Příklad 1
Uvažujme dvě absolutně spojité nezávislé náhodné veličiny X a Y se stejným rozdělením. Spočtěte a dokažte $P(X > Y)$.

Klíčovou informací je, že veličiny X a Y jsou spojité, nezávislé a stejně rozdělené.

$$P(X > Y) = P(Y < X) = a$$

(Toto víme, protože jsou veličiny **stejně rozdělené a nezávislé.**)

Dále víme

$$P(X = Y) = 0$$

(Toto víme, protože jsou veličiny **spojité a nezávislé.**)

Platí tedy, že

$$\begin{aligned} P(X = Y) + P(X > Y) + P(Y < X) &= 1 \\ 0 + a + a &= 1 \\ 2a &= 1 \\ a &= \underline{\underline{\frac{1}{2}}} \end{aligned}$$

Příklad 2

Vyjádřete výraz $P\left(X_{n+2}=j,\, X_{n+1}=k,\, X_{n}=i\right) | P\left(X_{n}=i\right)$ pomocí přechodových pravděpodobností $P_{i,j}$.

Příklad 3
Dokažte, že geometrické rozdělení nemá paměť.

- Nemá paměť = nezáleží na historii nějakého jevu.

Předpokládejme dobu čekání $T \sim \text{Geom}(p)$. Pokud jsme již čekali s jednotek času, zbývající čas čekání $T - s$ má opět $\text{Geom}(p)$, tedy stejný, jako kdybychom zatím vůbec nečekali. Snažíme se tedy dokázat

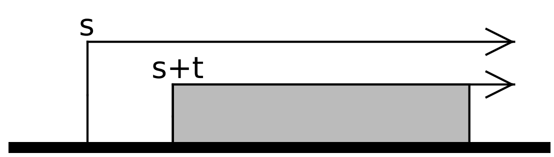
$$P(T - s > k | T > s) \stackrel{?}{=} P(T > k) \text{ pro } k,\, s > 0.$$

Upravme nerovnici a použijme vzorec pro podmíněnou pravděpodobnost a vzorec funkce přežití geometrického rozdělení

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)},\, P(B) \neq 0 \\ P(T > n) &= (1-p)^n \end{aligned}$$

$$P(T - s > k | T > s) = P(T > k + s | T > S) = \frac{P(T > k + s \cap T > s)}{P(T > s)} = \cdots$$

Průnik jevu v čitateli je následující:



celý výraz pokračuje vyjádřením pomocí funkce přežití:

$$\cdots = \frac{P(T > k + s)}{P(T > s)} = \frac{(1-p)^{k+s}}{(1-p)^s} = \frac{\cancel{(1-p)^s}(1-p)^k}{\cancel{(1-p)^s}} = \boxed{(1-p)^k = P(T > k)}$$

Geometrické rozdělení tedy skutečně **nemá paměť**.

Příklad 4
Dokažte, že exponenciální rozdělení nemá paměť.

Struktura tohoto příkladu je stejná jako u důkazu, že geometrické rozdělení nemá paměť.

$$X \sim \text{Exp}(\lambda)$$

$$P(X > u + t | X > u) = \frac{P(X > u + t \cap X > u)}{P(X > u)} = \frac{P(X > u + t)}{P(X > u)} = \cdots$$

Funkce přežití exponenciálního rozdělení je

$$P(X > k) = e^{-\lambda x}$$

$$\cdots = \frac{e^{-\lambda(u+t)}}{e^{-\lambda u}} = \frac{\cancel{e^{-\lambda u}} e^{-\lambda t}}{\cancel{e^{-\lambda u}}} = \boxed{e^{-\lambda t} = P(X > t)}$$

Exponenciální rozdělení tedy skutečně nemá paměť.

Příklad 5

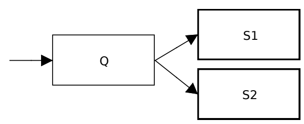
Mějme server, který dokáže zpracovat paralelně až dva požadavky. Pokud je server zahlcen, ukládají se příchozí požadavky do fronty o kapacitě 5. Zpracování požadavků probíhá náhodně a nezávisle. Střední doba zpracování požadavku na serveru je $25ms$.

(A) Server zpracovává dva požadavky. Jaká je průměrná doba odbavení třetího požadavku.

(B) Jaká je průměrná doba zpracování všech tří požadavků.

(C) Jaká je pravděpodobnost, že třetí požadavek bude zpracován jako poslední?

Graf exponenciálních závodů:



Důležité vzorce

$$\begin{aligned} Emin\{S,T\} &= \frac{1}{\lambda+\mu} \\ Emax\{S,T\} &= E(S)+E(T)-Emin\{S,T\} \end{aligned}$$

Za předpokladu

$$S \sim \mathrm{Exp}(\lambda),\, T \sim \mathrm{Exp}(\mu)$$

$$S_1 \sim \mathrm{Exp}(\lambda),\, E(S_1) = 25ms,\, \lambda = \frac{1}{E(S_1)} \rightarrow \lambda = \frac{1}{25}$$

$$S_2 \sim \mathrm{Exp}(\mu),\, E(S_2) = 25ms,\, \mu = \frac{1}{E(S_2)} \rightarrow \mu = \frac{1}{25}$$

A)

S_1a S_2 jsou obsazené, čekáme tedy, až se jeden z nich uvolní (ten nejrychlejší).

