Statistica Sinica 2(1992), 439-452

ORTHOGONAL ARRAYS FOR COMPUTER EXPERIMENTS, INTEGRATION AND VISUALIZATION

Art B. Owen

Stanford University

Abstract: This paper uses orthogonal arrays to define generalizations of Latin hypercube sampling and of lattice sampling in the d dimensional unit cube. These are proposed as suitable designs for computer experiments, numerical integration and visualization. The orthogonal array based designs extend to t dimensional margins the univariate stratification properties of Latin hypercube and lattice sampling. As a consequence, the variance reduction property of Latin hypercube and lattice sampling also extends to orthogonal array based samples. We give a sample based estimate of the error variance in the case of bivariate stratification.

Key words and phrases: Monte Carlo, variance reduction, quadrature.

1. Introduction

Let $X \in [0,1]^d$ and $Y = f(X) \in R$, where f is a measurable function. Computer experiments, as described by Sacks, Mitchell, Welch and Wynn (1989) and by Currin, Mitchell, Morris and Ylvisaker (1988) are conducted by computing f at a set of points $X_1, \ldots, X_n$ and making a Bayesian interpolation to $[0,1]^d$. The interpolated function can be orders of magnitude faster to compute than f itself, and may therefore be more suitable for exploration. Owen (1992) describes computer experiments in which regression models are fit to pairs (X_i, Y_i) using randomness in the selection of the X_i to assess the models. A sort of informal computer experimentation in which one evaluates f at $X_1, \ldots, X_n$ and then uses techniques of interactive data analysis to visualize relationships among X and Y can also be informative. In numerical integration, many methods estimate $\int f(X)dX$ by $\bar{Y} = n^{-1}\sum_{i=1}^n f(X_i)$ for carefully chosen X_i.

In all of these problems there is a design issue in choosing $X_1, \ldots, X_n$. This article presents a class of designs based on randomized orthogonal arrays that are suitable as designs for computer experimentation, integration and visualization. This class includes and generalizes Latin hypercube sampling (McKay, Conover and Beckman (1979)).

Section 2 of this paper discusses Latin hypercube sampling (McKay, Conover and Beckman (1979)) and lattice sampling (Patterson (1954)). Section 3 introduces orthogonal arrays and shows how they can be used to generalize Latin hypercube and lattice sampling. The result is a set of sampling plans that stratify on all bivariate or trivariate (or t-variate) margins simultaneously. When used as numerical integration schemes these methods remove the contribution to the Monte Carlo variance from low order interactions in the integrand. Patterson's (1954) results are used to give an expression for the error variance. Section 4 shows how to estimate the Monte Carlo variance from the data in a sample with bivariate stratification.

The rest of this section introduces notation, defines an anova decomposition for square integrable functions on $[0,1]^d$ and presents some elementary facts used in the other sections.

We shall use X to denote a row vector in $[0,1]^d$ with components $X^1,\dots,X^d$. Similarly $X_i = (X_i^1,\dots,X_i^d) \in [0,1]^d$ for $1 \le i \le n$. A matrix with elements X_i^j has rows X_i and columns X^j.

Following Efron and Stein (1981) and others cited by them, we introduce an anova decomposition for continuous functions of independent random variables. First we assume that $\int f^2 dF < \infty$. The grand mean is the integral

$$\mu = \int f(X) dF$$

where $dF = \prod_{j=1}^d dX^j$ is uniform measure on $[0,1]^d$. We use dF_j, dF_{jk}, dF_{-j} and dF_{-jk} for dX^j, $dX^j dX^k$, $\prod_{m\neq j} dX^m$ and $\prod_{m\neq j, m\neq k} dF^m$ respectively.

The main effects are

$$\alpha_j(X^j) = \int (f-\mu) dF_{-j}, \quad 1 \le j \le d.$$

The two factor interactions are

$$\alpha_{jk}(X^j, X^k) = \int (f - \mu - \alpha_j - \alpha_k) dF_{-jk}, \quad j \neq k.$$

Continuing this process we may expand f as

$$f(X) = \mu + \sum_j \alpha_j(X^j) + \sum_{j<k} \alpha_{jk}(X^j, X^k) + \cdots + \alpha_{1\dots d}(X^1,\dots,X^d).$$

In general, let $u, v, w \subseteq D = \{1,\dots,d\}$ denote subsets of the axes of $[0,1]^d$. We use dF_u for integration with respect to the axes in u, leaving a function defined over the axes in $D-u$. That is $dF_u = \prod_{j\in u} dX^j$. Then

$$f = \sum_{u \subseteq D} \alpha_u$$

where

$$\alpha_u = \int \left(f - \sum_{v \subset u} \alpha_v \right) dF_{D-u}. \tag{1.1}$$

The sum in (1.1) over $v \subset u$ is understood to be over proper subsets only, that is $v \neq u$. For $u = \emptyset$ the empty set, $\alpha_u = \mu$ and an integral with respect to dF_u is the integrand as a function over $[0,1]^d$.

