JOURNAL OF NUMBER THEORY 39, 367–373 (1991)

A New Measure of Irregularity of Distribution

ILYA M. SOBOL' AND OLEG V. NUZHDIN

Institute of Applied Mathematics, USSR Academy of Sciences, 4, Miusskaya Square, Moscow 125047, USSR

Communicated by E. Hlawka

Received May 15, 1990

We introduce a new measure of irregularity of distribution—the range, ψ, that is similar to the nonuniformity φ_∞ but much easier to compute. It is shown that for P_τ-nets and for initial segments of LP_τ-sequences $\psi \leqslant 2^\tau$ and as the number of points increases this is the lowest possible order of magnitude $\psi = O(1)$.

1. DEFINITIONS

The unit interval $[0, 1]$ is denoted by I so that I^n is the n-dimensional unit cube. Subintervals $[(j-1)\,2^{-m}, j2^{-m})$ are called dyadic intervals; here j and m are integers, $1 \leqslant j \leqslant 2^m$, $m \geqslant 0$. At $j = 2^m$ the dyadic interval is closed by definition. So a fixed integer m defines a partition of I into a sum of 2^m equal dyadic intervals.

A dyadic box (parallelepiped) Π is the Cartesian product of dyadic intervals. A set of integers $M = (m_1, ..., m_n) \neq (0, ..., 0)$, $m = m_1 + \cdots + m_n$, defines a partition of I^n into a sum of equal dyadic boxes Π_α whose volume is 2^{-m}, $1 \leqslant \alpha \leqslant 2^m$.

Given a set of points $x_1, ..., x_N \in I^n$ and a subset $G \subset I^n$, we introduce the counting function $S_N(G)$ as the number of points $x_i \in G$ while $1 \leqslant i \leqslant N$.

Now consider a fixed set of points $x_1, ..., x_N \in I^n$. For an arbitrary partition M of I^n denote

$$\psi_M = \max_\alpha S_N(\Pi_\alpha) - \min_\alpha S_N(\Pi_\alpha) \tag{1}$$

and define the range of the set as

$$\psi = \psi(x_1, ..., x_N) = \sup_M \psi_M, \tag{2}$$

where the supremum is extended over all such partitions of I^n.

0022-314X/91 $3.00

641/39/3-9

To list all these partitions one has to consider all integer solutions $m_1, ..., m_n$ of the equation

$$m_1 + \cdots + m_n = m$$

at $m = 1, 2, 3, \ldots$. In fact, only a finite number of partitions must be taken into consideration.

Indeed, if the projections of $x_1, ..., x_N$ onto each coordinate axis are distinct then for all partitions with sufficiently large m both $\max_\alpha S_N(\Pi_\alpha) = 1$ and $\min_\alpha S_N(\Pi_\alpha) = 0$. Clearly, $\psi \geqslant 1$.

If there are k points having identical groups of Cartesian coordinates one may easily find a partition with $\max_\alpha S_N(\Pi_\alpha) = k$, $\min_\alpha S_N(\Pi_\alpha) = 0$. Clearly, $\psi \geqslant k$, and more detailed partitions can be ignored.

In general,

$$1 \leqslant \psi(x_1, ..., x_N) \leqslant N. \tag{3}$$

EXAMPLE. For a rectangular lattice containing $N = c^n$ points (c an integer) the range is

$$\psi = c^{n-1} = N^{1-1/n};$$

maximal values $\psi_M = c^{n-1}$ correspond, e.g., to all partitions $M = (m, 0, 0, ..., 0)$ with sufficiently large m. In the case where $n = 1$ the range is the best possible, but for $n > 1$ the ranges of rectangular lattices $\psi \to \infty$ as $N \to \infty$.

2. THE RANGE—A MEASURE OF IRREGULARITY OF DISTRIBUTION

THEOREM. 1. *Let $x_1, x_2, \ldots$ be an arbitrary infinite sequence of points in I^n. The sequence is uniformly distributed (ud) if and only if*

$$\lim_{N \to \infty} \frac{\psi(x_1, ..., x_N)}{N} = 0. \tag{4}$$

The proof of the theorem follows immediately from the lemmas below:

LEMMA 1. *For an arbitrary set of points $x_1, ..., x_N \in I^n$*

$$\psi \leqslant 2^{n+1} D, \tag{5}$$

where D is the discrepancy of the set.

LEMMA 2. *For an arbitrary set of points* $x_1, \ldots, x_N \in I^n$

$$\varphi_\infty \leqslant 2^{n-1}\psi, \tag{6}$$

where φ_∞ *is the nonuniformity of the set.*

Proof of Theorem 1. Each of the relations $D(x_1, \ldots, x_N)/N \to 0$ and $\varphi_\infty(x_1, \ldots, x_N)/N \to 0$ as $N \to \infty$ is a necessary and sufficient condition of ud. So (4) is implied by (5) and (6).

