The Mean Square Discrepancy of Randomized Nets

FRED J. HICKERNELL
Hong Kong Baptist University

One popular family of low discrepancy sets is the (t, m, s)-nets. Recently a randomization of these nets that preserves their net property has been introduced. In this article a formula for the mean square $\mathscr{L}^2$-discrepancy of $(0, m, s)$-nets in base b is derived. This formula has a computational complexity of only $O(s \log(N) + s^2)$ for large N or s, where $N = b^m$ is the number of points. Moreover, the root mean square $\mathscr{L}^2$-discrepancy of $(0, m, s)$-nets is shown to be $O(N^{-1}[\log(N)]^{(s-1)/2})$ as N tends to infinity, the same asymptotic order as the known lower bound for the $\mathscr{L}^2$-discrepancy of an arbitrary set.

Categories and Subject Descriptors: G.1.4 [**Numerical Analysis**]: Quadrature and Numerical Differentiation—*error analysis; multiple quadrature*; G.3 [**Probability and Statistics**]: Probabilistic Algorithms (including Monte Carlo); I.6.8 [**Simulation and Modeling**]: Types of Simulation—*Monte Carlo*

General Terms: Algorithms, Performance, Theory

Additional Key Words and Phrases: Multidimensional integration, number-theoretic nets and sequences, quadrature, quasi-Monte Carlo methods, quasi-random sets

1. INTRODUCTION

Multidimensional integrals over the s-dimensional unit cube $C^s = [0, 1)^s$ may be approximated by the sample mean of the integrand evaluated on a point set, P, with N points. (Here, in contrast to ordinary sets, P may have multiple copies of the same point [Niederreiter 1992, p. 14].) The quadrature error depends on how uniformly the points in P are distributed on the unit cube and on how much the integrand varies from a constant. For example, if $D(P)$ is the $\mathscr{L}^\infty$-star discrepancy [Niederreiter 1992, Definition 2.1], and $V(f)$ is the variation of f on $\bar{C}^s = [0, 1]^s$ in the sense of Hardy and Krause, then the Koksma–Hlawka inequality [Niederreiter 1992,

This research was partially supported by a Hong Kong RGC grant 94-95/38 and HKBU FRG grant 95-96/II-01.
Author's address: Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, email: fred@hkbu.edu.hk, URL: http://www.math.hkbu.edu.hk/~fred.

Theorem 2.11] is

$$\mathrm{Err}(f) \equiv \left| \int_{C^s} f(x)dx - \frac{1}{N} \sum_{z \in P} f(z) \right| \leq D(P)V(f). \tag{1.1}$$

Error bounds of this form with other definitions of $D(P)$ and $V(f)$ appear in the literature as well. A good set P for quadrature is one that has a small discrepancy, $D(P)$. Several deterministic sets have been found that have smaller discrepancies than simple random points. These sets are often called quasi-random points.

Calculating the $\mathcal{L}^\infty$-star discrepancy of a particular set is impractical unless both N and s are small because it requires $O(N^s)$ operations. Much effort has been directed towards finding quasi-random sequences that have asymptotically small $\mathcal{L}^\infty$-star discrepancies as N tends to infinity. The (t, m, s)-nets [Niederreiter 1992, chap. 4] are one example. By comparison, the $\mathcal{L}^2$-star discrepancy is easier to compute, requiring only $O(N^2)$ operations using a naive algorithm, or $O(N[\log(N)]^s)$ operations using the algorithm of Heinrich [1996].

Owen [1995; 1997a; 1997b] has proposed a randomization of (t, m, s)-nets that preserves their net properties. His aim is to obtain probabilistic quadrature error estimates in a similar manner as one would for simple Monte Carlo quadrature. In this article, a formula for the mean square $\mathcal{L}^2$-discrepancy of randomized $(0, m, s)$-nets is derived that requires only $O(s \log(N) + s^2)$ mathematical operations to evaluate as N and/or s tend to infinity. The $\mathcal{L}^2$-discrepancy for these randomized nets is shown to decay like $O(N^{-1}[\log(N)]^{(s-1)/2})$ for large N—a result matching the asymptotic lower bound obtained by Roth [1954].

In the remainder of this section we define (t, m, s)-nets and describe Owen's randomization. Also, a generalized $\mathcal{L}^2$-discrepancy that arises in quadrature error bounds, recently derived by the author, is defined. In Section 2 the mean square $\mathcal{L}^2$-discrepancy is computed for simple random samples and randomized $(0, m, s)$-nets. The asymptotic behavior of the $\mathcal{L}^2$-discrepancy is studied in Section 3. This article concludes with some discussion.

Any point $z = (z_1, \ldots, z_s) \in C^s$ may be written *in base b* as

$$z = (0.z_{11}z_{21}z_{31} \ldots, 0.z_{12}z_{22}z_{32} \ldots, \ldots\ldots, 0.z_{1s}z_{2s}z_{3s} \ldots),$$

where the b-nary digits z_{ij} range from 0 to $b - 1$. Let $\mathbf{Z}_+^s$ denote the space of s-dimensional non-negative integer vectors. For any $k = (k_1, \ldots, k_s) \in \mathbf{Z}_+^s$, let $\sigma(k) = k_1 + \cdots + k_s$. There are $b^{k_1} \cdots b^{k_s} = b^{\sigma(k)}$ different ways to choose the first k b-nary digits of a point z:

$$z_{11}, \ldots, z_{k_1 1}, z_{12}, \ldots, z_{k_2 2}, \ldots, z_{1s}, \ldots, z_{k_s s}. \tag{1.2}$$

(If $k_j = 0$, then no digit z_{ij} is being specified.) A (t, m, s)-net contains at least one point with every possible choice of the first k digits, provided that $\sigma(k)$ is small enough.

Definition 1.1. Let t and m be non-negative integers with $0 \le t \le m$, and let s be a positive integer. A (t, m, s)-net in base b is a set, P, containing $N = b^m$ points in C^s. For any possible choice of the first k b-nary digits (1.2) there exist $b^{m-\sigma(k)}$ points in P with these digits provided that $\sigma(k) \le m - t$.

A smaller value of t tends to imply a net with better uniformity (smaller discrepancy). Any (t_1, m, s)-net is also a (t_2, m, s)-net for $t_1 < t_2$, and any set is an (m, m, s)-net. (For a fuller discussion of (t, m, s)-nets, see Niederreiter [1992, chap. 4].)

Owen randomizes the digits of the points in a given (t, m, s)-net P_0 to obtain a new (t, m, s)-net P. For every digit index $i = 1, 2, \ldots,$ every coordinate index $j = 1, \ldots, s$, and every $y \in P_0$, one obtains a random digit z_{ij} that contributes to a random point $z \in P$. The randomized net, P, satisfies the following assumptions:

Assumption 1.2

a. For any $z \in P$ each digit z_{ij} is uniformly distributed on the set $\{0, \ldots, b - 1\}$.
b. For any two points $z, z' \in P$ the random vectors $(z_1, z'_1), \ldots, (z_s, z'_s)$ are *mutually independent*.
c. For any two points $y, y' \in P_0$ let $z, z' \in P$ be the corresponding points in the randomized net. Suppose that y_j and y'_j share the same first k_j digits, but that their $k_j + 1$st digits are different. Then
   i. $z_{ij} = z'_{ij}$ for $i = 1, \ldots, k_j$,
   ii. the random vector $(z_{k_j+1,j}, z'_{k_j+1,j})$ is uniformly distributed on the set $\{(n, n') : n \ne n'; n, n' = 0, \ldots, b - 1\}$, and
   iii. $z_{k_j+2,j}, z_{k_j+3,j}, \ldots z'_{k_j+2,j}, z'_{k_j+3,j}, \ldots$ are *mutually independent*.

Assumptions 1a and 1b imply that the marginal probability distribution of any $z = (z_1, \ldots, z_s) \in P$ is uniform on C^s and that the $z_1, \ldots, z_s$ are mutually independent. Assumption 1c maintains the correlation between different points in P that is necessary for retaining its net properties and thus a low discrepancy. For a simple random sample, Assumptions 1a and 1b are also satisfied, but instead of Assumption 1c one has $z_{1j}, z_{2j}, \ldots, z'_{1j}, z'_{2j}, \ldots$ mutually independent for $z \ne z'$.

Let $S = \{1, \ldots, s\}$ be the set of coordinate indices. For any $u \subseteq S$ let $|u|$ denote the number of points in u. Let $C^u = [0, 1)^u$ denote the $|u|$-dimensional unit cube involving the coordinates in u, and likewise, let $\mathbf{Z}_+^u$ be the $|u|$-dimensional non-negative integer vectors. This notation allows us to distinguish spaces of the same dimension in different coordinate directions.

