5. cvičení z PST

23. října 2019

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 (spojitá náhodná veličina)

Bod se pohybuje v rovině po kružnici (se středem v počátku a poloměrem 1) stálou úhlovou rychlostí. Náhodná veličina X je průmět bodu na osu x. Pravděpodobnost nalezení X v intervalu $\langle a, b\rangle \subseteq \langle -1, 1\rangle$ je dána jako

$$P\big(X \in \langle a, b\rangle\big) = \frac{\text{“}\textit{čas, co stráví průmět bodu v } \langle a, b\rangle\text{”}}{\text{“}\textit{čas, co stráví průmět bodu v } \langle -1, 1\rangle\text{”}}$$

Určete:

(a) distribuční funkci F_X a hustotu pravděpodobnosti f_X.

(b) pravděpodobnost $P(\frac{1}{2} \leq |X| < 3)$

(c) $\varepsilon > 0$ takové, že $P\big(X \in \langle -\varepsilon, \varepsilon\rangle\big) = \frac{1}{2}$.

(d) střední hodnotu $E(X)$ a rozptyl $D(X)$.

Řešení:

Pohyb bodu po dané kružnici je popsán v čase t jako $x(t) = \cos(\omega t)$ a $y(t) = \sin(\omega t)$, kde $\omega > 0$ je úhlová rychlost. Stačí uvažovat jen oběh bodu po půlkružnici a proto si pro t zvolíme časový interval např. $\langle 0, \frac{\pi}{\omega}\rangle$ (s rostoucím t nám pak souřadnice $x(t)$ bude klesat).

Pro interval $\langle a, b\rangle \subseteq \langle -1, 1\rangle$ mějme teď takové (jednoznačně určené) $t_1, t_2 \in \langle 0, \frac{\pi}{\omega}\rangle$, $t_2 \leq t_1$, že $a = \cos(\omega t_1)$ a $b = \cos(\omega t_2)$. Vzhledem k volbě časového intervalu máme

$$t_1 = \tfrac{\arccos a}{\omega}$$

$$t_2 = \tfrac{\arccos b}{\omega}$$

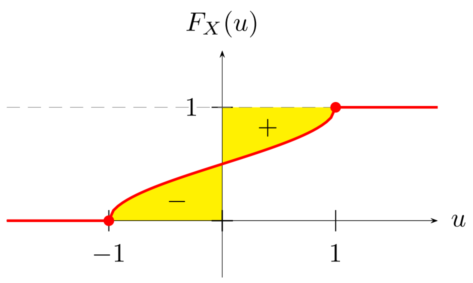
Speciálně pro $a = -1$ a $b = 1$ je $t_1 = \frac{\pi}{\omega}$ a $t_2 = 0$.

Pak máme

$$P\big(X \in \langle a, b\rangle\big) = \frac{\text{“}\textit{čas, co stráví průmět bodu v } \langle a, b\rangle\text{”}}{\text{“}\textit{čas, co stráví průmět bodu v } \langle -1, 1\rangle\text{”}} = \frac{t_1 - t_2}{\frac{\pi}{\omega} - 0} = \frac{\arccos(a) - \arccos(b)}{\pi}$$

Protože obor hodnot veličiny X je zřejmě $\langle -1, 1\rangle$, dostáváme distribuční funkci

$$F_X(u) = P(X \leq u) = \begin{cases} 0 & , u \leq -1 \\ P\big(X \in \langle -1, u\rangle\big) = \quad 1 - \frac{\arccos(u)}{\pi} & , u \in \langle -1, 1\rangle \\ 1 & , u \geq 1\ . \end{cases}$$



