

SCRAMBLED NET VARIANCE FOR INTEGRALS OF SMOOTH FUNCTIONS¹

BY ART B. OWEN

Stanford University

Hybrids of equidistribution and Monte Carlo methods of integration can achieve the superior accuracy of the former while allowing the simple error estimation methods of the latter. One version, randomized (t, m, s) -nets, has the property that the integral estimates are unbiased and that the variance is $o(1/n)$, for any square integrable integrand.

Stronger assumptions on the integrand allow one to find rates of convergence. This paper shows that for smooth integrands over s dimensions, the variance is of order $n^{-3}(\log n)^{s-1}$, compared to n^{-1} for ordinary Monte Carlo. Thus the integration errors are of order $n^{-3/2}(\log n)^{(s-1)/2}$ in probability. This compares favorably with the rate $n^{-1}(\log n)^{s-1}$ for unrandomized (t, m, s) -nets.

1. Introduction. We consider the problem of integrating a function f over the unit cube of dimension s . We assume that f is measurable without pointing out the places below where this assumption matters, and that f^2 is also integrable. For large enough s , Monte Carlo methods and equidistribution methods are most widely used. Recently Owen (1995, 1997) proposed a hybrid of these two techniques based on scrambling the digits in a (t, m, s) -net in base b . The resulting method provides unbiased estimates of $I = \int_{[0,1]^s} f(X) dX$ having a variance that is $o(1/n)$ along the sequence $n = \lambda b^m$, $1 \leq \lambda < b$, $0 \leq m$. Thus, for any nonconstant f , the ratio of the scrambled net variance to the ordinary Monte Carlo variance tends to zero as $n \rightarrow \infty$.

The main purpose of this paper is to find rates for the convergence, under stronger assumptions on the integrand f . Theorem 2 shows that if f is smooth enough then the scrambled net variance is of order $n^{-3}(\log n)^{s-1}$ as $n = \lambda b^m \rightarrow \infty$. High powers of $\log n$ might not be negligible until n is very large, and this raises the possibility that the scrambled net variance might be worse than the Monte Carlo variance for finite n . Owen (1997) shows that for $n = b^m$ the scrambled net variance is never more than a constant times the Monte Carlo variance, and that this constant is never more than $e \doteq 2.718$ for any f in any dimension. Theorem 1 here improves that result showing that for $n = \lambda b^m$ the scrambled net variance is never more than $1 + e \doteq 3.718$ times the Monte Carlo variance.

Received May 1996; revised October 1996.

¹ Research supported by National Science Foundation and Alcoa Foundation.

AMS 1991 subject classifications. Primary 65C05, 68O20; secondary 62D05, 62K05.

Key words and phrases. Integration, Latin hypercube, multiresolution, orthogonal array sampling, quasi-Monte Carlo, wavelets.

The paper proceeds as follows: Section 2 provides background material, and the proof of Theorem 1. Section 3 gives a simple proof that the scrambled net variance is of order n^{-3} for smooth univariate functions and $n = b^m$. Section 4 considers the general s -dimensional case and sample sizes $n = \lambda b^m$. The variance rate $n^{-3}(\log n)^{s-1}$ is established in Theorem 2. Section 5 considers an example integrand for which it is possible to compute both the scrambled net variance and the Monte Carlo variance and compare them to an asymptotic formula. Section 6 discusses for what n the asymptotic rate might be expected to take hold and what sort of integrands are likely to be handled well by scrambled net Monte Carlo. Section 6 also includes a discussion of several slightly different definitions of the error rate and compares the rates of convergence seen here to worst case rates obtained for unscrambled nets and to some rates for the average case over random integrands.

2. Background. This section introduces the notation and definitions used. Here and below, the integer $s \geq 1$ denotes the dimension of the space in which we work. The space is always the half-open cube $[0, 1)^s$. The integrand is $f: [0, 1)^s \rightarrow R$. The goal is to find $I = \int_{[0, 1)^s} f(X) dX$.

Any domain that can be expressed as the image of the unit cube under some function can be handled by subsuming the imaging function into f . The half-open cube is chosen because it partitions easily into half-open subcubes and hyperrectangles.

A point in the cube is denoted by $X = (X^1, \dots, X^s)$ or by $X_i = (X_i^1, \dots, X_i^s)$. The estimate of I is $\hat{I} = \hat{I}_n = n^{-1} \sum_{i=1}^n f(X_i)$ for carefully chosen points $X_i \in [0, 1)^s$.

The set $\mathcal{A} = \{1, 2, \dots, s\}$ denotes the coordinate axes of $[0, 1)^s$. The letter u denotes a subset of \mathcal{A} and $|u|$ is the cardinality of u . These subsets often appear as superscripts: $[0, 1)^u$ denotes the space of values for components X^j of X with $j \in u$, X^u denotes the coordinate projection of X onto $[0, 1)^u$ and, in integrals, $dX^u = \prod_{j \in u} dX^j$. The case $u = \emptyset$ can require special attention, either by a natural convention, or by restricting some operations to $|u| > 0$.

The integer $b \geq 2$ is used throughout as a base for representing points in $[0, 1)$. Thus $X_i^j = \sum_{k=1}^{\infty} x_{ijk} b^{-k}$ where x_{ijk} are integers with $0 \leq x_{ijk} < b$.

2.1. (t, m, s) -nets, (t, s) -sequences and (λ, t, m, s) -nets. Here we introduce equidistribution methods known as (t, m, s) -nets and (t, s) -sequences. These have been developed by Sobol', Faure, and Niederreiter and a comprehensive discussion of them appears in the monograph by Niederreiter [(1992), Chapter 4] on which this subsection is based.

Equidistribution methods produce sequences $X_1, \dots, X_n \in [0, 1)^s$ such that the discrete uniform distribution on X_1, \dots, X_n closely approximates the continuous uniform distribution on $[0, 1)^s$. This typically involves showing that for a class of subsets $E \subseteq [0, 1)^s$, the number of points $X_i \in E$ is close to n times the volume of E .

DEFINITION 1. An elementary interval of $[0, 1]^s$ in base b is a set of the form

$$E = \prod_{j=1}^s \left[\frac{t_j}{b^{k_j}}, \frac{t_j + 1}{b^{k_j}} \right)$$

for nonnegative integers k_j and $t_j < b^{k_j}$.

This elementary interval is a hyperrectangle of volume $b^{-(k_1 + \dots + k_s)}$. If a finite sequence of points X_1, \dots, X_n is to be nearly equidistributed in $[0, 1]^s$ there should be nearly $nb^{-(k_1 + \dots + k_s)}$ points in E .

DEFINITION 2. Let t and m be nonnegative integers. A finite sequence $X_1, \dots, X_n \in [0, 1]^s$ with $n = b^m$ is a (t, m, s) -net in base b if every elementary interval in base b of volume b^{t-m} contains exactly b^t points of the sequence.

Clearly smaller values of t imply better equidistribution properties for the net. In the best case with $t = 0$, every elementary interval of volume $1/n$ has one of the n points in the sequence.

DEFINITION 3. Let t be a nonnegative integer. An infinite sequence $X_1, X_2, \dots \in [0, 1]^s$ is a (t, s) -sequence in base b if for all $m \geq 0$ and all $k \geq 0$ the finite sequence $X_{kb^{m+1}}, \dots, X_{(k+1)b^m}$ is a (t, m, s) -net in base b .

An advantage of using nets taken from (t, s) -sequences is that one can increase n through a sequence of values $n = \lambda b^m$, $1 \leq \lambda < b$, and find that all of the points used in $\hat{I}_{\lambda b^m}$ are also used in $\hat{I}_{(\lambda+1)b^m}$. As n increases through this sequence of values, every elementary interval in base b of volume V eventually contains nV of the points, and once such an interval is balanced this way, it remains balanced as n increases. Niederreiter (1992) discusses existence and construction of (t, m, s) -nets and (t, s) -sequences.

The initial λb^m points of a (t, s) -sequence are well equidistributed but are not ordinarily a (t, m, s) -net. Owen (1997) introduces the following definition to describe such point sets.

DEFINITION 4. Let m, t, λ be integers with $m \geq 0$, $0 \leq t \leq m$, and $1 \leq \lambda < b$. A sequence (X_i) of λb^m points is called a (λ, t, m, s) -net in base b if every elementary interval in base b of volume b^{t-m} contains λb^t points of the sequence and no elementary interval in base b of volume b^{t-m-1} contains more than b^t points of the sequence.

Numerical integration by averaging over the points of a (t, m, s) -net has an error of order $n^{-1}(\log n)^{s-1}$, for functions of bounded variation in the sense of Hardy and Krause. See Niederreiter [(1992), Chapter 4] for this result and some related ones. The rate attained along initial subsequences of a fixed (t, s) -sequence is $n^{-1}(\log n)^s$.

2.2. *Base b scrambling of the unit cube.* Suppose that X_1, \dots, X_n is a (t, m, s) -net in base b . Write $X_i^j = \sum_{k=1}^{\infty} x_{ijk} b^{-k}$. It is possible to apply some permutations to the digits x_{ijk} while retaining the net property for X_1, \dots, X_n .