The expression dF_j above means dF_u where $u = \{j\}$. Whether the subscripting of dF and α is by indices or sets will be clear from context. A term like $\alpha_u(X)$ means the function α_u applied to the components of X in u, ignoring the components in $D - u$.

The continuous anova generalizes the familiar discrete anova. It is easy to show by induction that

$$\int \alpha_u dF_j = 0, \quad \text{if} \quad j \in u,$$

from which it follows that

$$\int \alpha_u \alpha_v dF = 0, \quad u \neq v$$

by writing $dF = dF_j dF_{-j}$ for $j \in u \cup v - u \cap v$ and so

$$\int f^2 dF = \sum_{u \subseteq D} \int \alpha_u^2 dF.$$

Measurability of f implies measurability of each α_u by Fubini's theorem. See Royden (1968, p.269) where the definition of integrability implies measurability. So if X is a random vector, $\alpha_u(X)$ is a random variable. We make no further mention of measurability.

Here we record some useful facts about integrals over hypercubes. For more details see Davis and Rabinowitz (1984). Let $C = [0,s]^d$ be a hypercube of side s in d dimensions. Let $x_0 = (s/2, \ldots, s/2)$ be the center of C, and let $X \sim \text{unif}(C)$. Then if f is sufficiently smooth, we have as $s \downarrow 0$

$$E(f(X)) = f(x_0) + \frac{s^2}{24} \nabla^2 f(x_0) + O(s^4) \tag{1.2}$$

and

$$\text{var}(f(X)) = \frac{s^2}{12} \|\nabla f(x_0)\|^2 + O(s^4) \tag{1.3}$$

where

$$\nabla f(x_0) = \left(\frac{\partial f(x_0)}{\partial X^1}, \ldots, \frac{\partial f(x_0)}{\partial X^d} \right)'$$

is the gradient of f at x_0 and

$$\nabla^2 f(x_0) = \sum_{j=1}^{d} \frac{\partial^2}{\partial X^j \partial X^j} f(x_0)$$

is the Laplacian of f at x_0. The smoothness needed in (1.2) is boundedness of the fourth derivatives of f. For (1.3) bounded third derivatives are needed. If only a bounded third derivative exists, the error in (1.2) becomes $O(s^3)$. If only a bounded second derivative exists, the error in (1.3) becomes $O(s^2)$ and the right side of (1.2) becomes $f(x_0) + O(s^2)$.

Now we consider the accuracy of some midpoint integration rules. Partition the cube $[0,1]^d$ into a q^d grid of subcubes of side $s = q^{-1}$. Now let G be a q^d by d matrix with rows given by the centers of the subcubes. The rows of G are the points of a regular q^d grid. To construct G, write each $i = 0, \ldots, q^d - 1$ in base q as $i = a_1(i) + a_2(i)q + \cdots + a_d(i)q^{d-1}$ where each $a_j(i) \in \{0, \ldots, q-1\}$ and take

$$G_i^j = \frac{a_j(i-1) + 1/2}{q}, \quad 1 \le i \le q^d,\ 1 \le j \le d. \tag{1.4}$$

The d dimensional midpoint rule approximates $\int f dF$ by $q^{-d} \sum f(G_i)$. From (1.2) we have

$$\begin{aligned} \int f dF &= q^{-d} \sum_{1 \le i \le q^d} \left(f(G_i) + (24q^2)^{-1} \nabla^2 f(G_i) + O(q^{-4}) \right) \\ &= q^{-d} \sum_{1 \le i \le q^d} f(G_i) + (24q^2)^{-1} \int \nabla^2 f dF + O(q^{-4}). \end{aligned} \tag{1.5}$$

