5. cvičení z PST

21. října 2020

Příklady **4.7**, **4.8**.

Co je směs: Mějme Kolmogorův model $(\Omega, \mathcal{A}, P)$, kde množina výsledků Ω se skládá ze dvou *disjunktních* jevů Ω_1 a Ω_2. Na každé z částí Ω_i vznikne Kolmogorův model s pravděpodobností $P_i(\,\cdot\,) := P(\,\cdot\,|\Omega_i)$ pro $i = 1, 2$.

Jestliže nyní máme náhodné veličiny $X_i : \Omega_i \to \mathbb{R}$ na každém z odvozených Kolmogorových modelů, pak veličina $X : \Omega \to \mathbb{R}$ definovaná sloučením obou veličin, tedy jako

$$X(\omega) = \begin{cases} X_1(\omega) & , \omega \in \Omega_1 \\ X_2(\omega) & , \omega \in \Omega_2 \end{cases}$$

se nazývá *směsí* veličin X_1 a X_2 a označuje se jako $X = \mathrm{Mix}_c(X_1, X_2)$, kde $c = P(\Omega_1)$.

V této chvíli se může zdát zbytečné vypisovat ještě konstantu c, která označuje pravděpodobnost výsledku z Ω_1. Způsob, který jsme teď popsali, je vlastně rozdělením původního modelu na dva odvozené.

Můžeme však také začít opačně - tedy vzít dva "nesouvisející" modely, tj. disjunktní množiny výsledků Ω_1 a Ω_2 s příslušnými pravděpodobnostmi P_1 a P_2 a z nich složit nový Kolmogorův model. Tento nový model bude přirozeně mít množinu výsledků $\Omega := \Omega_1 \cup \Omega_2$. Pravděpodobnost P na novém modelu ovšem nebude určena, dokud si nestanovíme, jakou chceme hodnotu $c = P(\Omega_1)$ (a tím i doplňkovou hodnotu $1 - c = P(\Omega_2)$). Protože opět chceme, abychom měli $P(\,\cdot\,|\Omega_i) = P_i(\,\cdot\,)$ pro $i = 1, 2$, tak nyní bude už pravděpodobnost P určená (z věty o úplné pravděpodobnosti) pro jevy $A_1 \subseteq \Omega_1$ a $A_2 \subseteq \Omega_2$ jako

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) = P(\Omega_1) \cdot P(A_1|\Omega_1) + P(\Omega_2) \cdot P(A_2|\Omega_2) =$$
$$= c \cdot P_1(A_1) + (1 - c) \cdot P_2(A_2)$$

tedy zde nutně *musíme* používat konstantu c, která pro různé hodnoty vytvoří různé Kolmogorovy modely a tím i různá rozdělení veličiny

$$X = \mathrm{Mix}_c(X_1, X_2)$$

jejíž předpis ovšem zůstane stále stejný (bez ohledu na c)!

(Rozdělení veličiny X se bude ale pochopitelně měnit podle toho, jaké pravděpodobnosti vzniknou na základě volby c.)

Připomenutí: Nechť $a \in \mathbb{R}$ je bod *spojitosti distribuční funkce* F_X náhodné veličiny X. Pak máme

$$P(X = a) = P(X \leq a) - P(X < a) = F_X(a) - \lim_{x \to a_-} F_X(t) = 0$$

a tedy

$$P(X \leq a) = P(X < a)$$

neboli v bodech spojitosti nezáleží na typu nerovnosti (neostré vs. ostré).

Speciálně pro veličinu X se spojitým rozdělením je $P(X = a) = 0$ pro každé $a \in \mathbb{R}$.

5.1 (rozdělení veličiny vytvořené jako směs)

Program generuje čísla z intervalu $(0, 1)$, která jsou rovnoměrně rozdělena. S pravděpodobností $c = \frac{1}{4}$ jsou ale někdy výstupem i čísla 0 a 1, a to v poměru 1 : 2. Náhodná veličina Z udává hodnoty výstupu generátoru. Určete:

(a) distribuční funkci F_Z.

(b) pravděpodobnosti $P(0 \leq Z \leq \frac{1}{2})$ a $P(Z > \frac{1}{3})$.

(c) střední hodnotu $E(Z)$.

(d) $t \in \mathbb{R}$ takové, že $P(Z \leq t) = 0.75$.

Řešení:
V tomto případě bude Ω množina všech výstupů z programu (výstupy v různých časech navzájem rozlišujeme). Rozdělíme ji na

$$\Omega_1 = \text{"}\textit{výstupy, které mají hodnotu } 0 \textit{ nebo } 1\text{"}$$

a

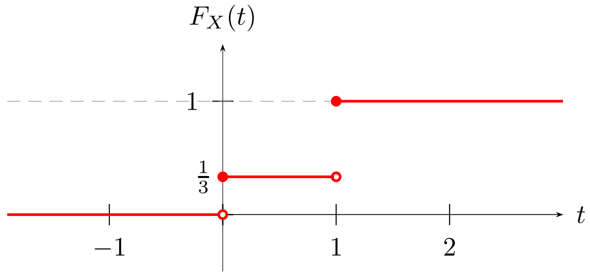
$$\Omega_2 = \text{“}\textit{výstupy, které mají hodnotu v } (0,1)\text{”}$$

kde $P(\Omega_1) = \frac{1}{4}$.

