

Метод Метрополиса (добавление к главе 2 вводного курса “Метод Монте-Карло в вычислительной математике”)

С.М. Ермаков

Одним из наиболее употребительных методов моделирования случайных величин с заданным законом распределения является метод отбора (мажорант). Предполагается, что имеется программный датчик (псевдо) случайных чисел, равномерно распределенных на промежутке $(0,1)$.

Рассмотрим методы моделирования случайного вектора $(\xi_1, \dots, \xi_s) = \vec{\xi}$, заданного плотностью совместного распределения компонент

$$\varphi(x_1, \dots, x_s) = \varphi(X).$$

Метод отбора легко обосновать с помощью известных формул для условных вероятностей распределения компонент вектора. Пусть $\psi(X)$ – плотность. Рассмотрим в $\mathbb{R}^{s+1}$ область

$$D_{s+1} = \left\{ (X, x_{s+1}) : \begin{array}{ll} 0 \leq x_{s+1} \leq C\psi(X), & C > 0 \text{ (константа)}, \\ 0, & x_{s+1} < 0, x_{s+1} > C\psi(X) \end{array} \right\}.$$

В D_{s+1} определим плотность (равномерного распределения)

$$\lambda(X, x_{s+1}) = \begin{cases} 1/C, & (X, x_{s+1}) \in D_{s+1}, \\ 0 & \text{в противном случае.} \end{cases}$$

Теперь, если $(\xi_1, \dots, \xi_{s+1})$ – вектор с плотностью λ , то маргинальная плотность распределения $\Xi = (\xi_1, \dots, \xi_{s+1})$ есть

$$\int \lambda(X, x_{s+1}) dx_{s+1} = \int_0^{C\psi(X)} 1/C dt = \psi(X)$$

и, если получена реализация $\tilde{\Xi}$ вектора Ξ , то плотность условного распределения ξ_{s+1} при условии $\Xi = \tilde{\Xi}$ есть

$$\frac{\lambda(\tilde{\Xi}, x_{s+1})}{\int \lambda(\tilde{\Xi}, x_{s+1}) dx_{s+1}} = \frac{1/C}{\int_0^{C\psi(\tilde{\Xi})} 1/C dt} = \frac{1}{C\psi(\tilde{\Xi})},$$

т.е. ξ_{s+1} равномерно распределен на отрезке $[0, C\psi(\tilde{\Xi})]$.

Таким образом, справедлива следующая

Лемма 0.1 Если D_{s+1} – область в $\mathbb{R}^{s+1}$,

$$D_{s+1} = \left\{ \begin{array}{ll} (X, x_{s+1}) : & 0 \leq x_{s+1} \leq C\psi(X), \quad C > 0 \text{ (константа)}, \\ & 0, \quad x_{s+1} < 0, x_{s+1} > C\psi(X), \quad X = (x_1, \dots, x_{s+1}) \end{array} \right\},$$

$\psi(X)$ является плотностью и вектор (Ξ, ξ_{s+1}) равномерно распределен в D_{s+1} , то $\Xi = (\xi_1, \dots, \xi_{s+1})$ распределен с плотностью $\psi(X)$, а ξ_{s+1} – случайная величина, равномерно распределенная на промежутке $[0, C\psi(\Xi)]$.

Имеет место обратное утверждение.

Если Ξ – случайный вектор, распределенный с плотностью $\psi(X)$, а ξ_{s+1} при каждом Ξ распределен равномерно на промежутке $[0, C\psi(\Xi)]$, то вектор (Ξ, ξ_{s+1}) равномерно распределен в D_{s+1} .

Лемма служит основанием для алгоритмов метода отбора, которые требуют использования следующей, почти очевидной леммы.

Лемма 0.2 Пусть G и G_1 , $G_1 \subset G$ – области в $\mathbb{R}^{s+1}$ с объемом конечным и отличным от нуля. Если Ξ_{s+1} – случайный вектор, равномерно распределенный в G , то его условное распределение при условии $\Xi_{s+1} \in G_1$ является равномерным в G_1 .

Простейший метод отбора (метод Неймана) относится к случаю, когда моделируемая плотность $\varphi(X)$ ограничена и сосредоточена на ограниченном множестве в $\mathbb{R}^s$. В этом случае существует $(s+1)$ -мерный гиперпараллелепипед Π_{s+1} такой, что $D_{s+1} \subset \Pi_{s+1}$. Моделирование вектора Ξ_{s+1} , равномерно распределенного в Π_{s+1} , не представляет труда, и алгоритм состоит в проверке принадлежности Ξ_{s+1} области D_s . Это частный случай метода мажорант, когда мажоранта равна константе. В общем случае, если моделируется плотность $\varphi(X)$ и есть другая плотность $\psi(X)$, обладающая следующими свойствами:

1. Алгоритм моделирования $\psi(X)$ существенно проще, чем алгоритм моделирования $\varphi(X)$.
2. Существует константа C такая, что $\varphi(X) \leq C\psi(X)$, то с помощью описанного выше алгоритма моделируется равномерное распределение в

$$D_{s+1}(C, \psi) = \{(X, x_{s+1}) : 0 \leq x_{s+1} \leq C\psi(X)\}$$

и отбираются реализации, попадающие в $D_{s+1}(\varphi)$ (остальные отбрасываются). Эффективность алгоритма равна $1/C$. Далее этот алгоритм подробно описан.

