ADAPTIVE SAMPLING WITH APPLICATION IN ENVIRONMENTAL STUDIES AND COMPUTER EXPERIMENTS

BY HUIJUAN LI

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics

Written under the direction of

Dr. Ying Hung

and approved by

New Brunswick, New Jersey

May, 2015

ABSTRACT OF THE DISSERTATION

Adaptive Sampling With Application in Environmental Studies and Computer Experiments

by Huijuan Li

Dissertation Director: Dr. Ying Hung

Adaptive sampling, which select samples sequentially, is known to be more efficient than traditional non-adaptive sampling and fixed design procedures. However, most of the methods are developed based on relatively small and well-defined regions. These assumptions are often violated in environmental studies we faced today because they invariably involve populations distributed over a large space with irregular sampling frame. A new sampling plan is proposed which enhances the estimation efficiency by taking into account the shape of the sampling region and incorporating a novel adaptive procedure. Unbiased estimators, an optimal sampling criterion, and a heuristic search algorithm is introduced. Applications to real examples are presented, which show remarkable improvement in estimation efficiency using the proposed plan over existing methods.

Unlike environmental studies, design of computer experiments has been widely investigated, however, most of the designs are chosen in advance without utilizing any information from the response, which results in insufficient information. We introduce a new class of sequential designs for computer experiments. It is model-free and constructed based on space-filling designs. The construction procedure, design-unbiased estimators, and some improvements using Rao-Blackwellization are proposed. More importantly, we introduce a refinement that provides better control over sample size and avoids replicates in the final sample. We demonstrate this new class of sequential designs are sampling-wise efficient by a simulation study and a IBM data center thermal management example.

Acknowledgements

I would like to express my great thanks to my advisor, Dr. Ying Hung, for her extraordinary guidance, ideas and encouragement. Without her supervision and support, the completion of this dissertation would not have been possible. Dr. Ying Hung is also the best role model in my life. I wish I could list all of her virtues and the spirits that I am learning from her or trying to, with my poor writing skills but a truly grateful heart.

I am greatly indebted for the loving faculty and staff members in the Rutgers Statistics Department for the continuous support throughout my PhD study.

Finally, I would like to thank my parents, Jifang Wu and Changan Li, for their tremendous love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Adaptive Probability-based Sampling for Environmental Studies

1.1 Introduction

The objective of this paper is to construct an efficient and flexible sampling plan that adaptively conduct samples in environmental studies. It is well-known that, for a given sample size and cost, more valuable information can be obtained by incorporating information from previous observations (Chernoff 1959; Cochran 1977; Ghosh and Sen 1991; Dryver 2005; Gramacy 2009; Salehi 1997; Sacks 1989; Santner 2003; Smith 1995; Fedorov 1972). Based on this idea, a sampling approach called adaptive cluster sampling is widely used in survey sampling literature. The idea is to first construct an initial design, if the selected samples satisfy some prespecified condition, follow-up samples are collected adaptively from their neighborhood. Adaptive cluster sampling is attractive not only because of its sequential property but also because it takes into account the neighborhood information which is an important feature in environmental studies. For example, in ecologial studies, it is common that the subjects of interest usually have spatially clustered patterns. If we observe one winter waterfowl in a place, the probability of having more winter waterfowls in the nearby sites becomes higher. Other examples include global warming, habitat alteration, and water pollution, in which subjects in the neighborhood of each other often share similar properties. As a result, the sampling efficiency can be enhanced by adaptively incorporating such a neighborhood information.

Although adaptive sampling is desirable, most of them are constructed based on relatively small and well-defined regions (Thompson 1991, 1992). These assumptions are often violated in environmental studies because the problems usually involve population distributed over space with irregular sampling frame in practice (Stehman and Overton 1994; Stevens and Olsen 2004; Roesch 1993). For example in the study of upper illinois basin stream networks (Stevens and Olsen 1995), not only the basin's shape, but also layouts of stream networks, are highly irregular (Figure 1.1). Given the irregularity of sampling regions, direct applications of the conventional approaches often result in the loss of efficiency and desirable properties (Stevens and Olsen

2004).

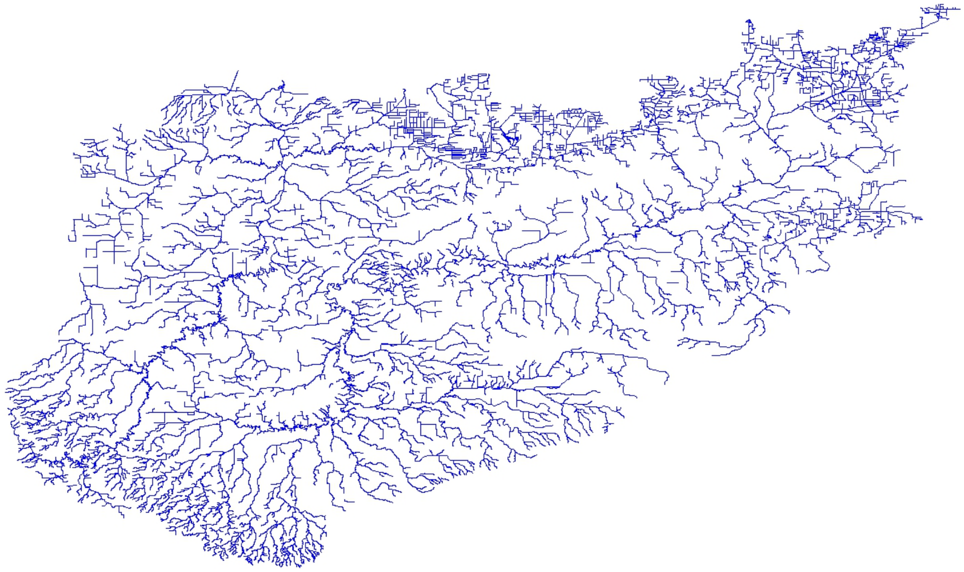

Figure 1.1: Upper Illinois Basin Stream Network, Indiana

A new adaptive sampling plan, called adaptive probability-based sampling (APS), is introduced to take into account the irregularity of experimental regions in environmental sampling. APS consists of two important components, an efficient initial design which spreads out sampling points uniformly over irrgular regions and a novel adaptive sampling plan which collects samples sequentially based on neighborhood information. APS is flexible and easy to implement. Unbiased estimators, optimal criteria, and a search algorithm is introduced. The merits of the proposed approach are demonstrated in two ecologic sampling studies.

The remainder of this paper is organized as follows. The construction procedure of APS and its unbiased estimators are introduced in Section 2. In Section 3, an optimal sampling criterion is proposed and an efficient heuristic algorithm is introduced for the search of optimal sampling plan. The proposed methods are illustrated by two real examples in environmental studies in Section 4. Conclusions are given in Section 5.

1.2 Adaptive Probability-based Sampling

1.2.1 Initial sample

Simple random sampling (SRS) is a widely used method to construct initial samples in adaptive sampling. Such an initial sample is easy to obtain, but it is well known that the efficiency can be improved by space-filling designs such as Latin hypercube sampling (LHS) (McKay et al. 1979). With the space-filling properties, such as the one-dimensional balance property in which the sampling points are evenly spread out in each dimension, it can be shown that LHS provides improvements over SRS (Iman 2008; Fang 2002; McKay 1979). However, when the sampling frame is in an irregular shape, LHS can no longer maintain the space-filling property (Hung et al. 2010). There are approaches developed to handle the irregularity of the sampling regions (Draguljić et al. 2012); however, most of them are chosen in advance (i.e., non-adaptive) which can be inefficient, especially when the spatially clustered patterns exist.

An important component of APS is a new initial sampling plan called probability-based Latin hypercube sampling (PLHS). It is inspired by the idea of space-filling designs and extended from the adaptive probability-based Latin hypercube designs (Hung 2011), which is developed for a particular type of irregular regions in engineering applications. The idea is to maintain the desirable space-filling properties in the irregular sampling regions so that the initial samples can be spread out uniformly regardless of the shape. A n-run PLHS can be described as follows.

Step 1: Specify two coordinates such as its latitude and longitude and define them as x_1 and x_2. Without loss of generality, assume that the length of x_1 is larger than x_2. Denote the n samples by (x_{1i}, x_{2i}), where $i = 1, ..., n$.

Step 2: Define k levels for x_2. For the jth level of x_2, the sampling region of x_1 is denoted by $E_j = (A_j, B_j)$ and therefore, the range of x_1 are located irregularly on the interval $[A, B]$, where $A = \min\{A_j\}$ and $B = \max\{B_j\}$.

Step 3: Divide the interval $[A, B]$ into n equally spaced subintervals and assign the n levels of x_1 at the middle of these subintervals. That is, $x_{1i} = i$, for all i. For the ith level of x_1, the corresponding level of x_2, denoted by x_{2i}, is assigned with probability

$$\mathrm{pr}(x_{2i} = j) = \begin{cases} [\sum_{j=1}^{k} I(j \in C_i)]^{-1}, & \text{if } j \in C_i, \\ 0, & \text{otherwise,} \end{cases} \tag{1.1}$$

where C_i is the range of x_2 when $x_{1i} = i$.