Owen (1995) describes such a scheme using uniform random permutations of the integers $0, \dots, b - 1$. There are $b!$ permutations of these integers and a uniform random permutation is one in which all permutations have the same probability.

This scheme proceeds as follows, and may be described for a generic point $A \in [0, 1]^s$. Suppose $A = (A^1, \dots, A^s)$ and $A^j = \sum_{k=1}^{\infty} a_{ijk} b^{-k}$. Now let $X = (X^1, \dots, X^s)$ with $X^j = \sum_{k=1}^{\infty} x_{ijk} b^{-k}$. The scrambled version of A is the point X , obtained by taking x_{ijk} , to be permutations of the digits a_{ijk} as described below.

The permutation applied to a_{ijk} depends on the values of a_{ijh} for $h < k$. Specifically $x_{ij1} = \pi_j(a_{ij1})$, $x_{ij2} = \pi_{j a_{ij1}}(a_{ij2})$, $x_{ij3} = \pi_{j a_{ij1} a_{ij2}}(a_{ij3})$, and in general

$$x_{ijk} = \pi_{j a_{ij1} a_{ij2} \dots a_{ijk-1}}(a_{ijk}),$$

where $\pi_{j a_{ij1} a_{ij2} \dots a_{ijk-1}}$ is a random permutation of $\{0, 1, \dots, b - 1\}$. The permutations are mutually independent uniform random permutations.

The following geometrical description may help the reader visualize this scrambling. The rule for choosing x_{ij1} is like cutting the unit cube into b equal (congruent) parts along the X^j axis and then reassembling these parts in random order to reform the cube. The rule for choosing x_{ij2} is like cutting the unit cube into b^2 equal parts along the X^j axis, taking them as b groups of b consecutive parts, and reassembling the b parts within each group in random order. The rule for x_{ijk} involves cutting the cube into b^k equal parts along the X^j axis, forming b^{k-1} groups of b equal parts, and reassembling the b parts within each group in random order.

The sequence (X_i) inherits certain equidistribution properties of (A_i) and the individual points in it are uniformly distributed on $[0, 1]^s$. Owen (1995, 1997) proves the following two propositions.

PROPOSITION 1. *If (A_i) is a (λ, t, m, s) -net in base b then (X_i) is a (λ, t, m, s) -net in base b with probability 1.*

PROPOSITION 2. *Let A be a point in $[0, 1]^s$ and let X be the scrambled version of A as described above. Then X has the uniform distribution on $[0, 1]^s$.*

2.3. *ANOVA and wavelet decompositions of $L^2[0, 1]^s$.* For $f, g \in L^2[0, 1]^s$, we take the usual inner product $\langle f, g \rangle = \int f(X)g(X) dX$ and norm $\|f\| = \langle f, f \rangle^{1/2}$.

It is possible to construct an ANOVA decomposition of f over $[0, 1]^s$ that mimics the usual ANOVA decomposition used for the discrete product domains widely used in experimental design. This functional ANOVA appears in Efron and Stein (1981), and in an operator form, in Wahba (1990). Hickernell (1996a) presents a whole family of ANOVAs useful for studying properties of numerical integration. The version below is from Owen (1992).

For $u \subseteq \mathcal{A}$, define the effect of u by

$$\alpha_u = \int \left(f - \sum_{v \subset u} \alpha_v \right) dX^{\mathcal{A}-u},$$

where the sum is over strict subsets $v \neq u$. For $u = \emptyset$ this produces $\alpha_u = I$. For $|u| > 0$, the subeffects α_v are first subtracted from f and then the residual is averaged over the coordinate axes that are not in u . The result is a function α_u defined on $[0, 1]^u$ which extends to $[0, 1]^s$ by $\alpha_u(X) = \alpha_u(X^u)$.

The ANOVA functions are analogous to the usual discrete ANOVA. Here $I = \alpha_{\emptyset}$ is the analogue of the grand mean, $\alpha_{\{j\}}$ is the analogue of the main effect for variable j and for $|u| > 1$, α_u is the analogue of the interaction among $X^j, j \in u$.

The following properties will also be familiar: $j \in u$ implies $\int \alpha_u dX^j = 0$, $u \neq v$ implies $\int \alpha_u \alpha_v dX = 0$, $f = \sum_{u \subseteq \mathcal{A}} \alpha_u$ and $\int (f - I)^2 dX = \sum_{|u| > 0} \sigma_u^2$ where $\sigma_u^2 = \int \alpha_u^2 dX$.

A further decomposition of the ANOVA is needed to study the variance of scrambled nets. Owen (1997) presents a base b Haar-like multiresolution analysis of $[0, 1]^s$ using ideas from Jawerth and Sweldens (1994), Daubechies (1992) and Madych (1992).

For $s = 1$, define the functions

$$\psi_c(x) = b^{1/2} 1_{[bx]=c} - b^{-1/2} 1_{[x]=0}, \quad c = 0, 1, \dots, b - 1.$$

The functions ψ_c are the mother wavelets in this multiresolution. They feature a narrow positive spike of width b^{-1} centered over $(c + 0.5)/b$ and they integrate to zero. Now for integers $k \geq 0$ and $0 \leq t < b^k$, define dilated and translated versions

$$\begin{aligned} \psi_{ktc}(x) &= b^{k/2} \psi_c(b^k x - t) \\ &= b^{(k+1)/2} 1_{[b^{k+1}x]=bt+c} - b^{(k-1)/2} 1_{[b^kx]=t} \end{aligned}$$

of these wavelets. Because the dilation is by factors of b it would be usual to have $b - 1$ mother wavelets in the family. Many readers will be familiar with diadic ($b = 2$) wavelets where only one mother wavelet is used. The Haar wavelets correspond to base $b = 2$ and just one mother wavelet. Here b wavelets are used, in order to preserve some symmetry. Notice that the ψ_c are not orthogonal:

$$\langle \psi_c, \psi_{c'} \rangle = 1_{c=c'} - b^{-1}.$$

More generally

$$\langle \psi_{ktc}, \psi_{k't'c'} \rangle = 1_{k=k'} 1_{t=t'} b^k (1_{c=c'} - b^{-1}).$$

All of the ψ_{ktc} integrate to zero, so they can only provide a basis for functions that integrate to zero. Let us therefore also use the function $\phi(x) = 1$ on $[0, 1]$. Owen [(1997), Section 6] develops the representation

$$f(x) = \langle \phi, f \rangle \phi(x) + \sum_{k \geq 0} \sum_{0 \leq t < b^k} \langle \psi_{ktc}, f \rangle \psi_{ktc}(x)$$

for square integrable f . Notice that although the $\psi_{k_{tc}}$ are not orthogonal the representation takes the same simple form that it would for an orthogonal basis. See Owen (1997), Lemma 3.

This decomposition extends to $[0, 1]^s$ by taking tensor products. For each $u \subseteq \mathcal{A}$, take a product over $j \in u$ of scaled and translated wavelets. Such a product requires $|u|$ scales k_j , $j \in u$ and similarly $|u|$ translations t_j and $|u|$ wavelets c_j . It is convenient to bundle these together into vectors of $|u|$ components, κ for the k_j , τ for the t_j and γ for the c_j . Then define

$$\psi_{u\kappa\tau\gamma}(X) = \psi_{u\kappa\tau\gamma}(X^u) = \prod_{j \in u} \psi_{k_j t_j c_j}(X^j).$$

When $u = \emptyset$ we take by convention $\psi_{u\kappa\tau\gamma}(X) = \psi_{\emptyset(\emptyset)(\emptyset)}(X) = 1$ where the subscripts $()$ denote vectors of zero components. This convention is natural as empty products are usually taken to be 1. The multiresolution decomposition of $f \in L^2[0, 1]^s$ is

$$f(X) = \sum_u \sum_\kappa \sum_\tau \sum_\gamma \langle \psi_{u\kappa\tau\gamma}, f \rangle \psi_{u\kappa\tau\gamma}(X).$$

This sum is taken over all subsets $u \subseteq \mathcal{A}$, over all $|u|$ vectors κ of nonnegative integers k_j , over all $|u|$ vectors τ of integers t_j with $0 \leq t_j < b^{k_j}$ and over all $|u|$ vectors γ of integers c_j with $0 \leq c_j < b$.

