Statistica Sinica **9**(1999), 451-473

VARIANCE OF QUADRATURE OVER SCRAMBLED UNIONS OF NETS

Rong-Xian Yue

Hong Kong Baptist University and East China Normal University

Abstract: Based on the work of Owen (1997a,b) who studied the variance of quadrature under a scrambled net with sample size $n = \lambda b^m$, this paper investigates scrambled sequences with sample sizes other than λb^m. First, the variance of quadrature under a scrambled sequence which is a union of two nets in base b is found. The scrambling schemes applied to the two nets can be independent or simultaneous. The results can be extended to the union of more than two nets. For finite sample sizes, the scrambled net-union variance is bounded by a small constant multiple of the Monte Carlo variance. Second, it is shown that for any Lipschitz integrand on [0,1), the variance is $O(n^{-3})$ for a scrambled net, and $O(n^{-3+\alpha})$ for a union of two scrambled nets in base b, for a certain $\alpha \in [0,1]$. For any multivariate smooth integrand on $[0,1)^s$, the scrambled net-union variance is $O(n^{-3+\alpha}(\log n)^{(s-1)1_{\alpha<1}})$ for a certain $\alpha \in [0,1]$. It turns out that adding some additional points may sometimes cause a large loss of efficiency.

Key words and phrases: Integration, multiresolution, quasi-Monte Carlo.

1. Introduction

We consider the problem of approximating an integral $I = \int_{[0,1)^s} f(\boldsymbol{x})d\boldsymbol{x}$ for $s \geq 1$ by the sample mean $\hat{I}_n = n^{-1}\sum_{i=1}^n f(\boldsymbol{x}_i)$, where the $\boldsymbol{x}_i$ are n points carefully chosen from the unit cube, $[0,1)^s$. We assume that f is square integrable, that is, $f \in \mathcal{L}^2[0,1)^s$.

For the high dimensional case, Monte Carlo methods and equidistribution or Quasi-Monte Carlo methods are most widely used. Under simple Monte Carlo methods, n points $\boldsymbol{x}_i$ are independently drawn from the uniform distribution on $[0,1)^s$, and so the estimator $\hat{I}_n$ is a random variable with mean I and variance σ^2/n where $\sigma^2 = \int_{[0,1)^s}[f(\boldsymbol{x}) - I]^2 d\boldsymbol{x}$. Thus the error of simple Monte Carlo integration is of order $n^{-1/2}$ in probability. Equidistribution methods use deterministic sequences of n points $\boldsymbol{x}_i$, such as good lattice points, (t,m,s)-nets and (t,s)-sequences, that are constructed to avoid gaps and clusters among the $\boldsymbol{x}_i$. See Hua and Wang (1981), Niederreiter (1992), Sloan and Joe (1994) and Fang and Wang (1994) for the general theory, applications, and further developments in equidistribution methods. These methods are usually more accurate than simple Monte Carlo methods, but it is much easier to estimate the accuracy under Monte Carlo methods. Owen (1995, 1997a) proposed a hybrid of these two techniques based on scrambling the digits in a (t, m, s)-net or (t, s)-sequence in base b. The resulting method provides unbiased estimates of I with a variance that is $o(n^{-1})$ along the sequence $n = \lambda b^m$, $1 \leq \lambda < b$, $0 \leq m$, for every integrand in $\mathcal{L}^2[0,1)^s$. Further, Owen (1997b) shows that under mild smoothness conditions on f, the variance of $\hat{I}_n$ for a scrambled $(\lambda, 0, m, s)$-net in base b is of order $n^{-3}(\log n)^{s-1}$ as $n = \lambda b^m \to \infty$.

Based on the work of Owen (1997a,b), one may ask about the variance of $\hat{I}_n$ for a scrambled sequence where the sample size n is not equal to λb^m, but is, for example, $\lambda_0 b^{m_0} + \lambda_1 b^{m_1}$. For such a sequence, of what order is the scrambled variance as $n \to \infty$? On the other hand, it is well known that under simple Monte Carlo methods adding additional points improves the accuracy of the estimate. Is this true for the estimate under scrambled equidistribution methods? In addition, do the results of Owen (1997b) hold for the integrands satisfying a weaker smoothness condition? This paper is concerned with these problems in understanding the base b scrambling scheme of Owen.

This paper proceeds as follows: in Section 2 we briefly review the equidistribution methods known as (t, m, s)-nets and (t, s)-sequences, and the related material in Owen (1997a). In Section 3 we derive the variance of $\hat{I}_n$ based on a union of two scrambled $(\lambda_j, 0, m_j, s)$-nets in base b, $j = 0, 1$. Such a variance is called a *scrambled net-union variance*. The sample size n used here is equal to $\lambda_0 b^{m_0} + \lambda_1 b^{m_1}$. We assume that $m_0 \geq m_1$ and $n < b^{m_0+1}$. It is reasonable to require the two nets to be disjoint and chosen from a $(0, m_0 + 1, s)$-net in base b, in order to preserve some equidistribution properties. The base b scrambling schemes applied to the two nets may be *independent* or *simultaneous*. For finite n the scrambled net-union variance is bounded by a small constant multiple of the simple Monte Carlo variance. Note that the results can be extended immediately to the union of more than two scrambled nets. In Section 4 we consider the order of the scrambled net-union variance as $n \to \infty$. For the one-dimensional case, we find the order of the scrambled net or net-union variance for any integrand satisfying a Lipschitz condition on [0,1). The results seem to be sharper than any similar result in the literature. For the multidimensional case, we find the order of the scrambled net-union variance for smooth integrands in the sense of Owen (1997b). Section 5 presents an example of an integrand for which it is possible to compute both the scrambled net-union variance and the simple Monte Carlo variance. The numerical results show that adding $n_1 = \lambda_1 b^{m_1}$ additional points to $n_0 = \lambda_0 b^{m_0}$ points may often inflate the variance and cause a large loss of efficiency for most $m_0 \leq s$ and $n_1 < n_0$. However, if $m_0 > s$ then adding additional points improves accuracy of the estimate. We also give a comparison between the variances under unions of independently and simultaneously scrambled nets, respectively. Section 6 contains discussion.

2. Preliminaries

This section briefly reviews background material. We begin with some notation. A typical point in the unit cube $[0,1)^s$ is denoted by $\boldsymbol{x} = (x^1, \ldots, x^s)$. A finite sequence of points is denoted by $\{\boldsymbol{x}_i\}_{i=1}^n$, an infinite sequence of points is denoted by $\{\boldsymbol{x}_i\}_{i\geq 1}$.

Let u be a subset of $\{1,\ldots,s\}$, $|u|$ be the cardinality of u, and $\bar{u}$ be the complement $\{1,\ldots,s\} - u$. By $[0,1)^u$ we denote the $|u|$-dimensional unit cube involving the coordinates in u, by $\boldsymbol{x}^u$ we denote the coordinate projection of $\boldsymbol{x}$ onto $[0,1)^u$, and $d\boldsymbol{x}^u = \prod_{r\in u} dx^r$.

An integer $b \geq 2$ is used throughout this paper as a base for representing points in $[0,1)$. Thus the rth component of a point $\boldsymbol{x}_i$ can be uniquely represented as $x_i^r = \sum_{k=1}^{\infty} x_{irk} b^{-k}$, where x_{irk} is an integer with $0 \leq x_{irk} < b$.

2.1. Equidistribution and its randomization

Here we briefly introduce equidistribution methods known as (t,m,s)-nets and (t,s)-sequences. See Niederreiter (1992) for other methods. In principle, equidistribution methods produce sequences $\{\boldsymbol{x}_i\}_{i=1}^n$ from $[0,1)^s$ such that the discrete uniform distribution on the $\boldsymbol{x}_i$ closely approximates the continuous uniform distribution on $[0,1)^s$.

An *elementary interval of* $[0,1)^s$ *in base* b is a set of the form

$$E = \prod_{r=1}^{s} \Big[\frac{\ell_r}{b^{k_r}}, \frac{\ell_r + 1}{b^{k_r}}\Big)$$

with integers $k_r \geq 0$, $0 \leq \ell_r < b^{k_r}$ for $1 \leq r \leq s$. Let t and m be nonnegative integers. A finite sequence $\{\boldsymbol{x}_i\}_{i=1}^n$ of points 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 stronger equidistribution properties of the net. When $t = 0$, every elementary interval of volume $1/n$ contains one of the n points in the sequence. We will confine our consideration to this particular case.

An infinite sequence $\{\boldsymbol{x}_i\}_{i\geq 1}$ of points in $[0,1)^s$ is a (t,s)-*sequence in base* b if for all integers $k \geq 0$ and $m \geq t$ the finite sequence $\{\boldsymbol{x}_i\}_{i=kb^m+1}^{(k+1)b^m}$ 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 $n = \lambda b^m$, $1 \leq \lambda < b$, so that all of the points used in $\hat{I}_{\lambda b^m}$ are also used in $\hat{I}_{(\lambda+1)b^m}$. Note that the initial

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

Let s, m, t, b, λ be integers with $s \geq 1$, $m \geq 0$, $0 \leq t \leq m$, $b \geq 2$ and $1 \leq \lambda < b$. A finite sequence $\{\boldsymbol{x}_i\}_{i=1}^n$ of points in $[0, 1)^s$ with $n = \lambda b^m$ 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.

From the above definitions, a (t, m, s)-net in base b is a $(1, t, m, s)$-net. If $\{\boldsymbol{x}_i\}_{i=1}^{b^{m+1}}$ is a (t, m, s)-net in base b, then $\{\boldsymbol{x}_i\}_{i=1}^{\lambda b^m}$ is a (λ, t, m, s)-net in base b, for $1 \leq \lambda < b$. In particular, if $\{\boldsymbol{x}_i\}_{i \geq 1}$ is a (t, s)-sequence in base b then $\{\boldsymbol{x}_i\}_{i=kb^{m+1}+1}^{kb^{m+1}+\lambda b^m}$ is a (λ, t, m, s)-net in base b, for integers $k \geq 0$ and $1 \leq \lambda < b$.