A uniform rectangle rule is formed by taking

$$\tilde{G}_i^j = G_i^j + \frac{U_i^j - 1/2}{q}, \quad 1 \le i \le q^d,\ 1 \le j \le d \tag{1.6}$$

where the U_i^j are independent $U[0,1]$ random variables and G_i^j is given by (1.4). From here on we shall use the tilde, as in $\tilde{X}_i$, to designate design points with a uniform random component. Now

$$E\left(q^{-d} \sum_{1 \le i \le q^d} f(\tilde{G}_i) \right) = \int f dF$$

and from (1.3) we have

$$\operatorname{var}\left(q^{-d} \sum_{1 \le i \le q^d} f(\tilde{G}_i) \right) = q^{-d-2} \left(\frac{1}{12q^d} \sum_{1 \le i \le q^d} \|\nabla f(G_i)\|^2 \right) + O(q^{-d-4}). \tag{1.7}$$

As an estimate of μ, the midpoint rule has bias $O(q^{-2})$ and variance 0 whereas the uniform rectangle rule has bias 0 and variance $O(q^{-d-2})$. The root-mean square error is therefore $O(q^{-2})$ under the midpoint rule and $O_p(q^{-(d+2)/2})$ under the uniform rectangle rule. A simple Monte Carlo scheme with independent $X_i \sim U[0,1]^d, i = 1, \ldots, n$, has root mean square error $O_p(n^{-1/2})$. For comparison take $n = q^d$. The midpoint rule attains a better rate of accuracy than Monte Carlo for $d < 4$; it attains the same rate if $d = 4$ and Monte Carlo is superior for $d > 4$. The uniform rectangle rule has root mean square error of order $O_p(n^{-1/2-1/d})$ which is better than the Monte Carlo rate for any d, but the advantage is negligible for large d and requires a very large sample. For $d = 1$ the midpoint rule has a better order of accuracy than the uniform rectangle rule, for $d = 2$ the rates are the same and for $d > 2$ the uniform rectangle rule has a better rate than the midpoint rule.

If q is even one can improve on the midpoint rule by taking

$$G_i^j = \begin{cases} (a_j(i-1) + 1 - 3^{-1/2})/q, & \text{if } a_j(i-1) \text{ is odd,} \\ (a_j(i-1) + 3^{-1/2})/q, & \text{if } a_j(i-1) \text{ is even,} \end{cases} \tag{1.8}$$

where $a_j(i)$ are given above for (1.4). Then $\bar{Y}$ becomes a product Gauss rule of order two (Davis and Rabinowitz (1984, p.95)) and has error $O(q^{-3}) = O(n^{-3/d})$ which is better than Monte Carlo for $d < 6$ and better than the uniform rectangle rule for $d < 4$.

2. Latin Hypercube Samples and Lattice Samples

A uniform Latin hypercube sample of size n has

$$\tilde{X}_i^j = \frac{\pi_j(i) - U_i^j}{n}, \quad 1 \le i \le n, \ 1 \le j \le d \tag{2.1}$$

where $\pi_j(1), \ldots, \pi_j(n)$ is a random permutation of the integers $1, \ldots, n$ (uniform over the $n!$ permutations), $U_i^j \sim U[0,1]$ and the d permutations and nd uniform variates are mutually independent. In the general case of a Latin hypercube sample, an inverse cdf transformation is applied to each column $\tilde{X}^j$. We shall use only the uniform version, subsuming the transformations in f. McKay, Conover and Beckman (1979) introduce Latin hypercube sampling in order to sample the input space of computer programs.

Stein (1987) shows that Latin hypercube sampling provides asymptotically more accurate estimates of integrals than i.i.d. sampling. His result is that

$$\mathrm{var}_{LHS}(\bar{Y}) = \frac{1}{n} \int e(X)^2 dF + o(n^{-1}) \tag{2.2}$$

where $e(X) = f(X) - \mu - \alpha_1(X^1) - \cdots - \alpha_d(X^d)$ is the non-additive part of f. By contrast, for a design of n independent $U[0,1]^d$ random vectors

$$\begin{aligned}\mathrm{var}_{IID}(\bar{Y}) &= \frac{1}{n}\int (f(X)-\mu)^2 dF \\ &= \frac{1}{n}\left(\int e(X)^2 dF + \int \alpha_1^2 dF_1 + \cdots + \int \alpha_d^2 dF_d\right)\end{aligned}$$

Owen (1992) shows that for bounded f

$$n^{1/2}(\bar{Y}-\mu) \to N\left(0, \int e(X)^2 dF\right)$$

in distribution as $n \to \infty$ under Latin hypercube sampling. The proof actually shows that for all integers $p \geq 1$

$$E_{LHS}\left(n^{1/2}(\bar{Y}-\mu)\right)^p = E_{IID}\left(n^{1/2}\bar{e}\right)^p + O\left(n^{-1}\right)$$

so that to this order of accuracy $\bar{Y}-\mu$ is distributed as the mean, $\bar{e}$, of an iid sample of size n from $e(X)$.