We can recover the usual functional ANOVA via

$$\alpha_u(X) = \sum_\kappa \sum_\tau \sum_\gamma \langle \psi_{u\kappa\tau\gamma}, f \rangle \psi_{u\kappa\tau\gamma}(X).$$

Owen (1997) introduces the terms

$$\nu_{u,\kappa}(X) = \sum_\tau \sum_\gamma \langle \psi_{u\kappa\tau\gamma}, f \rangle \psi_{u\kappa\tau\gamma}(X).$$

The function $\nu_{u,\kappa}$ is a step function, constant within elementary intervals

$$\prod_{j \in u} \left[\frac{a_j}{b^{k_j+1}}, \frac{a_j + 1}{b^{k_j+1}} \right)$$

for integers $0 \leq a_j < b^{k_j+1}$. Thus $\nu_{u,\kappa}$ is constant within each of $\prod_{j \in u} b^{k_j+1} = b^{|u|+|\kappa|}$ elementary intervals where $|\kappa|$ denotes $\sum_{j \in u} k_j$. Here $a_j = bt_j + c_j$. The $\nu_{u,\kappa}$ are mutually orthogonal because

$$\langle \psi_{u\kappa\tau\gamma}, \psi_{u'\kappa'\tau'\gamma'} \rangle = 1_{u=u'} 1_{\kappa=\kappa'} 1_{\tau=\tau'} b^{|\kappa|} \prod_{j \in u} (1_{c_j=c'_j} - b^{-1}).$$

2.4. *Variance over scrambled nets.* Suppose that X_i is obtained by scrambling the base b digits of A_i for $i = 1, \dots, n$. Then by Theorem 2 of Owen (1997), the variance of $\hat{I} = n^{-1} \sum_{i=1}^n f(X_i)$ is

$$(2.1) \quad \frac{1}{n} \sum_{|u|>0} \sum_\kappa \Gamma_{u,\kappa} \sigma_{u,\kappa}^2,$$

where $\sigma_{u,\kappa}^2 = \int \nu_{u,\kappa}^2(X) dX$ and $\Gamma_{u,\kappa}$ is determined by balance properties among A_1, \dots, A_n . The constants $\Gamma_{u,\kappa}$ are interpreted as “gains” that multiply the variance contribution of $\nu_{u,\kappa}$. If all of the $\Gamma_{u,\kappa}$ are equal to one, then the variance is the same as under simple Monte Carlo integration with independent $X_i \sim U[0, 1)^s$.

The value of $\Gamma_{u,\kappa}$ depends on the number and arrangement of matches among the various base b digits of the A_i^j . It follows from Section 7 of Owen (1997) that

$$(2.2) \quad \Gamma_{u,\kappa} = \frac{1}{n(b-1)^{|u|}} \sum_{i=1}^n \sum_{i'=1}^n \prod_{j \in u} (b \mathbf{1}_{\lfloor b^{k_j+1} A_i^j \rfloor = \lfloor b^{k_j+1} A_{i'}^j \rfloor} - \mathbf{1}_{\lfloor b^{k_j} A_i^j \rfloor = \lfloor b^{k_j} A_{i'}^j \rfloor}).$$

By following some counting arguments, Owen (1997) shows that if A_i are a $(\lambda, 0, m, s)$ -net in base b and $m \geq |u| + |\kappa|$, then $\Gamma_{u,\kappa} = 0$. This is intuitively obvious because $\nu_{u,\kappa}$ is constant within each of $b^{|u|+|\kappa|}$ elementary intervals and by placing the same number of points in all of those intervals, the scrambled $(\lambda, 0, m, s)$ -net therefore integrates $\nu_{u,\kappa}$ without error. Further counting arguments show that $\Gamma_{u,\kappa} = 1$ for $|\kappa| > m$ (or $|\kappa| = m$ and $\lambda = 1$). Thus very coarse effects do not contribute to the integration variance and very fine effects contribute the same as under simple Monte Carlo.

When $|u| > 1$ or $\lambda > 1$, there are also some intermediate effects for which Γ can be larger or smaller than one. In this case (2.2) reduces, for $(\lambda, 0, m, s)$ -nets, to

$$(2.3) \quad \Gamma_{u,\kappa} = (b-1)^{-|u|} \sum_{l=0}^{|u|} \binom{|u|}{l} b^l (-1)^{|u|-l} \max(1, \lambda b^{m-|\kappa|-l})$$

which can be reexpressed as

$$(2.4) \quad \Gamma_{u,\kappa} = 1 + (b-1)^{-|u|} (-1)^{|u|} \times \left[\lambda b^{m-|\kappa|} \binom{|u|-1}{m-|\kappa|} (-1)^{m-|\kappa|} - \sum_{l=0}^{m-|\kappa|} \binom{|u|}{l} (-b)^l \right].$$

For $(0, m, s)$ -nets, that is, for $\lambda = 1$,

$$0 \leq \Gamma_{u,\kappa} \leq \left(\frac{b}{b-1} \right)^{\min(s-1, m)} \leq \left(\frac{b}{b-1} \right)^{b-1} \leq e \doteq 2.718$$

holds by Theorem 3 of Owen (1997). The second inequality is tight, being attained for $m < s$ at $|u| = m + 1, |\kappa| = 0$ and for $m \geq s$ at $|u| = s, |\kappa| = m - s + 1$. The third inequality holds because a $(0, m, s)$ -net can only exist for $b \geq s$. Here we improve that theorem as follows.

THEOREM 1. *In the above notation,*

$$\Gamma_{u,\kappa} \leq \left(\frac{b}{b-1} \right)^{\min(s-1, m)} \leq e \doteq 2.718$$

for a scrambled $(0, m, s)$ -net in base b and

$$\Gamma_{u, \kappa} \leq 1 + e \doteq 3.718$$

for a scrambled $(\lambda, 0, m, s)$ -net in base b .

PROOF. Theorem 3 of Owen (1997) covers the first claim and so we need only consider the second claim for the case where $2 \leq \lambda \leq b - 1$ and $m - |u| < |\kappa| \leq m$. If $m - |\kappa| + |u|$ is odd then Γ is decreasing in λ . Therefore it is maximized at $\lambda = 1$, a case already covered by Theorem 3 of Owen (1997).

Suppose that $m - |\kappa| + |u|$ is even. Then, using (2.4),

$$\begin{aligned} \Gamma_{u, \kappa} &\leq 1 + (b - 1)^{-|u|} \lambda b^{m - |\kappa|} \binom{|u| - 1}{m - |\kappa|} \\ &\leq 1 + (b - 1)^{1 - |u|} b^{m - |\kappa|} \binom{|u| - 1}{m - |\kappa|} \\ &\leq 1 + (b - 1)^{1 - |u|} b^{|u| - 1} \\ &\leq 1 + (b / (b - 1))^{b - 1} \\ &\leq 1 + e \doteq 3.718. \end{aligned}$$

The first inequality follows because the excluded sum is negative by an argument based on alternating signs and monotone magnitudes of its terms. The third inequality follows because for $b \geq |u| - 1$ the expression $b^{m - |\kappa|} \binom{|u| - 1}{m - |\kappa|}$ is increasing in $m - |\kappa|$. \square

The second bound in Theorem 1 does not appear to be tight. Numerical investigations for $s \leq 100$ find $\Gamma_{u, \kappa}$ no larger than 2.678, and so it may be true that $\Gamma_{u, \kappa} \leq e$ for all u, κ and all $(\lambda, 0, m, s)$ -nets.

2.5. Elementary lemmas. This subsection gathers some elementary calculus results that are used below. Readers uninterested in the details of the proofs to follow can safely skip this subsection. Rules C1, C2 and C3 apply to univariate functions whose first derivative satisfies a Lipschitz condition.

RULE C1. Suppose that $|f'(x) - f'(x^*)| \leq B|x - x^*|^\beta$ for some finite $B \geq 0$ and $\beta \in (0, 1]$, for all real x, x^* . Then

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + C|x - x^*|^{1 + \beta},$$

where $|C| \leq B(1 + \beta)^{-1} \leq B$.

PROOF. Without loss of generality, $x^* \leq x$. Now

$$\begin{aligned} f(x) &= f(x^*) + \int_{x^*}^x f'(v) dv \\ &= f(x^*) + f'(x^*)(x - x^*) + \int_{x^*}^x f'(v) - f'(x^*) dv \end{aligned}$$

and

$$\begin{aligned} \left| \int_{x^*}^x f'(v) - f'(x^*) \, dv \right| &\leq \int_{x^*}^x |f'(v) - f'(x^*)| \, dv \\ &\leq B \int_{x^*}^x |v - x^*|^\beta \, dv \\ &= B(1 + \beta)^{-1} |x - x^*|^{1+\beta}. \quad \square \end{aligned}$$

RULE C2. Let f be as in Rule C1. Then for $a_0 < a_1$ and $b_0 < b_1$,

$$\begin{aligned} (a_1 - a_0)^{-1} \int_{a_0}^{a_1} f(x) \, dx - (b_1 - b_0)^{-1} \int_{b_0}^{b_1} f(x) \, dx \\ = (\bar{a} - \bar{b}) f'(\bar{b}) + C_{ab} |\bar{a} - \bar{b}|^{1+\beta} + C_a (a_1 - a_0)^{1+\beta} + C_b (b_1 - b_0)^{1+\beta}, \end{aligned}$$

where $\bar{a} = (a_0 + a_1)/2$, $\bar{b} = (b_0 + b_1)/2$ and $|C_a|, |C_b| \leq B2^{-1-\beta}(1 + \beta)^{-1}(2 + \beta)^{-1} \leq B/4$ and $|C_{ab}| \leq B(1 + \beta)^{-1} < B$.

