6. cvičení z PST

7. listopadu 2018

6.1 (geometrické rozdělení)

Alice a Bob hází na koš. Alice se trefí s pravděpodobností $p_1 = 0.05$ a Bob s pravděpodobností $p_2 = 0.08$. Začíná Alice a střídají se. S každým hodem Alice se začíná nové kolo. Náhodná veličina X je počet kol, které uběhnou předtím, než se někdo trefí.

(a) Určete rozdělení veličiny X.

(b) V kterém kole se průměrně někdo poprvé trefí?

(c) Jaký nejmenší počet kol musí začít, aby se s pravděpodobností alespoň 90% někdo trefil?

Řešení:
Postupujeme podobně jako v příkladu **5.5**.

(a) V rámci kola označme jevy

$A =$ "Alice se trefí"

$B =$ "Bob se trefí"

$C =$ "alespoň jeden se trefí"

Pak $C = A \cup B$ a protože A a B jsou nezávislé, máme $p := P(C) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = p_1 + (1 - p_1)p_2 = 0.05 + 0.95 \cdot 0.08 = 0.126$.

Veličina X je diskrétní s oborem hodnot $\{0, 1, 2, \dots\}$ a má geometrické rozdělení $p_X(k) = p(1 - p)^k$ pro $k = 0, 1, 2, \dots$.

(b) Pro veličinu $Y =$ "pořadí kola, ve kterém se někdo poprvé trefí" je $Y = X + 1$. Tedy $E(Y) = E(X + 1) = E(X) + 1 = \frac{1-p}{p} + 1 = \frac{1}{p} = \frac{1}{0.126} \doteq 7.94$.

Pro určení $E(X)$ (a případně i $D(X)$) buď můžeme sčítat mocninnou řadu (viz opět příklad **5.5**) nebo využijeme vyjádření veličiny X jako směs:

Označme si jev $D =$ "v 1.kole se alespoň jeden trefí". Pak D představuje právě množinu "$X = 0$", tedy $P(D) = p$ a $X_{|D} = 0$. Současně veličina $X' := X_{|\overline{D}}$ představuje počet kol, které uběhnou předtím, než se někdo trefí, za předpokladu, že v 1. kole se nikdo netrefí. Neboli, veličina $X' - 1$ znamená, že s počítáním kol (které uběhnou předtím, než se někdo trefí) začneme až po prvním (neúspěšném) kole. Po prvním neúspěšném kole jsme tedy ve stejné situaci, jako na začátku. Proto veličina $X' - 1$ má opět geometrické rozdělení s parametrem p, tj. stejné rozdělení jako má X.

Můžeme si to dokázat i explicitně. Množina $\overline{D}$ představuje právě množinu "$X \neq 0$". Pro $k \in \mathbb{N}$ je

$$p_{X'}(k) = P(X' = k|\overline{D}) = \frac{P\Big((X' = k) \ \cap \ (X \neq 0)\Big)}{P(X \neq 0)} = \frac{P(X = k)}{P(X \neq 0)} = \frac{p(1-p)^k}{1-p} = p(1-p)^{k-1} = p_X(k-1)$$

což znamená právě to, že $X' - 1$ a X mají stejná rozdělení. Z tohoto důvodu tedy speciálně máme $E(X) = E(X' - 1) = E(X') - 1$ a tedy $E(X') = E(X) + 1$.

Díky směsi

$$X = \mathrm{Mix}_p\Big(X_{|D}, X_{|\overline{D}}\Big) = \mathrm{Mix}_p\Big(0, X'\Big)$$

tedy dostáváme

$$E(X) = pE(0) + (1-p)E(X') = p \cdot 0 + (1-p)(E(X) + 1)$$

neboli

$$E(X) = \frac{1-p}{p} \ .$$

Pro rozptyl použijeme formuli pro rozptyl obecné směsi

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

kterou si odvodíme. Nejdříve použijeme, že

$$X^2 = \mathrm{Mix}_c(U^2, V^2)$$

a tedy

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

a

$$E(X^2) = cE(U^2) + (1-c)E(V^2)\ .$$

Tudíž

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

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

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

Tedy v našem případě ($U = 0$ a $V = X'$) máme (díky stejnému rozdělení X a $X'-1$) opět $D(X) = D(X'-1) = D(X')$, takže

