

МАТЕМАТИКА

УДК 519.644.2

**ЭМПИРИЧЕСКАЯ ОЦЕНКА ПОГРЕШНОСТИ ИНТЕГРИРОВАНИЯ
МЕТОДОМ КВАЗИ МОНТЕ-КАРЛО****А. А. Антонов, С. М. Ермаков*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. Известное неравенство Коксмы—Хлавки включает в качестве константы вариацию подынтегральной функции. Вычисление же этой вариации является более трудной задачей, чем исходная. Поскольку метод квази Монте-Карло использует в качестве оценки интеграла среднее арифметическое значение интегральной функции, то можно было бы ожидать, что распределение (в теоретико-числовом смысле) остатка приближенного интегрирования подчиняется нормальному закону. Дополнительная трудность состоит, однако, в том, что квазислучайные точки, рассматриваемые с теоретико-вероятностной точки зрения, являются зависимыми, и численная оценка 2-го момента остатка тем самым затруднена.

Авторы предлагают новый подход к проблеме оценки 2-го момента погрешности, основанный на полученных авторами результатах из теории случайных кубатурных формул. Приводимые в конце работы численные примеры свидетельствуют о перспективности разработанного метода оценки погрешности. Библиогр. 7 назв.

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

1. Введение. Известно, что важной особенностью метода Монте-Карло при вычислении интегралов является возможность эмпирической оценки погрешности (построение доверительного интервала). Метод квази Монте-Карло, строго говоря, такой особенностью не обладает. Известное неравенство Коксмы—Хлавки (см. [7]) требует оценки вариации интегрируемой функции и величины дискрепанса квазислучайной последовательности. Эти задачи существенно более трудные, чем исходная задача численного интегрирования. Последнее обстоятельство привело к использованию так называемых рандомизированных квазислучайных последовательностей (см. [5]). Их использование может, однако, существенно увеличить вычислительную работу.

Рассмотрим в качестве примера один из способов построения рандомизированных квазислучайных последовательностей (см. [6]). Пусть X_1, X_2, \dots, X_N — ква-

*Работа выполнена при финансовой поддержке РФФИ (грант № 11-01-00769-а).

зислучайные d -мерные векторы, расположенные в d -мерном единичном кубе D_d , а R_1, R_2, \dots, R_M — набор случайных векторов, равномерно распределенных в D_d . Если $\{Y\}$ обозначает вектор, компонентами которого являются дробные доли компонент вектора Y , то сумма

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

служит оценкой интеграла $\int_{D_d} f(X)dX$. Порядок погрешности будет равен $O(\frac{N}{\sqrt{M \ln(N)}})$, математическое ожидание величины $\Sigma_{MN}(f)$ есть искомый интеграл, и ее дисперсия может быть использована для построения доверительного интервала. Очевидно, что без использования рандомизации при том же объеме вычислительной работы мы имели бы погрешность порядка $O(\frac{\ln(MN)}{MN})$, что в $\frac{\ln(M)}{\sqrt{M}}$ раз лучше рандомизированной оценки.

Возникает вопрос: нельзя ли обойтись без рандомизации при оценке погрешности? Поскольку квазислучайные последовательности имеют некоторые свойства, сходные со свойствами случайных чисел (аналог закона больших чисел, например, см. [4]), 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 $\int_{D_d} f(X)dX$. Что касается дисперсии, то ее 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 объема d , содержащийся в единичном гиперкубе D_s с вероятностью d , то из N квазислучайных точек такому параллелепипеду принадлежит $\lfloor N/d \rfloor$ точек с вероятностью, близкой к единице. Для последовательностей Соболя, связанных с функциями Хаара (см. [1, 2]), 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 в этом случае зависимость между квазислучайными узлами интегрирования выражается в принадлежности с вероятностью 1 некоторых подмножеств узлов различным гиперпараллелепипедам. Указанные соображения делают оправданным привлечение к оценке погрешности квазислучайных методов интегрирования теории кубатурных формул со случайными узлами.

