PRAVDEPODOBNOSŤ a ŠTATISTIKA

Podmienená pravdepodobnosť

doc. RNDr. Štefan Peško

Katedra matematických metód a operačnej analýzy, FRI ŽU

4. marca 2019

Podmienená pravdepodobnosť

Nech $(\Omega, \mathcal{F}, \mathcal{P})$ je pravdepodobnostný priestor a A, B náhodné udalosti také, že $\mathcal{P}(B) > 0$. Potom pravdepodobnosťou udalosti A podmienenou udalosťou B rozumieme podiel

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}. \tag{1}$$

Poznámka:

Podmienená pravdepodobnosť slúži na určenie pp. náhodnej udalosti A, pričom je známa doplňujúca informácia o výskyte inej náhodnej udalosti B.

Príklad 2.1

V rodine sú dve deti. Určme pravdepodobnosť že obe deti sú dievčatá ak je jedno z nich dievča.

V výberový priestor $\Omega = \{(d, d), (d, ch), (ch, d), (ch, ch)\}$, kde dieťa „d“ je dievča a „ch“ je chlapec. Usporiadanie v dvojici hovorí o tom, ktoré dieťa sa narodilo skôr.

Nech náhodná udalosť $A =$ „obe deti sú dievčatá“ $= \{(d, d)\}$ a $B =$ „aspoň jedno z detí je dievča“ $= \{(d, d), (d, ch), (ch, d)\}$. Zrejme $\mathcal{P}(A) = \frac{|A|}{|\Omega|} = \frac{1}{4}$ a $\mathcal{P}(B) = \frac{|B|}{|\Omega|} = \frac{3}{4}$. Potom zo vzťahu (1)

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)} = \frac{\mathcal{P}(A)}{\mathcal{P}(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Príklad 2.2

Overme, že podmienená pravdepodobnosť definovaná vzťahom (1) pri pevne zvolenej podmieňujúcej náhodnej udalosti v pp. priestore $(\Omega, \mathcal{F}, \mathcal{P})$ je pp. v zmysle axiomatickej definície pravdepodobnosti.

Treba overiť platnosť axióm pp. podmienených udalostí $A|B$, kde $\mathcal{P}(B) > 0$.

Axióma 1. Pre každé $A, B \in \mathcal{F}$ je $A \cap B \subset B$ a tak $0 \leq \mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)} \leq 1.$

Axióma 2. $\mathcal{P}(\Omega|B) = \frac{\mathcal{P}(\Omega \cap B)}{\mathcal{P}(B)} = \frac{\mathcal{P}(B)}{\mathcal{P}(B)} = 1.$

Axióma 3. Pre ľubovoľné $A_1, A_2, A_3, \dots \in \mathcal{F}$ pre ktoré platí $A_i \cap A_j = \emptyset, 1 \leq i < j$ je

$$\mathcal{P}\left(\bigcup_{i=1}^{\infty} A_i \Big| B\right) = \frac{\mathcal{P}\left(\bigcup_{i=1}^{\infty} A_i \cap B\right)}{\mathcal{P}(B)} = \frac{\mathcal{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{\mathcal{P}(B)} =$$

$$= \frac{1}{\mathcal{P}(B)} \sum_{i=1}^{\infty} \mathcal{P}(A_i \cap B) = \sum_{i=1}^{\infty} \mathcal{P}(A_i|B).$$

Príklad 2.3 (Kozubík 2014)

Podľa úmrtnostných tabuliek mužov SR za rok 2011 a sa zo 100 000 mužov dožije veku 30 rokov 97 977 a veku 65 rokov sa dožije 73 595. Aká je pravdepodobnosť, že muž, ktorý sa dožil veku 30 rokov sa dožije aj veku 65 rokov.

