



---

Soviet-era science, translated into English

# Reports of the Academy of Sciences of the USSR

1957

SovietRxiv

---

View the original and related papers at <https://sovietrxiv.org/items/ru-195701.99702>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

**Abstract**

**Full Text**

## **Reports of the Academy of Sciences of the USSR**

1957. Volume 114, No. 4

**MATHEMATICS**

**I. M. SOBOL**

### **MULTIDIMENSIONAL INTEGRALS AND THE MONTE CARLO METHOD**

*(Presented by Academician M. V. Keldysh on 25 XII 1956)*

This paper investigates the error of the simplest multidimensional integration formula (1). The computation of integrals by the Monte Carlo method can be reduced to this same formula with random integration nodes (1, 3).

In § 4 an error estimate is obtained for arbitrary integration nodes (formula (8)). This formula makes it possible to explain why uniform grids in multidimensional spaces give low accuracy—lower than the accuracy of the Monte Carlo method.

#### **§ 1. Integration formula.**

**1,1.** Suppose that the function  $f(P)$ , where  $P = (x_1, \dots, x_d)$ , is holomorphic in a ball containing the unit cube  $K$  of  $d$ -dimensional real space:  $0 < x_s < 1$  ( $s = 1, 2, \dots, d$ ). We shall compute the integral  $I = \int_K f(P) dV$  by the simplest formula of arithmetic means:

$$I = \frac{1}{N} \sum_{\mu=1}^N f(P_\mu) + \Delta_N, \quad (1)$$

where  $P_1, P_2, \dots, P_N$  are the integration nodes;  $\Delta_N$  is the error\*.

**1,2.** Let  $\{P_\mu\}$  be an arbitrary sequence of points in  $K$ . Denote by  $S_N(\Pi)$  the number of points of the sequence with indices  $1 \leq \mu \leq N$  that lie in the parallelepiped  $\Pi$ .

**Definition.** The sequence  $\{P_\mu\}$  is called **uniformly distributed** in  $K$  if for every  $\Pi \subseteq K$

$$\lim_{N \rightarrow \infty} \frac{S_N(\Pi)}{N} = |\Pi|$$

(here  $|\Pi|$  is the volume of  $\Pi$ ).

**Theorem** (1). For all Riemann-integrable functions

$$\lim_{N \rightarrow \infty} \Delta_N = 0 \quad (2)$$

if and only if  $\{P_\mu\}$  is uniformly distributed in  $K$ .

This theorem resolves the question of convergence of formula (1).

**1,3.** Suppose that  $Z$  is a random point uniformly distributed in  $K$  (i.e. the distribution density  $p(Z) = 1$  if  $Z \in K$ , and  $p(Z) = 0$  if  $Z \notin K$ ). The mathematical expectation is

$$E\{f(Z)\} = \int f(P) p(P) dV = \int_K f(P) dV = I. \quad (3)$$

\* Formula (1), despite its crudeness, has a number of advantages from the standpoint of computation on high-speed computing machines: all points are computed identically; there are no “weights” loading the internal memory; convergence is guaranteed for a very broad class of functions (see 1, 2).

Therefore, to compute  $I$  we can apply the Monte Carlo method: if  $P_1, P_2, \dots$  are values of  $Z$ , then for large  $N$  the arithmetic mean

$$\frac{1}{N} \sum_{\mu=1}^N f(P_\mu) \approx I.$$

Thus, the Monte Carlo method also leads to formula (1), but with random nodes\*.

## § 2. Uniform grids and the Monte Carlo method.

**2.1.** Let us take a uniform grid of  $N = n^d$  points  $P_\mu$  with coordinates

$$x_{\mu s} = \frac{m_{\mu s} + l_s}{n} \quad (s = 1, 2, \dots, d).$$

Here  $m_{\mu s}$  are natural numbers  $0, 1, 2, \dots, n - 1$ ;  $0 < l_s \leq 1$ .

**Theorem.** If  $l_s \neq 1/2$ , then

$$\Delta_N = A_1 N^{-1/d} + O(N^{-2/d}); \quad (4)$$

whereas if  $l_1 = l_2 = \dots = l_d = 1/2$ , then

$$\Delta_N = A_2 N^{-2/d} + O(N^{-4/d}). \quad (5)$$

The constants  $A_1$  and  $A_2$  depend on  $f(x_1, \dots, x_d)$  and on  $l_s$ .

It is easy to verify that for  $f = x_s$  in formula (4)  $\Delta_N = (1/2 - l_s)N^{-1/d}$ , and for  $f = x_s^2$  in formula (5)  $\Delta_N = 1/12N^{-2/d}$ . Thus, the estimates (4) and (5) are sharp.

