14. cvičení z PSI

8. ledna 2020

14.1 (maximálně věrohodné odhady)

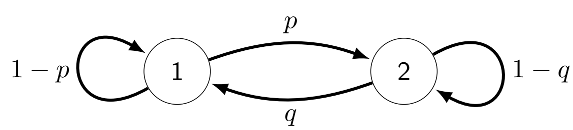
Markovův řetězec má dva stavy 1 a 2. Pravděpodobnost přechodu ze stavu 1 do stavu 2 je p, pravděpodobnost přechodu ze stavu 2 do stavu 1 je q. Z pozorované posloupnosti stavů

$$(1, 2, 2, 1, 1, 1, 2, 1, 1, 1, 2, 2, 1, 1, 1, 1)$$

odhadněte parametry p, q.

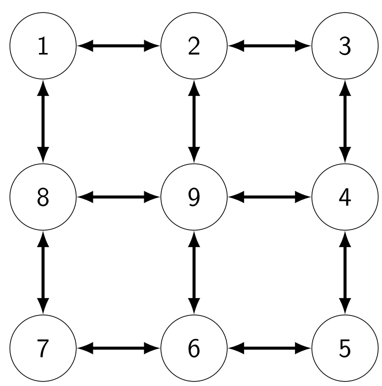
Řešení:

Cvičení 3.3: `http://cmp.felk.cvut.cz/~navara/stat/MarkovCvic_print.pdf`



14.2 (rozdělení po mnoha krocích)

Markovův řetězec je dán obrázkem:



Pro každý stav platí, že všechny hrany z něj vycházející mají stejnou pravděpodobnost.

(a) Klasifikujte všechny stavy a stanovte všechny komponenty.

(b) Konverguje rozdělení stavů ke stacionárnímu? Za jakých podmínek? Odůvodněte.

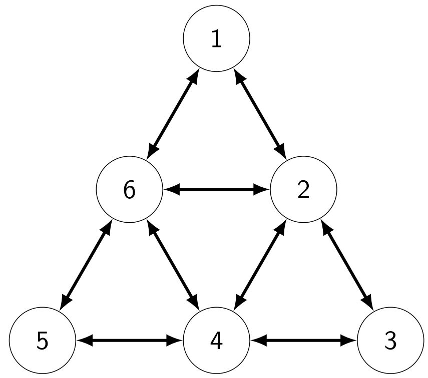
(c) Stanovte přibližně rozdělení pravděpodobností stavů po 10^5 krocích, pokud jsme vyšli se stavu 1.

Řešení:

Příklad 3: `http://cmp.felk.cvut.cz/~navara/stat/pisemky/PST180206res.pdf`

14.3 (rozdělení po mnoha krocích)

Markovův řetězec je dán obrázkem:



Pro každý stav platí, že všechny hrany z něj vycházející mají stejnou pravděpodobnost.

(a) Klasifikujte všechny stavy a stanovte všechny komponenty.

(b) Vypočtěte pravděpodobnost přechodu ze stavu 1 do stavu 4 v právě třech krocích.

(c) Stanovte rozdělení pravděpodobností stavů po 1000 krocích, pokud jsme vyšli ze stavu 1.

(d) Konverguje rozdělení stavů ke stacionárnímu?

Řešení:

Příklad 3: `http://cmp.felk.cvut.cz/~navara/stat/pisemky/PST180116res.pdf`

14.4 (Aplikace Markovových řetězců - asymptotické pravděpodobnosti stavů)

Alice trefí terč s pravděpodobností 1/3, Bob s pravděpodobností 1/2. Pokud hráč zasáhne terč, střílí dále, pokud mine, je na řadě druhý hráč. Začíná Alice. Alice vyhrává, pokud trefí terč $2\times$ za sebou, Bob vyhrává, pokud trefí terč $3\times$ za sebou. Pro oba hráče stanovte pravděpodobnosti výhry.

