4. cvičení z PST

11. - 15. října 2021

Důležitá poznámka: Náhodná veličina je funkce, která náhodnému výsledku (tj. elementárnímu jevu) přiřadí konkrétní hodnotu. Např. vybranému člověku přiřadí jeho tělesnou výšku. Jak je vidět, náhodná veličina vůbec není náhodná, co do hodnoty, kterou přiřazuje (ta je určena naprosto jasně). Náhodnost výstupu veličiny je dána náhodností jejího vstupu!

Abychom mohli s náhodnou veličinou X vůbec pracovat, potřebujeme umět určovat pravděpodobnosti toho, že hodnoty X budou např. v intervalu $\langle 1.5,\ 7.2) \subseteq \mathbb{R}$, což znamená, že množině $X^{-1}\langle 1.5,\ 7.2) = \{\omega \in \Omega \mid 1.5 \leq X(\omega) < 7.2\}$ musíme umět přiřadit její pravděpodobnost. To půjde jen tehdy, jestliže to bude množina měřitelná v daném systému jevů. Proto následující definice:

Definice: Náhodná veličina X v Kolmogorově modelu $(\Omega, \mathcal{A}, P)$ je takové zobrazení

$$X : \Omega \to \mathbb{R}$$

že vzor každého intervalu I v $\mathbb{R}$ je "přípustná" množina (neboli jev), tj. $X^{-1}(I) \in \mathcal{A}$. Tuto vlastnost stačí ověřit jen pro určité typy intervalů v $\mathbb{R}$:

$$X \text{ je náhodná veličina} \quad \Leftrightarrow \quad (\forall t \in \mathbb{R}) \;\; X^{-1}\Big((-\infty, t\rangle\Big) \in \mathcal{A}$$

4.1 (náhodná veličina v Kolmogorově modelu)

Zjistěte, zda $X : \Omega \to \mathbb{R}$ je náhodná veličina, pokud Kolmogorův model $(\Omega, \mathcal{A}, P)$ je

- $\Omega = \{a, b, c\}$,
- $\mathcal{A} = \big\{\emptyset, \{a\}, \{b, c\}, \{a, b, c\}\big\}$,
- $P(A) = \frac{1}{3}|A|$, $A \in \mathcal{A}$.

a platí-li, že $X(a) = 1$, $X(b) = 2$, $X(c) = 3$, a pokud není, navrhněte jiný (nekonstantní) předpis $Y : \Omega \to \mathbb{R}$, který veličinou bude.

Řešení:

Nejdříve bychom měli zkontrolovat, jestli máme opravdu Kolmogorův model:

Systém $\mathcal{A}$ je zřejmě uzavřen na sjednocení i na doplňky. Tedy $\mathcal{A}$ je σ-algebra, která je navíc podalgebrou $\exp(\{a, b, c\})$. Z tohoto důvodu budou splněny i požadavky na pravděpodobnost, protože ta má stejný předpis jako v jednom z předchozích příkladu, kde už to, jak víme, funguje. Tím spíše to musí platit i pro menší systém množin. Tedy $(\Omega, \mathcal{A}, P)$ je opravdu Kolmogorův model.

Podíváme se, jak vypadají vzory všech potřebných intervalů:

$$X^{-1}\Big((-\infty, t\rangle\Big) = \{\omega \in \Omega \mid X(\omega) \leq t\} = \begin{cases} \emptyset & ,\ t \in (-\infty, 1) \\ \{a\} & ,\ t \in \langle 1, 2) \\ \{a, b\} & ,\ t \in \langle 2, 3) \\ \{a, b, c\} & ,\ t \in \langle 3, +\infty) \end{cases}$$

X tedy není náhodná veličina, protože např.

$$X^{-1}\Big((-\infty, 2.5\rangle\Big) = \{\omega \in \Omega \mid X(\omega) \leq 2.5\} = \{a, b\} \notin \mathcal{A} \ .$$

Problematické jsou hodnoty $t \in \langle 2, 3)$. K této situaci došlo proto, že zobrazení X oddělilo svými hodnotami prvky b a c, které jsou v rámci σ-algebry $\mathcal{A}$ nerozlišitelné (není už žádná menší množina z $\mathcal{A}$, která by obsahovala b a neobsahovala c nebo naopak).

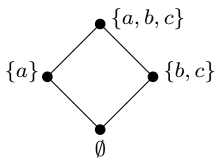
Jak tedy zvolit nějakou jinou (nekonstantní) náhodnou veličinu Y na Ω? Stačí např. položit

$$Y(a) = 1,\ Y(b) = Y(c) = 2 \ .$$

(Problémového intervalu jsme se zbavili tak, že všem prvkům z množiny $\{\omega \in \Omega \mid X(\omega) \in \langle 2, 3)\}$ jsme "nastavili" nějakou stejnou společnou hodnotu.)