Náhodná veličina Z je směsí $\mathrm{Mix}_{1/4}(X, Y)$, kde veličina $X : \Omega_1 \to \mathbb{R}$ je hodnota diskrétního výstupu (hodnoty $\{0, 1\}$) a $Y : \Omega_2 \to \mathbb{R}$ je hodnota spojitého výstupu (hodnoty z intervalu $(0, 1)$).

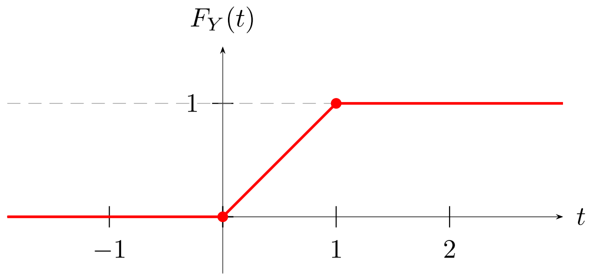
(a) Veličina X má alternativní rozdělení $\mathrm{Alt}(\frac{2}{3})$ s pravděpodobnostní funkcí $p_X(0) = \frac{1}{3}$, $p_X(1) = \frac{2}{3}$ a distribuční funkcí

$$F_X(t) = \sum_{u \le t} p_X(u) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1}{3}\,, & 0 \le t < 1\,, \\ 1\,, & t \ge 1\,. \end{cases}$$



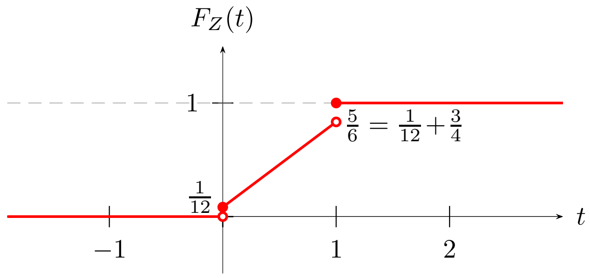
Veličina Y má spojité rozdělení s hustotou $f_Y(t) = \begin{cases} 1\,, & t \in (0,1)\,, \\ 0\,, & \text{jinak}\,, \end{cases}$ a distribuční funkcí

$$F_Y(t) = \int_{-\infty}^{t} f_Y(u)\ \mathrm{d}u = \begin{cases} & 0\,, & t < 0\,, \\ \int_0^t 1\ \mathrm{d}u = & t\,, & 0 \le t < 1\,, \\ & 1\,, & t \ge 1\,. \end{cases}$$



Pro distribuční funkci veličiny $Z = \mathrm{Mix}_{1/4}(X, Y)$ pak platí

$$F_Z(t) = \frac{1}{4} F_X(t) + \frac{3}{4} F_Y(t) = \begin{cases} 0 & , t < 0\,, \\ \frac{1}{12} + \frac{3}{4}t & , 0 \le t < 1\,, \\ 1 & , t \ge 1\,. \end{cases}$$



(b) $P(0 \leq Z \leq \frac{1}{2}) = P(Z \leq \frac{1}{2}) - P(Z < 0) = F_Z(\frac{1}{2}) - \lim_{t \to 0_-} F_Z(t) = \frac{1}{12} + \frac{3}{4} \cdot \frac{1}{2} - 0 = \frac{11}{24}$
$P(Z > \frac{1}{3}) = 1 - P(Z \leq \frac{1}{3}) = 1 - F_Z(\frac{1}{3}) = 1 - \left(\frac{1}{12} + \frac{1}{4}\right) = \frac{2}{3}$.

(c) Pro alternativní rozdělení veličiny X je $E(X) = 1 \cdot p_X(1) = \frac{2}{3}$ a pro rovnoměrné rozdělení veličiny Y na intervalu $(a, b) = (0, 1)$ je $E(Y) = \frac{a+b}{2} = \frac{1}{2}$. Takže pro $Z = \text{Mix}_{1/4}(X, Y)$ je

$$E(Z) = \frac{1}{4}E(X) + \frac{3}{4}E(Y) = \frac{1}{4} \cdot \frac{2}{3} + \frac{3}{4} \cdot \frac{1}{2} = \frac{13}{24} \; .$$

(d) Hledáme $t \in \mathbb{R}$ tak, že $0.75 = P(Z \leq t) = F_Z(t)$. K tomu potřebujeme vědět, kterou část předpisu pro F_Z máme použít. Protože $\frac{1}{12} \leq 0.75 < \frac{5}{6}$, což je rozmezí hodnot na poslední rostoucí části předpisu, tak musíme použít právě tuto část:

$$\frac{3}{4} = F_Z(t) = \frac{1}{12} + \frac{3}{4}t \qquad \Rightarrow \qquad t = \frac{8}{9} \in \langle 0, 1) \; .$$

5.2 (rozdělení veličiny vytvořené jako směs)
Náhodná veličina X má diskrétní rozdělení s pravděpodobnostní funkcí $p_X(0) = 0.4$ a $p_X(1) = 0.6$. Náhodná veličina Y má spojité rovnoměrné rozdělení na intervalu $\langle -1, 2\rangle$. Náhodná veličina Z je směsí $Z = \text{Mix}_{1/4}(X, Y)$. Určete:

(a) distribuční funkci F_Z.