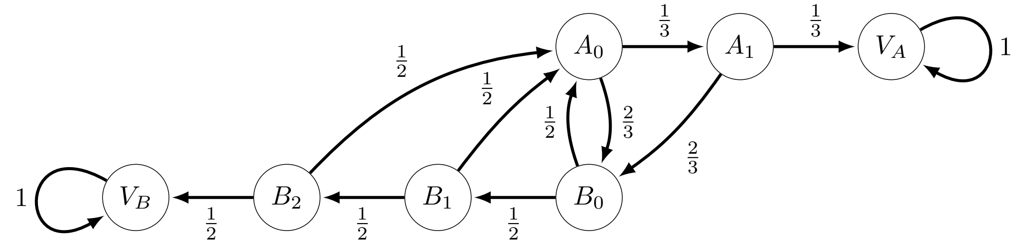
Řešení:

Cvičení 2.6: `http://cmp.felk.cvut.cz/~navara/stat/MarkovCvic_print.pdf`

Pokud bychom rozlišovali nejen to, který hráč je na řadě, ale i kolik již má úspěšných pokusů, potřebovali bychom 7 stavů:

- V_A - vyhrála Alice,
- V_B - vyhrál Bob,
- A_i - Alice má právě za sebou i úspěšných pokusů $i \in \{0, 1\}$,
- B_i - Bob má právě za sebou i úspěšných pokusů $i \in \{0, 1, 2\}$.

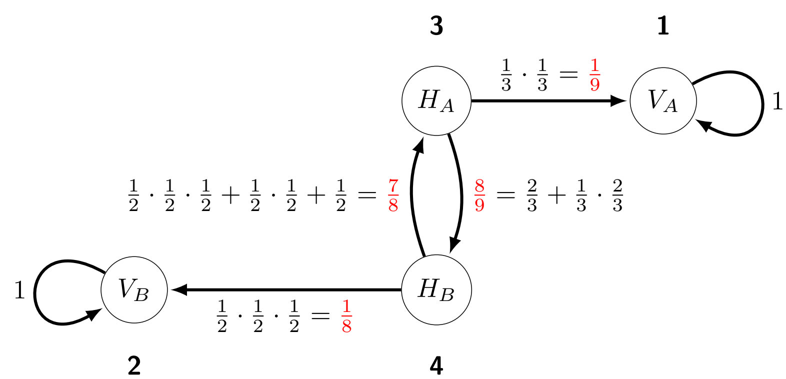
Odpovídající diagram by byl tento:



Protože nás zajímají pouze pravděpodobnosti výhry obou hráčů, můžeme si situaci popsat jednodušším způsobem a to tak, že rozlišíme pouze stavy:

- V_A - vyhrála Alice,
- V_B - vyhrál Bob,
- H_A - na řadě je Alice,
- H_B - na řadě je Bob,

kde celou sérii úspěšných pokusů daného hráče považujeme za jeden krok. Tento krok pak končí výhrou hráče s pravděpodobností $\left(\frac{1}{3}\right)^2$ pro Alici, $\left(\frac{1}{2}\right)^3$ pro Boba, nebo se na řadu dostává druhý hráč:



Stavy si opět očíslujeme tak, aby první byly ty absorpční a teprve po nich následovali ty přechodné:

$1 := V_A$ (vyhrála Alice), $2 := V_B$ (vyhrál Bob), $3 := H_A$ (na řadě je Alice), $4 := H_B$ (na řadě je Bob).

pravděpodobnosti výher Alice a Boba opět zjistíme z asymptotického rozdělení pravděpodobnosti $\boldsymbol{p}(\infty)$ s počátečním rozdělením

$$\boldsymbol{p}(0) = (0, 0, 1, 0)\ .$$

Je tedy opět potřeba spočítat $\mathbf{P}^{\infty}$ pro matici přechodu

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1/9 & 0 & 0 & 8/9 \\ 0 & 1/8 & 7/8 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{pmatrix},$$

kde

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} 1/9 & 0 \\ 0 & 1/8 \end{pmatrix}, \qquad \mathbf{Q} = \begin{pmatrix} 0 & 8/9 \\ 7/8 & 0 \end{pmatrix}.$$

Fundamentální matice tohoto řetězce je

$$\mathbf{F} = \left(\mathbf{I}_2 - \mathbf{Q}\right)^{-1} = \begin{pmatrix} 1 & -8/9 \\ -7/8 & 1 \end{pmatrix}^{-1} = (9/2) \cdot \begin{pmatrix} 1 & 8/9 \\ 7/8 & 1 \end{pmatrix} = \begin{pmatrix} 9/2 & 4 \\ 63/16 & 9/2 \end{pmatrix}$$

a tedy

$$\left(\mathbf{I}_2 - \mathbf{Q}\right)^{-1}\mathbf{R} = \begin{pmatrix} 9/2 & 4 \\ 63/16 & 9/2 \end{pmatrix} \cdot \begin{pmatrix} 1/9 & 0 \\ 0 & 1/8 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 7/16 & 9/16 \end{pmatrix}.$$

