Matematika IV – 7. přednáška
Pravděpodobnost – opakování a zobecnění pojmů

Michal Bulant

Masarykova univerzita
Fakulta informatiky

7. 4. 2008

Obsah přednášky

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.

Motto:

42,35 procenta všech statistik je nesmyslných.

Statistika v širším slova smyslu je jakékoliv zpracování číselných dat o nějakém souboru objektů a jejich více či méně přehledná prezentace.

Podstatou **matematické statistiky** je pro daná data zjišťovat, jaké vlastnosti mají objekty, které jsou daty popisovány. Zpravidla jde o sběr dat o části souboru objektů, jejich následnou analýzu a konečně o vyslovení důsledků pozorování pro celý soubor. Výsledkem práce matematického statistika je sdělení o velkém souboru objektů na základě studia malé (zpravidla náhodně vybrané) části z nich, **společně s kvalitativním odhadem věrohodnosti výsledného sdělení.**
Teorie pravděpodobnosti studuje modely popisující chování abstraktních souborů (pravděpodobnost jevů z jevového pole), statistika studuje skutečné náhodné výběry z nějakého základního souboru a zdůvodňuje výběr teoretického pravděpodobnostního modelu, resp. kvalitativní informace o jeho parametrech.

Příklad

Za soubor objektů vezměme všechny studenty přednášky Matematika III, jako číselný údaj můžeme uvažovat

1. průměrné bodové hodnocení studenta u zkoušky,
2. průměrnou známku u zkoušky z tohoto $(2,92)$ a z jiných pevně vybraných předmětů (`IB000` – $2,95$; `IB102` – $2,89$),
3. nejčastější známku (resp. úspěšnou známku) z tohoto předmětu (`F` – 92 krát, `E` – 91 krát), nejméně častou známku (`B` – 15 krát),
4. průměrný počet bodů dosažených na jednotlivých termínech zkoušky (1. – $16,8$; 2. – $8,9$; 3. – $8,1$; příklad, za nějž bylo uděleno nejvíce (nejméně) procent možných bodů – min. kostra (1B, $82,5\%$), resp. rekurence (2A, $3,6\%$)
5. počet pracovních hodin týdně odpracovaných mimo fakultu,
6. číselná data vypovídající o historii dřívějšího studia

a mnoho dalších údajů.

Zastavme se u prvního údaje. Samotný aritmetický průměr bodů nám mnoho neřekne nejen o kvalitě přednášky a o kvalitě přednášejícího, ale ani o samotném hodnocení. Zajímá nás také hodnota, která bude ,,uprostřed souboru", tj. počet bodů, pro které je stejně studentů pod ní a nad ní.
Obdobně první a poslední čtvrtina, desetina apod. Všem takovým údajům říkáme **statistiky** posuzované veličiny. V uvedených příkladech se jim říká **medián**, **kvartil**, **decil** apod.
Z obecné zkušenosti nebo jako výsledek úvah mimo matematiku víme, že rozumné hodnocení by na mělo mít tzv. **normální rozdělení** (odpovídá tzv. *Gaussově křivce*). Tento pojem patří do teorie pravděpodobnosti a k jeho zavedení potřebujeme poměrně dost matematiky.
Porovnáním výsledku třeba i docela malého náhodného výběru studentů s teoretickým modelem můžeme zjistit odhad parametrů takového rozdělení a činit závěry, zda je hodnocení ,,rozumné".
Zároveň lze popsat věrohodnost našich závěrů.

Daleko zajímavější vývody ovšem můžeme činit, když porovnáním statistik pro různé veličiny budeme moci dovozovat informace o souvislostech. Pokud např. neexistuje žádná doložitelná souvislost mezi historií předchozího studia a výsledky v dané přednášce, je jedním z možných vysvětlení vývod, že je přednáška (nebo její hodnocení) prostě špatná.

Závěr úvodních úvah:

- V matematice pracujeme s abstraktním matematickým popisem pravděpodobnosti.
- Vývody pro konktrétní soubory dat, pro které je zvolený model relevantní dává matematická statistika.
- To, zda je takový popis adekvátní pro konkrétní výběr dat, je také možné podpořit nebo zavrhnout pomocí metod matematické statistiky.

Připomeneme (a trochu zobecníme) pojmy a výsledky z prvního semestru.