In a lattice sample we have

$$X_i^j = \frac{\pi_i^j - 1/2}{n}, \quad 1 \leq i \leq n,\ 1 \leq j \leq d. \tag{2.3}$$

That is each U_i^j in the Latin hypercube sample (2.1) has been replaced by $1/2$. Patterson (1954) attributes the term "lattice sample" to Yates. Lattice samples also remove the additive part of f from the error.

In a lattice sample we may write

$$\begin{aligned}\bar{Y} &= \frac{1}{n}\sum_{i=1}^n \left(\mu + \alpha_1(X_i^1) + \cdots + \alpha_d(X_i^d) + e(X_i)\right) \\ &= \mu + \sum_{j=1}^d \frac{1}{n}\sum_{i=1}^n \alpha_j(X_i^j) + \bar{e} \\ &= \mu + (24n^2)^{-1}\sum_{j=1}^d (\alpha_j'(0) - \alpha_j'(1)) + \bar{e} + O(n^{-4})\end{aligned}$$

by applying (1.5) with $f = \alpha_j$, $d = 1$ and $q = n$, where $\alpha_j' = d\alpha_j/dX^j$. The lattice sample integrates the main effects of f by a midpoint rule with $d = 1$ and $q = n$ with an error that is negligible compared to the Monte Carlo error in $\bar{e}$. Similarly the Latin hypercube sample uses a uniform rectangle rule with $d = 1$ and $q = n$ on the main effects, and this has an error that is negligible by

comparison to $\bar{e}$. For main effects $d = 1$, and the lattice method eliminates them more effectively than does the Latin hypercube sample.

By closely matching the sample univariate margins to the corresponding population univariate margins, the terms α_j are "filtered out" of the error. This naturally raises the issue: can we filter out higher order terms by matching higher order population and sample margins? The answer is yes, at least for some combinations of n and d.

3. Orthogonal Arrays

An orthogonal array of strength t is a matrix of n rows and k columns with elements taken from a set of q symbols, such that in any n by t submatrix each of the q^t possible rows occurs the same number λ of times. Clearly $\lambda q^t = n$. Such an array is denoted by $OA(n, k, q, t)$. Orthogonal arrays are described in Raghavarao (1971). In the standard definition, $OA(n, k, q, t)$ has n columns and k rows (called constraints) and is the transpose of the matrix described above. We have transposed the arrays so that rows and columns of the array correspond to rows and columns of design matrices constructed from them.

The lattice sample X given by (2.3) is an orthogonal array of strength 1, $OA(n, d, n, 1)$. The grid G given by (1.4) is an orthogonal array of strength $d, OA(q^d, d, q, d)$.

The symbols of an orthogonal array A are ordinarily taken to be $0, 1, \ldots, q-1$. The elementwise mappings $X_i^j = (A_i^j + 1/2)/n$ and $G_i^j = (A_i^j + 1/2)/q$ produce a lattice sample and a grid from $OA(n, d, n, 1)$ and $OA(q^d, d, q, d)$ respectively.

The arrays $OA(q^2, k, q, 2)$ have $n = q^2$ points that plot as a q by q grid in each bivariate margin. This makes them a good choice for exploration of functions when it is thought that important bivariate effects might exist. In VLSI applications oxidation time and temperature might interact as might an ion implant dose and energy. It is valuable to have some points in each "corner" that might be interesting. There are $4\binom{k}{2} = 2k(k-1)$ bivariate corners. Random samples and even Latin hypercube samples often fail to have points near all such corners. These strength 2 designs exercise the program computing f in a more thorough way than does Latin hypercube sampling. They may also make good designs for fitting models such as MARS (Friedman (1991)) by obviating the need to interpolate bivariate effects estimated at the data points onto a plotting grid. We show below that these arrays are good for integration in that they can be used to filter out bivariate effects.