Celkem tedy dostaneme

$$\mathbf{P}^\infty = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0} \\ \left(\mathbf{I}_2 - \mathbf{Q}\right)^{-1}\mathbf{R} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 7/16 & 9/16 & 0 & 0 \end{pmatrix}$$

a asymptotické rozdělení tak je

$$\boldsymbol{p}(\infty) = \boldsymbol{p}(0) \cdot \mathbf{P}^\infty = (0, 0, 1, 0) \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 7/16 & 9/16 & 0 & 0 \end{pmatrix} = (1/2, 1/2, 0, 0)\ .$$

Zjistili jsme tak (vcelku překvapivě), že pravděpodobnosti výhry Alice i Boba jsou stejné (a sice $\frac{1}{2}$), pokud bude začínat Alice jako první.

Poznámka: Uvažujme následující obecnější případ. Pro $n, m, a, b \in \mathbb{N}$ předpokládejme, že

- Alice má pravděpodobnost zásahu $\frac{1}{n}$ a k výhře musí mít sérii a úspěšných pokusů a podobně
- Bob má pravděpodobnost zásahu $\frac{1}{m}$ a k výhře musí mít sérii b úspěšných pokusů a dále, že
- $\left(\frac{1}{n}\right)^a < \left(\frac{1}{m}\right)^b$ a proto opět necháme začít Alici.

Kdybychom opět chtěli, aby Alice a Bob měli stejné šance na výhru, zjistíme, že to nastane právě když bude platit

$$n^a - m^b = 1\ .$$

V rámci teorie čísel se řešeními této rovnice zabýval Eugène Charles Catalan a v roce 1844 vyslovil hypotézu (tzv. Catalan's conjecture), že jediné řešení této rovnice v kladných přirozených číslech je právě jen $3^2 - 2^3 = 1$. Hypotézu potvrdil Preda Mihăilescu v roce 2002.

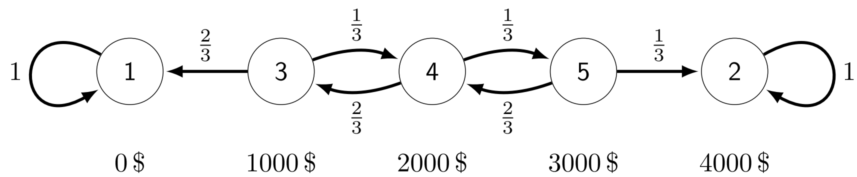
14.5 (Aplikace Markovových řetězců - asymptotické pravděpodobnosti stavů)

Alice hraje v kasinu hru, kde s pravděpodobností 1/3 vyhraje. V každém kole vsadí 1000 dolarů. V případě výhry získá 1000 dolarů, v případě prohry o 1000 dolarů přijde. Alice odejde z kasina, jestliže prohraje všechny své peníze nebo bude mít 4000 dolarů. Jaká je pravděpodobnost, že Alice odejde s prázdnou, měla-li na začátku 3000 dolarů?

Řešení:

Příklad 3: `http://cmp.felk.cvut.cz/~navara/psi/pisemky/PSI150106res.pdf`

Nakreslíme si příslušný orientovaný graf:



Pro Alici uvažujme stavy

1 - odchází s prázdnou, 2 - má 4000 dolarů (a tedy odchází), 3 - má 1000 dolarů, 4 - má 2000 dolarů a 5 - má 3000 dolarů.

Stavy jsme si očíslovali tak, aby první byly ty absorpční a teprve po nich následovali ty přechodné. Na začátku má Alice 3000 dolarů, tedy je ve stavu číslo 5 a počáteční rozdělení pravděpodobnosti tak je

$$\boldsymbol{p}(0) = (0, 0, 0, 0, 1)\ .$$

pravděpodobnost, že Alice odejde s prázdnou odpovídá složce pro stav 1 v asymptotickém rozdělení pravděpodobnosti $\boldsymbol{p}(\infty)$.

