

Monte Carlo metody

Jan Kracík

jan.kratick@vsb.cz

Součástí prakticky všech bayesovských metod je aposteriorní hustota

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}.$$

reprezentující celkovou informaci o parametru θ .

V úlohách se velice často vyskytují integrály obsahující (ne)normalizovanou aposteriorní hustotu.

- normalizace aposteriorní hustoty

$$\int f(x|\theta)f(\theta)d\theta$$

- marginalizace aposteriorní hustoty

$$f(\theta_1|x) = \int f(\theta_1, \theta_2|x)$$

- střední hodnoty (ztráty, momenty, aposteriorní pravděpodobnosti náhodných jevů)

$$\int g(\theta)f(\theta|x)d\theta$$

- predikce

$$f(x_{t+1}|x_1, \dots, x_t) = \int f(x_{t+1}|\theta)f(\theta|x_1, \dots, x_t)d\theta$$

- Až na výjimky nelze integrál $\int f(x|\theta)f(\theta)d\theta$ spočítat analyticky
(Př: omezené normální rozdělení).
- Pro diskrétní parametr nelze často sčítat ani hrubou silou
(Př: segmentace obrazu 40x40 - 10^{481} operací)
- Často nelze kvůli vysoké dimenzi využít běžné numerické metody jako Simpsonovo pravidlo apod.
(Př: Vícerozměrné normální rozdělení)

Řešením jsou simulační metody (Monte Carlo).

Cílem je (přibližný) výpočet integrálu

$$I(g) = E_f[g(X)] = \int g(x)f(x)dx. \quad (1)$$

Umíme-li generovat nezávislé vzorky $x^{(1)}, x^{(2)}, \dots$ z rozdělení s hustotou pravděpodobnosti $f(x)$, lze (1) aproximovat

$$I_N(g) = \frac{1}{N} \sum_{i=1}^N g(x^{(i)}).$$

Podle silného zákona velkých čísel platí

$$I_N(g) \xrightarrow{\text{a.s.}} I(g).$$

Krok strannou (konvergence skoro jistě)

Nechť X, X_1, X_2, \dots jsou náhodné veličiny na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Řekneme, že posloupnost X_1, X_2, \dots konverguje k náhodné veličině X *skoro jistě*, jestliže platí

$$P(\{\omega \in \Omega : \lim_n X_n(\omega) = X(\omega)\}) = 1.$$

Pro rozptyl aproximace střední hodnoty platí

$$D(I_N(g)) = \frac{1}{N} D_f(g(X)),$$

kde

$$D_f(g(X)) = \int (g(x) - E_f[g(X)])^2 f(x) dx.$$

Podle centrální limitní věty pak platí

$$\sqrt{N}(I_N(g) - I(g)) \xrightarrow{D} \mathcal{N}(0, D_f(g(X))).$$

Rozptyl $D_f(g(X))$ lze aproximovat podobně jako integrál $I(g)$:

$$d_N = \frac{1}{N} \sum_{i=1}^N (g(x^{(i)}) - I_N(g))^2.$$

Pro velký počet vzorků N má pak

$$\frac{I_N(g) - I(g)}{\sqrt{d_N}}$$

přibližně rozdělení $\mathcal{N}(0, 1)$.

Toho lze využít pro testy konvergence, intervaly spolehlivosti apod.

Podmínkou proveditelnosti všech MC metod je možnost generovat vzorky z daného rozdělení.

Základem všech algoritmů jsou pseudonáhodné generátory z rovnoměrného rozdělení. (Pozor na kvalitu generátoru!)

Tvrzení (Integrální transformace pravděpodobnosti)

Nechť F je libovolná distribuční funkce a F^{-} je pseudoinverze definovaná vztahem

$$F^{-}(u) = \inf\{x : F(x) \geq u\}.$$

Jestliže $U \sim \mathcal{U}_{(0,1)}$, pak $F^{-}(U)$ má distribuční funkci F .

Příklad (Diskrétní náhodná veličina)

Nechť pro $i = 1, \dots, n$ platí $P(X = i) = p_i$.

Vzorky náhodné veličiny X lze generovat podle následujícího algoritmu:

- Označíme

$$\pi_0 = 0, \quad \pi_i = \sum_{k=1}^i p_k.$$

- Vygenerujeme

$$U \sim U_{\langle 0,1 \rangle}.$$

- Zvolíme

$$X = k \Leftrightarrow \pi_{k-1} < U \leq \pi_k.$$

Obecně je integrální transformace málo efektivní.

Pro generování vzorků z konkrétních rozdělení se využívají transformace náhodných veličin.

Problém

Jak generovat vzorky z libovolného rozdělení?