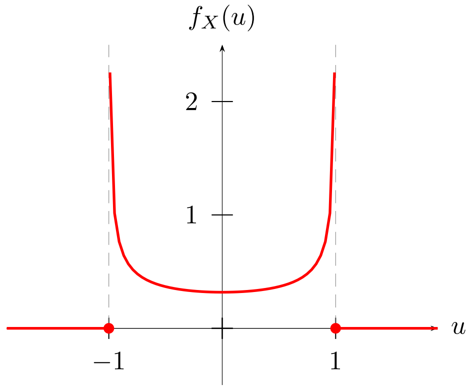
Hustotu f_X určíme z derivace F_X tam, kde tato derivace existuje (v jiných bodech si hodnotu f_X můžeme zvolit libovolně):

$$\frac{\mathrm{d}}{\mathrm{d}u} F_X(u) = \begin{cases} \frac{\mathrm{d}}{\mathrm{d}u}\left(1 - \frac{\arccos(u)}{\pi}\right) = \frac{1}{\pi} \cdot \frac{1}{\sqrt{1-u^2}} & , u \in (-1,1) \\ 0 & , u < -1 \text{ nebo } u > 1 \ . \end{cases}$$

Takže funkce

$$f_X(u) = \begin{cases} \frac{1}{\pi} \cdot \frac{1}{\sqrt{1-u^2}} & , u \in (-1,1) \\ 0 & , \text{jinak} \ . \end{cases}$$

je hustotou pravděpodobnosti veličiny X.



Jak je vidět, je f_X sudá funkce. Z toho např. vyplývá, že pro interval $\langle -b, -a\rangle \subseteq (-\infty, 0\rangle$ máme

$$P\big(X \in \langle -b, -a\rangle\big) = \int_{-b}^{-a} f_X(t)\ \mathrm{d}t = \left\{ \begin{smallmatrix} t=-u \\ dt=-du \\ u \in \langle a,b\rangle \end{smallmatrix} \right\} = \int_a^b f_X(u)\ \mathrm{d}u = P\big(X \in \langle a, b\rangle\big)$$

díky čemuž máme

$$P\left(\tfrac{1}{2} \le |X| < 3\right) = P\Big(X \in (-3, -\tfrac{1}{2}\rangle\Big) + P\Big(X \in \langle \tfrac{1}{2}, 3)\Big) = 2 \cdot P\left(\tfrac{1}{2} \le X < 3\right) =$$

$$= 2 \cdot \Big(P\left(X < 3\right) - P\left(X < \tfrac{1}{2}\right)\Big) = 2 \cdot \Big(F_X(3) - F_X\left(\tfrac{1}{2}\right)\Big) = 2 \cdot \Big(1 - 1 + \tfrac{1}{\pi} \cdot \arccos\left(\tfrac{1}{2}\right)\Big) = \tfrac{2}{\pi} \cdot \tfrac{\pi}{3} = \tfrac{2}{3} \ .$$

Podobně hledáme $0 < \varepsilon < 1$, aby

$$\tfrac{1}{2} = P\big(X \in \langle -\varepsilon, \varepsilon\rangle\big) = 2 \cdot P\big(0 \le X \le \varepsilon\big) = 2\big(F_X(\varepsilon) - F_X(0)\big) = \tfrac{2}{\pi}\big(\tfrac{\pi}{2} - \arccos(\varepsilon)\big)$$

tedy

$$\arccos(\varepsilon) = \tfrac{\pi}{4} \quad \Rightarrow \quad \varepsilon = \cos\tfrac{\pi}{4} = \tfrac{\sqrt{2}}{2} \ .$$

Střední hodnotu spočítáme pomocí hustoty pravděpodobnosti

$$E(X) = \underbrace{\int_{-\infty}^{\infty} t \cdot \underbrace{f_X(t)}_{\text{sudá funkce}}}_{\text{lichá funkce}} \, dt = \frac{1}{\pi} \int_{-1}^{1} \frac{t}{\sqrt{1-t^2}} \, dt = \frac{1}{\pi} \Big[-\sqrt{1-t^2} \Big]_{-1}^{1} = 0 \; .$$

Pro sudou hustotu tedy bude střední hodnota veličiny (pokud existuje!) vždy nulová. (POZOR: střední hodnota veličiny X obecně existovat nemusí!!). Existenci střední hodnoty (i bez počítání) v našem případě můžeme zjistit i z tvaru distribuční funkce F_X a sice pomocí tohoto vztahu:

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

jehož geometrická interpretace je, že střední hodnota je rovna součtu velikosti ploch s daným znaménkem v grafu F_X (viz žluté plochy na obrázku). Protože plochy jsou omezené, oba integrály existují, a protože obě plochy jsou stejně velké, je skutečně $E(X) = 0$.

