Bayesovská analýza

- má v dnešní době široké uplatnění – lékařská diagnostika, kriminalistika, pojistná matematika apod.

- vychází z Bayesova vzorce

Definice

Nechť $(\Omega, \mathcal{A}, \mathrm{P})$ je pravděpodobnostní prostor, $B \in \mathcal{A}$ a $\mathrm{P}(B) > 0$. Potom číslo

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \tag{1}$$

nazýváme podmíněnou pravděpodobností jevu A

Bayesův vzorec pro náhodné jevy

– úpravou získáme vztah pro výpočet pravděpodobnosti průniku náhodných jevů A a B

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

,

– pravděpodobnost průniku náhodných jevů A a B můžeme psát také ve tvaru

$$\mathrm{P}(A \cap B) = \mathrm{P}(B|A) \cdot \mathrm{P}(A) \tag{2}$$

– dosazením do vztahu pro podmíněnou pravděpodobnost obdržíme **Bayesův vzorec**

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A) \cdot \mathrm{P}(A)}{\mathrm{P}(B)} \tag{3}$$

– Zákon inverzní pravděpodobnosti (Bernoulli, Bayes, Laplace)

– Přímá pravděpodobnost: známe mechanismus, z kombinatoriky vypočteme pravděpodobnosti výsledků

– Inverzní pravděpodobnost: vidíme výsledky, chceme informaci o mechanismu, který je generuje

– $P(A)$ je apriorní pravděpodobnost jevu A

– P$(B|A)$ je věrohodnost (likelihood)

– $P(A|B)$ je aposteriorní pravděpodobnost jevu A, tedy "nová" pravděpodobnost po pozorování jevu B (nová informace).

Věta (Vzorec pro úplnou pravděpodobnost)

Nechť $\Omega = \bigcup_{i=1}^{n} A_i$ *je disjunktní rozklad, tj.* $\{A_i\}_{i=1}^{n}$ *je posloupnost po dvou neslučitelných (disjunktních) náhodných jevů s* $\mathrm{P}(A_i) > 0$ *pro* $i = 1, 2, \ldots, n$. *Potom*

$$\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(B|A_i)\mathrm{P}(A_i). \tag{4}$$

Věta (Bayesův vzorec - 2.verze)

Nechť $\Omega = \bigcup_{i=1}^{n} A_i$ *je disjunktní rozklad s* $\mathrm{P}(A_i) > 0$ *pro* $i = 1, 2, \ldots, n$, *nechť* $B \in \mathcal{A}$ *a* $\mathrm{P}(B) > 0$. *Pak*

$$\mathrm{P}(A_j|B) = \frac{\mathrm{P}(B|A_j) \cdot \mathrm{P}(A_j)}{\sum_{i=1}^{n} \mathrm{P}(B|A_i)\mathrm{P}(A_i)}, \quad j = 1, 2, \ldots, n. \tag{5}$$

Bayesův vzorec se používá v případě, kdy

- máme úplný systém hypotéz $A_1, A_2, \ldots, A_n$, které se navzájem vylučují a vyčerpávají všechny možnosti; přitom známe jejich apriorní pravděpodobnosti $\mathrm{P}(A_i)$
- nastal jev B a známe podmíněné pravděpodobnosti $\mathrm{P}(B|A_i)$
- nás zajímají nové aposteriorní pravděpodobnosti $\mathrm{P}(A_j|B)$, jež berou v úvahu, že nastal jev B

Bayesův vzorec pro náhodné veličiny

Podmíněné rozdělení

- nechť X a Y jsou spojité náhodné veličiny definované na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathrm{P})$ se ***sdruženou hustotou*** $f_{X,Y}(x, y)$ a ***marginálními hustotami*** $f_X(x)$, $f_Y(y)$

- ***podmíněná hustota*** náhodné veličiny X při daném $Y = y$ je vyjádřena ve tvaru

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \tag{6}$$

- jsou-li X a Y nezávislé, pak platí

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y),$$

a podmíněná hustota je shodná s marginální hustotou

- ze vztahu (6) lze ***sdruženou hustotu*** vyjádřit jako součin podmíněné a marginální hustoty

$$f_{X,Y}(x,y) = f_{X|Y}(x|y) \cdot f_Y(y), \tag{7}$$

analogicky

$$f_{X,Y}(x,y) = f_{Y|X}(y|x) \cdot f_X(x) \tag{8}$$

- ***marginální hustotu*** náhodné veličiny X získáme integrováním sdružené hustoty přes všechny možné hodnoty y

$$f_X(x) = \int f_{X,Y}(x,y), \tag{9}$$

analogicky odvodíme marginální hustotu náhodné veličiny Y

$$f_Y(y) = \int f_{X,Y}(x,y)\,\mathrm{d}x \tag{10}$$

- z (9) a (7) vidíme, že

$$f_X(x) = \int f_{X|Y}(x|y) \cdot f_Y(y), \tag{11}$$

obdobně

$$f_Y(y) = \int f_{Y|X}(y|x) \cdot f_X(x)\,\mathrm{d}x \tag{12}$$

- využijeme vztahů (6) a (8) k odvození ***Bayesova vzorce***

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) \cdot f_X(x)}{f_Y(y)}, \tag{13}$$

ten můžeme také získat, položíme-li do rovnosti výrazy na pravé straně u obou rovnic (7) a (8). Typicky y jsou data a x jsou parametry