Označme A_{30} a A_{65} náhodné udalosti, že sa muž dožije veku 30 a 65 rokov. Platí $A_{30} \cap A_{65} = A_{65}$. Teraz už môžeme podľa vzťahu (1) vypočítať pp.

$$\mathcal{P}(A_{65}|A_{30}) = \frac{\mathcal{P}(A_{30} \cap A_{65})}{\mathcal{P}(A_{30})} = \frac{\mathcal{P}(A_{65})}{\mathcal{P}(A_{30})} = \frac{0.73595}{0.97977} \approx 0.75.$$

Tvrdenie 2

Nech $(\Omega, \mathcal{F}, \mathcal{P})$ je pravdepodobnostný priestor, A a B náhodné udalosti s kladnými pravdepodobnosťami.

a) Platí $\mathcal{P}(A \cap B) = \mathcal{P}(A|B) \cdot \mathcal{P}(B)$.

b) Platí $\mathcal{P}(A|\Omega) = \mathcal{P}(A)$.

c) Nech $A_i \in \mathcal{F}$ pre $i = 1, 2, \ldots, n$ sú po dvoch disjunktné náhodné udalosti. Potom platí
$$\mathcal{P}\left(\bigcup_{i=1}^{n} A_i|B\right) = \sum_{i=1}^{n} \mathcal{P}(A_i|B).$$

d) Nech $A_i \in \mathcal{F}$ pre $i = 1, 2, \ldots, n$ sú náhodné udalosti také, že $\mathcal{P}\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$. Potom platí
$$\mathcal{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathcal{P}(A_1) \cdot \mathcal{P}(A_2|A_1) \cdots \mathcal{P}\left(A_n|\bigcap_{i=1}^{n-1} A_i\right).$$

Príklad 2.4

Zo šiestich batérií sú dve vybité. Aká je pp., že pri náhodnom výbere dvoch batérií vyberieme vybité: žiadne, jedno alebo dve.

Označme náhodné udalosti $A =$„prvá vybraná batéria nie je vybitá" a $B =$„druhá vybraná batéria nie je vybitá". Potom pp., že

- „ani jedna batéria nie je vybitá" je

$$\mathcal{P}(A \cap B) = \mathcal{P}(A) \cdot \mathcal{P}(B|A) = \frac{4}{6} \cdot \frac{3}{5} = \frac{2}{5},$$

- „jedna batéria je vybitá" je

$$\mathcal{P}((A \cap B^c) \cup (A^c \cap B)) = \frac{4 \cdot 2}{6 \cdot 5} + \frac{2 \cdot 4}{6 \cdot 5} = \frac{8}{15},$$

- „obe batérie sú vybitá" je

$$\mathcal{P}(A^c \cap B^c) = \frac{2 \cdot 1}{6 \cdot 5} = \frac{1}{15}.$$

Nech $(\Omega, \mathcal{F}, \mathcal{P})$ je pravdepodobnostný priestor, A a B náhodné udalosti. Hovoríme, že náhodné udalosti sú nezávislé, ak platí

$$\mathcal{P}(A \cap B) = \mathcal{P}(A) \cdot \mathcal{P}(B). \tag{2}$$

Nech sú $A_1, A_2, \ldots, A_n$ náhodné udalosti. Hovoríme, že tieto udalosti sú nezávislé, ak platí pre každú najmenej dvojprvkovú podmnožinu $\{i_1, i_2, \ldots, i_k\}$ množiny $\{1, 2, \ldots, n\}$

$$\mathcal{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathcal{P}(A_{i_1}) \cdot \mathcal{P}(A_{i_2}) \cdots \mathcal{P}(A_{i_k}). \tag{3}$$

Poznámka: Definícia nezávislosti (3) vyžaduje aby mimo iné platilo $\mathcal{P}(A_1 \cap \cdots \cap A_n) = \mathcal{P}(A_1) \cdots \mathcal{P}(A_n)$. Na druhej strane samotná platnosť tohoto vzťahu ešte nezaručuje nazávislosť náhodných udalostí $A_1, \ldots, A_n$ pre $n > 2$.

Príklad 2.5

Spolužiaci Jano a Fero nie sú najlepší počtári. Pravdepodobnosť, že daný problém vyrieši správne Jano je $\frac{1}{8}$ a v pripade Fera je len $\frac{1}{12}$. Pretože obaja počítajú zvlášť, potom ak počítaju nesprávne získajú rovnaký nesprávny výsledok s pp. $\frac{1}{1000}$. Aká je pp., že ak získajú rovnaký výsledok, je výsledok správny?

