

Часть I Параметрически разделимые алгоритмы (С.М. Ермаков)

Введение

Хорошо известно, что эффективное использование современных суперкомпьютеров налагает определенные требования на выбор алгоритмов решения задач. Разумеется, возможен подход, при котором любой алгоритм подвергается анализу, осуществляемому программистом или транслятором, и имеющиеся параллельные структуры используются при вычислениях. Такой подход, однако, в большинстве случаев приводит к недостаточно эффективному использованию оборудования. При этом часто оказывается, что алгоритм, обладающий худшими свойствами с точки зрения последовательных вычислительных структур Дж. фон Неймана, оказывается более эффективным при использовании многопроцессорной техники. Очевидно, представляет интерес изучение “плохих” (в указанном выше смысле) алгоритмов, способных эффективно загрузить многопроцессорное оборудование, и разработка новых алгоритмов такого сорта. При этом, очевидно, наиболее интересны “не очень плохие” алгоритмы.

Далее мы рассмотрим один из классов алгоритмов, способных эффективно загрузить систему с распределенной памятью, состоящую из центрального (управляющего) процессора и k процессоров, обладающих автономной памятью. Кластеры и облака могут служить примерами таких систем. Предполагается, что время обмена между процессорами существенно больше (на несколько порядков) времени выполнения операций каждым процессором. В связи с этим последним предположением заметим следующее.

Пусть оптимальный последовательный (требующий минимального среднего числа операций) алгоритм решения некоторой задачи требует времени T_0 . В качестве альтернативы мы можем использовать алгоритм, требующий для исполнения в последовательном варианте времени T , $T > T_0$, и привлечь k процессоров, каждый из которых будет занят время T/k . При этом потребуется дополнительное время t на обмен информацией в процессе решения задачи, t может быть значительным.

Очевидным требованием является выполнение неравенства

$$T_0 \geq \frac{T}{k} + t,$$

то есть альтернативный алгоритм должен обладать свойством

$$T \leq k(T_0 - t). \quad (1)$$

Ясно, что должно быть $t < T_0$. В противном случае многопроцессорность не улучшает дело. Мы видим также, что рост k и уменьшение t позволяет привлечь к решению задачи широкий круг алгоритмов со свойствами, далекими от оптимальности в традиционном смысле. Алгоритмы могут быть “плохими”, но иметь параллельную структуру.

К числу алгоритмов с параллельной структурой и редкими обходами между процессорами относятся параметрически разделимые ПР-алгоритмы [1].

ПР-алгоритм состоит из трёх частей, выполняемых последовательно

$$B_1 \rightarrow A(\theta) \rightarrow B_2,$$

и обладает следующими свойствами:

- 1) Алгоритм $A(\theta)$ зависит от параметра θ (имеет θ в качестве входных данных). θ принимает значения из дискретного множества Θ достаточно общей природы.
- 2) Выполнение алгоритмов $A(\theta_i)$, $\theta_i \in \Theta$, $i = 1, \dots, K$ может быть поручено k -независимым процессорам.

- 3) Алгоритм B_1 сообщает задания алгоритму A , а B_2 вычисляет решение исходной задачи, имея результаты, полученные $A(\theta_i)$, $i = 1, \dots, K$.

Если времена выполнения алгоритмов $A(\theta_i)$ при различных i разнятся незначительно, то ПР-алгоритм называется *однородным*.

В качестве одного из простейших примеров ПР-алгоритмов можно сослаться на ВВР-алгоритмы [2]. Алгоритм такого типа вычисляет (например) независимо каждый из разрядов числа π . Здесь $\Theta = \mathbb{N}$ — множество натуральных чисел. Вычислительная система из k процессоров вычисляет k разрядов или k групп разрядов независимо, результаты поступают в один из процессоров (обмен), и формируется приближенное с заданной точностью значение числа π .

Многие алгоритмы вычислительной математики являются ПР-алгоритмами. Так, если вычисление значений $f(X)$, где $X \in D \subset \mathbf{R}^s$ связано со значительными трудностями, то алгоритм построения интерполяционного многочлена $P(X) = \sum_{i=1}^n l_i(X)f(X_i)$, $l_i(X)$ — многочлены, $X_i \in D$, и вычисления интеграла с помощью кубатурной суммы

$$K_n[f] = \sum_{i=1}^n A_i f(X_i) \quad (2)$$

являются ПР-алгоритмами, где совокупность точек X_i составляют параметрическое множество Θ .

Очевидно, изучение ПР-алгоритмов представляет значительный интерес в свете основных тенденций развития современной вычислительной техники. Таковыми являются — создание компьютеров с параллельной структурой на базе большого числа процессоров, использование дополнительных вычислительных устройств типа SIMD (видеокарты, например) и другие.

Далее мы рассмотрим некоторые типы (ПР)-алгоритмов, тесно связанные с методами численного интегрирования. Оказывается, что класс таких алгоритмов очень широк. Это, в первую очередь, алгоритмы решения задач, приводимых к вычислению интегралов по траекториям. Представление решения в виде интеграла часто требует специальных исследований. Поскольку математические ожидания случайных величин являются интегралами по вероятностным мерам, аппарат теории вероятностей находит естественное применение в этих исследованиях. Вычислительным аппаратом являются метод Монте-Карло и кубатурные формулы для интегрирования функций малой гладкости.

В первую очередь нами обсуждается метод Монте-Карло. Метод широко используется для решения задач, имеющих вероятностную трактовку. В соответствии с развиваемой нами концепцией мы приводим новые результаты относительно расширения сферы его применения — вычисления интегралов в смысле обобщенного главного значения, уменьшения дисперсии в знакопеременном случае и др. Значительное место уделяется вопросам стохастической устойчивости, её роли при распараллеливании алгоритмов решения нелинейных задач, возможности использования детерминированных последовательностей (квази Монте-Карло).

В качестве приложений рассмотрены алгоритмы решения разностных аналогов уравнений Навье-Стокса, задачи с подвижной границей для расчёта цены американского опциона, алгоритмы решения некоторых экстремальных задач (построение D-оптимальных планов регрессионных экспериментов). Вторая часть монографии посвящена вероятностным методам решения краевых задач математической физики.

Глава 1

Моделирование распределений

1.1 Параметрическое множество метода Монте-Карло

Как мы уже отмечали, параметрическим множеством метода Монте-Карло является совокупность псевдослучайных чисел

(как правило, физические датчики случайных чисел при решении задач вычислительной математики не используются). Элементы множества могут быть вычислены с помощью одной из рекуррентных процедур. Наиболее исследованной является процедура вида

$$Z_{n+1} \equiv MZ_n + C, \pmod{P}, \quad \alpha_n = \frac{Z_n}{P}, \quad n = 0, 1, 2, \dots, \quad (1.1.1)$$

где M , C и Z_0 — заданные целые положительные числа.

Последовательность Z_n очевидным образом периодична. Если L — длина периода, то это означает, что

$$Z_0 \equiv MZ_L + C, \pmod{P}.$$

Параметрическое множество, таким образом, состоит из $L+1$ числа $\alpha_0, \dots, \alpha_{L-1}$. Числа α_n являются рациональными P -ичными дробями.

Понятно, что последовательность (1.1.1) вполне детерминирована, но при $P \rightarrow \infty$ оказывается, что она становится для почти всех α_0 последовательностью дробных долей показательной функции

$$\alpha_{n+1} = \{M\alpha_n + C\}, \quad (1.1.2)$$

а эта последовательность обладает рядом свойств, присущих реализациям равномерно распределённых на отрезке $(0, 1)$ случайных величин. Здесь $\{a\}$ обозначает дробную долю числа a , а C и все α_n являются вещественными числами.

Напомним, что для независимых реализаций случайной величины $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, равномерно распределённых на промежутке $[0, 1]$, имеют место равенства

$$\frac{1}{N} \sum_{l=0}^{N-1} f(Y_l) \xrightarrow[N \rightarrow \infty]{P=1} \int_{D_s}^1 f(X) dX, \quad X = (x_1, \dots, x_s), \quad Y_l = (\varepsilon_{sl+1}, \dots, \varepsilon_{(s+1)l}), \quad (1.1.3)$$

f — любая измеримая функция s переменных, D_s — единичный гиперкуб, $s = 1, 2, \dots$ — любое, и сходимость к пределу происходит с вероятностью 1.

Также имеет место частный случай Центральной предельной теоремы

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{N}}{\sigma} \left| \int_0^1 f(X) dX - \frac{1}{N} \sum_{l=0}^N f(Y_l) \right| < y\right) = \sqrt{\frac{2}{\pi}} \int_0^y e^{-\frac{u^2}{2}} du. \quad (1.1.4)$$

Очевидно, мы не можем найти детерминированную последовательность $\alpha_1, \alpha_2, \dots, \alpha_n$, удовлетворяющую аналогичным соотношениям. Достаточно заметить, что измеримая f может быть не определена на множестве меры нуль, а всякая счётная последовательность имеет лебегову меру нуль. То есть наверняка существует f , для которой (1.1.3) не выполнимо при любом выборе $\alpha_1, \dots, \alpha_n$.

Обоснование вероятностных свойств последовательности (1.1.1) возможно с помощью метрических теорем теории чисел. Мы воспользуемся теоремой [3], которая гласит.

Теорема 1 Пусть $a_1, a_2, \dots$ — последовательность различных целых чисел, $x \in (0, 1)$ и

$$\varepsilon_n = \{a_n x\}, \quad n = 1, 2, \dots, \quad (1.1.5)$$

тогда для любой интегрируемой по Риману на $[0, 1]$ функции $f(x)$ и почти всех x из $(0, 1)$ справедливо равенство

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(\varepsilon_n) = \int_0^1 f(x) dx \quad (1.1.6)$$

□.

Последовательности, для которых выполняется (1.1.6), называют равномерно распределёнными в смысле Вейля.

Также справедлива теорема [4].

Теорема 2 (Лемма Сю) Пусть $x \in (0, 1)$,

$$\alpha_n = \{M^n x\}, \quad n = 0, 1, \dots, \quad (1.1.7)$$

$f(X)$, $X = (x_1, \dots, x_s)$ — интегрируемая по Риману в единичном гиперкубе D_s функция и

$$\varphi_M(x) = f(x, \{Mx\}, \dots, \{M^{s-1}x\}),$$

тогда

$$\lim_{M \rightarrow \infty} \int_0^1 \varphi_M(x) dx = \int_{D_s} f(X) dX \quad (1.1.8)$$

□.

(1.1.7) называют последовательностью дробных долей показательной функции.

Составим из элементов последовательности (1.1.7) векторы

$$Y_l = (\alpha_{sl+1}, \dots, \alpha_{(s+1)l}). \quad (1.1.9)$$

Из (1.1.6) и (1.1.8) легко получить следующее утверждение [5].

Теорема 3 Для любого конечного s и $f(X)$, интегрируемой по Риману в D_s (и почти всех x из $(0, 1)$), справедливо предельное равенство

$$\lim_{M \rightarrow \infty} \left[\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^N f(Y_l) \right] = \int_{D_s} f(X) dX \quad (1.1.10)$$

□.

Таким образом, последовательность (1.1.9) асимптотически равномерно распределена в смысле Вейля при любом s . Формулы (1.1.5) сходятся с (1.1.1) и являются их предельным вариантом при $P \rightarrow \infty$. Однако, при конечном P , когда α_0 — рациональная дробь, существуют функции f , для которых (1.1.10) не выполняется.

Действительно, пусть $P = 2^m$, $\alpha = \bar{\alpha}$ рационально и $f(x) = \tilde{f}(X) = \exp 2\pi i(k_1 x_1 + k_2 x_2 + \dots + k_s x_s)$, где $k_1, \dots, k_s$ — заданные числа, не равные нулю одновременно. Интеграл в правой части равенства (1.1.3) для этой f равен нулю, а l -тое слагаемое суммы в левой части есть

$$\begin{aligned} & \tilde{f}(\bar{\alpha}_{sl+1}, \dots, \bar{\alpha}_{(s+1)l}) = \\ & = \exp 2\pi i(k_1 M^{sl+1} + k_2 M^{sl+2} + \dots + k_s M^{(s+1)l})\alpha_0 = \\ & = \exp 2\pi i M^{sl+1}(k_1 + k_2 M + \dots + k_s M^{s-l})\frac{Z_0}{P}, \quad (\text{здесь } \bar{\alpha}_{n+1} = \{M\bar{\alpha}_n\}) \end{aligned} \quad (1.1.11)$$

и, если окажется, что

$$k_1 + k_2 M + \dots + k_s M^{s-l} \equiv 0 \pmod{P}, \quad (1.1.12)$$

то сумма $\frac{1}{N} \sum_{l=1}^N f(Y_l)$ будет равна единице.

При каждом конечном P легко указать такие наборы $(k_1, \dots, k_s)$, что сравнение (1.1.12) будет иметь место, но с ростом P класс “плохих” функций будет сужаться.

Обозначим $\mathcal{K} = \{\vec{k} : \|\vec{k}\| \leq L\}$, где $\vec{k} = (k_1, \dots, k_s)$ удовлетворяет (1.1.11), $\|\vec{k}\|$ обозначает некоторую норму $\vec{k}$, L — заданная константа. Множество $\mathcal{K}$ есть множество тригонометрических многочленов степени L от s переменных.

Справедлива следующая лемма.

Лемма 1 Для заданного $L > 0$ существуют такие M и P , что множество $\mathcal{K}$ пусто.

Доказательство:

Сравнение

$$k_1 + k_2 M + \dots + k_s M^{s-1} \equiv 0 \pmod{P},$$

равносильно множеству равенств

$$k_1 + k_2 M + \dots + k_s M^{s-1} = lP, \quad l = 0, \pm 1, \pm 2, \dots \quad (1.1.13)$$

В силу эквивалентности норм конечномерных векторов существует константа $L' > 0$, такая что $\max_i |k_i| \leq L'$, и при достаточно больших M значение многочлена относительно M в левой части (1.1.13) определяется членом $k_s M^{s-1}$. Таким образом, при $l = 0$ найдется такое $M_0 > 0$, что при $M > M_0$ многочлен не будет менять знак. С другой стороны, этот член ограничен величиной $L' M^{s-1}$, и при заданном s найдется такое достаточно большое $P = P(M)$, что многочлен

$$k_s M^{s-1} + \dots + k_2 M + k_1 - lP$$

не имеет вещественных корней при любом значении $l = \pm 1, \dots$ (напомним, что все $|k_i|$ ограничены!). Лемма доказана. ■

Таким образом, для любого фиксированного s существуют столь большие L , M и P , $M < P$, что равенство (1.1.10) выполняется для любой непрерывной функции $f(X)$. Последнее утверждение следует из теоремы Вейерштрасса о приближении непрерывных функций тригонометрическими многочленами.

Это означает также, что совместные распределения последовательностей векторов $(\bar{\alpha}_{sl+1}, \dots, \bar{\alpha}_{(s+1)l})$, $l = 0, 1, \dots$ становятся близкими к распределениям дробных долей показательной функции при соответствующем M и некоторых начальных значениях Z_0 , обеспечивающих максимальность периода последовательности (1.1.1). Подробнее об этом сказано в [6].

Из сказанного следует важность изучения поведения последовательностей дробных долей показательной функции при больших M .

Равенство (1.1.10) является детерминистским аналогом Закона больших чисел. В [7] был также указан аналог Центральной предельной теоремы для дробных долей показательной функции. Имеет место

Теорема 4 *Для любой интегрируемой по Риману в D_s $f(X)$ имеет место равенство*

$$\begin{aligned} \lim_{M \rightarrow \infty} \left(\lim_{N \rightarrow \infty} \text{mes} \left\{ \alpha_0 : \left| \frac{1}{N} \sum_{l=1}^N f(X_l) - \int_{D_s} f(X) dX \right| \cdot \frac{\sqrt{N}}{\sigma_f} < y \right\} \right) = \\ = \sqrt{\frac{2}{\pi}} \int_0^y e^{-\frac{u^2}{2}} du, \end{aligned} \quad (1.1.14)$$

$$\text{где } \sigma_f = \left[\int_{D_s} f^2(X) dX - \left(\int_{D_s} f(X) dX \right)^2 \right]^{1/2}.$$

Доказательство:

Доказательство основано на теореме 3.

Рассмотрим остаток (погрешность)

$$R_N(f, \alpha_0) = \int_{D_s} f(X) dX - \frac{1}{N} \sum_{l=1}^N f(Y_l(\varepsilon)) = \frac{1}{N} \sum_{l=1}^N \tilde{f}(Y_l(\varepsilon)),$$

где $\tilde{f}(X) = \int_{D_s} f(X) dX - f(X)$, и изучим предельное поведение $R_N(f, \alpha_0)$ при $N \rightarrow \infty$,

$M \rightarrow \infty$. Заметим, что $\text{mes} \{x : R_N(f, x) < y\}$ есть функция распределения случайной величины $R_N(f, x)$ при равномерно распределённом на $(0, 1)$ случайном x . Тогда моменты $R_N(f, x)$ равны

$$\begin{aligned} I_p(R_N(f, \varepsilon)) &= \frac{1}{N^p} \int_0^1 \left(\sum_{l=1}^N \tilde{f}(Y_l(\varepsilon)) \right)^p d\varepsilon = \\ &= \sum_{m_1+m_2+\dots+m_k=p} C_N^k \frac{p!}{m_1! \dots m_k!} \int_0^1 \tilde{f}^{m_1}(Y_{l_1}(\varepsilon)) \dots \tilde{f}^{m_k}(Y_{l_k}(\varepsilon)) d\varepsilon, \end{aligned} \quad (1.1.15)$$

где $l_1 \neq l_2 \neq \dots \neq l_k$, $0 < i_r \leq N$, $r = 1, \dots, k$ (обобщенная формула бинома).

На основании (1.1.10) имеем

$$\begin{aligned} \lim_{M \rightarrow \infty} I_p(R_N(f, \alpha_0)) &= \sum_{m_1 + \dots + m_k = P} C_N^k \frac{p!}{m_1! \dots m_k!} \int_{D_s} \tilde{f}^{m_1}(X) dX \dots \int_{D_s} \tilde{f}^{m_k}(X) dX = \\ &= \mathbb{E} \left(\frac{1}{N} \sum_{l=1}^N f(\vec{\alpha}_l) \right)^p, \end{aligned}$$

где $\vec{\alpha}_l$ — независимые реализации s -мерного вектора с равномерно распределёнными и независимыми в D_s компонентами. Но при наших предположениях относительно f

$$\lim_{N \rightarrow \infty} P \left(\frac{1}{\sqrt{N\sigma_f}} \left| \sum_{l=1}^N f(\vec{\alpha}_l) \right| < y \right) = \sqrt{\frac{2}{\pi}} \int_0^y e^{-\frac{u^2}{2}} du,$$

$$\sigma_f^2 = \int_{D_s} \tilde{f}^2(X) dX = \int_{D_s} \tilde{f}^2(X) dX - \left(\int_{D_s} \tilde{f}(X) dX \right)^2.$$

■

Таким образом, при больших $M < p$ мультипликативный датчик порождает последовательность, близкую в указанном выше смысле (близость распределений) к последовательностям независимых равномерно распределённых случайных величин.

Литература по выбору приемлемых значений M при данных P весьма обширна. Мы сошлёмся лишь на монографии [8] и [9].

Далее мы рассмотрим в необходимом для наших целей объёме вопросы моделирования случайных величин с заданным законом распределения. Эффект конечности M и P не будет анализироваться. Нужно отметить, что анализ этого эффекта практически не исследовался.

1.2 Моделирование неравномерных распределений

Мультипликативный датчик (псевдо) случайных чисел порождает последовательности, близкие по своим свойствам к последовательностям независимых реализаций равномерно распределённых случайных чисел. Мы видели, что некоторые, и именно какие, свойства имеют место. Известно, что проверка последовательностей такого рода с помощью статистических тестов также даёт удовлетворительные результаты.

В настоящее время построены и частично исследованы другие датчики ([10] и [11]) с очень большим периодом, но период мультипликативного датчика при $P \geq 2^{64}$ и соответствующим образом выбранном M удовлетворителен для прикладных целей.

Таким образом обосновывается подход к численному интегрированию широкого класса интегрируемых по Риману функций. При этом одна и та же последовательность обслуживает любую размерность s области интегрирования в виде единичного гиперкуба с погрешностью $O(\frac{1}{\sqrt{N}})$. Это одна из сторон метода Монте-Карло. Другая, не менее важная, сторона состоит в возможности получать ответ в виде распределения (шкалы). Во многих случаях алгоритм получения распределения оказывается значительно менее трудоёмким, чем алгоритм получения ответа в виде числа или таблицы чисел со значительным числом значащих цифр.

Данный раздел будет посвящён отдельным вопросам моделирования распределений в предположении, что имеется качественный алгоритм моделирования равномерно распределённых случайных величин. При этом мы остановимся лишь на некоторых принципиально важных вопросах. Проблемы влияния конечности числа разрядов и выбора

числа M на качество моделирования отдельных распределений мало изучены и нами практически не рассматриваются, если не считать обсуждения влияния равномерности распределения на методы отбора. Во всех случаях предполагается, что приближённая в смысле предыдущего s реализация α случайной величины, равномерно распределённой на промежутке $[0, 1]$, может быть получена в виде r -ичной дроби с необходимым числом разрядов.

Моделирование дискретных распределений. Следующая лемма устанавливает связь между равномерным непрерывным и дискретным распределением.

Лемма 2 *Для того, чтобы случайная величина α , имеющая представление в виде r -ичной дроби*

$$\alpha = \sum_{j=1}^{\infty} \frac{\varepsilon_j}{r^j}, \quad 0 \leq \varepsilon_j < r$$

была равномерно распределённой на промежутке $[0, 1]$, необходимо и достаточно, чтобы ε_j были независимыми в совокупности случайными величинами с распределением

$$P(\varepsilon_j = l) = 1/r, \quad R = 0, \dots, r-1,$$

Доказательство:

Обозначим $\Delta(r, k) = (\frac{k}{r}, \frac{k+1}{r}]$, $k = 0, 1, \dots, r-1$ и рассмотрим поведение ε_1 . Поскольку $\alpha \in \Delta(r, k)$ с равной вероятностью, то это и означает, что ε_1 не зависит от остальных ε_l . Предположим теперь, что наше утверждение доказано для совокупности $\varepsilon_1, \dots, \varepsilon_L$, $L > 1$. Чтобы доказать его для $L+1$, достаточно перейти к системе счисления по основанию r^{L+1} и повторить рассуждения, проведённые ранее для ε_1 .

Достаточность. Пусть теперь ε_k независимы в совокупности и $P(\varepsilon_k = l) = 1/r$, $l = 0, \dots, r-1$. Покажем, что

$$P(\alpha < x) = \begin{cases} 0, & x \leq 0 \\ x, & x \in [0, 1] \\ 1, & x > 1 \end{cases}.$$

Пусть $x = \sum_{j=1}^{\infty} x_j/r^{j+1}$ — r -ичное представление числа x , при $0 \leq x \leq 1$. В силу независимости ε_k имеем

$$P(\alpha < x) = \sum_{j=1}^{\infty} P(\varepsilon_j/r^j \in \delta_j), \text{ где } \delta_j = \left(\sum_{l=1}^{j-1} x_l/r^l, \sum_{l=1}^j x_l/r^l \right).$$

Очевидно

$$P(\varepsilon_j/r^j \in \delta_j) = x_j/r^j \text{ и } P(\alpha < x) = \sum_{j=1}^{\infty} x_j/r^j = x.$$

■

Отметим, что результат легко обобщается на случай представления числа α в смешанной системе счисления.

Лемма позволяет рассматривать последовательность случайных испытаний с конечным числом исходов в качестве последовательности значащих цифр случайной величины, равномерно распределённой на $[0, 1]$ в соответствующей системе счисления. Также позволяет моделировать равномерное распределение на множестве n объектов путём использования операции выделения целой части числа $n\alpha$.

Основные принципы моделирования случайной величины ξ , распределение которой задано таблицей (n конечно)

k	1	2	$\dots$	n
$P(\xi = k)$	p_1	p_2	$\dots$	p_n

хорошо известны. Мы остановимся на этом вопросе кратко с целью выяснения некоторых общих тенденций моделирования распределений.

Первая группа методов основана на проверке неравенства

$$\sum_{i=1}^{j-1} p_i \leq \alpha < \sum_{i=1}^j p_i. \quad (1.2.1)$$

Если это неравенство выполняется, то j есть нужное значение ξ . Суммы $S_j = \sum_{i=1}^j p_i$, $S_0 = 0$, могут быть вычислены заранее или последовательно вычисляться для $j = 1, 2, \dots$ до выполнения неравенства. В первом случае возможно использование бисекции, и трудоёмкость алгоритма оценивается величиной $\log_2 n$. Во втором трудоёмкость зависит от нумерации p_i и их конкретных значений и должна оцениваться в каждом конкретном случае. В худшем случае среднее число операций может достичь величины $\frac{n+1}{2}$.

Формула (1.2.1) определяет одно значение ξ по одной реализации α . Алгоритм с меньшей трудоёмкостью может быть построен, если реализация ξ будет находиться по двум или более реализациям α . Действительно, если область S площади $|S| > 1$ будет содержать фигуры S_i , имеющие площади $p_1, p_2, \dots, p_n$ соответственно, так что $\sum_{i=1}^n p_i = 1$, и мы будем “бросать” внутрь S точку β, β' , равномерно распределённую в S , то возможны следующие исходы:

- а) β, β' попадает в одну из S_i (вероятность события p_i);
- б) β, β' не попадает ни в одну из S_i (вероятность события $q = |S| - 1$).

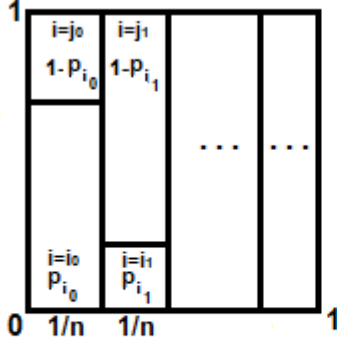
В случае а) мы полагаем $\xi = i$. В случае б) считаем, что произошло “непредусмотренное” событие и его нужно пропустить (метод отбора).

Очевидно, эффективность такого алгоритма тем выше, чем меньше q (формально эффективность определяют как $1/|S|$), и хотелось бы иметь алгоритм моделирования такой, чтобы каждое испытание было полезным. Можно попытаться расположить в единичном квадрате фигуры S_i с площадью p_i так, чтобы они полностью заполнили квадрат, и тогда каждая пробная точка, “брошенная” в квадрат, будет давать нужную реализацию, но сам алгоритм определения номера фигуры S_i может оказаться достаточно сложным, и всё обстоит не так просто.

Наиболее эффективный алгоритм, позволяющий избежать холостых проб, принадлежит Уокеру [12]. Конструкция покрытия единичного квадрата прямоугольниками с площадью p_i состоит в следующем. При $n = 2$ алгоритм очевиден. Достаточно отложить на одной из сторон отрезок длины p_1 , как это сделано на рисунке.