Если $X \in \mathbb{R}^s$, $\varphi(X)$ – плотность, и мы хотим получить реализации случайного вектора $\vec{\xi}$, распределенного с этой плотностью, то мы можем поступать следующим образом. Пусть $\phi(X)$ – плотность, алгоритм моделирования которой достаточно прост, и существует константа C такая, что для всех X выполняется неравенство

$$\varphi(X) \leq C\phi(X), \tag{1}$$

тогда алгоритм реализации $\vec{\xi}$ состоит в следующем:

1. Получаем реализацию $\vec{\eta}$ случайной величины с плотностью $\phi(X)$.

2. Получаем реализацию α случайной величины, равномерно распределенной на $[0, 1]$.
3. Проверяем неравенство $\alpha C\phi(\eta) \leq \varphi(\eta)$ (или $\alpha \frac{C\phi(\eta)}{\varphi(\eta)} \leq 1$).
4. Если неравенство 1 выполнено, полагаем $\vec{\xi} = \vec{\eta}$, в противном случае повторяем процедуру (переход к пункту 1).

Далее мы рассмотрим гипотетическую ситуацию, которая приблизит нас к пониманию алгоритма Метрополиса. Предположим ($s > 1$), что мы ошиблись в выборе константы C , то есть выбрали ее слишком большой, тогда описанный выше алгоритм будет давать правильные результаты, но эффективность его может быть недопустимо низкой. Если же мы ошиблись в меньшую сторону, то это означает, что $\mathbb{R}^s$ разбито на две части:

$$\mathbb{R}^s = \mathfrak{X} \cup \mathcal{Y}, \quad \mathfrak{X} = \{X : \varphi(X) \leq C\phi(X)\}, \quad \mathcal{Y} = \{X : \varphi(X) > C\phi(X)\}.$$

Как и ранее получим $\vec{\eta}$ – реализацию $\phi(X)$. Возможны два случая:

- $\vec{\eta} \in \mathfrak{X}$ (это устанавливается проверкой неравенства $\varphi(\vec{\eta}) \leq C\phi(\vec{\eta})$). Тогда мы можем поступать также, как и в случае правильного C , то есть выполнять пункты 2 и 3. Однако, мы не можем положить $\vec{\xi} = \vec{\eta}$, это было бы неверно.
- $\vec{\eta} \in \mathcal{Y}$, проверка показывает нам, что $C\phi(X)$ не является мажорантой. Можно произвести коррекцию, полагая, например, $\frac{\phi(\eta)}{\varphi(\eta)}$ новой константой C .

Возможен следующий динамический алгоритм:

1. Выбираем C_0 .
2. Получаем $\vec{\eta}$
3. Если $\vec{\eta} \in \mathcal{Y}$, то производим коррекцию $C_t = \frac{\phi(\eta)}{\varphi(\eta)}$ и переходим к пункту 2; если $\vec{\eta} \in \mathfrak{X}$, то проверяем неравенство

$$\alpha C_t \phi(\vec{\eta}) \leq \varphi(\vec{\eta}). \quad (2)$$

4. Если неравенство 2 выполнено, то полагаем $\vec{\eta}_t = \vec{\eta}$.
5. Переходим к пункту 2, не изменяя C и полагая $t = t + 1$.
6. Заметим, что в этом случае будет проверяться неравенство

$$\alpha \frac{\varphi(\vec{\eta}_t)\phi(\vec{\eta}_{t+1})}{\phi(\vec{\eta}_t)\varphi(\vec{\eta}_{t+1})} \leq 1. \quad (3)$$

Можно доказать, что распределение $\vec{\eta}_t$ сходится к распределению $\vec{\xi}$ при $t \rightarrow \infty$. Процесс марковский.

Этот динамический алгоритм коррекции константы C в неравенстве 1 обладает тем недостатком, что форма функции ϕ остается неизменной. Известный [1] алгоритм Метрополиса и его обобщение – алгоритм Метрополиса-Хастингса несколько исправляют положение.

Алгоритм Метрополиса-Хастингса (М-Х алгоритм) предполагает функцию ϕ , зависящей от двух переменных переходной плотностью марковского процесса. Вместо независимых реализаций случайной величины с плотностью ϕ осуществляется моделирование цепи Маркова с переходной плотностью $\phi(X, Y)$.

