \sim \mathcal{N}(0, 1)$, па вредностите за функцијата $\Phi_0(x)$ ги читаме од таблицата за стандардна нормалана распределба. Тогаш, Пирсоновата χ^2 статистика е

$$\overline{\chi}^2 = \sum_{k=1}^{5} \frac{(m_k - np_k)^2}{np_k} = 5,25482.$$

Критичната област го има обликот

$$C = \{x \mid \overline{\chi}^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i} > \chi^2_{r-j-1,\alpha}\},$$

каде $r = 5$ е бројот на подинтервали, $j = 2$ е бројот на параметри кои требаше да се оценат, $\alpha = 0,05$, па од таблица наоѓаме $\chi^2_{r-j-1,\alpha} = \chi^2_{5-2-1;0,05} = \chi^2_{2;0,05} = 5,991$. Бидејќи, $\overline{\chi}^2 = 5,25482 < 5,991 = \chi^2_{r-j-1,\alpha}$ следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Задача 5.23. Во една таблица на 200 случајни едноцифрени броеви, цифрите 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 се појавуваат 18, 23, 22, 17, 20, 25, 14, 19, 22, 20 пати соодветно. Со ниво на значајност $\alpha = 0,05$, провери ја хипотезата дека во таблицата на случајни едноцифрени броеви, цифрите 0-9 се рамномерно распределени.

Решение. Нека X е вредност на случајно избрана цифра од 0 до 9. Врз основа на дадените податоци, треба да ја тестираме хипотезата $H_0 : X$ има

дискретна рамномерна распределба на множеството $\{0, 1, ..., 9\}$, против H_1 : X нема дискретна рамномерна распределба на множеството $\{0, 1, ..., 9\}$.
Ја пресметуваме Пирсоновата χ^2 тест статистика со пополнување на следната табела:

x_k	0	1	2	3	4	5	6	7	8	9
m_k	18	23	22	17	20	25	14	19	22	20
p_k	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1
np_k	20	20	20	20	20	20	20	20	20	20
$(m_k - np_k)^2$	4	9	4	9	0	25	36	1	4	0
$\frac{(m_k-np_k)^2}{np_k}$	0,2	0,45	0,2	0,45	0	1,25	1,8	0,05	0,2	0

пришто $n = 200$, а за да ги пресметаме веројатностите p_k, $k = 1, ..., 10$, просторот $\mathbb{R}$ го разбиваме на подинтервали соодветни на дадените податоци т.е. $(\infty, 0]$, $(0, 1]$, $(1, 2]$, ..., $(7, 8]$, $(8, +\infty)$, па затоа

$$p_1 = P\{X \in (-\infty; 0] | H_0\} = P\{X = 0 | H_0\} = 0,1,$$
$$p_2 = P\{X \in (0, 1] | H_0\} = P\{X = 1 | H_0\} = 0,1,$$

и.т.н. односно $p_k = 0,1$, $k = 1, ..., 10$. Тогаш, Пирсоновата χ^2 статистика е

$$\overline{\chi}^2 = \sum_{k=1}^{10} \frac{(m_k - np_k)^2}{np_k} = 4,6.$$

Критичната област го има обликот

$$C = \{x \mid \overline{\chi}^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i} > \chi^2_{r-1,\alpha}\},$$

каде $r = 10$ е бројот на подинтервали, $\alpha = 0,05$, па од таблица наоѓаме $\chi^2_{r-1,\alpha} = \chi^2_{10-1;0,05} = \chi^2_{9;0,05} = 16,919$. Бидејќи, $\overline{\chi}^2 = 4,6 < 16,919 = \chi^2_{r-1,\alpha}$ следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Задача 5.24. Со критериумот на Колмогоров и ниво на значајност $\alpha = 0,01$, врз основа на податоците од табелата:

x_k	(0, 20)	(20, 40)	(40, 60)	(60, 80)	(80, 100)
m_k	2	8	13	17	10

тестирај ја хипотезата дека податоците одговараат на обележје со $\mathcal{U}(0, 100)$ распределба.

