Pravděpodobnost a popisná statistika

Helena Durnová

22. března 2012

Obsah

Kapitola 1

Úvodní poznámky

Toto je pracovní verze studijního textu pro předmět *Pravděpodobnost a popisná statistika* pro studenty matematiky Pedagogické fakulty Masarykovy univerzity v Brně.

Obtížnější příklady jsou označeny hvězdičkou (*).

Připomínky vítám. Pište, prosím, na adresu hdurnova@ped.muni.cz

1.1 Náhodné jevy

Motivační úloha Jaká je pravděpodobnost toho, že při současném hodu dvěma kostkami padne číslo 10?

Možné součty: $10 = 6 + 4 = 5 + 5$

Jaká je pravděpodobnost toho, že při současném hodu dvěma kostkami padne číslo 9?

Možné součty: $9 = 6 + 3 = 5 + 4$

Možných rozkladů čísel 9 a 10 je stejně, přesto součet 9 padá častěji než součet 10 (empiricky zjištěno).

Náhodný pokus

Definice 1.1 Neprázdnou množinu všech možných výsledků náhodného pokusu nazýváme *základní prostor* a označujeme Ω.

Prvky množiny Ω označujeme ω_t, kde $t \in T$ je vhodný index.

Definice 1.2 Systém podmnožin $\mathcal{A}$ základního prostoru Ω, který

a) obsahuje základní prostor;

b) s každými dvěma podmnožinami obsahuje i jejich rozdíl; a

c) s každým konečným [spočetným] systémem množin obsahuje i jejich sjednocení

nazýváme *jevové pole*.

1.2 Závislé a nezávislé jevy

Závislé jevy: opakované výběry bez vracení
- používáme větu o násobení pravděpodobností

Nezávislé jevy: opakované výběry s vracením, opakované hody kostkou
- pravděpodobnosti násobíme

1.3 Příklady: jevové pole

Náhodný pokus: házení kostkou

Možné výsledky: 1, 2, 3, 4, 5, 6; tj. $\Omega = \{1, 2, 3, 4, 5, 6\}$ (jednotlivá čísla označují skutečnost, že padlo dané číslo)

$\mathcal{A} = \{\emptyset, \Omega\}$ — triviální jevové pole

[Otázka: musí být prázdná množina prvkem $\mathcal{A}$?]

$\mathcal{B} = \{\emptyset, \Omega, \{2, 4, 6\}\}$

$\mathcal{C} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$

$\mathcal{D} = \{\emptyset, \Omega, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$

Náhodný jev: libovolný prvek jevového pole (množina).

[Úkol: nazvěte některé náhodné jevy podle $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$.]

Definice 1.3 Dvojice $(\Omega, \mathcal{A})$ se nazývá *měřitelný prostor.*

Označení: *Jev jistý:* Ω

Jev nemožný: $\emptyset$

Jev elementární: ω pro $\omega \in \Omega$

Společné nastoupení jevů A_i*:* $\bigcap_i \in IA_i$

Nastoupení alespoň jednoho z jevů A_i*:* $\bigcup_i \in IA_i$

Jev opačný k jevu A: $\overline{A_i} = \Omega \setminus A_i$

Definice 1.4 Pravděpodobností *rozumíme reálnou množinovou funkci* $P : \mathcal{A} \to \mathbb{R}$*, která je*

a) nezáporná;

b) spočetně aditivní; a

c) normovaná.

Trojici $(\Omega, \mathcal{A}, P)$ *(tj. základní prostor, jevové pole, pravděpodobnostní funkce) nazýváme* pravděpodobnostní prostor*.*

Kapitola 2

Klasická pravděpodobnost

$$P(A) = \frac{m(A)}{m(\Omega)}$$

(počet příznivých jevů lomeno počet všech možných jevů).

Lze použít pouze tehdy, jsou-li pravděpodobnosti všech elementárních jevů stejné.

Věta 2.1 (Věta o sčítání pravděpodobností) Nechť $A_1, A_2, \ldots, A_n \in \mathcal{A}$ jsou libovolné jevy. Pak platí:

$$P(\bigcup_{i \in I}^{n} A_i = \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i \cap A_j) + \cdots$$
$$\cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

Příklad 2.2 Z množiny zvané základní soubor rozsahu n vybereme k-krát po jednom prvku, který vždy vrátíme zpět. Získáme uspořádanou k-tici, která se nazývá *uspořádaný výběrový soubor k prvků s vracením.* Předpokládejme, že v základním souboru je právě r prvků označeno.

