Základy pravděpodobnosti

David Hampel

12235@mail.muni.cz

Přednáška Statistika 1 (BKMSTAI)

23. říjen 2011, Brno

Motivace

- Teorie pravděpodobnosti se snaží matematicky popsat činnosti („pokusy“), jejichž výsledek není předem jistý.
- Matematická statistika (odhady parametrů a testování hypotéz o nich) je založena na výsledcích teorie pravděpodobnosti.
- Ačkoliv není teorie pravděpodobnosti mnohdy přímo aplikovatelná, pro celkové pochopení statistiky je její znalost nutná.

Pokus

- *Pokusem* rozumíme jednorázové uskutečnění konstantně vymezeného souboru definičních podmínek. Předpokládáme, že pokus můžeme mnohonásobně nezávisle opakovat za dodržení definičních podmínek (ostatní podmínky se mohou měnit, proto různá opakování pokusu mohou vést k různým výsledkům).
- *Deterministickým pokusem* nazýváme takový pokus, jehož každé opakování vede k jedinému možnému výsledku. (Např. zahřívání vody na $100\,^{\circ}\mathrm{C}$ při atmosférickém tlaku $1015\,\mathrm{hPa}$ vede k varu vody.)
- *Náhodným pokusem* nazýváme takový pokus, jehož každé opakování vede k právě jednomu z více *možných výsledků*, které jsou vzájemně neslučitelné.

Náhodný jev a jeho pravděpodobnost

- Neprázdnou množinu možných výsledků náhodného pokusu značíme Ω a nazýváme ji **základní prostor**. Možné výsledky značíme ω_t, $t \in T$ kde T je indexová množina.
    - Příklad: házeme jedenkrát jednou pravidelnou šestistěnnou kostkou. Základním prostorem je množina $\{1, 2, 3, 4, 5, 6\}$.
- **Jevem** nazveme kombinace prvků základního prostoru, které můžeme přesně popsat.
  Jevem je:
    - Padne pětka.
    - Padne jednička nebo dvojka.
    - Padne liché číslo.
    - Padne číslo vyšší nebo rovné třem.

  Jevem není:
    - Dva.
    - Padne vysoké číslo.

Náhodný jev a jeho pravděpodobnost

- Zvláštním případem je **jev nemožný** – sice jsme přesně popsali možný výsledek pokusu, ale tento výsledek nemůže nastat (není kombinací prvků základního prostoru).
  - Padne číslo devět.
- Označení používaná v souvislosti s jevy

Ω	jev jistý	$\emptyset$	jev nemožný
$\bigcap_{i\in I} A_i$	společné nastoupení jevů A_i, $i \in I$	$\bigcup_{i\in I} A_i$	nastoupení alespoň jednoho z jevů A_i, $i \in I$
$\bar{A}_i$, A_i'	opačný jev k jevu A_i	$A_1 \backslash A_2$	nastoupení jevu A_1 za nenastoupení jevu A_2
$A_1 \subseteq A_2$	jev A_1 má za důsledek jev A_2	$A_1 \cap A_2 = \emptyset$	jevy A_1 a A_2 jsou neslučitelné

Náhodný jev a jeho pravděpodobnost

- **Pravděpodobností** rozumíme funkci, která každému jevu přiřadí reálné číslo tak, aby platily následující podmínky:
  - nezápornost
    $$P(A) \geq 0 \quad \forall A,$$
  - spočetná aditivita $\forall i, j, i \neq j \; P(A_i \cap A_j) = 0 \Rightarrow$
    $$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i),$$
  - normovanost
    $$P(\Omega) = 1.$$

Klasická pravděpodobnost

Označme $m(\Omega)$ počet všech možných výsledků a $m(A)$ počet výsledků příznivých nastoupení jevu A. Pak funkci

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

nazveme **klasickou pravděpodobností**. Předpokladem použití klasické pravděpodobnosti je, aby každý výsledek pokusu nastal se stejnou pravděpodobností.

Příklad: házeme jedenkrát jednou pravidelnou šestistěnnou kostkou. Díky pravidelnosti kostky je padnutí každého z čísel stejně pravděpodobné. Chceme spočítat pravděpodobnost jevu A: padne sudé číslo. Máme $m(\Omega) = 6$, $m(A) = 3$ a $P(A) = \frac{m(A)}{m(\Omega)} = \frac{3}{6} = \frac{1}{2}$.

