2. cvičení z PST

2. října 2019

2.0 Řešení příkladů ze vstupního testu.

Poznámka: Při používání klasické (tj. Laplaceovy) pravděpodobnosti potřebujeme, abychom všechny (elementární) výsledky mohli považovat za rovnocenné a tím jim mohli přiřadit stejnou pravděpodobnost.

Jako ukázkový příklad si vezměme situaci, kdy máme dvě nerozlišitelné mince, které mají stejnou pravděpodobnost, že padne kterákoliv ze stran (líc nebo rub). Pak jsou možné výsledky jednoho hodu oběma mincemi jen tyto kombinace bez opakování: $\{líc,\ líc\}$, $\{rub,\ rub\}$, $\{líc,\ rub\}$. Nemůžeme je ale považovat za rovnocenné, protože výsledek $\{líc,\ rub\}$ je (fyzicky) realizován dvěma stavy (variacemi s opakováním): $(rub,\ líc)$ a $(líc,\ rub)$. To, že tyto uspořádané stavy neumíme rozlišit ovšem nesouvisí s tím, že případ $\{líc,\ rub\}$ se (fyzicky) objeví dvakrát častěji než kterýkoliv z případů $\{líc,\ líc\}$ nebo $\{rub,\ rub\}$. Má tedy dvojnásobnou pravděpodobnost.

Proto kombinace s opakováním (tj. neuspořádané posloupnosti) v tomto případě nemůžou sloužit jako základ pro Laplaceovu pravděpodobnost a my si nutně musíme vzít variace s opakováním (tj. uspořádané posloupnosti). Na druhé straně i výsledky popisované kombinacemi lze použít, ovšem s tím, že daná neuspořádaná posloupnost bude mít pravděpodobnost úměrnou počtu jejích uspořádaných verzí:

- $P(\{líc,\ líc\}) = \frac{1}{4}$
- $P(\{rub,\ rub\}) = \frac{1}{4}$
- $P(\{líc,\ rub\}) = 2 \times \frac{1}{4} = \frac{1}{2}$.

V případě, že se tedy rozhodneme popisovat výsledky pomocí neuspořádaných výběrů s opakováním, budeme tak jako tak nakonec počítat s jejich uspořádanými verzemi (viz následující příklad).

2.1 (srovnání výběru s opakováním a bez opakování)

V loterii je $k = 500$ výher a $n = 10^7$ účastníků. Jaká je pravděpodobnost, že Alice získá nějakou výhru, pokud

(a) každý může vyhrát nejvýše jednou,

(b) každý může vyhrát opakovaně.

Řešení:

(a) Uvažujme neuspořádaný výběr bez opakování (ale klidně si můžeme vzít i uspořádané výběry a výsledek bude stejný - protože každý neuspořádaný výběr délky k zde bude mít stejný počet uspořádaných verzí, a sice právě $k!$).

Pravděpodobnost spočítáme přes doplňkový jev. Počet všech možností je $\binom{n}{k}$. Počet nepříznivých možností je $\binom{n-1}{k}$ (vynecháme Alici). Pravděpodobnost p_1 je tedy

$$p_1 = 1 - \frac{\binom{n-1}{k}}{\binom{n}{k}} = 1 - \frac{n-k}{n} = \frac{k}{n} = \frac{500}{10^7} = 5 \cdot 10^{-5} .$$

(b) Protože se jedná o výběr s opakováním, zvolíme si (kvůli snadnějšímu výpočtu) uspořádaný výběr (viz poznámka výše). Pravděpodobnost opět spočítáme přes doplňkový jev. Počet všech možností je n^k. Počet nepříznivých možností je $(n-1)^k$. Pravděpodobnost p_2 je tedy

$$p_2 = 1 - \frac{(n-1)^k}{n^k} = 1 - \left(1 - \frac{1}{n}\right)^k = 1 - (0.9999999)^{500} \doteq 1 - 0.999950001 = 4.9999 \cdot 10^{-5} .$$

