1. (15 bodů) Pojišťovna sleduje škody na automobilech po nahlášených haváriích. Výše pojistného plnění, tj. platby klientům, kteří utrpěli škodu, byly u pěti náhodně vybraných klientů následující (uvedeno v tisících eur): 2, 2.5, 1, 10 a 5. Předpokládejme, že data pocházejí z rozdělení s hustotou $f_X(x) = c\,x\,e^{-\beta x}$ pro $x > 0$ (jinak $f(x) = 0$). Určete konstantu $c$ a odhadněte parametr $\beta$ metodou momentů nebo metodou maximální věrohodnosti.

**Řešení:**

Pro hustotu pravděpodobnosti musí platit

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1,$$

tedy v tomto případě

$$\int_{-\infty}^{\infty} f_X(x) = \int_{-\infty}^{0} 0\,\mathrm{d}x + \int_{0}^{\infty} c\,x\,e^{-\beta x}\,\mathrm{d}x = \frac{c}{\beta^2} = 1 \Rightarrow c = \beta^2.$$

Pro odhad metodou momentů spočteme střední hodnotu náhodné veličiny udávající výši pojistného plnění (v tisících eur), tj.

$$\mathrm{E}X = \int_{-\infty}^{\infty} x\,f_X(x)\,\mathrm{d}x = \int_{0}^{\infty} \beta^2\,x^2\,e^{-\beta x}\,\mathrm{d}x = \frac{2}{\beta},$$

a průměr

$$\bar{\boldsymbol{x}} = \frac{2 + 2.5 + 1 + 10 + 5}{5} = 4.1.$$

Odhad pak získáme z rovnosti

$$\frac{2}{\widehat{\beta}} = 4.1 \Rightarrow \widehat{\beta} \doteq 0.488.$$

Pro odhad metodou maximální věrohodnosti sestrojíme věrohodnostní funkci

$$L(\beta) = \prod_{i=1}^{5} \beta^2\,x_i\,e^{-\beta x_i} = 2\,\beta^2\,e^{-2\beta} \cdot 2.5\,\beta^2\,e^{-2.5\beta} \cdot \ldots \cdot 5\,\beta^2\,e^{-5\beta} = 250\,\beta^{10}\,e^{-20.5\beta}.$$

Logaritmicko-věrohodnostní funkce je pak

$$l(\beta) = \ln L(\beta) = \ln 250 + 10 \ln \beta - 20.5\,\beta,$$

její derivace je

$$l'(\beta) = \frac{10}{\beta} - 20.5,$$

a tedy řešením je

$$\frac{10}{\widehat{\beta}} - 20.5 = 0 \Rightarrow \widehat{\beta} = \frac{10}{20.5} \doteq 0.488.$$

2. (15 bodů) Sčítáme $n$ čísel zaokrouhlených na jedno desetinné místo.

a) Jaká je pravděpodobnost, že pro $n = 300$ bude chyba výsledku (tj. rozdíl mezi součtem zaokrouhlených čísel a součtem skutečných čísel) v absolutní hodnotě menší než 1?

b) Chceme, aby pravděpodobnost toho, že chyba výsledku bude v absolutní hodnotě menší než 0.5, byla alespoň 0.9. Kolik čísel $n$ můžeme maximálně sečíst?

**Řešení:**

Pro $i = 1, 2, ..., n$, jsou veličiny $X_i =$ *"hodnota i-té zaokrouhlovací chyby"* nezávislé a mají rovnoměrné rozdělení $\mathrm{R}(-0.05, 0.05)$. Tedy $\mathrm{E}X_i = 0$ a

$$\mathrm{D}X_i = \frac{\big(0.05 - (-0.05)\big)^2}{12} = \frac{0.1^2}{12} = \frac{0.01}{12}\,.$$

Chyba výsledku je $Z_n = \sum_{i=1}^{n} X_i$. Pro ni je $\mathrm{E}Z_n = 0$ a $\mathrm{D}Z_n = \frac{0.01\,n}{12}$.

