Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский университет
«Высшая школа экономики»

Факультет математики

Имамутдинова Лилия Рафаильевна

Метод Монте-Карло по схеме марковских цепей

Курсовая работа студентки 1 курса
образовательной программы бакалавриата «Математика»

Научный руководитель:
Дымов Андрей Викторович

Москва 2024

Аннотация

Методы Монте-Карло с цепями Маркова — это класс алгоритмов для семплирования, которые основаны на создании марковской цепи, со стационарным распределением желаемого вида. Одним из представителей класса является алгоритм Метрополиса-Гастинга. В работе приведён пример использования данного метода для решения задачи из области криптографии. Для анализа времени работы алгоритма были рассмотрены оценки на среднее время прохода из стартового состояния в конечное для однородной Марковской цепи.

1 Введение

Метод Монте-Карло по схеме марковской цепей (MCMC) используется в широком спектре задач из различных областей. Например, в машинном обучении, генетике и биоинформатике, финансовой математике. В контексте байесовских методов машинного обучения, MCMC применяется для аппроксимации апостериорных распределений параметров обучаемой модели, таких как веса нейронной сети. Также метод используется для анализа генетических последовательностей и расчета вероятностей генетических событий, например, мутаций.

Рассмотрим ключевые условия и задачи, при которых MCMC демонстрирует высокую эффективность. Данные метод обладает преимуществом в случаях, связанных с высокой сложностью вычислений, анализом больших данных. В задачах с большим количеством параметров, где традиционные методы численной оптимизации или интегрирования сталкиваются с "проклятием размерности", MCMC позволяет эффективно семплировать из сложных распределений и делать статистические выводы. Методу не нужны предварительные предположения о форме распределения, что делает его универсальным инструментом для изучения априорно неизвестных или сложных распределений.

Методы Монте-Карло по схеме марковской цепей можно использовать для работы как с непрерывными, так и с дискретными распределениями. В данной работе внимание будет уделено дискретным случаям. Это особенно важно в таких приложениях, как криптография, где объектом исследования являются дискретные коды и шифры, или в биоинформатике, где работа ведется с дискретными генетическими последовательностями.

2 Метод Монте-Карло по схеме марковских цепей

Методы MCMC позволяют сгенерировать выборку $X_1, ..., X_n$ из распределения, с вероятностной мерой $p(x)$, заданной на конечном пространстве χ . Для формирования выборки используется однородная марковская цепь, с конечным числом состояний. Вероятность перехода из состояния x в состояние y равна $Q(x, y)$, и априорное распределение равно $p_0(x)$. Выборка генерируется следующим образом:

$$X_1 \sim p_0(x),$$
$$X_n \sim Q(X_{n-1}, X_n)$$

Задача выбрать марковскую цепь так, чтобы описанный выше метод давал выборку с заданным распределением.

Обозначим за $p_n(x)$ распределение n-ой точки выборки, тогда согласно определению марковской цепи.

$$p_n(x) = \sum_{y \in \chi} Q(y, x) p_{n-1}(y),$$

где Y – множество всевозможных состояний.

Определение 1. *Вероятностное распределение $\pi(x)$ называется стационарным относительно цепи с вероятность перехода из состояния x в состояние y равной $Q(x, y)$, если выполняется равенство*

$$\pi(x) = \sum_{y \in \chi} Q(y, x) \pi(y). \tag{2.1}$$

В MCMC нужно, чтобы $p(x)$ было стационарно. Рассмотрим условия, которые гарантируют стационарность распределения.

Лемма 2.1. *Пусть $\pi(x)$ – распределение, удовлетворяющее уравнению детального баланса*

$$\pi(y) Q(y, x) = \pi(x) Q(x, y). \tag{2.2}$$

Тогда распределение $\pi(x)$ стационарно.

Доказательство. Проверим равенство 2.1:

$$\sum_{y \in Y} \pi(y) Q(x, y) dY \overset{(2.2)}{=} \sum_{y \in Y} \pi(x) Q(y, x) dy = \pi(x) \sum_{y \in Y} Q(y, x) = \pi(x)$$

□

Определение 2. *Однородная марковская цепь с матрицей переходов Q называется перемешивающей, если она имеет единственное стационарное распределение $\pi(x)$ и*

$$\lim_{n \to \infty} Q^n(x, y) = \pi(y),$$

где $Q^n(x, y)$ – элемент находящий на пересечении x-той строки и y-того столбца матрицы Q^n.

