# 5. cvičení z PST

1. listopadu 2017

**5.1** (diskrétní veličina - geometrické rozdělení)

Sdělovací kanál přenáší symboly a při přenosu je pravděpodobnost výskytu chyby $p \in (0,1)$. Náhodná veličina $X$ je rovna délce úseků bezchybně přenesených symbolů mezi jednotlivými chybami, tj. jakmile se vyskytne chyba, začínáme počítat, kolik se za ní vyskytne bezchybně přenesených symbolů než nastane další chyba (a to bude hodnota $X$).

(a) Určete rozdělení náhodné veličiny $X$ a vypočtěte její střední hodnotu $E(X)$.

(b) Pro $p = 10^{-3}$ vypočtěte minimální $n$ takové, že $P(X \leq n) \geq 0.9$.

**Řešení:**

Příklad 5: `https://math.feld.cvut.cz/prucha/psi/wcv3.pdf`

**5.2** (diskrétní veličina - hypergeometrické rozdělení)

V osudí máme 3 bílé, 4 modré a 2 červené koule. Náhodně vytáhneme 4 z nich. Náhodná veličina $X$ je určena počtem bílých koulí.

(a) Určete pravděpodobnostní funkci náhodné veličiny $X$.

(b) Vypočtěte její střední hodnotu $E(X)$.

**Řešení:**

Příklad 2: `https://math.feld.cvut.cz/prucha/psi/wcv3.pdf`

**Poznámka:** Náhodná veličiny $X$ v tomto příkladu má tzv. *hypergeometrické* rozdělení $\mathrm{Hyp}(K, M, m)$, kde $M$ je počet prvků dané množiny (např. výrobků), která obsahuje $K$ prvků dané vlastnosti (např. vadné výrobky), a z níž vybíráme $m$ prvků (kde $m \leq M$). Zajímá nás, jaká je pravděpodobnost, že z vybraných $m$ prvků je právě $k$ prvků uvedené vlastnosti.

Pravděpodobnost je opět dána podílem příznivých možností ku všem. Příznivé jsou dány počtem způsobů jak vybrat $k$ výrobků z $K$ vadných násobeno počtem způsobů jak vybrat zbytek, tj. $m-k$ výrobků z $M-K$ bezvadných. Celkem tedy

$$p_{\mathrm{Hyp}(K,M,m)}(k) = \frac{\binom{K}{k} \cdot \binom{M-K}{m-k}}{\binom{M}{m}}$$

přičemž obor hodnot pro $k$ je

$$\max\{0, m + K - M\} \leq k \leq \min\{m, K\} \ .$$

Tuto podmínku dostaneme ihned z nerovností

$$k \leq K, \quad m - k \leq M - K,$$

$$k \leq m, \quad m - k \leq m \ .$$

Současně ale používáme dohodu, že pro $i, j \in \mathbb{Z}$ definujeme kombinační čísla jako

$$\binom{i}{j} = \begin{cases} \frac{i(i-1)\cdots(i-j+1)}{j!} & \text{pro } 0 \leq j \\ 0 & \text{pro } 0 > j \end{cases}$$

takže pro rozdělení $\mathrm{Hyp}(K, M, m)$ můžeme uvažovat pro jednoduchost obor hodnot všechna celá čísla $\mathbb{Z}$. Speciálně tedy platí, že

$$\sum_{k \in \mathbb{Z}} p_{\mathrm{Hyp}(K,M,m)}(k) = 1 \ .$$

Díky tomuto si můžeme ještě snadno spočítat střední hodnotu veličiny $X \sim \mathrm{Hyp}(K, M, m)$:

$$\color{red}{E(X) =} \sum_{k\in\mathbb{Z}} k \cdot p_{\mathrm{Hyp}(K,M,m)}(k) = \sum_{k\in\mathbb{Z}} k \cdot \frac{\binom{K}{k}\cdot\binom{M-K}{m-k}}{\binom{M}{m}} = \{\text{zdůvodnění níže}\} = \sum_{k\in\mathbb{Z}} K \cdot \frac{\binom{K-1}{k-1}\cdot\binom{M-K}{m-k}}{\binom{M}{m}} =$$