a) Pro $n = 300$ hledáme pravděpodobnost

$$P(|Z_n| < 1) = P\Big( \underbrace{\Big|\frac{Z_n - 0}{\sqrt{\mathrm{D}Z_n}}\Big|}_{\mathrm{norm}(Z_n)} < \frac{1}{\sqrt{\frac{300\cdot 0.01}{12}}} \Big) = P\big(|\,\mathrm{norm}(Z_n)| < 2\big) =$$
$$= 2 \cdot \Phi(2) - 1 \doteq 2 \cdot 0.9772 - 1 = \mathbf{0.9544}\,.$$

Odhad pomocí Čebyševovy nerovnosti:

$$P(|Z_n| < 1) = P\big(|\,\mathrm{norm}(Z_n)| < 2\big) \geq 1 - \frac{1}{2^2} = \mathbf{0.75}\,.$$

b) Hledáme největší $n$ tak, aby $P(|Z_n| \leq 0.5) \geq 0.9$:

$$0.9 \leq P(|Z_n| < 0.5) = P\Big( \Big|\,\mathrm{norm}(Z_n)\Big| \leq \frac{0.5}{\sqrt{\frac{n\cdot 0.01}{12}}} \Big) =$$
$$= P\Big(|\,\mathrm{norm}(Z_n)| \leq \sqrt{\frac{300}{n}}\Big) \doteq 2 \cdot \Phi\left(\sqrt{\frac{300}{n}}\right) - 1\,.$$

Máme přibližně

$$2 \cdot \Phi\left(\sqrt{\frac{300}{n}}\right) - 1 \geq 0.9$$

$$\Phi\left(\sqrt{\frac{300}{n}}\right) \geq \frac{1 + 0.9}{2} = 0.95$$

$$\sqrt{\frac{300}{n}} \geq \Phi^{-1}(0.95) \doteq 1.645$$

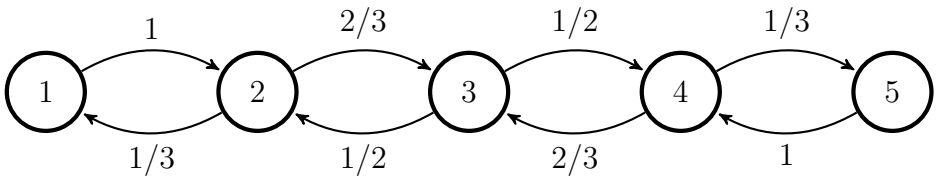
$$n \leq \frac{300}{1.645^2} \doteq \mathbf{110}$$

Odhad pomocí Čebyševovy nerovnosti:

$$P(|Z_n| < 0.5) = P\Big(|\operatorname{norm}(Z_n)| < \sqrt{\frac{300}{n}}\Big) \geq 1 - \frac{1}{\frac{300}{n}} = 1 - \frac{n}{300}\,.$$

Při požadavku $1 - \frac{n}{300} \geq 0.9$ máme $n \leq 300 \cdot 0.1 = \mathbf{30}$.

3. (15 bodů) Markovův řetězec je popsán přechodovým diagramem:




a) Klasifikujte stavy (včetně případné periody).

b) Najděte stacionární rozdělení pravděpodobností stavů a rozhodněte, zda k němu řetězec konverguje.

c) Odhadněte počáteční stav, jestliže po 4 krocích byl řetězec ve stavu 1.

d) Odhadněte rozdělení pravděpodobností stavů po 1000 krocích po startu ze stavu 1.

