Matematika I – 2. přednáška
Pravděpodobnost

Martin Panák

Masarykova univerzita
Fakulta informatiky

25. 2. 2009

Doporučené zdroje

- Martin Panák, Jan Slovák, **Drsná matematika**, *e*-text.
- Karel Zvára, Josef Štěpán, **Pravděpodobnost a matematická statistika**, Matfyzpress, 4. vydání, 2006, 230 stran, ISBN 80-867-3271-1.
- Marie Budíková, Štěpán Mikoláš, Pavel Osecký, **Teorie pravděpodobnosti a matematická statistika (sbírka příkladů)**, Masarykova univerzita, 3. vydání, 2004, 117 stran, ISBN 80-210-3313-4.
- Marie Budíková, Štěpán Mikoláš, Pavel Osecký, **Popisná statistika**, Masarykova univerzita, 3. vydání, 2002, 48 stran, ISBN 80-210-1831-3.
- Marie Budíková, Tomáš Lerch, Štěpán Mikoláš, **Základní statistické metody**, Masarykova univerzita, 2005, 170 stran, ISBN 80-210-3886-1.

Princip inkluze a exkluze

$$
\begin{aligned}
|A_1 \cup A_2 \cup \cdots \cup A_n| &= |A_1| + |A_2| + \cdots + |A_n| \\
&\quad -|A_1 \cap A_2| - |A_1 \cap A_3| - \cdots - |A_{n-1} \cap A_n| \\
&\quad +|A_1 \cap A_2 \cap A_3| + \cdots + |A_{n-2} \cap A_{n-1} \cap A_n| \ldots \\
&\quad +(-1)^{n+1}|A_1 \cap \cdots A_n| \\
&= \sum_{r=1}^{n} (-1)^{r+1} \left(\sum_{i_1 < \cdots < i_r} |A_{i_1} \cap \cdots \cap A_{i_r}| \right)
\end{aligned}
$$

Klasická pravděpodobnost

Připomeňme si stručně konečnou pravděpodobnost, kterou jsme si definovali již na minulé přednášce.

Klasická pravděpodobnost

Připomeňme si stručně konečnou pravděpodobnost, kterou jsme si definovali již na minulé přednášce.

Definition

Nechť Ω je konečný základní prostor a nechť jevové pole $\mathcal{A}$ je právě systém všech podmnožin v Ω. *Klasická pravděpodobnost* je pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ s pravděpodobnostní funkcí $P : \mathcal{A} \to \mathbb{R}$,

$$P(A) = \frac{|A|}{|\Omega|}.$$

Zjevně takto zadaná funkce skutečně definuje pravděpodobnost, kdy všem elementárním jevům přiřazujeme stejnou pravděpodobnost.

Příklad. *Vrhneme dvě šestiboké kostky. Jaká je pravděpodobnost, že padne součet šest?*

Příklad. *Vrhneme dvě šestiboké kostky. Jaká je pravděpodobnost, že padne součet šest?*

Příklad. *Šest lidí vhodí svoje peněženky do kloubouku a poté si každý vylosuje jednu peněženku zpět. Jaká je pravděpodobnost, že si nikdo nevylosuje zpět svoji peněženku?*

Podmíněná pravděpodobnost a nezávislost

Motto:

Je dokázáno, že slavení narozenin je zdraví prospěšné. Statistika ukazuje, že lidé, kteří oslavili nejvíce narozenin, se dožívají nejvyššího věku.

Podmíněná pravděpodobnost a nezávislost

Motto:

Je dokázáno, že slavení narozenin je zdraví prospěšné. Statistika ukazuje, že lidé, kteří oslavili nejvíce narozenin, se dožívají nejvyššího věku.

Obvyklé je také klást dotazy s dodatečnou podmínkou. Např.

- Jaká je pravděpodobnost, že při hodu dvěma kostkami padly dvě pětky, je-li součet hodnot deset?

Podmíněná pravděpodobnost a nezávislost

Motto:

Je dokázáno, že slavení narozenin je zdraví prospěšné. Statistika ukazuje, že lidé, kteří oslavili nejvíce narozenin, se dožívají nejvyššího věku.

Obvyklé je také klást dotazy s dodatečnou podmínkou. Např.