Jak je vidět, pravděpodobnosti jsou téměř stejné. To můžeme vysvětlit buď pomocí binomického rozvoje

$$p_2 = 1 - \left(1 - \frac{1}{n}\right)^k = 1 - \left(1 - \frac{k}{n} + \binom{k}{2}\frac{1}{n^2} - \cdots\right) = \underbrace{\frac{k}{n}}_{p_1} - \binom{k}{2}\frac{1}{n^2} + \cdots$$

nebo pomocí limity (pro k pevné a n rostoucí nade všechny meze)

$$\lim_{n\to\infty}\frac{p_2}{p_1} = \lim_{n\to\infty}\frac{1-\left(1-\frac{1}{n}\right)^k}{\frac{k}{n}} = \left[x = -\frac{1}{n}\right] = \frac{1}{k}\cdot\lim_{x\to 0}\frac{(1+x)^k - 1}{x} = \frac{1}{k}\cdot\frac{\mathrm{d}}{\mathrm{d}x}(1+x)^k_{|x=0} = \frac{1}{k}\cdot k = 1\ .$$

Současně tento výsledek odpovídá intuitivní představě, že v obrovském množství účastníků n se už v limitě (tj. pro $k << n$) ztratí to, jestli je losujeme opakovaně (tj. vlastně neměníme podmínky) nebo ne (tj. daného výherce vždy vyřadíme).

V rámci geometrické pravděpodobnosti pracujeme vždy v $\mathbb{R}^n$, kde máme obvyklý n-rozměrný objem $\mathrm{vol}(\cdot)$. Potřebujeme tedy nějak vymezit jevy jako množiny, kterým umíme přiřadit objem (tzv. borelovsky měřitelné množiny) a tím následně i pravděpodobnost. Systém takovýchto množin tvoří tzv. σ-algebru, která právě díky své struktuře umožňuje objem množin definovat. Odsud vidíme, že pojmu σ-algebry (a dalším definicím spojeným s pravděpodobností) se prostě nelze vyhnout, pokud hodláme pracovat např. s geometrickou nebo jinou pravděpodobností.

V případě geometrické pravděpodobnosti bude prostorem všech možných výsledků nějaká množina $\Omega \subseteq \mathbb{R}^n$ s konečným objemem $\mathrm{vol}(\Omega) < \infty$, jev bude její měřitelná podmnožina $A \subseteq \Omega$ a jeho pravděpodobnost bude určena jako $P(A) = \frac{\mathrm{vol}(A)}{\mathrm{vol}(\Omega)}$. Tím, že pravděpodobnost je úměrná jen objemu množiny A a nikoliv jejímu tvaru nebo umístění, chceme opět vyjádřit to, že všechny výsledky (tedy body množiny Ω) považujeme za rovnocenné.

2.2 (geometrická pravděpodobnost)
Dva přátelé A a B si domluví schůzku mezi 9.00 a 10.00. Jejich příchody na dané místo jsou náhodné v rámci smluveného časového intervalu. Každý bude čekat 10 minut a pak odchází. Jaká je pravděpodobnost, že dojde k setkání?

Řešení:
Jako elementární jev si zvolíme dvojici (t_1, t_2), která znamená příchody jednotlivých osob v jednotkách *hodin*. Tedy $\Omega = \langle 9, 10\rangle \times \langle 9, 10\rangle$.

