

Bayesova věta

Jednoduchá forma Bayesovy věty

Bayesova věta se zabývá podmíněnými pravděpodobnostmi jevů. Pro formální zápis podmíněné pravděpodobnosti jevu **A** za předpokladu výskytu jevu **B** používáme zápis **P (A/B)**. Anglický duchovní Thomas Bayes (1702–1761) objevil, že podmíněná pravděpodobnost **P (A/B)** souvisí s opačně podmíněnou pravděpodobností **P (B/A)** takto (jednoduchá forma Bayesovy věty):

Máme-li dva náhodné jevy **A** a **B**, jejichž pravděpodobnosti jsou **P (A)** a **P (B)**, a pokud **P (B) > 0**, potom platí

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)} \quad (\text{jednoduchý Bayesův vzorec})$$

Příklady použití jednoduché Bayesovy věty

Příklad č. 1

Předpokládejme, že máme školu s 60 % chlapců a 40 % dívek. Všichni chlapci nosí kalhoty. Z dívek nosí kalhoty polovina. Pozorovatel vidí z dálky studenta v kalhotách. Jaká je pravděpodobnost, že tento student je dívka?

Jako událost **G** označíme, že pozorovaný student je dívka. Jako událost **T**, že pozorovaný student nosí kalhoty. Pro výpočet podmíněné pravděpodobnosti **P (G/T)**, že student v kalhotách je dívka, potřebujeme vědět:

- **P (T/G)**, neboli pravděpodobnost, že náhodně vybraná dívka nosí kalhoty. Vzhledem k tomu, že dívky nosí kalhoty a sukně stejně často, je tato pravděpodobnost 0,5.
- **P (G)**, neboli pravděpodobnost, že student je dívka. Vzhledem k tomu, že pozorovatel vidí náhodného studenta a zastoupení dívek je 40 %, je tato pravděpodobnost 0,4.
- **P (T)**, pravděpodobnost, že náhodně vybraný student bude nosit kalhoty. Víme přitom, že polovina dívek a všichni chlapci nosí kalhoty, takže to bude $0,4 \times 0,5 + 0,6 \times 1,0 = 0,8$.

Pak můžeme použít Bayesův vzorec a dostáváme: $P(G|T) = \frac{P(T|G)P(G)}{P(T)} = \frac{0.5 \times 0.4}{0.8} = 0.25$

Další, v podstatě rovnocenný způsob, jak získat stejný výsledek je následující: Předpokládejme, abychom byli konkrétní, že žáků je celkem 100, 60 chlapců a 40 dívek. Z toho 60 chlapců a 20 dívek nosí kalhoty. Dohromady tedy 80 nositelů kalhot, z toho 20 dívek. Proto pravděpodobnost, že náhodný nositel kalhot je dívka, je rovna $20 / 80 = 0,25$.

Při výpočtech podmíněné pravděpodobnosti je často užitečné vytvořit si jednoduchou čtyřpolní tabulku, která obsahuje počty jednotlivých případů, nebo jejich relativní četnosti pro každou z nezávisle proměnných. Tedy např.:

	dívky	chlapci	celkem
kalhoty	20	60	80
sukně	20	0	20
celkem	40	60	100

Příklad č. 2

Podívejme se na příklad použití Bayesovy věty při testování na drogy. Předpokládejme, že test na zjištění drog má senzitivitu 99 % a specifitu 99 %. To znamená, že test správně identifikuje skutečného uživatele drog v 99 % případů a že test vyloučí osobu, která drogy neužívá rovněž v 99 % případů. Test se zdá být docela přesný, ale pomocí Bayesovy věty můžeme ukázat, že poměrně velké procento nevinných může být chybně označeno za uživatele drog. Předpokládejme, že ve společnosti, která se rozhodla testovat své zaměstnance na užívání drog je prevalence 0,5 %. Tj. jen 0,5 % ze všech zaměstnanců drogy skutečně bere.

Jaká potom bude pravděpodobnost, že ten kdo měl pozitivní test je opravdu uživatel drog?

Nechť "**D**", označuje uživatele drog, a "**N**" ostatní. Nechť současně "+" znamená pozitivní výsledek testu. Budeme potřebovat znát některé veličiny:

- **P (D)**, pravděpodobnost, že zaměstnanec je uživatelem drog, tedy prevalence. Tu známe, činí 0,5 % (nebo 0,005).
- **P (N)**, pravděpodobnost, že zaměstnanec není uživatelem drog je $1 - P (D)$, tedy 0,995.
- **P (+/D)** pravděpodobnost, že test je pozitivní, když je zaměstnanec uživatelem drog. Je to 0,99, protože senzitivita je 99 %.

- $P(+|N)$ je pravděpodobnost, že test bude pozitivní, i když zaměstnanec drogy neužívá. Je to 0,01, protože je to doplněk k specificitě, která je 99 %.
- $P(+)$ je pravděpodobnost pozitivního výsledku testu. U zaměstnanců, kteří užívají drogy je to $0,99 \times 0,005$ a u neužívajících je to $0,01 \times 0,995$. Tj. $0,00495 + 0,00995 = 0,0149$ (tj. 1,49 %)

Na základě těchto údajů můžeme vypočítat příslušnou podmíněnou pravděpodobnost, že ten, kdo měl pozitivní test je opravdu uživatel drog $P(D|+)$

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{0,99 \times 0,005}{0,0149} = 0,3322.$$

Přes vysokou senzitivitu a selektivitu je využitelnost testu nízká. U zaměstnance s pozitivním testem je jen 33% pravděpodobnost, že je skutečně uživatelem drog. Pravděpodobnost, že není uživatelem drog je větší (67%). **Čím vzácnější je onemocnění (nižší prevalence), tím větší bude procento falešně pozitivních testů (planý poplach).**

Bayesova věta

Předpokládejme, že náhodné jevy B_i (kde $i = 1, 2, 3, \dots, k$) jsou vzájemně neslučitelné a v každém pokusu nastává

právě jeden z nich, takže součet jejich pravděpodobností je roven jedné: $\sum_{i=1}^k P(B_i) = 1$

Známe-li podmíněné pravděpodobnosti $P(A | B_i)$ jevu A , za podmínky výskytu jevu B_i , pro všechna i a *apriorní* pravděpodobnosti jevu $P(B_i)$, potom můžeme pravděpodobnost jevu $P(B_j | A)$ vypočítat podle **Bayesova vzorce** [1]:

$$P(B_j|A) = \frac{P(A|B_j) P(B_j)}{P(A)} = \frac{P(A|B_j) P(B_j)}{\sum_{i=1}^k P(A|B_i) P(B_i)}. \quad (\text{Bayesův vzorec})$$

Kde jsme pro vyjádření $P(A)$ použili pravidlo o úplné pravděpodobnosti.

Odkazy

Související články

- ROC křivka
- Pravidlo o úplné pravděpodobnosti

Použitá literatura

- ZVÁROVÁ, Jana. *Základy statistiky pro biomedicínské obory*. 1. vydání. Praha : Karolinum, 1998. ISBN 80-7184-786-0.
- WIKIPEDIA. *Bayes' theorem* [online]. ©2002. Poslední revize 2011-07-14, [cit. 2011-07-17]. <https://en.wikipedia.org/wiki/Bayes%27_theorem>.

Reference

1. ZVÁROVÁ, Jana. *Základy statistiky pro biomedicínské obory*. 1. vydání. Praha : Karolinum, 1998. ISBN 80-7184-786-0.