Adaptive Probability-based Latin Hypercube Designs

Ying Hung

Department of Statistics and Biostatistics,
Rutgers, the State University of New Jersey, Piscataway, NJ

Abstract

Adaptive sampling is an effective method developed mainly for regular regions. However, experimental regions in irregular shapes are commonly observed in practice. Motivated by a data center thermal management study, a new class of adaptive designs is proposed to accommodate a specific type of irregular regions. Because the adaptive procedure introduces biases into conventional estimators, several design-unbiased estimators are given for estimating the population mean. Efficient and easy-to-compute unbiased estimators are also introduced. The proposed method is applied to obtain an adaptive sensor placement plan to monitor and study the thermal distribution in a data center.

KEY WORDS: Adaptive cluster sampling; Computer experiment; Latin hypercube design; Rao-Balckwell method; Sequential design; Space-filling design.

1. INTRODUCTION

A data center is an integrated facility housing multiple-unit servers, providing application services or management for data processing. Data center facilities constantly generate large amounts of heat to the room (See Figure 1 for an example of heat map in a data center), which must be maintained at an acceptable temperature for reliable operation of the equipment. A significant fraction of the total power consumption is for heat removal. Therefore, efficient cooling mechanism of a data center has become a major challenge. The objective for the thermal management study is to model the thermal distribution in data center and the final goal is to design a data center with efficient heat removal mechanism (Schmidt et al., 2005).
A challenging design issue arises in the data center thermal study. To monitor and study the thermal distribution, especially area close to data center facilities, sensors are attached to surfaces of facilities to measure the temperatures. An important question is how to allocate the sensors over the data center such that the temperature can be estimated accurately. This includes an initial plan that can place sensors uniformly over the experimental region and an adaptive plan that can adaptively increase sample effort in the vicinity of observed values that are high based on the initial sensor observations.

Despite the prevalence of space-filling designs and adaptive designs in many applications (McKay et al., 1979; Thompson and Seber, 1996, Santner et al., 2003; Fang et al., 2006), they are mainly developed under the assumption that the experimental region is rectangular. This assumption, however, does not hold in this study. First, data center may not be in a rectangular shape. This leads to non-rectangular allocation of racks (grey area in Figure 1), where data center facilities are designed to be stored in. Second, even if the racks are located in a rectangular region, the data center facilities can be stored irregularly because of the usage limitation. Such a non-rectangular region is called slid-rectangular (Hung et al., 2010), where the desirable range of one factor depends on the level of another factor. For example, a slid-rectangular region is shown in Figure 2 where the range of factor $x_1$ depends on the level of factor $x_2$. Because of the violation of the standard assumption, direct application of the
existing methods to slid-rectangular regions can easily fail. Therefore, new experimental
designs are called for to address the foregoing questions in the data center study.

In this article, a new class of adaptive design is proposed for slid-rectangular regions. It
is an adaptive sampling plan with initial samples following probability-based Latin hypercube
designs (PLHDs) (Hung et al., 2010). PLHDs is a class of designs which takes into account
the slid-rectangular structure in the construction of design so that the resulting designs still
maintain desirable space-filling properties, such as the one-dimensional balance and propor-
tional allocation properties. Even though PLHDs successfully provide space-filling designs
for slid-rectangular regions, how to adaptively sample design points based on PLHDs and the
resulting inferences of the adaptive design remain unsolved. The existing adaptive sampling
techniques (Thompson and Seber, 1996; Thompson, 1990, 1991) cannot be utilized. This is
because they are developed for independent and equal inclusion probabilities, but the inclu-
sion probabilities are no longer equal in order to achieve one-dimensional balance in PLHDs.
Furthermore, the desirable space-filling properties of PLHDs, such as proportional balance,
result in the dependence of the inclusion probabilities. These features pose challenges in
the study of adaptive design with slid-rectangular regions, especially the construction of
unbiased estimators.

Although the new class of adaptive designs is motivated from a data center study, slid-
rectangular experimental regions are generally observed in practice. Therefore, the proposed
designs can be applied in many areas, such as computer experiments (Kjell, 2000; Stinstra,
2003), network sampling (Kalton and Anderson, 1986), and ecological study (Czaja et al.,
1986; Seber, 1986; Sudman et al., 1988). The remainder of this article is organized as follows.
A new class of adaptive designs and detailed construction procedure is introduced in Section
2. In Section 3, new unbiased estimators are proposed and their variances are studied.
Furthermore, efficient estimators that successfully reduce variations are introduced. These
estimators are illustrated and compared via simulation studies in Section 4. In Section 5,
data center example is revisited and the proposed methods are applied to demonstrate the
performance. Conclusions and remarks are given in Section 6.
2. ADAPTIVE PROBABILITY-BASED LATIN HYPEUCUBE DESIGNS

2.1 Probability-based Latin hypercube designs

Space-filling designs (Santner et al., 2003; Fang et al., 2006), such as Latin hypercube design (McKay et al., 1979), are widely used to spread out the design points in the experimental regions. However, they have limitations because of the rectangular assumption of the experimental regions. Application of the existing methods to slid-rectangular regions can no longer maintain good space-filling properties. For instance, examples that the desirable space-filling properties, such as one-dimensional balance property, does not hold when LHDs are applied to slid-rectangular regions can be easily found. Therefore, PLHDs are introduced to provide space-filling designs in slid-rectangular regions.