При $n > 2$

- а) Находим i_0 , для которого $p_{i_0} < \frac{1}{n}$. Такое i_0 существует, если не все $p_i = 1/n$ (в последнем случае задача тривиальна).



- б) Находим p_{j_0} , $i_0 \neq j_0$ такое, что $p_{i_0} + p_{j_0} > 1/n$. Такое j_0 существует. Если для всех $j \neq i_0$, $p_{i_0} + p_j < 1/n$, то

$$\begin{aligned} \sum_{j \neq i_0} (p_{i_0} + p_j) &< \frac{n-1}{n} < 1, \quad \text{но} \\ \sum_{j \neq i_0} (p_{i_0} + p_j) &= (n-1)p_{i_0} + \sum_{j \neq i_0} p_j = \\ &= (n-1)p_{i_0} + 1 - p_{i_0} = 1 + (n-2)p_{i_0} > 1. \end{aligned}$$

- в) Первую полосу квадрата заполняем прямоугольниками $(1/n \times p_{i_0}) + (1/n \times 1/n - p_{i_0})$. Таким образом, p_{i_0} оказывается полностью использованным, и для покрытия остальной части остаются прямоугольники со сторонами

$$(p_{i_0} + p_{j_0} - 1/n, \text{ и все } p_i, \text{ кроме } p_{i_0} \text{ и } p_{j_0}) \times 1/n.$$

Обозначив $p_{i_0} + p_{j_0} - 1/n = p'_1$ и остальные $p'_2, \dots, p'_{n-1}$, мы вернулись к исходной ситуации, но $\sum_{i=1}^{n-1} p'_i = 1 - 1/n$, а у исходного квадрата должна быть исключена первая полоса. Изменяя масштаб вдоль оси x , мы можем найти нужные i_1 и j_1 для покрытия 2-ой полосы и так далее. Каждый шаг уменьшает n на единицу. Случай $n = 2$ обсуждался выше (очевиден), найдены $(i_0, j_0) \dots (i_{n-1}, j_{n-1})$.

Если покрытие сконструировано, то алгоритм моделирования исходного распределения состоит в следующем:

- 1) Выбираем α -равномерно распределённую псевдослучайную величину и полагаем $k = \lfloor n\alpha \rfloor$.
- 2) Выбираем α' . Если $\alpha' \leq p_{i_k}$, то $\xi = i_k$. В противном случае $\xi = j_k$.

Таким образом, после соответствующей подготовки, имея две независимые реализации равномерно распределённой случайной величины α и α' , мы находим реализацию ξ за $O(1)$ операций. Подготовка, однако, может в общем случае потребовать до $O(n^2)$ операций.

О случае бесконечного n будет упомянуто позже.

В непрерывном случае, как известно, если случайная величина ξ задана своей функцией распределения $F(x)$, задача моделирования принципиально решается с помощью формулы обращения $\xi = F^{-1}(\alpha)$ или $\int_{-\infty}^{\xi} \varphi(x) dx = \alpha$, где $\varphi(x)$ — плотность распределения. Для условных функций распределения случайного вектора $\Xi = (\xi_1, \dots, \xi_s)$ задача моделирования также решается с помощью цепочки формул обращения. Реализация Ξ выражается через вектор $(\alpha_1, \dots, \alpha_s)$, составленный из независимых реализаций случайных величин, равномерно распределённых на $(0, 1]$. Как правило, однако, алгоритмы вычисления реализаций случайной величины с помощью формул обращения излишне сложны. Упрощение возможно за счёт увеличения размерности (использования большего числа равномерно распределённых случайных величин).

Основными методами, использующими этот приём, являются метод отбора и метод композиции. Остановимся кратко на этих методах, считая, что моделируемый случайный вектор Ξ задан своей плотностью $\varphi(x_1, \dots, x_s)$.

Лемма 3 Если $D \subset \mathbf{R}^{s+1}$, $D = \{0 \leq x_{s+1} \leq c\varphi(x_1, \dots, x_s)\}$, $c > 0$, то для того, чтобы точка (Ξ, ξ_{s+1}) , $\Xi = (\xi_1, \dots, \xi_s)$ была равномерно распределена в области D , необходимо и достаточно, чтобы вектор Ξ был распределён с плотностью $\varphi(x)$ и условное распределение ξ_{s+1} при фиксированном Ξ было равномерным на промежутке $[0, \varphi(\Xi)]$.

Доказательство сразу становится очевидным, если учесть, что плотность $\psi(x_1, \dots, x_{s+1})$ распределения (Ξ, ξ_{s+1}) в D постоянна (равна 1). Тогда плотность распределения Ξ есть $\int \psi(x_1, \dots, x_s, x_{s+1}) dx_{s+1} = \varphi(X)$, а условная плотность распределения ξ_{s+1} при фиксированном Ξ есть константа (подробнее см. [6]).

На основе этой леммы конструируются следующие алгоритмы.

- А. Моделирование равномерного распределения в области $D = \{0 \leq x_{s+1} \leq f(x_1, \dots, x_s)\}$, где f — интегрируемая в $\mathbf{R}^s$ функция. Если $C = \left(\int_{\mathbf{R}^s} f(X) dX \right)^{-1}$ — константа нормировки, то алгоритм полностью определяется леммой в предположении, что существует удобный алгоритм моделирования Ξ с плотностью $Cf(x)$, затем моделируется равномерное распределение на $[0, f(\Xi)]$. Требуется не менее $s + 1$ случайных чисел.
- Б. Если D ограничена и может быть помещена в $(s+1)$ -мерный параллелепипед с рёбрами, параллельными координатным осям, то точка, равномерно распределённая в D , может быть получена методом отбора. Получаем точку внутри параллелепипеда с равномерно распределёнными координатами $(\beta_1, \dots, \beta_{s+1})$. Проверяем её принадлежность D . Если “да”, то эта точка и является нужной реализацией, если “нет”, то вычисляется новая реализация, равномерно распределённого случайного вектора в параллелепипеде.
Из леммы следует также, что отобранные $(\beta_1, \dots, \beta_s)$ распределены с плотностью $Cf(X)$.
- В. Метод мажорант. Если моделирование плотности $\varphi(X)$ очень трудоёмкая задача, но мы умеем хорошо моделировать плотность $\psi(X)$, причём $\varphi(X) \leq C\psi(X)$, то алгоритм А) позволяет строить равномерное распределение в области $D = \{0 \leq X_{s+1} \leq C\psi(X)\}$. Если среди точек, равномерно распределённых в D , отобрать

точки $(\beta_1^i, \dots, \beta_{s+1}^i)$, $i = 1, 2, \dots$, попадающие в $D' = \{0 \leq X_{s+1} \leq \varphi(X)\}$, то $(\beta_1^i, \dots, \beta_s^i)$ будут векторами, распределёнными с плотностью $\varphi(X)$. Окончательно алгоритм, использующий A), выглядит следующим образом:

- В1. Находим реализацию Ξ плотности $\psi(X)$.
- В2. Проверяем условие $\varphi(\Xi)/C\psi(\Xi) < \alpha$, где α — реализация равномерно распределённой случайной величины.
- В3. Если “да”, то Ξ — нужная реализация, если “нет”, то В1).

Замечание о дискретных распределениях. Метод мажорант очевидным образом применим к дискретным распределениям (в том числе при $n = \infty$), если они представлены в виде набора плоских фигур площади p_i (увеличение размерности). Мажорантой может быть любая $C\psi(X)$.

Метод композиции в простейшей форме предполагает представление плотности в виде (смеси)

$$\varphi(X) = \sum_{i=1}^n p_i \varphi_i(X), \quad p_i > 0, \quad \sum_{i=1}^n p_i = 1,$$

$\varphi_i(X)$ являются плотностями. Моделирование состоит из двух этапов.

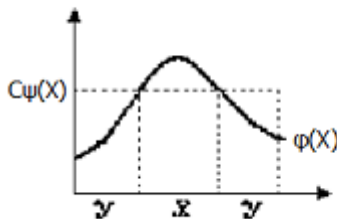
- А) Моделирование распределения $p_1, \dots, p_n$. Определяем $i = i_0$.
- Б) Моделирование плотности $\varphi_{i_0}(X)$. Находим реализацию Ξ . В более общей форме предполагается представление $\varphi(X)$ в виде $\varphi(X) = \int p(X, y) dF(y)$, где $p(X, y)$ — плотность по X при любом фиксированном y , а $F(y)$ — функция распределения.

Легко видеть, что метод композиции также предполагает увеличение размерности в понятном смысле. Приведённые приёмы позволяют строить алгоритмы, существенно менее трудоёмкие по сравнению с формулами обращения. Однако для ряда задач моделирования плотностей, многомерных в особенности, перечисленных приёмов может быть недостаточно.

Во-первых, вычисление константы нормировки при больших s может быть очень трудной задачей.

Во-вторых, если даже эта константа известна, бывает трудно подобрать мажоранту, обеспечивающую приемлемую эффективность алгоритма моделирования.

Адаптивный метод подбора мажоранты. Обсудим сначала вопрос о ситуации, возникающей при ошибке в выборе мажоранты. Если мы выберем константу C неоправданно большой, то это, строго говоря, ошибкой не будет, но сильно уменьшит эффективность алгоритма. Ошибка имеет место в случае, указанном на рисунке.



Обозначим

$$\begin{aligned}\mathcal{X} &= \{X : \varphi(X) > C\psi(X)\}, \\ \mathcal{Y} &= \{X : \varphi(X) \leq C\psi(X)\}\end{aligned}$$

и построим следующую адаптивную процедуру выбора $C = C_t$. Выбираем C_0 , которое, по-нашему мнению, удовлетворяет условию мажорирования, и X_0 из области определения $\varphi(X)$ ($\varphi(X_0) \neq 0$). Далее, если имеется X_t , $t = 0, 1, \dots$, полагаем $C_{t+1} = \frac{\varphi(X_t)}{\psi(X_t)}$, находим X' , распределённую с плотностью $\varphi(X)$, и вычисляем величину

$$\omega = \frac{\psi(X_t)\varphi(X')}{\psi(X')\varphi(X_t)} = \frac{\varphi(X')}{C_t\psi(X')}.$$

Если $\omega < 1$, то с вероятностью ω полагаем $X_{t+1} = X'$

и с вероятностью $1 - \omega$ полагаем $X_{t+1} = X_t$.

Если $\omega > 1$, то полагаем $X_{t+1} = X'$.

Алгоритм “исправляет” константу нормировки, и при некотором T , достаточно большом для всех $t > T$, X_T будут независимыми реализациями случайного вектора, распределённого с плотностью $\varphi(X)$.

Этот простой динамический алгоритм мало эффективен, поскольку изменяется константа, но не форма мажоранты. Ниже описан сходный по форме и широко используемый алгоритм Метрополиса-Хастингса [13], [14]. Алгоритм служит основой так называемого метода Монте-Карло на марковских цепях, которому посвящена обширная литература (например, [15]).

Формальное описание метода.

1. Выбираем X_0 и некую переходную плотность $\psi(X, Y)$, удовлетворяющую условию

$$\psi(X, Y) > 0, \text{ если } \psi(Y, X) > 0.$$

2. Если известно X_t , $t = 0, 1, \dots$, то моделируем $\psi(X_t, X')$, находим X' .

3. Вычисляем

$$\omega = \frac{\varphi(X') \psi(X', X_t)}{\varphi(X_t) \psi(X_t, X')}.$$

4. Если $\omega \geq 1$, то $X_{t+1} = X'$.

5. Если $\omega < 1$, то с вероятностью ω $X_{t+1} = X'$, а с вероятностью $1 - \omega$ $X_{t+1} = X_t$ (без изменений).

Очевидно, случайный процесс с X_t с дискретным временем $t = 0, 1, \dots$ является марковским. Справедливо следующее утверждение.

Теорема 5 Если случайный процесс, описываемый в предшествующих $n^01 - n^05$, является эргодическим, то его стационарное распределение имеет плотность $\varphi(X)$.

Доказательство:

Отметим следующее.

- а) Если цепь Маркова с переходной плотностью $p(X, Y)$ для некоторой функции $\varphi(X)$ удовлетворяет соотношению

$$p(X, Y)\varphi(Y) = p(Y, X)\varphi(X) \quad (1.2.2)$$

для всех X и Y (уравнение детального баланса), то $\varphi(X)$ является плотностью стационарного распределения для $p(X, Y)$. Действительно, интегрирование (1.2.2) по Y даёт $\int p(X, Y)\varphi(Y) dY = \varphi(X)$, поскольку $\int p(Y, X) dY = 1$.

- б) Переходная плотность случайного процесса 1)-5) есть

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

что следует из описания процесса.

- в) Но

$$\begin{aligned} \varphi(X)A(X, Y) &= \varphi(X)\psi(X, Y) \cdot \min \left(1, \frac{\varphi(Y)\psi(Y, X)}{\varphi(X)\psi(X, Y)} \right) = \\ &= \min (\varphi(X)\psi(X, Y), \varphi(Y)\psi(Y, X)). \end{aligned}$$

Последнее выражение не изменяется при замене X на Y . Отсюда

$$A(X, Y)\varphi(Y) = A(Y, X)\varphi(X),$$

что и доказывает, с учётом $n^0a)$ нашу теорему. ■

Метод Метрополиса-Хастингса относится к числу адаптивных и обладает высокой степенью универсальности. Он может быть особенно эффективным при моделировании многомерных плотностей, для которых трудно подобрать мажоранту.

Отдельный интерес может представлять специальный выбор плотности $\psi(X, Y)$. Так, если $\psi(X, Y) = f_1(X)f_2(Y)$, где f_1 и f_2 — плотности, то f_2 является стационарным распределением процесса с переходной плотностью $\psi(X, Y)$, а величина ω есть

$$\omega = \frac{\varphi(X') f_1(X')f_2(X_t)}{\varphi(X_t) f_1(X_t)f_2(X')}.$$

Очевидно, мы имеем достаточно простое обобщение метода мажорант.

В общем случае вопросы обоснования эргодичности и ускорения сходимости к стационарному распределению, а также выбора переходной плотности, породили обширную литературу.

1.3 Примеры

Строго говоря, речь идёт об одном примере, связанном с интересными приложениями. Ряд других примеров содержит 2-ая часть книги. Таким примером, использующим методы моделирования, описанных в §1.2, является метод решения одной из важных экспериментальных задач планирования эксперимента, которая состоит в нахождении глобального экстремума функции большого числа переменных. Методы случайного поиска глобального экстремума функции широко применяются и описаны в литературе. Заметим только, что возможность эффективного моделирования плотности

$\varphi(X) = Cf(X)$, ($f(X) \geq 0$ (C — константа нормировки)), может существенно облегчить задачу определения глобального максимума функции $f(X)$. Искомая точка максимума является модой распределения с плотностью $\varphi(X)$, и можно использовать методы оценки моды, например, [16], [17].

В рассматриваемом случае речь идёт о распределении Δ^2 , плотность которого по отношению к заданной вероятностной мере $\mu^n(dQ) = \bigotimes_{i=1}^n \mu(dX_i)$, $X_i \in D$ есть

$$\Delta_{m,n}^2 = C \det \left\| \sum_{j=1}^n \varphi_i(X_j) \varphi_k(X_j) \right\|_{i,k=1}^m, \quad (1.3.1)$$

где $\varphi_i(X_j)$ линейно независимы по отношению к мере μ в области D , $D \subset \mathbf{R}^s$ функции. $\Delta_{m,n}^2$ является функцией ns переменных, а точки её экстремума имеют важное значение в теории планирования эксперимента и носят название точного D -оптимального плана регрессионного эксперимента. Кроме того, реализации случайного вектора с Δ^2 -распределением могут использоваться для повышения эффективности метода Монте-Карло при вычислении многократных интегралов.

Использование Δ^2 -распределения требует его моделирования, а это, как правило, непростая задача. Самым простым представляется метод отбора. Он требует, однако, при каждом испытании вычисления определителя, а эффективность в общем случае определяется оценкой максимума $\Delta^2(Q)$. При этом, неравенство Адамара дает грубую оценку.

Случай, когда X есть отрезок, а функции φ_i , $i = 1, \dots, n$ — многочлены одной переменной изучался в работе [19]. В этом случае

$$\Delta_{n,m}^2(Q) = c \prod_{i < j} (x_i - x_j)^2. \quad (1.3.2)$$

Константа c служит очевидной оценкой сверху для $\Delta^2(Q)$, но при $n > 2$ довольно грубой. В этом простейшем случае до $n = 12$ вычисления по методу отбора были достаточно эффективными.

Что касается других функций φ_i , $i = 1, \dots, n$ и больших значениях n , то до недавнего времени считалось, что моделирование Δ^2 -распределения — задача безнадежно трудная.

Далее, в случае $n = m$ и ортонормированных (что не умаляет общности) по отношению к μ функций, мы покажем, как различные методы моделирования могут повысить эффективность алгоритма.

В нашем случае $\Delta_{m,n}^2$ имеет вид

$$\Delta_{m,n}^2 = p_n(X_1, \dots, X_n) = \left(\det \left\| \varphi_i(X_k) \right\|_{i,k=1}^n \right)^2 \frac{1}{n!}.$$

Справедливы следующие леммы.

Лемма 4 (Разложение Лапласа) Пусть D — определитель n -ого порядка. Фиксируем любые m его строк $i_1, \dots, i_m$, $m < n$. Тогда

$$D = \sum_{(j_1, \dots, j_m) \in (1, \dots, n)} M_{i_1, \dots, i_m; j_1, \dots, j_m} M'_{i_1, \dots, i_m; j_1, \dots, j_m},$$

где $M_{i_1, \dots, i_m; j_1, \dots, j_m}$ — минор D , получающийся при пересечении строк с номерами $i_1, \dots, i_m$ и столбцов с номерами $j_1, \dots, j_m$, а $M'_{i_1, \dots, i_m; j_1, \dots, j_m}$ — алгебраическое дополнение $M_{i_1, \dots, i_m; j_1, \dots, j_m}$.

Утверждение леммы верно, если вместо m строк фиксировать m столбцов.

Лемма 5 Пусть функции $\varphi_1, \varphi_2, \dots, \varphi_n$ образуют ортонормированную систему в пространстве $L^2(X, \mu)$, где X — измеримое множество, а μ — мера Лебега. Обозначим

$$p_n(x_1, \dots, x_n) = \frac{1}{n!} \left| \begin{array}{ccc} \varphi_1(x_1) & \dots & \varphi_1(x_n) \\ \vdots & \ddots & \vdots \\ \varphi_n(x_1) & \dots & \varphi_n(x_n) \end{array} \right|^2,$$

а

$$p_{n-k}(x_1, \dots, x_{n-k}) = \int_X p_n(x_1, \dots, x_n) dx_{n-k+1} \dots dx_n.$$

Тогда

$$p_{n-k}(x_1, \dots, x_{n-k}) = \frac{k!}{n!} \sum_{1 \leq i_1 < \dots < i_{n-k} \leq n} \left| \begin{array}{ccc} \varphi_{i_1}(x_1) & \dots & \varphi_{i_1}(x_{n-k}) \\ \vdots & \ddots & \vdots \\ \varphi_{i_{n-k}}(x_1) & \dots & \varphi_{i_{n-k}}(x_{n-k}) \end{array} \right|^2.$$

Доказательство:

Воспользуемся разложением Лапласа для определителя $p_n(x_1 \dots x_n)$, фиксируя первые $n - k$ его столбцов:

$$\begin{aligned} p_n(x_1 \dots x_n) &= \frac{1}{n!} \left| \begin{array}{ccc} \varphi_1(x_1) & \dots & \varphi_1(x_n) \\ \vdots & \ddots & \vdots \\ \varphi_n(x_1) & \dots & \varphi_n(x_n) \end{array} \right|^2 = \\ &= \frac{1}{n!} \left(\sum_{1 \leq i_1 < \dots < i_{n-k} \leq n} \left| \begin{array}{ccc} \varphi_{i_1}(x_1) & \dots & \varphi_{i_1}(x_{n-k}) \\ \vdots & \ddots & \vdots \\ \varphi_{i_{n-k}}(x_1) & \dots & \varphi_{i_{n-k}}(x_{n-k}) \end{array} \right| \left| \begin{array}{ccc} \varphi_{j_1}(x_{n-k+1}) & \dots & \varphi_{j_1}(x_n) \\ \vdots & \ddots & \vdots \\ \varphi_{j_k}(x_{n-k+1}) & \dots & \varphi_{j_k}(x_n) \end{array} \right| \right)^2, \end{aligned}$$

где $j_1, \dots, j_k$ — дополнение набора индексов $\{i_1 \dots i_{n-k}\}$ до $\{1, 2, \dots, n\}$.

$$\text{Обозначим } \left| \begin{array}{ccc} \varphi_{i_1}(x_1) & \dots & \varphi_{i_1}(x_{n-k}) \\ \vdots & \ddots & \vdots \\ \varphi_{i_{n-k}}(x_1) & \dots & \varphi_{i_{n-k}}(x_{n-k}) \end{array} \right| =: \Delta_{i_1 \dots i_{n-k}}.$$

Имеем

$$\begin{aligned} p_{n-k}(x_1, \dots, x_{n-k}) &= \int_X p_n(x_1, \dots, x_n) dx_{n-k+1} \dots dx_n = \\ &= \frac{1}{n!} \int_X \left(\sum_{1 \leq i_1 < \dots < i_{n-k} \leq n} \Delta_{i_1 \dots i_{n-k}} \left| \begin{array}{ccc} \varphi_{j_1}(x_{n-k+1}) & \dots & \varphi_{j_1}(x_n) \\ \vdots & \ddots & \vdots \\ \varphi_{j_k}(x_{n-k+1}) & \dots & \varphi_{j_k}(x_n) \end{array} \right| \right)^2 dx_{n-k+1} \dots dx_n = \\ &= \frac{1}{n!} \int_X \left(\sum_{1 \leq i_1 < \dots < i_{n-k} \leq n} \Delta_{i_1 \dots i_{n-k}} \left(\sum_{l=1}^k (-1)^{n+l} \varphi_{j_l}(x_n) \Delta'_{j_1 \dots j_{l-1} j_{l+1} \dots j_k} \right) \right)^2 dx_{n-k+1} \dots dx_n, \end{aligned}$$

$$\text{где } \Delta'_{j_1 \dots j_{l-1} j_{l+1} \dots j_k} = \begin{vmatrix} \varphi_{j_1}(x_{n-k+1}) & \dots & \varphi_{j_1}(x_n) \\ \vdots & \ddots & \vdots \\ \varphi_{j_{l-1}}(x_{n-k+1}) & \dots & \varphi_{j_{l-1}}(x_n) \\ \varphi_{j_{l+1}}(x_{n-k+1}) & \dots & \varphi_{j_{l+1}}(x_n) \\ \vdots & \ddots & \vdots \\ \varphi_{j_k}(x_{n-k+1}) & \dots & \varphi_{j_k}(x_n) \end{vmatrix}.$$

Проинтегрируем сначала по x_n . Возведем в квадрат подынтегральное выражение и воспользуемся тем, что функции $\varphi_1, \varphi_2, \dots, \varphi_n$ образуют ортонормированную систему в пространстве $\mathbf{L}^2(X)$. Благодаря этому попарные произведения в подынтегральном выражении, содержащие множители $\varphi_i(x_n)\varphi_j(x_n)$, $i \neq j$, обратятся в ноль. Более того, определитель $\Delta_{i_1 \dots i_{n-k}}$ под знаком внешней суммы не зависит от x_n . Поэтому его можно вынести за знак интеграла.

Наборов $(j_1, \dots, j_k)$ C_n^k штук. Каждому из них соответствует k слагаемых, состоящих из членов с элементами $(s_1, \dots, s_{k-1}) \subset (j_1, \dots, j_k)$ в подынтегральном выражении. Следовательно, общее количество членов становится равным kC_n^k . Из них ровно C_n^k различных. Таким образом, пользуясь ортогональностью функций $\varphi_1, \varphi_2, \dots, \varphi_n$ в $\mathbf{L}^2(X)$, получим

$$\int_X \left(\sum_{1 \leq i_1 < \dots < i_{n-k} \leq n} \Delta_{i_1 \dots i_{n-k}} \begin{vmatrix} p_{n-k}(x_1, \dots, x_{n-k}) = \frac{k}{n!} \cdot \\ \varphi_{s_1}(x_{n-k+1}) & \dots & \varphi_{s_1}(x_{n-1}) \\ \vdots & \ddots & \vdots \\ \varphi_{s_{k-1}}(x_{n-k+1}) & \dots & \varphi_{s_{k-1}}(x_{n-1}) \end{vmatrix} \right)^2 dx_{n-k+1} \dots dx_{n-1}.$$

Интегрируя по остальным переменным и пользуясь аналогичными рассуждениями приходим к утверждению леммы. ■

Укажем теперь вид условных плотностей $p_k(x_k | x_1 \dots x_{k-1})$, $k = 1, \dots, n$.

Следствие 1 Для условных плотностей $p_k(x_k | x_1 \dots x_{k-1})$, $k = 1, \dots, n$ справедливы формулы

$$p_k(x_k | x_1 \dots x_{k-1}) = \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{\Delta_{i_1 \dots i_k}^2}{(n-k+1) \sum_{1 \leq j_1 < \dots < j_{k-1} \leq n} \Delta_{j_1 \dots j_{k-1}}^2},$$

где

$$\Delta_{l_1 \dots l_s}^2 = \begin{vmatrix} \varphi_{l_1}(x_1) & \dots & \varphi_{l_1}(x_s) \\ \vdots & \ddots & \vdots \\ \varphi_{l_s}(x_1) & \dots & \varphi_{l_s}(x_s) \end{vmatrix}^2.$$

Доказательство:

Действительно, для любого $k = 1 \dots n$ имеем:

$$p_k(x_k | x_1 \dots x_{k-1}) = \frac{p_k(x_1 \dots x_k)}{p_{k-1}(x_1 \dots x_{k-1})}.$$

Осталось лишь воспользоваться леммой 5. ■

Согласно следствию 1 условные плотности $p_k(x_k | x_1 \dots x_{k-1})$, $k = 1, \dots, n$ представимы в виде линейной комбинации определителей k -ого порядка. Однако это еще не композиция распределений, поскольку

$$\int_X \Delta_{i_1 \dots i_k}^2 dx_k = \sum_{(m_1 < \dots < m_{k-1}) \subset (i_1 \dots i_k)} \Delta_{m_1 \dots m_{k-1}}^2.$$

Вводя нормировку в соответствии с этим соотношением мы получим представление условных плотностей в виде смеси:

$$p_k(x_k | x_1 \dots x_{k-1}) = \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{\sum_{(m_1 < \dots < m_{k-1}) \subset (i_1 \dots i_k)} \Delta_{m_1 \dots m_{k-1}}^2}{(n-k+1) \sum_{1 \leq j_1 < \dots < j_{k-1} \leq n} \Delta_{j_1 \dots j_{k-1}}^2} \frac{\Delta_{i_1 \dots i_k}^2}{\sum_{(m_1 < \dots < m_{k-1}) \subset (i_1 \dots i_k)} \Delta_{m_1 \dots m_{k-1}}^2}.$$

Далее мы можем использовать вместо “простого” метода, в котором используется оценка максимального значения определителя и в качестве мажоранты используется константа (это значение), — использовать комбинации различных методов, а именно:

а) Метод обращения: представляем

$$\Delta^2(x_1, x_2, \dots, x_n) = p_1(x_1) p_2(x_2 | x_1) p_3(x_3 | x_1 x_2) \dots p_n(x_n | x_1 x_2 \dots x_{n-1})$$

б) Метод композиций: каждая из плотностей

$$p_1(x_1), p_2(x_2 | x_1), p_3(x_3 | x_1 x_2), \dots, p_n(x_n | x_1 x_2 \dots x_{n-1})$$

моделируется как смесь распределений с плотностями

$$\frac{\Delta_{i_1 i_2 \dots i_k}^2}{\sum_{(m_1 < \dots < m_{k-1}) \subset (i_1 \dots i_k)} \Delta_{m_1 \dots m_{k-1}}^2}, \quad k = (1, \dots, n),$$

которые имеют вид

$$\frac{1}{\sum_{(m_1 < \dots < m_{k-1}) \subset (i_1 \dots i_k)} \Delta_{m_1 \dots m_{k-1}}^2} \cdot \left(\sum_{i \in \{i_1, i_2, \dots, i_k\}, 1 \leq i_1 < \dots < i_k \leq n} c_{\{i_1, i_2, \dots, i_k\}/\{i\}}^{(k)} \varphi_i(x_k) \right)^2,$$

$k=(1, \dots, n)$;

в) Метод отбора: каждая из последних плотностей моделируется методом отбора. Поэтому сложность их моделирования равна $(\max_{x_k \in X, i=\overline{1, k}} |\varphi_i(x_k)|)^2 C_k$, где C_k — сложность нахождения коэффициентов $c_{\{i_1, i_2, \dots, i_k\}/\{i\}}^{(k)}$.