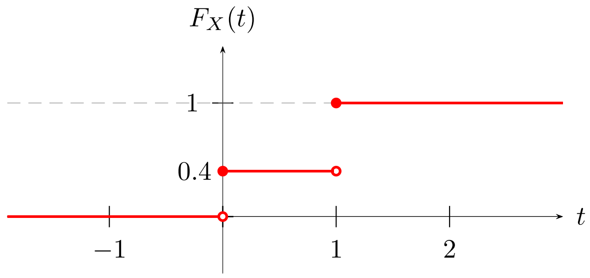
(b) pravděpodobnosti $P(0 \leq Z \leq 1)$ a $P(Z \geq 0.5)$.

(c) střední hodnotu $E(Z)$.

(d) $t \in \mathbb{R}$ takové, že $P(Z \leq t) = 0.9$.

Řešení:
(a) Veličina X s alternativním rozdělením Alt(0.6) má distribuční funkci

$$F_X(t) = \sum_{u \leq t} p_X(u) = \begin{cases} 0\,, & t < 0\,, \\ 0.4\,, & 0 \leq t < 1\,, \\ 1\,, & t \geq 1\,. \end{cases}$$

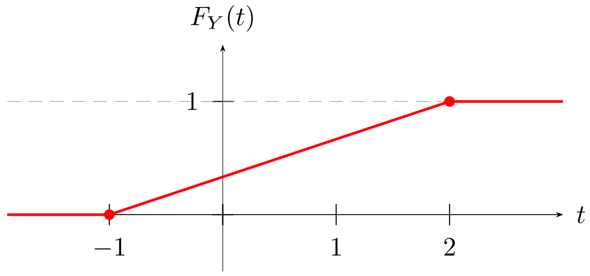


Veličina Y má spojité rovnoměrné rozdělení na $\langle -1, 2\rangle$ s hustotou

$$f_Y(t) = \begin{cases} \frac{1}{2-(-1)} = \frac{1}{3}\,, & t \in \langle -1, 2\rangle\,, \\ 0\,, & \text{jinak}\,, \end{cases}$$

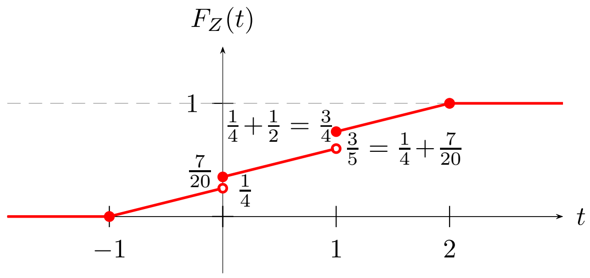
a distribuční funkcí

$$F_Y(t) = \int_{-\infty}^{t} f_Y(u)\,\mathrm{d}u = \begin{cases} & 0, & t < -1\,, \\ \int_{-1}^{t} \frac{1}{3}\,\mathrm{d}u = & \frac{t+1}{3}, & -1 \le t < 2\,, \\ & 1, & t \ge 2\,. \end{cases}$$



Pro distribuční funkci veličiny $Z = \mathrm{Mix}_{1/4}(X, Y)$ pak platí

$$F_Z(t) = \frac{1}{4}F_X(t) + \frac{3}{4}F_Y(t) = \begin{cases} & 0, & t < -1\,, \\ \frac{1}{4}\cdot 0 + \frac{3}{4}\cdot\frac{t+1}{3} = & \frac{1}{4}t + \frac{1}{4}, & -1 \le t < 0\,, \\ \frac{1}{4}\cdot 0.4 + \frac{3}{4}\cdot\frac{t+1}{3} = & \frac{1}{4}t + \frac{7}{20}, & 0 \le t < 1\,, \\ \frac{1}{4}\cdot 1 + \frac{3}{4}\cdot\frac{t+1}{3} = & \frac{1}{4}t + \frac{1}{2}, & 1 \le t < 2\,, \\ & 1, & t \ge 2\,. \end{cases}$$



(b) $P(0 \le Z \le 1) = P(Z \le 1) - P(Z < 0) = F_Z(1) - \lim\limits_{t \to 0_-} F_Z(t) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$

$P(Z \ge 0.5) = 1 - P(Z < 0.5) = 1 - \lim\limits_{t \to 0.5_-} F_Z(t) = 1 - \left(\frac{1}{4} \cdot 0.5 + \frac{7}{20}\right) = \frac{21}{40}$.

(c) Pro alternativní rozdělení veličiny X je $E(X) = 1 \cdot p_X(1) = 0.6 = \frac{3}{5}$ a pro rovnoměrné rozdělení veličiny Y na intervalu $(a, b) = (-1, 2)$ je $E(Y) = \frac{a+b}{2} = \frac{1}{2}$. Takže pro $Z = \mathrm{Mix}_{1/4}(X, Y)$ je

$$E(Z) = \tfrac{1}{4}E(X) + \tfrac{3}{4}E(Y) = \tfrac{1}{4} \cdot \tfrac{3}{5} + \tfrac{3}{4} \cdot \tfrac{1}{2} = \tfrac{21}{40}\ .$$

(d) Hledáme $t \in \mathbb{R}$ tak, že $0.9 = P(Z \le t) = F_Z(t)$. K tomu potřebujeme vědět, kterou část předpisu pro F_Z máme použít. Protože $\frac{3}{4} \le 0.9 \le 1$, což je rozmezí hodnot na poslední rostoucí části předpisu, tak musíme použít právě tuto část:

$$0.9 = F_Z(t) = \tfrac{1}{4}t + \tfrac{1}{2} \qquad \Rightarrow \qquad t = 3.6 - 2 = 1.6 \in \langle 1, 2 \rangle\ .$$