A simple example of a 10-run PLHS is illustrated in Figure 1.2. The grey areas, including the dark grey and light grey areas, form the irrigular sampling frame. The experimental ranges are $C_1 = \{4\}$, $C_2 = \{3,4\}$, $C_3 = \{1,3,4,5\}$, etc. The level of x_2 is assigned by (3.5). For example, $pr(x_{21} = 4) = 1$, $pr(x_{22} = 3) = pr(x_{22} = 4) = 1/2$, and $pr(x_{23} = 1) = pr(x_{23} = 3) = pr(x_{23} = 4) = pr(x_{23} = 5) = 1/4$. The circles are the initial sample of PLHS. It is clear that PLHS has one initial sample for each level of x_1; therefore, the one-dimensional balance property is maintained.

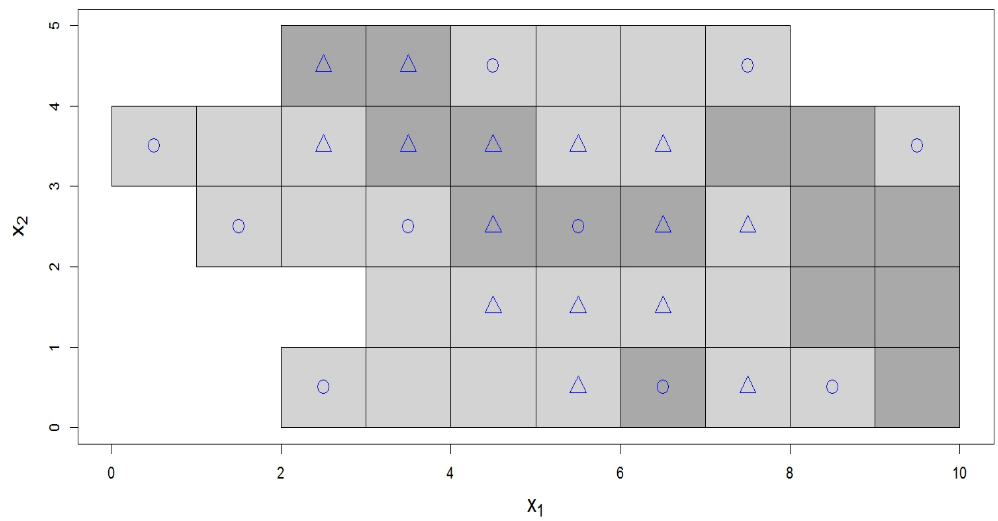

Figure 1.2: Blue-winged teal population example

1.2.2 Adaptive sampling procedure

Based on the n-run PLHS as an initial, APS introduces an adaptive sampling procedure with the following idea. According to a pre-specified condition of interest, additional neighborhood samples are collected whenever the response of a selected sample satisfies the condition. If any of these subsequently added sample satisfies the condition, then the units of its neighborhood are also added to the sample, so that finally the sample contains every unit in the neighborhood of any sample unit satisfying the condition. The neighborhood of each unit consists of, in addition to itself, the spatially adjacent units in the experimental region.

The population maybe partitioned into K sets of units, termed networks, such that selection in the initial sample of any unit in a network will result in inclusion in the final sample of all units in that network. A unit not satisfying the condition belongs to a network consisting just of itself. Any unit not satisfying the condition but in the neighborhood of one that does is called edge unit.

A simple example of APS is illustrated in Figure 1.2. Different colors in the figure represent the true response if they are selected. Dark grey indicates that the true response satisfies the prespecified criterion and light grey indicates that the response does not satisfy the criterion. Based on the 10-run PLHS as initial, the adaptively added points are denoted by triangles. The final sample includes both circles and triangles.

1.2.3 Estimators

For APS, conventional unbiased estimators, such as the sample mean, are no longer unbiased. A Horvitz-Thompson type of design-unbiased estimator (Cochran, 1977; Horvitz and Thompson, 1952; Thompson, 1990, 1991) is recommended here which is analogue to the one introduced by Hung (2011). We focus on the unbiased estimators for the population mean defined by $\mu = N^{-1}\sum_{i=1}^{N} y_i$, where N is the number of units in the population and $y_1, \cdots, y_N$ are the corresponding responses. Let Ψ_k be the set of units in the kth network. Because of the non-regular shape of the experimental region, further partition the units in Ψ_k by their coordinates in the x_1-axis. Let Ψ_k ranges in $\psi_k = (t_1, \ldots, t_k) \in (1, \ldots, n)$ in the x_1-axis. For each observation in ψ_k, the associated units in the network are Ψ_{kl}, where $l \in \psi_k$, and the number of units in Ψ_{kl} is denoted by d_{kl}. Using the foregoing notation, the estimator can be written as

$$\hat{\mu} = \frac{1}{N}\sum_{k}^{K} \frac{y_k^* I(n_k > 0)}{P(n_k > 0)}, \tag{1.2}$$

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 kth network is in the initial sample s_0, and takes 0 otherwise. The number of units selected from the kth network in the initial sample is $n_k = \sum_{i \in \Psi_k} I(i \in s_0)$, and the inclusion probability of network k is

$$P(n_k > 0) = 1 - \frac{\prod_{l \in \psi_k}(c_l - d_{kl})}{\prod_{l \in \psi_k} c_l}. \tag{1.3}$$

The variance of (2.1) can be calculated by

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

where

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

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

$$v\hat{a}r(\hat{\mu}) = N^{-2} \sum_{k=1}^{K} \sum_{h=1}^{K} \frac{y_k^* y_h^* [P(n_k > 0, n_h > 0) - P(n_k > 0)P(n_h > 0)]}{P(n_k > 0)P(n_h > 0)P(n_k > 0, n_h > 0)} I(n_k > 0)I(n_h > 0). \tag{1.4}$$

Based on the Rao-Blackwell theorem, the unbiased estimator can be further improved by reducing its variance. The idea is to calculate conditional expectation of the original estimator, given a sufficient statistics. The most efficient choice is the minimal sufficient statistics but it is computationally intensive which is often experienced in the conventional adaptive sampling (Dryver and Thompson, 2005). Therefore, with a similar argument in Hung (2011), an improved unbiased estimator $\hat{\mu}^*$ can be constructed by conditioning on a carefully chosen sufficient statistics but not the minimum so that the computation can be simplfied.

Let s denotes 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 $s_{\bar{c}}$. 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 $m^* = \{(i, y_i, f_i), V, (j, y_j) : i \in s_c, j \in s_{\bar{c}}\}$, and the sample space for m^* is defined by M^*. Hence, the improved unbiased estimator is obtained by conditioning on m^* as

$$\begin{aligned} \hat{\mu}^* = \;& E(\mu | M^* = m^*) \\ = \;& \tfrac{1}{N} \textstyle\sum_{k=1}^{K} \frac{y_k^* I(n_k>0)\left(1-e_k^*\right)}{P(n_k>0)} + \tfrac{1}{N} \sum_{l \in V} \frac{\sum_{i\in s} e_i y_i t_l(i)}{e_{s_l} c_l^{-1}}, \end{aligned} \tag{1.5}$$

where $t_l(i)$ is an indicator variable taking 1 if the unit i belongs to level l in factor x_1 and 0 otherwise and $e_{s_l} = \sum_{i \in s} e_i t_l(i)$.

The variance of the improved unbiased estimator is

$$\begin{aligned} var(\hat{\mu}^*) = \;& var(\hat{\mu}) - \textstyle\sum_{m^* \in M^*} \frac{P(m^*)}{L} \sum_{s_0' \in S} \Big\{ I(g(s_0') = m^*) \\ & \Big[\textstyle\sum_{l=1}^{n} \frac{c_l}{N} \Big(\sum_{i \in s_0', e_i = 1} y_i t_l(i) - \frac{1}{e_{s_l}} \sum_{i \in s} e_i y_i t_l(i) \Big) \Big]^2 \Big\}, \end{aligned} \tag{1.6}$$

where L is the number of all initial samples that can lead to the same final sample.