Definice (Náhodné jevy)

Budeme pracovat s neprázdnou pevně zvolenou množinou Ω všech možných výsledků, kterou nazýváme **základní prostor**.
Prvky $\omega \in \Omega$ představují jednotlivé **možné výsledky**.
Systém podmnožin $\mathcal{A}$ základního prostoru se nazývá **jevové pole** a jeho prvky se nazývají **jevy**, jestliže

- $\Omega \in \mathcal{A}$, tj. základní prostor, je jevem,
- je-li $A, B \in \mathcal{A}$, pak $A \setminus B \in \mathcal{A}$, tj. pro každé dva jevy je jevem i jejich množinový rozdíl,
- je-li $A_i \in \mathcal{A}$, $i \in I$ nejvýše spočetný systém jevů, pak také jejich sjednocení je jevem, tj. $\cup_{i \in I} A_i \in \mathcal{A}$.

Důsledek

- *Komplement $A^c = \Omega \setminus A$ jevu A je jevem, který nazýváme* opačný jev *k jevu A.*
- *Průnik dvou jevů opět jevem, protože pro každé dvě podmnožiny $A, B \subset \Omega$ platí*

$$A \setminus (\Omega \setminus B) = A \cap B.$$

Takový systém množin $\mathcal{A}$ se pak nazývá σ-algebra [1]
Jevové pole je tedy systém podmnožin základního prostoru uzavřený na konečné průniky, spočetná sjednocení a množinové rozdíly. Jednotlivé množiny $A \in \mathcal{A}$ nazýváme **náhodné jevy** (vzhledem k $\mathcal{A}$).

[1] Srovnej s pojmem množinová algebra v části o Booleovských algebrách, σ zde znamená *spočetnost*.

Terminologie připomíná souvislosti s popisem skutečných jevů a jejich statistickým popisem:

- celý základní prostor Ω se nazývá **jistý jev**, prázdná podmnožina $\emptyset \in \mathcal{A}$ se nazývá **nemožný jev**,
- jednoprvkové podmnožiny $\{\omega\} \in \Omega$ se nazývají **elementární jevy**,
- **společné nastoupení jevů** A_i, $i \in I$, odpovídá jevu $\cap_{i\in I} A_i$, **nastoupení alespoň jednoho z jevů** A_i, $i \in I$, odpovídá jevu $\cup_{i\in I} A_i$,
- $A, B \in \mathcal{A}$ jsou **neslučitelné jevy**, je-li $A \cap B = \emptyset$,
- jev A má za **důsledek** jev B, když $A \subset B$,
- je-li $A \in \mathcal{A}$, pak se jev $B = \Omega \setminus A$ nazývá **opačný jev k jevu** A, píšeme $B = A^c$.

Definice (Kolmogorovova definice pravděpodobnosti)

Pravděpodobnostní prostor je jevové pole $\mathcal{A}$ podmnožin (konečného) základního prostoru Ω, na kterém je definována funkce $P : \mathcal{A} \to \mathbb{R}$ s následujícími vlastnosti:

- je nezáporná, tj. $P(A) \geq 0$ pro všechny jevy A,
- je aditivní, tj. $P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$, pro každý nejvýše spočetný systém po dvou neslučitelných jevů,
- pravděpodobnost jistého jevu je 1.

Funkci P nazýváme **pravděpodobností** na jevovém poli $(\Omega, \mathcal{A})$.

Důsledek

Pro všechny jevy $A, B \in \mathcal{A}$ platí

- $P(\emptyset) = 0, \ 0 \leq P(A) \leq 1$,
- $P(A^c) = 1 - P(A)$,
- $A \subseteq B \implies P(A) \leq P(B), \ P(B \setminus A) = P(B) - P(A)$,
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Podobná tvrzení platí i pro nekonečné posloupnosti jevů:

Tvrzení

Pro libovolnou nejvýše spočetnou množinu jevů $(A_i)_{i=1}^{\infty}$ *platí:*

- *Je-li* $A_1 \subseteq A_2 \subseteq \cdots$, *pak*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} P(A_i),$$

- *Je-li* $A_1 \supseteq A_2 \supseteq \cdots$, *pak*

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} P(A_i),$$

- $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$,
- $P(\bigcap_{i=1}^{\infty} A_i) \geq 1 - \sum_{i=1}^{\infty}(1 - P(A_i))$.

Klasická pravděpodobnost

Připomeňme si klasickou konečnou pravděpodobnost.

Definice

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.

Že s klasickou pravděpodobností nevystačíme, ukazují následující příklady:

Příklad

- Cestou z Kotlářské na Botanickou jsem ztratil zadání písemky. Určete pravděpodobnost jevu ω_X slovně vyjádřeného: *ztracená písemka se nachází nejblíže k zastávce trolejbusu* X.
- Určete pravděpodobnost, jevu ω_k: *při opakovaném hodu mincí padne hlava poprvé při k-tém pokusu*.

