The Annals of Statistics
1994, Vol. 22, No. 2, 930–945

LATTICE SAMPLING REVISITED: MONTE CARLO VARIANCE OF MEANS OVER RANDOMIZED ORTHOGONAL ARRAYS[1]

BY ART OWEN

Stanford University

Randomized orthogonal arrays provide good sets of input points for exploration of computer programs and for Monte Carlo integration. In 1954, Patterson gave a formula for the randomization variance of the sample mean of a function evaluated at the points of an orthogonal array. That formula is incorrect for most of the arrays that are practical for computer experiments. In this paper we correct Patterson's formula.

We also remark on a defect, related to coincidences, in some orthogonal arrays. These are arrays of the form $\mathrm{OA}(2q^2, 2q+1, q, 2)$, where q is a prime power, obtained by constructions due to Bose and Bush and to Addelman and Kempthorne. We conjecture that subarrays of the form $\mathrm{OA}(2q^2, 2q, q, 2)$ may be constructed to avoid this defect.

1. Introduction. Patterson (1954) describes an interesting class of sampling techniques based on randomizations of orthogonal arrays. Most of that article covers sampling from a space described by two or three discrete variables, each taking the same number q of levels, but the final section considers the general case of d variables. The orthogonal array provides some multiple of q points describing a subset of the q^d possible settings and enjoying a balance property described below. Randomization applied to the levels of each variable may be used as a basis of inference.

While such designs are unwieldy for physical experimentation (e.g., 17 varieties of pig in 17 pens trying 17 diets and 17 antibiotics over 17 time periods), they are well suited for computer experiments in which, for example, 5 continuously varying inputs are to be examined at 17 levels each. Patterson's lattice sampling schemes predate by many years the closely related work on Latin hypercube sampling [McKay, Conover and Beckman (1979), Stein (1987), Tang (1993) and Owen (1992a)], which drew its impetus from computer experiments. Owen (1992b) considers the use of randomized orthogonal arrays for integration, visualization and computer experiments on functions defined over $[0,1]^d$.

Patterson (1954) gives a formula for the variance of the sample average of a function evaluated at the points of a randomized orthogonal array. Owen (1992b) uses Patterson's formula to show how the balance property of orthogonal arrays essentially removes certain low-order terms from the sampling variance. Patterson's formula turns out to be incorrect in most of the cases of interest

Received July 1992; revised October 1993.
[1]This research was supported by Semiconductor Research Corporation Grant SRC 92-MC-106 and by NSF Grant DMS-90-11074.
AMS 1991 *subject classifications.* Primary 62K15, 65C05; secondary 65D30, 05B15.
Key words and phrases. Coincidence, computer experiment, Latin hypercube sampling, orthogonal array, randomization.

for computer experiments. This paper gives the correct variance formula. The same low-order terms are removed from the corrected variance formula, so the main point in Owen (1992b) is restored to a firm footing.

Patterson's formula was also presented assuming that the array is free of a certain coincidence defect (described below). The variance formula given here does not require a defect-free array, although absence of defects may greatly simplify it.

We conclude this section by introducing orthogonal arrays, remarking on the coincidence defects mentioned above and defining continuous ANOVA's and randomized orthogonal arrays. Section 2 gives the formula for the randomization variance of the mean of a function evaluated on the points of an orthogonal array. Section 3 considers some special cases of interest to computer experiments. Section 4 compares randomized orthogonal arrays to equidistribution methods and mentions some potential applications.

1.1. *Orthogonal arrays.* Let A be a matrix with elements A_i^j, rows A_i and columns A^j, for $i = 1, \dots, n$ and $j = 1, \dots, d$. Suppose that each $A_i^j \in Q = \{0, 1, \dots, q-1\}$. A is called an orthogonal array of strength $t \le d$ if in each n-row-by-t-column submatrix of A, all q^t possible distinct rows occur the same number λ of times. By counting rows two ways it is clear that $n = \lambda q^t$. We denote such an array by $\mathrm{OA}(n, d, q, t)$. The d columns are called constraints. The array in this definition is the transpose of an orthogonal array under the usual definition, given in Raghavarao (1971). The notation used here is more natural when one thinks of rows as observations and columns as variables. A geometric interpretation of the defining property of an orthogonal array is as follows. Consider plotting an n-by-s, $1 \le s \le t$, submatrix of A, using one point for each of the n rows and one plotting axis for each of the s columns. The plot consists of a q^s grid with λq^{t-s} "overstrikes" at each grid point.

When asymptotic rates such as $O(q^{-4})$ or $O(n^{-2})$ are given in this paper it will be assumed that the limiting operation is $q \to \infty$ with t, λ and d fixed. It is assumed that q is tending to infinity through a subsequence of values for which the corresponding orthogonal arrays exist. In some instances this subsequence is all sufficiently large integers, in others it might be all sufficiently large prime powers.

In the above definition of array strength, if A has strength $t \ge 2$, then it also has strength $t - 1$. The largest integer t such that A has strength t will be referred to below as the maximum strength of A.

1.2. *Coincidence defect.* Consider an n-by-$t + 1$ submatrix of A, where t is the maximum strength of A. When $\lambda < q$, there are fewer than q^{t+1} rows in the submatrix. It is desirable for all these λq^t rows of the submatrix to be distinct, whatever subset of $t + 1$ columns is chosen. Stated another way, the desirable condition is that "no two rows of A agree in any $t + 1$ columns." When there exist two rows of A that do agree in $t + 1$ columns, we say the array A has the "coincidence defect."

