Úvod do pravděpodobnosti

Kombinatorika pro bioinformatiky (NDMI089)

RNDr. Ondřej Pangrác, Ph.D.

Pravděpodobnostní prostory

Definice

Konečný pravděpodobnostní prostor je trojice $(\Omega, \mathcal{F}, P)$*, kde* Ω *je konečná množina,* $\mathcal{F} = 2^\Omega$ *a* $P : \mathcal{F} \mapsto [0, 1]$ *splňující pro každou* $A \subseteq \Omega$

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

a $P(\Omega) = 1$*.*

Příklady pravděpodobnostních prostorů:

Hod (spravedlivou) mincí: $\Omega = \{R, L\}$, $P(R) = P(L) = \frac{1}{2}$.
Hod kostkou: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $P(k) = \frac{1}{6}$ pro $k = 1, 2, 3, 4, 5, 6$.
Opakovaný hod mincí: $\Omega = \{R, L\}^n$, $P(\omega) = \frac{1}{2^n}$ pro $\forall \omega \in \{R, L\}^n$.
Náhodná permutace: $\Omega = S_n$, $P(\pi) = \frac{1}{n!}$ pro všechny $\pi \in S_n$.

Pravděpodobnostní prostory

Definice

Pravděpodonostní prostor $(\Omega, 2^\Omega, P)$ *se nazývá uniformní (též klasický) pravděpodobnostní prostor, pokud* $P(A) = \frac{|A|}{|\Omega|}$ *pro všechny jevy* $A \subseteq \Omega$.

Definice

Součin pravděpodobnostních prostorů $(\Omega_1, 2^{\Omega_1}, P_1)$ *a* $(\Omega_2, 2^{\Omega_2}, P_2)$ *je prostor* $(\Omega, 2^\Omega, P)$, *kde* $\Omega = \Omega_1 \times \Omega_2$ *a pro* $A \subseteq \Omega$ *je*

$$P(A) = \sum_{(\omega_1,\omega_2)\in A} P_1(\{\omega_1\}) \cdot P_2(\{\omega_2\}) .$$

Podmíněná pravděpodobnost

Definice

Jevy A *a* B *se nazývají nezávislé, pokud* $P(A \cap B) = P(A) \cdot P(B)$. *Obecně, jevy* $A_1, A_2, \ldots, A_n$ *jsou nezávislé, pokud pro každou (neprázdnou) podmnožinu* $I \subseteq \{1, 2, \ldots, n\}$ *platí* $P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$.

Definice

Podmíněná pravděpodobnost jevu A *za předpokladu, že nastal jev* B, *se definuje jako*

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

(je def. pouze pro $P(B) > 0$*).*

Pozorování

Jsou-li jevy A, B *nezávislé a* $P(B) > 0$, *pak* $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$. *A naopak, pokud* $P(A|B) = P(A)$, *jedná se o nezávislé jevy.*

Věta o úplné pravděpodobnosti

Věta

Nechť A je nějaký jev a $B_1, \dots, B_n$ jsou disjunktní jevy pokrývající celé Ω, navíc $P(B_i) > 0$ pro $i = 1, \dots, n$. Pak

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i) \ .$$

Důkaz: Protože jevy $B_1, \dots, B_n$ pokrývají Ω, platí pro každé A

$$A = \bigcup_{i=1}^{n} (A \cap B_i) \ .$$

Z jejich disjunktnosti plyne, že i členy $A \cap B_i$ jsou disjunktní a tedy $P(A) = \sum_{i=1}^{n} P(A \cap B_i)$, což dále upravujeme s využitím vztahu

$$P(A \cap B_i) = P(A|B_i) \cdot P(B_i) \ .$$

A dotáváme

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

Bayesova věta

Věta

Nechť A je nějaký jev a $B_1, \dots, B_n$ jsou disjunktní jevy pokrývající celé Ω, navíc $P(B_i) > 0$ pro $i = 1, \dots, n$. Pak

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)} .$$

Důkaz: Z def. podmíněné pravděpodonosti vyjádříme pravděpodobnost průniku dvou jevů oběma možnými způsoby

$$P(A \cap B_k) = P(A|B_k) \cdot P(B_k) = P(B_k|A) \cdot P(A)$$

Z toho

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

a za jmenovatele zlomku dosadíme z Věty o úplné pravděpodobnosti.

Názorný příklad

Uvažujme, že v nějaké populaci je četnost bezejmenné choroby 1 osoba z 1000. K dispozici máme test, jehož spolehlivost je 95% (u skutečně nemocné osoby vyjde test s pravděpodobností 95% pozitivně, a taktéž u zdravé osoby vyjde test s pravděpodobností 95% negativně). Jaké je pravděpodobnost, že testovaná osoba je skutečně nemocná, pokud jí vyšel pozitivní test?

Zavedeme si značení: N bude označovat jev, že daná osoba je nemocná, a $\overline{N}$ jev, že nemocná není. Analogicky, T značí pozitivní test a $\overline{T}$ negativní test. Zadání pak znamená $P(N) = \frac{1}{1000}$, $P(T|N) = P(\overline{T}|\overline{N}) = \frac{19}{20}$ a zajímá nás $P(N|T)$.
Podle definice podmíněné pravděpodobnosti je

$$P(N|T) = \frac{P(N \cap T)}{P(T)} \; .$$

Názorný příklad

Čitatele lze vyjádřit z

$$P(T|N) = \frac{P(T \cap N)}{P(N)}$$

jako

$$P(N \cap T) = P(T|N) \cdot P(N) = \frac{19}{20} \cdot \frac{1}{1000} = \frac{19}{20000} \; .$$

Jmenovatele v řešeném výrazu spočteme

$$\begin{aligned} P(T) &= P(T|N) \cdot P(N) + P(T|\overline{N}) \cdot P(\overline{N}) = \\ &= P(T|N) \cdot P(N) + (1 - P(\overline{T}|\overline{N})) \cdot P(\overline{N}) = \frac{19}{20} \cdot \frac{1}{1000} + \frac{1}{20} \cdot \frac{999}{1000} = \frac{1018}{20000} \; . \end{aligned}$$

Nyní již jen dosadíme

$$P(N|T) = \frac{P(N \cap T)}{P(T)} = \frac{\frac{19}{20000}}{\frac{1018}{20000}} = \frac{19}{1018} < \frac{20}{1000} = 0.02 \; .$$