Poznámka: Pro *libovolnou* veličinu X a její distribuční funkci F_X si můžeme obecně definovat funkci

$$p_X(t) := F_X(t) - F_X(t_-) \ \Big(= P(X = t) \Big) \quad \text{pro všechna } t \in \mathbb{R}$$

(kde $F_X(t_-)$ je limita zleva funkce F_X v bodě t) a funkci

$$f_X(t) := \begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} F_X(t) & \text{, pro taková } t \in \mathbb{R}, \text{ kde konečná derivace existuje} \\ 0 & \text{, jinak.} \end{cases}$$

Nyní platí toto: Jestliže rozdělení X je

- *diskrétní* $\Rightarrow$ p_X je její pravděpodobnostní funkce a $f_X(t) = 0$ pro všechna $t \in \mathbb{R}$,
- *(absolutně) spojité* $\Rightarrow$ f_X je její hustota pravděpodobnosti a $p_X(t) = 0$ pro všechna $t \in \mathbb{R}$.

Mějme veličinu $X : \Omega \to \mathbb{R}$ na Kolmogorově modelu s pravděpodobností P a nechť X je směs veličin

$$X = \mathrm{Mix}_c(D, S)$$

kde D má diskrétní rozdělení a S má (absolutně) spojité rozdělení. Pak platí

$$F_X(t) = c \cdot F_D(t) + (1 - c) \cdot F_S(t) \quad \text{pro všechna} \ \ t \in \mathbb{R}$$

a odsud máme

$$P(X = t) = p_X(t) = c \cdot p_D(t) + (1 - c) \cdot \underbrace{p_S(t)}_{=0} \quad \Rightarrow \quad p_D(t) = \frac{P(X = t)}{c}$$

a také

$$\sum_{t \in \mathbb{R}} P(X = t) = c \underbrace{\sum_{t \in \mathbb{R}} p_D(t)}_{=1} \quad \Rightarrow \quad c = \sum_{t \in \mathbb{R}} P(X = t)$$

a konečně také

$$f_X(t) = c \cdot \underbrace{f_D(t)}_{=0} + (1 - c) \cdot f_S(t) \quad \Rightarrow \quad f_S(t) = \begin{cases} \frac{F_X'(t)}{1-c} & \text{, pro taková } t \in \mathbb{R}, \text{ kde konečná derivace existuje} \\ 0 & \text{, jinak.} \end{cases}$$

5.3 (rozklad na směs)

Náhodná veličina X má distribuční funkci

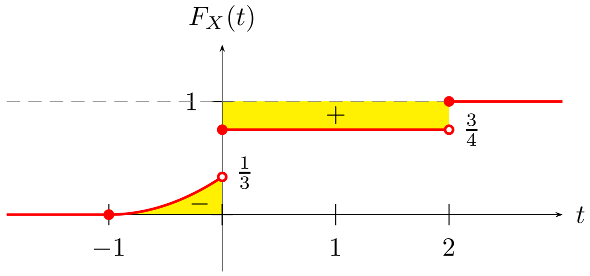
$$F_X(t) = \begin{cases} 0 & , t < -1 \\ \frac{1}{3}(t + 1)^2 & , t \in \langle -1, 0) \\ \frac{3}{4} & , t \in \langle 0, 2) \\ 1 & , t \ge 2\,. \end{cases}$$

(a) Určete $E(X)$.

(b) Najděte kvantil q_X.

(c) Vyjádřete X jako směs náhodných veličin D a S, z nichž D je diskrétní a S spojitá; popište a znázorněte jejich rozdělení.

Řešení:
(a) Znázorníme si graf F_X:



Střední hodnotu můžeme spočítat pomocí F_X jako

$$E(X) = -\int_{-\infty}^{0} F_X(t)\ \mathrm{d}t + \int_{0}^{\infty} (1 - F_X(t))\ \mathrm{d}t =$$

$$= -\int_{-1}^{0} \tfrac{1}{3}(t+1)^2\ \mathrm{d}t + \int_{0}^{2} \tfrac{1}{4}\ \mathrm{d}t = -\tfrac{1}{9}\left[(t+1)^3\right]_{-1}^{0} + \tfrac{1}{2} = \tfrac{7}{18}\ .$$

Geometrický význam střední hodnoty $E(X)$ je rozdíl velikosti plochy nad grafem F_X v intervalu $(0, +\infty)$ a plochy pod grafem funkce F_X v intervalu $(-\infty, 0)$ (viz žlutá plocha se znaménky v obrázku).

(b) Kvantilová funkce q_X pro veličinu X je v jistém smyslu "inverzí" k F_X. Je definována jako

$$q_X(\alpha) := \frac{1}{2}\Big(\sup\{t \in \mathbb{R} \mid F_X(t) \leq \alpha\} + \inf\{t \in \mathbb{R} \mid F_X(t) \geq \alpha\}\Big) \quad \text{pro} \quad \alpha \in (0,1).$$

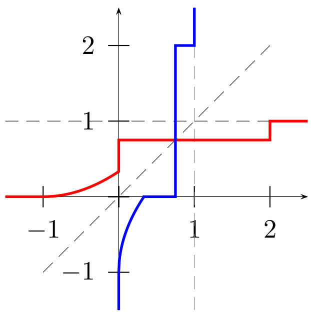
Tato definice vypadá poněkud složitě, ale je to jen kvůli případným úsekům, kde funkce F_X je konstantní. Vždy ale platí toto:

Věta: *Kvantilová funkce q_X je inverzní funkce k distribuční funkci F_X tam, kde F_X je na otevřeném intervalu $I \subseteq \mathbb{R}$ spojitá a ostře rostoucí.*

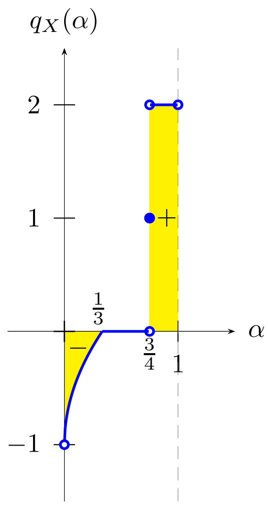
Obecněji pak platí následující: graf kvantilové funkce q_X získáme z grafu distribuční funkce F_X takto

- graf F_X doplníme na "souvislou čáru", tj. skoky funkce F_X nahradíme spojitou svislou úsečkou,
- tento útvar převrátíme podle osy 1. a 3. kvadrantu (tj. podle přímky "$x = y$"),
- tam, kde převrácený útvar není funkcí (tj. obsahuje svislé čáry) tyto úseky odstraníme a nahradíme jedinou hodnotou, a sice průměrem limit zprava a zleva (případné krajní úseky v bodech $\alpha = 0$ a $\alpha = 1$ odstraníme úplně, protože tam se kvantil nedefinuje).

V našem případě tedy graf F_X přejde na



a dostaneme graf q_X:



Kvantilovou funkci určíme také explicitně. Kvantil q_X je inverzní funkcí k ostře rostoucí funkci $F_X(t) = \frac{1}{3}(t+1)^2$ pro $t \in (-1, 0)$. Tedy

$$\alpha = \tfrac{1}{3}(t+1)^2 \quad \Leftrightarrow \quad t = -1 + \sqrt{3\alpha}$$

a tudíž

$$q_X(\alpha) = -1 + \sqrt{3\alpha} \quad \text{pro} \quad \alpha \in \left(0, \tfrac{1}{3}\right) \ .$$

Celkově:

$$q_X(\alpha) = \begin{cases} -1 + \sqrt{3\alpha} & , \ \alpha \in (0, \frac{1}{3}) \\ 0 & , \ \alpha \in \langle \frac{1}{3}, \frac{3}{4}) \\ 1 & , \ \alpha = \frac{3}{4} \\ 2 & , \ \alpha \in (\frac{3}{4}, 1). \end{cases}$$

Jestliže nyní budeme uvažovat Kolmogorův model na intervalu $\Omega = (0, 1)$ s rovnoměrným rozdělením pravděpodobnosti, můžeme kvantilovou funkci $q_X : \Omega \to \mathbb{R}$ chápat jako jeden ze způsobů, jak si představit veličinu X, tj. q_X je nyní náhodná veličina se stejnou distribuční funkci jako má X (skutečně je $F_{q_X} = F_X$) - tj. q_X je jeden z modelů veličiny X.

Tato představa má tu výhodu, že na $\Omega = (0, 1)$ můžeme "běžně"integrovat. Díky tomu, že interval $(0, 1)$ má délku jedna, bude střední hodnota z q_X jednoduše integrál z této funkce (viz žluté plochy se znaménky na obrázku). Střední hodnota $E(X)$ se tak dá spočítat také jednoduše jako $E(X) = \int_0^1 q_X(\alpha)\, d\alpha$ (viz žluté plochy se znaménky na obrázku).

(c) Připomeňme si, jak se hledá diskrétní a spojitá část veličiny $X : \Omega \to \mathbb{R}$. Diskrétní část D je zodpovědná za skoky v distribuční funkci F_X. Nejdříve si tedy určíme množinu všech bodů nespojitosti funkce F_X

$$A := \text{"}\textit{množina všech bodů nespojitosti funkce } F_X\text{"} = \{t \in \mathbb{R} \mid P(X = t) \neq 0\} = \{0, 2\}.$$

Diskrétní část D je zúžení X na množinu

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

tedy

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

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

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

Pak máme

$$X = \mathrm{Mix}_c(D, S),$$

kde

$$c = P(\Omega_1) = P(X \in A) = \sum_{t \in A} P(X = t) \ = \ \text{"}\textit{součet všech skoků funkce } F_X\text{"} \ =$$
$$= P(X = 0) + P(X = 2) \ .$$

Skoky dostaneme pomocí vztahu

$$P(X = 0) = P(X \leq 0) - P(X < 0) = F_X(0) - \lim_{t \to 0_-} F_X(t) = \tfrac{3}{4} - \tfrac{1}{3} = \tfrac{5}{12}$$

a podobně

$$P(X = 2) = 1 - \tfrac{3}{4} = \tfrac{1}{4}.$$

Takže

$$c = P(X = 0) + P(X = 2) = \tfrac{5}{12} + \tfrac{1}{4} = \tfrac{2}{3}.$$

Pro distribuční funkce máme vztah

$$F_X(t) = cF_D(t) + (1 - c)F_S(t)$$

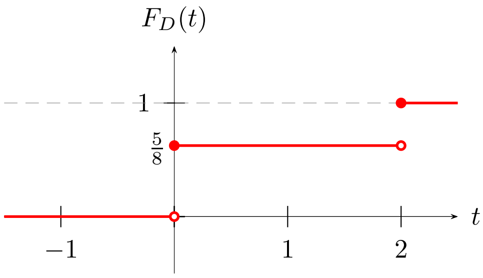
• **Popis diskrétní části D:**
Pro distribuční funkci F_D diskrétní veličiny D platí

$$cF_D(t) = \sum_{a \leq t} P(X = a) = \begin{cases} 0 & , \ t < 0 \\ \frac{5}{12} & , \ t \in \langle 0, 2) \\ \frac{5}{12} + \frac{1}{4} = \frac{2}{3} & , \ t \geq 2 \end{cases}$$

takže pro $\frac{1}{c} = \frac{3}{2}$ máme

$$F_D(t) = \begin{cases} 0 & , \ t < 0 \\ \frac{3}{2} \cdot \frac{5}{12} = \frac{5}{8} & , \ t \in \langle 0, 2) \\ \frac{3}{2} \cdot \frac{2}{3} = 1 & , \ t \geq 2 \end{cases}$$