Далее известный результат (см. [3]) относительно дисперсии квадратурной суммы в случае точности соответствующей формулы для двух функций Хаара распространяется на общий случай и приводятся численные примеры, свидетельствующие о перспективности данного подхода.

2. Вывод формулы и анализ дисперсии. Пусть пространство \mathfrak{X} наделено мерой μ , причем $\mu(\mathfrak{X}) = 1$. Разобьем \mathfrak{X} на N непересекающихся частей равной меры: $\mathfrak{X} = \mathfrak{X}_1 \cup \mathfrak{X}_2 \cup \dots \cup \mathfrak{X}_N$, $\mathfrak{X}_i \cap \mathfrak{X}_j = \emptyset \forall i \neq j$, $\mu(\mathfrak{X}_1) = \mu(\mathfrak{X}_2) = \dots = \mu(\mathfrak{X}_N) = \frac{1}{N}$. Пусть

функция $f : \mathfrak{X} \rightarrow \mathbb{R}$ μ -интегрируема с квадратом ($f \in L^2(d\mu)$). Далее, пусть $N = 2^s$. Построим N функций Хаара для $\mathfrak{X}_1, \mathfrak{X}_2, \dots, \mathfrak{X}_N$:

$$\chi_1 = 1;$$

$$\chi_i = \begin{cases} 2^{\frac{q_i-1}{2}}, & x \in \mathfrak{X}_j, j \in \{i_{st}, \dots, i_{st} + 2^{s-q_i} - 1\}; \\ -2^{\frac{q_i-1}{2}}, & x \in \mathfrak{X}_k, k \in \{i_{st} + 2^{s-q_i}, \dots, i_{st} + 2^{s-q_i+1} - 1\}; \\ 0, & \text{иначе;} \end{cases}$$

где для $i > 1$ $q_i = \lceil \log_2 i \rceil$, $i_{st} = (i-1) \cdot 2^{s-q_i+1} - 2^s + 1$.

Утверждение. $\{\chi_i\}_{i=1}^N$ — ортонормированная система в $L^2(d\mu)$:

- 1) $\int_{\mathfrak{X}} \chi_i^2 d\mu = 1 \quad \forall i$;
- 2) $\int_{\mathfrak{X}} \chi_i \chi_j d\mu = 0 \quad \forall i \neq j$.

Используя аппарат случайных кубатурных формул, построим формулу, точную для N функций Хаара. В силу теоремы Ермакова—Грановского [3] она будет выглядеть, как $S_N[f] = \frac{\Delta(f, Q)}{\Delta(Q)}$ в обозначениях $\Delta(f, Q) = \det |f(x_i), \chi_2(x_i), \dots, \chi_N(x_i)|_{i=1}^N$ и $\Delta(Q) = \Delta(\chi_1, Q)$. Плотность совместного распределения узлов этой формулы будет равна $\frac{1}{N!}(\Delta(Q))^2$.

Определитель $\Delta(Q)$ равен сумме всевозможных произведений вида $\pm \chi_1(x_{i_1}) \chi_2(x_{i_2}) \dots \chi_N(x_{i_N})$, где (i_1, i_2, \dots, i_N) есть перестановка $(1, 2, \dots, N)$, причем каждое такое произведение однозначно определяется своей перестановкой. Заметим, что если хотя бы две точки попадают в одно и тоже подмножество \mathfrak{X}_k , то, поскольку для любого i значение $\chi_i(x)$ постоянно $\forall x \in \mathfrak{X}_k$, определитель $\Delta(Q)$ имеет две одинаковые строки и, следовательно, обращается в ноль. Это означает, что плотность совместного распределения узлов может быть ненулевой только тогда, когда все точки x_i попадают в разные подмножества \mathfrak{X}_k . Дадим определение таким множествам.