Теорема 2.2. *Пусть задана марковская цепь с матрицей переходов Q и существует n такое, что каждый элемент матрицы Q^n положителен. Тогда марковская цепь перемешивает.*

Таким образом, для генерации выборки с заданным распределением достаточно, чтобы марковская цепь была перемешивающей, а распределение $p(x)$ было стационарно. Однако полученные случайные величины не будут независимыми, поэтому нужно проредить набор, взяв каждую m-тую величину, где m велико. При этих условиях и достаточно больших n распределение генерируемой выборки будет близко к желаемому.

3 Метод Метрополиса-Гастинга

Один из самых популярных представителей MCMC – это алгоритм Метрополиса-Гастинга. Данный метод был предложен в статье [5] 1953 года Николасом Метрополисом с соавторами. Изначально алгоритм был представлен только для случая симметричного распределения, но в 1970 году У.К. Гастингс расширил его на более общий случай.

Перейдём к описанию алгоритма. Пусть необходимо семплировать выборку из распределения $p(x)$. В основе метода Метрополиса-Гастинга лежат два основных шага:

1. Генерация следующего значения цепи.
   Зафиксировали марковскую цепь с матрицей переходов M, тогда $M(y, x)$ – вероятность предложения состояния y при нахождении в состоянии x.

2. Принятие или отвержение сгенерированного значения, основанное на некотором правиле.
   Пусть $A(x, y)$ – вероятность принять предложенное состояние y. Таким образом, получаем, что вероятность перехода из состояний x в состояние y равна $M(x, y)A(x, y)$

Описанный алгоритм создаёт новую марковскую цепь с матрицей переходов K, такую, что

$$K(x, y) = \begin{cases} M(x, y)A(x, y), & \text{если } x \neq y \\ M(x, y) + \sum\limits_{z: A(x,z)<1} M(x, z)(1 - A(x, z)), & \text{если } x = y. \end{cases} \tag{3.1}$$

Лемма 3.1. *Если матрица M стохастическая, то и K стохастическая.*

Доказательство. Очевидно, все элементы матрицы K неотрицательны. Докажем, что сумма элементов в каждой строке равна 1.

$$\sum_z K(x,z) =$$

$$= \sum_{z:A(x,z)=1} M(x,z) + \sum_{z:A(x,z)<1} M(x,z)A(x,z) +$$

$$+ M(x,x) + \sum_{z:A(x,z)<1} M(x,z)(1-A(x,z)) =$$

$$= \sum_{z:A(x,z)\geqslant 1} M(x,z) + M(x,x) + \sum_{z:A(x,z)<1} M(x,z) = \sum_z M(x,z) = 1$$

□

Лемма 3.2. *Пусть существует состояние i, такое что $p_{ii} > 0$ и из i достижимо любое состояние j и из любого j достижимо i. Тогда марковская цепь перемешивает.*

Доказательство. Рассмотрим состояние марковской цепи как вершины графа, а рёбра как переход с вероятностью равной весу ребра. Докажем, что существует n такое, что матрица переходных вероятностей за n шагов P^n имеет строго положительные элементы. Найдём кратчайшие пути от вершины i до других вершин. Обозначим длину максимального из этих кратчайших путей за d. Аналогично обозначим длину максимального из кратчайших путей от других вершин до i за g. Не более, чем за $g+d$ шагов можно дойти из любого состояния x до любого другого состояния y, сначала из x дойдя до вершины i, а дальше из i в y. Вероятность остаться в состоянии i при переходе ненулевая, значит путь меньшей длины можно удлинить до $g+d$ за счет остановки в состоянии i. Таким образом, матрица переходных вероятностей за $g+d$ шагов P^{g+d} имеет строго положительные элементы. Тогда по теореме 2.2 марковская цепь с матрицей переходов P перемешивает. □

Лемма 3.3. *Если марковская цепь M удовлетворяет условиям леммы 3.2 и $A(x,y) > 0$, тогда марковская цепь K перемешивает.*

Доказательство. Рассмотрим граф, соответствующий марковской цепи M, перестроим его в граф для марковской цепи K. Рассмотрим состояние i, такое что $p_{ii} > 0$ и из i достижимо любое состояние j. Из определения марковской цепи 3.1 и условия $A(x,y) > 0$ следует, что если ребро имеет положительный вес, то после преобразования вес также положительный, а значит любое состояние j остаётся достижимым. Аналогично, $K(i,i) \geqslant M(i,i) > 0$. Таким образом, марковская цепь с матрицей переходов K также удовлетворяет условиям леммы 3.2, а значит цепь перемешивает. □

Остаётся выбрать функцию принятия A так, чтобы распределение $\pi(x)$ было стационарно для полученной марковской цепи.

