

On reduction of finite-sample variance by extended Latin hypercube sampling

NOBUAKI HOSHINO¹ and AKIMICHI TAKEMURA²

¹*Faculty of Economics, Kanazawa University, Kakuma-machi, Kanazawa-shi, Ishikawa, Japan.
E-mail: hoshino@kenroku.kanazawa-u.ac.jp*

²*Faculty of Economics, University of Tokyo, 3-1, Hongo 7-chome, Bunkyo-ku, Tokyo, Japan*

McKay, Conover and Beckman introduced Latin hypercube sampling (LHS) for reducing the variance of Monte Carlo simulations. LHS is a method for stratifying a univariate margin. We consider an extension of LHS to stratify an m -variate margin with orthogonal arrays, after Owen and Tang. We define extended Latin hypercube sampling of strength m (henceforth denoted by ELHS(m)), such that ELHS(1) reduces to LHS. Using the results obtained by Owen, we first derive an explicit formula for the finite-sample variance of ELHS(m). Based on this formula, we give a sufficient condition for variance reduction by ELHS(m), generalizing similar results from McKay *et al.* for $m = 1$. Actually, our sufficient condition for $m = 1$ contains the sufficient condition of McKay *et al.* and thus strengthens their result.

Keywords: computer experiments; Monte Carlo simulation; numerical integration; orthogonal arrays; variance reduction

1. Introduction

Monte Carlo simulation is often used to evaluate the expectation of a statistic $W = g(X_1, \dots, X_K)$, which is not analytically tractable. In most Monte Carlo simulations simple random sampling (SRS) is used to generate sample points. SRS is widely applicable because of its simplicity. However, its sampling variance is often large and many replications are needed to achieve the desired precision. Therefore methods for reducing the sampling variance of SRS are of great importance.

One way of reducing SRS variance is to scatter the sample points more uniformly than in SRS. This is the basic idea of various techniques known as quasi-Monte Carlo methods (for a review, see Niederreiter 1992). We can achieve uniformity by stratifying the sample space. Latin hypercube sampling (LHS) is a method for stratifying each univariate margin simultaneously. It is natural to define extended Latin hypercube sampling ELHS(m) of strength m , $m \leq K$, by stratifying an m -variate margin simultaneously, such that ELHS(1) is equivalent to LHS. The desired stratification property can be achieved by sampling designs based on orthogonal arrays (OAs) proposed by Owen (1992a) and independently by Tang (1993). ELHS is a particular version of OA-based sampling.

The reduction of variance by LHS and ELHS is closely related to analysis of variance (ANOVA) decomposition of the statistic W . Stein (1987) showed that LHS asymptotically

filters out main effects of W . Therefore LHS asymptotically achieves variance reduction for any statistic W . Similarly, OA-based sampling asymptotically filters some higher-order interactions as well, and hence asymptotically achieves further variance reduction for any statistic. However, for the finite-sample case, LHS and its generalizations do not necessarily lead to variance reduction due to combinatorial complications.

In order to investigate the reduction of finite-sample variance, we first derive an explicit expression for the finite-sample variance of ELHS(m) based on the results obtained by Owen (1994a). Using this expression, we derive a sufficient condition for the variance reduction of ELHS(m) over SRS. Our sufficient condition is given in terms of m -variate monotonicity of the statistic $g(X_1, \dots, X_K)$. For the case of $m = 1$, our sufficient condition requires that g is monotone in any $K - 1$ out of K variables X_1, \dots, X_K . The sufficient condition given by McKay *et al.* (1979) requires that g is monotone in each X_i , $i = 1, \dots, K$. Thus our result for $m = 1$ strengthens the results of McKay *et al.* (1979).

The organization of this paper is as follows. In Section 2 we define ELHS(m) and introduce appropriate notational conventions. In Section 3 we derive explicit expression for the finite-sample variance of ELHS(m). Based on this expression, we give a sufficient condition for reduction of finite-sample variance of ELHS(m) over SRS in Section 4. In Section 5 we perform some numerical simulations to confirm our theoretical results. In Section 6 we make some additional comments on LHS and OA-based sampling.

2. Construction of extended Latin hypercube sampling

In this section we explicitly describe our ELHS(m), which is a particular version of Owen's generalization of LHS. We evaluate the expectation of a statistic $W = g(X_1, \dots, X_K)$:

$$\mu = E(W) = E[g(X_1, \dots, X_K)],$$

where $(X_1, \dots, X_K) \in \mathbb{R}^K$. We assume that X_1, \dots, X_K are independent continuous random variables with known distribution functions F_j , $j = 1, \dots, K$. The vector $(X_1, \dots, X_K) = \mathbf{X}$ has the joint distribution function $F = F_1 \cdots F_K$.

Suppose that the evaluation of this expectation is analytically intractable and we use a Monte Carlo method. In our extended Latin hypercube sampling defined below, the sample points are generated in two steps. For the first step, we partition the sample space \mathbb{R}^K into N^K cells of equal probability $1/N^K$ and choose N^m , $m \leq K$, cells of N^K cells using a random OA. For the second step the actual sample points are generated according to the conditional distribution on the chosen cells.

We now describe the first step. A $\lambda N^m \times K$ matrix \mathbf{D} , with elements taken from a set of N symbols, is called an orthogonal array of strength m ($m \leq K$), size λN^m , K constraints, N levels and frequency λ , if in any $\lambda N^m \times m$ submatrix of \mathbf{D} each of the possible $1 \times m$ row vectors occurs the same number λ of times. Such an array is denoted by OA($\lambda N^m, K, N, m$). Without essential loss of generality, we only consider the case $\lambda = 1$ as in the original LHS of McKay *et al.* (1979).

The orthogonal array is a natural generalization of orthogonal Latin squares. Plackett and Burman (1946) generated OAs of strength 2 by combining mutually orthogonal Latin

squares. Rao (1947) formulated the concept of OAs in general form and gave a lower bound of N for fixed m, K . Bose and Bush (1952) explained how to construct OAs and gave sufficient conditions for the existence of OAs, since they do not always exist.

