Základy teorie pravděpodobnosti

Informační a znalostní systémy

Teorie pravděpodobnosti není v podstatě nic jiného než vyjádření obecného povědomí počítáním.

P. S. de Laplace

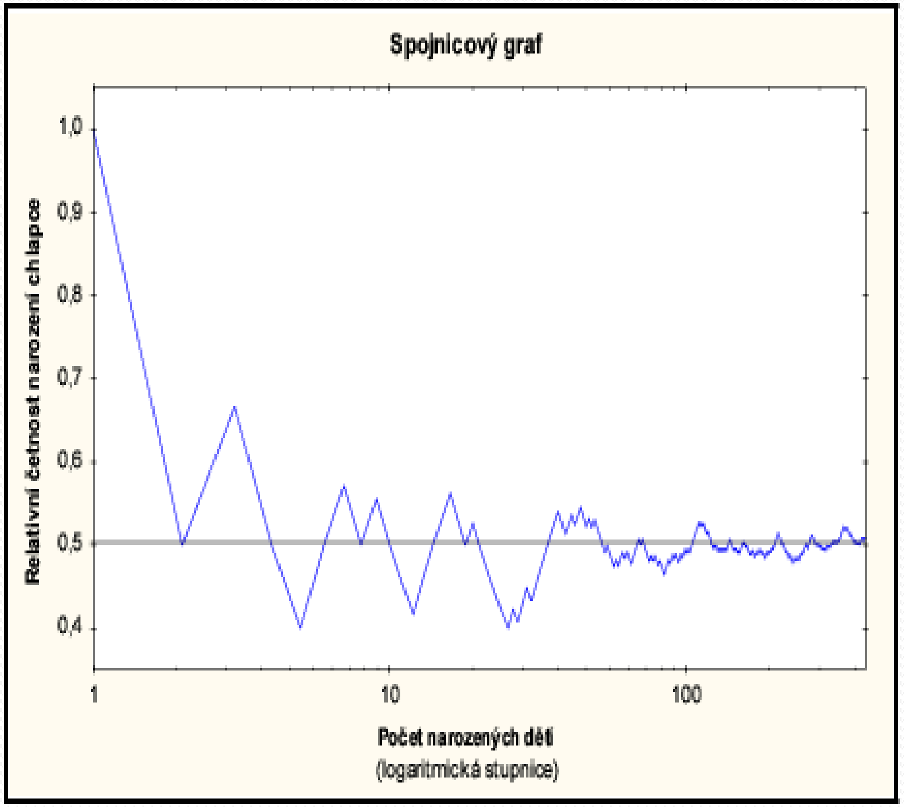
Pravděpodobnost a relativní četnost

- Pokusy, výsledky nejsou jednoznačně předurčeny podmínkami, za kterých probíhají.
- Pokusy, které jsou (teoreticky) neomezeně mnohokrát opakovatelné, nazýváme *náhodné pokusy*.
  - *házení hracími kostkami nebo mincemi, tahání losů z osudí, míchání karet*
  - Praxe: výběrová šetření lze realizovat náhodný výběr losováním -> lze určit pravděp., s jakými dostáváme výběry strukturou odlišné od *základního souboru* (*populace*)
- Význam počtu pravděpodobnosti: vhodné vyjádření či zachycení náhodné stránky experimentů např. v biologii a medicíně

Náhodný jev

Jakékoli tvrzení o výsledku, o kterém lze po uskutečnění pokusu či pozorování rozhodnout, zda je či není pravdivé

- Např. náhodný jev *A* "narození chlapce": v okamžiku početí, i těsně před porodem není předpověď pohlaví budoucího novorozence příliš spolehlivá
- Zaznamenávání jevu do posloupnosti
  - Četnost, s jakou nastává *A* pro lib. dlouhou posl. charakterizuje,
  - *n* ... délka posloupnosti (rozsah výběru)
  - *r* ... počet narozených chlapců
  - *r ... absolutní četnost*
  - *r/n ... relativní četnost* výskytu náhodného jevu *A* ve výběru o rozsahu *n*
  - S rostoucím rozsahem výběru se rel. četnosti ustalují kolem 0,5



Rel. četnost "narození chlapce" (log. stupnice)