Podmíněné očekávání

- mějme podmíněnou hustotu náhodné veličiny X za podmínky, že $Y = y$

$$f_{X|Y}(x|y)$$

- potom můžeme podmíněné očekávání vyjádřit ve tvaru

$$\mathrm{E}(X|Y = y) = \int x \cdot f_{X|Y}(x|y)\,\mathrm{d}x \tag{14}$$

  - je funkcí y
  - lze ji chápat jako náhodnou veličinu, nahradíme-li y za Y na pravé straně předchozí rovnice, $\mathrm{E}(X|Y)$ je tedy náhodná veličina, která je funkcí Y

- střední hodnotu náhodné veličiny $\mathrm{E}(X|Y)$ lze získat

$$\begin{aligned}
\mathrm{E}\big[\mathrm{E}(X|Y)\big] &= \int \mathrm{E}(X|Y=y)\cdot f_Y(y) = \\
&= \int\int x\cdot f_{X|Y}(x|y)\,\mathrm{d}x\cdot f_Y(y) = \\
&= \int x\int f_{X|Y}(x|y)\cdot f_Y(y)\,\mathrm{d}x = \\
&= \int x\cdot f_X(x)\,\mathrm{d}x = \mathrm{E}(X)
\end{aligned} \tag{15}$$

- Pro celkový rozptyl obdržíme

$$\mathrm{E}\big[\mathrm{Var}(X|Y)\big] + \mathrm{Var}\big[\mathrm{E}(X|Y)\big] = \mathrm{Var}(X) \qquad (16)$$

- výše odvozená formule říká, že rozptyl náhodné veličiny X lze vyjádřit jako součet střední hodnoty podmíněného rozptylu a rozptylu podmíněné střední hodnoty

Bayesovský a frekventistický přístup

Rozdíly mezi bayesovským a frekventistickým přístupem

1. pravděpodobnost

   - FP – pravděpodobnost chápeme jako relativní četnost výskytu události v případě, kdy se s počtem opakovaní náhodného pokusu blížíme k nekonečnu

   - BP – pravděpodobnost měří důvěryhodnost nějakého tvrzení, založeného na informacích, které v daném okamžiku máme

   událost může být nejistá z důvodu náhodnosti nebo neznalosti

   – FP řeší pouze první případ nejistoty, zatímco bayesovský přístup řeší oba typy nejistot

2. parametr(-y)

   - FP – s parametrem se počítá jako s neznámou, leč pevnou konstantou (odhad parametru provádíme momentovou metodou, metodou maximální věrohodnosti)

   - BP – k neznámému parametru se přistupuje jako k náhodné veličině (díky uvedenému faktu lze odhadnou celé pravděpodobnostní rozdělení)

3. používané funkce

   - FP – parametry odhadujeme z věrohodnostní funkce

   - BP – u bayesovských metod navíc pracujeme s apriorní funkcí, tím zahrneme všechny dostupné relevantní informace

Optimální teorie kredibility

- byla poprvé zformulována roku 1967 (Bühlmann)

Rizikový parametr

- při stanovení výše pojistného u klienta se nejdříve vyhodnocují ratingová kritéria, na jejichž základě je klient zařazen do jedné z tarifních tříd, podle níž pak pojistitel určí sazbu

- každá ratingová třída je homogenní s ohledem na použitá kritéria, ve skutečnosti ale v každé z nich zůstává jistá míra heterogenity – existuje totiž možnost, že se klient bude odlišovat od toho, co očekáváme

- předpokládejme, že úroveň rizika každého klienta lze charakterizovat nezáporným rizikovým parametrem θ, resp. θ, který se u jednotlivých klientů liší – díky tomu můžeme klienty odlišit vzhledem k jejich rizikovému profilu

- parametr θ v sobě zahrnuje skryté rizikové faktory jednotlivých pojištěnců, nelze jej vypozorovat, jeho přesnou hodnotu tedy neznáme