Osoba A bude čekat na daném místě v intervalu $\langle t_1, t_1 + \frac{1}{6}\rangle$ a podobně osoba B bude čekat na daném místě v intervalu $\langle t_2, t_2 + \frac{1}{6}\rangle$. Jev

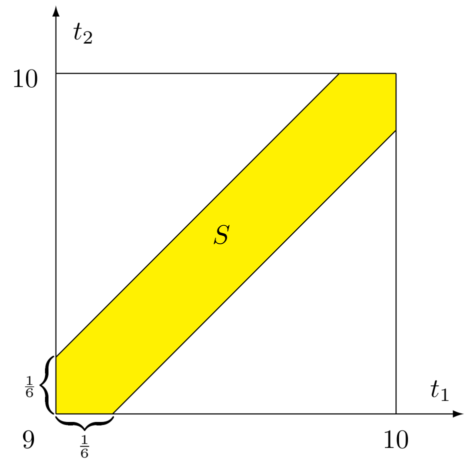
$$S = \textit{“přátelé se setkají”}$$

pak bude vyjádřen jako

$$S = \left\{(t_1, t_2) \in \Omega \mid \textit{intervaly } \langle t_1, t_1 + \tfrac{1}{6}\rangle \textit{ a } \langle t_2, t_2 + \tfrac{1}{6}\rangle \textit{ mají neprázdný průnik}\right\} =$$

$$= \left\{(t_1, t_2) \in \Omega \mid\ |t_1 - t_2| \leq \tfrac{1}{6}\right\}$$

(za jednotku jsme si zvolili hodinu, takže 10 min $= \frac{1}{6}$ hod). Z grafického znázornění množin v $\mathbb{R}^2$



snadno zjistíme, že $\mathrm{vol}(S) = 1 - (\frac{5}{6})^2 = \frac{11}{36}$ a $\mathrm{vol}(\Omega) = 1$, takže

$$P(S) = \frac{11}{36} \ .$$

Poznamenejme, že pokud by každá z osob volila jinou délku doby čekání, úloha by se řešila velmi podobně (tj. hledaly by se opět neprázdné průniky intervalů vyjadřujících dobu pobytu).

2.3 (geometrická pravděpodobnost)
Tyč délky ℓ je náhodně rozdělena na 3 části. Jaká je pravděpodobnost, že z částí lze sestavit trojúhelník?

Řešení:
V našem případě je tyč rozdělena na části o délkách a, b a c, kde $0 < a, b, c$ a $a + b + c = \ell$. Za jevové pole si zvolíme

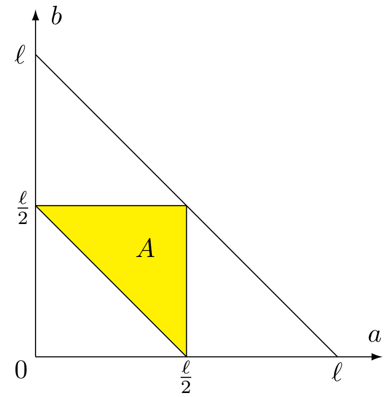
$$\Omega = \left\{(a, b) \in \mathbb{R}^2 \mid 0 < a, b \ \ \& \ \ a + b < \ell\right\},$$

kde původní elementární jev (a, b, c) popíšeme pouze prvními dvěma složkami (a, b) a třetí je jednoznačně určena jako $c = \ell - (a + b)$. Množina Ω je rovnoramenný pravoúhlý trojúhelník s odvěsnami délky ℓ. Jeho plocha tak je $\mathrm{vol}(\Omega) = \frac{\ell^2}{2}$. Abychom mohli sestavit trojúhelník, musí platit trojúhelníková nerovnost. Zajímá nás tedy tudíž jev

$$A = \left\{(a, b) \in \Omega \mid a + b > \underbrace{\ell - (a + b)}_{c} \ \ \wedge \ \ b + \underbrace{\ell - (a + b)}_{c} > a \ \ \wedge \ \ a + \underbrace{\ell - (a + b)}_{c} > b\right\} =$$

$$= \left\{(a, b) \in \Omega \mid a + b > \tfrac{\ell}{2} \ \ \wedge \ \ a, b < \tfrac{\ell}{2}\right\}.$$

Množina A vytváří v množině Ω trojúhelník, jehož vrcholy jsou středy stran trojúhelníku Ω.



