5. cvičení z STP

16. - 20. březen 2020

Poznámky k binomickému rozdělení: Mějme n nezávislých opakování daného pokusu, jehož úspěšnost je $0 < p < 1$. Veličina

$$X = \text{“\textit{počet úspěchů během} } n \text{ \textit{pokusů}”}$$

s hodnotami $X \in \{0, 1, \ldots, n\}$ má pak tzv. *binomické* rozdělení $\mathrm{Bi}(n, p)$. Pro $k = 0, 1, \ldots, n$ pak je

$$P(X = k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

což je jeden z členů v binomické větě (odtud také ten název): $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \cdot b^{n-k}$

Důvod, proč tu máme kombinační číslo, je v tom, že daných k úspěchů je rozmístěno mezi n pokusů právě $\binom{n}{k}$ způsoby.

Odvození: Vezmeme si jevy

$$A_i = \text{“\textit{úspěch v} } i\text{\textit{-tém pokusu}”}$$

Ty jsou nezávislé a mají pravděpodobnost $P(A_i) = p$. Pak jev

$$B_k = \text{“\textit{nastane právě} } k \text{ \textit{úspěchů v} } n \text{ \textit{pokusech}”}$$

vyjádříme jako sjednocení disjunktních jevů (ty v hranaté závorce)

$$B_k = \bigcup_{\substack{K \subseteq \{1,\ldots,n\} \\ |K|=k}} \left[\left(\bigcap_{i \in K} A_i \right) \cap \left(\bigcap_{j \in \{1,\ldots,n\} \setminus K} A_j^c \right) \right]$$

odkud máme ihned

$$P(X = k) = P(B_k) = \sum_{\substack{K \subseteq \{1,\ldots,n\} \\ |K|=k}} \left[\left(\prod_{i \in K} P(A_i) \right) \cdot \left(\prod_{j \in \{1,\ldots,n\} \setminus K} P(A_j^c) \right) \right] =$$

$$= \sum_{\substack{K \subseteq \{1,\ldots,n\} \\ |K|=k}} p^k \cdot (1-p)^{n-k} = \binom{n}{k} p^k \cdot (1-p)^{n-k} .$$

Příklad 5.1 *Pravděpodobnost výhry hráče A nad hráčem B je* 0.7*. Jaká je pravděpodobnost, že během deseti po sobě jdoucích zápasů*

(a) alespoň jednou vyhrál B,

(b) maximálně dvakrát vyhrál A?

Řešení:
Veličina

$$X = \text{“\textit{počet výher A během} 10 \textit{zápasů}”}$$

má binomické rozdělení $X \sim \mathrm{Bi}(10,\ 0.7)$.

(a) Máme

$$P(\text{“\textit{alespoň jednou vyhrál B}”}) = P(X < 10) = 1 - P(X \geq 10) =$$

$$= 1 - P(X = 10) = 1 - (0.7)^{10} \doteq 0.972 .$$

(b) Máme

$$P(\text{“}\textit{maximálně dvakrát vyhrál A}\text{”}) = P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) =$$

$$= (0.3)^{10} + \underbrace{\binom{10}{1}}_{=10} 0.7 \cdot (0.3)^9 + \underbrace{\binom{10}{2}}_{=45} (0.7)^2 \cdot (0.3)^8 \doteq 0.0016 \ .$$

Poznámka: Jestliže pro veličinu X provedeme n měření, ve kterých se budou vyskytovat číselné hodnoty $a_1, \dots, a_k$ (ne nutně navzájem různé), kde každá hodnota a_i se bude opakovat n_i-krát, pak za jejich průměrnou hodnotu $\overline{\boldsymbol{x}}$ (v rámci tohoto našeho měření) budeme považovat jejich aritmetický průměr (kde se započítá i opakování dané hodnoty a_i), tj.

$$\overline{\boldsymbol{x}} = \frac{\sum_i n_i \cdot a_i}{\sum_i n_i} = \sum_i a_i \cdot \frac{n_i}{n}$$

kde $n = \sum_i n_i$ je celkový počet měření. (Ve statistice se tato hodnota $\overline{\boldsymbol{x}}$ nazývá *výběrový průměr* - viz později). Výraz $\frac{n_i}{n}$ přitom vyjadřuje podíl výskytu hodnoty a_i při daném počtu měření. Uvědomme si ještě, že k výpočtu $\overline{\boldsymbol{x}}$ musíme mít k dispozici konkrétní soubor naměřených hodnot!