$$= \frac{K\cdot\binom{M-1}{m-1}}{\binom{M}{m}} \sum_{k\in\mathbb{Z}} \underbrace{\frac{\binom{K-1}{k-1}\cdot\binom{(M-1)-(K-1)}{(m-1)-(k-1)}}{\binom{M-1}{m-1}}}_{p_{\mathrm{Hyp}(K-1,M-1,m-1)}(k-1)} = \frac{K\cdot\binom{M-1}{m-1}}{\binom{M}{m}} \underbrace{\sum_{k\in\mathbb{Z}} p_{\mathrm{Hyp}(K-1,M-1,m-1)}(k-1)}_{=1} = K \cdot \frac{\binom{M-1}{m-1}}{\binom{M}{m}} =$$

$$= K \cdot \frac{(M-1)\cdots(M-m+1)}{(m-1)!} \cdot \frac{m!}{M(M-1)\cdots(M-m+1)} = \color{red}{m \cdot \frac{K}{M}}$$

Použili jsme přitom, že pro $k \geq 1$ je

$$k \cdot \binom{K}{k} = k \cdot \frac{K(K-1)\cdots(K-k+1)}{k!} = K \cdot \frac{(K-1)\cdots(K-k+1)}{(k-1)!} = K \cdot \binom{K-1}{k-1}$$

přičemž rovnost $k\cdot\binom{K}{k} = K\cdot\binom{K-1}{k-1}$ zřejmě platí pro všechna $k \in \mathbb{Z}$ podle naší úmluvy.

Výsledný tvar střední hodnoty je tedy velmi jednoduchý

$$E(X) = m \cdot \frac{K}{M}$$

a intuitivně odpovídá i následující limitní situaci $M \to \infty$, kdy $k$ a $m$ považujeme za pevné a předpokládáme, že $K/M \to q$ (tj. poměr počtu vadných výrobků $K$ ku všem výrobkům $M$ se blíží ke konstantní hodnotě $q \in (0,1)$):

$$\lim_{M\to\infty} p_{\mathrm{Hyp}(K,M,m)}(k) = \lim_{M\to\infty} \frac{m!}{k!(m-k)!} \frac{\prod_{i=0}^{k-1}(K-i)\cdot\prod_{j=0}^{m-k-1}(M-K-j)}{\prod_{i=0}^{m-1}(M-i)} =$$

$$= \binom{m}{k} \cdot \lim_{M\to\infty} \prod_{i=0}^{k-1} \frac{K-i}{M-i} \cdot \prod_{j=0}^{m-k-1} \frac{M-K-j}{M-k-j} = \binom{m}{k} q^k (1-q)^{m-k} = p_{\mathrm{Bi}(m,q)}(k)\ .$$

Dostáváme tedy známe binomické rozdělení $\mathrm{Bi}(m,q)$, které v tomto případě odpovídá limitní situaci, kdy taháme výrobky z nekonečného množství s určeným podílem $q$ vadných. Při $m$ pokusech jsme předpokládali, ze výrobky NEVRACÍME (anebo je prostě vytáhneme všechny naraz). Jak ale vidíme, i přesto jsme dostali rozdělení pravděpodobnosti, které používáme, když opakovaně uskutečňujeme tentýž pokus vždy za STEJNÝCH podmínek (např. opakovaně vytahujeme z osudí koule a VRACÍME je vždy zpátky). To je tím, že v obrovském množství výrobků $M$ se už v limite ztratí to, jestli tam těch pár výrobků $m$ vrátíme nebo ne (protože $m << M$).

Kromě toho vidíme, že limita platí i pro střední hodnoty, tj.

$$E\Big(\mathrm{Bi}(m,q)\Big) = mq = \lim_{M\to\infty} m \cdot \frac{K}{M} = \lim_{M\to\infty} E\Big(\mathrm{Hyp}(K,M,m)\Big)\ .$$

Celkově tedy hypergeometrické rozdělení přechází TÍMTO ZPŮSOBEM limitně v binomické.

**5.3** (smíšené rozdělení)

Náhodná veličina $X$ má rozdělení určené distribuční funkcí $F_X$ kde

$$F_X(t) = \begin{cases} 0 & \text{pro } \ t < 0, \\ t + \frac{1}{2} & \text{pro } \ 0 \leq t < \frac{1}{2}, \\ 1 & \text{pro } \ t \geq \frac{1}{2}. \end{cases}$$

(a) Nakreslete graf funkce $F_X$ a určete obor hodnot náhodné veličiny $X$.

(b) Vypočtěte pravděpodobnosti $P(0 < X < 1/3)$ a $P(0 \leq X \leq 1/3)$.