Some very useful arrays of the form $\mathrm{OA}(2q^2, d, q, 2)$ are available for $q = p^r$ and $d \leq 2q + 1$. Bose and Bush (1952) describe a construction for $p = 2$ and Addelman and Kempthorne (1961) give a construction that can be used for odd primes p. These arrays (at least the ones the author has constructed) suffer from the coincidence defect. They contain pairs of rows that agree in three columns. The question arises: How large can d be for an orthogonal array $\mathrm{OA}(2q^2, d, q, 2)$ in which no two rows agree in three columns? One can construct such an array by taking a subarray of $\mathrm{OA}(q^3, d+1, q, 3)$, but this does not lead to very large d. [Bounds are given in Bush (1952).] By computer inspection, the author has found that one can construct $\mathrm{OA}(2q^2, 2q, q, 2)$ with no rows matching in any three columns for $q \in \{2, 3, 4, 5, 7, 9, 11, 13\}$ by the method of Addelman and Kempthorne (1961) and for $q \in \{2, 4, 8, 16\}$ by the method of Bose and Bush (1952). In each case it turns out that there is one column that is involved in all the triple coincidences. Removing that column results in an array of strength 2 with no two rows agreeing in any three columns. We conjecture that all arrays $\mathrm{OA}(2q^2, 2q+1, q, 2)$ obtained by the Bose–Bush or Addelman–Kempthorne constructions have subarrays of the form $\mathrm{OA}(2q^2, 2q, q, 2)$ free of the coincidence defect.

1.3. *Continuous ANOVA's.* Following Efron and Stein (1981), we define a continuous ANOVA decomposition for functions on the unit cube $[0,1]^d$. The notation given here is that of Owen (1992b). Let $X \in [0,1]^d$ be a row vector in the cube with components $X^1, \dots, X^d$. Let $D = \{1, 2, \dots, d\}$ represent the set of axes of the unit cube.

Let $u \subseteq D$ be any subset of the axes of the unit cube. Each u is a "source" of variation in f. Let $|u|$ denote the cardinality of u. We will define an effect and a mean square for each source of variation. As in the usual ANOVA, the effects are orthogonal. They might be set out in an ANOVA table with 2^d rows, and columns labeled "source," "effect" and "mean square."

For each $u \subseteq D$, let $dF_u = \prod_{j \in u} dX^j$ be uniform measure on a $|u|$-cube $[0,1]^{|u|}$. We denote this cube $[0,1]^u$. This notation allows us to distinguish the margins for $u \neq v$ with $|u| = |v|$. Integration with respect to dF_u averages over the axes in u, leaving a function defined over the axes in $D - u$. Integration with respect to $dF_{\emptyset}$ leaves a function unchanged. We use $dF = \prod_{j=1}^d dX^j = dF_D$ to denote uniform measure on the original cube.

Let $f\colon [0,1]^d \to R$ be measurable with $\int f(X)^2 \, dF < \infty$. We may then write an ANOVA decomposition of f as a sum of effects α_u via $f(X) = \sum_{u \subseteq D} \alpha_u(X)$ with $\int f(X)^2 \, dF = \sum_{u \subseteq D} \int \alpha_u^2(X) \, dF$ and $\int \alpha_u(X) \alpha_v(X) \, dF = 0$, $u \neq v$. The effects are defined inductively via

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

where $v \subset u$ is understood to include only proper subsets $v \neq u$. Thus $\alpha_{\emptyset} = \int f(X) \, dF$ is the "grand mean," $\alpha_{\{j\}} = \int (f(X) - \alpha_{\emptyset}) \, dF_{D - \{j\}}$ is the main effect for axis j and so on. We abuse notation somewhat and let α_u denote either a function defined on $[0,1]^u$ or a function on $[0,1]^D$ which is constant along axes in $D-u$. Thus $\int \alpha_u^2\,dF_u = \int \alpha_u^2\,dF$.

1.4. *Embedded discrete ANOVA's.* We approximate the continuous ANOVA by a sequence of discrete ANOVA's. Let $E = E(q) = \{q^{-1}(i+1/2) \mid i = 0, 1, \dots, q-1\}$ denote a set of q equispaced points embedded in $[0,1]$. Let $dG = dG(q)$ be a discrete probability measure that puts equal mass q^{-d} on every point in E^d. The dependence of E and dG on q is suppressed for convenience below.

Now define measures dG_u analogously to dF_u above and effects β_u analogously to α_u above, using dG_u in place of dF_u in (1.1). The embedded ANOVA closely matches the continuous one, for smooth f.

In a discrete ANOVA decomposition the customary mean square for u is

$$\mathrm{MS}_u = \left(\frac{q}{q-1}\right)^{|u|} \int \beta_u^2\,dG_u, \tag{1.2}$$

obtained by dividing the sum of squares of β_u by the degrees of freedom $(q-1)^{|u|}$. This is commonly resolved into variance components σ_u^2 via

$$\mathrm{MS}_u = \sum_{v \supseteq u} q^{d-|v|}\sigma_v^2. \tag{1.3}$$

The 2^d mean squares may be written in natural way as a triangular linear system of equations in the variance components. The inverse linear transformation is also triangular and may be expressed as

$$\sigma_u^2 = q^{|u|-d} \sum_{v \supseteq u} (-1)^{|v-u|}\mathrm{MS}_v. \tag{1.4}$$

For finer embeddings, $q \to \infty$, and we have $\int \beta_u^2\,dG \to \int \alpha_u^2\,dF$, $\mathrm{MS}_u \to \int \alpha_u^2\,dF$ and $\sigma_u^2 = O(q^{|u|-d})$. This makes all variance components except σ_D^2 vanishingly small as q increases.

