1. (15 bodů) Chceme otestovat minci, zda je pravděpodobnost toho, že při hodu padne rub, rovna $p = 0.5$. Určete, kolikrát musíme hodit mincí, aby se relativní počet rubů lišil od 0.5 nejvýše o 0.1 s pravděpodobností $P \geq 0.95$. Interval spolehlivosti odhadněte pomocí

a) Čebyševovy nerovnosti *(7 bodů)*;
b) centrální limitní věty *(8 bodů)*.

Řešení:

Jestliže označíme K náhodnou veličinu „počet rubů po n hodech mincí," pak náhodná veličina $X = K/n$ je relativním počtem rubů. Náhodná veličina K má binomické rozdělení $\mathrm{Bi}(n, p)$ s parametry n a $p = 0.5$, tedy

$$\mathrm{E}K = n\,p = \frac{n}{2}\,, \quad \mathrm{D}K = np(1-p) = \frac{n}{4}\,.$$

Pro náhodnou veličinu $X = K/n$ jsou střední hodnota, rozptyl a směrodatná odchylka

$$\mathrm{E}X = \frac{1}{n}\,\mathrm{E}K = 0.5\,, \quad \mathrm{D}X = \frac{1}{n^2}\,\mathrm{D}K = \frac{1}{4n}\,, \quad \sigma_X = \frac{1}{2\sqrt{n}}\,.$$

a) Odhad počtu pokusů z Čebyševovy nerovnosti

$$P(|X - \mathrm{E}X| < \varepsilon) \geq 1 - \frac{\mathrm{D}X}{\varepsilon^2} \Rightarrow P\left(\left|\frac{K}{n} - \frac{1}{2}\right| < 0.1\right) \geq 1 - \frac{1}{4n\,10^{-2}} = 1 - \frac{25}{n}\,.$$

Požadavek z úlohy bude určitě splněn, pokud bude odhad pravděpodobností větší nebo roven požadované pravděpodobnosti P. Dostaneme tudíž nerovnici a z ní odhad počtu pokusů

$$1 - \frac{25}{n} \geq 0.95 \Rightarrow 0.05 \geq \frac{25}{n} \Rightarrow n \geq \frac{25}{0.05} = \mathbf{500}\,.$$

b) Binomické rozdělení nahradíme normálním rozdělením se stejnými parametry, střední hodnotou $\mathrm{E}X = p = \mu = 0.5$ a rozptylem $\mathrm{D}X = p(1-p)/n \Rightarrow \sigma^2 = 1/4n$. Z podmínky úlohy dostaneme

$$\begin{aligned}
&0.95 = P(|X - \mathrm{E}X| < \varepsilon) = P(\mathrm{E}X - \varepsilon < X < \mathrm{E}X + \varepsilon) \Rightarrow \\
&0.95 = P(0.4 < X < 0.6) = F_X(0.6) - F_X(0.4)\,.
\end{aligned}$$

Pro distribuční funkci náhodné veličiny máme podle centrální limitní věty vyjádření

$$F_X(u) = \varPhi\left(\frac{u-\mu}{\sigma}\right) = \varPhi(2(u-0.5)\sqrt{n}).$$

Je tedy

$$\begin{aligned}
0.95 &= \varPhi\left((0.6-0.5)2\sqrt{n}\right) - \varPhi\left((0.4-0.5)2\sqrt{n}\right) = \\
&= \varPhi\left(0.2\sqrt{n}\right) - \varPhi\left(-0.2\sqrt{n}\right) = 2\,\varPhi\left(0.2\,\sqrt{n}\right) - 1 \Rightarrow \\
\varPhi\left(0.2\,\sqrt{n}\right) &= \frac{1+0.95}{2} = 0.975 \Rightarrow 0.2\,\sqrt{n} = \varPhi^{-1}(0.975) \Rightarrow \\
\sqrt{n} &= \frac{\varPhi^{-1}(0.975)}{0.2} = 5\,\varPhi^{-1}(0.975) \Rightarrow n \geq 25\,(\varPhi^{-1}(0.975))^2.
\end{aligned}$$

Po dosazení z tabulky kvantilů $\varPhi^{-1}(0.975) \doteq 1.96$ dostaneme pro minimální počet hodů:

$$n \geq 25 \cdot 1.96^2 \doteq 96.04 \Rightarrow n \geq \mathbf{97}.$$

2. (15 bodů) Ve vstupním testu z předmětu B0B01PST bylo dosaženo následující bodové hodnocení

body	0	1	2	3	4	5	6	7
počet	6	2	7	5	9	14	7	14

Na hladině významnosti $\alpha = 0.05$ testujte hypotézu

H_0 : výsledky testu mají rovnoměrné rozdělení

proti hypotéze

H_1 : výsledky testu nemají rovnoměrné rozdělení.