We arbitrarily choose and fix an orthogonal array $OA(N^m, K, N, m)$ and denote it by \mathbf{D} . The ij th element of \mathbf{D} is denoted by d_{ij} . We call \mathbf{D} a *generator array*. Actually we will show that the selection of \mathbf{D} does not affect the variance of $ELHS(m)$. We generate random OAs by random permutations of elements of \mathbf{D} . For simplicity of notation, we take the set of N symbols as

$$Z_N = \{1, 2, 3, \dots, N\}.$$

Let \mathcal{S}_N denote the symmetric group of Z_N , that is, the set of permutations of $\{1, \dots, N\}$. Let $\pi \in \mathcal{S}_N$; then $(\pi(1), \dots, \pi(N))$ is a particular permutation of $(1, \dots, N)$. We choose K permutations $\pi_j, j = 1, \dots, K$, independently and uniformly from \mathcal{S}_N and we apply $\pi_j \in \mathcal{S}_N$ to the j th column of $\mathbf{D}, j = 1, \dots, K$. The resulting array is denoted by $\Pi(\mathbf{D})$, with ij th element $\pi_j(d_{ij})$. This constitutes the first step of generating the sample points.

Let $F_j^{-1}(u) = \inf\{x | F_j(x) \geq u\}, j = 1, \dots, K$, be the quantile functions, and for $\mathbf{z} = (z_1, z_2, \dots, z_K) \in Z_N^K$ let

$$P^{(\mathbf{z})} = F_1^{-1}\left(\left(\frac{z_1 - 1}{N}, \frac{z_1}{N}\right]\right) \times \dots \times F_K^{-1}\left(\left(\frac{z_K - 1}{N}, \frac{z_K}{N}\right]\right) \subset \mathbb{R}^K,$$

which is a cell of \mathbb{R}^K with probability $1/N^K$. When $\pi_j, j = 1, \dots, K$, are given we denote the i th row of $\Pi(\mathbf{D})$ by

$$z_{ij} = \pi_j(d_{ij}), \quad z_i = (z_{i1}, z_{i2}, \dots, z_{iK}) \in Z_N^K.$$

For the second step we generate a random vector (X_{i1}, \dots, X_{iK}) in the cell $P^{(z_i)}$ according to the conditional distribution of F on the cell. Let $0 < U_{ij} \leq 1, i = 1, \dots, N^m, j = 1, \dots, K$, be independent uniform random variables. Then X_{ij} can be generated as

$$X_{ij} = F_j^{-1}\{(z_{ij} - U_{ij})/N\}.$$

We now define our estimator T_{EL} of μ based on $ELHS(m)$ by

$$T_{EL} = \frac{1}{N^m} \sum_{i=1}^{N^m} g(X_{i1}, \dots, X_{iK}).$$

Obviously T_{EL} is an unbiased estimator of μ .

Furthermore we denote the usual estimator of μ based on SRS by T_R . We are interested in the comparison of variances of T_{EL} and T_R .

3. Finite-sample variance of ELHS

3.1. ANOVA decomposition of the cell mean function

In order to investigate the variance of T_{EL} , we introduce the *cell mean function* and its ANOVA decomposition. The cell $P^{(\mathbf{z})}$ is indexed by $\mathbf{z} = (z_1, z_2, \dots, z_K) \in Z_N^K$. In an abuse of

notation, we simply denote the cell as $\mathbf{z} = (z_1, z_2, \dots, z_K) \in Z_N^K$. Suppose that a sample point \mathbf{X} is obtained from a cell \mathbf{z} . We call the conditional expectation $E[g(\mathbf{X})|\mathbf{z}]$ the cell mean function and denote it by

$$\mu_{\mathbf{z}} = E[g(\mathbf{X})|\mathbf{z}] = N^K \int_{p(\mathbf{z})} g(\mathbf{X})f(\mathbf{X}) d\mathbf{X},$$

where $f(\mathbf{X})$ is the density function of \mathbf{X} .

We denote the usual ANOVA decomposition of the cell mean function by

$$\begin{aligned} &\mu_{z_1 z_2 \dots z_K} - \mu \\ &= \sum_{i=1}^K \alpha_1(z_i) + \sum_{i_1 < i_2} \alpha_2(z_{i_1}, z_{i_2}) + \sum_{i_1 < i_2 < i_3} \alpha_3(z_{i_1}, z_{i_2}, z_{i_3}) + \dots + \alpha_K(z_{i_1}, \dots, z_{i_K}), \end{aligned}$$

where

$$\begin{aligned} \alpha_1(z_i) &= \frac{1}{N^{K-1}} \sum_{z_1=1}^N \dots \sum_{z_{i-1}=1}^N \sum_{z_{i+1}=1}^N \dots \sum_{z_K=1}^N (\mu_{z_1 z_2 \dots z_K} - \mu), \\ \alpha_2(z_{i_1}, z_{i_2}) &= \frac{1}{N^{K-2}} \sum_{\substack{j=1 \\ j \neq i_1, i_2}}^K \sum_{z_j=1}^N (\mu_{z_1 z_2 \dots z_K} - \mu - \alpha_1(z_{i_1}) - \alpha_1(z_{i_2})), \end{aligned}$$

and we continue this process to α_K . In each $\alpha_s(z_{i_1}, \dots, z_{i_s})$, (z_1, \dots, z_K) is to be summed over all K -vectors in Z_N^K with the values of z_{i_1}, \dots, z_{i_s} fixed and all other coordinates free. Summation over z_i is always taken from 1 to N , and we henceforth omit the range $1 \leq z_i \leq N$ from the summation signs.

We denote the sum of squares of s th-order interaction effects by

$$\varphi_s^2 = \frac{1}{N^s} \sum_{i_1 < \dots < i_s} \sum_{z_{i_1}} \dots \sum_{z_{i_s}} \alpha_s(z_{i_1}, \dots, z_{i_s})^2.$$

Then $\text{var}(T_R)$ can be written as

$$\begin{aligned} \text{var}(T_R) &= \frac{1}{N^m} \{ \text{var}(\mu_{\mathbf{z}}) + E_{\mathbf{z}}[\text{var}(g(\mathbf{X})|\mathbf{z})] \} \\ &= \frac{1}{N^m} \sum_{s=1}^K \varphi_s^2 + \nabla, \end{aligned} \tag{1}$$

where

$$\nabla = \frac{1}{N^m} E_{\mathbf{z}}[\text{var}(g(\mathbf{X})|\mathbf{z})]$$

(see Serfling 1980, Section 5.2).