1.5. *Randomization variance.* Owen (1992b) recommends the use of randomized orthogonal arrays in computer experiments, integration and visualization. Suppose that one is experimenting on, integrating or visualizing a function f defined on the unit cube $[0,1]^d$. Then one might evaluate f at n points $X_i \in [0,1]^d$ given by $X_i^j = (\pi_j(A_i^j) + \frac{1}{2})/q$, where A_i^j are the elements of an orthogonal array and $\pi_1, \dots, \pi_d$ are independent uniform random permutations on Q. (By uniform, we mean that all $q!$ possible permutations are equally probable for π_j.) The orthogonal array structure of A is preserved by the permutations.

If one is interested in $\mu = \alpha_\emptyset = \int f(X)\,dF$, a natural estimate is formed by setting $Y_i = f(X_i)$ and then $\hat{\mu} = \overline{Y} = n^{-1}\sum Y_i$. Decomposing this sum, one gets

$$\hat{\mu} = \sum_{u \subseteq D} \frac{1}{n} \sum_{i=1}^{n} \alpha_u(X_i). \tag{1.5}$$

Each effect α_u with $|u| \le t$ is integrated in (1.5) by a product midpoint rule, which typically has accuracy $O(q^{-2})$, and each effect α_u with $|u| > t$ is integrated with Monte Carlo accuracy $O_p(n^{-1/2})$. See Owen (1992b) for details. The sampling variance of $\overline{Y}$ is more easily expressed using the discrete ANOVA

$$\hat{\mu} = \overline{Y} = \sum_{u \subseteq D} \frac{1}{n} \sum_{i=1}^{n} \beta_u(X_i). \tag{1.6}$$

The sums (1.5) and (1.6) match, but the individual terms do not ordinarily match. We will show in Section 2 that $\overline{Y}$ is unbiased for $\mu^{(q)} = \beta_{\emptyset}$, the mean of Y over the q^d-point embedded grid. Typically, $\mu^{(q)}$ differs from $\mu = \alpha_{\emptyset}$ by $O(q^{-2})$.

Patterson (1954) gives a formula for the Monte Carlo variance of (1.6). Suppose that A is an orthogonal array of strength t without the coincidence defect. Patterson's formula (4.3) for $V(\overline{Y})$ is

$$\frac{1}{n} \sum_{|u| > t} \sigma_u^2 \left(1 - \lambda q^{t - |u|}\right). \tag{1.7}$$

The multiplier $1 - \lambda q^{t - |u|}$ may be interpreted as a finite sample correction, reasoning that for effect α_u we have a sample of size λq^t from a population of size $q^{|u|}$. Patterson's other variance formulas are special cases of (1.7).

Formula (1.7) cannot be correct. From (1.4), all components of variance except σ_D^2 become vanishingly small as $q \to \infty$. Therefore, for large q, (1.7) gives $V(\overline{Y}) \simeq \sigma_D^2/n$, but $\sigma_D^2 = \mathrm{MS}_D$ vanishes if f only depends on the first $d-1$ axes. This implies that adjoining an irrelevant dimension to the domain of f would be quite advantageous. Each irrelevant dimension reduces the asymptotic order of (1.7) by a factor of q. (The sequence along which q tends to infinity for such larger d might be the tail of the corresponding sequence for the smaller value of d.)

In Section 2 we find a variance formula that, in Section 3, is seen to match Patterson's for some examples with $t = d - 1$. For $t = d - 1$, only the largest component of variance contributes to the sampling variance of $\overline{Y}$ and the component of variance equals the means square, $\sigma_D^2 = \mathrm{MS}_D$.

The variance formula given in Section 2 does not require the coincidence constraint, although it simplifies for matrices satisfying the constraint. The formula does not match Patterson's formula given as (1.7) above.

2. Moments. Let A be an n-by-d array with elements $A_i^j \in Q = \{0, \dots, q-1\}$. Let t be the strength of A as an orthogonal array, taking $t = 0$ if A is not an orthogonal array.

Let

$$X_i^j = \frac{1}{q}\left(\pi_j(A_i^j) + \frac{1}{2}\right), \tag{2.1}$$

where the π_j are independent uniform random permutations on Q. Let $Y_i = f(X_i)$, where f is a real-valued function defined on $[0,1]^d$. Let $D = \{1, \dots, d\}$ and, for

$u \subseteq D$, let β_u be the effect of u in a discrete ANOVA of f. Let dG be uniform probability measure on the embedded grid E^d.

Recall that, for $j \in u$, $\int \beta_u \, dG_{\{j\}} = 0$; that, for $u \neq v$, $\int \beta_u \beta_v \, dG = 0$; and that $\beta_\emptyset = \int f(X)\, dG$. Adopt the shorthand $\beta_u(i) = \beta_u(X_i)$ and, for $u = \{r_1, \ldots, r_{|u|}\}$, let $u(i)$ denote $(X_i^{r_1}, \ldots, X_i^{r_{|u|}})$, the center of the cell determining the value of β_u for observation X_i.

The estimated mean is

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n}\sum_{u \subseteq D} \beta_u(i).$$

We will find the first two moments of $\overline{Y}$ under this sampling scheme.

First we introduce more notation. For each pair of observations $1 \le i, j \le n$ and each source of variation u, let $w = w_{ij} = w_{ij}(u) = \{r \in u \mid A_i^r = A_j^r\}$. This is the maximal "subsource" of u on which $u(i)$ and $u(j)$ match. Now, for each u and each $r = 0, 1, \ldots, |u|$, let

$$M(u,r) = \sum_{i=1}^{n}\sum_{j=1}^{n} 1_{|w_{ij}(u)| = r}$$

count the number of observation pairs that match on exactly r of the axes in u. We do not need to disaggregate the $M(u,r)$ according to *which* subsource w of size r is the maximal match. Note that $M(u,r)$ and $w_{ij}(u)$ are properties of A that are unaffected by the permutations used to generate X.