Pravděpodobnost náhod. jevu

- Náhodný jev A je charakterizován číslem P(A), které je mírou „častosti“ výskytu tohoto jevu a nazývá se *pravděpodobnost náhodného jevu A*.
- Základní vlastnosti
    - P(A) nabývá hodnot z <0;1>, V případě, že A je jistý jev (nastává vždycky), je P(A)=1. V případě, že A je nemožný jev (nikdy nenastane), je P(A)=0.
    - Mnohonásobné nezávislé opakování náh. pokusu: prakticky jisté, že se relativní četnost výskytu náhodného jevu A jen nepatrně liší od P(A)
    - P(A) nepatrně odlišný od 0: prakticky jisté, že při jediném pokusu A nenastane
    - P(A) nepatrně odlišný od 1: prakticky jisté, že při jediném pokusu A nastane

    - V praxi pravděpodobnosti náh. jevů odhadujeme pomocí rel. četností. Kvalita odhadů vzrůstá s rostoucím počtem provedených pokusů

Odhady: statistické údaje

- Odhad pravděp. náh. jevu "narození chlapce" v naší populaci
- Rel. četnosti "narození chlapce" z celkového počtu živě narozených dětí v ČSSR (1966-1975)
- Rel. četnosti "narození chlapce" se v jedn. letech téměř neliší
- Lze usuzovat: P(A) se nemění - o něco vyšší než 0,51

Rok	Počet živě narozených dětí	Relativní četnost jevu "narození chlapce"
1966	222 615	0,5131
1967	215 985	0,5146
1968	213 807	0,5138
1969	222 934	0,5145
1970	228 531	0,5125
1971	237 242	0,5134
1972	251 455	0,5133
1973	274 703	0,5144
1974	291 367	0,5135
1975	289 425	0,5108
Celkem	2 448 064	0,5133

Pravidla pro počítání s pravděp.

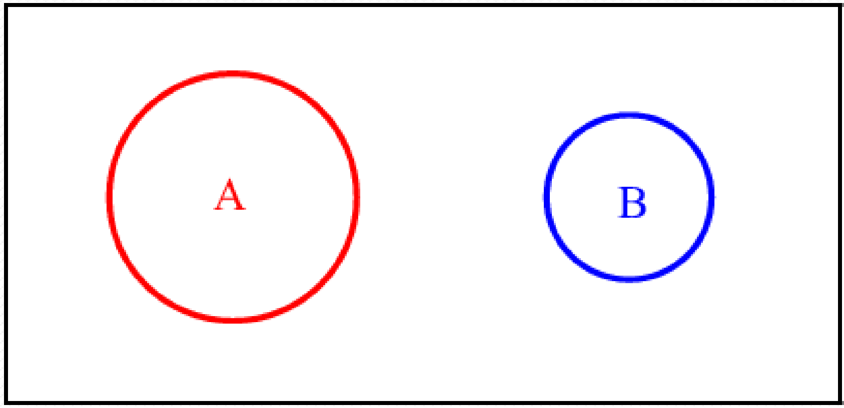
- V praxi se zajímáme o více jevů současně a o jejich vzájemné interakce.
- $C = (A,B)$ … nastanou oba jevy A a B současně
- $D = (A$ nebo $B)$ … nastane alespoň jeden z jevů A a B, tj. buď jev A, nebo jev B, či oba jevy A a B současně (tj. jev C).
- Pomocí pravděpodobností jevů A, B a $C = (A,B)$:
  P(*A* nebo *B*) = P(A) + P(B) – P(A,B)

Příklad 1

- V náhodně vybrané skupině 140 mužů věku 40-50 let ohrožených ateriální hypertenzí se vyskytl rizikový faktor "**zvýšený cholesterol**" (**jev** A) ve 37 případech a rizikový faktor "**kouření**" (**jev**B) v 98 případech. Ve 31 případech jsme zjistili současný výskyt obou rizikových faktorů. Odhadněte pomocí relativních četností pravděpodobnosti výskytu jevů A, B, $C = (A,B)$ a $D = (A$ nebo $B)$.