Numerical integration by averaging over the points of a (t, m, s)-net has an error $|\hat{I}_n - I| = O(n^{-1}(\log n)^{s-1})$, for integrands of bounded variation in the sense of Hardy and Krause. See Niederreiter (1992) for this result and some sharper versions of it.

The randomization of (t, m, s)-nets proposed by Owen (1995, 1997a, 1997b) preserves equidistribution properties of the nets. The randomization scheme can be briefly described as follows: Suppose that $\{\boldsymbol{a}_i\}_{i=1}^n$ is a (t, m, s)-net in base b. Write the components of $\boldsymbol{a}_i$ as $a_i^r = \sum_{k=1}^{\infty} a_{irk} b^{-k}$. For $i = 1, \ldots, n$, let $\boldsymbol{x}_i = (x_i^1, \ldots, x_i^s)$ with $x_i^r = \sum_{k=1}^{\infty} x_{irk} b^{-k}$, where x_{irk} is a random permutation applied to a_{irk}. The $\boldsymbol{x}_i$'s satisfy the following rules:

(1) Each digit x_{irk} is uniformly distributed on the set $\{0, 1, \ldots, b-1\}$;

(2) For any two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ the s pairs $(x_i^1, x_j^1), \ldots, (x_i^s, x_j^s)$ are mutually independent;

(3) If a_i^r and a_j^r share the same first k digits, but their $k+1^{\text{st}}$ digits are different, then

   (a) $x_{irh} = x_{jrh}$ for $h = 1, \ldots, k$;

   (b) the pair $(x_{ir\,k+1}, x_{jr\,k+1})$ is uniformly distributed on the set $\{(d_i, d_j) : d_i \neq d_j;\ d_i, d_j \in \{0, 1, \ldots, b-1\}\}$ and

   (c) $x_{ir\,k+2}, x_{ir\,k+3}, \cdots, x_{jr\,k+2}, x_{jr\,k+3}, \cdots$ are mutually independent.

We call this a base b scrambling scheme and call the sequence $\{\boldsymbol{x}_i\}_{i=1}^n$ a scrambled version of $\{\boldsymbol{a}_i\}_{i=1}^n$. A geometrical description of this scheme is given in Owen (1997b) which may help us visualize the randomization. Owen (1995, 1997a) proves the following two propositions.

Proposition 1. *If* $\{\boldsymbol{a}_i\}$ *is a* (λ, t, m, s)*-net in base* b*, then the scrambled version* $\{\boldsymbol{x}_i\}$ *is a* (λ, t, m, s)*-net in base* b *with probability* 1.

Proposition 2. *Let* $\boldsymbol{a}$ *be a point in* $[0, 1)^s$ *and* $\boldsymbol{x}$ *be the scrambled version of* $\boldsymbol{a}$ *as described above. Then* $\boldsymbol{x}$ *has the uniform distribution on* $[0, 1)^s$.

2.2. ANOVA and Haar-like decomposition of $\mathcal{L}^2[0,1)^s$

The ANOVA decomposition approach has been widely used in statistics and quadrature since it was introduced in Efron and Stein (1981); see Wahba (1990), Owen (1992), and Hickernell (1996). For $f \in \mathcal{L}^2[0,1)^s$, define $\alpha_\emptyset = \int_{[0,1)^s} f(\boldsymbol{x})d\boldsymbol{x} = I$ and

$$\alpha_u(\boldsymbol{x}) = \int_{[0,1)^{\bar{u}}} \Big[f(\boldsymbol{x}) - \sum_{v \subset u} \alpha_v(\boldsymbol{x})\Big] d\boldsymbol{x}^{\bar{u}}, \quad \emptyset \neq u \subseteq \{1, \ldots, s\},$$

where the sum is over strict subsets $v \neq u$. The crossed ANOVA decomposition of f is

$$f(\boldsymbol{x}) = \sum_{u \subseteq \{1,\ldots,s\}} \alpha_u(\boldsymbol{x}),$$

where the sum is over all 2^s subsets of $\{1, \ldots, s\}$. The following properties are well known:

$$\int_{[0,1)} \alpha_u dx^r = 0, \quad r \in u; \qquad \int_{[0,1)^s} \alpha_u \alpha_v d\boldsymbol{x} = 0, \quad u \neq v.$$

It follows that

$$\sigma^2 = \int_{[0,1)^s} (f - I)^2 d\boldsymbol{x} = \sum_{|u|>0} \int_{[0,1)^s} \alpha_u^2 d\boldsymbol{x} \equiv \sum_{|u|>0} \sigma_u^2,$$

which is the usual ANOVA decomposition.

Another kind of ANOVA decomposition, called nested ANOVA decomposition, is used to derive a formula for the variance over one-dimensional scrambled nets. The terms of the nested ANOVA are as follows: $\beta_0 = \int_{[0,1)} f(x)dx = I$, and

$$\beta_k(x) = b^k \int_{\lfloor b^k z \rfloor = \lfloor b^k x \rfloor} \Big[f(x) - \sum_{0 \le h < k} \beta_h(z)\Big] dz, \quad k \geq 1,$$

where $\lfloor z \rfloor$ denotes the greatest integer less than or equal to z. The equality $\lfloor b^k z \rfloor = \lfloor b^k x \rfloor$ means that z and x agree to $k \geq 0$ places past the decimal point in base $b \geq 2$. Each β_k is a constant on intervals of the form $[\ell b^{-k}, (\ell+1)b^{-k})$ for integers $0 \leq \ell < b^k$, that is, $\beta_k(x) = \beta_k(\lfloor b^k x \rfloor b^{-k})$. For $k \geq 1$, $\beta_k(x)$ may be expressed as

$$\beta_k(x) = b^k \int_{\lfloor b^k z \rfloor = \lfloor b^k x \rfloor} f(z)dz - b^{k-1} \int_{\lfloor b^{k-1} z \rfloor = \lfloor b^{k-1} x \rfloor} f(z)dz$$

with $\sum_{c=0}^{b-1} \beta_k((b\ell + c)b^{-k}) = 0$ for $0 \leq \ell < b^{k-1}$ and

$$\sum_{k=0}^{K} \beta_k(x) = b^K \int_{\lfloor b^K z \rfloor = \lfloor b^K x \rfloor} f(z)dz.$$

Furthermore f may be expressed as

$$f(x) = \sum_{k=0}^{\infty} \beta_k(x); \tag{2.1}$$

then

$$\sigma^2 = \int_{[0,1)} [f(x) - \beta_0]^2 dx = \sum_{k=1}^{\infty} \int_{[0,1)} [\beta_k(x)]^2 dx \equiv \sum_{k=1}^{\infty} \sigma_k^2. \tag{2.2}$$

As to the multidimensional case, Owen (1997a) develops a multivariate base b Haar-like multiresolution of $\mathcal{L}^2[0,1)^s$ using ideas from Jawerth and Sweldens (1994), Daubechies (1992) and Madych (1992). For $u \subseteq \{1,\ldots,s\}$, let $\boldsymbol{\kappa}$ be a vector of $|u|$ nonnegative integers k_r, $r \in u$, and let $|\boldsymbol{\kappa}|$ denote $\sum_{r\in u} k_r$. Then there are $b^{|u|+|\boldsymbol{\kappa}|}$ elementary intervals

$$E_{u,\boldsymbol{\kappa},\boldsymbol{\tau}} = \prod_{r\in u} \Big[\frac{\ell_r}{b^{k_r+1}}, \frac{\ell_r+1}{b^{k_r+1}}\Big),$$

where $\boldsymbol{\tau}$ is a $|u|$ vector of nonnegative integers $\ell_r < b^{k_r+1}$. Define

$$\nu_{\emptyset,0} = I, \quad \nu_{u,\boldsymbol{\kappa}}(\boldsymbol{x}) = \sum_{\boldsymbol{\tau}(u,\boldsymbol{\kappa})} \sum_{\boldsymbol{\gamma}(u)} \langle f, \psi_{u\boldsymbol{\kappa}\boldsymbol{\tau}\boldsymbol{\gamma}}\rangle \psi_{u\boldsymbol{\kappa}\boldsymbol{\tau}\boldsymbol{\gamma}}(\boldsymbol{x}), \tag{2.3}$$

where $\psi_{u\boldsymbol{\kappa}\boldsymbol{\tau}\boldsymbol{\gamma}}$ is

$$\psi_{u\boldsymbol{\kappa}\boldsymbol{\tau}\boldsymbol{\gamma}}(\boldsymbol{x}) = \prod_{r\in u} \Big(b^{(k_r+1)/2} 1_{\lfloor b^{k_r+1}x^r\rfloor = b\ell_r + c_r} - b^{(k_r-1)/2} 1_{\lfloor b^{k_r}x^r\rfloor = \ell_r}\Big),$$

$\langle\cdot,\cdot\rangle$ is the $\mathcal{L}^2$-inner product, and $\boldsymbol{\gamma}(u)$ is a $|u|$ vector of nonnegative integers $0 \le c_r < b$. Each $\nu_{u,\boldsymbol{\kappa}}$ is constant within each of $b^{|u|+|\boldsymbol{\kappa}|}$ elementary intervals $E_{u,\boldsymbol{\kappa},\boldsymbol{\tau}}$. Moreover, the $\nu_{u,\boldsymbol{\kappa}}$ are mutually orthogonal. The multiresolution decomposition of $f \in \mathcal{L}^2[0,1)^s$ is

$$f(\boldsymbol{x}) = I + \sum_{|u|>0} \sum_{\boldsymbol{\kappa}} \nu_{u,\boldsymbol{\kappa}}(\boldsymbol{x}). \tag{2.4}$$

The ANOVA decomposition in terms of the $\nu_{u,\boldsymbol{\kappa}}$ is

$$\sigma^2 = \int_{[0,1)^s} [f(\boldsymbol{x}) - I]^2 d\boldsymbol{x} = \sum_{|u|>0} \sum_{\boldsymbol{\kappa}} \int_{[0,1)^s} [\nu_{u,\boldsymbol{\kappa}}(\boldsymbol{x})]^2 d\boldsymbol{x} \equiv \sum_{|u|>0} \sum_{\boldsymbol{\kappa}} \sigma_{u,\boldsymbol{\kappa}}^2. \tag{2.5}$$

3. Variance over a Scrambled Union of Nets

Suppose that f is in $\mathcal{L}^2[0,1)^s$. Let P_{n_0} and P_{n_1} be two sequences of n_0 and n_1 points in $[0,1)^s$, respectively. We write $P_{n_0} = \{\boldsymbol{a}_i\}_{i=1}^{n_0}$ and $P_{n_1} = \{\tilde{\boldsymbol{a}}_j\}_{j=1}^{n_1}$.