Klasická pravděpodobnost – příklad

Jaká je pravděpodobnost, že při současném hodu šesti kostkami padne

- na každé kostce jiné číslo

$$P(A) = \frac{6!}{6^6} = \frac{720}{46656} = 0.01543$$

- právě šest šestek

$$P(B) = \frac{1}{6^6} = \frac{1}{46656} = 0.0000021$$

- právě pět šestek

$$P(C) = \frac{6 \cdot 5}{6^6} = \frac{30}{46656} = 0.000643$$

- právě čtyři šestky

$$P(D) = \frac{\binom{6}{2} \cdot 5 \cdot 5}{6^6} = \frac{375}{46656} = 0.008037$$

- samá sudá čísla

$$P(E) = \frac{3^6}{6^6} = \frac{729}{46656} = 0.015625$$

Klasická pravděpodobnost – příklad

Mezi N výrobky je M zmetků. Náhodně bez vracení vybereme n výrobků. Jaká je pravděpodobnost, že vybereme právě k zmetků?

Základní prostor Ω je tvořen všemi neuspořádanými n-ticemi vytvořenými z N prvků. Tedy $m(\Omega) = \binom{N}{n}$. Jev A spočívá v tom, že vybereme právě k zmetků z M zmetků (ty lze vybrat $\binom{M}{k}$ způsoby) a výběr doplníme $n-k$ kvalitními výrobky vybranými z $N-M$ kvalitních výrobků (tento výběr lze provést $\binom{N-M}{n-k}$ způsoby). Podle kombinatorického pravidla součinu dostáváme

$$m(A) = \binom{M}{k}\binom{N-M}{n-k}, \quad \text{tedy} \quad P(A) = \frac{m(A)}{m(\Omega)} = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

Podmíněná pravděpodobnost

Nechť H je jev s nenulovou pravděpodobností. Pak definujeme **podmíněnou pravděpodobnost** vzorcem

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

Interpretace této pravděpodobnosti může být následující: víme, že jev H již nastal, a ptáme se na pravděpodobnost, s jakou za této podmínky nastane ještě jev A.

Důležitou aplikací podmíněné pravděpodobnosti je věta o násobení pravděpodobností:

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1})$$

Podmíněná pravděpodobnost – příklad

Z 5 výrobků, mezi nimiž jsou 3 zmetky, vybíráme bez vracení po jednom výrobku. Označíme A_1: první vybraný výrobek byl kvalitní, A_2: druhý vybraný výrobek byl zmetek, A_3: třetí vybraný výrobek byl zmetek. Hledáme $P(A_1 \cap A_2 \cap A_3)$.

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) = \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} = 0.2$$

Formule úplné pravděpodobnosti

Nechť je dán rozklad $\{H_i, i \in I\}$ základního prostoru na nejvýše spočetně mnoho jevů H_i o nenulových pravděpodobnostech $P(H_i)$. Říkáme, že je dán **úplný systém hypotéz**.

Potom pro libovolný jev A platí **formule úplné pravděpodobnosti**

$$P(A) = \sum_{i \in I} P(H_i) \cdot P(A|H_i).$$

Formule úplné pravděpodobnosti – příklad

V první sadě výrobků je 12 výrobků, z toho 1 zmetek. Ve druhé sadě výrobků je 10 výrobků, z toho 1 zmetek. Náhodně zvolený výrobek jsme přemístili z první sady do druhé. Poté jsme ze druhé sady náhodně vybrali jeden výrobek. Jaká je pravděpodobnost, že to byl zmetek?

Označíme
A: výrobek vybraný ze druhé sady byl zmetek
H_1: výrobek přemístěný z první sady do druhé byl kvalitní
H_2: výrobek přemístěný z první sady do druhé byl zmetek
Máme

$$P(A) = P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) = \frac{11}{12} \cdot \frac{1}{11} + \frac{1}{12} \cdot \frac{2}{11} = \frac{13}{132}.$$

Bayesův vzorec

Nechť je dán úplný systém hypotéz $\{H_i, i \in I\}$. Potom pro jev A s nenulovou pravděpodobností a pro libovolný jev H_k platí tzv.
I. Bayesův vzorec

$$P(H_k|A) = \frac{P(H_k) \cdot P(A|H_k)}{\sum\limits_{i \in I} P(H_i) \cdot P(A|H_i)}.$$

Bayesův vzorec – příklad

U jistého druhu elektrického spotřebiče se s pravděpodobností 0.1 vyskytuje výrobní vada. U spotřebiče s touto výrobní vadou dochází v záruční lhůtě k poruše s pravděpodobností 0.5. Výrobky, které tuto vadu nemají, se v záruční lhůtě porouchají s pravděpodobností 0.01. Jaká je pravděpodobnost, že výrobek, který se v záruční době porouchá, bude mít dotyčnou výrobní vadu?