Часто выбирают

$$A(y,x) = \min(1, \frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}). \tag{3.2}$$

Если $p(x)Q(x,y) = 0$, то подразумевается, что $A(y,x) = 1$.

В таком случае либо $A(y,x) = 1$, либо $A(x,y) = 1$. Покажем, что при таком выборе распределение действительно стационарно.

Лемма 3.4. *Если вероятность принятия задана формулой 3.2, то распределение $p(x)$ стационарно для марковской цепи, полученной алгоритмом Метрополиса-Гастинга.*

Доказательство. Результат следует из леммы 2.1:

$$p(x)Q(x,y)A(x,y) = \min(p(x)Q(x,y), p(y)Q(y,x)) = p(y)Q(y,x)A(y,x).$$

□

Метод Метрополиса-Гастинга помогает в случае, когда распределение, из котого нужно делать выборку, известно с точностью до константы, так как в формуле 3.2 она сокращается.

4 Использование MCMC в криптографии

Методы Монте-Карло с марковскими цепями часто используются в компьютерных науках, в частности в криптографии. Они позволяют сократить вычисления, так как часто машины не способны сделать полный перебор случаев за приемлемое время. К примеру, если имеется текст, зашифрованный заменой, то есть символы шифруемого текста заменяются символами того же алфавита по заранее установленным правилам замены. Тогда дешифровать сообщение можно с помощью перебора всевозможных перестановок символов в алфавите. Алфавиты большинства современных языков мира содержат в среднем 30 букв, значит перебрать необходимо $30! \approx 2.65 10^{32}$ перестановок. Подобные вычисления оказываются времязатратными и требуют значительных ресурсов. Метод Метрополиса-Гастингса может оказаться полезным в данном случае.

Если количество символов шифра и алфавита совпадают, то задача расшифровки текста может быть поставлена как поиск функции

$$f : \{\text{символы шифра}\} \to \{\text{алфавит}\}. \tag{4.1}$$

Данная функция является биективной, так как каждому символу шифра сопоставляется единственный символ алфавита. Функцию 4.1 можно рассматривать как перестановку. Занумеруем символы шифра и алфавита, получим два набора чисел от 1 до n, где n размер алфавита. Перестановку можно определить как биекцию, отображающую множество S в себя. В данном случае в качестве множества S будет рассматриваться полученный набор чисел.

тогда можно рассматривать функцию 4.1 как нахождение перестановки.

Рассмотрим матрицу F, полученную с помощью частотного анализа текста. Пусть x и y символы алфавита, тогда $F(x, y)$ – частота встречаемости упорядоченной пары букв (x, y), стоящих в тексте на соседних позициях. Определим функцию правдоподобия как

$$Pl(f) = \prod_i F(f(s_i), f(s_{i+1})), \tag{4.2}$$

где s_i – i-тый символ зашифрованного текста.

Данная функция характеризует вероятность принадлежности расшифрованного текста к естественному языку в заданном алфавите. Нужно найти f на котором $Pl(f)$ приближена к максимальному значению.

Построим перемешивающую марковскую цепь, состояниями которой будут являться всевозможные перестановки, а в качестве стационарного состояния будет $\pi(f) = \frac{1}{z}Pl(f)$, где z – нормировочная константа. Алгоритм Метрополиса-Гастинга преобразует эргодическую марковскую цепь с матрицей переходов M в новую цепь с матрицей переходов K согласно равенству (3.1). Стационарное распределение новой цепи с матрицей перехода K равно π. После того, как цепь выйдет на стационарное распределение, можно будет делать выводы о функции $Pl(f)$, например на основе того, в каком состоянии чаще всего оказывалась марковская цепь или на каком из посещенных состояний данная функция принимает наибольшее значение.