By P_n we denote the union of the two sequences, $P_n = P_{n_0} \cup P_{n_1}$. Then the number of points in P_n is $n = n_0 + n_1$. The components of $\boldsymbol{a}_i$ and $\tilde{\boldsymbol{a}}_j$ are written $a_i^r = \sum_{k=1}^{\infty} a_{irk} b^{-k}$ and $\tilde{a}_j^r = \sum_{k=1}^{\infty} \tilde{a}_{jrk} b^{-k}$ for $r = 1, \ldots, s$, respectively. We consider two kinds of randomizations on P_n: One is to apply a scrambling scheme to each of P_{n_0} and P_{n_1} *independently*; another is to apply a single scrambling scheme to the n points of P_n, which implies that the randomizations performed on P_{n_0} and P_{n_1} are *not independent*. We say that the randomizations on P_{n_0} and P_{n_1} in the latter case are *simultaneous*.

Now let Q_{n_0} and Q_{n_1} be the scrambled versions of P_{n_0} and P_{n_1}, respectively. Let

$$Q_{n_0} = \{\boldsymbol{x}_i\}_{i=1}^{n_0}, \quad Q_{n_1} = \{\tilde{\boldsymbol{x}}_j\}_{j=1}^{n_1} \quad \text{and} \quad Q_n \equiv Q_{n_0} \cup Q_{n_1}.$$

Denote the estimates of the integral $I = \int f(\boldsymbol{x}) d\boldsymbol{x}$ based on Q_{n_0}, Q_{n_1} and Q_n by $\hat{I}_{n_0}$, $\hat{I}_{n_1}$ and $\hat{I}_n$, respectively, that is,

$$\hat{I}_{n_0} = \frac{1}{n_0} \sum_{i=1}^{n_0} f(\boldsymbol{x}_i), \quad \hat{I}_{n_1} = \frac{1}{n_1} \sum_{j=1}^{n_1} f(\tilde{\boldsymbol{x}}_j), \quad \hat{I}_n = \frac{1}{n} \Big[n_0 \hat{I}_{n_0} + n_1 \hat{I}_{n_1} \Big]. \tag{3.1}$$

Proposition 2 implies that these estimates are all unbiased. Now we consider the scrambled net-union variance. We have

$$\mathrm{Var}\,(\hat{I}_n) = \frac{1}{n^2} \Big[n_0^2 \mathrm{Var}\,(\hat{I}_{n_0}) + n_1^2 \mathrm{Var}\,(\hat{I}_{n_1}) + 2 n_0 n_1 \,\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1}) \Big]. \tag{3.2}$$

Since $\mathrm{Var}\,(\hat{I}_{n_j})$, $j = 0, 1$, has been investigated by Owen (1997a,b) for square integrable and smooth integrands, it is enough to find $\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1})$ for the case where the scrambling schemes performed on P_{n_0} and P_{n_1} are *not independent*. We shall only consider simultaneous scrambling and call the Q_{n_j}'s the simultaneously scrambled versions of the P_{n_j}'s. We will require that both P_{n_0} and P_{n_1} satisfy the following equidistribution properties:

Assumption 1. The P_{n_j}'s are $(\lambda_j, 0, m_j, s)$-nets in base b, $j = 0, 1$, with integers $0 \le m_1 \le m_0$, $1 \le \lambda_0, \lambda_1 < b$, and $n = \lambda_0 b^{m_0} + \lambda_1 b^{m_1} < b^{m_0+1}$.

Assumption 2. No elementary interval of volume b^{-m_0-1} contains more than one point of the union $P_n = P_{n_0} \cup P_{n_1}$.

Remark 1. Note that the union P_n may not be a net. However, if $m_1 = m_0$, then the two assumptions ensure that P_n is a $(\lambda_0 + \lambda_1, 0, m_0, s)$-net in base b. On the other hand, Assumption 2 implies that P_{n_0} and P_{n_1} are disjoint. Such nets P_{n_j}, $j = 0, 1$, may be disjoint pieces of a $(0, m_0 + 1, s)$-net in base b which is a subsequence of a $(0, s)$-sequence in base b.

3.1. One-dimensional case

First we consider the one-dimensional case. It is assumed that the integrand f is in $\mathcal{L}^2[0,1)$. Two sequences of points in [0,1) are written $P_{n_0}=\{a_i\}_{i=1}^{n_0}$ and $P_{n_1}=\{\tilde{a}_j\}_{j=1}^{n_1}$. We write

$$a_i=\sum_{k=1}^{\infty}a_{ik}b^{-k}\quad\text{and}\quad\tilde{a}_j=\sum_{k=1}^{\infty}\tilde{a}_{jk}b^{-k}$$

for $i=1,\ldots,n_0$ and $j=1,\ldots,n_1$. At first, we do not assume that both P_{n_0} and P_{n_1} have any nontrivial equidistribution properties. We treat the case in which the base b scrambling schemes are *simultaneously* applied to both P_{n_0} and P_{n_1}. Write the corresponding simultaneously scrambled versions by $Q_{n_0}=\{x_i\}_{i=1}^{n_0}$ and $Q_{n_1}=\{\tilde{x}_j\}_{j=1}^{n_1}$, respectively.

Using the nested ANOVA of f given by (2.1) and Proposition 2, we have

$$\mathrm{Cov}\,(\hat{I}_{n_0},\hat{I}_{n_1})=\frac{1}{n_0n_1}\sum_{i=1}^{n_0}\sum_{j=1}^{n_1}\sum_{k>0}\sum_{h>0}E\{\beta_k(x_i)\beta_h(\tilde{x}_j)\}.$$

From Lemmas 1 and 2 of Owen (1997a), we get

$$E\{\beta_k(x_i)\beta_h(\tilde{x}_j)\}=0,\quad k\neq h$$

and

$$E\{\beta_h(\tilde{x}_j)|x_i\}=\beta_k(x_i)\Big(\frac{b}{b-1}1_{\lfloor b^ka_i\rfloor=\lfloor b^k\tilde{a}_j\rfloor}-\frac{1}{b-1}1_{\lfloor b^{k-1}a_i\rfloor=\lfloor b^{k-1}\tilde{a}_j\rfloor}\Big)$$

for all $x_i\in Q_{n_0}$ and $\tilde{x}_j\in Q_{n_1}$. It follows that

$$\begin{aligned}\mathrm{Cov}\,(\hat{I}_{n_0},\hat{I}_{n_1})&=\frac{1}{n_0n_1}\sum_{k>0}\sum_{i=1}^{n_0}\sum_{j=1}^{n_1}E[\beta_k(x_i)E\{\beta_h(\tilde{x}_j)|x_i\}]\\&=\frac{1}{n_0n_1}\sum_{k>0}\sum_{i=1}^{n_0}\sum_{j=1}^{n_1}\Big(\frac{b}{b-1}1_{\lfloor b^ka_i\rfloor=\lfloor b^k\tilde{a}_j\rfloor}-\frac{1}{b-1}1_{\lfloor b^{k-1}a_i\rfloor=\lfloor b^{k-1}\tilde{a}_j\rfloor}\Big)\sigma_k^2,\end{aligned}$$

where $\sigma_k^2=E\{[\beta_k(x_i)]^2\}=\int_{[0,1)^s}[\beta_k(x)]^2dx$, see (2.2). Define for $k\geq 0$

$$M_k^*=M_k^*(P_{n_0},P_{n_1})=\sum_{i=1}^{n_0}\sum_{j=1}^{n_1}1_{\lfloor b^kx_i\rfloor=\lfloor b^k\tilde{x}_j\rfloor}=\sum_{i=1}^{n_0}\sum_{j=1}^{n_1}1_{\lfloor b^ka_i\rfloor=\lfloor b^k\tilde{a}_j\rfloor}.\tag{3.3}$$

Note that the meaning of such an M_k^* is different from that of M_k defined by Owen (1997a), since $P_{n_0}\neq P_{n_1}$. $M_k^*(P_{n_0},P_{n_1})$ counts the number of times that an $x_i \in Q_{n_0}$ and an $\tilde{x}_j \in Q_{n_1}$ share the same first k (or more) digits, and reflects how close Q_{n_0} is to Q_{n_1}. For $k \geq 0$ define

$$G_k = \frac{bM_k^* - M_{k-1}^*}{b-1}. \tag{3.4}$$

We then obtain the following result:

Lemma 1. *Suppose that f is in $\mathcal{L}^2[0,1)$. Let Q_{n_j}, $j = 0,1$, be simultaneously scrambled versions of two sequences P_{n_j}, $j = 0,1$, and $\hat{I}_{n_0}$, $\hat{I}_{n_1}$, M_k^* and σ_k^2 be as described above. Then*

$$\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1}) = \frac{1}{n_0 n_1} \sum_{k>0} G_k \sigma_k^2. \tag{3.5}$$

Now assume that P_{n_j}, $j = 0,1$, satisfies Assumptions 1 and 2 with $s = 1$. For $0 \leq k \leq m_0$ and each $\tilde{a}_j \in P_{n_1}$ there are $\lambda_0 b^{m_0-k}$ points $a_i \in P_{n_0}$ with $\lfloor b^k a_i \rfloor = \lfloor b^k \tilde{a}_j \rfloor$. From the definition of M_k^* in (3.3) it follows that $M_k^* = n_1 \lambda_0 b^{m_0-k}$ for $0 \leq k \leq m_0$. For $k \geq m_0 + 1$ and each $\tilde{a}_j \in P_{n_1}$ no point in P_{n_0} shares the first k digits with $\tilde{a}_j \in P_{n_1}$, by Assumption 2. Therefore, (3.4) becomes

$$G_k = \begin{cases} -\dfrac{\lambda_0 n_1}{b-1}, & k = m_0 + 1, \\ 0, & \text{otherwise.} \end{cases}$$