The $\mathscr{L}^p$-star discrepancy is sometimes defined as the $\mathscr{L}^p$-norm of the difference between the empirical distribution associated with the sample P and the uniform distribution on the unit cube:

$$D^*_{p,S}(P) \equiv \left\| \frac{|P \cap [0, x)|}{N} - \mathrm{Vol}([0, x)) \right\|_p, \tag{1.3}$$

where $|\ |$ denotes the number of points in a set, and $\|\ \|_p$ denotes the $\mathcal{L}^p$-norm. Although this definition is appropriate for $p = \infty$, it does not admit an error bound like (1.1) for other values of p. A more suitable definition [Hickernell to appear] is

$$D^*_p(P) \equiv \left\{ \sum_{\emptyset \subset u \subseteq S} \left\| \frac{|P_u \cap [0, x_u)|}{N} - \mathrm{Vol}([0, x_u)) \right\|_p^p \right\}^{1/p} \quad (1 \le p < \infty),$$

$$D^*_\infty(P) \equiv \max_{\emptyset \subset u \subseteq S} \left\| \frac{|P_u \cap [0, x_u)|}{N} - \mathrm{Vol}([0, x_u)) \right\|_\infty$$

$$= \left\| \frac{|P \cap [0, x)|}{N} - \mathrm{Vol}([0, x)) \right\|_\infty,$$

where P_u denotes the projection of the sample P into the cube C^u. This means that $D^*_{p,S}(P)$ is only one term in the definition of $D^*_p(P)$ for $p < \infty$. An error bound of the form (1.1) based on this definition was derived by Zaremba [1968] for $p = 2$ and Sobol' [1969, Ch. 8] for all p. The $\mathcal{L}^p$-star discrepancy is in general difficult to compute, except in the case $p = 2$ where it reduces to a double sum:

$$[D^*_2(P)]^2 = \left(\frac{4}{3}\right)^s - \frac{2}{N} \sum_{z \in P} \prod_{j=1}^{s} \left(\frac{3 - z_j^2}{2}\right) + \frac{1}{N^2} \sum_{z, z' \in P} \prod_{j=1}^{s} [2 - \max(z_j, z'_j)].$$

Hickernell [to appear] has derived a family of quadrature error bounds of form (1.1) and $\mathcal{L}^p$-discrepancies, $D_p(P)$, that include the star discrepancy as a special case. The $\mathcal{L}^2$-discrepancy is the simplest to compute. Let β be an arbitrary positive constant and μ be an arbitrary function on [0, 1) whose first derivative is essentially bounded and that satisfies $\int_0^1 \mu(x_j) dx_j = 0$. Furthermore, let

$$\bar{M} = \int_0^1 \left(\frac{d\mu}{dx}\right)^2 dx, \quad M = 1 + \beta^2 \bar{M}.$$

The generalized $\mathcal{L}^2$-discrepancy defined in Hickernell [to appear] is

$$[D_2(P)]^2 = M^s - \frac{2}{N} \sum_{z \in P} \prod_{j=1}^{s} [M + \beta^2 \mu(z_j)] + \frac{1}{N^2} \sum_{z,z' \in P} \prod_{j=1}^{s}$$

$$\cdot \left\{ M + \beta^2 \left[\mu(z_j) + \mu(z'_j) + \frac{1}{2} B_2(z_j) + \frac{1}{2} B_2(z'_j) + \frac{1}{6} - \frac{1}{2} \left| z_j - z'_j \right| \right] \right\}, \quad (1.4)$$

where B_2 denotes the quadratic Bernoulli polynomial [Abramowitz and Stegun 1964, chap. 23]. The star discrepancy is obtained by choosing

$$\mu(x) = \frac{1}{6} - \frac{x^2}{2}, \quad \beta = 1, \quad \bar{M} = \frac{1}{3}, \quad M = \frac{4}{3}. \quad (1.5)$$

Another choice of μ and β yields a discrepancy derived by Hickernell [1996] that is similar to the figure of merit, P_α, used in the study of lattice rules [Sloan and Joe 1994, Eq. (4.8)]. Several different choices of μ and β are discussed by Hickernell [to appear].

The $\mathcal{L}^2$-discrepancy defined in (1.4) can also be written as

$$[D_2(P)]^2 = \sum_{\emptyset \subset u \subseteq S} \beta^{2|u|} [D_{2,u}(P)]^2 \quad (1.6)$$

where

$$[D_{2,u}(P)]^2 = \bar{M}^{|u|} - \frac{2}{N} \sum_{z \in P} \prod_{j \in u} [\bar{M} + \mu(z_j)]$$

$$+ \frac{1}{N^2} \sum_{z,z' \in P} \prod_{j \in u} \left[\bar{M} + \mu(z_j) + \mu(z'_j) + \frac{1}{2} B_2(z_j) + \frac{1}{2} B_2(z'_j) + \frac{1}{6} - \frac{1}{2} \left| z_j - z'_j \right| \right]. \quad (1.7)$$

Choosing μ and β according to (1.5) and setting $u = S$ in the above equation yields a formula for $D^*_{2,S}$ of (1.3) originally derived by Warnock [1972]:

$$[D^*_{2,S}(P)]^2 = \left(\frac{1}{3}\right)^s - \frac{2}{N} \sum_{z \in P} \prod_{j=1}^{s} \left(\frac{1 - z_j^2}{2} \right) + \frac{1}{N^2} \sum_{z,z' \in P} \prod_{j=1}^{s} [1 - \max(z_j, z'_j),].$$

Not only does the $\mathcal{L}^2$-discrepancy appear in worst-case quadrature error bounds, such as (1.1), it also arises in average-case quadrature error analysis. Let E_f denote the expected value over some space of integrands. Then

$$E_f[\mathrm{Err}(f)]^2 = [D_2(P)]^2,$$

where the choice of the space of integrands determines the specific form of the discrepancy [Sacks and Ylvisacker 1970; Ritter 1995]. The case of the star discrepancy has been studied by Woźniakowski [1991] and Morokoff and Caflisch [1994].

2. COMPUTING THE MEAN SQUARE $\mathscr{L}^2$-DISCREPANCY

Let E denote the expected value over a space of random sets P. In this section we compute $E\{[D_2(P)]^2\}$ for simple random samples and for randomized $(0, m, s)$-nets. Both kinds of samples satisfy Assumptions 1a and 1b. This allows a simple treatment of terms in (1.4) involving only z_j or only z'_j. The difficult term is $|z_j - z'_j|$ since its expected value depends on the correlation of z and z'.

LEMMA 2.1. *If P is a random set satisfying Assumptions* 1a *and* 1b, *then*

$$E\{[D_2(P)]^2\}$$
$$= \left(M + \frac{\beta^2}{6}\right)^s \sum_{\emptyset \subset u \subseteq S} \left(\frac{-\beta^2}{6M + \beta^2}\right)^{|u|} \left\{-1 + \frac{1}{N^2} \sum_{z,z' \in P} \prod_{j \in u} [3E|z_j - z'_j|]\right\}. \quad (2.1)$$

PROOF. By Assumption 1.2(a), each z_j is uniformly distributed on $[0, 1)$, so

$$E[B_2(z_j)] = \int_0^1 B_2(x_j)dx_j = 0, \quad E[\mu(z_j)] = \int_0^1 \mu(x_j)dx_j = 0.$$

By Assumption 1.2(b), the components $(z_1, z'_1), \ldots, (z_s, z'_s)$ are mutually independent so that the expectation of the product over j is the product of the expectations. Combining these two results implies that (1.4) can be written as

$$E\{[D_2(P)]^2\} = -M^s + \frac{1}{N^2} \sum_{z,z' \in P} \prod_{j=1}^{s} \left\{\left(M + \frac{\beta^2}{6}\right) - \frac{\beta^2}{2} E\left|z_j - z'_j\right|\right\}.$$

Using the binomial theorem to rewrite the product completes the proof of this lemma. □

To simplify (2.1) further one must calculate $3E|z_j - z'_j|$. In the case where P is a simple random sample, it is straightforward to show that

$$3E|z_j - z_j'| = \begin{cases} 0 & z = z', \\ 1, & z \neq z'. \end{cases} \tag{2.2}$$

Substituting this expression into (2.1) calculation leads to

$$E\{[D_2(P)]^2\} = \left(M + \frac{\beta^2}{6}\right)^s \sum_{\emptyset \subset u \subseteq S} \left(\frac{-\beta^2}{6M + \beta^2}\right)^{|u|} \left\{-1 + \frac{N^2 - N}{N^2}\right\},$$

which after further simplification gives the following theorem:

THEOREM 2.2. *If P is a simple random sample, then*

$$E\{[D_2(P)]^2\} = \frac{1}{N}\left[\left(M + \frac{\beta^2}{6}\right)^s - M^s\right]. \tag{2.3}$$

This formula serves as a benchmark for other (presumably superior) quasi-random sets, P. Since the mean square generalized $\mathscr{L}^2$-discrepancy is $O(N^{-1})$, the generalized $\mathscr{L}^2$-discrepancy itself is typically $O(N^{-1/2})$ for a simple random sample.