Označme náhodné udalosti:

- $C =$„obaja dostali správny výsledok", $\mathcal{P}(C) = \frac{1}{8} \cdot \frac{1}{12}$,
- $N =$„obaja dostali rovnaký nesprávny výsledok", $\mathcal{P}(N) = \frac{7}{8} \cdot \frac{11}{12} \cdot \frac{1}{1000}$,
- $R =$„obaja dostali rovnaký výsledok", $\mathcal{P}(R) = \mathcal{P}(C) + \mathcal{P}(N)$.

Jano s Ferom dostanú správny výsledok s pp.

$$\mathcal{P}(C|R) = \frac{\mathcal{P}(C \cap R)}{\mathcal{P}(R)} = \frac{\mathcal{P}(C)}{\mathcal{P}(R)} = \frac{1000}{1000 + 7 \cdot 11} \approx 0.93.$$

Príklad 2.6 (O roztržitom profesorovi)

Roztržitý profesor zabúda v obchodoch dáždnik s pp. $\frac{1}{4}$. Vyšiel z domu s dáždnikom, navštívil tri obchody a cestou domov zistil, že dáždnik už nemá. Aká je pp., že dáždnik zabudol v i-tom obchode $(i = 1, 2, 3)$?

Označme udalosť A_i =„profesor zabudne dáždnik v i-tom obchode". Potom $A = A_1 \cup A_2 \cup A_3$=„profesor zabudne dáždnik v niektorom z troch obchodov". Postupne dostaneme

$$\mathcal{P}(A_1) = \frac{1}{4},\ \mathcal{P}(A_2) = \frac{3}{4} \cdot \frac{1}{4},\ \mathcal{P}(A_3) = \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4}.$$

Ale udalosti sú disjunktné, $\mathcal{P}(A) = \mathcal{P}(A_1) + \mathcal{P}(A_2) + \mathcal{P}(A_3) = \frac{37}{64}$. Pretože $\mathcal{P}(A_i|A) = \frac{\mathcal{P}(A_i \cap A)}{\mathcal{P}(A)} = \frac{\mathcal{P}(A_i)}{\mathcal{P}(A)}$, je

$$\mathcal{P}(A_1|A) = \frac{16}{37},\ \mathcal{P}(A_2|A) = \frac{12}{37},\ \mathcal{P}(A_3|A) = \frac{9}{37}.$$

A tak si profesor asi zabudol dáždnik už v prvom obchode.

Úplný systém udalostí

Nech $(\Omega, \mathcal{F}, \mathcal{P})$ je pravdepodobnostný priestor a $H_1, H_2, \ldots, H_n$ také náhodné udalosti, že

- sú po dvoch disjunktné t.j. $H_i \cap H_j = \emptyset, 1 \leq i < j \leq n$,
- $\bigcup\limits_{i=1}^{n} H_i = \Omega$.

Potom $H_1, H_2, \ldots, H_n$ nazývame úplný systém udalostí.

Poznámka:
Definíciu úplného systému udalostí možno prirodzeným spôsobom rozšíriť na spočitateľne veľa udalostí.

Tvrdenie 3 (Veta o úplnej pravdepodobnosti)

Nech $(\Omega, \mathcal{F}, \mathcal{P})$ je pravdepodobnostný priestor a $H_1, H_2, \ldots, H_n$ tvoria taký úplný systém udalostí, že $\mathcal{P}(H_i) > 0$. Nech je $A \in \mathcal{F}$, potom

$$\mathcal{P}(A) = \sum_{i=1}^{n} \mathcal{P}(A|H_i)\mathcal{P}(H_i). \tag{4}$$

Dôkaz. Pretože

$$A = A \cap \Omega = A \cap (H_1 \cup H_2 \cup \cdots \cup H_n) = (A \cap H_1) \cup \cdots \cup (A \cap H_n)$$

a náhodné udalosti $A \cap H_i$, $i = 1, 2, \ldots, n$ sú disjunktné

$$\mathcal{P}(A) = \sum_{i=1}^{n} \mathcal{P}(A \cap H_i) = \sum_{i=1}^{n} \mathcal{P}(A|H_i)\mathcal{P}(H_i).$$

Príklad 2.7

Výrobca monitorov má svoje závody v piatich rôznych krajinách. V prvej až piatej z nich pokrýva 25 %, 15 %, 18 %, 20 % a 22 % svojej celkovej produkcie. Pri tom je známe, že v záručnej dobe dochádza k reklamácii u 4 %, 5 %, 3 %, 6 % a 3 % monitorov vyrobených v prvej až piatej krajine. Určme pravdepodobnosť, že predaný monitor bude počas záručnej doby reklamovaný.