Ирена Стојковска

Решение. Нека X е обележјето, треба врз основа на дадените податоци да се тестира хипотезата $H_0 : X \sim \mathcal{U}(0, 100)$, против $H_1 : X$ нема $\mathcal{U}(0, 100)$ распределба. Ако H_0 е точна, густината на распределба X е

$$p(x) = \frac{1}{100},\ 0 < x < 100,$$

а функцијата на распределба на X е

$$F(x) = \begin{cases} 0 & , x \leq 0 \\ \frac{x}{100} & , 0 < x < 100 \\ 1 & , x \geq 100 \end{cases}.$$

За дадените податоци имаме $n = 50$, па емпириската функција на распределба $F_n(x)$ е

$$F_n(x) = \begin{cases} 0 & , x < 20 \\ 0,04 & , 20 \leq x < 40 \\ 0,2 & , 40 \leq x < 60 \\ 0,46 & , 60 \leq x < 80 \\ 0,8 & , 80 \leq x < 100 \\ 1 & , x \geq 100 \end{cases}.$$

Статистиката на Комогоров ја пресметуваме со помош на следната табела:

x_k	(0, 20)	(20, 40)	(40, 60)	(60, 80)	(80, 100)
m_k	2	8	13	17	10
a_k	20	40	60	80	100
$F_n(a_k)$	0,04	0,2	0,46	0,8	1
$F(a_k)$	0,2	0,4	0,6	0,8	1
$\lvert F_n(a_k) - F(a_k)\rvert$	0,16	0,2	0,14	0	0

каде a_k е десните граници на интервалите кај интервално зададените податоци. Тогаш, вредноста на статистиката на Колмогоров е (најголемата вредност во последниот ред од табелата)

$$d_n = \max_{1 \leq k \leq r} |F_n(a_k) - F(a_k)| = 0,2.$$

Критичната област го има обликот

$$C = \{x \mid d_n > d_{n,\alpha}\},$$

каде вредноста $d_{n,\alpha} = d_{50;0,01} = 0,226$ се чита од таблица. За дадените податоци имаме дека $d_n = 0,2 < 0,226 = d_{n,\alpha}$, од каде следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Ирена Стојковска

Задача 5.25. Извршени се мерења на висините на учениците во две различни училишта А и Б и при тоа се добиени следните резултати:

висина (во cm)	(160, 170)	(170, 180)	(180, 190)	(190, 200)
училиште А	25	65	8	2
училиште Б	16	34	6	4

Со ниво на значајност $\alpha = 0,01$, тестирај ја хипотезата дека распределбите на висините на учениците во училиштата А и Б се еднакви.

Решение. Нека X е висината на учениците во училиштето А, Y е висината на учениците во училиштето Б. Ја тестираме хипотезата $H_0 : X$ и Y имаат иста распределба, против $H_0 : X$ и Y немаат иста распределба. За дадените податоци имаме $n_1 = 100$ и $n_2 = 60$, па нивните емпириски функции на распределба се:

$$F_{n_1}(x) = \begin{cases} 0 & ,x < 170 \\ 0,25 & ,170 \le x < 180 \\ 0,90 & ,180 \le x < 190 \\ 0,98 & ,190 \le x < 200 \\ 1 & ,x \ge 200 \end{cases}, \quad F_{n_2}(x) = \begin{cases} 0 & ,x < 170 \\ 0,27 & ,170 \le x < 180 \\ 0,83 & ,180 \le x < 190 \\ 0,93 & ,190 \le x < 200 \\ 1 & ,x \ge 200 \end{cases}$$

соодветно. Статистиката на Комогоров-Смирнов ја пресметуваме со помош на следната табела:

x_k	(160, 170)	(170, 180)	(180, 190)	(190, 200)
m_k^1	25	65	8	2
m_k^2	16	34	6	4
a_k	170	180	190	200
$F_{n_1}(a_k)$	0,25	0,90	0,98	1
$F_{n_2}(a_k)$	0,27	0,83	0,93	1
$\lvert F_{n_1}(a_k) - F_{n_2}(a_k)\rvert$	0,02	0,07	0,05	0

