6. cvičení z PSI

7. - 11. listopadu 2016

6.1 (kvantil, střední hodnota, rozptyl - pokračování příkladu z minula)
Náhodná veličina X má distribuční funkci

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

a je směsí diskrétní náhodné veličiny U a spojité náhodné veličiny V, konkrétně $X = \mathrm{Mix}_{\frac{7}{12}}(U, V)$.
Pro X, U a V najděte kvantilové funkce, střední hodnoty a rozptyly.

Řešení:
Kvantilová funkce pro veličinu X je definována jako

$$q_X(\alpha) = \frac{1}{2}\Big(\sup\{x \in \mathbb{R} \mid F(x) \le \alpha\} + \inf\{x \in \mathbb{R} \mid F(x) \ge \alpha\}\Big)$$

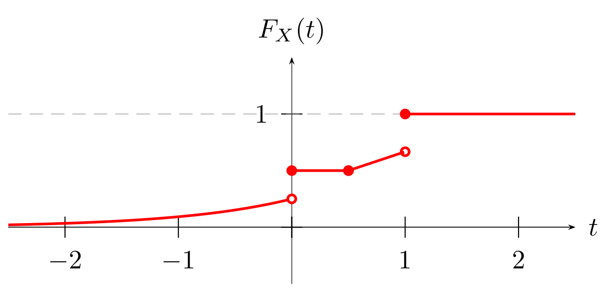
pro $\alpha \in (0, 1)$. Tato definice vypadá poněkud složitě, ale je to jen kvůli případným bodům nespojitosti funkce $F(x)$. Pro zbylé body platí následující:

Věta: *Distribuční funkce F_X a kvantilová funkce q_X jsou vůči sobě navzájem inverzní tam, kde jsou spojité a ostře rostoucí.*

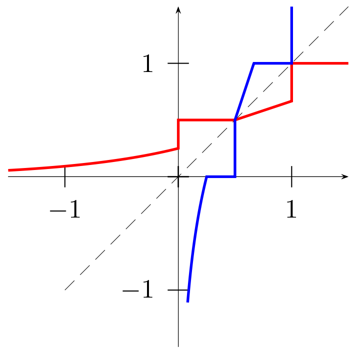
Obecněji pak platí následující: graf kvantilové funkce q_X získáme z grafu distribuční funkce F_X takto

- 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 z práva a zleva (případné krajní úseky v bodech $\alpha = 0$ a $\alpha = 1$ odstraníme úplně, protože tam se kvantil nedefinuje).

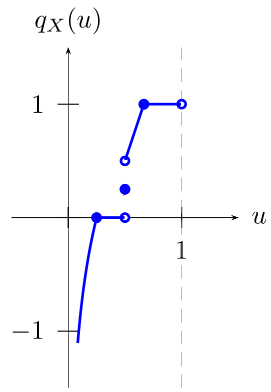
V našem případě tedy graf F_X:



přejde na



a dostaneme graf q_X:



Kvantilovou funkci určíme také explicitně a to tak, že najdeme příslušné inverze, tj. vyjádříme u z rovnic $\frac{e^t}{4} = u$ a $\frac{t+1}{3} = u$:

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

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. Dostáváme tak jiný pohled na to, proč je $E(X) = \int_0^1 q_X(u)\ du$.

Navíc pro každou integrabilní (tj. "rozumnou") funkci $h : \mathbb{R} \to \mathbb{R}$ máme, že $E(h(X)) = \int_0^1 h(q_X(u))\ du$. Neboli s kvantilem můžeme zacházet opravdu úplně stejně jako s původní veličinou.