(c) Vypočtěte střední hodnotu $E(X)$, rozptyl $D(X)$ a směrodatnou odchylku $\sigma(X)$.

(d) Vyjádřete náhodnou veličinu $X$ jako směs spojité a diskrétní náhodné veličiny.

(e) Vypočtěte střední hodnotu $E(X)$ ze směsi náhodných veličin a kvantilové funkce $q_X$.

**Řešení:**

Příklad 6: `https://math.feld.cvut.cz/prucha/psi/wcv3.pdf`

**Poznámka:** Střední hodnotu $E(X)$ a rozptyl $D(X)$ také můžeme spočítat podle vzorce

$$E(X^n) = -\int_{-\infty}^{0} nt^{n-1} \cdot F_X(t)\ \mathrm{d}t + \int_{0}^{\infty} nt^{n-1} \cdot \big(1 - F_X(t)\big)\ \mathrm{d}t$$
$$\left( = \int_0^1 \Big(q_X(\alpha)\Big)^n\ \mathrm{d}\alpha \right).$$

Konkrétně

$$E(X) = -\int_{-\infty}^{0} F_X(t)\ \mathrm{d}t + \int_0^{\infty} 1 - F_X(t)\ \mathrm{d}t = \int_0^{\frac{1}{2}} \left(\tfrac{1}{2} - t\right)\ \mathrm{d}t = \tfrac{1}{4} - \tfrac{1}{8} = \tfrac{1}{8}$$

a

$$E(X^2) = -\int_{-\infty}^{0} 2t \cdot F_X(t)\ \mathrm{d}t + \int_0^{\infty} 2t \cdot \big(1 - F_X(t)\big)\ \mathrm{d}t = \int_0^{\frac{1}{2}} t - 2t^2\ \mathrm{d}t = \tfrac{1}{8} - \tfrac{1}{12} = \tfrac{1}{24}$$

tedy

$$D(X) = E(X^2) - \Big(E(X)\Big)^2 = \tfrac{1}{24} - \left(\tfrac{1}{8}\right)^2 = \tfrac{5}{192} \doteq 0.02604\ .$$

**5.4** (Poissonovo a exponenciální rozdělení)

Do pojišťovny přijdou průměrně 2 hlášení škody denně. Jaká je pravděpodobnost, že do pojišťovny přijde nejbližší hlášení škody nejdříve třetí den?

**Řešení:**

Tuto úlohu můžeme řešit jak s využitím exponenciálního rozdělení, tak Poissonova rozdělení.

Exponenciální rozdělení popisuje pravděpodobnost veličiny

$X =$ “*doba mezi dvěma následnými výskyty události*”,

pokud události nemají paměť. Tedy to, co se stane od určitého okamžiku nezávisí na tom, co bylo předtím. V praxi jde např. o to, že zařízení, které se ”neopotřebovává”(např. polovodičové součástky), bude mít poruchu nebo o dobu radioaktivního rozpadu atd. Exponenciální rozdělení je jediné, které splňuje následující rovnici:

$$P(X > x + y | X > y) = P(X > x)$$

pro všechna $x, y > 0$. Rovnice vyjadřuje to, že pravděpodobnost, že zařízení bude bez poruchy pracovat $x$ hodin je stejná v případě, že jsme jej právě zapnuli jako za předpokladu, že předtím už bez poruchy pracovalo $y$ hodin.

Exponenciální rozdělení je charakterizováno parametrem $\tau > 0$ (s fyzikálním rozměrem času), který představuje střední dobou čekání, tedy $E(X) = \tau$ a dále ještě platí $D(X) = \tau^2$. Hustota je pak dána jako

$$f_X(t) = \begin{cases} 0 & ,\ t \le 0 \\ \frac{1}{\tau} e^{-\frac{t}{\tau}} & ,\ t > 0\ . \end{cases}$$

a distribucni funkce je

$$F_X(t) = \begin{cases} 0 & ,\ t \le 0 \\ 1 - e^{-\frac{t}{\tau}} & ,\ t > 0\ . \end{cases}$$

Diskrétní analogií exponenciálního rozdělení je *geometrické* rozdělení, které neměří čas spojitě ale pouze diskrétně. Lze ho také chápat jako:

$$X = \text{"}\textit{počet neúspěšných pokusů než nastane první úspěch}\text{"},$$

např. házení míče na koš atd. Je to opět jediné takové diskrétní rozdělení splňující analogickou rovnici:

$$P(X > k + n | X > n) = P(X > k)$$

pro všechna $k, n \in \mathbb{N}_0$ s podobným významem jako u exponenciálního rozdělení.