Tento intuitivní přístup vede přirozeně k definici *střední hodnoty* $E(X)$ pro *diskrétní* veličinu X (tj. pro takovou veličinu X, že existuje *nejvýše spočetná* množina $A \subseteq \mathbb{R}$, že $P(X \in A) = 1$) v podobě

$$E(X) = \sum_{k \in A} k \cdot P(X = k)$$

samozřejmě pokud suma vůbec existuje. Spočetnost nebo konečnost množiny A tu je ovšem podstatná - jestliže sčítáme nespočetně mnoho nenulových čísel, pak tento součet buď neexistuje nebo bude vždy nekonečný. Tedy šanci na to, abychom dostali konečně číslo mají jen nejvýše spočetně součty nenulových čísel. No a protože pro hodnoty $k \in \mathbb{R} \setminus A$ je vždy $P(X = k) = 0$ (pro naši diskrétní veličinu X), tak můžeme použít i vyjádření ve tvaru:

$$E(X) = \sum_{k \in \mathbb{R}} k \cdot P(X = k) \ .$$

Poznámka: Pro veličinu s tzv. spojitým rozdělením (viz později) se střední hodnota definuje sice obdobně, ale pomocí pojmu hustoty pravděpodobnosti. Tento jistý nesoulad v přístupech k různým typům veličin se ale dá elegantně odstranit tím, že lze ukázat, že ve všech případech nakonec bude platit následující vztah

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

Připomenutí: Rozptyl $\mathrm{var}(X)$ (libovolné) náhodné veličiny X určuje, jak moc se hodnoty odchylují od střední hodnoty. Rozptyl je proto definován jako střední hodnota veličiny $\big(X - E(X)\big)^2$, která vyjadřuje kvadrát odchylky X od své střední hodnoty, tedy předpisem:

$$\mathrm{var}(X) = E\Big(\big(X - E(X)\big)^2\Big) = \cdots = E\big(X^2\big) - \Big(E\big(X\big)\Big)^2$$

(ovšem opět jen pokud uvedené střední hodnoty existují.)

Věta: Pro (borelovsky) měřitelnou funkci $h : \mathbb{R} \to \mathbb{R}$ (v praxi to je téměř libovolná funkce, např. po částech spojitá) a diskrétní veličinu X platí, že

$$E\big(h(X)\big) = \sum_{k \in \mathbb{R}} h(k) \cdot P(X = k) \; .$$

Příklad 5.2 *Z urny, v níž jsou bílé a černé koule v poměru $9 : 1$, vybíráme třikrát po jedné kouli, kterou vždy zase vracíme zpět. Náhodná veličina X je počet vybraných černých koulí. Určete:*

(a) typ rozdělení veličiny X a distribuční funkci F_X náhodné veličiny X.

(b) pravděpodobnost, že vytáhneme alespoň jednu černou kouli.

(c) průměrný počet vybraných černých koulí a rozptyl $\operatorname{var}(X)$ *veličiny X.*

Řešení:
Náhodná veličina X měří *počet úspěšných pokusů (vybrání černé koule) během n pokusů.* Zde je $n = 3$ a pravděpodobnost každého (z nezávislých) pokusů je $p = 0.1$.