каде a_k е десните граници на интервалите кај интервално зададените податоци. Тогаш, вредноста на статистиката на Колмогоров-Смирнов е (најголемата вредност во последниот ред од табелата)

$$d_{n_1,n_2} = \max_{1\le k\le r} |F_{n_1}(a_k) - F_{n_2}(a_k)| = 0,07.$$

Критичната област го има обликот

$$C = \{x \mid d_{n_1,n_2} > d_{\frac{n_1 n_2}{n_1+n_2},\alpha}\},$$

каде вредноста $d_{\frac{n_1 n_2}{n_1+n_2},\alpha} = d_{37,5;0,01} = 0,2605$ се чита од таблица. За дадените податоци имаме дека $d_{n_1,n_2} = 0,07 < 0,2605 = d_{\frac{n_1 n_2}{n_1+n_2},\alpha}$, од каде следи дека $x \notin C$, па хипотезата H_0 се прифаќа.

Ирена Стојковска

Задача 5.26. Врз основа на податоците дадени во табелата, каде X е оценката по математика во последната година од средното образование и Y е оценката на испитот по математика на факултет, добиени од анкета спроведена на 150 студенти:

$X \backslash Y$	5	6	7	8	9	10
2	2	1	0	1	0	0
3	5	35	2	5	0	1
4	3	1	10	15	6	6
5	1	0	0	6	18	32

со ниво на значајност $\alpha = 0,05$, тестирај ја хипотезата дека обележјата X и Y се независни.

Решение. Треба да ја тестираме хипотезата $H_0 : X$ и Y се независни, против $H_1 : X$ и Y не се независни. Критичната област има облик

$$C = \{(x, y) \mid d = nf^2 > \chi^2_{(r-1)(s-1),\alpha}\},$$

каде $n = 150$ е вкупниот број на студенти, f^2 е отстапувањето од статистичката независност кое се пресметува според $f^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{f_{ij}^2}{g_i h_j} - 1 = 1,077947$, каде r е бројот на различни вредности кои ги прима X, тоа се $a_1, ..., a_r$, потоа s е бројот на различни вредности кои ги прима Y, тоа се $b_1, ..., b_s$, додека f_{ij}, g_i, h_j се честотите на (a_i, b_j), a_i, b_j соодветно. Вредноста $\chi^2_{(r-1)(s-1),\alpha} = \chi^2_{(4-1)(6-1);0,05} = \chi^2_{15;0,05} = 24,996$ ја читаме од таблица за χ^2 распределба. Бидејќи

$$d = nf^2 = 150 \cdot 1,077947 = 161,69205 > 24,996 = \chi^2_{(r-1)(s-1),\alpha},$$

следи дека $(x, y) \in C$, па хипотезата H_0 за независност на X и Y ја отфрламе.

Задачи за самостојна работа

Задача 5.27. Седум монети се фрлаат истовремено 1526 пати при што секој пат се разгледува бројот на паднати "грбови" и при тоа добиени се следните резултати:

број на паднати "грбови"	0	1	2	3	4	5	6	7
број на фрлања	12	78	270	456	386	252	69	3

Со Пирсоновиот χ^2 критериум, со ниво на значајност $\alpha = 5\%$, провери ја хипотезата дека бројот на паднати "грбови" при едно фрлање на 7 монети има биномна $\mathcal{B}(7, \frac{1}{2})$ распределба.

Ирена Стојковска

Задача 5.28. При заокружување на 45 броеви, направени се следните апсолутни грешки:

$$0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.07, 0.08, 0.09, 0.09, 0.10, 0.11, 0.13, 0.14, 0.14,$$
$$0.15, 0.16, 0.17, 0.17, 0.18, 0.21, 0.22, 0.22, 0.26, 0.27, 0.31, 0.32, 0.32, 0.33, 0.33,$$
$$0.34, 0.34, 0.35, 0.35, 0.36, 0.37, 0.37, 0.39, 0.41, 0.41, 0.42, 0.43, 0.44, 0.45, 0.48.$$

Со Пирсоновиот χ^2 критериум и ниво на значајност $\alpha = 0,05$, тестирај ја хипотезата дека апсолутната грешка има непрекината рамномерна распределба.

