PRAVDĚPODOBNOST A STATISTIKA OPAKOVÁNÍ, pro rozpoznávání

Václav Hlaváč

Fakulta elektrotechnická ČVUT v Praze
katedra kybernetiky, **Centrum strojového vnímání**
hlavac@fel.cvut.cz, http://cmp.felk.cvut.cz/~hlavac

OBSAH PŘEDNÁŠKY

- Pravěpodobnost (definice, náhodná veličina)
- Statistika (náhodný výběr, odhad parametrů)

Poděkování Ing. Martinovi Urbanovi, PhD. za první verzi z podzimu 2005; prof. Ing. Mirko Navarovi, DrSc. za několik průsvitek.

DOPORUČENÉ ČTENÍ

- M. Navara: Pravděpodobnost a matematická statistika, skriptum FEL ČVUT, Vydavatelství ČVUT, Praha 2007.
- J. Novovičová: Pravděpodobnost a matematická statistika. skriptum Fakulty dopravní ČVUT, Vydavatelství ČVUT, Praha 2002.
- A. Papoulis: Probability, Random Variables and Stochastic Processes, McGraw Hill, Edition 4, 2002.
- http://mathworld.wolfram.com/
- http://www.statsoft.com/textbook/stathome.html

PRAVDĚPODOBNOST, STATISTIKA

- Pravděpodobnost
  - Abstraktní matematický model neurčitosti.
  - Modeluje děje, v nichž hraje roli náhodnost.
- Statistika
  - Popisná nebo inferenční.
  - Sběr, organizace a analýza dat.
  - Zobecňuje z omezených / konečných vzorků.
  - Odhad parametrů, testování hypotéz, atd.

PRAVDĚPODOBNOST
Navarův motivační příklad

- Los loterie se prodává za **cenu** 2 EUR.
- 1 los z 1000 vyhrává 1000 EUR, ostatní nic. (Tím je dána **hodnota** losu po losování.)
- Za kolik máme prodat los *před* losováním?

- Za 2 EUR by ho koupil jen hlupák. (Nebo ne?)
- Hodnota losu před losováním je $\frac{1}{1000}1000 = 1$ EUR $=$ průměrná hodnota po losování.

Na to je teorie pravděpodobnosti.

Otázka "Loterie": Proč se přesto kupují losy a loterie fungují?

STATISTIKA, Navarův motivační příklad

Dosud jsme předpokládali, že parametry pravděpodobnostního modelu jsou známy. To je málokdy splněno.

Příklad (Sportka): Na Sportce se normálně prodělává; jelikož jsou výhry stanoveny podle počtu výherců, je výhodnější sázet jinak než ostatní. K tomu potřebujeme vědět, podle jakého modelu sázejí ostatní.

Příklad (ruleta): U rulety se obě strany zajímají, zda padají všechna čísla se stejnou pravděpodobností, přesněji, jak velké jsou odchylky od rovnoměrného rozdělení. Ale jak to zjistit a jaké je riziko chybných závěrů?

Na to je statistika.

Statistika poskytuje daleko víc: nástroj pro zkoumání světa, pro hledání a ověřování závislostí, které nejsou zjevné.

NÁHODNÉ JEVY, POJMY

Náhodný pokus

Prostor elementárních jevů je neprázdná množina Ω všech možných výsledků daného pokusu.

Elementární jevy $\omega \in \Omega$ jsou prvky prostoru elementárních jevů (výsledky pokusu).

Jevové pole $\mathcal{A}$ je tvořeno systémem všech podmnožin prostoru elementárních jevů Ω.

Náhodný jev $A \in \mathcal{A}$ je prvkem jevového pole.

Poznámka: pojem náhodného jevu byl zaveden proto, aby bylo možné definovat pravděpodobnost, rozdělení pravděpodobnosti, atd.

PRAVDĚPODOBNOST, ZAVEDENÍ

- Klasická (dnes se již nepovažuje za definici pravděpodobnosti, ale metodu odhadu pravděpodobnosti)

$$P(A) \approx \frac{N_A}{N}$$

- Limitní (četnostní)

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

- Axiomatická (Andreje Kolmogorova 1930)

AXIOMATICKÁ DEFINICE PRAVDĚPODOBNOSTI

Ω - prostor elementárních jevů

A - jevové pole

1. $P(A) \geq 0, \quad A \in \mathcal{A}$.
2. $P(\Omega) = 1$.
3. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$, $A \in \mathcal{A}$, $B \in \mathcal{B}$.