(a) Veličina X má tedy obor hodnot $\{0, 1, 2, 3\}$ a (diskrétní) binomické rozdělení $\text{Bi}(n, p)$ s pravděpodobnostmi

$$P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & , \ k \in \{0, 1, 2, 3\} \\ 0 & , \ \text{jinak.} \end{cases}$$

Konkrétně:

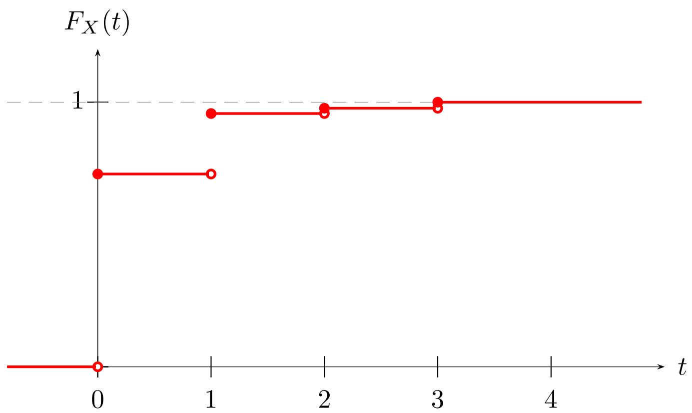
k	0	1	2	3
$P(X = k)$	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3
$p = 0.1$	0.729	0.243	0.027	0.001

Distribuční funkce

$$F_X(t) = P(X \leq t) = \sum_{k \leq t} P(X = k)$$

pro diskrétní veličinu X má hodnoty

$$F_X(t) = \begin{cases} 0 & , t < 0 \\ P(X = 0) = 0.729 & , t \in \langle 0, 1) \\ P(X = 0) + P(X = 1) = 0.972 & , t \in \langle 1, 2) \\ P(X = 0) + P(X = 1) + P(X = 2) = 0.999 & , t \in \langle 2, 3) \\ 1 & , t \geq 3. \end{cases}$$



(b) Pravděpodobnost jevu

$$B = \text{"}\textit{vytáhneme alespoň jednu černou kouli}\text{"}$$

je

$$P(B) = P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - 0.729 = 0.271 \ .$$

(c) Průměrný počet vybraných černých koulí je střední hodnota naší diskrétní veličiny, tedy

$$E(X) = \sum_{k \in \mathbb{R}} k \cdot P(X = k) = 0 \cdot 0.729 + 1 \cdot 0.243 + 2 \cdot 0.027 + 3 \cdot 0.001 = 0.3$$

nebo podle vzorce pro binomické rozdělení $\mathrm{Bi}(n, p)$:

$$E(X) = np = 3 \cdot 0.1 = 0.3 \ .$$

Pro rozptyl $\mathrm{var}(X) = E(X^2) - \big(E(X)\big)^2$ potřebujeme spočítat ještě

$$E(X^2) = \sum_{k \in \mathbb{R}} k^2 \cdot P(X = k) = 0^2 \cdot 0.729 + 1^2 \cdot 0.243 + 2^2 \cdot 0.027 + 3^2 \cdot 0.001 = 0.36$$

tedy

$$\mathrm{var}(X) = E(X^2) - \big(E(X)\big)^2 = 0.36 - (0.3)^2 = 0.27$$

nebo podle vzorce pro binomické rozdělení $\mathrm{Bi}(n, p)$:

$$\mathrm{var}(X) = np(1 - p) = 3 \cdot 0.1 \cdot 0.9 = 0.27 \ .$$

Připomeňme si **definici nezávislosti veličin:** Veličiny $X_1, \ldots, X_n$ jsou *nezávislé* $\Leftrightarrow$ pro libovolné intervaly $I_1, \ldots, I_n \subseteq \mathbb{R}$ platí, že

$$P(X_1 \in I_1, \ldots, \ X_n \in I_n) = \prod_{i=1}^{n} P(X_i \in I_i)$$

konkrétně se lze (ekvivalentně) omezit jen na určité intervaly a pak máme:

Veličiny $X_1, \ldots, X_n$ jsou *nezávislé* $\Leftrightarrow$ pro libovolné $a_1, \ldots, a_n \in \mathbb{R}$ platí, že

$$P(X_1 \leq a_1, \ldots, \ X_n \leq a_n) = \prod_{i=1}^{n} P(X_i \leq a_i)$$

Pro rozptyl jejich součtu pak platí:
Nechť $X_1, \ldots, X_n$ jsou nezávislé veličiny, pak