В качестве матрицы M рассмотрим

$$M(f, f_{new}) = \begin{cases} \frac{1}{\binom{n}{2}}, & \text{если } f_{new} \text{ отличается от } f \text{ одной транспозицией} \\ 0, & \text{иначе} \end{cases}$$

Алгоритм может быть интерпретирован как блуждание по графу, где вершины – перестановки, а рёбра – транспозиции. Находясь в вершине x, нужно выбрать в какую вершину совершится переход. Выбираем вершину y, смежную с x с вероятностью $M(x, y)$. Далее проверяем условие, если $A(x, y), \geqslant 1$ то перемещаемся в y, иначе подбрасываем перемещаемся в y с вероятностью $A(x, y)$. То есть $A(x, y)$ является некоторой функций, показывающей выгодно ли нам переместится в конкретную вершину или лучше остаться на текущей позиции и проверить другие.

Положим

$$A(x, y) = \min(1, \frac{\pi(y)M(y, x)}{\pi(x)M(x, y)}) = \min(1, \frac{\pi(y)}{\pi(x)}).$$

Согласно лемме 3.4, распределение $\pi(x)$ инвариантно относительно марковской цепи K. Теперь, осталось доказать, что цепь K перемешивает.

Определение 3. *Неразложимый класс — класс эквивалентности множества состояний по отношению сообщаемости. Если представить марковскую цепь как граф, неразложимый класс будет аналогичен компоненте сильной связности.*

Определение 4. *Неразложимая цепь — цепь Маркова, в которой все состояния образуют один неразложимый класс.*

Определение 5. *Число d_i, равное $\gcd(\{n : p_{ii}(n) > 0\})$ называется периодом состояния i.*

Определение 6. *Цепь называется апериодической, если все ее состояния апериодичны, то есть их периоды равны единице.*

Теорема 4.1 (эргодическая теорема для цепей Маркова с конечным числом состояний)**.** *Неразложимая апериодическая цепь Маркова перемешивает.*

Лемма 4.2. *Пусть для любого x выполняется $\pi(x) > 0$. Тогда цепь K удовлетворяет условиям теоремы 4.1*

Доказательство. Из условия следует, что $\frac{\pi(y)}{\pi(x)} > 0$, а значит и $A(x, y) > 0$

Докажем, что цепь неразложима. Каждые два состояния, в данном случае функции перестановки, являются сообщающимися. Рассмотрим две перестановки f и f^*. Из любой перестановки можно получить любую другую перестановку с помощью k транспозиций. Поэтому в графе, где ребро это транспозиция всегда найдётся путь из f в f^*. В графе для марковской цепи на этих рёбрах будут находится значения $K(x, y) = \frac{1}{\binom{n}{2}} A(x, y) > 0$. В таком случае вероятность перехода из f в f^* равна произведению весов рёбер на пути и будет положительной.

Теперь докажем апериодичность. Вероятность перейти из состояние в себя же за один шаг равна:

$$p_{xx}(1) = K(x, x) = \frac{1}{\binom{n}{2}} \sum_{z:A(x,z)<1} (1 - \frac{\pi(z)}{\pi(x)}) > 0$$

Откуда получаем, что если у состояния x есть хотя бы одно сообщающееся состояние z, для которого $A(x, z) < 1$, то период состояния равен 1. Теперь рассмотрим случай, когда для любого x $A(x, z) \geqslant 1$. В таком случаем перейдём в любую из смежных вершин z, и так как $A(x, z) = \frac{\pi(z)}{\pi(x)} \geqslant 1$, то $A(z, x) = \frac{\pi(z)}{\pi(x)} \leqslant 1$, а значит $p_{zz}(n) \geqslant (p_{zz}(1))^n > 0$ для любого n. Тогда

$$p_{xx}(k) = p_{xz}(1)p_{zz}(k - 2)p_{xz}(1) > 0$$

для $k \geqslant 3$, а значит период состояния равен

$$\gcd(\{n : p_{xx}(n) > 0\}) = \gcd(\{k : k \geqslant 3\}) = 1.$$

Таким образом, для любое состояние является апериодическим, а значит цепь апериодична.