Задача 5.29. На случајно избрани 50 студенти измерена им е висината во cm и добиени се следните резултати:

$$146, 151, 151, 155, 156, 157, 158, 158, 158, 159, 160, 160, 161, 162, 163, 163, 164,$$
$$164, 165, 165, 167, 169, 169, 170, 171, 172, 172, 174, 174, 174, 175, 175, 175, 175,$$
$$177, 177, 178, 179, 179, 179, 179, 180, 180, 182, 185, 186, 187, 187, 189, 202.$$

Спореди ги, со ниво на значајност $\alpha = 0,05$, Пирсоновиот и критериумот на Колмогоров за тестирање на хипотезата дека висината на еден студент има нормална $\mathcal{N}(170, 10^2)$ распределба.

Задача 5.30. Случајно се избрани испитаници од две различни држави А и Б, кои се изјасниле за тоа колку часови дневно гледаат телевизија и добиени се следните резултати:

време на гледање TV (во h)	0	1	2	3	4	5	6	7	8
држава А	3	10	25	40	15	5	0	1	1
држава Б	1	8	30	44	10	6	1	0	0

Со ниво на значајност $\alpha = 0,05$, тестирај ја хипотезата дека распределбите на бројот на часови дневно поминати пред телевизорот за жителите од државите А и Б се еднакви.

Задача 5.31. При едно испитување направено на пазарот за посетеноста на супермаркетите, во еден супермаркет, во 40 случајно избрани денови, измерени се просечниот процент на намалување на цените и бројот на посетители во минута, при што се добиени следните резултати:

$$(0, 12), (0, 13), (0, 13), (1, 14), (1, 14), (2, 14), (2, 14), (2, 16), (3, 17), (3, 17),$$
$$(4, 18), (4, 18), (4, 18), (4, 18), (5, 19), (5, 19), (6, 19), (6, 19), (6, 19), (7, 19),$$
$$(7, 20), (7, 20), (7, 20), (7, 20), (7, 20), (7, 20), (8, 21), (9, 21), (10, 21), (11, 21),$$
$$(11, 22), (13, 22), (13, 22), (14, 23), (15, 25), (16, 26), (17, 26), (18, 27), (22, 29), (26, 30).$$

Со 5% ниво на значајност тестирај ја хипотезата за независност на просечниот процент на намалување на цената и бројот на посетители во минута.

6

Регресиона анализа

Задача 6.1. За време на едно истражување за врската меѓу приходот и писменоста на жителите во афричките држави, забележани се следните податоци за пет афрички држави:

годишен приход по жител (во долари)	110	370	380	500	500
неписменост (во проценти)	85	75	73	63	61

а) Врз основа на дадените податоци најди ги оценките на параметрите на моделот на проста линеарна регресија кој ја опишува зависноста на неписменоста од годишниот приход по жител.

б) Предвиди го процентот на неписменост при годишен приход од 200 и од 1000 долари.

в) Најди 95% интервал на доверба за стапката на промена на процентот на неписменост, при единица промена на годишниот приход по жител.

г) Со 5% ниво на значајност, тестирај ја хипотезата дека националниот приход по жител не влијае на процентот на неписменост.

Решение. а) Годишниот приход по жител е независната променлива x, а процентот на неписменост е зависната променлива Y. За дадените податоци имаме

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5} \cdot 1860 = 372,$$

$$\overline{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = \frac{1}{5} \cdot 357 = 71,4,$$

$$s_x^2 = \frac{1}{5}\sum_{i=1}^{5} x_i^2 - \overline{x}^2 = \frac{1}{5} \cdot 793400 - 372^2 = 20296,$$

$$s_{xy} = \frac{1}{5}\sum_{i=1}^{5} x_i y_i - \overline{xy} = \frac{1}{5} \cdot 126840 - 372 \cdot 71,4 = -1192,8.$$