$$E(C_A) = Emin\{S_1,S_2\} + E(C) = \frac{1}{\mu+\lambda} + E(C) = \frac{1}{\frac{1}{25}+\frac{1}{25}} + 25 = \ldots = \underline{\underline{37,5ms}}$$

B)

Všechny tři požadavky budou zpracovány, když bude zpracován nejpomalejší z nich.

$$\begin{aligned} E(C_B) &= Emin\{S_1,S_2\} + Emax\{S_1,S_2\} = Emin\{S_1,S_2\} + E(S_1) + E(S_2) - Emin\{S_1,S_2\} = \\ &= E(S_1) + E(S_2) = 25 + 25 = \underline{\underline{50ms}} \end{aligned}$$

C)

$$\begin{aligned} P(C_L) &= P(C_L \cap A_F) + P(C_L \cap B_F) = \frac{\mu}{\mu+\lambda} * \frac{\lambda}{\mu+\lambda} + \frac{\mu}{\mu+\lambda} * \frac{\lambda}{\mu+\lambda} = \\ &= \frac{\frac{1}{25}}{\frac{1}{25}+\frac{1}{25}} * \frac{\frac{1}{25}}{\frac{1}{25}+\frac{1}{25}} + \frac{\frac{1}{25}}{\frac{1}{25}+\frac{1}{25}} * \frac{\frac{1}{25}}{\frac{1}{25}+\frac{1}{25}} = \cdots = \underline{\underline{\frac{1}{2}}} \end{aligned}$$

Vzorec vychází z

$$P(S<T) * P(S>T) + \ldots$$

Příklad 6
Spočtěte entropii geometrického rozdělení, $p = \frac{1}{2}$.

Příklad 7

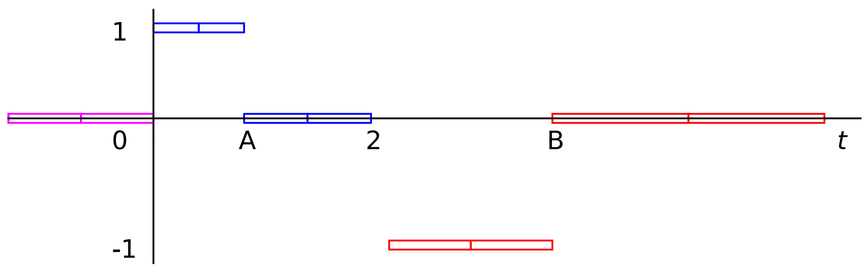
Náhodný proces $X(t)$ je definován jako: $X(t) = 1$ pro $0 \leq t < A$ a $X(t) = -1$ pro $2 \leq t \leq B$. A je rovnoměrně rozdělená veličina na intervalu $(0, 2)$. B je náhodná veličiny s exponenciálním rozdělením s intenzitou $\lambda = 2$.

(A) Nakreslete graf náhodného procesu.

(B) Spočtěte $E(X(t))$.

(C) Určete, zda je proces $X(t)$ stacionární ve střední hodnotě.

A)

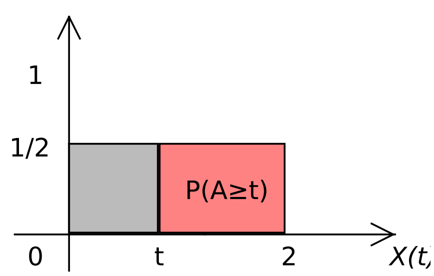


B)

Veličiny A a B jsou jinak rozdělené. Výpočet $E(X)$ tedy rozdělíme na tři části (první interval je jen pro formálnost).

$$\begin{array}{lll} \text{I.} & t \leq 0 & E(X(t)) = 0 \\ \text{II.} & t \in (0,2) & E(X(t)) = 1 * P(0 \leq t \leq A) + \cancel{0 * P(A \leq t \leq 2)} = P(0 \leq t \leq A) = P(A \geq t) \\ \text{III.} & t \geq 2 & E(X(t)) = (-1) * P(t \leq B) + \cancel{0 * P(t \geq B)} = -P(B \geq t) \end{array}$$

Interval II.



Chceme spočítat pravděpodobnost (plocha pod křivkou), že $P(A \geq t)$. Jedná se o jednoduchý výpočet obsahu:

$$P(A \geq t) = (2 - t) * \frac{1}{2} = \boxed{\frac{2-t}{2}}$$

Interval III.

Funkce přežití exponenciálního rozdělení je

$$P(X > k) = e^{-\lambda k}$$

$$-P(B \geq t) = -\left(e^{-\lambda t}\right)[\lambda = 2] = \boxed{-e^{-2t}}$$

Celkový výsledek je tedy:

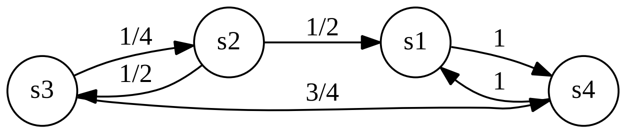
$$E(X(t)) = \underline{\underline{\frac{2-t}{2} - e^{-2t}}}$$

C)

Střední hodnota není ustálená, závisí na čase t. Proces tedy **není stacionární ve střední hodnotě.**

Příklad 8

U markovského řetězce



(A) Definujte rekurentní stav.

(B) Najděte stacionární rozdělení.

(C) Vypište rekurentní stavy.