$\square$

5 Среднее время достижения состояния Марковской цепи

В примере, изложенном в параграфе 4, одним из подходов к определению дешифрующей перестановки является выбор состояния, на котором функция $Pl(f)$ 4.2 принимает наибольшее значение, относительно других посещённых состояний. Важно понимать количество шагов необходимых для посещения состояния, на котором значение функции $Pl(f)$ близко к её глобального максимуму.

Среднее время достижения, это ожидаемое количество шагов, необходимых для того, чтобы марковская цепь впервые достигла определённого состояния, начиная из другого состояния. Этот параметр важен для понимания того, как быстро марковская цепь может "перемещаться" по своему пространству состояний.

Рассмотрим среднее время прохода из состояний i в состоянии j однородной Марковской цепи.

$$m_{ij} = \mathbb{E}\min\{n \geqslant 1 : X_n = j | X_0 = i\}$$

Лемма 5.1. *Пусть p_{ij} – вероятность перехода из i-того состояния в состояние j. Тогда среднее время прохода задаётся рекуррентной формулой*

$$m_{ij} = p_{ij} + \sum_{k \neq j} p_{ik}(m_{kj} + 1). \tag{5.1}$$

Доказательство. Применим формулу полного математического ожидания ожидания

$$m_{ij} = \sum_k \Pr(X_1 = k | X_0 = i) \cdot \mathbb{E}(\min\{n \geqslant 1 : X_n = j | X_0 = i\} | X_1 = k).$$

Рассмотрим отдельно случаи, когда $k = j$ и когда $k \neq j$. Если $k = j$,

$$\Pr(X_1 = j | X_0 = i) \cdot \mathbb{E}(min\{n \geqslant 1 : X_n = j | X_0 = i\} | X_1 = j) = p_{ij} \cdot \mathbb{E}1 = p_{ij}.$$

Если $k \neq j$,

$$\begin{aligned}
&\Pr(X_1 = k | X_0 = i) \cdot \mathbb{E}(\min\{n \geqslant 1 : X_n = j | X_0 = i\} | X_1 = k) = \\
&\qquad = \Pr(X_1 = k | X_0 = i) \cdot \mathbb{E}(\min\{n \geqslant 1 : X_n = j | X_0 = k\} + 1) = \\
&\qquad = p_{ik}(m_{kj} + 1)
\end{aligned}$$

Таким образом, подставляя полученные результаты в формулу полного математического ожидания, получаем

$$m_{ij} = p_{ij} + \sum_{k \neq j} p_{ik}(m_{kj} + 1).$$

□

Заметим, что $\sum_k p_{ik} = 1$, значит формулу (5.1) можем переписать как

$$m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj}.$$

Пусть матрица $M = \{m_{ij}\}$, $M_d = \mathrm{diag}(m_{ii})$, J – матрица единиц, E – единичная, P – матрица переходов. Получаем равенство

$$M = P(M - M_d) + J \tag{5.2}$$

Лемма 5.2. *Пусть π – стационарное распределение цепи Маркова с матрицей переходов P. Тогда при всех i выполняется*

$$m_{ii} = \frac{1}{\pi_i}.$$

Доказательство. Умножим левую и правую части формулы (5.2) слева на вектор-строку стационарного распределения π

$$\pi M = \pi M - \pi M_d + \pi J.$$

Следовательно,

$$\pi M_d = \pi J.$$

Отсюда получаем

$$\pi_i m_{ii} = \sum_k \pi_k = 1,$$

то есть

$$m_{ii} = \frac{1}{\pi_i}$$

□

Теорема 5.3. [2, Proposition 11.1] Пусть P – матрица переходов эргодической цепи Маркова с конечным числом состояний, то есть $P^n \to W$ при $n \to \infty$. Тогда матрица $E - P + W$ обратима.

Доказательство. Покажем, что уравнение

$$(E - P + W)x = 0 \tag{5.3}$$

имеет единственное решение $x = 0$.

π – стационарное распределение, то есть собственный вектор матрицы P, тогда

$$\pi(E - P) = \pi - \pi P = \pi - \pi = 0.$$

Так как $P^n \to W$ при $n \to \infty$, а π не зависит от n, то устремив n к бесконечности получаем равенство $\pi W = \pi$.