**Věta:** Nechť

$X =$ *"doba čekání na událost"*

je veličina s exponenciálním rozdělením $\text{Exp}(\tau)$ se střední hodnotou $E(X) = \tau$ (s jednotkou "čas").

Pak veličina

$Y =$ *"počet událostí během doby $T$"*

má Poissonovo rozdělení $\text{Poiss}(\lambda)$ se střední hodnotou $E(Y) = \lambda$ (s bezrozměrnou jednotkou) a platí $\frac{T}{\lambda} = \tau$.

V našem případě máme

$X=$ *"doba čekání na příchod dalšího hlášení"*

$Y=$ *"počet hlášení během 1 dne"*

$Y'=$ *"počet hlášení během 2 dnů"*

(a) *Pomocí exponenciálního:* Podle zadání má veličina $Y$ Poissonovo rozdělení s parametrem $\lambda = 2$. Tudíž náhodná veličina $X$ s exponenciálním rozdělením ma parametr $\tau = \frac{T}{\lambda} = \frac{1}{2}$, kde $T = 1$. Hledaná pravděpodobnost je pak

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) = 1 - \left(1 - e^{-\frac{2}{\tau}}\right) = e^{-4}$$

nebo také počítáno jako

$$P(X > 2) = \int_2^\infty f(x)dx = \int_2^\infty \tfrac{1}{\tau} \cdot e^{-\frac{x}{\tau}} dx = \int_2^\infty 2e^{-2x} dx = e^{-4}.$$

(b) *Pomocí Poissonova:* Hledaná pravděpodobnost je dána jako

$$P(Y' = 0) = \frac{(\lambda')^0}{0!} \cdot e^{-\lambda'}$$

kde $\lambda'$ je parametr veličiny $Y'$ s Poissonovým rozdělením. K určení parametru $\lambda'$ můžeme použít

- buď vztah mezi veličinami $X$ a $Y'$ a už známou hodnotu parametru veličiny $X$, tj. platí $\lambda' = \frac{T'}{\tau}$, kde $T' = 2$ a $\tau = \frac{1}{2}$ a tudíž je $\lambda' = 4$,
- nebo to, ze platí $Y' = Z_1 + Z_2$, kde veličiny

  $Z_1=$ *"počet hlášení během prvního dne"*

  $Z_2=$ *"počet hlášení během druhého dne"*

  mají Poissonovo rozdělení s parametry $\lambda_1 = 2 = \lambda_2$ a tudiz

$$\lambda' = E(Y') = E(Z_1) + E(Z_2) = \lambda_1 + \lambda_2 = 2 + 2 = 4\ .$$

Hledaná pravděpodobnost je pak

$$P(Y = 0) = \frac{4^0}{0!} \cdot e^{-4} = e^{-4}.$$

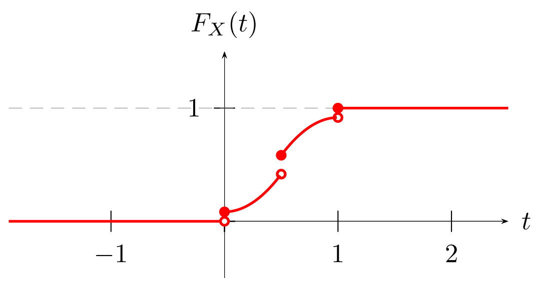
**5.5** (rozklad na směs)

Náhodná veličina $X$ má distribuční funkci

$$F_X(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1}{12} + \frac{4t^2}{3}\,, & 0 \le t < 1/2\,, \\ \frac{11}{12} - \frac{4(1-t)^2}{3}\,, & 1/2 \le t < 1\,, \\ 1\,, & t \ge 1\,. \end{cases}$$

Vyjádřete ji jako směs náhodných veličin $U, V$, z nichž $U$ je diskrétní a $V$ spojitá; popište a znázorněte jejich rozdělení.