**2.2.** For the Monte Carlo method 1.3 one can obtain a probabilistic estimate of the error from the classical limit theorem (2): with probability greater than 0.99,

$$|\Delta_N| \leq 3\sqrt{D} N^{-1/2}. \quad (6)$$

(Here  $D = D\{f(Z)\}$  is the variance.) By increasing the constant in (6), one can increase the probability of the estimate.

**2.3.** Comparing (6) with (4) or (5), we arrive at the well-known paradoxical conclusion: for  $d \gg 1$ ,  $N$  random integration nodes with uniform distribution laws give a better result than a uniform grid with the same number of nodes.

Attempts to explain this paradox can be found in the survey (3). A new explanation will be given in 6.1.

### § 3. Definition of the function $\varphi_q(N)$ .

For simplicity, in this section we assume  $d = 3$ .

**3.1.** We construct a sequence of dyadic-rational intervals  $\{\Delta_k\}$ . They are defined by groups. The zero group ( $m = 0$ ) consists of one interval  $\Delta_1 = (0, 1]$ . In group number  $m$  ( $m = 1, 2, \dots$ ) there are altogether  $2^{m-1}$  intervals  $\Delta_{mj}$  ( $j = 1, 2, \dots, 2^{m-1}$ ), which are obtained by dividing  $(0, 1]$  into  $2^{m-1}$  equal parts. Thus,

$$\Delta_{mj} = \left( \frac{j-1}{2^{m-1}}, \frac{j}{2^{m-1}} \right];$$

$$|\Delta_{mj}| = \frac{1}{2^{m-1}}.$$

The intervals are numbered so that for  $m > 0$ ,  $k = 2^{m-1} + j$ .

**3.2.** Consider the parallelepiped

$$\Pi_{k_1 k_2 k_3} = \Delta_{k_1} \times \Delta_{k_2} \times \Delta_{k_3}.$$

Transfer

\* From theorem 1.2 it follows that, when computing integrals by the Monte Carlo method, one may, instead of random points  $P_\mu$ , take any nonrandom uniformly distributed sequence.

The use of “true” random numbers in high-speed computing machines entails great technical difficulties. In practice, “pseudorandom” numbers are always used (see, for example, (3)).

the origin of coordinates to the center  $C$  of the parallelepiped:  $\xi_s = x_s - x_{C_s}$ . The new coordinate planes  $\xi_s = 0$  divide  $\Pi_{k_1 k_2 k_3}$  into  $2^3 = 8$  equal parallelepipeds.

If all  $k_s > 1$ , then by  $V_{k_1 k_2 k_3}^+$  we denote the sum of those parallelepipeds of the subdivision in which  $\xi_1 \xi_2 \xi_3 > 0$ . If one of the  $k_s$ , for example  $k_1 = 1$ , then  $V_{1 k_1 k_3}^+$  is the sum of the parallelepipeds in which  $\xi_2 \xi_3 > 0$ . Similarly,  $V_{k_1 1 1}^+$  is the parallelepiped  $\xi_1 > 0$ .

Let  $V_{k_1 k_2 k_3}^- = \Pi_{k_1 k_2 k_3} - V_{k_1 k_2 k_3}^+$ . Obviously,  $|V^+| = |V^-|$ .

**3.3.** Fix an arbitrary natural number  $q \geq 2$ . Let  $\varepsilon = 1/q$ . A triple of numbers  $(m_1, m_2, m_3)$  determines a subdivision  $K$  into equal parallelepipeds  $\Pi_{k_1 k_2 k_3}$ , with all possible  $j_1, j_2, j_3$ .

**Definition.**

$$\varphi_q(N) = \sup_{(m_1 m_2 m_3)} \left\{ \sum_{j_1, j_2, j_3} |S_N(V_{k_1 k_2 k_3}^+) - S_N(V_{k_1 k_2 k_3}^-)|^q \right\}^{1/q}. \quad (7)$$

(The sum is over all parallelepipeds of the given subdivision; the least upper bound is over all subdivisions. The case  $m_1 = m_2 = m_3 = 0$  is excluded. For the definition of  $S_N(V)$ , see 1.2.)

Obviously, the value  $\varphi_q(N)$  depends on the arrangement of the points  $P_1, P_2, \dots, P_N$  (see also 4.2).

#### § 4. Estimate of $\Delta_N$ for arbitrary nodes of integration.

**4.1.** With the aid of the expansion of  $f(x_1, \dots, x_d)$  into a Haar series (4), the following theorem is proved.

**Theorem.** *Whatever the points  $P_1, \dots, P_N$ , for the error of formula (1) the estimate*

$$|\Delta_N| \leq A_q \frac{\varphi_q(N)}{N}, \quad (8)$$

*is valid, where the constant  $A_q$  depends only on  $f(x_1, \dots, x_d)$ .*

**4.2.** In fact, the least upper bound in formula (7) is taken over a finite number of subdivisions, since for a sufficiently fine subdivision each cell will contain no more than one point; all still finer subdivisions will give one and the same result:

$$\sum |S_N(V^+) - S_N(V^-)|^q = N.$$

If  $k_s = 1$ , then the subdivision is continued until the points lying in one plane  $x_s = \text{const}$  become isolated.