Označme udalosti $H_i =$„monitor bol vyrobený v i-tej krajine“, ktoré tvoria úplný systém udalostí a $A =$„monitor bude reklamovaný“.

Zo zadania máme: $\mathcal{P}(H_1) = 0.25, \mathcal{P}(H_2) = 0.15, \mathcal{P}(H_3) = 0.18,$ $\mathcal{P}(H_4) = 0.2, \mathcal{P}(H_5) = 0.22, \mathcal{P}(A|H_1) = 0.04, \mathcal{P}(A|H_2) = 0.05,$ $\mathcal{P}(A|H_3) = 0.03, \mathcal{P}(A|H_4) = 0.06, \mathcal{P}(A|H_5) = 0.03.$

Po dosadení do vzťahu (4) dostaneme

$$\mathcal{P}(A) = \sum_{i=1}^{5} \mathcal{P}(A|H_i)\mathcal{P}(H_i) = 0.0415.$$

Príklad 2.8 (O výstrednom žalárnikovi)

V žalári je väzeň odsúdený na smrť. Výstredný žalárnik však dá väzňovi šancu. Prinesie mu 12 čiernych 12 bielych guličiek a dve prázdne urny. Povie mu, že zajtra príde kat, náhodne vyberie urnu a z nej jednu guličku. Ak bude gulička biela, dostane väzeň milosť, inak ho čaká smrť. Ako má väzeň rozdeliť guličky do urien, aby maximalizoval svoje oslobodenie.

Označme udalosti: $U_1 =$„v 1. urne je n guličiek“, $U_2 =$„v 2. urne je $24 - n$ guličiek“ a $A =$„vybraná gulička je biela“. Kat vyberie urnu náhodne s $\mathcal{P}(U_1) = \mathcal{P}(U_2) = \frac{1}{2}; \mathcal{P}(A|U_1) = \frac{i}{n}, \mathcal{P}(A|U_2) = \frac{12-i}{24-n}$, kde i je počet bielych guličiek v 1.urne. Zo vzťahu (4) vety o úplnej pp. máme pre $n = 1, \ldots, 12$ a $i = 0, \ldots, 12$

$$p_{n,i} = \mathcal{P}(A) = \mathcal{P}(A|U_1)\mathcal{P}(U_1) + \mathcal{P}(A|U_2)\mathcal{P}(U_2) = \frac{1}{2}\left(\frac{i}{n} + \frac{12 - i}{24 - n}\right).$$

Možno overiť, že $p_{1,1} = 0.739$ je maximálna pp. výberu bielej guličky, čo znamená, že v 1.urne má byť jediná biela gulička a všetky ostatné v 2.urne.

Tvrdenie 4 (Bayesova veta)

Nech $(\Omega, \mathcal{F}, \mathcal{P})$ je pravdepodobnostný priestor a $H_1, H_2, \ldots, H_n$ tvoria taký úplný systém udalostí, že $\mathcal{P}(H_i) > 0$. Nech je $A \in \mathcal{F}$ a $\mathcal{P}(A) > 0$, potom

$$\mathcal{P}(H_k|A) = \frac{\mathcal{P}(A|H_k)\mathcal{P}(H_k)}{\sum_{i=1}^{n} \mathcal{P}(A|H_i)\mathcal{P}(H_i)}. \tag{5}$$

Dôkaz. Podľa definície podmienenej pp. platí

$$\mathcal{P}(H_k|A) = \frac{\mathcal{P}(A \cap H_k)}{\mathcal{P}(A)} \quad (\clubsuit)$$

a vieme, že $\mathcal{P}(A \cap H_k) = \mathcal{P}(A|H_k)\mathcal{P}(H_k)$. Podľa vety o úplnej pp. platí

$$\mathcal{P}(A) = \mathcal{P}(A|H_1)\mathcal{P}(H_1) + \cdots + \mathcal{P}(A|H_n)\mathcal{P}(H_n),$$

čo po dosadení do $(\clubsuit)$ dáva tvrdenie (5).