**Řešení:**




Nespojitosti distribuční funkce jsou v bodech $0, 1/2, 1$,

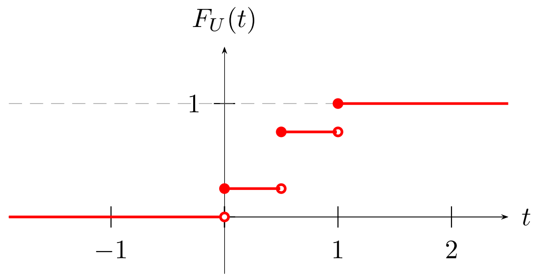
$$F_X(0) - F_X(0-) = F_X(1) - F_X(1-) = \frac{1}{12}\,,$$

$$F_X(1/2) - F_X(1/2-) = \frac{1}{6}\,,$$

$$c\,F_U(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1}{12}\,, & 0 \le t < \frac{1}{2}\,, \\ \frac{1}{4}\,, & \frac{1}{2} \le t < 1\,, \\ \frac{1}{3}\,, & t \ge 1\,, \end{cases}$$

$$c = \lim_{t\to\infty} F_U(t) = \frac{1}{3}\,,$$

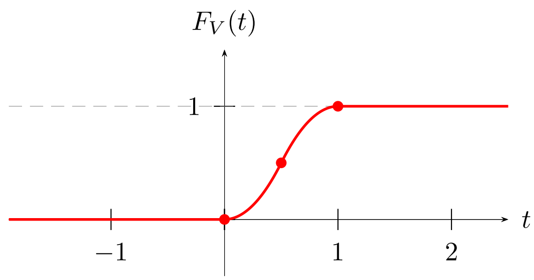
$$F_U(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1}{4}\,, & 0 \le t < \frac{1}{2}\,, \\ \frac{3}{4}\,, & \frac{1}{2} \le t < 1\,, \\ 1\,, & t \ge 1\,, \end{cases}$$




$$p_U(t) = \begin{cases} \frac{1}{4}\,, & t \in \{0,1\}\,, \\ \frac{1}{2}\,, & t = \frac{1}{2}\,, \\ 0\,, & \text{jinde}\,. \end{cases}$$

$$(1-c)F_V(t) = F_X(t) - cF_U(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{4\,t^2}{3}\,, & 0 \le t < 1/2\,, \\ \frac{2}{3} - \frac{4\,(1-t)^2}{3}\,, & 1/2 \le t < 1\,, \\ \frac{2}{3}\,, & t \ge 1\,, \end{cases}$$

$$F_V(t) = \begin{cases} 0\,, & t < 0\,, \\ 2t^2\,, & 0 \le t < 1/2\,, \\ 1 - 2\,(1-t)^2\,, & 1/2 \le t < 1\,, \\ 1\,, & t \ge 1\,, \end{cases}$$



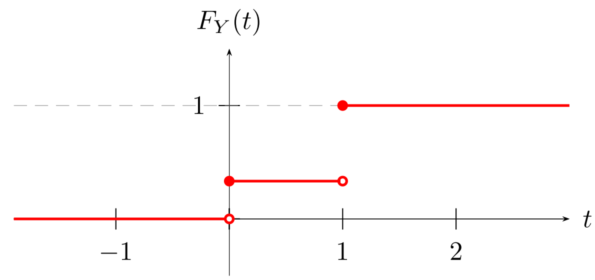

$$f_V(t) = \begin{cases} 4\,t\,, & 0 \le t < 1/2\,, \\ 4\,(1-t)\,, & 1/2 \le t < 1\,, \\ 0\,, & \text{jinak}\,. \end{cases}$$

**5.6** (rozdělení veličiny vytvořené jako směs)

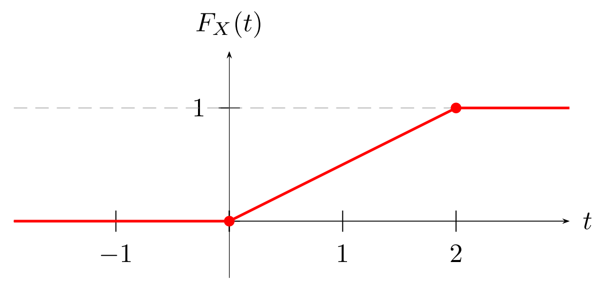
Náhodná veličina $X$ má alternativní rozdělení (s hodnotami $0,1$), $P(X = 1) = 2/3$. Náhodná veličina $Y$ má spojité rovnoměrné rozdělení na intervalu $\langle 0, 2\rangle$. Popište rozdělení jejich směsi $Z = \mathrm{Mix}_{2/3}(X,Y)$.

