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.

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

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

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.