- *Řešení:*
  - P(A)=37/140=0,2643
  - P(B)=98/140=0,7
  - P(C) =31/140=0,2214
  - P(D)=0,2643+0.7-0,2214=0,7429

Neslučitelné jevy



- Nemohou nastat současně
- P(A,B) = 0
- pravidlo o sčítání pravděpodobností:

  P(A nebo B)=P(A)+P(B)

  obecněji:

  $P(A_1$ nebo A_2 nebo … nebo $A_k) = \sum^k_1 P(A_i)$

- Speciální případ: **opačné jevy (doplňkové jevy)**
  - A, ¬A
  - např. "narození chlapce" a "narození dívky„

  P(¬A)=1-P(A)

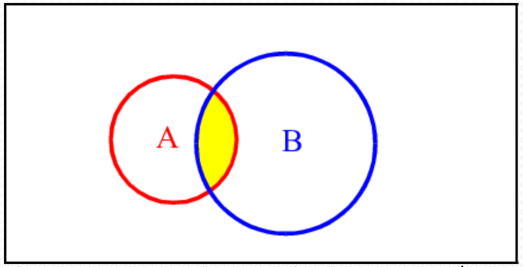
Příklad 2

- Pravděpodobnost jevu *A* "narození chlapce" je 0,51, spočtěte pravděpodobnost jevu ¬A "narození dívky"

- *Řešení:* Jev ¬A "narození dívky" je opačným jevem k jevu *A* "narození chlapce„

$$P(\neg A) = 1 - P(A) = 1 - 0{,}51 = 0{,}49$$

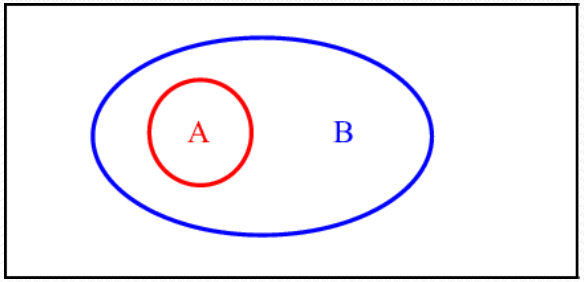
Podmíněná pravděpodobnost



- **Slučitelné jevy** mohou nastat současně
- Nelze použít pravidlo o sčítání pravděpodobností, ale:

  P(*A* nebo *B*) = P(A) + P(B) – P(A,B)

- Někdy se zajímáme o výskyt *A* jen, pokud nastal určitý jev *B* s kladnou pravděpodobností (tj. může nastat)
- Víme-li že nastal *B*, může se tím změnit i pravděpodobnost výskytu *A*:
  - Jevy neslučitelné s *B* se stanou nemožnými
  - Jevy deterministicky určené *B* se stanou jistými (obr.: *A* nastává vždy, když nastane jev *B*)



  - Ostatní jevy se mohou vyskytnout s pravděpodobnostmi obecně odlišnými od původních
- Pravděpodobnosti jevů, zjištěné za podmínky výskytu jevu *B*, se nazývají *podmíněné pravděpodobnosti* vzhledem k jevu *B*.
- ***Podmíněná pravděpodobnost*** jevu *A* vzhledem k jevu *B* je definována: $\mathrm{P}(A|B) = \frac{\mathrm{P}(A,B)}{\mathrm{P}(B)}.$
  - Tudíž pravděpodobnost současného výskytu dvou jevů *A* a *B* $\mathrm{P}(A,B) = \mathrm{P}(A|B)\mathrm{P}(B)$

Příklad 3

- Odhadněte podmíněnou pravděpodobnost výskytu faktoru "**zvýšený cholesterol**" (**jev *A***) *za podmínky* výskytu faktoru "**kouření**" (**jev *B***) - viz příklad 1

- *Řešení:*

  Odhad P(A|B) spočteme, jestliže za pravděpodobnosti jevů *A* a *B* dosadíme jejich odhady pomocí relativních četností. Dostaneme

  P(A|B) = 0,2214/0,7 = 0,3163

  Stejný výsledek musíme dostat, jestliže z celkového počtu 98 případů, ve kterých nastal *B*, stanovíme počet případů, ve kterých zároveň nastal jev *A*: 31.

  Odhad je tedy **P(A|B) = 31/98 = 0,3163** a vyjadřuje relativní četnost jevu *A* mezi případy, kdy nastal jev *B*

Nezávislé jevy

- Jevy *A* a *B* jsou *nezávislé*, jestliže výskyt jednoho jevu neovlivňuje výskyt druhého jevu. Matematické vyjádření tohoto faktu zapíšeme pomocí podmíněné pravděpodobnosti jako $P(A|B) = P(A)$ nebo obdobně $P(B|A) = P(B)$.
- Vidíme tedy, že pro nezávislé jevy *A*, *B* platí ***pravidlo o násobení pravděpodobností***:

$$\mathbf{P(A, B) = P(A)P(B)}$$

Obecněji:

$$\mathbf{P(A_1, A_2, \dots, A_k) = P(A_1)P(A_2) \dots P(A_k)}$$

Příklad 4

- Zjistěte, zda faktory *A* a *B* uvedené v příkladu 1 se vyskytují nezávisle, jestliže vypočtené relativní četnosti považujeme za skutečné pravděpodobnosti.