Vypočtěte pravděpodobnosti jevů A, B, C, které jsou definovány takto:

A - každý z prvků základního souboru se ve výběrovém souboru vyskytne nejvýše jedenkrát

B - předem daný prvek základního souboru se ve výběrovém souboru vyskytne nejvýše jedenkrát

C - ve výběrovém souboru se vyskytne právě x označených prvků

2.1 Geometrická pravděpodobnost

$$P(A) = \frac{V(A)}{V(E)}$$

Buffonova úloha o jehle – řešení i s obrázky viz http://mant.upol.cz/soubory/OdevzdanePrace/B08/b08-20-ls.pdf

Kapitola 3

Bayesův vzorec

3.1 Podmíněná pravděpodobnost

Nechť je dán základní prostor Ω, jev A a jev B. Pravděpodobnost toho, že za podmínky B nastal jev A, vypočteme takto:

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

Např. $\Omega\{1, 2, 3, 4, 5, 6\}$ - výsledky hodu kostkou
jev $A = \{3, 4, 5, 6\}$ - padne číslo větší než 2
jev $B = \{1, 6\}$ - padne 1 nebo 6

$$P(A) = \frac{2}{3}, P(B) = \frac{1}{3}$$

$$P(A|B) = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2}$$

Násobení pravděpodobností Z klasické definice pravděpodobnosti víme, že

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

Problém: jak určit $|A \cap B|$?
Ze vzorce pro podmíněnou pravděpodobnost plyne, že

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

a analogicky

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

Pro pravděpodobnost průniku 3 a více jevů $A_1, A_2, \ldots, A_n$ pak platí:

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

Definice 3.1 (Nezávislé jevy) *Říkáme, že jev A nezávisí na jevu B, pokud platí*

$$P(A|B) = P(A)$$

a pak tedy také

$$P(B|A) = P(B)$$

Zřejmě potom platí, že

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

Sčítání pravděpodobností Vyjdeme-li z klasické definice pravděpodobnosti, zřejmě

$$A = (A \cap B) \cup (A \cap \overline{B}),$$

kde $\overline{B}$ je doplněk jevu B. Tedy

$$P(A) = P(A \cap B) + P(A \cap \overline{B})$$

Dále zřejmě

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Pro pravděpodobnost sjednocení více jevů použijeme princip inkluze a exkluze.

3.2 Úplná pravděpodobnost

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_n)P(A|B_n)$$

Bayesova věta odpovídá na otázku, jaká je pravděpodobnost, že nastal jev B_i, víme-li, že nastal jev A:

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{k=1}^{n} P(B_k)P(A|B_k)}$$

Kapitola 4

Popisná statistika

Kdy nestačí klasická definice pravděpodobnosti?
- geometrická pravděpodobnost je jen model klasické pravděpodobnosti
- podstatné: nastoupení lib. jevu má stejnou možnost, tj. žádný jev nemá přednost před ostatními

Statistická definice pravděpodobnosti - nazývaná také frekvenční či empirická

Opakujeme-li n-krát nezávisle daný pokus a nastane-li v těchto pokusech sledovaný jev A m-krát, potom jeho relativní četnost je rovna zlomku m/n. Bude-li při rostoucím počtu pokusů relativní četnost kolísat ve stále užších mezích kolem určitého čísla, můžeme předpokládat, že toto číslo je pravděpodobností jevu A.
- absolutní četnost
- relativní četnost

Borelovské pole a borelovská množina: Minimální jevové pole na $\mathcal{R}^n$ obsahující třídu všech intervalů $(-\infty, x_1 > \times(-\infty, x_2 > \times(-\infty, x_3 > \cdots \times (-\infty, x_{n-1} > \times(-\infty, x_{n-1} >$ se nazývá borelovské pole, jeho prvky borelovské množiny.

Náhodná veličina formálně: Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Zobrazení $X : \Omega \to \mathcal{R}$ se nazývá *náhodná veličina* (vzhledem k jevovému poli $\mathcal{A}$), právě tehdy, když úplný vzor každé borelovské množiny je jevem, tj. $\forall B\ \mathcal{B} : \{\omega \in \Omega : X\omega \in B\} \in \mathcal{A}$.

Náhodná veličina může být diskrétní nebo spojitá.