A)

Stav j se nazývá rekurentní, jestliže řetězec, který z něj vychází. se do j vrátí s pravděpodobností 1 po konečně mnoha krocích. Tj. $P\left(T_j < \infty\right) = 1$.

B)

Sestrojme matici přechodů P

$$P = \begin{array}{c c} & \begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} & \begin{pmatrix} 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Protože je markovským řetězcem popsán jev diskrétní, platí pro stacionární rozdělení:

$$\pi : \pi * P = \pi$$

Sestavme soustavu rovnic

$$\begin{aligned} \frac{1}{2}\pi_2 + \pi_4 &= \pi_1 \\ \frac{1}{4}\pi_3 &= \pi_2 \\ \frac{1}{2}\pi_2 &= \pi_3 \\ \pi_1 + \frac{3}{4}\pi_3 &= \pi_4 \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 &= 1 \end{aligned}$$

Nyní vyřešme soustavu rovnic, pro tento případ použijme GEM.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & \frac{1}{2} & 0 & 1 & 0 \\ 0 & -1 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & -1 & 0 & 0 \\ 1 & 0 & \frac{3}{4} & -1 & 0 \end{pmatrix} \sim \cdots \sim \cdots \Rightarrow \pi = \left(\frac{1}{2}, 0, 0, \frac{1}{2}\right)$$

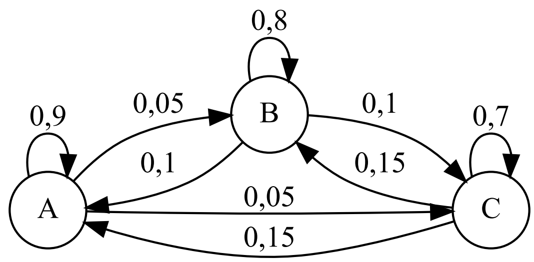
C)

Rekurentní stavy jsou s_1a s_2. Přijít na to můžeme pomocí *silně souvislých komponent* z teorie grafů.

Def.: Pro všechny dvojice vrcholů u a v existuje cesta z u do v a zároveň z v do u.

Příklad 9

Tři výrobky A, B, C mají následující pravděpodobnosti nákupu: $A = 0,9$, $B = 0,8$ a $C = 0,7$. Pro jiný než oblíbený výrobek se zákazníci rozhodnou ve zbytku procent rovnoměrně mezi zbylýma dvěma. Spočtěte oblíbenost jednotlivých výrobků z dlouhodobého hlediska.



Sestavme matici přechodů Q:

$$Q = \begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} \frac{9}{10} & \frac{1}{20} & \frac{1}{20} \\ \frac{1}{10} & \frac{8}{10} & \frac{1}{10} \\ \frac{3}{20} & \frac{3}{20} & \frac{7}{10} \end{pmatrix} \end{array}$$

Pro stacionární rozdělení (*tedy oblíbenost z dlouhodobého hlediska*) platí:

$$\pi * Q = \pi \tag{1}$$

$$\pi_1 + \pi_2 + \ldots + \pi_n = 1 \tag{2}$$

Sestavme rovnice pro výpočet stacionárního rozdělení (po sloupcích):

$$\begin{aligned} \frac{9}{10}\pi_1 + \frac{1}{20}\pi_2 + \frac{1}{20}\pi_3 &= \pi_1 \\ \frac{1}{10}\pi_1 + \frac{8}{10}\pi_2 + \frac{1}{10}\pi_3 &= \pi_2 \\ \frac{3}{20}\pi_1 + \frac{3}{20}\pi_2 + \frac{7}{10}\pi_3 &= \pi_3 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

Soustavu upravíme pro použití v GEM:

$$\begin{aligned} -2\pi_1 + 2\pi_2 + 3\pi_3 &= 0 \\ \pi_1 - 4\pi_2 + 3\pi_3 &= 0 \\ \pi_1 + 2\pi_2 - 6\pi_3 &= 0 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

A vyřešíme (nesmíme zapomenout počítat s pravou stranou matice a s rovnicí $\pi_1 + \pi_2 + \pi_3 = 1$):

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & -6 & 0 \\ 1 & -4 & 3 & 0 \\ -2 & 2 & 3 & 0 \end{pmatrix} \sim \cdots \sim \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -7 & -1 \\ 0 & 0 & -33 & -6 \\ 0 & 0 & 33 & 6 \end{pmatrix}$$

$$\boxed{\pi = \left(\frac{6}{11}, \frac{3}{11}, \frac{2}{11}\right)}$$

Příklad 10
Mějme zadány tabulky rozdělení X a Y a *codeword.*

X, Y	-12	5
-12	$\frac{1}{8}$	$\frac{1}{8}$
4	$\frac{1}{2}$	$\frac{1}{2}$

X, Y	-12	5
-12	110	110
5	10	0

(A) Najděte nejmenší možnou očekávanou délku kódového slova jakéhokoliv binárního kódu pro náhodný vektor X, Y s výše uvedeným rozdělením.

(B) Dosahuje výše uvedené kódování tohoto minima? Proč?

(C) Jsou náhodné veličiny X a Y nezávislé? Proč?