THEOREM 1. *Using the definitions above*, $E(\overline{Y}) = \beta_\emptyset$ *and*

$$V(\overline{Y}) = \frac{1}{n^2}\sum_{|u|>t}\sum_{r=0}^{|u|} M(u,r)(1-q)^{r-|u|}\int \beta_u^2 \, dG. \tag{2.2}$$

PROOF. By the ANOVA decomposition

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}\sum_{u}\beta_u(i) = \beta_\emptyset + \frac{1}{n}\sum_{i=1}^{n}\sum_{|u|>t}\beta_u(i)$$

since effects β_u with $0 < |u| \le t$ sum to zero over the design.

Now $E(\overline{Y} - \beta_\emptyset) = n^{-1}\sum_{i=1}^{n}\sum_{|u|>t} E(\beta_u(i)) = 0$ by uniformity of the π's, and this holds even if $t = 0$. Therefore

$$V(\overline{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{|u|>t}\sum_{|v|>t} E\big(\beta_u(i)\beta_v(j)\big). \tag{2.3}$$

To evaluate $E(\beta_u(i)\beta_v(j))$ for all i, j, u, v we consider four cases depending on whether $i = j$ and on whether $u = v$.

For $i = j$ and $u = v$, we get $E(\beta_u(i)\beta_u(i)) = \int \beta_u^2 \, dG$. For $i = j$ and $u \neq v$, orthogonality of effects implies that $E(\beta_u(i)\beta_v(i)) = \int \beta_u \beta_v \, dG = 0$.

For $i \neq j$ and $u \neq v$, there is an axis r in $v - u$ or in $u - v$. Without loss of generality take $r \in v - u$. Then, since $r \in v$,

$$\begin{aligned} E\big(\beta_u(i)\beta_v(j)\big) &= E\big(\beta_u(i)\,(\beta_v(j) \mid \pi_1 \cdots \pi_{r-1}, \pi_{r+1} \cdots \pi_d)\big) \\ &= E\left(\beta_u(i) \int \beta_v \, dG_{\{r\}}\right) = 0. \end{aligned}$$

For the remaining case $i \neq j$ and $u = v$, write

$$E\big(\beta_u(i)\beta_u(j)\big) = E\Big(\beta_u(i) E\big(\beta_u(j) \mid u(i)\big)\Big). \tag{2.4}$$

The distribution of $u(j)$ given $u(i)$ is uniform over $(q-1)^{|u-w|}$ cell centers by the definition of the π's. The cells in question are the ones that match $u(i)$ on all axes in w and do not match $u(i)$ on any axes in $u - w$.

For $v \subseteq u$, let dH_v be counting measure on the $q^{|v|}$ cell centers that match $u(i)$ on the axes $u - v$ and may or may not match $u(i)$ on the axes in v. We suppress the dependence of dH_v on $u(i)$ from our notation.

By an inclusion–exclusion argument,

$$\begin{aligned} &E\big(\beta_u(j) \mid u(i)\big) \\ &\quad = (q-1)^{-|u-w|}\Bigg(\int \beta_u \, dH_{u-w} - \sum_{r \in u-w} \int \beta_u \, dH_{u-w-\{r\}} \\ &\qquad\qquad + \sum_{r,s \in u-w, r \neq s} \int \beta_u \, dH_{u-w-\{r,s\}} \\ &\qquad\qquad\qquad + \cdots + (-1)^{|u-w|} \int \beta_u \, dG_{\varnothing}\Bigg) \\ &\quad = (q-1)^{-|u-w|}(-1)^{|u-w|}\beta_u(i) \end{aligned}$$

since β_u sums to zero over any nonempty set of axes contained in u. Substituting into (2.4) yields

$$E\big(\beta_u(i)\beta_u(j)\big) = (1-q)^{-|u-w_{ij}|} \int \beta_u^2 \, dG.$$

Applying the results of these four cases to (2.3), one finds

$$\begin{aligned} V(\overline{Y}) &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{|u|>t} E\big(\beta_u(i)\beta_v(j)\big) \\ &= \frac{1}{n^2} \sum_{|u|>t} \sum_{r=0}^{|u|} M(u,r)(1-q)^{r-|u|} \int \beta_u^2 \, dG. \qquad \square \end{aligned}$$

The maximum strength of the array A was not used in the proof of Theorem 1. An array of strength 2 is also an array of strength 1. For any array A of elements from Q, we have

$$V(\overline{Y}) = \frac{1}{n^2} \sum_{|u|>0} \sum_{r=0}^{|u|} M(u,r)(1-q)^{r-|u|} \int \beta_u^2 \, dG.$$

This leads to the following combinatorial identity.

COROLLARY. *If A has strength $t \geq |u| > 0$, then $\sum_{r=0}^{|u|} M(u,r)(1-q)^{r-|u|} = 0$.*

In many important cases we have the following convenient approximation

$$V(\overline{Y}) \doteq V_t = \frac{1}{n} \sum_{|u|>t} \int \beta_u^2 \, dG, \tag{2.5}$$

where t is the maximum strength of A, taking $t = 0$ if A is not an orthogonal array. The experimental dictum "control what you can, randomize the rest" as embodied by (2.1) has the consequence in terms of (2.4) that sources $|u| \leq t$, controlled by balancing, do not contribute to V_t and that sources $|u| > t$, from which we have a random sample, contribute $n^{-1} \int \beta_u^2 \, dG$ to V_t. The accuracy of (2.5) will be investigated for some special cases in Section 3.