Výchozí myšlenka:

$$f(x) = \int_0^{f(x)} du$$

Hustotu $f(x)$ tak můžeme chápat jako marginální hustotu ze sdružené hustoty náhodného vektoru (X, U) , kde

$$(X, U) \sim \mathcal{U}(\{(x, u) : 0 < u < f(x)\})$$

Tvrzení (Základní věta simulace)

Simulace $X \sim f(x)$ je ekvivalentní simulaci

$$(X, U) \sim \mathcal{U}(\{(x, u) : 0 < u < f(x)\})$$

Jak generovat $(X, U) \sim \mathcal{U}(\{(x, u) : 0 < u < f(x)\})$?

Po částech $(X, U|X)$ nemá smysl. Snadnější je generovat (X, U) najednou na větší oblasti.

Předpokládejme, že

$$\int_a^b f(x) dx = 1,$$

$$f(x) \leq m \text{ na } \langle a, b \rangle$$

pro nějaké $m \in \mathbb{R}$.

Generujeme

$$(Y, U) \sim \mathcal{U}(\langle a, b \rangle \times \langle 0, m \rangle),$$

tj.

$$Y \sim \mathcal{U}(\langle a, b \rangle), U|Y \sim \mathcal{U}(\langle 0, m \rangle).$$

Vzorek (y, u) považujeme za vzorek $(X, U) \sim \mathcal{U}(\{(x, u) : 0 < u < f(x)\})$, jestliže

$$u < f(x).$$

V opačném případě vzorek zahazujeme.

Pro vygenerované vzorky platí

$$\begin{aligned}P(X \leq x) &= P(Y \leq x | U \leq f(Y)) \\&= \frac{\int_a^x \int_0^{f(y)} du dy}{\int_a^b \int_0^{f(y)} du dy} \\&= \int_a^x f(y) dy,\end{aligned}$$

tj. náhodná veličina X má hustotu pravděpodobnosti f .

Pro generování vzorků X s hustotou f stačí umět hustotu f vyčíslit v libovolném bodě.

Zamítací metoda

Metodu lze snadno zobecnit pro obecnější oblast. Podmínkou je umět simulovat z rovnoměrného rozdělení na oblasti.

Předpokládejme, že hustotu $f(x)$ lze shora omezit integrovatelnou funkcí $m(x)$, tj. $f(x) \leq m(x)$.

Označme

$$M = \int m(x) dx.$$

Pak existuje hustota pravděpodobnosti $g(x)$ taková, že

$$m(x) = Mg(x).$$

Zamítací metoda

Vektor $(Y, U) \sim \mathcal{U}((y, u) : 0 < u \leq m(y))$ lze simulovat po částech:

$$Y \sim g(y)$$
$$U|Y = y \sim \mathcal{U}(0, Mg(y)).$$

Opět přijímáme pouze vzorky, pro které platí $u < f(y)$.

Podmínkou pro praktickou použitelnost je schopnost generovat vzorky z $g(y)$.

$$\begin{aligned}P(X \in A) &= P(Y \in A | U < f(Y)) \\&= \frac{\int_A \int_0^{f(y)} \frac{1}{Mg(y)} du g(y) dy}{\int_{\mathcal{X}} \int_0^{f(y)} \frac{1}{Mg(y)} du g(y) dy} \\&= \int_A f(y) dy.\end{aligned}$$

Nechť $X \sim f(x)$ a pro hustotu $g(x)$ platí $f(x) \leq Mg(x)$ pro nějaké $M \geq 1$. Simulaci $X \sim f$ lze provést následovně:
Generujeme

$$Y \sim g, \quad U|Y = y \sim \mathcal{U}(0, Mg(y))$$

dokud

$$0 < u < f(y).$$

Pro toto y položíme $x = y$.

Poznámky

- Pro zamítací metodu stačí znát hustotu $f(x)$ až na normalizační konstantu. Výhoda pro simulaci z aposteriorní hustoty:

$$f(\theta) \propto f(x|\theta)f(\theta).$$

- Pro normalizovanou hustotu $f(x)$ je pravděpodobnost přijetí vzorku $\frac{1}{M}$. Odtud se odvíjí efektivita algoritmu.
- Zůstává problém nalezení vhodného $g(x)$ a M .
- S rostoucí dimenzí veličiny X zpravidla rychle klesá efektivita (pravděpodobnost přijetí vzorku).

Algoritmus

Vstup: $f(x)$, $g(x)$, M

- 1 Generuj $Y \sim g$.
- 2 Generuj $U \sim \mathcal{U}(\langle 0, 1 \rangle)$.
- 3 Jestliže $U \leq \frac{f(Y)}{Mg(Y)}$, polož $X = Y$. Jinak jdi na. 1

Výstup: X

Poznámky

- Pro aposteriorní hustoty je počet operací potřebných k vyhodnocení

$$f(x_1, \dots, x_t | \theta) f(\theta)$$

úměrný počtu dat t . Lze urychlit vhodným dolním odhadem.