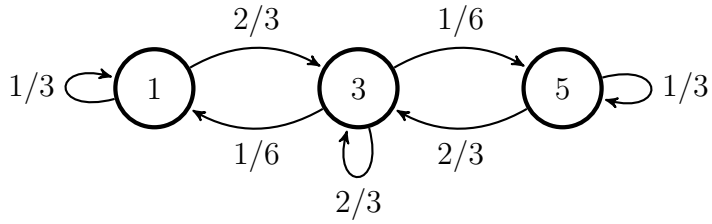
**Řešení:**

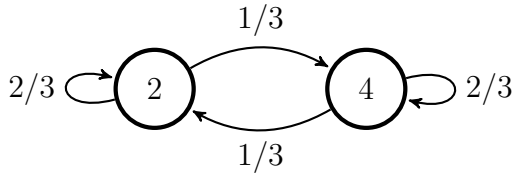
a) Všechny stavy jsou trvalé s periodou 2, tvoří jedinou komponentu.

b) Pokud je rozdělení $\boldsymbol{p}$ stacionární, pak

$$\boldsymbol{p} = \boldsymbol{p}\,\boldsymbol{P} = \boldsymbol{p}\,\boldsymbol{P}^2\,. \tag{1}$$

Matice $\boldsymbol{P}^2$ reprezentuje pravděpodobnosti přechodu řetězce s dvojnásobným krokem; ten má dvě komponenty, jednu ze sudých, druhou z lichých stavů:



Stacionární rozdělení původního řetězce musí být i stacionárním rozdělením řetězce s dvojnásobným krokem. Snadno určíme, že stav 3 musí mít pravděpodobnost 4× větší než 1 a 5, zatímco stavy 2 a 4 musí mít stejnou pravděpodobnost. (Zde lze využít též symetrie úlohy, permutací $(5, 4, 3, 2, 1)$ dostáváme tentýž řetězec.) V původním řetězci musí být stacionární rozdělení takové, že součet pravděpodobností přes sudé i přes liché stavy je stejný, tj. 1/2, tedy

$$\boldsymbol{p} = \left(\frac{1}{12}, \frac{1}{4}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12}\right).$$

Je to jediné stacionární rozdělení. (Lze použít i standardní postup založený na řešení soustavy lineárních rovnic vyjádřených v (1), kteou lze rozdělit na dvě soustavy, jednu s 3, druhou s 2 rovnicemi a neznámými.)

V důsledku netriviální periody řetězec k tomuto rozdělení nekonverguje.

c) Stačí sledovat dva kroky komponenty řetězce s dvojnásobným krokem, která obsahuje liché stavy. Věrohodnosti přechodů z lichých stavů do stavu 1 ve dvou dvojkrocích jsou

$$\begin{aligned} p_{1,1}(4) &= \frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{6} = \frac{2}{9}, \\ p_{3,1}(4) &= \frac{1}{6} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{6} = \frac{1}{6}, \\ p_{5,1}(4) &= \frac{2}{3} \cdot \frac{1}{6} = \frac{1}{9}, \end{aligned}$$

největší věrohodnost má počáteční stav 1. Tyto věrohodnosti lze najít též v 1. sloupci matice $\boldsymbol{P}^4$.

d) Po 1000 krocích po startu ze stavu 1 bude řetězec v některém z lichých stavů, a to s pravděpodobností blízkou stacionární (pro tuto komponentu), která je $(1/6, 2/3, 1/6)$. Pravděpodobnosti všech pěti stavů budou

$$\boldsymbol{p}(1000) \doteq \left(\frac{1}{6}, 0, \frac{2}{3}, 0, \frac{1}{6}\right).$$

4. (5 bodů) „Moji přátelé mají víc přátel než já.“ Takový dojem mají mnozí lidé. Je to klam, nebo má reálnou podstatu?

**Řešení:**

Něco na tom je, i když pomineme vágnost pojmu „přítel“. Představme si to jako neorientovaný graf. Uzly vysokého řádu (s mnoha sousedy) sousedí s více jinými uzly a jsou tedy započítány víckrát. Odpovídající jedinci (s mnoha přáteli) se víc projeví.

Kromě toho se může uplatnit výběrový efekt; kdo si takto stýská, je víc slyšet než ten, koho to netrápí.