3.2. The variance of estimators using ELHS

For any $N^m \times s$ subarray $\mathbf{A} = \{a_{ij}\}$, $i = 1, \dots, N^m$, $j = 1, \dots, s$, of the generator array \mathbf{D} , define

$$L(s, h) = \sum_{i=1}^{N^m} \sum_{l=1}^{N^m} I\left(\sum_{j=1}^s I(a_{ij} = a_{lj}) = h\right), \quad 0 \leq h \leq s,$$

where I is an indicator function. We can interpret $L(s, h)$ as the number of pairs of rows of \mathbf{A} with exactly h common elements. By symmetry among the columns of \mathbf{D} , $L(s, h)$ is common to all $N^m \times s$ subarrays \mathbf{A} . Theorem 1 of Owen (1994a) implies that

$$\text{var}(T_{\text{EL}}) = \frac{1}{N^{2m}} \sum_{s=m+1}^K \sum_{h=0}^s L(s, h)(1 - N)^{h-s} \varphi_s^2 + \nabla. \tag{2}$$

Owen (1994a) does not give an explicit form for $L(s, h)$. We obtain an explicit form of $L(s, h)$, which leads to Theorem 1 below. We have given an alternative proof of (3) in the preprint form of this paper (Hoshino and Takemura 1997) available from us.

Theorem 1.

$$\text{var}(T_{\text{EL}}) = \frac{1}{N^m} \sum_{s=m+1}^K \varphi_s^2 \left\{ 1 - (1 - N)^{1-s} \sum_{u=0}^{m-1} (-N)^u \binom{s-1}{u} \right\} + \nabla. \tag{3}$$

Note that in (3) interaction effects up to order m , that is, φ_s^2 , $s = 1, \dots, m$, are cancelled out. This has to be the case for ELHS(m). In ELHS(m) we stratify an m -variate margin, such that for $s \leq m$ all elements of Z_N^s appear an equal number of times in each combination of s axes. In view of $\sum_{z_{i_1}} \dots \sum_{z_{i_s}} \alpha_s(z_{i_1}, \dots, z_{i_s}) = 0$, interaction effects of the cell mean function up to order m vanish for each realization of T_{EL} . Expression (3) may also be written as

$$\text{var}(T_{\text{EL}}) = \frac{1}{N^m} \sum_{s=m+1}^K \varphi_s^2 \left\{ (1 - N^{-1})^{1-s} \sum_{u=0}^{s-m-1} \binom{s-1}{u} (-N)^{-u} \right\} + \nabla,$$

using the fact that $1 = (1 - N)^{1-s} \sum_{u=0}^{s-1} (-N)^u \binom{s-1}{u}$. This expression was suggested to us by a referee.

Theorem 1 is an immediate consequence of the following two lemmas, whose proofs are given in Appendix A.

Lemma 1. *Let $m < s \leq K$. For $0 \leq h \leq m - 1$,*

$$L(s, h) = \binom{s}{h} \sum_{t=0}^{m-h-1} (-1)^t (N^{m-h-t} - 1) \binom{s-h}{t} N^m. \tag{4}$$

$L(s, h) = 0$ for $m \leq h \leq s - 1$, and $L(s, s) = N^m$.

Combining Lemma 1 and (2),

$$\text{var}(T_{\text{EL}}) = \frac{1}{N^m} \sum_{s>m} \left(\sum_{h=0}^{m-1} (1-N)^{h-s} \binom{s}{h} \sum_{t=0}^{m-h-1} (-1)^t (N^{m-h-t} - 1) \binom{s-h}{t} + 1 \right) \varphi_s^2. \tag{5}$$

The coefficients of φ_s^2 in (5) can be simplified as follows.

Lemma 2. *For $m < s \leq K$,*

$$\begin{aligned} & \sum_{h=0}^{m-1} (1-N)^{h-s} \binom{s}{h} \sum_{t=0}^{m-h-1} (-1)^t (N^{m-h-t} - 1) \binom{s-h}{t} + 1 \\ &= 1 - (1-N)^{1-s} \sum_{u=0}^{m-1} (-N)^u \binom{s-1}{u}. \end{aligned} \tag{6}$$

4. A sufficient condition for variance reduction

In this section we obtain a sufficient condition for the variance reduction $\text{var}(T_{\text{EL}}) \leq \text{var}(T_{\text{R}})$. McKay *et al.* (1979) showed that monotonicity of $g(X_1, \dots, X_K)$ in each argument X_i , $i = 1, \dots, K$, is a sufficient condition for the variance reduction at $m = 1$. Here $g(\mathbf{X})$ is monotone in X_i if $g(\mathbf{X})$, regarded as a function of the i th argument (with all other arguments arbitrarily held fixed), is monotone in the usual sense. The direction of the monotonicity (increasing or decreasing) must not depend on the values of the other arguments. Their proof of the variance reduction depends on the results of Lehmann (1966). Our proof is completely different and we use the monotonicity in a different way. In the following, we clarify their condition and generalize the condition to $m > 1$.

First, let us introduce the difference operator Δ_{x_i, x'_i} , defined by the formula

$$\Delta_{x_i, x'_i} g(\mathbf{X}) = g(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_K) - g(X_1, \dots, X_{i-1}, x'_i, X_{i+1}, \dots, X_K).$$

Then

$$\Delta_{x_1, x'_1} \Delta_{x_2, x'_2} \cdots \Delta_{x_t, x'_t} g(\mathbf{X}) \tag{7}$$

corresponds to the mixed partial derivative with respect to X_1, \dots, X_t . We make the following definitions. *Monotonicity of $g(\mathbf{X})$ in (X_1, \dots, X_t)* holds if the sign of (7) remains the same, when (7) is regarded as a function of the t arguments with all other arguments arbitrarily held fixed. Also *t th-order monotonicity of $g(\mathbf{X})$ in X_1* holds if $g(\mathbf{X})$ is monotone in $(X_1, X_{i_2}, \dots, X_{i_t})$ for all $\binom{K-1}{t-1}$ combinations $(1, i_2, \dots, i_t)$, $1 < i_2 < \dots < i_t \leq K$. The monotonicity in other arguments is similarly defined. Our generalized monotonicity implies that for differentiable functions the sign of the partial mixed derivative remains the same regardless of the values of the variables.