First, we recall the definition of discrepancy. For a given point $y \in I^n$ with Cartesian coordinates $(y^1, \ldots, y^n)$, a box $G_y = [0, y^1) \times \cdots \times [0, y^n)$ with n-dimensional volume $V(G_y) = y^1 \cdots y^n$ is introduced. The discrepancy of the points $x_1, \ldots, x_N$ is defined by

$$D = D(x_1, \ldots, x_N) = \sup_y |S_N(G_y) - NV(G_y)|,$$

where the supremum is extended over all $y \in I^n$.

Proof of Lemma 1. It follows from the definition of D that for all boxes G_y the counting functions $S_N(G_y) = NV(G_y) + h_y$ with $|h_y| \leqslant D$. A box Π is a sum of 2^n boxes G_y, where all the vertices of Π play the role of y. Therefore $S_N(\Pi) = NV(\Pi) + h'$ with $|h'| \leqslant 2^n D$.

For a fixed partition M of I^n the volumes of all dyadic boxes Π_α are equal. So,

$$\max_\alpha S_N(\Pi_\alpha) - \min_\alpha S_N(\Pi_\alpha) \leqslant 2 \cdot 2^n D = 2^{n+1} D.$$

And the range ψ does not exceed $2^{n+1}D$ also.

Second, we recall the definition of nonuniformity [7, 8, 4]. Let Π be an arbitrary dyadic box. If we move the origin of the coordinate system to the center of Π and denote by $\xi_1, \ldots, \xi_n$ the new coordinates, Π may be split into two parts: Π^+, in which the product $\xi_1 \cdots \xi_n > 0$, and Π^-, in which $\xi_1 \cdots \xi_n < 0$. More precisely, each of these parts is the union of 2^{n-1} "hyperquadrants" of Π that are again dyadic boxes.

Let $x_1, \ldots, x_N$ be N given points in I^n. The n-dimensional nonuniformity of these points is defined to be

$$\sup_\Pi |S_N(\Pi^+) - S_N(\Pi^-)|, \tag{7}$$

where the supremum is extended over all dyadic boxes Π.

Furthermore, we consider the projections of $x_1, \ldots, x_N$ onto various s-dimensional faces of I^n and calculate the s-dimensional nonuniformities of these projections. The largest value among all $2^n - 1$ nonuniformities $(1 \leqslant s \leqslant n)$ is called the nonuniformity of $x_1, \ldots, x_N$ and denoted by $\varphi_\infty(x_1, \ldots, x_n)$.

Proof of Lemma 2. Consider an arbitrary dyadic box $\Pi = \Pi^+ \cup \Pi^-$. All hyperquadrants of Π are members of the same partition of I^n. Therefore

$$2^{n-1} \min_\alpha S_N(\Pi_\alpha) \leqslant S_N(\Pi^+) \leqslant 2^{n-1} \max_\alpha S_N(\Pi_\alpha)$$

and the same inequalities are true for $S_N(\Pi^-)$. Hence,

$$|S_N(\Pi^+) - S_N(\Pi^-)| \leqslant 2^{n-1}[\max_\alpha S_N(\Pi_\alpha) - \min_\alpha S_N(\Pi_\alpha)] = 2^{n-1}\psi_M.$$

And the n-dimensional nonuniformity (7) does not exceed $2^{n-1}\psi$.

A remarkable point is that the s-dimensional nonuniformities do not exceed $2^{s-1}\psi$: an s-dimensional dyadic hyperquadrant may be replaced by an n-dimensional dyadic box corresponding to a partition $(m_1, ..., m_n)$ containing $n-s$ zero values of m_k. So all nonuniformities do not exceed $2^{n-1}\psi$ and (6) is true.

3. P_τ-NETS AND LP_τ-SEQUENCES

Let $0 \leqslant \tau < v$ be integers. A point set of $N = 2^v$ points in I^n is called a P_τ-net if every dyadic box Π with $V(\Pi) = 2^\tau/N$ contains exactly 2^τ points of the set.

The smaller τ is the better the uniformity of the P_τ-nets. It was shown in [7, 8] that P_0-nets exist only in I^1, I^2, and I^3. If $\tau(n)$ is the least value of τ having the property that in I^n, P_τ-nets exist for arbitrarily large N, then

$$\tau(1) = \tau(2) = \tau(3) = 0, \qquad \tau(4) = 1, \qquad \tau(5) \leqslant 3.$$

As $n \to \infty$ the value $\tau(n) < n \log_2 n + \cdots$.