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{var}(X_i)$$

- -

Důležité pozorování: Veličina s binomickým rozdělením $\mathrm{Bi}(n, p)$ je součtem n nezávislých veličin X_i, které mají alternativní rozdělení $\mathrm{Alt}(p)$ a kde X_i představuje počet úspěchů (tj. 0 nebo 1) v i-tém pokusu. Více viz následující příklad. Díky tomuto si lze celkem snadno vždy rychle odvodit střední hodnotu i rozptyl binomického rozdělení (a to právě s pomocí vzorců uvedených výše). Více opět viz následující příklad.

Příklad 5.3 *Pravděpodobnost, že atlet v oddíle skočí do dálky přes* $5\,m$*, je* 0.7*. V oddíle je* 6 *atletů.*

(a) *Určete rozdělení náhodné veličiny*

$$X = \begin{cases} 1, & \text{\textit{atlet skočil přes} } 5\,m \\ 0, & \text{\textit{atlet neskočil přes} } 5\,m, \end{cases}$$

její střední hodnotu $E(X)$ *a rozptyl* $\mathrm{var}(X)$*.*

(b) *Určete rozdělení náhodné veličiny*

$$Y = \text{“ \textit{počet atletů v oddíle, kteří skočili přes} } 5\,m \text{ ”}$$

její střední hodnotu $E(Y)$ *a rozptyl* $\mathrm{var}(Y)$*.*

(c) *Jaká je pravděpodobnost, že přes* $5\,m$ *skočí v oddíle alespoň* 4 *atleti?*

Řešení:

(a) Veličina X má *alternativní* rozdělení, tj. $X \sim \mathrm{Alt}(p)$, kde $p = P(X = 1) = 0.7$ je pravděpodobnost úspěšného pokusu a $P(X = 0) = 1 - p = 0.3$.

A dále máme

$$E(X) = \sum_{i \in \mathbb{R}} i \cdot P(X = i) = 0 \cdot (1 - p) + 1 \cdot p = p = 0.7$$

$$E(X^2) = \sum_{i \in \mathbb{R}} i^2 \cdot P(X = i) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p = 0.7$$

a

$$\mathrm{var}(X) = E(X^2) - \big(E(X)\big)^2 = p - p^2 = p(1 - p) = 0.7 \cdot 0.3 = 0.21\ .$$

(b) Veličina Y představuje počet úspěchů při $n = 6$ nezávislých pokusech, s pravděpodobností úspěchu $p = 0.7$, takže Y má *binomické rozdělení* $\mathrm{Binom}(n, p)$. Hodnoty veličiny Y jsou $k = 0, 1, \ldots, n$ a jejich pravděpodobnosti jsou

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k} = \binom{6}{k} 0.7^k \cdot 0.3^{6-k}\ .$$

Pro další výpočty se hodí všimnout si, že $Y = \sum_{i=1}^{n} X_i$ kde

$$X_i = \begin{cases} 1, & i\text{-tý atlet skočil přes } 5\,\mathrm{m} \\ 0, & i\text{-tý atlet neskočil přes } 5\,\mathrm{m}, \end{cases}$$

jsou navzájem nezávislé náhodné veličiny a $X_i \sim \mathrm{Alt}(p)$.

Pro střední hodnotu veličiny Y pak máme

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \underbrace{E(X_i)}_{p} = n \cdot p = 6 \cdot 0.7 = 4.2$$

a z nezávislosti X_i pak pro rozptyl máme

$$\mathrm{var}(Y) = \sum_{i=1}^{n} \underbrace{\mathrm{var}(X_i)}_{p(1-p)} = np(1 - p) = 6 \cdot 0.7 \cdot 0.3 = 1.26\ .$$

(c)

$$P(Y \geq 4) = \sum_{k=4}^{6} \binom{6}{k} 0.7^k \cdot 0.3^{6-k} = 15 \cdot 0.7^4 \cdot 0.3^2 + 6 \cdot 0.7^5 \cdot 0.3^1 + 1 \cdot 0.7^6 \cdot 1 =$$
$$= 0.324135 + 0.302526 + 0.117649 = 0.74431 \ .$$

Připomenutí: Veličina

$$X = \text{“\textit{počet neúspěchů než nastane první úspěch}”}$$

má geometrické rozdělení Geom(p), pokud můžeme opakovat libovolné množství nezávislých pokusů, které mají všechny stejnou pravděpodobnost úspěchu p. Příkladem je třeba situace, že se chceme trefit míčem do koše apod.