PROOF. Taking $x^* = \bar{a}$ in C1 and integrating the first term over x from a_0 to a_1 yields

$$(a_1 - a_0)^{-1} \int_{a_0}^{a_1} f(x) \, dx = f(\bar{a}) + C_a (a_1 - a_0)^{1+\beta},$$

where $|C_a| \leq B2^{-1-\beta}(1 + \beta)^{-1}(2 + \beta)^{-1}$. A similar result holds for the integral from b_0 to b_1 with constant $|C_b| \leq B2^{-1-\beta}(1 + \beta)^{-1}(2 + \beta)^{-1}$. Applying C1 to the difference of the integrals, with $x^* = \bar{b}$ and $x = \bar{a}$, yields the result. \square

RULE C3. Let f be as in Rule C1, for $x, x^* \in [0, 1]$. Then

$$\frac{1}{n} \sum_{t=0}^{n-1} f'(\bar{x}_t)^2 = \int_0^1 f'(x)^2 \, dx + O(n^{-\beta}),$$

where $\bar{x}_t = (t + 0.5)/n$.

PROOF. Note that $f'(x) \leq |f'(0)| + B$ for $0 \leq x \leq 1$. Now

$$\begin{aligned} \left| \int_0^1 f'(x)^2 \, dx - \frac{1}{n} \sum_{t=0}^{n-1} f'(\bar{x}_t)^2 \right| \\ \leq \sum_{t=0}^{n-1} \int_{t/n}^{(t+1)/n} |f'(x) + f'(\bar{x}_t)| |f'(x) - f'(\bar{x}_t)| \, dx \\ \leq 2(|f'(0)| + B) \sum_{t=0}^{n-1} \int_{t/n}^{(t+1)/n} B|x - x_t|^\beta \, dx \\ \leq 2B(|f'(0)| + B)n^{-\beta}. \quad \square \end{aligned}$$

3. Smooth integrands, $s = 1$. The one-dimensional case is much simpler to understand and so it is worth describing as a special case. We further simplify by taking $\lambda = 1$. That is, we suppose that X_i are a scrambled $(0, m, 1)$ -net in base b . Then each interval of the form $[tb^{-m}, (t+1)b^{-m})$, $0 \leq t < b^m$ contains exactly one of the X_i , let us call it Z_t , and the Z_t are independent random variables having the uniform distribution over the interval $[tb^{-m}, (t+1)b^{-m})$. Let $\bar{x}_t = (t + 0.5)/n$.

Suppose that $|f'(x) - f'(x^*)| \leq B|x - x^*|^\beta$ for finite $B \geq 0$ and $\beta \in (0, 1]$. Then by Rule C1, $f(Z_t) = f(\bar{x}_t) + f'(\bar{x}_t)(Z_t - \bar{x}_t) + C_t|Z_t - \bar{x}_t|^{1+\beta}$ for a random variable C_t satisfying $|C_t| \leq B$. Now

$$\begin{aligned} V(\hat{I}) &= n^{-2} \sum_{t=0}^{n-1} V(f'(\bar{x}_t)(Z_t - \bar{x}_t) + C_t|Z_t - \bar{x}_t|^{1+\beta}) \\ &= n^{-2} \sum_{t=0}^{n-1} f'(\bar{x}_t)^2 (12n^2)^{-1} + E_1 + E_2, \end{aligned}$$

where

$$|E_1| \leq 2Bn^{-2} \sum_{t=0}^{n-1} |f'(\bar{x}_t)| E(|Z_t - \bar{x}_t|^{2+\beta}) = O(n^{-3-\beta})$$

and

$$|E_2| \leq B^2 n^{-2} \sum_{t=0}^{n-1} E(|Z_t - \bar{x}_t|^{2+2\beta}) = O(n^{-3-2\beta}).$$

Thus by Rule C3,

$$\begin{aligned} V(\hat{I}) &= \frac{1}{12n^3} \left(\int_0^1 f'(x)^2 dx + O(n^{-\beta}) \right) + O(n^{-3-\beta}) \\ &= \frac{1}{12n^3} \int_0^1 f'(x)^2 dx + O(n^{-3-\beta}). \end{aligned}$$

The result is integration with errors of order $n^{-3/2}$ in probability. This is a big improvement over the corresponding $n^{-1/2}$ rate for ordinary Monte Carlo, but is not as good as some basic one-dimensional integration schemes. For example, a midpoint rule attains $O(n^{-1-\beta})$ accuracy, which for $\beta > 1/2$ is better than scrambled nets.

This rate $n^{-3/2}$ in probability is better than the worst case rate n^{-1} obtained for unrandomized nets. Interestingly, the worst-case rate is commonly seen in applications of one dimensional nets, which are usually generalizations of the van der Corput sequence. Fred Hickernell has provided a convincing explanation of this phenomenon. The unscrambled van der Corput schemes are, for $n = b^m$, essentially rectangle rules with the data points taken at the left ends of each of n intervals of width $1/n$. With large n and well-behaved f , most intervals will have either their largest or smallest value of f at the left endpoint. It cannot be assumed that the resulting positive and negative errors will cancel. The error in such a rule is typically of order $1/n$.

Scrambled nets take the data points at random in each of these intervals and so there is a tendency for some cancellation to take place among the errors from each interval.

A worst-case integrand for a scrambled net might be one which takes its minimum over the interval $[t/n, (t + 1)/n)$ at the point Z_t randomly generated by the scrambled net. It is unduly pessimistic to expect such integrands in practice.

4. Smooth integrands, $s \geq 1$. The situation for $s > 1$ is more complicated. A simple analysis based on breaking $[0, 1]^s$ into n subcubes does not capture all of the balance properties in the net. The analysis below proceeds through the multiresolution described in Section 2.

We have that

$$V(\hat{I}) = \frac{1}{n} \sum_{|u|>0} \sum_{\kappa} \Gamma_{u,\kappa} \sigma_{u,\kappa}^2,$$

where

$$\sigma_{u,\kappa}^2 = \int \nu_{u,\kappa}^2(X) dX.$$

The coefficients $\Gamma_{u,\kappa}$ are zero if n is a multiple of $b^{|u|+|\kappa|}$ and otherwise, for $n = \lambda b^m$, they are between 0 and $1 + e$. Therefore

$$V(\hat{I}) = O\left(\frac{1}{n} \sum_{|u|>0} \sum_{|\kappa|>m-|u|} \sigma_{u,\kappa}^2\right),$$

and so a rate of convergence for $V(\hat{I})$ can be attained by finding a rate of decay for $\sigma_{u,\kappa}^2$.

The proof goes as follows: Lemma 1 gives an approximation to $\langle f, \psi_{u\kappa\tau\gamma} \rangle$ for smooth f . Using this formula, Lemma 2 shows that $\sigma_{u,\kappa}^2$ is of order $b^{-2|\kappa|}$ for $|\kappa| > m - |u|$. Thus the largest $\sigma_{u,\kappa}^2$ contributing to $V(\hat{I})$ is of order n^{-2} . At first sight this would suggest that $V(\hat{I}) = O(n^{-3})$, but the number of terms $\sigma_{u,\kappa}^2$ for a given value $|\kappa|$ grows with $|\kappa|$ and the result is that $V(\hat{I}) = O(n^{-3}(\log n)^{s-1})$.

For the purposes of this paper a smooth function is one for which the order s mixed partial derivative satisfies the following Lipschitz condition.

DEFINITION 5. A real-valued function f on $[0, 1]^s$ is smooth if there exists finite $B \geq 0$ and $\beta \in (0, 1]$ such that

$$\left| \frac{\partial^s}{\partial X} f(x) - \frac{\partial^s}{\partial X} f(x^*) \right| \leq B \|x - x^*\|^\beta,$$

where $\|z\|$ is the Euclidean norm of z .

REMARK 1. If f is smooth, then every mixed partial derivative $\partial^{|u|}f/\partial X^u$ with $|u| \leq s$ satisfies a Lipschitz condition with the same β and a possibly larger B . To show this for $|u| = s - 1$, connect a point x^* to x by s steps each

of which changes exactly one coordinate and bound the difference over each step. Make this argument $s - |u|$ times for the general case. In particular, f itself satisfies a Lipschitz condition with the same β .