A)
Délku kódového slova spočteme jako

$$\begin{aligned} L(C) &= \mathbb{E}\ell(X) = \sum_i \ell(x_i) * \mathbb{P}(X = x_i) \\ L(C) &\geq H_D(X) \end{aligned}$$

Tedy:

$$L(C) = \frac{1}{8} * 3 + \frac{1}{8} * 3 + \frac{1}{2} * 2 + \frac{1}{2} * 1 = \underline{\underline{2}}$$

B)
Kódování je optimální tehdy, když platí $L(C) = H(X)$.

$$H(X, Y) = -\sum p_i \log_b p_i = -\left(\frac{1}{8}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{2}\log_2\frac{1}{2} + \frac{1}{4}\log_2\frac{1}{4}\right) = -\frac{7}{4}$$

$L(C) \geq H(X)$ kódování tedy **není optimální** (délka slova není optimální)
C)
Pro nezávislost jevů obecně platí

$$P(A \cap B) = P(A) * P(B)$$

V tabulce jsou uvedené sdružené pravděpodobnosti ($P(A \cap B)$), spočteme tedy samostatné pravděpodobnosti a ty porovnáme s tabulkou.

$$\begin{aligned} P(X = -12) &= \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \text{ (součet řádku)} \\ P(Y = -12) &= \frac{1}{8} + \frac{1}{2} = \frac{5}{8} \text{ (součet sloupce)} \\ P(X = -12 \wedge Y = -12) &= \frac{1}{4} * \frac{5}{8} = \frac{5}{32} \end{aligned}$$

$$\frac{5}{32} \neq \underbrace{\frac{1}{8}}_{\text{z tabulky}}$$

Veličiny jsou tedy závislé, tedy nejsou nezávislé.

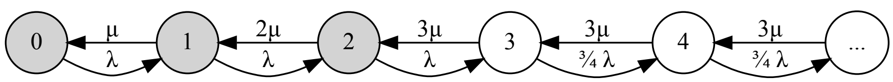
Příklad 11
Systém serverů přijímá požadavky s intenzitou $\lambda = 20$ požadavků za sekundu. V systému jsou 3 servery, jeden požadavek se zpracovává $E(T) = 50ms$. Pokud jsou servery vytížené, s pravděpodobností $P = 25\%$ jsou požadavky zahozeny, jinak se ukládají do fronty.

(A) Nakreslete diagram systému.

(B) Nalezněte stacionární rozdělení.

(C) Jaká je pravděpodobnost, že ve frontě není žádný požadavek?

A)



B)
Řešení tohoto příkladu má **tři fáze:** I. rekurentní předpis, II. explicitní předpis a III. obecný předpis.

$$E(T) = \frac{1}{\mu} \rightarrow 50 = \frac{1}{\mu} \rightarrow \mu = \frac{1}{50} \text{ požadavků na ms}$$

μ musíme převést na stejné jednotky jako λ. $\frac{1}{50} * 1000 \rightarrow \boxed{\mu = 20}$ požadavků za s.
Stacionární rozdělení nalezneme pomocí *detailní rovnováhy* (řetězec je typu „*birth-and-death*“).
I. rekurentní předpis:

$$\begin{array}{ccccccc} \lambda\pi_0 & = & \mu\pi_1 & \rightarrow & \pi_1 & = & \frac{\lambda}{\mu}\pi_0 \\ \lambda\pi_1 & = & 2\mu\pi_2 & \rightarrow & \pi_2 & = & \frac{\lambda}{2\mu}\pi_1 \\ \lambda\pi_2 & = & 3\mu\pi_3 & \rightarrow & \pi_3 & = & \frac{\lambda}{3\mu}\pi_2 \\ \frac{3}{4}\lambda\pi_3 & = & 3\mu\pi_4 & \rightarrow & \pi_4 & = & \frac{\lambda}{4\mu}\pi_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \rightarrow & \pi_{n+1} & = & \frac{\lambda}{4\mu}\pi_n \quad \text{pro } \forall n \geq 3 \end{array}$$

Pomocí zpětného dosazování (π_0) vytvoříme **II. explicitní předpis:**

$$\begin{array}{ccc} \pi_0 & = & \pi_0 \\ \pi_1 & = & \dfrac{\lambda}{\mu}\pi_0 \\ \pi_2 & = & \dfrac{\lambda^2}{2\mu^2}\pi_0 \\ \pi_3 & = & \boxed{\dfrac{\lambda^3}{6\mu^3}\pi_0} \\ \pi_4 & = & \dfrac{\lambda^3}{24\mu^4}\pi_0 \\ \vdots & \vdots & \vdots \end{array}$$

Ve stavu π_3 dojde k ustálení systému a graf bude vypadat až do ∞ stejně. Vyjádříme tedy obecně π_n (**III. obecný předpis**):

$$\pi_n = \pi_3 * \left(\frac{\frac{3}{4}\lambda}{3\mu}\right)^{n-3} = \left(\frac{\lambda^3}{6\mu^3} * \pi_0\right) * \left(\frac{\frac{3}{4}\lambda}{3\mu}\right)^{n-3} = \boxed{\frac{\lambda^3}{6\mu^3} * \pi_0 * \left(\frac{\lambda}{4\mu}\right)^{n-3}} \text{ pro } n \geq 3$$