Velikost plochy A tak zřejmě je $\mathrm{vol}(A) = \frac{1}{4}\mathrm{vol}(\Omega)$. Proto máme

$$P(A) = \frac{\mathrm{vol}(A)}{\mathrm{vol}(\Omega)} = \frac{1}{4} = 0.25\ .$$

K výpočtu lze použít také jevové pole

$$\Omega' = \left\{(a,b,c) \in \mathbb{R}^3 \mid 0 < a,b,c \quad \wedge \quad a+b+c = \ell\right\}\ .$$

To je rovnostranný trojúhelník o straně délky ℓ. Jeho plocha tak je $\mathrm{vol}(\Omega') = \frac{\sqrt{3}}{4}\ell^2$. Odpovídající jev sestrojení trojúhelníku pak je

$$A' = \left\{(a,b,c) \in \Omega' \mid a+b > c \quad \wedge \quad b+c > a \quad \wedge \quad a+c > b\right\}\ ,$$

který opět v množině Ω' vytváří trojúhelník, jehož vrcholy jsou středy stran trojúhelníku Ω'. Velikost plochy A' je opět $\mathrm{vol}(A') = \frac{1}{4}\mathrm{vol}(\Omega')$. Proto znovu máme $P(A') = \frac{\mathrm{vol}(A')}{\mathrm{vol}(\Omega')} = \frac{1}{4}$.

Co není u podobných příkladů ihned zřejmé, je to, zda a proč budou různé přístupy dávat stejný výsledek. Obecně tomu tak být nemusí. Zde ale zřejmě ano. Důvod je obecněji ten, že máme zobrazení $\varphi : \Omega \to \Omega'$,

$$\varphi(a,b) = (a, b, \ell - a - b)$$

které parametrizuje množinu Ω' pomocí původní množiny Ω a přitom platí $\varphi(A) = A'$. Toto zobrazení je vlastně "natažení" trojúhelníku Ω do podoby trojúhelníku Ω'. Množina A se přitom natáhne stejným způsobem (do množiny A') a proto poměry velikostí zůstanou zachovány. Tedy pravděpodobnost vyjde stejně.

Jak přirozeně definovat nezávislost jevů: Nejdříve si zavedeme podmíněnou pravděpodobnost $P(A|B)$, tj. pravděpodobnost, že nastane jev A za předpokladu, že výsledky se budou omezovat jen na jev B (také to můžeme chápat tak, že nastal jev B a my se *zpětně* ptáme, jaká byla za tohoto předpokladu pravděpodobnost jevu A). Přirozeně to bude $P(A|B) = \frac{P(A\cap B)}{P(B)}$, pokud $P(B) \neq 0$.

To, že jev A nebude záviset na jevu B, si pak přirozeně určíme podmínkou $P(A|B) = P(A)$ a podobně B nebude záviset na jevu A pokud $P(B) = P(B|A)$. Takže jevy A a B budou nezávislé, pokud platí podmínky $P(A|B) = P(A)$ a $P(B) = P(B|A)$ (a také bychom ještě mohli uvažovat i nezávislost na doplňcích $P(A) = P(A|\overline{B})$ atd.).

Jak je ale vidět, všechny tyto podmínky odpovídají jediné rovnici $P(A \cap B) = P(A) \cdot P(B)$, samozřejmě za předpokladu, že $P(A) \neq 0$ a $P(B) \neq 0$.

Proto se nezávislost jevů A a B definuje jako $P(A \cap B) = P(A) \cdot P(B)$ (ať už jsou $P(A)$ nebo $P(B)$ nulové nebo ne).

Podobným způsobem dojdeme k definici pro více jevů jako:
jevy $A_1, \ldots, A_n$ jsou *nezávislé* $\overset{\text{def.}}{\Longleftrightarrow}$ pro každou indexovou podmnožinu $K \subseteq \{1, \ldots, n\}$ platí

$$P\left(\bigcap_{i\in K} A_i\right) = \prod_{i\in K} P(A_i)\ .$$

Tj. pravděpodobnost libovolných průniků je součin pravděpodobností příslušných jevů.