- Jaká je pravděpodobnost, že při hodu dvěma kostkami padly dvě pětky, je-li součet hodnot deset?
- Mějme urnu s 10 koulemi. Desetkrát jsem vytáhl kouli, zkontroloval její barvu a vrátil do urny. Jestliže byla vždy bílé barvy, s jakou pravděpodobností jsou všechny koule v urně bílé?

Podmíněná pravděpodobnost a nezávislost

Motto:

Je dokázáno, že slavení narozenin je zdraví prospěšné. Statistika ukazuje, že lidé, kteří oslavili nejvíce narozenin, se dožívají nejvyššího věku.

Obvyklé je také klást dotazy s dodatečnou podmínkou. Např.

- Jaká je pravděpodobnost, že při hodu dvěma kostkami padly dvě pětky, je-li součet hodnot deset?
- Mějme urnu s 10 koulemi. Desetkrát jsem vytáhl kouli, zkontroloval její barvu a vrátil do urny. Jestliže byla vždy bílé barvy, s jakou pravděpodobností jsou všechny koule v urně bílé?
- Na dostizích jsou známy pravděpodobnosti vítězství jednotlivých koní. Jak se tyto pravděpodobnosti změní, pokud uprostřed závodu spadne jezdec jednoho z koní ze sedla?

Připomeneme, že formalizovat takové úvahy umíme následovně.

Definition

Nechť H je jev s nenulovou pravděpodobností v jevovém poli $\mathcal{A}$ v pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$. *Podmíněná pravděpodobnost* $P(A|H)$ jevu $A \in \mathcal{A}$ vzhledem k jevu H je definována vztahem

$$P(A|H) = \frac{P(A \cap H)}{P(H)}.$$

Připomeneme, že formalizovat takové úvahy umíme následovně.

Definition

Nechť H je jev s nenulovou pravděpodobností v jevovém poli $\mathcal{A}$ v pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$. *Podmíněná pravděpodobnost* $P(A|H)$ jevu $A \in \mathcal{A}$ vzhledem k jevu H je definována vztahem

$$P(A|H) = \frac{P(A \cap H)}{P(H)}.$$

Přirozená definice nezávislosti je, že hypotéza H a jev A jsou nezávislé tehdy, je-li $P(A) = P(A|H)$.
Z výše uvedeného snadno vyplývá *symetričtější* definice:

Připomeneme, že formalizovat takové úvahy umíme následovně.

Definition

Nechť H je jev s nenulovou pravděpodobností v jevovém poli $\mathcal{A}$ v pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$. *Podmíněná pravděpodobnost* $P(A|H)$ jevu $A \in \mathcal{A}$ vzhledem k jevu H je definována vztahem

$$P(A|H) = \frac{P(A \cap H)}{P(H)}.$$

Přirozená definice nezávislosti je, že hypotéza H a jev A jsou nezávislé tehdy, je-li $P(A) = P(A|H)$.
Z výše uvedeného snadno vyplývá *symetričtější* definice:

Definition

Říkáme, že jevy A a B jsou nezávislé, jestliže

$$P(A \cap B) = P(A)P(B).$$

Definition

Říkáme, že jevy $A_1, A_2, \ldots$ jsou nezávislé, jestliže pro každou k-tici $A_{i_1}, \cdots, A_{i_k}$ z nich platí

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Definition

Říkáme, že jevy $A_1, A_2, \ldots$ jsou nezávislé, jestliže pro každou k-tici $A_{i_1}, \cdots, A_{i_k}$ z nich platí

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Example

V urně jsou 4 lístky označené 000, 110, 101, 011. Uvažujme pro $i = 1, 2, 3$ náhodné jevy
$A_i = \{$náhodně vytažený lístek má na i-tém místě 1$\}$.

Definition

Říkáme, že jevy $A_1, A_2, \dots$ jsou nezávislé, jestliže pro každou k-tici $A_{i_1}, \cdots, A_{i_k}$ z nich platí

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Example

V urně jsou 4 lístky označené 000, 110, 101, 011. Uvažujme pro $i = 1, 2, 3$ náhodné jevy
$A_i = \{$náhodně vytažený lístek má na i-tém místě 1$\}$.
Snadno se vidí, že $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$, dále, že $P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}$ a že $P(A_1 \cap A_2 \cap A_3) = 0$.