Protože máme určen kvantil, můžeme ho využít pro výpočet střední hodnoty $E(X)$:

$$E(X) = \int_0^1 q_X(u)\ du = \int_0^{\frac{1}{4}} \ln(4u)\ du + \int_{\frac{1}{2}}^{\frac{2}{3}} 3u - 1\ du + \int_{\frac{2}{3}}^1 1\ du =$$

$$= \frac{1}{4}\int_0^1 \ln v\ dv + \left[\frac{3u^2}{2} - u\right]_{\frac{1}{2}}^{\frac{2}{3}} + \left(1 - \frac{2}{3}\right) = \frac{1}{4}\Big[v(\ln v - 1)\Big]_0^1 + \frac{11}{24} = -\frac{1}{4} + \frac{11}{24} = \frac{5}{24} \doteq 0.208\ .$$

Střední hodnota odpovídá tedy ploše určené kvantilem a ta se dá při překlopení podle osy 1. a 3. kvadrantu spočítat i z distribuční funkce jako

$$E(X) = -\int_{-\infty}^0 F_X(t)\ dt + \int_0^\infty \big(1 - F_X(t)\big)\ dt\ .$$

To má tu výhodu, že nemusíme určovat kvantil, ani veličinu rozkládat na diskrétní a spojitou složku.

Pro rozptyl platí

$$D(X) = E\Big(\big(X - E(X)\big)^2\Big) = E(X^2) - \Big(E(X)\Big)^2\ .$$

Můžeme sice opět využít kvantil, ale tentokrát si vyzkoušíme jiný vzorec:

$$E(X^2) = -\int_{-\infty}^0 2t \cdot F_X(t)\ dt + \int_0^\infty 2t \cdot \big(1 - F_X(t)\big)\ dt =$$

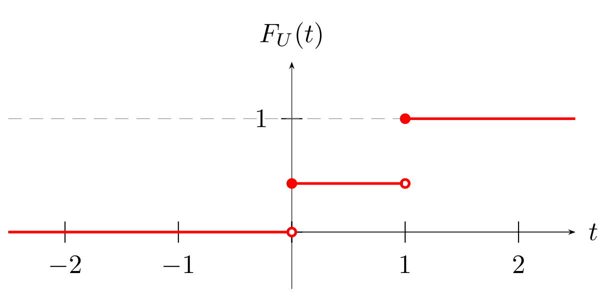
$$= -\int_{-\infty}^0 2t \cdot \frac{e^t}{4}\ dt + \int_0^{\frac{1}{2}} 2t \cdot \left(1 - \frac{1}{2}\right)\ dt + \int_{\frac{1}{2}}^1 2t \cdot \left(1 - \frac{t+1}{3}\right)\ dt + \int_1^\infty 0\ dt =$$

$$= -\left[\frac{(t-1)e^t}{2}\right]_{-\infty}^0 + \left[\frac{t^2}{2}\right]_0^{\frac{1}{2}} + \left[\frac{2}{3}\left(t^2 - \frac{t^3}{3}\right)\right]_{\frac{1}{2}}^1 = \frac{1}{2} + \frac{1}{8} + \frac{4}{9} - \frac{2}{3}\left(\frac{1}{4} - \frac{1}{24}\right) = \frac{67}{72} \doteq 0.931\ .$$

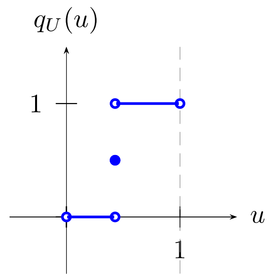
a rozptyl tedy bude

$$D(X) = E(X^2) - \Big(E(X)\Big)^2 = \frac{67}{72} - \left(\frac{5}{24}\right)^2 = \frac{511}{576} \doteq 0.887\ .$$

Výpočet pro *diskrétní část* U: z distribuční funkce F_U



určíme graf q_U:



Explicitní tvar q_U:

$$q_U(u) = \begin{cases} 0 & , \; u \in (0, \frac{3}{7}) \\ \frac{1}{2} & , \; u = \frac{3}{7} \\ 1 & , \; u \in (\frac{3}{7}, 1). \end{cases}$$