The result of Theorem 1 extends naturally to cases in which each of the axes is subdivided at a different scale of fineness. Let $Q^j = \{0, \ldots, q_j - 1\}$, for positive integers q_j, $1 \leq j \leq d$. Let π_j be a uniform random permutation on Q^j, with all d permutations independent. Let $dG_u = \prod_{j \in u} dG_{\{j\}}$ where $G_{\{j\}}$ is the uniform distribution on Q^j, and define effects β_u as above.

Now let

$$X_i^j = \frac{1}{q_j}\left(\pi_j(A_i^j) + \frac{1}{2}\right) \tag{2.6}$$

and $\overline{Y} = n^{-1} \sum_{i=1}^n f(X_i)$ as before.

The match counts $M(u,r)$ do not give enough information to allow us to write the variance. For each source u and subsource $v \subseteq u$, let $\widetilde{M}(u,v) = \sum_{i=1}^n \sum_{j=1}^n 1_{w_{ij}(u)=v}$ count the number of pairs of rows on which the maximal matching subsource of u is v.

THEOREM 2. *Let X_i^j be defined by (2.6) and let $\overline{Y}$ be defined as above. Then $E(\overline{Y}) = \beta_{\emptyset}$ and*

$$V(\overline{Y}) = \frac{1}{n^2} \sum_{|u|>0} \sum_{v \subseteq u} \widetilde{M}(u,v) \prod_{j \in u-v} (1-q_j)^{-1} \int \beta_u^2 \, dG.$$

PROOF. Arguments like those used to prove Theorem 1 show that $E(\overline{Y}) = \beta_{\emptyset}$ and

$$\begin{aligned}
V(\overline{Y}) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{|u|>0}\sum_{|v|>0} E\big(\beta_u(i)\beta_v(j)\big)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{|u|>0} E\big(\beta_u(i)\beta_u(j)\big)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{|u|>0} E\Big(\beta_u(i)E\big(\beta_u(j)\mid u(i)\big)\Big)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{|u|>0} E\left(\beta_u(i)^2 \prod_{r\in u - w_{ij}(u)} (1-q_r)^{-1}\right)\\
&= \frac{1}{n^2}\sum_{|u|>0}\sum_{v\subseteq u} \widetilde{M}(u,v)\prod_{j\in u-v}(1-q_j)^{-1}\int \beta_u^2\,dG. \qquad \square
\end{aligned}$$

Theorem 2 is used in Example 5 of the next section, which combines a strength-2 array on q levels with a strength-1 array on q^2 levels. Another potential use for Theorem 2 is for applications in which some input variables vary continuously and others are dichotomous. In these settings one might employ an array with two levels for the dichotomous variables and an even number q of levels for each continuous variable.

3. Examples. In this section we evaluate the variance formula from Theorems 1 and 2 for various orthogonal arrays. We consider arrays of strength 1 and 2, with $\lambda \in \{1,2\}$ since these seem to be the most useful in the sort of problems that motivated this work. Theorems 1 and 2 reduce the variance calculations to finding the appropriate match counts, $M(u,r)$ and $\widetilde{M}(u,v)$. Typically $M(u,0)$ or $\widetilde{M}(u,\emptyset)$ may be found by subtraction since

$$\sum_{r=0}^{|u|} M(u,r) = \sum_{v\subseteq u}\widetilde{M}(u,v) = n^2. \tag{3.1}$$

When (3.1) is solved for one of the match counts, it cannot also be used to check correctness. Simple inspection by a computer is helpful in establishing the correctness of a formula.

EXAMPLE 1 [OA$(q,d,q,1)$]. For this array we have $X_i^j = (\pi_j(i-1)+\frac{1}{2})/q$. The closely related Latin hypercube sample [McKay, Conover and Beckman, (1979)] has $X_i^j = (\pi_j(i-1) + U_i^j)/q$, where the U_i^j are $U[0,1]$ independently of each other and of the π_j. This array is especially convenient for computer experiments, since there are no a priori constraints linking the number of rows and columns. Stein (1987) finds the asymptotic variance of the mean of a function from a Latin hypercube sample. See also Owen (1992a) for a central limit theorem, and Owen (1992b) for higher-strength analogues of Latin hypercube sampling.

For $|u| > 1$, $M(u,|u|) = n = q$, $M(u,0) = n^2 - n = q^2 - q$ and so

$$V(\overline{Y}) = \frac{1}{n^2} \sum_{|u| \geq 2} \left((q^2 - q)(1-q)^{-|u|} + q\right) \int \beta_u^2 \, dG$$

$$= \frac{1}{n} \sum_{|u| \geq 2} \left(1 + (-1)^{|u|}(q-1)^{1-|u|}\right) \int \beta_u^2 \, dG \tag{3.2}$$

$$= V_1 + O\left(\frac{1}{n^2}\right).$$

Compared to V_1, the variance $V(\overline{Y})$ has a larger coefficient for $\int \beta_u^2 \, dG$ with even $|u|$, and a smaller coefficient with odd $|u|$. From (3.2) we have $V(\overline{Y}) \leq V_1 q/(q-1)$, which may be of use in finding conservative estimates of $V(\overline{Y})$.

For $d = 2$, (3.2) can be shown to reduce to $V(\overline{Y}) = (1/n)(1 - q^{-1})\sigma_D^2$, which matches Patterson's (1954) formulas (2.1, 4.1 and 4.3), which are all special cases of (1.7). For $d = 3$, however, we get a contradiction to Patterson's formulas.

