

О методах Монте-Карло с цепями Маркова

Шведов А.С.

В работе рассматриваются два метода Монте-Карло с цепями Маркова, широко применяемые в эконометрических исследованиях. Это алгоритм Метрополиса и гиббсовский выбор. Приводится описание обоих методов. Методы Монте-Карло с цепями Маркова предназначены для симулирования наборов векторов, отвечающих многомерным распределениям вероятностей. В частности, эти методы применяются в байесовской статистике для исследования апостериорных распределений. Существенное значение имеет соблюдение условия инвариантности, доказательства, что это условие выполняется, приводятся для обоих методов. Для обоснования и изучения методов используется теория цепей Маркова с конечным числом состояний. На нескольких примерах исследуется точность рассматриваемых методов Монте-Карло с цепями Маркова. Эти примеры включают двумерное нормальное распределение с высокой корреляцией, двумерное экспоненциальное распределение, смесь двумерных нормальных распределений.

Ключевые слова: симулирование многомерных распределений; алгоритм Метрополиса; гиббсовский выбор.

1. Введение

В последние два десятилетия в эконометрическую науку прочно вошли методы Монте-Карло с цепями Маркова, так называемые МСМС (Markov chain Monte Carlo) методы. Большое значение для развития этого направления эконометрики имеют современные высокоскоростные компьютеры. Математическая модель любой системы должна быть оправданно сложной, чтобы включать в себя основные черты этой системы, но не должна быть излишне сложной. Как правильно определить границу между оправданной и излишней сложностью? И в естественных науках, и в науках о человеке и обществе эта граница при решении многих трудных задач смещается в сторону все более сложных математических моделей. Во многом это определяется увеличением вычислительных мощностей, используемых при анализе.

Но чем более сложная эконометрическая модель используется, тем труднее ее исследовать аналитическими методами и тем чаще приходится прибегать к чис-

Шведов А.С. – д. физ.-мат. н., профессор кафедры математической экономики и эконометрики Государственного университета – Высшей школы экономики, e-mail: ashvedov@hse.ru

Статья поступила в Редакцию в январе 2010 г.

ленным методом. Примеры интересных и важных эконометрических моделей, использование которых становится возможным благодаря привлечению методов Монте-Карло с цепями Маркова, можно найти, скажем, в книгах [5] и [9]. Так, в [5] приводится пример, относящийся к аграрной экономике, в [9] – пример из портфельной теории. Число подобных примеров достаточно велико.

При применении метода Монте-Карло требуется построить набор случайных чисел, отвечающих тому или иному распределению вероятностей. Для одномерных функций распределения существующие подходы к построению наборов случайных чисел излагаются, например, в [1, гл. 4]. Такую же задачу построения набора теперь уже не случайных чисел, а векторов необходимо решать и для многомерных функций распределения. Для решения этой задачи и используются методы Монте-Карло с цепями Маркова.

Многие применения методов Монте-Карло с цепями Маркова в эконометрике связаны с байесовским подходом, когда эти методы используются для изучения апостериорного распределения параметров модели. Именно с таких позиций чаще всего ведется изложение и в обзорах, посвященных методам Монте-Карло с цепями Маркова. Но суть этих методов никак не связана с байесовским подходом. Одна из целей настоящей работы состоит в том, чтобы дать обзор соответствующих результатов в чистом виде, не увязывая с их возможным применением в байесовской эконометрике. Непростые в идейном отношении методы заслуживают такого аккуратного рассмотрения.

Кроме того, работа содержит некоторые новые соображения, касающиеся обоснования методов Монте-Карло с цепями Маркова.

Еще одна цель данной работы связана с достаточно длительной дискуссией о так называемой диагностике сходимости при применении методов Монте-Карло с цепями Маркова (см., например, [9]). Дискуссия вызвана практически важными вопросами о том, какой длины брать цепь, или, что лучше – использовать одну длинную цепь или несколько более коротких. В настоящей работе предлагается ответ с позиций, принятых в вычислительной математике. Бессмысленно говорить о какой-то диагностике сходимости, например, при приближенном вычислении интеграла с помощью квадратурной формулы. Чем больше используется узлов, тем точнее интеграл можно вычислить. Несколько подробно разобранных в данной статье примеров показывают, что в методах Монте-Карло с цепями Маркова ситуация совершенно такая же. Чем длиннее цепь, тем выше точность. На этих же примерах, для которых известно либо аналитическое решение, либо решение, посчитанное с большой точностью другим численным методом, исследуется, какой длины необходима цепь, чтобы обеспечить ту или иную точность ответа.

Пусть $f(x)$ – совместная функция плотности компонент m -мерного случайного вектора. Рассматривается задача построения набора m -мерных векторов, отвечающих распределению вероятностей с функцией плотности f . Обозначим через B множество точек $x \in R^m$ таких, что $f(x) > 0$ при $x \in B$ и $f(x) = 0$ при $x \notin B$.

Рассмотрим последовательность m -мерных случайных векторов $X_0, X_1, \dots, X_n, \dots$, которые составляют цепь Маркова. Начальные сведения о цепях Маркова можно найти, например, в [1, гл. 6]. Для этой цепи Маркова и для борелевского множества $A \subseteq R^m$ обозначим через $P^{(n)}(x, A)$ вероятность перехода из состояния x в множество A ровно за n шагов. (Условие, что множество A является борелевским, накла-

дывается для того, чтобы по этому множеству можно было интегрировать. Параллелепипеды, шары, цилиндрические, конические множества – далеко не полное перечисление видов борелевских множеств. Примеры неборелевских множеств весьма искусственны и имеют отдаленное отношение к экономическим исследованиям. Но все же брать в качестве A произвольное множество из R^m нельзя.)