THEOREM 2. *For an arbitrary P_τ-net in I^n*

$$\psi \leqslant 2^\tau. \tag{8}$$

Proof. Consider a partition M of I^n into dyadic boxes Π_α with $V(\Pi_\alpha) = 2^\tau/N$; then every Π_α contains 2^τ points and $\psi_M = 0$. The same will be true for all partitions with $V(\Pi_\alpha) > 2^\tau/N$ since every large Π_α is a sum of several boxes with volumes $2^\tau/N$. Finally, for more detailed partitions, $S_N(\Pi_\alpha) \leqslant 2^\tau$ and $\max_\alpha S_N(\Pi_\alpha) - \min_\alpha S_N(\Pi_\alpha) \leqslant 2^\tau$,

COROLLARY. *For an arbitrary P_0-net, $\psi = 1$.*

Thus we have proved that whenever P_0-nets exist they belong to the most uniform sets of points.

The following is a kind of inverse statement: if $\psi(x_1, ..., x_N) = 1$ and $N = 2^\nu$, ν an integer, then $x_1, ..., x_N$ is a P_0-net.

Indeed, consider an arbitrary partition of I^n into dyadic boxes with $V(\Pi_\alpha) = 1/N$. The requirement $\psi = 1$ implies that

$$\max_\alpha S_N(\Pi_\alpha) - \min_\alpha S_N(\Pi_\alpha) \leqslant 1.$$

But there are N points and N boxes; if one of the Π_α is empty then another must contain at least two points, which contradicts the last inequality. Thus every Π_α contains exactly one point.

Now we turn to infinite sequences $x_0, x_1, ... \in I^n$.

An initial segment of the sequence is the set of points $x_0, ..., x_{N-1}$. A binary segment is the set of points x_i with indices satisfying $(k-1)\, 2^p \leqslant i < k2^p$ for some integers $k \geqslant 1$, $p \geqslant 1$.

The sequence $x_0, x_1, ...$ is called an LP_τ-sequence if every binary segment with $p > \tau$ is a P_τ-net [7, 8].

THEOREM 3. *For an arbitrary initial segment of an LP_τ-sequence in I^n*

$$\psi(x_0, ..., x_{N-1}) \leqslant 2^\tau. \tag{9}$$

Proof. Consider an arbitrary partition M of I^n into dyadic boxes Π_α and denote the volume $V(\Pi_\alpha) = 2^{-m}$. Let $p = m + \tau$ and define k_0 by the condition $(k_0 - 1)\, 2^p \leqslant N - 1 < k_0 2^p$. Then the initial segment $0 \leqslant i \leqslant N-1$ consists of $k_0 - 1$ full binary segments that are P_τ-nets, and may be of one extra segment $(k_0 - 1)\, 2^p \leqslant i \leqslant N-1$ that is a part of a P_τ-net. Therefore $(k_0 - 1)\, 2^\tau \leqslant S_N(\Pi_\alpha) \leqslant k_0 2^\tau$. It follows from (1) that $\psi_M \leqslant 2^\tau$. Hence (9) is true.

From (3) one can easily conclude that the best asymptotic behavior of $\psi(x_0, ..., x_{N-1})$ as $N \to \infty$ is $O(1)$. Thus all LP_τ-sequences can be regarded as asymptotically best ud sequences.

For LP_0-sequences even the best numerical values $\psi(x_0, ..., x_{N-1}) = 1$ are attained but such sequences exist only in I^1 and I^2. In I^3 only LP_1-sequences exist and Theorem 3 provides the estimate $\psi \leqslant 2$.

From Theorems 2 and 3 and Lemma 2 two known statements [7, 8] follow: in I^n for arbitrary P_τ-nets and for arbitrary initial segments of LP_τ-sequences, $\varphi_\infty \leqslant 2^{n-1+\tau}$.

4. GENERALIZATIONS

Let $r > 2$ be an arbitrary integer. If one substitutes r for 2 in Section 1 then, mutatis mutandis, r-adic intervals and r-adic boxes Π_α with $V(\Pi_\alpha) = r^{-m}$ may be defined. The r-adic range may be introduced similarly, and Theorem 1 remains true.

Turning from 2 to r in Section 3 leads to definitions of r-adic P_τ-nets, r-adic segments of a sequence, and r-adic LP_τ-sequences. Here Theorems 2 and 3 can be easily generalized: in I^n for arbitrary r-adic P_τ-nets and for arbitrary initial segments of r-adic LP_τ-sequences the r-adic ranges do not exceed r^τ.