- jelikož θ je různá pro různé klienty, jsme schopni v každé tarifní třídě určit pravděpodobnostní rozdělení $\pi_{\Theta}(\theta)$ udávající pravděpodobnost jednotlivých hodnot θ uvnitř dané třídy

- distribuční funkce $F_\Theta(\theta) = \mathrm{P}(\Theta \leq \theta)$ náhodné veličiny Θ udává pravděpodobnost, že náhodně vybraný pojistník z dané tarifní třídy bude mít hodnotu rizikového parametru menší nebo rovnu θ

- zkušenost jednotlivých pojistníků je jistým způsobem ovlivněna hodnotou θ

- škody, resp. ztráty X vycházejí z podmíněného rozdělení $f_{X|\Theta}(x|\theta)$ náhodné veličiny X při daném θ

Bayesovská metodologie

- mějme pro konkrétního klienta tato pozorování $X = x$, kde $X = (X_1, X_2, \ldots, X_n)$ a $x = (x_1, x_2, \ldots, x_n)$
- naším cílem je stanovit takovou sazbu, abychom pokryli škody, resp. ztráty v nadcházejícím období X_{n+1}
- předpokládejme, že θ je klientův rizikový parametr, jehož hodnotu neznáme a že škody, resp. ztráty $X_1, X_2, \ldots, X_n, X_{n+1}$ jsou při daném θ nezávislé
- nechť X_j, kde $j = 1, 2, \ldots, n, n+1$, má podmíněné rozdělení $f_{X_j|\Theta}(x_j|\theta)$

- pro známé θ bychom mohli použít $f_{X_{n+1}|\Theta}(x_{n+1}|\theta)$ pro předpověď škod, resp. ztrát v nadcházejícím období X_{n+1}

- θ bohužel neznáme, ale známe historii pojistných nároků x, využijeme ji k určení ***prediktivního rozdělení***, což je podmíněné rozdělení X_{n+1} při daném $X = x$

- za předpokladu nezávislosti X_j při daném $\Theta = \theta$ obdržíme ***sdružené rozdělení*** X a Θ v následujícím tvaru

$$f_{X,\Theta}(x,\theta) = f_{X|\Theta}(x|\theta) \cdot \pi_\Theta(\theta) \overset{nezáv.}{=} \left[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)\right] \pi_\Theta(\theta) \tag{17}$$

- ***marginální rozdělení*** X získáme z předchozího vztahu marginalizací – integrováním přes všechny možné hodnoty parametru θ

$$f_X(x) = \int f_{X,\Theta}(x,\theta)\mathrm{d}\theta = \int \left[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)\right] \pi_\Theta(\theta)\mathrm{d}\theta \tag{18}$$

- má-li Θ diskrétní rozdělení, nahradíme ve (18) integrál sumou

- analogicky odvodíme sdružené rozdělení náhodného vektoru $(X_1, X_2, \ldots, X_n, X_{n+1})$, značíme $f_{X,X_{n+1}}(x, x_{n+1})$ – na pravé straně (18) zaměníme v součinu n na $n+1$
- z předchozích vztahů dostaneme ***prediktivní rozdělení***

$$f_{X_{n+1}|X}(x_{n+1}|x) = \frac{f_{X,X_{n+1}}(x, x_{n+1})}{f_X(x)} = \\ = \frac{1}{f_X(x)} \cdot \int \left[\prod_{j=1}^{n+1} f_{X_j|\Theta}(x_j|\theta) \right] \pi_\Theta(\theta) \mathrm{d}\theta \tag{19}$$

- dále

$$\pi_{\Theta|X}(\theta|x) = \frac{f_{X,\Theta}(x, \theta)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta) \cdot \pi_\Theta(\theta)}{f_X(x)} = \\ = \frac{1}{f_X(x)} \cdot \left[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta) \right] \pi_\Theta(\theta) \tag{20}$$

- dosazením (20) do vztahu (19) pro prediktivní rozdělení máme

$$f_{X_{n+1}|X}(x_{n+1}|x) = \int f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \cdot \pi_{\Theta|X}(\theta|x)\mathrm{d}\theta \qquad (21)$$

- $\pi_{\Theta|X}(\theta|x)$ – ***posterior***, ***aposteriorní rozdělení***

  $f_{X|\Theta}(x|\theta)$ – ***věrohodnostní funkce***

  $\pi_{\Theta}(\theta)$ – ***prior***, ***apriorní rozdělení***

  $f_X(x)$ – ***normalizační faktor***, ***evidence***

- ve vztahu (20) se normalizační faktor někdy vynechává (např. při odhadu parametrů), jelikož nezávisí na θ

$$\pi_{\Theta|X}(\theta|x) \propto f_{X|\Theta}(x|\theta) \cdot \pi_{\Theta}(\theta) \qquad (22)$$