Hence the **stability** of estimate (8) follows at once: if the nodes  $P_\mu$  are changed slightly (so that they do not leave the cells of the smallest essential subdivision), then the value of  $\varphi_q(N)$  will not change and estimate (8) will be preserved.

**4.3.** It was not possible to pass in (8) to the simpler function

$$\varphi_\infty(N) = \lim_{q \rightarrow \infty} \varphi_q(N) :$$

for  $q = \infty$  the series majorizing  $A_q$  became divergent.

**4.4.** The requirement that  $f(x_1, \dots, x_d)$  be holomorphic is probably too strong. In the one-dimensional case estimate (8) is valid for all continuous functions satisfying a Lipschitz condition. For this class of functions it has been proved <sup>(5)</sup> that the best one-dimensional quadrature formula gives the error  $\Delta_N = O(N^{-1})$ .

## § 5. Some properties of $\varphi_q(N)$ .

These properties are proved directly from Definition 3.3.

**5.1. Theorem.** For any  $P_1, P_2, \dots, P_N$

$$N^\varepsilon \leq \varphi_q(N) \leq N. \quad (9)$$

The upper bound (9) is sharp; the lower one is attained for  $d = 1$  (see 6.2).

One may expect that, if  $d > 1$ , then in the exact lower bound for  $N^\varepsilon$  there should be the coefficient  $2^{(d-D)(1-\varepsilon)}$  (cf. 6.3).

**5.2. Theorem.** If  $M$  nodes lie in the plane  $x_s = \text{const}$ , then  $\varphi_q(N) \geq M$ .

For example, for the uniform grid (2.1) it is easy to compute

$$\varphi_q(N) = N^{1 - \frac{1-\varepsilon}{d}}. \quad (10)$$

Comparison of (10) with (9) shows that, for  $d \gg 1$ , uniform grids are bad grids.

**5.3. Theorem.** An infinite sequence  $\{P_\mu\}$  is uniformly distributed in  $K$  (see 1.2) if and only if

$$\varphi_q(N) = o(N) \quad \text{as } N \rightarrow \infty.$$

This result establishes a connection between Theorems 1.2 and 4.1.

## § 6. Consequences and examples.

**6.1.** From the standpoint of what has been set forth, it is easy to understand the paradox 2.3: the orders of the estimates (4) and (5) are low because uniform grids are bad. If the nodes are shifted, these grids can be improved, and random grids are always “shifted.”

The order  $N^{-1/2}$  of estimate (6) is not the best: it is average. Greater accuracy can be achieved by using nonrandom and at the same time nonuniform grids.

**6.2. Example.**  $d = 1$ ;  $N$  arbitrary. A sequence  $\{p_i\}$  for which  $\varphi_q(N) = N^\varepsilon$  can be constructed from dyadic-rational points:

$$1, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \frac{9}{16}, \frac{5}{16}, \frac{13}{16}, \frac{3}{16}, \frac{11}{16}, \dots$$

**6.3. Example.**  $d = 2$ ;  $N = n^2$ , where  $n = 2^{n_1}$ . Let  $\alpha_i = np_i$  ( $p_i$  from 6.2). We define the coordinates of the node  $P_{ij}$  in the  $x, y$ -plane by the formulas

$$x_{ij} = \frac{i-1}{n} + \frac{1}{n^2} \left[ (i + \alpha_j) \bmod n + \frac{1}{2} \right],$$

$$y_{ij} = \frac{j-1}{n} + \frac{1}{n^2} \left[ (j + \alpha_i) \bmod n + \frac{1}{2} \right]. \quad (i, j = 1, \dots, n)$$

For this system of points  $\varphi_q(N) = 2^{1-\varepsilon} N^\varepsilon$ .

Department of Applied Mathematics  
V. A. Steklov Mathematical Institute  
Academy of Sciences of the USSR

Received  
18 XII 1956

## References Cited

1. H. Weyl, *Math. Ann.*, **77**, 3, 313 (1916).
2. J. H. Curtiss, *J. Math. Phys.*, **32**, 4, 209 (1954).
3. K. D. Tocher, *J. Roy. Stat. Soc.*, ser. B, **16**, 1, 39 (1954).
4. S. Kaczmarz, H. Steinhaus, *Theorie der Orthogonalreihen*, 1935, p. 120.

5. A. Kh. Turetskii, *Uspekhi Mat. Nauk*, **6**, 5(45), 166 (1951).

*Note: Figure translations are in progress. See original paper for figures.*

*Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.*