Pro výpočet rozptylu $D(X)$ opět budeme potřebovat:

$$E(X^2) = \int_{-\infty}^{\infty} t^2 \cdot f_X(t) \, dt = \frac{1}{\pi} \int_{-1}^{1} \frac{t^2}{\sqrt{1-t^2}} \, dt = \left\{ \begin{smallmatrix} t=\cos(\alpha) \\ dt=-\sin(\alpha)d\alpha \\ \alpha\in\langle 0,\pi\rangle \end{smallmatrix} \right\} =$$

$$= \frac{1}{\pi} \int_{\pi}^{0} \frac{\cos^2(\alpha)}{\sin(\alpha)} \Big(-\sin(\alpha) \Big) \, d\alpha = \frac{1}{\pi} \int_{0}^{\pi} \cos^2(\alpha) \, d\alpha = \frac{1}{\pi} \cdot \frac{\pi}{2} = \frac{1}{2} \; .$$

Takže $D(X) = E(X^2) - \Big(E(X) \Big)^2 = \frac{1}{2}$.

Připomenutí: 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.)

5.2 (rozdělení veličiny vytvořené jako směs)

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

(a) distribuční funkci F_Z.

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

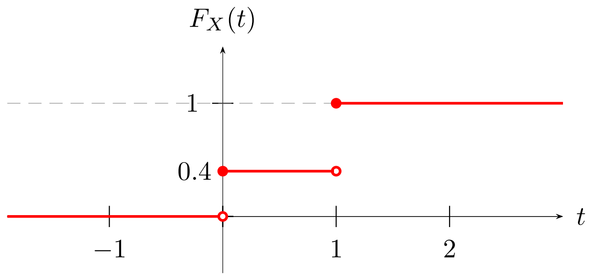
(c) střední hodnotu $E(Z)$.

(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

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

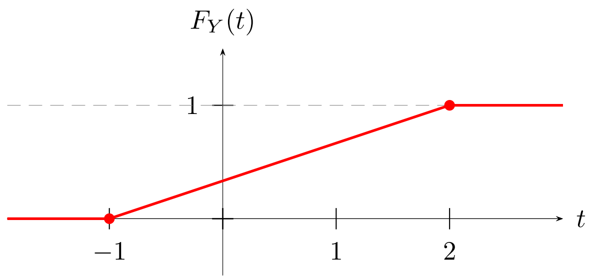


Veličina Y má spojité rovnoměrné rozdělení na $\langle -1, 2\rangle$ s hustotou

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

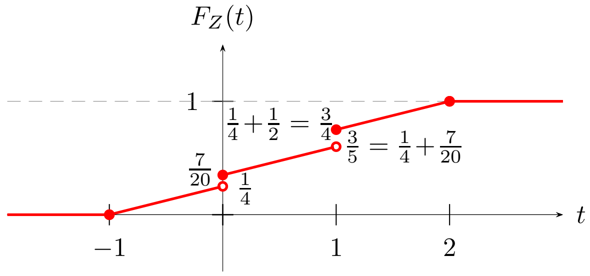
a distribuční funkcí

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



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

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



(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:

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

5.3 (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 $P(0 \le Z \le \frac{1}{2})$ a $P(Z > \frac{1}{3})$.

(c) střední hodnotu $E(Z)$.