Если для любого борелевского множества $A \subseteq R^m$ при любом или почти любом x

$$(1) \quad P^{(n)}(x, A) \rightarrow \int_A f(y) dy \text{ при } n \rightarrow \infty,$$

то, реализовав цепь Маркова, можно получить набор m -мерных векторов, отвечающих распределению вероятностей с функцией плотности f . Обычное для одномерных распределений требование независимости случайных чисел, в данном случае, не накладывает. Об условии, при которых имеет место (1), см., например, статью [2] и, особенно, теорему 6 из указанной статьи. Разумеется, для выполнения условия (1) цепь Маркова $X_0, X_1, \dots, X_n, \dots$ должна каким-то образом строиться по функции f .

Методы Монте-Карло с цепями Маркова применяются при работе не только с непрерывными, но и с дискретными и смешанными многомерными распределениями. В качестве примера здесь можно назвать статью [11]. Но в данной работе мы ограничиваемся непрерывными многомерными распределениями. В разделе 2 настоящей работы рассматривается алгоритм Метрополиса, в разделе 3 – гиббсовский выбор.

2. Алгоритм Метрополиса

Этот алгоритм построения набора m -мерных векторов, отвечающих распределению вероятностей с функцией плотности f , был предложен (в несколько меньшей общности, чем излагается ниже) в работе [10]. Затем он изучался и развивался в работах [3], [8], [12], где и был доведен до современного уровня общности. Название алгоритма по фамилии только первого из авторов работы [10] является установившимся, хотя, видимо, не совсем справедливым.

Цепь Маркова $X_0, X_1, \dots, X_n, \dots$ ищется среди тех, для которых вероятности перехода за один шаг для любого $x \in B$ и для любого борелевского множества $A \subseteq R^m$ имеют вид

$$(2) \quad P^{(1)}(x, A) = g(x)I_A(x) + \int_A h(x, y) dy.$$

Здесь $I_A(x)$ – индикатор множества A , т.е. $I_A(x) = 1$, если $x \in A$, $I_A(x) = 0$, если $x \notin A$. Очевидно, что

$$(3) \quad g(x) = 1 - \int_{R^m} h(x, y) dy,$$

поскольку $P^{(1)}(x, R^m) = 1$.

Функция $h(x, y)$ строится так, чтобы выполнялось условие реверсивности:

$$(4) \quad f(x) h(x, y) = f(y) h(y, x)$$

для любых $x, y \in R^m$.

Покажем, что из выполнения условия реверсивности (4) следует выполнение условия инвариантности:

$$(5) \quad \int_A f(y) dy = \int_{R^m} P^{(1)}(x, A) f(x) dx$$

для любого борелевского множества $A \subseteq R^m$.

Действительно, используя (4) при втором переходе и (3) – при четвертом, получаем

$$\begin{aligned} \int_{R^m} P^{(1)}(x, A) f(x) dx &= \int_{R^m} f(x) g(x) I_A(x) dx + \int_{R^m} \int_A f(x) h(x, y) dy dx = \\ &= \int_A f(x) g(x) dx + \int_{R^m} \int_A f(y) h(y, x) dy dx = \\ &= \int_A f(x) g(x) dx + \int_A \int_{R^m} f(y) h(y, x) dx dy = \\ &= \int_A f(x) g(x) dx + \int_A f(y) (1 - g(y)) dy = \int_A f(y) dy, \end{aligned}$$

т.е. имеет место (5).

Как показано в [2], именно условие инвариантности (5) является основным для выполнения (1).

«Рычагом управления» в алгоритме Метрополиса является неотрицательная функция $c(x, y)$, которую следует выбрать так, чтобы

$$(6) \quad \int_{R^m} c(x, y) dy = 1$$

для любого $x \in R^m$. Далее принимается

$$\alpha(x, y) = \begin{cases} \min\left(1, \frac{f(y)c(y, x)}{f(x)c(x, y)}\right) & \text{при } f(x)c(x, y) > 0 \\ 1 & \text{при } f(x)c(x, y) = 0. \end{cases}$$

Существенно отметить, что если $x \in B$, $y \notin B$ и $c(x, y) > 0$, то $\alpha(x, y) = 0$.

Проверим, что функция

$$(7) \quad h(x, y) = c(x, y) \alpha(x, y)$$

удовлетворяет условию реверсивности (4). Выберем точки x и y .

Пусть $f(x)c(x,y)=0$. В этом случае левая часть в (4) равна нулю. Если $f(y)c(y,x)=0$, то и правая часть в (4) равна нулю. Если $f(y)c(y,x)>0$, то $\alpha(y,x)=0$, и вновь правая часть в (4) равна нулю. Таким образом, при $f(x)c(x,y)=0$ условие 4 выполняется.