The mean square discrepancy of a randomized $(0, m, s)$-net is given by the following theorem. Two major steps in the calculation are contained in Lemmas 2.4 and 2.5 below.

THEOREM 2.3. *Let P be a $(0, m, s)$-net randomized according to Assumption 1.2. Let $R(l, \tau)$ be defined as the partial binomial sum:*

$$R(l, \tau) \equiv (1 - b^{-1})^{1-l} \sum_{r=0}^{\tau-1} \binom{l-1}{r} (-b^{-1})^r. \tag{2.4}$$

The mean square discrepancy of P is

$$E\{[D_2(P)]^2\} = \left(M + \frac{\beta^2}{6}\right)^s \sum_{l=1}^{s} \left\{ \binom{s}{l} \left(\frac{-\beta^2}{6M + \beta^2}\right)^l \times \left[-1 + (1 - b^{-2})^l \sum_{\tau=0}^{m-1} \binom{l + \tau - 1}{l - 1} b^{-2\tau} R(l, m - \tau)\right]\right\}. \tag{2.5}$$

The first step in the proof of this theorem is to calculate $3E|z_j - z_j'|$ for a randomized net and obtain a formula analogous to (2.2).

LEMMA 2.4. *Suppose that P is a randomized (t, m, s)-net satisfying*

Assumption 1.2, *and that z and z' are two points in P, such that the components z_j and z'_j share the same first k_j digits, but no more. It follows that*

$$3E|z_j - z'_j| = (1 + b^{-1})b^{-k_j}, \quad \prod_{j\in u}[3E|z_j - z'_j|] = (1 + b^{-1})^{|u|}b^{-\sigma(k_u)}.$$

PROOF. The jth components z_j and z'_j share the same first k_j digits, but have different $k_j + 1$st digits, if and only if the original y_j and y'_j from which they came have the same first k_j digits but different $k_j + 1$st digits. In this case, Assumption 1.2(c) implies that

$$|z_j - z'_j| = [|z_{k_j+1,j} - z'_{k_j+1,j}| + (\delta - \delta')]b^{-k_j-1},$$

where $(z_{k_j+1,j}, z'_{k_j+1,j})$ is distributed according to Assumption 1.2(c)(ii), and where δ and δ' are uniformly distributed on [0, 1). Then

$$3E|z_j - z'_j| = \frac{6b^{-k_j-1}}{N^2 - N}\sum_{n=1}^{b-1}\sum_{n'=0}^{n-1}(n - n') = (1 + b^{-1})b^{-k_j}. \qquad \square$$

To compute the mean square $\mathcal{L}^2$-discrepancy for a randomized net one must count how many points share exactly the same first k_j digits for all possible $k_u \in \mathbf{Z}_+^u$. The result for $(0, m, s)$-nets is contained in the lemma below. It seems impossible to extend this calculation to (t, m, s)-nets for $t > 0$ because a (t, m, s)-net is not defined precisely enough. For example, a $(1, m, s)$-net may also be a $(0, m, s)$-net.

LEMMA 2.5. *Suppose that P is any $(0, m, s)$-net (randomized or not). For any fixed z let $\lambda(k_u)$ denote the number of points $z' \in P$ such that z_j and z'_j share exactly the same k_j digits (but no more) for $j \in u$. Then, for finite k_u*

$$\lambda(k_u) = \lambda_{|u|,\sigma(k_u)}, \tag{2.6a}$$

where

$$\lambda_{l,\tau} \equiv b^{m-\tau}(1 - b^{-1})\sum_{r=0}^{m-\tau-1}\binom{l-1}{r}(-b^{-1})^r = b^{m-\tau}(1 - b^{-1})^l R(l, m - \tau). \tag{2.6b}$$

PROOF. As a consequence of Definition 1.1 there are no points (other than z itself) that have the same k_u digits as z for $\sigma(k_u) > m$, and formula (2.6) holds in this case. Now assume that (2.6) holds for all k_u with $\sigma(k_u)$ strictly greater than some $\tau' \le m$. This is shown to imply that (2.6) holds for $\sigma(k_u) = \tau'$, and therefore (2.6) holds for all $\sigma(k_u)$ by induction.

For any $k'_u \in \mathbf{Z}_+^u$ with $\sigma(k'_u) = \tau'$, Definition 1.1 implies that there are $b^{m-\tau'}$ points $z' \in P$ which have the same first k'_u (or more) digits as z. This can be written as

$$b^{m-\tau'} = 1 + \sum_{k_u \geq k'_u} \lambda(k_u),$$

or

$$\lambda(k'_u) = b^{m-\tau'} - 1 - \sum_{k_u > k'_u} \lambda(k_u), \tag{2.7}$$

where $k_u > k'_u$ means that $k_j \geq k'_j$ for all $j \in u$ and $\sigma(k_u) > \sigma(k'_u)$. The term 1 above represents $z' = z$. Since (2.6) is assumed to be true for all $\sigma(k_u) > \sigma(k'_u) = \tau'$ and since there are $\binom{|u|+\tau-\tau'-1}{\tau-\tau'}$ different $k_u > k'_u$ with $\sigma(k_u) = \tau$, the sum over $k_u > k'_u$ may be rewritten as

$$\sum_{k_u > k'_u} \lambda(k_u) = \sum_{\tau=\tau'+1}^{m-1} \binom{|u| + \tau - \tau' - 1}{\tau - \tau'} \lambda_{|u|,\tau} = \sum_{\tau=1}^{m-\tau'-1} \binom{|u| + \tau - 1}{\tau} \lambda_{|u|,\tau+\tau'}$$

$$= (1 - b^{-1}) \sum_{\tau=1}^{m-\tau'-1} \sum_{r=0}^{m-\tau-1} \binom{|u| + \tau - 1}{\tau} \binom{|u| - 1}{r} (-1)^r b^{m-\tau-\tau'-r}.$$

To simplify this double sum, the index r is replaced by $r' = r + \tau$ and the order of summation is reversed:

$$\sum_{k_u > k'_u} \lambda(k_u) = (1 - b^{-1}) \sum_{r'=1}^{m-\tau-1} (-1)^{r'} b^{m-\tau'-r'} \sum_{\tau=1}^{r'} \binom{|u| + \tau - 1}{\tau} \binom{|u| - 1}{r' - \tau} (-1)^{\tau}.$$

An identity for binomial coefficients [Prudnikov et al. 1986, sec. 4.2.5, eq. 47] simplifies the inner sum:

$$\sum_{k_u > k'_u} \lambda(k_u) = (1 - b^{-1}) \sum_{r'=1}^{m-\tau-1} (-1)^{r'} b^{m-\tau'-r'} \left[(-1)^{r'} - \binom{|u| - 1}{r'} \right]$$

$$= (1 - b^{-1}) b^{m-\tau'} \sum_{r'=0}^{m-\tau-1} \left[b^{-r'} - \binom{|u| - 1}{r'} (-b)^{-r'} \right]$$

$$= 1 - b^{m-\tau'} - \lambda_{|u|,\tau'}.$$

Substituting this formula for the sum back into (2.7) gives $\lambda(k'_u) = \lambda_{|u|,\tau'}$, which completes the proof of (2.6). □

PROOF OF THEOREM 2.3. Combining the results of Lemmas 2.4 and 2.5, the sum over $z, z' \in P$ in (2.1) can be written as:

$$\frac{1}{N^2} \sum_{z,z' \in P} \prod_{j \in u} [3E|z_j - z'_j|] = b^{-m}(1 + b^{-1})^{|u|} \sum_{k_u \in \mathrm{Z}_+^u} b^{-\sigma(k_u)} \lambda_{|u|,\sigma(k_u)}$$

$$= (1 - b^{-2})^{|u|} \sum_{k_u \in \mathrm{Z}_+^u} b^{-2\sigma(k_u)} R(|u|, m - \sigma(k_u)).$$

For every $\tau = 0, \ldots, m - 1$, there are $\binom{|u| + \tau - 1}{|u| - 1}$ distinct $k_u \in \mathbf{Z}_+^u$ with $\sigma(k_u) = \tau$, so

$$\frac{1}{N^2} \sum_{z,z' \in P} \prod_{j \in u} [3E|z_j - z'_j|] = (1 - b^{-2})^{|u|} \sum_{\tau=0}^{m-1} \binom{|u| + \tau - 1}{|u| - 1} b^{-2\tau} R(|u|, m - \tau)$$

Also, for any positive integer l, there are $\binom{s}{l}$ distinct $u \subseteq S$ with $|u| = l$, which allows (2.1) to be written as:

$$E\{[D_2(P)]^2\} = \left(M + \frac{\beta^2}{6}\right)^s \sum_{l=1}^{s} \left\{ \binom{s}{l} \left(\frac{-\beta^2}{6M + \beta^2}\right)^l \right.$$

$$\left. \times \left[-1 + (1 - b^{-2})^l \sum_{\tau=0}^{m-1} \binom{l + \tau - 1}{l - 1} b^{-2\tau} R(l, m - \tau) \right] \right\},$$

thereby completing the proof. □

From (2.4), it follows that

$$R(l, \tau) = \begin{cases} 0 & \tau \leq 0, \\ 1 & \tau \geq l. \end{cases} \tag{2.8}$$

Therefore, only $O(s^2)$ operations are required to calculate the $O(s^2)$ nontrivial values of $R(l, m - t)$ for $1 \leq m - t < l \leq s$. The sums over l and τ in (2.5) involve a total of $O(ms)$ terms, so the mean square $\mathcal{L}^2$-discrepancy of randomized $(0, m, s)$-nets can be calculated in $O(s \log(N) + s^2)$ operations as N and/or s tend to infinity. This order is much smaller than that required to compute the $\mathcal{L}^2$-discrepancy of an arbitrary set P.