REMARK 2. If f is smooth, then every ANOVA effect α_u , regarded as a function on $[0, 1]^s$, is also smooth. Clearly this is true for the constant function α_{\emptyset} . For the general case, write $\alpha_u(X^u) = \int_{Z^u=X^u} f(Z) dZ^{s-u} - \sum_{v \subset u} \alpha_v(X^u)$. Trivially, $\partial^{|u|} \alpha_v / \partial X^u = 0$. Finally

$$\frac{\partial^{|u|} \alpha_u(X)}{\partial X^u} = \frac{\partial^{|u|}}{\partial X^u} \int_{Z^u=X^u} f(Z) dZ^{s-u} = \int_{Z^u=X^u} \frac{\partial^{|u|}}{\partial X^u} f(Z) dZ^{s-u}$$

which satisfies the Lipschitz condition because $\partial^{|u|} f / \partial X^u$ does. Below we use the equivalence

$$(4.1) \quad \frac{\partial^s \alpha_{\mathcal{A}}}{\partial X} = \frac{\partial^s f}{\partial X}.$$

Before proving the rate of convergence, we introduce some univariate indicator functions N^j and W^j through

$$\begin{aligned} \psi_{u\kappa\tau\gamma}(X) &= \prod_{j \in u} \psi_{k_j t_j c_j}(X^j) \\ &= \prod_{j \in u} b^{(k_j+1)/2} (N^j(X^j) - b^{-1}W^j(X^j)) \\ &= b^{(|\kappa|+|u|)/2} \prod_{j \in u} (N^j(X^j) - b^{-1}W^j(X^j)), \end{aligned}$$

where $N^j(x) = \mathbf{1}_{\lfloor b^{k_j+1}x \rfloor = bt_j + c_j}$ is the indicator function of the j th “narrow” interval and $W^j(x) = \mathbf{1}_{\lfloor b^{k_j}x \rfloor = t_j}$ is the indicator function of the j th “wide” interval. Let $\omega = \omega_{u, \kappa, \tau}$ denote the point in $[0, 1]^u$ with coordinates equal to the centers of all the wide intervals. That is $\omega^j = b^{-k_j}(t_j + 0.5)$. We also use

$$\bar{c} = \frac{1}{b} \sum_{c=0}^{b-1} c = \frac{b-1}{2}$$

and $d_j = d_j(\gamma) = (c_j - \bar{c})/b$.

The error magnitude $O(\sum_{j \in u} b^{-(1+\beta)k_j})$ appears often in the calculations below. For fixed s , it is equivalent to $O(b^{-(1+\beta)\min_{j \in u} k_j})$. For brevity we write this as $O(b^{-(1+\beta)\underline{\varepsilon}})$ within proofs.

LEMMA 1. *Suppose that f is smooth; then*

$$\langle f, \psi_{u\kappa\tau\gamma} \rangle = b^{-(3|\kappa|+|u|)/2} \frac{\partial^{|u|} \alpha_u(\omega_{u, \kappa, \tau})}{\partial X^u} \left(\prod_j d_j \right) \left(1 + O\left(\sum_{j \in u} b^{-(1+\beta)k_j} \right) \right).$$

PROOF. The proof uses $|u|$ applications of Rule C2 to approximate the inner product with $\psi_{u\kappa\tau\gamma}$:

$$\begin{aligned} \langle f, \psi_{u\kappa\tau\gamma} \rangle &= \langle \alpha_u, \psi_{u\kappa\tau\gamma} \rangle \\ &= \int \alpha_u(X) \psi_{u\kappa\tau\gamma}(X) dX^u \\ &= b^{-(|\kappa|+|u|)/2} \int \prod_{j \in u} b^{k_j+1} (N^j(X^j) - b^{-1}W^j(X^j)) \alpha_u(X) dX^u \\ &= b^{-(|\kappa|+|u|)/2} \frac{\partial^{|u|} \alpha_u(\omega_u, \kappa, \tau)}{\partial X^u} \left(\prod_j \frac{c_j - \bar{c}}{b^{k_j+1}} \right) (1 + O(b^{-(1+\beta)\kappa})) \\ &= b^{-(3|\kappa|+|u|)/2} \frac{\partial^{|u|} \alpha_u(\omega_u, \kappa, \tau)}{\partial X^u} \left(\prod_j d_j \right) (1 + O(b^{-(1+\beta)\kappa})). \quad \square \end{aligned}$$

LEMMA 2. Suppose f is smooth, then:

$$\sigma_{u, \kappa}^2 = b^{-2|\kappa|} \left(\frac{b^2 - 1}{12b^2} \right)^{|u|} \left\| \frac{\partial^{|u|} \alpha_u}{\partial X^u} \right\|^2 \left(1 + O\left(\sum_{j \in u} b^{-(1+\beta)k_j} \right) \right).$$

PROOF. The supports of $\psi_{u\kappa\tau\gamma}$ and $\psi_{u\kappa\tau'\gamma'}$ are disjoint unless $\tau = \tau'$ and so

$$\begin{aligned} \nu_{u, \kappa}^2 &= \sum_{\tau} \sum_{\gamma} \sum_{\tau'} \sum_{\gamma'} \langle f, \psi_{u\kappa\tau\gamma} \rangle \langle f, \psi_{u\kappa\tau'\gamma'} \rangle \psi_{u\kappa\tau\gamma} \psi_{u\kappa\tau'\gamma'} \\ &= \sum_{\tau} \sum_{\gamma} \sum_{\gamma'} \langle f, \psi_{u\kappa\tau\gamma} \rangle \langle f, \psi_{u\kappa\tau\gamma'} \rangle \psi_{u\kappa\tau\gamma} \psi_{u\kappa\tau\gamma'}. \end{aligned}$$

Applying Lemma 1 and letting $d'_j = d_j(\gamma')$,

$$\begin{aligned} \nu_{u, \kappa}^2 &= b^{-3|\kappa|-|u|} \sum_{\tau} \left(\frac{\partial^{|u|} \alpha_u(\omega_u, \kappa, \tau)}{\partial X^u} \right)^2 \sum_{\gamma} \sum_{\gamma'} \psi_{u\kappa\tau\gamma} \psi_{u\kappa\tau\gamma'} \left(\prod_j d_j \right) \left(\prod_j d'_j \right) \\ &\quad \times (1 + O(b^{-(1+\beta)\kappa})). \end{aligned}$$

Now

$$\begin{aligned} &\sum_{\gamma} \sum_{\gamma'} \int \psi_{u\kappa\tau\gamma} \psi_{u\kappa\tau\gamma'} dX \left(\prod_j d_j \right) \left(\prod_j d'_j \right) \\ &= \sum_{\gamma} \sum_{\gamma'} \prod_j (1_{c_j=c'_j} - b^{-1}) \left(\prod_j d_j \right) \left(\prod_j d'_j \right) \\ &= \left(\sum_{c=0}^{b-1} \sum_{c'=0}^{b-1} (1_{c=c'} - b^{-1})(c - \bar{c})(c' - \bar{c})/b^2 \right)^{|u|} \\ &= \left(\frac{b^2 - 1}{12b} \right)^{|u|}. \end{aligned}$$

Therefore from the definition of $\sigma_{u, \kappa}^2$ and making use of Rule C3,

$$\begin{aligned} \sigma_{u, \kappa}^2 &= b^{-3|\kappa|} \left(\frac{b^2 - 1}{12b^2} \right)^{|u|} \sum_{\tau} \left(\frac{\partial^{|u|} \alpha_u(\omega_{u, \kappa, \tau})}{\partial X^u} \right)^2 (1 + O(b^{-(1+\beta)\underline{\kappa}})) \\ &= b^{-3|\kappa|} \left(\frac{b^2 - 1}{12b} \right)^{|u|} \left(b^{|\kappa|} \int \left(\frac{\partial^{|u|} \alpha_u(X)}{\partial X^u} \right)^2 dX^u (1 + O(b^{-(1+\beta)\underline{\kappa}})) \right)^2 \\ &\quad \times (1 + O(b^{-(1+\beta)\underline{\kappa}})) \\ &= b^{-2|\kappa|} \left(\frac{b^2 - 1}{12b^2} \right)^{|u|} \left\| \frac{\partial^{|u|} \alpha_u}{\partial X^u} \right\|^2 (1 + O(b^{-(1+\beta)\underline{\kappa}})). \quad \square \end{aligned}$$

The constants implicit in the error terms of Lemmas 1 and 2 depend on the smoothness constants, B and β for each α_u . These lemmas become ever sharper as $\underline{\kappa} = \min_{j \in u} k_j$ increases. Where $\underline{\kappa}$ vanishes or is bounded, the lemmas are less sharp, reducing to statements that the expression on the left-hand side of the equality is bounded, uniformly in κ , by some multiple of the lead expression on the right-hand side. This is sufficient accuracy to get the rate, but not the lead constant, in Theorem 2 below.

THEOREM 2. *Suppose that f is smooth, and X_1, \dots, X_n are points of a scrambled $(\lambda, 0, m, s)$ -net in base b . Then:*

$$V(\hat{I}) = O(n^{-3}(\log n)^{s-1}).$$

PROOF. For $n = \lambda b^m$, by Lemma 2 we have,

$$\begin{aligned} V(\hat{I}) &= O \left(\frac{1}{n} \sum_{|u|>0} \left\| \frac{\partial^{|u|} \alpha_u}{\partial X^u} \right\|^2 \left(\frac{b^2 - 1}{12b^2} \right)^{|u|} \sum_{|\kappa| \geq (m - |u| + 1)_+} b^{-2|\kappa|} \right) \\ &= O \left(\frac{1}{n} \sum_{|u|>0} \left\| \frac{\partial^{|u|} \alpha_u}{\partial X^u} \right\|^2 \left(\frac{b^2 - 1}{12b^2} \right)^{|u|} \sum_{l=(m - |u| + 1)_+}^{\infty} b^{-2l} \binom{|u| + l - 1}{|u| - 1} \right) \end{aligned}$$

by counting the number of vectors κ with a given value of $|\kappa|$ and using $z_+ = \max(z, 0)$.