Пусть $f(x)c(x,y)>0$. Если $f(x)c(x,y)<f(y)c(y,x)$, то $\alpha(x,y)=1$,

$$\alpha(y,x) = \frac{f(x)c(x,y)}{f(y)c(y,x)}.$$

Поэтому $f(x)h(x,y) = f(x)c(x,y) = f(y)c(y,x)\alpha(y,x) = f(y)h(y,x)$, и условие (4) выполняется. Если $f(x)c(x,y) \geq f(y)c(y,x)$, то $\alpha(y,x)=1$,

$$\alpha(x,y) = \frac{f(y)c(y,x)}{f(x)c(x,y)}.$$

Тогда $f(x)h(x,y) = f(x)c(x,y)\alpha(x,y) = f(y)c(y,x) = f(y)c(y,x)\alpha(y,x) = f(y)h(y,x)$.

Показано, что функция $h(x,y)$, задаваемая формулой (7), удовлетворяет условию реверсивности (4). Функция $g(x)$ строится по формуле (3), заметим, что эта функция может принимать значения лишь между нулем и единицей.

Реализацию цепи Маркова $X_0, X_1, \dots, X_n, \dots$ обозначим $x^{(0)}, x^{(1)}, \dots, x^{(n)}, \dots$. Пусть как-то выбран вектор $x^{(0)} \in B$. В алгоритме Метрополиса вектор $x^{(n+1)}$ строится по вектору $x^{(n)}$ следующим образом.

Шаг 1. Симулируется m -мерный вектор y из распределения с плотностью $c(x^{(n)}, y)$.

Шаг 2. Симулируется число u из равномерного на отрезке $[0, 1]$ распределения.

Шаг 3. При $u < \alpha(x^{(n)}, y)$ принимается $x^{(n+1)} = y$; при $u \geq \alpha(x^{(n)}, y)$ принимается $x^{(n+1)} = x^{(n)}$.

Не при любом выборе функции $c(x,y)$, удовлетворяющей условию (6), построенный набор m -мерных векторов $x^{(0)}, x^{(1)}, \dots, x^{(n)}, \dots$ отвечает распределению вероятностей с функцией плотности f . Например, если $c(x,y)=0$ при любых $x, y \in B$, то очевидно, что это не так. Приводимые ниже рассуждения помогают понять, каким условиям должна удовлетворять функция $c(x,y)$.

В статье [2] доказательство того, что имеет место (1), – это сложная теорема. Нет сомнений, что должно существовать более простое, пусть и более грубое обоснование алгоритма Метрополиса. Какие соображения привели авторов работ [3, 10] к их алгоритму? Далее приводится интуитивное объяснение алгоритма Метрополиса. Чтобы понять это интуитивное объяснение, достаточно знать теорию цепей Маркова с конечным числом состояний. А для чтения статьи [2] необходимо знание существенно более сложной общей теории цепей Маркова.

Будем считать, что множество B ограничено. Предположим также, что это множество можно разбить на m -мерные кубы A_1, \dots, A_M , длина стороны каждого из кубов ε – малое положительное число.

Положим

$$(8) \quad \pi_r = \int_{A_r} f(x) dx, \quad r=1, \dots, M.$$

Пусть $y^{(1)}, \dots, y^{(M)}$ – центры кубов A_1, \dots, A_M . Будем считать, что $x^{(0)}$ – это одна из точек $y^{(1)}, \dots, y^{(M)}$. Далее предположим, что любая симулированная на шаге 1 точка y , если она оказалась принадлежащей кубу A_r , заменяется на $y^{(r)}$. Это приводит к новой цепи Маркова, имеющей лишь конечное число состояний. Эти состояния можно занумеровать числами $1, \dots, M$.

Покажем, что для этой новой цепи Маркова с конечным числом состояний вероятности (8) приближенно являются инвариантными. (Определение инвариантных вероятностей для цепи Маркова с конечным числом состояний дается, например, в [1, гл. 6].)

При $x^{(n)} = y^{(r)}$ в цепи Маркова с конечным числом состояний новое значение $y^{(s)}$ получается на шаге 1 с вероятностью, приближенно равной $\varepsilon^m c(y^{(r)}, y^{(s)})$. Это значение оставляется с вероятностью $\alpha(y^{(r)}, y^{(s)})$. Поэтому с учетом (7) вероятность получить на шаге 3 новое значение $y^{(s)}$ приближенно равна

$$\varepsilon^m c(y^{(r)}, y^{(s)}) \alpha(y^{(r)}, y^{(s)}) = \varepsilon^m h(y^{(r)}, y^{(s)}).$$

Если новое значение не принимается, то цепь остается в состоянии $y^{(r)}$. Это происходит с вероятностью, приближенно равной

$$1 - \sum_{s=1}^M \varepsilon^m h(y^{(r)}, y^{(s)}).$$

Заменяя сумму на интеграл, получаем другое приближенное выражение для этой вероятности:

$$1 - \int_{R^m} h(y^{(r)}, y) dy.$$

В соответствии с (3) последнее выражение равно $g(y^{(r)})$.