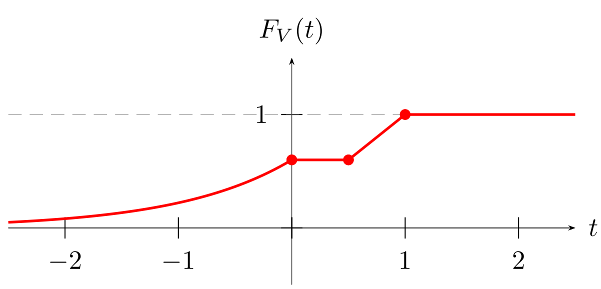
Střední hodnotu $E(U)$ a rozptyl $D(U)$ teď pro změnu spočítáme způsobem, který je obvyklejší pro diskrétní veličiny, tedy přes pravděpodobnostní funkci (která má nenulové hodnoty $p_U(0) = \frac{3}{7}$ a $p_U(1) = \frac{4}{7}$):

$$E(U) = \sum_{t \in \mathbb{R}} t \cdot p_U(t) = 0 \cdot \frac{3}{7} + 1 \cdot \frac{4}{7} = \frac{4}{7}$$

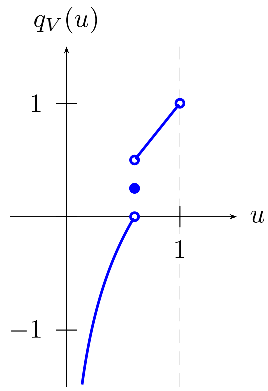
$$E(U^2) = \sum_{t \in \mathbb{R}} t^2 \cdot p_U(t) = 0^2 \cdot \frac{3}{7} + 1^2 \cdot \frac{4}{7} = \frac{4}{7}$$

$$D(U) = E(U^2) - \big(E(U)\big)^2 = \frac{4}{7} - \left(\frac{4}{7}\right)^2 = \frac{12}{49} \; .$$

Výpočet pro *spojitou část* V: z distribuční funkce F_V



určíme graf q_V:



Explicitní tvar q_V:

$$q_V(u) = \begin{cases} \ln(\frac{5}{3}u) & , \ t \in (0, \frac{3}{5}) \\ \frac{1}{4} & , \ t = \frac{3}{5} \\ \frac{5u-1}{4} & , \ t \in (\frac{3}{5}, 1). \end{cases}$$

Střední hodnotu $E(V)$ a rozptyl $D(V)$ opět spočítáme obvyklým způsobem přes hustotu (která má nenulové hodnoty $f_V(t) = \frac{3}{5}e^t$ pro $t \le 0$ a $f_V(t) = \frac{4}{5}$ pro $t \in \langle \frac{1}{2}, 1 \rangle$):

$$E(V) = \int_{-\infty}^{\infty} t \cdot f_V(t)\ dt = \int_{-\infty}^{0} t \cdot \frac{3}{5}e^t\ dt + \int_{\frac{1}{2}}^{1} t \cdot \frac{4}{5}\ dt = \frac{3}{5}\Big[(t-1)e^t\Big]_{-\infty}^{0} + \left[\frac{2t^2}{5}\right]_{\frac{1}{2}}^{1} =$$

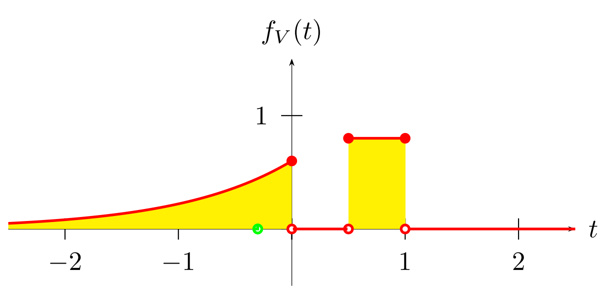
$$= -\frac{3}{5} + \frac{3}{10} = -\frac{3}{10} = -0.3$$

$$E(V^2) = \int_{-\infty}^{\infty} t^2 \cdot f_V(t)\ dt = \int_{-\infty}^{0} t^2 \cdot \frac{3}{5}e^t\ dt + \int_{\frac{1}{2}}^{1} t^2 \cdot \frac{4}{5}\ dt = \frac{3}{5}\Big[(t^2 - 2t + 2)e^t\Big]_{-\infty}^{0} + \left[\frac{4t^3}{15}\right]_{\frac{1}{2}}^{1} =$$

$$= \frac{6}{5} + \frac{7}{30} = \frac{43}{30} \doteq 1.433$$

$$D(V) = E(V^2) - \big(E(V)\big)^2 = \frac{43}{30} - \left(\frac{3}{10}\right)^2 = \frac{403}{300} \doteq 1.343\ .$$