An efficient unbiased estimator of the variance can be obtained by

$$\begin{aligned} v\hat{a}r(\hat{\mu}^*) = \;& E[v\tilde{a}r(\hat{\mu}^*) \mid M^* = m^*] \\ = \;& \tfrac{1}{L} \textstyle\sum_{s_0' \in S} I(g(s_0') = m^*) v\hat{a}r(\hat{\mu}(s_0')) - \frac{1}{L} \sum_{s_0' \in S} \Big\{ I(g(s_0') = m^*) \\ & \Big[\textstyle\sum_{l=1}^{n} \frac{c_l}{N} \Big(\sum_{i \in s_0', e_i = 1} y_i t_l(i) - \frac{1}{e_{s_l}} \sum_{i \in s} e_i y_i t_l(i) \Big) \Big]^2 \Big\}. \end{aligned} \tag{1.7}$$

1.3 Optimal initial design and search algorithm

Samples constructed by PLHS are not unique and they are not equally good. Although they have the one-dimensional balance property on x_1, they can often be highly structured. For example, in Figure 1.3, PLHS can generate an initial sample with all the points located in the same row, i.e., $x_{2i} = 4$ for all i, which is not desirable because it is not spread out as uniform as it should be on x_2. Therefore, an optimal criterion is proposed and a search algorithm is introduced in this section to obtain an initial PLHS that has better space filling properties.

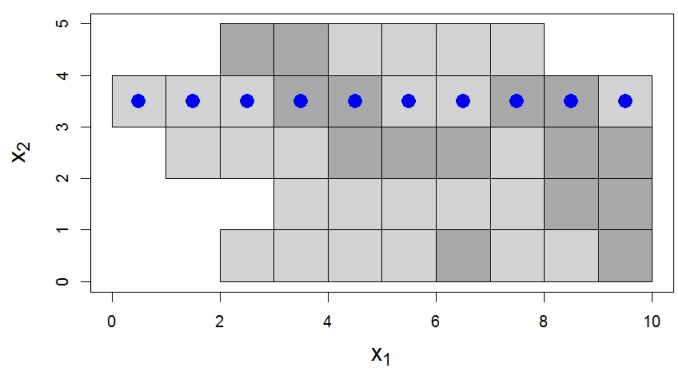

Figure 1.3: An initial sample

In order to spread out design points over the experimental region, many design criteria are proposed for space-filling designs (Iman and Conover, 1982; Johnson et al., 1990; Owen, 1994; Tang, 1998; Ye et al., 2000; Joseph and Hung, 2008). Here we focus on a maximin distance criterion proposed by Morris et al.(1995) and other criteria can be easily adopted to the proposed framework. The idea is to maximize the minimum inter-site distances among samples. For any two sample points (x_{1i}, x_{2i}) and (x_{1j}, x_{2j}), we define the distance by $d((x_{1i}, x_{2i}), (x_{1j}, x_{2j})) = [(x_{1i} - x_{1j})^p + (x_{2i} - x_{2j})^p]^{1/p}$, where $p > 0$. Then a maximin PLHS is obtained by minimizing the sum of all pairs of design points $\phi_p = \{1/[\sum_{i=1}^n \sum_{j=1,j\neq i}^n d((x_{1i}, x_{2i}), (x_{1j}, x_{2j}))]\}^{1/p}$.

Because of the combinatorial nature of the problem, finding the optimal PLHS can be a difficult task. For example, there are $(\prod_{i=1}^n c_i)$ different PLHSs with sample size n. Complete search cannot always be possible, especially when the number of runs or the values of c_i are large. Instead, we proposed an efficient heuristic algorithm that can quickly identify optimal designs.

A design is called feasible if the points lie in the experimental region. In a widely used search algorithm for optimal Latin hypercube designs, Morris and Mitchell (1995) proposed a columnwise-pairwise exchange approach in generating candidate designs. This approach, however, cannot be directly applied here because arbitrary exchange of two elements within a column does not always lead to a feasible design. A naive approach is to modify the algorithm by adding an additional verification step for the first two factors, where the slid-rectangular region is defined. Instead of exchanging any two randomly selected elements, we only allow random exchanges that are capable of producing a feasible solution. Using this procedure, a sequence of feasible designs are generated and examined by the optimality criterion. Because of the combinatorial nature of the problem, finding optimal probability-based Latin hypercube designs can be computationally difficult. For example, there are $(\prod_{i=1}^{n} c_i) \times (n!)^{p-2}$ probability-based Latin hypercube designs with n runs and p factors. Complete search can be computationally prohibitive. Therefore an efficient heuristic algorithm is needed.

To improve the efficiency of the naive search, a new algorithm is introduced. The idea is to broaden search by preventing visits of neighborhood designs that have been visited before. Neighborhood designs are those that differ from each other by a small number of columnwise-pairwise exchanges. Because the neighborhood designs are similar, avoiding them would allow the search to move to other parts of the region with more promising values. To do so, the naive search is modified by keeping track of the previous q feasible settings of x_2 visited, which are called forbidden settings, with q being a tuning parameter. The forbidden setting is also called memory in the tabu search literature (Glover, 1986). Since a visit of the forbidden settings would lead to a movement toward the neighborhood designs, identification of the forbidden settings/memory can be effectively used to prevent the current design from moving toward neighborhoods. This makes the search more efficient. In fact the forbidden settings can be defined based on any factor. However, a more desirable choice is x_2 because the verification of feasible exchanges in x_2 is more time-consuming. If the x_2 candidates lie in the forbidden set, they are removed from consideration immediately without verifying their feasibility. Thus, some computation can be saved and the x_2 settings are explored more efficiently. Typically the tuning parameter q should be small in order to maintain a small neighborhood. This algorithm is called a columnwise-pairwise exchange tabu algorithm.

After specifying a design criterion, the proposed algorithm begins with a randomly chosen design X, and proceeds with the examination of a sequence of designs. Each design is generated as follows. First, a column from x_2 to x_p is randomly selected. If the selected column is one of the last $p-2$ factors, a new design is obtained by exchanging two randomly selected elements within the column. This is similar to the procedure in Morris and Mitchell (1995). If the column of x_2 is selected, one has to check whether the exchange of two randomly selected x_2 levels leads to a setting that is in the forbidden set, i.e., if it is among the last q feasible settings visited. If not, check the feasibility, i.e., if the resulting design lies in the irregular region. The exchange is allowed to proceed only if the resulting new setting is non-forbidden and feasible. Otherwise, random exchanges within the x_2 column continue to be examined until a new feasible setting is obtained. Following this procedure, new designs X^{try} are generated. In each iteration, X is replaced by X^{try} if it leads to an improvement with respect to the design criterion. Once values of the design criterion are stabilized, the algorithm is terminated and the resulting X is the optimal design. This algorithm is easy to implement and the simulated annealing approach in Morris and Mitchell (1995) can be applied to further improve the search efficiency. Optimal probability-based Latin hypercube designs can be obtained by this procedure. It also works for balanced probability-based Latin hypercube designs with carefully constructed initial designs because the proposed procedure maintains the balance property throughout the search if the initial design is balanced.

1.4 Application to Environmental Studies

1.4.1 Example 1

Example 1 focuses on the study of blue-winged teal population introduced by Smith et al.(1995).The population region is gridded into fifty 100 km^2 quadrates shown in Table 1.1.Note that, IA In Table 1.1 indicates inaccessible sites which lead to an irregular accessible region.In order to identify sites with teals, we define a quadrate that satisfies the condition if the observed number of teals y is nonzero, i.e. $y \geq 1$.This condition results in more quadrates sampled in the neighborhood of the existing sampled sites that have nonzero observations, which is desirable because it is believed that blue-winged teals are gregacious animals.So once they show up, with larger probability we can observe other teals in nearby sites.

IA	IA	3	5	0	0	0	0	IA	IA
0	0	0	24	14	0	0	10	103	0
IA	0	0	0	2	3	2	0	13639	1
IA	IA	IA	0	0	0	0	0	14	122
IA	IA	0	0	0	0	2	0	0	177

Table 1.1: Numbers of blue-winged teal in fifty 100 km^2 quadrats

To demonstrate the performance of the proposed adaptive sampling method, 100000 simulations are conducted for the adaptive probability-based LHS and simple random sampling with the same sample size. In each simulation, a randomly generated 10-run PLHS was used for as an initial design, and adaptive designs were proceeded accordingly. The estimated means and variances are summarized in Table 1.2, where $\hat{\mu}$ is calculated based on the adaptive PLHS and $\hat{\mu}_{srs}$ denotes the result based on simple random sampling. Efficiency is the ratio of variance of APS divided by that of SRS. As shown in Table 1.2, the unbiased estimator $\hat{\mu}$ had much better performances than simple random sampling estimator $\hat{\mu}_{srs}$. The average final sample size was 28.96.