Замечание. В некоторых случаях более удобным может быть моделирование последних плотностей методом обращения.

Подробная оценка указанных алгоритмов двух типов, в зависимости от размерности определителя, содержится в работе [18]. Мы приведём лишь таблицу отношений $T_{\text{комб.}} / T_{\text{отб.}}$, где $T_{\text{отб.}}$ — трудоёмкость метода простого отбора, а $T_{\text{комб.}}$ — трудоёмкость описанного комбинированного метода.

n	2	3	4	5	6	7	8	9	...
$T_{\text{комб.}} / T_{\text{отб.}}$	1,7388	1,4685	1,0082	0,6563	0,4230	0,2743	0,1798	0,1194	...
...	10	20	30	40	50	100	200		
...	0,0801	0,0021	$7,50 \cdot 10^{-5}$	$2,89 \cdot 10^{-6}$	$1,17 \cdot 10^{-7}$	$1,70 \cdot 10^{-14}$	$5,52 \cdot 10^{-28}$		

Если моделирование $\varphi(X) = Cf(X)$ позволяет определять наибольшее значение $f(X)$, то, очевидно, что возможность моделирования плотности $\varphi_m(X) = C_m f^m(X)$ могло бы сильно упростить задачу.

Действительно, имеет место следующий факт. Плотность распределения

$$\lim_{m \rightarrow \infty} (f^m(X) / \int f^m(X) dX)$$

есть равномерная плотность на множестве функций $f(X)$ (то есть имеет постоянное значение на всех подмножествах носителя плотности, имеющих положительную меру).

Для случая единственного экстремума вопрос обсуждался впервые в [7]. Общий случай можно свести к задаче о минимуме $\exp(-f(X)/T)$, при $T \rightarrow \infty$ (см., например, [20]).

Моделирование плотности $f^m(X) / \int f^m(X) dX$ при больших m и оценивание её моды может дать хорошие результаты. Проблема в вычислении константы нормировки $(f^m(X) / \int f^m(X) dX)^{-1}$. К счастью, описанный нами ранее метод Метрополиса не требует знания константы нормировки.

Имеются многочисленные связи описанного метода с другими методами случайного поиска (см., например, [21]). Вопрос о сравнении качества различных модификаций случайного поиска находится в процессе изучения. Однако в многоэкстремальном случае метод Метрополиса находится вне конкуренции, так как позволяет иметь выборку из распределения на множестве экстремумов.

Вернёмся к задаче о точных D -оптимальных планах.

Простейший двумерный случай приводит нас к случаю бесконечного множества экстремумов. При определении коэффициентов линейной регрессии в единичном квадрате с минимальным числом измерений D -оптимальный план есть множество из трёх точек (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , доставляющих наибольшее значение квадрату определителя

$$\Delta^2(x_1, y_1; x_2, y_2; x_3, y_3) = (\det \|1, x_i, y_i\|_{i=1}^3)^2, \quad 0 \leq x_i \leq 1; \quad 0 \leq y_i \leq 1; \quad i = 1, 2, 3.$$

Несложные исследования показывают, что наибольшие значения достигаются при

$$x_1 = y_1 = 0, \quad x_2 = 0, \quad y_2 = 1 \quad \text{и при} \quad x_3 = 1, \quad 0 \leq y_3 \leq 1,$$

а также при других конфигурациях, получаемых из этой поворотами на $\frac{\pi}{2}$, π , $3/2\pi$. Применение метода Метрополиса для моделирования плотности

$C_m = (\det \|1, x_i, y_i\|_{i=1}^3)^{2m}$ может с хорошим приближением давать реализации точек, равномерно распределённых на множестве D -оптимальных планов. Это позволяет выделить подмножество D -оптимальных планов, удовлетворяющих дополнительным условиям.

Нужно заметить, что известные [22] методы численного определения непрерывных D -оптимальных планов (которые можно считать приближением к точным планам) сталкиваются со значительными трудностями в случае неединственности решения.

Глава 2

Интегралы и линейные уравнения

2.1 Вычисление интегралов

Если решение нужной задачи представлено в виде $J = \int f(x) \mu(dx)$, где μ — σ -конечная мера, то, не умаляя общности, можно считать μ вероятностной мерой, а J — математическим ожиданием случайной величины $\xi = f(\eta)$, где η — случайная величина, распределённая в соответствии с мерой μ . f естественным образом предполагается μ -интегрируемой, что равносильно предположению о существовании и конечности интеграла

$$\int |f(x)| \mu(dx) < +\infty.$$

Наиболее распространённая вычислительная схема метода Монте-Карло состоит в получении независимых реализаций $\eta_1, \eta_2, \dots, \eta_N$ случайной величины η средствами предыдущей главы и вычислении

$$S_N(f) = \frac{1}{N} \sum_{i=1}^N f(\eta_i). \quad (2.1.1)$$

Усиленный закон больших чисел гарантирует с вероятностью 1 сходимость $S_N(f)$ к величине J .

Если также конечен интеграл $\int f^2(x) \mu(dx)$, ($f \in \mathbf{L}_2(\mu)$), то Центральная предельная теорема даёт оценку погрешности

$$|J - S_N(f)| \leq \frac{\sigma_f \cdot y}{\sqrt{N}}, \quad (2.1.2)$$

где y определяется из равенства $\int_0^y e^{-u^2/2} du = \sqrt{\frac{\pi}{2}} \cdot p$, p — доверительная вероятность, с которой (2.1.2) имеет место, а $\sigma_f^2 = \int (f(x) - J)^2 \mu(dx)$.

Возможность моделирования распределений позволяет использовать другие предельные соотношения теории вероятностей, но задача оценки математического ожидания наиболее типична. Другим примером может служить рассмотренная выше задача оценивания моды распределения (оба алгоритма являются ПР-алгоритмами). Оценка (2.1.2) показывает, что погрешность метода Монте-Карло убывает сравнительно медленно с ростом N , но можно попытаться улучшить ситуацию за счёт уменьшения величины σ_f .

Существует много приёмов повышения эффективности метода Монте-Карло, основанных на уменьшении второго момента оценок вычисляемого интеграла. Большинство

таких методов достаточно полно описаны в стандартных учебниках, и мы напомним лишь о некоторых из них, представляющих интерес для дальнейшего изложения.

Метод существенной выборки в общей форме основан на представлении интеграла J в форме

$$J = \int f(x) \frac{d\mu}{d\nu}(x) \nu(dx), \quad (2.1.3)$$

где $\frac{d\mu}{d\nu}(x)$ — производная Радона-Никодима меры μ по мере ν (ν предполагается абсолютно непрерывной по отношению к μ). J может оцениваться суммами

$$\kappa_\nu(f) = \frac{1}{N} \sum_{j=1}^N f(Y_j) \frac{d\mu}{d\nu}(Y_j), \quad (2.1.4)$$

где Y_j — независимые случайные величины, распределённые в соответствии с мерой ν . Рассмотрим второй момент случайной величины $\xi_j(\nu) = f(Y_j) \frac{d\mu}{d\nu}(Y_j)$.

$$\text{Имеем } \mathbb{E} \xi_j^2(\nu) = \int f^2(x) \left(\frac{d\mu}{d\nu}(x) \right)^2 \nu(dx) \leq \left(\int |f(x)| \mu(dx) \right)^2$$

по неравенству Коши-Буняковского. Легко проверить, что знак равенства в полученном неравенстве достигается при

$$\nu(dx) = \nu_0(dx) = \frac{1}{\bar{J}} |f(x)| \mu(dx), \quad \text{где } \bar{J} = \int |f(x)| \mu(dx).$$

Отсюда следует, что дисперсия оценки интеграла J имеет нижнюю грань

$$\mathbb{D} \kappa_{\nu_0}(f) = \frac{1}{N} (\bar{J}^2 - J^2). \quad (2.1.5)$$

Для положительных f дисперсия обращается в нуль, и следует выбирать меру ν , по возможности близкой к ν_0 . Понятно, что моделирование ν_0 — задача не менее трудная, чем задача вычисления интеграла при моделировании исходной меры μ . Знакопеременный случай связан с дополнительными трудностями, о которых будет далее сказано особо.

Другими важными приёмами построения оценок интеграла J с уменьшенной дисперсией является точное интегрирование по отдельным переменным, осуществляемое предварительно или в процессе вычислений (интегрирование по дискретной составляющей меры), и использование сумм вида

$$\tilde{\kappa}_m(f) = \sum_{i=1}^m A_i(Q) f(X_i), \quad Q = (X_1, \dots, X_m),$$

где Q — случайный вектор со специально выбранной плотностью $\mathfrak{G}(Q)$ такой, что

$$\mathbb{E} \tilde{\kappa}_m(f) = J.$$

Частным случаем таких сумм являются суммы “antithetic variate” метода и метода расслоненной выборки.

О расслонении нужно сказать особо. В простейшем случае, когда интеграл $J = \int |f(X)| \mu(dX)$ вычисляется с помощью оценки

$$\kappa_1 = \frac{1}{2} (f(X_1) + f(X_2)), \quad (2.1.6)$$

где X_1 и X_2 независимы и распределены в соответствии с мерой μ и оценки

$$\varkappa_2 = \frac{1}{2}(f(Y_1) + f(Y_2)),$$

где Y_1 и Y_2 зависимы. Если Y_1 распределён по закону μ_1 , то Y_2 — по закону μ_2 и наоборот. μ_1 и μ_2 таковы, что $\int \mu_1(dX) = \int \mu_2(dX) = 1/2$ и $\mu_1(dX) + \mu_2(dX) = \mu(dX)$ (расслоение). Легко показать (см., например, [7]), что

$$D\varkappa_1 \geq D\varkappa_2 \quad \text{для всех } f \text{ из } L_2(\mu).$$

Это обстоятельство служит основанием для конструирования алгоритмов вычисления интеграла, в которых точкам (узлам) интегрирования надлежит быть с вероятностью 1 достаточно удалёнными друг от друга. Мы вернёмся к этому вопросу при обсуждении методов квази Монте-Карло.

2.2 Линейные интегральные уравнения

Обратимся теперь к уравнениям второго рода

$$\varphi(X) = \int k(X, Y)\varphi(Y) \mu(dY) + f(X) \pmod{\mu}. \quad (2.2.1)$$

Специальный выбор конструкции меры μ позволяет трактовать это уравнение как систему линейных алгебраических уравнений (дискретная мера) и как систему интегральных уравнений. Возможны различные смешанные конструкции. Общим предположением является предположение о сходимости в некоторой метрике итерационного процесса

$$\bar{\varphi}_{m+1}(X) = \int |k(X, Y)| \bar{\varphi}_m(Y) \mu(dY) + |f(X)|, \quad \bar{\varphi}_0(X) = |f(X)|, \quad m = 0, 1, \dots, \quad (2.2.2)$$

который мы будем называть *мажорантным*.

Нашей задачей будет вычисление некоторого множества функционалов $(\varphi, h_i) = \int h_i(X)\varphi(X) \mu(dX)$. Предполагается существование этих интегралов и, по крайней мере, слабая сходимость процесса (2.2.2) на множестве этих функционалов.

Сделанные предположения достаточны для представимости функционалов вида (h, φ) , где φ — решение (2.2.1) в виде интегралов по траекториям. Из (2.2.2) следует сходимость процесса

$$\varphi_{m+1}(X) = \int k(X, Y)\varphi_m(Y) \mu(dY) + f(X) \pmod{\mu}, \quad \varphi_0 = f, \quad m = 0, 1, \dots, \quad \text{и}$$

$$(h, \varphi) = \sum_{m=0}^{\infty} \int d\mu^m h(X_0) k(X_1, X_2) \cdots k(X_{m-1}, X_m) \varphi(X_m), \quad k(X_{-1}, X_0) \equiv 1.$$

Можно заметить, что m -тое слагаемое формально связано с набором $X_0, X_1, \dots, X_m$ -траекторией. На множестве всех траекторий можно определить меру, соотнося набору $X_0, \dots, X_m$ число

$$\frac{1}{(|h|, \bar{\varphi})} |h(X_0) K(X_0, X_1) \cdots K(X_{m-1}, X_m) \cdot f(X_m)|. \quad (2.2.3)$$

Легко проверить, что введённая таким образом мера является вероятностной. По своей структуре она сходна с мерой, индуцируемой цепью Маркова.

Пусть однородная цепь Маркова с дискретным временем $t = 0, 1, \dots$ задана плотностью $p^0(X)$ начального распределения по отношению к мере μ переходной плотностью $p(X \rightarrow Y)$,

$$\int p(X \rightarrow Y) \mu(dY) = 1 - g(X) \pmod{\mu}, \quad 0 \leq g(X) \leq 1. \quad (2.2.4)$$

$g(X)$ трактуется обычно как вероятность обрыва траектории. Траектории моделируются средствами §2 гл.1. В результате имеем траекторию $X_0 \rightarrow X_1 \rightarrow \dots \rightarrow X_\tau$, плотность вероятности которой $p_0(X_0)p(X_0 \rightarrow X_1) \dots p(X_{\tau-1} \rightarrow X_\tau)g(X_\tau)$.

Такое задание плотности вероятности по отношению к $\bigotimes_{j=0}^{\tau} \mu(dX_j)$ задаёт вероятностную меру на множестве траекторий. Если мера, задаваемая (2.2.3), абсолютно непрерывна по отношению к этой мере, то может быть определена соответствующая производная Радона-Никодима, и легко проверить, что случайная величина

$$J_a = \frac{h_0 k_{0,1} \dots k_{\tau-1,\tau} f_\tau}{p_0^0 p_{0,1} \dots p_{\tau-1,\tau} g_\tau} \quad (2.2.5)$$

является несмещённой оценкой функционала (h, φ)

$$E J_l = (h, \varphi). \quad (2.2.6)$$

Здесь и далее введены очевидные сокращённые обозначения $h_0 = h(X_0)$, $k_{i,j} = K(X_i, X_j)$ и так далее. Оценка (2.2.5) является простейшей и в вычислительном, и в теоретическом отношении. Достаточно простыми являются также оценки

$$J_c = \sum_{l=0}^{\tau} \frac{h_0 k_{0,1} \dots k_{l-1,l} f_l}{p_0^0 p_{0,1} \dots p_{l-1,l}} \quad (2.2.7)$$

и двойственные к ним

$$J_a^* = \frac{f_0 k_{1,0} \dots k_{\tau,\tau-1} h_\tau}{p_0^0 p_{0,1} \dots p_{\tau-1,\tau} g_\tau}, \quad (2.2.8)$$

$$J_c^* = \sum_{l=0}^{\tau} \frac{f_0 k_{1,0} \dots k_{l,l-1} f_l}{p_0^0 p_{0,1} \dots p_{\tau-1,\tau}}, \quad (2.2.9)$$

которые могут быть получены из соотношений

$$(h, \varphi) = (\varphi^*, f), \quad \text{где} \quad \varphi^*(X) = \int k(Y, X) \varphi^*(Y) \mu(dY) + h(X). \quad (2.2.10)$$

Условия абсолютной непрерывности мер, указанные нами для оценки (2.2.5), модифицируются соответствующим образом для других оценок [6], так что для каждого вида оценок выбираются цепи Маркова своего типа.

Другие, более сложные оценки, описаны в монографиях [7], [25]. Мы ограничимся использованием оценок, приведённых выше. Метод существенной выборки можно применить для конструирования цепи Маркова, минимизирующей дисперсии оценок J_c и J_c^* . В других случаях метод может быть применён, но оптимальная мера не является марковской. Простое доказательство приведённых ниже результатов можно найти в [6].

Для оценки J_c оптимальная цепь задаётся начальным распределением $p_{\text{опт.}}^0(X)$ и $p_{\text{опт.}}(X \rightarrow Y)$.

$$p_{\text{опт.}}^0(X) = \frac{|\bar{\varphi}(X)h(X)|}{(\varphi, h)}, \quad p_{\text{опт.}}(X \rightarrow Y) = \frac{|K(X, Y)|\bar{\varphi}(Y)}{\bar{\varphi}(X)}, \quad g_{\text{опт.}}^0(X) = \frac{|f(X)|}{\bar{\varphi}(X)},$$

где $\bar{\varphi}(X)$ есть решение мажорантного уравнения $(\bar{\varphi}(X) = \int |k(X, Y)|\bar{\varphi}(Y) \mu(dY) + |f(X)|)$. Дисперсия, соответствующая этой цепи, есть

$$D_{\text{опт.}} = (\bar{\varphi}, |h|)^2 - (\varphi, h)^2.$$

Для оценки J_c^* соответствующие выражения есть

$$p_{\text{опт.}}^0(X) = \frac{|f(X)|\bar{\varphi}^*(X)}{(f, \varphi^*)}, \quad p_{\text{опт.}}(X \rightarrow Y) = \frac{|K(Y, X)|\bar{\varphi}^*(Y)}{\bar{\varphi}^*(X)}, \quad g_{\text{опт.}}^0(X) = \frac{|h(X)|}{\bar{\varphi}^*(X)}, \quad (2.2.11)$$

где $\bar{\varphi}^*$ — решение двойственного мажорантного уравнения

$$\bar{\varphi}^*(X) = \int |K(Y, X)|\bar{\varphi}^*(Y) \mu(dY) + |h(X)|, \quad (2.2.12)$$

$$D_{\text{опт.}} = (\bar{\varphi}^*, |f|)^2 - (h, \varphi)^2.$$

Зависимость оптимальных цепей Маркова от точного решения исходной задачи не позволяет пользоваться результатами непосредственно, но позволяет судить о форме оптимальных плотностей и учесть априорную информацию, имеющуюся у исследователя. В случае, когда все функции, входящие в исходное уравнение, и функция $h(X)$ неотрицательны, оптимальные дисперсии оценок J_c и J_c^* обращаются в нуль. Ниже мы покажем, как в общем случае использование дополнительной информации позволяет получить оценки с нулевой дисперсией. Нужно сказать, что большинство задач, решаемых методом Монте-Карло сегодня, имеют неотрицательные параметры (уравнения переноса, уравнение Больцмана и др.), но наше исследование предполагает общий случай.

Как уже отмечалось, знакопеременность не позволяет обращать дисперсию в нуль за счет выбора марковской цепи (см. формулу (2.2.12)).

Ниже мы покажем, как строится цепь, обращающая в нуль дисперсию J_a и J_a^* в знакопеременном случае. Для этого мы используем некоторое преобразование уравнения (2.2.1) к системе уравнений, содержащей лишь неотрицательные функции.

Обозначим

$$f_+(X) = f(1, X) = \begin{cases} f(X), & f > 0 \\ 0, & f \leq 0, \end{cases}$$

$$f_-(X) = f(2, X) = \begin{cases} f(X), & f \leq 0 \\ 0, & f > 0, \end{cases} \quad (2.2.13)$$

$$k_+(X, Y) \equiv k(1, X; 1, Y) = k(2, X; 2, Y) = \begin{cases} k(X, Y), & k(X, Y) > 0 \\ 0, & k(X, Y) \leq 0, \end{cases}$$

$$k_-(X, Y) \equiv k(1, X; 2, Y) = k(2, X; 1, Y) = \begin{cases} k(X, Y), & k(X, Y) \leq 0 \\ 0, & k(X, Y) > 0, \end{cases}$$

и через $\mathcal{K}_+$ и $\mathcal{K}_-$ — операторы с соответствующими ядрами $f_l = f(l, X)$, $l = 1, 2$.

Справедлива следующая лемма.

Лемма 6 Система уравнений относительно ψ_1 и ψ_2

$$\begin{aligned}\psi_1 &= \mathcal{K}_{11}\psi_1 + \mathcal{K}_{12}\psi_2 + f_1, \\ \psi_2 &= \mathcal{K}_{21}\psi_1 + \mathcal{K}_{22}\psi_2 + f_2,\end{aligned}\tag{2.2.14}$$

где $\mathcal{K}_{11} = \mathcal{K}_{22} = \mathcal{K}_+$; $\mathcal{K}_{12} = \mathcal{K}_{21} = -\mathcal{K}_-$ содержит лишь неотрицательные функции, при этом функция φ

$$\varphi = \psi_1 - \psi_2,\tag{2.2.15}$$

удовлетворяет исходному уравнению (2.2.1).

Доказательство:

Первая часть утверждения очевидна — следует из (2.2.13). Для доказательства второй части вернемся к обозначениям (2.2.13). Система (2.2.14) будет иметь вид

$$\begin{aligned}\psi_1 &= \mathcal{K}_+\psi_1 - \mathcal{K}_-\psi_2 + f_+, \\ \psi_2 &= -\mathcal{K}_-\psi_1 + \mathcal{K}_+\psi_2 + f_-\end{aligned}\tag{2.2.16}$$

Вычтем почленно второе уравнение из 1-го, получим:

$$\psi_1 - \psi_2 = (\mathcal{K}_+ + \mathcal{K}_-)\psi_1 - (\mathcal{K}_+ + \mathcal{K}_-)\psi_2 + f_+ + f_-,$$

что равносильно

$$\psi_1 - \psi_2 = \mathcal{K}(\psi_1 - \psi_2) + f,$$

что и доказывает лемму. ■

Дальнейшие рассуждения проведем для оценки J_a^* . Случай оценки J_a рассматривается по аналогии, он даже проще случая J_a^* .

Очевидно, что все преобразования, проведенные при доказательстве леммы, могут быть осуществлены применительно к уравнению $\varphi^*(X) = \int k(Y, X)\varphi^*(Y)\mu(dY) + h(X)$. Нужно φ , ψ_1 , ψ_2 , $\mathcal{K}_+$, $\mathcal{K}_-$ снабдить звездочкам и задать функцию $h(X)$, обозначая

$$\begin{aligned}h_1(X) &= h_+(X) = \begin{cases} h(X), & h(X) > 0 \\ 0, & h(X) \leq 0, \end{cases} \\ h_2(X) &= -h_-(X) = \begin{cases} -h(X), & h(X) \leq 0 \\ 0, & h(X) > 0, \end{cases}\end{aligned}$$

Аналог системы (2.2.14) имеет вид

$$\begin{aligned}\psi_1^* &= \mathcal{K}_{11}^*\psi_1^* + \mathcal{K}_{12}^*\psi_2^* + h_1, \\ \psi_2^* &= \mathcal{K}_{21}^*\psi_1^* + \mathcal{K}_{22}^*\psi_2^* + h_2,\end{aligned}\tag{2.2.17}$$

где $\mathcal{K}_{11}^* = \mathcal{K}_{12}^*$ — оператор ядром $k_+(Y, X)$,

$\mathcal{K}_{12}^* = \mathcal{K}_{21}^*$ — оператор ядром $-k_-(Y, X)$. Нам также удобно будет пользоваться обозначениями

$$h_i(X) = h(X, i), \quad f_j(X) = f(Y, j),$$

а для ядер операторов $\mathcal{K}_{i,j}^*$ — обозначениями $k(Y, j; X, i)$, $i, j = 1, 2$. Для решения системы (2.2.17) моделируется цепь Маркова, задаваемая начальной плотностью относительно меры λ , или, что то же самое, вектор $(p^0(X, 1), p^0(X, 2))$ и плотность перехода $p(X, i \rightarrow Y, j)$, то есть матрица $\|p(X, i \rightarrow Y, j)\|_{i,j=1}^2$. При этом $\sum_{j=1}^2 p(X, i \rightarrow Y, j) =$

$$1 - g(X, i), i = 1, 2,$$

$0 \leq g(X, i) \leq 1$ и должны выполняться условия согласования, о которых сказано ниже.

Оценка J_a^* функционала $\psi_1^* f_1 + \psi_2^* f_2$, записанная в плотностях по отношению к мере λ , имеет вид:

$$J_a^* = \frac{f(X_0, i_0) k(X_0, i_0; X_1, i_1) \dots k(X_{\tau-1}, i_{\tau-1}; X_\tau, i_\tau) h(X_\tau, i_\tau)}{p^0(X_0, i_0) p(X_0, i_0 \rightarrow X_1, i_1) \dots p(X_{\tau-1}, i_{\tau-1} \rightarrow X_\tau, i_\tau) g(X_\tau, i_\tau)} \quad (2.2.18)$$

или

$$J_a^* = \frac{f_0 k_{01} \dots k_{\tau-1, \tau} h_\tau}{p_0^0 p_{01} \dots p_{\tau-1, \tau} g_\tau} \quad (2.2.19)$$

при очевидном упрощении обозначений.

Здесь $(X_0, 1) \rightarrow (X_1, i_1) \rightarrow (X_2, i_2) \rightarrow \dots \rightarrow (X_\tau, i_\tau)$ — траектория моделируемой цепи Маркова, а упоминавшиеся условия согласования предполагают такой выбор параметров цепи Маркова, при котором знаменатель положителен для всех траекторий, для которых отличен от нуля числитель.

В нашем случае $k(Y, j; X, i)$ неотрицательны, но функционал, который вычисляется, имеет вид

$$(h, \varphi) = (f, \varphi^*) = ((f_+, f_-), (\psi_1^* - \psi_2^*)).$$

Здесь присутствует знакопеременность, что препятствует непосредственному применению теоремы о существенной выборке для интегралов по вероятностной мере с целью обращения в нуль дисперсии оценки. Однако мы можем представить (h, φ) в виде разности

$$(h, \varphi) = [(f_+, \psi_1^*) + (-f_-, \psi_2^*)] - [(-f_-, \psi_1^*) + (-f_+, \psi_2^*)]$$

двух величин, каждая из которых есть сумма двух положительных слагаемых. Для каждой из этих величин можно построить цепь Маркова, обращающую в нуль дисперсию несмещенной оценки вида (2.2.8).