Můžeme si ještě pro názornost zakreslit Hasseův diagram pro $\mathcal{A}$:

$\mathcal{A}$:



Poznámka: Jak by mohla vypadat nějaká konkrétní fyzická realizace výše uvedené Kolmogorova modelu (zejména nás zajímá realizace výsledků Ω a jevů, tj. σ-algebry $\mathcal{A}$). Potřebujeme tudíž mít 3 výsledky a dva z nich nerozlišitelné:

- můžeme např. házet válečkem, který bude dopadat buď na hranu nebo na jednu z nerozlišitelných podstav (dlouhý válec bude padat spíše na hranu a plochý válec zase spíše na podstavy)
- nebo můžeme uvažovat urnu, která obsahuje koule dvou barev a kostky jedné barvy. Předměty pak vytahujeme potmě, takže umíme rozlišit jejich tvar ale ne barvu.

Důležitá poznámka (možná i nejdůležitější vůbec....): Co umožňuje náhodná veličina:

Náhodná veličina $X : \Omega \to \mathbb{R}$ umožňuje jednu podstatnou věc - zapomenout na původní prostor Ω všech výsledků (i s celou jeho strukturou jevů a jejich pravděpodobnostmi) a přitom stále umět vyčíslovat pravděpodobnosti pro X, které potřebujeme znát.

Veličina X totiž umí přenést a vytvořit rozdělení pravděpodobnosti tam, kde má své hodnoty - konkrétně vytvoří nový Kolmogorův model na reálné přímce $\mathbb{R}$. A to tak, že každý interval $I \subseteq \mathbb{R}$ bude mít prostě pravděpodobnost $P_X(I) := P\Big(X^{-1}(I)\Big)$.

Pravděpodobnosti P_X se říká rozdělení pravděpodobnosti veličiny X (na $\mathbb{R}$). Toto rozdělení můžeme úplně popsat pokud opět známe pravděpodobnosti jen některých speciálních intervalů: opět jsou to intervaly $(-\infty, t\rangle$ pro $t \in \mathbb{R}$. Jejich pravděpodobnosti pak přirozeně definují tzv. *distribuční funkci* $F_X : \mathbb{R} \to \langle 0, 1\rangle$ jako

$$F_X(t) := P(X \leq t) \qquad \Big(= P\big(X^{-1}(-\infty, t\rangle\big)\Big)$$

Toto je jeden z nejdůležitějších pojmů v celém kurzu! Proto se jej snažte pochopit.

Definice: Veličina $X : \Omega \to \mathbb{R}$ má tzv. *diskrétní rozdělení* $\Leftrightarrow$ existuje **nejvýše spočetná** množina hodnot $A \subseteq \mathbb{R}$ taková, že $P(X \in A) = 1$.

(Veličina může nabývat i jiných hodnot, ale pak se tyto hodnoty objevují s nulovou pravděpodobností. To ovšem neznamená, že by nemohly nastat!)

4.2 (diskrétní veličina)

Do terče, nezávisle na sobě, střílí 3 střelci. První se trefí s pravděpodobností 70%, druhý s pravděpodobností 80% a třetí s pravděpodobností 90%. Náhodná veličina X je počet zásahů v terči. Určete:

(a) rozdělení pravděpodobnosti veličiny X.

(b) distribuční funkci F_X náhodné veličiny X.

(c) pravděpodobnost, že v terči budou alespoň 2 zásahy.

(d) střední počet zásahů v terči.

Řešení:
Veličina X má obor hodnot $\{0,1,2,3\}$ a tedy diskrétní rozdělení. Abychom si mohli odvodit pravděpodobnosti jednotlivých hodnot, označme si pro $i=1,2,3$ jevy:

$A_i =$ *"i-tý střelec se trefí"*,

které jsou nezávislé a mají pravděpodobnosti $P(A_1)=0.7$, $P(A_1)=0.8$ a $P(A_1)=0.9$. Pravděpodobnostní funkce pak je

$$p_X(0) = P(X=0) = P(\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}) = (1-0.7)\cdot(1-0.8)\cdot(1-0.9) = 0.3\cdot 0.2\cdot 0.1 = \mathbf{0.006}$$