2.4 ((ne)závislost jevů)

Automat vyrábí podložky ve tvaru obdélníka. Tolerance v šířce není dodržena v 8%, tolerance v délce v 7% a v obou rozměrech ve 3% případů.

(a) Rozhodněte, zda jsou porušení tolerance v délce a v šířce závislé nebo nezávislé jevy.

(b) Vypočtěte pravděpodobnost toho, že náhodně vybraný kus má oba rozměry v toleranci.

Řešení:

Uvažujme jevy

$A =$ "*není dodržena tolerance v šířce*",

$B =$ "*není dodržena tolerance v délce*".

Ze zadání máme, že $P(A) = 0.08$, $P(B) = 0.07$ a $P(A \cap B) = 0.03$.

(a) Protože máme $P(A \cap B) = 0.03 \neq 0.08 \cdot 0.07 = P(A) \cdot P(B)$, jsou jevy A a B závislé.

(b) Pro jev

$C =$ "*oba rozměry jsou v toleranci*"

máme $C = \overline{A} \cap \overline{B} = \overline{A \cup B}$. Takže dostáváme

$$P(C) = P(\overline{A \cup B}) = 1 - P(A \cup B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.08 + 0.07 - 0.03 = 0.12$$

a tudíž

$$P(C) = 1 - P(A \cup B) = 1 - 0.12 = 0.88\ .$$

2.5 (nezávislé jevy)

Dva střelci střílí na terč po jedné ráně. Pravděpodobnost, že se první trefí je $p_1 = 0.7$. Pravděpodobnost, že se trefí druhý střelec je $p_2 = 0.8$. Jaká je pravděpodobnost, že

(a) alespoň jeden střelec zasáhne cíl?

(b) první střelec se trefí a druhý ne?

(c) pokud v terči byl právě jeden zásah, trefil se první střelec?

Řešení:

Uvažujme jevy

$S_1 =$ "*první střelec se trefí*",

$S_2 =$ "*druhý střelec se trefí*".

Tyto jevy považujeme (z podstaty zadání) za nezávislé a dále máme $P(S_1) = 0.7$ a $P(S_2) = 0.8$.

(a) Pro jev

$A =$ "*alespoň jeden střelec zasáhne cíl*" máme $A = S_1 \cup S_2$ a tedy

$$P(A) = P(S_1 \cup S_2) = P(S_1) + P(S_2) - P(S_1 \cap S_2) \overset{\text{(nezav.)}}{=}$$

$$\overset{\text{(nezav.)}}{=} P(S_1) + P(S_2) - P(S_1) \cdot P(S_2) = 0.7 + 0.8 - 0.7 \cdot 0.8 = 0.94 \ .$$

(b) Pro jev

$B =$ "*první střelec se trefí a druhý ne*"

máme $B = S_1 \cap \overline{S_2}$ a tedy

$$P(B) = P(S_1 \cap \overline{S_2}) \overset{\text{(nezav.)}}{=} P(S_1) \cdot P(\overline{S_2}) = 0.7 \cdot (1 - 0.8) = 0.14 \ .$$

(c) Zde budeme oproti předchozím příkladům hledat tzv. aposteriorní (tj. následnou) a tudíž podmíněnou pravděpodobnost. Označme si jev

$C =$ "*právě jeden ze střelců se trefí*"

a hledáme $P(S_1|C)$. Máme $C = (S_1 \cap \overline{S_2}) \cup (\overline{S_1} \cap S_2)$. Pro výpočet si ještě určíme, že

$$S_1 \cap C = \text{“ první střelec se trefí a druhý ne”} = S_1 \cap \overline{S_2} \ .$$

Tedy

$$P(S_1|C) = \frac{P(S_1 \cap C)}{P(C)} = \frac{P(S_1 \cap \overline{S_2})}{P(S_1 \cap \overline{S_2}) + P(\overline{S_1} \cap S_2)} = \frac{0.7 \cdot 0.2}{0.7 \cdot 0.2 + 0.3 \cdot 0.8} = \frac{7}{19} \doteq 0.3684 \ .$$

kde jsme použili, že jevy $S_1 \cap \overline{S_2}$ a $\overline{S_1} \cap S_2$ jsou neslučitelné (disjunktní).

