

## МАТЕМАТИКА

УДК 519.644.2

**АЛГОРИТМ ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ  
МЕТОДОМ КВАЗИ МОНТЕ-КАРЛО  
С АПОСТЕРИОРНОЙ ОЦЕНКОЙ ПОГРЕШНОСТИ\****А. А. Антонов*

Санкт-Петербургский государственный университет,  
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7-9

В работе рассматривается возможность применения теоретико-вероятностных методов к детерминированной процедуре оценки погрешности метода квази Монте-Карло вычисления многократных интегралов. Существующие методы оценки упомянутой погрешности являются неконструктивными. Ранее полученный результат, выведенный при помощи теории случайных кубатурных формул, дополнен новыми теоретическими результатами. Алгоритм  $Q_{int}$ , представленный в пошаговом виде, удобном для непосредственного применения, обсуждается в контексте параметризации. Продемонстрированы важные свойства  $Q_{int}$ : монотонность оценки по количеству разбиений и зависимость точности от правила этих разбиений.

Предлагаются новые численные результаты, полученные с использованием последовательностей Соболя и подынтегральных функций из программной библиотеки TESTPACK. Демонстрируется преимущество по сравнению с оценкой, полученной традиционным методом Монте-Карло. Библиогр. 14 назв. Ил. 4.

*Ключевые слова:* метод Монте-Карло, квази Монте-Карло, доверительный интервал, случайные кубатурные формулы, функции Хаара, последовательности Соболя.

**1. Введение.** Как известно, методы Монте-Карло основаны на результатах теории вероятностей и математической статистики. Методы же квази Монте-Карло, имеющие формальное сходство с методами Монте-Карло, основаны на детерминистских (теоретико-числовых, в основном) результатах. Вопрос о сходстве и различии этих методов представляет несомненный интерес как с теоретической, так и с прикладной точек зрения. Имеется ряд работ, где теоретико-вероятностные методы применяются для анализа детерминированных последовательностей чисел. Достаточно заметить, что так называемые датчики псевдослучайных чисел являются детерминированными алгоритмами и законность их использования подтверждается обычно системой статистических тестов (см. Кнут, [8]).

Рассмотрение квазислучайных последовательностей с вероятностной точки зрения показывает, что члены этих последовательностей статистически зависимы, а при

---

\*Работа выполнена под руководством д-ра физ.-мат. наук проф. С. М. Ермакова и при финансовой поддержке РФФИ (грант № 14-01-00271).

вычислении с их помощью интеграла для оценки погрешности не может быть использована центральная предельная теорема так, как это делается в методе Монте-Карло. Оценка погрешности метода квази Монте-Карло теоретически решается с помощью неравенства Коксмы—Хлавки [9]. Практически же его использование связано с рядом трудностей, что отмечалось многими авторами (см., например, работу Нидеррейтера [11], где указаны более эффективные подходы к анализу погрешности интегрирования). Таким образом, если при вычислении интегралов поведение средних  $\frac{1}{N} \sum_{j=1}^N f(X_j)$ , где  $X_j$  — квазислучайные векторы в единичном  $s$ -мерном гиперкубе  $U_s$ ,

во многом сходно с поведением средних  $\frac{1}{N} \sum_{j=1}^N f(\Xi_j)$ , где  $\Xi_j$  — равномерно распределённые случайные векторы в  $U_s$ , то средние значения сумм квадратов  $f^2(X_j)$  и  $f^2(\Xi_j)$  пригодны для оценки погрешности во втором случае и не пригодны в первом.

Задачи настоящей работы состоят в том, чтобы выяснить, что может служить формальным аналогом дисперсии при использовании метода квази Монте-Карло, и указать эффективные средства оценки погрешности интегрирования с использованием такого аналога.

Первая из задач использует связи квазислучайных последовательностей со случайными квадратурными формулами и расслоенной выборкой. Пусть гиперкуб  $U_s$  разбит на  $L$  непересекающихся подобластей  $U_1, \dots, U_L$ . Расслоенная выборка предполагает, что в каждой области выбирается  $k_j$  случайных точек, так что  $\sum_{j=1}^L k_j = N$ .

Известно выражение для дисперсии суммы  $\frac{1}{N} \sum_{j=1}^N f(\Xi_j)$  в этом случае (см. [5] в контексте генерации выборок из распределений, [4] — в контексте численного интегрирования). Если объемы всех  $U_j$  равны, а  $k_1 = k_2 = \dots = k_L$ , то такая выборка обладает улучшенными свойствами равномерности. Подход, основанный на теории случайных кубатурных формул, связывает расслоенную выборку с улучшенными свойствами равномерности с формулами, точными для кусочно-постоянных функций с постоянными значениями в областях  $U_1, \dots, U_L$  и с ортогональными разложениями по кусочно-постоянным функциям. В работе получен новый результат, обобщающий частный случай  $L = 2$ , исследованный ранее в [4].

Квазислучайные последовательности, предложенные И. М. Соболев, очевидным образом связаны с расслоением и ортогональными разложениями по кусочно-постоянным функциям (функции Хаара, см. [3]). Показано, что с квазислучайными последовательностями может быть связана дисперсия случайной квадратурной формулы. Приведённые в работе численные примеры показывают, что основанные на этой дисперсии доверительные интервалы могут оценивать остаток интегрирования.

Однако приведённые выше результаты не являются строго обоснованным методом оценки остатка. Строгое обоснование может быть получено на пути рандомизации квазислучайных последовательностей, который обсуждается далее.

Рассмотрим задачу численного интегрирования вещественнозначной функции  $f$  по единичному гиперкубу размерности  $s$ :  $\int_{U_s} f(\mathbf{u}) d\mathbf{u}$ . При этом мы предполагаем, что размерность  $s$  достаточно высока, а функция  $f$  устроена достаточно сложно<sup>1</sup>. Как

<sup>1</sup> В нашей работе мы сознательно ограничиваемся контекстом Монте-Карло и квази Монте-Карло. Сравнение с традиционными методами вычислительной математики остаётся за рамками

уже отмечено ранее, в подобных случаях на практике часто используется рандомизированный метод квази Монте-Карло (RQMC), который позволяет строить доверительные интервалы на основе оценки дисперсии. Такой подход ведет к вопросу выбора достаточного количества рандомизаций (см. [10]).

Так, пусть мы используем  $k$  рандомизаций  $\{\tilde{P}_{n,j}\}_{j=1}^k$  одного и того же множества квази Монте-Карло  $P_n$  мощности  $n$ . При этом дисперсия может иметь порядок вплоть до  $O(\frac{\log^{s-1} n}{kn^3})$  (при существенных ограничениях на функцию  $f$ , см. [13]), в то время как традиционный Монте-Карло имеет дисперсию порядка  $O(\frac{1}{kn})$ . Поэтому для получения более ощутимого выигрыша необходимо увеличивать  $n$ , что при фиксированных вычислительных затратах ведет к уменьшению  $k$ . Это, в свою очередь, приводит к недостоверной оценке дисперсии, поэтому для практических задач рекомендуется выбирать  $k \geq 10$ .

Пусть мы имеем фиксированное  $k$ , тогда конечное RQMC-множество узлов интегрирования представляет собой объединение всех рандомизаций:  $\mathfrak{P} = \cup_{j=1}^k \tilde{P}_{n,j}$ . Если исходный набор  $P_n$  обладает некоторыми хорошими квази Монте-Карло свойствами, а выбранный способ рандомизации их не нарушает, то результирующий набор такими свойствами, вообще говоря, уже не обладает. Например, если  $P_n$  есть  $(t, m, s)$ -сеть для некоторых параметров  $t$  и  $m$ , а в качестве рандомизации выбран digital shift (см., напр., [6]), сохраняющий структуру сети для каждого  $\tilde{P}_{n,j}$ , то в общем случае нельзя утверждать, что  $\mathfrak{P}$  является  $(t^*, m^*, s)$ -сетью для каких-то  $t^*$  и  $m^*$ . Вопрос о том, насколько сильно нарушается структура конечного RQMC-множества, исследован мало; очевидно лишь, что чем больше количество рандомизаций  $k$ , тем потенциально сильнее это нарушение.

Предлагаемый в статье подход позволяет, в частности, построить альтернативную оценку дисперсии рандомизированного метода квази Монте-Карло. Используя аппарат случайных квадратурных формул из [4], мы получаем формулу Монте-Карло, точную для ортонормированной системы Хаара в  $U_s$ . Полученная формула, по сути, эквивалентна одному частному случаю расслоения. Полученный результат, однако, не является тривиальным, поскольку он

- 1) показывает связь аппарата случайных квадратур с традиционными методами понижения дисперсии;
- 2) имеет естественную интерпретацию в терминах интегралов по подмножествам  $U_s$  определённого вида;
- 3) имеет дисперсию, не превосходящую дисперсии традиционного метода Монте-Карло с тем же количеством точек;
- 4) устанавливает свойства точности для хааровских систем, что дает возможность рассмотреть предложенный подход в рамках квазислучайных множеств, в частности,  $(t, m, s)$ -сетей.

В заключительной части статьи мы приводим численные примеры, свидетельствующие о перспективности предложенного гибридного метода.

**2. Теоретические основы Qint.** Рассмотрим произвольное вероятностное пространство  $(\mathfrak{X}, \mathfrak{B}, \mu)$ , где  $\mathfrak{X}$  — непустое множество,  $\mathfrak{B}$  —  $\sigma$ -алгебра для подмножеств  $\mathfrak{X}$ , снабжённая вероятностной мерой  $\mu$ . Выберем некоторое число  $P \in \mathbb{N}_0$  и возьмём  $N = 2^P$ . Разбив  $\mathfrak{X}$  на  $N$  дизъюнктивных частей равной меры  $\mathfrak{X}_1, \mathfrak{X}_2, \dots, \mathfrak{X}_N$ , покрывающих собой всё  $\mathfrak{X}$ , построим ортонормированную в  $L^2(d\mu)$  систему из  $N$  функций

---

данной статьи; среди подобных работ можно выделить [14], откуда, в частности, и взяты функции для вычислительных примеров, приведённых в заключительной части работы.

Хаара, порождённую  $\mathfrak{X}_1, \mathfrak{X}_2, \dots, \mathfrak{X}_N$ . Здесь и далее при проведении подобных построений неявно предполагается, что такое разбиение возможно для данного  $N$ . Свойства такого рода систем подробно описаны в [3], а явный вид функций и процесс построения указан в [1].

Чтобы построить оценку интеграла  $I = \int_{\mathfrak{X}} f(x)\mu(dx)$  для функции  $f \in L^2(d\mu)$ , воспользуемся аппаратом случайных квадратурных формул, изложенным в [4], в особенности теоремой Ермакова—Грановского. Эта теорема позволяет построить  $N$ -точечную формулу, обладающую двумя важными свойствами: свойством несмещённости для искомого интеграла  $I$  и свойством точности для рассмотренной системы Хаара. Такая формула (см. [1], теорема 1) имеет вид

$$S_N[f] = \frac{1}{N} \sum_{i=1}^N f(x_i), \quad (1)$$

при этом узлы формулы являются случайными величинами, совместная плотность распределения которых задаётся как

$$\varphi(x_1, x_2, \dots, x_N) = \begin{cases} \frac{N^N}{N!}, & \text{если } (x_1, x_2, \dots, x_N) \in \mathfrak{Lat}(i_1, i_2, \dots, i_N), \\ 0, & \text{иначе.} \end{cases}$$

Здесь использовано определение *латинского множества*, задаваемого перестановкой  $(i_1, i_2, \dots, i_N)$ :

$$(x_1, x_2, \dots, x_N) \in \mathfrak{Lat}(i_1, i_2, \dots, i_N) \Leftrightarrow \forall j \in \{1, 2, \dots, N\} x_j \in \mathfrak{X}_{i_j}.$$

Далее, в той же работе ([1], теорема 2) показано, что дисперсия формулы равна

$$\mathbb{D}S_N[f] = \frac{1}{N} \int_{\mathfrak{X}} f^2(x)\mu(dx) - \frac{1}{N} \left( \int_{\mathfrak{X}} f(x)\mu(dx) \right)^2 - \frac{1}{N} \sum_{i < j} (\alpha_i - \alpha_j)^2, \quad (2)$$

где  $\alpha_i = \int_{\mathfrak{X}_i} f(x)\mu(dx)$ ,  $i = 1, 2, \dots, N$ . Эта дисперсия ни при каких условиях не меньше нуля и не больше дисперсии обычного Монте-Карло:  $0 \leq \mathbb{D}S_N[f] \leq \mathbb{D}_{MC}$ .

Хотя полученная формула обладает рядом важных характеристик, её использование на практике затруднено из-за сложности моделирования распределения с плотностью  $\varphi(x_1, x_2, \dots, x_N)$ , особенно в случае произвольного пространства и произвольной схемы его разбиения. Поэтому мы будем использовать полученную формулу в варианте квази Монте-Карло с применением последовательностей Соболя.

В самом деле, пусть здесь и далее  $\mathfrak{X}$  есть  $s$ -мерный куб  $[0, 1]^s$ , снабжённый мерой Лебега. Известно, что  $s$ -мерная последовательность Соболя для некоторого известного  $t(s)$  является  $(t(s), s)$ -последовательностью<sup>2</sup> по основанию 2. Таким образом,

<sup>2</sup> Напомним, последовательность  $\{x_n\}$  называется  $(t, s)$ -последовательностью по основанию  $b$ , если для произвольных натуральных  $k \geq 0$ ,  $m \geq t$  отрезок последовательности  $\{x_{k \cdot b^m}, \dots, x_{(k+1) \cdot b^m - 1}\}$  содержит в точности  $b^t$  точек внутри произвольного элементарного подмножества вида  $\prod_{j=1}^s \left[ \frac{a_j}{b^{d_j}}, \frac{a_j+1}{b^{d_j}} \right]$  ( $a_j, d_j \in \mathbb{N}_0$ ,  $a_j < b^{d_j}$  для всех  $j$ ), имеющего объём  $b^{t-m}$ . Это замечательное свойство впервые установлено Соболевым [2] и впоследствии обобщено Нидеррейтером [12] на случай произвольного основания.

если строить  $\mathfrak{X}_1, \mathfrak{X}_2, \dots, \mathfrak{X}_N$  на основе элементарных подмножеств, мы гарантируем, что для достаточно большого  $P$  любой отрезок последовательности Соболя длины  $N = 2^P$ , имеющий вид  $\{x_{T \cdot N+1}, \dots, x_{(T+1) \cdot N}\}$ , принадлежит объединению латинских множеств, то есть в точности тех областей, которые не обращают в ноль плотность  $\varphi(x_1, x_2, \dots, x_N)$ .

Резюмируем выше приведённое рассуждение: мы будем использовать формулу  $S_N[f]$  в сочетании с последовательностью Соболя, а оценивая дисперсию в соответствии с (2), мы получим оценку погрешности интегрирования, аналогичную доверительному интервалу метода Монте-Карло. Таким образом, на каждом шаге алгоритма Qint мы будем получать одновременно оценку интеграла  $I$  и оценку погрешности  $\varepsilon = |I - S_N[f]|$ .

Опишем шаги алгоритма Qint и исследуем вопрос выбора параметров. Пусть дана  $s$ -мерная функция  $f \in L^2(d\mu)$ , определённая и принимающая конечные значения в единичном гиперкубе  $[0, 1]^s$  (рассмотрение функций с особенностями остаётся за рамками данной статьи; отметим лишь, что требование конечности значений можно существенно ослабить).

1. Выбираем *параметр подразбиений* (partition parameter)  $P \in \mathbb{N}_0$ .
2. Выбираем *параметр повторений* (repetition parameter)  $R \in \mathbb{N}$ .
3. Строим разбиение  $[0, 1]^s$  на  $N = 2^P$  подмножеств равного объёма на основе элементарных подмножеств.
4. Для  $r = 1, 2, \dots, R$  повторяем следующий вычислительный цикл:
  - (a) берём  $N$  следующих<sup>3</sup> точек  $s$ -мерной последовательности Соболя  $\{x_{(r-1) \cdot N+1}, \dots, x_{r \cdot N}\}$ ; вычисляем значения подынтегральной функции в них и дописываем в массив уже имеющихся значений длины  $rN$ ;
  - (b) обновляем оценку интеграла:  $\hat{I} = \frac{1}{rN} \sum_{i=1}^{rN} f(x_i)$ ;
  - (c) обновляем оценку стандартного отклонения  $\hat{S}$  в соответствии с (2), где коэффициенты  $\alpha_j$  оцениваются по  $r$  точкам каждый:  $\hat{\alpha}_j = \frac{1}{r} \sum_{i \in \mathfrak{X}_j} f(x_i)$ ;
  - (d) для количества вычислений подынтегральной функции, равного  $rN$ , место нахождения истинного значения оценивается при помощи *квазидоверительного интервала*  $(\hat{I} - 3\hat{S}; \hat{I} + 3\hat{S})$ .

Важно понимать, что чем больше параметр  $P$ , тем больше  $N$ , и тем реже будет обновляться оценка, поскольку она определена только для  $\{N, 2N, 3N, \dots, kN, \dots\}$ . Для промежуточных значений  $\{kN + 1, kN + 2, kN + 3, \dots, (k+1)N - 1\}$  мы будем использовать последнюю доступную оценку, полученную на итерации  $k$ . Таким образом, с ростом  $P$  оценка будет носить всё более ярко выраженный ступенчатый характер.

Отметим отдельно, что оценка интеграла  $I$  при помощи выборочного среднего  $\hat{I}$  есть стандартный подход квази Монте-Карло, поэтому требование сходимости оценки к истинному значению не требует никакого формального доказательства: оно выполнено автоматически. При этом основной результат алгоритма Qint состоит в том,

<sup>3</sup> Вообще говоря, можно брать  $N$  любых подряд идущих не использованных ранее точек; мы пренебрегаем этой возможностью для простоты.

что мы имеем альтернативную оценку погрешности  $|I - \hat{I}|$ . Отсутствие такого инструмента всегда являлось важным недостатком использования квази Монте-Карло в прикладных задачах; мы надеемся, что распространение и дальнейшее усовершенствование Qint позволит расширить область применения квази Монте-Карло.

**3. Параметризация Qint. 3.1. Двойственность параметров подразбиений и повторений.** Зачастую в прикладных вычислениях явный вид функции  $f$  неизвестен, а вычисление её значения может занимать значительное время. Поэтому для сравнения различных алгоритмов численного интегрирования часто используется ограничение на количество вычислений функции. Мы будем использовать этот подход для сравнения вариантов Qint с разными наборами параметров.

Пусть мы ограничены  $M$  вычислениями подынтегральной функции, причём  $M = 2^m$ , где  $m$  — достаточно большое натуральное число. Такое ограничение несущественно и вводится только для удобства демонстрации связи параметров  $P$  и  $R$ . Если представить  $M$  в виде  $M = R \cdot 2^P$ , то допустимые значения для этих параметров исчерпываются парами вида  $\{P = 0, R = 2^m\}, \{P = 1, R = 2^{m-1}\}, \{P = 2, R = 2^{m-2}\}, \dots, \{P = i, R = 2^{m-i}\}, \dots, \{P = m, R = 1\}$ . Мы уже упоминали, что с ростом  $P$  оценка становится более ступенчатой, но, что куда более важно, дисперсия  $S_M[f]$  монотонно убывает по  $P$ .

**Теорема 1.** Пусть для параметров  $\{P, R\}$ ,  $R \geq 2$ , есть разбиение  $\mathfrak{X}_1, \mathfrak{X}_2, \dots, \mathfrak{X}_{2^P}$ . Если для параметров  $\{P + 1, R/2\}$  взять такое разбиение  $\mathfrak{Y}_1, \mathfrak{Y}_2, \dots, \mathfrak{Y}_{2^{P+1}}$ , что  $\mathfrak{X}_i = \mathfrak{Y}_{2i-1} \cup \mathfrak{Y}_{2i}$  для любого  $i$ , то  $\mathbb{D}S_{M, \{P, R\}}[f] \geq \mathbb{D}S_{M, \{P+1, R/2\}}[f]$ .

**ДОКАЗАТЕЛЬСТВО.** Пусть для исходного разбиения определены коэффициенты  $\alpha_i = \int_{\mathfrak{X}_i} f(x)\mu(dx)$ ,  $i = 1, 2, \dots, 2^P$ , участвующие в (2). Для нового разбиения пусть  $\beta_j = \int_{\mathfrak{Y}_j} f(x)\mu(dx)$ ,  $j = 1, 2, \dots, 2^{P+1}$ . Достаточно доказать только  $\sum_{i_\alpha < j_\alpha} (\alpha_{i_\alpha} - \alpha_{j_\alpha})^2 \leq \sum_{i_\beta < j_\beta} (\beta_{i_\beta} - \beta_{j_\beta})^2$ . Этот факт докажем почленно: рассмотрим сначала  $i_\alpha = 1, j_\alpha = 2$  и соответствующие им  $i_\beta = 1, 2, j_\beta = 3, 4$ . Поскольку новое разбиение получено на основе исходного, верны соотношения  $\alpha_1 = \beta_1 + \beta_2, \alpha_2 = \beta_3 + \beta_4$ . Если неравенство выполнено, то аналогичное неравенство будет верно и для произвольных  $i_\alpha, j_\alpha$ , которыми исчерпывается исходная сумма, и соответствующих им непересекающихся наборов  $i_\beta, j_\beta$ . Итак, для фиксированных индексов имеем следующую цепочку неравенств:

$$\begin{aligned} (\alpha_1 - \alpha_2)^2 &\leq \\ &\leq (\beta_1 - \beta_2)^2 + (\beta_1 - \beta_3)^2 + (\beta_1 - \beta_4)^2 + (\beta_2 - \beta_3)^2 + (\beta_2 - \beta_4)^2 + (\beta_3 - \beta_4)^2; \end{aligned}$$

$$\begin{aligned} \alpha_1^2 - 2\alpha_1\alpha_2 + \alpha_2^2 &\leq \\ &\leq 3\beta_1^2 + 3\beta_2^2 + 3\beta_3^2 + 3\beta_4^2 - 2\beta_1\beta_2 - 2\beta_1\beta_3 - 2\beta_1\beta_4 - 2\beta_2\beta_3 - 2\beta_2\beta_4 - 2\beta_3\beta_4; \end{aligned}$$

$$\begin{aligned} 2\beta_1\beta_2 + 2\beta_3\beta_4 &\leq 2\beta_1^2 + 2\beta_2^2 + 2\beta_3^2 + 2\beta_4^2 - 2\beta_1\beta_2 - 2\beta_3\beta_4; \\ 0 &\leq 2(\beta_1 - \beta_2)^2 + 2(\beta_3 - \beta_4)^2. \end{aligned}$$

■

Итак, наращивая параметр подразбиений и уменьшая параметр повторений, мы вправе ожидать более точной оценки погрешности. Это довольно естественный результат, потому что вырожденный случай  $P = 0$  соответствует обычному Монте-Карло, в котором вместо псевдослучайных чисел используется последовательность Соболя. С другой стороны, чем больше  $P$ , тем меньше  $R$ , и тем потенциально более неустойчивы оценки  $\alpha_i$  (для них используются только те точки, которые лежат в  $\mathfrak{X}_i$ , которых в точности  $R$ ). Практика показывает, что максимально большое  $P$  (случай  $\{P = m, R = 1\}$ ) приводит к недостоверной оценке дисперсии, так что мы будем рассматривать только такие  $P$ , для которых  $R \geq 2$ .

Подведём итог рассуждениям о двойственности параметров Qint. Для постоянного количества вычислений подынтегральной функции мы можем варьировать соотношение между параметром подразбиений  $P$  и параметром повторений  $R$ . В силу монотонности по  $P$  чем больше  $P$ , тем более точной и ступенчатой будет оценка погрешности. С другой стороны, чем больше  $R$ , тем более оценка погрешности будет стабильной и плавной. Эти эффекты мы увидим в вычислительных примерах, приведённых в заключительной части статьи.

**3.2. Параметризация формы разбиений.** Алгоритм Qint не описывается однозначно парой параметров  $(P, R)$ , поскольку необходимо дополнительно указать правила формирования разбиений  $\mathfrak{X}_1, \mathfrak{X}_2, \dots, \mathfrak{X}_{r \cdot 2^P}$  для каждого  $r = 1, 2, \dots, R$ . Вообще говоря, для разных  $r$  могут быть совершенно разные наборы, но мы будем рассматривать только такие универсальные способы построения, для которых характерна естественная итеративность по  $P$ . В частности, в [1] описан один из таких способов, который устроен чрезвычайно просто: разбиение всегда проводится по первой координате гиперкуба.

В этой статье мы используем более нетривиальный способ, который мы назовём *правилом кубической формы* (*cubic shape rule*). Мы не будем его подробно описывать; скажем лишь, что гиперкуб разбивается равномерно по всем имеющимся размерностям. Шаги правила кубической формы в трёхмерном случае представлены на рис. 1.

**4. Численные примеры.** В предыдущей работе [1] рассмотрена функция довольно простого вида. В качестве более продвинутых демонстрационных примеров мы возьмём три функции: одну функцию специального вида, близкую к первой хараровской и две из набора TESTPACK (набор функций для тестирования алгоритмов многомерного интегрирования, составленный Аланом Генцем, [7]) — рис. 2–4 соответственно.

**4.1. Кусочно-линейная функция.** Рассмотрим функцию, задаваемую прямым произведением координатных функций, для произвольной размерности  $i$  задаваемых как

$$f_i(x) = 2 \cdot \begin{cases} 0, & x \leq 0.5 - c_i, \\ \frac{10+i}{i}x - \frac{5}{i}, & x \in (0.5 - c_i, 0.5 + c_i], \\ 1, & x > 0.5 + c_i, \end{cases}$$

где  $c_i = \frac{i}{2i+10}$ . Такая функция обладает интересным свойством: в низких размерностях она очень близка к первой функции Хаара, в высоких становится практически линейной  $f(x) = x$ . Интегрируя такую функцию при помощи Qint, мы рассчитываем на очень точную оценку дисперсии. Кроме того, мы сможем оценить преимущество, получаемое при помощи правила кубической формы.

Размерность пространства  $s = 8$ , истинное значение интеграла  $I = 1$ .

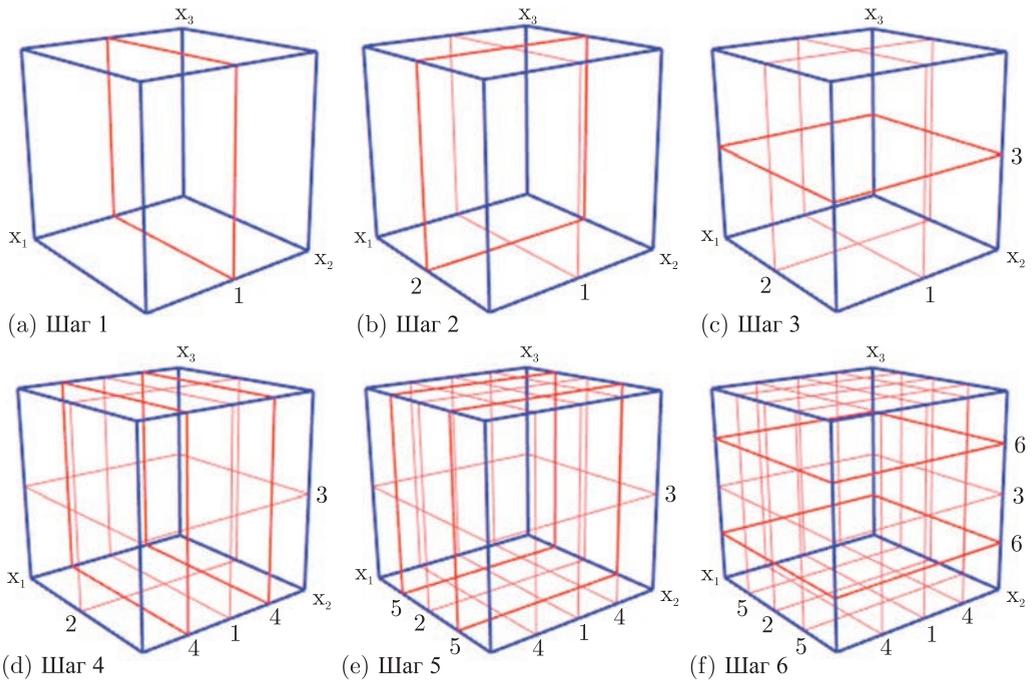


Рис. 1. Первые шесть шагов правила кубической формы,  $s = 3$ .

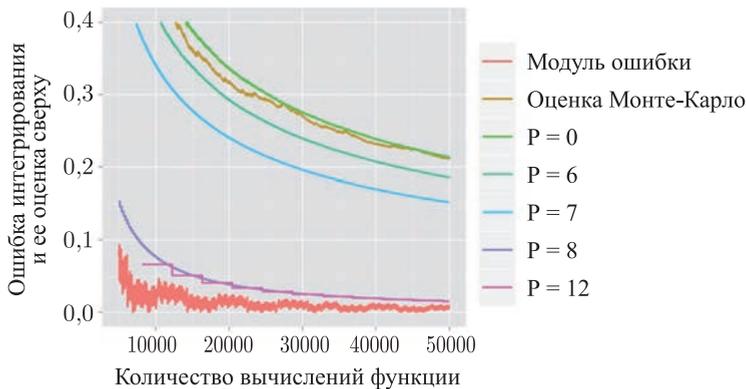


Рис. 2. Кусочно-линейная функция в размерности 8.

Стоит отметить тот факт, что качественный скачок в точности оценки дисперсии происходит тогда, когда параметр подразбиений становится равен размерности пространства. Это происходит ровно в тот момент, когда правило кубической формы задействует разбиения по всем имеющимся размерностям. Такого скачка не наблюдалось бы с разбиениями только по первой координате.

**4.2. Осциллирующая функция из TESTPACK.** В качестве второго примера возьмём функцию под кодовым именем «Oscillatory» из набора TESTPACK [7]:

$$f_{osc}(\mathbf{x}) = \cos \left( \sum_{i=1}^s x_i \right).$$

Такого рода функции часто встречаются в практических задачах и дают представление о преимуществах использования Qint в тех случаях, когда функция не подобрана специальным образом, как это было в первом примере.

Размерность пространства  $s = 20$ , истинное значение интеграла  $I \approx -0.362$ .

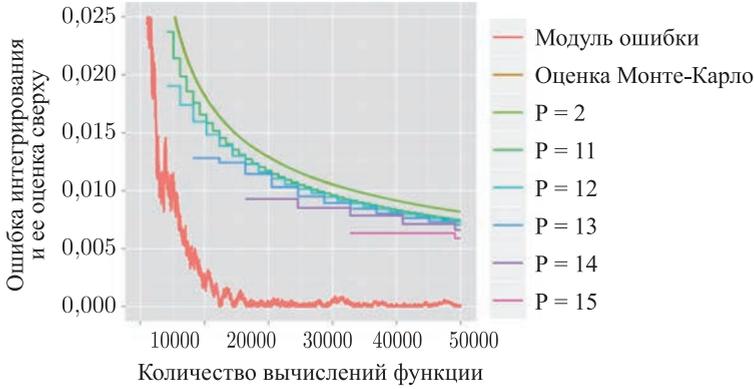


Рис. 3. Осциллирующая функция в размерности 20.

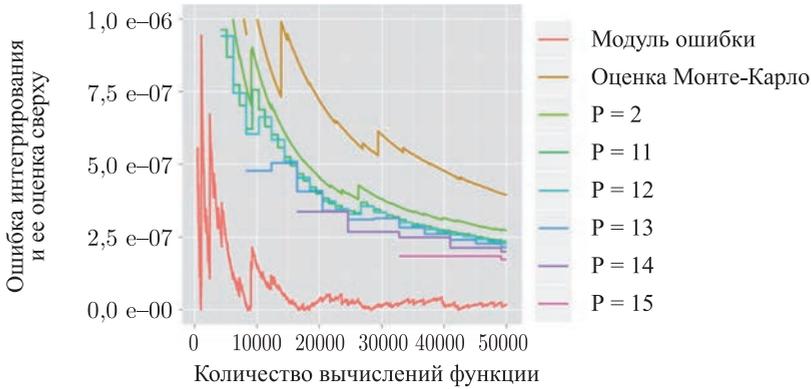


Рис. 4. Функция с пиком в размерности 8.

На этом примере хорошо видны все особенности оценок Qint: растущий вместе с параметром подразбиений выигрыш по сравнению с доверительным интервалом Монте-Карло; всё более выраженная ступенчатость оценок; монотонность по параметру подразбиений.

**4.3. Функция с явно выраженным пиком из TESTPACK.** Заключительным примером будет функция под кодовым именем «Corner peak» из того же набора TESTPACK:

$$f_{cp}(\mathbf{x}) = \left( 1 + \sum_{i=1}^s x_i \right)^{-s-1}.$$

Она интересна тем, что у неё есть область быстрого роста, и значения в этой области значительно выше, чем среднее значение по всему гиперкубу<sup>5</sup>. Это означает,

<sup>5</sup> Так, для рассматриваемого далее случая  $s = 8$  минимум функции составляет  $1 \times 10^{-9}$ , значение в центре куба примерно  $2.2 \times 10^{-7}$ , а пик находится в точке  $f(0, \dots, 0) = 1$ .

что дисперсия, посчитанная методом Монте-Карло, будет большой, а доверительный интервал будет нестабилен. Продемонстрируем, является ли это препятствием для Qint.

Размерность пространства  $s = 8$ , истинное значение интеграла  $I \approx 2.8 \times 10^{-6}$ .

Безусловно, скачки в оценке дисперсии портят не только доверительный интервал Монте-Карло, но и оценки Qint. Однако увеличение параметра подразбиений и связанное с этим усреднение помогает справиться с этой проблемой и получить стабильные оценки, пожертвовав при этом начальным отрезком. Так, для  $P = 14$  мы не сможем построить оценку раньше, чем при  $N$ , близком 16000; тем не менее, для остальных  $N$  оценка стабильна и значительно более точна по сравнению с обычным Монте-Карло.

## Литература

1. Антонов А. А., Ермаков С. М. Эмпирическая оценка погрешности интегрирования методом квази Монте-Карло // Вестн. С.-Петерб. ун-та. Сер. 1. Т. 1(59). 2014. С. 3–11.
2. Соболев И. М. О распределении точек в кубе и приближенном вычислении интегралов // Ж. вычисл. матем. и матем. физ. Т. 7(4). 1967. С. 784–802.
3. Соболев И. М. Многомерные квадратурные формулы и функции Хаара. М.: Наука, 1969.
4. Ермаков С. М. Метод Монте-Карло и смежные вопросы. М.: Наука, 1975.
5. Cochran W. G. Sampling Techniques, 3rd Edition. John Wiley, 1977.
6. Dick J., Pillichshammer F. Digital Nets and Sequences. New York: Cambridge University Press, 2010.
7. Genz A. A Package for Testing Multiple Integration Subroutines // Numer. Integr.: Recent Dev., Softw. and Appl. Vol. 203. 1987. P. 337–340.
8. Donald E. Knuth. Seminumerical Algorithms // The Art of Computer Programming. Vol. 2. Addison-Wesley, Reading, Massachusetts, Second, 1981.
9. Kuipers L., Niederreiter H. Uniform Distribution Of Sequences. Wiley. New York. 1974.
10. Lemieux C. Monte Carlo and Quasi-Monte Carlo Sampling. New York: Springer, 2009.
11. Niederreiter H. Error bounds for Quasi-Monte Carlo integration with uniform point sets // J. of Comput. and Appl. Math. 2003. 150. P. 283–292.
12. Niederreiter H. Random Number Generation and Quasi-Monte Carlo Methods. Dover: Dover Publications, 2006.
13. Owen A. B. Scrambled net variance for integrals of smooth functions // Annals of Statistics. 1997. Vol. 25, N 4. P. 1541–1562.
14. Schürer R. A Comparison between (Quasi-)Monte Carlo and Cubature Rule Based Methods for Solving High-dimensional Integration Problems // Mathematics and Computers in Simulation. 2003. Vol. 62, N 3–6. P. 509–517.

Статья поступила в редакцию 23 октября 2014 г.

Сведения об авторах

Антонов Антон Александрович — аспирант; [tonytonov@gmail.com](mailto:tonytonov@gmail.com)

## QUASI MONTE CARLO INTEGRATION ALGORITHM WITH A POSTERIORI ERROR ANALYSIS

Anton A. Antonov

St.Petersburg State University, Universitetskaya nab., 7-9, St.Petersburg, 199034, Russian Federation; [tonytonov@gmail.com](mailto:tonytonov@gmail.com)

A possibility of a probabilistic approach to a deterministic error estimation procedure of quasi Monte-Carlo method is presented. Existing estimation methods of the aforementioned estimation are non-constructive.

A previously obtained result, which was deduced from the theory of random cubature formulae, is expanded with new theoretical results. Qint algorithm, which is presented in a convenient for direct

application step-by-step form, is discussed in terms of parametrization. Several key features of Qint are discussed, including monotonicity by a partition parameter and accuracy dependency on a partition rule.

New practical results are presented, which are obtained with the help of Sobol sequences and TESTPACK integrand functions. In all cases a better estimate compared to a traditional Monte Carlo confidence interval is demonstrated. Refs 14. Figs 4.

*Keywords:* Monte Carlo and quasi Monte Carlo method, confidence interval, random cubature formulae, Haar functions, Sobol sequences.

## References

1. Antonov A. A., Ermakov S. M., “Empirically estimating error of integration by Quasi-Monte Carlo method”, *Vestnik St.Petersburg University. Mathematics* **47**(1), 1–8 (2014).
2. Sobol’ I. M., “Distribution of points in a cube and approximate evaluation of integrals”, *Zh. Vych. Mat. Mat. Fiz.* **7**(4), 784–802 (1967).
3. Sobol’ I. M., *Multidimens. Quadrature Formulas and Haar Funct.* (Nauka, Moscow, 1969) [in Russian].
4. Ermakov S. M., *Die Monte Carlo Methode und Verwandte Fragen* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1975).
5. Cochran W. G., *Sampling Techniques*, 3rd ed. (John Wiley, 1977).
6. Dick J., Pillichshammer F., *Digital Nets and Sequences* (Cambridge University Press, New York, 2010).
7. Genz A., “A Package for Testing Multiple Integration Subroutines”, *Numer. Integr.: Recent Dev., Softw. and Appl.* **203**, 337–340 (1987).
8. Donald E. Knuth., “Seminumerical Algorithms”, *The Art of Computer Programming 2* (Addison-Wesley, Reading, Massachusetts, Second, 1981).
9. Kuipers L., Niederreiter H., “Uniform Distribution Of Sequences” (Wiley, New York, 1974).
10. Lemieux C., “Monte Carlo and Quasi-Monte Carlo Sampling” (Springer, New York, 2009).
11. Niederreiter H., “Error bounds for Quasi-Monte Carlo integration with uniform point sets”, *J. of Comput. and Appl. Math.* **150**, 283–292 (2003).
12. Niederreiter H., “Random Number Generation and Quasi-Monte Carlo Methods” (Dover Publications, Dover, 2006).
13. Owen A. B., “Scrambled net variance for integrals of smooth functions”, *Annals of Statistics.* **25**(4), 1541–1562 (1997).
14. Schürer R., “A Comparison between (Quasi-)Monte Carlo and Cubature Rule Based Methods for Solving High-dimensional Integration Problems”, *Mathematics and Computers in Simulation.* **62**(3–6), 509–517 (2003).