$$\begin{aligned} p_X(1) = P(X=1) &= P\Big((A_1\cap\overline{A_2}\cap\overline{A_3}) \;\cup\; (\overline{A_1}\cap A_2\cap\overline{A_3}) \;\cup\; (\overline{A_1}\cap\overline{A_2}\cap A_3)\Big) = \\ &= 0.7\cdot 0.2\cdot 0.1 + 0.3\cdot 0.8\cdot 0.1 + 0.3\cdot 0.2\cdot 0.9 = \mathbf{0.092} \\ p_X(2) = P(X=2) &= P\Big((\overline{A_1}\cap A_2\cap A_3) \;\cup\; (A_1\cap\overline{A_2}\cap A_3) \;\cup\; (A_1\cap A_2\cap\overline{A_3})\Big) = \\ &= 0.3\cdot 0.8\cdot 0.9 + 0.7\cdot 0.2\cdot 0.9 + 0.7\cdot 0.8\cdot 0.1 = \mathbf{0.398} \end{aligned}$$

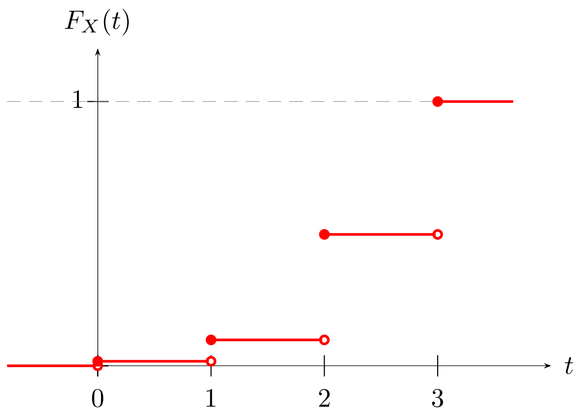
$$p_X(3) = P(X=3) = P(A_1\cap A_2\cap A_3) = 0.7\cdot 0.8\cdot 0.9 = 0.3\cdot 0.2\cdot 0.1 = \mathbf{0.504}$$

Konkrétně tedy:

k	0	1	2	3
$p_X(k)$	0.006	0.092	0.398	0.504

Distribuční funkce $F_X(t) = P(X\le t) = \sum_{k\le t} p_X(k)$ pro diskrétní veličinu X má hodnoty

$$F_X(t) = \begin{cases} 0 & , t<0 \\ p_X(0) = 0.006 & , t\in\langle 0,1) \\ p_X(0)+p_X(1) = 0.098 & , t\in\langle 1,2) \\ p_X(0)+p_X(1)+p_X(2) = 0.496 & , t\in\langle 2,3) \\ 1 & , t\ge 3. \end{cases}$$



A konečně, pravděpodobnost jevu

$$B = \text{"}v\ terči\ budou\ alespoň\ 2\ zásahy\text{"}$$

je

$$P(B) = P(X\ge 2) = P(X=2)+P(X=3) = 0.398+0.504 = 0.902\ .$$

Průměrný počet zásahů v terči neboli střední hodnota naší diskrétní veličiny je

$$E(X) = \sum_{k\in\mathbb{R}} k\cdot p_X(k) = 0\cdot 0.006 + 1\cdot 0.092 + 2\cdot 0.398 + 3\cdot 0.504 = 2.4\ .$$

4.3 (diskrétní veličina - hypergeometrické rozdělení)

V balíčku je 9 semen, z toho 3 už nevyklíčí. Náhodně vytáhneme 4 z nich a zasadíme. Náhodná veličina X je počet rostlin, které vyrostou. Určete:

(a) typ rozdělení veličiny X a její obor hodnot.

(b) pravděpodobnostní funkci p_X a distribuční funkci F_X náhodné veličiny X.

(c) pravděpodobnost, že budeme mít alespoň dvě rostliny.

(d) kolik průměrně vyroste rostlin.

Řešení:
Máme k dispozici 6 dobrých a 3 špatná semena. Z nich vybíráme 4. Možné hodnoty X jsou tak $\{1, 2, 3, 4\}$. Protože máme jen konečně mnoho hodnot, jde o veličinu s diskrétním rozdělením. Pravděpodobnostní funkce je

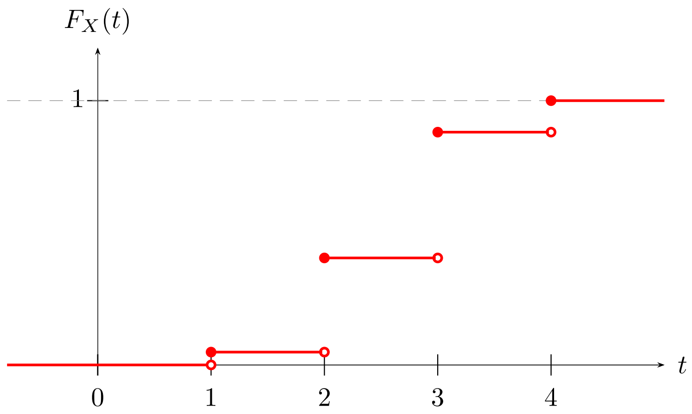
$$p_X(k) = P(X = k) = \begin{cases} \frac{\binom{6}{k}\cdot\binom{3}{4-k}}{\binom{9}{4}} & , k \in \{1, 2, 3, 4\} \\ 0 & , \text{jinak.} \end{cases} .$$