The basic idea of PLHDs is to maintain the one-dimensional balance property in slid-rectangular regions so that the design points can be spread out uniformly as they should be. A general construction procedure for PLHDs is given as follows. Hereafter, assume that factors, $x_1$ and $x_2$, form a slid-rectangular region. Here, $x_2$ can be a quantitative or qualitative factor with $k$ predetermined levels and the experimental ranges for $x_1$ are located irregularly on the interval $[A, B]$. Specifically, for the $j$th level of $x_2$, the feasible interval for $x_1$ is denoted by $E_j = (A_j, B_j)$. Thus, we have $A = \min\{A_j\}$ and $B = \max\{B_j\}$ for $j = 1, 2, \ldots, k$. Assume we plan for an $n$-run experiment and each design point $u_i$ can be represented by $u_i = (x_{1i}, x_{2i})'$, where $i = 1, \ldots, n$ and $n \geq k$. According to the PLHD procedure, the one-dimensional balance on $x_1$ is achieved by dividing the interval $[A, B]$ into $n$ equally spaced sub-intervals and assigning the $n$ levels of $x_1$ by the middle of these sub-intervals (i.e. $x_{11} = 1, \ldots, x_{1n} = n$). Then, for each level of $x_1$, the feasible range for $x_2$ is defined by $C_i$, $i = 1, \ldots, n$, and the level of $x_2$ is assigned by

$$P(x_{2i} = j) = \begin{cases} c_i^{-1} & \text{if } j \in C_i \\ 0 & \text{if } j \notin C_i \end{cases}$$

where $c_i = \sum_{j=1}^k I[j \in C_i]$, $j = 1, \ldots, k$, and $i = 1, \ldots, n$.

It is clear that the number of observations for each level of $x_2$ is not proportional to the length of the experimental interval based on the definition of PLHDs. Therefore, a
further improvement is to incorporate the proportional balance property, where the number of observations is proportional to the length of the interval. They are named balanced probability-based Latin hypercube designs (BPLHDs). Define the length of the gray area by 
\[ B_j - A_j = \sum_{i=1}^{n} I(j \in C_i) \text{ and } j = 1, \cdots, k. \] 
A BPLHD can be written as a modification of the \( n \)-run PLHD in (1) with the constraints
\[
\frac{\sum_{i=1}^{n} I[x_{2i} = j]}{n} = \frac{(B_j - A_j)}{\sum_{j=1}^{k} (B_j - A_j)} = p_j, \text{ for } j = 1, \cdots, k. \] (2)

In practice, the quantities \( n_j \) are not always integers for given \( n \) and \( p_j \). In that situation, an approximate balance with \( |n_j - np_j| < 1 \) should be considered. Figure 2(a) is an example of the BPLHD. Detailed procedure for the BPLHD construction can be found in Hung et al. (2010).

The choices of PLHDs and BPLHDs are not unique and they are not equally good. Therefore, some optimal design criteria are needed for further discrimination (Iman and Conover, 1982; Johnson et al., 1990; Owen, 1994; Tang, 1998; Joseph and Hung, 2008). More discussions on design criteria and efficient searching algorithms are referred to Hung et al. (2010).

2.2 Adaptive probability-based Latin hypercube designs

Based on an initial set of observations collected by PLHDs or BPLHDs, denoted by \( s_0 \), adaptive probability-based Latin hypercube designs or adaptive balanced probability-based Latin hypercube designs can be obtained. They refer to designs in which \( n \) initial observations are selected according to PLHDs or BPLHDs, and, whenever the response of a selected unit satisfies a given criterion, additional units in the neighborhood of that unit are added to the sample. If any of the additional-added units satisfies the condition, more units may be added to the sample. This procedure continues until no more units are found that meet the condition. Therefore, the final design contains every unit in the neighborhood of any sample unit satisfying the condition.

The neighborhood of a sample unit is defined within the slid-rectangular region and assumed to be independent to the response. The neighborhood relation is symmetric, which is similar to that in the adaptive cluster sampling (Thompson, 1990), i.e. if unit \( j \) is in the
neighborhood of unit \( i \), then unit \( i \) is in the neighborhood of unit \( j \). In this paper, geographically nearest neighbors are considered in the slid-rectangular region. More discussions on various neighborhood definitions can be found in Thompson (1990). Incorporating the slid-rectangular shapes, extensions to those definitions can be easily made. The condition for adding additional selection of neighborhood points can be defined based on the response. Here we define a design point satisfies the condition if the response (i.e., the variable of interest) is greater than or equal to some constant \( \nu \).

Figure 2 demonstrates a 22-run BPLHD and the resulting adaptive design. For the slid-rectangular region (denoted by the dashed lines), factor \( x_2 \) has five levels and the proportional lengths of \( x_1 \) at different levels of \( x_2 \) are 3 : 4 : 5 : 6 : 4. The feasible ranges (i.e., set of levels) are \( C_1 = \{1\} \), \( C_2 = \{1\} \), \( C_3 = \{1, 2\} \), etc. The triangles shown in Figure 2(a) are design points of the BPLHD. Clearly, both the proportional allocation property and the one-dimensional balance property hold. The bullet points indicate that the associated responses satisfy the condition. The neighborhood of each point in this example consists of, in addition to itself, the intersection of the four spatially adjacent points and the slid-rectangular region. Figure 2(b) illustrates the final sample resulting from initial design in Figure 2(a).

For the adaptive designs for slid-rectangular regions, several definitions are analogue to that in the adaptive sampling literature (Thompson and Seber, 1996). The set of all units satisfying the condition in the neighborhood of one another is called a network. According to the definition, selection in the initial design of any point in a network will result in the final sample of all units in that network. For example, the four adjacent bullet points in the middle of Figure 2(a) belong to the same network. For any unit that does not meet the condition, it forms a network with size one, i.e., consisting just of itself. The population can be partitioned into \( K \) networks. The units that do not meet the condition but in the neighborhood of some design points that satisfy the condition are called edge units.
Figure 2: (a) An example of balanced probability-based Latin hypercube Design. (b) The resulting adaptive BPLHD.
3. ESTIMATOR

Despite some commonly shared concepts, the adaptive designs for slid-rectangular regions differ from the conventional adaptive sampling (Thompson, 1990) in which initial samples are usually drawn independently with equal inclusion probabilities. First, unequal inclusion probabilities are assigned over the experimental region for both adaptive PLHDs and adaptive BPLHDs. Second, the independence of the inclusion probabilities is violated in order to achieve the one-dimensional balance in slid-rectangular regions. Third, it is desirable to have the number of design points proportional to the length of the experimental region in a slid-rectangular region. Maintaining such proportional balance in PLHDs further complicates the relationship of the inclusion probabilities. It is shown that the inclusion probability of each sample in a BLHD depends highly on the other points in the sample (Hung et al., 2010). Because of these distinct features, standard estimation methods for adaptive sampling cannot be applied and new estimators and related inference for slid-rectangular regions need to be rigorously established.

