

Bayesovské metody

Mnohorozměrná analýza dat

Podmíněná pravděpodobnost

Definice: Uvažujme náhodné jevy A a B takové, že $\mathbb{P}(B) > 0$.

Podmíněnou pravděpodobností jevu A za podmínky, že nastal jev B , nazýváme podíl

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

Poznámka: Podmíněná pravděpodobnost $\mathbb{P}(A|B)$ je vlastně pravděpodobnost jevu A na prostoru B .

Věta o úplné pravděpodobnosti: Necht' $\Omega = \cup_i E_i$, kde E_i jsou po dvou disjunktí jevy. Potom pro pravděpodobnost jevu A platí

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|E_i) \cdot \mathbb{P}(E_i).$$

Bayesova věta

Bayesova věta: Necht' $\mathbb{P}(A) > 0$ a $\Omega = \cup_i E_i$, kde E_i jsou po dvou disjunktní jevy, $\mathbb{P}(E_i) > 0$. Potom

$$\mathbb{P}(E_j|A) = \frac{\mathbb{P}(A|E_j)\mathbb{P}(E_j)}{\sum_i \mathbb{P}(A|E_i)\mathbb{P}(E_i)}$$

Příklad (Hebák, AOP, 2012): Tento příklad vychází z výuky základního kurzu z pravděpodobnosti na VŠE. Studentům se jako příprava ke zkoušce doporučuje spočítat 100 příkladů. Tohoto doporučení neuposlechne 95 % studentů. Pro úspěšné napsání testu je třeba získat alespoň polovinu bodů. Při prvním pokusu jsou pak dlouhodobě výsledky následující: 20 % ze studentů, kteří příklady nespočetli, uspěje a ze studentů, kteří se připravili, uspěje 99 %. Spočtete pravděpodobnost úspěchu v testu a pravděpodobnost, že náhodně vybraný student se připravil (resp. nepřipravil), víme-li, že v testu uspěl.

Bayesova věta pro spojité náhodné veličiny

Uvažujme náhodné veličiny X a Y . Potom

$$f_{Y|X}(y|x) = \frac{f_Y(y)f_{X|Y}(x|y)}{\int f_Y(z)f_{X|Y}(x|z)dz}.$$

Funkce $f_{Y|X}(y|x)$ a $f_{X|Y}(x|y)$ jsou tzv. **podmíněné hustoty**. Na podmíněnou hustotu $f_{Y|X}(y|x)$ se můžeme (hrubě řečeno) dívat jako na hustotu rozdělení s pevně daným parametrem x .

Funkce f_Y je tzv. **apriorní hustota** veličiny Y a funkce $f_{Y|X}(y|x)$ tzv. **aposteriorní hustota** veličiny Y .

Bayesovský přístup ve statistice

Uvažujme nyní, že máme veličinu X , jejíž rozdělení závisí na parametru y (obecně vícerozměrného).

▶ **Klasická statistika:**

Z náhodného výběru z rozdělení veličiny X odhadujeme hodnotu parametru y .

▶ **Bayesovský přístup:**

Parametr rozdělení veličiny X považujeme za náhodnou veličinu Y . Na základě nějaké předchozí (**apriorní**) informace či zkušenosti můžeme odhadnout aposteriori podmíněnou hustotu veličiny Y (hustota **po** realizaci X) pomocí Bayesovy věty (podmíněné rozdělení X za podmínky $Y = y$ považujeme za známé).

Rozdíl: Při klasickém přístupu použijeme k závěrům o hodnotě parametru y pouze hustotu $f_{X|Y=y}$, kdežto v případě Bayesovském také apriorní informaci f_Y .

Bayesovský přístup-diskuze

Bayesovské metody mají své příznivce i odpůrce.

Výhody

- ▶ Využití apriorní informace

Nevýhody:

- ▶ Parametr je náhodná veličina
- ▶ Volba a použití apriorního rozdělení - na něm závisí výsledné rozdělení $f_{Y|X}$!

Je-li apriorní informace neurčitá (nebo dokonce žádná), vznikají problémy.