PRAVDĚPODOBNOST

je funkce P, která jevům přiřazuje čísla z $\langle 0, 1 \rangle$ a splňuje podmínky

(P1) $P[true] = 1$,

(P2) $P\left[\bigvee_{n \in \mathbb{N}} A_n\right] = \sum_{n \in \mathbb{N}} P[A_n]$, pokud se jevy A_n, $n \in \mathbb{N}$, navzájem vylučují.

Z těchto podmínek vyplývá:

$P[false] = 0, \qquad P[\neg A] = 1 - P[A]$,

jestliže $A \Rightarrow B$, pak $P[A] \leq P[B]$.

Pro korektnost je potřeba, aby systém jevů splňoval určité další podmínky.

ODVOZENÉ VZTAHY

- If $A \subset B$ then $P(B \setminus A) = P(B) - P(A)$.
  Symbol $\setminus$ označuje množinový rozdíl.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

PODMÍNĚNÁ PRAVDĚPODOBNOST

Máme pravděpodobnostní popis systému. Dostaneme-li dodatečnou informaci, že nastal jev B, změní se naše znalost o pravděpodobnosti jevu A na

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

což je *podmíněná pravděpodobnost* jevu A za podmínky B. Je definována pouze pro $P(B) \neq 0$.

VLASTNOSTI PODMÍNĚNÉ PRAVDĚPODOBNOSTI

- $P(true|B) = 1$, $P(false|B) = 0$.
- Pokud $A = \bigvee_{n \in \mathbb{N}} A_n$ a jevy $A_1, A_2, \ldots$ se navzájem vylučují, pak $P(A|B) = \sum_{n \in \mathbb{N}} P(A_n|B)$.
- Jevy A, B jsou *nezávislé*, právě když $P(A|B) = P(A)$.
- Pokud $B \Rightarrow A$, pak $P(A|B) = 1$. Pokud $B \Rightarrow \neg A$, pak $P(A|B) = 0$.

Jevy B_i, $i \in I$, tvoří *úplný systém jevů*, jestliže se navzájem vylučují a $\bigvee_{i \in I} B_i = true$.

PŘÍKLAD: podmíněná pravděpodobnost

Jeden hod kostkou. Jaká je pravděpodobnost, že padne číslo větší než 3 (jev A) za podmínky, že padlo liché číslo (jev B).

$$\Omega = \{1, 2, 3, 4, 5, 6\}\,, \quad A = \{4, 5, 6\}\,, \quad B = \{1, 3, 5\}$$

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

$$P(A \cap B) = P(\{5\}) = \frac{1}{6}$$

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

VĚTA O ÚPLNÉ PRAVDĚPODOBNOSTI

Nechť B_i, $i \in I$, je úplný systém jevů a $\forall i \in I\colon P(B_i) \neq 0$.

Pak pro každý jev A platí

$$P(A) = \sum_{i \in I} P(B_i)\, P(A|B_i)\,.$$

Důkaz:

$$P(A) = P\left(\left(\bigvee_{i \in I} B_i\right) \wedge A\right) = P\left(\bigvee_{i \in I} (B_i \wedge A)\right)$$

$$= \sum_{i \in I} P(B_i \wedge A) = \sum_{i \in I} P(B_i)\, P(A|B_i)\,.$$

BAYESOVA VĚTA

Nechť B_i, $i \in I$, je úplný systém jevů a $\forall i \in I\colon P(B_i) \neq 0$.

Pak pro každý jev A splňující $P(A) \neq 0$ platí

$$P(B_i|A) = \frac{P(B_i)\, P(A|B_i)}{\sum_{i\in I} P(B_i)\, P(A|B_i)}\,.$$

Důkaz (s využitím věty o úplné pravděpodobnosti):

$$P(B_i|A) = \frac{P(B_i \wedge A)}{P(A)} = \frac{P(B_i)\, P(A|B_i)}{\sum_{i\in I} P(B_i)\, P(A|B_i)}\,.$$

VÝZNAM BAYESOVY VĚTY

- Pravděpodobnosti $P(A|B_i)$ odhadneme z pokusů nebo z modelu, pomocí nich určíme pravděpodobnosti $P(B_i|A)$, které slouží k "optimálnímu" odhadu, který z jevů B_i nastal.
- **Problém:** Ke stanovení *aposteriorní pravděpodobnosti* $P(B_i|A)$ potřebujeme znát i *apriorní pravděpodobnost* $P(B_i)$.
- Podobně definujeme podmíněné rozdělení náhodné veličiny, podmíněnou hustotu spojité náhodné veličiny apod.