Таким образом, переходные вероятности p_{rs} , т.е. вероятности $P^{(1)}(y^{(r)}, \{y^{(s)}\})$, для цепи Маркова с конечным числом состояний приближенно могут быть определены по следующей формуле:

$$(9) \quad p_{rs} = \begin{cases} g(y^{(r)}) + \varepsilon^m h(y^{(r)}, y^{(r)}) & \text{при } r = s \\ \varepsilon^m h(y^{(r)}, y^{(s)}) & \text{при } r \neq s. \end{cases}$$

Чтобы найти уравнения, которым приближенно удовлетворяют вероятности (8), вернемся к рассмотрению исходной цепи Маркова и воспользуемся формулой (5) при $A = A_s$ (свойство инвариантности для исходной цепи Маркова установлено!). При $A = A_s$ левая часть в формуле (5) равна π_s . Для правой части формулы (5) с учетом (2) и аппроксимации $\pi_r = \varepsilon^m f(y^{(r)})$ может быть записана следующая цепочка приближенных равенств:

$$\begin{aligned} & \int_{R^m} \left(g(x) I_{A_s}(x) + \int_{A_s} h(x, y) dy \right) f(x) dx = \\ & = \int_{A_s} g(x) f(x) dx + \int_{A_s} \int_{R^m} h(x, y) f(x) dx dy = \\ & = \varepsilon^m g(y^{(s)}) f(y^{(s)}) + \int_{A_s} \sum_{r=1}^M h(y^{(r)}, y) \varepsilon^m f(y^{(r)}) dy = \\ & = g(y^{(s)}) \pi_s + \varepsilon^m \sum_{r=1}^M h(y^{(r)}, y^{(s)}) \pi_r = \sum_{r=1}^M p_{rs} \pi_r. \end{aligned}$$

Итак, приближенно, для вероятностей π_r получена следующая система линейных уравнений:

$$(10) \quad \sum_{r=1}^M \pi_r = 1, \quad \pi_s = \sum_{r=1}^M p_{rs} \pi_r \quad \text{при } s = 1, \dots, M.$$

Если все переходные вероятности p_{rs} для цепи Маркова с конечным числом состояний положительны, то системе уравнений (10) удовлетворяют инвариантные вероятности π_s такие, что при любых r и s

$$P^{(n)}(y^{(r)}, \{y^{(s)}\}) \rightarrow \pi_s \quad \text{при } n \rightarrow \infty$$

(см., например, [1, гл. 6, теорема 6.1]). При условии положительности всех переходных вероятностей p_{rs} система линейных уравнений (10) имеет единственное решение. И если вероятности (8) этой системе линейных уравнений удовлетворяют, то они являются инвариантными для цепи Маркова с конечным числом состояний.

Это и требуется, но следует помнить, что соотношения, о которых говорится выше, имеют место не в точности, а приближенно.

Если обоснование алгоритма не является безупречно строгим (как, например, приведенное выше объяснение алгоритма Метрополиса), то на вопрос о пригодности или непригодности алгоритма, или насколько алгоритм хорош, ответ могут дать расчеты. В этой области математической науки велика роль вычислительного опыта.

Положительность всех переходных вероятностей p_{rs} , задаваемых формулой (9), обеспечивается условием $c(x, y) > 0$ при любых $x, y \in B$. Однако практически функцию $c(x, y)$ можно брать с некоторым «запасом», чтобы условие $c(x, y) > 0$ выполнялось и при каких-то $y \notin B$, если это помогает в проведении шага 1.

Можно указать на аналогию между парой «исходная цепь Маркова и аппроксимирующая ее цепь Маркова с конечным числом состояний» и парой «дифферен-

циальное уравнение и разностная схема». Однако эта аналогия не является полной. В данном случае для расчета используется исходная цепь Маркова, а не цепь Маркова с конечным числом состояний.

Пример 1. Пусть f – функция плотности двумерного нормального распределения

$$(11) \quad f(x) = \frac{1}{2\pi(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)' \Sigma^{-1}(x-\mu)\right),$$

где

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

(см., например, [1, гл. 1, § 6]). Штрих означает транспонирование.

Разумеется, чтобы построить набор векторов, отвечающих двумерному нормальному распределению, нет необходимости использовать метод Монте-Карло с цепью Маркова. Достаточно построить набор векторов $z^{(0)}, z^{(1)}, \dots, z^{(n)}, \dots$, отвечающих двумерному нормальному распределению с нулевым средним и единичной ковариационной матрицей. (Задача построения таких векторов сводится к задаче построения стандартных нормальных случайных чисел. Различные подходы к построению нормальных случайных чисел рассматриваются, например, в [1, гл. 4, § 2].) Затем, взяв матрицу

$$L = \begin{pmatrix} \sqrt{1-\rho^2}\sigma_1 & 0 \\ \rho\sigma_1 & \sigma_2 \end{pmatrix},$$

для которой выполняется соотношение $\Sigma = L'L$, надо положить

$$(12) \quad x^{(n)} = \mu + L' z^{(n)}, \quad n = 1, 2, \dots$$

Но мы используем функцию плотности (11), чтобы проиллюстрировать алгоритм Метрополиса.

При каждом $a \geq 0$ на плоскости R^2 с координатами x_1, x_2 можно рассмотреть эллипс рассеивания

$$(x-\mu)' \Sigma^{-1}(x-\mu) = a^2,$$

где $x = (x_1, x_2)'$. Тогда вероятность того, что двумерный нормальный вектор с функцией плотности (11) принимает значение внутри этого эллипса, равна

$$(13) \quad 1 - \exp(-0,5a^2)$$

(см., например, [1, гл. 1, § 6]).

В расчете принимаются значения $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 0,8$, $\sigma_2 = 1,2$, $\rho = 0,9$. Вероятность p_1 рассчитывается при каждом a по формуле (13).

Во-первых, при различных N по формуле (12) строится набор точек $x^{(1)}, \dots, x^{(N)}$. Через p_2 обозначается доля точек из этого набора, попавших внутрь эллипса рассеивания.