Opět si můžeme názorně interpretovat střední hodnotu $E(V)$ spojité veličiny V, která má hustotu f_V, jako x-ovou souřadnici (zelený bod) těžiště plochy, která je určena grafem hustoty f_V:



Celkově si výsledky ještě můžeme překontrolovat pomocí vztahu pro střední hodnotu a rozptyl směsi $X = \mathrm{Mix}_c(U, V)$:

$$E(X) = cE(U) + (1-c)E(V)$$

$$D(X) = cD(U) + (1-c)D(V) + c(1-c)\Big(E(U) - E(V)\Big)^2$$

Odvození pro rozptyl:

$$D(X) = E(X^2) - \big(E(X)\big)^2 = cE(U^2) + (1-c)E(V^2) - \Big(cE(U) + (1-c)E(V)\Big)^2 =$$
$$= c\Big(D(U) + \big(E(U)\big)^2\Big) + (1-c)\Big(D(V) + \big(E(V)\big)^2\Big) - \Big[c^2\big(E(U)\big)^2 + 2c(1-c)E(U)E(V) + (1-c)^2\big(E(V)\big)^2\Big] =$$
$$= cD(U) + (1-c)D(V) + c(1-c)\Big(E(U) - E(V)\Big)^2 .$$

V našem případě tedy skutečně pro $c = \frac{7}{12}$ dostaneme:

veličina	X	U	V
$E(\cdot)$	$\frac{5}{24}$	$\frac{4}{7}$	$-\frac{3}{10}$
$D(\cdot)$	$\frac{511}{576}$	$\frac{12}{49}$	$\frac{403}{300}$

$$E(X) = \frac{7}{12} \cdot \frac{4}{7} + \frac{5}{12} \cdot \Big(-\frac{3}{10}\Big) = \frac{5}{24}$$
$$D(X) = \frac{7}{12} \cdot \frac{12}{49} + \frac{5}{12} \cdot \frac{403}{300} + \frac{7}{12} \cdot \frac{5}{12}\Big(\frac{4}{7} + \frac{3}{10}\Big)^2 =$$
$$= \frac{1}{7} + \frac{403}{12^2 \cdot 5} + \frac{61^2}{12^2 \cdot 35 \cdot 4} = \frac{511}{576} .$$

6.2 (Hypergeometrické rozdělení)

Mezi M výrobky je K vadných. Jaká je pravděpodobnost, že mezi m náhodně vybranými výrobky je právě k vadných?

Určete, k čemu se blíží hodnota pravděpodobnosti pro pevné k a m pokud $M \to \infty$ a $K/M \to q$, kde $0 \le q \le 1$. Jak byste tento výsledek interpretovali?

Řešení:
Pravděpodobnost bude 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

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

Rozdělení náhodné veličiny X, která označuje počet vadných výrobků ve vybraném vzorku m výrobků (z množiny M obsahující K vadných výrobků), se nazývá hypergeometrické. Tedy

$$P(X = k) = p_{K,M}(k, m).$$

Určíme ještě limitu ze zadání:

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

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

protože uvažujeme, ze k a m jsou pevné a předpokládáme, ze $K/M \to q$ pro $M \to \infty$.

Dostáváme tedy známe binomické rozdělení, které v tomto případě odpovídá limitní situaci, kdy taháme výrobky z nekonečného množství s určeným podílem q vadných. Při n 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, ze v obrovském množství výrobků M se už v limite ztratí to, jestli tam těch pár výrobků m vrátíme nebo ne (protože $m << M$).

6.3 (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.