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

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

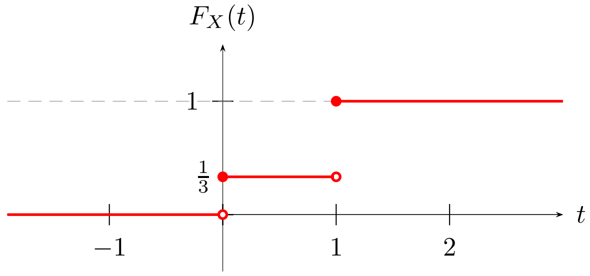
$$\Omega_2 = \text{“}\textit{výstupy, které mají hodnotu v } (0,1)\text{”}$$

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 $(0, 1)$).

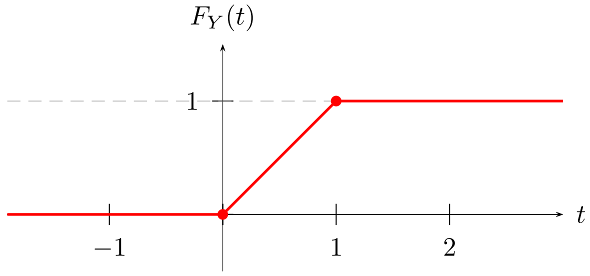
(a) Veličina X má alternativní rozdělení $\mathrm{Alt}(\frac{2}{3})$ s pravděpodobnostní funkcí $p_X(0) = \frac{1}{3}$, $p_X(1) = \frac{2}{3}$ a distribuční funkcí

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



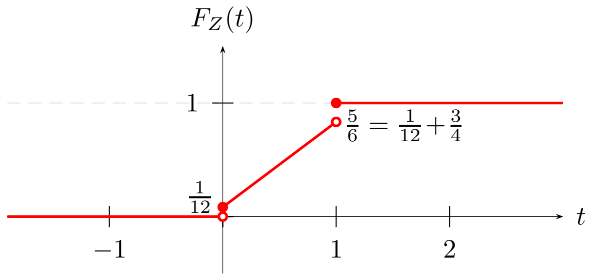
Veličina Y má spojité rozdělení s hustotou $f_Y(t) = \begin{cases} 1\,, & t \in (0,1)\,, \\ 0\,, & \text{jinak}\,, \end{cases}$ a distribuční funkcí

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



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

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



(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 $Z = \mathrm{Mix}_{1/4}(X, Y)$ 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:

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

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

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

5.4 (rozklad na směs)

Náhodná veličina X má spojité rovnoměrné rozdělení ve sjednocení intervalů $\langle -4, 0\rangle \cup \langle 2, 4\rangle$ (ve smyslu geometrické pravděpodobnosti) a náhodná veličina $Y = h(X)$ je definována pomocí "ořezávací" funkce

$$h(x) = \begin{cases} -1, & x \leq -1, \\ x, & -1 < x < 1, \\ 1, & x \geq 1. \end{cases}$$

(a) Určete F_Y a $E(Y)$.

(b) Najděte kvantilovou funkci q_Y.

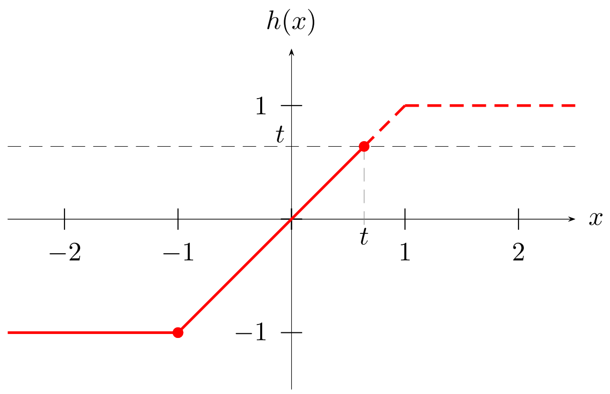
(c) Vyjádřete Y jako směs náhodných veličin D a S, z nichž D je diskrétní a S spojitá; popište a znázorněte jejich rozdělení.