Во-вторых, при помощи алгоритма Метрополиса реализуется цепь Маркова. Используется функция $s(x, y) = f_0(y - x)$, где f_0 – функция плотности двумерного нормального распределения с нулевым средним и с ковариационной матрицей

$$\begin{pmatrix} 0,36 & 0 \\ 0 & 0,36 \end{pmatrix}.$$

Принимается начальное значение $x^{(0)} = 0$, и первые 500 построенных точек игнорируются. Таким образом, рассматривается набор точек $x^{(500+1)}, \dots, x^{(500+N)}$. Через p_3 обозначается доля точек из этого набора, попавших внутрь эллипса рассеивания.

Результаты расчетов представлены в табл. 1.

Таблица 1.

Соответствие долей точек, попавших внутрь эллипсов рассеивания, построенных различными методами (p_2 – с использованием аналитической формулы (12), p_3 – с использованием алгоритма Метрополиса), и вероятностей p_1 попадания точки внутрь эллипса рассеивания, рассчитанных по формуле (13), при различных N

a	p_1	p_2	$p_1 - p_2$	p_3	$p_1 - p_3$
a) $N = 500$					
0,5	0,11750	0,09000	0,02750	0,12000	-0,00250
1,0	0,39347	0,37000	0,02347	0,45400	-0,06053
1,5	0,67535	0,66000	0,01535	0,73000	-0,05465
2,0	0,86466	0,87200	-0,00734	0,85200	0,01266
b) $N = 5000$					
0,5	0,11750	0,10900	0,00850	0,13640	-0,01890
1,0	0,39347	0,38060	0,01287	0,42780	-0,03433
1,5	0,67535	0,66960	0,00575	0,72160	-0,04625
2,0	0,86466	0,86040	0,00426	0,87600	-0,01134
c) $N = 50000$					
0,5	0,11750	0,11708	0,00042	0,12728	-0,00978
1,0	0,39347	0,39000	0,00347	0,39866	-0,00519
1,5	0,67535	0,67392	0,00143	0,68304	-0,00769
2,0	0,86466	0,86232	0,00234	0,86420	0,00046

Окончание табл. 1.

a	p_1	p_2	$p_1 - p_2$	p_3	$p_1 - p_3$
d) $N = 500000$					
0,5	0,11750	0,11722	0,00028	0,11779	-0,00029
1,0	0,39347	0,39268	0,00079	0,39470	-0,00123
1,5	0,67535	0,67561	-0,00027	0,67827	-0,00292
2,0	0,86466	0,86448	0,00018	0,86646	-0,00179

Из табл. 1 видно, что, как и следовало ожидать, построенная с помощью аналитической формулы (12) величина p_2 лучше аппроксимирует вероятность p_1 , чем построенная с помощью численного алгоритма величина p_3 . Однако и для алгоритма Метрополиса аппроксимация существенно улучшается при увеличении N .

Пример 2. Пусть функция плотности имеет вид

$$f(x) = 0,5f^{(1)}(x) + 0,5f^{(2)}(x),$$

где $f^{(1)}(x)$ и $f^{(2)}(x)$ – функции плотности двумерного нормального распределения (см. (11)) соответственно со средними $\mu^{(1)}$, $\mu^{(2)}$ и с ковариационными матрицами $\Sigma^{(1)}$, $\Sigma^{(2)}$. В расчете принимаются значения $\mu_1^{(1)} = -0,3$, $\mu_2^{(1)} = 0,5$, $\sigma_1^{(1)} = 0,7$, $\sigma_2^{(1)} = 1,1$, $\rho^{(1)} = 0,8$, $\mu_1^{(2)} = 0,8$, $\mu_2^{(2)} = -0,2$, $\sigma_1^{(2)} = 0,9$, $\sigma_2^{(2)} = 1,0$, $\rho^{(2)} = -0,1$. Для построения набора векторов, отвечающих распределению вероятностей с функцией плотности f , используется алгоритм Метрополиса. Берется та же функция $c(x, y)$, что и в примере 1, принимается начальное значение $x^{(0)} = 0$, и первые 500 построенных точек игнорируются.

Полное исследование того, насколько хорошо по набору точек $x^{(500+1)}, \dots, x^{(500+N)}$ можно судить о свойствах данного распределения, мы не проводим. Сравним лишь

$$p_1 = \int_0^1 \int_0^1 f(x_1, x_2) dx_1 dx_2$$

и долю точек из рассматриваемого набора, попавших в квадрат $[0,1] \times [0,1]$. Эту долю обозначим через p_2 . Применение кубатурной формулы с достаточно мелкой сеткой показывает, что p_1 приближенно равно 0,10614. Полученные в расчете значения p_2 при различных N приводятся в табл. 2.

Хотя самая высокая точность в данном расчете получена при $N = 5000$, расчеты с другими наборами исходных случайных чисел не дают повторения этого эффекта. Однако во всех расчетах видна примерно одна и та же тенденция к улучшению точности при увеличении N .

Таблица 2.