Označíme
A: výrobek se v záruční době porouchá
H_1: výrobek má výrobní vadu
H_2: výrobek nemá výrobní vadu

$$P(H_1|A) = \frac{P(H_1) \cdot P(A|H_1)}{P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2)} = \frac{0.1 \cdot 0.5}{0.059} = 0.847$$

Stochastická nezávislost

Řekneme, že jevy $A_1, A_2, \ldots, A_n$ jsou **stochasticky nezávislé**, právě když platí vztahy

$$\forall\ i < j \quad P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$$

$$\forall\ i < j < k \quad P(A_i \cap A_j \cap A_k) = P(A_i) \cdot P(A_j) \cdot P(A_k)$$

$$\vdots$$

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \cdots \cdot P(A_n)$$

Stochastické nezávislosti se využívá velmi často. Např. máme-li stanovit pravděpodobnost nastoupení alespoň jednoho z jevů $A_1, A_2, \ldots, A_n$, využijeme de Morganova pravidla:

$$P(\bigcup_{i=1}^{n} A_i) = P(\overline{\bigcap_{i=1}^{n} \bar{A}_i}) = 1 - P(\bigcap_{i=1}^{n} \bar{A}_i) = 1 - \prod_{i=1}^{n} P(\bar{A}_i) = 1 - \prod_{i=1}^{n} (1 - P(A_i))$$

Stochastická nezávislost - příklad

Pravděpodobnost, že semínko slunečnice vyklíčí, je 0.5. Zasejeme-li 7 semínek, jaká je pravděpodobnost, že alespoň jedno vyklíčí?
Označíme A_i, $i = 1, \ldots, 7$: i-té semínko vyklíčí
A: alespoň jedno semínko vyklíčí

$$A = \bigcup_{i=1}^{7} A_i$$

$$P(A) = P(\bigcup_{i=1}^{7} A_i) = \cdots = 1 - \prod_{i=1}^{7}(1 - P(A_i)) = 1 - \left(\frac{1}{2}\right)^7 = \frac{127}{128}$$

Opakované nezávislé pokusy

- Provádíme jeden náhodný pokus. Výsledkem tohoto pokusu může být jen „úspěch“ anebo „neúspěch“. Pravděpodobnost úspěchu bude θ. Pokud označíme úspěch 1 a neúspěch 0, můžeme tuto situaci popsat vztahem tzv. **alternativního rozdělení**

$$P(x) = \theta^x (1-\theta)^{1-x}, \quad x \in \{0, 1\}.$$

- Tento pokus budeme provádět n-krát nezávisle na sobě. Zajímat nás bude počet úspěchů $y = \sum_{i=1}^{n} x_i$, kde x_i je výsledek i-tého alternativního pokusu. Pravděpodobnost nastoupení právě y úspěchů z n pokusů má tzv. **binomické rozdělení** dané vztahem

$$P_n(y) = \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

Opakované nezávislé pokusy – příklad

Pětkrát nezávisle na sobě házeme třemi kostkami. Jaká je pravděpodobnost, že právě ve dvou hodech padnou tři jedničky?

Pravděpodobnost, že v jednom hodu třemi kostkami padnou tři jedničky je $\theta = \frac{1}{6^3} = \frac{1}{216}$. Dle vzorce binomického rozdělení máme

$$P_5(2) = \binom{5}{2} \left(\frac{1}{216}\right)^2 \left(1 - \frac{1}{216}\right)^{5-2} =$$

$$= 10 \cdot 0.00002143 \cdot 0.98617 = 0.0002.$$

Některé důsledky pro výpočet pravděpodobností

Podmíněná pravděpodobnost v různých situacích

Předpokládejme, že $P(H) \neq 0$. Čemu je rovna pravděpodobnost $P(A|H)$, jsou-li jevy A; H