Since we are interested in large n , we suppose that $(m - |u| + 1)_+ = m - |u| + 1$ for all $|u| \leq s$. That is, $m \geq s - 1$, and so, using $l = m - |u| + r$,

$$\begin{aligned} V(\hat{I}) &= O \left(\frac{1}{n} \sum_{|u|>0} \left\| \frac{\partial^{|u|} \alpha_u}{\partial X^u} \right\|^2 \left(\frac{b^2 - 1}{12b^2} \right)^{|u|} b^{2|u| - 2m} \sum_{r=1}^{\infty} b^{-2r} \binom{m + r - 1}{|u| - 1} \right) \\ &= O \left(\frac{\lambda^2}{n^3} \sum_{|u|>0} \left\| \frac{\partial^{|u|} \alpha_u}{\partial X^u} \right\|^2 \left(\frac{b^2 - 1}{12} \right)^{|u|} \sum_{r=1}^{\infty} b^{-2r} \binom{m + r - 1}{|u| - 1} \right). \end{aligned}$$

It remains to examine the binomial coefficients as $m \rightarrow \infty$:

$$\begin{aligned} \sum_{r=1}^{\infty} b^{-2r} \binom{m+r-1}{|u|-1} &= \frac{1}{(|u|-1)!} \sum_{r=1}^{\infty} b^{-2r} \prod_{j=0}^{|u|-2} (m+r-1-j) \\ &\sim \frac{m^{|u|-1}}{(|u|-1)!} \sum_{r=1}^{\infty} b^{-2r} \\ &= \frac{m^{|u|-1}}{(b^2-1)(|u|-1)!} \\ &= \frac{(\log n)^{|u|-1}}{(\log b)^{|u|-1}(b^2-1)(|u|-1)!}. \end{aligned}$$

Therefore

$$V(\hat{I}) = O\left(\frac{(\log n)^{s-1}}{n^3}\right). \quad \square$$

Keeping the term of largest order, ignoring the magnitudes of the gains $\Gamma_{u,\kappa}$ and using (4.1), we get the approximation

$$(4.2) \quad V(\hat{I}) \doteq \frac{(\log n)^{s-1}}{n^3} \frac{\lambda^2}{12^s(s-1)!} \left(\frac{b^2-1}{\log b}\right)^{s-1} \left\| \frac{\partial^s f}{\partial X} \right\|^2.$$

The factor λ^2 reflects the extra efficiency of a $(0, m, s)$ -net over a $(\lambda, 0, m, s)$ -net with $\lambda > 1$. The approximation (4.2) is only good to within a constant multiple because terms with large $|\kappa|$ can still have $\sum_{j \in u} b^{-(1+\beta)k_j}$ of order 1. At this level of accuracy the λ^2 term could be ignored, because $1 \leq \lambda^2 \leq (b-1)^2$. But the computational example in Section 5 and simulations in Owen (1995) show a certain wavy effect in error versus sample size plots that can be explained by the λ^2 factor.

It is noteworthy that the value of β does not affect the rate of convergence. The example in the next section considers a function for which the mixed partial is constant, and this extra smoothness does not improve the rate of convergence. It is an open question whether the rate in Theorem 2 holds under weaker smoothness conditions.

5. Example. The integrand

$$(5.1) \quad f(X) = 12^{s/2} \prod_{j=1}^s (X^j - 0.5)$$

has integral $I = 0$ and variance $\sigma^2 = 1$ for any s . It has only s -dimensional structure because $\sigma_u^2 = 1_{|u|=s}$. Scrambled nets are based on piecewise constant approximations to f , and this f is multilinear. Thus f is not artificially easy for scrambled nets, at least in terms of the rates of convergence to be expected. Of course, this f is easier than would be a highly oscillatory function with only s -dimensional structure.

The various approximations used in Lemmas 1 and 2 have no error for this integrand. Thus

$$\langle f, \psi_{u_{\kappa\tau\gamma}} \rangle = 1_{|u|=s} 12^{s/2} b^{-(3|\kappa|+s)/2} \prod_j d_j$$

and

$$\sigma_{u, \kappa}^2 = 1_{|u|=s} b^{-2|\kappa|} \left(\frac{b^2 - 1}{b^2} \right)^s.$$

For $n = \lambda b^m$,

$$V(\hat{I}) = \frac{1}{n} \left(\frac{b^2 - 1}{b^2} \right)^s \sum_{|\kappa| \geq (m-s+1)_+} \Gamma_{\mathcal{A}, \kappa} b^{-2|\kappa|}.$$

This variance formula has been evaluated numerically for $s = 1, \dots, 10$, b equal to all prime powers between s and 11 inclusive, and all $n = \lambda b^m$ from $n = 1$ to the smallest such n greater than or equal to 10^8 . Some of the variances are above the Monte Carlo variance but never by more than a factor of 2.331, so the root mean square error (RMSE) of scrambled nets was never more than 1.53 times as large as that of Monte Carlo.

Figures 1 through 3 show some of these calculations. In each figure the horizontal axis displays $\log_{10} n$ and the vertical axis displays $\log_{10} V(\hat{I})^{1/2}$. For each plot, reference lines $n^{-1/2}$ are given corresponding to the simple Monte Carlo RMSE for the example function and $n^{-3/2}$ corresponding to the asymptotic RMSE for the example function when $s = 1$. The reference lines meet at the upper left of each plot where $n = 1$ and the RMSE is 1.0. The complete set of RMSE values for scrambled nets plot as a nearly uniform gray haze (not shown) between the reference lines.

Figure 1 compares the RMSE values for $s = b = 4$ with a dashed third reference line corresponding to the asymptotic rate

$$\text{RMSE} \doteq \left(\frac{\lambda^2}{6} \left(\frac{15}{\log 4} \right)^3 \right)^{1/2} (\log n)^{3/2} n^{-3/2}$$

predicted by (4.2). The approximation is surprisingly accurate at least for $n \geq 4^4 = 256$.

Figure 2 shows results for dimensions $1 \leq s \leq 10$, using in each case the smallest prime power $b \geq s$. This choice usually, though not always, gives the best asymptotic discrepancy among unrandomized $(0, s)$ -sequences and seems to be a good choice for randomized nets as well. It is the smallest b for which a $(0, s)$ -sequence in base b exists. The points for $s = 1$, $b = 2$ overstrike the $n^{-3/2}$ reference line. For higher values of s , the line follows the Monte Carlo

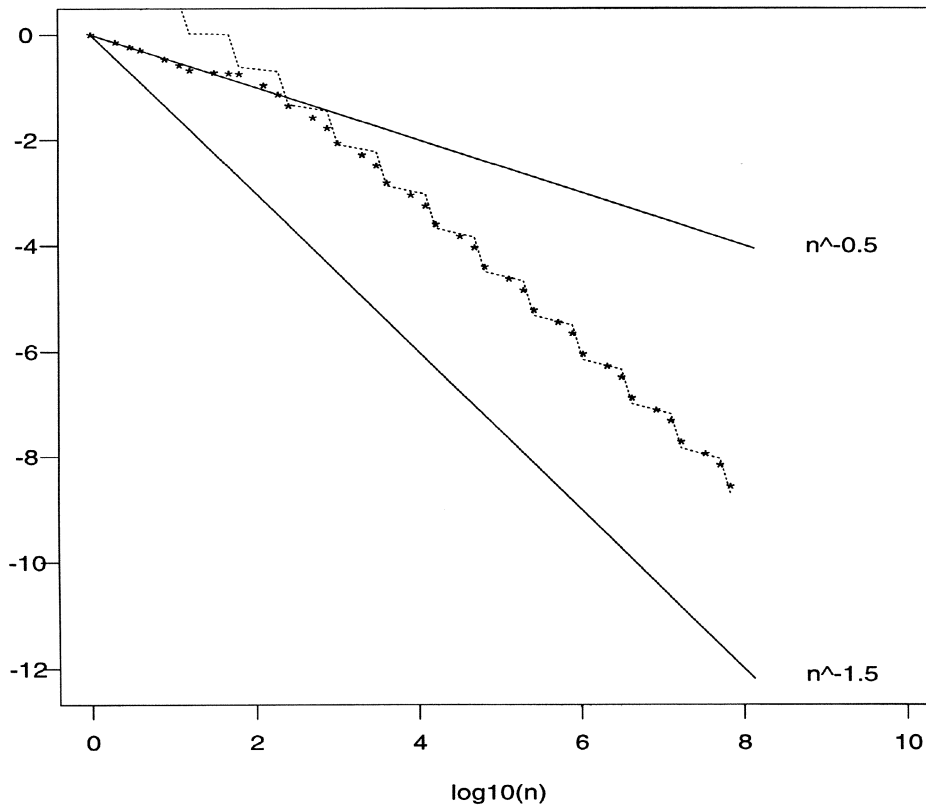


FIG. 1. This plot shows the root mean squared error (RMSE) of scrambled nets in dimension $s = 4$ generated with base $b = 4$, versus the sample size n . The exact RMSE values for integrating the multilinear function given in (5.1), are plotted as asterisks. The sample sizes n are all integers of the form $\lambda 4^m$ from 1 to the first such integer larger than 10^8 . The theoretical asymptote, proportional to $(\log n)^{3/2} n^{-3/2}$, is shown as a dashed line. Also shown are two solid reference lines, with $RMSE = n^{-1/2}$ corresponding to simple Monte Carlo and $RMSE = n^{-3/2}$ corresponding to scrambled nets in dimension $s = 1$.