In this section, unbiased estimators for the population mean are the main focus. Conventional unbiased estimators, such as the sample mean, are no longer unbiased with adaptive PLHDs and adaptive BPLHDs. Hence, new unbiased estimators are introduced and their variances and unbiased estimators for variances are discussed. A new class of modified unbiased estimators are proposed to further reduce the variation. Derivations of the unbiased estimators and the unbiased estimators of variance are given in the Appendix.

3.1 Estimators for adaptive PLHD

Assume that there are $N$ units in the population with the corresponding responses $y_1, \ldots, y_N$ and the population mean is defined by $\mu = N^{-1} \sum_{i=1}^{N} y_i$. For the adaptive PLHDs, a Horvitz-Thompson type of design-unbiased estimator (Cochran, 1977; Horvitz and Thompson, 1952; Thompson, 1990, 1991) is introduced based on the inclusion probabilities. To do so, the following notation is needed. Let $\Psi_k$ be the set of units in the $k$th network. Because of the non-regular shape of the experimental region, we further partition the units in $\Psi_k$ by their coordinates in the $x_1$-axis. Let $\Psi_k$ ranges in $\psi_k = (t_1, \ldots, t_k) \in (1, \ldots, n)$ in
the $x_1$-axis. For each observation in $\psi_k$, the associated units in the network are $\Psi_{kl}$, where $l \in \psi_k$, and the number of units in $\Psi_{kl}$ is denoted by $d_{kl}$. Using the foregoing notation, the number of units selected from the $k$th network in the initial sample can be written as

$$n_k = \sum_{i \in \Psi_k} I(i \in s_0),$$

and the inclusion probability of network $k$ is

$$P(n_k > 0) = 1 - \prod_{l \in \psi_k} \left(1 - \frac{(c_l - d_{kl})}{c_l}\right).$$

(3)

Based on the inclusion probability, the unbiased estimator can be written as

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^*I(n_k > 0)}{P(n_k > 0)},$$

(4)

where $K$ denotes the number of networks in the population and $y_k^* = \sum_{j \in \psi_k} y_j$. The indicator variable $I(n_k > 0)$ takes 1 if any unit of the $k$th network is in the initial sample $s_0$, and takes 0 otherwise.

The variance of the unbiased estimator can be calculated by

$$\operatorname{var}(\hat{\mu}) = N^{-2} \sum_{k=1}^{K} y_k^* y_h^* \frac{P(n_k > 0, n_h > 0) - p(n_k > 0)P(n_h > 0)}{P(n_k > 0)P(n_h > 0)},$$

where

$$P(n_k > 0, n_h > 0) = 1 - \prod_{l \in \psi_k} \left(1 - \frac{(c_l - d_{kl})}{c_l}\right) - \prod_{l \in \psi_h} \left(1 - \frac{(c_l - d_{hl})}{c_l}\right) + \prod_{l \in (\psi_k \cup \psi_h)} \left(1 - \frac{(c_l - d_{k\cup h, l})}{c_l}\right).$$

and $d_{k\cup h, l}$ is the number of units in $\Psi_{kl} \cup \Psi_{hl}$. An unbiased estimator of the variance of $\hat{\mu}$ is

$$\hat{\operatorname{var}}(\hat{\mu}) = N^{-2} \sum_{k=1}^{K} \sum_{h=1}^{K} y_k^* y_h^* \frac{P(n_k > 0, n_h > 0) - p(n_k > 0)P(n_h > 0)}{P(n_k > 0)P(n_h > 0)P(n_k > 0, n_h > 0)} I(n_k > 0)I(n_h > 0).$$

(5)

The foregoing unbiased estimator can be improved by incorporating more of the information obtained in the final sample. Particularly, according to (4), the observations from edge points are used in the estimator only if they are included in the initial sample. Therefore, using the Rao-Blackwell method, an improved unbiased estimator can be obtained by
calculating the conditional expectation of the original estimator, given a sufficient statistics. The most efficient choice is the minimal sufficient statistics.

Based on PLHDs, the minimum sufficient statistics \( m \) is the unordered set of distinct, labeled observations, i.e., \( m = \{ (i, y_i) : i \in s \} \). Define \( M \) as the sample space for \( m \), \( g(s'_0) \) as the function that maps an initial design \( s'_0 \) into a value of \( m \), and \( S \) as the sample space containing all possible samples. The resulting improved unbiased estimator for adaptive PLHD is

\[
\hat{\mu}^{RB} = E(\mu | M = m) = \frac{1}{N} \sum_{k=1}^{K} \frac{g_k(1-e_k^*)}{P(n_k > 0)} + \frac{1}{NL} \sum_{s'_0 \in S} \{ I(g(s'_0) = m) \left[ \sum_{i \in s_0, e_i = 1} y_i c_i \right] \},
\]

where \( e^* = \sum_{i \in \Psi_k} e_i \) and \( e_i = 1 \) if unit \( i \) is an edge point and \( e_i = 0 \) otherwise. The variance of this improved unbiased estimator can be written as

\[
\text{var}(\hat{\mu}^{RB}) = \text{var}(\hat{\mu}) - \sum_{m \in M} \frac{P(m)}{L} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = m) \left[ (\hat{\mu} - \hat{\mu}^{RB})^2 \right],
\]

where \( L \) is the number of initial designs that are \textit{compatible} with \( m \) and \( P(m) \) is the probability that \( M = m \). An unbiased estimator of the variance is

\[
\text{var}(\hat{\mu}^{RB}) = \text{var}(\hat{\mu}) - L^{-1} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = m) \left[ (\hat{\mu} - \hat{\mu}^{RB})^2 \right]
\]

and a more efficient estimator can be further obtained by conditioning on the minimum sufficient statistics as follows

\[
\text{var}(\hat{\mu}^{RB}) = E(\text{var}(\hat{\mu}^{RB}) | M = m) = \frac{1}{L} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = m) \text{var}(\hat{\mu}) - \frac{1}{L} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = m) \left[ (\hat{\mu} - \hat{\mu}^{RB})^2 \right].
\]