Řešení:

(a) Pro distribuční funkci veličiny $Y = h(X)$ máme

$$F_Y(t) = P(Y \leq t) = P(h(X) \leq t) = P\Big(h(X) \in (-\infty, t\rangle\Big) = P\Big(X \in h^{-1}(-\infty, t\rangle\Big)$$

Podíváme se proto, jak vypadají množiny $h^{-1}(-\infty, t\rangle$. To snadno uvidíme z grafu funkce h:



Když "uřízneme" z grafu všechno, co je nad danou hladinou t a zbytek promítneme na vodorovnou osu, dostaneme právě hledanou množinu. Tedy:

$$h^{-1}(-\infty, t\rangle = \begin{cases} \emptyset & , \ t < -1 \\ (-\infty, t\rangle & , \ -1 \leq t < 1 \\ \mathbb{R} & , \ t \geq 1. \end{cases}$$

Takže můžeme psát:

$$F_Y(t) = P\Big(X \in h^{-1}(-\infty, t\rangle\Big) = \begin{cases} P(X \in \emptyset) = & 0 & , \ t < -1 \\ P\Big(X \in (-\infty, t\rangle\Big) = & F_X(t) & , \ -1 \leq t < 1 \\ P(X \in \mathbb{R}) = & 1 & , \ t \geq 1. \end{cases}$$

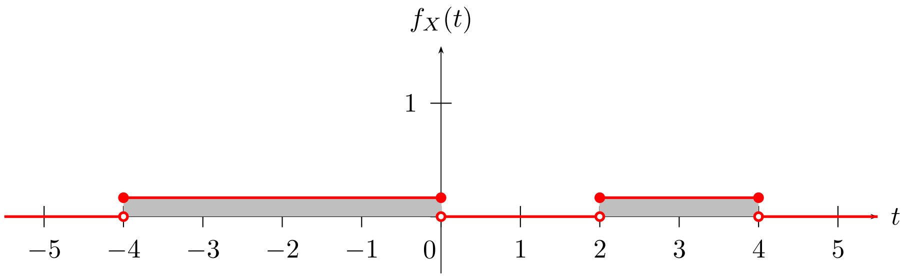
Teď tedy stačí jen určit funkci F_X. Hustota f_X je dána jako

$$f_X(t) = \begin{cases} c & , t \in \langle -4,\, 0\rangle \cup \langle 2, 4\rangle \\ 0 & , \text{jinak} \end{cases}$$

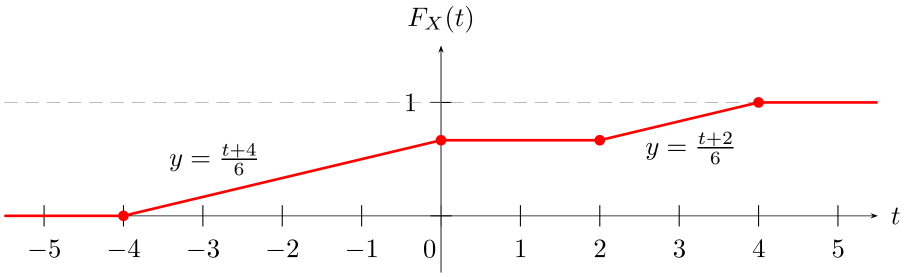
pro vhodné c, které určíme z podmínky, že (šedá) plocha pod hustotou je rovna 1:

$$1 = P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f_X(t)\ \mathrm{d}t = \int_{-4}^{0} c\ \mathrm{d}t + \int_{2}^{4} c\ \mathrm{d}t = 6c$$

tedy $c = \frac{1}{6}$.