Konkrétně:

k	1	2	3	4
$p_X(k)$	$\frac{\binom{6}{1}\cdot\binom{3}{3}}{\binom{9}{4}} = \frac{2}{42} \doteq 0.048$	$\frac{\binom{6}{2}\cdot\binom{3}{2}}{\binom{9}{4}} = \frac{15}{42} \doteq 0.357$	$\frac{\binom{6}{3}\cdot\binom{3}{1}}{\binom{9}{4}} = \frac{20}{42} \doteq 0.476$	$\frac{\binom{6}{4}\cdot\binom{3}{0}}{\binom{9}{4}} = \frac{5}{42} \doteq 0.119$

Distribuční funkce $F_X(t) = P(X \leq t) = \sum_{k \leq t} p_X(k)$ pro diskrétní veličinu X má hodnoty

$$F_X(t) = \begin{cases} 0 & , t < 1 \\ p_X(0) = \frac{2}{42} \doteq 0.048 & , t \in \langle 1, 2) \\ p_X(0) + p_X(1) = \frac{17}{42} \doteq 0.405 & , t \in \langle 2, 3) \\ p_X(0) + p_X(1) + p_X(2) = \frac{37}{42} \doteq 0.881 & , t \in \langle 3, 4) \\ 1 & , t \geq 4. \end{cases}$$



Pravděpodobnost jevu

$$A = \text{“}\textit{vybereme alespoň dvě dobrá semena}\text{”}$$

je

$$P(A) = P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 1) = 1 - \frac{1}{21} = \frac{20}{21} \doteq 0.952 \; .$$

Průměrný počet rostlin je střední hodnota naší diskrétní veličin X, což je

$$E(X) = \sum_{k \in \mathbb{R}} k \cdot p_X(k) = 1 \cdot \tfrac{2}{42} + 2 \cdot \tfrac{15}{42} + 3 \cdot \tfrac{20}{42} + 4 \cdot \tfrac{5}{42} = \tfrac{8}{3} \doteq 2.67 \; .$$

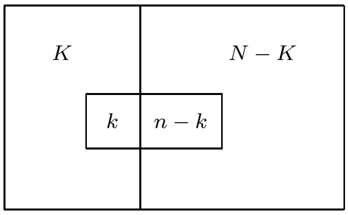
Střední hodnotu také můžeme spočítat na základě hypergeometrického rozdělení $\mathrm{Hyp}(n; K, N)$ jako

$$E(X) = n \cdot \frac{K}{N} = 4 \cdot \frac{6}{9} = \frac{8}{3} \; .$$

Poznámka: Náhodná veličiny X v tomto příkladu má tzv. *hypergeometrické* rozdělení

$$X \sim \mathrm{Hyp}(n; K, N)$$

kde N je počet prvků dané množiny (zde: počet semen), která obsahuje K prvků dané vlastnosti (zde: dobrá semena), a z níž vybíráme n prvků (kde $n \leq N$). Zajímá nás, jaká je pravděpodobnost, že z vybraných n prvků je právě k prvků uvedené vlastnosti.



V tomto případě je $n = 4$, $K = 6$ a $N = 9$.

Pravděpodobnost je dána podílem příznivých možností ku všem. Příznivé jsou dány počtem způsobů jak vybrat k semen z K dobrých semen (které fyzicky rozlišujeme) násobeno počtem způsobů jak vybrat zbytek, tj. $n - k$ semen z $N - K$ špatných (které také fyzicky rozlišujeme). Celkem tedy

$$p_{\mathrm{Hyp}(n;K,N)}(k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

přičemž obor hodnot pro k je

$$\max\{0, n + K - N\} \leq k \leq \min\{n, K\}$$

tedy

$$1 = \max\{0, 4 + 6 - 9\} \leq k \leq \min\{4, 6\} = 4 \; .$$

Tuto podmínku dostaneme ihned z nerovností

$$k \leq K, \quad n - k \leq N - K,$$

$$k \leq n, \quad n - k \leq n \; .$$

Můžeme si ještě pro zajímavost spočítat střední hodnotu naší veličiny X s rozdělením $\mathrm{Hyp}(4; 6, 9)$ způsobem, který není těžké zobecnit pro libovolné hypergeometrické rozdělení:

$$E(X) = \sum_{k=1}^{4} k \cdot \underbrace{\frac{\binom{6}{k} \cdot \binom{3}{4-k}}{\binom{9}{4}}}_{= p_{\mathrm{Hyp}(4;6,9)}(k)} = \sum_{k=1}^{4} \frac{k \cdot \frac{6!}{k!(6-k)!} \cdot \binom{3}{4-k}}{\binom{9}{4}} = \sum_{k=1}^{4} \frac{6 \cdot \frac{5!}{(k-1)!(5-(k-1))!} \cdot \binom{3}{4-k}}{\binom{9}{4}} =$$

$$= \frac{6 \cdot \binom{8}{3}}{\binom{9}{4}} \sum_{k=1}^{4} \frac{\binom{5}{k-1} \cdot \binom{3}{3-(k-1)}}{\binom{8}{3}} = \frac{6 \cdot \binom{8}{3}}{\binom{9}{4}} \sum_{\ell=0}^{3} \underbrace{\frac{\binom{5}{\ell} \cdot \binom{3}{3-\ell}}{\binom{8}{3}}}_{= p_{\mathrm{Hyp}(3;5,8)}(\ell)} = \frac{6 \cdot \binom{8}{3}}{\binom{9}{4}} \underbrace{\sum_{\ell=0}^{3} p_{\mathrm{Hyp}(3;5,8)}(\ell)}_{=1} =$$

$$= 6 \cdot \frac{\frac{8!}{3!5!}}{\frac{9!}{4!5!}} = 4 \cdot \frac{6}{9} = n \cdot \frac{K}{N} \ .$$

Limitně pro $N \to \infty$ a $K/N \to q$ a n pevně zvolené (tj. když poměr počtu dobrých semen K ku počtu všech semen N se blíží ke konstantní hodnotě $q \in (0,1)$) se hypergeometrické rozdělení blíží binomickému rozdělení $\mathrm{Bi}(n,q)$. Toto binomické rozdělení odpovídá limitní situaci, kdy vybíráme n-krát z nekonečného množství s určeným podílem q dobrých semen.

Hypergeometrické rozdělení odpovídá výběru *bez* vracení, binomické zase výběru *s* vracením. V limitě se ovšem tato vlastnost u hypergeometrického vytratí, což je tím, že v obrovském množství semen N se už ztratí to, jestli tam těch pár semen n (pro $n << N$) vrátíme nebo ne.

Všimněme si ještě, že střední hodnoty vycházejí pro hypergeometrické i binomické rozdělení stejně v tomto smyslu:

$$E\Big(\mathrm{Hyp}(n;K,N)\Big) = n \cdot \frac{K}{N} = E\Big(\mathrm{Bi}(n, \tfrac{K}{N})\Big)$$

4.4 (diskrétní veličina - binomické rozdělení)

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 její obor hodnot.

(b) pravděpodobnostní funkci p_X a distribuční funkci F_X náhodné veličiny X.

(c) pravděpodobnost, že vytáhneme nejvýše dvě černé koule.

(d) průměrný počet vybraných černých koulí.

Řešení:

Budeme postupovat podobně jako v příkladu **4.2**.

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

Veličina X má tedy obor hodnot $\{0,1,2,3\}$ a (diskrétní) binomické rozdělení $\mathrm{Bi}(n,p)$. Pravděpodobnostní funkce proto je

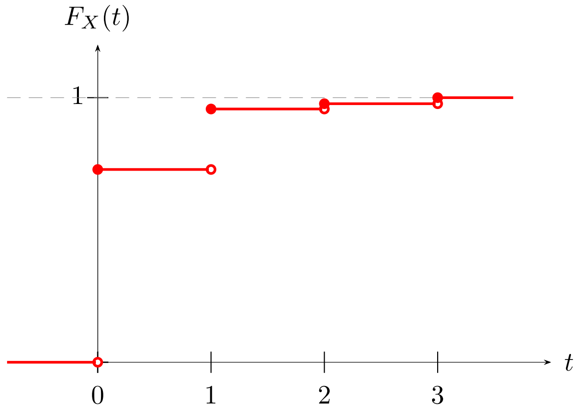
$$p_X(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & , k \in \{0,1,2,3\} \\ 0 & , \text{jinak.} \end{cases}$$

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

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



Pravděpodobnost jevu

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

je

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

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

Ještě si připomeneme, proč má veličina X vyjadřující počet úspěchů v n pokusech právě binomické rozdělení:

Jevy

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

považujeme za nezávislé se stejnou pravděpodobností $P(A_i) = p$. Pak jev

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

vyjádříme jako sjednocení disjunktních jevů (ty v největší závorce)

$$B_k = \bigcup_{\substack{K\subseteq\{1,\dots,n\}\\|K|=k}} \left(\left(\bigcap_{i\in K} A_i \right) \cap \left(\bigcap_{j\in\{1,\dots,n\}\setminus K} \overline{A_j} \right) \right)$$

odkud máme ihned

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

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

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). Pokud veličina X hustotu má, pak platí, že derivace funkce F_X, tj. $(F_X)'$ je hustotou veličiny X (tato derivace sice nemusí všude existovat, ale to není na překážku - tam, kde neexistuje, si prostě zvolíme libovolné nezáporné hodnoty).

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 (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.

4.5 (spojitá náhodná veličina)

Házíme kuličku (o zanedbatelném průměru) na kruh o poloměru $r = 1$. Náhodná veličina X je vzdálenost dopadu kuličky od středu kruhu. Určete:

(a) obor hodnot náhodné veličiny X.

(b) distribuční funkci F_X a hustotu pravděpodobnosti f_X.

(c) pravděpodobnost $P(X \geq \frac{1}{2})$.

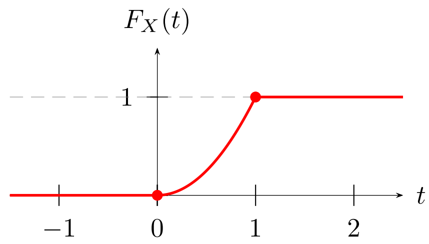
(d) $\varepsilon > 0$ takové, že $P\Big(|X - \frac{2}{3}| \leq \varepsilon\Big) = 0.9$.

Řešení:

Obor hodnot je interval $\langle 0, 1\rangle$. Množina všech možných výsledků Ω (tj. jevové pole) je kruh o poloměru 1. Předpokládáme Kolmogorův model daný geometrickou pravděpodobností.

Distribuční funkce je

$$F_X(t) = P(X \leq t) = \begin{cases} 0 & , t < 0 \\ \frac{\text{“obsah kruhu o poloměru } t\text{”}}{\text{“obsah kruhu o poloměru 1”}} = \frac{\pi t^2}{\pi 1^2} = t^2 & , t \in \langle 0, 1\rangle \\ 1 & , t > 1. \end{cases}$$



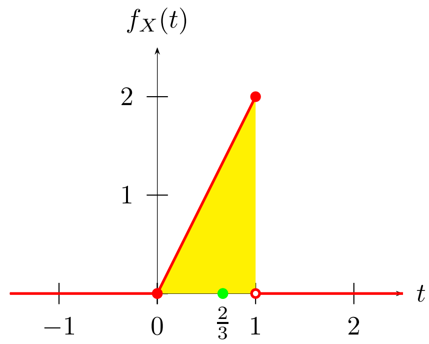
Veličina X má spojité rozdělení, protože distribuční funkce F_X je spojitá. Hustotu pravděpodobnosti (tj. funkci $f_X : \mathbb{R} \to \langle 0, +\infty)$, pro kterou je $F_X(t) = \int_{-\infty}^{t} f_X(u)\,\mathrm{d}u$ pro každé $t \in \mathbb{R}$) můžeme určit jako derivaci funkce F_X tam, kde derivace existuje (ve zbylých bodech si hustotu můžeme definovat libovolně nezáporně). Protože

$$\frac{\mathrm{d}}{\mathrm{d}t} F_X(t) = \begin{cases} 2t & , t \in (0,1) \\ 0 & , t < 0 \text{ nebo } t > 1 \ . \end{cases}$$

je funkce

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

hustotou pravděpodobnosti.



Dále máme

$$P(X \geq \tfrac{1}{2}) = 1 - P(X < \tfrac{1}{2}) = 1 - P(X \leq \tfrac{1}{2}) = 1 - F_X(\tfrac{1}{2}) = 1 - (\tfrac{1}{2})^2 = \tfrac{3}{4} \ .$$

A konečně, hledáme $\varepsilon > 0$ takové, že

$$0.9 = P\Big(X \in \langle \tfrac{2}{3} - \varepsilon, \tfrac{2}{3} + \varepsilon \rangle\Big) = P\Big(X \leq \tfrac{2}{3} + \varepsilon\Big) - P\Big(X < \tfrac{2}{3} - \varepsilon\Big) = F_X\left(\tfrac{2}{3} + \varepsilon\right) - F_X\left(\tfrac{2}{3} - \varepsilon\right) \ .$$