Now we state our main theorem.

Theorem 2. *If t th-order monotonicity holds in $K - t$ arguments of $g(X_1, \dots, X_K)$ for all t such that $1 \leq t \leq m$, then $\text{var}(T_{\text{EL}}) \leq \text{var}(T_{\text{R}})$.*

We denote T_{EL} under LHS, that is, $m = 1$, simply by T_L .

Corollary 1. *If $g(X_1, \dots, X_K)$ is monotone in $K - 1$ arguments, then $\text{var}(T_L) \leq \text{var}(T_R)$.*

The sufficient condition given by McKay *et al.* (1979) for LHS requires that g be monotone in all K arguments for variance reduction. Actually $K - 1$ monotonicities are sufficient.

Since (3) can alternatively be written as

$$\begin{aligned} \text{var}(T_{EL}) &= \frac{1}{N^m} \sum_{s=1}^K \varphi_s^2 \left\{ 1 - (1 - N)^{1-s} \sum_{u=0}^{m-1} (-N)^u \binom{s-1}{u} \right\} + \nabla \\ &= \text{var}(T_R) + \frac{1}{N^m} \sum_{s=1}^K \varphi_s^2 \left\{ -(1 - N)^{1-s} \sum_{u=0}^{m-1} (-N)^u \binom{s-1}{u} \right\}, \end{aligned} \tag{8}$$

the variance reduction holds if and only if

$$\sum_{s=1}^K \varphi_s^2 \left\{ (-1)^s (N - 1)^{1-s} \sum_{u=0}^{m-1} (-N)^u \binom{s-1}{u} \right\} \leq 0. \tag{9}$$

We rewrite (9) as

$$\sum_{u=0}^{m-1} N^u \sum_{s=1+u}^K \varphi_s^2 (-1)^{s+u} (N - 1)^{1-s} \binom{s-1}{u} \leq 0. \tag{10}$$

Our objective is to obtain a sufficient condition for (10). If, for $0 \leq u \leq m - 1$,

$$\sum_{s=1+u}^K \varphi_s^2 (-1)^{s+u} (N - 1)^{1-s} \binom{s-1}{u} \leq 0, \tag{11}$$

then (10) holds. Let $t = 1 + u$. In Appendix B, we show that t th-order monotonicity in $K - t$ axes is a sufficient condition for (11).

The monotonicity condition of Theorem 2 may be difficult to check in practice. If we can assume that higher-order interaction effects are small, then variance reduction holds without the full monotonicity condition of Theorem 2. For example, assume that $\varphi_1^2 \geq \varphi_i^2$ holds for all i . Then $\text{var}(T_L) \leq \text{var}(T_R)$ for $N \geq 3$ by

$$\varphi_1^2 \geq \varphi_1^2 \frac{N - 1}{(N - 1)^2 - 1} = \varphi_1^2 \sum_{i=0}^{\infty} (N - 1)^{-2i-1} \geq \sum_{s=2}^K \varphi_s^2 (-1)^s (N - 1)^{1-s}.$$

Hence LHS gives the variance reduction in quite a number of situations. Similarly, if we assume that $\varphi_2^2 \geq \varphi_i^2$ for $i \geq 2$, then (11) holds for $N \geq 4$ with $u = 1$ by

$$\frac{\varphi_2^2}{N - 1} \geq \varphi_2^2 \frac{2(N - 1)^2}{(N^2 - 2N)^2} = \varphi_2^2 \sum_{i=0}^{\infty} \frac{2 + 2i}{(N - 1)^{2i+2}} \geq \sum_{s=3}^K \varphi_s^2 (-1)^{s+1} (N - 1)^{1-s} \binom{s-1}{1}.$$

We summarize these simple sufficient conditions for variance reduction in the following proposition.

Proposition 1. *If $\varphi_1^2 \geq \varphi_i^2$ for all i , then the variance reduction of LHS holds for $N \geq 3$. If $\varphi_1^2 \geq \varphi_2^2 \geq \varphi_i^2$ for $i \geq 2$, then the variance reduction of ELHS(2) holds for $N \geq 4$.*

5. Some simulation results

In this section we confirm our theoretical results by simulation. We choose a rather simple situation where the exact variance of ELHS(m) can be computed. Note that this is not a realistic situation in which simulations are actually used.

Let $K = 4$, and consider

$$W = \exp(X_1 + X_2 + X_3 + X_4), \quad (12)$$

where X_i , $i = 1, \dots, 4$, are independently and identically distributed as $U[0, 1]$. In this example all mixed derivatives are positive and hence monotonicities of all orders hold.

We obtained the simulation variance with 1 000 000 replications for each T_R , T_L and T_{EL} of $m = 2$. The sample size was 25. Thus $N = 25$ for LHS and $N = 5$ for ELHS(2). The simulated sampling distributions of SRS, LHS and ELHS(2) are plotted in Figure 1. We see

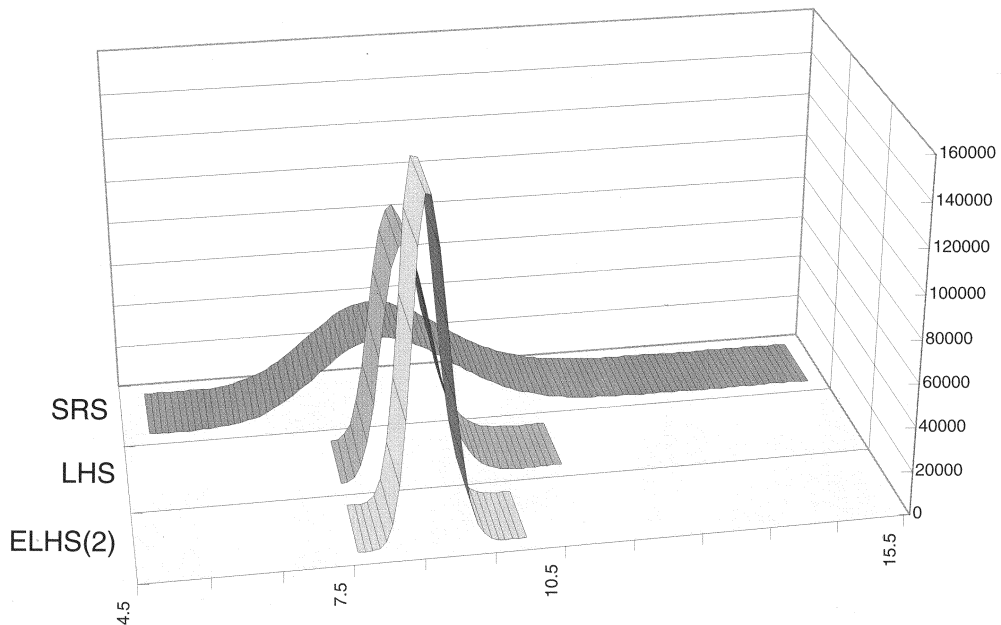


Figure 1. Simulated distributions of T_R , T_L and T_{EL} (Section 5).