Dále musí platit, že $\sum_{i=0}^{\infty} \pi_i = 1$ nebo-li $\pi_0 + \pi_1 + \ldots + \pi_n = 1$, tedy

$$\pi_0 + \frac{\lambda}{\mu}\pi_0 + \frac{\lambda^2}{2\mu^2}\pi_0 + \sum_{n=3}^{\infty}\left[\frac{\lambda^3}{6\mu^3} * \pi_0 * \left(\frac{\lambda}{4\mu}\right)^{n-3}\right] = \pi_0 * \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu} + \underbrace{\sum_{n=3}^{\infty}\left[\frac{\lambda^3}{6\mu^3} * \left(\frac{\lambda}{4\mu}\right)^{n-3}\right]}_{S}\right) = 1$$

$$S = \sum_{n=3}^{\infty}\left[\frac{\lambda^3}{6\mu^3} * \left(\frac{\lambda}{4\mu}\right)^{n-3}\right] [\lambda = \mu = 20] = \sum_{n=3}^{\infty} \frac{\cancel{20^3}}{6 * \cancel{20^3}} * \left(\frac{\cancel{20}}{4 * \cancel{20}}\right)^{n-3} = \frac{1}{6}\sum_{n=3}^{\infty}\left(\frac{1}{4}\right)^{n-3} = \ldots$$

Součet nekonečné geometrické řady spočteme jako $S_g = \frac{a_0}{1-q}$ pro $|q| < 1$, tedy:

$$\ldots = \frac{1}{6} * \frac{1}{1 - \frac{1}{4}} = \ldots = \boxed{\frac{2}{9}}$$

$$\pi_0 * \left(1 + 1 + \frac{1}{2} + \frac{2}{9}\right) = 1 \Rightarrow \pi_0 = \frac{18}{\underline{\underline{\underline{49}}}}$$

Výsledné stacionární rozdělení je

$$\pi = \left(\pi_0, \pi_0, \frac{1}{2}\pi_0, \frac{1}{6}\pi_0, \frac{1}{24}\pi_0, \ldots\right) = \boxed{\left(\frac{18}{49}, \frac{18}{49}, \frac{9}{49}, \frac{3}{49}, \frac{3}{196}, \ldots\right)}$$

C)

Fronta je prázdná tehdy, pokud se v systému nachází 0, 1, 2 nebo 3 požadavky.

$$P\left(X\left(t\right) \leq 3\right) = \pi_0 + \pi_1 + \pi_2 + \pi_3 = \frac{18}{49} + \frac{18}{49} + \frac{9}{49} + \frac{3}{49} = \frac{48}{\underline{\underline{\underline{49}}}}$$

Příklad 12
Uvažujte markovský řetězec s konečnou množinou stavů $S = \{s_1, s_2, \ldots, s_n\}$.

(A) Definujte Markovsou podmínku.

(B) Definujte stacionární rozdělení pro tento markovský řetězec.

(C) Pro stav $s_i \in S$ definujte čas návratu τ_i.

(D) Jak souvisí střední hodnota času návratu se stacionárním rozložením?

A) Markovova podmínka

$$P\left(X_n = s_j | X_0 = s_0,\, X_1 = s_1,\, \ldots,\, X_{n-1} = s_i\right) = P\left(X_n = s_j | X_{n-1} = s_i\right) = P_{ij}$$

B) Stacionární rozdělení
π je stacionární rozdělení, jestliže platí

$$\pi * P = \pi \text{ (přesněji zapsáno: } \sum_{\text{all } i} \pi_i P_{ij} = \pi_j)$$

$$(P \text{ je přechodová matice})$$

Pro stacionární rozdělení platí

$$\pi_i \geq 0 \text{ a } \sum_{i=1}^{n} \pi_i = 1$$

C) Čas návratu
Pro Markovův řetězec

$$(X_n)_{n \geq 0}$$

začínající ve stavu $s_i \in S$, definujme **náhodný čas návratu**

$$\tau_i = \inf\{n \geq 1 : \, X_n = s_i\}\,.$$

Náhodný čas návratu je nejmenší čas n, ve kterém se vrátíme do stavu s_i, přičemž počáteční stav se jako návrat nepočítá (proto $n \geq 1$).

Příklad 13
Náhodný výběr 16 hodnot z normálního rozdělení. Je zadán 98% konfidenční interval $(7,398;\, 12,602)$. Otestujte hypotézu $H_0 : \mu = 7,5$ vs. $H_A : \mu \neq 7,5$ s chybou 1. stupně $\alpha = 5\%$.
(Pozn. jsou dostupné tabulky studentova a chi-kvadrát rozdělení.)

Máme zadaný 98% CI, my však potřebujeme 95% interval.
σ^2 neznáme, použijeme tedy $\mathcal{T}$-rozdělení:

$$\mu \in \left(\overline{X}_n - \mathcal{T}_{\frac{\alpha}{2},n-1} \frac{\mathcal{S}_n}{\sqrt{n}};\, \overline{X}_n + \mathcal{T}_{\frac{\alpha}{2},n-1} \frac{\mathcal{S}_n}{\sqrt{n}} \right)$$

$$\overline{X}_n = \frac{7,398 + 12,602}{2} = \frac{20}{2} = 10$$

$$\mathcal{T}_{\frac{\alpha}{2},n-1} = \mathcal{T}_{0,01;,16-1} = \boxed{2,602}$$

(Máme 98% interval /jen na jedné straně/ $\Rightarrow \alpha = 2\% = \frac{0,02}{2} = 0,01$)
(Hodnotu $2,602$ jsme získali z tabulky Studentova rozdělení /$\mathcal{T}$-rozdělení/: řádek 15 a sloupec $0,99$.)