	μ	μ_{srs}
Mean	353.0499	353.6494
Variance	6509.15	4652191
Efficiency	0.001399	1

Table 1.2: Blue-Winged Waterfowl APS results

This example is used to further assess the efficiency of the proposed columnwise-pairwise exchange tabu algorithm in the search of optimal adaptive sampling procedure with maximin criteria. From Morris and Mitchell (1995), we know for large enough value of the distance parameter p, the maximin criterion function can rank designs in the same way that the more complex maximin criterion searching for an optimal maximin design; and when p is set to a relatively small value, it tends to facilitate the reliability of optimization process. Here we tried different distance parameter p to see whether p has an impact on the performance of search, while under p ranges from 1 to 25, columnwise-pairwise exchange tabu search does a better job than without tabu. As an example, we use $p = 13$ to show the efficiency of the design optimization method as in Figure 1.4. It is shown significant improvement on space-filling property by optimizting top initial design to bottom design. Figure 1.5 demonstrates the advantage of the tabu search idea in this optimizaiton algorithm. In a random 10-run PLHD design example, the convergence rate of optimization with tabu(red dashed line) is much faster than without tabu(blue solid line).

iteration	1	2	3	4	5	6
	1.1226806	1.1226806	1.1011936	1.0548158	1.0317525	1.0317364
iteration	7	8	9	10	11	
	1.0317299	1.0000802	0.9482473	0.9482473	0.9482473	

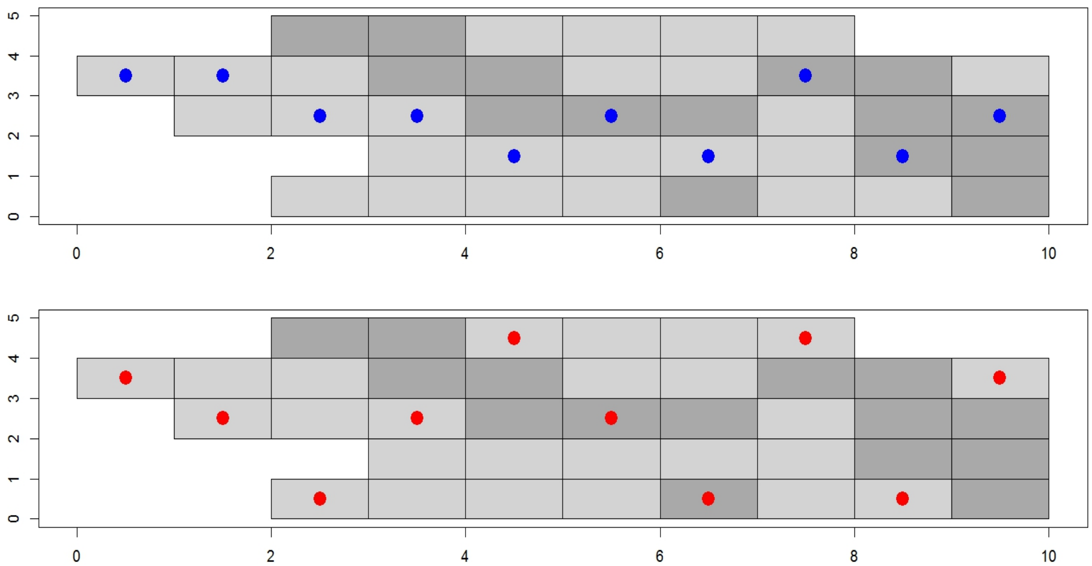

Figure 1.4: Design optimization

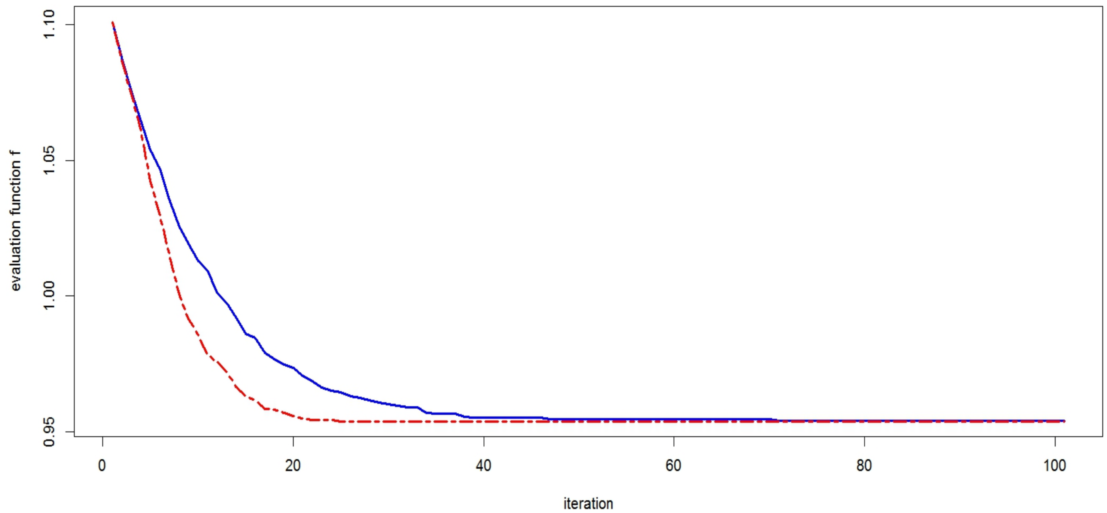

Figure 1.5: Comparison of design optimization with Tabu and without Tabu

1.4.2 Example 2

Reach File is a network-oriented, spatially referenced geographic data system developed by the U.S. environmental protection agency (EPA). Reach File describes surface-water features of the United States. It stores stream network information and facilitates water quality modeling. In this example, data are from the latest version of Reach File, RF3. Two indices of water quality standards, qualitative habitat evaluation index (QHEI) and Index of Biological Integrity (IBI) were observed in 85 stream sites in Upper Illinois Basin,Indiana (1.6). These two indices are the major water quality standard summary considered by EPA. The IBI score, which is a fish community index of biological integrity (Karr 1981) that assesses water quality using resident fish communities as a tool for monitoring the biological integrity of streams. The Qualitative Habitat Evaluation Index(QHEI) score (Rankin 1989) gives an estimate of the suitability of a stream segment to meet warm-water habitat for aquatic organisms. More details can be found in EPA aquatic resources monitoring website. During the survey of this example, there are several pre-targeted sites could not be sampled. This occur may due to (1) sites not being a member of the target population, (2) landowners deny access to a site, (3) a site is physically inaccessible, or (4) site not sampled for other reasons.Thus, the sampling frame is an irregular region.

In Upper illinois Basin stream sample area, the geographical coverage is gridded into 10×10

cells as plotted in Figure 1.6, where the x-axis and y-axis are derived from the latitude and longitude using Albers projection. Albers projection is a standard projection used by major environmental institutions, which puts a degree of longitude and a degree of latitude into a coordinate system where x and y mean same thing in terms of distance. As shown in the figure, there are 41 accessible sites for QHEI and 40 for IBI.

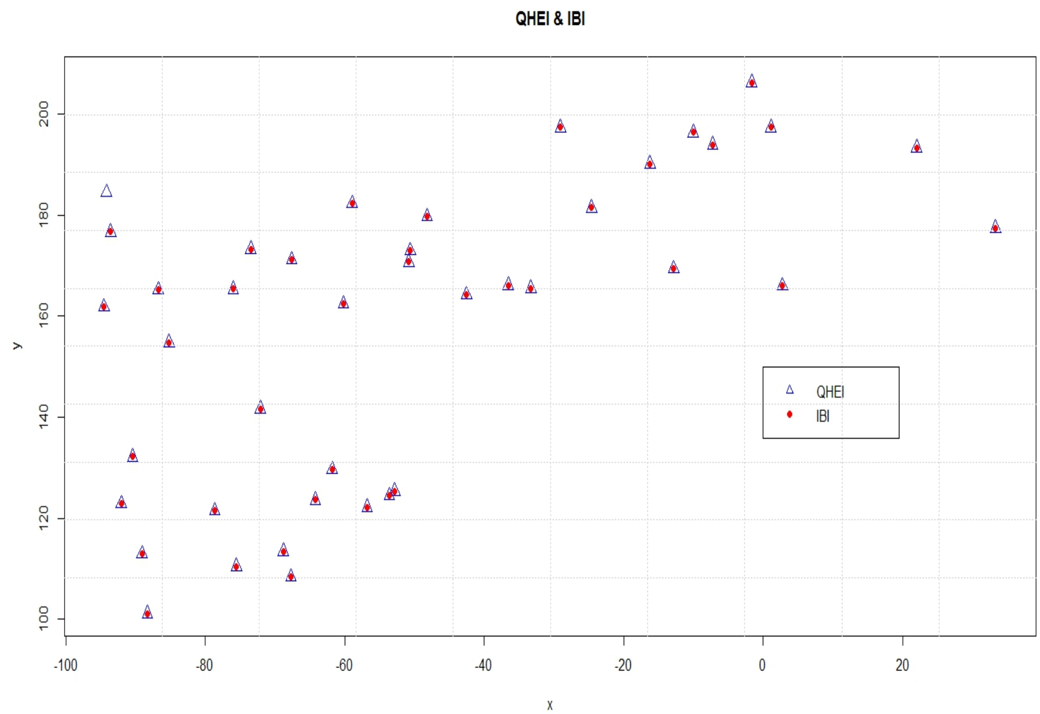

Figure 1.6: Upper Illinois Basin water quality standard observation

If QHEI score is no larger than 45, then streams are indicated as not suitable for warm-water habitat witout using impairment. Streams with IBI score lower or equal than 40 will be considered as poor water quality. Both situations imply alarming water habitat environment. Therefore, we correspondingly specify the adaptie sampling criteria as QHEI score $\leq$ 45 and IBI score $\leq$ 40. Similar to Example 1, 100000 simulations were conducted in this example to compare the performance of the adaptive PLHS with simple random sampling under same sample size. In each simulation, a randomly generated 10-run PLHD was used as initial design.