a) stochasticky nezávislé

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

Podmíněná pravděpodobnost v různých situacích

Předpokládejme, že $P(H) \neq 0$. Čemu je rovna pravděpodobnost $P(A|H)$, jsou-li jevy A; H

b) neslučitelné

$$P(A|H) = \frac{P(A \cap H)}{P(H)} = \frac{P(\emptyset)}{P(H)} = 0$$

Podmíněná pravděpodobnost v různých situacích

Předpokládejme, že $P(H) \neq 0$. Čemu je rovna pravděpodobnost $P(A|H)$, jsou-li jevy A; H

c) H má za důsledek A

$$P(A|H) = \frac{P(A \cap H)}{P(H)} = \frac{P(H)}{P(H)} = 1$$

Podmíněná pravděpodobnost v různých situacích

Předpokládejme, že $P(H) \neq 0$. Čemu je rovna pravděpodobnost $P(A|H)$, jsou-li jevy A; H

d) A má za důsledek H

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

Nastoupení alespoň jednoho z jevů

Vyjádřete pravděpodobnost

$$P(\bigcup_{i=1}^{n} A_i)$$

- za předpokladu, že $A_1, \ldots, A_n$ jsou neslučitelné

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

Nastoupení alespoň jednoho z jevů

Vyjádřete pravděpodobnost

$$P(\bigcup_{i=1}^{n} A_i)$$

- za předpokladu, že $A_1, \ldots, A_n$ jsou neslučitelné

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

- za předpokladu, že $A_1, \ldots, A_n$ jsou nezávislé

$$P(\bigcup_{i=1}^{n} A_i) = 1 - \prod_{i=1}^{n} (1 - P(A_i))$$

Společné nastoupení jevů

Vyjádřete pravděpodobnost

$$P(\bigcap_{i=1}^{n} A_i)$$

- za předpokladu, že $A_1, \ldots, A_n$ jsou neslučitelné

$$P(\bigcap_{i=1}^{n} A_i) = 0$$

Společné nastoupení jevů

Vyjádřete pravděpodobnost

$$P(\bigcap_{i=1}^{n} A_i)$$

- za předpokladu, že $A_1, \ldots, A_n$ jsou neslučitelné

$$P(\bigcap_{i=1}^{n} A_i) = 0$$

- za předpokladu, že $A_1, \ldots, A_n$ jsou nezávislé

$$P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$$

Aplikace pravděpodobnosti

Příklad o plachetnicích

Firma se rozhoduje, zda vstoupit na trh s novým typem malé sportovní plachetnice, což by se vyplatilo, kdyby ji aspoň 10 % koupěschopných zákazníků koupilo. Určitému počtu náhodně vylosovaných zákazníků udělá firma fiktivní nabídku, v níž požaduje zatržení jedné ze čtyř možností: určitě bych koupil, zřejmě bych koupil, snad bych koupil, nekoupil bych. Přitom z dřívějších zkušeností je známo, kolik procent zákazníků volících jednotlivé odpovědi skutečně přistoupí ke koupi.

Příklad o plachetnicích

odpověď	% odpovědí	% skutečných kupců
1. určitě bych koupil	12 %	40 %
2. zřejmě bych koupil	23 %	20 %
3. snad bych koupil	17 %	8 %
4. nekoupil bych	48 %	1 %

a) S jakou pravděpodobností koupí náhodně vybraný zákazník loď?

b) S jakou pravděpodobností ti, kteří skutečně koupili, předtím zaškrtli „nekoupil bych“.

Příklad o plachetnicích – řešení

A ... náhodně vybraný zákazník koupí plachetnici

H_i ... náhodně vybraný zákazník patří do i-té skupiny, $i = 1, \ldots, 4$.