Nyní nám chybí směrodatná odchylka $\mathcal{S}$, tu získáme ze zadaného 98% intervalu:

$$\begin{aligned} CI_L &= \overline{X}_n - \mathcal{T}_{\frac{\alpha}{2},n-1} * \frac{\mathcal{S}_n}{\sqrt{n}} \\ 7,398 &= 10 - 2,602 * \frac{\mathcal{S}_n}{4} \\ \mathcal{S}_n &= \boxed{4} \end{aligned}$$

V tuto chvíli máme všechny potřebné údaje na sestavení 95% CI.

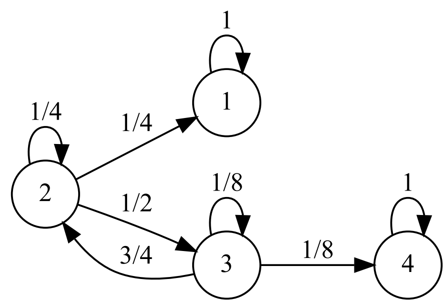
$$CI_{95} = \left(10 - \mathcal{T}_{\frac{0,05}{2},15};\, 10 + \mathcal{T}_{\frac{0,05}{2},15}\right) = (10 - 2,131;\, 10 + 2,131) = \boxed{(7,869;\, 12,131)}$$

(Opět hledáme v tabulce Studentova rozdělení, tentokrát na řádku 15 a sloupci $0,975$ – chceme 95%, na každé straně $2,5\%$)

Ptáme se

$$H_0 \overset{?}{\in} CI_{95}$$
$$7,5 \notin CI_{95} \quad \Rightarrow \quad \text{Hypotézu } H_0 \textbf{ zamítáme}.$$

Příklad 14
Mějme Markovův řetězec



se dvěma absorbujícími stavy (s_1, s_2) a dvěma neabsorbujícími stavy (s_2, s_3).

(A) Definujte fundamentální matici.

(B) Jaký bude střední počet kroků do pohlcení, pokud budeme začínat ve stavu s_3?

(C) Jaká je pravděpodobnost, že budeme pohlceni stavem s_4, pokud začínáme ve stavu s_2?

A)

Fundamentální matice je matice N, $N = (1 - Q)^{-1}$, kde Q je matice přechodů Markovova řetězce $P = \left(\begin{array}{c|c} \boxed{Q} & R \\ \hline 0 & 1 \end{array}\right)$,
kde Q zachycuje tranzientní stavy, R rekurentní stavy (absorbující), 0 je matice nul a 1 je jednotková matice.

B)
Markovský řetězec obsahuje absorbující stavy. Matici přechodů tedy sestavíme dle výše uvedeného.

$$P = \begin{array}{c} \\ s_2 \\ s_3 \\ s_1 \\ s_4 \end{array} \begin{array}{c} \begin{array}{cccc} s_2 & s_3 & s_1 & s_4 \end{array} \\ \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{3}{4} & \frac{1}{8} & 0 & \frac{1}{8} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Pro zjištění středního počtu návštěv potřebujeme fundamentální matici N. Definice říká, že N_{ij} je střední počet návštěv stavu j, začínáme-li ve stavu i a že součet i-tého řádku je počet kroků do pohlcení, začínáme-li ve stavu i.

$$N = (1 - Q)^{-1} = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{8} \end{pmatrix}\right]^{-1} = \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} \\ -\frac{3}{4} & \frac{7}{8} \end{pmatrix}^{-1}$$

Inverzní matici vypočítáme pomocí

$$A^{-1} = \frac{1}{\det A} * \mathrm{adj} A$$

(U adjungované matice nesmíme zapomenout na pronásobení „mřížkou“ $\begin{pmatrix} + & - & \cdots \\ - & + & \cdots \\ + & - & \cdots \end{pmatrix}$ a na závěrečnou **transponaci**)

$$\begin{pmatrix} \frac{3}{4} & -\frac{1}{2} \\ -\frac{3}{4} & \frac{7}{8} \end{pmatrix}^{-1} = \frac{32}{9} * \begin{pmatrix} \frac{28}{9} & \frac{8}{3} \\ \frac{16}{9} & \frac{8}{3} \end{pmatrix}^{T} = \begin{array}{cc} & \begin{array}{cc} s_2 & s_3 \end{array} \\ \begin{array}{c} s_2 \\ s_3 \end{array} & \begin{pmatrix} \frac{28}{9} & \frac{16}{9} \\ \frac{8}{3} & \frac{8}{3} \end{pmatrix} \end{array}$$

(Nezapomeňme, že stavy ve výsledné fundamentální matici N kopírují rozložení původní matice přechodů P)

Pro počet kroků do pohlcení, začínáme-li ve stavu s_3, sečteme druhý řádek: $E\left(\text{pohlcení}|s_3\right) = \frac{8}{3} + \frac{8}{3} = \underline{\underline{\frac{16}{3}}}$

C)

Pro určení pravděpodobnosti potřebujeme matici $B = N * R$.

$$B = \begin{pmatrix} \frac{28}{9} & \frac{16}{9} \\ \frac{8}{3} & \frac{8}{3} \end{pmatrix} * \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{8} \end{pmatrix} = \begin{array}{cc} & \begin{array}{cc} s_1 & s_4 \end{array} \\ \begin{array}{c} s_2 \\ s_3 \end{array} & \begin{pmatrix} \frac{7}{9} & \frac{2}{9} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} \end{array}$$