Действительно, мы можем построить несмещенную оценку $(J_a^*)_1$ для $(f_+, \psi_1^*) + (-f_-, \psi_2^*)$ следующим образом.

Моделируем $p^0(X, i)$, если $i = 1$, то полагаем $f(X_0, 1) = f_+(X_0)$ и требуем, чтобы обрыв траектории происходил с вероятностью $g(X_\tau, 1)$. Соответственно, $h_\tau = h(X_\tau, 1) = h_+(X_\tau)$.

Если же $i = 2$, то $f(X_0, 2) = f_-(X_0)$, обрыв траектории происходит с вероятностью $g(X_\tau, 2)$ и $h_\tau = h_-(X_\tau)$.

Аналогично вычисляется оценка $(J_a^*)_2$ второго слагаемого. Теперь, следуя [6], мы можем строить цепь, обращающую в нуль дисперсию $(J_a^*)_1$. Для этого

$$(1) \text{ Полагаем } p_{\text{опт}}^0(X, i) = \frac{f(X, i) \psi_i^*(X)}{(f_+, \psi_1^*) + (-f_-, \psi_2^*)}.$$

$$(2) \text{ Полагаем } p(X, i \rightarrow Y, j) = \frac{k(X, i; Y, j) \psi_j^*(Y)}{\psi_i^*(X)}.$$

Легко видеть, что при таком выборе цепи Маркова оценка $(J_a^*)_1$ остается постоянной, равной $[(f_+, \psi_1^*) + (-f_-, \psi_2^*)]$ на всех траекториях, и следовательно её дисперсия равна нулю. Если $(J_a^*)_1$ и $(J_a^*)_2$ вычисляются по независимым траекториям, то $D((J_a^*)_1 + (J_a^*)_2) = D(J_a^*)_1 + D(J_a^*)_2$ и при оптимальном выборе цепей Маркова дисперсия суммарной оценки будет нулевой.

Исследование оценки второго слагаемого и анализ аналога оценки J_a заинтересованный читатель может легко и с удовольствием проделать самостоятельно.

Таким образом, при решении интегральных уравнений $2^{\text{го}}$ рода со знакопеременными коэффициентами (оценивании заданного функционала (φ, h) от решения) оказывается возможным использовать известные приемы понижения дисперсии с тем же эффектом, что и в знакопеременном случае, если воспользоваться сведением к системе вида (2.2.14) и разбить функционал на два слагаемых, в каждое из которых войдут функции одного знака. Очевидно, что последнее замечание относится ко всем известным несмещенным оценкам.

Внимательный читатель мог заметить, что при вычислении интеграла $\int f(X) \mu(dX)$ от знакопеременной функции f для получения оценок с нулевой дисперсией достаточно разбить интеграл на два слагаемых

$$\int f(X) \mu(dX) = \int f(X) \mu_1(dX) - \int f(X) \mu_2(dX), \quad \text{где}$$

мера μ_1 совпадает с μ на множестве $\{X : f(X) \geq 0\}$ и равна нулю вне этого множества, а мера μ_2 совпадает с μ на множестве $\{X : f(X) < 0\}$, и к каждому слагаемому применить принцип существенной выборки. Однако случай интегральных уравнений обладает определённым своеобразием, что и было продемонстрировано выше.

Оценки J_a и J_c дают возможность получать значение конкретного функционала. Повторные вычисления (частично могут использоваться одни те же траектории для разных функционалов) позволяют получить оценки заданного числа функционалов. Оценки J_a^* и J_c^* дают возможность оценивать заданное число функционалов и получать непараметрические оценки искомой функции φ . Заметим, что многие виды непараметрических оценок также связаны с получением оценок некоторого набора функционалов [23].

Следующая задача состоит в определении значений искомой функции во всех точках носителя μ или на заданном подмножестве этих точек (обобщённая задача интерполяции).

Определение класса Φ_m . Мы будем говорить, что $\varphi(X)$ принадлежит классу функций $\phi_m = \Phi(\Omega, \delta, \varepsilon_l, h_l; l = 1, \dots, m)$, если знание функционалов (h_l, φ) с погрешностью ε_l позволяет восстановить $\varphi(X)$, $X \in \Omega$ с погрешностью $\delta(X)$, $\|\delta(X)\| \leq \delta$.

Можно уточнять это определение: погрешности предполагать относительными или абсолютными, детерминированными или случайными и др. Такого рода уточнения мы будем делать по мере необходимости в дальнейшем.

Нужно отметить также сходство сформулированной задачи с непараметрическими подходами в статистике и методами, развиваемыми Войтишкем [25].

Глава 3

Стохастическая устойчивость и метод квази Монте-Карло

3.1 Стохастическая устойчивость

Вопросы стохастической устойчивости в данной главе излагаются применительно к дискретному случаю — системам линейных алгебраических уравнений (разностным схемам, в частности). Рассматриваемые алгоритмы, как правило, легко переносятся на общий случай линейных операторов. Вместо дискретных распределений моделируются непрерывные. Однако, их обоснование требует в конкретных случаях существенно более развитого функционально-аналитического аппарата. Один из примеров подобных обобщений можно найти в [31].

Системы линейных алгебраических уравнений вида

$$X = AX + F, \quad X = (x_1, \dots, x_n)^T, \quad A = \|a_{i,j}\|_{i,j=1}^n, \quad F = (f_1, \dots, f_n)^T$$

в случае, когда $\lambda_1(|A|)$ меньше единицы, являются наиболее простым объектом, свойства которого могут иллюстрировать общую теорию, изложенную в разделе 2.2.

Использование оценок (2.2.5)–(2.2.9) предполагает задание однородной цепи Маркова $p^0 = (p_0^0, \dots, p_n^0)$ и переходной матрицы $P = \|p_{i,j}\|_{i,j=1}^n$, удовлетворяющих условиям согласования (абсолютная непрерывность мер).

Для оценки J_c это:

$$\begin{aligned} p_i^0 &> 0, \quad \text{если } h_i \neq 0; \\ p_{i,j} &> 0, \quad \text{если } a_{i,j} \neq 0 \text{ и } g_i > 0; \\ \text{если } h_i \neq 0, i, j &= 1, \dots, n, \quad g_i = 1 - \sum_{j=1}^n p_{i,j}; \quad p_{i,j} \geq 0. \end{aligned}$$

Аналогично выглядят условия согласования для других оценок (подробно см. [6]). Эти условия определяют общую структуру цепи Маркова.

Алгоритм вычисления оценок (2.2.5)–(2.2.9) отличается от описанного ранее общего случая лишь тем, что в данном случае моделируются дискретные распределения. В предположении, что с вероятностью 1 все траектории конечны, мы N раз независимо вычисляем последовательности $i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_\tau$ номеров состояний, где каждый номер есть результат моделирования дискретного распределения с $(n+1)$ исходом. Один из исходов (обычно $(n+1)$ -ый) соответствует обрыву траектории.

Оценка J_a , например, имеет вид

$$J_a = \frac{h_{i_0} a_{i_0, i_1} \cdots a_{i_{\tau-1}, i_{\tau}} f_{i_{\tau}}}{p_{i_0}^0 p_{i_0, i_1} \cdots p_{i_{\tau-1}, i_{\tau}} g_{i_{\tau}}}. \quad (3.1.1)$$

Остальные оценки читатель легко выпишет самостоятельно (также см. [6], [7], [25]).

В соответствии с общей концепцией, сформулированной в начале книги, мы имеем дело с ПР-алгоритмом, который может быть “хуже” классических алгоритмов решения систем линейных уравнений. Важный вопрос — если хуже, то насколько хуже? Мы проведём прикидочную сравнительную оценку трудоёмкости описанного алгоритма метода Монте-Карло и метода итераций. Будем вычислять функционал

$$(H, X_M) = \left(H, \sum_{j=0}^{M-1} A^j F \right), \quad (3.1.2)$$

при некотором фиксированном M . Несмещённой его оценкой служит

$$\begin{cases} J_a, J_a^*, J_c, J_c^* & \text{при } \tau \leq M-1 \\ 0, & \tau \geq M \end{cases}. \quad (3.1.3)$$

Если среднее число ненулевых элементов в строке матрицы A есть K , дисперсия J_M равна σ^2 , и мы хотим вычислить (H, X_M) с погрешностью, не превосходящей $\varepsilon > 0$, то трудоёмкость T (среднее число операций) алгоритма может быть выражена произведением

$$T = N_{\varepsilon} M t_{\text{mod}}, \quad (3.1.4)$$

где N_{ε} — число испытаний, $N_{\varepsilon} \approx C \frac{\sigma_n^2}{\varepsilon^2}$, σ_n^2 — дисперсия оценки.

t_{mod} — трудоёмкость вычисления оценки, состоящая из трудоёмкости моделирования дискретного распределения, равного 2 для метода Уокера и $\log_2(n+1)$ — для метода бисекции, и арифметических операций (умножение на $a_{i,j}/p_{i,j}$ и сложение). Константа C определяется уровнем доверия (обычно $C \leq 3$).

Если вычисляются m функционалов, то для “прямых” оценок J_a и J_c , если каждый функционал оценивается независимо, трудоёмкость увеличивается в m раз. Специальными приёмами, использующими общие участки траекторий, трудоёмкость можно уменьшить.

В случае оценок J_a^* и J_c^* оценить трудоёмкость сложнее. Относительная погрешность оценок каждой из компонент здесь зависит от n (растёт, как $\sqrt{n}$).

Если система линейных уравнений получена в результате дискретизации дифференциального уравнения, то, как правило, решение обладает свойствами гладкости (кусочной гладкости) и может быть восстановлено с достаточной точностью по значениям m функционалов (принадлежит классу Φ_m). В этом случае также обычно A слабо заполнена.

Сравнение с итерационными методами, требующими $\approx Mn^2$ операций, показывает, что в случае слабой заполненности при больших n и фиксированном ε метод Монте-Карло может выигрывать при малых m . Во всяком случае, m должно слабо расти с ростом n .

При $k \approx n$ асимптотически с ростом n при ε фиксированном метод Монте-Карло может оказаться предпочтительнее итерационных методов.

Наше обсуждение имеет весьма предварительный характер, но позволяет наметить границы применимости метода Монте-Карло.

Далее отметим:

1. Мы не обсуждали возможность использования практически неограниченного числа процессоров без какого либо изменения последовательного варианта программы.
2. Не обсуждали возможность использования различного вида предобуславливателей в методе итераций. Введение предобуславливателя часто ведёт к нарушению условия $\lambda_1(|A|) < 1$.

Рассмотрим случай, когда $\lambda_1(|A|) > 1$, но $|\lambda_1(A)| < 1$. В этом случае функционал (H, X) не может быть представлен в виде интеграла по траекториям, хотя сходится процесс

$$X_{l+1} = AX_l + F, \quad X_0 = F, \quad l = 0, 1, \dots \quad (3.1.5)$$

Если A — очень большая матрица, то может быть выгодным использовать метод Монте-Карло для вычисления произведения $Z = AY$ при известном векторе Y . Действительно, $Z = \left\| \sum_{i=1}^n a_{i,j} y_j \right\|_{i=1}^n$, и при больших n каждую компоненту вектора Z можно оценить с помощью выборочного среднего некоторого числа m ($m < n$) слагаемых суммы. С ростом m ошибка вычисления сумм будет убывать и можно надеяться, что алгоритм, у которого будут последовательно оцениваться X_l с запоминанием результатов, позволит получать решение системы в случае, когда $\lambda_1(|A|) \geq 1$, но $|\lambda_1(A)| < 1$.

Эти соображения нуждаются в строгом исследовании, которое будет выполнено ниже. Мы будем считать, что для решения системы (3.1.5) в рассматриваемом случае будет строиться последовательность случайных векторов

$$\Xi_{l+1} = B\Xi_l + F, \quad \Xi_0 = F, \quad l = 0, 1, \dots, \quad (3.1.6)$$

удовлетворяющих условиям

$$E\Xi_l = X_l \quad \text{и} \quad EB = A. \quad (3.1.7)$$

Предварительно заметим, что при решении разностных схем в задачах, где решение зависит от времени, также применим этот подход, но l ограничено, $l = 0, 1, \dots, T-1$.

Для неявных схем

$$X^{l+1} = A_1 X^{l+1} + A_2 X^l + F_l, \quad A = A_1 + A_2, \quad (3.1.8)$$

более общий случай

$$X^{l+1} = \sum_{k=0}^{M-1} A_{k+1} X^{l+1-k} + F_l$$

сводится к упомянутому выше замене переменной с переходом к большей размерности векторов. Если выполнено достаточное условие устойчивости схемы

$$|\lambda_1((I - A_1)^{-1} A_2)| < 1, \quad (3.1.9)$$

то для получения Ξ_{l+1} в схеме (3.1.8) могут быть использованы методы Монте-Карло решения систем линейных уравнений, использующие оценки J_c , J_c^* , J_a и J_a^* , как это было описано ранее. Этот случай с применениями к волновому уравнению подробно описан в работах [28] и [29].

Алгоритмы другого типа для вычисления оценок X_l основаны на аналоге схемы Неймана-Улама для умножения матрицы на вектор.

Пусть $Y = AX$, $Y = (y_1, \dots, y_n)^T$ и H — заданный вектор $H = (h_1, \dots, h_n)$. Алгоритм предполагает задание матрицы $\mathcal{P} = \|p_{ij}\|_{i,j=1}^n$, удовлетворяющей условиям согласования

(а) $p_{i,j} > 0$, если $a_{i,j} \neq 0$ и распределение $\vec{q} = (q_1, \dots, q_n)$ такое, что $q_i > 0$, если $h_i \neq 0$ или

(б) $p_{i,j} > 0$, если $a_{j,i} \neq 0$ и распределение $\vec{p} = (p_1, \dots, p_n)$ такое, что $p_i > 0$, если $x_i \neq 0$.

Введем случайные величины

$$\xi = \frac{h_\alpha a_{\alpha,\beta} x_\beta}{q_\alpha p_{\alpha\beta}}, \quad (3.1.10)$$

$$\xi^* = \frac{x_\gamma a_{\delta,\gamma} h_\delta}{p_\gamma p_{\gamma\delta}}. \quad (3.1.11)$$

Здесь α распределено по закону $\vec{q}$, γ — по закону $\vec{p}$, β — по закону $(p_{\alpha,1}, \dots, p_{\alpha,n})$ при фиксированном α и δ — по закону $(p_{\gamma,1}, \dots, p_{\gamma,n})$ при фиксированном γ . Легко проверить, что

$$E\xi = E\xi^* = (H, Y).$$

Столь же просто вычисляются дисперсии оценок (3.1.10) (3.1.11), если учесть, что

$$E\xi^2 = \sum_{\alpha} \sum_{\beta} \frac{h_\alpha^2 a_{\alpha,\beta}^2 x_\beta^2}{q_\alpha p_{\alpha\beta}}, \quad E(\xi^*)^2 = \sum_{\gamma} \sum_{\delta} \frac{x_\gamma^2 a_{\delta,\gamma}^2 h_\delta^2}{p_\gamma p_{\gamma\delta}}.$$

С помощью неравенства Бине-Коши, как при доказательстве теоремы о существенной выборке, легко установить, что оптимальный выбор $p, q, \mathcal{P}$ состоит в следующем

(а) Для оценки ξ

$$q_\alpha = \frac{|h_\alpha|}{\sum_{\alpha} |h_\alpha|}, \quad p_{\alpha,\beta} = \frac{|a_{\alpha,\beta}| |x_\beta|}{\sum_{\alpha,\beta} |a_{\alpha,\beta} x_\beta|}. \quad (3.1.12)$$

(б) Для оценки ξ^*

$$p_\gamma = \frac{|x_\gamma|}{\sum_{\gamma} |x_\gamma|}, \quad p_{\gamma,\delta} = \frac{|a_{\delta,\gamma}| |h_\delta|}{\sum_{\gamma,\delta} |a_{\delta,\gamma} h_\delta|}. \quad (3.1.13)$$

Второй случай интересен тем, что стохастическая матрица $\mathcal{P}$ не зависит от точного решения задачи.

Оценки (3.1.10), (3.1.11) могут служить основой алгоритмов получения решения системы $X = AX + F$ без ограничений $\lambda_1(|A|) < 1$. Алгоритмы состоят в последовательном вычислении оценок векторов X_l по формуле (3.1.6).

Скалярное произведение (H, X_l) может оцениваться также с помощью оценок вида (3.1.10). В этом случае

$$(H, \Xi_l) \sim \frac{h_\alpha a_{\alpha,\beta} \xi_\beta (l-1)}{q_\alpha p_{\alpha\beta}} + f_\alpha, \quad \Xi_l = (\xi_1(l), \dots, \xi_n(l)). \quad (3.1.14)$$

Или оценок вида (3.1.11) и в этом случае

$$(H, \Xi_l) \sim \frac{\xi_\gamma (l-1) a_{\delta,\gamma} h_\delta}{p_\gamma p_{\gamma\delta}} + f_\gamma. \quad (3.1.15)$$

В каждом случае берется среднее из некоторого количества N реализаций оценок. Оценка всех компонент X_l может быть получена, если полагать H равным векторам с одной из компонент равной 1 и остальными равными нулю.

Как уже отмечалось, схема (3.1.15) интересна тем, что локально оптимальный выбор параметров, участвующих в расчетах переходной матрицы цепи Маркова не зависит от l . Схема (3.1.14) интересна тем, что она определяет рекуррентную процедуру, приводящую к ветвящейся цепи Маркова (число ветвлений — это число повторных реплик при оценке каждой компоненты Ξ_l). Предположим, что матрица $A = A_l$ может зависеть от номера итерации l , $X_{l+1} = A_l X_l + F_l$, как это бывает при решении разностных схем.

Поведение дисперсии случайных величин Ξ_l определяет следующая лемма.

Лемма 7 Пусть случайные векторы $\Xi_l = (\xi(l), \dots, \xi_n(l))^T$, $l = 0, 1, 2, \dots$ связаны соотношением

$$\Xi_l = \mathbb{B}_l \Xi_{l-1} + F_l, \quad (3.1.16)$$

где $\mathbb{B}_l = \|\beta_{i,j}\|_{i,j=1}^n$ случайные матричные операторы, $F_l = (f_1(l), \dots, f_n(l))^T$ — детерминированные векторы, $\Xi_0 = F_0$. При этом $\mathbb{E}\Xi_l = X_l = (x_1(l), \dots, x_n(l))^T$, $\mathbb{E}\mathbb{B}_l = A_l = \|a_{i,j}\|_{i,j=1}^n$ и $\mathbb{B}_l$ при каждом фиксированном l не зависит от Ξ_k для $k < l$. Тогда для матрицы ковариации вектора $\varepsilon_l = \Xi_l - X_l$ справедливо соотношение

$$\text{vec cov} \varepsilon_l = A_l \otimes A_l^T \text{vec cov} \varepsilon_{l-1} + \mathbb{E}(\delta_l \otimes \delta_l^T) \text{vec cov} \varepsilon_{l-1} + \mathbb{E}(\delta_l \otimes \delta_l^T) \text{vec}(X_{l-1} X_{l-1}^T),$$

где $\delta_l = \mathbb{B}_l - A_l$, vec — операция векторизации матрицы, а $\otimes$ — операция кронекеровского произведения матриц.

Доказательство:

Подставим в (3.1.16) выражения для $\mathbb{B}_l$ и Ξ_l :

$$X_l + \varepsilon_l = (A_l + \delta_l)(X_{l-1} + \varepsilon_{l-1}) + F_l.$$

Поскольку X_l подчинен (3.1.5), получим выражения для ε_l и ε_l^T :

$$\varepsilon_l = A_l \varepsilon_{l-1} + \delta_l X_{l-1} + \delta_l \varepsilon_{l-1} \quad (3.1.17)$$

и

$$\varepsilon_l^T = \varepsilon_{l-1}^T A_l^T + X_{l-1}^T \delta_l^T + \varepsilon_{l-1}^T \delta_l^T. \quad (3.1.18)$$

После перемножения правых и левых частей равенств (3.1.17) и (3.1.18) получим

$$\begin{aligned} \varepsilon_l \varepsilon_l^T &= A_l \varepsilon_{l-1} \varepsilon_{l-1}^T A_l^T + A_l \varepsilon_{l-1} X_{l-1}^T \delta_l^T + A_l \varepsilon_{l-1} \varepsilon_{l-1}^T \delta_l^T + \\ &+ \delta_l X_{l-1} \varepsilon_{l-1}^T A_l^T + \delta_l X_{l-1} X_{l-1}^T \delta_l^T + \delta_l X_{l-1} \varepsilon_{l-1}^T \delta_l^T + \\ &+ \delta_l \varepsilon_{l-1} \varepsilon_{l-1}^T A_l^T + \delta_l \varepsilon_{l-1} X_{l-1}^T \delta_l^T + \delta_l \varepsilon_{l-1} \varepsilon_{l-1}^T \delta_l^T \end{aligned}$$

Далее вычислим математическое ожидание от всех членов получившегося равенства. При этом стоит отметить, что математические ожидания многих слагаемых правой части будут равны нулю. Так, например, $\mathbb{E} \delta_l \varepsilon_{l-1} X_{l-1}^T \delta_l^T = 0$ в силу того, что ε_{l-1}^T не зависит от δ_l и $\mathbb{E} \varepsilon^T = 0$. Учитывая эти соображения, имеем

$$\mathbb{E}(\varepsilon_l \varepsilon_l^T) = A_l \mathbb{E}(\varepsilon_{l-1} \varepsilon_{l-1}^T) A_l^T + \mathbb{E}(\delta_l \varepsilon_{l-1} \varepsilon_{l-1}^T \delta_l^T) + \mathbb{E}(\delta_l X_{l-1} X_{l-1}^T \delta_l^T)$$

или, что то же самое,

$$\text{cov} \varepsilon_l = A_l \text{cov} \varepsilon_{l-1} A_l^T + \mathbb{E}(\delta_l \text{cov} \varepsilon_{l-1} \delta_l^T) + \mathbb{E}(\delta_l X_{l-1} X_{l-1}^T \delta_l^T).$$

Теперь, применяя операцию векторизации к матрицам ε_l , ε_{l-1} и $X_{l-1} X_{l-1}^T$, получим

$$\text{vec cov} \varepsilon_l = A_l \otimes A_l^T \text{vec cov} \varepsilon_{l-1} + \mathbb{E}(\delta_l \otimes \delta_l^T) \text{vec cov} \varepsilon_{l-1} + \mathbb{E}(\delta_l \otimes \delta_l^T) \text{vec}(X_{l-1} X_{l-1}^T).$$

или

$$\begin{aligned} \text{vec cov } \varepsilon_l = & (A_l \otimes A_l^T + E(\delta_l \otimes \delta_l^T)) \text{vec cov } \varepsilon_{l-1} + \\ & + E(\delta_l \otimes \delta_l^T) \text{vec}(X_{l-1} X_{l-1}^T). \end{aligned} \quad (3.1.19)$$

■

По поводу равенства (3.1.19) заметим следующее. Поскольку при каждом l Ξ_l вычисляется как среднее N_l независимых оценок, то $\|\text{cov } \varepsilon_{l-1}\|$ имеет порядок $O(1/N_l)$. В соответствии со свойствами кронекеровского произведения имеем

$$\begin{aligned} e_l & \leq \|\mathfrak{A}\| e_{l-1} + C/N_l \cdot \|X_l\|^2, \quad \text{где} \\ e_l & = \|\text{vec cov } \varepsilon_l\|, \quad \mathfrak{A} = A_1 \otimes A_1 + E(\delta_1 \otimes \delta_1^T) \quad \text{и} \\ \|\mathfrak{A}\| & \leq |\lambda_1(A)|^2 + C_1/N_l. \end{aligned}$$

При $N_l = N$, не зависящим от номера итерации, дисперсия ошибки либо растёт экспоненциально, либо стремится с ростом l к постоянному пределу. В первом случае алгоритм называют стохастически неустойчивым и стохастически устойчивым — во втором. Примеры стохастической неустойчивости, возникающей при решении разностного аналога волнового уравнения, описаны в [29].

Рассмотрим вопрос о параллелизме и трудоёмкости алгоритма. Здесь можно различать “внутренний” и “внешний” параллелизм. “Внутренний” — это параллелизм при вычислении произведений X_l при фиксированном l . Поскольку это вычисление состоит из N независимых испытаний, то в вычислении естественным образом могут участвовать $k \leq N$ процессоров. “Внешний” параллелизм основан на равенстве $E\Xi_l = X_l$. Независимые реализации случайного вектора Ξ_l при любом l можно получать на различных процессорах, а затем находить среднее M таких реализаций (ПР-алгоритм). Получение оценки решения вычислением среднего при достаточно большом M позволяет также вычислить эмпирическую дисперсию.

Мы, таким образом, построили ПР-алгоритм для решения систем линейных алгебраических уравнений и в случае $\lambda_1(|A|) > 1$ и указали его особенности в связи с величиной $\lambda_1(A)$ и числом испытаний N при каждом l .

Рассмотрим в общих чертах вопрос о трудоёмкости метода. В рамках выбранной модели (N не зависит от l) мы полагаем, что N удовлетворяет условиям стохастической устойчивости. Здесь, как отмечалось ранее, возможно использование “прямых” оценок (2.2.5), (2.2.7) или двойственных (2.2.8), (2.2.9). В общем случае N зависит от n — растёт с ростом n , и для больших n алгоритм может быть непомерно трудоёмким. Особый интерес представляет случай, когда решение может быть восстановлено по некоторому числу t функционалов, t слабо растёт с ростом n , и используются эффективные методы понижения дисперсии.

Пусть $\max_l \|\mathcal{F}_l\| \leq F$ и наибольшее собственное число матрицы $\mathfrak{A}$ есть Λ , тогда для l , достаточно больших, $e_l \leq L \frac{\sigma^2}{N} \cdot \frac{\|X_l\|^2}{1-\Lambda}$, где L — константа, а σ^2 — максимальная дисперсия оценок X_l . При N_1 , достаточно большом, $\Lambda \approx |\lambda_1(A)|^2$. Трудоёмкость T оценивается при N_1 независимых испытаниях группами по N испытаний при переходе от l -той к $(l+1)$ -ой итерации.

$$T = N_1 \cdot NM t_{\text{mod}}, \quad t_{\text{mod}} \leq \begin{cases} 2 \text{ (метод Уокера)} \\ 2 \log_2 k + 4 \text{ (метод бисекции)}. \end{cases} \quad (3.1.20)$$

Можно утверждать, что наличие множителя $\mu = \frac{1}{1-\Lambda}$ ухудшает ситуацию в μ раз по сравнению со случаем $\lambda_1(|A|) < 1$.

В работе [1] для разностного аналога уравнения Лапласа оценивалась трудоёмкость метода Монте-Карло в случае вычисления интеграла по траекториям (аналог метода простых итераций) и метода Монте-Карло при введении предобуславливателя (метод верхней релаксации). В последнем случае оказывается $\lambda_1(|A_1|) > 1$. Результат сравнения оценок показывает, что введение предобуславливателя в этом случае практически не уменьшает трудоёмкость алгоритма. Вероятно, введение предобуславливателя целесообразно тогда, когда для исходной системы $\lambda_1(|A|) > 1$.

Обратимся далее к примерам применения описанной теории для конкретных алгоритмов. В предположении необходимой гладкости решения $U(x, t)$ начально-краевой задачи для волнового уравнения

$$\frac{\partial^2 U}{\partial t^2} = \frac{\partial^2 U}{\partial x^2} \quad (3.1.21)$$

в области $[0, T] \times [0, 1]$ при начальных условиях $\frac{\partial U}{\partial t}(x, 0) = \varphi_1(x)$, $U(x, 0) = \varphi_0(x)$ и граничных условиях $U(0, t) = \psi_0(t)$, $U(1, t) = \psi_1(t)$ рассмотрим некоторые разностные схемы, приближённо решающие эту задачу. А именно, на сетке с шагом $h = 1/n$ и $\Delta t = T/m$ явную схему — “крест”

$$U_{k,j+1} = U_{k+1,j} + U_{k-1,j} - U_{k,j-1}, \quad k = 1, \dots, n-1 \text{ и } j = 1, 2, \dots, m-1 \quad (3.1.22)$$

и неявную схему

$$U_{j+1,k} = \frac{1}{2 + \mu^2} (U_{j+1,k+1} + U_{j+1,k-1}) + \frac{\mu^2}{2 + \mu^2} (2U_{j,k} - U_{j-1,k}), \quad (3.1.23)$$

где $\mu^2 = \frac{h^2}{(\Delta t)^2}$.

Обе схемы устойчивы относительно ошибок округления (вторая при любом значении μ). Но в обоих случаях мы имеем дело с системами, для которых наибольшее собственное число модуля матрицы будет больше единицы, и попытки использования схемы Неймана-Улама приводят к стохастически неустойчивым алгоритмам. Тем не менее, можно построить как в случае (3.1.22), так и в случае (3.1.23) последовательность случайных векторов Ξ_j таких, что выполняется равенство

$$E\Xi_j = (U_{j,1}, \dots, U_{j,n-1})^T \quad j = 2, 3, \dots$$

Для этого достаточно в случае (3.1.22) воспользоваться оценками (3.1.3), вычисляя их на каждом слое N_j раз независимо, и оценками для решения систем линейных уравнений в случае (3.1.23). Результаты для последнего случая подробно описаны в [29]. Там же показано, что для достаточно больших N_j алгоритм вычисления Ξ_j стохастически устойчив.

Будем считать, что N_j от j не зависит, $N_j = N$, и остановимся подробно на схеме (3.1.22).

Пусть вычислены несмещённые оценки $\xi_{j,k}$ и $\xi_{j-1,k}$, $k = 1, \dots, n-1$ для $U_{j,k}$ и $U_{j-1,k}$ соответственно. В оценках (J_c) и (J_c^*) выберем $\mathcal{P}$ с нулевыми элементами, соответствующими нулевым элементам матрицы системы и равными для простоты $1/3$ для ненулевых.

“Прямой” алгоритм состоит в том, что для выбранного k $\xi_{j+1,k}$ полагаем равным $3/N \sum_{i=1}^N \xi_{j+1,k}^{(i)}$, где

$$\xi_{j+1,k}^{(i)} = \begin{cases} \xi_{j,k+1} & \text{с вероятностью } 1/3, \\ \xi_{j,k-1}, \\ \xi_{j-1,k}. \end{cases} \quad (3.1.24)$$

Таким образом, структура алгоритма представляет собой случайный граф, в котором при каждом переходе из точек с номерами $(j, k+1)$, $(j, k-1)$, $(j-1, k)$ в точку $(j+1, k)$ три ветви сливаются в одну. Переход от начальных данных к данным на линии $j = T$ может быть описан как траектория столкновительного процесса движения “частиц”, при котором три частицы могут сливаться в одну при переходе со слоя на слой. Процессы такого рода являются двойственными к ветвящимся, и, действительно, легко указать алгоритм моделирования процесса, двойственного к столкновительному.

Алгоритм (“сопряжённый”) для вычисления U_{T,k_0} связан со следующим ветвящимся процессом. Частица блуждает по сетке (j, k) . С равными вероятностями $p = 1/3$ она переходит из (j, k) в одну из точек $(j-1, k-1)$, $(j-1, k+1)$, $(j-2, k)$. После перехода в соответствующей точке рождается N аналогичных частиц, если точка не принадлежит границе области. В граничной точке траектория обрывается.

Оценка ξ_{T,k_0} решения U_{T,k_0} вычисляется следующим образом. Начальная частица (в точке (T, k_0)) имеет вес $\vartheta = 1$. При каждом переходе вес умножается на $3/N$. При достижении частицей границы имеем вклад $(3/N)^r U(r)$, где

r — число переходов до границы, а $U(r)$ — граничное (начальное) значение решения U , которое известно. Оценкой для U_{T,k_0} служит сумма вкладов всех образовавшихся частиц

$$\xi_{T,k_0} = \sum_{\text{всех вкладов } i} Q_r^{(i)} U^{(i)}(r). \quad (3.1.25)$$

Обработка траекторий ветвящегося процесса может, как обычно, вестись либо по поколениям (j) , либо лексикографически. В последнем случае требуется T ячеек для хранения сведений о траектории. Если N выбрано так, что обеспечивается стохастическая устойчивость, то оба алгоритма (прямой и сопряжённый) являются ПР-алгоритмами.

Легко заметить, что суммирование в (3.1.25) нужно проводить “по слоям”, то есть по поколениям — при каждом j находить общие веса для каждой точки, а затем продолжать ветвление. Вычислительная работа для каждого j будет пропорциональна $N \cdot n$ — произведению числа ветвей на число точек на каждом слое. При обработке траекторий ветвящегося процесса иными методами работа может расти экспоненциально с ростом j .

Подробное исследование связей прямых и двойственных процессов с ветвлением для решения уравнений выходит за рамки данной книги. Отметим только, что формула вида (3.1.19) для ковариаций двойственных оценок может быть получена аналогично тому, как это сделано в рассмотренном нами случае. Обобщение на случай операторов существенно более трудоёмкий процесс. По этому поводу см. [31], а также некоторые работы Г.А. Михайлова [32].

3.2 Метод квази Монте-Карло

В отличие от метода Монте-Карло метод квази Монте-Карло (КМК-метод) имеет отдалённое отношение к теории вероятностей и основан на результатах из теории чисел и вычислительной математики, и, в частности, теории кубатурных формул.

Коротко об этих результатах.

Кубатурной формулой принято называть формулу приближённого вычисления интеграла

$$\int_D p(X)f(X) dX \approx \sum_{i=1}^N f(X_i)A_i, \quad (3.2.1)$$

$D \subset \mathbf{R}^s$; $X_i \in D$, $i = 1, \dots, N$, $p(X) > 0$ — заданная функция (весовая функция). Числа A_i называют коэффициентами, а $X_i = (x_1^{(i)}, \dots, x_s^{(i)})$ — узлами формулы. Функционал

$$R_N(f) = \int_D p(X)f(X) dX - \sum_{i=1}^N A_i f(X_i) \quad (3.2.2)$$

называют остатком кубатурной формулы.

Для выбора параметров A_i и X_i кубатурной формулы обычно используют два подхода:

- (1) Из условия $R_N(\varphi_j) = 0$, где $\varphi_j(X)$, $(j = 1, 2, \dots, m)$ — заданные функции.
- (2) Для заданного линейного нормированного пространства $\mathbf{F}$ из условия минимума нормы функционала $R_N(f)$.

В одномерном случае ($s = 1$) широко известные квадратурные формулы Гаусса обеспечивают обращение в нуль остатка для алгебраических многочленов степени $\leq (2N - 1)$. К сожалению, распространить теорию этих формул на многомерный случай удаётся лишь в отдельных частных случаях. Наиболее успешной в этом отношении является знаменитая работа И. Радона [33], где полностью исследован случай полиномов двух переменных суммарной пятой степени ($N = 5$). И это, как это ни странно, единственный случай, когда для $s > 1$, $N > 1$ строится почти полный аналог гауссовских квадратур, но и в этом случае вещественность параметров формулы гарантируется лишь для областей, обладающих центром симметрии.

Свойство симметрии (инвариантность относительно групп преобразований) вообще, служит источником большинства удачных примеров формул, точных для большого числа функций φ_i . Большое разнообразие конфигураций областей интегрирования в многомерном случае объясняет недостаточную универсальность подхода (1), что не умаляет его значения при решении отдельных задач. Из весьма обширной литературы, посвящённой теории вопроса, сошлёмся на [31] и [34].

Подход (2) наиболее полно отражён в монографии [36], где изучены свойства нормы функционала $R_N(f)$ для класса функций W , имеющих частные производные заданного порядка. Построение оптимальных формул для этих классов с учётом свойств границы области оказывается достаточно сложной задачей, требующей специальных исследований в конкретных случаях.

Достаточно общий подход к вычислению многократных интегралов был впервые развит в работах Н.М. Коробова [37]. Используя мощный теоретико-числовой аппарат он для случая s -мерного гиперкуба впервые указал методы численного интегрирования, близкие к оптимальным в классах дифференцируемых функций. При этом использовались оценки снизу для $R_N[f]$ в классах дифференцируемых функций, полученные Н. Бахваловым, и предположение (явное или неявное) о сводимости с помощью замены

переменных любой области с гладкой границы к гиперкубу с помощью замены переменной. Формулы Н.М. Коробова также точны для тригонометрических многочленов.

Эти краткие и совершенно неполные (дилетантские) упоминания о результатах в области теории кубатурных формул приводятся здесь лишь для того, чтобы указать место методов квази Монте-Карло в этой теории. Формально методы квази Монте-Карло состоят в том, что в вычислительной схеме метода Монте-Карло используются вместо равномерно распределённых на промежутке $(0, 1)$ чисел используются компоненты X_i — узлов некоторой кубатурной формулы. Точнее, речь идёт о последовательности кубатурных формул

$$\int_{D_s} f(X) dX \approx \frac{1}{N} \sum_{i=1}^N f(X_i), \quad (3.2.3)$$

D_s — единичный s -мерный гиперкуб, $X_i \in D_s$. При увеличении N предыдущие узлы сохраняются!

Легко видеть из § 2 гл.1, что методы моделирования случайных величин определяют замену переменных, переводящую в единичный гиперкуб область интегрирования. Однако, s может быть достаточно большим (может быть $s = \infty$?).

Характер этой замены полезно проиллюстрировать на простом примере. Пусть $s = 2$ и $D = \{0 \leq x \leq 1; 0 \leq y \leq C\varphi(X)\}$, где $\varphi(X)$ является плотностью распределения, а $C > 0$ — константа. При вычислении интеграла

$$I = \int_D f(x, y) dx dy$$

методом Монте-Карло можно использовать оценку

$$\xi_N(f) = \frac{1}{N} \sum_{i=1}^N f(\xi_i, \eta_i),$$

где (ξ_i, η_i) — точка, равномерно распределённая в D . Если использовать описанный в Лемме 3 метод получения (ξ_i, η_i) , то нужно взять два равномерно распределённых случайных числа $\alpha_1^{(i)}$ и $\alpha_2^{(i)}$, и $\xi_i = F^{-1}(\alpha_1^{(i)})$, $\eta_i = \alpha_2^{(i)} C\varphi(\xi_i)$, где $F(X)$ — функция распределения, имеющая плотность φ . Очевидно, мы имеем дело с заменой переменных $x = F^{-1}(u)$, $y = vC\varphi(F^{-1}(u))$. Якобиан преобразования есть $C\varphi(F^{-1}(u))$.

Возможны другие замены переменной, в том числе, учитывающие свойства f (существенная выборка), но метод квази Монте-Карло связан, как правило, с заменами, имеющими вероятностный смысл. Конечно, можно было бы использовать для моделирования плотность $\varphi(x)$ (метод отбора), но именно тогда s является случайной (сколь угодно большой).

Это очень важное **замечание**: кратность вычисляемого методом квази Монте-Карло интеграла зависит от конструкции соответствующего Монте-Карловского алгоритма.

Указанную кратность мы будем называть, следуя И.М. Соболю, конструктивной размерностью интеграла.

Выбор узлов формулы (3.2.3) осуществляется на основе подхода (2). Строятся последовательности X_i , $i = 1, 2, \dots$, доставляющие оптимальный порядок убывания норм остатка $R_N(f)$ в классе функций ограниченной вариации (в смысле Харди-Крауде).

Соответствующее неравенство норм носит название неравенства Коксмы-Хлавки (подробное его доказательство, не использующее средства функционального анализа, см., например, в [38]).

Далее мы в двумерном случае поясним основные идеи, лежащие в основе выбора критерия оптимизации для выбора последовательности узлов $X_1, X_2, \dots$.

Пусть в единичном квадрате $D_2 = \{0 \leq x, y \leq 1\}$ задана функция $f(x, y)$, имеющая первые интегрируемые производные f'_x и f'_y и смешанную производную f''_{xy} в этом квадрате. Тогда, используя разложение в ряд Тейлора, имеем для любой точки (u, v) из D_2

$$f(x, y) = f(u, v) + \int_u^x f'_x(t_1, v) dt_1 + \int_v^y f'_y(u, t_2) dt_2 + \int_u^x \int_v^y f''_{xy}(t_1, t_2) dt_1 dt_2. \quad (3.2.4)$$

Полагая $u = v = 1$ и подставляя полученное выражение для $f(x, y)$ в (3.2.2) при $P(X) \equiv 1$, $s = 2$, $A_i = 1/N$, $i = 1, 2, \dots$, получим после несложных преобразований

$$R_N[f] = \int_0^1 f'_x(t, 1) K_1(t) dt + \int_0^1 f'_y(1, t) K_2(t) dt + \int_0^1 \int_0^1 f''_{xy}(t_1, t_2) K_3(t_1, t_2) dt_1 dt_2,$$

где

$$K_1(t) = \left(-t + 1/N \sum_{i=1}^N \Theta(x_i - t) \right), \quad K_2(t) = \left(-t + 1/N \sum_{i=1}^N \Theta(y_i - t) \right),$$

$$K_3(t_1, t_2) = \left(t_1 t_2 - 1/N \sum_{i=1}^N \Theta(x_i - t_1) \Theta(y_i - t_2) \right), \quad \Theta(\tau) = \begin{cases} 0, & \tau > 0 \\ 1/2, & \tau = 0 \\ 1, & \tau < 0. \end{cases}$$

Легко видеть, что суммы, входящие в выражение для K_1 , K_2 и K_3 имеют следующую наглядную интерпретацию.

$\sum_{i=1}^N \Theta(X_i - t) = n(t, x_1, \dots, x_N)$ — количество точек с координатами X_i , удовлетворяющими условию $x_i \leq t$.

$\sum_{i=1}^N \Theta(X_i - t_1) \Theta(y_i - t_2) = n(t_1, t_2, X_1, \dots, X_N)$ — количество точек с координатами (x_i, y_i) , удовлетворяющими неравенствам $x_i < t_1$, $y_i < t_2$.

Заметим, что используя аналогичное (3.2.4) разложение в ряд Тейлора функции s -переменных, можно получить представление остатка $R_N[f]$, содержащее при производных ядра $K_l(t_1, \dots, t_l)$ вида

$$K(t_1, \dots, t_l) = (-1)^l \left(\prod_{i=1}^l t_i - n(t_1, \dots, t_l; X_1^{(l)}, \dots, X_N^{(l)}) \right), \quad l \leq s,$$

где $X_i^{(l)}$ — проекция точки $X_i = (x_1, \dots, x_s)$ на одну из гиперграней гиперкуба, а $n(t_1, \dots, t_l; X_1^{(l)}, \dots, X_N^{(l)})$ — количество точек, для которых выполнены неравенства $x_{i,j}^{(l)} < t_i$, $i = 1, \dots, l$; $X_j^{(l)} = (x_{1,j}^{(l)}, \dots, x_{N,j}^{(l)})$.

Функции $D_l^*(X_1^{(l)}, \dots, X_N^{(l)}) = N \sup K(t_1, \dots, t_l)$ носят названия “дискрепанс со звёздочкой” (star discrepancy). Поскольку “дискрепанс без звёздочки” нами не используется,

мы будем говорить просто дискрепанс (иногда используется русский термин “отклонение множества точек $X_1^{(l)}, \dots, X_N^{(l)}$ ”).

Вернёмся к двумерному случаю. Функции $f'_x(x, 1)$, $f'_y(1, y)$ и $f''_{xy}(x, y)$ независимы и могут принадлежать различным классам интегрируемых функций. Неравенство

$$|R_N(f)| \leq M_{1,0} \sup_t |K_1(t)| + M_{0,1} \sup_t |K_2(t)| + M_{1,1} \sup_{t_1, t_2} |K_3(t_1, t_2)|$$

$$M_{1,0} = \int_0^1 |f'_x(u, 1)| dt, \quad M_{0,1} = \int_0^1 |f'_y(1, v)| dt, \quad M_{1,1} = \int_0^1 \int_0^1 |f''_{xy}(u, v)| dudv \quad (3.2.5)$$

является точным. Существует функция $f(x, y)$ с интегрируемыми производными, для которой достигается знак равенства (3.2.5).

Поскольку $K_1(t) = K_3(t, 1)$ и $K_2(t) = K_3(1, t)$, то $\sup_{t_1, t_2} K_3(t_1, t_2) \geq \sup_t K_1(t)$ и неравенство (3.2.5) можно переписать в виде

$$|R_N(f)| \leq \text{Var} f \cdot \frac{D_2(x_1^{(2)}, \dots, x_N^{(2)})}{N}, \quad (3.2.6)$$

где $\text{Var} f = M_{10} + M_{01} + M_{11}$ — вариация в смысле Харди-Краузе функции f . Это и есть частный случай неравенства Коксмы-Хлавки, которое в общем s -мерном случае имеет сходный вид

$$|R_N(f)| \leq \text{Var} f \cdot \frac{D_s(X_1, \dots, X_N)}{N}. \quad (3.2.7)$$

Отметим, что использование другой метрики, например, метрики $\mathbf{L}^2$ в предположении, что f'_x, f'_y, f''_{xy} интегрируемы с квадратом, не позволяет получить похожего неравенства, точного в упомянутом классе функций (см. приложение, где также обсуждается вопрос об оценке остатка в случае существования у f производных высших порядков).

Дискрепанс служит мерой отклонения множества точек $X_1, \dots, X_N$ от равномерного расположения в единичном s -мерном гиперкубе. Метод квази Монте-Карло использует последовательности точек $X_1, X_2, \dots$, для которых можно построить последовательность чисел $D_1(X_1, X_2), \dots, D_s(X_1, \dots, X_N)$. Известно, что эти числа возрастают с ростом N . Представляют интерес последовательности $X_1, X_2, \dots$ с наименее быстрым увеличением дискрепансов. Построенные сегодня с помощью теоретико-числовых методов последовательности обеспечивают порядок роста $O(\ln^s N)$. Наиболее удачными, по-видимому, являются последовательности И.М. Соболя [39], используются также последовательности Дж. Холтона, Форе и другие. Они достаточно полно описаны в литературе. Имеются программы (например, в пакете R), вычисляющие координаты соответствующих точек. По этой причине мы не будем останавливаться на вопросе их конструирования. Отметим лишь следующее.

1. В отличие от псевдослучайных чисел, квазислучайные числа (последовательности с малыми дискрепансами) настроены на определённую размерность интеграла.
2. Со статистической точки зрения квазислучайные числа сильно коррелированы и в задачах моделирования должны использоваться с осторожностью.

Таким образом, последовательности квазислучайных чисел обладают следующими полезными свойствами (преимуществами).

1. Позволяют вычислять интегралы по единичному s -мерному гиперкубу при фиксированном s с погрешностью порядка $\ln^s N / N$ от широкого класса функций.
2. Обладают улучшенными по сравнению с квазислучайными числами свойствами равномерной распределённости, что может быть очень важным для общих методов глобальной оптимизации.

Определёнными недостатками квазислучайного подхода можно считать.

1. Квазислучайные векторы X_i зависимы в статистическом смысле, и использование их в задачах имитации требует предварительного тщательного анализа. Их координаты приближённо независимы в совокупности. Действительно, пусть координаты точек X_i произвольным образом разбиты на две группы: $X_i = (Y_i, Z_i)$ и $f(X) = f_1(Y)f_2(Z)$, тогда имеет место асимптотическое равенство

$$\frac{1}{N} \sum_{i=1}^N f(X_i) = \frac{1}{N} \sum_{j=1}^N f_1(Y_j) \frac{1}{N} \sum_{i=1}^N f_2(Z_i).$$

2. При очень большой кратности s интеграла $\ln^s N$ может играть заметную роль, и асимптотика, при которой $\frac{\ln^s N}{N}$ будет существенно меньше $\frac{1}{\sqrt{N}}$, наступит лишь при очень больших значениях N . Ситуация осложняется в тех весьма распространённых случаях, когда s трудно или невозможно оценить априори. Как уже отмечалось, уже в простейших случаях, когда используется метод отбора, конструктивная размерность интеграла может быть неограниченной. При решении уравнений с помощью моделирования цепи Маркова даже без использования метода отбора, s может быть очень большим.
3. Неравенство Коксмы-Хлавки, как легко видеть, не может служить эффективным средством оценки погрешности интегрирования конкретной функции. Вычисление вариации — не менее трудная задача, чем вычисление самого интеграла. В то же время, если $f(x) \in \mathbf{L}_\mu^2$, обычный метод Монте-Карло позволяет строить доверительный интервал для значения интеграла. Для преодоления недостатка QMC -метода было предложено использовать рандомизованные квазислучайные числа (РКСЧ) или RQM — английское сокращение для названия метода. Наиболее распространённый приём введения случайности состоит в том, что вместо любой квазислучайной координаты x в вычислениях участвует $\{x + \alpha\}$, где α — псевдослучайное число, $\{a\}$ обозначает дробную долю числа a .

Пусть $X_1, X_2, \dots, X_N$ — квазислучайные s -мерные векторы, расположенные в s -мерном единичном кубе D_s , а $R_1, R_2, \dots, R_N$ — набор случайных векторов, равномерно распределённых в D_s . Если $\{Y\}$ обозначает вектор, компонентами которого являются дробные доли компонент вектора Y , то сумма

$$\sum_{MN} (f) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N f(\{X_n + R_m\})$$

служит оценкой интеграла $\int_{D_s} f(X) dX$. Порядок погрешности будет равен $O\left(\frac{N}{M \ln(N)}\right)$, математическое ожидание величины $\sum_{MN} (f)$ есть искомый интеграл, и её дисперсия может быть использована для построения доверительного интервала. Очевидно,

что без использования рандомизации при том же объёме вычислительной работы мы имели бы погрешность порядка $O\left(\frac{\ln(MN)}{MN}\right)$, что в $\frac{\ln(M)}{M}$ -раз лучше рандомизованной оценки. Попытка обойтись без рандомизации рассматривалась в работе [60].

Поскольку квазислучайные последовательности имеют некоторые свойства, сходные со свойствами случайных чисел (аналог закона больших чисел, например), то можно попробовать применить к ним формально статистические процедуры оценивания погрешности. Разумно предположить, что распределение в теоретико-числовом смысле средних арифметических $\frac{1}{N} \sum_{n=1}^N f(X_n)$ близко к нормальному со средним $\int_{D_s} f(X) dX$. Что касается дисперсии, то её оценка оказывается затруднительной в силу очевидной зависимости векторов $\{X_n\}_{n=1}^N$. Действительно, попытка формально оценить второй момент распределения средних приводит нас к необходимости оценивания ковариаций $f(X_n)$ и $f(X_{n+k})$ для $n = 1, 2, \dots, N$ и $k = 1, 2, \dots, N - n$. Очевидно, что корректная процедура оценивания требует привлечения более чем N значений подынтегральной функции, что опять же приводит к увеличению вычислительной сложности.

Заметим теперь, что сумма $\frac{1}{N} \sum_{j=1}^N f(X_j)$ может рассматриваться как реализация случайной кубатурной суммы $\frac{1}{N} \sum_{j=1}^N f(\{X_j + R\})$. Квазислучайные узлы, однако, обладают свойством расслоения. Если случайные точки попадают в гиперпараллелепипед объёма d , содержащийся в единичном гиперкубе D_s с вероятностью d , то из N квазислучайных точек такому параллелепипеду принадлежит N/d точек с вероятностью, близкой к единице. Для последовательностей Соболя, связанных с функциями Хаара, указанное свойство выполняется при делении гиперкуба на параллелепипеды со сторонами, равными 2^{-m_l} , $l = 1, 2, \dots, s$ при некоторых целых положительных m_s , и зависимость между квазислучайными узлами интегрирования выражается в принадлежности с вероятностью 1 некоторых подмножеств узлов различным гиперпараллелепипедам. Использование этих соображений позволяет привлечь теорию случайных квадратурных формул [7]. Для построения доверительного интервала привлекаются случайные квадратуры, точные для функций Хаара. Их дисперсия вычисляется без привлечения упоминавшихся ковариаций. Вычислительные примеры для интегралов кратности $s \leq 10$ показали удовлетворительные результаты.

Глава 4

Нелинейные задачи

4.1 Теоретические основы. Параллелизм

Решение нелинейных уравнений с использованием методов Монте-Карло и квази Монте-Карло — область сравнительно мало исследованная. Как обычно, широко используются методы приближённой линеаризации. Мы обсудим кратко основные подходы с точки зрения ПР-свойств соответствующих алгоритмов.

1. Для уравнений с полиномиальной нелинейностью вида

$$\begin{aligned}\varphi(X) &= \sum_{l=1}^{\infty} \int d\mu^l K_l(X, X_1, X_2, \dots, X_l) \prod_{m=1}^{\infty} \varphi(X_m) + f(X), \\ d\mu^l &= \bigotimes_{m=1}^l \mu(dX_m)\end{aligned}\tag{4.1.1}$$

при соответствующих предположениях об абсолютной сходимости рядов и существовании интегралов можно указать методы представления решения (функционалов от решения) в виде интеграла по траекториям ветвящегося марковского процесса.

Необходимым условием для этого является сходимость в некоторой метрике итерационного процесса

$$\begin{aligned}\bar{\varphi}_{n+1}(X) &= \sum_{l=1}^{\infty} \int d\mu^l |K_l(X, X_1, X_2, \dots, X_l)| \prod_{m=1}^{\infty} \bar{\varphi}(X_m) + |f(X)|, \\ n &= 0, 1, \dots, \quad \bar{\varphi}_0 = |f|.\end{aligned}\tag{4.1.2}$$

При вычислении функционалов $\int \varphi(X) h_j(X) \mu(dX) = (\varphi_n, h_j)$, $j = 1, 2, \dots$ достаточно предполагать слабую сходимость (для функционалов (φ_n, h_j)). При сделанных предположениях линеаризация (одна из простейших [42]), если положить

$\psi^j(X_1, \dots, X_j) = \prod_{l=1}^j \varphi(X_l)$, даст

$$\begin{aligned}\psi^{(j+1)}(X, X_1, \dots, X_j) &= \\ \sum_{l=1}^{\infty} \int d\mu^l K_l(X, y_1, X_2, \dots, X_l) \psi^{(l+j)}(X_1, \dots, X_j, y_1, \dots, y_j) + \\ &+ |f(X)| \psi^j(X_1, \dots, X_j), \quad j = 0, 1, \dots\end{aligned}$$

бесконечную систему линейных уравнений относительно ψ_j . Общая схема, описанная в § 2.2, позволяет строить ветвящийся марковский процесс для решения этой системы.

Известные методы обработки траекторий процесса (деревьев) — метод поколений и лексикографический, определяют структуру ПР-алгоритма решения уравнений вида (4.1.1).

Соответствующие алгоритмы для решения систем алгебраических уравнений с квадратичной нелинейностью довольно подробно исследованы в работах С.М. Ермакова и И.В. Калошина [43], [44]. Уравнения вида $\Delta U = F(u)$, где F — аналитическая функция, рассматривались Г.А. Михайловым [47], А.С. Расуловым [24]. В этих работах приводятся примеры Эффективного использования аппарата ветвящихся процессов для решения уравнений. Приложения к решению уравнения Больцмана рассматривались Н.М. Москалёвой [46], Г. А. Михайловым [47].

Как и в линейном случае, условие (4.1.2) является довольно ограничительным. Некоторые приёмы снятия этого ограничения, основанные на специальной замене переменных, указаны в работах Сережиной [48].

2. Значительно более широкий класс методов решения нелинейных задач использует последовательные приближения, основанные на линеаризации. Если при этом не выполняются условия мажорантного типа (вида (4.1.2)), то получить представление решения в виде интеграла по траекториям не удастся и необходимо запоминание промежуточных результатов.

Пусть для решения уравнения $\varphi = F(\varphi)$, где φ — функция $\varphi = \varphi(X)$, $X \in D \subset \mathbf{R}^s$, а F — нелинейный оператор, используется метод последовательных приближений

$$\varphi_{n+1} = F(\varphi_n), \quad n = 0, 1, \dots, \quad (4.1.3)$$

и при каждом n используется алгоритм метода Монте-Карло, позволяющий вычислить $\varphi_n(X)$ с заданной точностью (в смысле некоторой функциональной метрики). Как мы видели, параллелизм рассмотренных ранее для линейного случая алгоритмов был тесно связан с несмещённостью оценок решения при каждом m . Для линейных операторов широкого класса L выполняется соотношение

$$EL\left(\frac{1}{N} \sum_{i=1}^N \xi_i\right) = L(E\xi), \quad (4.1.4)$$

где ξ_i — независимые в совокупности реализации случайной величины ξ . Для нелинейного оператора K соотношение такого рода, вообще говоря, не выполняется. Однако, если F — нелинейная функция с нужными свойствами гладкости, то

$$F(\bar{\xi}) = F(E\xi) + F'(E\xi)(\bar{\xi} - E\xi) + F''(E\xi)\frac{(\bar{\xi} - E\xi)^2}{2!} + \dots, \quad \text{где } \bar{\xi} = \frac{1}{N} \sum_{i=1}^N \xi_i \quad \text{и}$$

$$EF(\bar{\xi}) = F(E\xi) + \frac{F''(E\xi)}{2!} \text{Var}\xi + O(N^{-3/2}). \quad (4.1.5)$$

Аналогичные результаты могут быть получены для оператора F , если производные понимать в смысле Фреше. Из (4.1.5) следует, что смещение имеет большой

порядок малости $O(1/N)$ по сравнению с погрешностью математического ожидания — $O(N^{-1/2})$.

Аналогично вычисляется главный член дисперсии $F(\bar{\xi})$. При достаточно широких предположениях распределение $F(\bar{\xi})$ асимптотически нормально [49].

Изложенные соображения позволяют исследовать различные методы последовательной линеаризации, в том числе метод Ньютона. Метод Ньютона и его модификации являются одними из наиболее популярных методов решения нелинейных уравнений. В случае операторных уравнений — это метод Ньютона-Канторовича.

Напомним некоторые определения, относящиеся к операторному случаю.

Определение 1 $\mathcal{B}$ и $\mathcal{C}$ — банаховы пространства. Будем обозначать $\mathcal{L}(\mathcal{B}, \mathcal{C})$ банахово пространство линейных операторов из $\mathcal{B}$ в $\mathcal{C}$, норма в котором является операторной нормой, порожденной пространствами $\mathcal{B}$ и $\mathcal{C}$. Положим $\mathcal{L}(\mathcal{B}) := \mathcal{L}(\mathcal{B}, \mathcal{B})$.

Определение 2 Пусть $\mathcal{B}$ и $\mathcal{C}$ — банаховы пространства. Пусть D — открытое подмножество в $\mathcal{B}$ и A — оператор, отображающий D в $\mathcal{C}$. Оператор A называется дифференцируемым по Фреше в точке $g \in D$, если существует такой оператор $L \in \mathcal{L}(\mathcal{B}, \mathcal{C})$, что

$$\lim_{\|h\| \rightarrow 0} \|A(g+h) - Ag - Lh\| / \|h\| = 0. \quad (4.1.6)$$

Оператор L , который обозначают обычно $A'(g)$, называют производной Фреше оператора A в g .

Определение 3 Последовательностью Ньютона будем называть последовательность величин

$$f_{n+1} = f_n - (B'(f_n))^{-1} B f_n, \quad n \geq 0. \quad (4.1.7)$$

Следующая теорема дает достаточные условия сходимости метода Ньютона в случае банаховых пространств (считаем, что метрика и топология порождаются нормой).

Следующая теорема, принадлежащая Канторовичу, дает критерий сходимости и устанавливает существование решения в общем случае.

Теорема 6 Пусть $\mathcal{B}$ — банахово пространство и заданы $r > 0$ и $f_0 \in \mathcal{B}$. Предположим, что оператор $B : \bar{S}(f_0, r) \rightarrow \mathcal{B}$ дифференцируем по Фреше и $(B'(f_0))^{-1} \in \mathcal{L}(\mathcal{B})$. Предположим, кроме того, что B удовлетворяет следующему условию Липшица в $\bar{S}(f_0, r)$:

$$\|B'(f) - B'(g)\| \leq p \|f - g\|. \quad (4.1.8)$$

Допустим, что

$$b_0 \geq \|B'(f_0)\|^{-1}, \quad \eta_0 \geq \|f_1 - f_0\| = \|(B'(f_0))^{-1} B f_0\|, \quad (4.1.9)$$

и положим $h_0 = b_0 p \eta_0$. Если

$$h_0 \leq 1/2, \quad (1 - (1 - 2h_0)^{1/2}) h_0^{-1} \eta_0 \leq r, \quad (4.1.10)$$

то последовательность Ньютона (4.1.7) с начальным приближением f_0 сходится к решению $\bar{f}$ уравнения $Bf = 0$ в $\bar{S}(f_0, r)$.

Основные недостатки метода Ньютона, как вычислительного средства, следующие:

- 1) чувствительность к выбору начального приближения, которая проявляется, даже в конечномерном случае. При этом часто теорема 6 дает неоправданно "пессимистические" условия сходимости;
- 2) необходимо на каждой итерации построения последовательности Ньютона вычислять производную Фреше;
- 3) необходимо на каждой итерации построения последовательности Ньютона решать линейное уравнение для обращения производной Фреше.

На практике процедуры расчета производной Фреше и ее обращения могут быть очень трудоемкими.

Для преодоления второй и третьей проблемы в ряде работ были предложены модификации метода Ньютона. В частности, в одной из модификаций используется последовательность, определяемая формулой

$$f_{n+1} = f_n - (B'(f_0))^{-1} B f_n, \quad n \geq 0, \quad (4.1.11)$$

и требуется только одно вычисление производной Фреше. При этом скорость сходимости уменьшается и становится геометрической.

Рандомизация и квази рандомизация метода Ньютона основана на применении метода (квази) Монте-Карло к обращению линейного оператора B' , входящего в (4.1.8). В конечномерном случае, когда $F(x) = (f_1(x), \dots, f_m(x))^T$, $B' \equiv F'$ есть матрица Якоби, и при каждом n задача сводится к решению системы линейных алгебраических уравнений. Как неоднократно отмечалось, при определённых условиях рандомизация позволяет строить ПР-алгоритм решения такой системы. Если случайный процесс с дискретным временем получения последовательности приближений к точному решению будет стохастически устойчивым, то можно моделировать различные независимые реализации процесса на отдельных вычислительных устройствах ("глобальный" или крупнозернистый параллелизм). Кроме полезного свойства параллелизма алгоритм позволяет также строить доверительные эллипсоиды для последовательных приближений к решению.

Проблема, которая возникает при наличии многих процессоров, состоит в следующем. Относительно малое число независимых испытаний на каждом из процессоров может привести к ухудшению порядка сходимости методов (1-ый порядок) и к потере стохастической устойчивости. Большое число испытаний приводит к использованию параллелизма в основном при решении линейной задачи. В любом случае важно понимать существо проблемы. О рандомизации метода Ньютона см. [50] и [51].

Как обычно, метод Монте-Карло может быть эффективным, если нас удовлетворяет сравнительно грубое приближённое решение и s очень велико. Предполагая, что выполнены достаточные условия применимости метода Ньютона (используем первые члены разложения в ряд Тейлора и строим метод последовательных приближений

$$x_i^{(l+1)} = f_i(X^{(l)}) + \sum_{j=1}^s \frac{\partial f_i}{\partial x_j}(X^{(l)})(x_j^{(l+1)} - x_j^{(l)}), \quad i = 1, \dots, s, \quad l = 0, 1, \dots \quad (4.1.12)$$

При каждом l полученную систему уравнений для X^{l+1} решаем методом Монте-Карло или квази Монте-Карло, используя описанные ранее методы.

Выбор соответствующей модификации зависит от свойств матрицы $A = \left\| \frac{\partial f_i(X^l)}{\partial x_j} \right\|_{i,j=1}^s$. Если окажется, что $\|A\| > 1$, то необходимо использовать специальные приёмы введения

псевдообуславливателя или применять рандомизованные методы вычисления определителей [52], а также случайный поиск (минимизация невязки).

Параллелизм. При фиксированном m используются параллельные технологии для решения систем линейных уравнений, описанные ранее. При определённых условиях (стохастическая устойчивость) вычисления по формулам (4.1.12) можно осуществлять на различных процессорах. Полученные оценки $\xi_i^{(m)}$ затем осредняются. Среднее при условии стохастической устойчивости распределено приблизительно нормально, и может быть подсчитана эмпирическая дисперсия, но $E\xi_i^{(m)} \neq x_i^{(m)}$ ввиду наличия смещения, которое имеет порядок $O(1/N_1)$, где N_1 — число независимых испытаний, осуществляемых каждым процессором при решении линейных систем. Для простоты мы считаем это число одинаковым для каждого процессора. Следует иметь в виду, что “внешнее” осреднение (между k процессорами) уменьшает случайную ошибку в $\frac{1}{\sqrt{k}}$ раз, но сохраняет порядок систематической. Это означает, что N_1 должно заметно превосходить число процессоров для того, чтобы систематической погрешностью можно было пренебречь.

4.2 Примеры

Пример 4.1. Ниже приводится пример конструирования и исследования ПР-алгоритма решения конкретной прикладной задачи расчёта цены американского опциона. Это задача с подвижной границей. Она сводится к нелинейной краевой задаче для параболического уравнения в частных производных и достаточно трудна, чтобы служить иллюстрацией упомянутых выше подходов. Американский опцион является частным случаем класса ценных бумаг, цена которых зависит от пройденного пути, то есть, нас интересует не только цена в конечный момент T , но и промежуточные состояния.

Стоимость акций, на которых построен опцион обозначим $S(t)$, а период существования опциона через $[0, T]$. Цена опциона удовлетворяет следующему уравнению в частных производных [53]

$$\frac{\partial P}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 P}{\partial S^2} + rS \frac{\partial P}{\partial S} - rP = 0, \quad (4.2.1)$$

где

- t — время
- S — цена акции
- $P = P(S, t)$ — цена опциона
- σ — постоянная волатильность
- r — безрисковая ставка.

Предположения и ограничения, при которых это уравнение имеет место, можно найти в доступной литературе (например, [59]).

Наиболее известными двумя подходами, позволяющими устранить подвижную границу, являются специальная замена переменных и метод штрафных функций [55]. После замены переменных полученное нелинейное уравнение решают методом последовательной линеаризации (Ньютона).

Мы рассмотрим метод штрафных функций совместно с дискретизацией задачи, что дает более простые алгоритмы сравнительно с методом Ньютона.

Суть метода заключается в добавлении к уравнению в частных производных небольшого слагаемого, которое представляет собой непрерывную функцию $f(P)$, нелинейно зависящую от цены опциона P . При этом получается нелинейное дифференциальное уравнение с фиксированной областью определения. Как показано в работе [55], решение этого уравнения будет обладать свойствам аналитического решения задачи нахождения цены Американского опциона при достаточной малости добавляемого слагаемого. Рассмотрим уравнение Блэка-Шоулса с такой добавкой:

$$\frac{\partial P_\varepsilon}{\partial t} + F + f_\varepsilon(P_\varepsilon) = 0, \quad (4.2.2)$$

с граничными условиями

$$\begin{aligned} P(S, T) &= \max(K - S, 0), \\ \lim_{S \rightarrow \infty} P(S, t) &= 0, \\ P(0, t) &= K. \end{aligned} \quad (4.2.3)$$

Здесь индексация призвана подчеркнуть зависимость от ε . Справедлива следующая теорема [55], [56]:

Теорема 7 Пусть P — единственное решение уравнения Блэка-Шоулса, а P_ε — единственное решение уравнения (4.2.2) для любого $\varepsilon > 0$. Тогда $P_\varepsilon \rightarrow P$ в $L_\infty^{loc}(\overline{Q}_T)$ при $\varepsilon \rightarrow 0$. Где $Q_T = [0, T] \times Q$, $Q \subset \mathbb{R}^m$ открытое и ограниченное.

Далее индекс ε будет опускаться. Пусть $0 < \varepsilon \ll 1$ и

$$\frac{\partial P}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 P}{\partial S^2} + rS \frac{\partial P}{\partial S} - rP + f(P) = 0, \quad S \geq 0, \quad 0 \leq t < T, \quad (4.2.4)$$

где

$$f(P) = \frac{\varepsilon C}{P + \varepsilon - K + S}. \quad (4.2.5)$$

Здесь $C \geq rK$ — положительная константа.

Как уже отмечалось, при стремлении ε к нулю, решение задачи (4.2.4), (4.2.5), (4.2.3) будет стремиться к цене Американского опциона. Получившаяся задача является нелинейной с фиксированной областью определения.

Для численного решения задачи (4.2.4), (4.2.5), (4.2.3) могут быть использованы различные методы, их обзор можно найти в [3]. Мы рассмотрим метод конечных разностей.

Обратим внимание также на следующий интересный факт. При дискретизации мы переходим от задачи, где время представлено непрерывно, к дискретному времени. Рассматривая список временных моментов, когда опцион может быть исполнен, мы переходим, таким образом, к формальному рассмотрению так называемого Бермудского опциона.

Для решения задачи (4.2.4), (4.2.5), (4.2.3) методом конечных разностей в области $S \geq 0$, $0 \leq t \leq T$, как и в [3], введем сетку:

$$\begin{aligned} \Delta S &= \frac{S_\infty}{M}, \quad \Delta t = \frac{T}{N}, \\ S_i &= i\Delta S, \quad i = 0, \dots, M, \\ t_j &= j\Delta t, \quad j = 0, \dots, N, \\ P_{i,j} &= P(S_i, t_j). \end{aligned} \quad (4.2.6)$$

Полунеявная разностная схема для уравнения (4.2.4) имеет вид

$$\begin{aligned} & \frac{P_{i,j+1} - P_{i,j}}{\Delta t} + \frac{(i\Delta S)^2 \sigma^2}{2} \frac{P_{i-1,j} - 2P_{i,j} + P_{i+1,j}}{(\Delta S)^2} + i\Delta S r \frac{P_{i+1,j} - P_{i-1,j}}{2\Delta S} - rP_{i,j} + \\ & + \frac{\varepsilon C}{P_{i,j+1} + \varepsilon - K + i\Delta S} = 0, \quad i = 1 \dots M-1, \quad j = 0 \dots N-1, \end{aligned} \quad (4.2.7)$$

$$\begin{aligned} P_{i,N} &= \max(K - i\Delta S, 0), \quad i = 1 \dots M-1, \\ P_{M,j} &= 0, \quad j = 0 \dots N-1, \\ P_{0,j} &= K, \quad j = 0 \dots N-1. \end{aligned}$$

После перегруппировки слагаемых получим

$$\begin{aligned} P_{i,j} &= \frac{i^2 \sigma^2 \Delta t - ri\Delta t}{2(1 + i^2 \sigma^2 \Delta t + r\Delta t)} P_{i-1,j} + \frac{i^2 \sigma^2 \Delta t + ri\Delta t}{2(1 + i^2 \sigma^2 \Delta t + r\Delta t)} P_{i+1,j} + \\ & + \frac{P_{i,j+1}}{(1 + i^2 \sigma^2 \Delta t + r\Delta t)} + \frac{\Delta t \varepsilon C}{(1 + i^2 \sigma^2 \Delta t + r\Delta t)(P_{i,j+1} + \varepsilon - K + i\Delta S)}, \end{aligned} \quad (4.2.8)$$

$$i = 1 \dots M-1, \quad j = 0 \dots N-1.$$

Заметим, что в общем эта система является нелинейной, однако на каждом шаге по времени получается линейная система вида

$$X_j = AX_j + F(X_{j+1}), \quad (4.2.9)$$

$$A = \begin{pmatrix} 0 & c_1 & 0 & \dots & 0 & 0 \\ a_2 & 0 & c_2 & \dots & 0 & 0 \\ 0 & a_3 & 0 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 0 & c_{M-2} \\ 0 & 0 & 0 & \dots & a_{M-1} & 0 \end{pmatrix}, \quad (4.2.10)$$

где

$$\begin{aligned} a_i &= \frac{i^2 \sigma^2 \Delta t - ri\Delta t}{2(1 + i^2 \sigma^2 \Delta t + r\Delta t)}, \\ c_i &= \frac{i^2 \sigma^2 \Delta t + ri\Delta t}{2(1 + i^2 \sigma^2 \Delta t + r\Delta t)}, \\ S_i &= i\Delta S, \\ i &= 1, \dots, M-1, \end{aligned} \quad (4.2.11)$$

а вектор-функция F :

$$F((x_1, x_2, \dots, x_{M-1})) = (f_1(x_1), f_2(x_2), \dots, f_{M-1}(x_{M-1}))^T, \quad (4.2.12)$$

$$f_i(x_i) = \frac{x_i}{(1 + i^2 \sigma^2 \Delta t + r\Delta t)} + \frac{\Delta t \varepsilon C}{(x_i + \varepsilon - K + i\Delta S)(1 + i^2 \sigma^2 \Delta t + r\Delta t)}, \quad i = 1, \dots, M-1. \quad (4.2.13)$$

Таким образом, при каждом фиксированном j система (4.2.9) является системой линейных алгебраических уравнения (X_{j+1} — вычислено на предыдущем шаге, X_j — неизвестно). Таким образом, последовательно решив N систем уравнений начиная с j равного $N-1$ и до 0, мы найдем цену опциона в начальный момент времени. Теперь мы перейдем к основному содержанию работы — применению стохастических методов и исследованию их устойчивости. Докажем следующую лемму

Лемма 8 При каждом фиксированном j существует $\Delta t > 0$ такое, что спектральный радиус матрицы $|A|$ (4.2.10) меньше единицы.

Доказательство:

Рассмотрим сумму модулей строки матрицы $|A|$:

$$|a_i| + |c_i| = \frac{\Delta t(|i^2\sigma^2 - ri| + i^2\sigma^2 + ri)}{2(1 + i^2\sigma^2\Delta t + r\Delta t)}, \quad i = 1, \dots, M-1. \quad (4.2.14)$$

При фиксированном ΔS , уменьшая Δt , можно сделать сумму меньше единицы для любого i , сделав норму (в качестве нормы берётся максимум из сумм модулей элементов строк) матрицы $|A|$, а тем самым и её спектральный радиус, меньше единицы. Можно получить оценку для Δt ,

$$\frac{\Delta t(|i^2\sigma^2 - ri| + i^2\sigma^2 + ri)}{2(1 + i^2\sigma^2\Delta t + r\Delta t)} = \frac{\Delta t \max(i^2\sigma^2, ri)}{1 + i^2\sigma^2\Delta t + r\Delta t} < 1. \quad (4.2.15)$$

Заметим, что неравенство

$$\frac{i^2\sigma^2\Delta t}{1 + i^2\sigma^2\Delta t + r\Delta t} < 1 \quad (4.2.16)$$

выполнено для любых Δt и ΔS . Рассмотрим второе неравенство

$$\frac{ri\Delta t}{1 + i^2\sigma^2\Delta t + r\Delta t} < 1. \quad (4.2.17)$$

Напомним, что $1 \leq i \leq M$, поэтому

$$\frac{ri\Delta t}{1 + i^2\sigma^2\Delta t + r\Delta t} \leq \frac{rM\Delta t}{1 + \sigma^2\Delta t + r\Delta t} < 1. \quad (4.2.18)$$

После преобразований получим

$$\Delta t(r(M-1) - \sigma^2) < 1. \quad (4.2.19)$$

При $r(M-1) - \sigma^2 \leq 0$ предыдущее неравенство, а следовательно и (4.2.15), выполнено. Если же $r(M-1) - \sigma^2 > 0$, то для выполнения неравенства (4.2.15), достаточно следующего условия на Δt :

$$\Delta t < \frac{1}{r(M-1) - \sigma^2}. \quad (4.2.20)$$

■

Доказанная лемма означает, что можем находить решение системы методом Монте-Карло (возможно также применение квази Монте-Карло).

В этом случае вычисления будут производиться по формуле

$$\xi_j = (\widetilde{I - A})^{-1} F(\hat{\xi}_{j+1}(N_{j+1})), \quad (4.2.21)$$

где волна обозначает, что соответствующий результат обращения матрицы и умножения её на вектор осуществляется посредством метода Монте-Карло, а $\hat{\xi}_{j+1}(N_{j+1})$ — среднее N_{j+1} испытаний на $(j+1)$ -ом слое. При переходе со слоя на слой мы будем иметь дело с двумя видами ошибок: случайной и детерминированной. При решении системы (4.2.9) используются несмещенные оценки, однако, в силу нелинейности функции $F(X)$, будет

возникать смещение при переходе со слоя на слой. Возьмем математическое ожидание от правой части (4.2.21):

$$\begin{aligned} E(\widetilde{(I-A)^{-1}F(\hat{\xi}_{j+1}(N_{j+1}))}) = \\ E(\widetilde{(I-A)^{-1}})E(F(\hat{\xi}_{j+1}(N_{j+1}))) = (I-A)^{-1}E(F(\hat{\xi}_{j+1}(N_{j+1}))). \end{aligned} \quad (4.2.22)$$

Первое равенство выполнено в силу того, что на разных слоях используются независимы траектории. Рассмотрим разложение $F(X)$ в ряд Тейлора:

$$\begin{aligned} F(\hat{\xi}_{j+1}(N_{j+1})) = \\ F\left(\frac{1}{N_{j+1}} \sum_{l=1}^{N_{j+1}} \xi_{j+1}^{(l)}\right) = F(E\xi_{j+1}) + F'(E\xi_{j+1})\left(\frac{1}{N_{j+1}} \sum_{l=1}^{N_{j+1}} \xi_{j+1}^{(l)} - E\xi_{j+1}\right) + \\ + \frac{1}{2}F''(E\xi_{j+1})\left(\frac{1}{N_{j+1}} \sum_{l=1}^{N_{j+1}} \xi_{j+1}^{(l)} - E\xi_{j+1}\right)^2 + \dots \end{aligned} \quad (4.2.23)$$

Возьмем математическое ожидание от левой и правой части получившегося равенства, предварительно отметив, что $E\xi_{j+1}^{(l)} = E\xi_{j+1}$

$$E(F(\hat{\xi}_{j+1}(N_{j+1}))) = F(E\xi_{j+1}) + F''(E\xi_{j+1})\frac{E\xi_{j+1}^2 - (E\xi_{j+1})^2}{2N_{j+1}} + \dots, \quad (4.2.24)$$

где

$$F''(X) = \begin{pmatrix} f_1'' & 0 & \dots & 0 \\ 0 & f_2'' & 0 & \dots \\ \vdots & & f_3'' & \vdots \\ & & & \ddots \\ 0 & \dots & & f_{M-2}'' & 0 \\ & & & 0 & f_{M-1}'' \end{pmatrix}, \quad (4.2.25)$$

$$f_i'' = \frac{\Delta t \varepsilon C}{(1 + i^2 \sigma^2 \Delta t + r \Delta t)(P_{i,j+1} + \varepsilon - K + i \Delta S)^3}, \quad i = 1, \dots, M-1. \quad (4.2.26)$$

В [5] было показано, что при $\Delta t \leq \varepsilon/C$ решение уравнения (4.2.9) удовлетворяет неравенству

$$P_{i,j} \geq \max(K - S_i, 0) \quad \forall i, j. \quad (4.2.27)$$

Оценим f_i'' :

$$\begin{aligned} f_i'' &\leq \frac{\Delta t \varepsilon C}{(1 + i^2 \sigma^2 \Delta t + r \Delta t)(\max(K - S_i, 0) + \varepsilon - K + i \Delta S)^3} \leq \frac{\Delta t C}{(1 + i^2 \sigma^2 \Delta t + r \Delta t) \varepsilon^2} \leq \\ &\leq \frac{1}{(1 + i^2 \sigma^2 \Delta t + r \Delta t) \varepsilon}. \end{aligned} \quad (4.2.28)$$

Из этого следует, что при достаточно больших N_j (порядка $1/\varepsilon$ и больше) смещением можно будет пренебречь.

Перейдем к рассмотрению случайной ошибки. При переходе по времени со слоя на слой эта ошибка может оставаться ограниченной, а может неограниченно расти с ростом N . Поэтому необходимо исследовать её поведение. Обозначим $\varepsilon_j = \xi_j - X_j$ — вектор-столбец случайных ошибок на каждом слое. Выразим $\text{cov} \varepsilon_j$ через $\text{cov} \varepsilon_{j+1}$ в удобной для дальнейшего анализа форме. Из (4.2.21) следует

$$X_j + \varepsilon_j = \widetilde{(I-A)^{-1}}(F(X_{j+1} + \varepsilon_{j+1})). \quad (4.2.29)$$

Напишем разложение функции F в ряд Тейлора:

$$F(X_j + \varepsilon_j) = F(X_j) + F'(X_j)\varepsilon_j + F''(X_j)\varepsilon_j^2 + \dots \quad (4.2.30)$$

Отметим, что при достаточно малых ε и Δt слагаемым $F''(X_j)\varepsilon_j^2$ и всеми последующими можно будет пренебречь. То есть $F(X_j + \varepsilon_j) \approx F(X_j) + F'(X_j)\varepsilon_j$. Положим также $(\widetilde{I - A})^{-1} = (I - A)^{-1} + \delta$, где $E\delta = 0$. Равенство (4.2.29) перепишется

$$X_j + \varepsilon_j = ((I - A)^{-1} + \delta)(F(X_{j+1}) + F'(X_{j+1})\varepsilon_{j+1}). \quad (4.2.31)$$

Учитывая, что X_j удовлетворяет (4.2.9), получим для ε_j следующее выражение

$$\varepsilon_j = \delta F(X_{j+1}) + (I - A)^{-1}F'(X_{j+1})\varepsilon_{j+1} + \delta F'(X_{j+1})\varepsilon_{j+1} \quad (4.2.32)$$

и соответственно для ε_j^T , заметив прежде, что F' — диагональная матрица и следовательно $(F')^T = F'$,

$$\varepsilon_j^T = F(X_{j+1})^T \delta^T + \varepsilon_{j+1}^T F'(X_{j+1})((I - A)^{-1})^T + \varepsilon_{j+1}^T F'(X_{j+1})\delta^T. \quad (4.2.33)$$

Методами, описанными в [6], можно прийти к

$$\|Y_L^j\| = (\mathcal{A} + \mathcal{D})\|Y_L^{j+1}\| + \mathcal{F}_{j+1}, \quad (4.2.34)$$

где Y_L^j — векторизация матрицы $\text{cov}\varepsilon_j$; $\mathcal{A} = \|\alpha_{i_0, i_1} \alpha_{i_3, i_2}\|$ — кронекеровское произведение, $\mathcal{D} = \|f'_{i_1} E(\delta_{i_0, i_1} \delta_{i_3, i_2}) f'_{i_2}\|$ — матрицы $(M-1)^2 \times (M-1)^2$, i_0, i_1, i_2, i_3 меняются от 1 до $M-1$, $\alpha_{l, k}$ — элементы матрицы $(I - A)^{-1}F'(X_{j+1})$; матрица $\mathcal{F}_{j+1}$ объединяет слагаемые, не зависящие от ε_{j+1} . При этом собственные векторы $\mathcal{A}$ есть $\varphi_i \varphi_j^T$, $i, j = 1, \dots, M-1$, где φ_i — собственный вектор-столбец матрицы $(I - A)^{-1}F'$, а собственные числа $\mathcal{A}$ есть $\lambda_i \lambda_j$, $i, j = 1, \dots, M-1$, где λ_i — собственные числа матрицы $(I - A)^{-1}F'$. Матрица $\mathcal{D}$ представляет собой произведение $\mathcal{M}\hat{F}$, где $\hat{F}$ — диагональная матрица с диагональю $\|f'_i f'_j\|_{(i,j)=(1,1)}^{(M-1, M-1)}$, а элементы матрицы $\mathcal{M}$ есть ковариации вектора, составленного из элементов матрицы погрешностей δ . Элементы матрицы $\mathcal{M}$ имеют порядок $1/N_j$, где N_j — количество независимых траекторий на слое с номером j . Оценим f'_i , выражение для которого имеет вид

$$f'_i = \frac{1}{1 + i^2 \sigma^2 \Delta t + r \Delta t} \left(1 - \frac{\Delta t \varepsilon C}{(P_{i, j+1} + \varepsilon - K + i \Delta S)^2} \right), \quad (4.2.35)$$

найдем условие на Δt , при котором разность, стоящая в скобках, больше нуля

$$1 - \frac{\Delta t \varepsilon C}{(P_{i, j+1} + \varepsilon - K + i \Delta S)^2} \geq 1 - \frac{\Delta t \varepsilon C}{(\varepsilon - K)^2} \geq 0, \quad (4.2.36)$$

откуда получаем условие

$$\Delta t \leq \frac{(\varepsilon - K)^2}{\varepsilon C}. \quad (4.2.37)$$

Таким образом, норма матрицы F' меньше единицы при Δt из (4.2.37).

Из (4.2.34) и вышесказанного следует следующие утверждения:

Утверждение 4.2.1 Для стохастической устойчивости алгоритма (4.2.21) необходимо и достаточно, чтобы максимум из модулей первых собственных чисел матриц $\mathcal{A} + \frac{1}{N_j} \mathcal{M}'$, $j = N-1 \dots 1$ при любых натуральных N был меньше единицы.

Утверждение 4.2.2 Если первые собственные числа матриц $(I - A)^{-1}F'_j$ по модулю строго меньше единицы, то существует такой набор N_j , $j = N - 1, \dots, 1$, что при всех $N_k > N_j$ алгоритм (4.2.21) будет стохастически устойчивым.

Необходимо оценить первое собственное число матрицы $(I - A)^{-1}F'_j$. Матрицу F'_j можно представить в виде произведения диагональных матриц $\tilde{F}_1 \tilde{F}_2$ с элементами, равными соответственно

$$\frac{1}{1 + i^2 \sigma^2 \Delta t + r \Delta t} \quad \text{и} \quad 1 - \frac{\Delta t \varepsilon C}{(P_{i,j+1} + \varepsilon - K + i \Delta S)^2}, \quad i = 1, \dots, M - 1. \quad (4.2.38)$$

Далее

$$(I - A)^{-1}F'_j = (\tilde{F}_1^{-1}(I - A))^{-1}\tilde{F}_2 = (I - \Delta t \tilde{A})^{-1}(I - \Delta t \tilde{F}), \quad (4.2.39)$$

где

$$\tilde{A} = \begin{pmatrix} b_1 & c_1 & 0 & \dots & 0 & 0 \\ a_2 & b_2 & c_2 & \dots & 0 & 0 \\ 0 & a_3 & b_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & b_{M-1} & c_{M-2} \\ 0 & 0 & 0 & \dots & a_{M-1} & b_{M-1} \end{pmatrix}, \quad \tilde{F} = \begin{pmatrix} \tilde{f}_1 & 0 & \dots & 0 \\ 0 & \tilde{f}_2 & 0 & \dots \\ \vdots & \tilde{f}_3 & & \vdots \\ & & \ddots & \\ & & & \tilde{f}_{M-2} & 0 \\ 0 & \dots & 0 & \tilde{f}_{M-1} \end{pmatrix} \quad (4.2.40)$$

$$a_i = \frac{i^2 \sigma^2 - ri}{2}, \quad b_i = -i^2 \sigma^2 - r, \quad c_i = \frac{i^2 \sigma^2 + ri}{2}, \quad (4.2.41)$$

$$\tilde{f}_i = \frac{\varepsilon C}{(P_{i,j+1} + \varepsilon - K + i \Delta S)^2}, \quad i = 1, \dots, M - 1.$$

При достаточно малых Δt можно считать $(I - \Delta t \tilde{A})^{-1} \approx I + \Delta t \tilde{A}$, следовательно

$$(I - \Delta t \tilde{A})^{-1}(I - \Delta t \tilde{F}) \approx (I + \Delta t \tilde{A})(I - \Delta t \tilde{F}) \approx I + (\tilde{A} - \tilde{F})\Delta t. \quad (4.2.42)$$

Если матрица $\tilde{A} - \tilde{F}$ отрицательно определённая, то при достаточно малых Δt первое собственное число матрицы $I + \Delta t(\tilde{A} - \tilde{F})$, а следовательно и матрицы $(I - A)^{-1}F'_j$, меньше единицы. Можно показать, что при $r > \sigma^2/3$ матрица $\tilde{A} - \tilde{F}$ отрицательно определённая. Также выполнение условия

$$\tilde{f}_i > \frac{(4i^2 + 4i + 1)\sigma^4 - (12i^2 + 12i + 6)\sigma^2 r - 3r^2}{4(i + 1)^2 \sigma^2 + 4i^2 \sigma^2 + 4r} \quad (4.2.43)$$

для любого i обеспечит отрицательную определённость. Это условие также легко проверять в процессе вычислений. Показать отрицательную определённость матрицы $\tilde{A} - \tilde{F}$ в общем случае не удаётся. Более того, в некоторых, далёких от реальных ситуаций, случаях возможно нарушение отрицательной определённости, что вызывает небольшие возмущения в поведении алгоритма, но сохраняет устойчивость в целом.

Результаты вычислений

Ниже приведены результаты решения систем (4.2.9) методом Монте-Карло и детерминированным методом. Расчеты производились для следующих параметров: $T = 0.75$, $S_\infty = 100$, $M = 100$, $N = 700$, $\sigma \equiv 0.15$, $r \equiv 0.055$, $K = 35$, $\varepsilon = 0.001$, $C = 2$.

На графиках изображена цена опциона в зависимости от цены акции в момент времени $t = 0$. Количество моделируемых траекторий для оценки каждого узла сетки: 50 — для первого графика, 100 — для второго. Сплошная линия соответствует решению, полученному методом Монте-Карло, (нанесены средние значения оценок. Доверительный интервал достаточно мал — имеет порядок колебаний среднего и на графиках не изображён). Пунктирная линия соответствует точно посчитанному детерминированным методом.

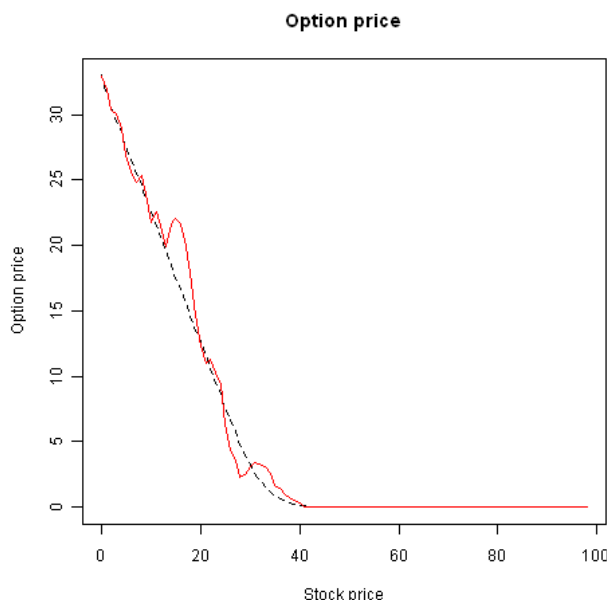


Рис. 4.1: Кол-во траекторий — 50

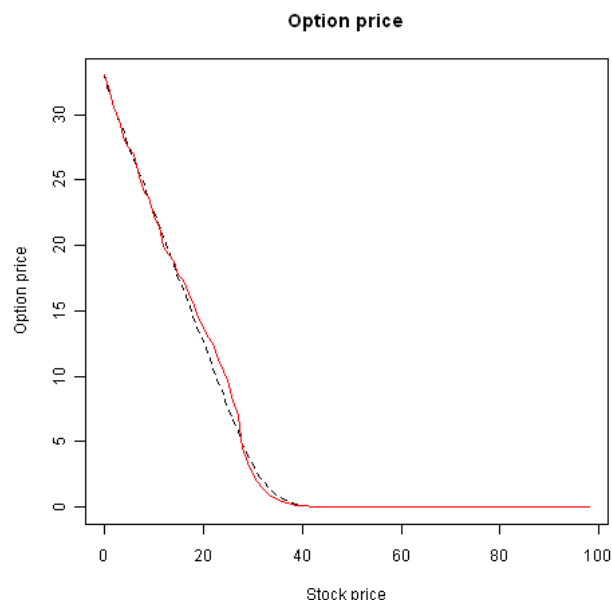


Рис. 4.2: Кол-во траекторий — 100

Также были проведены эксперименты для следующих параметров: $T = 0.75$, $S_\infty = 100$, $M = 100$, $N = 900$, $\sigma \equiv 0.7$, $r \equiv 10^{-4}$, $K = 35$, $\varepsilon = 10^{-4}$, $C = 2$. Количество моделируемых траекторий для оценки каждого узла сетки: 200 — для первого графика, 250 — для второго. Сплошная линия соответствует решению, полученному методом Монте-Карло, пунктирная — точно посчитанному детерминированным методом.

Условие (4.2.43) очевидным образом может быть наиболее ограничительным при малых r . Случай $r = 0$ не реален, но интересно было проследить поведение алгоритма при малых r .

Следующие примеры ($r = 10^{-4}$) обнаруживают нарушение устойчивости (даже детерминированной!) в начальной области (наличие “горба” на графиках) и сравнительное увеличение статистической ошибки (Рис.3, Рис.4).

Следующим примером распараллеливания достаточно сложного алгоритма решения нелинейной задачи (создания ПР-алгоритма) является пример исследования спектра линейного оператора.

Пример 4.2. Мы будем рассматривать линейный оператор A , действующий в линейном нормированном пространстве F вещественных функций (векторов). Спектр A будем предполагать дискретным. Собственные числа могут иметь комплексные значения и быть кратными. Они будут нумероваться в порядке убывания их модулей, $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n| \geq \dots$

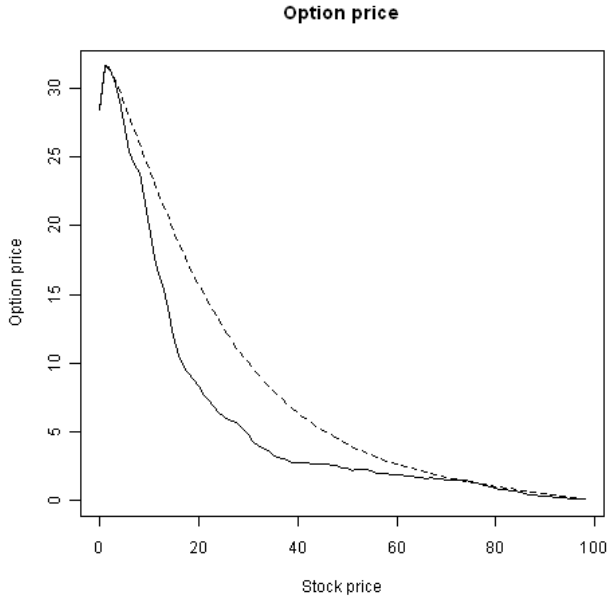


Рис. 4.3: Кол-во траекторий — 200

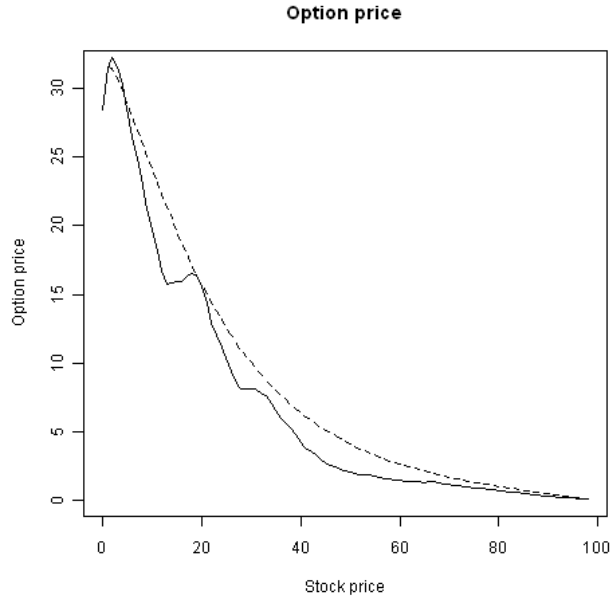


Рис. 4.4: Кол-во траекторий — 250

Предметом исследования далее будут спектральные свойства A при условии, что нам известны, может быть со случайной ошибкой (измерены, вычислены) итерации

$$\xi_m = A^m X, m = 0, 1, \dots, M \quad (4.2.44)$$

для некоторого X из F . Задачи такого рода возникают во многих приложениях. Это, например, исследования систем массового обслуживания, расчет ядерных реакторов и др.

Особенно интересным для приложений может быть случай, когда функции (векторы (4.2.44)) вычисляются с помощью метода Монте-Карло.

Обозначим через $\mu_1 = \lambda_1, \mu_2, \dots, \mu_l, \dots$ значения собственных чисел без учета их кратности, занумерованных в порядке убывания их модулей (для комплексно сопряженных чисел порядок нумерации чисел с равным модулем безразличен) и назовем многочлен степени k

$$P_k(x) = \prod_{j=1}^k (x - \mu_j), \quad (4.2.45)$$

k -минимальным многочленом оператора A .

Если F — конечномерное пространство векторов размерности n и A — матричный $n \times n$ оператор, то коэффициенты минимального многочлена A при известных итерациях (4.2.44) находятся известным методом Крылова [58]. Очевидно, что при очень больших значениях степени минимального многочлена метод Крылова практически не применим. Тем более он не применим в случае счетного спектра оператора. Далее мы опишем метод приближенного вычисления коэффициентов k -минимального многочлена для заданных, как правило небольших значений k . При этом мы будем опираться на Лемму 9. Ниже следует её формулировка и краткое доказательство.

Рассматривается последовательность вещественных чисел

$$\xi_t = \sum_{k=1}^T a_k b_k^t, \quad |\xi_t| < \infty, \quad (4.2.46)$$

где a_k и b_k — числа (м.б комплексные), все b_k различны

$$|b_1| \geq |b_2| \geq \dots \geq |b_T|, \quad (4.2.47)$$

но могут быть равными по модулю, $T > 0$, может быть бесконечным и в этом случае предполагается абсолютная сходимость ряда (4.2.46) (в последнем случае $|b_k| \leq 1$).

Лемма 9 Пусть ξ_t , $t = 1, 2, \dots$ заданы формулой (4.2.46) и пусть для некоторого L , $1 \leq L \leq T$ и возрастающей последовательности номеров t , $t = 1, 2, \dots$ при каждом t существует решение системы уравнений

$$\sum_{l=0}^L c_l^{(t)} \xi_{t+l+j} = 0, \quad j = 0, 1, \dots, L-1. \quad (4.2.48)$$

Тогда если

(1) b_k , $k = 1, \dots, T$, удовлетворяет неравенствам (4.2.47), но $|b_L| > |b_{L+1}|$, $|b_L| > 0$,

(2) $a_k \neq 0$, $k = 1, \dots, L$,

(3) для некоторого $l = l_0$, $0 \leq l_0 \leq L$, $c_{l_0}^{(t)} = 1$, $t = 0, 1, \dots$,

то существует предел

$$\lim_{t \rightarrow \infty} c_l^{(t)} = c_l, \quad l = 0, 1, \dots, L \quad (4.2.49)$$

и многочлен

$$Q(x) = \sum_{l=0}^L c_l x^l$$

имеет своими корнями числа b_k , $k = 1, \dots, L$:

$$Q(x) = c \prod_{k=1}^L (x - b_k), \quad c = \text{const.} \quad (4.2.50)$$

Доказательство:

Обозначим $Q_t(x) = \sum_{l=0}^L c_l^{(t)} x^l$ и положим $l_0 = L$ ($c_L^{(t)} \equiv 1$), что не умаляет общности.

Подставляя в (4.2.48) выражение (4.2.46) для ξ_{t+l+j} , имеем

$$\sum_{l=0}^L c_l^{(t)} \sum_{k=1}^T a_k b_k^{t+l+j} = \sum_{k=1}^T a_k b_k^{t+j} \sum_{l=0}^L c_l^{(t)} b_k^l = 0, \quad j = 0, \dots, L-1.$$

Или

$$\sum_{k=1}^L a_k b_k^{t+j} Q_t(b_k) + \sum_{k=L+1}^T a_k b_k^{t+j} Q_t(b_k) = 0, \quad j = 0, \dots, L-1.$$

Заметим далее, что $Q_t(b_k)$ ограничены $|Q_t(b_k)| \leq M_1 < \infty$, $k = 1, 2, \dots$,

в силу предположения о существовании решения системы (4.2.48) и условия $|b_k| < |b_L|$,

$k = L + 1, L + 2, \dots$ и $\sum_{k=L+1}^T |a_k| < M_2 < \infty$ (абсолютная сходимость ряда при $T \rightarrow \infty$).

Для каждого j разделим j -тое равенство на $|b_L|^{t+j}$.

Имеем

$$\sum_{k=1}^L a_k Q_t(b_k) \frac{b_k^{t+j}}{|b_L|^{t+j}} = - \sum_{k=L+1}^T a_k Q_t(b_k) \frac{b_k^{t+j}}{|b_L|^{t+j}}, \quad j = 1, 2, \dots, L. \quad (4.2.51)$$

Совокупность полученных равенств можно рассматривать как систему уравнений относительно $a_k Q_t(b_k)$ с определителем $\det \|b_k^{t+j}/|b_L|^{t+j}\|_{k,j=1}^L$. При этом

$$\left| \sum_{k=L+1}^T a_k Q_t(b_k) \frac{b_k^{t+j}}{|b_L|^{t+j}} \right| \leq M_1 \cdot M_2 \cdot q^{t+j}, \quad q = \max_{L+1 \leq k \leq T} \left| \frac{b_k}{b_L} \right|.$$

Определитель системы при каждом t отличен от нуля (определитель Вандермонда), а её правая часть стремится к нулю при $t \rightarrow \infty$. Таким образом, $a_k Q_t(b_k) \xrightarrow[t \rightarrow \infty]{} 0$, а так как

$a_k \neq 0$, $k = 1, \dots, L$, получим $\sum_{l=0}^L c_l^{(t)} b_k^l \xrightarrow[t \rightarrow \infty]{} 0$. Поскольку все b_k различны (по предположению), то многочлен $\sum_{l=0}^L c_l^{(t)} y^l$ кратных корней не имеет, и из существования предела

$\lim_{t \rightarrow \infty} \sum_{l=0}^L c_l^{(t)} b_k^l = 0$ следует существование предела $\lim_{t \rightarrow \infty} c_l^{(t)} = c_l$, что и доказывает Лемму. ■

Если A — линейный оператор в гильбертовом пространстве F , а X и Y — элементы F , то функционалы $(A^t X, Y)$, $t = 1, 2, \dots$ могут быть выбраны в качестве последовательности ξ_t , определяемой формулой (4.2.46). Если в разложении X и Y по элементам базиса F присутствуют представители всех инвариантных подпространств F , то b_k будут представлять все значения собственных чисел оператора A .

Доказанная Лемма может служить источником алгоритмов для вычисления коэффициентов k -минимального многочлена оператора. При этом следует заметить:

- при $T = L$ в случае матричного оператора вычислительный метод совпадает с методом Крылова вычисления коэффициентов минимального многочлена матрицы.
- лемма справедлива для вещественных чисел ξ_t , и реальные вычисления, вообще говоря, могут потребовать операции над числами с очень большим числом знаков.

Теоретическое исследование последнего вопроса является достаточно сложной задачей.

Если при некотором t вычислен (со случайной ошибкой) вектор $A^t X = (x_1^t, \dots, x_n^t)$, то $A^{t+1} X$ определяют с помощью многократного вычисления оценок, рассмотренных ранее. Например, оценки

$$\frac{x_{i_0}^t a_{i_0 i_1}}{p_{i_0}^0 p_{i_0 i_1}}, \quad (4.2.52)$$

где $\vec{p}^0 = (p_1^0, \dots, p_n^0)$ — вектор, согласованный с $A^t X$, а $\mathcal{P} = \|p_{ij}\|_{i,j=1}^n$ — стохастическая матрица, согласованная с A . Моделируется распределение $\vec{p}^0$ и переход от i_0 к i_1 с помощью матрицы $\mathcal{P}$. При очень большом n и относительно небольшой требуемой точности такой подход весьма эффективен.

Поскольку $\xi_t = (A^t X, Y)$ находятся со случайной ошибкой, то случайную ошибку будут иметь и коэффициенты $c_t^{(t)}$ и определяемые как решение нелинейного алгебраического уравнения, приближенные оценки $\hat{\lambda}_j$, $j = 1, \dots, L$ собственных чисел оператора A . $\hat{\lambda}_j$ будут иметь также систематическую погрешность ввиду нелинейности задачи. Если число N_t независимых испытаний при оценке ξ_t относительно невелико, то изучение поведения ошибок у $\hat{\lambda}_j$ очень трудная задача. При больших N_t работает асимптотическая теория, можно воспользоваться предельной теоремой (Утверждение 9.3.1 на стр. 271 книги [49]), из которой следует, в частности, что систематическая погрешность имеет порядок $O(N^{-1})$, а случайная — $O(N^{-1/2})$, где $N = \min_t N_t$. Следовательно, систематическая ошибка может быть, начиная с некоторого N , пренебрежимо малой. Если смещением можно пренебречь, то вычислительная процедура оценки результирующей погрешности состоит в следующем.

При некотором фиксированном $t, t+1, \dots, t+L-1$ и N_t вычисляется M независимых оценок $\hat{\lambda}_j$ и по этим оценкам строится доверительный интервал для λ_j . Конечность дисперсии оценок (4.2.52) гарантирует их асимптотическую нормальность.

Замечание относительно параллелизма. *Параллелизм при вычислении каждого ξ_t осуществляется очевидным образом. Получение каждой оценки $\hat{\lambda}_j$ или некоторой их частичной совокупности может быть поручено отдельному процессору. При этом следует иметь ввиду, что доверительный интервал, построенный по среднему M реализаций системы, как обычно уменьшится в $\sqrt{M}$ раз, но систематическая погрешность — среднее погрешностей каждой реализации — является средним систематический погрешностей каждой реализации и сохраняет порядок. Таким образом, порядок числа используемых процессоров не может превосходить $O(\sqrt{N})$. Особо следует отметить, что распараллеливание возможно при величине N , обеспечивающей правильный порядок получаемых оценок $\hat{\lambda}_j$.*

Ниже приводятся некоторые простейшие примеры вычислений. Метод показал себя безусловно эффективным при вычислении первых нескольких собственных чисел. Для матриц с нормой большей единицы требовалось масштабирование. В противном случае наблюдалось неустойчивость относительно ошибок округления.

Пусть дан линейный оператор A с дискретным спектром, собственные числа которого различны, но могут быть равны по модулю. Будем моделировать случайные величины $\xi_m = (A^m X, Y)$ с помощью метода Монте-Карло, где X и Y — векторы, компоненты которых получены с помощью датчика случайных чисел. Выбирая длину окна L и составляя систему линейных уравнений

$$\begin{aligned} c_0 \xi_m + c_1 \xi_{m+1} + \dots + c_L \xi_{m+L} &= 0 \\ c_0 \xi_{m+1} + c_1 \xi_{m+2} + \dots + c_L \xi_{m+L+1} &= 0 \\ \dots &\dots \dots \dots \\ c_0 \xi_{m+L-1} + c_1 \xi_{m+L} + \dots + c_L \xi_{m+L+1} &= 0, \end{aligned} \quad (4.2.53)$$

находим значения чисел $\hat{c}_0, \hat{c}_1, \dots, \hat{c}_L$. Так как c_j должны быть известны с точностью до общего множителя, фиксируем $c_L = 1$. Решив уравнение вида

$$\lambda^L + c_{L-1} \lambda^{L-1} + \dots + c_0 = 0, \quad (4.2.54)$$

коэффициентами которого являются найденные c_j , получим собственные числа линейного оператора A .

Сложность метода заключается в выборе исходных параметров: L — длины окна, m — степень ξ_m , с которой начинаем строить систему сдвигов, и N — число испытаний в методе Монте-Карло при моделировании ξ_m .

При выборе длины окна L равным размерности матрицы, метод совпадает с методом Крылова, погрешность вычислений зависит только от погрешности метода Монте-Карло и параметр m можно взять равным 1.

Если же L меньше размерности матрицы, то возникает дополнительная погрешность, погасить которую можно с помощью увеличения значения параметра m , погрешность метода Монте-Карло также присутствует в этом случае.

Рассматривались матрицы различных порядков, для которых подсчитывались математическое ожидание и дисперсия (ниже обозначенного как σ) в группе экспериментов с одинаковыми исходными параметрами (размер выборки фиксировался и был равен 10 для всех испытаний с фиксированными исходными параметрами). Как показала практика, выбором X и Y с помощью датчика случайных чисел мы можем обеспечить ненулевые коэффициенты a_k в разложении векторов $\xi_m = (A^m X, Y) = \sum_{k=1}^T a_k \lambda_k^m$ на составляющие, что позволяет получить достоверные результаты.

Результаты вычислений

Приведём сначала некоторые результаты численных экспериментов, в которых величины $(A^m X, Y)$ были вычислены без случайной ошибки.

Пример 4.3. Рассматривалась матрица восьмого порядка с вещественными собственными числами. Вычисления производились при различных значениях параметров L и m , где L – длина окна, а m – степень матрицы, то есть $\xi_m = (A^m X, Y)$. Приведена таблица результатов (Таблица 4.1) с наилучшими значениями параметров. В левой колонке записаны точные собственные числа исследуемой матрицы, в последующих колонках – результаты применения Алгоритма с параметрами L и m .

числа \ L, m	8,1	7,1	6,2	5,5	4,8	3,14	3,25	2,14
30,1947	30,1947	30,1947	30,1947	30,1947	30,1947	30,1947	30,1947	30,1947
13,2645	13,2645	13,2645	13,2645	13,2645	13,2645	13,2645	13,2641	-
-8,2408	-8,2408	-8,2408	-8,241	-8,2405	-8,2414	-8,2455	-	-
6,0134	6,0134	6,0498	5,9697	5,9532	5,8677	-	-	-
-3,7765	-3,7765	-3,6329	-3,5384	-4,0049	-	-	-	-
3,7439	3,7439	4,3031	3,1941	-	-	-	-	-
-2,5051	-2,5051	-1,4452	-	-	-	-	-	-
1,3058	1,3058	-	-	-	-	-	-	-

Таблица 4.1: Собственные числа матрицы 8-го порядка, найденные при различных значениях параметров L и m .

Пример 4.4. Рассматривалась стохастическая матрица 10-го порядка, второе и третье собственные числа которой являются комплексно-сопряженными. Результаты показывают, что точность вычисления собственных чисел зависит от обоих параметров, причем, чем меньше длина окна L , тем больше должно быть значение m . Следует отметить, ситуация когда число L попадает между двумя комплексно-сопряженными числами, модули которых равны, нарушает условие (1) Леммы, и последнее собственное число находится неверно, что и показывают приведенные ниже результаты (Таблица 4.2).

Перейдем к случаям, когда ξ_t вычисляются со случайной ошибкой. Здесь и далее рассматривались примеры, где все ξ_t вычислялись с помощью метода Монте-Карло, и оценивалось второе собственное число. Для стохастических матриц задача актуальна, когда второе собственное число является комплексным.

числа \ L, m	9,2	8,4	7,6
1.0000	1.0000	1.0000	1.0000
-0.1466 + 0.0862i	-0.1465 + 0.0862i	-0.1465 + 0.0862i	-0.1465 + 0.0862i
-0.1466 - 0.0862i	-0.1465 - 0.0862i	-0.1465 - 0.0862i	-0.1465 - 0.0862i
-0.0420 + 0.1387i	-0.0418 + 0.1387i	-0.0418 + 0.1386i	-0.0418 + 0.1387i
-0.0420 - 0.1387i	-0.0418 - 0.1387i	-0.0418 - 0.1386i	-0.0418 - 0.1387i
0.1397	0.1398	0.1398	0.1397
-0.0561 + 0.0595i	-0.0495 + 0.0343i	-0.0633 + 0.0399i	-0.0426
-0.0561 - 0.0595i	-0.0495 - 0.0343i	-0.0633 - 0.0399i	-
0.0486 + 0.0236i	0.0208	-	-
0.0486 - 0.0236i	-	-	-
6,8	5,10	4,13	3,16
1.0000	1.0000	1.0000	1.0000
-0.1465 + 0.0862i	-0.1418 + 0.0882i	-0.1459 + 0.0730i	-0.1556 + 0.0898i
-0.1465 - 0.0862i	-0.1418 - 0.0882i	-0.1459 - 0.0730i	-0.1556 - 0.0898i
-0.0418 + 0.1387i	0.0059 + 0.1533i	-0.0047	-
-0.0418 - 0.1387i	0.0059 - 0.1533i	-	-
0.1398	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-

Таблица 4.2: Собственные числа матрицы 10-го порядка, найденные при различных значениях параметров L и m .

Пример 4.5. Ставилась задача определить оптимальное соотношение между количеством итераций N , необходимых для нахождения ξ_m , и конечным результатом, то есть математическим ожиданием и среднеквадратичным отклонением второго собственного числа.

Рассматривалась стохастическая матрица пятого порядка, для которой фиксировались исходные вектора X , Y , длина окна L и m , исходя из результатов, основанных на точных значениях ξ_m , полученных в работе [41]. Параметры с наилучшими результатами фиксировались и вычисления производились по алгоритму, описанному выше, где ξ_m вычислялись рекурсивно с помощью метода Монте-Карло с числом итераций равным N .

$$A = \begin{pmatrix} 0.0625 & 0.125 & 0.25 & 0.25 & 0.3125 \\ 0.3 & 0.05 & 0.15 & 0.1 & 0.4 \\ 0.09 & 0.18 & 0.27 & 0.136 & 0.318 \\ 0.21 & 0.36 & 0.105 & 0.26 & 0.053 \\ 0.12 & 0.04 & 0.24 & 0.28 & 0.32 \end{pmatrix} \quad (4.2.55)$$

Матрица A – стохастическая, собственные числа которой равны

$$1.00000; \quad -0.04273 \pm 0.21713i; \quad 0.10905; \quad -0.05521;$$

При фиксированных X и Y , $L = 3$, $m = 2$, ($M = [m + 2 * (L + 1) - 3] = 7$ – количество подряд полученных ξ_m , необходимых для реализации алгоритма) и применении ортогонализации получены следующие результаты (см. Таблицу 4.3).

Величины ξ_m находятся рекурсивно, поэтому именно первые значения ξ_m должны находиться с наибольшей точностью. Практика показывает, в этом случае, при мень-

$N \times M$	$E\lambda_2$	σ_1	σ_2
$50\,000 \times 7$	-0.04194 + 0.20610i	0.00722	0.00279
$25\,000 \times 7 = 175\,000$	-0.04688 + 0.20564i	0.00545	0.00744i
$10\,000 \times 7$	-0.05456 + 0.20218i	0.03323	0.01393i
$50\,000\ 50\,000\ 10\,000 \dots = 150\,000$	-0.04454 + 0.20522i	0.00403	0.00704i
$5\,000 \times 7$	-0.02406 + 0.21819i	0.06061	0.02480i

Таблица 4.3: Зависимость погрешности второго собственного числа от числа испытаний в методе Монте-Карло.

шей вычислительной сложности, погрешность метода меньше, чем при большей вычислительной сложности, но равном числе итераций на каждом шаге вычисления ξ_m (сравните 4-ую строку таблицы со 2-ой, первый столбец таблицы показывает общее число итераций в методе Монте-Карло, необходимых для получения последовательности величин ξ_m).

Пример 4.6. Ставилась задача нахождения погрешности значения второго собственного числа от длины окна L и m – степень ξ_m , с которой начинаем строить систему сдвигов. Количество итераций в методе Монте-Карло для вычисления одного ξ_m фиксировалось ($N = 250000$). Результаты отображены в таблице 4.4. Через σ_1 и σ_2 обозначены средние квадратичные отклонения (на одно испытание) вещественной и мнимой части λ_2 соответственно.

	$E\lambda_2$	σ_1	σ_2
L=3, m=2	-0.04216 + 0.21949i	0.0039	0.0056i
L=3, m=3	-0.03665 + 0.22852i	0.00418	0.0164i
L=3, m=1	-0.04136 + 0.23641i	0.01065	0.0223i
L=4, m=1	-0.04556 + 0.21075i	0.00625	0.00598i

Таблица 4.4: Зависимость погрешности второго собственного числа от длины окна L и степени m .

Наилучшими параметрами для матрицы пятого порядка оказались $L = 3$, $m = 2$. Как видно из таблицы, с уменьшением длины окна L необходимо увеличивать число m , но это увеличение не должно быть слишком большим, иначе это приведет к обратным результатам. Параметры $L = 3$, $m = 3$ немного ухудшают вычисление второго собственного числа. При увеличении значения параметра m , возникает дополнительная погрешность метода, которая растет не линейно и с трудом может быть найдена аналитически. Результаты данного эксперимента показывают, что выбор длины L и m должен быть «сбалансирован».

Пример 4.7. Ставилась задача нахождения второго собственного числа для матриц, не являющихся стохастическими. Практика показала, что в этом случае погрешность метода Монте-Карло с ростом m растет экспоненциально. В таких ситуациях необходимо нормировать матрицу и свести её модуль к стохастической. Известно, что неотрицательная матрица A , имеющая положительное максимальное характеристическое число $r > 0$ и соответствующее этому числу положительный собственный вектор $z = (z_1, z_2, \dots, z_n) > 0$, простыми линейными преобразованиями может быть сведена к стохастической матрице $P = \frac{1}{r} z^{-1} A z$. Собственные числа исходной матрицы могут быть легко восстановлены по собственным числам полученной стохастической матрицы.

Результаты метода, применяемого к матрице без нормировки и с нормировкой, приведены в Таблице 4.5.

Рассматривалась матрица

$$A = \begin{pmatrix} 1 & 2 & 4 & 4 & 5 \\ 6 & 1 & 3 & 2 & 8 \\ 2 & 4 & 6 & 3 & 7 \\ 4 & 7 & 2 & 5 & 1 \\ 3 & 1 & 6 & 7 & 8 \end{pmatrix} \quad (4.2.56)$$

с собственными числами

$$20.98978, -0.66060 \pm 4.51289i, -0.97786, 2.3093,$$

фиксировались X, Y , производилась ортогонализация. Преимущества ортогонализации очевидны.

N = 50 000	$E\lambda_2$	σ_1	σ_2
без нормировки	$-0.48443 + 4.54973i$	0.26874	0.29891i
с нормировкой	$-0.65834 + 4.825788i$	0.03436	0.11946i

Таблица 4.5: Подсчет второго собственного числа для нестохастической матрицы.

Пример 4.8. Так как одной из важных областей приложений являлось массовое обслуживание, то интерес данной работы состоял в определении погрешности метода для матриц больших размерностей. Рассматривались простейшие системы массового обслуживания с одним прибором, время обслуживания которых распределено по показательному закону с параметром $\mu(t)$ и интенсивностью поступающего потока $\lambda(t)$. Для генерирования матрицы перехода такого рода использовался датчик случайных чисел RAND, который генерирует числа от 0 до 1. $\lambda = 0,21 + 0,2 * RAND$ и $\mu = 0,45 + 0,2 * RAND$. Для системы такого типа с десятью возможными состояниями при фиксированном значении $N = 50000$ (число испытаний в методе МК) хорошие результаты для λ_2 были получены уже при длине окна $L = 3$.

Собственные числа сгенерированной матрицы перехода равны:

$$1.00000; \mathbf{0.89539}; 0.78407; -0.67308; -0.59884; 0.57361; -0.42364; 0.36718; -0.42364; 0.05796.$$

Результаты, полученные в результате применения рассматриваемого метода, отображены в Таблице 4.6.

	$E\lambda_2$	σ
L=3, m=16	0.89059	0.01848
L=3, m= 6	0.92924	0.01310
L=2, m=25	0.88163	0.04796

Таблица 4.6: Подсчет второго собственного числа для СМО с 10 состояниями.

Для системы массового обслуживания подобного типа, но с количеством состояний равным 20 также были получены хорошие результаты (Таблица 4.7). Фиксировалось значение $N = 50000$, вектора X, Y . Точное значение второго собственного числа $\lambda_2 = \mathbf{0,96112}$. Вопрос устойчивости данного метода при увеличении погрешности метода Монте-Карло также был изучен для матрицы большого порядка $n = 20$ (см. Таблицу

	$E\lambda_2$	σ
L=3, m=16	0.96201	0.02255
L=3, m= 6	0.96102	0.01443
L=2, m=25	0.96894	0.01460

Таблица 4.7: Подсчет второго собственного числа для СМО с 20 состояниями.

4.8). Кроме того был рассмотрен алгоритм, использующий в своей основе метод наименьших квадратов [41].

Алгоритм состоит в следующем:

1⁰. Вычисляются $\xi_t = ((A^t(A - \lambda I)X, Y)$ с помощью метода Монте-Карло.

2⁰. Выбирается длина окна l и образуются линейные комбинации с элементами векторов

l - вложений $Z_i = (\xi_i, \dots, \xi_{i+l-1})$, с коэффициентами $c_0, c_1, \dots, c_{l-1}$, которые находятся из условия

$$\sum_{t=k}^{k+m} (c_0 \xi_t + c_1 \xi_{t+1} + c_2 \xi_{t+2} + \dots + c_{l-1} \xi_{t+l-1})^2 = \min.$$

Для этого решается система из l уравнений:

$$\begin{aligned} c_0 \sum_{t=k}^{k+m} \xi_t^2 + c_1 \sum_{t=k}^{k+m} \xi_t \xi_{t+1} + \dots + c_L \sum_{t=k}^{k+m} \xi_t \xi_{t+L} &= 0 \\ c_0 \sum_{t=k}^{k+m} \xi_t \xi_{t+1} + c_1 \sum_{t=k}^{k+m} \xi_{t+1}^2 + \dots + c_L \sum_{t=k}^{k+m} \xi_{t+1} \xi_{t+L+1} &= 0 \\ \dots \dots \dots \dots & \\ c_0 \sum_{t=k}^{k+m} \xi_t \xi_{t+L-1} + c_1 \sum_{t=k}^{k+m} \xi_{t+1} \xi_{t+L-1} + \dots + c_L \sum_{t=k}^{k+m} \xi_{t+L-1}^2 &= 0. \end{aligned} \quad (4.2.57)$$

Обозначим найденные коэффициенты $\hat{c}_0, \hat{c}_1, \dots, \hat{c}_{l-1}$.

3⁰. Составляется многочлен с найденными коэффициентами

$$Q(\lambda) = c_0 + c_1 \lambda + c_2 \lambda^2 + \dots + c_{l-1} \lambda^{l-1}.$$

Корни многочлена являются 2-м, 3-м ... (l-1)-м собственными числами оператора A .

Алгоритм 1	$E\lambda_2$	σ	Алгоритм 2	$E\lambda_2$	σ
L=3, m=16, N=50 000	0.94333	0.02189	L=4, m=25, k=15 N=50 000	0.95698	0.02380
L=3, m=16, N=25 000	0.96683	0.02573	L=4, m=25, k=15 N=25 000	0.96614	0.02524
L=3, m=16, N=10 000	0.92884	0.04781	L=4, m=25, k=15 N=10 000	0.96798	0.03798

Таблица 4.8: Таблица сравнения двух алгоритмов.

Данный алгоритм оказывается более устойчив к случайным ошибкам и должен использоваться в случаях недостаточной точности вычисления ξ_t , $n = 20$, $\lambda_2 = 0,95795$.

Как видно из приведённых примеров, предлагаемое обобщение метода Крылова может быть эффективным для определённого круга задач. Как обычно, случай близких собственных чисел может представлять значительные вычислительные трудности.

Как и ранее, в нелинейных задачах мы различаем “внутренний” параллелизм — при вычислении скалярных произведений и “внешний”, позволяющий при различных значениях параметров (множеств псевдо- или квази-случайных чисел) решать задачу полностью на различных вычислительных устройствах. Это позволяет, кроме всего прочего, строить доверительный эллипсоид для решения задачи.

Литература

- [1] Ермаков С. М. Параметрически разделимые алгоритмы // Вестник СПбГУ, Сер.1, вып. 4, 2010. С. 25–31.
- [2] Bailey, David H., Borwein, Peter B., and Plouffe, Simon. On the Rapid Computation of Various Polylogarithmic Constants // Mathematics of Computation, Vol. 66, №. 218, 1997. P. 903–913.
- [3] Кейперс Л., Нидеррейтер Г. Равномерное распределение последовательностей. М.: Наука, 1985. 408 с.
- [4] Hsu L. C. A general approximation method of evaluating multiple integrals // Tohohi math. J. 1957. Vol. 9. №1. P. 45–55.
- [5] Ермаков С. М. О мультипликативном датчике случайных чисел // Вестник Санкт-Петербургского университета, Сер.1. 2003. Вып. 3, № 17. С. 23–30.
- [6] Ермаков С.М. Метод Монте-Карло в вычислительной математике. Вводный курс. СПб: «Невский Диалект»; Москва: «Бином», 2009. 192 с.
- [7] Ермаков С. М. Метод Монте-Карло и смежные вопросы. 2-е изд. М.: Наука, 1975. 472 с.
- [8] Fishman G. S. Monte Carlo Concept, Algorithms and applications. Springer, 1996. 698 p.
- [9] Gerlovina V., Nekrutkin V. Asymptotical behavior of linear congruential generators // Monte Carlo Methods and Appl. 2005. Vol. 11, № 2. P. 135–162.
- [10] Matsumoto M. and Nishimura T. Mersenne Twister A 623-dimensionally equidistributed uniform pseudo-random number generator // ACM Transactions on Modeling and Computer Simulation. 1998. № 8. P.3–30.
- [11] Panneton F., L'Ecuyer P. Improved Long-Period Generators Based on Linear Recurrences Modulo 2 // ACM Transactions on Mathematical Software. 2006. , Vol. 32, № 1. P. 1–16.
- [12] Walker A. J. New Fast method for generating discrete random numbers with arbitrary friguency distributions // Elektronik Letters. 1974. Vol. 10. P. 127–128.
- [13] Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., & Teller E. Equations of state calculations by fast computing machines // Journal of Chemical Physics, Vol.21, 1953. P. 1087–1091.

- [14] Hastings, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*. 1970. № 57. P. 97–109.
- [15] Kendall W.S., Liang F., Wang J.S. Markov Chain Monte Carlo: Innovations and Applications. World Scientific Publishing Company. 2005. 239 p.
- [16] Tsybakov A. Adaptive estimation of the mode of a multivariate density // *J. of nonparametric statistics*. Vol. 17. 2009. № 1. P. 83–105.
- [17] Цыбаков А. Б. Рекуррентное оценивание моды многомерного распределения // *Проблемы передачи информации*. 1990. Т. 26, Выпуск 1. С. 38–45.
- [18] Ермаков С. М., Мисов Т. Моделирование δ^2 -распределения Вестник Санкт-Петербургского университета. Серия 1. 2005. Выпуск 4. С. 53–60
- [19] Ermakov S.M., Schwabe R. On Randomizing Estimators in Linear Regression Models // *Freie Universitaet Berlin Verlag, Serie A – Mathematik*. 1999. № 4. P.7–10.
- [20] Ермаков С. М., Леора С. Н. Метод Метрополиса в задачах поиска экстремума // *Математические модели. Теория и приложения: Сб. Вып. 11*. СПб.: Изд-во ВВМ, 2010. С. 72–82.
- [21] Price K.V., Storn R.M., Lampinen J.A. Differential Evolution: A Practical Approach to Global Optimization. Springer, 2005. 538 p.
- [22] Wynn H. P. (1970). The sequential generation of D-optimum experimental designs // *The Annals of Mathematical Statistics*. 1970. Vol.41, № 5. P. 1655–1664.
- [23] Ченцов Н.Н. Оценка неизвестной плотности распределения по наблюдениям. *ДАН СССР*. 1962. № 147, I. С. 45–48.
- [24] Ермаков С. М., Расулов А. С., Бакоев, Веселовская А. З. Избранные алгоритмы метода Монте-Карло. Ташкент, изд. Ташкентского госуниверситета. 1992. 132 с.
- [25] Михайлов Г. А., Войтишек А. В. Численное статистическое моделирование. Методы Монте-Карло. М.: Академия, 2006. 367 с.
- [26] Хисамутдинов А. И. Единичный класс оценок для вычисления по методу Монте-Карло функционалов от решения интегрального уравнения 2-го рода // *ЖВМ и МФ*. 1970. Вып. 10. № 5. С. 1269–1280.
- [27] G. M. Baudet, Asynchronous iterative methods for multiprocessors // *Journal of the ACM*, Vol. 25, №. 2, April 1978. P. 226–244.
- [28] Вагнер В., Ермаков С. М. Стохастическая устойчивость и параллелизм метода Монте-Карло // *ДАН*. 2001. Т. 379. № 4. С. 439–441.
- [29] Ermakov S. M., Wagner W. Monte Carlo difference schemes for the wave equations // *Monte Carlo Methods and Appl*. 2002. Vol. 8. № 1. P. 1–29.
- [30] Evans D. H. Applied Multiplex Sampling // *Technometrics*. 1963. Vol. 5, № 3. P. 341–359.

- [31] Ермаков С.М., Адамов А.В. О стохастической устойчивости метода Монте-Карло (случай операторов) // Вестник Санкт-Петербургского университета. Изд-во СПб-ГУ. 2004. Сер. 1, вып. 2. С. 3–8.
- [32] Михайлов Г. А., Медведев И. Н. Оптимизация весовых алгоритмов статистического моделирования: научное издание. Новосибирск : Омега Принт. 2011. 302 с.
- [33] Radon J. Cur mehanischen kubatur // Monatshefte für mathematik. 1948. Vol. 52, № 4. P. 286–300.
- [34] Мысовских И.П. Интерполяционные кубатурные формулы. М.: Наука, 1981. 336 с.
- [35] Мысовских И.П. Лекции по методам вычислений. М.: Физматгиз, 1962. 342 с.
- [36] Соболев С. Л. Введение в теорию кубатурных формул. М.: Наука, 1974. 808 с.
- [37] Коробов Н. М. Теоретико-числовые методы в приближенном анализе. М.: Физматгиз, 1963. 224 с.
- [38] Niederreiter H. Random Number Generation and Quasi-Monte Carlo Methods. CBMS-NSF Regional Conference Series in Applied Mathematics, 63. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. 244 p.
- [39] Соболев И. М. Многомерные квадратурные формулы и функции Хаара. М.: Наука, 1969. 288 с.
- [40] Liu J. S. Monte Carlo strategies in scientific Computing. Springer, 2003. 359 p.
- [41] Ермаков С.М., Видяева К.О. Об оценке спектра линейного оператора // Вестник Санкт-Петербургского университета, Сер. 1. 2012. Вып. 1. С. 11–17.
- [42] Некруткин В. В. Прямая и сопряженная схема Неймана–Улама для решения нелинейных интегральных уравнений // ЖВМ и МФ. 1974. Вып. 14. № 6. С. 1409–1415.
- [43] Ermakov S. M., Kaloshin I. Solving the Nonlinear Algebraic Equations with Monte Carlo Method //Advances in Stochastic Simulation Methods. 2000, Birkhauser: Boston-Basel-Berlin. P. 3–15.
- [44] Ермаков С. М., Калошин И.В. Метод Монте-Карло для решения систем уравнений с квадратичной нелинейностью // Нелинейные динамические системы под ред. доктора ф.-м. наук, профессора Г.А. Леонова. Изд-во СПбГУ. 2003. Выпуск 5.
- [45] Расулов А. С. Метод Монте-Карло для решения нелинейных задач. Ташкент: ФАН, 1992. 104 с.
- [46] Ермаков С. М., Москалева Н. М. Ветвящиеся процессы и уравнение Больцмана. Вычислительные аспекты // Вестник ЛГУ. Сер. 1. 1987. Вып. 3(15). С. 38–43.
- [47] Михайлов Г. А. Весовые методы Монте-Карло. Новосибирск, 2000, 248 с.
- [48] Ермаков С. М., Сережина В.Г. Об алгоритмах обобщенного интегрирования при решении нелинейных уравнений. Деп. в ВИНТИ 02.09.83 N 5021-83.
- [49] Уилкс С. Математическая статистика. М.: Наука, 1967. 632 с.

- [50] Halton J. H. Sequential Monte Carlo techniques for solving non-linear systems // Monte Carlo Methods And Applications. 2006. Vol. 12. № 2. P. 113–141.
- [51] Тимофеев К. А. Об одном классе методов Монте-Карло для решения уравнений с квадратичной нелинейностью уравнений // Вестник СПбГУ. Сер. 10. Вып. 3. 2008. С. 23–28.
- [52] Ермаков С.М., Карюкин В.В., Каштанов Ю.Н. Алгоритмы метода Монте-Карло для решения линейных уравнений в случае $\rho(|k|) \geq 1$ // Сб. VII Всесоюзное совещание “Методы Монте-Карло в вычислительной математике и математической физике”. Новосибирск. 1985. С. 121–124.
- [53] Black F. and Scholes M. The Pricing of Options and Corporate Liabilities // Journal of Political Economy. 1973. Vol. 81. С. 637–654.
- [54] Duffy D. Finite difference methods in financial engineering: a partial differential equation approach // John Wiley & Sons Ltd. 2006. 423 p.
- [55] R. Zvan and P.A. Forsyth and K.R. Vetzal Penalty methods for american options with stochastic volatility // Journal of Computational and Applied Mathematics, № 218. 1998. P. 91–199.
- [56] Nielson B.F, Skavhaug O. and Tvelto A. Penalty and front-fixing methods for the numerical solution of American option problems // J. Comp. Finan, № 4, 2002.
- [57] Видяева К.О., Ермаков С.М. К обобщению метода Крылова вычисления коэффициентов минимального многочлена // ЖВМ и МФ. 2013 . Том 53, № 5 . С. 1–9.
- [58] Крылов А.Н. О численном решении уравнения, которым в технических вопросах определяются частоты малых колебаний материальных систем // Известия АН СССР. Отделение математических и естественных наук. 1931. Серия VII, № 4. С. 491–539.
- [59] Люу Ю.Д. Методы и алгоритмы финансовой математики. М.: Бином; Лаборатория знаний, 2007. 751 с.
- [60] Антонов А.А., Ермаков С.М. Эмпирическая оценка погрешности квази Монте-Карло вычислений // Вестник СПбГУ, Сер.1. 2013. (В печати).