Formula (1.4) for $D_2(P)$ and formula (1.7) for $D_{2,u}(P)$ are quite similar. Thus, the root mean square of $D_{2,u}(P)$ can be calculated using the same arguments as for Theorems 2.2 and 2.3.

THEOREM 2.6. *If P is a simple random sample, then*

$$E\{[D_{2,u}(P)]^2\} = \frac{1}{N}\left[\left(\bar{M} + \frac{\beta^2}{6}\right)^{|u|} - \bar{M}^{|u|}\right].$$

THEOREM 2.7. *Let P be a $(0, m, s)$-net randomized according to Assumption 1 and u be some non-empty subset of S. Then*

$$E\{[D_{2,u}(P)]^2\} = \left(\bar{M} + \frac{1}{6}\right)^{|u|} \sum_{l=1}^{|u|} \left\{ \binom{|u|}{l} \left(\frac{-1}{6\bar{M} + 1} \right)^{l} \times \left[-1 + (1 - b^{-2})^{l} \sum_{\tau=0}^{m-1} \binom{l + \tau - 1}{l - 1} b^{-2\tau} R(l, m - \tau) \right] \right\}. \tag{2.9}$$

3. ASYMPTOTICS

As N (or equivalently m) approaches infinity, one would like to know how quickly the mean squared $\mathcal{L}^2$-discrepancy tends to zero. This can be calculated by an asymptotic analysis of the formulas in Theorems 2.3 and 2.7.

THEOREM 3.1. *Let P be a $(0, m, s)$-net randomized according to Assumption 1.2. Then*

$$E\{[D_2(P)]^2\} \sim \frac{\beta^{2s}(b - b^{-1})^{s-1}}{6^s(s - 1)![\log(b)]^{s-1}} N^{-2}[\log(N)]^{s-1} \text{ as } N \to \infty, \tag{3.1}$$

$$E\{[D_{2,u}(P)]^2\} \sim \frac{(b - b^{-1})^{|u|-1}}{6^{|u|}(|u| - 1)![\log(b)]^{|u|-1}} N^{-2}[\log(N)]^{|u|-1} \text{ as } N \to \infty. \tag{3.2}$$

PROOF. The dependence on m in (2.5) and (2.9) lies in the term in square brackets, which we call $T(l)$. Its asymptotic form can be obtained by rewriting the sum over τ as an infinite sum. Since

$$-1 = -(1 - b^{-2})^{l} \sum_{\tau=0}^{\infty} \binom{l + \tau - 1}{l - 1} b^{-2\tau},$$

it follows by (2.8) that

$$\begin{aligned} T(l) &\equiv -1 + (1 - b^{-2})^{l} \sum_{\tau=0}^{m-1} \binom{l + \tau - 1}{l - 1} b^{-2\tau} R(l, m - \tau) \\ &= (1 - b^{-2})^{l} \sum_{\tau=0}^{\infty} \binom{l + \tau - 1}{l - 1} b^{-2\tau} [R(l, m - \tau) - 1] \\ &= (1 - b^{-2})^{l} b^{2(l-1-m)} \sum_{\tau=\max(0, l-1-m)}^{\infty} \binom{m + \tau}{l - 1} b^{-2\tau} [R(l, l - 1 - \tau) - 1]. \end{aligned}$$

The definition of $R(l, \tau)$ in (2.4) can be used to simplify this expression further:

$$R(l, l-1-\tau) - 1 = -(1-b^{-1})^{1-l} \sum_{r=l-1}^{l-1-\tau} \binom{l-1}{r} (-b^{-1})^r$$

$$= -(1-b^{-1})^{1-l}(-b^{-1})^{l-1} \sum_{r=0}^{\tau} \binom{l-1}{r} (-b^{-1})^{-r},$$

which implies

$$T(l) = (-1)^l (1+b^{-1})^l (1-b^{-1})^{-1} b^{l-1-2m}$$

$$\times \sum_{\tau=\max(0,l-1-m)}^{\infty} \sum_{r=0}^{\tau} \binom{m+\tau}{l-1} \binom{l-1}{r} (-1)^r b^{-2\tau+r}.$$

For large m the binomial coefficient $\binom{m+\tau}{l-1}$ is asymptotic to $m^{l-1}/(l-1)!$ and $\max(0, l-1-m) = 0$. By reversing the order of the summation above one can compute an asymptotic form for $T(l)$:

$$T(l) \sim (-1)^l (1+b^{-1})^l (1-b^{-1})^{-1} b^{l-1-2m} \frac{m^{l-1}}{(l-1)!} \sum_{\tau=0}^{\infty} \sum_{r=0}^{\tau} \binom{l-1}{r} (-1)^r b^{-2\tau+r}$$

$$= (-1)^l (1+b^{-1})^l (1-b^{-1})^{-1} b^{l-1-2m} \frac{m^{l-1}}{(l-1)!} \sum_{r=0}^{\infty} \binom{l-1}{r} (-b)^r \sum_{\tau=r}^{\infty} b^{-2\tau}$$

$$= (-1)^l (1+b^{-1})^l (1-b^{-1})^{-1} b^{l-1-2m} \frac{m^{l-1}}{(l-1)!} \sum_{r=0}^{\infty} \binom{l-1}{r} (-b)^r \frac{b^{-2r}}{1-b^{-2}}$$

$$= (-1)^l (b-b^{-1})^{l-1} b^{-2m} \frac{m^{l-1}}{(l-1)!}$$

The asymptotic expected mean square discrepancy involves the sum of $T(l)$ over l. Because of the factor m^{l-1} above, the most significant term occurs when $l = s$. Therefore,

$$E\{[D_2(P)]^2 \sim \left(M + \frac{\beta^2}{6}\right)^s \left(\frac{-\beta^2}{6M+\beta^2}\right)^s T(s) \sim \frac{\beta^{2s}(b-b^{-1})^{s-1}}{6^s(s-1)!} m^{s-1} b^{-2m}$$

$$= \frac{\beta^{2s}(b-b^{-1})^{s-1}}{6^s(s-1)![\log(b)]^{s-1}} N^{-2}[\log(N)]^{s-1},$$

which completes the proof of (3.1). A similar argument is used to prove (3.2). □

Note that the asymptotic behavior of $E\{[D_2(P)]^2\}$ is independent of the function μ and the constant M; the asymptotic behavior of $E\{[D_{2,u}(P)]^2\}$ is independent of the constant β as well. For prime power dimensions s there exist $(0, m, s)$-nets with base $b = s$ [Faure 1982; Niederreiter 1992, th. 4.54]. For this case, we have

$$E\{[D_2(P)]^2\} \sim \beta^{2s}[A(s)]^2 N^{-2}[\log(N)]^{(s-1)} \qquad \text{as} \quad N \to \infty,$$

$$E\{[D_{2,S}(P)]^2\} \sim [A(s)]^2 N^{-2}[\log(N)]^{(s-1)} \qquad \text{as} \quad N \to \infty,$$

where

$$A(s) \equiv \left[\frac{(s - s^{-1})^{s-1}}{6^s(s-1)![\log(s)]^{(s-1)}}\right]^{1/2}.$$

For large s the coefficient $A(s)$ tends to zero faster than exponentially. By Stirling's formula

$$A(s) \sim \left[\frac{e^s}{\sqrt{2\pi s}\, 6^s[\log(s)]^{(s-1)}}\right]^{1/2} \qquad \text{as} \quad s \to \infty.$$