Although conditioning on the minimum sufficient statistics is most efficient, it is computationally difficult for large designs because one has to evaluate all the compatible designs in order to obtain the estimation. This is not surprising and the same difficulty is experienced in the conventional adaptive sampling (Salehi, 1999; Dryver and Thompson, 2005). To tackle the computational issue, a modification is proposed. The idea is to construct an unbiased estimator by conditioning on a carefully chosen sufficient statistics, instead of the minimum sufficient statistics.
Let $s$ denote the final sample and define $s_c$ as the set of all the distinct units in the sample for which the condition to sample adaptively is satisfied. The remaining part is denoted by $s_c$. Define $V$ as a collection of $x_1$ coordinates with which edge points occurs in the initial sample. For unit $i$, let $f_i$ be the number of times that the network to which unit $i$ belongs is intersected by the initial sample. Using the above notation, a sufficient statistics can be defined by

$$m^* = \{(i, y_i, f_i), V, (j, y_j) : i \in s_c, j \in s_c\},$$

and the sample space for $m^*$ is defined by $M^*$. Hence, the improved unbiased estimator by conditioning on the sufficient statistics $m^*$ can be obtained by

$$\hat{\mu}^* = E(\mu | M^* = m^*)$$

$$= \frac{1}{N} \sum_{k=1}^{K} y_k I(n_k > 0) \left(\frac{1 - e_k^*}{P(n_k > 0)}\right) + \frac{1}{N} \sum_{l \in V} \sum_{i \in S} e_i y_i t_l(i),$$

where $t_l(i)$ is an indicator variable taking 1 if the unit $i$ belongs to level $l$ in factor $x_1$ (column $l$) and 0 otherwise and $e_{s_l} = \sum_{i \in S} e_i t_l(i)$. Note that $\hat{\mu}^*$ has a smaller variance compared to the original unbiased estimator $\hat{\mu}$, and more importantly it is much easier to evaluate compared to $\hat{\mu}^{RB}$.

The variance of the improved unbiased estimator is

$$\text{var}(\hat{\mu}^*) = \text{var}(\hat{\mu}) - \frac{1}{L} \sum_{m^* \in M^*} \frac{P(m^*)}{L} \sum_{s_0^* \in S} \left\{ I(g(s_0^*) = m^*) \right\}$$

$$\left[ \sum_{i=1}^{n} \frac{c_i}{N} \left( \sum_{i \in S} e_i y_i t_l(i) - \frac{1}{e_{s_l}} \sum_{i \in S} e_i y_i t_l(i) \right) \right]^2,$$

An unbiased estimator of the variance is

$$\tilde{\text{var}}(\hat{\mu}^*) = \text{var}(\hat{\mu}) - \frac{1}{L} \sum_{s_0^* \in S} \left\{ I(g(s_0^*) = m^*) \right\}$$

$$\left[ \sum_{i=1}^{n} \frac{c_i}{N} \left( \sum_{i \in S} e_i y_i t_l(i) - \frac{1}{e_{s_l}} \sum_{i \in S} e_i y_i t_l(i) \right) \right]^2,$$

and a more efficient estimator of the variance can be obtained by

$$\hat{\text{var}}(\hat{\mu}^*) = E[\tilde{\text{var}}(\hat{\mu}^*) | M^* = m^*]$$

$$= \frac{1}{L} \sum_{s_0^* \in S} I(g(s_0^*) = m^*) \text{var}(\hat{\mu}(s_0^*)) - \frac{1}{L} \sum_{s_0^* \in S} \left\{ I(g(s_0^*) = m^*) \right\}$$

$$\left[ \sum_{i=1}^{n} \frac{c_i}{N} \left( \sum_{i \in S} e_i y_i t_l(i) - \frac{1}{e_{s_l}} \sum_{i \in S} e_i y_i t_l(i) \right) \right]^2.$$
3.2 Estimators for the balanced probability-based LHD

The BPLHDs are constrained PLHDs which achieve not only one-dimensional balance but also proportional balance. In order to achieve these desirable space-filling properties, the inclusion probabilities become highly correlated. To calculate both the inclusion probabilities and unbiased estimators, it will be most convenient to define the following notation. Define \( w_i \) as an indicator variable with \( w_i = 1 \) if unit \( i \) is selected in the sample and \( w_i = 0 \) otherwise. The unbiased estimator for BPLHDs, denoted by \( \hat{\mu}_B \), and its corresponding variance can be obtained using (4) and (5) with

\[
P(n_k > 0) = 1 - P(\cap_{i \in \Psi_k} w_i = 0)
\]

and

\[
P(n_k > 0, n_h > 0) = 1 - P(\cap_{i \in \Psi_k} w_i = 0) - P(\cap_{i \in \Psi_h} w_i = 0) + P(\cap_{i \in (\Psi_k \cup \Psi_h)} w_i = 0).
\]

Similar to the adaptive PLHDs, the unbiased estimator \( \hat{\mu}_B \) can be improved by taking the expected value of \( \hat{\mu}_B \) conditional on the minimum sufficient statistics. Therefore, a more efficient unbiased estimator for the adaptive BPLHDs can be written as

\[
\hat{\mu}_{RB} = E(\hat{\mu}_{rmB}|M = m) = \frac{1}{N} \sum_{k=1}^{K} \frac{g_k J_k (1-c_k^s)}{P(n_k > 0)} + \frac{1}{NL} \sum_{s' \in S} \sum_{k=1}^{K} \frac{y_k^* J_k e_k}{P(n_k > 0)}.
\]

The variance of this unbiased estimator can be defined by replacing \( \hat{\mu} \) and \( \hat{\mu}_{RB} \) by \( \hat{\mu}_B \) and \( \hat{\mu}_{RB} \) in (8).