Tang (1991) independently and contemporaneously had the idea of using orthogonal arrays to construct designs for computer experiments. He constructs Latin hypercube samples with q^2 runs that become orthogonal arrays of strength 2 after grouping observations using q bins on each axis. Tang further shows that

these designs achieve the same order of variance reduction as orthogonal arrays of strength 2, while their Latin hypercube property means that they use q^2 distinct values on each variable instead of the q used by the straight orthogonal array. This advantage becomes more important with the smaller computer experiments appropriate for the more expensive simulators. For larger experiments one might prefer the orthogonal arrays on account of their better balance.

A collection of orthogonal arrays of strength 2 has been contributed to statlib. A listing of the contents of statlib, and instructions on retrieving them may be obtained by sending an electronic mail message to *statlib@lib.stat.cmu.edu* with the line *send index* as the sole message. The contributed designs are of the form $OA(q^2, q+1, q, 2)$ for $q = 2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, 19, 25, 27, 32$ and $OA(2q^2, 2q+1, q, 2)$ for $q = 2, 3, 4, 5, 7, 8, 9, 11, 13, 16$. The former are $q^{(q+1)-(q-1)}$ fractional factorials of resolution III; that is, main effects are unconfounded with each other but are confounded with two factor interactions. Matrices with $q = 23, 29, 31$ are not included. They are very large and are much easier to generate than those for $q = 25, 27, 32$. It should be mentioned that the designs with $2q^2$ columns have some rows that agree in three columns. While they are still valuable for exploring functions and provide good estimates of integrals, the variance formula (3.7) below does not apply to them. From here on we consider only orthogonal arrays of strength t in which no two rows agree on any $t+1$ columns.

Arrays of strength $t > 2$ require quite large sample sizes for modest q and hence would seem to be of less practical use at present. This concern notwithstanding, much of the analysis below is given in terms of general strength t.

If the symbols $0, \ldots, q-1$ in a column of an orthogonal array are permuted, the result is an orthogonal array of the same type. Therefore we can generalize the lattice samples (2.3) by taking

$$X_i^j = \frac{\pi_j(A_i^j) + 1/2}{q}, \tag{3.1}$$

where the π_j are independent permutations of $0, \ldots, q-1$, all $q!$ permutations being equally probable. Similarly we can generalize the Latin hypercube samples (2.1) by taking

$$\tilde{X}_i^j = \frac{\pi_j(A_i^j) + U_i^j}{q}, \tag{3.2}$$

where $U_i^j \sim U[0,1]$ independently of each other and of the π_j. Patterson (1954) discusses the analysis of designs like $OA(\lambda q^t, k, q, t)$, but does not address their existence. The nonexistence of $OA(6^2, 4, 6, 2)$ is well known; no Graeco-Latin square of side 6 exists.

Let G_i^j be the q^d grid given by (1.4). Then for the lattice type integrals we can write

$$\bar{Y} - \mu = \left(\bar{Y} - q^{-d}\sum f(G_i)\right) + \left(q^{-d}\sum f(G_i) - \mu\right). \tag{3.3}$$

The first term in (3.3) is the error in estimating the mean of a finite population of size q^d by the mean over the sample of n observations given by X_i. The second term in (3.3) is the error in a midpoint rule of dimension d. These will be called the sampling error and quadrature error respectively. From (1.5) we know that the quadrature error is $O(q^{-2})$. If A_i^j is an array of strength t, then $n = \lambda q^t$ and so the quadrature error is $O(n^{-2/t})$. We assume here that λ is not increasing with n. The sampling error is $O_p(n^{-1/2})$ and will be considered below. Here we note that strength $t \le 3$ is necessary for the quadrature error be asymptotically negligible compared to the Monte Carlo error. Using a Gauss rule grid (1.8) makes the quadrature error $O(q^{-3}) = O(n^{-3/t})$, improving on Monte Carlo for $t \le 5$.

The sampling error in (3.3) is of a type studied by Patterson (1954). If X_i is a row of the design in (3.1) and G_m is a row in the grid (1.4) , then $P(X_i = G_m) = q^{-d}$ by the randomization. Since no two X_i are identical, the probability that G_m appears in the sample is λq^{t-d}. Since this probability is the same for all G_m, the mean sampling error is zero.