Results for QHEI and IBI are summarized in table 1.3 and table 1.4 respectively. According to the adaptive sampling procedure, the average final sample size for QHEI and IBI are 14.49 and 16.14. For QHEI observations, the improved unbiased estimator based on adaptive designs has 0.3523177 efficiency compared to simple random sampling. Moreover,it provides an approximate

	μ	μ^*	μ_{srs}
Mean	65.72352	65.89497	65.73458
Variance	431.7279	376.1401	911.0651
Efficiency	0.4738606	0.3523177	1

Table 1.3: Accessible water quality standard QHEI in Upper Illinois Basin

	μ	μ^*	μ_{srs}
Mean	47.167	46.82065	47.15551
Variance	243.9689	212.5425	508.4736
Efficiency	0.4798064	0.4180011	1

Table 1.4: Accessible water quality standard IBI in Upper Illinois Basin

55.5878 variance reduction compared with the original unbiased estimator. In IBI results, adaptive sampling based on PLHD has similar performance.

1.5 Concluding Remarks

Environmental Sampling usually occurs in irregular geographical coverage and often have clustered sampling pattern.Our proposed adaptive probability-based sampling efficiently sample on non-regular region. Through optimization with columnwise-pairwise tabu optimization, the initial one-dimensional balanced design maintains good space-filling property. And by adaptive sampling with pre-defined criteria, we observe more desired samplers. We proposed unbiased estimators for this particular methodology, and proved this method is more efficient than simple random sampling through real examples. Besides, for sampling issues other than environmental sampling, which having similar irregular coverage and clustered pattern, our APS approach could be easily extended. Hence, this approach can be implemented in multiple applications.

Sample size control is one disadvantage of this APS method. This sampling method is also highly depending on sampling framework, such as gridding size, gridding scheme, or general framework information. All of these need further exploration.

Chapter 2

Application of Adaptive Sampling in Computer Experiment

2.1 Introduction

Many complex phenomena are too costly or difficult to investigate directly using physical experiments. Instead, computer experiment becomes a widely used alternative to provide insight into such phenomena. However, computational expense of computer experiments often prohibit the naive approach of running the experiment over a dense grid of input configurations. Therefore, an efficient design is important in the study of complex phenomena using computer experiments. A popular approach is called space-filling designs (Santner et al. 2003, Fang et al. 2006), such as Latin hypercube designs (LHDs).

Despite the prevalence of space-filling designs, they are chosen in advance which results in insufficient information in parts of the space, particularly where the responses appear to be promising. It has been shown in conventional experimental designs that, for a given sample size and cost, more valuable information can be obtained by performing the experiments sequentially. However, the existing sequential designs cannot be directly extended to computer experiments because of their special features such as deterministic outputs and nonlinearity of the response surface. Studies of sequential design remain scant in computer experiments. The existing methods along this line rely heavily on model fitting (Williams et al. 2000, Gramacy and Lee 2009). The efficiency of these methods deteriorates as the fitted model deviates from the global response surface, which commonly occurs when a limited amount of data are collected with high dimensionality in complex experiments.

A new class of sequential designs is introduced for computer experiments. It is called adaptive Latin hypercube designs (ALHDs). ALHDs appear to be more efficient than nonadaptive design. Moreover, they are model-free and constructed based on space-filling designs. Thus, ALHD is attractive for complex computer experiments particularly in their early stage where model fitting tends to be unreliable.

The remainder of the paper is organized as follows. In Section 2, the design procedure of

ALHDs is introduced and three design-unbiased estimators are proposed. A further refinement of ALHD is proposed in Section 3, which provides better control over sample size and avoids replicates in the final sample. The two types of ALHDs are illustrated using a simple example in Section 4. In Section 5, the performance of the proposed designs are demonstrated based on two studies using computer experiments.

2.2 Adaptive Latin hypercube designs

2.2.1 Design procedure

To provide sequential designs which is robust to model assumptions, we incorporate the concept of adaptive sampling. Adaptive sampling is a sequential model-free procedure developed mainly for two-dimensional survey sampling. It is generally used in animal population sampling and shown to be efficient in sampling sparse but clustered populations such as rare species. For the design of computer experiments, the direct extension of conventional adaptive sampling is not desirable because it is based on a simple random sampling as an initial which is not ideal for spreading out design points (Thompson 1990,1991,1992; Salehi 1997; Seber and Thompson 1994; Smith 1995), particually for high dimensional space. It is shown that Latin hypercube designs can improve estimation efficiency over simple random sample, therefore, an intuitive idea is to conduct the adaptive design based on Latin hypercube designs. Such a design is called an adaptive Latin hypercube design.

The procedure to construct adaptive Latin hypercube designs can be described as follows. Assuming that there are p factors and each factor has n levels. An initial sample is collected based on Latin hypercube sampling over the experimental region. In general, a LHD can be easily obtained by a random permutation of $(1, \ldots, n)$ for each factor and optimal design criteria (Morris 1995; Johnson 1990; Hung 2008; Tsay 1976; Constantine 1981; Murty 1971; Ouwens, Tan and Berger 2002) can be applied. Given a fixed Latin hypercube design as an initial, whenever the response of a selected initial 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, then more units may be added to the sample. This procedure continues until no more units that meet the condition are found. Thus the final design contains every unit in the neighborhood of any sample unit satisfying the condition. In this paper, we define a neighborhood which consists of, in addition to itself, the spatially adjacent points. Discussions on various neighborhood definitions can be found in Thompson (1990) and they can be extended to adaptive Latin hypercube designs.

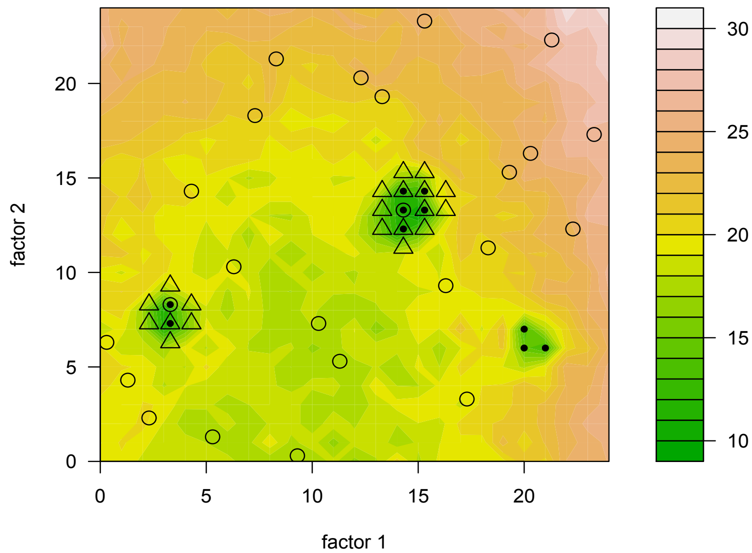

Figure 2.1: A data center thermal management example

An example of a two-factor adaptive Latin hypercube design is illustrated in Figure 2.1. Assume that two factors, the flow rate of an air condition and the percentage of tile open area, control the thermal distribution in a data center and the interest is to maintain the data center at an acceptable temperature for reliable operation of the equipment. Different colors represent the underlying maximum data center room temperature with respect to different settings of the two factors. The initial design is a 24-run Latin hypercube design represented by the circles. The bullet points, not necessarily selected by the initial design, indicate that the associated variable of interest satisfy the prespecified condition, i.e., the corresponding maximum room temparture is below 15°C. The triangles represent the design points selected adaptively to the final sample. So, the final adaptive Latin hypercube design includes the circles and the triangles.

For the adaptive Latin hypercube designs, 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 five adjacent bullet points in the middle of Figure 2.1 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 units that do not meet the condition but in the neighborhood of some design points that satisfy the condition are called edge units. The neighborhood of a sample unit is assumed to be independent to the response.

Apart from the space-filling property and the efficiency gained by the sequential procedure, the proposed designs are particularly attractive for computer experiments because they provide information that fulfills the needs in computer experiment modeling at a later stage, including the estimation of the mean function and the covariance structure (Sacks et al. 1989). Specifically, spreading out initial design points uniformly is useful in efficiently estimating the regression parameters in the mean function. On the other hand, adaptively taking into account the local information in the follow-up experiment is desirable for capturing the smoothness of the underlying system, which leads to better estimation of parameters in the covariance structure, such as the smoothness parameters in Matern class (Matern 1986; Handcock and Stein 1993; Handcock 1994; Palacios and Steel 2006; Zhang 2004). It is important typically for complex experiments with large number of inputs because the initial design points are very sparse in the high dimensional cases (Handcock 2004).