4. DISCUSSION

The mean of the square of a random variable is equal to the square of its mean plus the square of its standard deviation. In particular,

$$E\{[D_2(P)]^2\} = \{E[D_2(P)]\}^2 + \{\text{Std}[D_2(P)]\}^2,$$

so $\sqrt{E\{[D_2(P)]^2\}}$ is an upper bound on both the mean and the standard deviation of $D_2(P)$. Thus, by Theorem 3.1

$$E[D_2(P)] = \mathrm{O}(N^{-1}(\log(N))^{(s-1)/2}),$$

which is the same asymptotic order as a lower bound on $D^*_{2S}(P)$ derived by Roth [1954]. Kuipers and Niederreiter [1974, p. 102] refined this lower bound to give an explicit constant (see also Niederreiter [1978, eq. (3.10) and p. 972]):

$$D^*_{2,S}(P) \geq B(s) N^{-1}[\log(N)]^{(s-1)/2},$$

where

$$B(s) = \begin{cases} 12^{-1/2} & \text{for} \quad s = 1 \\ \{16^s[(s-1)\log(2)]^{(s-1)/2}\}^{-1} & \text{for} \quad s \geq 2. \end{cases}$$

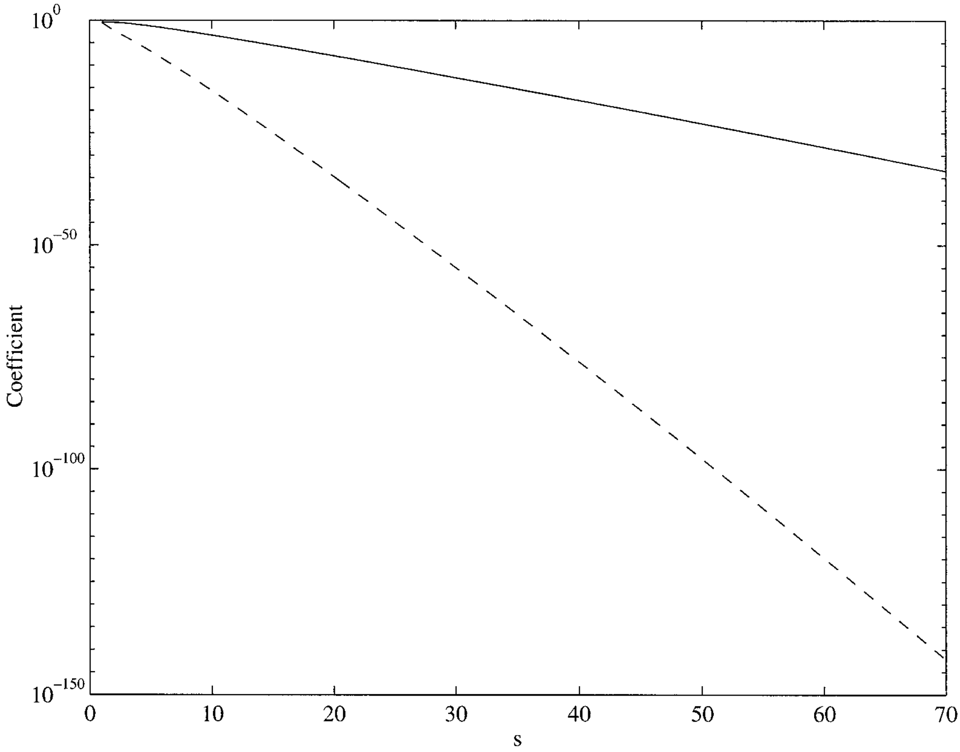


Fig. 1. A comparison of A(s) (solid) and B(s) (dashed).

A lower bound on $D_2^*(P)$ follows:

$$D_2^*(P) \geq \left\{ \sum_{l=1}^{s} \binom{s}{l} [B(l)]^2 [\log(N)]^{l-1} \right\}^{1/2} N^{-1}. \tag{4.1}$$

To compare the lower bound on the $\mathscr{L}^2$-star discrepancy for all P and the asymptotic behavior for randomized $(0, m, s)$-nets, the coefficients $A(s)$ and $B(s)$ are plotted in Figure 1. It is clear that $B(s)$ tends to zero much faster than $A(s)$ as the dimension increases. The reason could be that (i) the lower bound is not tight, (ii) the average discrepancy of $(0, m, s)$-nets is much worse than the discrepancies of some especially good $(0, m, s)$-nets, or (iii) the optimal discrepancy is obtained only by some other kind of point set P. It is not clear to the author which reason is more likely.

Given a lower bound, L, on $D_2(P)$, such as zero or that above, one can derive an upper bound on the likelihood that a randomized $(0, m, s)$-net has a large discrepancy. For any non-negative $y \geq L$

$$E\{[D_2(P)]^2\} \geq L^2 \operatorname{Prob}\{D_2(P) \leq y\} + y^2 \Pr\{D_2(P) > y\},$$

that is,

$$\Pr\{D_2(P) > y\} \geq \frac{E\{[D_2(P)]^2\} - L^2}{y^2 - L^2}.$$

Replacing y by a new constant α^2 equal to the right hand side of this equation gives

$$\Pr\left\{ D_2(P) > \frac{\sqrt{E\{[D_2(P)]^2\} - (1 - \alpha^2)L^2}}{\alpha} \right\} \leq \alpha^2.$$

In particular, no more than 1% of randomized $(0, m, s)$-nets can have an $\mathscr{L}^2$-discrepancy greater than $10\sqrt{E\{[D_2(P)]^2\} - 0.99L^2\}}$.

Figure 2 shows the root mean square $\mathscr{L}^2$-star discrepancy for randomized $(0, m, s)$-nets in base s plotted versus N from formula (2.5). The values of M and β are given in (1.5). The dimensions considered are the prime numbers 2 through 13, and N runs up to 10^{10}. For comparison, the root mean square asymptotic $\mathscr{L}^2$-discrepancy, (3.1), the root mean square $\mathscr{L}^2$-star discrepancy of a simple random sample, (2.3), and the lower bound on the $\mathscr{L}^2$-star discrepancy, (4.1), are also shown.

Figure 3 is similar to Figure 2 except that it shows the $\mathscr{L}^2$-symmetric discrepancy defined in Hickernell [to appear]. In contrast to (1.5), the symmetric discrepancy is defined by

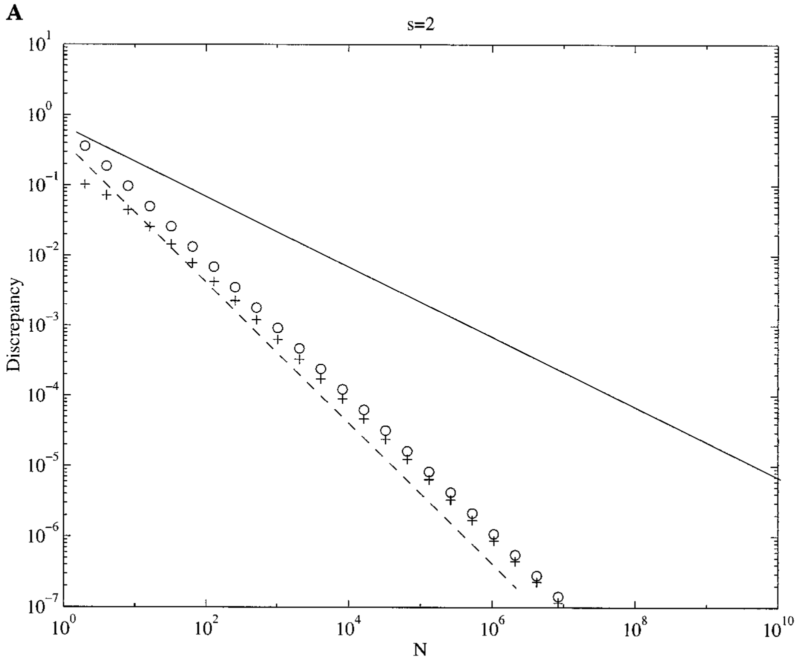
$$\mu(x) = \frac{1}{24} - \frac{(x - 1/2)^2}{2}, \quad \beta = 2, \quad \bar{M} = \frac{1}{3}, \quad M = \frac{4}{3}.$$

Because $\mu(x)$ is invariant when x is replaced by $1 - x$, the symmetric discrepancy does not change when the P is reflected about any plane $x_j = 1/2$. Like the star discrepancy, the symmetric discrepancy has a geometric interpretation. For any subset of coordinate indices u, the planes passing through a point x_u parallel to the faces of the cube C^u divide this cube into $2^{|u|}$ intervals (rectangular solids). Each interval consists of the points between a vertex of the cube and the point x_u. These intervals can be denoted as odd or even depending on the sum of the coordinates of the corresponding vertex. For example, $[(0, \ldots, 0), x_u)$ is an even interval, and $[x_u, (1, \ldots, 1))$ is even or odd according to whether $|u|$ is even or odd. Let $J_e(x_u)$ denote the union of the even intervals. The $\mathscr{L}^2$-symmetric discrepancy is defined as

$$[D_2(P)]^2 = 2 \sum_{\emptyset \subset u \subseteq S} \left| \frac{|P_u \cap J_e(x_u)|}{N} - \mathrm{Vol}(J_e(x_u)) \right|^2.$$