$$D(X) = pD(0) + (1-p)D(X') + p(1-p)\Big(E(0) - E(X')\Big)^2 =$$

$$= (1-p)D(X) + p(1-p)\Big(E(X) + 1\Big)^2 = (1-p)D(X) + p(1-p)\tfrac{1}{p^2}$$

a výsledek je

$$D(X) = \frac{1-p}{p^2}\ .$$

(c) Chceme tedy znát nejmenší $n \in \mathbb{N}$ takové, že $P\Big(Y \le n\Big) = P\Big(X \le n-1\Big) \ge 0.9$. Takže

$$0.9 \le P\Big(X \le n-1\Big) = 1 - (1-p)^n$$

neboli

$$0.1 \ge (1-p)^n = (1-0.126)^n = 0.874^n$$

$$n \ge \frac{\log 0.1}{\log 0.874} \doteq 17.097$$

Tedy musí začít alespoň $n = 18$ kol.

6.2 (Poissonovo a exponenciální rozdělení)

Během hodiny přijme telefonní operátor průměrně 5 hovoru.

(a) Jaká je pravděpodobnost, že jich za hodinu přijme méně než 3?

(b) Jaká je pravděpodobnost, že když si operátor na 10 min odskočí, nikdo mu nezavolá?

(c) Operátorovi je přidělena skupina $n = 300$ lidí. Každý z nich mu bude během hodiny volat s pravděpodobností $p = 0.01$ (nezávisle na ostatních a nejvýše jednou). Jaká je pravděpodobnost toho, že během hodiny zavolají právě 4 lidé z této skupiny?

Řešení:
(a) Máme tedy veličinu

$$X = \text{“\textit{počet hovorů během jedné hodiny}”}\ ,$$

U této veličiny můžeme předpokládat Poissonovo rozdělení

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

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

- události pocházejí z velkého počtu nezávislých zdrojů.

- jednotlivé události nenastávají současně (lze je časově oddělit) a z každého zdroje nastane událost nejvýše jednou,
- průměrný počet událostí v libovolném časovém podintervalu je úměrný pouze časové délce tohoto podintervalu a ne jeho umístění v původním intervalu (tj. lze říct, že četnosti událostí za jednotku času se s průběhem doby nemění),

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. $E(X) = \lambda$) protože:

$$E(X) = \sum_{k=0}^{\infty} k p_X(k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \lambda e^{-\lambda} \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{=e^{\lambda}} = \lambda$$

Poissonovo rozdělení je většinou spíše limitní případ a používá se jako aproximace binomického rozdělení $\mathrm{Bi}(n, p)$ u kterého sice neznáme n a p ale víme, že n je (dostatečně) velké a známe střední hodnotu dané veličiny.

Za výše uvedených předpokladů dojdeme ke tvaru Poissonova rozdělení takto:

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

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

se střední hodnotou $\lambda = E(X_n) = n \cdot p_n$, kterou si vezmeme jako pevnou (neboli vlastně položíme $p_n := \frac{\lambda}{n}$). Tedy X_n má binomické rozdělení $\mathrm{Bi}(n, p_n)$. Spočítáme si teď limitu:

$$\lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} \binom{n}{k} (p_n)^k (1 - p_n)^{n-k} = \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} =$$

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

V našem případě je tedy $\lambda = E(X) = 5$ (jde o bezrozměrné číslo).

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

$$= e^{-5}\left(1 + 5 + \frac{5^2}{2}\right) = \tfrac{37}{2} e^{-5} \doteq 0.12465 \ .$$

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

Exponenciální rozdělení popisuje pravděpodobnost veličiny

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

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:

$$P(Y > s + t | Y > s) = P(Y > t)$$

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

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

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

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

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

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

$$P(\widetilde{Y} > k + n | \widetilde{Y} > n) = P(\widetilde{Y} > k)$$

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

Věta: Nechť

$Y =$ *"doba čekání na událost"*

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

Pak veličina

$X =$ *"počet událostí během doby T"*

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

V našem případě máme

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

- *Pomocí exponenciálního:* Podle zadání má veličina X Poissonovo rozdělení s parametrem $\lambda = 5$. Tudíž náhodná veličina Y s exponenciálním rozdělením má parametr $\tau = \frac{T}{\lambda} = \frac{60\ min}{5} = 12\ min$, kde $T = 60\ min$. Hledaná pravděpodobnost je pak

$$P(Y > 10\ min) = 1 - P(Y \leq 10\ min) = 1 - F_Y(10) = 1 - \left(1 - e^{-\frac{10}{12}}\right) = e^{-\frac{5}{6}} \doteq 0.4346$$