PODMÍNĚNÁ NEZÁVISLOST

Náhodné jevy A, B jsou podmíněně nezávislé za podmínky C, jestliže

$$P(A \wedge B|C) = P(A|C)\,P(B|C)\,.$$

Podobně definujeme podmíněnou nezávislost více jevů, náhodných veličin apod.

SDRUŽENÁ PRAVDĚPODOBNOST

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

NEZÁVISLÉ JEVY

Jevy A, B jsou nezávislé $\Leftrightarrow P(A \cap B) = P(A)\, P(B)$.

Příklad

Jeden hod kostkou. Jevy $A > 3$, jev B liché číslo. Jsou jevy nezávislé?

$$\Omega = \{1,2,3,4,5,6\}\,, \quad A = \{4,5,6\}\,, \quad B = \{1,3,5\}$$

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

$$P(A \cap B) = P\left(\{5\}\right) = \frac{1}{6}$$

$$P(A)\, P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$P(A \cap B) \neq P(A)\, P(B) \Leftrightarrow$ jevy jsou závislé.

DISTRIBUČNÍ FUNKCE NÁHODNÉ VELIČINY

Distribuční funkce náhodné veličiny X je funkce $F\colon X \to \langle 0, 1\rangle$ je definovaná pomocí $F(x) = P(X \leq x)$, kde P je pravděpodobnost.

Vlastnosti:

1. $F(x)$ je neklesající funkce, tj. pro $\forall$ dvojici $x_1 < x_2$ platí $F(x_1) \leq F(x_2)$.
2. $F(X)$ je zprava spojitá, tj. platí $\lim\limits_{h \to 0^+} F(x + h) = F(x)$.
3. - Pro každou distribuční funkci platí $\lim\limits_{x \to -\infty} F(x) = 0$ a $\lim\limits_{x \to \infty} F(x) = 1$. Zapsáno zkráceně: $F(-\infty) = 0$, $F(\infty) = 1$.
   - Jestliže jsou možné hodnoty $F(x)$ z intervalu (a, b), pak $F(a) = 0$, $F(b) = 1$.

Každou funkci splňující předchozí tři vlastnosti můžeme pokládat za distribuční funkci.

ABSOLUTNĚ SPOJITÁ DISTRIBUČNÍ FUNKCE

- Distribuční funkce F se nazývá absolutně spojitá, jestliže existuje nezáporná funkce f (hustota pravděpodobnosti) a platí

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u \qquad \text{pro každé } x \in X.$$

- Hustota splňuje

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1\,.$$

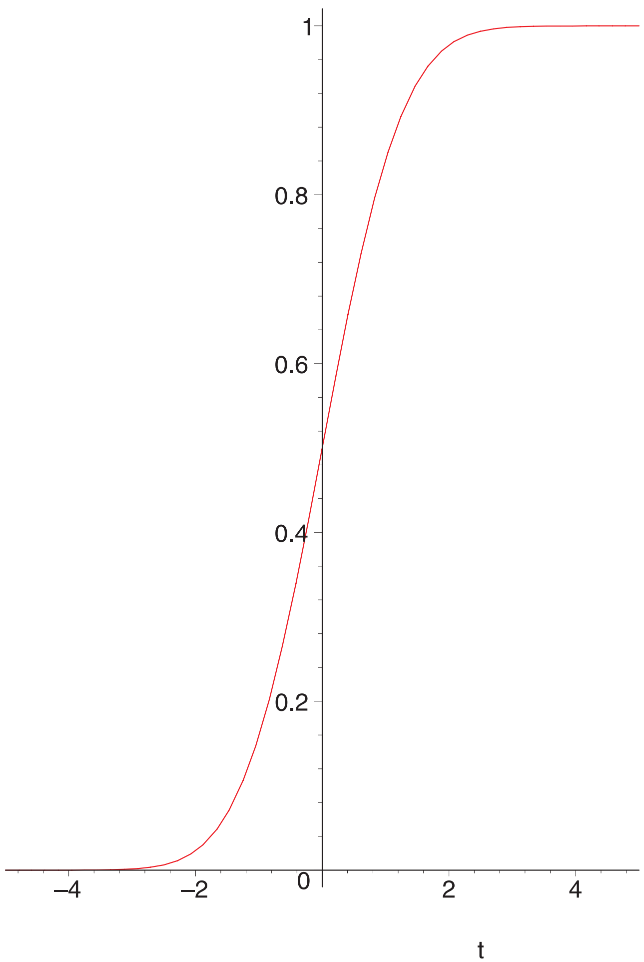
- Existuje-li derivace $F(x)$ v bodě x, potom $F'(x) = f(x)$.
- Pro $a, b \in \mathbb{R}$, $a < b$ platí