Определение. Латинским множеством $\mathfrak{Lat}(i_1, i_2, \dots, i_N)$, задаваемым перестановкой (i_1, i_2, \dots, i_N) , мы будем называть множество, задаваемое следующим образом:

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

и при этом $k_{j_1} \neq k_{j_2} \quad \forall j_1 \neq j_2$.

Определитель $\Delta(f, Q)$ также является суммой произведений вида $\pm f(x_{i_1}) \chi_2(x_{i_2}) \dots \chi_N(x_{i_N})$. Если такое произведение не есть ноль, то оно равно $\pm f(x_{i_1}) \cdot C_N$, где $C_N > 0$ — некая константа, не зависящая от перестановки (i_1, i_2, \dots, i_N) . Если $\mathbf{Inv}(i_1, i_2, \dots, i_N)$ обозначает количество инверсий в перестановке (i_1, i_2, \dots, i_N) , то для точек $(x_1, x_2, \dots, x_N) \in \mathfrak{Lat}(i_1, i_2, \dots, i_N)$ будет выполнено

$$\Delta(f, Q) = (-1)^{\mathbf{Inv}(i_1, i_2, \dots, i_N)} \cdot C_N \cdot (f(x_1) + f(x_2) + \dots + f(x_N)) \cdot C_z,$$

$$\Delta(Q) = (-1)^{\mathbf{Inv}(i_1, i_2, \dots, i_N)} \cdot C_N \cdot (1 + 1 + \dots + 1) \cdot C_z,$$

где константа $C_z > 0$ отвечает за количество ненулевых произведений.

Итак, кубатурная формула имеет вид

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

а плотность совместного распределения узлов —

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

Таким образом, нами доказана следующая

Теорема 1. *Кубатурная формула $S_N[f]$ с плотностью распределения узлов $\varphi(x_1, x_2, \dots, x_N)$ является случайной кубатурной формулой, точной для семейства функций Хаара $\chi_1, \chi_2, \dots, \chi_N$.*

Хотя факт несмещенности оценки непосредственно следует из вышеприведенной теоремы, мы можем доказать его отдельно в качестве проверки:

$$\begin{aligned} \mathbb{E}S_N[f] &= \int_{\mathfrak{X}^N} S_N[f] \cdot \frac{N^N}{N!} \cdot \mathbb{1}_{\mathfrak{Lat}(i_1, i_2, \dots, i_N)}(x_1, x_2, \dots, x_N) \mu(dx_1) \mu(dx_2) \dots \mu(dx_N) = \\ &= \frac{1}{N} \left(\int_{\mathfrak{X}} f(x_1) \mu(dx_1) + \dots + \int_{\mathfrak{X}} f(x_N) \mu(dx_N) \right) = \int_{\mathfrak{X}} f(x) \mu(dx). \end{aligned}$$

Ключевым вопросом здесь является величина дисперсии построенной квадратурной формулы. Заметим, что теорема Ермакова—Грановского дает точное значение для дисперсии в случае регулярных систем, но система Хаара таковой не является, что приводит нас к необходимости вычислять дисперсию непосредственно:

$$\mathbb{D}S_N[f] = \int_{\mathfrak{X}^N} S_N^2[f] d\varphi - \left(\int_{\mathfrak{X}^N} S_N[f] d\varphi \right)^2 = A - B,$$

где в силу несмещенности

$$B = \left(\int_{\mathfrak{X}} f(x) \mu(dx) \right)^2.$$

Вычислим A :