2.2.2 Design properties

Despite some commonly shared concepts, the adaptive Latin hypercube designs differ from the conventional adaptive sampling in which initial samples, such as simple random sampling, are drawn independently. The independence of the inclusion probabilities is violated in adaptive Latin hypercube designs. Therefore, standard estimation methods for adaptive sampling cannot be applied and new estimators and the related inference 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 the adaptive Latin hypercube designs. Hence, new unbiased estimators are introduced and their variances and unbiased estimators for variances are discussed. Derivations of the unbiased estimators and the unbiased estimators of variance are given in the Appendix.

Assume that there are $N = n^p$ units in the population with the corresponding responses $y_1, \cdots, y_N$ and the population mean is defined by $\mu = N^{-1} \sum_{i=1}^{N} y_i$. For the adaptive Latin hypercube designs, a Horvitz-Thompson type of design-unbiased estimator (Cochran, 1977; Horvitz and Thompson, 1952; Thompson 1990,1991) is introduced. To do so, the following notation is needed. Let Ψ_k be the set of units in the kth network and K be the number of networks in the population. So the number of units selected from the kth network in the initial sample can be written as $n_k = \sum_{j \in \Psi_k} I(j \in s_0)$. The unbiased estimator is developed based on Theorem 1 as follows.

Theorem 1: Define the size of network k by ψ_k. The number of designs having at least j elements from the kth netwrok in their initial Latin hypercube design is defined by $q_{k,j}$, where $j = 1, \cdots, \psi_k$ and $q_{k,1} = \psi_k$. A Horvitz-Thompson type of design-unbiased estimator can be written as

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{K}\frac{y_k^* I(n_k > 0)}{P(n_k > 0)}, \tag{2.1}$$

where $y_k^* = \sum_{i \in \Psi_k} y_i$, the inclusion probability of network k in the inital sample can be written as

$$P(n_k > 0) = \sum_{j=1}^{\min(\psi_k, n)} (-1)^{j+1} q_{k,j} \left[\prod_{l=0}^{j-1}(n-l)\right]^{1-p},$$

and $I(n_k > 0)$ takes 1 if any unit of the kth network is in the initial sample s_0, and 0 otherwise.

The variance of the unbiased estimator can be calculated by

$$var(\hat{\mu}) = N^{-2}\sum_{k=1}^{K}\sum_{h=1}^{K}\frac{y_k^* y_h^*[P(n_k > 0, n_h > 0) - P(n_k > 0)P(n_h > 0)]}{P(n_k > 0)P(n_h > 0)}, \tag{2.2}$$

where

$$P(n_k > 0, n_h > 0) = \sum_{j=2}^{\min(\psi_k \cup \psi_h, n)} (-1)^j \left\{\sum_{i=1}^{j-1}[q_{(k,i),(h,j-i)}]\right\}\left[\prod_{l=0}^{j-1}(n-l)\right]^{1-p},$$

and $q_{(k,i),(h,t)}$ is the number of designs having at least i elements from the kth network and at least t elements from the hth network in their initial Latin hypercube design, for $i = 1, \cdots, \psi_k$ and $t = 1, \cdots, \psi_h$. An unbiased estimator of (2.2) is

$$\widehat{var}(\hat{\mu}) = N^{-2}\sum_{k=1}^{K}\sum_{h=1}^{K}\frac{y_k^* y_h^*[P(n_k > 0, n_h > 0) - P(n_k > 0)P(n_h > 0)]}{P(n_k > 0)P(n_h > 0)P(n_k > 0, n_h > 0)} I(n_k > 0)I(n_h > 0). \tag{2.3}$$

A special case when the experimental region consists of only networks with size one, we have the following Lemma.

Lemma 1: Define $w_{ij} = 1$ if y_i and y_j have no coordinates in common and 0 otherwise. If all the networks have size one, i.e., $\psi_k = 1$ for all $k = 1, \cdots, K$, the variance in (2.2) can be written as

$$var(\hat{\mu}) = \frac{1}{n^2}\left\{\sum_{i=1}^{n^p} n^{1-p}\big(1 - n^{1-p}\big)y_i^2 + \sum_{i=1}^{n^p}\sum_{j=1}^{n^p} y_i y_j w_{ij} n^{1-p}\big[(n-1)^{1-p} - n^{1-p}\big]\right\},$$

which equals to the variance of a Latin hypercube design.

The efficiency of the foregoing unbiased estimator can be improved by incorporating more information of the edge points because the observations of edges points are used in the estimator only if they appear in the initial sample based on (2.1). The improvement is achieved by Rao-Blackwellization in which an improved unbiased estimator is obtained by the conditional expectation of the original estimator, given a sufficient statistics. The most efficient choice is the minimal sufficient statistics. For adaptive Latin hypercube designs, the minimum sufficient statistics is the unordered set of distinct, labeled observations, denoted by $m = \{(i, y_i) : i \in s\}$, where s denotes the final sample. 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. An improved unbiased estimator for adaptive LHD is

$$\begin{aligned} \hat{\mu}^{\mathrm{RB}} = & \quad E(\mu|M = m) \\ = & \quad \tfrac{1}{N} \textstyle\sum_{k=1}^{K} \frac{y_k^* J_k(1-e_k^*)}{P(n_k>0)} + \frac{1}{nL} \sum_{s_0' \in S} \big\{ I(g(s_0') = m) \big[\sum_{i \in s_0', e_i = 1} y_i \big] \big\}, \end{aligned} \tag{2.4}$$

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

$$var(\hat{\mu}^{\mathrm{RB}}) = var(\hat{\mu}) - \sum_{m \in M} \frac{P(m)}{L} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = m) \big[(\hat{\mu} - \hat{\mu}^{\mathrm{RB}})^2 \big], \tag{2.5}$$

where L is the number of initial designs that are *compatible* with m and $P(m)$ is the probability that $M = m$. An unbiased estimator of the variance is $\widetilde{var}(\hat{\mu}^{\mathrm{RB}}) = \widehat{var}(\hat{\mu}) - L^{-1} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = m) \big[(\hat{\mu} - \hat{\mu}^{\mathrm{RB}})^2 \big]$ and a more efficient estimator can be further obtained by conditioning on the minimum sufficient statistics as follows

$$\begin{aligned} \widehat{var}(\hat{\mu}^{\mathrm{RB}}) = & \quad E(\widetilde{var}(\hat{\mu}^{\mathrm{RB}}) \mid M = m) \\ = & \quad \tfrac{1}{L} \textstyle\sum_{\tilde{s}_0 = S} I(g(\tilde{s}_0) = m) \widehat{var}(\hat{\mu}) - \frac{1}{L} \sum_{\tilde{s}_0 \in S} I(g(\tilde{s}_0) = m) \big[(\hat{\mu} - \hat{\mu}^{\mathrm{RB}})^2 \big]. \end{aligned} \tag{2.6}$$

New estimator

Define a sufficient statistic but not the minimum. Decompose the final sample s into two parts. The first part, denoted by s_c, is the set of all distinct units in the sample for which the condition to sample adaptively is satisfied. The remaining units are denoted by $s_{\bar{c}}$. Let f_i be the number of times that the network to which point i belongs is intersected by the initial sample. Using these notation, a sufficient statistic can be defined by $m^* = \{(i, y_i, f_i), (j, y_j) : i \in s_c, j \in s_{\bar{c}}\}$.

$$\hat{\mu}^* = \tfrac{1}{N} \textstyle\sum_{k=1}^{K} \frac{y_k^* I(n_k>0)(1-e_k^*)}{P(n_k>0)} + \frac{e_{s_0}}{n e_s} \left(\sum_{i \in s} e_i t_i y_i \right) \tag{2.7}$$

The variance is

$$\begin{aligned} var(\hat{\mu}^*) = & \quad var(\hat{\mu}) - \textstyle\sum_{m^* \in M^*} \frac{P(m^*)}{L} \sum_{s_0' \in S} \Big\{ I(g(s_0') = m^*) \\ & \quad \Big[\tfrac{1}{n} \textstyle\sum_{i \in s_0, e_i = 1} (y_i - \bar{y}_e) \Big]^2 \Big\}. \end{aligned} \tag{2.8}$$

An unbiased estimator of the variance is

$$\widetilde{var}(\hat{\mu}^*) = \widehat{var}(\hat{\mu}) - L^{-1} \textstyle\sum_{s_0' \in S} \Big\{ I(g(s_0') = m^*) \Big[\frac{1}{n} \sum_{i \in s_0, e_i = 1} (y_i - \bar{y}_e) \Big]^2 \Big\}. \tag{2.9}$$

A more efficient estimator can be further obtained by conditioning on the sufficient statistics as follows:

$$\begin{aligned}\widehat{var}(\hat{\mu}^*) = \quad & E(\widetilde{var}(\hat{\mu}^*)|M^* = m^*) \\ = \quad & \tfrac{1}{L}\textstyle\sum_{s_0' \in S} I(g(s_0') = m^*)\widehat{var}(\hat{\mu}) - \tfrac{1}{L}\sum_{s_0' \in S}\Big\{ I(g(s_0') = m^*) \\ & \Big[\tfrac{1}{n}\textstyle\sum_{i \in s_0, e_i = 1}(y_i - \bar{y}_e)\Big]^2 \Big\}.\end{aligned} \tag{2.10}$$

2.3 Refined Adaptive Latin Hypercube Designs

2.3.1 Sequential ALHDs without replacement

Ideally, it is desirable to have design points sampled without replacement because computer experiments often generate deterministic outputs and replicates should be avoided. According the adaptive Latin hypercubde procedure, if more than one point from a network is selected in the Latin hypercube design, then the network will be selected more than once. Inspired by this observation, a modification is introduced to conduct the adaptive design without replacement of networks. The idea is to perform the adaptive design sequentially and in each iteration, the design space is updated by removing the selected network so that the networks are selected without replacement. Moreover, such a refinement provides better control over the size of the experiment. An inherent problem with the original adaptive Latin hypercube design is that the size of the experiment may exceed a predetermined cost limit because the neighborhood size increases linearly with the dimension of the experiments. The design procedure can be described as follows.