- *Pomocí Poissonova:* Uvažujme veličinu

  $X' =$ *"počet hovorů během 10 minut"*

  Hledaná pravděpodobnost je dána jako

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

  kde λ' je parametr veličiny X' s Poissonovým rozdělením. K určení parametru λ' použijeme vztah mezi veličinami Y a X' a už známou hodnotu parametru veličiny X, tj. platí $\lambda' = \frac{T'}{\tau} = \lambda \cdot \frac{T'}{T}$ neboli

$$\frac{E(X')}{E(X)} = \frac{T'}{T}$$

  což odpovídá i výše zmíněnému požadavku, že průměrný počet událostí v časovém intervalu je úměrný jeho délce. Konkrétně tedy $\lambda' = \frac{10\ min}{60\ min} \cdot 5 = \frac{5}{6}$ a hledaná pravděpodobnost je pak

$$P(X' = 0) = e^{-\frac{5}{6}} \doteq 0.4346.$$

(c) Zde máme veličinu

$Z_n =$ *"počet hovorů během jedné hodiny ze skupiny n lidí"*

Jde o binomické rozdělení $\text{Bi}(n, p)$ a pravděpodobnost bude

$$P(Z_n = 4) = \binom{n}{4} \cdot p^4 \cdot (1-p)^{n-4} = \binom{300}{4} \cdot 0.01^4 \cdot 0.99^{296} \doteq 0.168877\ .$$

Protože zde násobíme velká čísla malými a počítáme vysoké mocniny, nabízí se pro výpočet použít aproximaci pomocí Poissonova rozdělení se stejnou střední hodnotou jako má veličina Z_n, tj. parametr Poissonova rozdělení bude $\lambda = E(Z_n) = n \cdot p = 300 \cdot 0.01 = 3$.

Z výše spočítané limity pro Poissonovo rozdělení potřebujeme k aproximaci binomického rozdělení aby:

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

tedy stručně řečeno, aby

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

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

$$P(Z_n = 4) \doteq \frac{\lambda^4}{4!} \cdot e^{-\lambda} = \frac{3^4}{4!} \cdot e^{-3} \doteq 0.168031\ .$$

6.3 (Poissonovo a exponenciální rozdělení)

Do pojišťovny přijdou průměrně 2 hlášení škody denně.

(a) Jaká je pravděpodobnost, že za den přijdou alespoň 4?

(b) Jaká je pravděpodobnost, že do pojišťovny přijde nejbližší hlášení škody nejdříve třetí den?

(c) Agent pojišťovny spravuje skupinu $n = 500$ klientů. Každý z nich mu během dne nahlásí událost s pravděpodobností $p = 0.001$ (nezávisle na ostatních a nejvýše jednou). Jaká je pravděpodobnost toho, že se mu během dne ozve právě 5 klientů z této skupiny?

Řešení:
Postupujeme stejně jako v příkladu **6.2**.
(a) $X =$ *"počet hlášení za den"* , $X \sim \text{Poiss}(\lambda)$, $\lambda = 2$.

$$P(X \geq 4) = 1 - P(X < 4) = 1 - e^{-\lambda}\Big(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!}\Big) = 1 - e^{-2}\Big(1 + 2 + \frac{2^2}{2} + \frac{2^3}{3!}\Big) =$$
$$= 1 - \tfrac{19}{3}e^{-2} \doteq 1 - 0.857123 = 0.142877 \ .$$

(b) Pomocí exponenciálního:
$Y =$ *"doba čekání na příchod dalšího hlášení"*, $Y \sim \text{Exp}(\tau)$, $\tau = \frac{T}{\lambda} = \frac{1\ den}{2} = 0.5\ dne$, kde $T = 1\ den$.