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

Řešení:

Máme tedy veličinu

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

U této veličiny můžeme předpokládat Poissonovo rozdělení

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

s (bezrozměrným) parametrem $\lambda > 0$, pokud jsou splněny následující podmínky

- události je možné oddělit intervaly stejné délky (tj. pravděpodobnost, ze nastane více než jedna událost v těchto malých interválcích 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, chytání ryb, průjezd aut atd. během nějaké předem určené doby.

Parametr λ pak představuje střední hodnotu (tj. $E(X) = k$). Poissonovo rozdělení je většinou spíše limitní případ a používá se jako aproximace binomického rozdělení Bi(n, p) u kterého neznáme n a p a kde n je dostatečně velké.

V našem případě je tedy $\lambda = 6$ (jde o bezrozměrné číslo).

(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 $\lambda'=12$. Takže

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

(c) Protože pro rozdělení pravděpodobností platí

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

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

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

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.

Zdůvodníme si ještě, proč má Poissonovo rozdělení takový zápis, jaký má. 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

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

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

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

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

Odsud je i vidět, kdy se dá Poissonovo rozdělení použít jako aproximace binomického. Potřebujeme, aby:

$$1-\frac{\lambda}{n}\doteq 1,\qquad 1-\frac{i}{n}\doteq 1\quad\text{pro}\quad 0\le i\le k\quad\text{a}\quad\Big(1-\frac{\lambda}{n}\Big)^n\doteq e^{-\lambda}\quad(\text{neboli}\quad n\cdot\ln\Big(1-\frac{\lambda}{n}\Big)\doteq -\lambda)$$

tedy stručně řečeno, aby

$$\lambda<<n\quad\text{a}\quad k<<n\ .$$

Pokud tedy kolem okna projede za celý den velké množství aut n, pak během krátké doby bude jejich střední počet λ malý. Pokud nás budou zajímat jen pravděpodobnosti malých počtů k, můžeme s klidem použít Poissonovo rozdělení.

6.4 (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, nabízí se pro výpočet použít aproximaci pomoci Poissonova rozdělení. Parametrem bude střední hodnota binomického rozdělení, tj. $\lambda = E(X) = n\cdot p = 300\cdot 0.01 = 3$.

Hodnoty $\lambda = 3$ i $k = 4$ jsou malé ve srovnání s $n = 300$. To nás opravňuje použít tuto aproximaci.

Dostáváme tak

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

6.5 (exponenciální rozdělení, transformace veličiny)

Náhodná veličina X má exponenciální rozdělení s distribuční funkcí

$$F_X(t)=\begin{cases}0 & ,\ t\le 0\\ 1-e^{-2t} & ,\ t>0\ .\end{cases}$$

Popište rozdělení náhodné veličiny $Y = 2-2X$ a stanovte její střední hodnotu a rozptyl.

Řešení:

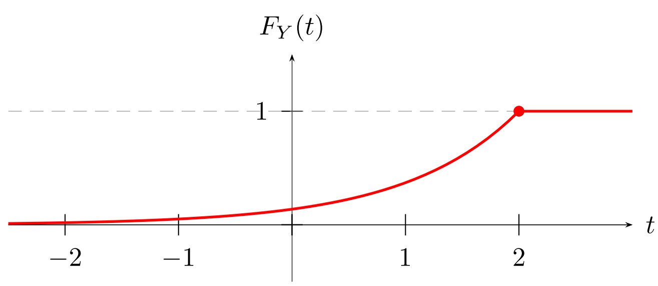
Najdeme rozdělení veličiny $Y = 2-2X$:

$$F_Y(t)=P(Y\le t)=P(2-2X\le t)=P\Big(1-\frac{t}{2}\le X\Big)=1-P\Big(X<1-\frac{t}{2}\Big)=1-F_X\Big(1-\frac{t}{2}\Big)$$