that T_{EL} is much more concentrated around the true value μ . The numerical results of our simulation are summarized in Table 1. Observe that the true values and the simulated values in Table 1 are in close agreement, and this confirms our theoretical results.

In this example $\varphi_1^2 = 23.825\ 679$, $\varphi_2^2 = 2.801\ 345$, $\varphi_3^2 = 0.146\ 388$, $\varphi_4^2 = 0.002\ 869$ for ELHS(2) with $N = 5$. For LHS with $N = 25$, $\varphi_1^2 = 24.873\ 724$, $\varphi_2^2 = 3.053\ 217$, $\varphi_3^2 = 0.166\ 568$, $\varphi_4^2 = 0.003\ 408$ and these are different from those for ELHS(2) with $N = 5$. Our choice of $m = 2$ seems to be appropriate in this example.

6. Further discussion

It is desirable to compare ELHS(m) schemes for different m under the same sample size. In the example (12) of the preceding section, ELHS(2) dominates ELHS(1). However, general monotone variance reduction in m might not hold for small sample size. The difficulty lies in the fact that $\varphi_{s,s}$ depend on N and signs in the summation of (3) are alternating. For large sample size φ_s becomes stable with respect to N , but the monotone variance reduction in m obviously holds asymptotically.

The method of mapping z_{ij} into X_{ij} is a controversial point. Although it produces bias, we can make $\nabla = 0$ by choosing x_{ij} deterministically given z_{ij} . Hence one needs to consider the trade-off between bias and variance. Owen (1992a) discusses the midpoint rule and rectangular rule, and Tang (1993) introduces a Latin hypercube structure to the cells, that is, the sample points in the given cells are generated by using a method like LHS. Under this stratification, ∇ is reduced when W is additive.

To achieve specific objectives some authors impose restrictions on the Latin hypercube design. Handcock (1991) proposed a ‘cascading’ Latin hypercube design. Sample points were obtained by using modified LHS with midpoint rule, and a few points from the same cells are added with LHS. In the literature of experimental design, optimality of permuted generator arrays is discussed. Uniformity of the sample points is improved by restricting permutations. Tang (1994) introduced a criterion for comparing design arrays. Shaw (1988) reviewed the criteria of uniformity.

Table 1. Simulation results (1 000 000 replications)

	SRS	LHS	ELHS(2)
True $E[W]$	8.717 212	8.717 212	8.717 212
Sample mean	8.716 336	8.716 541	8.717 124
True variance of μ_z	–	0.134 005	0.009 239
True ∇	–	0.002 221	0.055 047
True variance of T	1.126 098	0.136 226	0.064 286
Sample variance of T	1.123 594	0.136 028	0.064 175
Minimum T	4.991 916	7.612 060	7.613 056
Maximum T	15.304 256	10.909 609	10.060 986

In our assumption, each axis is independently distributed. Iman and Conover (1982) treated dependency in LHS, and Owen (1994b) proposed another algorithm for controlling correlations. Stein (1987) discussed the central limit theorem for LHS and Owen (1992b) gave a proof of the central limit theorem using the method of moments. Loh (1996) established a Berry–Esseen-type bound for LHS. It is natural to expect that the central limit theorem holds for T_{EL} under appropriate regularity conditions (see Figure 1).

Appendix A. Proofs of lemmas in Section 3

Proof of Lemma 1. We first show that $L(s, h) = 0$ for $m \leq h \leq s - 1 < K$. Note that any $N^m \times s$ submatrix of $OA(N^m, K, N, m)$ becomes $OA(N^m, s, N, m)$. In an orthogonal array $OA(N^m, s, N, m)$, the rows of each $N^m \times m$ submatrix are all distinct. This implies that two rows of $OA(N^m, s, N, m)$ can only have less than m elements in common. Therefore $L(s, h) = 0$ for $m \leq h \leq s - 1$, and $L(s, s) = N^m$.

Now let $h < m$. We consider a particular $OA(N^m, s, N, m)$, say \mathbf{A} , whose first row is assumed for the moment to be $\mathbf{y} = (1, 1, \dots, 1)$. With the remaining $N^m - 1$ rows of \mathbf{A} , we want to know the number of rows that have just h axes indexed as 1. There are $N^{m-h} - 1$ rows \mathbf{z} of \mathbf{A} such that $z_1 = z_2 = \dots = z_h = 1$. Using an inclusion–exclusion formula, we can count the number of rows that satisfy $z_1 = \dots = z_h = 1, z_{h+1} \neq 1, \dots, z_K \neq 1$ among these. Note that the first row $(1, 1, \dots, 1)$ has to be subtracted from the count. Therefore the number is

$$\sum_{t=0}^{m-h-1} (-1)^t (N^{m-h-t} - 1) \binom{s-h}{t}. \tag{13}$$

This times $\binom{s}{h}$ yields the total number of rows that have just h axes indexed as 1.

Note that the above argument does not depend on the assumption $\mathbf{y} = (1, 1, \dots, 1)$. Therefore, multiplying (13) by $N^m \binom{s}{h}$, we obtain the lemma. \square

The following identity is easily shown by induction on $c \geq 0$:

$$\sum_{b=0}^c (-1)^b \binom{a}{b} = (-1)^c \binom{a-1}{c}. \tag{14}$$

Now we need the following lemma to prove Lemma 2.