**Řešení:**

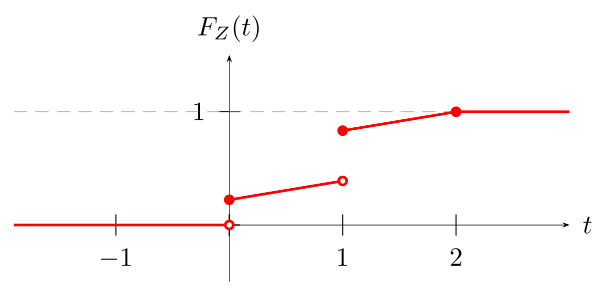
$$F_X(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1}{3}\,, & 0 \leq t < 1\,, \\ 1\,, & t \geq 1\,, \end{cases}$$




$$F_Y(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{t}{2}\,, & 0 \leq t < 2\,, \\ 1\,, & t \geq 2\,, \end{cases}$$




$$F_Z(t) = \frac{2}{3}\,F_X(t) + \frac{1}{3}\,F_Y(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{2}{9} + \frac{t}{6}\,, & 0 \leq t < 1\,, \\ \frac{2}{3} + \frac{t}{6}\,, & 1 \leq t < 2\,, \\ 1\,, & \text{pro } t \geq 2\,. \end{cases}$$



**5.7** (rozdělení veličiny vytvořené jako směs)

Diskrétní náhodná veličina $X$ má rovnoměrné rozdělení na množině $\{0,1\}$ a spojitá náhodná veličina $Y$ má rovnoměrné rozdělení na intervalu $[0,1]$. Veličiny $X$ a $Y$ jsou nezávislé. Určete rozdělení veličiny $W = X + Y$ a veličiny $Z = XY$.

**Řešení:**
Protože $X$ nabývá pouze konečně mnoha hodnot, bude výhodné použít větu o úplné pravděpodobnosti:

$$F_W(t) = P(X+Y \leq t) = P(X+Y \leq t|X=0)\cdot P(X=0) + P(X+Y \leq t|X=1)\cdot P(X=1)$$

Díky nezávislosti veličin $X$ a $Y$ máme

$$P(X+Y \leq t|X=1) = P(1+Y \leq t|X=1) = P(Y \leq t-1) = F_Y(t-1)$$

a podobně

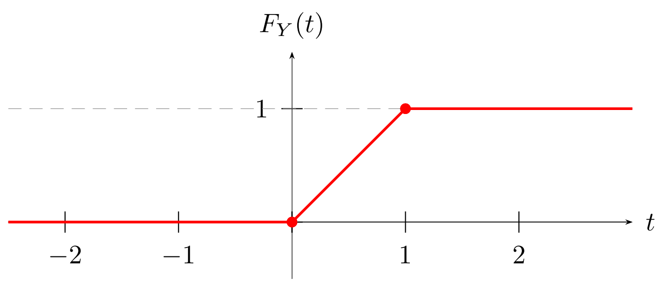
$$P(X+Y \leq t|X=0) = P(Y \leq t) = F_Y(t).$$

Protože ještě víme, že $P(X=0) = \frac{1}{2} = P(X=1)$, tak můžeme psát

$$F_W(t) = \frac{1}{2}\Big(F_Y(t) + F_Y(t-1)\Big)\ .$$

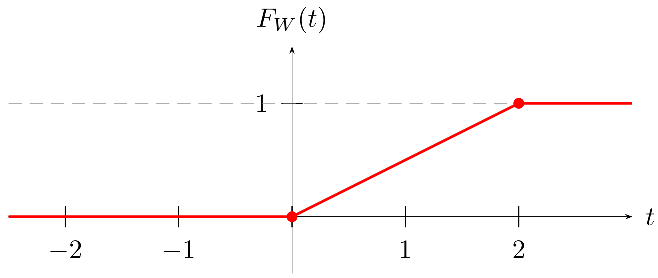
Nyní jen vyjádříme distribuční funkci veličiny $Y$

$$F_Y(t) = \begin{cases} 0 & ,\ t < 0 \\ t & ,\ t \in \langle 0, 1\rangle \\ 1 & ,\ t > 1\ . \end{cases}$$