Hodnoty veličiny X jsou $\{0, 1, 2, \dots\}$. Pro odvození rozdělení X si pro $i = 1, 2, 3 \dots$ označme jevy

$$A_i = \text{“\textit{i-tý pokus je úspěšný}”}$$

které budou nezávislé s budou mít pravděpodobnosti $P(A_i) = p$. Pak máme pravděpodobnosti

$$P(X = k) = P(A_1^c \cap \dots \cap A_k^c \cap A_{k+1}) = P(A_1^c) \cdots P(A_k^c) \cdot P(A_{k+1}) = (1-p)^k p$$

pro $k = 0, 1, 2, \dots$ Pro střední hodnotu a rozptyl pak máme:

$$E(X) = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=0}^{\infty} kp(1-p)^k = \dots = \tfrac{1-p}{p}$$

$$\mathrm{var}(X) = \tfrac{1-p}{p^2}$$

což si můžeme lépe zapamatovat jako

$$\mathrm{var}(X) = \frac{1}{p} \cdot E(X) \ .$$

Příklad 5.4 *Pravděpodobnost narození chlapce je* 0.51. *Jaká je pravděpodobnost, že v dané porodnici dnes bylo nejpozději (v časovém pořadí) čtvrté narozené dítě holka?*

Řešení:
Lze použít dva přístupy:

(a) Vezmeme si náhodnou veličinu

$$X = \text{“ \textit{počet narozených chlapců před první narozenou holkou.}”}$$

Ta má *geometrické* rozdělení Geom(p) s pravděpodobností $p = 1 - 0.51 = 0.49$ úspěšného pokusu (tj. narození holky). Hodnoty veličiny X jsou $k = 0, 1, 2, \dots$ s pravděpodobnostmi

$$P(X = k) = (1-p)^k \cdot p = 0.51^k \cdot 0.49$$

Pro jednodušší výpočet si ještě označme $q = 1 - p = 0.51$. Hledaná pravděpodobnost je tedy

$$P(X < 4) = \sum_{k=0}^{3} q^k \cdot (1-q) = (1-q) \cdot \underbrace{(1 + q + \dots + q^3)}_{\frac{1-q^4}{1-q}} = 1 - q^4 = 1 - 0.51^4 \doteq 0.9323.$$

(b) Vezmeme si náhodnou veličinu

$$Y = \text{“ } \textit{počet narozených holek mezi prvními 4 narozenými dětmi.”}$$

Ta má binomické rozdělení $\mathrm{Bi}(4, p)$, kde je opět $p = 0.49$. Hledaná pravděpodobnost je tedy

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{4}{0} p^0 \cdot (1-p)^4 = 1 - q^4 = 1 - 0.51^4.$$

Příklad 5.5 *Při hodu na koš se trefíme s pravděpodobností $p = 0.2$. Náhodná veličina X je počet hodů, než se trefíme.*

(a) Určete rozdělení veličiny X.

(b) V kterém kole se průměrně poprvé trefíme?

(c) Kolikrát musíme nejméně hodit, abychom se s pravděpodobností alespoň 90% alespoň jednou (v rámci těchto hodů) trefili?

Řešení:

(a) Veličina X počítá počet neúspěšných kol. Má tedy geometrické rozdělení $\mathrm{Geom}(0.2)$ s oborem hodnot $\{0, 1, 2, \dots\}$ a jejich pravděpodobnostmi

$$P(X = k) = p(1-p)^k = 0.2 \cdot 0.8^k$$

pro $k = 0, 1, 2, \dots$.