Vzorec (5) je mimoriadne dôležitý. Často sa stáva, že poznáme tzv. apriórnu pravdepodobnosť $\mathcal{P}(H_1), \ldots, \mathcal{P}(H_n)$. Môžu to byť pp. nejakých príčin alebo chorôb. Ďalej sú zo skúseností známe podmienené pp. nejakej udalosti A za podmienky, že nastala udalosť H_i tj. sú známe pp. $\mathcal{P}(A|H_1), \ldots, \mathcal{P}(A|H_n)$. Udalosťou tu môžu byť komplex príznakov (vysoká teplota, kašel, krv v moči, atď). Vzorec (5) umožňuje vypočítať z týchto údajov aposteriornú pravdepodobnosť $\mathcal{P}(H_k|A)$ pre každé k. Takto sa zistí, aká je pri danom komplexe príznakov pp. jednotlivých chorôb.

Príklad 2.9 (Test na rakovinu)

Predpokladajme, že určitý test na rakovinu má spoľahlivosť 0.95 v tomto zmysle: Ak nemá skúmaná osoba rakovinu, bude test negatívny tiež s pp. 0.95. Je známe, že v sledovanej populácii má rakovinu 0.5 % ľudí. Aká je pp., že náhodne vybraná osoba má rakovinu, ak dal test pozitívny výsledok?

Označme udalosť H_1 =„náhodne vybraná osoba má rakovinu“ a $H_2 = H_1^c$. Podľa zadania je $\mathcal{P}(H_1) = 0.005$ a tak $\mathcal{P}(H_2) = 0.995$. Nech „+“ znamená pozitívny a „−“ negatívny výsledok testu. Podľa našich predpokladov platí

$$\mathcal{P}(+|H_1) = 0.95,\ \mathcal{P}(-|H_1) = 0.05,$$

$$\mathcal{P}(+|H_2) = 0.05,\ \mathcal{P}(-|H_2) = 0.95.$$

Väčšina ľudí bez váhania odpovie, že taká osoba má rakovinu s pp. 0.95 – však taká je spoľahlivosť testu. Ale nie je to tak!

Platí

$$\mathcal{P}(H_1|+) = \frac{\mathcal{P}(+|H_1)\mathcal{P}(H_1)}{\mathcal{P}(+|H_1)\mathcal{P}(H_1) + \mathcal{P}(+|H_2)\mathcal{P}(H_2)} = 0.087.$$

$$\mathcal{P}(H_2|-) = \frac{\mathcal{P}(-|H_2)\mathcal{P}(H_2)}{\mathcal{P}(-|H_2)\mathcal{P}(H_2) + \mathcal{P}(-|H_1)\mathcal{P}(H_1)} \approx 0.9997.$$

Pravdepodobnosť 0.9997, že náhodne vybraná osoba nemá rakovinu, nech je nám útechou.

Cvičenie

2.1 Trikrát hodíme kockou. Aká je pp. že súčet bodov je väčší ako 10 ak prvá a posledná padla 5?

2.2 V prvej zásuvke sú dve zlaté mince, v druhej zlatá a strieborná a v tretej dve strieborné. Náhodne vyberieme jednu zásuvku a vyberieme mincu. Aká je pp. že v zásuvke zostanú zlaté mince ak sme vybrali striebornú mincu?

2.3 K výstupnej kontrole prichádzajú výrobky z ktorých 13% je nepodarkov. Výstupná kontrola pustí 1% nepodarkov a vyradí 2% dobrých výrobkov. Koľko percent výrobkov kontrola vyradí?

*2.4 Neprehľadné vrecko obsahuje x čiernych a y bielych guličiek. Náhodne vyberieme jednu guličku a nevrátime ju do vrecka. Aká je pravdepodobnosť, že druhá vybraná guličky bude čierna ak prvá bola biela?

*2.5 Traja strelci s pp. zásahu $0.6, 0.4, 0.5$ vystrelili 1x na cieľ. Zistilo sa , že dvaja z nich zasiahli. Aká je pp., že to boli dvojice 1,2; 2,3; 1,3?