Vzhledem k předpisu F_X teď máme dvě možnosti:

- $0 < \varepsilon \leq \frac{1}{3}$:
  Pak je $\frac{2}{3} + \varepsilon, \frac{2}{3} - \varepsilon \in \langle 0, 1 \rangle$ a tedy

  $$0.9 = F_X\left(\tfrac{2}{3} + \varepsilon\right) - F_X\left(\tfrac{2}{3} - \varepsilon\right) = \left(\tfrac{2}{3} + \varepsilon\right)^2 - \left(\tfrac{2}{3} - \varepsilon\right)^2 = \tfrac{8}{3}\varepsilon$$

  odkud máme $\varepsilon = 0.9 \cdot \frac{3}{8} = 0.3375 \not\leq \frac{1}{3}$, což je spor. Tato možnost tedy nedává řešení.

- $\frac{1}{3} < \varepsilon \leq \frac{2}{3}$:
  Pak je $\frac{2}{3} + \varepsilon > 1$ a $\frac{2}{3} - \varepsilon \in \langle 0, 1 \rangle$ a tedy

  $$0.9 = F_X\left(\tfrac{2}{3} + \varepsilon\right) - F_X\left(\tfrac{2}{3} - \varepsilon\right) = 1 - \left(\tfrac{2}{3} - \varepsilon\right)^2 \quad \Rightarrow \quad \left(\tfrac{2}{3} - \varepsilon\right)^2 = 0.1$$

takže $\varepsilon = \frac{2}{3} \pm \sqrt{0.1}$ a z omezujících podmínek pro ε je jediné řešení

$$\varepsilon = \frac{2}{3} - \sqrt{0.1} \doteq 0.6496\ .$$

4.6 (spojitá náhodná veličina)

Rozřežeme náhodně úsečku $\langle 0,1\rangle$ na dvě části. Náhodná veličina X je délka kratší části. Určete:

(a) obor hodnot náhodné veličiny X.

(b) distribuční funkci F_X a hustotu pravděpodobnosti f_X.

(c) pravděpodobnost $P(\frac{1}{3} \leq X < 1)$

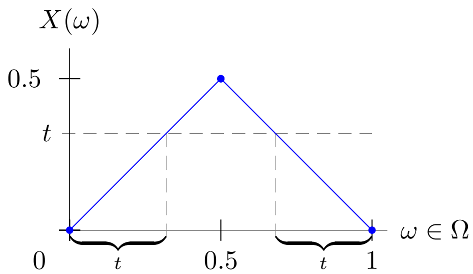
(d) $\varepsilon > 0$ takové, že $P\Big(|X - \frac{1}{4}| \geq \varepsilon\Big) = 0.2$.

Řešení:

Označme si $\Omega = \langle 0,1\rangle$, kde budeme uvažovat model geometrické pravděpodobnosti a nechť $\omega \in \Omega$ představuje délku levé části tyče. Veličina má nyní předpis

$$X(\omega) = \min\{\omega, 1-\omega\} = \begin{cases} \omega & ,\ \text{pro } 0 \leq \omega \leq 0.5, \\ 1-\omega & ,\ \text{pro } 0.5 \leq \omega \leq 1. \end{cases}$$

Veličinu X si můžeme znázornit grafem níže (modrá čára).

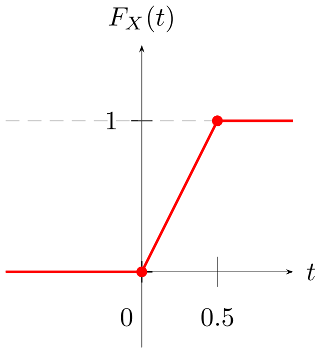


Veličina má tedy obor hodnot $\langle 0, 0.5\rangle$. Pro $t \in \langle 0, 0.5\rangle$ díky grafu máme, že

$$\text{“}X \leq t\,\text{”} = \{\omega \in \Omega \mid X(\omega) \leq t\} = \langle 0,t\rangle \cup \langle 1-t, 1\rangle.$$