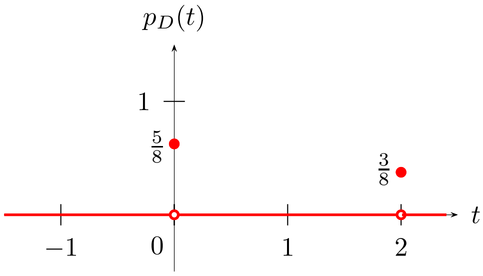
a graf F_D je:



Pro pravděpodobnostní funkci p_D diskrétní veličiny D platí

$$p_D(t) = \frac{1}{c} \cdot P(X = t) = \frac{3}{2} \cdot P(X = t) = \begin{cases} \frac{3}{2} \cdot \frac{5}{12} = \frac{5}{8} & , \ t = 0 \\ \frac{3}{2} \cdot \frac{1}{4} = \frac{3}{8} & , \ t = 2 \\ 0 & , \ \text{jinak} \end{cases}$$

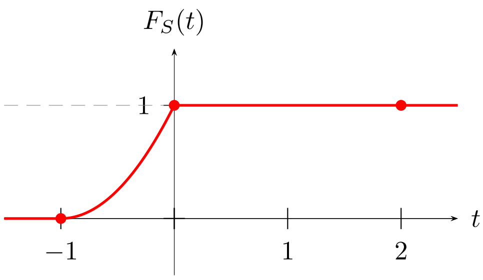
s grafem:



• **Popis spojité části S:**
Pro distribuční funkci spojité veličiny S pak ze vztahu $F_X = cF_D + (1-c)F_S$ dostáváme

$$F_S(t) = \frac{F_X(t) - cF_D(t)}{1-c} = 3\Big(F_X(t) - cF_D(t)\Big) = \begin{cases} & 0 & , \ t < -1 \\ 3 \cdot \frac{1}{3}(t+1)^2 = & (t+1)^2 & , \ t \in \langle -1, 0) \\ 3(\frac{3}{4} - \frac{5}{12}) = & 1 & , \ t \in \langle 0, 2) \\ 3(1 - \frac{2}{3}) = & 1 & , \ t \geq 2 \end{cases}$$

Graf funkce F_S tedy dostaneme jednoduše tak, že 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:

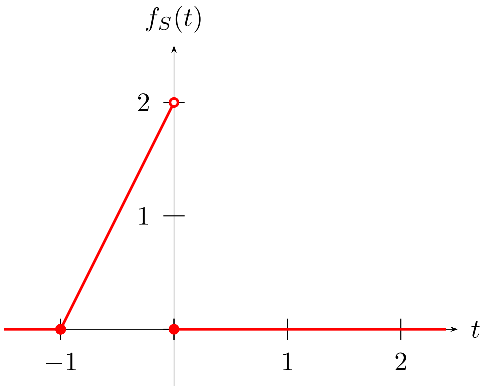


Hustotu f_S spojité veličiny S pak získáme derivací F_S pro body, kde derivace existuje. V ostatních (v tomto případě konečně mnoha bodech) na (nezáporných) hodnotách nezáleží. Takže můžeme psát toto:

$$f_S(t) = F_S'(t) = \left(\frac{F_X - cF_D}{1-c}\right)'(t) = \frac{1}{1-c} \cdot F_X'(t)$$

$$f_S(t) = \begin{cases} 2(t+1) & , \ t \in (-1, 0) \\ 0 & , \ \text{jinak.} \end{cases}$$

Graf f_S pak bude:



5.4 (rozklad na směs)
Náhodná veličina X má distribuční funkci

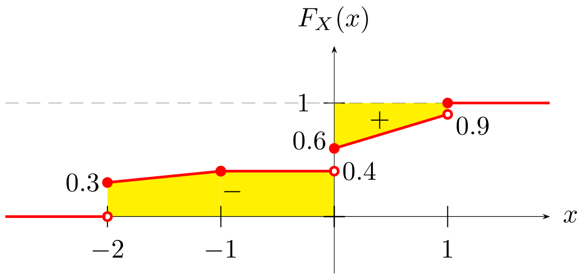
$$F_X(x) = \begin{cases} 0 & , \ x \in (-\infty, -2) \\ 0.1\,x + 0.5 & , \ x \in \langle -2, \ -1) \\ 0.4 & , \ x \in \langle -1, \ 0) \\ 0.3\,x + 0.6 & , \ x \in \langle 0, \ 1) \\ 1 & , \ x \in \langle 1, \infty) \end{cases}$$

(a) Určete $E(X)$.

(b) Najděte kvantilovou funkci q_X.

(c) Vyjádřete X jako směs náhodných veličin D a S, z nichž D je diskrétní a S spojitá; popište a znázorněte jejich rozdělení.