To study the variance, define discrete main effects and interactions on the grid. Let $F^{(q)}$ denote the discrete measure with an atom of probability q^{-d} on each of the q^d rows G_i, and so let

$$\nu = \int f(X) dF^{(q)}$$

denote the mean of $f(X)$ over the grid points $G_i \in [0,1]^d$. Similarly define $dF_u^{(q)}$ by analogy to dF_u and discrete effects

$$\beta_u = \int \left(f - \sum_{v \subset u} \beta_v\right) dF_{D-u}^{(q)}$$

by analogy with α_u in (1.1). It is clear that effects u of cardinality $|u| \le t$ are removed from the error, since we can write

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n}\sum_{u \subseteq D} \beta_u(X_i)$$

and note that $|u| \le t$ implies that

$$\sum_{i=1}^{n} \beta_u(X_i) = \lambda q^t \int \beta_u dF_u^{(q)} = 0.$$

If follows that if $t \le 3$ and f has an anova decomposition with $\alpha_u = 0$ for $|u| > t$, then $\bar{Y} - \mu$ has a smaller order of magnitude than it would have under Monte Carlo sampling. More practically, if the α_u for $|u| > t$ are small compared to those for $|u| \le t$, then $\bar{Y} - \mu$ has the same order of magnitude as it would have under Monte Carlo sampling but has a smaller asymptotic variance, which we study below.

Let SS_u denote the sum of squares in the analysis of variance of $f(G_i)$ associated with the effect u. This sum of squares has $(q-1)^{|u|}$ degrees of freedom. Patterson decomposes the corresponding mean square into variance components, via

$$MS_u = \sum_{v \supseteq u} q^{d-|v|}\sigma_v^2. \tag{3.4}$$

The decomposition does not guarantee that $\sigma_v^2 \ge 0$, but all the MS_u are nonnegative. This implies, for example, that $\sigma_{-j}^2 \ge -\sigma_D^2/q$. Then for X obtained from $OA(q^t, d, q, t)$ with independent randomizations applied to all d columns

$$\mathrm{var}(\bar{Y}) = q^{-t} \sum_{|u|>t} \sigma_u^2 \left(1 - q^{t-|u|}\right). \tag{3.5}$$

Formula (3.5) is Equation (4.1) of Patterson (1954) after translating an effective variance per unit into the variance of a mean. The factors $(1 - q^{t-|u|})$ are finite population corrections. It follows from (3.5) that

$$(1-q^{-1})n^{-1} \sum_{|u|>t} \sigma_u^2 \le \mathrm{var}(\bar{Y}) \le n^{-1} \sum_{|u|>t} \sigma_u^2. \tag{3.6}$$

Equation (3.5) is obtained for $\lambda = 1$. For $A = OA(\lambda q^t, d, q, t)$ Patterson's (4.3) translates into

$$\mathrm{var}(\bar{Y}) = n^{-1} \sum_{|u|>t} \sigma_u^2 \left(1 - \lambda q^{t-|u|}\right) \tag{3.7}$$

assuming that no two rows of A agree on any $t+1$ columns.

A referee points out that Patterson (1954) states his formulas (4.1) and (4.3) without proof. By contrast Tang (1991) gives a proof of the variance reduction property for Latin hypercube samples based on orthogonal arrays of strength 2. The author has been able to show by a direct argument that for $A = OA(q^2, d, q, 2)$

$$\mathrm{var}(\bar{Y}) = n^{-1} \sum_{|u|>2} \int \alpha_u^2 dF \left(1 + O(n^{-1/2})\right).$$

Tang's proof was obtained earlier.

Now consider uniform rectangle rules (3.2) instead of midpoint rules (3.1). Without loss of generality we may suppose that the uniform errors in $\tilde{G}$ as given by (1.6) and $\tilde{X}$ from (3.2) are coupled so that if $A_i = qG_m - 1/2$, where G is the grid used to define $\tilde{G}$ in (1.6), then U_i^j from (3.2) is equal to U_m^j from (1.6). We also couple the random permutations in (3.2) to those in (3.1) so that $\tilde{X}_i^j = X_i^j + (U_i^j - 1/2)/q$. In words, uniform errors are added to the points of the grid G, giving $\tilde{G}$ and then $\tilde{X}$ is sampled from $\tilde{G}$ according to the same randomization of the orthogonal array A that was used to define X. The only difference between integrals estimated using the uniform rectangle rule instead of the midpoint rule comes from the n random U_i^j's actually used in the sample.