EXAMPLE 2 [OA$(\lambda q, d, q, 1)$, $2 \leq \lambda \leq q$]. This array is formed from λ of the Latin hypercube-like arrays considered in Example 1. Sometimes the coincidence defect is avoidable and the array can be chosen with no two rows agreeing in any two columns. Usually, one would prefer to use OA$(\lambda q, d, \lambda q, 1)$ since it allows more distinct values for each input variable. However, if an experiment based on OA $(q, d, q, 1)$ has already been run, and one wishes to augment it with another q runs, then an OA$(2q, d, q, 1)$ design avoiding the coincidence defect is attractive.

For $\lambda = 2$, a "diagonal construction" with $A_i^j = (i-1)$, for $1 \leq i \leq q$, and $A_i^j = (i-1)+(j-1) \bmod q$, for $q < i \leq 2q$, works. A second construction is to take the subarray of $A = \text{OA}(q^2, d+1, q, 2)$ corresponding to the first d columns and those λq rows with $A_i^{d+1} < \lambda$. Neither construction makes the other redundant. For example, with $q = 10$, the diagonal construction works for $\lambda = 2$ and $d \leq 10$, while the OA construction works for $2 \leq \lambda \leq 10$ and $d \leq 3$, using a set of two mutually orthogonal Latin squares of order 10. To use the OA construction for $d = 4$ and $q = 10$ would require three mutually orthogonal Latin squares of order 10. The existence of such squares is still an open problem [Brouwer (1991) and Dénes and Keedwell (1991)].

Let us assume that in $A = \mathrm{OA}(\lambda q, d, q, 1)$, with $2 \le \lambda \le q$, that no two rows of A agree in any two columns. Then, for $|u| \ge 2$, $M(u,|u|) = n$, $M(u,1) = n|u|(\lambda - 1)$ and $M(u,0) = n^2 - n - n|u|(\lambda - 1)$. Therefore

$$V(\overline{Y}) = \frac{1}{n^2} \sum_{|u| \ge 2} \Big(n + n|u|(\lambda - 1)(1-q)^{1-|u|} + \big(n^2 - n - n|u|(\lambda - 1)\big)(1-q)^{-|u|}\Big) \int \beta_u^2 \, dG$$

$$(3.3) \qquad = \frac{1}{n} \sum_{|u| \ge 2} \Big(1 + (1-q)^{-|u|}\big(|u|(\lambda - 1)(1-q) + \lambda q - 1 - |u|(\lambda - 1)\big)\Big) \int \beta_u^2 \, dG$$

$$= \frac{1}{n} \sum_{|u| \ge 2} \Big(1 + (1-q)^{-|u|}\big(-q|u|(\lambda - 1) + \lambda q - 1\big)\Big) \int \beta_u^2 \, dG,$$

where $V_1 - V(\overline{Y}) = O(n^{-2})$. In the special case $\lambda = 2$ we have even better accuracy, $V_1 - V(\overline{Y}) = O(n^{-3})$. For $q \ge 4$ and $\lambda = 2$, we have $(1 - (q-1)^{-2})V_1 \le V(\overline{Y}) \le (1 + (q+1)(q-1)^{-3})V_1$. To illustrate how sharp these bounds are, note that when $q = 8$ (so $n = 128$) we have $0.979V_1 \le V(\overline{Y}) \le 1.027V_1$.

As in Example 1, we can show that, for the special case $d = 2$, (3.3) agrees with Patterson's formula (1.7).

If $\lambda = q$ and no two rows agree in any two columns, the array must have strength 2. Appropriately, the coefficient of $\int \beta_u^2 \, dG$ in (3.3) vanishes for $\lambda = q$ and $|u| = 2$.

EXAMPLE 3 [$\mathrm{OA}(q^2, d, q, 2)$]. For $d > 2$, $\mathrm{OA}(q^2, d, q, 2)$ can be constructed from $d - 2$ mutually orthogonal Latin squares of order q. (For $d = 2$ this OA is trivial.) Bose (1938) shows, using Galois field theory, that for prime powers $q = p^r$ one can construct this array for d as large as $q + 1$. In the experimental design literature these point sets are used as fractional factorials denoted by $q_{\mathrm{III}}^{(q+1)-(q-1)}$. The III designates that the designs are of resolution III, meaning that estimated main effects are confounded with interactions but not with each other.

We need $M(u,r)$ for sources of variation u with $|u| > 2$. For $|u| > 2$, the row pairs i, j counted in $M(u,|u|)$ are those with $i = j$. Therefore $M(u,|u|) = n = q^2$. For each point i and each of $|u|$ axes in u there are $q - 1$ points j matching i on that and only that axis. Therefore $M(u,1) = n(q-1)|u|$. For no $|u| > 2$ and $i \ne j$ is $|w_{ij}(u)| > 1$, so by subtraction $M(u,0) = n^2 - n - n(q-1)|u|$.

Therefore for this design

$$
\begin{aligned}
V(\overline{Y}) &= \frac{1}{n^2} \sum_{|u| \geq 3} \Big(n + (1-q)^{1-|u|} \big(n(q-1)|u| \big) \\
&\qquad + (1-q)^{-|u|} \big(n^2 - n - n(q-1)|u| \big) \Big) \int \beta_u^2 \, dG \\
&= \frac{1}{n} \sum_{|u| \geq 3} \Big(1 + (-1)^{|u|-1} |u| (q-1)^{2-|u|} \\
&\qquad + (-1)^{|u|} (q-1)^{1-|u|} \big(q + 1 - |u| \big) \Big) \int \beta_u^2 \, dG \qquad (3.4) \\
&= \frac{1}{n} \sum_{|u| \geq 3} \Big(1 + (-1)^{|u|-1} (q-1)^{1-|u|} \big(q|u| - q - 1 \big) \Big) \int \beta_u^2 \, dG \\
&= V_2 + O\big(n^{-3/2}\big).
\end{aligned}
$$

Moreover, for $q \geq 3$, $V(\overline{Y}) \leq V_2(q/(q-1))^2$.