Define T as a predetermined number of initial points, where $T \leq n$. Let the initial sample space for $(x_1, \cdots, x_p)$ be Ω_1, which includes n^p points. For $l = 1, \cdots, T$, perform steps 1 and 2 iteratively to obtain a sequential ALHD.

Step 1: Based on the sample space Ω_l, randomly select an initial design point, denoted by $u_l = (t_{1l}, \cdots, t_{pl})$,. Sample all points in the cluster of u_l and denote the network of u_l by W.

Step 2: To maintain the Latin hypercube structure of the initial points, the sample space Ω_{l+1} is updated by removing the design points sharing at least one of the coordinates with u_l (i.e., remove the points with $x_i = t_{il}$ for at least one i, where $i = 1, \cdots, p$). Furthermore, remove the units belonging to network W from Ω_{l+1} so that the networks are selected without replacement.

We use the same example in Figure 2.1 to illustrate the procedure of sequential ALHDs. Let $T = 9$. The initial sample space Ω_1 includes 576 ($= 24^2$) points. Assuming that the first

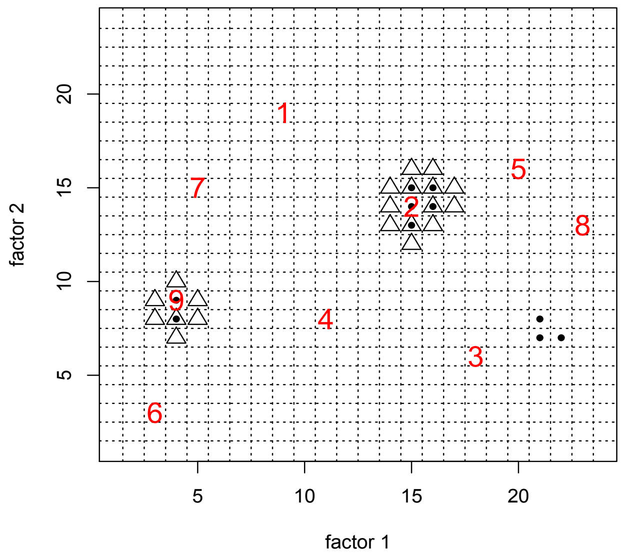

Figure 2.2: An example of sequential ALHD

point $u_1 = (9, 19)$ is randomly selected point from Ω_1, then the sample space Ω_2 is defined by removing the network of u_1 (i.e., a network with size one) and also the $9th$ column and 19th row from Ω_1. The second initial point is selected randomly from Ω_2. Assume $u_2 = (15, 14)$ is selected. Then all the units (marked by triangles) belonging to the network of u_2 are sampled. The sample space Ω_3 is updated by removing the network of u_2 and the $15th$ column and $14th$ row from Ω_2. Following the procedure, a set of nine initial design points are numbered in Figure 2.2 and the adaptively added points are represented by triangles. Clearly, the nine points is a subset of a 24-run LHD.

It is worth noting that ALHDs provide a better control over the size of the experiment because the final number of distinct networks selected, T, 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 $T = n$ because the resulting initial design is a Latin hypercube design and the one-dimensional balance property holds in this case. On the other hand, if the experiment is expensive or time-consuming, T can be a smaller number to reduce costs. It is possible that T cannot be equal to n. This happens only when there is a large network containing at least two adjacent levels of x_1 and at least one of them have all the units included in the network. In this situation, there is a possibility that $w \neq \emptyset$ at least once, and thus we have $N < n$. It, however, rarely occurs in practice because we are mainly focus on the experiments in which responses of interest are clustered in a relatively small area, i.e., the network sizes are small.

2.3.2 Design properties

To study the sampling property, we again focus on the unbiased estimators for the population mean. Denote $\tilde{y}_k^*$, $k = 1, \ldots, K$, be the set of ordered sample values with $\tilde{y}_k^* = \sum_{j \in \tilde{\Psi}_k} y_j$, where $\tilde{\Psi}_k$ is the set of units in the kth network. Raj (1956) introduced a general approach to construct unbiased estimator for unequal probability sampling without replacement. Following Raj's idea, we first consider

$$\begin{aligned} \omega_1 = &\quad \textstyle\sum_{k=1}^K \frac{\tilde{y}_k^* I_1(n_k>0)}{P_1(n_k>0)}, \\ \omega_2 = &\quad \tilde{y}_1^* + \textstyle\sum_{k=2}^K \frac{\tilde{y}_k^* I_2(n_k>0)}{P_2(n_k>0)}, \\ \omega_i = &\quad \tilde{y}_1^* + \ldots + \tilde{y}_{i-1}^* + \textstyle\sum_{k=i}^K \frac{\tilde{y}_k^* I_i(n_k>0)}{P_i(n_k>0)}, \end{aligned} \tag{2.11}$$

where $i = 3, \ldots, T$, $I_i(n_k > 0)$ 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. $P_i(n_k > 0)$ represents the probability that the kth network is selected as the ith sample. This probability is calculated based on the updated feasible region in the iteration corresponding to the ith sample. Let μ be the population mean. It is clear that $n^{-p}\mathrm{E}(\omega_i) = \mu$. Therefore, according to Raj (1956), an unbiased estimator of μ can be written as

$$\hat{\mu}_R = N^{-1} \sum_{i=1}^T u_i \omega_i, \tag{2.12}$$

where the u_i are constants and $\sum_{i=1}^T u_i = 1$.

Murthy (1957) proposed a modification of Raj's estimator, which is derived by constructing an unordered version of Raj's ordered unbiased estimator. Let s_0^* be an ordered final sample of the ν distinct networks selected according to the procedure discussed in Section 2, Γ 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 T network. Murthy introduced an unbiased estimator

$$\hat{\mu}_M = N^{-1} \frac{\sum_{s_0^* \in \Gamma} P(s_0^*) \hat{\mu}_R(s_0^*)}{P(s_0)} \tag{2.13}$$

for the population mean with its 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

$$\hat{\mu}_M = N^{-1} \sum_{i=1}^T \frac{P(s_0 \mid i)}{P(s_0)} y_i^*, \tag{2.14}$$

where $P(s_0 \mid i)$ is the conditional probability of choosing sample s_0 given network i has been chosen as the first network. The variance of (2.14) is given by (Murthy 1957; Cochran 1977)

$$\begin{aligned}\operatorname{var}(\hat{\mu}_M) = \quad & N^{-2}\Bigg[\sum_{i=1}^{K}\sum_{j<i}^{K}\left(1-\sum_{a,b\in s_0}\frac{P(s_0|a)P(s_0|b)}{P(s_0)}\right) \\ & \left(\frac{y_i^*}{P_1(n_i>0)}-\frac{y_j^*}{P_1(n_j>0)}\right)^2 P_1(n_i>0)P_1(n_j>0)\Bigg]\end{aligned} \tag{2.15}$$

and its unbiased estimator is

$$\begin{aligned}\widehat{\operatorname{var}}(\hat{\mu}_M) = \quad & N^{-2}\Bigg[\sum_{i=1}^{T}\sum_{j<i}^{T}\left(\frac{P(s_0|i,j)}{P(s_0)}-\frac{P(s_0|i)P(s_0|j)}{P(s_0)^2}\right) \\ & \left(\frac{y_i^*}{P_1(n_i>0)}-\frac{y_j^*}{P_1(n_j>0)}\right)^2 P_1(n_i>0)P_1(n_j>0)\Bigg],\end{aligned} \tag{2.16}$$

where $P(s_0 \mid i, j)$ 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.