4.1 Distribuční funkce

- neklesající

- zprava spojitá
- $\lim_{x\to\infty} \Phi(x) = 1$, $\lim_{x\to-\infty} \Phi(x) = 0$
- $0 \leq \Phi(x) \leq 1$
- pro $x_0 \in \mathbb{R}$ platí: $P(X = x) = \Phi(x_0) - \lim_{x\to x_0^-} \Phi(x)$
- pro $a, b \in \mathbb{R}, a < b$ platí $p(a < X \leq b) = \Phi(b) - \Phi(a)$

Diskrétní náhodná veličina

- pravděpodobnostní funkce $\pi(x)$
  - nezáporná
  - normovaná, tj. $\sum_{-\infty}^{\infty} \pi(x) = 1$
- distribuční funkce: $\Phi(x) = \sum_{t=-\infty}^{x} \pi(t)$

Spojitá náhodná veličina

- hustota pravděpodobnosti $\varphi(x)$
  - nezáporná
  - normovaná, tj. $\int_{-\infty}^{\infty} \varphi(x) = 1$
- distribuční funkce $\Phi(x) = \int_{t=-\infty}^{x} \varphi(t)dt$

Základní a výběrový soubor

- základní soubor E (neprázdná množina)
- podmnožina základního souboru G - prvky s danou vlastností
- výběrový soubor (neprázdná podmnožina výběrového souboru)
- rozsah výběrového souboru n
- absolutní četnost G ve výběrovém souboru $N(G)$
- relativní četnost G ve výběrovém souboru $p(G) = \frac{N(G)}{n}$

Vlastnosti relativní četnosti

- $p(\emptyset) = 0$
- $0 \leq p(G) \leq 1$
- $p(E) = 1$
- $p(G) \leq 1$
- $p(G) + p(\overline{G}) = 1$
- $p(G_1 \cup G_2) + (G_1 \cap G_2) = p(G_1) + p(G_2)$
- $1 + (G_1 \cap G_2) \geq p(G_1) + p(G_2)$
- $p(G_1 \cup G_2) + 0 \leq p(G_1) + p(G_2)$
- $G_1 \subseteq G_2 \Rightarrow p(G_2 \setminus G_2) = p(G_2) - p(G_1)$
- $G_1 \subseteq G_2 \Rightarrow p(G_1) \leq p(G_2)$

Podmíněná relativní četnost podobně jako podmíněná pravděpodobnost (klasická definice pravděpodobnosti)

$$p(G_1|G_2) = \frac{N(G_1 \cap G_2)}{N(G_2)}$$

Četnostně nezávislé podmnožiny G_1, G_2 Říkáme, že G_1 a G_2 jsou četnostně nezávislé, platí-li

$$p(G_1 \cap G_2) = p(G_1)p(G_2)$$

Nominální znaky - hodnoty znaku představují jen číselné kódy kvalitativních pojmenování (např. čísla tramvají)

Ordinální znaky - uspořádání znaků má smysl (např. známky ve škole)

$$x_1 \leq x_2 \leq \ldots \leq x_c \leq x_\vartheta \leq x_{c+1} \leq \ldots \leq x_n$$

Intervalové znaky - připouštějí uspořádání a navíc operaci rozdílu (např. teplota)

Poměrové znaky
- připouštějí uspořádání, operaci rozdílu a navíc i operaci podílu

Alternativní znaky
- nabývají pouze dvou hodnot (úspěch-neúspěch, žena-muž)

Charakteristiky znaků:

- modus: nejčetnější hodnota (pro nominální znaky)
- dolní kvartil
- medián
- horní kvartil
- percentily
- kvartilová odchylka

Charakteristiky polohy

- aritmetický průměr

Charakteristiky variability

- průměrná odchylka
- rozptyl
- směrodatná odchylka

Literatura

1. Marie Budíková, Štěpán Mikoláš, Pavel Osecký: *Popisná statistika.* PřF MU Brno 1998.
2. Marie Budíková, Štěpán Mikoláš, Pavel Osecký: *Teorie pravděpodobnosti a matematická statistika. Sbírka příkladů.* PřF MU Brno 1996.
3. Jaroslav Hátle, Jana Kahounová: *Úvod do teorie pravděpodobnosti.* Praha: SNTL, 1987.

Kapitola 5

Pravděpodobnost a popisná statistika 1 - otázky a příklady ke kolokviu

1. Jevové pole. Elementární jev, jev jistý a nemožný.
2. Pravděpodobnostní prostor.
3. Klasická definice pravděpodobnosti.
4. Věta o sčítání pravděpodobností.
5. Věta o násobení pravděpodobností.
6. Geometrická pravděpodobnost.
7. Podmíněná pravděpodobnost.
8. Úplná pravděpodobnost.
9. Bayesův vzorec.
10. Nezávislé jevy.
11. Statistická definice pravděpodobnosti.
12. Diskrétní a spojitá náhodná veličina, distribuční funkce, pravděpodobnostní funkce a hustota pravděpodobnosti.
13. Relativní četnost. Vlastnosti relativní četnosti.
14. Nominální, ordinální, intervalové, poměrové a alternativní znaky a jejich číselné charakteristiky.
15. Charakteristiky polohy a variability.