Тогаш, оценките на параметрите на моделот на проста линеарна регресија, кој ја опишува зависноста на неписменоста од годишниот приход по жител т.е. $Y_i = ax_i + b + \varepsilon_i$, каде $E(\varepsilon_i) = 0$ и $D(\varepsilon_i) = \sigma^2$ и ε_i се независни, се

$$a = \frac{s_{xy}}{s_x^2} = \frac{-1192,8}{20296} = -0,0587702,$$

$$b = \overline{y} - a\overline{x} = 71,4 - (-0,0587702) \cdot 372 = 93,2625,$$

од каде предвидувањата за процентот на неписменост y за познат годишен приход x се прават според формулата $y = ax + b = -0,0587702x + 93,2625$. Оценката на дисперзијата на случајната компонента, однсно средноквадратното отстапување на предвидените вредности според моделот од точните вредности, е

$$\begin{aligned}
\sigma^2 &= \frac{1}{5}\sum_{i=1}^{5}(y_i - (ax_i + b))^2 = \\
&= \frac{1}{5}\sum_{i=1}^{5} y_i^2 + a^2\frac{1}{5}\sum_{i=1}^{5} x_i^2 + b^2 - 2a\frac{1}{5}\sum_{i=1}^{5} x_i y_i - 2b\frac{1}{5}\sum_{i=1}^{5} y_i + 2ab\frac{1}{5}\sum_{i=1}^{5} x_i = \\
&= \frac{1}{5} \cdot 25869 + (-0,0587702)^2 \cdot \frac{1}{5} \cdot 793400 + 93,2625^2 - \\
&\quad -2 \cdot (-0,0587702) \cdot \frac{1}{5} \cdot 126840 - 2 \cdot 93,2625 \cdot \frac{1}{5} \cdot 357 + \\
&\quad +2 \cdot (-0,0587702) \cdot 93,2625 \cdot \frac{1}{5} \cdot 1860 = 5,7389.
\end{aligned}$$

б) Според моделот, предвидениот процент на неписменост при годишен приход од 200 долари е

$$a \cdot 200 + b = -0,0587702 \cdot 200 + 93,2625 = 81,5085\%,$$

додека пак предвидениот процент на неписменост при годишен приход од 1000 долари е

$$a \cdot 1000 + b = -0,0587702 \cdot 1000 + 93,2625 = 34,4923\%.$$

Да забележиме дека првото предвидување е во рамките на моделот (моделот е изграден врз основа на податоци за годишен приход од 110 до 500 долари), додека второто предвидување излегува од рамките на моделот, и може да биде подложно на други непознати фактори, па не е препорачливо да се користи.

Ирена Стојковска

в) Се бара 95% интервал на доверба за стапката на промена на процентот на неписменост, при единица промена на годишниот приход по жител т.е. 95% интервал на доверба за параметарот a од моделот.

При дополнителна претпоставка за нормално распределени случајни компоненти $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, ја користиме статистиката

$$T_1 = \frac{(\hat{a} - a)\sqrt{(n-2)s_x^2}}{\hat{\sigma}} \sim t_{n-2},$$

како централна статистика, каде $\hat{a}$ и $\hat{\sigma}^2$ се оценувачи на параметрите a и σ^2 соодветно.

За да го одредиме $(1-\alpha)100\%$ интервал на доверба за a, бараме $c, d \in \mathbb{R}$ така што $P\{c < T_1 < d\} = 1 - \alpha$. Статистиката T_1 има студентова распределба која е симетрична во однос на 0, па минимален интервал на доверба се добива за $c = -d$ и тогаш посленото равенство преминува во $P\{|T_1| < d\} = 1 - \alpha$, од каде $P\{|T_1| \geq d\} = 1 - (1-\alpha) = \alpha$, па $d = t_{n-2,\alpha}$ е број кој се чита од таблицата за студентова распределба. Тогаш $c = -d = -t_{n-2,\alpha}$.