Note that the idea of constructing the efficient and easy-to-compute estimator in adaptive PLHDs cannot be extended to adaptive BPLHDs because all of the inclusion probabilities in BLHDs are not necessarily the same and highly correlated. Therefore, inclusion-exclusion formulas derived by Salehi (1999) should be considered to systematically compute the Rao-Blackwell estimators. This method is computationally efficient provided that relatively few networks meeting the condition are intersected, regardless of the size of the network and the number of times the networks were intersected. However, it can be computationally difficult when the number of networks meeting the condition increases and calculating the inclusion
probabilities can be also computationally intensive for larger designs. Hence, computational aspects for adaptive BPLHDs deserve further study.

4. SIMULATION

The proposed adaptive designs are applied to two small examples with slid-rectangular regions in this section. Detail evaluations of the unbiased estimators are illustrated and the variance of these estimators is compared.

4.1 Example of an adaptive PLHD

Figure 3 illustrates a simple example of slid-rectangular region in which there are 3 levels for factor $x_2$ and the feasible regions of $x_1$ depend on the level of $x_2$. The numbers shown in the cells are the responses and the last row indicates the inclusion probability of each unit from the associated column. Based on this region, there are 8 possible 4-run PLHDs and the resulting adaptive designs are listed in Table 1. For each row in Table 1, the first four design points listed are the PLHDs and the rest of them are adaptively added when the responses are larger than or equal to 60, i.e., $\nu = 60$. The column $V$ indicates the collection of $x_1$ coordinates with which edge points occur in the initial samples. Three unbiased estimators, $\hat{\mu}$, $\hat{\mu}^*$, and $\hat{\mu}^{RB}$, and their corresponding variances are calculated for each design. The last row summarizes the average performances of these estimators over all possible designs.

Take the first adaptive PLHD as an example, the observations from the initial design are $y_{11}, y_{21}, y_{32},$ and $y_{42}$, where $y_{ij}$ represents the response observed in unit $(i, j)$. Additionally, three observations, $y_{33}, y_{22},$ and $y_{43},$ are added to the final sample. The observations $y_{32}$ and $y_{33}$ belong to the same network and the probability of including at least one point from this
Table 1: All possible adaptive PLHDs and a comparison of the unbiased estimators

<table>
<thead>
<tr>
<th>sampler</th>
<th>$V$</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\mu}^*$</th>
<th>$\hat{\mu}^{RB}$</th>
<th>$\text{vár}(\hat{\mu})$</th>
<th>$\text{vár}(\hat{\mu}^*)$</th>
<th>$\text{vár}(\hat{\mu}^{RB})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${11}, {21}, {32}, {42}; {33}, {22}, {43}$</td>
<td>4</td>
<td>32.29</td>
<td>26</td>
<td>26</td>
<td>591.76</td>
<td>277.47</td>
<td>320.24</td>
</tr>
<tr>
<td>${11}, {21}, {32}, {43}; {33}, {22}, {42}$</td>
<td>4</td>
<td>19.71</td>
<td>26</td>
<td>26</td>
<td>42.20</td>
<td>277.47</td>
<td>320.24</td>
</tr>
<tr>
<td>${11}, {21}, {33}, {42}; {32}, {42}, {22}$</td>
<td>4</td>
<td>32.29</td>
<td>26</td>
<td>26</td>
<td>591.76</td>
<td>277.47</td>
<td>320.24</td>
</tr>
<tr>
<td>${11}, {21}, {33}, {43}; {32}, {22}, {42}$</td>
<td>4</td>
<td>19.71</td>
<td>26</td>
<td>26</td>
<td>42.20</td>
<td>277.47</td>
<td>320.24</td>
</tr>
<tr>
<td>${11}, {22}, {32}, {42}; {33}, {43}$</td>
<td>2, 4</td>
<td>33.43</td>
<td>27.14</td>
<td>27.14</td>
<td>634.53</td>
<td>320.24</td>
<td>320.24</td>
</tr>
<tr>
<td>${11}, {22}, {32}, {43}; {33}, {42}$</td>
<td>2, 4</td>
<td>20.86</td>
<td>27.14</td>
<td>27.14</td>
<td>84.98</td>
<td>320.24</td>
<td>320.24</td>
</tr>
<tr>
<td>${11}, {22}, {33}, {42}; {32}, {43}$</td>
<td>2, 4</td>
<td>33.43</td>
<td>27.14</td>
<td>27.14</td>
<td>634.53</td>
<td>320.24</td>
<td>320.24</td>
</tr>
<tr>
<td>${11}, {22}, {33}, {43}; {32}, {42}$</td>
<td>2, 4</td>
<td>20.86</td>
<td>27.14</td>
<td>27.14</td>
<td>84.98</td>
<td>320.24</td>
<td>320.24</td>
</tr>
</tbody>
</table>

Mean: $26.57$ $26.57$ $26.57$ $338.37$ $298.86$ $298.86$

Network is 1. Therefore, the unbiased estimator is calculated by

$$
\hat{\mu} = \frac{1}{7} \left[ \frac{y_{11}}{1} + \frac{y_{21}}{1/2} + \frac{(y_{32} + y_{33})}{1} + \frac{y_{42}}{1/2} \right] \\
= \frac{1}{7} \left[ \frac{2}{1} + \frac{1}{1/2} + \frac{(62 + 66)}{1} + \frac{47}{1/2} \right] = 32.29.
$$

The edge points collected from the initial design is $y_{42}$ in this sample, hence $V = 4$ and

$$
\hat{\mu}^* = \frac{1}{7} \left[ \frac{y_{11}}{1} + \frac{y_{21}}{1/2} + \frac{(y_{32} + y_{33})}{1} + \frac{(y_{42} + y_{43})}{1/2} \right] \\
= \frac{1}{7} \left[ \frac{2}{1} + \frac{1}{1/2} + \frac{(62 + 66)}{1} + \frac{(47 + 3)}{1/2} \right] = 26.
$$