$$\begin{aligned} A &= \frac{1}{N^2} \int_{\mathfrak{X}^N} (f(x_1) + f(x_2) + \dots + f(x_N))^2 d\varphi = \\ &= \frac{N^{N-2}}{N!} \left\{ \int_{\mathfrak{X}^N} N \cdot f^2(x_1) \mathbb{1}_{\mathfrak{Lat}(i_1, i_2, \dots, i_N)}(x_1, x_2, \dots, x_N) \mu(dx_1) \mu(dx_2) \dots \mu(dx_N) + \right. \\ &\quad \left. + 2 \sum_{i < j} \int_{\mathfrak{X}^N} f(x_i) f(x_j) \mathbb{1}_{\mathfrak{Lat}(i_1, i_2, \dots, i_N)}(x_1, x_2, \dots, x_N) \mu(dx_1) \mu(dx_2) \dots \mu(dx_N) \right\} \\ &= \frac{N^{N-2}}{N!} \left\{ N \cdot \int_{\mathfrak{X}^N} f^2(x_1) \mathbb{1}_{\mathfrak{Lat}(i_1, i_2, \dots, i_N)}(x_1, x_2, \dots, x_N) \mu(dx_1) \mu(dx_2) \dots \mu(dx_N) + \right. \\ &\quad \left. + N(N-1) \int_{\mathfrak{X}^N} f(x_1) f(x_2) \mathbb{1}_{\mathfrak{Lat}(i_1, i_2, \dots, i_N)}(x_1, x_2, \dots, x_N) \mu(dx_1) \mu(dx_2) \dots \mu(dx_N) \right\} = \\ &= \frac{N^{N-2}}{N!} [N \cdot C + N(N-1) \cdot D]. \end{aligned}$$

Выражения C и D посчитаем отдельно:

$$C = \int_{\mathfrak{X}^N} f^2(x_1) \mathbb{1}_{\mathfrak{Lat}(i_1, i_2, \dots, i_N)}(x_1, x_2, \dots, x_N) \mu(dx_1) \mu(dx_2) \dots \mu(dx_N) = \\ = \frac{(N-1)!}{N^{N-1}} \int_{\mathfrak{X}} f^2(x) \mu(dx);$$

$$D = \int_{\mathfrak{X}^N} f(x_1) f(x_2) \mathbb{1}_{\mathfrak{Lat}(i_1, i_2, \dots, i_N)}(x_1, x_2, \dots, x_N) \mu(dx_1) \mu(dx_2) \dots \mu(dx_N) = \\ = \frac{(N-2)!}{N^{N-2}} \int_{\mathfrak{X}_1 \neq \mathfrak{X}_2} f(x_1) f(x_2) \mu(dx_1) \mu(dx_2) = \\ = \frac{(N-2)!}{N^{N-2}} \left\{ \int_{\mathfrak{X}^2} f(x_1) f(x_2) \mu(dx_1) \mu(dx_2) - \int_{\mathfrak{X}_1 = \mathfrak{X}_2} f(x_1) f(x_2) \mu(dx_1) \mu(dx_2) \right\} = \\ = \frac{(N-2)!}{N^{N-2}} \left\{ \left(\int_{\mathfrak{X}} f(x) \mu(dx) \right)^2 - \sum_{i=1}^N \int_{\mathfrak{X}_i} f(x_1) \mu(dx_1) \int_{\mathfrak{X}_i} f(x_2) \mu(dx_2) \right\} = \\ = \frac{(N-2)!}{N^{N-2}} \left\{ \left(\int_{\mathfrak{X}} f(x) \mu(dx) \right)^2 - \sum_{i=1}^N \left(\int_{\mathfrak{X}_i} f(x) \mu(dx) \right)^2 \right\}.$$

Таким образом,

$$A = \frac{N^{N-2}}{N!} [N \cdot C + N(N-1) \cdot D] = \\ = \frac{N^{N-2}}{N!} \left[\frac{N!}{N^{N-1}} \int_{\mathfrak{X}} f^2(x) \mu(dx) + \frac{N!}{N^{N-2}} \left\{ \left(\int_{\mathfrak{X}} f(x) \mu(dx) \right)^2 - \sum_{i=1}^N \left(\int_{\mathfrak{X}_i} f(x) \mu(dx) \right)^2 \right\} \right] = \\ = \frac{1}{N} \int_{\mathfrak{X}} f^2(x) \mu(dx) + \left(\int_{\mathfrak{X}} f(x) \mu(dx) \right)^2 - \sum_{i=1}^N \left(\int_{\mathfrak{X}_i} f(x) \mu(dx) \right)^2,$$