In both Figures 2 and 3, the root mean square $\mathscr{L}^2$-discrepancies of the nets approach their asymptotic behaviors more quickly for lower dimen-

A



B

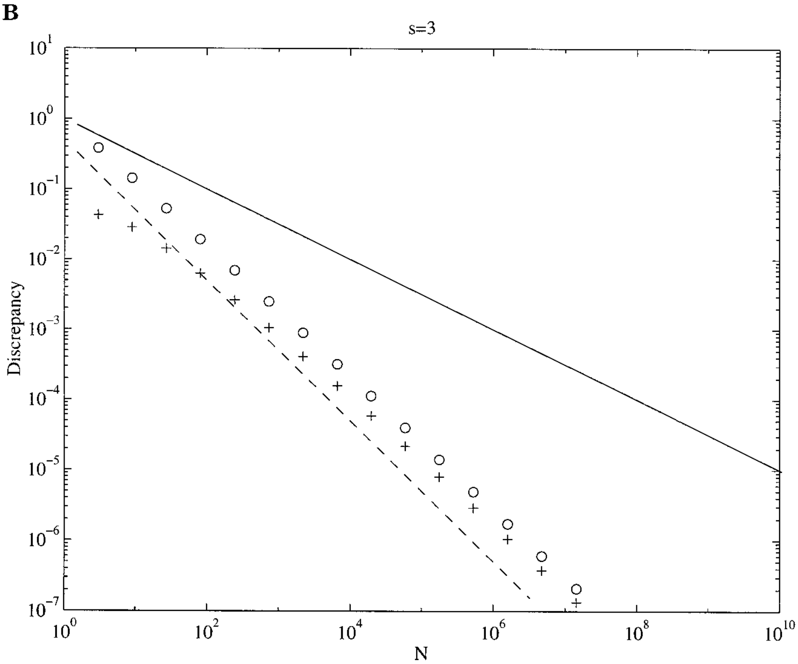
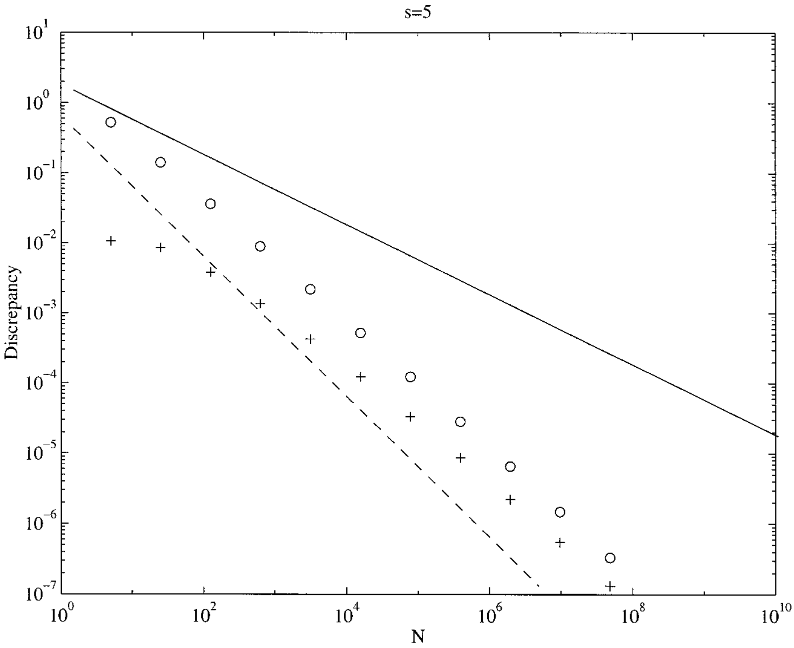


Fig. 2(a). Root mean square $\mathscr{L}^2$-star discrepancy for randomized $(0, m, s)$-nets base s versus N ($\circ$), asymptotic behavior ($+$), root mean square $\mathscr{L}^2$-star discrepancy of simple random sample (solid), and lower bound on $\mathscr{L}^2$-star discrepancy (dashed). $s = 2$. 2b. $s = 3$. 2c. $s = 5$. 2d. $s = 7$. 2e. $s = 11$. 2f. $s = 13$.

C



D

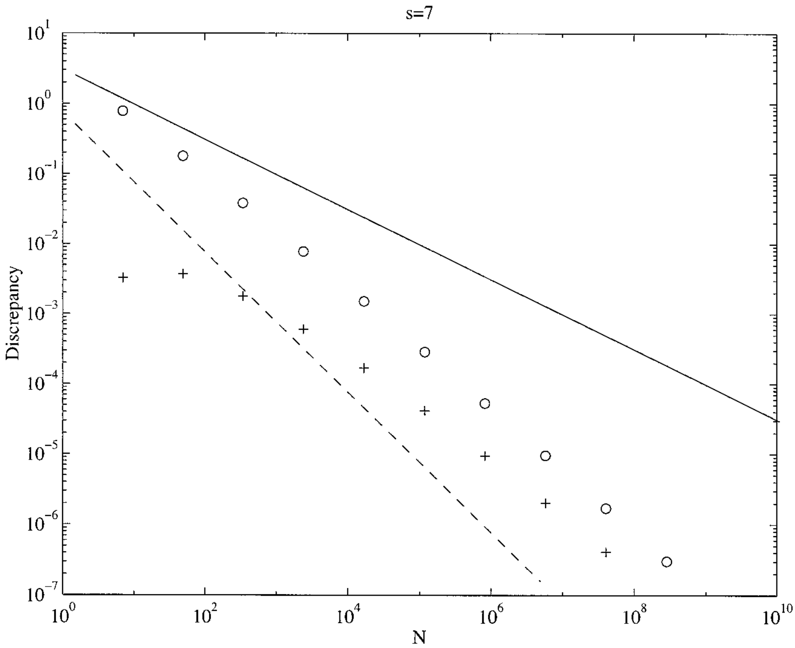
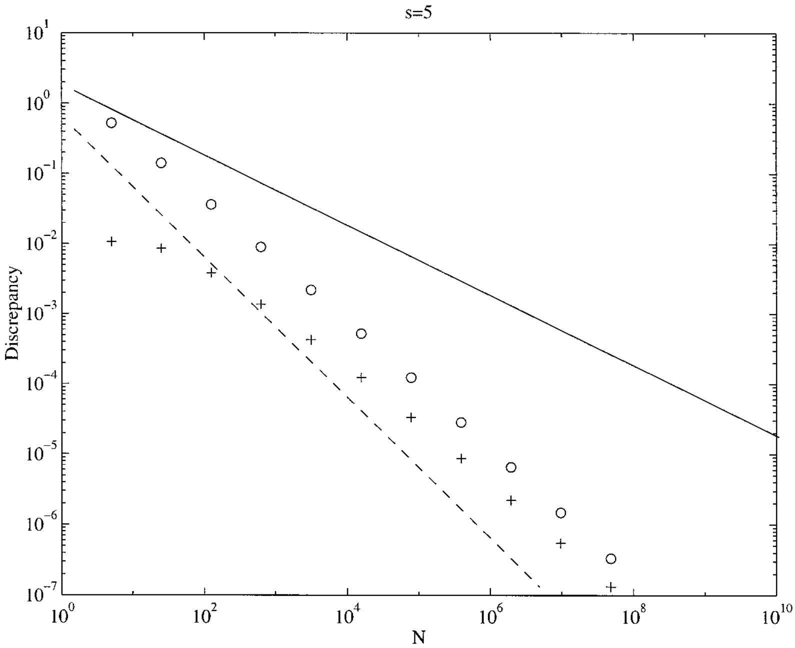


