

4. cvičení z PST

16. října 2019

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 $\langle 1.5, 7.2 \rangle \subseteq \mathbb{R}$, což znamená, že množině $X^{-1}\langle 1.5, 7.2 \rangle = \{\omega \in \Omega \mid 1.5 \leq X(\omega) < 7.2\}$ 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 (Ω, \mathcal{A}, P) je takové zobrazení

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

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} :

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

4.1 (náhodná veličina v Kolmogorově modelu)

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

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

a platí-li, že $X(a) = 1$, $X(b) = 2$, $X(c) = 3$, a pokud není, navrhněte jiný (nekonstantní) předpis $Y : \Omega \rightarrow \mathbb{R}$, 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 (Ω, \mathcal{A}, P) je opravdu Kolmogorův model.

Podíváme se, jak vypadají vzory všech potřebných intervalů:

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

X tedy není náhodná veličina, protože např.

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

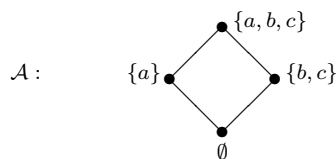
Problematické jsou hodnoty $t \in \langle 2, 3 \rangle$. K této situaci došlo proto, že zobrazení X oddělilo svými hodnotami prvky b a c , které jsou v rámci σ -algebry \mathcal{A} nerozlišitelné (není už žádná menší množina z \mathcal{A} , která by obsahovala b a neobsahovala c nebo naopak).

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 $\{\omega \in \Omega \mid X(\omega) \in \langle 2, 3 \rangle\}$ jsme "nastavili" nějakou stejnou společnou hodnotu.)

Můžeme si ještě pro názornost zakreslit Hasseův diagram pro \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 váleček bude padat spíše na hranu a plochý váleček zase spíše na podstavu)
- 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 $X : \Omega \rightarrow \mathbb{R}$ 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 $P_X(I) := P(X^{-1}(I))$.

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 $(-\infty, t)$ pro $t \in \mathbb{R}$. Jejich pravděpodobnosti pak přirozeně definují tzv. *distribuční funkci* $F_X : \mathbb{R} \rightarrow \langle 0, 1 \rangle$ jako

$$F_X(t) := P(X \leq t) \quad \left(= P(X^{-1}(-\infty, t)) \right)$$

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

Definice: Veličina $X : \Omega \rightarrow \mathbb{R}$ 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 - 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:

- typ rozdělení veličiny X a její obor hodnot.
- pravděpodobnostní funkci p_X a distribuční funkci F_X náhodné veličiny X .
- kolik průměrně vyrostou rostlin.
- rozptyl $D(X)$ veličiny X .
- pravděpodobnost, že budeme mít alespoň dvě rostliny.

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

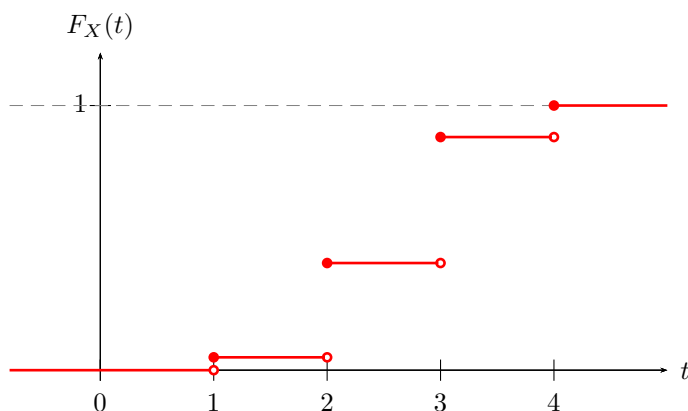
$$p_X(k) = P(X = k) = \begin{cases} \frac{\binom{6}{k} \cdot \binom{3}{4-k}}{\binom{9}{4}} & , k \in \{1, 2, 3, 4\} \\ 0 & , \text{jinak.} \end{cases}$$