что приводит к окончательному виду дисперсии:

$$\mathbb{D}S_N[f] = A - B = \frac{1}{N} \int_{\mathfrak{X}} f^2(x) \mu(dx) - \sum_{i=1}^N \left(\int_{\mathfrak{X}_i} f(x) \mu(dx) \right)^2 = \\ = \frac{1}{N} \int_{\mathfrak{X}} f^2(x) \mu(dx) - \frac{1}{N} \left(\int_{\mathfrak{X}} f(x) \mu(dx) \right)^2 + \frac{1}{N} \left(\int_{\mathfrak{X}} f(x) \mu(dx) \right)^2 - \sum_{i=1}^N \left(\int_{\mathfrak{X}_i} f(x) \mu(dx) \right)^2 = \\ = \mathbb{D}_{MC}[f] + \frac{1}{N} (\alpha_1 + \alpha_2 + \dots + \alpha_N)^2 - \alpha_1^2 - \alpha_2^2 - \dots - \alpha_N^2 = \mathbb{D}_{MC}[f] - \frac{1}{N} \sum_{i < j} (\alpha_i - \alpha_j)^2,$$

где $\mathbb{D}_{MC}[f] = \frac{1}{N} \int_{\mathfrak{X}} f^2(x) \mu(dx) - \frac{1}{N} \left(\int_{\mathfrak{X}} f(x) \mu(dx) \right)^2$ — дисперсия традиционного метода

Монте-Карло, в то время как $\alpha_i = \int_{\mathfrak{X}_i} f(x) \mu(dx)$, $i = 1, 2, \dots, N$.

Таким образом, нами получен следующий результат.

Теорема 2. *Кубатурная формула $S_N[f]$ с плотностью распределения узлов $\varphi(x_1, x_2, \dots, x_N)$ имеет дисперсию, равную $\mathbb{D}S_N[f]$. Эта дисперсия всегда не больше, чем дисперсия традиционного метода Монте-Карло.*

Отметим также, что выполняется формальное требование $\mathbb{D}S_N[f] \geq 0$, что немаловажно для описанного далее алгоритма.

3. Схема эксперимента и численные результаты. Полученная нами формула $S_N[f]$ имеет весьма простой вид. Сложность в ее практическом применении, однако, кроется в моделировании плотности $\varphi(x_1, x_2, \dots, x_N)$, что особенно актуально в случае высокой размерности. Алгоритм такого моделирования может иметь следующий вид:

1) компоненты вектора Y моделируются равномерно в каждом из множеств $\mathfrak{X}_1, \mathfrak{X}_2, \dots, \mathfrak{X}_N$: $y_i \in U(\mathfrak{X}_i)$;

2) случайная перестановка (i_1, i_2, \dots, i_N) выбирается равномерно из множества всевозможных перестановок чисел $(1, 2, \dots, N)$;

3) финальный вектор X образуется под действием (i_1, i_2, \dots, i_N) на вектор Y .

Полученный вектор X будет распределен с искомой плотностью $\varphi(x_1, x_2, \dots, x_N)$. Этому факту, безусловно, требуется некое формальное доказательство. Мы не будем его приводить здесь, поскольку не будем использовать этот алгоритм. Вместо этого мы остановимся на одном простом частном случае, который, тем не менее, исчерпывает собой подавляющее количество прикладных вычислений.