RMSE $n^{-1/2}$ until $n = b^s$ at which point it appears to take on the rate $n^{-3/2}$, at least approximately. The lines for $s = 9$ and $s = 10$ do not break below the $n^{-1/2}$ line because the sample size is not sufficient. They crisscross each other, and near $n = 10^8$ it appears that the $s = 10$ -dimensional problem is slightly easier than the $s = 9$ -dimensional version.

Figure 3 shows the RMSE values for $1 \leq s \leq 10$ and $b = 11$. These curves are relevant for 10- or 11-dimensional integrands dominated by an s -dimensional multilinear component. The RMSE's for $s = 1$ return to the $n^{-3/2}$ line at powers of 11. In between powers of 11 the RMSE initially decreases parallel to the Monte Carlo reference curve, but remains within bands

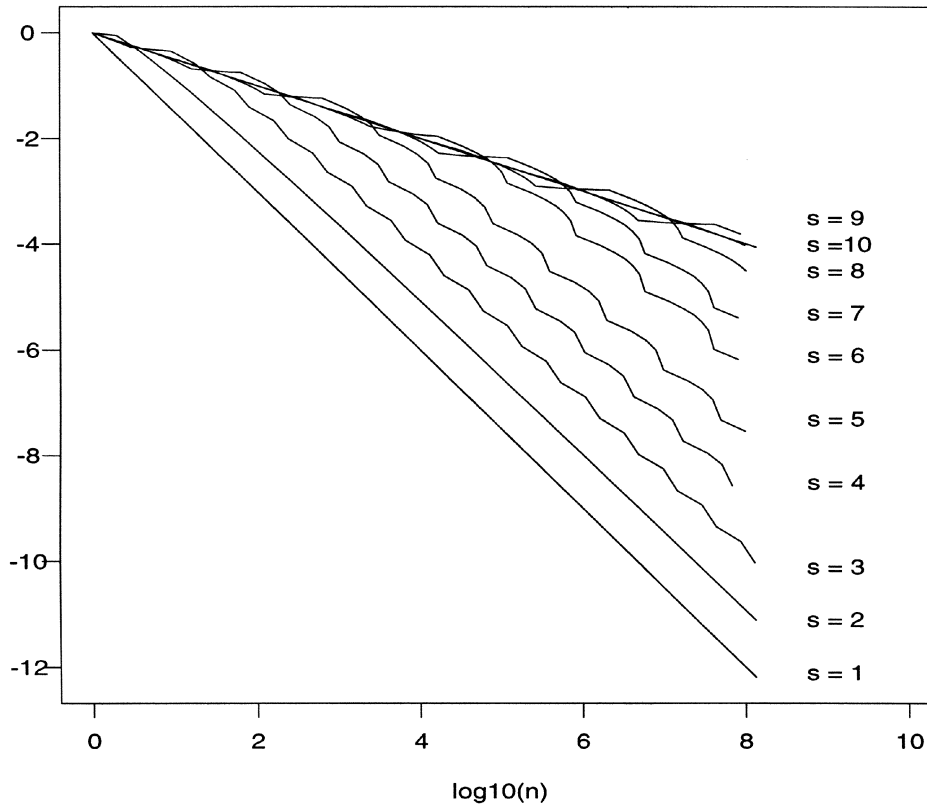


FIG. 2. This plot shows the RMSE for scrambled net integration of the multilinear integrand given in (5.1), versus the sample size n . For each dimension $s = 1, \dots, 10$, a curve is plotted of the exact RMSE for the first n points of a scrambled $(0, s)$ -sequence in base b , where b is the smallest prime power for which $b \geq s$. The sample sizes n are all integers of the form λb^m from 1 to the first such integer larger than 10^8 . Also shown are two solid reference lines, with $\text{RMSE} = n^{-1/2}$ corresponding to simple Monte Carlo and $\text{RMSE} = n^{-3/2}$ corresponding to scrambled nets in dimension $s = 1$. The dimension $s = 1$ values overstrike the $\text{RMSE} = n^{-3/2}$ line, and the $s = 9, 10$ values fluctuate around the $\text{RMSE} = n^{-1/2}$ line. The improvement over Monte Carlo sets in around $n = b^s$.

parallel to the $n^{-3/2}$ line. The curves for $s > 7$ do not break below the Monte Carlo reference curve in this data because the sample size is not sufficient.

There is no practical way to include the errors from unrandomized nets in the figures. The true error depends on which (t, m, s) -net one uses, not just on t, m, s , and b . Numerical enumeration is unsuitable for this reason, and is impractical for n as large as 10^8 . The widely quoted Koksma–Hlawka bound, using the best known constants in Niederreiter and Xing (1996), is not sharp enough. For example, even with $s = 4$ and $n \doteq 10^8$ this bound is larger than the Monte Carlo RMSE.

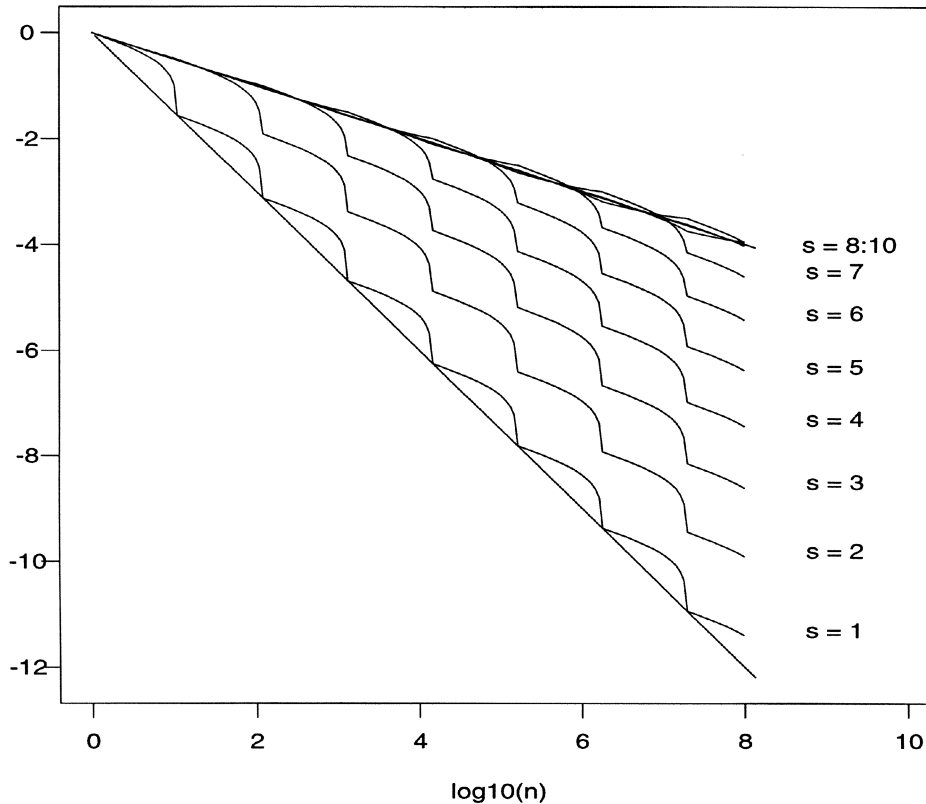


FIG. 3. This plot shows the RMSE of scrambled nets for the multilinear integrand given in equation (5.1), versus the sample size n . For each dimension $s = 1, \dots, 10$, a curve is plotted of the exact RMSE for the first n points of a scrambled $(0, s)$ -net in base $b = 11$. The sample sizes n are all integers of the form $\lambda 11^m$ from 1 to the first such integer larger than 10^8 . Also shown are two solid reference lines, with $RMSE = n^{-1/2}$ corresponding to simple Monte Carlo and $RMSE = n^{-3/2}$ corresponding to scrambled nets in dimension $s = 1$. The dimension $s = 1$ values overstrike the $RMSE = n^{-3/2}$ line, and the $s = 8, 9, 10$ values fluctuate around the $RMSE = n^{-1/2}$ line.

