Введение в Markov chain Monte Carlo

Рассмотрим выборку $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ из параметрического семейства распределений $\{F_\theta, \theta \in \Theta\}$. Будем предполагать, что задано некоторое априорное распределение $\pi(\theta)$ параметра θ. Один из способов построения байесовских оценок параметра θ состоит в вычислении среднего значения $\mathbf{E}\,(\theta|\boldsymbol{X})$ для апостериорного распределения параметра θ:

$$\pi\,(\theta|\boldsymbol{X}) = \frac{L(\boldsymbol{X};\theta)\pi(\theta)}{\int_\Theta L(\boldsymbol{X};\theta)\pi(\theta)d\theta},$$

которое считается известным с точностью до константы в знаменателе.

Мы выделили две основные задачи MCMC методов:

(i) Сгенерировать выборку из распределения $\pi(\theta|\boldsymbol{X})$, известного с точностью до константы.

(ii) Используя методы Монте Карло, численно оценить апостериорное среднее значение $\mathbf{E}\,(\theta|\boldsymbol{X})$.

Семинар был посвящен разбору двух классических MCMC методов: метода Метрополиса — Гастингса и метода сэмплирования по Гиббсу.

Алгоритм Метрополиса — Гастингса. Предположим, что $\pi(\theta|\boldsymbol{X}) \propto f(\theta)$, причем $f(\theta)$ — полностью известная функция. Рассмотрим также некоторую вспомогательную цепь Маркова с переходным ядром $g(\theta_2|\theta_1)$. Мы будем считать, что при любом $\theta_1 \in \Theta$ существуют реализуемые алгоритмы сэмплирования из условного распределения $g(\,\cdot\,|\theta_1)$.

Алгоритм 1 (Метрополиса — Гастингса)**.**

1. *Выбираем произвольное* $\theta_0 \in \Theta$.

2. *Для* $t \geq 1$, *генерируем* Y_t *согласно переходной плотности* $g(\,\cdot\,|\theta_{t-1})$. *Вычисляем «коэффициент принятия»*

$$R(Y_t|\theta_{t-1}) = \frac{f(Y_t)}{f(\theta_{t-1})} \times \frac{g(\theta_{t-1}|Y_t)}{g(Y_t|\theta_{t-1})}.$$

Генерируем случайную величину U_t *из равномерного на отрезке* $[0,1]$ *распределения. Определяем* θ_t *по правилу:*

$$\theta_t = \begin{cases} Y_t, & \text{если } R(Y_t|\theta_{t-1}) > U_t, \\ \theta_{t-1}, & \text{иначе.} \end{cases}$$

Мы показали, что результатом работы алгоритма Метрополиса — Гастингса является цепь Маркова θ_t с переходным ядром

$$g^*(\theta_2|\theta_1) = g(\theta_2|\theta_1) \cdot \min\{1, R(\theta_2|\theta_1)\} + \delta_{\theta_1}(\theta_2) \cdot \widetilde{R}(\theta_1),$$

где $\widetilde{R}(\theta_1) = \int_\Theta g(\theta|\theta_1)(1 - \min\{1, R(\theta|\theta_1)\})d\theta$, и $\delta_{\theta_1}(\theta)$ — дельта-функция Дирака.

Используя это представление, мы показали, что цепь $\{\theta_t\}$ удовлетворяет уравнению детального баланса:

$$\pi(\theta_1|\boldsymbol{X})g^*(\theta_2|\theta_1) = \pi(\theta_2|\boldsymbol{X})g^*(\theta_1|\theta_2) \text{ при всех } \theta_1, \theta_2 \in \Theta, \tag{1}$$

что влечет за собой тот факт, что $\pi(\theta|\boldsymbol{X})$ является стационарным распределением цепи $\{\theta_t\}$. Сходимость цепи к стационарному распределению гарантирует следующая теорема (которую мы привели без доказательства):

Теорема 1. *Пусть*

$$\mathbf{P}\left(R(Y_{t+1}|Z_t) \geq 1\right) < 1 \qquad \text{и} \qquad g(\theta_2|\theta_1) > 0 \text{ при всех } (\theta_1, \theta_2) \in \Theta^2.$$

Тогда

(i) Для любой функции h такой, что $\mathbf{E}\left(h(\theta)|\boldsymbol{X}\right) < \infty$,

$$\frac{1}{T}\sum_{t=1}^{T} h(\theta_t) \to \mathbb{E}\left(h(\theta)|\boldsymbol{X}\right) \text{ п.н. при } T \to \infty.$$

(ii) Для любых начальных данных $\theta_0 \in \Theta$,

$$d_{TV}\left(g^{*T}(\,\cdot\,|\theta_0) - \pi(\,\cdot\,|\boldsymbol{X})\right) \to 0 \text{ при } T \to \infty.$$

Отметим, что первое утверждение теоремы обеспечивает сходимость выборочного среднего, построенного по $\{\theta_t\}$, к апостериорному среднему $\mathbf{E}\left(\theta|\boldsymbol{X}\right)$.

Алгоритм сэмплирования по Гиббсу. Предположим, что апостериорная плотность $\pi(\theta, \varphi|\boldsymbol{X}) \propto f(\theta, \varphi)$, где $(\theta, \varphi) \in \Theta \times \Phi$, причем функция $f(\theta, \varphi)$ такова, что полностью известны условные плотности

$$g_1(\theta|\varphi) = \frac{f(\theta, \varphi)}{f_2(\varphi)} \qquad \text{и} \qquad g_2(\varphi|\theta) = \frac{f(\theta, \varphi)}{f_1(\theta)}.$$

Здесь,

$$f_1(\theta) = \int_\Phi f(\theta, \varphi)d\varphi \qquad \text{и} \qquad f_2(\varphi) = \int_\Theta f(\theta, \varphi)d\theta.$$

По аналогии с алгоритмом Метрополиса — Гастингса, мы будем предполагать, что для $g_1(\theta|\varphi)$ и $g_2(\varphi|\theta)$ реализуемы какие-либо алгоритмы сэмплирования.

Алгоритм 2 (Сэмплирования по Гиббсу)**.**

1. *Выбираем произвольное* $\theta_0 \in \Theta$. *Генерируем* φ_0 *согласно условной плотности* $g_2(\,\cdot\,|\theta_0)$.

2. *Для* $t \geq 1$, *генерируем* θ_t *согласно условной плотности* $g_1(\,\cdot\,|\varphi_{t-1})$ *и* φ_t *согласно условной плотности* $g_2(\,\cdot\,|\theta_t)$. *Добавляем пару* (θ_t, φ_t) *в выборку.*

На семинаре мы рассмотрели пример использования алгоритма сэмплирования по Гиббсу для моделирования двумерного нормального вектора (θ, φ) с нулевым вектором средних и ковариационной матрицей

$$\Sigma = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix},$$

где $r \in (-1, 1)$.

Наконец, мы обсудили некоторые классические способы уменьшения погрешности оценок, построенных с использованием MCMC методов: использование интервала «разогрева», «просеивание» получающейся цепи Маркова и т.д.