Lemma 3. For $0 \leq l < m - q, q \geq 1, s \geq m$,

$$\sum_{h=l}^{m-q} (-1)^{m-h-q} \binom{s-h-1}{m-h-q} \binom{s-l}{h-l} = 1.$$

Proof. Note that $s - l \geq 0$, $s - m + q \geq 0$. Consider the series expansion of

$$\frac{(1+x)^{s-l}}{(1+x)^{s-m+q}} = (1+x)^{m-q-l}.$$

We obtain the lemma by comparing the coefficients of x^{m-q-l} . □

Proof of Lemma 2. The left-hand side of (6) equals

$$1 - (1-N)^{1-s} \sum_{h=0}^{m-1} (1-N)^h \binom{s}{h} \sum_{t=0}^{m-h-1} (-1)^t \frac{N^{m-h-t} - 1}{N-1} \binom{s-h}{t}.$$

Thus we need to show that

$$\sum_{h=0}^{m-1} (1-N)^h \binom{s}{h} \sum_{t=0}^{m-h-1} (-1)^t \frac{N^{m-h-t} - 1}{N-1} \binom{s-h}{t} = \sum_{u=0}^{m-1} (-N)^u \binom{s-1}{u}. \quad (15)$$

The left-hand side of (15) reduces to

$$\begin{aligned} & \sum_{h=0}^{m-1} (1-N)^h \binom{s}{h} \sum_{t=0}^{m-h-1} (-1)^t \sum_{j=0}^{m-h-t-1} N^j \binom{s-h}{t} \\ &= \sum_{h=0}^{m-1} \sum_{u=0}^h \binom{h}{u} (-N)^u \binom{s}{h} \sum_{j=0}^{m-h-1} N^j (-1)^{m-h-1-j} \binom{s-h-1}{m-h-1-j} \end{aligned} \quad (16)$$

by (14). Furthermore, (16) equals

$$\begin{aligned} & \sum_{h=0}^{m-1} \sum_{l=0}^h (-N)^l \binom{s}{l} \binom{s-l}{h-l} \sum_{j=0}^{m-h-1} N^j (-1)^{m-h-1-j} \binom{s-h-1}{m-h-1-j} \\ &= \sum_{u=0}^{m-1} \sum_{l=0}^u \sum_{h=l}^{m-q} \binom{s}{l} \binom{s-l}{h-l} \binom{s-h-1}{m-h-q} (-1)^{m-h-q} (-1)^l N^u, \end{aligned} \quad (17)$$

where $q = 1 + u - l$. Using Lemma 3, (17) equals

$$\sum_{u=0}^{m-1} \sum_{l=0}^u \binom{s}{l} (-1)^l N^u = \sum_{u=0}^{m-1} (-N)^u \binom{s-1}{u}$$

by (14). This is the right-hand side of (15). □

Appendix B. Proofs of results required for section 4

It is straightforward to show that

$$\text{var}(T_{\text{EL}}) = \text{var}(T_{\text{R}}) + \frac{N^m - 1}{N^m} \text{cov}(\mu_{\mathbf{y}}, \mu_{\mathbf{z}}),$$

where $\text{cov}(\mu_y, \mu_z) = E[(\mu_y - \mu)(\mu_z - \mu)]$ is the covariance of two cell mean functions under the cell selection. Comparing this with (8), we have

$$(N^m - 1)\text{cov}(\mu_y, \mu_z) = \sum_{s=1}^K \varphi_s^2 \left\{ (-1)^s (N - 1)^{1-s} \sum_{u=0}^{m-1} (-N)^u \binom{s-1}{u} \right\}.$$

We are interested in a sufficient condition for this covariance term to be non-positive: $(N^m - 1)\text{cov}(\mu_y, \mu_z) \leq 0$.

Here we prefer to examine the case $m = 1$ first and generalizing the result to $m > 1$ later.

B.1. LHS

For the case $m = 1$, we can rewrite $(N - 1)\text{cov}(\mu_y, \mu_z)$ to be non-positive as

$$\begin{aligned} 0 \geq & - \sum_{i=1}^K \sum_{z_i} \alpha_1(z_i)^2 N^{-1} (N - 1)^0 + \sum_{i_1 < i_2} \sum_{z_{i_1}} \sum_{z_{i_2}} \alpha_2(z_{i_1}, z_{i_2})^2 N^{-2} (N - 1)^{-1} \\ & - \sum_{i_1 < i_2 < i_3} \sum_{z_{i_1}} \sum_{z_{i_2}} \sum_{z_{i_3}} \alpha_3(z_{i_1}, z_{i_2}, z_{i_3})^2 N^{-3} (N - 1)^{-2} \\ & + \dots + (-1)^K \sum_{z_1} \dots \sum_{z_K} \alpha_K(z_1, \dots, z_K)^2 N^{-K} (N - 1)^{1-K}. \end{aligned}$$

Lemma 4. *If μ_z is monotone in z_1 , then*

$$\begin{aligned} 0 \leq & \sum_{z_1} \alpha_1(z_1)^2 N^{-1} (N - 1)^0 (-1)^0 + \sum_{1 < i} \sum_{z_1} \sum_{z_i} \alpha_2(z_1, z_i)^2 N^{-2} (N - 1)^{-1} (-1)^1 \\ & + \sum_{1 < i_1 < i_2} \sum_{z_1} \sum_{z_{i_1}} \sum_{z_{i_2}} \alpha_3(z_1, z_{i_1}, z_{i_2})^2 N^{-3} (N - 1)^{-2} (-1)^2 \\ & + \dots + \sum_{z_1} \dots \sum_{z_K} \alpha_K(z_1, \dots, z_K)^2 N^{-K} (N - 1)^{1-K} (-1)^{K-1}. \end{aligned}$$

Proof. By the monotonicity in z_1 ,

$$(\mu_{z_1 z_2 \dots z_K} - \mu_{1 z_2 \dots z_K})(\mu_{z_1 y_2 \dots y_K} - \mu_{1 y_2 \dots y_K}) \geq 0, \quad \forall z_1, z'_1, z_2, \dots, z_K, y_2 \neq z_2, \dots, y_K \neq z_K.$$

Therefore

$$\sum_{z_1} \sum_{z'_1} \sum_{z_2} \dots \sum_{z_K} \sum_{y_2 \neq z_2} \dots \sum_{y_K \neq z_K} (\mu_{z_1 z_2 \dots z_K} - \mu_{z'_1 z_2 \dots z_K})(\mu_{z_1 y_2 \dots y_K} - \mu_{z'_1 y_2 \dots y_K}) \geq 0. \tag{18}$$