Definition

Říkáme, že jevy $A_1, A_2, \ldots$ jsou nezávislé, jestliže pro každou k-tici $A_{i_1}, \cdots, A_{i_k}$ z nich platí

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Example

V urně jsou 4 lístky označené 000, 110, 101, 011. Uvažujme pro $i = 1, 2, 3$ náhodné jevy
$A_i = \{$náhodně vytažený lístek má na i-tém místě 1$\}$.
Snadno se vidí, že $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$, dále, že $P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}$ a že $P(A_1 \cap A_2 \cap A_3) = 0$. Jevy A_1, A_2, A_3 jsou tedy po dvou nezávislé, ale nejsou nezávislé.

Bayesovy věty

Přepsáním formule pro podmíněnou pravděpodobnost dostáváme

$$P(A \cap B) = P(B \cap A) = P(A)P(B|A) = P(B)P(A|B).$$

Theorem (Bayesovy věty)

Pro pravděpodobnost jevů A a B platí

1. $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$.
2. $P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^c)P(B|A^c)}$.

Bayesovy věty

Přepsáním formule pro podmíněnou pravděpodobnost dostáváme

$$P(A \cap B) = P(B \cap A) = P(A)P(B|A) = P(B)P(A|B).$$

Theorem (Bayesovy věty)

Pro pravděpodobnost jevů A a B platí

1. $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$.
2. $P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^c)P(B|A^c)}$.

Důkaz.

První tvrzení je přepsáním předchozí formule, druhé z prvého plyne doszením $P(B) = P(A)P(B|A) + P(A^c)P(B|A^c)$. $\square$

Příklad – preventivní screening

Předpokládejme, že krevní test na HIV pozitivní osoby má 99% správnost v případě osoby skutečně HIV pozitivní (*vysoká citilivost – sensitivity*). Zároveň předpokládejme, že u HIV negativní osoby dopadně test pozitivně v $0,2\%$ případů (*relativně vysoká specifičnost – specificity*).

Příklad – preventivní screening

Předpokládejme, že krevní test na HIV pozitivní osoby má 99% správnost v případě osoby skutečně HIV pozitivní (*vysoká citilivost – sensitivity*). Zároveň předpokládejme, že u HIV negativní osoby dopadně test pozitivně v $0,2\%$ případů (*relativně vysoká specifičnost – specificity*).
Náhodně z populace vyberem osobu a otestujeme pozitivně.
S jakou pravděpodobností je skutečně HIV pozitivní, jestliže četnost výskytu HIV v populaci je p promile (tj. p osob z tisíce je skutečně HIV pozitivní).

Příklad – preventivní screening

Předpokládejme, že krevní test na HIV pozitivní osoby má 99% správnost v případě osoby skutečně HIV pozitivní (*vysoká citilivost – sensitivity*). Zároveň předpokládejme, že u HIV negativní osoby dopadně test pozitivně v $0,2\%$ případů (*relativně vysoká specifičnost – specificity*).
Náhodně z populace vyberem osobu a otestujeme pozitivně.
S jakou pravděpodobností je skutečně HIV pozitivní, jestliže četnost výskytu HIV v populaci je p promile (tj. p osob z tisíce je skutečně HIV pozitivní).
Označme A jev, že je daná osoba HIV pozitivní, a B jev, že daná osoba má pozitivní test. Dle druhé Bayesovy věty je hledaná pravděpodobnost

$$P(A|B) = \frac{p/1000 \cdot 99/100}{p/1000 \cdot 99/100 + (1000 - p)/1000 \cdot 2/1000}$$

Příklad – preventivní screening, pokr.

Jestliže zvolíme za p nějaké konkrétní četnosti, dostaneme příslušné očekávatelné spolehlivosti testu. V následující tabulce je spočten výsledek pro několik p:

p	100	10	1	0,1
$P(A\|B)$	0,982	0,8333	0,3313	0,0471

Výsledek asi neodpovídá naší intuici a může se zdát šokující ve vztahu k použití takovýchto testů.

Geometrická pravděpodobnost

V praktických problémech se často setkáváme s daleko složitějšími modely, kde základní prostor není konečnou množinou. Nemáme momentálně k dispozici ani základní nástroje pro dostatečné zobecnění pojmu pravděpodobnosti, nicméně můžeme uvést alespoň jednoduchou ilustraci.