In the special case $d = 3$, only $u = D$ appears in (3.4), and the result is

$$
\begin{aligned}
V(\overline{Y}) &= \frac{1}{n} \big(1 + (q-1)^{-2}(2q-1) \big) \int \beta_D^2 \, dG \\
&= \frac{1}{n} \big(1 + (q-1)^{-2}(2q-1) \big) \left(\frac{q-1}{q} \right)^3 \mathrm{MS}_D \qquad (3.5) \\
&= \frac{1}{n} \left(1 - \frac{1}{q} \right) \mathrm{MS}_D,
\end{aligned}
$$

recovering Patterson's formula (1.7), since $\sigma_D^2 = \mathrm{MS}_D$.

EXAMPLE 4 [$\mathrm{OA}(2q^2, d, q, 2)$]. These designs are described in Section 1.2. It is conjectured there that they may be constructed without a coincidence defect for prime powers q and $d \leq 2q$. We assume below that a defect-free array is used.

These designs often allow one to get a strength-2 array with a smaller number of runs than would be possible using the standard construction of Example 3. With $d = 21$, the smallest design of the sort given in Example 3 requires $q = 23$ and hence $n = 529$ runs. The designs in this example can handle $d = 21$ with $q = 11$ and $n = 242$ runs.

For $|u| \geq 3$, $M(u, |u|) = n$, since no distinct rows can agree in three or more places. For each row A_i and each bivariate subsource $v \subset u$, $|v| = 2$, there is one $j \neq i$ with A_i and A_j matching on v. These must be distinct for differing v, since otherwise there would be a triple of axes in u and a row A_j matching A_i on that triple. Therefore $M(u, 2) = n\binom{|u|}{2}$. For each row A_i and each univariate subsource $v \subset u, |v| = 1$, there are $2q - 1$ rows A_j with $j \neq i$ and A_i matching A_j on v. Therefore $M(u, 1) \leq n(2q-1)|u|$. The reason for the inequality is that rows $j \neq i$ with A_j matching A_i on two axes have been counted twice in $(2q - 1)|u|$, but should not be counted at all in $M(u,1)$ since they are counted in $M(u,2)$. Therefore $M(u,1) = n(2q-1)|u| - 2n\binom{|u|}{2}$. By subtraction $M(u,0) = n^2 - n - n(2q-1)|u| + n\binom{|u|}{2}$.

Substituting in (2.3) and simplifying we find

$$
\begin{aligned}
V(\overline{Y}) &= \frac{1}{n}\sum_{|u|\geq 3}\Bigg(1+(-1)^{|u|}(q-1)^{-|u|} \\
&\qquad \times\left(\frac{q^2(|u|-1)(|u|-4)}{2}+q|u|-1\right)\Bigg)\int \beta_u^2\,dG \qquad (3.6)\\
&= V_2 + O(n^{-3/2}).
\end{aligned}
$$

For $q \geq 4$, $V(\overline{Y}) \leq V_2(1+(q^2-3q+1)/(q-1)^3)$. For $d = 3$, (3.6) reduces to Patterson's formula (1.7).

EXAMPLE 5 [$\mathrm{OA}(q^2, d_1, q, 2) \times \mathrm{OA}(q^2, d_2, q^2, 1)$]. In this example the first d_1 columns are taken from a strength-2 design and then d_2 Latin hypercube, or strength-1 columns are added. In applications, the variables among which interactions are strongly suspected may be sampled by an orthogonal array of strength-2, while d_2 other variables that are either less important or less likely to interact are sampled by a strength-1 array. This may allow a smaller experiment than would be possible with a strength-2 array with $d_1 + d_2$ columns and q symbols. Another application is design augmentation: long after some function of the first d_1 variables has been computed, it may be desirable to compute another function that also depends on the next d_2 variables, perhaps using the values computed for the first function. Using independent strength-1 columns in augmentation may be more convenient than trying to find d_2 more columns than preserve the strength-2 structure.

We need some extra notation for this example. Let $D_1 = \{1,\dots,d_1\}$, $D_2 = \{d_1+1,\dots,d_1+d_2\}$ and $D = D_1 \cup D_2 = D_1 + D_2$ be the disjoint (hence the plus sign) union of the two axis sets. For any source $u \subseteq D$, we can find subsources $u_1 \subseteq D_1$ and $u_2 \subseteq D_2$ such that $u = u_1 + u_2$. Where u_1 and u_2 appear below they are understood to be subsets of D_1 and D_2, respectively.

It is natural to expect that $V(\overline{Y})$ would be approximately equal to

$$
\frac{1}{n}\sum_{|u_1|>2}\int \beta_{u_1}^2\,dG + \frac{1}{n}\sum_{|u_2|>1}\int \beta_{u_2}^2\,dG + \frac{1}{n}\sum_{\substack{|u_1|>0\\|u_2|>0}}\int \beta_{u_1+u_2}^2\,dG \qquad (3.7)
$$

because the design is balanced with respect to sources $|u_1| \leq 2$ and $|u_2| \leq 1$.

We can use Theorem 2 to investigate how accurate (3.7) is. A lengthy argument based on partitioning the possible u_1, u_2, v_1, v_2 subset into cases, carefully counting each $\widetilde{M}(u_1+u_2, v_1+v_2)$ and then applying Theorem 2 verifies that (3.7) matches the true variance to $O(n^{-3/2})$.

4. Discussion. Davis and Rabinowitz (1984) say that, for integration over high-dimensional domains, Monte Carlo and equidistribution methods are best.