It follows that

$$\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1}) = -\frac{\sigma_{m_0+1}^2}{b^{m_0}(b-1)} \equiv C_1(m_0). \tag{3.6}$$

Here, the subscript in C_1 indicates the one-dimensional case. Applying (3.6) and the result of Owen (1997a), we obtain the following theorem:

Theorem 1. *Suppose that P_{n_j} satisfies Assumptions 1 and 2 with $s = 1$, $j = 0,1$. Let Q_n be a union of simultaneously scrambled versions of P_{n_j}, $j = 0,1$, and $\hat{I}_n$ be defined as in (3.1). Then*

$$\mathrm{Var}\,(\hat{I}_n) = \frac{1}{n^2}\Big[n_0^2 V_1(\lambda_0, m_0) + n_1^2 V_1(\lambda_1, m_1) + 2n_0 n_1 C_1(m_0)\Big], \tag{3.7}$$

where $C_1(m_0)$ is given by (3.6), and

$$V_1(\lambda_j, m_j) = \frac{1}{\lambda_j b^{m_j}}\Big(\frac{b-\lambda_j}{b-1}\sigma_{m_j+1}^2 + \sum_{k \geq m_j+2} \sigma_k^2\Big). \tag{3.8}$$

Remark 2a. We can extend to a union of more than two simultaneously scrambled nets. Suppose that the P_{n_j}, $j = 0,1,\ldots,p$, are $(\lambda_j, 0, m_j, 1)$-nets in base b, with integers $0 \le m_p \le \cdots \le m_0$. Assume that each pair of P_{n_j} and P_{n_k}, $j<k$, satisfy Assumptions 1 and 2 with $s=1$. Let Q_n be a union of simultaneously scrambled versions of the P_{n_j}'s, and $\hat{I}_n$ be the estimate of I based on Q_n, where $n=\sum_{j=0}^{p}\lambda_j b^{m_j}$. Then we have

$$\mathrm{Var}\,(\hat{I}_n) = \frac{1}{n^2}\Big[\sum_{j=0}^{p} n_j^2 V_1(\lambda_j, m_j) + 2\sum_{j<k} n_j n_k C_1(m_j)\Big], \tag{3.9}$$

where $C_1(m_j)$ is given by (3.6), with m_0 repalced by m_j.

Remark 2b. Under the conditions given in Theorem 1 or Remark 2a, the variance of $\hat{I}_n$ based on a union of simultaneously scrambled nets should be smaller than that based on a union of independently scrambled nets, since the covariance in (3.6) is negative. Furthermore, the covariance for two simultaneously scrambled $(\lambda_j, 0, m_j, 1)$-nets in base b, $j=0,1$, only depends on the value of m_0 but not on λ_0, λ_1 and m_1. Here is a simple example. Suppose that $\{a_i\}_{i\ge 1}$ is a $(0,1)$-sequence in base $b=5$. Then the sequence $P_n=\{a_i\}_{i=1}^{125}$ is a $(0,3,1)$-net in base b. One may write

$$P_n = \{a_i\}_{i=1}^{100} \cup \{a_j\}_{j=101}^{125}, \tag{3.10a}$$

the union of a $(4,0,2,1)$-net and a $(1,0,2,1)$-net in base b. From (3.7), the simultaneously scrambled net-union variance is

$$\mathrm{Var}\,(\hat{I}_n) = \frac{1}{125}\sum_{k>3}\sigma_k^2. \tag{3.10b}$$

Alternatively, one may write

$$P_n = \{a_i\}_{i=1}^{75} \cup \{a_j\}_{j=76}^{100} \cup \{a_h\}_{h=101}^{125}, \tag{3.10c}$$

which is the union of a $(3,0,2,1)$-net, a $(1,0,2,1)$-net and a $(1,0,2,1)$-net in base b. Straightforward calculation in (3.9) yields the same scrambled variance for (3.10c) as that for (3.10a). On the other hand, the variance for the scrambled $(0,3,1)$-net in base b can be obtained from (3.8), which is identical to (3.10b).

Remark 2c. It is easy to see from (3.7) that

$$\mathrm{Var}\,(\hat{I}_n) \le \frac{1}{n}\sum_{k\ge m_1+1}\sigma_k^2 \le \frac{\sigma^2}{n}.$$

This means that the scrambled net-union variance is less than the simple Monte Carlo variance. On the other hand, it can be verified from (3.8) that

$$\mathrm{Var}\,(\hat{I}_{(\lambda_0+\lambda_1)b^{m_0}}) \le \mathrm{Var}\,(\hat{I}_{\lambda_0 b^{m_0}+\lambda_1 b^{m_1}}), \quad 0 \le m_1 \le m_0.$$

Furthermore, if m_0 tends to ∞ and m_1 is bounded or also tends to ∞, then we have $n\mathrm{Var}\,(\hat{I}_n) \to 0$ since $n_j\mathrm{Var}\,(\hat{I}_{n_j}) \leq \sigma^2$ and $n_j\mathrm{Var}\,(\hat{I}_{n_j}) \to 0$ as $m_j \to \infty$ (Owen (1997a)). It turns out that for any non-constant integrand f the ratio of the scrambled net-union variance to the simple Monte Carlo variance also tends to zero as $n \to \infty$.

3.2. Multidimensional case

Suppose that f is in $\mathcal{L}^2[0,1)^s$, $s > 1$. Let P_{n_j}, Q_{n_j} and $\hat{I}_{n_j}$, $\hat{I}_n$ be as described at the beginning of this section. We begin with the covariance of $\hat{I}_{n_0}$ and $\hat{I}_{n_1}$ for simultaneously scrambled versions Q_{n_0} and Q_{n_1}. At first, we do not assume that both P_{n_0} and P_{n_1} have any nontrivial equidistribution properties.

For each $\boldsymbol{a}_i = (a_i^1, \ldots, a_i^s) \in P_{n_0}$ and each $\tilde{\boldsymbol{a}}_j = (\tilde{a}_j^1, \ldots, \tilde{a}_j^s) \in P_{n_1}$, define

$$N_{ijrk_r} = 1_{\lfloor b^{k_r+1} a_i^r \rfloor = \lfloor b^{k_r+1} \tilde{a}_j^r \rfloor}, \quad W_{ijrk_r} = 1_{\lfloor b^{k_r} a_i^r \rfloor = \lfloor b^{k_r} \tilde{a}_j^r \rfloor}.$$

These are indicator functions designating "narrow" and "wide" matches, respectively, between the components a_i^r and $\tilde{a}_j^r$. For each $u \subseteq \{1, \ldots, s\}$ with $|u| > 0$ and $\boldsymbol{\kappa} = \boldsymbol{\kappa}(u)$, define

$$G_{u,\boldsymbol{\kappa}} = G_{u,\boldsymbol{\kappa}}(P_{n_0}, P_{n_1}) = \frac{1}{(b-1)^{|u|}} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \prod_{r \in u} (bN_{ijrk_r} - W_{ijrk_r}). \tag{3.11}$$

Lemma 2. *Let f be in $\mathcal{L}^2[0,1)^s$. Suppose that Q_{n_0} and Q_{n_1} are simultaneously scrambled versions of sequences P_{n_0} and P_{n_1}, respectively. Let $\hat{I}_{n_0}$, $\hat{I}_{n_1}$ and $G_{u,\boldsymbol{\kappa}}$ be as described above, and $\sigma_{u,\boldsymbol{\kappa}}^2$ be as in* (2.5). *Then we have*

$$\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1}) = \frac{1}{n_0 n_1} \sum_{|u|>0} \sum_{\boldsymbol{\kappa}(u)} G_{u,\boldsymbol{\kappa}} \sigma_{u,\boldsymbol{\kappa}}^2. \tag{3.12}$$

Proof. By using the multiresolution of f shown in (2.4), we have

$$\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1}) = \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \sum_{|u|>0} \sum_{\boldsymbol{\kappa}(u)} \sum_{|u'|>0} \sum_{\boldsymbol{\kappa}'(u')} E\{\nu_{u,\boldsymbol{\kappa}}(\boldsymbol{x}_i) \nu_{u',\boldsymbol{\kappa}'}(\tilde{\boldsymbol{x}}_j)\},$$

where $\nu_{u,\boldsymbol{\kappa}}$ and $\nu_{u',\boldsymbol{\kappa}'}$ are as in (2.3). From Lemmas 4 and 5 of Owen (1997a), we get

$$E\{\nu_{u,\boldsymbol{\kappa}}(\boldsymbol{x}_i) \nu_{u',\boldsymbol{\kappa}'}(\tilde{\boldsymbol{x}}_j)\} = 0$$

if $u \neq u'$ or $\boldsymbol{\kappa} \neq \boldsymbol{\kappa}'$, and

$$E\{\nu_{u,\boldsymbol{\kappa}}(\tilde{\boldsymbol{x}}_j) | \boldsymbol{x}_i\} = \nu_{u,\boldsymbol{\kappa}}(\boldsymbol{x}_i) \prod_{r \in u} \Big(\frac{b}{b-1} N_{ijrk_r} - \frac{1}{b-1} W_{ijrk_r} \Big).$$

It follows from the definition of $G_{u,\boldsymbol{\kappa}}$ in (3.11) that

$$\begin{aligned}\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1}) &= \frac{1}{n_0 n_1}\sum_{i=1}^{n_0}\sum_{j=1}^{n_1}\sum_{|u|>0}\sum_{\boldsymbol{\kappa}(u)} E\{\nu_{u,\boldsymbol{\kappa}}(\boldsymbol{x}_i)E[\nu_{u,\boldsymbol{\kappa}}(\tilde{\boldsymbol{x}}_j)|\boldsymbol{x}_i]\}\\ &= \frac{1}{n_0 n_1}\sum_{|u|>0}\sum_{\boldsymbol{\kappa}(u)} G_{u,\boldsymbol{\kappa}} E\{\nu_{u,\boldsymbol{\kappa}}(\boldsymbol{x}_i)\}^2.\end{aligned}$$

Then (3.12) follows from the definition of $\sigma^2_{u,\boldsymbol{\kappa}}$ in (2.5).

