Моделирование методом отбора (Rejection Sampling)

422 группа

Tue 07 Mar 2017

Содержание

1 Схема метода

1.1 Мотивация

Пусть $p(x) = \mathrm{pdf}_{\mathcal{P}}(x)$, H — подграфик.

Предложение. *Если* $(\xi_1, \xi_2) \sim \mathrm{U}(H)$, *то* $\xi_1 \sim \mathcal{P}$.

Доказательство. Действительно,

$$\mathsf{P}(\xi_1 < z) = \int_{-\infty}^{z} \mathrm{d}x \int_0^{p(x)} \mathrm{d}y = \int_{-\infty}^{z} p(x)\,\mathrm{d}x.$$

□

Алгоритм. *Пусть* $g(x)$ *— мажоранта,* $|\mathrm{supp}\,\mathcal{P}| < \infty$.

1. $x \leftarrow \mathrm{U}(\mathrm{supp}\,\mathcal{P})$, $y \leftarrow \mathrm{U}(0, g(x))$
2. *Если* $y < p(x)$, $\xi \leftarrow x$, *конец; иначе 1.*

Алгоритм (Частный случай)**.** *Пусть* $p(x)$ *сосредоточена на* $[0, 1]$, $g(x) \equiv M$.

1. *Для* $t \in 1, 2, \ldots$

$$\begin{aligned} \alpha_1^{(t)} &\sim \mathrm{U}(0,1) \\ \alpha_2^{(t)} &\sim \mathrm{U}(0,1). \end{aligned}$$

2. *Если*

$$p(\alpha_1^{(t)}) \geq M\alpha_2^{(t)},$$

$\xi \leftarrow \alpha_1^{(t)}$, *конец; иначе 1.*

1.2 Формальный алгоритм

Пусть $\mathcal{Q}$ — мажорирующее распределение, т.е. $p(x) \le Mq(x) = g(x)\ \forall x$ и существует производная Радона–Никодима $r = \mathrm{d}\mathcal{P}/\,\mathrm{d}\mathcal{Q} \le M$.

Алгоритм (Общий)**.**

1. *Для* $t \in 1, 2, \dots$

$$\begin{array}{rcl} \eta^{(t)} & \sim & \mathcal{Q} \\ \alpha^{(t)} & \sim & \mathrm{U}(0,1). \end{array}$$

2. *Если*

$$r(\eta^{(t)}) \ge M\alpha^{(t)},$$

или (поскольку $r = p/q$*) что то же самое,*

$$\frac{p(\eta^{(t)})}{g(\eta^{(t)})} \ge \alpha^{(t)},$$

то $\tau = t$*;* $\xi \leftarrow \eta^{(\tau)}$*; конец; иначе 1.*

Утверждение. Справедливо:

1. $\tau - 1 \sim \mathrm{Geom}(1/M)$;
2. $\xi := \eta_\tau \sim \mathcal{P}$;
3. $\tau \perp\!\!\!\perp \eta_\tau$.

Трудоемкость По смыслу, $p = \mathrm{pdf}_{\mathcal{P}}$ сложная в вычислении[1], $q = \mathrm{pdf}_{\mathcal{Q}}$ простая. Чтобы получить одну реализацию нужной случайной величины η_τ, нужно τ раз вычислить r; в среднем это, поскольку $\tau - 1 \sim \mathrm{Geom}(1/M)$,

$$\mathsf{E}(\tau - 1) = \mathsf{E}(\tau) - 1 = \frac{1 - 1/M}{1/M} = M - 1 \implies \mathsf{E}(\tau) = M.$$

Но M — то, насколько одна плотность подогнана под другую ($M = p/q = 1$ соответствует с случаю $p = q$).

Так что важно правильно выбрать мажорирующее распределение, ведь когда отвергается величина $r(\eta^{(t)})$ процессорные такты, потраченные на её вычисление, пропадают.

1.3 Трюки

Следующие трюки позволяют выбирать мажорирующую плотность так, что $M \to 1$.