Poznámka: Apriorní hustota dokonce nemusí být ani hustota. Je-li f_Y obecně nezáporná integrovatelná funkce, aposteriorní rozdělení se nezmění. Někdy dokonce volíme za apriorní hustotu i nezápornou neintegrovatelnou funkci, ale s tím je třeba zacházet velice opatrně (někdy výsledky dávají dobrý smysl, někdy ne).

Některá využití Bayesovské statistiky

- ▶ Odhad parametru rozdělení y
- ▶ Simulace z cílového rozdělení, odhady parametrů (sekvenční MCMC metody)

Simulace z rozdělení - přímé metody

- ▶ **Simulace NV s diskretním rozdělením** (x_k, p_k) , $k = 1, 2, \dots$

Označme $I_1 = [0, p_1]$, $I_n = (\sum_{k=1}^{n-1} p_k, \sum_{k=1}^n p_k]$. Vygenerujeme náhodnou veličinu $U \sim R(0, 1)$. Najdeme n takové, že $U \in I_n$. Potom x_n výběr z daného rozdělení.

Poznámka: Pro některá rozdělení můžeme využít jejich interpretace (Binomické, Poissonovo, ...)

- ▶ **Simulace NV se spojitým rozdělením**

- (a) Inverzní metoda: Generujeme z rozdělení s distribuční funkcí F takovou, že k ní existuje F^{-1} .
Generuj $U \sim R(0, 1)$. Pak $F^{-1}(U)$ má rozdělení s distribuční funkcí F .

Poznámka: Normální rozdělení takhle nenagenerujeme :-).

(b) Transformační metoda - příklad (normální rozdělení)

Generujeme nezávisle náhodné veličiny $U_1 \sim R(0, 1)$ a $U_2 \sim R(0, 1)$. Potom náhodná veličiny

$$Z_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$$

$$Z_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

mají normované normální rozdělení.

► Simulace náhodného vektoru**(a) Transformací**

Náhodný vektor, jehož realizaci chceme nasimulovat, získáme transformací náhodného vektoru s rozdělením, ze kterého umíme generovat.

Příklad: Uvažujme náhodný vektor $\mathbb{Y} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Pokud umíme najít matici A takovou, že $\boldsymbol{\Sigma} = A \cdot A^T$, potom

$$\mathbb{Y} = \boldsymbol{\mu} + A \cdot \mathbb{X},$$

kde $\mathbb{X} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$.

(b) Interpretací

Viz. např. Wishartovo rozdělení:

$\mathbb{X}_1, \dots, \mathbb{X}_n$ výběr z $\mathbb{Y} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$, potom

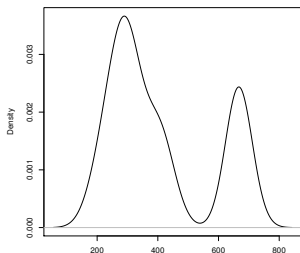
$$\sum_{i=1}^n (\mathbb{X}_i - \boldsymbol{\mu})(\mathbb{X}_i - \boldsymbol{\mu})^T$$

má rozdělení $W_d(\frac{n}{2}, \frac{n}{2}\Sigma^{-1})$.

Simulace z rozdělení - zamítací metoda

Uvažujme náhodnou veličinu X s hustotou f . Potom simulace z rozdělení s hustotou f je ekvivalentní simulaci vektoru (X, U) z rovnoměrného rozdělení na množině $A = \{(x, u) : 0 < u < f(x)\}$.

Postup: Generujeme z obdélníku, který "obalí" množinu A . Padne-li bod do této množiny, považujeme jeho x -ovou souřadnici za realizaci X , pokud ne, pokračujeme dále, dokud se tak nestane.



Simulace z rozdělení - zamítací metoda

Co když ale neznáme integrační konstantu hustoty, tj. místo f známe

$$f^* = c \cdot f,$$

kde konstantu c neznáme. Pomůžeme si hustotou g rozdělení, ze kterého umíme generovat a platí $f^* \leq Mg(x)$, $\forall x$, M známé.

Algoritmus:

1. generuj $X \sim g$,
2. generuj $U \sim R(0, 1)$ nezávisle na X ,
3. pokud $U \leq \frac{f^*(X)}{Mg(X)}$, pak $Z = X \sim f$. Jinak zpět na krok 1.