- *Řešení:*

  V případě nezávislosti faktorů *A* a *B* platí $P(A|B)=P(A)$. Z dat příkladu dostáváme $P(A|B)=0{,}3163$, což se liší od pravděpodobnosti $P(A)=0{,}2643$. Jevy *A* a *B* tedy **nejsou nezávislé.**

Příklad 5

- Jev *A*: "první novorozenec narozený v příštím kalendářním roce v ČR je chlapec“
- Jev *B*: "druhý novorozenec narozený v příštím kalendářním roce v ČR je chlapec"
- Vyloučíme-li vícečetné porody, spočtěte pravděpodobnost jevu *C*, že "oba novorozenci jsou chlapci" za předpokladu, že pravděpodobnost narození chlapce je 0,51

- *Řešení:*

  Výskyt jevu *A* neovlivňuje výskyt jevu *B*, tudíž jevy *A* a *B* jsou nezávislé. Pravděpodobnost jevu $C = (A,B)$, že oba novorozenci jsou chlapci:

$$\mathbf{P(C) = P(A)P(B) = 0{,}51 * 0{,}51 = 0{,}2601}$$

Příklad 6

- Pravděp. jevu A "osoba má pravé oko modré" je 0,3 a pravděp. jevu B "osoba má levé oko modré" je také 0,3.
- Jestliže pravděpodobnost, že "osoba má pravé oko modré" za podmínky, že nastal jev "osoba má levé oko modré" je rovna 1, spočtěte pravděpodobnost jevu C "osoba má obě oči modré".

- *Řešení:* Jevy A a B nejsou nezávislé, neboť

  $P(A) = 0{,}3$ a $P(A|B)=1$

  Pravděpodobnost jevu $C = (A, B)$:

  $$\mathbf{P(C) = P(A|B)P(B) = 1 * 0{,}3 = 0{,}3}$$

Příklad 7

- Za předpokladu, že pravděpodobnost narození chlapce je 0,51, spočtěte, jaká je pravděpodobnost, že v sérii čtyř po sobě narozených dětí (vícečetné porody vyloučíme), bude právě jeden chlapec.

- *Řešení:*
- Označme *C* jev, že mezi čtyřmi novorozenci je právě jeden chlapec. Konkrétní možnosti, které vytvářejí jev *C*, jsou:
  $C_1 = (A, \neg A, \neg A, \neg A)$ … chlapec se narodí jako první, $C_2 = (\neg A, A, \neg A, \neg A)$ … druhý, $C_3 = (\neg A, \neg A, A, \neg A)$ … třetí, $C_4 = (\neg A, \neg A, \neg A, A)$ … čtvrtý
- C_1, C_2, C_3 a C_4 … vzájemně neslučitelné.
  Z pravidla o sčítání pravděpodobností:
  $$P(C) = P(C_1) + P(C_2) + P(C_3) + P(C_4)$$
- Pravděpodobnosti jevů C_1, C_2, C_3 a C_4 jsou všechny stejné a jsou vypočteny pomocí pravidla o násobení pravděpodobností. Např.:
  $$P(C_1) = P(A)P(\neg A)P(\neg A)P(\neg A) = 0{,}51 * 0{,}49^3 = 0{,}06$$

  $$\mathbf{P(C) = 4 * 0{,}06 = 0{,}24}$$

Bayesův vzorec

- Předpokládejme, že náhodné jevy B_i, kde i=1, 2, 3, …, k, jsou vzájemně neslučitelné a v každém pokusu nastává právě jeden z nich, takže musí platit:
  $P(B_1 \text{ nebo } B_2 \text{ nebo } \dots \text{ nebo } B_k)\ \sum^{k}_{1} P(B_i) = 1$
- Známe-li podmíněné pravděpodobnosti jevu A za podmínky výskytu jevu B_i: $P(A|B_i)$ pro i=1, 2, 3, …, k, potom pravděpodobnost jevu A lze vyjádřit:

$$\mathbf{P(A)=\sum^{k}_{1} P(A, B_i) = \sum^{k}_{1} P(A|B_i)P(B_i)}$$

\- ***Pravidlo o úplné pravděpodobnosti***

Příklad 8

- Předpoklad: pravděpodobnost "**úrazu**" (**jev *A***)
  - u "**dítěte**" (**jev B_1**) je $P(A|B_1) = 0,2$,
  - u "**osoby v reprodukčním věku**" (**jev B_2**) je $P(A|B_2) = 0,2$
  - u "**osoby v postreprodukčním věku**" (**jev B_3**) je $P(A|B_3) = 0,4$
- Pravděpodobnosti, že osoba bude patřit do některé z těchto skupin, jsou $P(B_1)=0,25$, $P(B_2)=0,6$ a $P(B_3)=0,15$
- Spočtěte pravděpodobnost úrazu v dané populaci.