$\mathsf{P}(H_1)$=0,12	$\mathsf{P}(A\vert H_1)$=0,4
$\mathsf{P}(H_2)$=0,23	$\mathsf{P}(A\vert H_2)$=0,2
$\mathsf{P}(H_3)$=0,17	$\mathsf{P}(A\vert H_3)$=0,08
$\mathsf{P}(H_4)$=0,48	$\mathsf{P}(A\vert H_4)$=0,01

Příklad o plachetnicích – řešení

a)

$$P(A) = \sum_{i=1}^{4} P(H_i)P(A|H_i) = 0,12 \cdot 0,4 + 0,23 \cdot 0,2 + 0,17 \cdot 0,08 + 0,48 \cdot 0,01 = 0,1124 > 0,10$$

Firmě se s novou plachetnicí vyplatí vstoupit na trh.

b)

$$P(H_4|A) = \frac{P(H_4)P(A|H_4)}{P(A)} = \frac{0,01 \cdot 0,48}{0,1124} = 0,0427$$

Ti zákazníci, kteří skutečně plachetnici koupili, předtím zaškrtli „nekoupil bych“ s pravděpodobností 0,0427.

Zloději ve firmě

Firma pokládá náhodně vybraným zaměstnancům otázku: „Odnesli jste si během minulého roku některý z našich výrobků bez zaplacení?“ Aby je ujistila o anonymitě dotazníku, připojuje tuto instrukci: Hoďte si v soukromí mincí a padne-li líc, odpovězte bez ohledu na skutečnost „ano“, jestliže padne rub odpovězte ve shodě se skutečností „ano“, nebo „ne“.

a) Zvolte matematický model.
b) Vypočtěte pravděpodobnost odpovědi „ano“ a pravděpodobnost vylosování zloděje.
c) Aproximujte pravděpodobnost vylosování zloděje, bylo-li dotázáno n osob, z nichž a osob odpovědělo „ano“.

Zloději ve firmě – řešení

a) Zvolte matematický model.

Náhodný pokus spočívá ve vylosování jedné osoby a v jednom hodu mincí.

L . . . padl líc

K . . . byl vylosován zloděj

A . . . vylosovaný odpověděl „ano“

Zloději ve firmě – řešení

b) Vypočtěte pravděpodobnost odpovědi „ano“ a pravděpodobnost vylosování zloděje.

$$
\begin{aligned}
P(A) &= P(L \cup (L' \cap K)) \\
&= P(L) + P(L') \cdot P(K) \\
&= \frac{1}{2} + \frac{1}{2} P(K)
\end{aligned}
$$

$$P(K) = 2P(A) - 1$$

Zloději ve firmě – řešení

c) Aproximujte pravděpodobnost vylosování zloděje, bylo-li dotázáno n osob, z nichž a osob odpovědělo „ano“.

$$P(A) \approx \frac{a}{n}$$

$$P(K) \approx 2\frac{a}{n} - 1$$

Antidrogový test

Jistá společnost podrobuje uchazeče o zaměstnání antidrogovému testu. Uchazeči pochází z oblasti, kde pouze půl procenta obyvatel užívá drogy. Citlivost užívaného testu je 99 % (tzn., že u 99 % narkomanů test vyjde pozitivně) a dále test je specifický také na 99 % (Tzn., že test vyjde negativně u 99 % „ne-narkomanů“). Zdá se tedy, že test je relativně přesný.

Určete pravděpodobnost, že osoba, jejíž výsledek testu je pozitivní, skutečně užívá drogy.

Antidrogový test – řešení

P(A) . . . u náhodně vybraného uchazeče test vyjde pozitivně

P(H_1) . . . uchazeč bere drogy

P(H_2) . . . uchazeč nebere drogy

$$
\begin{array}{ll}
\mathrm{P}(H_1)=0{,}005 & \mathrm{P}(A|H_1)=0{,}99 \\
\mathrm{P}(H_2)=0{,}995 & \mathrm{P}(A|H_2)=1-\mathrm{P}(A'|H_2)=0{,}01
\end{array}
$$

Antidrogový test – řešení

$$\begin{aligned} P(A) &= P(H_1)P(A|H_1) + P(H_2)P(A|H_2) \\ &= 0,005 \cdot 0,99 + 0,995 \cdot 0,01 \\ &= 0,0149 \end{aligned}$$

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(A)} = \frac{0,005 \cdot 0,99}{0,0149} = 0,3322$$

Pravděpodobnost, že uchazeč, jemuž vyšel test pozitivně, skutečně bere drogy, je pouze 33,22 %. Je tedy více pravděpodobné, že drogy nebere!

Čím menší je pravděpodobnost zkoumaného jevu (H_1), tím větší je pravděpodobnost, že test bude „falešně" pozitivní.

$$P(A \cap H_1) = 0,00495 < P(A \cap H_2) = 0,00995$$