Соответствие долей точек p_2 , попавших внутрь квадрата $[0,1] \times [0,1]$, построенных с использованием алгоритма Метрополиса, и вероятности p_1 попадания точки внутрь данного квадрата, рассчитанной по кубатурной формуле с большим числом узлов

N	p_2	$p_1 - p_2$
500	0,08200	0,02414
5000	0,10600	0,00014
50000	0,10528	0,00086
500000	0,10640	-0,00026

3. Гиббсовский выбор

Здесь описывается еще один алгоритм построения набора m -мерных векторов, отвечающих распределению вероятностей с функцией плотности f . Название «гиббсовский выбор» для данного алгоритма предложено в статье [7], поскольку используется он в этой работе для распределения, носящего имя выдающегося физика конца XIX – начала XX вв. Дж.У. Гиббса. Данное название закрепилось после того, как этот алгоритм был представлен в работе [6]. О более ранних работах, где используются сходные алгоритмы, см. [5, с. 20] или [9, с. 12].

Пусть $x = (x_1, \dots, x_k)$, где $x_i \in R^{m_i}$, $m_1 + \dots + m_k = m$, и пусть множество B имеет вид

$$B = B_1 \times \dots \times B_k,$$

где $B_i \subseteq R^{m_i}$ – борелевское множество положительной меры.

При $i = 1, \dots, k$ рассмотрим маргинальную функцию плотности

$$f_i(x_1, \dots, x_k) = \int_{B_i} f(x_1, \dots, x_{i-1}, q_i, x_{i+1}, \dots, x_k) dq_i$$

(при $i=1$ отсутствуют аргументы x_1, \dots, x_{i-1} ; при $i=k$ отсутствуют аргументы x_{i+1}, \dots, x_k). Фактически f_i является функцией аргументов $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k$.

Тогда при $x \in B$ можно определить условную функцию плотности $g_i(x)$ соотношением

$$f(x) = g_i(x) f_i(x), \quad i = 1, \dots, k.$$

Используя распространенное обозначение, можно сказать, что $g_i(x_1, \dots, x_k)$ – это условная плотность

$$x_i \mid x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k$$

(см., например, [1, гл. 1, § 3]). Будем считать, что $g_i(x) = 0$ при $x \notin B$.

Цепь Маркова $X_0, X_1, \dots, X_n, \dots$ ищется среди тех, для которых вероятности перехода за один шаг для любого $x \in B$ и для любого борелевского множества $A \subseteq R^m$ имеют вид

$$P^{(1)}(x, A) = \int_A h(x, y) dy$$

(ср. (2)). Положим

$$(14) \quad h(x, y) = \prod_{i=1}^k g_i(y_1, \dots, y_i, x_{i+1}, \dots, x_k)$$

и проверим, что выполняется условие инвариантности (5).

Имеем

$$f(y_1, \dots, y_{i-1}, x_i, x_{i+1}, \dots, x_k) = g_i(y_1, \dots, y_{i-1}, x_i, x_{i+1}, \dots, x_k) f_i(y_1, \dots, y_{i-1}, x_i, x_{i+1}, \dots, x_k)$$

и аналогично

$$f(y_1, \dots, y_{i-1}, y_i, x_{i+1}, \dots, x_k) = g_i(y_1, \dots, y_{i-1}, y_i, x_{i+1}, \dots, x_k) f_i(y_1, \dots, y_{i-1}, y_i, x_{i+1}, \dots, x_k).$$

Тогда при условии, что аргументы функций f , g_i и f_i являются точками множества B

$$\frac{f(y_1, \dots, y_{i-1}, x_i, x_{i+1}, \dots, x_k)}{g_i(y_1, \dots, y_{i-1}, x_i, x_{i+1}, \dots, x_k)} = \frac{f(y_1, \dots, y_{i-1}, y_i, x_{i+1}, \dots, x_k)}{g_i(y_1, \dots, y_{i-1}, y_i, x_{i+1}, \dots, x_k)}.$$

Отсюда следует, что

$$(15) \quad \begin{aligned} & g_i(y_1, \dots, y_{i-1}, y_i, x_{i+1}, \dots, x_k) f(y_1, \dots, y_{i-1}, x_i, x_{i+1}, \dots, x_k) = \\ & = g_i(y_1, \dots, y_{i-1}, x_i, x_{i+1}, \dots, x_k) f(y_1, \dots, y_{i-1}, y_i, x_{i+1}, \dots, x_k). \end{aligned}$$

Данный прием исключения маргинальных плотностей с целью установить связь между условными и совместными плотностями, по-видимому, впервые был использован в [4, с. 195, 196]. Из формулы (15) нетрудно увидеть, что двум различным совместным функциям плотности $f^{(1)}$ и $f^{(2)}$ (при одном и том же множестве B) не может соответствовать один набор условных функций плотности g_1, \dots, g_k .

Имеем

$$\int_B P^{(1)}(x, A) f(x) dx = \int_B \int_A f(x) h(x, y) dy dx = \int_A J(y) dy,$$

где

$$J(y) = \int_B f(x) h(x, y) dx.$$

Из выражения (14) для $h(x, y)$ находим

$$J(y) = \int_{B_k} g_k(y_1, \dots, y_k) \int_{B_{k-1}} g_{k-1}(y_1, \dots, y_{k-1}, x_k) \dots \int_{B_1} g_1(y_1, x_2, \dots, x_k) f(x_1, \dots, x_k) dx_1 \dots dx_k.$$