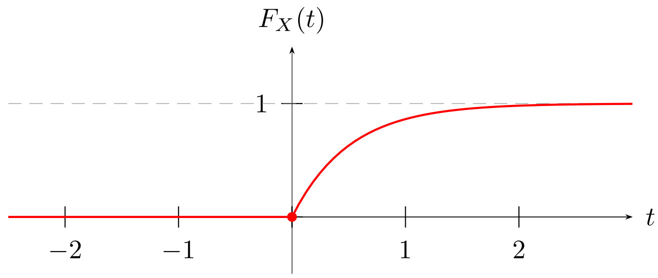
Pro $1-\frac{t}{2}\le 0$ je tedy $F_Y(t)=1$ a pro $1-\frac{t}{2}>0$ máme $F_Y(t)=1-(1-e^{-2(1-\frac{t}{2})})=e^{t-2}$. Celkově tedy dostáváme

$$F_Y(t)=\begin{cases}e^{t-2} & ,\ t<2\\ 1 & ,\ t\ge 2\ .\end{cases}$$

Graf F_Y si snadno nakreslíme:



ale je dobré si uvědomit, že pro tvar $F_Y(t)=1-F_X\Big(-\frac{1}{2}(t-2)\Big)$ lze F_Y získat z grafu F_X:



následující posloupností transformací:

- $g_0(a) := F_X(a)$
- $g_1(b) := g_0(-\frac{1}{2}b) = F_X(-\frac{1}{2}b)$ (graf g_0 se otočí kolem osy y a 2-krát se roztáhne podle středu 0 ve směru osy x)
- $g_2(c) := g_1(c-2) = F_X(-\frac{1}{2}(c-2))$ (graf g_1 se posune o vektor $(2,0)$, tj. doprava o 2)
- $g_3(d) := 1 - g_2(d) = 1 - F_X\Big(-\frac{1}{2}(d-2)\Big) = F_Y(d)$ (graf g_2 se otočí kolem osy x a posune nahoru o 1)

Protože transformace nezávisle t proměnné lze vyjádřit také jako $F_X\big(-\frac{1}{2}(t-2)\big) = F_X\big(1 + (-\frac{1}{2}t)\big)$, můžeme na začátku alternativně použít tento postup:

- $\widetilde{g}_1(b) := g_0(1+b) = F_X(1+b)$ (graf g_0 se posune o vektor $(-1,0)$, tj. doleva o 1)
- $\widetilde{g}_2(c) := \widetilde{g}_1(-\frac{1}{2}c) = F_X\big(1 + (-\frac{1}{2}c)\big)$ (graf $\widetilde{g}_1$ se otočí kolem osy y a 2-krát se roztáhne podle středu 0 ve směru osy x)

Evidentně máme $\widetilde{g}_2(c) = g_2(c)$. Poslední transformace (tj. že graf g_2 se otočí kolem osy x a posune nahoru o 1) zůstává stejná.

Protože veličina Y se získá z X jako $Y = h(X)$, kde $h(x) = 2 - 2x$ je (ostře) *klesající* spojitá funkce, mohli jsme k odvození F_Y a jejího grafu použít také vztahu známého z přednášky, a sice:

$$F_Y(t) = F_{h(X)}(t) = 1 - F_X(h^{-1}(t))$$

kde $h^{-1}(y) = 1 - \frac{1}{2}y$ je inverze k h (tj. vypočítaná ze vztahu $y = 2 - 2x$.)

Určíme ještě střední hodnotu (tentokrát pro změnu z distribuční funkce):

$$E(Y) = -\int_{-\infty}^{0} F_Y(t)\,dt + \int_0^\infty \big(1 - F_Y(t)\big)\,dt = -\int_{-\infty}^{0} e^{t-2}\,dt + \int_0^2 \big(1 - e^{t-2}\big)\,dt = 2 - \Big[e^{t-2}\Big]_{-\infty}^{2} = 1$$

Mohli jsme také využít, že X má exponenciální rozdělení s parametrem $\tau = \frac{1}{2}$ a střední hodnotou $E(X) = \tau$. Pak bychom opět měli

$$E(Y) = E(2 - 2X) = 2 - 2E(X) = 1\ .$$

Pro rozptyl můžeme jednoduše využít, že $D(X) = \tau^2 = \frac{1}{4}$, takže

$$D(Y) = D(2 - 2X) = D(-2X) = (-2)^2 D(X) = 1\ .$$

Exponenciální rozdělení popisuje pravděpodobnost veličiny

$$X = \text{“}\textit{doba mezi dvěma následnými výskyty události}\text{”},$$

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

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

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.