Proč tomu tak je: Platí:

- (Podmíněná) pravděpodobnost $P\big(\underbrace{i \to \cdots \to j}_{n \text{ kroků}}\big)$ toho, že se po *právě* n krocích ze stavu i přesuneme do stavu j je
$$P\big(\underbrace{i \to \cdots \to j}_{n \text{ kroků}}\big) = \Big(\mathbf{P}^n\Big)_{i,j}\ .$$
To se snadno ukáže indukcí.
- Jestliže i_* je *absorpční* stav, pak (podmíněná) pravděpodobnost $P\big(\underbrace{i \to \cdots \to i_*}_{\leq n \text{ kroků}}\big)$ toho, že se po *nejvýše* n krocích ze stavu i přesuneme do stavu i_* je
$$P\big(\underbrace{i \to \cdots \to i_*}_{\leq n \text{ kroků}}\big) = P\big(\underbrace{i \to \cdots \to i_*}_{n \text{ kroků}}\big) = \Big(\mathbf{P}^n\Big)_{i,i_*}\ .$$
To je proto, že jakoukoliv posloupnost kratší než n můžeme nastavit opakovaným přidáním stavu i_* (protože je absorpční).
- Jestliže i_* je *absorpční* stav, pak (podmíněná) pravděpodobnost $P\big(\underbrace{i \to \cdots \to i_*}_{\text{konečně kroků}}\big)$ toho, že se po *konečně* mnoha krocích ze stavu i přesuneme do stavu i_* je
$$P\big(\underbrace{i \to \cdots \to i_*}_{\text{konečně kroků}}\big) = P\Big(\bigcup_n \underbrace{i \to \cdots \to i_*}_{\leq n \text{ kroků}}\Big) = \lim_{n\to\infty} P\Big(\underbrace{i \to \cdots \to i_*}_{\leq n \text{ kroků}}\Big) = \lim_{n\to\infty}\Big(\mathbf{P}^n\Big)_{i,i_*} = \Big(\mathbf{P}^\infty\Big)_{i,i_*}\ .$$

A protože $\boldsymbol{p}(\infty) = \boldsymbol{p}(0)\cdot\mathbf{P}^\infty$, můžeme tento závěr ekvivalentně vyjádřit přes rozdělení pravděpodobnosti.

Pro výpočet asymptotického rozdělení pravděpodobnosti si opět zapíšeme matici přechodu

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 2/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 2/3 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{pmatrix},$$

kde

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} 2/3 & 0 \\ 0 & 0 \\ 0 & 1/3 \end{pmatrix}, \qquad \mathbf{Q} = \begin{pmatrix} 0 & 1/3 & 0 \\ 2/3 & 0 & 1/3 \\ 0 & 2/3 & 0 \end{pmatrix}.$$

Opět si určíme matici

$$\mathbf{P}^\infty := \lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0} \\ (\mathbf{I}_3 - \mathbf{Q})^{-1}\mathbf{R} & \mathbf{0} \end{pmatrix} .$$

Spočítáme fundamentální matici $\mathbf{F} = \big(\mathbf{I}_3 - \mathbf{Q}\big)^{-1}$:

$$\Big(\mathbf{I}_3 - \mathbf{Q} \;\Big|\; \mathbf{I}_3\Big) = \left(\begin{array}{ccc|ccc} 1 & -1/3 & 0 & 1 & 0 & 0 \\ -2/3 & 1 & -1/3 & 0 & 1 & 0 \\ 0 & -2/3 & 1 & 0 & 0 & 1 \end{array}\right) \sim \cdots \sim$$

$$\sim \cdots \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & \\ 0 & 1 & 0 & \frac{1}{5}\begin{pmatrix} 7 & 3 & 1 \\ 6 & 9 & 3 \\ 4 & 6 & 7 \end{pmatrix} \\ 0 & 0 & 1 & \end{array}\right) = \Big(\mathbf{I}_3 \;\Big|\; \big(\mathbf{I}_3 - \mathbf{Q}\big)^{-1}\Big) .$$

Tedy

$$\mathbf{F} = \big(\mathbf{I}_3 - \mathbf{Q}\big)^{-1} = \begin{pmatrix} 1 & -1/3 & 0 \\ -2/3 & 1 & -1/3 \\ 0 & -2/3 & 1 \end{pmatrix}^{-1} = \frac{1}{5}\begin{pmatrix} 7 & 3 & 1 \\ 6 & 9 & 3 \\ 4 & 6 & 7 \end{pmatrix}$$

a

$$\big(\mathbf{I}_3 - \mathbf{Q}\big)^{-1}\mathbf{R} = \frac{1}{5}\begin{pmatrix} 7 & 3 & 1 \\ 6 & 9 & 3 \\ 4 & 6 & 7 \end{pmatrix} \cdot \frac{1}{3}\begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{15}\begin{pmatrix} 14 & 1 \\ 12 & 3 \\ 8 & 7 \end{pmatrix} .$$

Celkem tedy dostaneme

$$\mathbf{P}^\infty = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 14/15 & 1/15 & 0 & 0 & 0 \\ 12/15 & 3/15 & 0 & 0 & 0 \\ 8/15 & 7/15 & 0 & 0 & 0 \end{pmatrix}$$