Формально алгоритм описывается следующим образом. Для моделирования плотности $\varphi(X)$, $X \in \mathbb{R}^s$

1. Выбираем (произвольно) X_0 и переходную плотность $\phi(X, Y)$. Если известно X_t ($t = 0, 1, \dots$), то
2. Моделируем плотность $\phi(X_t, X)$. Получаем нужную реализацию $X = X'$.
3. Вычисляем величину $a = \frac{\varphi(X')\phi(X_t, X')}{\varphi(X_t)\phi(X', X_t)}$ (сравните с 3).
4. Если $a \geq 1$, то полагаем $X_{t+1} = X'$.
5. Если $a < 1$, то с вероятностью a $X_{t+1} = X'$ и с вероятностью $1 - a$ $X_{t+1} = X_t$ (отсутствие изменений).

Первоначально предложенный Метрополисом алгоритм предполагал цепь Маркова обратимой, то есть

$$\phi(X, Y) = \phi(Y, X).$$

В этом случае $a = \varphi(X')/\varphi(X_t)$. Полученный в результате случайный процесс X_t также является цепью Маркова с переходной плотностью

$$A(X, Y) = \phi(X, Y) \cdot \min\left(1, \frac{\varphi(Y)\phi(Y, X)}{\varphi(X)\phi(X, Y)}\right) \quad \text{при } X \neq Y. \quad (4)$$

Если эта цепь является эргодической, что чаще всего бывает в прикладных задачах, то ее стационарным распределением будет распределение с плотностью $\varphi(X)$. Это означает, что при достаточно больших $t = N$, $X_N, X_{N+1}, \dots$ будут независимыми реализациями случайной величины с плотностью $\varphi(X)$.

Обоснование М-Х метода связано со свойствами обратимых цепей Маркова. Стационарность плотности $\varphi(X)$ может, в частности, следовать из более общего соотношения – уравнения детального баланса.

Если для переходной плотности $p(X, Y)$ будет выполнено равенство

$$p(X, Y)\varphi(Y) = p(Y, X)\varphi(X) \quad \text{для всех } X \text{ и } Y,$$

то после интегрирования по Y получим

$$\int p(X, Y)\varphi(Y) dY = \varphi(X), \quad \text{поскольку} \quad \int p(Y, X) dY = 1.$$

Из равенства 4 следует

$$\phi(X)A(X, Y) = \varphi(X)\phi(X, Y) \cdot \min\left(1, \frac{\varphi(Y)\phi(Y, X)}{\varphi(X)\phi(X, Y)}\right) = \min(\varphi(X)\phi(X, Y), \varphi(Y)\phi(Y, X)).$$

Полученная функция симметрична относительно X и Y , что и обозначает выполнение условия детального баланса

$$\phi(X)A(X, Y) = \varphi(\psi)A(Y, X).$$

В литературе можно найти ряд обобщений М-Х алгоритма.

М-Х алгоритм обладает важными преимуществами. Он универсален и не требует знания константы нормировки. В случае большого s прямой выбор эффективной мажоранты может быть трудной и практически неразрешимой задачей, и М-Х алгоритм является единственным средством моделирования плотности.

Многочисленные применения М-Х алгоритма описаны в литературе (например, [2],[3]). Это, прежде всего, задачи статистической физики (моделирование равновесных систем и др.), задачи восстановления искаженных изображений, задачи дискретной оптимизации.

Методы поиска глобального экстремума, связанные с М-Х алгоритмом (имитации отжига) обычно излагаются в терминах статистической физики – гибсовские распределения, модели Изинга и др.

Напомним одну из процедур поиска глобального экстремума, не связанную с физическими представлениями ([4], стр.252).

Легко показать, что для любой ограниченной в некоторой области $D \subset \mathbb{R}^s$ функции $f(X)$ предел

$$\psi_n(X) = \frac{f^n(X)}{\int_D f^n(X) dX} \quad \text{при } n \rightarrow \infty.$$

есть δ мера, равномерно распределенная на множестве S

$$S = \{X : f(X) = \sup_X f(X)\}.$$

Обычные, требующие вычисления $\int_D f^n(X) dX$ методы моделирования, очевидно, трудно реализуемы.

Однако, М-Х алгоритм не требует вычисления константы нормировки, и при некотором n моделирование $\psi_n(X)$ оказывается теоретически, по крайней мере, простой задачей. Трудность может представлять эффективный выбор цепи Маркова $p(X, Y)$. Для получения качественных алгоритмов поиска глобального экстремума здесь следует привлекать марковские процедуры случайного поиска.

Список литературы

- [1] *N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller* Equations of state Calculations by fast computing machines // J. Chem. Phys. Vol. 21, pp. 1087-1092, 1953.
- [2] *Liu J.S.* Monte Carlo Strategies in Scientific Computing. – Springer, 2001.
- [3] *Винклер Г.* Анализ изображений, случайные поля и методы Монте-Карло на цепях Маркова. Математические основы. – Новосибирск, ГЕО, 2008.
- [4] *Ермаков С.М.* Метод Монте-Карло и смежные вопросы. – Москва: Наука, 1975.