$$P(Y > 2\ dny) = 1 - P(Y \leq 2\ dny) = 1 - F_Y(2) = 1 - \Big(1 - e^{-\frac{2}{0.5}}\Big) = e^{-4} \doteq 0.01831.$$

Pomocí Poissonova:
$X' =$ *"počet hlášení během 2 dnů"*, $X \sim \text{Poiss}(\lambda')$, $\lambda' = \lambda \cdot \frac{T'}{T} = 2 \cdot \frac{2\ dny}{1\ den} = 4$, kde $T' = 2\ dny$.

$$P(X' = 0) = \frac{(\lambda')^0}{0!} \cdot e^{-\lambda'} = e^{-\lambda'} = e^{-4} \doteq 0.01831.$$

(c) $Z_n =$ *"počet hlášení ze skupiny n lidí za den"*, $Z_n \sim \text{Bi}(n,p)$,

$$P(Z_n = 5) = \binom{n}{5} \cdot p^5 \cdot (1-p)^{n-5} = \binom{500}{5} \cdot 0.001^5 \cdot 0.999^{495} \doteq 1.5555 \cdot 10^{-4} \ .$$

Pomocí Poissonova rozdělení s $\lambda = E(Z_n) = n \cdot p = 500 \cdot 0.001 = 0.5$:

$$P(Z_n = 5) \doteq \frac{\lambda^5}{5!} \cdot e^{-\lambda} = \frac{0.5^5}{5!} \cdot e^{-0.5} \doteq 1.5795 \cdot 10^{-4} \ .$$

6.4 (spojitá veličina - střední hodnota, rozptyl, transformace, kvantil)
Nezáporná náhodná veličina X má hustotu

$$f_X(t) = \begin{cases} 0 & , \ t < 1 \\ \frac{1}{t^2} & , \ t \geq 1. \end{cases}$$

Uvažujme veličinu $Y = \frac{1}{X}$. Určete

(a) distribuční funkci F_X,

(b) střední hodnotu $E(X)$ a rozptyl $D(X)$,

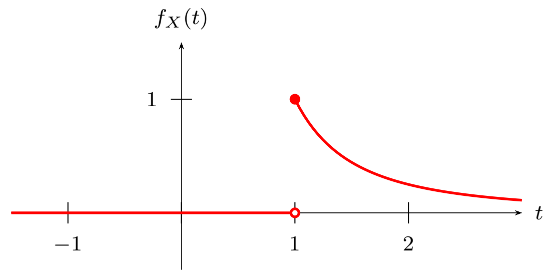
(c) hustotu pravděpodobnosti f_Y,

(d) kvantilové funkce q_X a q_Y.

Řešení:
Můžeme si zkontrolovat, že f je skutečně hustota, tj. že f je nezáporná funkce a platí, že $\int_{-\infty}^{\infty} f_X(t)\ dt = 1$:

$$\int_{-\infty}^{\infty} f_X(t)\ dt = \int_1^{\infty} t^{-2}\ dt = \Big[-\frac{1}{t}\Big]_1^{\infty} = 1 \ .$$

Graf f_X je



(a) Distribuční funkce F_X je rovna

$$F_X(u) = \int_{-\infty}^{u} f_X(t)\ dt$$

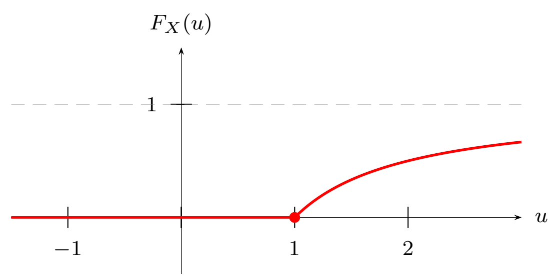
a pro $u \geq 1$ tedy máme

$$F_X(u) = \int_1^u \tfrac{1}{t^2}\ dt = \Big[-\frac{1}{t} \Big]_1^u = 1 - \tfrac{1}{u}$$

celkově pak

$$F_X(u) = \begin{cases} 0 & ,\ u \leq 1 \\ 1 - \frac{1}{u} & ,\ u \geq 1 \end{cases}$$

s grafem



(b) Střední hodnotu a rozptyl určíme nejlépe z hustoty pravděpodobnosti:

$$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)\ dt = \int_1^{\infty} \tfrac{1}{t}\ dt = \Big[\ln t \Big]_1^{\infty} = +\infty$$

Tedy $E(X) = +\infty$, a protože rozptyl je definován jako $D(X) = E\Big(\big(X - E(X)\big)^2\Big)$, dostáváme prostě, že rozptyl neexistuje.