H. Faure [2] was the first to introduce r-adic P_τ-nets and LP_τ-sequences with $\tau = 0$ (in [2] they are called $P^v_{r,n}$-réseau and $P_{r,n}$-suite) that exist in I^n for sufficiently large r (in fact, for $r \geqslant n$). H. Niederreiter [5, 6] investigated the general case and, among such nets and sequences (in [5, 6] they are called (τ, v, n)-nets and (τ, n)-sequences in base r), found those having the smallest discrepancy estimates that are currently known.

Computational experiments with dyadic and r-adic LP_τ-sequences are presented in [1, 3]. Various applications of these sequences are discussed in [10].

5. Numerical Examples

A program has been written for computing ψ for a given set of N points in I^n. We assume that a maximum value m^* is prescribed and only partitions with $m \leqslant m^*$ are considered in (2). For the following examples we put $m^* = [\log_2 N] + 1$.

Let $x_0, x_1, \ldots$ be the LP_τ-sequence in I^n defined in [9] that has additional uniformity properties. Programs for generating these sequences are available in [11, 1]. We have computed values $\psi = \psi(x_0, \ldots, x_{N-1})$ for various n and N.

For the case of $n = 3$ where $\tau = 1$ we have obtained $\psi = 2$ for all $3 \leqslant N \leqslant 50$.

TABLE I

Values of $\psi(x_0, \ldots, x_{N-1})$ in I^4

N:	4	5–6	7–9	10–11	12–16	17–20	21–32	33–40
ψ	2	3	4	3	2	3	4	5

N:	41–48	49–56	57–71	72–79	80–87	88–95	96–160	161–168
ψ	6	7	8	7	6	5	4	5

N:	169–176	177–184	185–199	200–207	208–215	216–223	224–257
ψ	6	7	8	7	6	5	4

TABLE II

Values of $\psi(z_1, ..., z_N)$ in I^4

N:	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64
ψ	2	4	8	12	14	14	12	12	14	14	12	14	14	16	16	14
N:	68	72	76	80	84	88	92	96	100	104	108	112	116	120	124	128
ψ	14	16	16	16	14	16	14	18	16	18	18	18	17	16	18	20

For the case of $n=4$ where $\tau=3$ the values of ψ for $4 \leqslant N \leqslant 257$ are listed in Table I.

We have generated a sequence $z_1, z_2, ...$ of independent random points uniformly distributed in I^4 and computed values of $\psi(z_1, ..., z_N)$ that are listed in Table II for $N = 4(4)\,128$.

We have carried out several computations for random sequences and we think that at large N the rate of growth of $\psi(z_1, ..., z_N)$ is about $\sqrt{N}$.

References

1. P. Bratley and B. L. Fox, Implementing Sobol's quasirandom sequence generator, *ACM Trans. Math. Software* **14** (1988), 88–100.
2. H. Faure, Discrépance de suites associées à un système de numération (en dimension s), *Acta Arith.* **41** (1982), 337–351.
3. Yu. L. Levitan, N. I. Markovich, S. G. Rozin, and I. M. Sobol', On quasi-random sequences for numerical computations, *Zh. Vychisl. Mat. i Mat. Fiz.* **28** (1988), 755–759; *USSR Comput. Math. and Math. Phys.* **28** (1988), 88–92.
4. H. Niederreiter, Quasi-Monte Carlo methods and pseudo-random numbers, *Bull. Amer. Math. Soc.* **84** (1978), 957–1041.
5. H. Niederreiter, Point sets and sequences with small discrepancy, *Monatsh. Math.* **104** (1987), 273–337.
6. H. Niederreiter, Low-discrepancy and low-dispersion sequences, *J. Number Theory* **30** (1988), 51–70.
7. I. M. Sobol', The distribution of points in a cube and the approximate evaluation of integrals, *Zh. Vychisl. Mat. i Mat. Fiz.* **7** (1967), 784–802; *USSR Comput. Math. and Math. Phys.* **7** (1967), 86–112.
8. I. M. Sobol', "Multidimensional Quadrature Formulas and Haar Functions," Nauka, Moscow, 1969. [In Russian]
9. I. M. Sobol', Uniformly distributed sequences with an additional uniformity property, *Zh. Vychisl. Mat. i. Mat. Fiz.* **16** (1976), 1332–1337; *USSR Comput. Math. and Math. Phys.* **16** (1976), 236–242.
10. I. M. Sobol', Points which uniformly fill a multidimensional cube, "Mathematics, Cybernetics" Vol. 2, Znanie, Moscow, 1985. [In Russian]
11. I. M. Sobol' and R. B. Statnikov, "Selection of Optimal Parameters in Problems with Several Criteria," Nauka, Moscow, 1981. [In Russian]