Сега, од $P\{c < T_1 < d\} = 1 - \alpha$, имаме

$$P\{c < \frac{(\hat{a} - a)\sqrt{(n-2)s_x^2}}{\hat{\sigma}} < d\} = 1 - \alpha,$$

$$P\{\frac{c\hat{\sigma}}{\sqrt{(n-2)s_x^2}} < \hat{a} - a < \frac{d\hat{\sigma}}{\sqrt{(n-2)s_x^2}}\} = 1 - \alpha,$$

$$P\{-\hat{a} + \frac{c\hat{\sigma}}{\sqrt{(n-2)s_x^2}} < -a < -\hat{a} + \frac{d\hat{\sigma}}{\sqrt{(n-2)s_x^2}}\} = 1 - \alpha,$$

$$P\{\hat{a} - \frac{d\hat{\sigma}}{\sqrt{(n-2)s_x^2}} < a < \hat{a} - \frac{c\hat{\sigma}}{\sqrt{(n-2)s_x^2}}\} = 1 - \alpha,$$

$$P\{\hat{a} - \frac{t_{n-2,\alpha}\ \hat{\sigma}}{\sqrt{(n-2)s_x^2}} < a < \hat{a} + \frac{t_{n-2,\alpha}\ \hat{\sigma}}{\sqrt{(n-2)s_x^2}}\} = 1 - \alpha,$$

односно

$$I_a = \left(\hat{a} - \frac{t_{n-2,\alpha}\ \hat{\sigma}}{\sqrt{(n-2)s_x^2}}, \hat{a} + \frac{t_{n-2,\alpha}\ \hat{\sigma}}{\sqrt{(n-2)s_x^2}}\right)$$

е $(1-\alpha)100\%$ интервал на доверба за a. За дадените податоци $n = 5$ и $\alpha = 0,05$, па од таблицата се наоѓа $t_{n-2,\alpha} = t_{3;0,05} = 3,182$. Врз основа на дадените податоци, реализациите на оценувачите $\hat{a}$ и $\hat{\sigma}^2$ се најдени под а) и изнесуваат $a = -0,0587702$ и $\sigma^2 = 5,7389$ соодветно, од каде $\sigma = 2,3956$, а $s_x^2 = 20296$ е исто така пресметано под а). Овие вредности ги заменуваме во интервалот на добверба и добиваме дека 95% интервал на доверба за a е

$$I_a = (-0,0896624; -0,027878).$$

Ирена Стојковска

г) Се бара да се тестира хипотезата $H_0 : a = 0$ против алтернативната $H_1 : a \neq 0$, со ниво на значајност $\alpha = 5\% = 0,05$. Од в) најдовме дека $(1-\alpha)100\%$ интервал на доверба за a е $I_a = \left(\hat{a} - \frac{t_{n-2,\alpha}\,\hat{\sigma}}{\sqrt{(n-2)s_x^2}}, \hat{a} + \frac{t_{n-2,\alpha}\,\hat{\sigma}}{\sqrt{(n-2)s_x^2}}\right)$, тогаш критичната област за тестирање на $H_0 : a = a_0$, против $H_1 : a \neq a_0$, со ниво на значајност α, е

$$C = \{(x,y) \mid a - \frac{t_{n-2,\alpha}\,\sigma}{\sqrt{(n-2)s_x^2}} \geq a_0 \text{ или } a + \frac{t_{n-2,\alpha}\,\sigma}{\sqrt{(n-2)s_x^2}} \leq a_0\},$$

каде a и σ^2 се реализации на оценувачите $\hat{a}$ и $\hat{\sigma}^2$ соодветно. За $a_0 = 0$, добиваме дека критичната област за тестирање на $H_0 : a = 0$ против $H_1 : a \neq 0$, со ниво на значајност α, е

$$C = \{(x,y) \mid a - \frac{t_{n-2,\alpha}\,\sigma}{\sqrt{(n-2)s_x^2}} \geq 0 \text{ или } a + \frac{t_{n-2,\alpha}\,\sigma}{\sqrt{(n-2)s_x^2}} \leq 0\},$$

односно

$$C = \{(x,y) \mid \frac{a\sqrt{(n-2)s_x^2}}{\sigma} \geq t_{n-2,\alpha} \text{ или } \frac{a\sqrt{(n-2)s_x^2}}{\sigma} \leq -t_{n-2,\alpha}\},$$

што преминува во

$$C = \{(x,y) \mid \left|\frac{a\sqrt{(n-2)s_x^2}}{\sigma}\right| \geq t_{n-2,\alpha}\},$$