V prvním případě je třeba pracovat s nekonečně mnoha stejně pravděpodobnými elementárními jevy: *písemku jsem ztratil v bodě* (x, y) , ve druhém pak musíme připustit teoretickou možnost, že hlava nepadne nikdy, a prostorem jevů tedy bude $\mathbb{N} \cup \{\infty\}$.

Petersburgský „paradox" (Bernoulli, 1738)

Typický příklad klasické pravděpodobnosti jsou jevy související s házením mincí. Představme si následující pravidla kasina:

Návštěvník zaplatí vklad C a poté hází mincí. V banku je na začátku dolar a při každém hodu se bank zdvojnásobí. Padne-li hlava, hráč získá obsah banku. Je-li tedy T počet hodů potřebných k první hlavě, hráč obdrží výhru 2^T. Jaká je „fér hodnota" pro vklad C?

A co vy? Zaplatili byste za možnost zahrát si tuto hru třeba 20$?

Pravděpodobnost, že padne hlava je u férové mince $1/2$, je proto $P(T = k) = 2^{-k}$. Sečteme-li všechny pravděpodobnosti výsledků vynásobených výhrami 2^k, dostaneme očekávanou výhru

$$\frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 2^2 + \cdots = \sum_{1}^{\infty} 1 = \infty.$$

Zdá se proto, že se vyplatí vložit i velký vklad, protože libovolný vklad C se nám ,,časem" vrátí.
Ve skutečnosti simulací hry zjistíme, že nezávisle na počtu pokusů se prakticky všechny výhry budou pohybovat v rozmezí malých hodnot. Důvodem je, že vysoké výhry jsou velice nepravděpodobné a proto je při reálných úvahách nelze brát vážně.
Tento paradox je vysvětlován nelinearitou funkce *užitečnosti peněz* (utility function), případně nezbytností diskontování jejich hodnoty.

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.

Definice

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:

Definice

Říkáme, že jevy A a B jsou nezávislé, jestliže

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

Definice

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

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

Příklad

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

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

Věta (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*).
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.

Poznámka

Sami si můžete podobný výpočet udělat pro tzv. triple test na Downův syndrom, prováděný ve 2. trimestru těhotenství s 90% citlivostí a 5% ,,false-positive rate" či pro statistiky svého oblíbeného spamfilteru (např. SpamAssassin s někde udávanou citlivostí 99,64% a specifičností 98.23%).

Evidentně prostý výběr náhodné osoby a použití jediného testu, byť velmi citlivého a specifického, nejsou vhodné ani na otestování skutečného stavu populace, ani na preventivní vyšetření jednotlivců, pokud nemáme další podpůrné informace a lepší nástroje.
Právě matematická statistika dává nástroje na kvalifikovanější postupy v medicínské i průmyslové diagnostice, ekonomických modelech, vyhodnocování experimentálních dat atd.

Vraťme se k jednoduchému a názornému příkladu statistik kolem výsledků studentů v daném předmětu, který je a není podobný klasické pravděpodobnosti a s ní související statistice při házení kostkou.
Na jedné straně jsme připustili pouze konečný počet možných bodových hodnocení (celá čísla od 0 do 30), zároveň ale není patrně vhodné představovat si výsledky jednotlivých studentů jako analogii nezávislého házení kostkou (to by byla skutečně divně vedená přednáška).
Místo toho máme na základním prostoru Ω všech studentů definovánu funkci bodového ohodnocení $X : \Omega \to \mathbb{R}$. Je to typický příklad **náhodné veličiny**.
U každé náhodné veličiny potřebujeme umět pracovat s vhodnou množinou jevů. Zpravidla požadujeme, abychom mohli pracovat s pravděpodobnostmi příslušnosti hodnoty X do předem zadaného intervalu.

Přirozenější interpretací výsledku pokusu je totiž často spíše než zjištění, zda náhodný jev *nastal či nenastal*, nějaká hodnota:

- součet bodů na dvou kostkách,
- počet bakterií v daném množství roztoku nebo
- počet studentů, kteří uspěli u zkoušky.

Od pravděpodobnostního prostoru $(\Omega, \mathcal{A}, P)$ tedy potřebujeme přejít k obdobné dvojici $(\mathbb{R}, \mathcal{B})$ tak, abychom podmnožinám $\mathbb{R}$, ležícím v σ-algebře $\mathcal{B}$ byli schopni přiřadit pravděpodobnost odvozenou z $(\Omega, \mathcal{A}, P)$.