Poznámka: Očekávaný počet iterací je $\frac{M}{c}$.

Metody Monte Carlo

Třída algoritmů k simulaci systémů. Nejčastěji slouží k

- ▶ počítání (zejména) mnohorozměrných integrálů
- ▶ simulacím z daného rozdělení
- ▶ dále například k numerickému řešení diferenciálních rovnic

Monte Carlo integrace

Budeme chtít odhadnout střední hodnotu funkce h náhodné veličiny X s rozdělením daným hustotou f . Víme:

$$\mathbb{E}h(X) = \int_{\mathbb{R}} h(x)f(x)dx.$$

Odhad: Budeme-li mít náhodný výběr o rozsahu n z rozdělení veličiny X , potom

$$\mathbb{E}h(X) \approx \frac{1}{n} \sum_{i=1}^n h(x_i).$$

My ovšem výběr nemáme, jen známe rozdělení $X \Rightarrow$ výběr si můžeme nagerovat.

Problém nastává, když z f neumíme generovat.

Importance Sampling - výběr na základě důležitosti

Neumíme generovat přímo z rozdělení s hustotou f (nebo se nám to nehodí, např. je to zdlouhavé), ale umíme generovat z rozdělení s hustotou g .

Myšlenka:

$$\mathbb{E}h(X) = \int_{\mathbb{R}} h(x)f(x) \frac{g(x)}{g(x)} dx = \mathbb{E}h(X) = \int_{\mathbb{R}} \left(h(x) \frac{f(x)}{g(x)} \right) g(x) dx.$$

Stačí nám tedy generovat z rozdělení s hustotou g (tzv. **importance density**) a poté odhadovat $\mathbb{E}h(X)$ sumou

$$\frac{1}{n} \sum_{i=1}^n h(x_i) \frac{f(x_i)}{g(x_i)}.$$

Hustotu g volit s rozvahou!

Náhodný proces

Uvažujme množinu $T \subset \mathbb{R}$. **Náhodný proces** je množina $\{X_t, t \in T\}$, kde X_t jsou náhodné veličiny.

Index t většinou interpretujeme jako čas.

Pro $T = \mathbb{Z}$ nebo $T = \mathbb{N} \cup \{0\}$ mluvíme o náhodném procesu s diskrétním časem, popř. o **časové řadě**.

Markovský řetězec $\{X_n, n \in \mathbb{N}_0\}$ je takový náhodný proces s diskrétním časem, že

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n)$$

Metody Markov Chain Monte Carlo

Třída Monte Carlo algoritmů využívající Markovské vlastnosti řetězců.

Uvedeme si Metropolisův–Hastingsův algoritmus ke generování z cílového rozdělení s hustotou f .

Metropolisův–Hastingsův algoritmus

Algoritmus sloužící k simulaci z cílového rozdělení s hustotou f .

Chceme generovat z rozdělení s hustotou f , ale nemůžeme to, nebo je to z nějakého důvodu nevýhodné.

Místo toho generujeme z rozdělení s hustotou q , ze kterého již generovat umíme.

Algoritmus probíhá v $t = 1, \dots, T$ iteracích.

Metropolisův–Hastingsův algoritmus

ALGORITMUS:

1. Zvol počáteční stav x_0 libovolně. Polož $t = 0$.
2. Generuj y z rozdělení s hustotou $q(\cdot|x_t)$.
3. Spočítej tzv. **Hastingsův poměr** (pravděpodobnost přijetí návrhu y):

$$\alpha = \min \left\{ \frac{f(y)q(y|x_t)}{f(x_t)q(x_t|y)}, 1 \right\},$$

pokud $f(x_t)q(x_t|y) > 0$, jinak $\alpha = 1$.

4. S pravděpodobností α přijmi návrh y , tj. $x_{t+1} = y$.
S pravděpodobností $1 - \alpha$ polož $x_{t+1} = x_t$.
5. Pokud $t < T$ polož $t = t + 1$.

Použitá literatura

Zbyněk Pawlas: skripta k p'v redmětu MCMC.