Пусть \mathfrak{X} есть d -мерный куб $[0, 1]^d$, а μ — мера Лебега на нем. В качестве $\mathfrak{X}_1, \mathfrak{X}_2, \dots, \mathfrak{X}_N$ мы будем рассматривать параллелепипеды вида $[\frac{i-1}{N}, \frac{i}{N}] \times [0, 1] \times \dots \times [0, 1]$ (формально $\mathfrak{X}_N = [\frac{N-1}{N}, 1] \times [0, 1] \times \dots \times [0, 1]$). Это означает, что для N точек с плотностью совместного распределения $\varphi(x_1, x_2, \dots, x_N)$ ровно одна должна иметь первую координату в отрезке $[0, \frac{1}{N})$, еще одна — в отрезке $[\frac{1}{N}, \frac{2}{N})$, и так далее, наконец, ровно одна — в отрезке $[\frac{N-1}{N}, 1]$. Таким замечательным свойством обладают точки Соболя. Это означает, что мы можем использовать любые N подряд идущих точек из обобщенной d -мерной последовательности Соболя.

Зафиксируем общее количество вычислений функции, обозначив его за M . Для удобства примем $M = k \cdot N = k \cdot 2^s$. Заметим, что для корректного сравнения результатов алгоритма с различными параметрами необходимо обеспечить постоянство M . Это означает, что при заданном M у нас есть два переменных параметра, k и s , но менять их мы можем только так, чтобы всегда выполнялось $M = k \cdot 2^s$. Далее, при выбранных k и s алгоритм оценивания будет выглядеть следующим образом:

1) возьмем M точек обобщенной последовательности Соболя x_i с номерами $i \in I_{k,s} = \{1, 2, \dots, k \cdot 2^s\}$;

2) построим оценку искомого интеграла $S_M[f] = \frac{1}{M} \sum_{i \in I_{k,s}} f(x_i)$;

3) построим оценку дисперсии $\mathbb{D}S_{k,2^s}[f]$ на основе величин $\hat{\alpha}_{k,j} = \frac{1}{k} \sum_{x_l \in \mathfrak{X}_j} f(x_l)$, где $j = 1, 2, \dots, N$, то есть мы оцениваем N коэффициентов $\alpha_{k,j}$ по k точкам каждый. На основе этой дисперсии строим доверительный интервал с уровнем доверия 95%.

Отметим двойственность выбора параметров k и s . С одной стороны, при $s = 0$ мы получаем не что иное, как традиционный метод Монте-Карло, а с ростом s мы ожидаем сокращения дисперсии. Следовательно, стоит выбирать s как можно большим. С другой стороны, при максимальном s имеем $k = 1$, то есть каждый из коэффициентов $\alpha_{k,j}$ оценивается всего по одной точке, и встает вопрос о достоверности такой оценки. Компромиссным решением в этой ситуации, по-видимому, будет использование неких средних величин k и s .

Продемонстрируем рассмотренный алгоритм и его чувствительность к изменениям параметров k и s на численных примерах.

В качестве тестовой функции рассмотрим функцию вида $f(x_1, x_2, \dots, x_d) = \frac{1}{(\ln 2)^d} \prod_{i=1}^d \frac{1}{1+x_i}$. Константа выбрана таким образом, что значение интеграла от этой функции для любой размерности d равно единице. Для наглядности мы будем рассматривать модуль истинного остатка в силу того, что оценки сверху симметричны относительно оси абсцисс. Мы будем сравнивать оценки, полученные при помощи формулы $S_M[f]$ для различных k и s . Для всех последующих графиков приняты следующие обозначения:

- по оси ординат отложены значения модуля остатка интегрирования и полученных оценок;
- по оси абсцисс — количество вычислений подынтегральной функции;
- красным цветом выделен истинный остаток интегрирования;
- синим — эмпирические оценки, полученные предложенным методом с различными параметрами;
- зеленым — доверительный интервал Монте-Карло с уровнем доверия 95%.