Řešení:

Data jsou setříděna do $k = 8$ tříd a celkový rozsah souboru je $n = 64$. K testování použijeme test dobré shody s testovou statistikou

$$T = \sum_{i=1}^{8} \frac{(N_i - np_i)^2}{np_i} \sim \chi^2(7),$$

kde $p_i = \frac{1}{8} = 0.125,\ 1 \le i \le 8$, jsou teoretické pravděpodobnosti tříd. Podmínka použitelnosti testu $np_i = 64/8 = 8 > 5$ je splněna, test lze použít.

Pro realizaci testovací statistiky dostaneme

$$t = \frac{1}{8}\left((6-8)^2 + (2-8)^2 + (7-8)^2 + (5-8)^2 + (9-8)^2 + (14-8)^2 + (7-8)^2 + (14-8)^2\right) =$$
$$= \frac{1}{8}(4 + 36 + 1 + 9 + 1 + 36 + 1 + 36) = \frac{124}{8} \doteq 15.5\,.$$

Kritický obor testu je $W = \{u:\ u > q_{\chi^2(7)}(0.95)\}$. Z tabulek nalezneme kritickou hodnotu testu, kvantil $q_{\chi^2(7)}(0.95) \doteq 14.07$. Protože je

$$t \doteq 15.5 > q_{\chi^2(7)}(0.95) \doteq 14.07\,,$$

$t \in W$, je realizace testovací statistiky v kritickém oboru, testovanou hypotézu H_0 **zamítáme** ve prospěch hypotézy H_1.

3. (15 bodů) Markovův řetězec má stavy $1, 2, \ldots, k$, $k \ge 3$. V každém kroku lze ze současného stavu (i) přejít pouze do sousedního ($i-1$ nebo $i+1$, pokud existují), a to se stejnou pravděpodobností (1/2, pokud existují oba). Během 4 kroků se změnil stav z 1 na 3. Pro jaké k model nejlépe vyhovuje tomuto pozorování? *(12 bodů)*
Pro tuto hodnotu oklasifikujte všechny stavy. *(3 body)*

Řešení:

Tabulka ukazuje všechny posloupnosti stavů, které přicházejí v úvahu, a jejich pravděpodobnosti přechodů pro $k = 3, 4, 5$ (pro všechna $k \ge 5$ jsou pravděpodobnosti stejné). První dva kroky mají vždy pravděpodobnosti 1, resp. 1/2, což na *porovnání* věrohodností nemá vliv; analyzujeme jen zbývající dva kroky mezi stavy x_2, x_3, x_4.

	$k=3$			$k=4$			$k=5$		
posloupnost stavů	$p_{x_2x_3}$	$p_{x_3x_4}$	$p_{x_2x_3}$ $\cdot p_{x_3x_4}$	$p_{x_2x_3}$	$p_{x_3x_4}$	$p_{x_2x_3}$ $\cdot p_{x_3x_4}$	$p_{x_2x_3}$	$p_{x_3x_4}$	$p_{x_2x_3}$ $\cdot p_{x_3x_4}$
$(1,2,1,2,3)$	1	1/2	1/2	1	1/2	1/2	1	1/2	1/2
$(1,2,3,2,3)$	1	1/2	1/2	1/2	1/2	1/4	1/2	1/2	1/4
$(1,2,3,4,3)$	0	0	0	1/2	1	1/2	1/2	1/2	1/4
$\sum$			1			5/4			1
věrohodnost			**1/2**			**5/8**			**1/2**

Největší věrohodnost má pozorovaný jev pro $k = 4$. V tom případě jsou všechny 4 stavy trvalé s periodou 2 a tvoří jedinou komponentu.