Nyní už snadno napočítáme F_X pomocí integrování f_X (anebo ještě jednodušeji z obrázku hustoty: F_X poroste s lineárním přírůstkem o směrnici $\frac{1}{6}$ tam, kde je nenulová hustota a jinde F_X nebude rostoucí):

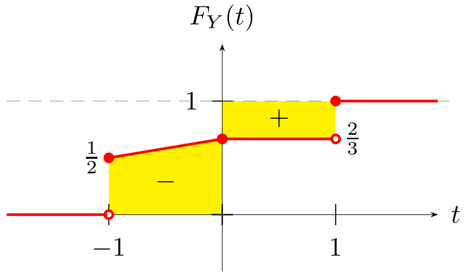


$$F_X(t) = \int_{-\infty}^{t} f_X(u)\ \mathrm{d}u = \begin{cases} \int_{-\infty}^{t} 0\ \mathrm{d}u = & 0 & , t < -4 \\ \int_{-4}^{t} \frac{1}{6}\ \mathrm{d}u = & \frac{t+4}{6} & , t \in \langle -4,\, 0\rangle \\ \frac{4}{6} + \int_{0}^{t} 0\ \mathrm{d}u = & \frac{2}{3} & , t \in \langle 0,\, 2\rangle \\ \frac{2}{3} + \int_{2}^{t} \frac{1}{6}\ \mathrm{d}u = & \frac{t+2}{6} & , t \in \langle 2, 4\rangle \\ & 1 & , t \geq 4 \end{cases}$$

Funkci F_Y teď získáme výše spočítaným oříznutím na interval $\langle -1, 1)$:

$$F_Y(t) = \begin{cases} 0 & , t < -1 \\ \frac{t+4}{6} & , t \in \langle -1,\, 0) \\ \frac{2}{3} & , t \in \langle 0,\, 1) \\ 1 & , t \geq 1 \end{cases}$$

a její graf je:



Střední hodnotu můžeme spočítat pomocí F_Y jako

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

$$= -\int_{-1}^{0} \tfrac{t+4}{6}\ \mathrm{d}t + \int_{0}^{1} (1 - \tfrac{2}{3})\ \mathrm{d}t = -\left[\tfrac{(t+4)^2}{12}\right]_{-1}^{0} + \tfrac{1}{3} = -\tfrac{1}{4}\ .$$

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. Je definována jako

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

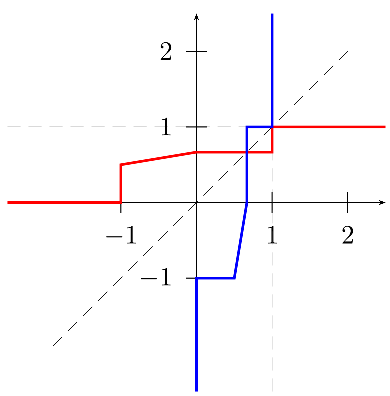
Tato definice vypadá poněkud složitě, ale je to jen kvůli případným bodům nespojitosti funkce F_X. Vždy ale platí toto:

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.*

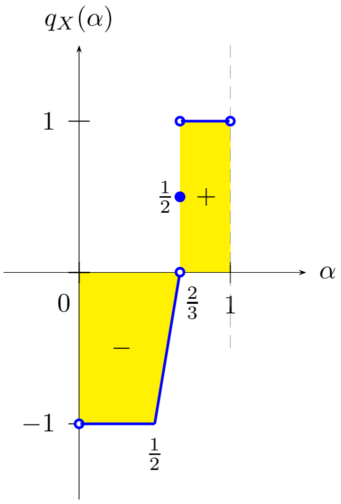
Obecněji pak platí následující: graf kvantilové funkce q_X získáme z grafu distribuční funkce F_X takto

- 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 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 $F_X(t) = \frac{t+4}{6}$ pro $t \in (-1, 0)$. Tedy

$$\alpha = \tfrac{t+4}{6} \quad \Leftrightarrow \quad t = 6\alpha - 4$$

a tudíž

$$q_Y(\alpha) = 6\alpha - 4 \quad \text{pro} \quad \alpha \in \left(\tfrac{1}{2}, \tfrac{2}{3}\right) \ .$$

Celkově:

$$q_X(\alpha) = \begin{cases} -1 & , \ \alpha \in (0, \frac{1}{2}\rangle \\ 6\alpha - 4 & , \ \alpha \in (\frac{1}{2}, \frac{2}{3}) \\ \frac{1}{2} & , \ \alpha = \frac{2}{3} \\ 1 & , \ \alpha \in (\frac{2}{3}, 1). \end{cases}$$

