

МСМС, Метод Метрополиса-Гастингса

Звонарев Никита

Слушатели: 4-й курс бакалавриата

Санкт-Петербургский Государственный Университет
Кафедра Статистического моделирования



Санкт-Петербург – 2025 г.

Markov Chain Monte-Carlo — класс алгоритмов, занимающихся генерацией случайных величин с помощью моделирования марковской цепи с требуемым стационарным распределением.

Приложения:

- Сэмплинг случайных величин с невыпуклым логарифмом функции плотности
- Сэмплинг многомерных случайных векторов
- Оценивание интегралов
- В том числе решение СЛАУ
- Оптимизация методом имитации отжига
- ...

Последовательность случайных величин X_0, X_1, X_2 называется марковским процессом, если

$\mathbb{P}(X_{n+1} \in A | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} \in A | X_n = x_n)$
для любого n и измеримого множества A (марковское свойство).

Марковский процесс является стационарным, если

$$\mathbb{P}(X_0 \in A_0, X_1 \in A_1, \dots, X_n \in A_n) = \\ \mathbb{P}(X_k \in A_0, X_{k+1} \in A_1, \dots, X_{n+k} \in A_n)$$

для любых n и k .

Стационарное распределение

Чтобы задать марковский процесс, нужны π_0 — распределение X_0 и $P(x'|x)$ — вероятности [плотности] перехода из прежнего состояния x в любое новое состояние x' . Для стационарного процесса π_0 — распределение произвольного X_i .

Распределение π является стационарным распределением для марковской цепи с условными вероятностями перехода $P(x'|x)$, если для его вероятностей [плотности в точке] выполнено:

$$\pi(x) = \int P(x|y) dP_\pi(y)$$

для любой x .

π является предельным распределением X_i при $i \rightarrow \infty$ при выполнении некоторых условий . . .

Стационарное распределение

Чтобы задать марковский процесс, нужны π_0 — распределение X_0 и $P(x'|x)$ — вероятности [плотности] перехода из прежнего состояния x в любое новое состояние x' . Для стационарного процесса π_0 — распределение произвольного X_i .

Распределение π является стационарным распределением для марковской цепи с условными вероятностями перехода $P(x'|x)$, если для его вероятностей [плотности в точке] выполнено:

$$\pi(x) = \int P(x|y) dP_\pi(y)$$

для любой x .

π является предельным распределением X_i при $i \rightarrow \infty$ при выполнении некоторых условий ...

Марковский процесс называется обратимым, если для его вероятностей перехода $P(x'|x)$ выполнено *уравнение детального баланса*:

$$\pi(x)P(x'|x) = \pi(x')P(x|x').$$

Дискретный случай: для каждой пары состояний x, x' вероятность нахождения в x и перехода в x' равна вероятности нахождения в x' и перехода в x .

Замечание

Уравнения детального баланса достаточно для существования стационарного распределения.

Марковский процесс называется (топологически) *неприводимым*, если для любой точки x и любого открытого множества A существует такое $t \geq 0$, что $P_t(x, A) > 0$, где $P_t(x, A) = \mathbb{P}(x_{t+1} \in A | X_1 = x)$.

Периодом состояния x называется $d(x) = \text{НОД}(n \in \mathbb{N} | P_n(x|x) > 0)$. Если $d(x) = 1$, то состояние x называется аperiodическим.

Марковский процесс называется *aperiodическим*, если все его состояния аperiodичны.

Первое время возврата состояния x :

$$\tau(x) = \{\min \tau \in \mathbb{N} : X_{\tau+1} = x | X_1 = x\}.$$

Если $\mathbb{E}\tau(x) < +\infty$ для всех x , то такой процесс называется *положительно рекуррентным*.

Эргодичность = неприводимость + апериодичность + положительная рекуррентность!

Замечание

Эргодичности достаточно для единственности стационарного распределения π у марковского процесса.

Хотим научиться моделировать распределение P с плотностью [вероятностями] $P(x)$.

Подход: построим марковский процесс, у которого стационарное распределение π — это требуемое распределение P .

Уравнение детального баланса для P перепишем как:

$$\frac{P(x'|x)}{P(x|x')} = \frac{P(x')}{P(x)}.$$

Затем представим $P(x'|x) = g(x'|x)A(x', x)$, где $g(x'|x)$ — т.н. *proposal distribution*, $A(x', x)$ — *acceptance distribution* (вероятность принять предложенную точку x').

Подставляем, получаем:

$$\frac{A(x', x)}{A(x, x')} = \frac{P(x')}{P(x)} \frac{g(x|x')}{g(x'|x)}.$$

Выбор Метрополиса для того, чтобы это условие выполнялось:

$$A(x', x) = \min \left(1, \frac{P(x')}{P(x)} \frac{g(x|x')}{g(x'|x)} \right),$$

т.о., либо $A(x', x) = 1$, либо $A(x, x') = 1$.

Частные случаи: $g(x'|x) = g(x' - x)$ (Random Walk), $g(x'|x) = g(x|x')$ (симметрия, оригинальная работа Метрополиса), $g(x'|x) = g(x')$.

Метрополис-Гастингс: алгоритм

- ❶ Положим x_0 в соответствии с распределением π_0 .
- ❷ Положим $t = 0$
- ❸ Для $t = 0, 1, \dots, T - 1$:
 - Промоделировать x' в соответствии с proposal distribution $g(x'|x)$.
 - Вычислим вероятность принятия новой точки
$$A(x', x_t) = \min \left(1, \frac{P(x')}{P(x_t)} \frac{g(x_t|x')}{g(x'|x_t)} \right).$$
 - Промоделируем $\alpha \sim U_{[0,1]}$.
 - Если $\alpha \leq A(x', x_t)$, то мы принимаем новое состояние, и $x_{t+1} = x'$, иначе оставляем прежнее $x_{t+1} = x_t$.

В случае выполнения всех вышеописанных требований к марковской цепи алгоритм даёт последовательность реализаций случайных величин $x_0, x_1, \dots, x_T$, при этом распределение x_t сходится к P при $t \rightarrow +\infty$.

Вычисление интегралов с помощью марковских процессов

Предположим, что X_1, X_2, X_3 — марковский процесс, при этом X_1 имеет стационарное распределение π , g — измеримая вещественная функция, $\mathbb{D}(X_1) < +\infty$.

Теорема (ЦПТ для марковских процессов)

Положим:

- $\mu = \mathbb{E}(g(X_1))$,
- $\hat{\mu}_n = \frac{1}{n} \sum_{k=1}^n g(X_k)$,
- $\sigma^2 = \lim_{n \rightarrow \infty} n \mathbb{D}(\hat{\mu}_n) = \mathbb{D}(g(X_1)) + 2 \sum_{k=1}^{\infty} \text{cov}(g(X_1), g(X_{k+1}))$.

Тогда $\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$ [сходимость по распределению].

При наличии эргодичности требование на стационарность процесса можно смягчить.

- 1 Burn-in — убираем сколько-то первых шагов из последовательности "до выхода на стационарность" (избавляемся от $x_1, x_2, \dots$, у которых распределения могут быть далекими от стационарного).
Практический трюк, чтобы снизить возможное большое смещение на первых шагах.
- 2 Прореживание — оставляем только каждый k -ый элемент цепи для уменьшения закореллированности цепи (*для какой задачи это важно, для какой нет?*).
- 3 Посмотреть на автокорреляционную функцию процесса (acf в R), на траекторию, на acceptance rate...