The former are almost always done through pseudorandom number generators, and the latter methods are often called quasirandom. Quasirandom point sets are chosen so that the uniform measure on the n points used is suitably close to uniform measure on the input domain $[0, 1]^d$. Means over randomized orthogonal arrays may be viewed as a hybrid of pseudorandom and quasirandom methods. The rows of A are constructed similarly to many quasirandom point sets and the permutations π_j are ordinarily pseudorandom.

Pseudorandom integration has errors that are $O_p(n^{-1/2})$. Randomized orthogonal arrays provide a way to reduce the variance of pseudorandom integrals, but do not change the underlying $n^{-1/2}$ rate. Quasirandom integrals can attain error bounds of $O((\log n)^{d-1}/n)$. This is asymptotically smaller than $O(n^{-1/2})$, but even for d as small as 10, n must be impractically large, or the constants in the rates must be extreme, for this bound to be smaller than the standard error of a pseudorandom estimate. For instance $(\log n)^9/n > n^{-1/2}$ for $n = 10^{34}$.

A practical drawback of many quasirandom techniques is that the accuracy of the integral is hard to assess. The constant in the quasirandom error bound depends on the total variation (in the sense of Hardy and Krause) of f and is not easy to estimate from function values. By incorporating pseudorandom sampling, one can use statistical techniques to assess the accuracy of an estimated integral.

Various statistical techniques may be used to assess the accuracy of means over randomized orthogonal arrays. The simplest approach is to use a small number (say, five) of replicates. All replicates share the same orthogonal array A, but each replicate has its own independently generated permutations π_j. This provides independent estimates of $\int f$ with bias $O(q^{-2}) = O(n^{-2/t})$, where t is the maximum strength of A, and variance estimable by the sample variance over replicates. For low-strength arrays the bias is asymptotically negligible. The bias may also be reduced to $O(q^{-3}) = O(n^{-3/t})$ by using non-equispaced sets of q points in $[0, 1]$. Owen (1992b) uses a Gauss rule for this with even q; the idea also works for odd q. Bias may also be reduced to $O(n^{-2})$ by using Tang's (1993) OA-based Latin hypercube samples. These combine features of strength-1 and higher arrays. Bias may be eliminated entirely by using $X_i^j = q^{-1}(\pi(A_i^h) + U_i^j)$, where the U_i^j are independent $U[0, 1]$ random variables.

Variance estimates may be obtained from a single replicate, by estimating the approximate variance V_t given by (2.4). A simple way to use (2.4) is to pick k basis functions $\phi_1(X), \dots, \phi_k(X)$ that all depend on t or fewer of the components of X. One then computes

$$\frac{1}{n-k-1}\sum_{i=1}^{n}\left(Y_i - \widehat{\beta}_{\varnothing} - \sum_{j=1}^{k}\widehat{\beta}_j\phi_j(X_i)\right)^2, \tag{4.1}$$

where the $\widehat{\beta}_j$ are estimated by least squares regression of the Y_i on $\phi_j(X_i)$. Formula (4.1) will usually have a positive bias as an estimate of V_t, since the functions ϕ_j will only span a subspace of the effects α_u for $|u| \le t$. If one has chosen the k basis functions from a larger set of candidates, by finding which ones most reduce the sum of squared errors, then formula (4.1) will have a negative selection bias that may well be larger than the positive bias mentioned above.

For Latin hypercube samples, Stein (1987) discusses both replication and the basis function method. Replication is also considered in Iman and Conover (1980). Owen (1992a) considers estimates of V_1 based on using nonparametric regressions to estimate each main effect $\alpha_{\{j\}}$.

Potential statistical applications of randomized orthogonal array integrals include computer experiments as described in Owen (1992b), Bayesian posterior expectations, simulation studies and the bootstrap.

Some C programs for generating and randomizing orthogonal arrays are available by anonymous ftp. They are in the directory /pub/oa on the host playfair.stanford.edu reachable over the Internet.

Acknowledgments. An Associate Editor and two referees made valuable comments that helped to improve the exposition in this paper.

REFERENCES

ADDELMAN, S. and KEMPTHORNE, O. (1961). Some main-effect plans and orthogonal arrays of strength two. *Ann. Math. Statist.* **32** 1167–1176.

BOSE, R. (1938). On the application of the theory of Galois fields to the problem of construction of hyper-Graeco-Latin squares. *Sankhyā* **3** 323–338.

BOSE, R. and BUSH, K. (1952). Orthogonal arrays of strength two and three. *Ann. Math. Statist.* **23** 508–524.

BUSH, K. (1952). Orthogonal arrays of index unity. *Ann. Math. Statist.* **23** 426–434.

BROUWER, A. E. (1991). Recursive constructions of mutually orthogonal Latin squares. In *Latin Squares: New Developments in the Theory and Applications* (J. Dénes and A. D. Keedwell, eds.) 149–168. North-Holland, Amsterdam.

DAVIS, P. J. and RABINOWITZ, P. (1984). *Methods of Numerical Integration,* 2nd ed. Academic, New York.

DÉNES J. and KEEDWELL, A. D. (1991). Latin squares and geometry. In *Latin Squares: New Developments in the Theory and Applications* (J. Dénes and A. D. Keedwell, eds.) 343–380. North-Holland, Amsterdam.

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

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

IMAN, R. L. and CONOVER, W. J. (1980). Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Comm. Statist. A—Theory Methods* **9** 1749–1874.

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. (1992a). A central limit theorem for Latin hypercube sampling. *J. Roy. Statist. Soc. Ser. B* **54** 541–551.

OWEN, A. B. (1992b). Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2** 439–452.

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.* Wiley, New York.

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.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CALIFORNIA 94305