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

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



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

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

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

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

Rozptyl náhodné veličiny určuje, jak moc se hodnoty odchylojí od střední hodnoty. Rozptyl je proto definován jako střední hodnota veličiny $(X - E(X))^2$, která vyjadřuje kvadrát odchylky X od své střední hodnoty, tedy předpisem:

$$D(X) = E\left((X - E(X))^2\right) = E(X^2) - (E(X))^2$$

K tomu tedy potřebujeme ještě určit $E(X^2)$, což uděláme pomocí vzorce

$$E(h(X)) = \sum_{k \in \mathbb{R}} h(k) \cdot p_X(k),$$

kde $h : \mathbb{R} \rightarrow \mathbb{R}$ je (borelovsky) měřitelná funkce, např. spojitá. Takže máme

$$E(X^2) = \sum_{k \in \mathbb{R}} k^2 \cdot p_X(k) = 1^2 \cdot \frac{2}{42} + 2^2 \cdot \frac{15}{42} + 3^2 \cdot \frac{20}{42} + 4^2 \cdot \frac{5}{42} = \frac{23}{3}.$$

Tedy

$$D(X) = E(X^2) - (E(X))^2 = \frac{23}{3} - \left(\frac{8}{3}\right)^2 = \frac{5}{9} \doteq 0.556.$$

Rozptýl také můžeme počítat podle vzorce pro hypergeometrického rozdělení $\text{Hyp}(n; K, N)$ jako

$$D(X) = n \cdot \frac{K}{N} \cdot \left(1 - \frac{K}{N}\right) \cdot \frac{N-n}{N-1} = 4 \cdot \frac{6}{9} \cdot \left(1 - \frac{6}{9}\right) \cdot \frac{9-4}{9-1} = \frac{5}{9}.$$

A konečně, pravděpodobnost jevu

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

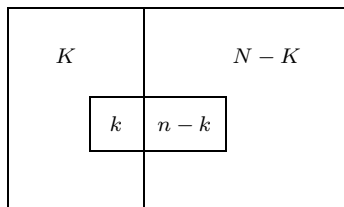
je

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

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

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

kde N je počet prvků dané množiny (zde: počet semen), která obsahuje K prvků dané vlastnosti (zde: dobrá semena), a z ní 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

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

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 $\text{Hyp}(4; 6, 9)$ způsobem, který není těžké zobecnit pro libovolné hypergeometrické rozdělení:

$$\begin{aligned} E(X) &= \sum_{k=1}^4 k \cdot \underbrace{\frac{\binom{6}{k} \cdot \binom{3}{4-k}}{\binom{9}{4}}}_{=p_{\text{Hyp}(4;6,9)}(k)} = \sum_{k=1}^4 k \cdot \frac{6!}{k!(6-k)!} \cdot \frac{\binom{3}{4-k}}{\binom{9}{4}} = \sum_{k=1}^4 \frac{6 \cdot \frac{5!}{(k-1)!(5-(k-1))!} \cdot \binom{3}{4-k}}{\binom{9}{4}} = \\ &= \frac{6 \cdot \binom{8}{3}}{\binom{9}{4}} \sum_{k=1}^4 \frac{\binom{5}{k-1} \cdot \binom{3}{3-(k-1)}}{\binom{8}{3}} = \frac{6 \cdot \binom{8}{3}}{\binom{9}{4}} \sum_{\ell=0}^3 \underbrace{\frac{\binom{5}{\ell} \cdot \binom{3}{3-\ell}}{\binom{8}{3}}}_{=p_{\text{Hyp}(3;5,8)}(\ell)} = \frac{6 \cdot \binom{8}{3}}{\binom{9}{4}} \underbrace{\sum_{\ell=0}^3 p_{\text{Hyp}(3;5,8)}(\ell)}_{=1} = \\ &= 6 \cdot \frac{8!}{9!} = 4 \cdot \frac{6}{9} = n \cdot \frac{K}{N}. \end{aligned}$$

Limitně pro $N \rightarrow \infty$ a $K/N \rightarrow q$ 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 $\text{Bi}(n, q)$. 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 $n \ll N$) vrátíme nebo ne.