After straightforward calculation, $G_{u,\boldsymbol{\kappa}}$ can be expressed as

$$G_{u,\boldsymbol{\kappa}} = \frac{1}{(b-1)^{|u|}}\sum_{v\subseteq u}(-1)^{|u|-|v|}b^{|v|}\sum_{i=1}^{n_0}\sum_{j=1}^{n_1}\Big(\prod_{r\in v} N_{ijrk_r}\Big)\Big(\prod_{r\in u-v} W_{ijrk_r}\Big). \tag{3.13}$$

Now assume that P_{n_0} and P_{n_1} satisfy Assumptions 1 and 2. Then for a given $u \subset \{1,\ldots,s\}$ and $\boldsymbol{\kappa} = \boldsymbol{\kappa}(u)$ we can find a convenient expression for $G_{u,\boldsymbol{\kappa}}$. From the equidistribution properties, for each $\tilde{\boldsymbol{a}}_j \in P_{n_1}$, if $|\boldsymbol{\kappa}| + |v| \le m_0$, there are $\lambda_0 b^{m_0-|\boldsymbol{\kappa}|-|v|}$ points $\boldsymbol{a}_i \in P_{n_0}$ such that

$$\begin{cases} \lfloor b^{k_r+1}a_i^r\rfloor = \lfloor b^{k_r+1}\tilde{a}_j^r\rfloor & \text{for all } r \in v,\\ \lfloor b^{k_r}a_i^r\rfloor = \lfloor b^{k_r}\tilde{a}_j^r\rfloor & \text{for all } r \in u-v,\end{cases}$$

else no point in P_{n_0} meets these conditions. Then expression (3.13) becomes

$$\begin{aligned}G_{u,\boldsymbol{\kappa}} &= \frac{1}{(b-1)^{|u|}}\sum_{v\subseteq u}(-1)^{|u|-|v|}b^{|v|}n_1\lambda_0 b^{m_0-|\boldsymbol{\kappa}|-|v|}1_{|\boldsymbol{\kappa}|+|v|\le m_0}\\ &= \frac{(-1)^{|u|}}{(b-1)^{|u|}}\lambda_0 n_1 b^{m_0-|\boldsymbol{\kappa}|}\sum_{q=0}^{|u|}\binom{|u|}{q}(-1)^q 1_{|\boldsymbol{\kappa}|+|v|\le m_0}.\end{aligned}$$

Therefore, for $|\boldsymbol{\kappa}| \le m_0 - |u|$ or $|\boldsymbol{\kappa}| \ge m_0+1$ we have $G_{u,\boldsymbol{\kappa}} = 0$, and for $m_0 - |u| < |\boldsymbol{\kappa}| \le m_0$ we have

$$\begin{aligned}G_{u,\boldsymbol{\kappa}} &= \frac{(-1)^{|u|}}{(b-1)^{|u|}}\lambda_0 n_1 b^{m_0-|\boldsymbol{\kappa}|}\sum_{q=0}^{m_0-|\boldsymbol{\kappa}|}\binom{|u|}{q}(-1)^q\\ &= \frac{(-1)^{m_0+|u|-|\boldsymbol{\kappa}|}}{(b-1)^{|u|}}\lambda_0 n_1 b^{m_0-|\boldsymbol{\kappa}|}\binom{|u|-1}{m_0-|\boldsymbol{\kappa}|}\end{aligned}$$

and so (3.12) becomes

$$\begin{aligned}\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1}) &= \frac{1}{b^{m_0}}\sum_{|u|>0}\ \sum_{m_0-|u|<|\boldsymbol{\kappa}|\le m_0}\frac{(-1)^{m_0+|u|-|\boldsymbol{\kappa}|}}{(b-1)^{|u|}}b^{m_0-|\boldsymbol{\kappa}|}\binom{|u|-1}{m_0-|\boldsymbol{\kappa}|}\sigma^2_{u,\boldsymbol{\kappa}}\\ &\equiv C_s(m_0). \end{aligned} \tag{3.14}$$

Here the subscript in C_s stands for the s-dimensional case. Combining (3.14) with Corollary 4 of Owen (1997a), we get

Theorem 2. *Suppose that f is in $\mathcal{L}^2[0,1)^s$. Let P_{n_j}, $j = 0,1$, satisfies Assumptions* 1 *and* 2. *Then the variance for a union of simultaneously scrambled versions of the P_{n_j}'s is*

$$\mathrm{Var}\,(\hat{I}_n) = \frac{1}{n^2}\Big[n_0^2 V_s(\lambda_0, m_0) + n_1^2 V_s(\lambda_1, m_1) + 2n_0 n_1 C_s(m_0)\Big], \tag{3.15}$$

where $C_s(m_0)$ is given in (3.14), *and*

$$V_s(\lambda_j, m_j) = \frac{1}{\lambda_j b^{m_j}} \sum_{|u|>0} \Big[\sum_{m_j - |u| < |\boldsymbol{\kappa}| \le m_j} \Gamma_{u,\boldsymbol{\kappa}}(\lambda_j, m_j)\sigma^2_{u,\boldsymbol{\kappa}} + \sum_{|\boldsymbol{\kappa}| \ge m_j+1} \sigma^2_{u,\boldsymbol{\kappa}}\Big], \tag{3.16}$$

where

$$\Gamma_{u,\boldsymbol{\kappa}}(\lambda_j, m_j) = 1 + \frac{(-1)^{|u|}}{(b-1)^{|u|}}\Big[\lambda_j b^{m_j - |\boldsymbol{\kappa}|}\binom{|u|-1}{m_j - |\boldsymbol{\kappa}|}(-1)^{m_j - |\boldsymbol{\kappa}|} - \sum_{q=0}^{m_j - |\boldsymbol{\kappa}|}\binom{|u|}{q}(-b)^q\Big]. \tag{3.17}$$

Remark 2d. As in the case $s = 1$, under Assumptions 1 and 2, the covariance for two simultaneously scrambled $(\lambda_0, 0, m_0, s)$-net and $(\lambda_1, 0, m_1, s)$-net in base b with $m_1 \le m_0$ only depends on m_0 but not on λ_0, λ_1 and m_1. Moreover, since the terms

$$b^{m_0 - |\boldsymbol{\kappa}|}\binom{|u|-1}{m_0 - |\boldsymbol{\kappa}|}$$

are nondecreasing as $m_0 - |\boldsymbol{\kappa}|$ increases from 0 to $|u| - 1$, for large m_0 the most significant term in (3.14) occurs at $|\boldsymbol{\kappa}| = m_0 - |u| + 1$. It follows that for large m_0

$$C_s(m_0) \sim -\frac{1}{b^{m_0}} \sum_{|u|>0} \sum_{|\boldsymbol{\kappa}| = m_0 - |u| + 1} \Big(\frac{b}{b-1}\Big)^{|u|} \sigma^2_{u,\boldsymbol{\kappa}}.$$

That is, $C_s(m_0)$ is negative for large m_0. Therefore, for large m_0 the simultaneously scrambled net-union variance should be smaller than the independently scrambled net-union variance. Furthermore, we have the following bounds for $C_s(m_0)$:

$$-\frac{\sigma^2}{n_0} e \le -\frac{\sigma^2}{n_0}\Big(\frac{b}{b-1}\Big)^{s-1} \le C_s(m_0) \le \frac{\sigma^2}{n_0}\Big(\frac{b}{b-1}\Big)^{s-2} \le \frac{\sigma^2}{n_0} e. \tag{3.18}$$

On the other hand, formula (3.15) can be extended to the case where the estimate is based on a union of more than two simultaneously scrambled nets in base b.

Remark 2e. The constants $\Gamma_{u,\boldsymbol{\kappa}}$ are interpreted as "gains" that multiply the variance contribution of $\nu_{u,\boldsymbol{\kappa}}$. See Owen (1997a,b) for a full discussion of these gain factors. From Theorem 1 of Owen (1997b) and the bounds in (3.18), it is easy to show that

$$\mathrm{Var}\,(\hat{I}_n) \le \frac{\sigma^2}{n}(1+2e)$$

for a union of simultaneously scrambled nets in base b. That is, the simultaneously scrambled-net union variance is never more than a constant multiple the simple Monte Carlo variance for any square integrable f in any dimension.

4. The Order of the Variance

In this section, we consider the variance of the sample mean based on a union of two scrambled $(\lambda_0, 0, m_0, s)$-net and $(\lambda_1, 0, m_1, s)$-net in base b, as $n = \lambda_0 b^{m_0} + \lambda_1 b^{m_1} \to \infty$, with $m_1 \le m_0$. We do not need both nets to satisfy Assumption 2, and the scrambling schemes applied to both nets may or may not be independent. For the one-dimensional case we assume that f satisfies a Lipschitz condition on $[0,1)$, while for the multidimensional case we assume that f is smooth on $[0,1)^s$, as was done in Owen (1997b).

4.1. One-dimensional case

Here we take $\lambda_0 = \lambda_1 = 1$ for simplicity, that is, we assume that P_{n_0} and P_{n_1} are a $(0, m_0, 1)$-net and a $(0, m_1, 1)$-net in base b, respectively, with $m_1 \le m_0$ and $n_j = b^{m_j}$, $j = 0, 1$. Let Q_{n_j}, $j = 0, 1$, be scrambled versions of the P_{n_j}'s, which may or may not be independent. First we give the following lemma:

Lemma 3. *Suppose f satisfies the Lipschitz condition*

$$|f(x') - f(x'')| \le B|x' - x''| \tag{4.1}$$

for a finite $B \ge 0$ and any $x', x'' \in [0,1)$. Then the variance of $\hat{I}_n$ is $O(n^{-3})$ as $n = b^m \to \infty$.