For $2 \leq t \leq s$,

$$\begin{aligned} & \sum_{y_i \neq z_i} (\alpha_s(z_1, y_2, \dots, y_s) - \alpha_s(z'_1, y_2, \dots, y_s)) \\ &= (-1)(\alpha_s(z_1, y_2, \dots, z_t, \dots, y_s) - \alpha_s(z'_1, y_2, \dots, z_t, \dots, y_s)). \end{aligned}$$

Further,

$$\sum_{z_1} \sum_{z'_1} (\alpha_s(z_1, \dots) - \alpha_s(z'_1, \dots))^2 = 2N \sum_{z_1} (\alpha_s(z_1, \dots))^2.$$

Using these facts, (18) divided by $2N$ is

$$\begin{aligned} 0 &\leq N^{K-1}(N-1)^{K-1} \sum_{z_1} \alpha_1(z_1)^2 + N^{K-2}(N-1)^{K-2} \sum_{2 \leq i \leq K} \sum_{z_1} \sum_{z_i} \alpha_2(z_1, z_i)^2 (-1) \\ &\quad + N^{K-3}(N-1)^{K-3} \sum_{2 \leq i_1 < i_2 \leq K} \sum_{z_1} \sum_{z_{i_1}} \sum_{z_{i_2}} \alpha_3(z_1, z_{i_1}, z_{i_2})^2 (-1)^2 \\ &\quad + \dots + N^0(N-1)^0 \sum_{z_1} \dots \sum_{z_K} \alpha_K(z_1, \dots, z_K)^2 (-1)^{K-1}. \end{aligned} \tag{19}$$

Multiplying by $N^{-K}(N-1)^{1-K}$ proves the lemma. □

By Lemma 4, under the monotonicity of μ_z in z_1 , our new sufficient condition for the variance reduction of LHS is given by

$$\begin{aligned} 0 &\leq \sum_{1 < i} \sum_{z_i} \alpha_1(z_i)^2 \cdot N^{-1}(N-1)^0(-1)^0 + \sum_{1 < i_1 < i_2} \sum_{z_{i_1}} \sum_{z_{i_2}} \alpha_2(z_{i_1}, z_{i_2})^2 \cdot N^{-2}(N-1)^{-1}(-1)^1 \\ &\quad + \sum_{1 < i_1 < i_2 < i_3} \sum_{z_{i_1}} \sum_{z_{i_2}} \sum_{z_{i_3}} \alpha_3(z_{i_1}, z_{i_2}, z_{i_3})^2 \cdot N^{-3}(N-1)^{-2}(-1)^2 \\ &\quad + \dots + \sum_{z_2} \dots \sum_{z_K} \alpha_{K-1}(z_2, \dots, z_K)^2 \cdot N^{1-K}(N-1)^{2-K}(-1)^{K-2}. \end{aligned} \tag{20}$$

Note that -1 times the right-hand side of this inequality corresponds to the covariance of partially averaged cell mean function:

$$\mu_{z_2 z_3 \dots z_K} = \frac{1}{N} \sum_{z_1=1}^N \mu_{z_1 z_2 \dots z_K},$$

which can be regarded as a function with $K-1$ arguments. In fact (20) is equivalent to $\text{cov}(\mu_{z_2 \dots z_K}, \mu_{y_2 \dots y_K}) \leq 0$ under LHS.

Therefore by assuming the monotonicity in any one axis, we have reduced our problem from K arguments to $K-1$ arguments. Employing the monotonicity in z_2 , we can further reduce the number of arguments. To complete this inductive argument we need the following initial condition.

Lemma 5. *Let $m = 1, K = 2$. Then monotonicity in one axis implies*

$$\text{cov}(\mu_{\mathbf{z}}, \mu_{\mathbf{y}}) \leq 0.$$

Proof.

$$(N - 1)\text{cov}(\mu_{\mathbf{z}}, \mu_{\mathbf{y}}) = -N^{-1} \left\{ \sum_{z_1} \alpha_1(z_1)^2 + \sum_{z_2} \alpha_1(z_2)^2 \right\} + N^{-2}(N - 1)^{-1} \sum_{z_1} \sum_{z_2} \alpha_2(z_1, z_2)^2.$$

As in (19) the monotonicity in z_1 implies

$$0 \geq -N^{-1} \left\{ \sum_{z_1} \alpha_1(z_1)^2 \right\} + N^{-2}(N - 1)^{-1} \sum_{z_1} \sum_{z_2} \alpha_2(z_1, z_2)^2.$$

Now the lemma follows from $\sum_{z_2} \alpha_1(z_2)^2 \geq 0$. □

Summarizing the above arguments we have established the following Proposition.

Proposition 2. *If the cell mean function $\mu_{\mathbf{z}}$ is monotone in any $K - 1$ of the K axes, then $\text{var}(T_L) \leq \text{var}(T_R)$.*

Since the ranges of $F_i^{-1}(((z_i - 1)/N, z_i/N])$, $i = 1, \dots, K$, are monotone in z_i 's, if g is monotone in the i -th axis then the cell mean function is also monotone in the i -th axis. Hence Corollary 1 holds as a consequence of Proposition 2.

B.2. ELHS

For general m we establish the proposition below.

Proposition 3. *Suppose that for all $1 \leq t \leq m$, t th-order monotonicities of $\mu_{\mathbf{z}}$ concerning $K - t$ axes hold, then $\text{var}(T_{EL}) \leq \text{var}(T_R)$.*