Řešení:
(a) Znázorníme si graf F_X:



Střední hodnotu můžeme spočítat pomocí F_X jako

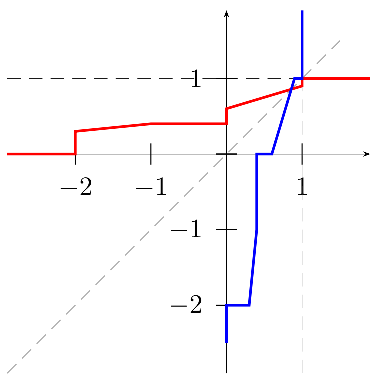
$$E(X) = -\int_{-\infty}^{0} F_X(x)\ \mathrm{d}x + \int_{0}^{\infty} (1 - F_X(x))\ \mathrm{d}x =$$

$$= -\int_{-2}^{-1} (0.1\,x + 0.5)\ \mathrm{d}x - \int_{-1}^{0} 0.4\ \mathrm{d}x + \int_{0}^{1} (-0.3\,x + 0.4)\ \mathrm{d}x =$$

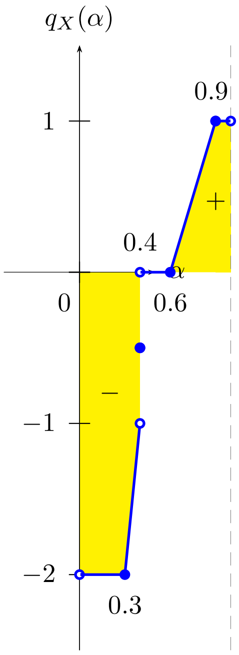
$$= \underbrace{-\left[\tfrac{0.1\,x^2}{2}\right]_{-2}^{-1}}_{0.15} - 0.5 - 0.4 + \underbrace{\left[-\tfrac{0.3\,x^2}{2}\right]_{0}^{1}}_{-0.15} + 0.4 = -0.5\ .$$

Geometrický význam střední hodnoty $E(X)$ je rozdíl velikosti plochy nad grafem F_X v intervalu $(0, +\infty)$ a plochy pod grafem funkce F_X v intervalu $(-\infty, 0)$ (viz žluté plochy se znaménky v obrázku).

(b) Kvantilová funkce q_X pro veličinu X je v jistém smyslu "inverzí" k F_X. V našem případě graf F_X přejde na



a dostaneme graf q_X:



Kvantilovou funkci určíme také explicitně. Kvantil q_X je inverzní funkcí k tam, kde F_X je ostře rostoucí funkce, tj. pak $q_X(\alpha) = x$ právě když $F_X(x) = \alpha$. Tedy pro $x \in (-2, -1)$ je $F_X(x) = 0.1\,x + 0.5$ a tudíž

$$\alpha = 0.1\,x + 0.5 \quad \Leftrightarrow \quad x = 10\alpha - 5$$

a tudíž

$$q_X(\alpha) = 10\alpha - 5 \quad \text{pro} \quad \alpha \in (0,\ 0.3)$$

Podobně pro $x \in (0, 1)$ je $F_X(x) = 0.3\,x + 0.6$. Tedy

$$\alpha = 0.3\,x + 0.6 \quad \Leftrightarrow \quad x = \frac{10}{3}\alpha - 2$$

a tudíž

$$q_X(\alpha) = \frac{10}{3}\alpha - 2 \quad \text{pro} \quad \alpha \in (0.6,\ 0.9) \ .$$

Celkově:

$$q_X(\alpha) = \begin{cases} -2 & ,\ \alpha \in (0,\ 0.3) \\ 10\alpha - 5 & ,\ \alpha \in \langle 0.3,\ 0.4) \\ -0.5 & ,\ \alpha = 0.4 \\ 0 & ,\ \alpha \in (0.4,\ 0.6) \\ \frac{10}{3}\alpha - 2 & ,\ \alpha \in \langle 0.6,\ 0.9) \\ 1 & ,\ \alpha \in \langle 0.9,\ 1). \end{cases}$$

(c) Připomeňme si, jak se hledá diskrétní a spojité části směsi $X = \mathrm{Mix}_c(D, S)$. Diskrétní část D je zodpovědná za skoky v distribuční funkci F_X. Pro koeficient c ve směsi platí

$$c = P(X \in \underbrace{\text{"}\textit{množina všech bodů nespojitosti funkce } F_X\text{"}}_{=\{-2,0,1\}}) =$$

$$= P\Big(X \in \{-2, 0, 1\}\Big) = P(X = -2) + P(X = 0) + P(X = 1)\ .$$

Hodnotu skoku dostaneme pomocí vztahu

$$P(X = -2) = P(X \leq -2) - P(X < -2) = F_X(-2) - \lim_{t \to -2_-} F_X(t) = 0.3 - 0 = 0.3$$

a podobně

$$P(X = 0) = 0.6 - 0.4 = 0.2 \ \text{ a } \ P(X = 1) = 1 - 0.9 = 0.1$$

Takže

$$c = P(X = -2) + P(X = 0) + P(X = 1) = 0.3 + 0.2 + 0.1 = \mathbf{0.6}.$$

Pro distribuční funkce máme vztah

$$F_X(t) = cF_D(t) + (1 - c)F_S(t)$$

- **Popis diskrétní části D:**

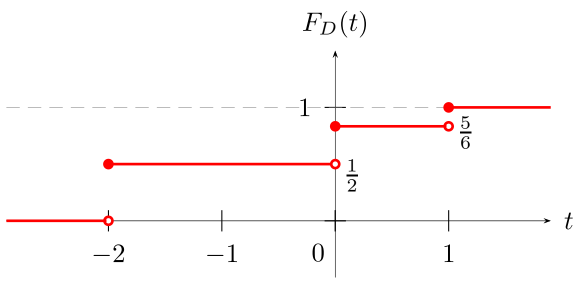
Pro distribuční funkci F_D diskrétní veličiny D platí