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

$$E\left(\text{Hyp}(n; K, N)\right) = n \cdot \frac{K}{N} = E\left(\text{Bi}\left(n, \frac{K}{N}\right)\right)$$

Rozptyly se pak už liší:

$$D\left(\text{Hyp}(n; K, N)\right) = n \cdot \frac{K}{N} \cdot \left(1 - \frac{K}{N}\right) \cdot \frac{N-n}{N-1}$$

$$D\left(\text{Bi}\left(n, \frac{K}{N}\right)\right) = n \cdot \frac{K}{N} \cdot \left(1 - \frac{K}{N}\right)$$

4.3 (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:

- typ rozdělení veličiny X a její obor hodnot.
- pravděpodobnostní funkci p_X a distribuční funkci F_X náhodné veličiny X .
- průměrný počet vybraných černých koulí.
- rozptyl $D(X)$ veličiny X .
- pravděpodobnost, že vytáhneme alespoň jednu černou kouli.

Řešení:

Budeme postupovat podobně jako v příkladu 4.3.

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 $\text{Bi}(n, p)$. Pravděpodobnostní funkce proto je

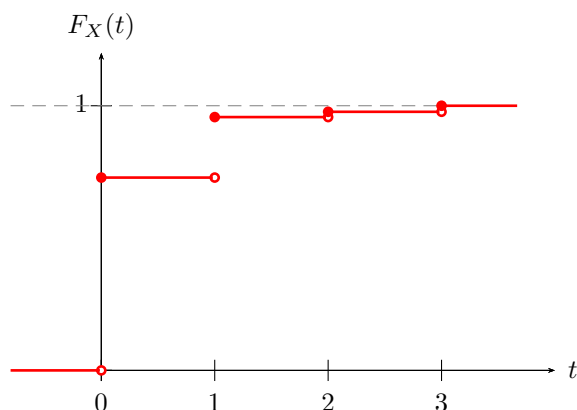
$$p_X(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & , k \in \{0, 1, 2, 3\} \\ 0 & , \text{jinak.} \end{cases}$$

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

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



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

Pro rozptyl $D(X) = E(X^2) - (E(X))^2$ 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

$$D(X) = E(X^2) - (E(X))^2 = 0.36 - (0.3)^2 = 0.27$$

nebo podle vzorce pro binomické rozdělení $D(X) = np(1-p) = 3 \cdot 0.1 \cdot 0.9 = 0.27$.

A konečně, pravděpodobnost jevu

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

je

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

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

Jevy

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

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

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

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

$$\begin{aligned} P(X = k) = P(B_k) &= \sum_{\substack{K \subseteq \{1, \dots, n\} \\ |K|=k}} \left(\left(\prod_{i \in K} P(A_i) \right) \cdot \left(\prod_{j \in \{1, \dots, n\} \setminus K} P(\overline{A_j}) \right) \right) = \\ &= \sum_{\substack{K \subseteq \{1, \dots, n\} \\ |K|=k}} p^k \cdot (1-p)^{n-k} = \binom{n}{k} p^k \cdot (1-p)^{n-k} \end{aligned}$$

4.4 (diskrétní veličina - binomické rozdělení)

Měli jsme tři stejné páry rukavic. Každá z nich se, nezávisle na ostatních, ztratila s pravděpodobností 0.1.

- Určete rozdělení veličin $X = \text{“počet zbylých levých rukavic”}$ a $Y = \text{“počet zbylých pravých rukavic”}$.
- Jaká je pravděpodobnost, že pravých a levých rukavic je stejný počet?
- Jestliže víme, že pravých i levých rukavic zbylo stejně, jaká je pravděpodobnost, že se žádná rukavice neztratila?