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 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

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

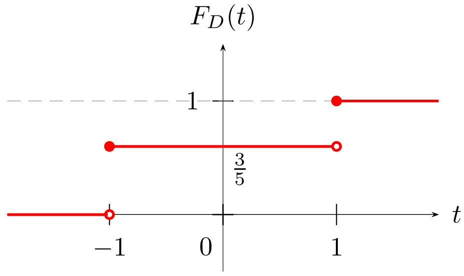
$$= P\Big(X \in \{-1, 1\}\Big) = P(X = -1) + P(X = 1) \ .$$

Hodnotu skoku dostaneme pomocí vztahu

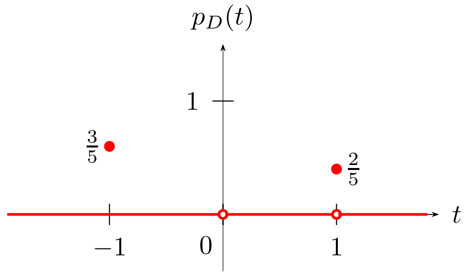
$$P(X = -1) = P(X \le -1) - P(X < -1) = F_X(-1) - \lim_{t \to -1_-} F_X(t) = \tfrac{1}{2} - 0 = \tfrac{1}{2}$$

a podobně
$$P(X=1) = 1 - \tfrac{2}{3} = \tfrac{1}{3}\ .$$
Takže
$$c = P(X=-1) + P(X=1) = \tfrac{1}{2} + \tfrac{1}{3} = \tfrac{5}{6}.$$
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
$$cF_D(t) = \sum_{a \le t} P(X=a) = \begin{cases} 0 & ,\ t < -1 \\ \frac{1}{2} & ,\ t \in \langle -1, 1) \\ \frac{1}{2} + \frac{1}{3} = \frac{5}{6} & ,\ t \ge 1 \end{cases}$$
takže pro $\frac{1}{c} = \frac{6}{5}$ máme
$$F_U(t) = \begin{cases} 0 & ,\ t < 0 \\ \frac{6}{5} \cdot \frac{1}{2} = \frac{3}{5} & ,\ t \in \langle -1, 1) \\ \frac{6}{5} \cdot \frac{5}{6} = 1 & ,\ t \ge 1 \end{cases}$$
a graf F_D je:



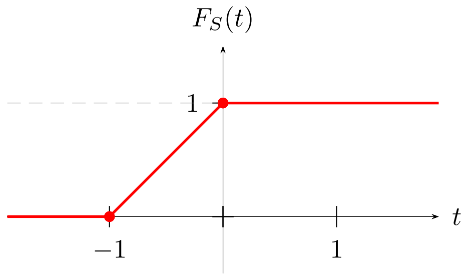
Pro pravděpodobnostní funkci p_D diskrétní veličiny D platí
$$p_D(t) = \tfrac{1}{c} \cdot P(X=t) = \tfrac{6}{5} \cdot P(X=t) = \begin{cases} \frac{6}{5} \cdot \frac{1}{2} = \frac{3}{5} & ,\ t = -1 \\ \frac{6}{5} \cdot \frac{1}{3} = \frac{2}{5} & ,\ t = 1 \\ 0 & ,\ \text{jinak} \end{cases}$$
s grafem:



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

$$F_S(t) = \frac{F_X(t) - cF_D(t)}{1-c} = 6\Big(F_X(t) - cF_D(t)\Big) = \begin{cases} & 0 & , t < -1 \\ 6\left(\frac{t+4}{6} - \frac{1}{2}\right) = & t+1 & , t \in \langle -1, 0) \\ 6(\frac{2}{3} - \frac{1}{2}) = & 1 & , t \in \langle 0, 1) \\ 6(1 - \frac{5}{6}) = & 1 & , t \geq 1 \end{cases}$$

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:

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

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

Graf f_V pak bude: