4. cvičení z PSI

19. - 23. října 2015

4.1 (Poissonovo rozdělení) Kolem okna projíždí auta. Průměrně jich projede 6 za minutu. Určete pravděpodobnost, že

(a) jich za minutu projede méně než 3,

(b) během dvou minut neprojede žádné.

Pokud průměrný počet aut bude 6.8 za minutu, která celočíselná hodnota $k \in \mathbb{N}_0$ počtu aut bude nejpravděpodobnější?

Řešení:
Použijeme Poissonovo rozdělení. Toto rozdělení popisuje pravděpodobnost veličiny

$$X = \textit{počet událostí v daném časovém úseku},$$

pokud jsou splněny následující podmínky

- události je možné oddělit intervaly stejné délky (tj. pravděpodobnost, ze nastane více než jedna událost v daných rozděleních je nulová),
- výskyt dané události je nezávislý na minulých výskytech,
- události pocházejí z velkého poctu zdrojů.

V praxi jde např. o příchod zákazníka do fronty, průjezd aut, chytání ryb atd. během nějaké předem určené doby. Odvodíme si jeho rozdělení. Časový interval si rozdělíme na n dílků tak malých, aby v každém byla maximálně jedna událost se stejnou pravděpodobností p_n. Dostaneme tak binomické rozdělení veličiny

$$X_n = \textit{počet událostí v daném časovém úseku rozděleném na } n \textit{ dílků}$$

se střední hodnotou $\lambda = E(X_n) = n \cdot p_n$, kterou si vezmeme jako pevnou a Poissonovo rozdělení si definujeme jako limitu:

$$P(X=k) = \lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} \binom{n}{k} (p_n)^k (1-p_n)^{n-k} = \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \Big(\frac{\lambda}{n}\Big)^k \Big(1-\frac{\lambda}{n}\Big)^{n-k} =$$

$$= \frac{\lambda^k}{k!} \lim_{n\to\infty} \Big(1-\frac{\lambda}{n}\Big)^n \Big(1-\frac{\lambda}{n}\Big)^{-k} \prod_{i=0}^{k} \Big(\frac{n-i}{n}\Big) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \ .$$

Parametr λ tedy představuje střední hodnotu (skutečně je $E(X) = k$). Poissonovo rozdělení je vetšinou spíše limitní případ a používá se jako aproximace binomického rozdělení $\mathrm{Bi}(n,p)$ u kterého neznáme n a p a kde n je dostatečně velké.

V našem případě je tedy $\lambda = 6\ min^{-1}$ (rozměrově je to minuta na minus první).

(a)

$$P(X<3) = P(X=0) + P(X=1) + P(X=2) = e^{-\lambda}\Big(1+\lambda+\frac{\lambda^2}{2!}\Big) =$$

$$= e^{-6}\Big(1+6+\frac{6^2}{2}\Big) = 25e^{-6} \doteq 0.062$$

(b) Zde je interval delší a během něj je očekávaná střední hodnota $\lambda' = 12\ min^{-1}$. Takže

$$P(X' = 0) = e^{-\lambda'} = e^{-12} \doteq 6.14 \cdot 10^{-6}.$$

Protože pro rozdělení pravděpodobností platí

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \frac{\lambda}{k} \cdot P(X = k-1)$$

pro $k = 1, 2, \ldots$, tak hodnoty s nejvyšší pravděpodobností budou:

$$k = \begin{cases} \lfloor \lambda \rfloor & , \lambda \notin \mathbb{N} \\ \lambda, \lambda - 1 & , \lambda \in \mathbb{N} \end{cases}$$

Pokud tedy máme $\lambda = 6.8$, tak kupodivu hodnota, která má nejvyšší pravděpodobnost bude $k = \lfloor 6.8 \rfloor = 6$ a nikoliv bližší zaokrouhlena střední hodnota $E(X) \doteq 7$. To ale není žádný rozpor. Střední hodnota prostě nemusí odpovídat (v zaokrouhlení) hodnotě s nejvyšší pravděpodobností.

4.2 (aproximace binomického rozdělení Poissonovým) Na telefonní ústřednu je napojeno $n = 300$ účastníků. Každý z nich bude volat během hodiny s pravděpodobností $p = 0.01$. Jaká je pravděpodobnost toho, že během hodiny zavolají právě 4 účastníci?

Řešení:
Budeme předpokládat, ze účastníci volají nezávisle. Jde o binomické rozdělení a pravděpodobnost bude

$$P(X = 4) = \binom{n}{4} \cdot p^4 \cdot (1-p)^{n-4} = \binom{300}{4} \cdot 0.01^4 \cdot 0.99^{296} .$$

Protože zde násobíme velká čísla malými, případně počítáme vysoké mocniny, bude výhodnější pro výpočet použít aproximaci pomoci Poissonova rozdělení, pro něž máme splněny předpoklady. Parametrem bude střední hodnota binomického rozdělení, tj. $\lambda = E(X) = n \cdot p = 300 \cdot 0.01 = 3$. Pak máme

$$P(X = 4) \doteq \frac{\lambda^4}{4!} \cdot e^{-\lambda} = \frac{3^4}{4!} \cdot e^{-3} \doteq 0.168 .$$

4.3 (směs náhodných veličin) V urně je 15 hracích kostek, z toho 10 správných, na nichž padají všechna čísla se stejnou pravděpodobností, a 5 vadných, na nichž padá šestka s pravděpodobností $\frac{1}{2}$, ostatní čísla s pravděpodobností $\frac{1}{10}$. Náhodně vybereme jednu kostku a hodíme. Jaká je pravděpodobnost možných výsledků?

Řešení:
Použijeme jev

$A =$ "vytažená kostka je správná",

a veličinu

$X =$ "hodnota, která padla na vytažené kostce".

Pak pro $i \in \{1, \ldots, 6\}$ dostáváme z Bayesovy vety

$$P(X = i) = P(X = i|A) \cdot P(A) + P(X = i|\overline{A}) \cdot P(\overline{A}).$$

Z předpokladu pak máme $P(A) = \frac{10}{15} = \frac{2}{3}$, $P(X = i|A) = \frac{1}{6}$ pro všechna i a

$$P(X = i|\overline{A}) = \begin{cases} \frac{1}{2} & , i = 6 \\ \frac{1}{10} & , i \neq 6 \end{cases} .$$

Celkem tedy dostáváme pro $i = 6$:

$$P(X = i) = \frac{1}{6} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{5}{18}$$

a pro $i \in \{1, \ldots, 5\}$:

$$P(X = i) = \frac{1}{6} \cdot \frac{2}{3} + \frac{1}{10} \cdot \frac{1}{3} = \frac{13}{90}.$$

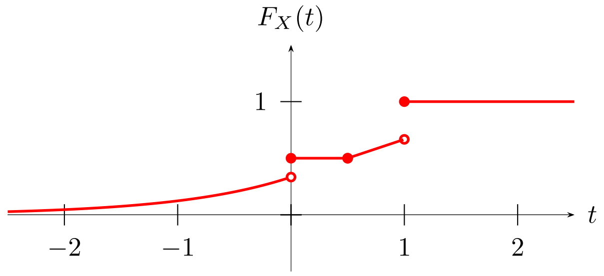
Veličinu X také můžeme chápat jako směs $\mathrm{Mix}_c(X_1, X_2)$, kde veličiny $X_1 : A \to \mathbb{R}$ a $X_2 : \overline{A} \to \mathbb{R}$ znamenají hodnoty, které padnou na správných, resp. na vadných kostkách. Koeficient c pak odpovídá poměru správných kostek ku všem kostkám, tj. $c = \frac{10}{15} = P(A)$. Při tomto postupu bychom dále už jen kopírovali Bayesovu vetu.

4.4 (rozklad náhodné veličiny) Náhodná veličina X má distribuční funkci

$$F_X(t) = \begin{cases} \frac{e^t}{3} & , \ t < 0 \\ \frac{1}{2} & , \ t \in \langle 0, \frac{1}{2}) \\ \frac{t+1}{3} & , \ t \in \langle \frac{1}{2}, 1) \\ 1 & , \ 1 \leq t. \end{cases}$$

Rozložte X na směs diskrétní náhodné veličiny U a spojité náhodné veličiny V. Pro U a V potom najděte distribuční funkce a pravděpodobnostní funkci, resp. hustotu.

Řešení:
Abychom o distribuční funkci F_X měli lepší představu, znázorníme si její graf:



Připomeňme si, jak se hledá diskrétní a spojitá část veličiny X. Nejdříve si určíme množinu

$$O_X := \{a \in \mathbb{R} \mid P(X = a) \neq 0\} = \text{"}\textit{množina všech bodů nespojitosti funkce } F_X\text{"} = \{0, 1\}.$$

Pak máme

$$X = \mathrm{Mix}_c(U, V),$$

kde

$$c = P(X \in O_X) = \sum_{a \in O_X} P(X = a) = \text{"}\textit{součet všech skoků funkce } F_X\text{"} = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}.$$

To dostaneme pomocí vztahu

$$P(X = a) = F_X(a) - F_X(a-),$$

a jednotlivé skoky jsou tedy

$$P(X = 0) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \qquad \text{a} \qquad P(X = 1) = 1 - \frac{2}{3} = \frac{1}{3}.$$

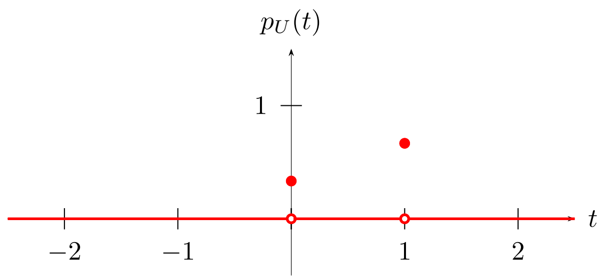
Z definice směsi veličin $X = \mathrm{Mix}_c(U, V)$ pak pro pravděpodobnostní míry platí

$$P_X = cP_U + (1 - c)P_V$$

a tedy pravděpodobnostní funkce diskrétní veličiny U bude

$$p_U(t) = P_U\big(\{t\}\big) = \frac{1}{c} P_X\big(\{t\}\big) = 2 \cdot P(X = t) = \begin{cases} \frac{1}{3} & , \ t = 0 \\ \frac{2}{3} & , \ t = 1 \\ 0 & , \ \text{jinak} \end{cases}$$

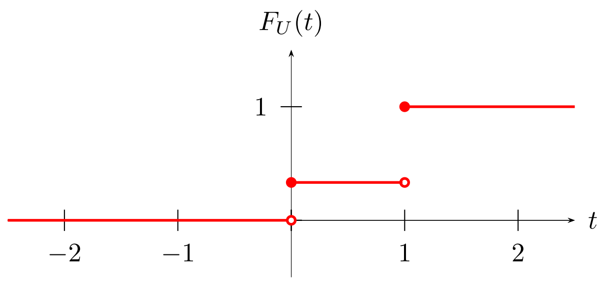
s grafem:



Pro distribuční funkci F_U pak nasčítáním dostaneme

$$F_U(t) = \sum_{x \leq t} p_U(x) = \begin{cases} 0 & , \ t < 0 \\ \frac{1}{3} & , \ 0 \leq t < 1 \\ 1 & , \ 1 \leq t \end{cases}$$

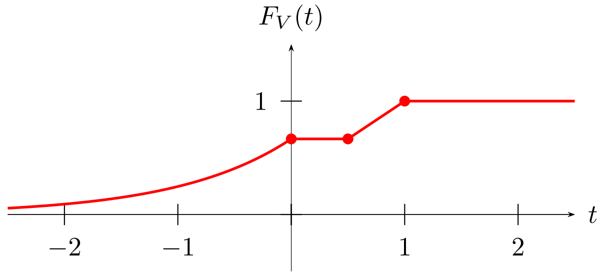
s grafem:



Pro distribuční funkci spojité veličiny V pak ze vztahu $F_X = cF_U + (1 - c)F_V$ dostáváme

$$F_V(t) = \frac{F_X(t) - cF_U(t)}{1 - c} = 2F_X(t) - F_U(t) = \begin{cases} \frac{2}{3}e^t & , \ t < 0 \\ \frac{2}{3} & , \ t \in \langle 0, \frac{1}{2}) \\ \frac{2t+1}{3} & , \ t \in \langle \frac{1}{2}, 1) \\ 1 & , \ 1 \leq t \end{cases}$$

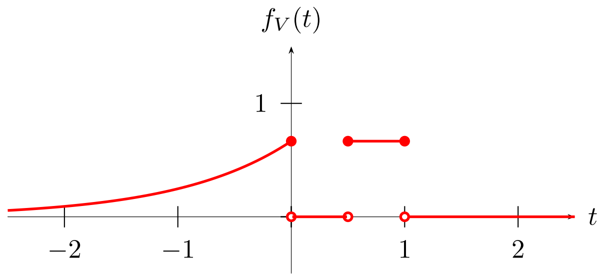
Graf funkce F_V dostaneme jednoduše tak, ze části grafu F_X, které na sebe nenavazují, posuneme dolů tak, aby výsledek byl spojitý. Celý graf pak ve směru y natáhneme tak, aby v nekonečnu měl limitu rovnou 1:



Hustotu f_V spojité veličiny V pak získáme derivací F_V pro body, kde derivace existuje. V ostatních (konečně mnoha bodech) na hodnotách nezáleží. Takže můžeme psát např. toto:

$$f_V(t) = F_V'(t) = \begin{cases} \frac{2}{3}e^t & , \ t \leq 0 \\ \frac{2}{3} & , \ t \in \langle \frac{1}{2}, 1 \rangle \\ 0 & , \ \text{jinak.} \end{cases}$$

Graf f_V pak bude:



Ještě pro další připomenutí: Diskrétní část U veličiny $X : \Omega \to \mathbb{R}$ je definovaná jako zúžení X na množinu

$$\Omega_1 := X^{-1}(O_X) \subseteq \Omega,$$

tedy

$$U := X|_{\Omega_1},$$

spojitá část je pak zúžení X na zbytek prostoru, tj.

$$V := X|_{\overline{\Omega_1}} .$$

Vztah $c = P(X \in O_X) = P_X(O_X)$ dostaneme pomocí pravděpodobnostních měr z definice směsi

$$P_X = cP_U + (1-c)P_V$$

Z toho, ze V je spojitá a množina O_X je nejvýše spočetná, máme $P_V(O_X) = \sum_{a \in O_X} P_V(\{a\}) = 0$. Z definice U pak plyne, ze množina všech jejich hodnot je právě O_X, tedy $U^{-1}(O_X) = \Omega_1 = U^{-1}(\mathbb{R})$. Proto dostáváme $P_U(O_X) = P_U(\mathbb{R}) = 1$. Takže po dosazení získáme

$$P_X(O_X) = cP_U(O_X) + (1-c)P_V(O_X) = c .$$