Pravděpodobnost, že Alice vše prohraje, pokud na začátku měla 3000 dolarů, nyní odpovídá hodnotě

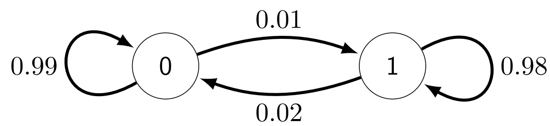
$$\Big(\mathbf{P}^\infty\Big)_{5,1} = \frac{8}{15} .$$

14.6 (Asymptotické pravděpodobnosti)

Při obnovování paměti přepisujeme binární informaci, přičemž s pravděpodobností 1% přepíšeme 0 jako 1, s pravděpodobností 2% přepíšeme 1 jako 0. Jaké bude rozdělení pravděpodobností po velkém počtu kroků?

Řešení:

Cvičení 2.10: `http://cmp.felk.cvut.cz/~navara/stat/MarkovCvic_print.pdf`



14.7 (maximálně věrohodné odhady)

Znaky (A, B, C) jsou permutací stavů $(1, 2, 3)$ Markovova řetězce s maticí přechodu

$$\mathbf{P} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0.1 & 0.8 & 0.1 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$

Všechny stavy mají na počátku stejnou pravděpodobnost.

(a) Odhadněte tuto permutaci z pozorované posloupnosti znaků $(B, C, C, C, A, A, B, A, C)$.

(b) Mějme nyní permutaci, která měla v části (a) největší věrohodnost. Posloupnost znaků z části (a) skončila ve stavu C. Najděte její nejpravděpodobnější pokračování z následujících možností (první uvedený stav C je počátečním stavem této posloupnosti):

(i) (C, A, C, C, B),

(ii) (C, C, C, B, A),

(iii) (C, B, A, A, C).

Řešení:
Příklad 3: `http://cmp.felk.cvut.cz/~navara/stat/pisemky/PST180123res.pdf`

14.8 (maximálně věrohodné odhady)

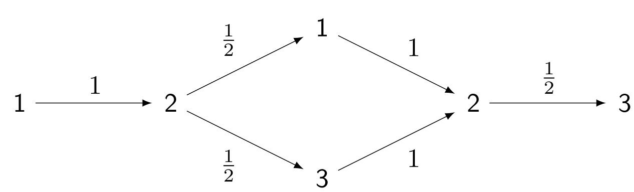
Markovův řetězec má stavy $1, 2, ..., k$, pro $k \geq 3$. V každém kroku lze ze současného stavu i přejít pouze do sousedního (tj. $i - 1$ nebo $i + 1$, pokud daný sousední stav existuje), a to se stejnou pravděpodobností (tj. 1, pokud existuje jen jeden a $1/2$, pokud existují oba). Během 4 kroků se změnil stav z 1 na 3. Pro jaké k model nejlépe vyhovuje tomuto pozorování?

Řešení:

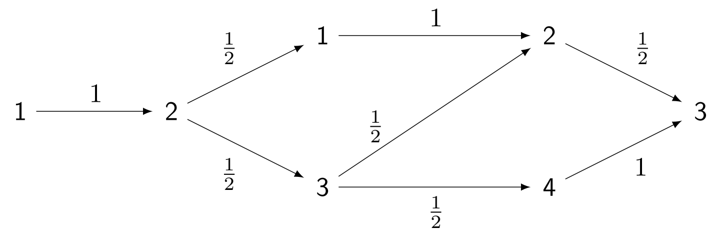
Příklad 3: `http://cmp.felk.cvut.cz/~navara/stat/pisemky/PST190110res.pdf`

K výpočtu hodnot věrohodnosti lze použít i tyto diagramy cest délky čtyři, které jdou ze stavu 1 do stavu 3:

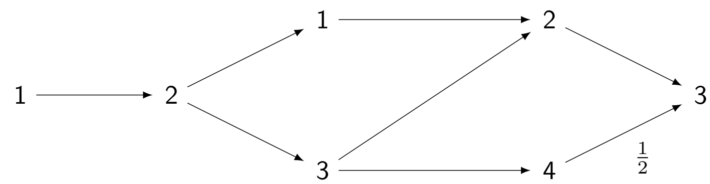
$k = 3:$



$k = 4$:



$k \geq 5$:



U posledního případu $k \geq 5$ jsou zakresleny jen ty hodnoty u šipek, ve kterých se liší od případu $k = 4$.