POZOR: Nepleťte si *geometrickou pravděpodobnost*, což je způsob počítání pravděpodobnosti jevu (název je odvozen od geometrických obrazců) a *geometrické rozdělení*, které zase přísluší náhodné veličině (název je odvozen od geometrické posloupnosti, kterou tvoří hodnoty pravděpodobnostní funkce dané veličiny).

(b) Hledáme střední hodnotu veličiny

$$Y = \text{“}\textit{pořadí hodu, při kterém se poprvé trefíme}\text{”} .$$

Zřejmě je $Y = X + 1$ a tedy

$$E(Y) = E(X+1) = E(X) + 1 = \frac{1-p}{p} + 1 = \frac{1}{p} = \frac{1}{0.2} = 5 .$$

(c) Abychom si lépe představili situaci, předpokládejme, že i po té, co se trefíme, pokračujeme v házení (a veličina Y zaznamená pouze ten první úspěšný pokus). Ptáme se tedy, jaký je nejmenší počet hodů $n \in \mathbb{N}$, abychom se v jejich průběhu alespoň jednou trefili. Chceme tudíž znát nejmenší $n \in \mathbb{N}$ takové, že $P\Big(Y \leq n\Big) \geq 0.9$, neboli

$$0.9 \leq P\Big(X + 1 \leq n\Big) = P\Big(X \leq n - 1\Big) = F_X(n-1) .$$

K tomu potřebujeme tudíž znát distribuční funkci X pro $k \in \mathbb{N}_0$:

$$F_X(k) = \sum_{i \leq k} P(X = i) = \sum_{i=0}^{k} p(1-p)^i = p \cdot \frac{1 - (1-p)^{k+1}}{1 - (1-p)} = 1 - (1-p)^{k+1} .$$

Po dosazení tedy dostaneme podmínku

$$0.9 \leq F_X(n-1) = 1 - (1-p)^n$$

neboli

$$\begin{aligned} 0.1 &\geq (1-p)^n = (1-0.2)^n = 0.8^n \\ \log 0.1 &\geq n \log 0.8 \\ n &\geq \tfrac{\log 0.1}{\log 0.8} \doteq 10.32 \end{aligned}$$

Pozor: logaritmus hodnoty 0.8 je záporný! Tedy musíme hodit alespoň $n = 11$ -krát.

\- -

Úlohu (c) můžeme vyřešit i s pomocí binomického rozdělení. Pro $n \in \mathbb{N}$ uvažujme veličinu

$$Z_n = \text{“}\textit{počet tolika hodů (z } n \textit{ možných), ve kterých se trefíme}\text{”}$$

která zřejmě má binomické rozdělení $\mathrm{Bi}(n,p)$. Hledáme nyní nejmenší $n \in \mathbb{N}$ tak, aby

$$P(Z_n \geq 1) \geq 0.9 \ .$$

Máme tedy

$$0.9 \leq P(Z_n \geq 1) = 1 - P(Z_n = 0) = 1 - (1-p)^n$$

což je stejná nerovnost jako výše a tím dostaneme i stejné řešení.

Příklad 5.6 *Revizor najde v dané tramvaji alespoň jednoho černého pasažéra s pravděpodobností p. Náhodná veličina X je počet tramvají, které revizor projde před tím, než najde černého pasažéra.*

(a) Určete rozdělení veličiny X a její distribuční funkci F_X.

(b) Kolik ***dalších*** *tramvají musí revizor průměrně projít než opět nalezne dalšího černého pasažéra? Jaký je rozptyl veličiny X? Spočítejte obecně a pak pro $p = 26\%$.*

(c) Jestliže je $p = 26\%$, kolik musí revizor zkontrolovat minimálně tramvají, aby s pravděpodobností alespoň 95% našel alespoň jednoho černého pasažéra?