6. Discussion. From the proof of Theorem 2, the $n^{-3}(\log n)^{s-1}$ rate doesn't really set in until $m \geq s$ at the earliest. For smaller n the variance is a multiple of n^{-1} , though this multiple decreases each time n passes another power of b . Since we work with $b \geq s$ this means that the full benefits do not appear until $n \doteq s^s$. For $s \geq 10$ this is very expensive, but for $s \leq 8$ it is more acceptable. The reason that scrambled nets may be useful on high-dimensional problems is that the integrand f may have a lot of low-dimensional structure by which is meant large values of σ_u^2 for small $|u|$. The better rate of convergence starts to set in for all effects α_u of size $|u| \leq m$,

and if these dominate the integrand, then the result is a much smaller variance.

An effect like this was seen in Owen (1995) for scrambled net integration of some 10 dimensional integrands using $b = 11$ and $m \leq 4$. In some of those examples, taken from Genz (1984), scrambled nets had much greater accuracy than Monte Carlo and the accuracy ratio increased as n increased through the powers of 11. Had the integrands been fully 10-dimensional with no 1 through 4-dimensional structure, the scrambled net results would have been essentially the same as the ordinary Monte Carlo results, for the sample sizes used.

It is not necessary for f to be smooth for scrambled net Monte Carlo to show a big improvement over ordinary Monte Carlo. The reason is that α_u with $|u| < s$ can be smooth even when f is not, because α_u is defined through integrals of f . For example if f is 1 for $\sum_{j=1}^s (X^j - 0.5)^2 \leq 1$ and 0 otherwise, then all of the α_u are continuous except for the highest order one, α_s . If a high-dimensional function is dominated by smooth α_u for $|u| \leq m$ then scrambled net Monte Carlo should perform well.

The rate for scrambled net Monte Carlo is $n^{-3/2}(\log n)^{(s-1)/2}$ in probability while the rate for unscrambled nets is $n^{-1}(\log n)^{s-1}$ or $n^{-1}(\log n)^s$ along (t, s) sequences. Of course great care should be taken comparing these rates. The first rate is an average case result for a fixed function f , taken over random permutations. The other results describe the worst case over functions, for a fixed set of integration points. Because scrambled nets remain nets, the worst-case bounds also apply to them. Interestingly, by the argument of Hickernell mentioned in the end of Section 3, we can expect that for commonly encountered functions that the main effect $\alpha_{(1)}$ for the first input variable will be integrated by unscrambled nets at the rate n^{-1} and for large enough n this alone would cause the errors to be larger than the average case with scrambled nets. Owen (1995) observes that on some test integrands scrambling improves accuracy, but that other integrands show little difference. Furthermore, the staircase effect from the factor λ^2/n^3 appears there, and an even larger effect of this kind usually appears for the unscrambled nets.

The worst-case bounds typically involve the total variation of f in the sense of Hardy and Krause, as described in Niederreiter [(1992), Chapter 2]. In practice it is probably harder to estimate this total variation than to estimate the integral I . Thus the worst-case bounds are not useful for estimating or bounding $|\hat{I} - I|$ in practice. Owen (1997) describes two techniques to estimate the variance of scrambled net Monte Carlo. The first technique is to take a small number r of independent replicates of the scrambled sequence and to use the observed variation among the r answers obtained. The second technique breaks one randomized $(\lambda, 0, m, s)$ -net into a number of smaller nets and uses the observed variation among the answers from subnets.

Hickernell (1996b) has found that the average over permutations of a net, of the worst case over functions f , of the absolute error $|\hat{I} - I|$ decreases at the rate $n^{-1}(\log n)^{(s-1)/2}$. Thus in an extreme analysis, where the function f

is chosen pessimally after the random permutations have been drawn, no real improvement is obtained by scrambling.

Yet another kind of error has been studied in the complexity literature. This is the average case performance of integration with respect to random integrands f . Clearly the outcome depends on the distribution used for the random functions. Some well-known results are Woźniakowski (1991) who shows that when f is drawn from a Brownian sheet measure, the best possible rate is $n^{-1}(\log n)^{(s-1)/2}$ which the Hammersley sequence attains. Wasilkowski (1993) shows that $n^{-(1+s^{-1})/2}$ is the best possible rate for f drawn from isotropic Brownian motion. Hickernell (1996b) shows that scrambled nets attain the same best possible rate as the Hammersley sequence, for f drawn from Brownian sheet measure. Morokoff and Cafisch (1994) found a simpler proof of Woźniakowski's (1991) result. Paskov (1993) considers integrated Brownian sheets. Ritter (1995) includes an up-to-date survey of this area.

The rate $n^{-3/2}(\log n)^{(s-1)/2}$ obtained for scrambled net quadrature, assuming $\partial^s f / \partial X^u$ is Lipschitz with exponent β , is almost as good as the rate $n^{-3/2-\beta/s}$ that Ritter, Wasilkowski and Woźniakowski (1993) show can be obtained for f drawn from any mean 0 process for which *all* partial derivatives of order s exist in quadratic mean, and for which a certain second-order difference applied to the covariance kernel of any such derivative satisfies a Lipschitz condition of order 2β .

Most of this paper has considered the case of randomized $(\lambda, 0, m, s)$ -nets. For the case with $t > 0$ it is not possible to compute $\Gamma_{u, \kappa}$ from the defining properties of (λ, t, m, s) -nets. While λb^t points must be in each elementary interval of volume b^{t-m} , an elementary interval of volume b^{t-m-1} has somewhere between 0 and b^t points. For example, merging b^t identical copies of a $(0, m, s)$ -net produces a (t, m, s) -net and at the other extreme a (t, m, s) -net with $t > 0$ might also be a $(t-1, m, s)$ -net.

It is however clear that $m \geq |\kappa| + |u| + t$ implies that $\Gamma_{u, \kappa} = 0$. But for arbitrarily large $|\kappa|$, it is possible to have $\Gamma_{u, \kappa} = b^t$, corresponding to a b^t -fold $(\lambda, 0, m, s)$ -net. A similar factor of b^t appears in bounds for other $\Gamma_{u, \kappa}$. Because of this, the case with $t = 0$ appeared most promising. Recent work by Niederreiter and Xing (1995) appears to change the picture dramatically. Using global function fields, they have found new constructions of (t, s) -sequences in base 2, where for large s , t is nearly equal to s . This reduces the sample size at which the s -dimensional effect starts to be balanced from roughly s^s to $b^{t+s} \simeq 4^s$. This suggests that enormous improvements in Monte Carlo accuracy may be possible by scrambling these new nets. It is not yet clear how much improvement will be seen. The bounds in Owen (1997) suggest a multiplicative variance penalty of at most b^t . A penalty this large would reduce the practical importance of the new nets, but the true penalty might be smaller.

Acknowledgment. This paper was improved as a result of the comments of anonymous referees.

REFERENCES

- DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.
- EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596.
- GENZ, A. (1994). Testing multidimensional integration routines. In *Tools, Methods and Languages for Scientific and Engineering Computation* (B. Ford, J. C. Rault and F. Thomasset, eds.) 81–94. North Holland, Amsterdam.
- HICKERNELL, F. J. (1996a). Quadrature error bounds and figures of merit for quasi-random points. Technical Report 111, Dept. Mathematics, Hong Kong Baptist Univ.
- HICKERNELL, F. J. (1996b). The mean square discrepancy of randomized nets. Technical Report 112, Dept. Mathematics, Hong Kong Baptist Univ.
- JAWERTH, B. and SWELDENS, W. (1994). An overview of wavelet based multiresolution analyses. *SIAM Review* **30** 377–412.
- MADYCH, W. R. (1992). Some elementary properties of multiresolution analyses of $L^2(R^n)$. In *Wavelets: A Tutorial in Theory and Applications* (C. K. Chui, ed.) 259–294. Academic Press, 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. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia.
- NIEDERREITER, H. and XING, C. (1995). Low discrepancy sequences obtained from global function fields. In *Finite Fields and Applications* 241–273. Cambridge Univ. Press.
- OWEN, A. B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2** 439–452.
- OWEN, A. B. (1995). Randomly permuted (t, m, s) -nets and (t, s) -sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and J.-S. Shiu, eds.) 299–317. Springer, New York.
- OWEN, A. B. (1997). Monte Carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.* **34**. To appear.
- PASKOV, S. H. (1993). Average case complexity of multivariate integration for smooth functions. *J. Complexity* **9** 291–312.
- RITTER, K. (1995). Average case analysis of numerical problems. Ph.D. dissertation, Univ. Erlangen, Germany.
- RITTER, K., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1993). On multivariate integration for stochastic processes. In *Numerical Integration* (H. Brass and G. Hämmerlin, eds.) **112** 331–347. Birkhäuser, Basel.
- WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.
- WASILKOWSKI, G. W. (1993). Integration and approximation of multivariate functions: average case complexity with isotropic Wiener measure. *Bull. Amer. Math. Soc.* **28** 308–314. [Full version (1994) *J. Approx. Theory* **77** 212–227.]
- WOŹNIAKOWSKI, H. (1991). Average case complexity of multivariate integration. *Bull. Amer. Math. Soc.* **24** 185–194.

DEPARTMENT OF STATISTICS
 SEQUOIA HALL
 STANFORD UNIVERSITY
 STANFORD, CALIFORNIA 94305
 E-MAIL: owen@playfair.stanford.edu