The bias in $\tilde{Y} = n^{-1}\sum_{i=1}^n f(\tilde{X}_i)$ is zero because in $E(E(f(\tilde{X}_i)|\pi_1,\ldots,\pi_d))$ the inner expectation is uniform over one of the q^d subcubes of $[0,1]^d$ and in the outer expectation each of the subcubes has the same probability nq^{-d}. To investigate the variance we write

$$\begin{aligned}&\operatorname{var}\left(\frac{1}{n}\sum f(\tilde{X}_i) - f(X_i)\right)\\ &= \operatorname{var}\left(E\left(\frac{1}{n}\sum f(\tilde{X}_i) - f(X_i)\mid \tilde{G}\right)\right) + E\left(\operatorname{var}\left(\frac{1}{n}\sum f(\tilde{X}_i) - f(X_i)\mid \tilde{G}\right)\right) \quad (3.8)\end{aligned}$$

where conditioning on $\tilde{G}$ fixes the q^d uniform variates but leaves the permutations π_j random. The first term in (3.8) is

$$\operatorname{var}\left(q^{-d}\sum_{1\le i\le q^d} f(\tilde{G}_i) - f(G_i)\right) = O\left(q^{-d-2}\right)$$

by (1.7). Since $n = \lambda q^t$ and λ is assumed to be $O(1)$, this becomes $O(n^{-(d+2)/t})$. The variance in the second term of (3.8) is asymptotically smaller than what it would be if the sampling given $\tilde{G}$ were iid, meaning sampling with replacement from the population of q^d subcubes. Under Lipschitz continuity of f, $f(\tilde{X}_i) - f(X_i) = O(q^{-1})$. So the second term in (3.8) is $O(n^{-1}q^{-2})$, provided only that f is Lipschitz continuous.

It follows that the difference between estimates based on (3.2) and (3.1) has standard deviation $o(n^{-1/2})$. Method (3.2) will be asymptotically superior in cases where the bias of method (3.1) is not $O(n^{-1/2})$. Method (3.1) can be asymptotically superior only in those uncommon cases with $n = q^d$ and hence no Monte Carlo error. Though (3.2) will usually have a better rate in most common asymptotic settings, it should be borne in mind that where the additive terms in f are dominant, method (3.1) may be much better for small n.

4. Variance Estimation

Owen (1992) gives a way of estimating the error variance in Latin hypercube sampling. It is possible to estimate $\int(f-\mu-\sum_j \alpha_j)^2 dF$ at the $n^{1/2}$ rate even though this rate is not generally available for the α_j themselves.

The method is as follows. For $M \geq 1$ let

$$\hat{\alpha}_j(X_i^j) = (2M)^{-1} \sum_{m=\pm 1,\ldots,\pm M} N_j^m Y_i,$$

where for $m > 0, N_j^m Y_i$ is the value of Y observed m observations to the right of X_i along axis j. That is $N_j^m Y_i = Y_k$ when $X_k^j = X_i^j + m/q$. For negative m the observation $|m|$ units to the left is taken. If X_i^j is so near to 0 or 1 that some of the $N_j^m Y_i$ do not exist, an arbitrary value like $\bar{Y}$ or Y_i may be substituted for the missing neighbors. Then Lipschitz continuity of the α_j and $M = O(n^{1/2})$ as $n \to \infty$ leads to the estimate

$$\hat{\tau}_j^2 = (1+2M)^{-1} n^{-1} \sum_{i=1}^n \left(Y_i - \hat{\alpha}_j(X_i^j)\right)^2 = \int (f-\mu-\alpha_j)^2 dF + O_p\left(n^{-1/2}\right)$$

of $\tau_j^2 = \int(f-\mu-\alpha_j)^2 dF$, the residual variance from a nonparametric regression on X^j. $M = 1$ is adequate. Integration with respect to dF_j gives additional smoothing allowing estimation of τ_j^2 at the $n^{1/2}$ rate, even though that rate is not achievable for α_j. For a general discussion of estimation of residual variance in nonparametric regression see Buckley, Eagleson and Silverman (1988). We can estimate $\int(f-\mu-\sum_j \alpha_j)^2 dF$ by

$$\sum_{j=1}^d \hat{\tau}_j^2 - \frac{d-1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

In this section we extend this idea to residual variance in randomized orthogonal array samples of strength 2 and $\lambda = 1$. We consider only the midpoint type rules.