Řešení:

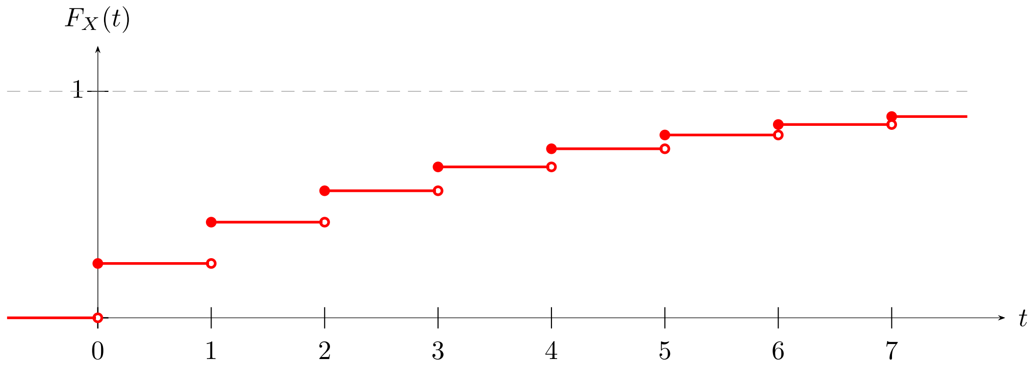
(a) Veličina X počítá počet neúspěchů (nenalezení černého pasažéra) než nastane úspěch. Má tedy (viz výše) geometrického rozdělení s oborem hodnot $\{0, 1, 2, \dots\}$ a pravděpodobnostmi

$$P(X = k) = \begin{cases} p(1-p)^k & , k \in \mathbb{N}_0 \\ 0 & , \text{jinak} \ . \end{cases}$$

Distribuční funkce je po částech konstantní. Pro $t < 0$ je $F_X(t) = 0$ a pro $t > 0$ máme

$$F_X(t) = \sum_{k \leq t} P(X = k) = \sum_{k=0}^{[t]} p(1-p)^k = p \cdot \frac{1 - (1-p)^{[t]+1}}{1 - (1-p)} = 1 - (1-p)^{[t]+1}$$

kde $[t]$ je celá část t (tj. zaokrouhlení desetinného čísla t směrem dolů). Pro $p = 0.26$ je graf F_X naznačen níže.



(b) Ptáme se na střední hodnotu veličiny X (tj. otázku zde chápeme jako střední počet tramvají *předtím* než se najde černý pasažér. Pokud bychom chtěli započítat i tu tramvaj, kde buče černý pasažér, vezmeme veličinu $X+1$).

Tedy

$$E(X) = \frac{1-p}{p} = \frac{0.74}{0.26} \doteq 2.85$$

$$\mathrm{var}(X) = \frac{1-p}{p^2} = \frac{0.74}{0.26^2} \doteq 10.95$$

(c) Pro $n \in \mathbb{N}$ zřejmě máme, že

"revizor v prvních n tramvajích najde alespoň jednoho černého pasažéra"

právě když platí

$$X \leq n-1 \ .$$

Chceme tedy znát nejmenší $n_0 \in \mathbb{N}$ takové, že $P\Big(X \leq n_0 - 1\Big) \geq 0.95$. Takže

$$0.95 \leq P\big(X \leq n_0 - 1\big) = F_X(n_0 - 1) = 1 - (1-p)^{n_0}$$

neboli

$$0.05 \geq (1-p)^{n_0} = (1-0.26)^{n_0} = 0.74^{n_0}$$

$$\log 0.05 \geq n_0 \cdot \log 0.74$$

$$9.95 \doteq \frac{\log 0.05}{\log 0.74} \leq n_0$$

Pozor, logaritmus je záporný pro hodnoty menší než 1! Revizor tedy musí projít alespoň $n_0 = 10$ tramvají.

Uvědomte si rozdíl mezi tímto číslem $n_0 = 10$ a středním počtem tramvají $E(X) \doteq 3$.

Definice: Veličina $X : \Omega \to \mathbb{R}$ má tzv. *spojité rozdělení* $\Leftrightarrow$ její distribuční funkce F_X je spojitá.