To calculate the last unbiased estimator, information is required from all possible designs that have the same minimum sufficient statistics, $m$. Clearly the first four designs have the same minimum sufficient statistics, and the edge points in these four initial designs are $y_{42}$, $y_{43}$, $y_{42}$, and $y_{43}$ respectively. Therefore,

$$
\hat{\mu}^{RB} = \frac{1}{7} \left[ \frac{y_{11}}{1} + \frac{y_{21}}{1/2} + \frac{(y_{32} + y_{33})}{1} + \frac{2y_{42} + 2y_{43} + 2y_{42} + 2y_{43}}{4} \right] \\
= \frac{1}{7} \left[ \frac{2}{1} + \frac{1}{1/2} + \frac{(62 + 66)}{1} + \frac{94 + 6 + 94 + 6}{4} \right] = 26.
$$
Note that even though $\hat{\mu}^*$ and $\hat{\mu}^{\text{RB}}$ have the same estimations in this example, the way they are calculated is different and $\hat{\mu}^*$ is easier to obtain because only the information from the sample itself is needed. The reason of having the same estimation for both $\hat{\mu}^*$ and $\hat{\mu}^{\text{RB}}$ is that the sufficient statistics $m^*$ and the minimum sufficient statistics $m$ lead to the same partition of designs in this example. These results, including the estimation and variance, can be different in general. The three estimators are unbiased and the average variance of the improved estimators, $\hat{\mu}^*$ and $\hat{\mu}^{\text{RB}}$, are 12% ($= (338.37 - 298.86)/338.37$) smaller than that from the original estimator, $\hat{\mu}$.

### 4.2 Example of an adaptive BPLHD

Figure 4 demonstrates another slid-rectangular region with 3 levels for $x_2$ and the same length of $x_1$ at each level of $x_2$. BPLHDs with 6 design points are applied. Two values are listed in each cell. The first value is the response and the one in the parentheses is the inclusion probability. There are in total seven 6-run BPLHDs and the resulting adaptive designs are listed in Table 2. Similar to Table 1, each row in the table corresponds to an adaptive BPLHD with the BPLHD listed as the first six design points and the rests adaptively selected by $\nu = 20$. For the adaptive BPLHDs, two types of unbiased estimators are calculated and their variances are compared in Table 2. The average performances of these estimators are summarized in the last row. It is clear that both estimators are unbiased and the average variance of the Rao-Blackwell estimator has 16% ($= (4.79 - 4.02)/4.79$) variance reduction compared to the original estimator.

To illustrate how the estimators in the adaptive BPLHDs are calculated, we consider the first design as an example in which nine points are selected in the final sample. Note that
Table 2: All possible adaptive BPLHDs and a comparison of the unbiased estimators

<table>
<thead>
<tr>
<th>sampler</th>
<th>$\hat{\mu}_B$</th>
<th>$\hat{\mu}_B^{RB}$</th>
<th>$\text{vár}(\hat{\mu}_B)$</th>
<th>$\text{vár}(\hat{\mu}_B^{RB})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${11}, {21}, {32}, {42}, {53}, {63}; {52}, {43}, {41}$</td>
<td>6.99</td>
<td>8.15</td>
<td>2.23</td>
<td>5.36</td>
</tr>
<tr>
<td>${11}, {21}, {32}, {43}, {52}, {63}; {42}, {41}, {53}$</td>
<td>9.32</td>
<td>8.15</td>
<td>11.22</td>
<td>5.36</td>
</tr>
<tr>
<td>${11}, {21}, {33}, {42}, {52}, {63}; {32}, {41}, {43}, {53}$</td>
<td>4.85</td>
<td>4.85</td>
<td>3.24</td>
<td>3.24</td>
</tr>
<tr>
<td>${11}, {22}, {31}, {42}, {53}, {63}; {52}, {32}, {41}, {43}$</td>
<td>5.72</td>
<td>6.89</td>
<td>2.50</td>
<td>5.40</td>
</tr>
<tr>
<td>${11}, {22}, {31}, {43}, {52}, {63}; {42}, {32}, {41}, {53}$</td>
<td>8.06</td>
<td>6.89</td>
<td>11.02</td>
<td>5.40</td>
</tr>
<tr>
<td>${11}, {22}, {32}, {41}, {53}, {63}$</td>
<td>3.68</td>
<td>3.68</td>
<td>3.11</td>
<td>3.11</td>
</tr>
<tr>
<td>${11}, {22}, {33}, {41}, {52}, {63}; {42}, {32}, {43}, {53}$</td>
<td>6.42</td>
<td>6.31</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>mean</td>
<td>6.42</td>
<td>6.42</td>
<td>4.79</td>
<td>4.02</td>
</tr>
</tbody>
</table>

$y_{42}$ and $y_{52}$ belong to the same network and among the seven initial BPLHDs, there is only one design (i.e., the sixth design) not including this network. Therefore, $P(w_{42} = 0 \cap w_{52} = 0) = 1/7$ and thus the inclusion probability of this network is $6/7$ and

$$\hat{\mu} = \frac{1}{12} \left[ \frac{y_{11}}{7/7} + \frac{y_{21}}{3/7} + \frac{y_{32}}{3/7} + \frac{(y_{42}+y_{52})}{6/7} + \frac{y_{53}}{3/7} + \frac{y_{63}}{7/7} \right]$$

$$= \frac{1}{12} \left[ \frac{1}{7/7} + \frac{0}{3/7} + \frac{8}{3/7} + \frac{(24+25)}{6/7} + \frac{3}{3/7} + \frac{0}{7/7} \right] = 6.99.$$  

To obtain the improved unbiased estimator, the first two designs are needed because they have the same minimum sufficient statistics. The edge units from the initial sample are $y_{32}$ and $y_{53}$ in the first design, while they are $y_{32}$ and $y_{43}$ in the second design. Therefore,

$$\hat{\mu}_B^{RB} = \frac{1}{12} \left[ \frac{y_{11}}{7/7} + \frac{y_{21}}{3/7} + \frac{(y_{42}+y_{52})}{6/7} + \frac{y_{63}}{3/7} + \frac{1}{2} \left( \frac{y_{32}}{3/7} + \frac{y_{53}}{3/7} + \frac{y_{43}}{3/7} \right) \right]$$

$$= \frac{1}{12} \left[ \frac{1}{7/7} + \frac{0}{3/7} + \frac{(24+25)}{6/7} + \frac{0}{7/7} + \frac{1}{2} \left( \frac{8}{3/7} + \frac{3}{3/7} + \frac{8}{3/7} + \frac{10}{2/7} \right) \right] = 8.15.$$  

### 5. APPLICATION IN DATA CENTER THERMAL MANAGEMENT

In this section, we revisit the data center thermal management study and apply the proposed method to place the sensors in a data center located in the IBM T. J. Watson
The objective is to obtain an initial plan so that the sensors can be allocated uniformly over the data center. Based on the initial sensor observations, more sensors are adaptively placed in the neighborhood sites of the existing sensors that have higher temperature observations because these sites are more likely to have higher temperature.