- *Řešení:*

  B_1, B_2 a B_3 jsou vzájemně neslučitelné a v každém případě nastává právě jeden z nich.

  Ze znalosti podmíněných pravděpodobností výskytu úrazu v jednotlivých věkových kategoriích obyvatelstva a ze znalostí pravděpodobností těchto kategorií spočteme pravděpodobnost úrazu v populaci jako:

$$\mathbf{P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) =}$$
$$\mathbf{= 0,2*0,25 + 0,2*0,6 + 0,4*0,15 = 0,23}$$

Bayesův vzorec

… udává, jakým způsobem vypočítáme pravděpodobnosti $\mathrm{P}(B_j|A)$ jevu B_j za podmínky, že nastal jev A, jestliže známe *apriorní* pravděpodobnosti $\mathrm{P}(B_i)$ a *podmíněné* pravděpodobnosti pro všechny jevy B_i,

i=1, 2, 3, …, k. Bayesův vzorec má tvar

$$\mathrm{P}(B_j|A) = \frac{\mathrm{P}(A|B_j)\mathrm{P}(B_j)}{\sum_{i=1}^{k} \mathrm{P}(A|B_i)\mathrm{P}(B_i)}.$$

- Odvození Bayesova vzorce:

$$\begin{aligned}
\mathrm{P}(B_j, A) &= \mathrm{P}(A, B_j) = \mathrm{P}(A|B_j)\mathrm{P}(B_j), \\
\mathrm{P}(B_j|A) &= \frac{\mathrm{P}(B_j, A)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A|B_j)\mathrm{P}(B_j)}{\mathrm{P}(A)}, \\
\text{kde } \mathrm{P}(A) &= \sum_{i=1}^{k} \mathrm{P}(A, B_i) = \sum_{i=1}^{k} \mathrm{P}(A|B_i)\mathrm{P}(B_i).
\end{aligned}$$

Příklad 9

- Pravděpodobnost, že "**osoba je kuřák**" **(jev *A*)** ve skupině "**osob s chronickou bronchitidou**" **(jev *B*$_1$)** je $P(A|B_1) = 0{,}75$
- Pravděpodobnost, že "osoba je kuřák" ve skupině "**osob bez chronické bronchitidy**" **(jev *B*$_2$)** je $P(A|B_2) = 0{,}5$
- Pravděpodobnost "výskytu osoby s chronickou bronchitidou" v populaci $P(B_1) = 0{,}4$
- Pravděpodobnost "výskytu osoby bez chronické bronchitidy" v populaci $P(B_2) = 0{,}6$.
- Spočtěte pravděpodobnost výskytu chronické bronchitidy u kuřáka.

- *Řešení:* Pomocí Bayesova vzorce dostaneme, že pravděpodobnost výskytu chronické bronchitidy u kuřáka je

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = \\ &= \frac{0{,}75 \cdot 0{,}40}{0{,}75 \cdot 0{,}40 + 0{,}50 \cdot 0{,}60} = 0{,}50. \end{aligned}$$

Šance a pravděpodobnost

- Řekneme-li že **šance O(A) (odds)** závodního koně na první místo v dostihovém závodě (jev *A*) je 1 ku 4, znamená to, že kůň závod vyhraje s P(A) = 1/5 = 0,2.
- Převod šancí na pravděpodobnosti: sečtením 1 + 4 = 5 dostaneme jmenovatel zlomku pro vyjádření pravděpodobnosti výhry, tj. 1/5.
- Pro libovolný náhodný jev *A* tedy platí: *šance* O(A) výskytu jevu *A* je

$$\mathrm{O}(A) = \frac{\mathrm{P}(A)}{\mathrm{P}(\neg A)} = \frac{\mathrm{P}(A)}{1 - \mathrm{P}(A)} \qquad \mathrm{P}(A) = \frac{\mathrm{O}(A)}{1 + \mathrm{O}(A)}.$$