(Nezapomeňme, že stavy ve výsledné matici B opět kopírují rozložení původní matice přechodů P)

Definice říká, že $B_{ij} = P\left(j|i\right)$, tedy pravděpodobnost, že spadneme do stavu j, začínáme-li ve stavu i.

$$P\left(s_4|s_2\right) = \underline{\underline{\frac{2}{9}}}$$

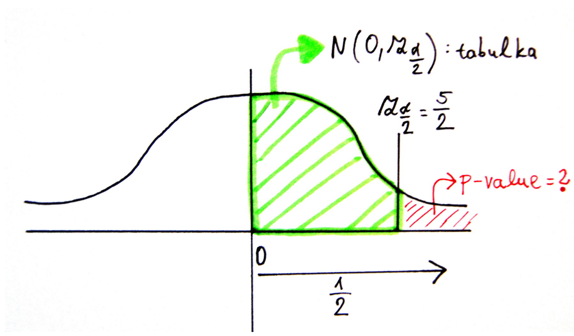
(Tedy řádek s_2a sloupec s_4)

Příklad 15

Mějme N nezávislých jevů $X_1 = \mathrm{Exp}(\lambda_1), \cdots, X_n = \mathrm{Exp}(\lambda_n)$ a jev $Y = \min(X_1, \ldots, X_n)$. Najděte hustotu a distribuční funkci Y.

Příklad 16

Mějme 25 naměřených hodnot. Jejich součet je 250, $\sigma^2 = 16$. Otestujte hypotézu $H_0 : \mu = 12$ oproti $H_A : \mu \neq 12$. Při jaké nejmenší chybě můžeme H_0 ještě zamítnout?



Hledáme p-value,tedy pravděpodobnost, že H_A nastane.

$$\overline{X}_n = \frac{\sum}{N} = \frac{250}{25} = 10$$

Pracujeme jen s jednou stranou intervalu, tedy

$$\begin{aligned}
\mu_0 &= \overline{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}} \\
z_{\frac{\alpha}{2}} &= \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{N}}} \\
z_{\frac{\alpha}{2}} &= \frac{10 - 12}{\frac{\sqrt{16}}{\sqrt{25}}} \\
z_{\frac{\alpha}{2}} &= -\frac{5}{2}
\end{aligned}$$

V tabulce normálního rozdělení nalezneme příslušnou hodnotu

$$N\left(0,\, z_{\frac{\alpha}{2}}\right) = N\left(0,\, \frac{5}{2}\right) = 0,4938$$

$$\frac{\alpha}{2} = \frac{1}{2} - 0,4938 = 0,0062$$

(Chceme zjistit hodnotu pod křivkou jen na jedné straně intervalu)

$$p = 2 * \frac{p}{2} = 2 * 0,0062 = \underline{\underline{0,0124}}$$

(Výsledná pravděpodobnost je však pod křivkou na obou stranách grafu)

p-value je tedy $0,0124 = 1,24\%$. V tomto bodě se přesně láme zamítnout/nezamítnout hypotézu. Pro zamítnutí potřebujeme hodnotu větší, tedy výsledek

$$p\text{-value} = \underline{\underline{0,0124 + \epsilon}}$$

Příklad 17
Mějme dvě veličiny $X_1 \sim \text{Geom}\left(p_1\right)$a $X_2 \sim \text{Geom}\left(p_2\right)$. Určete rozdělení $Y = \min\left\{X_1,\, X_2\right\}$.

Budeme zkoumat

$$P\left(\min\left\{X_1,\, X_2\right\} > k\right) = P\left(X_1 > k,\, X_2 > k\right) = P\left(X_1 > k\right) P\left(X_2 > k\right) = \dots$$

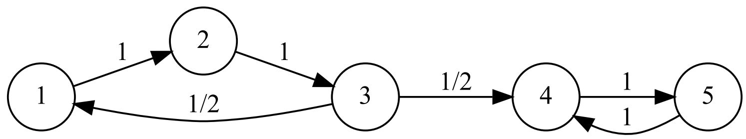
Funkce přežití geometrického rozdělení je

$$P\left(X > k\right) = \left(1 - p\right)^k$$

$$\dots = \left(1 - p\right)^k \left(1 - p\right)^k = \left[\left(1 - p\right)^2\right]^k = \left[1 - \underbrace{\left(2p - p^2\right)}_{p'}\right]^k = \boxed{\left(1 - p'\right)^k}$$

Výsledek má stejný tvar jako funkce přežití geometrického rozdělení, tedy $\min\left\{X_1,\, X_2\right\} \sim Geom\left(p'\right)$, kde $p' = 2p - p^2$.

Příklad 18
Mějme Markovský řetězec v diskrétním čase



(A) Definujte tranzientní stav.

(B) Spočtěte stacionární rozdělení.

(C) Nalezněte transientní stavy. Zdůvodněte.

A)
Tranzientní stav je takový stav, kde existuje nenulová pravděpodobnost, že se do nich nikdy nevrátím.

B)
Sestavíme standardní matici přechodů P.

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Pro výpočet stacionárního rozdělení použijeme GEM. Sestavíme tedy po sloupcích rovnice ($\pi * P = \pi$), pravé strany převedeme tak, aby se rovnice rovnaly 0 a doplníme o $\sum_{\text{all}\, i} = 1$.

$$S = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{pmatrix} \sim \cdots \sim \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 4 & -2 & 1 \\ 0 & 0 & 0 & 4 & -2 & 1 \\ 0 & 0 & 0 & 0 & 2 & 1 \end{pmatrix} \Rightarrow \underline{\underline{\pi = \left(0,\, 0,\, 0,\, \frac{1}{2},\, \frac{1}{2}\right)}}$$

C)
Tranzientní stavy jsou stavy s_1, s_2 a s_3.