Figure 5 is the two-dimensional layout of the data center. Sensors can be located in six rows (8 racks) with different lengths, denoted by the dashed lines. This layout results in a slid-rectangular experimental region which can be specified by $x_1$ and $x_2$. There is another coordinate, height, which determines the exact sensor locations. In this study, we fixed all sensors to the top of the rack because the temperature increases theoretically with height and monitoring the highest temperature is more desirable for the safety of a data center. There are 24 sensors available for the initial design. Therefore, denoted by circles in Figure 5, a 24-run PLHD is considered. Sensors appear to be spread out uniformly over the experimental region according to this design. Based on these initial observations, three more sensors are added by the proposed adaptive sampling strategies with $\nu = 34^\circ \text{C}$ and the estimated temperature is $\hat{\mu} = \hat{\mu}^* = 24.47^\circ \text{C}$. 
Table 3: Comparison of estimators in the data center study

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\mu}$</th>
<th>$\hat{\mu}^*$</th>
<th>$\hat{\mu}_{SRS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>24.20</td>
<td>24.20</td>
<td>24.21</td>
</tr>
<tr>
<td>variance</td>
<td>0.53</td>
<td>0.47</td>
<td>0.78</td>
</tr>
</tbody>
</table>

To study the advantage of the adaptive designs, simulations are conducted in this example to compare the performance of the adaptive designs with that of nonadaptive designs. In particular, the nonadaptive simple random sample (without replacement) with sample size equal to the adaptive design is considered as a comparison. Simulations are performed based on a snapshot of the data center temperatures collected using the mobile measurement technology (Hamann, 2008). The temperature map is provided in Figure 5 where the green areas indicate the non-experimental regions. Using the snapshot data, detailed temperature observations are available for all 166 sites and therefore the comparison can be performed based on different choices of designs. The simulations consist of 3000 iterations and the temperature in each site is generated independently from normal distribution with the snapshot observation as the mean value and variance 0.1. The estimated mean temperatures and estimated variances are summarized in Table 3, where $\hat{\mu}$ and $\hat{\mu}^*$ are calculated based on the adaptive designs with 24-run PLHDs as initial designs, and $\hat{\mu}_{SRS}$ represents the estimator based on simple random sampling with sample size equal to the adaptive designs. As shown in the table, the improved unbiased estimator $\hat{\mu}^*$ based on adaptive designs has about 40% ($= (0.78 - 0.47)/0.78$) variance reduction compared to $\hat{\mu}_{SRS}$. Moreover, it provides about 11% ($= (0.53 - 0.47)/0.53$) variance reduction compared to the original unbiased estimator $\hat{\mu}$. The average final sample size is 28.73 following the adaptive procedure.

6. SUMMARY AND CONCLUDING REMARKS

Adaptive sampling has been shown to be a useful sampling method; however, it is generally developed based on simple random sampling or stratified sampling where experimental
regions are assumed to be rectangular. This assumption can be violated in practice and therefore the conventional methods may not be suitable. A new class of designs, adaptive probability-based Latin hypercube designs and adaptive balanced probability-based Latin hypercube designs, is introduced in this paper to accommodate a specific type of irregular regions, called slid-rectangular regions.

Adaptive probability-based Latin hypercube designs refer to designs in which initial samples are based on probability-based Latin hypercube designs, a new class of space-filling designs for slid-rectangular regions. As the adaptive procedure introduces biases into the conventional estimators, new estimators are proposed which are design-unbiased for the population mean with adaptive probability-based Latin hypercube designs. Furthermore, improvements are achieved by introducing a modified unbiased estimator that is easy-to-compute and with smaller variation. Small examples are used to illustrate the detail calculation and compare the performance of the unbiased estimators. The proposed adaptive procedure is successfully applied in a data center thermal management study.

The proposed adaptive designs outperform the nonadaptive designs according to the simulation study. Further investigation is called for to explore the conditions under which adaptive sampling are more efficient than nonadaptive ones. Moreover, the proposed adaptive designs do not have control on the final sample size; therefore, modifications in this regard deserves further studies.

ACKNOWLEDGEMENTS

The authors would like to thank Y. Amemiya for helpful discussions. This research is supported by NSF grants DMS 0905753 and CMMI 0927572.

APPENDIX A: DERIVATION OF (3) AND (4)

Define \( w_{ij} \) as an indicator variable taking 1 if unit \((l, j)\) is selected in the sample and 0 otherwise. Therefore, we have

\[
P(n_k > 0) = 1 - P(\cap_{l,j \in \Psi_k} w_{ij} = 0) = 1 - \prod_{l \in \psi_k} P(\cap_{j \in \Psi_{kl}} w_{ij} = 0) = 1 - \prod_{l \in \psi_k} (1 - \frac{d_{kl}}{c_l}).
\]
To proof the unbiaseness of $\hat{\mu}$, we rewrite $\hat{\mu}$ as follows

$$\hat{\mu} = \frac{1}{N} \sum_{k}^{K} \left( \sum_{j \in \Psi_k} y_j \right) I(n_k > 0) \frac{P(n_k > 0)}{P(n_k > 0)} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i w_i}{P(w_i = 1)}$$

and thus, the result follows.

**APPENDIX B: DERIVATION OF (6) AND (7)**

Let $S_0$ be the sample space for the initial design, thus

$$\hat{\mu}^{RB} = E(\mu|M = m) = \sum_{s'_0 \in S} \hat{\mu} P(S_0 = s'_0|M = m)$$

and the conditional probability can be written as

$$P(S_0 = s'_0|M = m) = \frac{I(g(s'_0) = m)}{\sum_{s'_0 \in S} I(g(s'_0) = m)}.$$

We first decompose $\hat{\mu}$ into two parts. The first part excludes the sample edge units and the second part includes the sample edge units. Recall that $e_k = 1$ if the initial sample $k$ is an edge point and $e^*_k = \sum_{i \in \Psi_k} e_i$. Thus, we have

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{K} \frac{y^*_k J_k (1-e^*_k)}{P(n_k > 0)} + \frac{1}{N} \sum_{k=1}^{K} \frac{y^*_k J_k e_k}{P(n_k > 0)} = \frac{1}{N} \sum_{k=1}^{K} \frac{y^*_k J_k (1-e^*_k)}{P(n_k > 0)} + \frac{1}{N} \sum_{i \in s_0, e_i = 1} \frac{y_i}{c_i}.$$  \hfill (11)

The first term on the right hand side of (11) is fixed given $M = m$. For the second term, $J_k e_k = 1$ implies that the network size is 1 (i.e. $n_k = 1$) and thus $P(n_k > 0) = P(n_k = 1) = c_k^{-1}$. Therefore, we have

$$\hat{\mu}^{RB} = \sum_{s'_0 \in S} \hat{\mu} \frac{I(g(s'_0) = m)}{\sum_{s'_0 \in S} I(g(s'_0) = m)} = \frac{1}{N} \sum_{k=1}^{K} \frac{y^*_k J_k (1-e^*_k)}{P(n_k > 0)} + \frac{1}{N} \sum_{s'_0 \in S} I(g(s'_0) = m) \left\{ \sum_{i \in s_0, e_i = 1} y_i c_i \right\},$$

where $L = \sum_{s'_0 \in S} I(g(s'_0) = m)$.

The variance of $\hat{\mu}^{RB}$ can be calculated by $\text{var}(\hat{\mu}^{RB}) = \text{var}(\hat{\mu}) - E[(\hat{\mu} - \hat{\mu}^{RB})^2]$ and the detail derivation is omitted because it is similar to that in (10) below.
Appendix C: Derivation of (9) and (10)

Define $S_0$ as the set of initial design.

$$\hat{\mu}^* = E(\mu|M^* = m^*) = \sum_{s_0' \in S} \hat{\mu}P(S_0 = s_0'|M^* = m^*)$$  \hspace{1cm} (12)$$

Because the conditional probability can be calculated by

$$P(S_0 = s_0'|M^* = m^*) = \frac{I(g(s_0') = m^*) \prod_{i} c_t^{-1}}{\sum_{s_0' \in S} I(g(s_0') = m^*) \prod_{i} c_t^{-1}} = \frac{I(g(s_0') = m^*)}{\sum_{s_0' \in S} I(g(s_0') = m^*)}.$$  

Similarly, we decompose $\hat{\mu}$ into two parts. But note that the second part for the sample edge units now is written as a summation of $V$ terms. Thus, we have

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k I(n_k > 0)(1 - e_k)}{P(n_k > 0)} + \frac{1}{N} \sum_{k=1}^{K} \frac{y_k I(n_k > 0)e_k}{P(n_k > 0)} t_i(k)$$  

$$= \frac{1}{N} \sum_{k=1}^{K} \frac{y_k I(n_k > 0)(1 - e_k)}{P(n_k > 0)} + \frac{1}{N} \sum_{k=1}^{K} \frac{y_k I(n_k > 0)e_k}{c_t} t_i(k)$$  \hspace{1cm} (13)$$

Based on (12) and (13), it follows that

$$\hat{\mu}^* = \sum_{s_0' \in S} \hat{\mu} I(g(s_0') = m^*)$$  

$$= \frac{1}{N} \sum_{k=1}^{K} \frac{y_k I(n_k > 0)(1 - e_k)}{P(n_k > 0)} + (NL)^{-1} \sum_{s_0' \in S} I(g(s_0') = m^*) \left( \sum_{i \in V} \left[ \frac{y_i t_i(i)}{c_t} \right] \right)$$  

$$= \frac{1}{N} \sum_{k=1}^{K} \frac{y_k I(n_k > 0)(1 - e_k)}{P(n_k > 0)} + (NL)^{-1} \sum_{i \in V} \left( \sum_{s_0' \in S} I(g(s_0') = m^*) \left[ \frac{y_i t_i(i)}{c_t} \right] \right).$$

Because each edge unit with the same $x_1$ level (i.e., in the same column) has an equal probability of being selected in the initial design and there is only one being selected in the initial design by the definition of PLHD, we have

$$\sum_{s_0' \in S} I(g(s_0') = m^*) \left[ \frac{y_i t_i(i)}{c_t} \right] = \frac{1}{\bar{e}_{x_1}} \sum_{i \in s_0, e_1} \frac{y_i t_i(i)}{c_t}$$  

$$= \frac{L}{e_{x_1}} \sum_{i \in s_0, e_1} \frac{y_i t_i(i)}{c_t}$$  

$$= \frac{L \sum_{i \in s_0, e_1} y_i t_i(i)}{e_{x_1} c_t},$$

where $e_{x_1}$ is the total number of edge points with level $l$ in $x_1$ in the sample. Therefore (9) holds.
For the variance, we have

\[
\text{var}(\hat{\mu}^*) = \text{var}(\hat{\mu}) - E[(\hat{\mu} - \hat{\mu}^*)^2]
\]

\[
= \text{var}(\hat{\mu}) - \sum_{m^* \in M^*} \frac{P(m^*)}{L} \sum_{s'_0 \in S} I(g(s'_0) = m^*) (\hat{\mu} - \hat{\mu}^*)^2
\]

\[
= \text{var}(\hat{\mu}) - \sum_{m^* \in M^*} \frac{P(m^*)}{L} \sum_{s'_0 \in S} \left\{ I(g(s'_0) = m^*) \right\}
\]

\[
\sum_1^n \frac{c}{N} \left( \sum_{i \in s'_0, e_i = 1} y_i t_l(i) - \frac{1}{c_l} \sum_{i \in s} e_i y_i t_l(i) \right)^2
\}\}

Therefore, (10) holds.

REFERENCES