Fig. 2.—*continued*

E



F

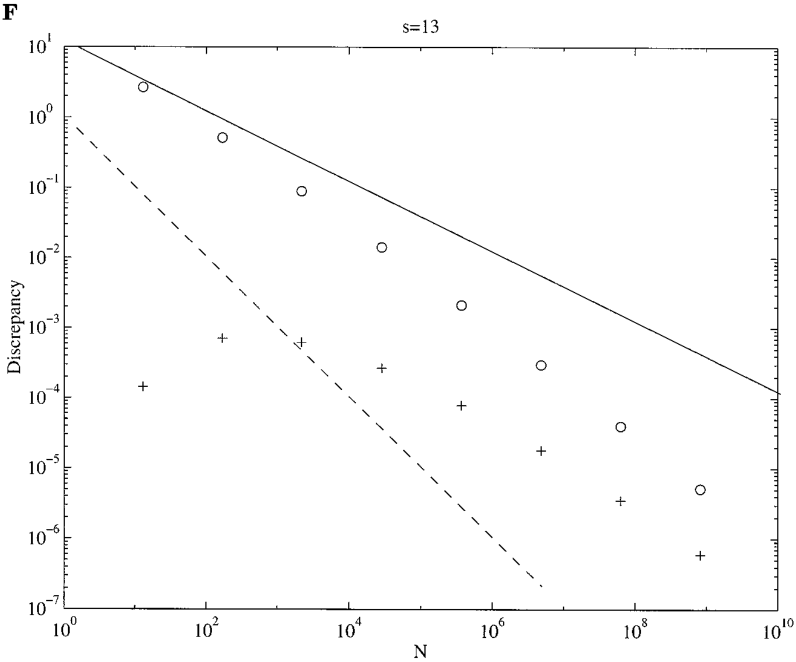
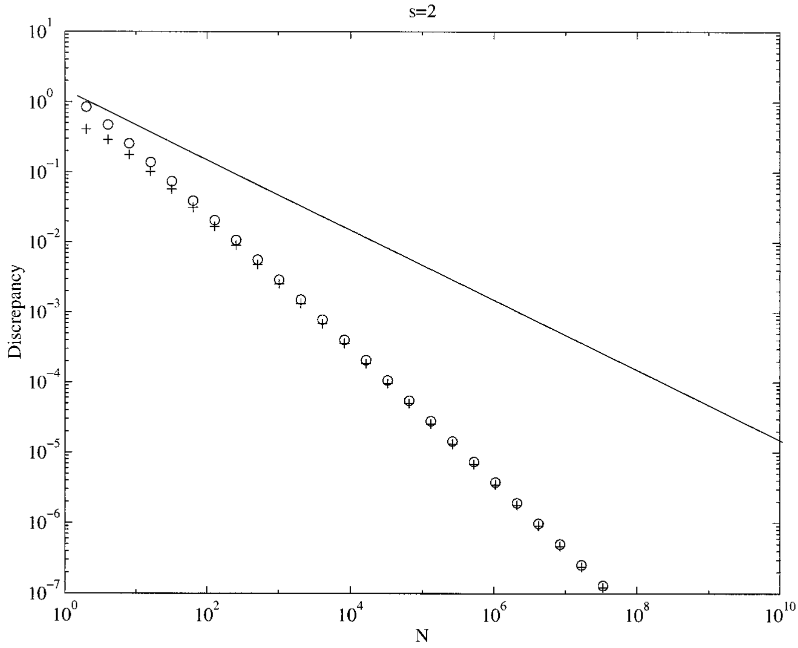


Fig. 2.—*continued*

A



B

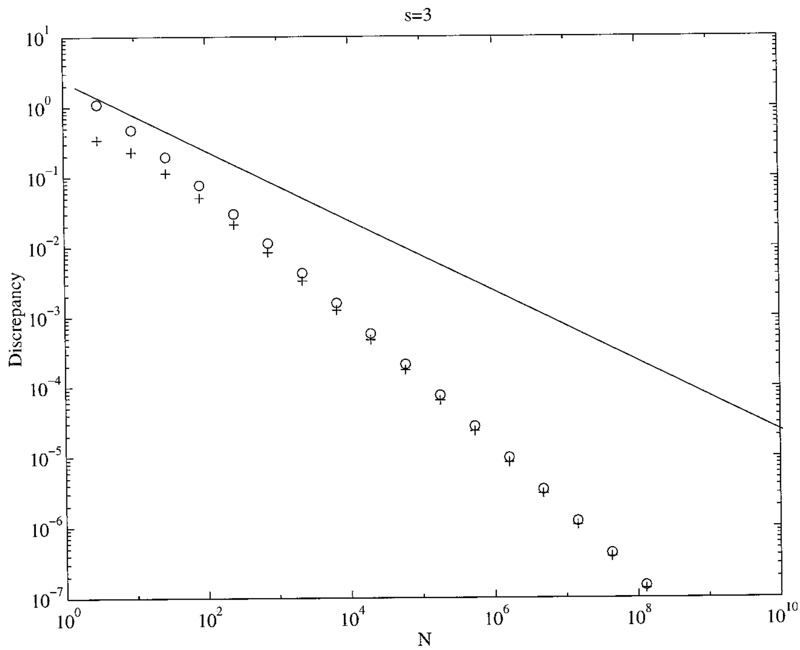
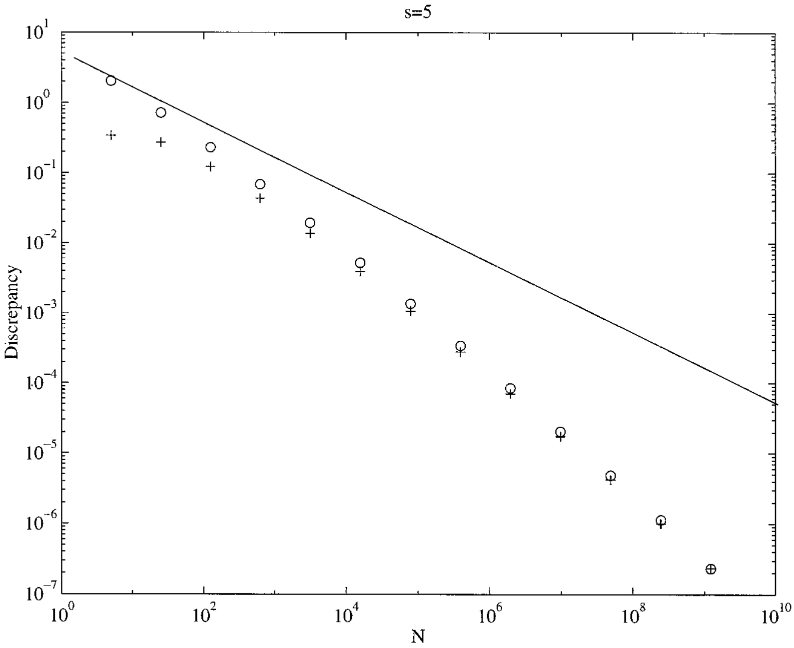


Fig. 3(a). Root mean square $\mathscr{L}^2$-symmetric discrepancy of randomized $(0, m, s)$-nets base s versus N ($\circ$), asymptotic behavior ($+$), and root mean square $\mathscr{L}^2$-symmetric discrepancy of simple random sample (solid). $s = 2$. 3(b). $s = 3$. 3(c). $s = 5$. 3(d). $s = 7$. 3(e). $s = 11$. 3(f). $s = 13$.

C



D

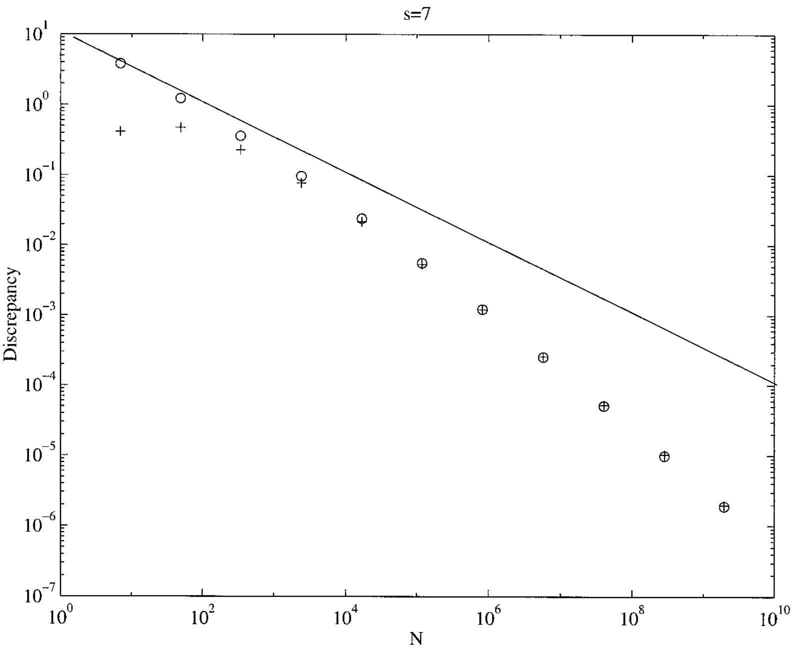
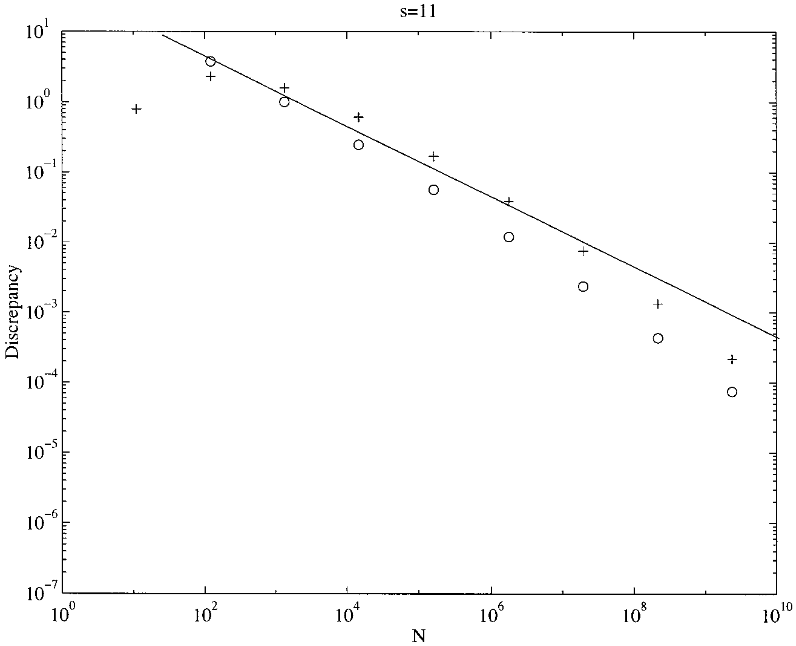