Geometrická pravděpodobnost

V praktických problémech se často setkáváme s daleko složitějšími modely, kde základní prostor není konečnou množinou. Nemáme momentálně k dispozici ani základní nástroje pro dostatečné zobecnění pojmu pravděpodobnosti, nicméně můžeme uvést alespoň jednoduchou ilustraci.

Uvažme rovinu $\mathbb{R}^2$ dvojic reálných čísel a v ní podmnožinu Ω se známým obsahem $\operatorname{vol}\Omega$ (symbol „vol" od anglického „volume", tj. obsah/objem). Příkladem může sloužit třeba jednotkový čtverec. Náhodné jevy budou reprezentovány podmnožinami $A \subset \Omega$ za jevové pole $\mathcal{A}$ bereme systém podmnožin, u kterých umíme určit jejich obsah. Třeba všechna konečná sjednocení trojůhelníků. Nastoupení nebo nenastoupení jevu je dáno výběrem bodu v Ω, kterým se trefíme nebo netrefíme do množiny reprezentující jev A.

Podobně jako u klasické pravděpodobnosti pak definujeme pravděpodobnostní funkci $P : \mathcal{A} \to \mathbb{R}$ vztahem

$$P(A) = \frac{\operatorname{vol} A}{\operatorname{vol} \Omega}.$$

Uvažme jako příklad problém, kdy náhodně vyberem dvě hodnoty $a < b$ v intervalu $(0, 1) \subset \mathbb{R}$. Všechny hodnoty a i b jsou stejně pravděpodobné a otázka zní „jaká je pravděpodobnost, že interval (a, b) bude mít velikost alespoň jedna polovina?“.

Uvažme jako příklad problém, kdy náhodně vyberem dvě hodnoty $a < b$ v intervalu $(0, 1) \subset \mathbb{R}$. Všechny hodnoty a i b jsou stejně pravděpodobné a otázka zní „jaká je pravděpodobnost, že interval (a, b) bude mít velikost alespoň jedna polovina?".

Odpověď je docela jednoduchá: volba čísel a, b je volbou libovolného bodu (a, b) ve vnitřku trojúhelníku Ω s hraničními vrcholy $[0, 0]$, $[0, 1]$, $[1, 1]$ (načrtněte si obrázek!). Potřebujeme znát plochu podmnožiny, která odpovídá bodům s $b > a + \frac{1}{2}$, tj. vnitřku trojúhelníku A ohraničeného vrcholy $[0, \frac{1}{2}]$, $[0, 1]$, $[\frac{1}{2}, 1]$. Evidentně dostáváme $P(A) = \frac{1}{4}$. Zkuste si samostatně odpovědět na otázku „pro jakou požadovanou minimální délku intervalu (a, b) dostaneme pravděpodobnost jedna polovina?".

Jednou z účinných výpočetních metod přibližných hodnot je naopak simulace známé takovéto pravděpodobnosti pomocí relativní četnosti nastoupení vhodně zvoleného jevu. Např. známá formule pro obsah kruhu o daném poloměru říká, že obsah jednotkového kruhu je roven právě konstantě $\pi = 3,1415\dots$, která vyjadřuje poměr obsahu a čtverce poloměru. Pokud zvolíme za Ω jednotkový čtverec a za A průnik Ω a jednotkového kruhu se středem v počátku, pak $\operatorname{vol} A = \frac{1}{4}\pi$. Máme-li tedy spolehlivý generátor náhodných čísel mezi nulou a jedničkou a počítáme relativní četnosti, jak často bude vzdálenost vygenerované dvojice (a, b) menší než jedna, tj. $\sqrt{a^2 + b^2} < 1$, pak výsledek bude při velkém počtu pokusů s velikou jistotou dobře aproximovat číslo $\frac{1}{4}\pi$. Numerickým postupům založeným na tomto principu se říká *metody Monte Carlo*.

Příklad. *Mirek a Marek chodí na obědy do univerzitní menzy. Do menzy je možné přijít kdykoliv mezi jedenáctou a čtrnáctou hodinou. Každý z nich stráví na obědě půl hodiny a dobu příchodu (mezi 11h a 14h) si vybírá náhodně. Jaká je pravděpodobnost, že se na obědě v daný den potkají, sedávají-li oba u stejného stolu?*