затоа што $t_{n-2,\alpha} > 0$. За дадените податоци имаме $n = 5$, $\alpha = 0,05$, од каде $t_{n-2,\alpha} = t_{3;0,05} = 3,182$. Од а) имаме дека $a = -0,0587702$ и $\sigma^2 = 5,7389$, од каде $\sigma = 2,3956$, и $s_x^2 = 20296$. Тогаш,

$$\left|\frac{a\sqrt{(n-2)s_x^2}}{\sigma}\right| = \left|\frac{-0,0587702 \cdot \sqrt{(5-2) \cdot 20296}}{2,3956}\right| = 6,05353 > 3,182 = t_{n-2,\alpha},$$

од каде $(x,y) \in C$, па хипотезата H_0 се отфрла, односно се отфрла хипотезата дека националниот проход по жител не влијае на процентот на писменост.

Забелешка. Интерпретацијата на параметрите кај моделот на проста линеарна регресија $Y = ax + b + \varepsilon$ е следна: Параметарот a е стапката на промена на зависната променлива Y, при единица промена на независната променлива x. Параметарот b е фиксниот дел од Y кој не зависи од x. Вредноста $ax + b$ е очекуваната вредност на Y за дадена вредност на x.

Задачи за самостојна работа

Задача 6.2. На 10 студенти измерени им се следните вредности за масата (во kg) и висината (во cm):

маса (во kg)	90	65	76	49	85	58	64	73	83	93
висина (во cm)	194	164	162	155	174	164	170	184	185	183

За дадените податоци важи $\sum x_i = 736$, $\sum x_i^2 = 56054$, $\sum y_i = 1735$, $\sum y_i^2 = 302443$, $\sum x_i y_i = 129015$.

а) Врз основа на дадените податоци најди ги оценките на параметрите на моделот на простата линеарна регресија кој ја опишува зависноста на масата од висината.

б) Најди ја очекуваната маса на лице кое има висина од 160 cm.

в) Најди 95% интервал на доверба за очекуваната маса на лице кое има висина од 160 cm.

г) Со 5% ниво на значајност тестирај ја хипотезата дека очекуваната маса на лице кое има висина од 160 cm е 70 kg, против алтернативната дека помала од 70 kg.

Задача 6.3. Се претпосатува дека меѓу процентот на неотплатени кредити и висината на камтната стапка постои линеарна стохастичка врска. Од евиденцијата на кредитното одделение извадени се следните податоци:

висина на кам. стапка (во %)	3,50	3,60	8,75	9,50	10,00	11,50	12,00	18,00	20,00
неотплатени кредити (во %)	4,0	3,5	6,2	6,8	6,7	7,4	7,9	9,2	10,2

За дадените податоци важи $\sum x_i = 96,85$, $\sum x_i^2 = 1292,27$, $\sum y_i = 61,9$, $\sum y_i^2 = 463,67$, $\sum x_i y_i = 769,95$.

а) Врз основа на дадените податоци најди ги оценките на параметрите на моделот на простата линеарна регресија кој ја опишува зависноста на процентот неотплатени кредити од висината на каматната стапка.

б) Предвиди го процентот на неотплатени кредити при висина на каматна стапка од 7%.

в) Најди 95% интервал на доверба за фиксниот дел од неотплатените кредити кој не зависи од висината на каматната стапка.

г) Со 5% ниво на значајност тестирај ја хипотезата дека фиксниот дел од неотплатените кредити кој не зависи од висината на каматната стапка е 2%.

Ирена Стојковска

Прилози

Литература

[1] Р. Малчески, *Статистика за бизнис,* Факултет за општествени науки, Скопје (2006)

[2] П. Младеновић, *Вероватноћа и статистика,* Математички факултет, Београд (2005)

[3] Ж. Попеска, *Статистика - предавања,* Институт за информатика, Природно-математички факултет, Скопје

[4] И. Стојковска, *Основи на статистика - вежби,* Институт за математика, Природно-математички факултет, Скопје

[5] Z. A. Ivković, *Matematička statistika,* Naučna knjiga, Beograd (1975)

[6] K. Knight, *Mathematical statistics,* Chapman & Hall/CRC, Boca Raton, Florida (2000)

[7] Ž. Pauše, *Uvod u matematičku statistiku,* Školska knjiga, Zagreb (1993)