Často je v tomto případě F_X tzv. *absolutně spojitá* $\Leftrightarrow$ ex. $f_X : \mathbb{R} \to \langle 0, +\infty)$, která je integrabilní a platí

$$F_X(t) = \int\limits_{-\infty}^{t} f_X(u)\ \mathrm{d}u$$

pro každé $t \in \mathbb{R}$.

Funkce f_X se nazývá *hustotou pravděpodobnosti* veličiny X.

Odsud snadno máme, že pokud $I \subseteq \mathbb{R}$ je interval (nebo i nějaká množina poskládaná z intervalů), pak

$$P(X \in I) = \int\limits_{I} f_X(u)\ \mathrm{d}u$$

tedy pravděpodobnost, že hodnoty veličiny X padnou do I, zjistíme prostě zintegrováním hustoty přes I (podobně zjišťujeme např. hmotnost nějaké křivky, když zintegrujeme hustotu (hmotnosti) přes danou křivku).

Poznamenejme ještě důležitou věc a sice, že hustota f_X NENÍ zdaleka určena jednoznačně jako funkce (např. její změnou v konečně mnoha bodech se nezmění příslušné integrály, takže i změněná funkce bude také hustotou).

Věta (o existenci hustoty):

Nechť F_X je spojitá a existují reálná čísla $a_0 < a_1 < \cdots < a_n$ taková, že na otevřených intervalech $(-\infty,\ a_1)$, $(a_1,\ a_2)$, $\ldots$, $(a_{n-1},\ a_n)$, $(a_n,\ +\infty)$ je F_X spojitě diferencovatelná. Pak veličina X má hustotu a funkce

$$f_X(t) = \begin{cases} (F_X)'(t) & \text{pokud tato derivace v bodě } t \text{ existuje a je konečná} \\ 0 & \text{jinak.} \end{cases}$$

je hustotou pravděpodobnosti veličiny X.

Poznámka: Stojí za to ještě poznamenat, že vlastnost absolutní spojitosti není samozřejmá pro spojité distribuční funkce - příkladem je tzv. *Cantorova funkce c*, také známá jako *"ďábelské schodiště"* (viz např. `https://en.wikipedia.org/wiki/Cantor_function`). Tato funkce (rozšířená z intervalu $\langle 0, 1\rangle$ přirozeně na celé $\mathbb{R}$) má nulovou derivaci c' až na (z hlediska integrálu) nepodstatnou množinu bodů. Přesto však Cantorova funkce není konstantní (a tudíž nelze získat integrováním své derivace - příslušné integrály $\int_{-\infty}^{t} c'(t)\ \mathrm{d}t = 0$ jsou všechny nulové). Tedy pro veličinu s touto distribuční funkcí neexistuje hustota pravděpodobnosti.

Příklad 5.7 *Určete konstantu $c \in \mathbb{R}$ tak, aby funkce*

$$f(x) = \begin{cases} c \cdot xe^{-2x} & , x \in (0, 1) \\ 0 & , \textit{jinak.} \end{cases}$$

byla hustotou nějaké náhodné veličiny.

Řešení:

Máme charakterizační větu:

> Nezáporná integrabilní funkce $f : \mathbb{R} \to \mathbb{R}$ je hustotou pravděpodobnosti nějaké náhodné veličiny X právě když $\int\limits_{\mathbb{R}} f(x)\ dx = 1$.

Použitím integrace per partes dostaneme

$$1 = \int_{-\infty}^{\infty} f(x)\ dx = \int_{-\infty}^{0} 0\ dx + \int_{0}^{1} c \cdot xe^{-2x}\ dx + \int_{1}^{\infty} 0\ dx =$$

$$= c\left[x \cdot \frac{e^{-2x}}{-2}\right]_0^1 + c\int_0^1 \frac{e^{-2x}}{2}\ dx = -\frac{c}{2}e^{-2} + c\left[\frac{e^{-2x}}{-4}\right]_0^1 = c \cdot \frac{1-3e^{-2}}{4},$$

a tudíž $c = \frac{4}{1-3e^{-2}}$. Protože $c > 0$, je splněna i nezápornost funkce f.