Pro distribuční funkci pak dostáváme

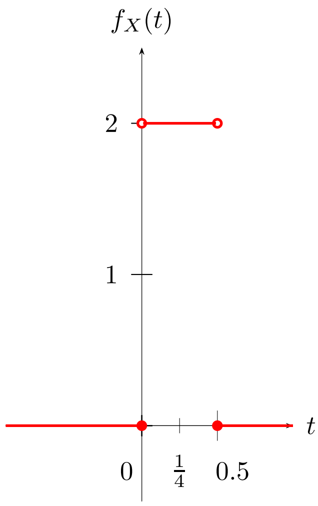
$$F_X(t) = P(X \leq t) = \begin{cases} P(\emptyset) = & 0 & ,t < 0 \\ P\big(\langle 0,t\rangle \cup \langle 1-t,1\rangle\big) = & \text{“poměr délek množin”} = \frac{t+t}{1} = 2t & ,t \in \langle 0, 0.5\rangle \\ P(\Omega) = & 1 & ,t > 0.5\ . \end{cases}$$



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

$$f_X(t) = \frac{\mathrm{d}}{\mathrm{d}t} F_X(t) = \begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}(2t) = 2 & , t \in (0, 0.5) \\ 0 & , \text{ jinak} \end{cases} .$$

Veličina X má tzv. rovnoměrné rozdělení na intervalu $\langle 0, 0.5 \rangle$.



Dále máme (díky spojitosti funkce F_X)

$$P\left(\tfrac{1}{3} \le X < 1\right) = P(X < 1) - P(X < \tfrac{1}{3}) \overset{(spojitost)}{=} F_X(1) - F_X(\tfrac{1}{3}) = 1 - 2 \cdot \tfrac{1}{3} = \tfrac{1}{3} .$$

Pro výpočet můžeme použít i hustotu:

$$P\left(\tfrac{1}{3} \le X < 1\right) = \int_{\frac{1}{3}}^{1} f_X(t)\, \mathrm{d}t = \int_{\frac{1}{3}}^{0.5} 2\, \mathrm{d}t + \int_{0.5}^{1} 0\, \mathrm{d}t = 2 \cdot (0.5 - \tfrac{1}{3}) = 2 \cdot \tfrac{1}{6} = \tfrac{1}{3} .$$

Nakonec hledáme $0 < \varepsilon$, aby $0.2 = P\Big(|X - \frac{1}{4}| \ge \varepsilon\Big)$. Je vidět, že pro $\varepsilon > 0.25 = \frac{1}{4}$ je

$$P\Big(|X - \tfrac{1}{4}| \ge \varepsilon\Big) \le P\Big(|X - \tfrac{1}{4}| \ge \tfrac{1}{4}\Big) = P\Big(X \in (-\infty, 0\rangle \cup \langle 0.5, \infty)\Big) = 0$$

a tedy řešení nenajdeme.

Proto budeme předpokládat, že $0 < \varepsilon < \frac{1}{4}$. Pak z předpisu F_X máme

$$0.2 = P\Big(|X - \tfrac{1}{4}| \geq \varepsilon\Big) = 1 - P\Big(|X - \tfrac{1}{4}| < \varepsilon\Big) = 1 - P\big(\tfrac{1}{4} - \varepsilon < X < \tfrac{1}{4} + \varepsilon\big) =$$

$$= 1 - \big(F_X(\tfrac{1}{4} + \varepsilon) - F_X(\tfrac{1}{4} - \varepsilon)\big) = 1 - 2\Big(\tfrac{1}{4} + \varepsilon - (\tfrac{1}{4} - \varepsilon)\Big) = 1 - 4\varepsilon$$

neboli

$$0.2 = 1 - 4\varepsilon \quad \Rightarrow \quad \varepsilon = 0.2 < \frac{1}{4}\ .$$

Podobně můžeme k výpočtu použít hustotu, kde podmínka $0 < \varepsilon < \frac{1}{4}$ bude ihned zřejmá z integrálu:

$$0.2 = P\Big(|X - \tfrac{1}{4}| \geq \varepsilon\Big) = P\Big(X \in (-\infty, \tfrac{1}{4} - \varepsilon\rangle \cup \langle \tfrac{1}{4} + \varepsilon, \infty)\Big) =$$

$$= \int\limits_{-\infty}^{\frac{1}{4}-\varepsilon} f_X(t)\ \mathrm{d}t + \int\limits_{\frac{1}{4}+\varepsilon}^{\infty} f_X(t)\ \mathrm{d}t = \int\limits_{0}^{\frac{1}{4}-\varepsilon} 2\ \mathrm{d}t + \int\limits_{\frac{1}{4}+\varepsilon}^{0.5} 2\ \mathrm{d}t = 2\Big(\tfrac{1}{4} - \varepsilon + (0.5 - \tfrac{1}{4} - \varepsilon)\Big) = 1 - 4\varepsilon$$

Ještě poznamenejme, že výše uvedená pravděpodobnost představuje vlastně jen velikost níže znázorněné (šedé) plochy pod grafem hustoty: