Statistica Sinica **24** (2014), 985-1000
doi:http://dx.doi.org/10.5705/ss.2012-335

SEQUENTIAL PROBABILITY-BASED LATIN HYPERCUBE DESIGNS WITHOUT REPLACEMENT

Ying Hung

Rutgers, the State University of New Jersey

Abstract: Despite the prevalence of space-filling designs in many applications, most of the existing approaches are unsuitable for irregular regions. A new sequential design procedure is introduced for a specific type of irregular region. This procedure constructs designs with desirable space-filling properties and without replacement of design points. Moreover, it provides flexibility on the size of the experiments. Efficient and easy-to-compute unbiased estimators are introduced. The various estimators are compared using a simulation and a data center thermal management example.

Key words and phrases: Adaptive cluster sampling, computer experiment, Latin hypercube design, Rao-Blackwell method, sequential design, space-filling design.

1. Introduction

Space-filling designs, such as Latin hypercube designs (LHDs) (McKay, Beckman, and Conover (1979)), are widely used to spread out the design points in an experimental region (Santner, Williams, and Notz (2003); Fang, Li, and Sudjianto (2006)), but are limited by the rectangular assumption on the experimental region. When the design region is irregular, direct application of the conventional designs can lose such desirable space-filling properties as one-dimensional balance. A specific type of irregular region commonly occurring in practice is a slid-rectangular region (Hung, Amemiya, and Wu (2010)), a two-dimensional region where the feasible range of one factor depends on the level of the second factor. For example, Figure 1 (Hung (2011)) concerns design for sensor placement over a slid-rectangular region (gray area) in a data center thermal management problem. Other examples occur in ecological study, environmental sampling (Stevens and Olsen (2004)), and computer experiments (Hung (2011)).

Challenging issues arise in designing experiments with slid-rectangular regions: how to spread out the initial design points uniformly; how to select the follow-up design of experiments to efficiently improve estimation. As to the first issue, probability-based LHDs (Hung, Amemiya, and Wu (2010)) can maintain the one-dimensional balance property with design points evenly spread out over the region. Figure 2 shows a data center example with a slid-rectangular region

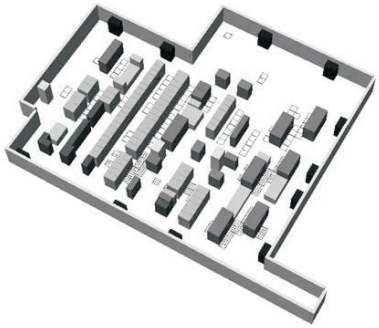

Figure 1. A data center.

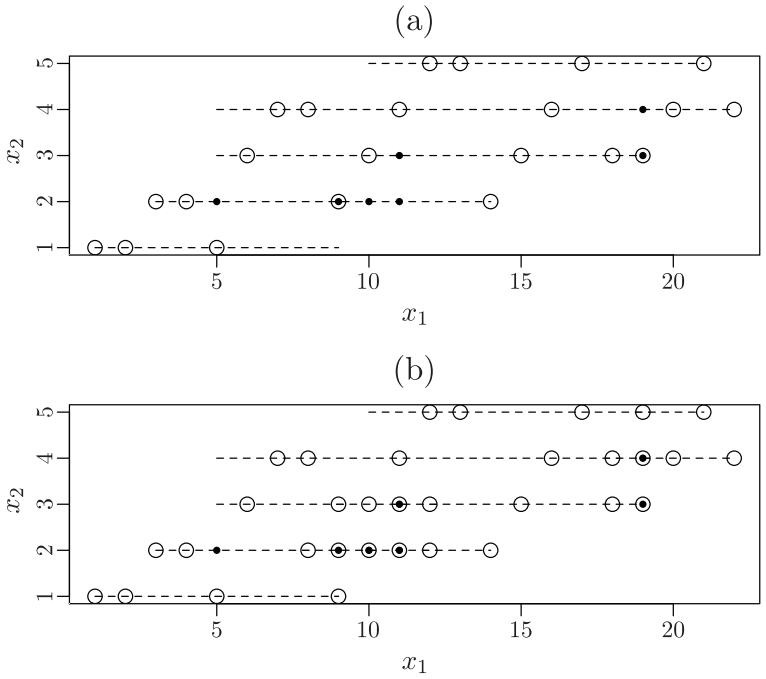

Figure 2. (a)A 22-run PLHD. (b) The resulting APLHD.

in which sensors can be placed. The circles shown in Figure 2(a) are design points of a 22-run probability-based LHD.

Adaptive probability-based Latin hypercube designs (APLHDs) have been proposed (Hung (2011)) to select follow-up design points. Here a criterion is defined for sequentially adding design points based on the response. When the response at an initial design point satisfies the criterion, additional points in the neighborhood of that point are added to the sample; if any of the additionally

added points satisfy the condition, then more points may be added to the sample. This procedure continues until no more points meet the condition. In Figure 2(a), the bullet points satisfy the prespecified condition, but are not necessarily included in the initial sample. The final sample is denoted by the circles in Figure 2(b). The set of points satisfying the condition in the neighborhood of one another is called a network. For example, in Figure 2(b), the four bullet points in the middle form a network of size four.

APLHD is intuitive and works well in general, but design points may be sampled more than once. Then, the final number of distinct networks selected is random, and the size of the experiment may exceed a predetermined cost limit. Further, the procedure allows the selection of networks with replacement, so design points cannot be spread out as uniformly as they should be and for some experiments, observations with the same design setting provide little or no information. This issue is particularly important for computer experiments (Santner, Williams, and Notz (2003)) in which replicates should be avoided because the outputs are deterministic (i.e., the same input produces exactly the same output.)

We discuss a new sequential sampling procedure in this paper. Design points are selected without replacement and the one-dimensional balance property is maintained. This approach can efficiently reduce estimation variance, and the size of the experiment can be fixed in advance. This procedure is different from the existing sampling approaches (Thompson (1990, 1991); Thompson and Seber (1996); Salehi and Seber (1997)) that are developed for rectangular regions with independent and equal individual inclusion probabilities. For slid-rectangular regions, the inclusion probabilities are no longer equal, they are updated sequentially according to the shape of the region and the design points selected earlier. There are challenges in the development of the new approach, especially in the construction of unbiased estimators.

The remainder of the article is organized as follows. A sequential design procedure is introduced in Section 2. In Section 3, unbiased estimators are proposed and their variances are studied. Further improvements are achieved by reducing the variance and simplifying the calculation. These estimators are illustrated and compared via a toy example in Section 4. In Section 5, the performance of the proposed methods is demonstrated in a data center thermal management example. Conclusions are given in Section 6.

2. Design Procedure

Assume that factors x_1 and x_2 form a slid-rectangular region. Factor x_2 has, for now, t levels. For the jth level of x_2, the feasible interval for x_1 is$E_j = (A_j, B_j)$, $j = 1, \ldots, t$, with $A = \min\{A_j\}$ and $B = \max\{B_j\}$. The interval $[A, B]$ is divided into n equally spaced subintervals and n levels of x_1 are assigned to

the middle of these subintervals, $x_{11} = 1, \ldots, x_{1n} = n$. For each level of x_1, the feasible range for x_2 is C_i, $i = 1, \ldots, n$.

A neighborhood is defined for each design point, consisting of, in addition to itself, the intersection of the four spatially adjacent points and the slid-rectangular region. For design point that satisfies a certain condition, its neighborhood points are added to the sample; if any of these additional points satisfy the condition, their neighborhood is added to the sample as well. Continually adding points leads to a cluster of units whose edge points do not satisfy the condition. The cluster minus its edge points is a network. Any point that does not meet the condition forms a network of size 1.

Design points, together with their networks, are sampled sequentially with the inclusion probabilities updated at every iteration. After selection of an initial point and its network, they are removed from the population and inclusion probabilities are adjusted accordingly. Suppose we plan to select N initial design points sequentially, N predetermined with $N \leq n$, represented by $u_i = (x_{1i}, x_{2i})'$. For each level of x_1, the feasible ranges for x_2 are different and change at each iteration, say C_{il} for $x_1 = i$ in the lth iteration, $C_{i1} = C_i$.

Step 0: Let the initial sample space for x_1 be $\Omega_1 = (1, \ldots, n)$. For $l = 1, \ldots, N$, perform steps 1 to 3 iteratively.

Step 1: Randomly select a level of x_1, denoted by t_l, from the sample space Ω_l, and sample the design point u_l by selecting a level of x_2 with probability $P(x_{2l} = j) = c_{t_l l}^{-1}$ if $j \in C_{t_l l}$ and 0 otherwise, where $c_{t_l l} = \sum_{j=1}^{t} I[j \in C_{t_l l}]$. Thus, $x_{1l} = t_l$.

Step 2: Sample the initial design point u_l and its network W. If this network contains all the units from some other levels of x_1, denote the collection of these x_1 levels by w.

Step 3: Update the sample space of x_1 by removing the set $v_l = \{t_l \cup w\}$, $\Omega_{l+1} = \Omega_l \backslash v_l$. Remove the selected network from the population and update the feasible regions C_{il+1}, $i \in \Omega_{l+1}$, by removing the units belonging to network W.

Figure 3 depicts an example of a slid-rectangular region in which there are three levels for factor x_2 and the feasible regions of x_1 depend on the levels of x_2. Factor x_1 has three levels and the sample space is $\Omega_1 = \{1, 2, 3\}$. For each level of x_1, the feasible ranges of x_2 are $C_{11} = \{1\}$, $C_{21} = \{1, 2\}$, and $C_{31} = \{2, 3\}$ respectively. The numbers shown in the cells are the responses. We consider adding points adaptively when the observed response is larger than 8. Based on this condition, it is clear that points $(2, 2)$ and $(3, 2)$ form a network of size two. Let $N = 3$. Assume level 2 is selected randomly from Ω_1, $t_1 = 2$ in the first iteration. The corresponding x_2 coordinate is chosen with probability

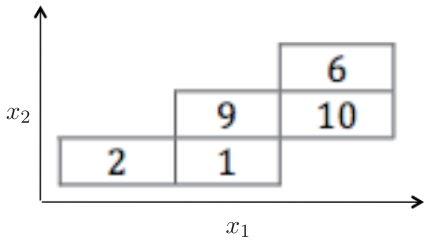

Figure 3. Illustrative example

$P(x_{21} = j) = 1/2$ for $j = 1, 2$. Let $x_{21} = 2$ be selected, so the first design point is $u_1 = (2, 2)$. The selection of u_1 leads to the selection of $(3, 2)$, because they form a network. Now remove this network from the population and update $\Omega_2 = \{1, 3\}$. The new feasible ranges are $C_{12} = \{1\}$ and $C_{32} = \{3\}$. In the second iteration, suppose $t_2 = 3$ is selected from Ω_2. There is only one point available for this level, therefore $P(x_{22} = 3) = 1$ and $u_2 = (3, 3)$. Update $\Omega_3 = \{1\}$, $C_{13} = \{1\}$ and $P(x_{23} = 1) = 1$. Hence, in the third iteration we have $u_3 = (1, 1)$.

The sequential design procedure has another advantage in that the final number of distinct networks selected, N, can be determined based on the objective of the experiment. If the focus is to explore the experimental region uniformly, it is desirable to have $N = n$ because the resulting initial design is a probability-based LHD and the one-dimensional balance property holds. On the other hand, if the experiment is expensive or time-consuming, N can be a smaller number to reduce costs. Compared with APLHD, this approach provides a better control over the size of the experiment. It is possible that N is less than n; this happens when there is a large network containing at least two adjacent levels of x_1 and at least one of them has all the units included in the network. In this situation, there is a possibility that $w \neq \emptyset$ at least once, and thus $N < n$. This rarely occurs in practice because we mainly focus on experiments in which responses of interest are clustered in a relatively small area.

3. Estimator

In this section, we focus on unbiased estimators for the population mean. The individual inclusion probabilities are no longer equal, so the conventional unbiased estimators cannot be directly used here. Since the new procedure features unequal probability sampling without replacement, a Murthy-type unbiased estimator (Murthy (1957)) is constructed. This estimator can be improved based on the Rao-Blackwell Theorem to reduce variance. However, being different from the conventional adaptive sampling (Salehi and Seber (1997)), the Rao-Blackwell estimator is difficult to compute due to the unequal inclusion probabilities of the

edge points. Thus, an easy-to-compute unbiased estimator is introduced that provides a smaller variance with a simpler computation. Derivations of the unbiased estimators of mean and variance are given in the Appendix.

3.1. Murthy's estimator

Let K denote the number of networks in the population and y_k^*, $k = 1, \ldots, K$, be the set of ordered sample values with $y_k^* = \sum_{j \in \Psi_k} y_j$, where Ψ_k is the set of units in the kth network. Raj (1956) introduced a general approach to unbiased estimators for unequal probability sampling without replacement based on the fact that the scheme of selection of a unit at a particular draw depends on the units already drawn in the sample, but not on the order in which they were drawn. Following Raj, we consider

$$\begin{aligned} \omega_1 &= \sum_{k=1}^{K} \frac{y_k^* I_1(n_k > 0)}{P_1(n_k > 0)}, \\ \omega_2 &= y_1^* + \sum_{k=2}^{K} \frac{y_k^* I_2(n_k > 0)}{P_2(n_k > 0)}, \\ \omega_i &= y_1^* + \cdots + y_{i-1}^* + \sum_{k=i}^{K} \frac{y_k^* I_i(n_k > 0)}{P_i(n_k > 0)}, \end{aligned} \tag{3.1}$$

where $i = 3, \ldots, N$, $I_i(n_k > 0)$ is an indicator variable taking value 1 when the kth network of the population is the ith network selected in the sample, and 0 otherwise. $P_i(n_k > 0)$ is the probability that the kth network is selected as the ith sample, based on the updated feasible region in the iteration corresponding to the ith sample. Let μ be the population mean. It is clear that $\mathrm{E}(\omega_i)/\sum_j c_j = \mu$, where $c_j = \sum_{l=1}^{t} I[l \in C_j]$ is the number of feasible points for the jth level of x_1. Therefore, according to Raj (1956), an unbiased estimator of μ is

$$\hat{\mu}_R = (\sum_{j=1}^{n} c_j)^{-1} \sum_{i=1}^{N} u_i \omega_i, \tag{3.2}$$

where u_i can be any constant satisfying $\sum_{i=1}^{N} u_i = 1$.

Murthy (1957) proposed a modification of Raj's estimator, by constructing an unordered version of Raj's ordered unbiased estimator. Let s_0^* be an ordered sample of the N networks selected according to the procedure in Section 2, T be the set of all samples obtained by permuting the coordinates of the elements of s_0^*, and s_0 be the unordered sample set of the N network. Murthy introduced the unbiased estimator

$$\hat{\mu}_M = (\sum_{j} c_j)^{-1} \frac{\sum_{s_0^* \in T} P(s_0^*) \hat{\mu}_R(s_0^*)}{P(s_0)} \tag{3.3}$$

for the population mean, with a variance less than that of $\hat{\mu}_R$. By choosing $u_1 = 1$ and $u_i = 0$ for $i > 1$, Murthy's estimator can be written as

$$\hat{\mu}_M = (\sum_j c_j)^{-1} \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^*, \tag{3.4}$$

where $P(s_0|i)$ is the conditional probability of choosing sample s given network i has been chosen as the first network. The variance of (3.4) is (Murthy (1957); Cochran (1977))

$$\begin{aligned} \operatorname{var}(\hat{\mu}_M) = \Big(\sum_j c_j\Big)^{-2} \Bigg[\sum_{i=1}^{K} \sum_{j<i}^{K} \Bigg(1 - \sum_{a,b \in s_0} \frac{P(s_0|a)P(s_0|b)}{P(s_0)}\Bigg) \\ \left(\frac{y_i^*}{P_1(n_i > 0)} - \frac{y_j^*}{P_1(n_j > 0)}\right)^2 P_1(n_i > 0) P_1(n_j > 0)\Bigg], \end{aligned} \tag{3.5}$$

and its unbiased estimator is

$$\begin{aligned} \hat{\operatorname{var}}(\hat{\mu}_M) = \Big(\sum_j c_j\Big)^{-2} \Bigg[\sum_{i=1}^{N} \sum_{j<i}^{N} \left(\frac{P(s_0|i,j)}{P(s_0)} - \frac{P(s_0|i)P(s_0|j)}{P(s_0)^2}\right) \\ \left(\frac{y_i^*}{P_1(n_i > 0)} - \frac{y_j^*}{P_1(n_j > 0)}\right)^2 P_1(n_i > 0) P_1(n_j > 0)\Bigg], \end{aligned} \tag{3.6}$$

where $P(s_0|i,j)$ denotes the probability of the sample s_0 given that the points i and j are selected in either order in the first two draws.

3.2. Rao-Blackwell estimator

The estimator $\hat{\mu}_M$ can be improved by incorporating more of the information obtained in the final sample. In particular, the observations from edge points are used in the estimator only if they are included in the first N samples. Thus, according to the Rao-Blackwell method, an improved unbiased estimator can be obtained by calculating the conditional expectation of $\hat{\mu}_M$ given the minimum sufficient statistics.

According to the new procedure, the minimum sufficient statistic, d, is the final unordered sample of the ν distinct networks, with labels denoted by $d = \{(i_1, y_{i_1}), \ldots, (i_\nu, y_{i_\nu})\}$. Define D as the sample space for d, $g(s_0')$ as the function that maps an initial design s_0' into a value of d, and S as the sample space containing all possible initial samples. An improved unbiased estimator, $\hat{\mu}_{\mathrm{RB}} = E(\hat{\mu}_M \mid D = d)$, can be written as

$$\begin{aligned} \hat{\mu}_{\mathrm{RB}} = \Big(\sum_j c_j\Big)^{-1} \Bigg\{ \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) \\ + L^{-1} \sum_{s_0' \in S} \Bigg[I(g(s_0') = d) \sum_{i \in s_0',\, e_i = 1} \frac{P(s_0'|i)}{P(s_0')} y_i \Bigg] \Bigg\}, \end{aligned} \tag{3.7}$$

where $L = \sum_{s_0' \in S} I(g(s_0') = d)$, $e_i^* = \sum_{j \in \Psi_i} e_j$, and $e_j = 1$ if unit j is an edge point and $e_j = 0$ otherwise. The variance of this improved unbiased estimator is

$$\mathrm{var}(\hat{\mu}_{\mathrm{RB}}) = \mathrm{var}(\hat{\mu}_M) - \sum_{d \in D} \frac{P(d)}{L} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = d)[\hat{\mu}_M - \hat{\mu}_{\mathrm{RB}}]^2, \tag{3.8}$$

where $P(d)$ is the probability that $D = d$. An unbiased estimator of the variance is $\tilde{\mathrm{var}}(\hat{\mu}_{\mathrm{RB}}) = \hat{\mathrm{var}}(\hat{\mu}_M) - L^{-1} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = d)[\hat{\mu}_M - \hat{\mu}_{\mathrm{RB}}]^2$. A more efficient estimator can be further obtained by conditioning on the minimum sufficient statistic

$$\begin{aligned} \hat{\mathrm{var}}(\hat{\mu}_{RB}) &= E(\tilde{\mathrm{var}}(\hat{\mu}_{RB}) | D = d) \\ &= \frac{1}{L} \sum_{\tilde{s}_0 = S} I(g(\tilde{s}_0) = d) \hat{\mathrm{var}}(\hat{\mu}_M) - \frac{1}{L} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = d)[\hat{\mu}_M - \hat{\mu}_{RB}]^2. \end{aligned} \tag{3.9}$$

3.3. Easy-to-compute Rao-Blackwell estimator

The RB estimator is computationally difficult for large designs because all compatible designs must be evaluated to obtain the estimation. This is not surprising, given that the same difficulty is experienced in some adaptive designs (Dryver and Thompson (2005); Hung (2011)). We construct a new unbiased estimator obtained by conditioning on a carefully chosen sufficient statistic, not the minimum.

Decompose the final sample s into s_c and $s_{\bar{c}}$, s_c the set of units in the sample for which the condition to sample adaptively is satisfied. Take V as a collection of x_1 coordinates at which edge points occur in the initial sample. Using the notation of Section 3.2, a sufficient statistic is $d^* = \{(i, y_i), V, (j, y_j) : i \in s_c, j \in s_{\bar{c}}\}$. The sample space for d^* is defined by D^* and an improved unbiased estimator can be derived by conditioning on d^* as $\hat{\mu}^* = E(\hat{\mu}_M \mid D^* = d^*)$. Let $e_{s_l} = \sum_{i \in s} e_i t_l(i)$ and $t_l(i)$ be an indicator variable taking 1 if the unit i belongs to level l in factor x_1 and 0 otherwise. This estimator can be written as

$$\mu^* = \Big(\sum_j c_j \Big)^{-1} \Big\{ \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) + \sum_{l \in V} \frac{P(s_0 | x_{11} = l)}{P(s_0)} \sum_{i \in s} \frac{e_i y_i t_l(i)}{e_{s_l}} \Big\}. \tag{3.10}$$

The variance is

$$\begin{aligned} \mathrm{var}(\hat{\mu}^*) = \mathrm{var}(\hat{\mu}_M) - \sum_{d^* \in D^*} \frac{P(d^*)}{L^*} \sum_{s_0' \in S} \Big\{ & I(g(s_0') = d^*) \\ & \Big[\sum_{l \in V} \Big(\sum_{i \in s_0',\, e_i = 1} \frac{P(s_0'|i)}{P(s_0')} y_i t_l(i) - \frac{P(s_0' | x_{11} = l)}{P(s_0')} \sum_{i \in s} \frac{e_i y_i t_l(i)}{e_{s_l}} \Big) \Big]^2 \Big\}. \end{aligned} \tag{3.11}$$

An unbiased estimator of the variance is

$$\tilde{\text{var}}(\hat{\mu}^*) = \hat{\text{var}}(\hat{\mu}_M) - \frac{1}{L^*}\sum_{s_0'\in S}\Bigg\{I(g(s_0')=d^*) \Bigg[\sum_{l\in V}\Bigg(\sum_{i\in s_0',\, e_i=1}\frac{P(s_0'|i)}{P(s_0')}y_i t_l(i) - \frac{P(s_0'|x_{11}=l)}{P(s_0')}\sum_{i\in s}\frac{e_i y_i t_l(i)}{e_{s_l}}\Bigg)\Bigg]^2\Bigg\},$$

and a more efficient estimator for the variance is

$$\begin{aligned}\hat{\text{var}}(\hat{\mu}^*) &= E[\tilde{\text{var}}(\hat{\mu}^*) \mid D^* = d^*] \\ &= \frac{1}{L^*}\sum_{s_0'\in S} I(g(s_0')=d^*)\hat{\text{var}}(\hat{\mu}_M) - \frac{1}{L^*}\sum_{s_0'\in S}\Bigg\{I(g(s_0')=d^*) \\ &\quad \Bigg[\sum_{l\in V}\Bigg(\sum_{i\in s_0',\, e_i=1}\frac{P(s_0'|i)}{P(s_0')}y_i t_l(i) - \frac{P(s_0'|x_{11}=l)}{P(s_0')}\sum_{i\in s}\frac{e_i y_i t_l(i)}{e_{s_l}}\Bigg)\Bigg]^2\Bigg\}.\end{aligned}$$

4. Simulation Study

The small population example in Figure 3 is revisited in order to show the computations and properties of the unbiased estimators in relation to the estimators in APLHDs. APLHDs partition the population into distinct networks and selection of any point within the network leads to inclusion in the sample of every other design point in the network. More than one of the initial design points can fall in the same network and be selected more than once.

The population in Figure 3 consists of five points. Based on APLHD, there are four possible designs listed in Table 1. For each row in the table, the first three points form the initial design according to probability-based LHD and the rest of them are added adaptively. For each design, we evaluated two unbiased estimators introduced by Hung (2011), $\hat{\delta}$ and its Rao-Blackwell estimator $\hat{\delta}_{\text{RB}}$. The estimator $\hat{\delta}_{\text{RB}}$ is derived conditional on the minimum sufficient statistics, and therefore it is the most efficient. Hung (2011) proposed another Rao-Blackwell estimator, $\hat{\delta}^*$. It may not be as efficient as $\hat{\delta}_{\text{RB}}$ but it is easier to compute. This estimator has the same performance as $\hat{\delta}$ in this example so it is not listed in the table. Variances of the unbiased estimators are also reported. The last row summarizes the average performance of these estimators over all possible designs. Note that the third design in the table has two initial design points falling in the same network, and therefore the network $\{(2,2),(3,2)\}$ is selected twice.

With the new procedure, there are 18 possible ordered samples listed in Table 2, along with their probabilities reported in column p. Three unbiased estimators, $\hat{\mu}_M$, $\hat{\mu}_{\text{RB}}$, and $\hat{\mu}^*$, and their variances were evaluated for each design. The value of V for each design is listed for the calculation of $\hat{\mu}^*$. Similar to

Table 1. All possible APLHDs and the unbiased estimators

sampler	$\hat{\delta}$	$\hat{\delta}_{\mathrm{RB}}$	$\hat{\mathrm{var}}(\hat{\delta})$	$\hat{\mathrm{var}}(\hat{\delta}_{\mathrm{RB}})$
$(1,1),(2,1),(3,2);(2,2),(3,3)$	5.87	6.40	6.04	3.16
$(1,1),(2,1),(3,3);$	3.20	3.20	2.96	2.96
$(1,1),(2,2),(3,2);(2,1),(3,3)$	5.47	6.40	6.42	3.16
$(1,1),(2,2),(3,3);(2,1),(3,2)$	7.87	6.40	0.34	3.16
Mean	5.60	5.60	3.94	3.11

Table 2. All possible sequential designs without replacement and the unbiased estimators

ordered sampler	p	V	$\hat{\mu}_M$	$\hat{\mu}^*$	$\hat{\mu}_{\mathrm{RB}}$	$\hat{\mathrm{var}}(\hat{\mu})$	$\hat{\mathrm{var}}(\hat{\mu}^*)$	$\hat{\mathrm{var}}(\hat{\mu}_{\mathrm{RB}})$
$(1,1),(2,1),(3,2);(2,2),(3,3)$	1/24	2	5.73	5.73	6.40	5.14	5.14	2.56
$(1,1),(3,2),(2,1);(2,2),(3,3)$	1/12	2	5.73	5.73	6.40	5.14	5.14	2.56
$(2,1),(1,1),(3,2);(2,2),(3,3)$	1/24	2	5.73	5.73	6.40	5.14	5.14	2.56
$(2,1),(3,2),(1,1);(2,2),(3,3)$	1/24	2	5.73	5.73	6.40	5.14	5.14	2.56
$(3,2),(1,1),(2,1);(2,2),(3,3)$	1/12	2	5.73	5.73	6.40	5.14	5.14	2.56
$(3,2),(2,1),(1,1);(2,2),(3,3)$	1/12	2	5.73	5.73	6.40	5.14	5.14	2.56
$(1,1),(2,1),(3,3);$	1/24	2	3.20	3.20	3.20	0	0	0
$(1,1),(3,3),(2,1);$	1/24	2	3.20	3.20	3.20	0	0	0
$(2,1),(1,1),(3,3);$	1/24	2	3.20	3.20	3.20	0	0	0
$(2,1),(3,3),(1,1);$	1/24	2	3.20	3.20	3.20	0	0	0
$(3,3),(1,1),(2,1);$	1/24	2	3.20	3.20	3.20	0	0	0
$(3,3),(2,1),(1,1);$	1/24	2	3.20	3.20	3.20	0	0	0
$(1,1),(2,2),(3,3);(2,1),(3,2)$	1/12	3	7.07	7.07	6.40	0.87	0.87	2.56
$(1,1),(3,3),(2,2);(2,1),(3,2)$	1/24	3	7.07	7.07	6.40	0.87	0.87	2.56
$(2,2),(1,1),(3,3);(2,1),(3,2)$	1/12	3	7.07	7.07	6.40	0.87	0.87	2.56
$(2,2),(3,3),(1,1);(2,1),(3,2)$	1/12	3	7.07	7.07	6.40	0.87	0.87	2.56
$(3,3),(1,1),(2,2);(2,1),(3,2)$	1/24	3	7.07	7.07	6.40	0.87	0.87	2.56
$(3,3),(2,2),(1,1);(2,1),(3,2)$	1/24	3	7.07	7.07	6.40	0.87	0.87	2.56
Mean			5.60	5.60	5.60	2.00	2.00	1.71

the previous table, the last row summarizes the average performances of these estimators over all designs.

The first design in Table 2 is spelled out as an example. The ordered observations are denoted by y_{11}, y_{21}, and y_{32}, respectively. Additional points y_{22} and y_{33} are added sequentially to the final sample. For Murthy's unbiased estimator, we first evaluate probabilities $P(s_0|\{11\}) = 3/8$, $P(s_0|\{21\}) = 1/2$, $P(s_0|\{22,32\}) = 1/2$, and $P(s_0) = 9/24$, where $P(s_0|\{11\})$ is the conditional probability of choosing the unordered sample given network $(1,1)$ has been chosen as the first network. The unbiased estimator is

$$\hat{\mu}_M = \frac{1}{5}\left[y_{11}\frac{3/8}{9/24} + y_{21}\frac{1/2}{9/24} + (y_{22}+y_{32})\frac{1/2}{9/24}\right] = 5.73.$$

To calculate $\hat{\mu}_{RB}$ for the first design, we need the information from twelve designs, including the first and last six designs, because they share the same final unordered sample. We need probabilities $P(s_0|\{33\}) = 1/2$ and $P(s_0) = 9/24$ for the last six designs, because $(3, 3)$ is the edge point collected in the initial sample. Thus, we have

$$\hat{\mu}_{RB} = \frac{1}{5}\left\{ y_{11}\frac{3/8}{9/24} + (y_{22} + y_{32})\frac{1/2}{9/24} + \frac{1}{12}\left[y_{21}\frac{1/2}{9/24} \times 6 + y_{33}\frac{1/2}{9/24} \times 6 \right] \right\} = 6.40.$$

In this example, $\hat{\mu}^*$ is the same as $\hat{\mu}_M$ because $V = 2$ and there is only one edge point at the second level of x_1.

Comparing Tables 1 and 2, it is clear that the design constructed without replacement is more efficient than APLHD in terms of variance reduction. The average variance of $\hat{\mu}_M$ is 36% (=$(3.11 - 2)/3.11$) smaller than $\hat{\delta}_{\mathrm{RB}}$, and the variance is further reduced by 15% (=$(2 - 1.71)/2$) using $\hat{\mu}_{\mathrm{RB}}$.

5. Application in Data Center Thermal Management

We consider the data center thermal management example in Hung (2011). The objective is to model the thermal distribution in the data center and the final goal is to design a data center with an efficient heat-removal mechanism (Schmidt, Cruz, and Iyengar (2005); Hamann (2008)). To monitor and study the thermal distribution, sensors are attached to the surfaces of facilities to measure temperature. The facilities are located in six rows with different lengths, which result in a slid-rectangular experimental region.

Sequential designs are desirable in reducing estimation variances because hot spots are usually clustered in a few locations of the data center, with large portions of the region remaining cool. In this example, we consider adding design points if the observed temperature is at least $34°C$.

To compare performance with APLHDs, we conducted simulations based on the snapshot of the data center temperatures reported in Hung (2011). With the snapshot data, detailed temperature observations are available for all 166 sites, allowing comparisons based on different choices of design. The simulations consisted of 1,000 iterations. The x_1 coordinate was divided into 24 equally spaced intervals and we assumed $N = 24$. For the Rao-Blackwell estimators, we considered $\hat{\delta}^*$ and $\hat{\mu}^*$ because they are easy to compute and recommended in practice. The estimated mean temperatures and estimated variances for all the unbiased estimators are summarized in Table 3. The proposed sequential design without replacement outperforms $\hat{\delta}$ and $\hat{\delta}^*$ based on APLHDs. In particular, $\hat{\mu}^*$ has an approximate 12.5% variance reduction compared with $\hat{\delta}^*$ using APLHDs. The average sample size for the sequential design without replacement was 27.53, which is slightly smaller than that for APLHDs, 28.81.

Table 3. Comparison of estimators in the data center study

	$\hat{\delta}$	$\hat{\delta}^*$	$\hat{\mu}_M$	$\hat{\mu}^*$	$\hat{\mu}_{\text{SRS}}$
mean	24.208	24.211	24.198	24.215	24.219
variance	0.538	0.497	0.451	0.435	0.782

In this study, we also compared the performance of the proposed sequential designs with simple random sampling, a non-sequential design, given the same sample size. In each simulation, simple random samples were generated with the same sample size as the corresponding sequential design without replacement, and the unbiased estimator is denoted by $\hat{\mu}_{\text{SRS}}$. The result in Table 3 shows that the proposed design obtains 44.4% variance reduction compared with simple random sampling.

6. Summary and Concluding Remarks

We focus on experimental design for a specific type of irregular region called slid-rectangular region, where the desirable range of one factor depends on the level of another factor. For slid-rectangular regions, we introduce a sequential design procedure that achieves space-filling properties for design points chosen without replacement. This procedure provides flexibility on the size of the experiment so that it can be determined according to the objective of the experiment. Unbiased estimators are introduced for the procedure and improvements are obtained by the Rao-Blackwell Theorem. An unbiased estimator is recommended which is efficient and easy to compute. Examples show significant improvement in variance reduction over existing approaches.

The computation of $P(s_0)$ can be intensive when N is large because it requires the consideration of $N!$ permutations. The approach is mainly recommended for the experiments in which observations of interest are usually clustered in a few locations; if the number of network with size greater than one in the sample is small, computation is reduced. We are currently developing algorithms to efficiently evaluate $P(s_0)$ under a general setting. The results will be reported elsewhere.

Acknowledgements

The author is grateful to the Editor, an Associate Editor and three referees for their helpful comments and suggestions. This research is supported by NSF grants DMS 0905753 and CMMI 0927572.

Appendix A: Derivation of (3.7) and (3.8)

Because S is the sample space containing all possible initial samples, we have

$\hat{\mu}_{\mathrm{RB}} = E(\hat{\mu}_M \mid D = d)$, which can be written as

$$\hat{\mu}_{\mathrm{RB}} = \sum_{s_0' \in S} \hat{\mu}_M P(S = s_0' \mid D = d) = \sum_{s_0' \in S} \hat{\mu}_M \frac{I(g(s_0') = d)}{\sum_{s_0' \in S} I(g(s_0') = d)}. \tag{A.1}$$

We decompose $\hat{\mu}_M$ into a part that excludes the sample edge units and another that includes the sample edge units. With e_i the indicator variable taking value 1 if the initial sample i is an edge point and $e_i^* = \sum_{j \in \Psi_i} e_j$, we have

$$\begin{aligned}\sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* &= \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) + \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* e_i \\ &= \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) + \sum_{i \in s_0,\ e_i = 1} \frac{P(s_0|i)}{P(s_0)} y_i. \end{aligned} \tag{A.2}$$

The first term on the right side of (A.2) is fixed given $D = d$, so based on (A.1) and (A.2) we have

$$\begin{aligned}\hat{\mu}_{RB} = \Big(\sum_j c_j \Big)^{-1} \Bigg\{ & \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) \\ & + L^{-1} \sum_{s_0' \in S} \left[I(g(s_0') = d) \sum_{i \in s_0',\ e_i = 1} \frac{P(s_0'|i)}{P(s_0')} y_i \right] \Bigg\},\end{aligned}$$

where $L = \sum_{s_0' \in S} I(g(s_0') = d)$.

The variance of $\hat{\mu}_{\mathrm{RB}}$ can be easily derived as $\mathrm{var}(\hat{\mu}_{\mathrm{RB}}) - E[(\hat{\mu}_M - \hat{\mu}_{\mathrm{RB}})^2]$, details are omitted.

Appendix B: Derivation of (3.10) and (3.11)

The easy to compute unbiased estimator is

$$\hat{\mu}^* = E(\hat{\mu}_M \mid D^* = d^*) = \sum_{s_0' \in S} \hat{\mu}_M P(S = s_0' \mid D^* = d^*). \tag{B.1}$$

Similar to Appendix A, we decompose $\hat{\mu}_M$ as

$$\sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* = \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) + \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* e_i. \tag{B.2}$$

With $t_l(i)$ an indicator variable taking 1 if the unit i belongs to level l in factor

x_1 and 0 otherwise, the second term on the right side of (B.2) can be written as

$$\sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* e_i = \sum_{l\in V} \Big[\sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* e_i t_l(i) \Big] = \sum_{l\in V} \sum_{i\in s_0,\ e_i=1} \frac{P(s_0|i)}{P(s_0)} y_i t_l(i).$$

The first term on the right side of (B.2) remains the same given $D^* = d^*$, therefore we have

$$\begin{aligned} &\sum_{s_0'\in S} \Big(\sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* \Big) \frac{I(g(s_0') = d^*)}{\sum_{s_0'\in S} I(g(s_0') = d^*)} \\ &= \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) + (\frac{1}{L^*}) \sum_{s_0'\in S} \Big\{ I(g(s_0') = d^*) \sum_{l\in V} \Big[\sum_{i\in s_0',\ e_i=1} \frac{P(s_0'|i)}{P(s_0')} y_i t_l(i) \Big] \Big\} \\ &= \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) + (\frac{1}{L^*}) \sum_{l\in V} \Big\{ \sum_{s_0'\in S} I(g(s_0') = d^*) \Big[\sum_{i\in s_0',\ e_i=1} \frac{P(s_0'|i)}{P(s_0')} y_i t_l(i) \Big] \Big\}. \end{aligned} \tag{B.3}$$

Because each edge point with the same level of x_1 has an equal probability of being selected as the first unit, and only one is selected, we have $P(s_0'|x_1 = l)/P(s_0')$ the same given l and $D^* = d^*$. Denote this value by $P(s_0|x_1 = l)/P(s_0)$. Hence, the second term on the right side of (B.3) can be written as

$$(\frac{1}{L^*}) \sum_{l\in V} \frac{P(s_0|x_1 = l)}{P(s_0)} \Big\{ \sum_{s_0'\in S} I(g(s_0') = d^*) \Big[\sum_{i\in s_0',\ e_i=1} y_i t_l(i) \Big] \Big\}$$

and since

$$\begin{aligned} \sum_{s_0'\in S} I(g(s_0') = d^*) \Big[\sum_{i\in s_0',\ e_i=1} y_i t_l(i) \Big] &= \frac{L^*}{e_{s_l}} \sum_{i\in s,\ e_i=1} y_i t_l(i) \\ &= \frac{L^*}{e_{s_l}} \sum_{i\in s} e_i y_i t_l(i), \end{aligned}$$

we have

$$\mu^* = \Big(\sum_j c_j \Big)^{-1} \Big\{ \sum_{i=1}^{N} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) + \sum_{l\in V} \frac{P(s_0|x_{11} = l)}{P(s_0)} \sum_{i\in s} \frac{e_i y_i t_l(i)}{e_{s_l}} \Big\},$$

where s is the final sample and e_{s_l} represents the total number of edge points with level l in x_1 in the sample. Therefore, (3.10) holds.

For the variance, we have

$$\begin{aligned}\operatorname{var}(\hat{\mu}^*) &= \operatorname{var}(\hat{\mu}_M) - E[(\hat{\mu}_M - \hat{\mu}^*)^2] \\ &= \operatorname{var}(\hat{\mu}_M) - \sum_{d^* \in D^*} \frac{P(d^*)}{L^*} \sum_{s_0' \in S} I(g(s_0') = d^*)(\hat{\mu}_M - \hat{\mu}^*)^2 \\ &= \operatorname{var}(\hat{\mu}_M) - \sum_{d^* \in D^*} \frac{P(d^*)}{L^*} \sum_{s_0' \in S} \Bigg\{ I(g(s_0') = d^*) \\ &\qquad \Bigg[\sum_{l \in V} \Bigg(\sum_{i \in s_0',\ e_i = 1} \frac{P(s_0'|i)}{P(s_0')} y_i t_l(i) - \frac{P(s_0'|x_{11} = l)}{P(s_0')} \sum_{i \in s} \frac{e_i y_i t_l(i)}{e_{s_l}} \Bigg) \Bigg]^2 \Bigg\}.\end{aligned}$$

References

Cochran, W. G. (1977). *Sampling Techniques*. (3rd edition. Wiley, New York.

Dryver, A. L. and Thompson, S. K. (2005). Improved unbiased estimators in adaptive cluster sampling. *J. Roy. Statist. Soc. Ser. B* **67**, 157-166.

Fang, K. T., Li, R. and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. CRC Press, New York.

Hamann, H. F. (2008). A measurement-based method for improving data center energy efficiency. *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, 312-313.

Hung, Y. (2011). Adaptive probability-based Latin hypercube designs. *J. Amer. Statist. Assoc.* **106**, 213-219.

Hung, Y., Amemiya, Y., and Wu, C. F. J. (2010). Probability-based Latin hypercube designs for slid-rectangular regions. *Biometrika*, to appear.

McKay, M. D., Beckman, R. J. and Conover, W. 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.

Murthy, M. N. (1957). Ordered and unordered estimators in sampling without replacement. *Sankhyā* **18**, 379-390.

Raj, D. (1956). Some estimators in sampling with varying probabilities without replacement. *J. Amer. Statist. Assoc.* **51**, 269-284.

Salehi, M. S. M. and Seber, G. A. F. (1997). Adaptive cluster sampling with networks selected without replacement. *Biometrika* **86**, 209-219.

Santner, T., Williams, B., and Notz, W. (2003), *The Design and Analysis of Computer Experiments*. Springer, New York.

Schmidt, R. R., Cruz, E. E., and Iyengar, M. K. (2005). Challenges of data center thermal management. *IBM Journal of Research and Development* **49**, 709-723.

Stevens, D. L. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *J. Amer. Statist. Assoc.* **99**, 262-278.

Thompson, S. K. (1990). Adaptive cluster sampling. *J. Amer. Statist. Assoc.* **85**, 1050-1059.
Thompson, S. K. (1991). Stratified adaptive cluster sampling. *Biometrika* **78**, 389-397.
Thompson, S. K. and Seber, G. A. F. (1996). *Adaptive Sampling*. Wiley, New York.

Department of Statistics and Biostatistics, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, USA.
E-mail: yhung@stat.rutgers.edu

(Received November 2012; accepted June 2013)