Rejection with squeezing Пусть умеем строить не только мажорнату g_u, но и миноранту g_ℓ. Для отбора нужно брать только те, что ниже p; значит, сначала вычислим миноранту, и если значение ниже нее, то принимать значение.

Adaptive Rejection Sampling (ARS) См. далее.

[1] Например, в случае условного распределения — тогда это интеграл $h(x) = \int \ell(x, y)\,\mathrm{d}y$.

2 Адаптивный метод отбора

(Рис. 1) Нужно, чтобы

1. p была непрерывно дифференцируема на связном множестве D.
2. $h(x) = \log p(x)$ была бы выпукла.

Большая часть стандартных плотностей этим требованиям удовлетворяют; интегралы тоже, потому что неравенство Єнсена.

2.1 Мажоранта

Мажоранту можно построить кусочно-линейную — по касательным. Пусть выбрано разбиение $x_1 \le \cdots \le x_k$. Пусть также $z_1, \ldots, z_{k-1}$ — точки пересечения касательных:

$$\begin{aligned}
z_j &= \frac{h(x_{j+1}) - h(x_j) - x_{j+1}h'(x_{j+1}) + x_j h'(x_j)}{h'(x_j) - h'(x_{j+1})}, \quad j \in 1:(k-1) \\
z_0 &= -\infty \text{ или } \inf D \\
z_k &= +\infty \text{ или } \sup D.
\end{aligned}$$

Тогда можно задать кусочно-линейную мажоранту

$$u(x) = h(x_j) + h'(x_j)(x - x_j), \quad x \in (z_{j-1}, z_j],\ j \in 1:k.$$

Потенциировав, получают $\exp u(x)$ — мажоранту исходного распределения $\mathcal{Q}$, каковая мажоранта задает кусочно-экспоненциальную плотность

$$s(x) = \frac{\exp u(x)}{C}, \quad C = \int_D \exp u(w)\,\mathrm{d}w,$$

где

$$\begin{aligned}
\int_D \exp u(w)\,\mathrm{d}w &= \sum_{j=1}^{k} \int_{z_{j-1}}^{z_j} \exp\left\{h(x_j) + h'(x_j)(w - x_j)\right\}\,\mathrm{d}w \\
&= \sum_{j=1}^{k} \exp\left\{h(x_j) - h'(x_j)x_j\right\} \int_{z_{j-1}}^{z_j} \exp\left\{h'(x_j)w\right\}\,\mathrm{d}w \\
&= \sum_{j=1}^{k} \frac{1}{h'(x_j)} \exp\left\{h(x_j) - h'(x_j)x_j\right\} \left(\exp\left\{h'(x_j)z_j\right\} - \exp\left\{h'(x_j)z_{j-1}\right\}\right) \\
&= \sum_{j=1}^{k} \frac{1}{h'(x_j)} \left(\exp\left\{h(x_j) + h'(x_j)(z_j - x_j)\right\} - \exp\left\{h(x_j) + h'(x_j)(z_{j-1} - x_j)\right\}\right) \\
&= \sum_{j=1}^{k} \frac{1}{h'(x_j)} \left(\exp u(z_j) - \exp u(z_{j-1})\right).
\end{aligned}$$

2.1.1 Моделирование кусочно-экспоненциальной плотности

Функция распределения кусочно-экспоненциальной плотности $\mathcal{Q}$ есть

$$F_{\mathcal{Q}}(z) = \int_{-\infty}^{z} s(x)\,\mathrm{d}x = \frac{1}{C}\left(\sum_{j=1}^{i:z_i \ge z} \frac{1}{h'(x_j)} \left(\exp u(z_j) - \exp u(z_{j-1})\right) + \exp u(z) - \exp u(z_i)\right).$$