(c) Hustotu f_Y si spočítáme pomocí distribuční funkce F_Y (ale lze to určit i přímo z f_X - viz níže). Protože $P\big(X \leq 0\big) = F_X(0) = 0$ tak speciálně pro $t \leq 0$ máme

$$F_Y(t) \leq F_Y(0) = P\left(\tfrac{1}{X} \leq 0\right) = P\big(X < 0\big) = 0\ .$$

Poznámka: Pokud $P(A) = 0$ pro nějaký jev A, pak dostaneme ihned $P(B) = P(B \cap \overline{A}) + P(B \cap A) = P(B \cap \overline{A})$, protože $0 \leq P(B \cap A) \leq P(A) = 0$.

Pro $t > 0$ teď, opět díky $P\big(X \leq 0\big) = 0$, můžeme psát

$$F_Y(t) = P\left(\tfrac{1}{X} \leq t\right) = P\Big(\big(\tfrac{1}{X} \leq t\big) \cap (X > 0)\Big) = P\left(\tfrac{1}{t} \leq X\right) = 1 - P(X < \tfrac{1}{t}) = 1 - F_X\left(\tfrac{1}{t}\right)$$

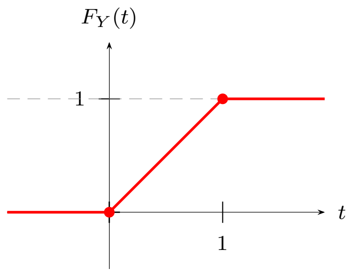
Po dosazení $u = \frac{1}{t}$ pro $t > 0$ do F_X tak dostáváme

$$F_Y(t) = 1 - F_X\left(\tfrac{1}{t}\right) = \begin{cases} 1 - 0 = 1 & ,\ \frac{1}{t} \leq 1\ \wedge\ t > 0 \\ 1 - \Big(1 - \frac{1}{t^{-1}}\Big) = t & ,\ \frac{1}{t} \geq 1\ \wedge\ t > 0 \end{cases}$$

Celkem tedy dostáváme

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

s grafem



což znamená, že Y má rovnoměrné rozdělení na intervalu $(0,1)$, tj.

$$f_Y(t) = F_Y'(t) = \begin{cases} 1 & , t \in (0,1) \\ 0 & , \text{jinak.} \end{cases}$$

K hustotě f_Y se lze bez přímého výpočtu F_Y dostat pomocí vzorce:

$$f_{h(X)}(t) = -\frac{f_X\Big(h^{-1}(t)\Big)}{h'\Big(h^{-1}(t)\Big)}$$

kde h je ostře *klesající* diferencovatelná funkce a t je z definičního oboru h^{-1}.

V našem případě máme $h(x) = \frac{1}{x}$ s definičním oborem $D(h) = (0, +\infty)$. Její inverze h^{-1} má definiční obor $D(h^{-1}) = (0, +\infty)$. Pro $Y = h(X)$ dosazením opět máme

$$f_Y(t) = -\frac{f_X\Big(h^{-1}(t)\Big)}{h'\Big(h^{-1}(t)\Big)} = -\frac{f_X(t^{-1})}{h'(t^{-1})} = \Big\{h'(x) = -\tfrac{1}{x^2}\Big\} = \frac{f_X(t^{-1})}{t^2} = \begin{cases} 0 & , \ 0 < t^{-1} < 1 \ \Leftrightarrow \ t > 1, \\ \frac{t^2}{t^2} = 1 & , \ t^{-1} > 1 \ \Leftrightarrow \ t \in (0,1). \end{cases}$$