Řešení:

(a) Veličina X označuje počet úspěchů (levá rukavice se neztratila s pravděpodobností $p = 0.9$) při $n = 3$ pokusech. Má tedy binomické rozdělení $\text{Bi}(3, 0.9)$ (viz příklad 4.3). Podobně to platí pro veličinu Y , která má tudíž také binomické rozdělení $\text{Bi}(3, 0.9)$. Konkrétní hodnoty pravděpodobnostních funkcí

$$p_X(k) = p_Y(k) = \binom{n}{k} p^k \cdot (1-p)^{n-k} = \binom{3}{k} 0.9^k \cdot 0.1^{3-k} \text{ pro } k = 0, 1, 2, 3$$

udává tabulka

k	0	1	2	3
$p_X(k) = p_Y(k)$	0.001	0.027	0.243	0.729

(b) Obě diskrétní veličiny X a Y jsou tzv. *nezávislé*, což speciálně znamená, že

$$P(X = k \wedge Y = \ell) = P(X = k) \cdot P(Y = \ell) \text{ pro všechna } k, \ell \in \{0, 1, 2, 3\}.$$

To, že uvedené rovnosti platí, zdůvodníme pomocí jevů

$$A_i = \text{“}i\text{-tá levá rukavice se neztratila”}$$

$$B_j = \text{“}j\text{-tá pravá rukavice se neztratila”}$$

které jsou všechny dohromady nezávislé. Jev “ $X = k$ ” pak vznikne pomocí průniků, sjednocení a doplňků pouze z jevů A_i , $i = 1, 2, 3$. Podobně jev “ $Y = \ell$ ” vznikne pomocí průniků, sjednocení a doplňků pouze z jevů B_j , $j = 1, 2, 3$. Podle pravidel o manipulaci s nezávislými jevy tak musí být jevy “ $X = k$ ” a “ $Y = \ell$ ” nezávislé, tudíž výše uvedená rovnost platí.

S využitím této znalosti pak pro jev

$$C = \text{“pravých i levých rukavic zbyl stejný počet”}$$

máme

$$\begin{aligned} P(C) &= P(X = Y) = \sum_{k=0}^3 P(X = k \wedge Y = k) = \sum_{k=0}^3 P(X = k) \cdot P(Y = k) = \\ &= \sum_{k=0}^3 (p_X(k))^2 = 0.001^2 + 0.027^2 + 0.243^2 + 0.729^2 = 0.59122. \end{aligned}$$

(c) Podmíněná pravděpodobnost jevu, že se žádná rukavice neztratila (tj. $X = Y = 3$), pokud obou typu rukavic zbylo stejně (tj. $X = Y$), je

$$\begin{aligned} P(X = Y = 3 | X = Y) &= \frac{P(X = Y = 3 \wedge X = Y)}{P(X = Y)} = \frac{P(X = Y = 3)}{P(X = Y)} = \\ &= \frac{P(X = 3) \cdot P(Y = 3)}{P(X = Y)} \doteq \frac{0.729^2}{0.591} \doteq 0.899. \end{aligned}$$

4.5 (diskrétní veličina - hypergeometrické rozdělení)

V krabici jsou 4 nové a 3 použité tenisové míčky. Pro *první* hru se náhodně se vyberou 2 míčky, které se po skončení hry zase vrátí do krabice.

- Pro veličinu $X = \text{“počet nových míčeků vybraných pro (první) hru”}$ určete její rozdělení a střední počet nových míčeků mezi vybranými.
- Pro *druhou* hru se znovu náhodně vyberou opět 2 míčky. Určete pravděpodobnost toho, že všechny míčky, použité při této *druhé* hře, jsou nové.

Řešení:

(a) Ze $4 + 3 = 7$ míčeků, z toho 4 nových, vybíráme 2 míčky. Veličina X má proto hodnoty $\{0, 1, 2\}$ a pravděpodobnostní funkci

$$p_X(k) = P(X = k) = \begin{cases} \frac{\binom{4}{k} \cdot \binom{3}{2-k}}{\binom{7}{2}} & , k \in \{0, 1, 2\} \\ 0 & , \text{jinak.} \end{cases}$$

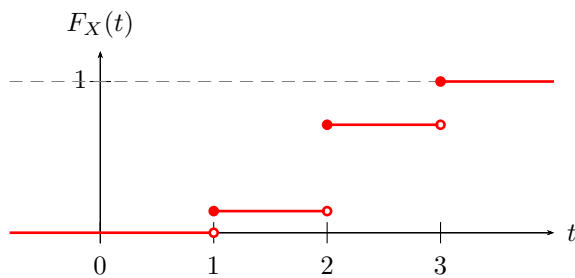
Je to tzv. hypergeometrické rozdělení (viz příklad 4.2).