Множитель $g_i(y_1, \dots, y_{i-1}, y_i, x_{i+1}, \dots, x_k)$ намеренно не выносится из под знака интеграла по B_i . На основании (15) подынтегральная функция самого внутреннего интеграла равна

$$g_1(x_1, x_2, \dots, x_k) f(y_1, x_2, \dots, x_k).$$

Теперь второй множитель можно вынести из под знака интеграла по B_1 и воспользоваться тем, что

$$\int_{B_1} g_1(x_1, x_2, \dots, x_k) dx_1 = 1.$$

Тогда

$$J(y) = \int_{B_k} g_k(y_1, \dots, y_k) \int_{B_{k-1}} g_{k-1}(y_1, \dots, y_{k-1}, x_k) \dots \int_{B_2} g_2(y_1, y_2, x_3, \dots, x_k) f(y_1, x_2, \dots, x_k) dx_2 \dots dx_k.$$

На основании (15) подынтегральная функция самого внутреннего интеграла равна

$$g_2(y_1, x_2, x_3, \dots, x_k) f(y_1, y_2, x_3, \dots, x_k).$$

Поэтому данный интеграл равен $f(y_1, y_2, x_3, \dots, x_k)$. Проведя аналогичные преобразования и для других подынтегральных функций, получаем $J(y) = f(y)$. Свойство инвариантности (5) установлено.

Как и в предыдущем разделе, реализацию цепи Маркова $X_0, X_1, \dots, X_n, \dots$ обозначим $x^{(0)}, x^{(1)}, \dots, x^{(n)}, \dots$. Пусть как-то выбран вектор $x^{(0)} \in B$. При гиббсовском выборе вектор $x^{(n+1)}$ строится по вектору $x^{(n)}$ следующим образом.

Шаг 1. Симулируется m_1 -мерный вектор y_1 из распределения с функцией плотности $g_1(y_1, x_2^{(n)}, \dots, x_k^{(n)})$.

⋮

Шаг i . Симулируется m_i -мерный вектор y_i из распределения с функцией плотности $g_i(y_1, \dots, y_{i-1}, y_i, x_{i+1}^{(n)}, \dots, x_k^{(n)})$.

⋮

Шаг k . Симулируется m_k -мерный вектор y_k из распределения с функцией плотности $g_k(y_1, \dots, y_{k-1}, y_k)$.

Принимается $x^{(n+1)} = (y_1, \dots, y_k)$.

Интуитивное объяснение гиббсовского выбора аналогично объяснению алгоритма Метрополиса.

Пример 3. Рассмотрим ту же функцию плотности двумерного нормального распределения (11), что и в примере 1. Условные плотности имеют вид

$$g_1(x_1, x_2) = \frac{1}{\sigma_{1,c}\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{1,c}^2}(x_1 - \mu_{1,c})^2\right),$$

$$g_2(x_1, x_2) = \frac{1}{\sigma_{2,c}\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{2,c}^2}(x_2 - \mu_{2,c})^2\right),$$

где

$$\sigma_{1,c} = \sigma_1\sqrt{1-\rho^2}, \quad \sigma_{2,c} = \sigma_2\sqrt{1-\rho^2},$$

$$\mu_{1,c} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \quad \mu_{2,c} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$$

(см., например, [1, гл. 1, § 6]).

Вектора $x^{(1)}, \dots, x^{(n)}, \dots$ строятся по следующим формулам. Пусть $z_1^{(1)}, \dots, z_1^{(n)}, \dots, z_2^{(1)}, \dots, z_2^{(n)}, \dots$ – стандартные нормальные случайные числа. Тогда

$$(16) \quad \begin{aligned} x_1^{(n+1)} &= \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2^{(n)} - \mu_2) + \sigma_{1,c}z_1^{(n+1)} \\ x_2^{(n+1)} &= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1^{(n+1)} - \mu_1) + \sigma_{2,c}z_2^{(n+1)}. \end{aligned}$$

Интересно отметить, что отсюда следуют уравнения авторегрессии первого порядка

$$x_1^{(n+1)} = \mu_1(1-\rho^2) + \rho^2x_1^{(n)} + \varepsilon_1^{(n+1)}, \quad x_2^{(n+1)} = \mu_2(1-\rho^2) + \rho^2x_2^{(n)} + \varepsilon_2^{(n+1)}.$$

Выражения для $\varepsilon_1^{(n+1)}$ и $\varepsilon_2^{(n+1)}$ легко выводятся из (16).

Как и в примере 1, принимается начальное значение $x^{(0)} = 0$, и первые 500 построенных точек игнорируются. Таким образом, рассматривается набор точек $x^{(500+1)}, \dots, x^{(500+N)}$. Через p_4 обозначается доля точек из этого набора, попавших внутрь эллипса рассеивания. Через p_1 обозначается та же вероятность, что и в примере 1.

Результаты расчетов представлены в табл. 3.

Точность, с которой величина p_4 аппроксимирует p_1 , улучшается с увеличением N . В расчетах, результаты которых приведены в данном примере и в примере 1, гиббсовский выбор дает более высокую точность, чем алгоритм Метрополиса. Гиббсовский выбор дает точность, которая близка к точности результатов, полученных при использовании аналитической формулы (12). Это можно объяснить тем, что в

данном примере гиббсовский выбор сводится к применению аналитических формул (16).

Таблица 3.