Proof. Let $\{x_i\}_{i=1}^n$ be a scrambled $(0, m, 1)$-net in base b with $n = b^m$. Then from the definition of a net, each interval of the form $[\ell b^{-m}, (\ell+1)b^{-m})$, $0 \le \ell < b^m$, contains exactly one of the x_i. We denote it by z_ℓ. Then the z_ℓ are independent random variables with the uniform distribution over $[\ell b^{-m}, (\ell+1)b^{-m})$. Put $\bar{x}_\ell = (\ell + 0.5)/n$. Then we have

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} f(x_i) = \frac{1}{n}\sum_{\ell=0}^{n-1} f(z_\ell) = \frac{1}{n}\sum_{\ell=0}^{n-1} f(\bar{x}_\ell) + \frac{1}{n}\sum_{\ell=0}^{n-1}[f(z_\ell) - f(\bar{x}_\ell)].$$

It follows that

$$\begin{aligned}\mathrm{Var}\,(\hat{I}_n) &= \frac{1}{n^2}\sum_{\ell=0}^{n-1}\mathrm{Var}\,\{f(z_\ell) - f(\bar{x}_\ell)\}\\ &\leq \frac{1}{n^2}\sum_{\ell=0}^{n-1}E\{|f(z_\ell) - f(\bar{x}_\ell)|^2\}\\ &\leq \frac{B^2}{n^2}\sum_{\ell=0}^{n-1}E\{|z_\ell - \bar{x}_\ell|^2\} = \frac{B^2}{12n^3}. \end{aligned} \tag{4.2}$$

The last inequality follows from the Lipschitz condition (4.1), and the last equality holds since $\mathrm{Var}\,(z_\ell) = 1/(12n^2)$.

Remark 3. Owen (1997b) shows that for a smooth integrand f on $[0,1)$, (the derivative df/dx satisfies a Lipschitz condition on $[0,1)$), the variance over a scrambled $(0,m,1)$-net in base b is $O(n^{-3})$ as $n = b^m \to \infty$. Now Lemma 3 shows that this result is also true under a weaker condition on the integrand.

Theorem 3. *Suppose that f satisfies a Lipschitz condition on $[0,1)$. Consider the union of two scrambled $(0,m_j,1)$-nets in base b with $m_1 \leq m_0$. If there is a constant $\alpha \in [0,1]$ such that $(1-\alpha)m_0 - m_1$ is bounded as $m_0 \to \infty$, then*

$$\mathrm{Var}\,(\hat{I}_n) = O(n^{-3+\alpha})$$

as $n = b^{m_0} + b^{m_1} \to \infty$, regardless of whether the scrambling schemes are independent or not.

Proof. Note that

$$|\,\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1})| \leq [\mathrm{Var}\,(\hat{I}_{n_0})\mathrm{Var}\,(\hat{I}_{n_1})]^{1/2}. \tag{4.3}$$

From the expression for $\mathrm{Var}\,(\hat{I}_n)$ in (3.2) and the last inequality in (4.2) we have

$$\begin{aligned}\mathrm{Var}\,(\hat{I}_n) &\leq \frac{1}{n^2}\Big\{n_0^2\mathrm{Var}\,(\hat{I}_{n_0}) + n_1^2\mathrm{Var}\,(\hat{I}_{n_1}) + 2n_0n_1[\mathrm{Var}\,(\hat{I}_{n_0})\mathrm{Var}\,(\hat{I}_{n_1})]^{1/2}\Big\}\\ &\leq \frac{B^2}{12n^2}\Big[\frac{1}{n_0} + \frac{1}{n_1} + \frac{2}{(n_0n_1)^{1/2}}\Big]\\ &= \frac{B^2}{12n^{3-\alpha}}b^{(1-\alpha)m_0-m_1}(1+b^{m_1-m_0})^{1-\alpha}(1+b^{m_1-m_0}+2b^{(m_1-m_0)/2}).\end{aligned}$$

It follows from the boundedness of $(1-\alpha)m_0 - m_1$ and $m_1 - m_0 \leq 0$ that $\mathrm{Var}\,(\hat{I}_n) = O(n^{-3+\alpha})$ as $n = b^{m_0} + b^{m_1} \to \infty$.

Remark 4. If $m_1 = m_0 - d$ where d is positive and bounded as $n \to \infty$, then $\mathrm{Var}\,(\hat{I}_n) = O(n^{-3})$. If $m_1 = \lfloor \omega m_0 \rfloor$ for a fixed $0 \leq \omega \leq 1$, then $\mathrm{Var}\,(\hat{I}_n) = O(n^{-(2+\omega)})$. However, if m_1 is bounded, then $\mathrm{Var}\,(\hat{I}_n) = O(n^{-2})$.

4.2. Multidimensional case

For the multidimensional case, Owen (1997b) shows that if there exists finite $B \geq 0$ and $\beta \in (0,1]$ such that

$$\Big|\frac{\partial^s}{\partial \boldsymbol{x}} f(\boldsymbol{x}') - \frac{\partial^s}{\partial \boldsymbol{x}} f(\boldsymbol{x}'')\Big| \leq B\|\boldsymbol{x}' - \boldsymbol{x}''\|^\beta \tag{4.4}$$

for any $\boldsymbol{x}'$, $\boldsymbol{x}'' \in [0,1)^s$, where $\|\cdot\|$ is the Euclidean norm, then the variance of $\hat{I}_n$ based on a scrambled $(\lambda, 0, m, s)$-net in base b is $O(n^{-3}(\log n)^{s-1})$ as $n = \lambda b^m \to \infty$. High powers of $\log n$ might not be negligible until n is very large, so this raises the possibility that the scrambled net variance might be worse than the Monte Carlo variance for finite n.

Under the smoothness condition (4.4), we have the following theorem:

Theorem 4. *Suppose that f satisfies* (4.4). *Consider the union of two scrambled $(\lambda_j, 0, m_j, s)$-nets in base b with $1 \leq \lambda_j < b$ and $m_1 \leq m_0$. If there is a constant $\alpha \in [0,1]$ such that $(1-\alpha)m_0 - m_1$ is bounded as $m_0 \to \infty$, then*

$$\mathrm{Var}\,(\hat{I}_n) = \begin{cases} O(n^{-2}), & \text{if } \alpha = 1 \\ O(n^{-3+\alpha}(\log n)^{s-1}), & \text{if } \alpha < 1 \end{cases}$$

as $n = \lambda_0 b^{m_0} + \lambda_1 b^{m_1} \to \infty$, regardless of whether the scrambling schemes are independent or not.

Proof. Consider each of the three terms in expression (3.2). From Theorem 2 of Owen (1997b), for large n_0 we have

$$\frac{n_0^2}{n^2}\mathrm{Var}\,(\hat{I}_{n_0}) = O\left(\frac{(\log n_0)^{s-1}}{n^2 n_0}\right) = O\left(\frac{(\log n)^{s-1}}{n^3}\right). \tag{4.5}$$

Consider the second term in the right side of (3.2). Suppose that there is a constant $\alpha \in [0,1]$ such that $(1-\alpha)m_0 - m_1$ is bounded as $m_0 \to \infty$. If $\alpha = 1$, that is, m_1 is bounded as $m_0 \to \infty$, then n_1 is finite and so

$$\frac{n_1^2}{n^2}\mathrm{Var}\,(\hat{I}_{n_1}) \leq \frac{n_1^2}{n^2}\,\frac{\sigma^2}{n_1}(1+e) = O(n^{-2}), \tag{4.6a}$$

since $\mathrm{Var}\,(\hat{I}_{n_1}) \leq n_1^{-1}\sigma^2(1+e)$. If $\alpha < 1$, then $m_1 \to \infty$ as $m_0 \to \infty$, and so

$$\frac{n_1^2}{n^2}\mathrm{Var}\,(\hat{I}_{n_1}) = O\left(\frac{(\log n_1)^{s-1}}{n^2 n_1}\right) = O\left(\frac{(\log n)^{s-1}}{n^{3-\alpha}}\,\frac{n^{1-\alpha}}{n_1}\,\Big(\frac{\log n_1}{\log n}\Big)^{s-1}\right).$$

Note that

$$\frac{n^{1-\alpha}}{n_1} = \lambda_1^{-1} b^{(1-\alpha)m_0 - m_1}\Big(\lambda_0 + \lambda_1 b^{m_1 - m_0}\Big)^{1-\alpha}$$

and

$$\frac{\log n_1}{\log n} = \frac{\log \lambda_1 + m_1 \log b}{\log \lambda_0 + m_0 \log b} \; \frac{\log n_0}{\log n_0 + \log(1 + \lambda_0^{-1}\lambda_1 b^{m_1 - m_0})}$$

are both bounded for $m_1 \leq m_0$. It follows that

$$\frac{n_1^2}{n^2}\mathrm{Var}\,(\hat{I}_{n_1}) = O(n^{-3+\alpha}(\log n)^{s-1}). \tag{4.6b}$$

Furthermore, applying (4.3), (4.5), (4.6a) and (4.6b) yields

$$\frac{2n_0 n_1}{n^2}|\,\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1})| = \begin{cases} O(n^{-5/2}(\log n)^{(s-1)/2}) & \text{if } \alpha = 1, \\ O(n^{-3+\alpha/2}(\log n)^{s-1}) & \text{if } \alpha < 1. \end{cases} \tag{4.7}$$

The desired result follows from

$$\mathrm{Var}\,(\hat{I}_n) \leq \frac{1}{n^2}\Big\{n_0^2 \mathrm{Var}\,(\hat{I}_{n_0}) + n_1^2 \mathrm{Var}\,(\hat{I}_{n_1}) + 2n_0 n_1 |\,\mathrm{Cov}\,(\hat{I}_{n_0}, \hat{I}_{n_1})|\Big\}$$

and (4.5), (4.6a), (4.6b) and (4.7).

5. An Illustrative Example

We consider the following multilinear integrand

$$f(\boldsymbol{x}) = 12^{s/2} \prod_{r=1}^{s} (x^r - 0.5) \tag{5.1}$$

which has integral $I = 0$ and variance $\sigma^2 = 1$ for any $s \geq 1$. Owen (1997b) shows that it only has s-dimensional structure, and

$$\sigma^2_{u,\boldsymbol{\kappa}} = 1_{|u|=s} b^{-2|\boldsymbol{\kappa}|}\Big(\frac{b^2-1}{b^2}\Big)^s.$$

The variance formula given by Theorem 2 has been evaluated numerically for $s = 1, \ldots, 11$, b equal to all primes between s and 11 inclusive and all $n = \lambda_0 b^{m_0} + \lambda_1 b^{m_1}$ from $n = 1$ to the smallest such n greater than or equal to 10^7, where $1 \leq \lambda_0, \lambda_1 < b$, $0 \leq m_1 \leq m_0$, and $\lambda_1 b^{m_1} \leq b^{m_0}$. Some of the computation results are shown in Figure 1. In each plot, the horizontal axis displays sample size n and the vertical axis displays the square root of the variance, $[\mathrm{Var}\,(\hat{I}_n)]^{1/2}$. Two reference lines are also given, one is $n^{-1/2}$, corresponding to the simple Monte Carlo rate for the integrand (5.1), and the other is $n^{-3/2}$, corresponding to the asymptotic rate for (5.1) when $s = 1$ and $n = \lambda b^m$ with $1 \leq \lambda < b$ and $m \geq 0$. The dots in each plot correspond to the square root of variance along $n = \lambda b^m$.