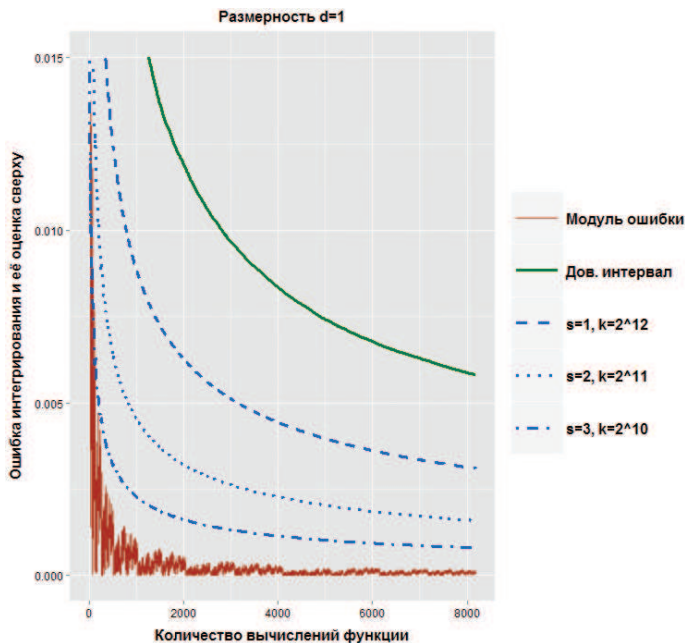


Рис. 1. Эмпирические оценки метода квази Монте-Карло, одномерный случай.

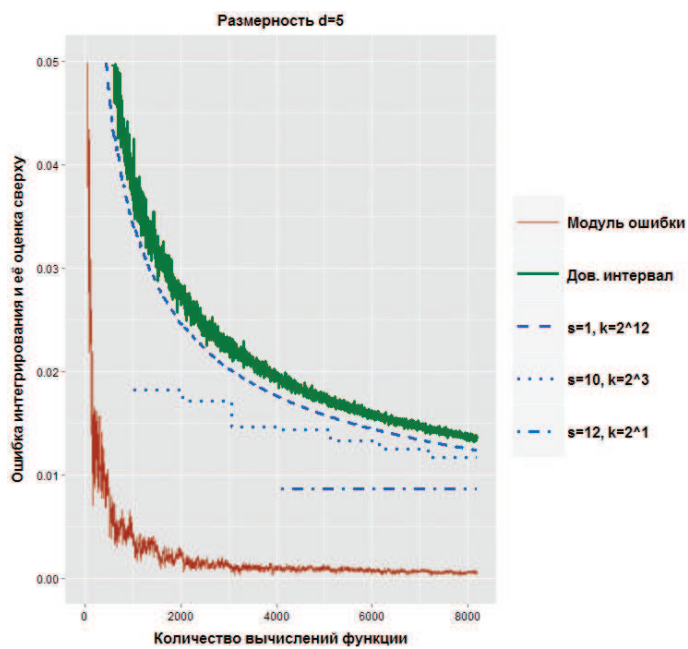


Рис. 2. Эмпирические оценки метода квази Монте-Карло, пятимерный случай.