Соответствие долей точек p_4 , попавших внутрь эллипсов рассеивания, построенных с использованием гиббсовского выбора, и вероятностей p_1 попадания точки внутрь эллипса рассеивания, рассчитанных по формуле (13), при различных N

a	p_1	p_4	$p_1 - p_4$
a) $N = 500$			
0,5	0,11750	0,12200	-0,00450
1,0	0,39347	0,39600	-0,00253
1,5	0,67535	0,69800	-0,02265
2,0	0,86466	0,88200	-0,01734
b) $N = 5000$			
0,5	0,11750	0,12040	-0,00290
1,0	0,39347	0,39540	-0,00193
1,5	0,67535	0,67600	-0,00065
2,0	0,86466	0,87000	-0,00534
c) $N = 50000$			
0,5	0,11750	0,11768	-0,00018
1,0	0,39347	0,38888	0,00459
1,5	0,67535	0,66866	0,00669
2,0	0,86466	0,86064	0,00402
d) $N = 500000$			
0,5	0,11750	0,11741	0,00010
1,0	0,39347	0,39379	-0,00032
1,5	0,67535	0,67486	0,00048
2,0	0,86466	0,86445	0,00021

Пример 4. Пусть $m = 2$, множество B состоит из тех точек $x = (x_1, x_2)$, для которых $x_1 \geq 0$, $x_2 \geq 0$. Совместная функция плотности имеет вид

$$f(x_1, x_2) = C \exp(-x_1 x_2 - x_1 - x_2),$$

где C – нормирующая константа. В этом примере гиббсовский выбор используется для расчета маргинальной функции плотности.

При $m = 2$ удобнее использовать другие обозначения для маргинальных функций плотности, чем те, которые применялись выше. Пусть

$$f_1(x_1) = \int_0^{\infty} f(x_1, x_2) dx_2, \quad f_2(x_2) = \int_0^{\infty} f(x_1, x_2) dx_1.$$

Из вида $f(x_1, x_2)$ следует, что

$$(17) \quad f_1(x_1) = \frac{C}{x_1 + 1} \exp(-x_1),$$

и аналогичная формула имеет место для $f_2(x_2)$.

Тогда, как нетрудно увидеть, условные функции плотности

$$g_1(x_1, x_2) = (x_2 + 1) \exp(-(x_2 + 1)x_1), \quad g_2(x_1, x_2) = (x_1 + 1) \exp(-(x_1 + 1)x_2)$$

являются функциями плотности экспоненциального распределения.

Для определения значения функции f_1 при некотором фиксированном x_1 по набору точек $x^{(500+1)}, \dots, x^{(500+N)}$, построенных при помощи гиббсовского выбора (как и в предыдущих примерах первые 500 точек игнорируются), используется следующее выражение:

$$(18) \quad \hat{f}_1(x_1) = \frac{1}{N} \sum_{n=1}^N g_1(x_1, x_2^{(500+n)}).$$

Поскольку

$$f_1(x_1) = \int_0^{\infty} g_1(x_1, x_2) f_2(x_2) dx_2,$$

а правая часть в (18) – это приближенное выражение для правой части последнего равенства. Преимущества такого подхода к нахождению маргинальной функции плотности перед прямым подходом, когда расчеты производились бы по $x_1^{(500+1)}, \dots, x_1^{(500+N)}$, объясняются, например, в [6].

В табл. 4 при различных N показаны значения

$$\max_{x_1 \in Q} |f_1(x_1) - \hat{f}_1(x_1)|,$$

где множество Q состоит из точек $0,04 \cdot j$ при $j = 1, \dots, 100$; функция f_1 определяется по формуле (17), а функция \hat{f}_1 – по формуле (18); принимается начальное значение $x^{(0)} = (0,5, 0,5)$.

Расчеты показывают, что точность аппроксимации улучшается с увеличением N .

Таблица 4.

Точность расчета маргинальной функции плотности
при использовании гиббсовского выбора

$N = 500$	$N = 5000$	$N = 50000$	$N = 500000$
0,01875	0,00554	0,00360	0,00024

* *

*

СПИСОК ЛИТЕРАТУРЫ

1. Шведов А.С. Теория вероятностей и математическая статистика – 2 (промежуточный уровень). М.: Изд. дом ГУ ВШЭ, 2007.
2. Athreya K.B., Doss H., Sethuraman J. On the Convergence of the Markov Chain Simulation Method // Annals of Statistics. 1996. 24. P. 69–100.
3. Barker A.A. Monte Carlo Calculations of the Radial Distribution Functions for a Proton–electron Plasma // Australian Journal of Physics. 1965. 18. P. 119–133.
4. Besag J. Spatial Interaction and Statistical Analysis of Lattice Systems // Journal of the Royal Statistical Society. Series B. 1974. 36. P. 192–225.
5. Chen M.-H., Shao Q.-M., Ibragim J.G. Monte Carlo Methods in Bayesian Computation. Springer, 2000.
6. Gelfand A.E., Smith A.F.M. Sampling-based Approaches to Calculating Marginal Densities // Journal of the American Statistical Association. 1990. 85. P. 398–409.
7. Geman S., Geman D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984. 6. P. 721–741.
8. Hastings W.K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications // Biometrika. 1970. 57. P. 97–109.
9. Markov Chain Monte Carlo in Practice / W.R. Gilks, W.R.S. Richardson, D. Spiegelhalter (eds.) L.: Chapman and Hall, 1996.
10. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E. Equation of State Calculations by Fast Computing Machines // Journal of Chemical Physics. 1953. 21. P. 1087–1092.
11. Neal P., Rao T.S. MCMC for Integer-valued ARMA Processes // Journal of Time Series Analysis. 2007. 28. P. 92–110.
12. Peskun P.N. Optimum Monte Carlo Sampling Using Markov Chains // Biometrika. 1973. 60. P. 607–612.