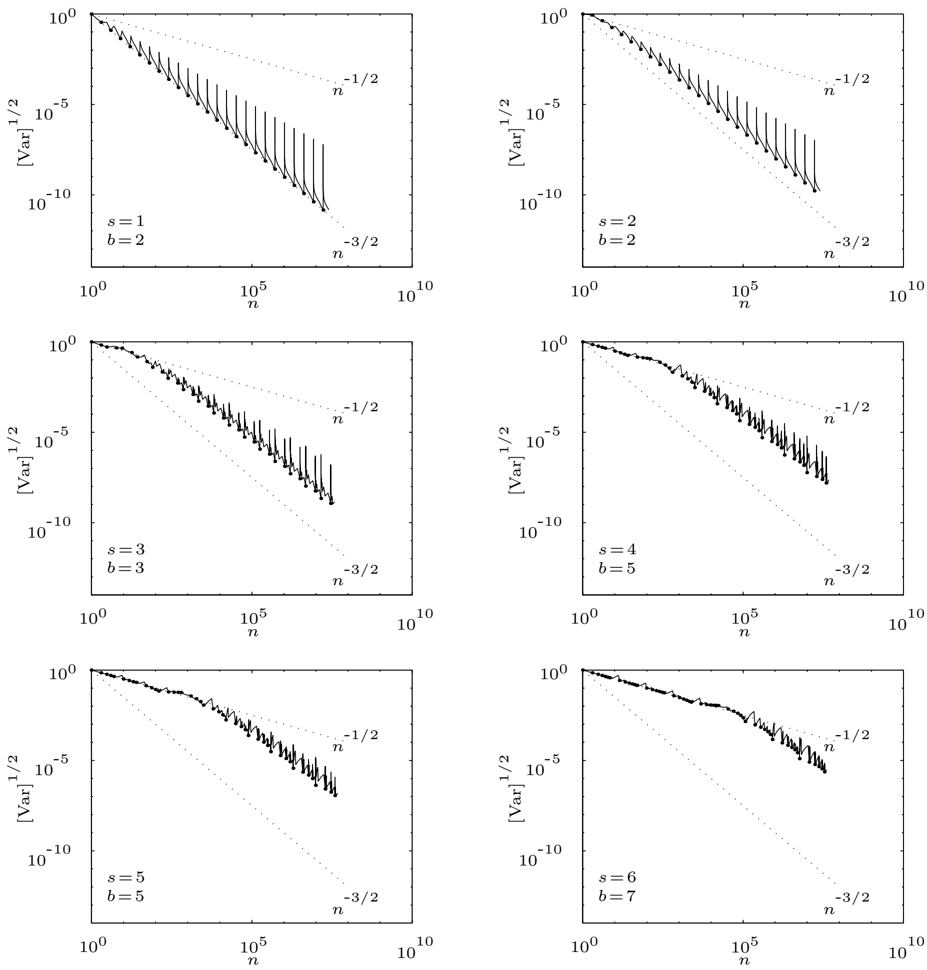

Figure 1. The square root of the variance along simultaneously scrambled unions of two nets for the integrand (5.1).

Suppose that $n_0 = \lambda_0 b^{m_0}$, the number of points in P_{n_0}, is fixed. We consider how adding $n_1 = \lambda_1 b^{m_1}$ points to P_{n_0} affects the estimate. We define the efficiency of the sample of $n = n_0 + n_1$ points with respect to the sample of n_0 points as follows

$$E_{n_0,n} = \frac{\mathrm{Var}\,(\hat{I}_{n_0})}{\mathrm{Var}\,(\hat{I}_n)}. \tag{5.2}$$

Then $E_{n_0,n} \geq 1$ implies that combining two scrambled nets increases the accuracy of the estimate; in contrast, $E_{n_0,n} < 1$ implies that combining two scrambled nets causes some loss of the accuracy. The values of $E_{n_0,n}$ for various cases are calculated. Figure 2 shows $E_{n_0,n}$ for $n_0 = b^{m_0}$ and $n = n_0 + \lambda_1 b^{m_1}$ where $1 \le \lambda_1 < b$, $0 \le m_1 < m_0$, and $\lambda_1 b^{m_1} < b^{m_0}$.

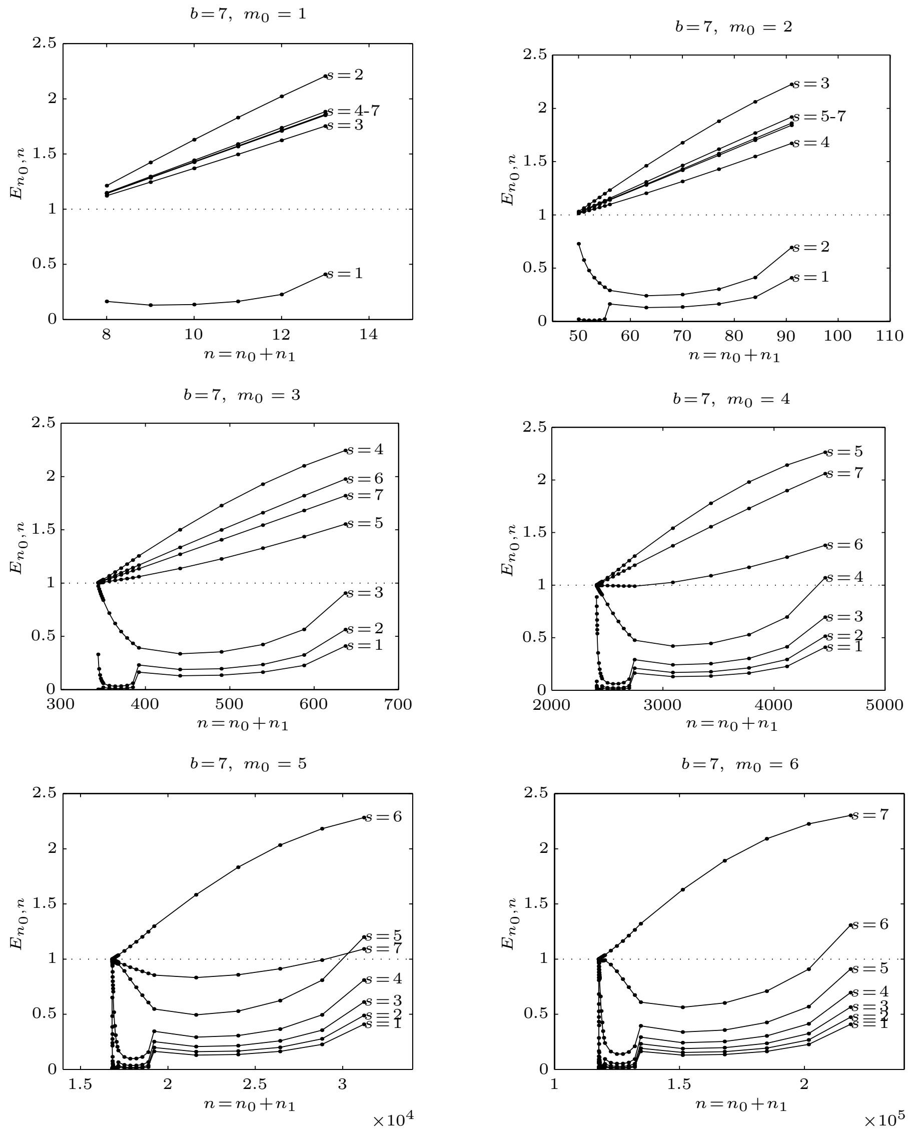

Figure 2. Values of $E_{n_0,n}$ along simultaneously scrambled unions of two nets for the integrand (5.1), where $n_0 = b^{m_0}$, $n_1 = \lambda_1 b^{m_1}$ with $1 \le \lambda_1 < b$, $0 \le m_1 < m_0$, and $n_1 < n_0$.

These results show that among sequences with sample sizes ranging from $\lambda_0 b^{m_0}$ to $(\lambda_0+1)b^{m_0}$, the one with $(\lambda_0+1)b^{m_0}$ points, that is a $(\lambda_0+1, 0, m_0, s)$-net in base b, has the smallest variance. In general, if $m_0 < s$, then increasing sample size $n = \lambda_0 b^{m_0} + \lambda_1 b^{m_1}$ for fixed λ_0 and m_0 may improve the accuracy of the estimate. In contrast, if $m_0 \geq s$, then adding additional $n_1 = \lambda_1 b^{m_1}$ points with $n_1 < b^{m_0}$ may cause an increase in variance. This is different from the case of simple Monte Carlo methods.