Konkrétně:

k	0	1	2
$p_X(k)$	$\frac{\binom{4}{0} \cdot \binom{3}{2}}{\binom{7}{2}} = \frac{1}{7} \doteq 0.143$	$\frac{\binom{4}{1} \cdot \binom{3}{1}}{\binom{7}{2}} = \frac{4}{7} \doteq 0.571$	$\frac{\binom{4}{2} \cdot \binom{3}{0}}{\binom{7}{2}} = \frac{2}{7} \doteq 0.286$

Distribuční funkce pro toto diskrétní rozdělení je

$$F_X(t) = P(X \leq t) = \sum_{k \leq t} p_X(k) = \begin{cases} 0 & , t < 0 \\ p_X(0) = \frac{1}{7} \doteq 0.143 & , t \in \langle 0, 1 \rangle \\ p_X(0) + p_X(1) = \frac{5}{7} \doteq 0.714 & , t \in \langle 1, 2 \rangle \\ 1 & , t \geq 2. \end{cases}$$



Střední počet nových míčeků je střední hodnota veličiny X , tedy

$$E(X) = \sum_{k \in \mathbb{R}} k \cdot p_X(k) = 0 \cdot \frac{1}{7} + 1 \cdot \frac{4}{7} + 2 \cdot \frac{2}{7} = \frac{8}{7} \doteq 1.14 .$$

Střední hodnotu také můžeme spočítat na základě hypergeometrického rozdělení $\text{Hyp}(n; K, N) = \text{Hyp}(2; 4, 7)$ jako

$$E(X) = n \cdot \frac{K}{N} = 2 \cdot \frac{4}{7} = \frac{8}{7} .$$

(b) Pro $i = 0, 1, 2$ si označme jevy:

$A_i =$ "pro první hru bylo vybráno i nových a $2 - i$ použitých míčeků" = " $X=i$ ",
 $B =$ "pro druhou hru byly vybrány 2 nové míčky".

Víme, že A_0, A_1 a A_2 je úplný disjunktivní systém jevů. Hledáme pravděpodobnost $P(B)$. Z věty o úplné pravděpodobnosti máme

$$P(B) = \sum_{i=0}^2 P(B|A_i) \cdot P(A_i)$$

Zbývá určit podmíněné pravděpodobnosti. Jestliže nastal jev A_i , je v krabici $4 - i$ nových míčků. Takže máme

$$P(B|A_i) = \frac{\binom{4-i}{2}}{\binom{7}{2}}$$

s hodnotami v tabulce

i	0	1	2
$P(B A_i)$	$\frac{\binom{4}{2}}{\binom{7}{2}} = \frac{6}{21}$	$\frac{\binom{3}{2}}{\binom{7}{2}} = \frac{3}{21}$	$\frac{\binom{2}{2}}{\binom{7}{2}} = \frac{1}{21}$

Výsledek tak je

$$P(B) = \sum_{i=0}^2 P(B|A_i) \cdot P(A_i) = \sum_{i=0}^2 P(B|A_i) \cdot P(X = i) = \frac{6}{21} \cdot \frac{1}{7} + \frac{3}{21} \cdot \frac{4}{7} + \frac{1}{21} \cdot \frac{2}{7} = \frac{20}{147} \doteq 0.136 .$$

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

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

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

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

$$P(X \in I) = \int_I f_X(u) du$$

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).

Poznámka: Stojí za to ještě poznamenat, že vlastnost absolutní spojitosti není samozřejmá pro spojitě distribuční funkce - příkladem je tzv. *Cantorova funkce* c , také známá jako "*d'ábelské schodiště*" (viz např. https://en.wikipedia.org/wiki/Cantor_function). Tato funkce (rozšířená z intervalu $(0, 1)$ 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 $\int_{-\infty}^t c'(t) dt = 0$ jsou všechny nulové). Tedy pro veličinu s touto distribuční funkcí neexistuje hustota pravděpodobnosti.