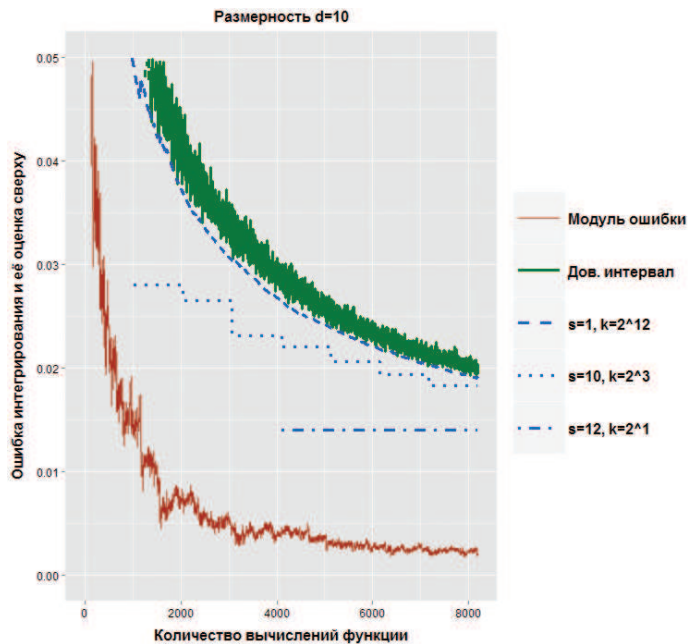


Рис. 3. Эмпирические оценки метода квази Монте-Карло, десятимерный случай.

Как видно из иллюстрации 1, даже при $s = 1$ мы имеем значительно более точную оценку остатка по сравнению с Монте-Карло. Увеличение s ведет к дальнейшему уточнению оценки и лучше отражает поведение истинного остатка интегрирования.

С ростом размерности эффективность метода несколько падает, и выигрыш уже не выглядит настолько значительным. Так, на иллюстрациях 2 и 3 представлены случаи $d = 5$ и $d = 10$. Однако становится заметным другое преимущество предложенного метода: оценка более устойчива при даже небольших s по сравнению с доверительным интервалом. Максимальное увеличение параметра s дает больший выигрыш, однако оценка приобретает ступенчатый вид, что не всегда желательно. Такой вид оценки может быть полезен, если нас устраивает оценивание всего в нескольких контрольных точках.

Литература

1. *Соболь И. М.* О распределении точек в кубе и приближенном вычислении интегралов // Журнал вычисл. матем. и матем. физ. 1967. 7(4). С. 784–802.
2. *Соболь И. М.* Многомерные квадратурные формулы и функции Хаара. М.: Наука, 1969.
3. *Ермаков С. М.* Метод Монте-Карло и смежные вопросы. М.: Наука, 1975.
4. *Halton J. H.* Quasi-probability — why quasi-monte-carlo methods are statistically valid and how their errors can be estimated statistically // Monte Carlo Methods & Applications. 2005. 11. P. 203–350.
5. *Kuipers L., Niederreiter H.* Uniform Distribution Of Sequences. New York: Wiley, 1974.
6. *L'Ecuyer P., Munger D., Tuffin B.* On the distribution of integration error by randomly-shifted lattice rules // Electronic Journal of Statistics. 2010. 4. P. 950–993.
7. *Niederreiter H.* Random Number Generation and Quasi-Monte Carlo Methods. Dover: Dover Publications, 2006.

Статья поступила в редакцию 24 октября 2013 г.

Сведения об авторах

Антонов Антон Александрович — аспирант; anton.antonov@math.spbu.ru

Ермаков Сергей Михайлович — доктор физико-математических наук, профессор; sergej.ermakov@gmail.com

EMPIRICAL CONVERGENCE BOUNDS FOR QUASI-MONTE CARLO INTEGRATION

Anton A. Antonov, Sergej M. Ermakov

St.Petersburg State University, Universitetskaya nab., 7/9, St.Petersburg, 199034, Russian Federation; anton.antonov@math.spbu.ru, sergej.ermakov@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 well-known Koksma–Hlawka inequality includes a constant integrand variation, estimation of which is a more complex problem. Since quasi Monte-Carlo method estimates the integral with the mean integrand value, we can expect that the error distribution (in a probabilistic sense) converges to normal. An additional difficulty, however, is that quasi-random sequences, being treated as randomly generated, are statistically dependent, which complicates the numerical estimation of second moments.

Authors propose a new approach to the problem of estimating second order moments, based on new results, derived from a theory of random cubature formulas. Presented numerical examples indicate the superiority of the discussed error estimation method over classical ones. Refs 0. Figs 0. Tables 0.

Keywords: Monte Carlo and Quasi-Monte Carlo method, confidence interval, random cubature formulas, Haar functions, Sobol sequences.