2.6 (výběr s proměnnými podmínkami, Stirlingův vzorec)

V urně jsou dvě koule, bílá a černá. Provádíme výběr po jedné kouli do doby, než vytáhneme černou kouli. Kdykoliv se vytáhne bílá koule, vrátí se do urny a přidají se ještě dvě bílé koule. Určete pravděpodobnost toho, že se při prvních 50 tazích vytáhnou pouze bílé koule.

Řešení:
Použijeme následující vztah, který si obecně hodí, pokud provádíme sérii pokusů:

Mějme posloupnost jevů $A_1 \supseteq A_2 \cdots \supseteq A_n$ kde $P(A_n) \neq 0$. Protože máme rovnost $A_i = A_i \cap A_{i-1}$, tak dostáváme

$$P(A_i) = \frac{P(A_i \cap A_{i-1})}{P(A_{i-1})} \cdot P(A_{i-1}) = P(A_i|A_{i-1}) \cdot P(A_{i-1}) \ .$$

Postupně tak iterací dostaneme

$$P(A_n) = P(A_1) \cdot P(A_2|A_1) \cdots P(A_n|A_{n-1}) \ .$$

Při praktickém použití pak jev A_i znamená výsledky série prvních i pokusů. Podmíněné pravděpodobnosti $P(A_i|A_{i-1})$ pak znamenají, jaká je pravděpodobnost výsledků i-tého pokusu za předpokladu, že už nastaly dané výsledky prvních $i-1$ pokusů.

V našem případě budeme mít

$A_i =$ "*v prvních i tazích se vytáhne bílá koule*",

podmíněné pravděpodobnosti pak budou

$$P(A_i|A_{i-1}) = \frac{2i-1}{2i}$$

protože po prvních $i-1$ pokusech v urně je $2i-1$ bílých koulí a 1 černá koule. A samozřejmě $P(A_1) = \frac{1}{2}$. Pro $n = 50$ tedy máme

$$P(A_n) = \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdots \frac{2n-3}{2n-2} \cdot \frac{2n-1}{2n} = \frac{1 \cdot 2 \cdots (2n-1) \cdot 2n}{\big(2 \cdot 4 \cdots (2n-2) \cdot 2n\big)^2} = \frac{(2n)!}{(n!)^2 \cdot 2^{2n}} = \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} .$$

Použitím Stirlingova vzorce $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ dostaneme přibližnou hodnotu

$$P(A_n) = \frac{(2n)!}{(n!)^2 \cdot 2^{2n}} \doteq \frac{\sqrt{4\pi n}\left(\frac{2n}{e}\right)^{2n}}{2\pi n\left(\frac{n}{e}\right)^{2n} \cdot 2^{2n}} = \frac{1}{\sqrt{\pi n}} ,$$

tedy pro $n = 50$ to je

$$P(A_{50}) = \frac{100!}{(50!)^2 \cdot 2^{100}} \doteq \frac{1}{\sqrt{50\pi}} \doteq 0.0798 .$$

Všimněme si, ze výsledek je v jistém smyslu v rozporu s intuicí: Jestliže máme úspěšnou sérii n vytažených bílých koulí, tak pravděpodobnost, že ji prodloužíme o další tah se blíží k jedné

$$P(A_{n+1}|A_n) = \tfrac{2n+1}{2(n+1)} \overset{n\to\infty}{\longrightarrow} 1$$

ale současně pravděpodobnost, že takováto úspěšná série vznikne se postupně blíží k nule

$$P(A_n) \doteq \frac{1}{\sqrt{\pi n}} \overset{n\to\infty}{\longrightarrow} 0 .$$