Na prostoru $\mathbb{R}^k$ uvažujme nejmenší jevové pole $\mathcal{B}$ obsahující všechny k–rozměrné intervaly. Množinám v $\mathcal{B}$ říkáme **borelovské množiny** (nebo také měřitelné množiny) na $\mathbb{R}^k$.
Speciálně pro $k = 1$ jde o množiny, které obdržíme **z intervalů konečnými průniky a nejvýše spočetnými sjednoceními**.

Definice

Náhodná veličina X na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$ je taková funkce $X : \Omega \to \mathbb{R}$, že vzor $X^{-1}(B)$ patří do $\mathcal{A}$ pro každou Borelovskou množinu $B \in \mathcal{B}$ na $\mathbb{R}$ (tj. $X : \Omega \to \mathbb{R}$ je tzv. borelovsky měřitelná).
Množinová funkce

$$P_X(B) = P(X^{-1}(B))$$

se nazývá **rozdělení pravděpodobnosti** náhodné veličiny X.
Náhodný vektor $(X_1, \ldots, X_k)$ na $(\Omega, \mathcal{A}, P)$ je k–tice náhodných veličin.

Definice náhodné veličiny zajišťuje, že pro všechny $-\infty \leq a \leq b \leq \infty$ existuje pravděpodobnost $P(a < X \leq b)$, kde používáme stručné značení pro jev $A = (\omega \in \Omega;\ a < X(\omega) \leq b))$.

Definice

Distribuční funkcí (*distribution, cumulative density function*) náhodné veličiny X je funkce $F : \mathbb{R} \to \mathbb{R}$ definovaná pro všechny $x \in \mathbb{R}$ vztahem

$$F(x) = P(X \leq x).$$

Distribuční funkcí náhodného vektoru $(X_1, \dots, X_k)$ je funkce $F : \mathbb{R}^k \to \mathbb{R}$ definovaná pro všechny $(x_1, \dots, x_k) \in \mathbb{R}^k$ vztahem

$$F(x) = P(X_1 \leq x_1 \wedge \cdots \wedge X_k \leq x_k).$$

Diskrétní náhodné veličiny

Předpokládejme, že náhodná veličina X na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$ nabývá jen konečně mnoha hodnot $x_1, x_2, \dots, x_n \in \mathbb{R}$. Pak existuje tzv. **pravděpodobnostní funkce** $f(x)$ taková, že

$$f(x) = \begin{cases} P(X = x_i) & \text{pro } x = x_i \\ 0 & \text{jinak.} \end{cases}$$

Evidentně $\sum_1^n f(x_i) = 1$.
Takové náhodné veličině se říká **diskrétní**.
Každá náhodná veličina definovaná pro klasickou pravděpodobnost je diskrétní. Obdobně lze definici pravděpodobnostní funkce rozšířit na veličiny se spočetně mnoha hodnotami (pracujeme pak s nekonečnými řadami)

Spojité náhodné veličiny

I když hodnoty náhodné veličiny X nejsou diskrétní, můžeme postupovat podobně s užitím ideí diferenciálního a integrálního počtu. Intuitivně lze uvažovat takto: **hustotu** $f(x)$ **pravděpodobnosti** pro X si představíme jako

$$P(x < X \leq x + dx) = f(x)dx.$$

To znamená, že chceme pro $-\infty \leq a \leq b \leq \infty$

$$P(a < X \leq b) = \int_a^b f(x)dx. \qquad (*)$$

Definice

Náhodná veličina X, pro kterou existuje její **hustota pravděpodobnosti** splňující $(*)$, se nazývá **spojitá**.

Vlastnosti distribuční funkce

Věta

Nechť X *je náhodná veličina,* $F(x)$ *je její distribuční funkce.*

1. F *je neklesající.*
2. F *je zprava spojitá,* $\lim_{x\to-\infty} = 0$ *a* $\lim_{x\to\infty} = 1$.
3. *Je-li* X *diskrétní s hodnotami* $x_1, \ldots, x_n$*, pak je* $F(x)$ *po částech konstantní,* $F(x) = \sum_{x_i \leq x} P(X = x_i)$ *a* $F(x) = 1$ *kdykoliv* $x \geq x_n$.
4. *Je-li* X *spojitá, pak je* $F(x)$ *diferencovatelná a její derivace se rovná hustotě* X*, tj. platí* $F'(x) = f(x)$.

Distribuční funkce