Příklad 19
Uvažujme databázový server, na který požadavky přicházejí dle Poissonova procesu s intenzitou 140 požadavků za sekundu. Předpokládejme, že doba obsluhy T má exponenciální rozdělení a trvá v průměru $5ms$. Předpokládejme dále, že server má nekonečnou frontu pro čekající požadavky a že časy příchodů požadavků jsou nezávislé od délky obsluhy. V následujících otázkách uvažujeme stav serveru z dlouhodobého hlediska poté, když se již stabilizoval.

(A) Kolik požadavků je celkem průměrně v systému, kolik jich průměrně čeká na zpracování a kolik jich je v průměru zpracováno?

(B) Jaká je pravděpodobnost, že v systému bude přesně 6 požadavků?

(C) Jaká je průměrná doba čekání ve frontě a jaká je průměrná doba zpracování požadavku?

(D) Za předpokladu, že příchozí požadavek není ihned zpracován, jaká je pravděpodobnost, že bude zpracování trvat délek než $20ms$?

A)
Počet požadavků v systému spočteme jako:

$$N = \rho * \frac{1}{1-\rho}$$

$$\lambda = 140\,rps,\ T_S = 5\,ms = 5 * \frac{1}{1000}\,s$$

Intenzita obsluhy

$$\mu = \frac{1}{T_S} = \frac{1}{\frac{5}{1000}} = 200\,rps$$

Míra vytížení systému ρ je

$$\rho = \frac{\lambda}{\mu} = \frac{140}{200} = \frac{7}{10}$$

Nyní máme vše pro vypočítání počtu požadavků v systému (N):

$$N = \rho * \frac{1}{1-\rho} = \frac{7}{10} * \frac{1}{1-\frac{7}{10}} = \underline{\frac{7}{3}}$$

Nyní použijeme vztah pro celkový průměrný počet požadavků v systému N: (pozn.: Q – queue, S – service)

$$\begin{aligned} N &= N_Q + N_S \\ \frac{7}{3} &= N_Q + \frac{7}{10} \text{ (protože } N_S = \rho) \\ N_Q &= \underline{\underline{\frac{49}{30} \doteq 1,63}} \end{aligned}$$

B)

$$P(n=6) = \rho^n (1-\rho) = \left(\frac{7}{10}\right)^6 \left(1 - \frac{7}{10}\right) \doteq \underline{\underline{0,035}}$$

C)

Využijeme vztahu

$$\begin{aligned}\mathbf{T} &= \mathbf{T_Q} + \mathbf{T_S} \\ T_Q &= T - T_S \\ T_Q &= T - \frac{5}{1000}\end{aligned}$$

Celkový čas požadavku v systému T neznáme, využijeme vztahu

$$\begin{aligned}\mathbf{N} &= \lambda * \mathbf{T} \\ \frac{7}{3} &= 140 * T \\ T &= \frac{1}{60} s\end{aligned}$$

$$T_Q = \frac{1}{60} - \frac{5}{1000} = \underline{\underline{\frac{7}{600}}} s$$

D)

$$2ms = \frac{2}{1000} s$$

$$P\left(X > \frac{2}{1000}\right) = e^{-x*T'} = \dots$$

$$T' = \mu - \lambda = 200 - 140 = 60$$

$$\dots = e^{-\frac{2}{1000}*60} \doteq \underline{\underline{0,89}}$$

Příklad 20
Mějme $S \sim Exp\left(\lambda\right)$ a $T \sim Exp\left(\mu\right)$ a $V = min\left\{S, T\right\}$.

(A) Spočtěte $E\left(V\right)$.

(B) JNajděte distribuční funkci minima.

A)
Nejprve musíme zjistit, jak je funkce $min\left\{S, T\right\}$ rozdělena a jaké má parametry:

$$P\left(min\left\{S, T\right\} > k\right) = P\left(S > k, T > k\right) = P\left(S > k\right) P\left(T > k\right) = \left(1 - F_S\left(k\right)\right) * \left(1 - F_T\left(k\right)\right) = \dots$$

$$1 - F_{Exp}\left(k\right) = 1 - \left(1 - e^{-\lambda k}\right) = e^{-\lambda k}$$

$$\dots = e^{-\lambda k} * e^{-\mu k} = \underline{\underline{e^{-k(\lambda+\mu)}}}$$

Nyní víme, že $min\left\{S, T\right\}$ je rozděleno exponenciálně s parametrem $(\lambda + \mu)$. Pro střední hodnotu exponenciálního rozdělení platí

$$E\left(X \sim Exp\left(\lambda\right)\right) = \frac{1}{\lambda}$$

pro V tedy platí

$$E\left(V\right) = \frac{1}{\lambda + \mu}$$

B)
Již víme, že

$$V \sim Exp\left(\lambda + \mu\right)$$

Distribuční funkce exponenciálního rozdělení F je

$$F = 1 - e^{-\lambda x}$$

Distribuční funkce V je tedy

$$F_V\left(k\right) = \underline{\underline{1 - e^{-k(\lambda+\mu)}}}$$