Let X_i^j be given by (3.1), where A is $OA(q^2, d, q, 2)$. From Equation (3.5)

$$\begin{aligned}
\mathrm{var}(\bar{Y}) &= q^{-2} \sum_{|u|>2} \sigma_u^2 \left(1 - q^{|u|-2}\right) = q^{-2} \sum_{|u|>2} \sigma_u^2 + O\left(q^{-3}\right) \\
&= q^{-2} \sum_{|u|>2} MS_u + O\left(q^{-3}\right) = q^{-2} \sum_{|u|>2} \int \beta_u^2 dF^{(q)} + O\left(q^{-3}\right) \\
&= q^{-2} \left(\int (f-\nu)^2 dF^{(q)} - \sum_j \int \beta_j^2 dF^{(q)} - \sum_{j<k} \int \beta_{jk}^2 dF^{(q)} \right) + O\left(q^{-3}\right).
\end{aligned}$$

Of the q^2 values of X_i, $(q-2)^2$ of them are not on an edge or corner of the design in the jk plane. They therefore have 4 nearest neighbors in the jk plane and we take their average as an undersmoothed estimate $\hat{\alpha}_{jk}(X_i)$. For points on an edge in the jk plane but not in a corner there are two neighbors in the edge whose average we take to be $\hat{\alpha}_{jk}$. For corner points we simply take $\hat{\alpha}_{jk}(X_i) = Y_i$ although we could also use the sum of their edge neighbors minus their diagonal neighbor in the jk plane. If α_{jk} has a bounded first derivative, then

$$\hat{\tau}^2_{jk} = \frac{4}{5n}\sum_{i=1}^{n}(Y_i - \hat{\alpha}(X_i))^2$$

is $n^{-1/2}$ consistent for

$$\sum_{u \not\subseteq \{j,k\}} \int \alpha_u^2 dF.$$

The multiplier $4/5n$ is within a relative error $O(q^{-1})$ of the reciprocal of the trace of the quadratic form $\sum_{i=1}^{n}(Y_i - \hat{\alpha}(X_i))^2$. A more closely calculated denominator should be used in practice.

Now let MSE be the mean square for error from a main effects analysis of variance of the Y_i. This estimates $\sum_{|u|\geq 2}\int \alpha_u^2 dF$ at the $n^{1/2}$ rate. Finally we take

$$\widehat{\mathrm{var}}(\bar{Y}) = q^{-2}\left(\sum_{j<k}\hat{\tau}^2_{jk} - \left(\binom{d}{2} - 1\right) MSE\right)$$

We hasten to add the caveat that while the above method shows the possibility in theory of getting $n^{1/2}$ consistent estimates of the error variance, in applications certain inconsistent estimators might be preferred. The main issue is that estimates based on differences in sums of squares can take negative values. Raising these to zero is not satisfactory in applications. This estimate also requires that q be large compared to d since MSE has $q^2 - d(q-1) - 1$ degrees of freedom. A simple conservative, but inconsistent, estimate is obtained by using a set of k regressors which are smooth functions of univariate and bivariate margins of X. One simply takes the mean square residual from a regression of Y on said basis and divides it by $n-k$ for an estimate of the sampling variance of $\bar{Y}$. This estimate is asymptotically too large by a factor that depends on what proportion of $\sum_{0<|u|\leq 2}\int \alpha_u^2 dF$ is not captured by the chosen regressors. In many applications a variance estimate that is conservative is preferable to one that is consistent but sometimes nonsensical. The challenge of finding a $n^{1/2}$ consistent and practical estimator remains open.

Acknowledgement

This research was supported by the National Science Foundation and the Semiconductor Research Corporation.

References

Buckley, M. J., Eagleson, G. K. and Silverman, B. W. (1988). The estimation of residual variance in nonparametric regression. *Biometrika* **75**, 189–199.

Currin, C., Mitchell, T., Morris, M. and Ylvisaker, D. (1988). A Bayesian approach to the design and analysis of computer experiments. Technical Report 6498, Oak Ridge National Laboratory.

Davis, P. J. and Rabinowitz, P. (1984). *Methods of Numerical Integration*, 2nd Edition. Academic Press Inc, Orlando FL.

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

Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist.* **19**, 1–141.

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.

Owen, A. B. (1992). A central limit theorem for Latin hypercube sampling. *J. Roy. Statist. Soc. Ser.B.*

Patterson, H. D. (1954). The errors of lattice sampling. *J. Roy. Statist. Soc. Ser.B* **16**, 140–149.

Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Design of Experiments.* John Wiley, New York.

Royden, H. L. (1968). *Real Analysis.* Macmillan, New York.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4**, 409–435.

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29**, 143–151.

Tang, B. (1991). OA-based Latin hypercubes. IIQP Research Report RR-91-10, The Institute for Improvement in Quality and Productivity, University of Waterloo. Submitted to JASA.

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

(Received June 1991; accepted February 1992)