Proof. In view of (11) the variance reduction holds if

$$c_t = (-1)^{t-1} \sum_{s=t}^K \varphi_s^2 (-1)^s (N - 1)^{1-s} \binom{s-1}{t-1} \leq 0$$

for $1 \leq t \leq m$. We claim that t th-order monotonicity in any $K - t$ axes implies $c_t \leq 0$. The monotonicity in (z_1, z_2, \dots, z_t) implies

$$(\Delta_{z_1, z'_1} \dots \Delta_{z_t, z'_t} \mu_{\mathbf{z}})(\Delta_{z_1, z'_1} \dots \Delta_{z_t, z'_t} \mu_{\mathbf{y}}) \geq 0,$$

for all $\mathbf{z}, z'_1, \dots, z'_t, y_{t+1}, \dots, y_K$. Summing these as in Lemma 4 yields

$$\begin{aligned}
 0 &\leq \sum_{z_1} \sum_{z_2} \cdots \sum_{z_t} \alpha_t(z_1, z_2, \dots, z_t)^2 (N(N-1))^{-1} \\
 &\quad - \sum_{t < i} \sum_{z_1} \cdots \sum_{z_t} \sum_{z_i} \alpha_{t+1}(z_1, \dots, z_t, z_i)^2 (N(N-1))^{-2} \\
 &\quad + \sum_{t < i_1 < i_2} \sum_{z_1} \cdots \sum_{z_t} \sum_{z_{i_1}} \sum_{z_{i_2}} \alpha_{t+2}(z_1, \dots, z_t, z_{i_1}, z_{i_2})^2 (N(N-1))^{-3} \\
 &\quad + \dots + (-1)^{K-t} \sum_{z_1} \cdots \sum_{z_K} \alpha_K(z_1, \dots, z_K)^2 (N(N-1))^{-(K-t+1)}.
 \end{aligned}$$

In order to use induction on z_1 , we utilize $\binom{K-1}{t-1}$ monotonicities concerning z_1 . By symmetry and

$$\frac{\binom{K-1}{t-1} \binom{K-t}{s-t}}{\binom{K-1}{s-1}} = \binom{s-1}{t-1},$$

t th-order of the monotonicity in z_1 implies

$$\begin{aligned}
 0 &\leq \sum_{1 < i_1 < \dots < i_{t-1}} \sum_{z_1} \sum_{z_{i_1}} \cdots \sum_{z_{i_{t-1}}} \alpha_t(z_1, z_{i_1}, \dots, z_{i_{t-1}})^2 (N(N-1))^{-1} \binom{t-1}{t-1} \\
 &\quad - \sum_{1 < i_1 < \dots < i_t} \sum_{z_1} \sum_{z_{i_1}} \cdots \sum_{z_{i_t}} \alpha_{t+1}(z_1, z_{i_1}, \dots, z_{i_t})^2 (N(N-1))^{-2} \binom{t}{t-1} \\
 &\quad + \sum_{1 < i_1 < \dots < i_{t+1}} \sum_{z_1} \sum_{z_{i_1}} \cdots \sum_{z_{i_{t+1}}} \alpha_{t+2}(z_1, z_{i_1}, \dots, z_{i_{t+1}})^2 (N(N-1))^{-3} \binom{t+1}{t-1} \\
 &\quad + \dots + (-1)^{K-t} \sum_{z_1} \cdots \sum_{z_K} \alpha_K(z_1, \dots, z_K)^2 (N(N-1))^{-(K-t+1)} \binom{K-1}{t-1}.
 \end{aligned}$$

We see that the monotonicity leads to the reduction of one argument by $N^{t-1}(N-1)^{t-2} \geq 0$. By induction $K - (t + 1)$ t th-order monotonicities lead to the initial condition

$$\begin{aligned}
 0 &\leq \sum_{i_1 < \dots < i_t} \sum_{z_{i_1}} \cdots \sum_{z_{i_t}} \alpha_t(z_{i_1}, \dots, z_{i_t})^2 (N(N-1))^{-1} \binom{t-1}{t-1} \\
 &\quad - \sum_{z_1} \cdots \sum_{z_{t+1}} \alpha_{t+1}(z_1, \dots, z_{t+1})^2 (N(N-1))^{-2} \binom{t}{t-1},
 \end{aligned}$$

which is implied by t th-order monotonicity in z_1 . This proves the proposition. □

Then, as in Corollary 1, Theorem 2 is an immediate consequence of Proposition 3.

REFERENCES

- Bose, R.C. and Bush, K.A. (1952) Orthogonal arrays of strength two and three. *Ann. Math. Statist.*, **23**, 508–524.
- Handcock, M.S. (1991) On cascading Latin hypercube designs. *Comm. Statist. Theory Methods*, **20**(2), 417–439.
- Hoshino, N. and Takemura, A. (1997) On reduction of finite sample variance by extended Latin hypercube sampling. Discussion Paper 97-F-7, Faculty of Economics, University of Tokyo. (Available electronically from <ftp://lms.e.u-tokyo.ac.jp/pub/takemura/dp97f7.ps>)
- Iman, R.L. and Conover, W.J. (1982) A distribution-free approach to inducing rank correlation among input variables. *Comm. Statist. Simulation Comput.*, **11**, 311–334.
- Lehmann, E.L. (1966) Some concepts of dependence. *Ann. Math. Statist.*, **35**, 1137–1153.
- Loh, W.L. (1996) On Latin hypercube sampling. *Ann. Statist.*, **24**(5), 2058–2080.
- McKay, M.D., Conover, W.J. and Beckman, R.J. (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**, 239–245.
- Niederreiter, H. (1992) *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia: SIAM.
- Owen, A.B. (1992a) Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica*, **2**, 439–452.
- Owen, A.B. (1992b) A central limit theorem for Latin hypercube sampling. *J. Roy. Statist. Soc. Ser. B*, **54**, 541–551.
- Owen, A.B. (1994a) Lattice sampling revisited: Monte Carlo variance of means over randomized orthogonal arrays. *Ann. Statist.*, **22**(2), 930–945.
- Owen, A.B. (1994b) Controlling correlations in Latin hypercube samples. *J. Amer. Statist. Assoc.*, **89**, 1517–1522.
- Plackett, R.L. and Burman, J.P. (1946) The design of optimum multifactorial experiments. *Biometrika*, **33**, 305–325.
- Rao, C.R. (1947) Factorial experiments derivable from combinatorial arrangements of arrays. *J. Roy. Statist. Soc.*, **9**, Suppl., 128–139.
- Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
- Shaw, J.E.H. (1988) A quasirandom approach to integration in Bayesian statistics. *Ann. Statist.*, **16**(2), 895–914.
- Stein, M. (1987) Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, **29**, 143–151.
- Tang, B. (1993) Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.*, **88**, 1392–1397.
- Tang, B. (1994) A theorem for selecting OA-based Latin hypercubes using a distance criterion. *Comm. Statist. Theory Methods*, **23**(7), 2047–2058.

Received June 1997 and revised July 1999.