Докажем, что $Wx = 0$. Строками матрицы W является вектор π, поэтому достаточно показать, что $\pi x = 0$. Для этого домножим уравнение (5.3) слева на вектор π

$$\pi(E - P + W)x = \pi(E - P)x + \pi W x = \pi x = 0. \tag{5.4}$$

Подставив $Wx = 0$ в уравнение (5.3), получаем

$$(E - P + W)x = (E - P)x + Wx = (E - P)x = 0.$$

Раскроем скобки и перенесём Px в правую часть

$$x = Px.$$

Домножая слева на P и приравнивая Px к x получаем, что для любого n выполняется $x = P^n x$, а значит можем устремить n к бесконечности и получить равенство

$$x = Wx.$$

Таким образом, элементы вектора x совпадают друг с другом и равны πx, воспользовавшись равенством (5.4) получаем, что $x = 0$. А значит матрица $E - P + W$ невырождена.

□

Теорема 5.4. [2, Theorem 11.16] Пусть $Z = (E - P + W)^{-1}$. Тогда

$$m_{ij} = \frac{z_{jj} - z_{ij}}{\pi_j},$$

где $i \neq j$.

Доказательство. Обозначим за M_0 матрицу $M - M_d$, то есть матрицу M, диагональные элементы которой заменили нулями.

$$M_0 = PM_0 + J - M_d.$$

Перенесём матрицу PM_0 в левую сторону и домножим обе части на матрицу Z

$$Z(E - P)M_0 = ZJ - ZM_d. \tag{5.5}$$

Докажем, что $ZJ = J$. Домножим матрицы P и W справа на J. Элементы полученных матриц PJ и WJ равны $\sum_j p_{ij} = 1$ и $\sum_j \pi_j = 1$ соответственно. Таким образом, $PJ = WJ = J$.

Домножим матрицу $Z^{-1} = E - P + W$ справа на J

$$(E - P + W)J = J - PJ + WJ = J$$

Домножая полученное выражение справа на матрицу Z, доказываем равенство $ZJ = J$.

Возвращаясь к равенству (5.5), получаем

$$Z(E-P)M_0 = J - ZM_d$$

Матрицу $Z(E-P)$ можно выразить как $E-W$. Это следует из свойств матрицы W, которая является предельной для матрицы P. Домножая матрицу $E-W$ слева на обратную к Z, получаем матрицу $E-P$

$$(E-P+W)(E-W) = E-P+W-W+PW-W^2 = E-P+W-W = E-P.$$

Тогда подставляя $E-W$ вместо $Z(E-P)$ и раскрыв скобки можем выразить матрицу M_0

$$M_0 = J - ZM_d + WM_0.$$

Посмотрим на диагональный элемент, находящийся на пересечении j столбца и строки

$$0 = 1 - \frac{z_{jj}}{\pi_j} + \sum_{k\neq j} \pi_k m_{kj}$$

Недиагональные элементы дают равенство:

$$m_{ij} = 1 - \frac{z_{ij}}{\pi_j} + \sum_{k\neq j} \pi_k m_{kj}$$

Таким образом, вычитая из второго равенство первое, получаем

$$m_{ij} = \frac{z_{jj} - z_{ij}}{\pi_j}$$

□

Хотим получить некоторое обоснование, что в состояние с наибольшим значением π, независимо от начального состояния, в среднем будем приходить быстрее, чем в остальные

Теорема 5.5. [3, Proposition 1.2.] Пусть P – матрица перехода эргодической неприводимой марковской цепи, с конечным числом состояний. Тогда

$$\max_{\substack{1\leqslant i\leqslant n\\ i\neq j}} m_{ij} \geqslant \frac{1}{\pi_j} - 1$$

Доказательство. Не умаляя общности предположим, что $j = n$. Разобьём матрицу P, отделив блок, состоящий из p_{nn}. Сумма элементов в строке равна единице, поэтому $p_{nn} = 1 - \sum_{i\neq n} p_{ni} = 1 - r^T e$, а разбиение выглядит следующим образом

$$P = \left(\begin{array}{c|c} T & (E-T)e \\ \hline r^T & 1 - r^T e \end{array}\right),$$

где e – вектор-столбец из единиц, T – матрица, образованная $n-1$ первыми строками и столбцами матрицы P, r^T – вектор-строка, образованный $n-1$ первыми элементами n-той строки матрицы P.