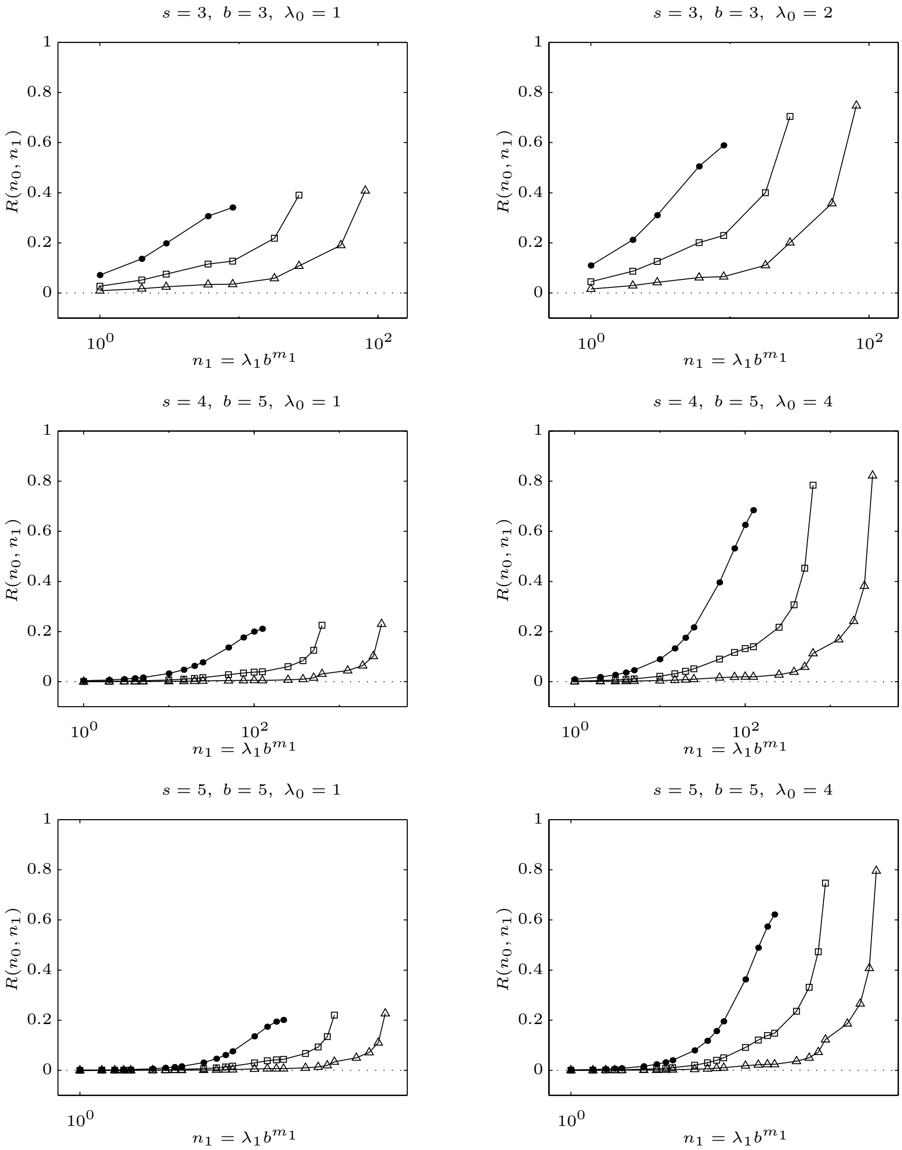

Figure 3. Values of $R(n_0, n_1)$ defined by (5.3) for the integrand (5.1), where $n_0 = \lambda_0 b^{m_0}$ and $n_1 = \lambda_1 b^{m_1}$ with $1 \leq \lambda_1 < b$, $0 \leq m_1 \leq m_0$, and $n_1 \leq b^{m_0}$, while $m_0 = s - 1$ (•), s (□) and $s + 1$ (△), respectively.

Concluding this section, we give a numerical comparison of the variances of the integration for a union of simultaneously scrambled nets and a union of independently scrambled nets. The nets used here are a $(\lambda_0, 0, m_0, s)$-net and a $(\lambda_1, 0, m_1, s)$-net in base b, where $0 \le m_1 \le m_0$, $1 \le \lambda_0, \lambda_1 < b$, and $\lambda_1 b^{m_1} \le b^{m_0}$. Here, we denote the estimate based on simultaneously scrambled nets by $\hat{I}^{\text{sim}}_{n_0+n_1}$, whose variance is calculated from (3.15), and denote the estimate based on independently scrambled nets by $\hat{I}^{\text{ind}}_{n_0+n_1}$, whose variance is calculated from (3.15) ignoring the third term on the right side. Define

$$R(n_0, n_1) = \frac{\mathrm{Var}\,(\hat{I}^{\text{ind}}_{n_0+n_1}) - \mathrm{Var}\,(\hat{I}^{\text{sim}}_{n_0+n_1})}{\mathrm{Var}\,(\hat{I}^{\text{ind}}_{n_0+n_1})}. \tag{5.3}$$

Clearly, the sign of such an $R(n_0, n_1)$ indicates $\mathrm{Var}\,(\hat{I}^{\text{sim}}_{n_0+n_1})$ being larger or smaller than $\mathrm{Var}\,(\hat{I}^{\text{ind}}_{n_0+n_1})$. The closer R is to 1, the more accurate is $\hat{I}^{\text{sim}}_{n_0+n_1}$. Figure 3 shows the values of $R(n_0, n_1)$ for (s, b)=(3,3), (4,5) and (5,5), and $m_0 = s-1, s, s+1$, respectively. The results show that $\hat{I}^{\text{sim}}_{n_0+n_1}$ is much more accurate than $\hat{I}^{\text{ind}}_{n_0+n_1}$ for large λ_0 and large n_1.

6. Discussion

This paper has considered the variance of the sample mean of a deterministic response function over a randomized sequence with $n = \lambda_0 b^{m_0} + \lambda_1 b^{m_1}$ points. This sequence may be obtained by scrambling the union of a $(\lambda_0, 0, m_0, s)$-net and a $(\lambda_1, 0, m_1, s)$-net in base b. Without loss of generality, we assume $m_1 \le m_0$. It turns out that if $m_0 - m_1$ is bounded as $m_0 \to \infty$ the variance is of order $o(n^{-1})$ for any square integrable integrand, $O(n^{-3})$ for a univariate Lipschitz integrand and $O(n^{-3}(\log n)^{s-1})$ for a smooth multivariate integrand, which is the same as the variance of a randomized $(\lambda, 0, m, s)$-net in base b with $n = \lambda b^m$. However, if m_1 is bounded then the variance is $O(n^{-2})$. In general, if there is a constant $\alpha \in [0, 1]$ such that $(1-\alpha)m_0 - m_1$ is bounded as $m_0 \to \infty$, then the variance is $O(n^{-3+\alpha}(\log n)^{(s-1)1_{\alpha<1}})$. It seems that the rate of decay of the error $|\hat{I}_n - I|$ for the case $m_1 = m_0$, that is, for randomized $(\lambda_0 + \lambda_1, 0, m_0, s)$-nets, has a big improvement over nonrandomized nets and simple Monte Carlo.

Our numerical results show that for any fixed $n_0 = \lambda_0 b^{m_0}$, increasing sample size n through $n = n_0 + \lambda_1 b^{m_1}$ does not guarantee the reduction of the variance. This is different from the case of simple Monte Carlo. However, the variance over a randomized union of two nets is never more than a constant multiple of the simple Monte Carlo variance. It is interesting to compare the rate of decay of the error $|\hat{I}_n - I|$ for randomized nets or sequences and for nonrandomized ones. The rate considered in Owen (1995), (1997a,b) and in this paper is an average case result for a fixed function f, taken over random permutations of nets or sequences with low discrepancy. The rate for nonrandomized nets or sequences (Niederreiter (1992)) applies to the worst case over functions, for a fixed set of points. On the other hand, Hickernell (1996) and (1998) has studied the behaviour of randomized nets for the worst case over functions. It is found that the root mean square discrepancy is $n^{-1}(\log n)^{(s-1)/2}$ for large n for the $\mathcal{L}^2$-star, centered symmetric and other similar discrepancies. Thus in an analysis where the function f is chosen pessimistically after the random permutations have been drawn, no real improvement is obtained by scrambling.

This paper considers (λ, t, m, s)-nets in base b with $t = 0$. It would be desirable to deal with the case with $t > 0$. Much less is known about randomized versions of the nets with $t > 0$. Recently, Owen (1998) has studied the variance in this case. Another direction is to study the variance of quadrature under a scrambled net or a union of scrambled nets for a multivariate integrand satisfying some weaker smoothness condition, for instance, a Lipschitz condition on $[0, 1)^s$.

Acknowledgements

I am very grateful to Professor Art B. Owen and Dr. Fred J. Hickernell for stimulating discussions and valuable suggestions on this work. Thanks go also to the Department of Statistics of Stanford University for their hospitality during the visit. This paper was improved as a result of the comments of the referees.

References

Daubechies, I. (1992). *Ten Lectures on Wavelets.* CBMS-NSF, SIAM, Philadelphia.

Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9**, 586-596.

Fang, K. T. and Wang, Y. (1994). *Number-Theoretic Methods in Statistics.* Chapman and Hall, London.

Hickernell, F. J. (1996). Quadrature error bounds with applications to lattice rules. *SIAM J. Num. Anal.* **33**, 1995-2016.

Hickernell, F. J. (1998). A generalized discrepancy and quadrature error bound. *Math. Comp.* **67**, 229-322.

Hua, L. K. and Wang, Y. (1981). *Applications of Number Theory to Numerical Analysis.* Springer Verlag, Berlin.

Jawerth, B. and Sweldens, W. (1994). An overview of wavelet based multiresolution analyses. Technical Report, Department of Mathematics, University of South Carolina.

Madych, W. R. (1992). Some elementary properties of multiresolution analyses of $\mathcal{L}^2(R^m)$. In *Wavelets: A Tutorial in Theory and Applications* (Edited by C. K. Chui). Academic Press, New York.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods.* SIAM, Philadelphia.

Owen, A. B. (1992). Randomized 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* (Edited by H. Niederreiter and P. J. S. Shiue), 299-317. Lecture Notes in Statistics, Vol. 106, Springer-Verlag, Berlin.

Owen, A. B. (1997a). Monte Carlo variance of scrambled equidistribution quadrature. *SIAM, J. Num. Anal.* **34**, 1884-1910.

Owen, A. B. (1997b). Scrambled net variance for integrals of smooth functions. *Ann. Statist.* **25**, 1541-1562.

Owen, A. B. (1998). Scrambled Sobol and Niederreiter-Xing points. *J. Complexity* **14**, 466-489.

Sloan, I. H. and Joe, S. (1994). *Lattice Methods for Multiple Integration.* Clarendon Press, Oxford.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

Department of Mathematics, Hong Kong Baptist University, Hong Kong, and Department of Statistics, East China Normal University, Shanghai 200026.

E-mail: rxyue@online.sh.cn

(Received March 1997; accepted July 1998)