6.6 (transformace spojité veličiny)

Náhodná veličina X má rovnoměrné rozdělení na intervalu $(-2, 2)$. Zobrazíme ji funkci h, definovanou následovně:

$$h(x) = \begin{cases} 0 & , \ x < 0 \\ x^2 & , \ 0 \le x \le 1 \\ 1 & , \ x > 1. \end{cases}$$

Nalezněte rozdělení náhodné veličiny $Y = h(X)$.

Řešení:
(a) **Krátké řešení:** Protože $0 \le Y = h(X) \le 1$, tak

$$F_Y(t) = P(h(X) \le t) = \begin{cases} 0 & , \ t < 0 \\ 1 & , \ t \ge 1. \end{cases}$$

Pro $t \in \langle 0, 1)$ máme

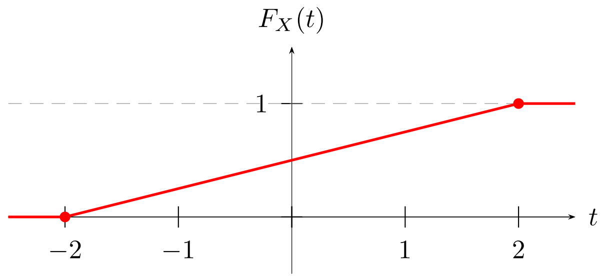
$$h(X) \le t \ \Leftrightarrow \ \Big(X < 0 \ \vee \big(X \in \langle 0,1\rangle \ \& \ X^2 \le t\big)\Big) \ \Leftrightarrow$$

$$\Leftrightarrow \ \Big(X < 0 \ \vee \big(X \in \langle 0,1\rangle \ \& \ X \le \sqrt{t}\big)\Big) \ \Leftrightarrow \ X \le \sqrt{t}$$

a tudíž

$$F_Y(t) = P(h(X) \le t) = P(X \le \sqrt{t}) = F_X(\sqrt{t})\ .$$

Teď už stačí jen dosadit za F_X. Protože veličina X je spojitá s konstantní hustotou na intervalu $(-2, 2)$, tak její distribuční funkce F_X bude lineární na tomto intervalu. Jinak bude konstantní a celkově spojitá:

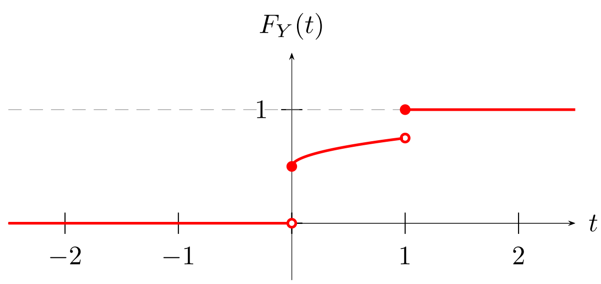


Tedy

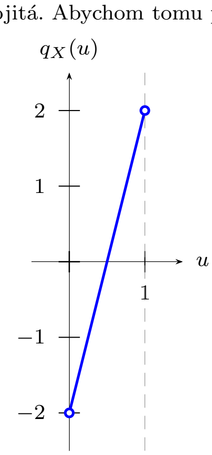
$$F_X(t) = \begin{cases} 0 & ,\ t < -2 \\ \frac{t+2}{4} & ,\ t \in \langle -2, 2 \rangle \\ 1 & ,\ t > 2. \end{cases}$$

a po dosazení dostaneme:

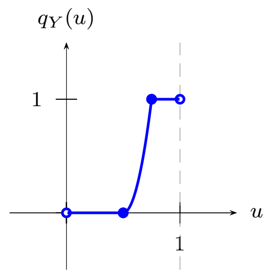
$$F_Y(t) = \begin{cases} 0 & ,\ t < 0 \\ \frac{\sqrt{t}+2}{4} & ,\ t \in \langle 0, 1) \\ 1 & ,\ t \geq 1. \end{cases}$$



Veličina $Y = h(X)$ tedy kupodivu nemá spojité rozdělení ale smíšené, přestože veličina X je (absolutně) spojitá a i funkce h je spojitá. Abychom tomu porozuměli, podíváme se na kvantil. Kvantil q_X je spojitý:



A tedy i kvantil $q_Y = q_{h(X)} = h(q_X)$ je spojitý:

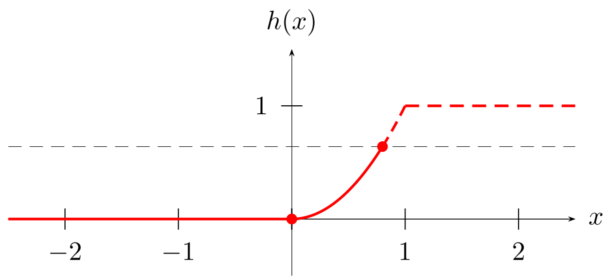


Ale protože je místy konstantní, tak F_Y má skoky.

(b) **Řešení použitelné i pro obecnější případy:** 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:

$$h^{-1}(-\infty, t\rangle = \begin{cases} \mathbb{R} & , \ t \geq 1 \\ (-\infty, \sqrt{t}\rangle & , \ 0 \leq t < 1 \\ \emptyset & , \ t < 0. \end{cases}$$

Takže můžeme psát:

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

Zbytek postupu už pak bude stejný jako v předchozím případě.

6.7 (nezávislé diskrétní veličiny)

Náhodné veličiny X a Y jsou nezávislé a mají diskrétní rozdělení s pravděpodobnostními funkcemi

a	-1	2
$p_X(a)$	0.3	0.7

b	0	1	3
$p_Y(b)$	0.2	0.45	0.35

Vypočtete střední hodnotu součinu $E(XY)$, pravděpodobnostní funkci náhodné veličiny $Z = X+Y$ a koeficient korelace $\rho(X, Y)$.

Řešení:
Veličiny jsou *nezávislé*, proto můžeme psát:

$$E(XY) = E(X) \cdot E(Y)$$

$$E(X) = \sum_{a \in \mathbb{R}} a \cdot p_X(a) = (-1) \cdot 0.3 + 3 \cdot 0.7 = 1.8$$

$$E(Y) = 0 \cdot 0.2 + 1 \cdot 0.45 + 3 \cdot 0.35 = 1.5$$

takže $E(XY) = 1.8 \cdot 1.5 = 2.7$.
Pro pravděpodobnostní funkci náhodné veličiny $Z = X + Y$ máme:

$$p_Z(c) = P(X + Y = c) = \sum_{a+b=c} P(X = a \,\&\, Y = b) = \sum_{a+b=c} p_X(a) \cdot p_Y(b)$$

Hodnoty c, kterých nabývá veličina Z, jsou všechny možné součty hodnot od veličiny X a od veličiny Y. Takže dostáváme

c	-1	0	2	3	5
možné rozklady $a + b = c$	-1+0	-1+1	2+0, -1+3	2+1	2+3
$\sum p_X(a)p_Y(b)$	$0.3 \cdot 0.2$	$0.3 \cdot 0.45$	$0.7 \cdot 0.2 + 0.3 \cdot 0.35$	$0.7 \cdot 0.45$	$0.7 \cdot 0.35$
$p_Z(c)$	0.06	0.135	0.245	0.315	0.245

Protože veličiny jsou *nezávislé*, koeficient korelace je $\rho(X, Y) = 0$ (plyne ihned ze vztahu $\rho(X, Y) = \frac{E(XY)-E(X)\cdot E(Y)}{\sqrt{D(X)\cdot D(Y)}}$).