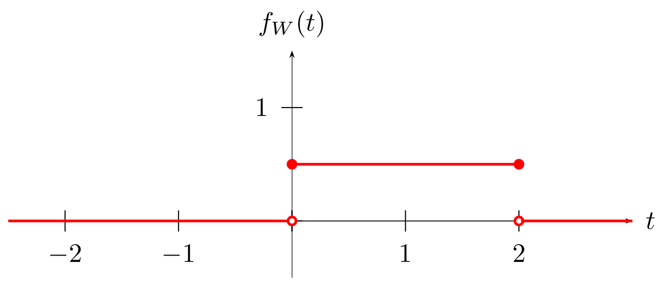

a dosadíme spolu s jejím posunutím do vzorce pro $F_W$, čímž získáme:

$$F_W(t) = \begin{cases} 0 & ,\ t < 0 \\ \frac{t}{2} & ,\ t \in \langle 0, 2\rangle \\ 1 & ,\ t > 2\ . \end{cases}$$




Veličina $W$ je tedy spojitá s rovnoměrným rozdělením na intervalu $\langle 0, 2\rangle$ a hustotou:

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

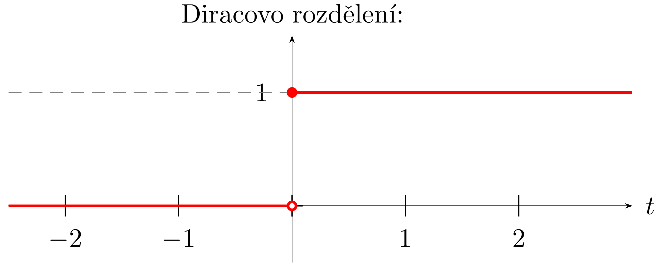



Pro veličinu $Z = XY$ budeme postupovat obdobně

$$F_Z(t) = P(XY \leq t) = P(XY \leq t|X = 0) \cdot P(X = 0) + P(XY \leq t|X = 1) \cdot P(X = 1) =$$
$$= \frac{1}{2}\Big(P(Z = 0 \leq t|X = 0) + P(Y \leq t|X = 1)\Big)$$

Na množině $A = X^{-1}(\{0\})$ je veličina $Z$ konstantně nulová, takže výraz $P(Z = 0 \leq t|X = 0)$ je buď roven 1, když $0 \leq t$ anebo roven 0, když $t < 0$. Na množině $A$ jde tedy o Diracovo rozdělení s distribuční funkci formálně psanou jako $P(0 \leq t)$.

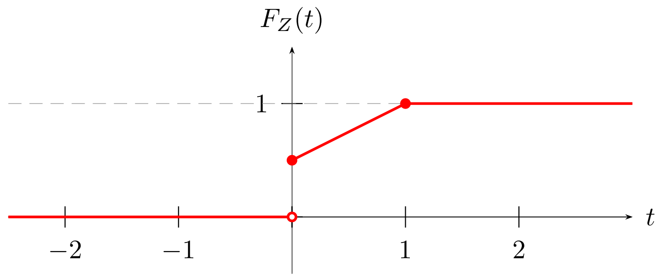
Diracovo rozdělení:




Nyní tedy máme

$$F_Z(t) = \frac{1}{2}\Big(P(0 \leq t) + F_Y(t)\Big) = \begin{cases} 0 & , \ t < 0 \\ \frac{x+1}{2} & , \ t \in \langle 0, 1\rangle \\ 1 & , \ t > 1 \ . \end{cases}$$

takže veličina $Z$ má smíšené rozdělení a graf $F_Z$ je:




Na začátku jsme také mohli veličiny vyjádřit jako směsi, kde by jedna ze složek odpovídala případu $X = 0$ a druhá případu $X = 1$ (tj. šlo by o zúžení veličin $W$ a $Z$ za daných podmínek). Pro $A = X^{-1}(\{0\})$ a $c = P(A) = P(X = 0) = \frac{1}{2}$ tak máme

$$W|_A = Y|_A, \quad W|_{\overline{A}} = Y + 1|_{\overline{A}}$$
$$Z|_A = 0|_A, \quad Z|_{\overline{A}} = Y|_{\overline{A}}$$

a pro vyjádření směsi tak dostáváme

$$W = \mathrm{Mix}_c(W|_A, W|_{\overline{A}}) = \mathrm{Mix}_{\frac{1}{2}}(Y|_A, Y + 1|_{\overline{A}})$$
$$Z = \mathrm{Mix}_c(Z|_A, Z|_{\overline{A}}) = \mathrm{Mix}_{\frac{1}{2}}(0|_A, Y|_{\overline{A}})$$

Zbytek postupu by pak už ale byl stejný.