Алгоритм (Моделирование $\eta \sim \mathcal{Q}$)**.**

1. $\alpha \sim \mathrm{U}(0,1)$

2. $z_i \leftarrow \max\{z_j : F_{\mathcal{Q}}(z_j) < \alpha\}$

3.
$$\eta \leftarrow z_i + \frac{1}{h'(x_{i+1})} \log\left[1 + \frac{h'(x_{i+1})C(\alpha - F_{\mathcal{Q}}(z_i))}{\exp u(z_i)}\right].$$

2.2 Миноранта

Миноранта — из хорд составленная. Тогда для $x \in [x_j, x_{j+1}]$ миноранта задается как
$$\ell(x) = \frac{(x_{j+1} - x)h(x_j) + (x - x_j)h(x_{j+1})}{x_{j+1} - x_j}, \quad j \in 1 : (k-1).$$

Для $x < x_1$ или $x > x_k$ $\ell(x) := -\infty$.
Утверждение. Раз h выпукла, то автоматически $e^{\ell(x)} \le p(x) \le e^{u(x)}$.

2.3 Алгоритм

Алгоритм. *Таким образом,*

1. *Инициализация:*
   a) *Выберем* $T_k = \{x_1, \ldots, x_k\}$ *— количество начальных точек.*
   Замечание. Нужно выбрать по крайней мере по разным сторонам от моды — $h'(x_1) > 0$, $h'(x_k) < 0$, чтобы можно было построить мажоранту (минимум две; три с третьей в моде уже гораздо лучше).
   b) *Посчитаем* $h(x_1), \ldots, h(x_k)$, $h'(x_1), \ldots, h'(x_k)$, $z_1, \ldots, z_{k-1}$. *Посчитаем* u_j, ℓ_j, s_j.
2. *Моделирование (отбор):*
   a) $\eta \leftarrow \mathcal{Q}$, $\alpha \leftarrow \mathrm{U}[0,1]$.
   b) *если*
   $$\alpha \le \exp\{\ell(\eta) - u(\eta)\},$$
   то $\xi \leftarrow \eta$*; конец.*
   c) *если*
   $$\alpha \le \exp\{h(\eta) - u(\eta)\},$$
   то $\xi \leftarrow \eta$*; конец.*
3. *Хочется переиспользовать вычисленное значение* h *(потому что она дорогая) — вычислим производную* $h' \leftarrow h'(\eta)$, $T_{k+1} \leftarrow T_k \cup \{\eta\}$ *и переупорядочим* T_{k+1}. *По полученному можно пересчитать* $u_{j+1}, \ell_{j+1}, s_{j+1}$.

   Замечание. Если дошли до этого шага, значит находимся с большой вероятностью в месте большого зазора мажоранты и h; значит добавление точки в данном месте очень хорошо этот зазор уменьшает.

Данный алгоритм хорошо аппроксимирует исходную плотностью кусочно-экспоненциальным распределением, так что $M \to 1$.

Замечание. Если функция оказывается выпуклой после монотонного преобразования, то плотность будет не кусочно-экспоненциальная, а кусочно-какая-то.

Замечание. Если плотность вогнутая, то поменяются хорды и касательные (плюс нужно будет разбираться с асимптотами).

Замечание. Любую хорошую функцию можно разложить на выпуклую и вогнутую части. Тогда можно посчитать как смесь, для чего необходимо почти всегда знание интегралов компонентов смести.

2.4 ARS с ушами

Пусть не умеем считать производную. Тогда мажоранту можно делать из ушей продолжением хорд (Рис. 2). Пусть $L_{ij}(x)$ — прямая, соединяющая $(x_i, h(x_i))$ и $(x_j, h(x_j))$. Тогда на $x_i \le x \le x_{i+1}$, $u_k(x) = \min(L_{i-1,i}(x), L_{i+1,i+2}(x))$. Крайние интервалы нужно, понятно, обрабатывать отдельно.

Adaptive Rejection Metropolis Sampling Пусть есть функция с особенностью — перегибом. Тогда в окрестности перегиба значение мажоранты будет меньше значения функции. Стандартный трюк — отказаться от независимости и в качестве результата выдавать предыдущий результат.