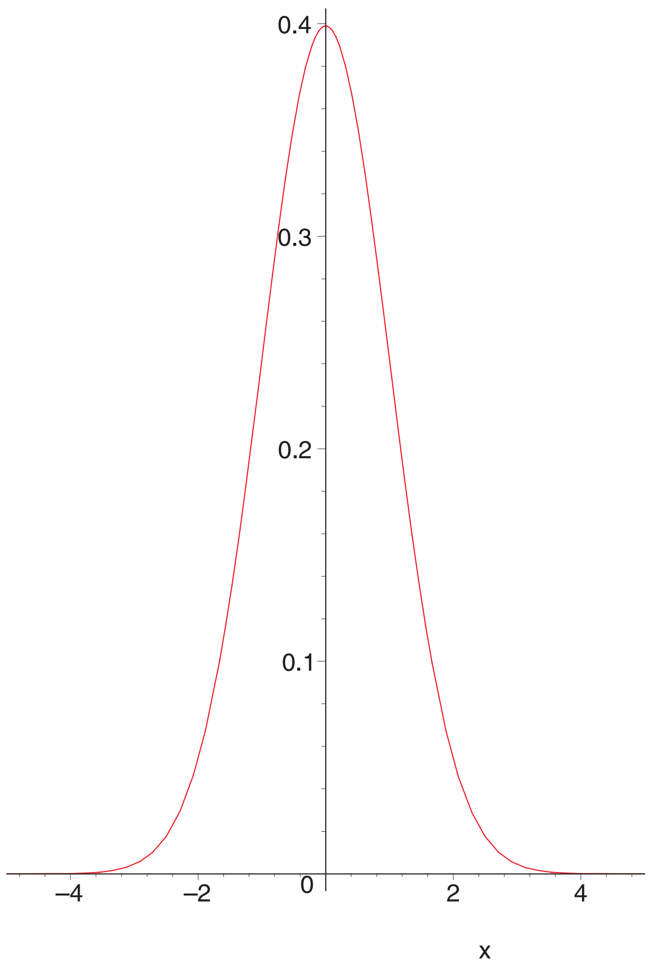
$$P(a < X < b) = \int_{a}^{b} f(x)\,\mathrm{d}x = F(b) - F(a)\,.$$

PŘÍKLAD, NORMÁLNÍ ROZDĚLENÍ

$$F(x)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{-x^2}{2\sigma^2}}$$





Distribuční funkce

Hustota

ZÁKON VELKÝCH ČÍSEL

Zákon velkých čísel říká, že při velkém počtu nezávislých pokusů je možné téměř jistě očekávat, že relativní četnost se bude blížit teoretické hodnotě pravděpodobnosti.

Jakob Bernoulli, Ars Conjectandi: Usum & Applicationem Praecedentis Doctrinae in Civilibus, Moralibus & Oeconomicis, 1713, Chapter 4.

ZÁKLADNÍ CHARAKTERISTIKY NÁHODNÉ VELIČINY

Spojité rozdělení	Diskrétní rozdělení

Střední hodnota

$$E(x) = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x \qquad E(x) = \sum_x x\, P(x)$$

k-tý obecný moment

$$E(x^k) = \int_{-\infty}^{\infty} x^k\, f(x)\, \mathrm{d}x \qquad E(x) = \sum_x x^k\, P(x)$$

k-tý centrální moment

$$E(x^k) = \int_{-\infty}^{\infty} (x-\mu_x)^k\, f(x)\, \mathrm{d}x \qquad E(x) = \sum_x x^k\, P(x)$$

Rozptyl, též disperze

$$D(x) = \int_{-\infty}^{\infty} (x-\mu_x)^2\, f(x)\, \mathrm{d}x \qquad E(x) = \sum_x x^2\, P(x)$$

KOVARIANCE

Vzájemná kovariance dvou veličin X, Y

$$\sigma_{xy} = E\left((X - \mu_x)(Y - \mu_y)\right)$$

Kovarianční matice n veličin $X_1, \ldots, X_n$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \ldots & \sigma_{1n} \\ & \ddots & \\ \sigma_{n_1} & \ldots & \sigma_n^2 \end{bmatrix}$$

Kovarianční matice je symmetrická, pozitivně semidefinitní.

KVANTILY, MEDIÁN

- p-kvantil Q_p

$$P(X < Q_p) = p$$

- Medián je p-kvantil pro $p = 0,5$

$$P(X < Q_p) = 0,5$$

Poznámka: Medián se používá jako náhrada střední hodnoty v robustní statistice.