Odvození vzorce:
Pro $t \in D(h^{-1})$ a $Y = h(X)$ teď máme

$$F_Y(t) = P(h(X) \le t) = P(X \ge h^{-1}(t)) = 1 - P(X < h^{-1}(t)) = 1 - F_X\Big(h^{-1}(t)\Big)\ .$$

Derivaci tudíž dostáváme

$$f_Y(t) = F_Y'(t) = \left(1 - F_X\Big(h^{-1}(t)\Big)\right)' = \{\text{derivace složené a inverzní funkce}\} = -\frac{F_X'\Big(h^{-1}(t)\Big)}{h'\Big(h^{-1}(t)\Big)} = -\frac{f_X\Big(h^{-1}(t)\Big)}{h'\Big(h^{-1}(t)\Big)}\ .$$

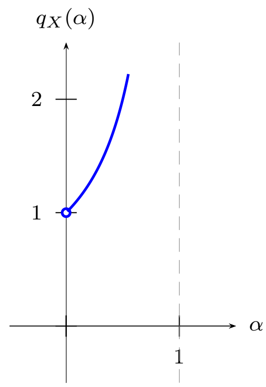
(d) Kvantil q_X je inverzní funkcí k ostře rostoucí funkci
$F_X(t) = 1 - \frac{1}{t}$ pro $t \in (1, +\infty)$. Tedy

$$\alpha = 1 - \tfrac{1}{t} \quad \Leftrightarrow \quad t = \tfrac{1}{1-\alpha}$$

a tudíž

$$q_X(\alpha) = \tfrac{1}{1-\alpha} \quad \text{pro} \quad \alpha \in (0,1)$$

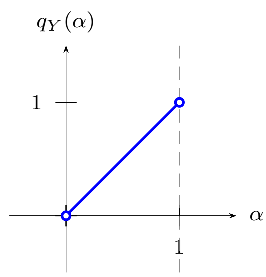
s grafem



Kvantil q_Y je inverzní funkcí k ostře rostoucí funkci
$F_Y(t) = t$ pro $t \in (0,1)$. Tedy

$$q_Y(\alpha) = \alpha \quad \text{pro} \quad \alpha \in (0,1)$$

s grafem



Opět můžeme q_Y získat z q_X pomocí ostře *klesající* spojité funkce h jako

$$q_{h(X)}(\alpha) = h(q_X(1-\alpha))$$

tedy po dosazení

$$q_Y(\alpha) = h(q_X(1-\alpha)) = \frac{1}{q_X(1-\alpha)} = \frac{1}{\alpha^{-1}} = \alpha\ .$$

6.5 (transformace spojité veličiny, kvantil)

Nezáporná náhodná veličina X má hustotu

$$f_X(t) = \begin{cases} 0 & , \ t < 1 \\ \frac{2}{t^3} & , \ t \geq 1 \end{cases}$$

Uvažujme veličinu $Y = \frac{1}{X+1}$. Určete

(a) distribuční funkci F_Y,

(b) hustotu pravděpodobnosti f_Y,

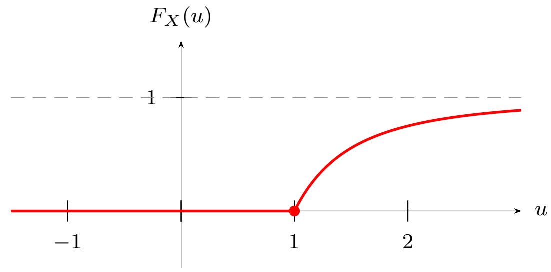
(c) kvantilové funkce q_X a q_Y.

Řešení:

Z příkladu **5.6** už máme distribuční funkci

$$F_X(u) = \begin{cases} 0 & , \ u \leq 1 \\ 1 - \frac{1}{u^2} & , \ u \geq 1 \end{cases}$$

s grafem



(a) Protože X je nezáporná spojitá veličina, tak $Y = \frac{1}{X+1} > 0$ a pro $t \leq 0$ tudíž je

$$F_Y(t) = P\left(\frac{1}{X+1} \leq t\right) = 0\ .$$

,

Pro $t > 0$ pak máme

$$F_Y(t) = P\Big(\frac{1}{X+1} \leq t\Big) = P\Big(\frac{1}{t} \leq X + 1\Big) = P\Big(\frac{1}{t} - 1 \leq X\Big) =$$
$$= 1 - P\Big(X < \frac{1}{t} - 1\Big) = 1 - F_X\left(\frac{1}{t} - 1\right)\ .$$