Fig. 3.—*continued*

E



F

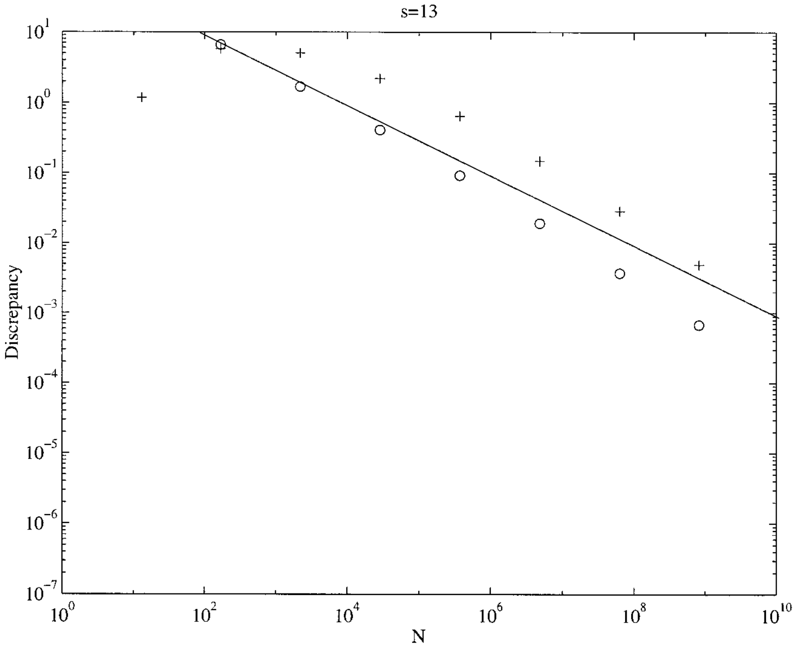


Fig. 3.—*continued*

sions. The asymptotic values of the discrepancies increase with N initially because the term $N^{-1}[\log(N)]^{(s-1)/2}$ is increasing for $N < e^{(s-1)/2}$.

The discrepancies of the nets decay to zero faster with increasing sample size than the discrepancies of the simple random samples. The number of $\log(N)$ factors in the asymptotic behavior of the discrepancy for nets increases with dimension (th. 3.1), thus reducing their advantage over simple random samples. This effect is more pronounced for the symmetric discrepancy than the star discrepancy because of the larger value of β in the former. Formula (1.6) shows how $[D_2(P)]^2$ is a sum of the $[D_{2,u}(P)]^2$ with weights $\beta^{2|u|}$. A larger value of β accentuates the importance of terms with large $|u|$, that is, those terms that decay to zero more slowly according to Theorem 3.1. A large or small value of β is not inherently better, but rather reflects a personal choice. Although decreasing β decreases the $\mathscr{L}^2$-discrepancy, it increases the variation (or norm) of the integrand in the corresponding quadrature error bound. Thus, there is a trade-off. The effect of varying β is discussed further in Hickernell [to appear]. Our numerical experiments indicate that the root mean square discrepancy of a randomized $(0, m, s)$-net base s is never greater than that of a simple random sample, regardless of the choice of μ and β. However, this conjecture has not yet been proven.

Equations (1.1) and (3.1) imply that for randomized $(0, m, s)$-nets

$$E\{\sup_{V(f)=1} [\mathrm{Err}(f)]^2\} = E\{[D_2(P)]^2\} = \mathrm{O}(N^{-2}[\log(N)]^{(s-1)}) \text{ as } N \to \infty,$$

where the variation of the integrand, $V(f)$, is defined in Hickernell [to appear]. For a fixed integrand f under similar smoothness conditions, Owen [1997b] showed that

$$E\{[\mathrm{Err}(f)]^2\} = \mathrm{O}(N^{-3}[\log(N)]^{(s-1)} \quad \text{as} \quad N \to \infty$$

for randomized $(0, m, s)$-nets. The additional power of N^{-1} in the latter result is due to the integrand being fixed in advance of choosing a randomized net for quadrature. In the former result, the integrand is chosen pessimistically after picking a specific randomized net.

One may wonder why we have computed the mean square $\mathscr{L}^2$-discrepancy and not the mean $\mathscr{L}^2$-discrepancy itself or the mean $\mathscr{L}^p$-discrepancy in general. The answer is convenience. Only the square $\mathscr{L}^2$-discrepancy appears to have a simple enough form to allow its mean value to be computed.

ACKNOWLEDGMENTS

The author would like to thank Professors Harald Niederreiter and Art Owen for their valuable comments and suggestions.

REFERENCES

ABRAMOWITZ, M., AND STEGUN, I. A. Eds. 1964. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. U.S. Government Printing Office, Washington, DC.

FAURE, H. 1982. Discrépance de suites associées à un système de numération (en dimension s). *Acta Arith. 41*, 337–351.

HEINRICH, S. 1996. Efficient algorithms for computing the L_2 discrepancy. *Math. Comp.*

HICKERNELL, F. J. 1996. Quadrature error bounds with applications to lattice rules. *SIAM J. Numer. Anal. 33,* 65, 1621–1633.

HICKERNELL, F. J. A generalized discrepancy and quadrature error bound. *Math. Comp.*, to appear.

KUIPERS, L., AND NIEDERREITER, H. 1974. *Uniform Distribution of Sequences.* Wiley, New York.

MOROKOFF, W. J., AND CAFLISCH, R. E. 1994. Quasi-random sequences and their discrepancies. *SIAM J. Sci. Comput. 15*, 1251–1279.

NIEDERREITER, H. 1978. Quasi-Monte Carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc. 84*, 957–1041.

NIEDERREITER, H. 1992. Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia, Pa.

OWEN, A. B. 1995. Equidistributed Latin hypercube samples. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*. Lecture Notes in Statistics, Vol. 106. Springer-Verlag, New York, 299–317.

OWEN, A. B. 1997a. Monte carlo variance of scrambled equidistribution quadrature. *SIAM J. Numer. Anal. 34*.

OWEN, A. B. 1997b. Scrambled net variance for integrals of smooth functions. *Ann. Stat. 25* (to appear).

PRUDNIKOV, A. P., BRYCHKOV, Y. A., AND MARICHEV, O. I. 1986. *Integrals and Series*. Gordon and Breach, New York.

RITTER, K. 1995. Average case analysis of numerical problems. Ph.D. dissertation, Universität Erlangen-Nürnberg, Erlangen, Germany.

ROTH, K. F. 1954. On irregularities of distribution. *Mathematika 1*, 73–79.

SACKS, J., AND YLVISACKER, D. 1970. Statistical design and integral approximation. In *Proceedings of the 12th Biennial Canadian Math Congress*. Canadian Math Society, Montreal, Canada, 115–136.

SLOAN, I. H., AND JOE, S. 1994. *Lattice Methods for Multiple Integration*. Oxford University Press, Oxford, England.

SOBOL', I. M. 1969. Multidimensional quadrature formulas and Hoar functions. Izdat "Nauka," Moscow (in Russian).

WARNOCK, T. T. 1972. Computational investigations of low discrepancy point sets. In *Applications of Number Theory to Numerical Analysis*. Academic Press, Orlando, Fla., pp. 319–343.

WOZNIAKOWSKI, H. 1991. Average case complexity of multivariate integration. *Bull. Amer. Math. Soc. 24*, 185–194.

ZAREMBA, S. K. 1968. Some applications of multidimensional integration by parts. *Ann. Polon. Math. 21*, 85–96.

Received April 1996; revised August 1996; accepted August 1996