Пусть π – это собственный вектор-строка, тогда получаем равенство

$$\pi P = \pi$$

Перенесём вектор-строку π в левую часть и вынесем за скобки

$$\pi(P-E) = 0$$

Разобьём получившуюся матрицу $P-E$ на блоки, отделив элемент, стоящий на пересечении n-той строки и n-того столбца, в отдельный блок, получим

$$(\pi_1, ..., \pi_{n-1}, \pi_n) \left(\begin{array}{c|c} T-E & (E-T)e \\ \hline r^T & -r^T e \end{array}\right) = 0$$

Рассмотрим вектор-столбец, полученный перемножением вектора-строки π и подматрицы, состоящей из блока $E-T$ и вектора строки r^T. Исходя из равенства выше, получаем, что такое произведение равно нулевому вектору-строке. Тогда

$$(\pi_1, ..., \pi_{n-1})(T-E) + \pi_n r^T = 0.$$

Следовательно

$$(\pi_1, ..., \pi_{n-1})(E-T) = \pi_n r^T.$$

Матрица $E-T$, образованная $n-1$ первыми строками и столбцами матрицы $E-P$, обратима [1]

$$(\pi_1, ..., \pi_{n-1}) = \pi_n r^T (E-T)^{-1}$$

Домножим обе части справа на вектор-столбец e

$$(\pi_1, ..., \pi_{n-1})e = \pi_n r^T (E-T)^{-1} e$$

Заметим, что $\sum_{k=1..n} \pi_k = 1$, тогда левую часть можно представить как $\sum_{k \neq n} \pi_k = 1 - \pi_n$. Подставим это в предыдущее выражение, тогда выражая π_n получаем

$$\pi_n = \frac{1}{1 + r^T (E-T)^{-1} e} \tag{5.6}$$

Аналогичным образом разобьём на блоки матрицу M

$$M = \left(\begin{array}{c|c} \hat{M} & y \\ \hline s^T & m_{nn} \end{array}\right)$$

Подставим полученные разбиения в формулу (5.2)

$$\left(\begin{array}{c|c} \hat{M} & y \\ \hline s^T & m_{nn} \end{array}\right) = \left(\begin{array}{c|c} T & (E-T)e \\ \hline r^T & 1 - r^T e \end{array}\right) \left(\begin{array}{c|c} \hat{M}_0 & y \\ \hline s^T & 0 \end{array}\right) + J$$

Перемножим матрицы справа и рассмотрим блок, состоящий из вектора-столбца y

$$y = Ty + (E - T)e \cdot 0 + e$$

Перенесём вектор-столбец Ty в левую часть и домножим обе части равенства на обратную матрицу $(E - T)^{-1}$

$$y = (E - T)^{-1}e$$

Домножим обе части на вектор-строку r^T

$$r^T(E - T)^{-1}e = r^T y = \sum_{i=1}^{n-1} r_i m_{in} \leqslant \sum_{i=1}^{n-1} r_i(\max_{1 \leqslant i \leqslant n-1} m_{in}) \leqslant \max_{1 \leqslant i \leqslant n-1} m_{in}$$

Выразим $r^T(E - T)^{-1}e$ из равенства (5.6)

$$\frac{1}{\pi_n} - 1 \leqslant \max_{1 \leqslant i \leqslant n-1} m_{in}$$

□

Таким образом, при уменьшении стационарного распределения оценка снизу на среднее количество шагов, требуемое для достижения состояния, будет увеличиваться. Часто начальное состояние выбирается случайно, алгоритм может стартовать из самой удалённой вершины, поэтому важно смотреть на максимум среднего времени достижения. Данная оценка в среднем позволяет понять сколько минимум шагов необходимо сделать, чтобы выйти на нужное состояние.

Список литературы

[1] Г. И. Фалин, “Об эргодичности блужданий в полуполосе”, Матем. заметки, 1988, том 44, выпуск 2, 225–230.

[2] Charles M. Grinstead and J. Laurie Snell. Introduction to Probability. The American Mathematical Society, 1997.

[3] J. Breen and S. Kirkland, “Minimising the largest mean first passage time of a Markov chain: the influence of directed graphs”, Linear Algebra and its Applications 520, 2017, 306-334.

[4] P. Diaconis, “The Markov chain Monte Carlo revolution”, Bull. American Math. Soc., 2008, 179-205.

[5] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth and A. H. Teller, “Equation of State Calculations by Fast Computing Machines”, The Journal of Chemical Physics, 1953, 1087–1092.