$$cF_D(t) = \sum_{a \leq t} P(X = a) = \begin{cases} 0 & , \ t < -2 \\ 0.3 & , \ t \in \langle -2, 0) \\ 0.3 + 0.2 = 0.5 & , \ t \in \langle 0, 1) \\ 0.3 + 0.2 + 0.1 = 0.6 & , \ t \geq 1 \end{cases}$$

takže pro $\frac{1}{c} = \frac{1}{0.6}$ máme

$$F_U(t) = \begin{cases} 0 & , \ t < -2 \\ \frac{1}{0.6} \cdot 0.3 = \frac{1}{2} & , \ t \in \langle -2, 0) \\ \frac{1}{0.6} \cdot 0.5 = \frac{5}{6} & , \ t \in \langle 0, 1) \\ \frac{1}{0.6} \cdot 0.6 = 1 & , \ t \geq 1 \end{cases}$$

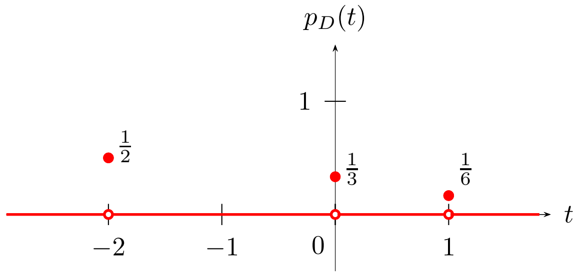
a graf F_D je:



Pro pravděpodobnostní funkci p_D diskrétní veličiny D platí

$$p_D(t) = \frac{1}{c} \cdot P(X = t) = \frac{1}{0.6} \cdot P(X = t) = \begin{cases} \frac{1}{0.6} \cdot 0.3 = \frac{1}{2} & , \ t = -2 \\ \frac{1}{0.6} \cdot 0.2 = \frac{1}{3} & , \ t = 0 \\ \frac{1}{0.6} \cdot 0.1 = \frac{1}{6} & , \ t = 1 \\ 0 & , \ \text{jinak} \end{cases}$$

s grafem:

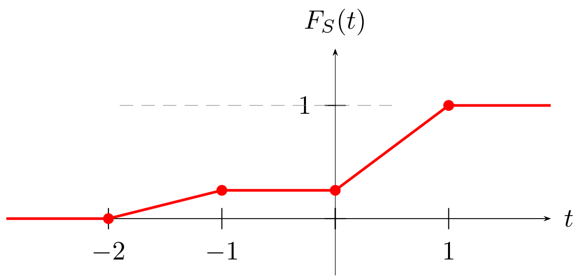


• **Popis spojité části S:**

Pro distribuční funkci spojité veličiny S pak ze vztahu $F_X = cF_D + (1-c)F_S$ dostáváme

$$F_S(t) = \frac{F_X(t) - cF_D(t)}{1-c} = \frac{1}{0.4}\Big(F_X(t) - cF_D(t)\Big) = \begin{cases} & 0 & , \ t < -1 \\ \frac{1}{0.4}\Big(0.1\ t + 0.5 - 0.3\Big) = & \frac{t+2}{4} & , \ t \in \langle -2, -1) \\ \frac{1}{0.4}\Big(0.4 - 0.3\Big) = & \frac{1}{4} & , \ t \in \langle -1, 0) \\ \frac{1}{0.4}(0.3\ t + 0.6 - 0.5) = & \frac{3t+1}{4} & , \ t \in \langle 0, 1) \\ \frac{1}{0.4}(1 - 0.6) = & 1 & , \ t \geq 1 \end{cases}$$

Graf funkce F_S tedy dostaneme jednoduše tak, že 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:

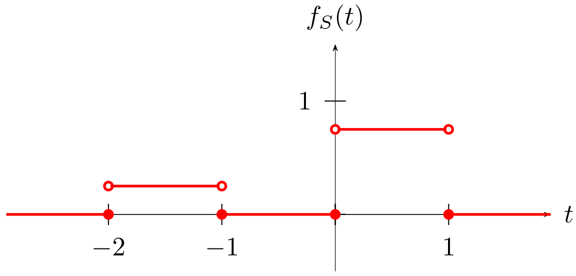


Hustotu f_S spojité veličiny S pak získáme derivací F_S pro body, kde derivace existuje. V ostatních (v tomto případě konečně mnoha bodech) na (nezáporných) hodnotách nezáleží. Takže můžeme psát toto:

$$f_S(t) = F'_S(t) = \left(\frac{F_X - cF_D}{1-c}\right)'(t) = \frac{1}{1-c} \cdot F'_X(t)$$

$$f_S(t) = \begin{cases} \frac{1}{4} & , \ t \in (-2, -1) \\ \frac{3}{4} & , \ t \in (0, 1) \\ 0 & , \ \text{jinak.} \end{cases}$$

Graf f_S pak bude:



$f_S(t)$
1
t
-2
-1
1