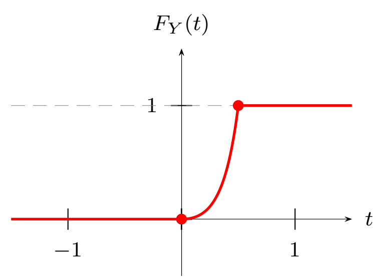
Dosazením $u = \frac{1}{t} - 1$ pro $t > 0$ do F_X tak dostáváme

$$F_Y(t) = 1 - F_X\left(\frac{1-t}{t}\right) = \begin{cases} 1 - \left(1 - \frac{1}{\left(\frac{1-t}{t}\right)^2}\right) = \frac{t^2}{(1-t)^2} & , \ \frac{1}{t} - 1 \geq 1 \ \wedge \ t > 0 \\ 1 - 0 = 1 & , \ \frac{1}{t} - 1 \leq 1 \ \wedge \ t > 0 \end{cases}$$

Celkově pak máme

$$F_Y(t) = \begin{cases} 0 & , \ t \leq 0 \\ \frac{t^2}{(1-t)^2} & , \ t \in (0, \frac{1}{2}\rangle \\ 1 & , \ t \geq \frac{1}{2}. \end{cases}$$

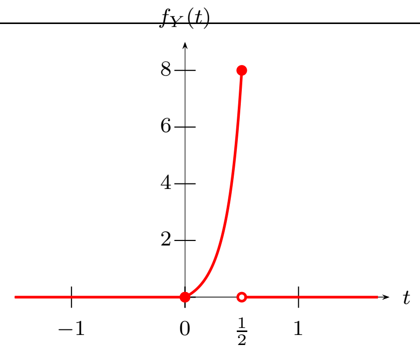
s grafem



(b) Distribuční funkce F_Y je opět (absolutně) spojitá.
Hustotu pravděpodobnosti f_Y veličiny Y dostaneme jako

$$f_Y(t) = F_Y'(t) = \begin{cases} \frac{2t}{(1-t)^3} & , \ t \in \langle 0, \frac{1}{2}\rangle \\ 0 & , \ \text{jinak}\ . \end{cases}$$

Její graf je:



Při výpočtu f_Y jsme také mohli postupovat podle vzorce z příkladu **6.4**:

$$f_{h(X)}(t) = -\frac{f_X\Big(h^{-1}(t)\Big)}{h'\Big(h^{-1}(t)\Big)}$$

pro $h(x) = \frac{1}{1+x}$ ostře *klesající* diferencovatelnou funkci na $D(h) = (0, +\infty)$ a $t \in D(h^{-1}) = (0, 1)$. Nyní je $h^{-1}(t) = \frac{1}{t} - 1$ a dosazením opět máme

$$f_Y(t) = -\frac{f_X(\frac{1}{t} - 1)}{h'(\frac{1}{t} - 1)} = \Big\{h'(x) = -\frac{1}{(1+x)^2}\Big\} =$$

$$= \frac{f_X\left(\frac{1-t}{t}\right)}{t^2} = \begin{cases} 0 & , \frac{1}{t} - 1 < 1 \ \wedge \ t \in (0,1) \ \Leftrightarrow \ t \in \left(\frac{1}{2}, 1\right), \\ \frac{2}{t^2\left(\frac{1-t}{t}\right)^3} = \frac{2t}{(1-t)^3} & , \frac{1}{t} - 1 > 1 \ \wedge \ t \in (0,1) \ \Leftrightarrow \ t \in \left(0, \frac{1}{2}\right). \end{cases}$$