4.6 (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:

- obor hodnot náhodné veličiny X .
- distribuční funkci F_X a hustotu pravděpodobnosti f_X .

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

(d) střední hodnotu $E(X)$ a rozptyl $D(X)$.

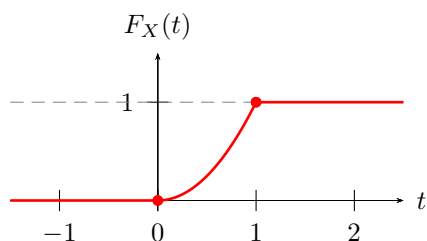
(e) 90% interval spolehlivosti se středem v $E(X)$, tedy interval $\langle E(X) - \varepsilon, E(X) + \varepsilon \rangle$ takový, že $P(|X - E(X)| \leq \varepsilon) = 0.9$.

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

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



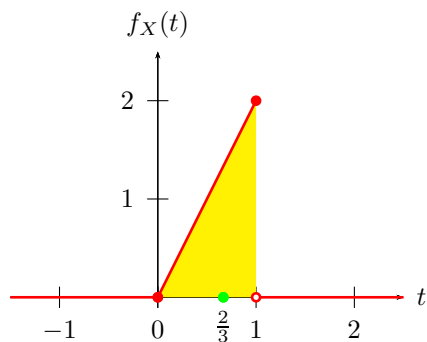
Veličina X má spojité rozdělení, protože distribuční funkce F_X je spojitá. Hustotu pravděpodobnosti (tj. funkci $f_X : \mathbb{R} \rightarrow \langle 0, +\infty \rangle$, pro kterou je $F_X(t) = \int_{-\infty}^t f_X(u) du$ pro každé $t \in \mathbb{R}$) můžeme určit jako derivaci funkce F_X tam, kde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

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

je funkce

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

hustotou pravděpodobnosti.



Dále máme

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

Střední hodnotu pro spojitou náhodnou veličinu s hustotou pravděpodobnosti můžeme spočítat jako

$$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t) dt = \int_0^1 t \cdot 2t dt = \frac{2}{3}.$$

Názornou interpretací střední hodnoty spojitě veličiny X , která má hustotu, je, že $E(X)$ je vodorovná souřadnice (zelený bod) těžiště plochy, která je určena grafem hustoty f_X (žlutá plocha). V tomto případě speciálně musí platit, že $0 < E(X) < 1$, což je evidentně splněno. V případě, že by vypočtené hodnoty $E(X)$ tyto nerovnosti nesplňovala, znamenalo by to, že jsme někde ve výpočtu udělali chybu.

Dále, pro rozptyl $D(X) = E(X^2) - (E(X))^2$ potřebujeme ještě určit $E(X^2)$, což uděláme pomocí vzorce analogického tomu pro diskrétní rozdělení:

$$E(h(X)) = \int_{-\infty}^{\infty} h(t) \cdot f_X(t) dt,$$

kde $h : \mathbb{R} \rightarrow \mathbb{R}$ je (borelovsky) měřitelná funkce, např. po částech spojitá. (Měřitelnost znamená, že h je "obyčejná" náhodná veličina ovšem vzhledem k σ -algebře borelovských množin, tedy $h^{-1}(I)$ je borelovská pro každý interval $I \subseteq \mathbb{R}$. No a k tomu stačí např. částečná spojitost.)

Tedy

$$E(X^2) = \int_{-\infty}^{\infty} t^2 \cdot f_X(t) dt = \int_0^1 t^2 \cdot 2t dt = \left[\frac{2}{4} t^4 \right]_0^1 = \frac{1}{2}$$

a proto máme

$$D(X) = E(X^2) - (E(X))^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

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

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

Vzhledem k předpisu F_X teď máme dvě možnosti:

- $0 < \varepsilon \leq \frac{1}{3}$:

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

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

- $\frac{1}{3} < \varepsilon \leq \frac{2}{3}$:

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

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

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