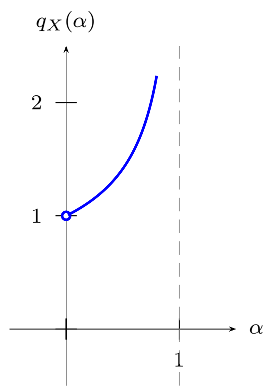
(c) Kvantil q_X je inverzní funkcí k ostře rostoucí funkci
$F_X(t) = 1 - \frac{1}{t^2}$ pro $t \in (1, +\infty)$. Tedy

$$\alpha = 1 - \frac{1}{t^2} \quad \Leftrightarrow \quad t = \frac{1}{\sqrt{1-\alpha}}$$

a tudíž

$$q_X(\alpha) = \frac{1}{\sqrt{1-\alpha}} \quad \text{pro} \quad \alpha \in (0, 1)$$

s grafem



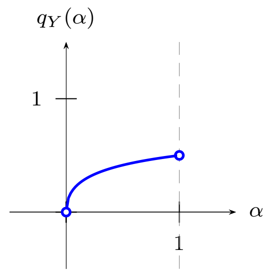
Kvantil q_Y je inverzní funkcí k ostře rostoucí funkci
$F_Y(u) = \frac{u^2}{(1-u)^2}$ pro $u \in (0, \frac{1}{2})$. Tedy

$$\alpha = \frac{u^2}{(1-u)^2} \quad \Leftrightarrow \quad \sqrt{\alpha} = \frac{u}{1-u} = \frac{1}{1-u} - 1 \quad \Leftrightarrow \quad u = 1 - \frac{1}{1+\sqrt{\alpha}} = \frac{\sqrt{\alpha}}{1+\sqrt{\alpha}}$$

a tudíž

$$q_Y(\alpha) = \tfrac{\sqrt{\alpha}}{1+\sqrt{\alpha}} \quad \text{pro} \quad \alpha \in (0,1)$$

s grafem



Opět můžeme q_Y získat z q_X pomocí ostře *klesající* spojité funkce $h(x) = \frac{1}{1+x}$ jako

$$q_{h(X)}(\alpha) = h(q_X(1-\alpha))$$

tedy po dosazení

$$q_Y(\alpha) = h(q_X(1-\alpha)) = \frac{1}{1+q_X(1-\alpha)} = \frac{1}{1+\frac{1}{\sqrt{\alpha}}} = \tfrac{\sqrt{\alpha}}{1+\sqrt{\alpha}}\ .$$

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

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

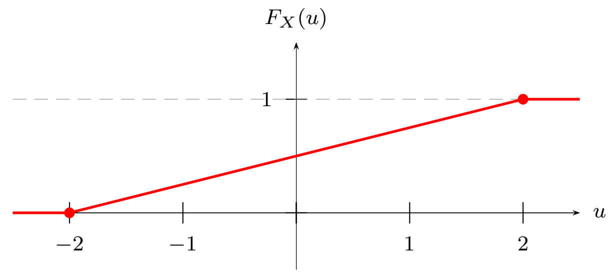
Pro $t \in \langle 0, 1)$ máme

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

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:

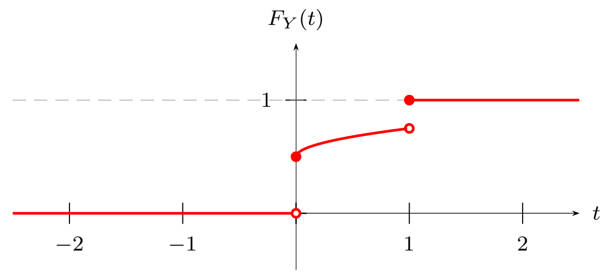


Tedy

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

a po dosazení $u = \sqrt{t}$ pro $t \in \langle 0, 1)$ dostaneme:

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

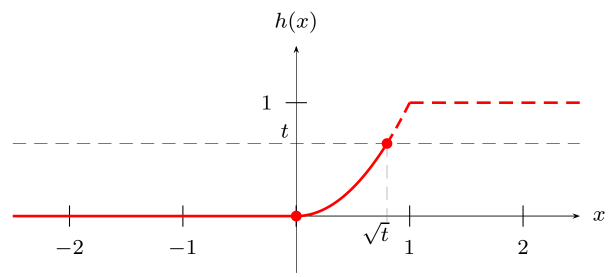


Veličina $Y = h(X)$ tedy nemá spojité rozdělení ale smíšené, přestože veličina X je (absolutně) spojitá a i funkce h je spojitá.

(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 výše.