Similar to the discussion in Section 2, the estimator $\hat{\mu}_M$ can be improved by Rao-Blackwellization and an improved unbiased estimator is obtained by the conditional expectation of $\hat{\mu}_M$ given the minimum sufficient statistics. According to the sequential ALHD procedure, the minimum sufficient statistics, d, is the final unordered sample of the ν distinct networks with their 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 is given by $\hat{\mu}_M^{\mathrm{RB}} = E(\hat{\mu}_M \mid D = d)$. It can be written as

$$\hat{\mu}_M^{\mathrm{RB}} = N^{-1}\left\{\sum_{i=1}^{T}\frac{P(s_0 \mid i)}{P(s_0)}y_i^*(1-e_i^*) + \frac{\sum_{s_0'\in S}\left[I(g(s_0')=d)P(s_0')\sum_{i\in s_0',\ e_i=1}\frac{P(s_0'|i)}{P(s_0')}y_i^*\right]}{\sum_{s_0'\in S} I(g(s_0'=d))P(s_0')}\right\}, \tag{2.17}$$

Applying the same derivation, an efficient estimator for the variance is obtained as follows:

$$\widehat{\operatorname{var}}(\hat{\mu}_M^{RB}) = \frac{\sum_{\tilde{s}_0=S} I(g(\tilde{s}_0)=d)P(s_0')\widehat{\operatorname{var}}(\hat{\mu}_M) - \sum_{\tilde{s}_0\in S} I(g(\tilde{s}_0)=d)P(s_0')[\hat{\mu}_M-\hat{\mu}_M^{RB}]^2}{\sum_{s_0'\in S} I(g(s_0'=d))P(s_0')}. \tag{2.18}$$

New estimator Based on a sufficient statistics that contains the unordered final sample, the number of also share the same $P(g(s_0) = m)$

Define a sufficient statistic but not the minimum. Decompose the final unordered sample of the ν distinct networks into two parts. The first part, denoted by s_c, is the set of all distinct units in the sample for which the condition to sample adaptively is satisfied. The remaining units are denoted by $s_{\bar{c}}$. Using these notation, a sufficient statistic can be defined by $d^* = \{(i, y_i), (j, y_j), P(s_0) : i \in s_c, j \in s_{\bar{c}}\}$.

The new unbiased estimator can be written as

$$\mu_M^* = N^{-1}\left\{\sum_{i=1}^{T}\frac{P(s_0|i)}{P(s_0)}y_i^*(1-e_i^*) + \frac{P(s_0|e)e_{s_0}}{P(s_0)e_s^*}\sum_{i\in s} e_i y_i \delta_i\right\}. \tag{2.19}$$

The variance is given by

$$\begin{aligned}\operatorname{var}(\hat{\mu}^*_M) = \quad & \operatorname{var}(\hat{\mu}_M) - \textstyle\sum_{d^*\in D^*} \frac{P(d^*)}{L^* N^2} \sum_{s'_0 \in S} \Big\{ I(g(s'_0) = d^*) \\ & \left[\left(\textstyle\sum_{i\in s'_0,\ e_i=1} \frac{P(s'_0|i)}{P(s'_0)} y_i \right) - \left(\frac{P(s'_0|e)e_{s'_0}}{P(s'_0)e^*_s} \textstyle\sum_{i\in s} e_i y_i \delta_i \right) \right] \Big\}^2 . \end{aligned} \tag{2.20}$$

An unbiased estimator of the variance is

$$\begin{aligned}\widetilde{\operatorname{var}}(\hat{\mu}^*_M) = \quad & \widehat{\operatorname{var}}(\hat{\mu}_M) - \tfrac{1}{L^* N^2} \textstyle\sum_{s'_0 \in S} \Big\{ I(g(s'_0) = d^*) \\ & \left[\left(\textstyle\sum_{i\in s'_0,\ e_i=1} \frac{P(s'_0|i)}{P(s'_0)} y_i \right) - \left(\frac{P(s'_0|e)e_{s'_0}}{P(s'_0)e^*_s} \textstyle\sum_{i\in s} e_i y_i \delta_i \right) \right] \Big\}^2 , \end{aligned} \tag{2.21}$$

and a more efficient estimator for the variance can be written as

$$\begin{aligned}\widehat{\operatorname{var}}(\hat{\mu}^*_M) = \quad & E[\widetilde{\operatorname{var}}(\hat{\mu}^*) \mid D^* = d^*] \\ = \quad & \tfrac{1}{L^*} \textstyle\sum_{s'_0 \in S} I(g(s'_0) = d^*) \widehat{\operatorname{var}}(\hat{\mu}_M) - \tfrac{1}{L^* N^2} \textstyle\sum_{s'_0 \in S} \Big\{ I(g(s'_0) = d^*) \\ & \left[\left(\textstyle\sum_{i\in s'_0,\ e_i=1} \frac{P(s'_0|i)}{P(s'_0)} y_i \right) - \left(\frac{P(s'_0|e)e_{s'_0}}{P(s'_0)e^*_s} \textstyle\sum_{i\in s} e_i y_i \delta_i \right) \right] \Big\}^2 , \end{aligned} \tag{2.22}$$

2.4 A toy example

In this section, the proposed design procedures are implemented using a small example that allows us to demonstrate all the possible design combinations. The example is shown in Figure 2.3, which involves two factors at three levels. The numbers shown in the cells are the responses. We first illustrate ALHDs. There are six possible ALHDs with $n = 3$ and the final designs are listed in Table 2.1. For each row in the table, the first three design points are the initial LHDs and the rest of them are adaptively added when the responses are larger than or equal to 8. Three unbiased estimators, $\hat{\mu}, \hat{\mu}^*$, and $\hat{\mu}^{RB}$, and their estimated variances are calculated for each design. The last row summarizes the average performance of these estimators over all possible designs. The three estimators are unbiased and the average variance of $\hat{\mu}^{RB}$ and $\hat{\mu}^*$ are 8.4% and 1.5% smaller than that from the original estimator, $\hat{\mu}$. The average final sample size is 7.67.

Using the sequential ALHDs procedure with $T = 2$, there are 34 possible designs listed in Table 2.2 along with their probabilities reported in the column "P". For each row, the first two points are the ordered initial design points and the rests are the adaptively added points according to the same criterion in Table 2.1. Similar to the previous table, the last row summarized the average performances of the unbiased estimators over all designs. Note that as shown in this toy example, the unbiased estimators of the variance are not guaranteed to be non-negative. To obtain a reasonable average of variance, negative variances are replaced by zeros in the calculation. Conditions for non-negativity of the estimators and new estimators of

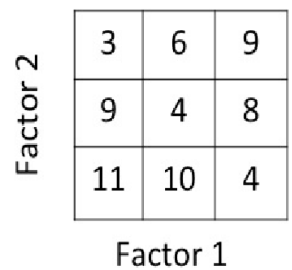

Figure 2.3: A toy example with two factors

Table 2.1: Toy Example: Adaptive Latin hypercube designs

Design points	$\hat{\mu}$	$\hat{\mu}^{RB}$	$\hat{\mu}^*$	$\widehat{\mathrm{var}}(\hat{\mu})$	$\widehat{\mathrm{var}}(\hat{\mu}^{RB})$	$\widehat{\mathrm{var}}(\hat{\mu}^*)$
(1,2)(2,2)(3,3);all	8.17	7.92	8.28	0.68	5.24	4.47
(1,1)(2,3)(3,2);all	8.83	7.92	8.28	10.68	5.24	4.47
(1,2)(2,1)(3,3);all	6.83	7.92	6.83	9.12	5.24	9.12
(1,3)(2,1)(3,2);all	7.83	7.92	8.28	2.56	5.24	4.47
(1,2)(2,3)(3,1);(1,1)(1,3)(2,1)(2,2)	7.33	7.33	7.33	3.85	3.85	3.85
(1,3)(2,2)(3,1);	3.67	3.67	3.67	6	6	6
mean	7.11	7.11	7.11	5.48	5.13	5.40

the variance that can maintain the non-negativity deserve further investigation. According to results in Table 2.2, the average variance is reduced by 39% (= (5.48 − 3.36)/5.48) compare with the original ALHDs and the average final sample size is 6.04, which is 21.3% ((7.67 − 6.04)/7.67) smaller than that in Table 2.1.

2.5 Applications

2.5.1 Borehole model

To study the advantage of the proposed designs, we conducted simulations based on a computer model to compare the performance of the adaptive designs and the nonadaptive simple random sampling (without replacement) with sample size equal to the adaptive design. The model considered is a widely used borehole model (Morris 1993), which describes the cause of the flow rate through a borehole as follows:

$$y = \frac{2\pi T_u(H_u - H_l)}{ln(r/r_w)[1 + \frac{2LT_u}{ln(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l}]}$$

The simulation is conduced based on four factors at four levels. The detailed setting are given below:

$$H_u - H_l = (170, 270, 400), r_w = (0.05, 0.1, 0.15),$$
$$L = (1120, 1440, 1680), r = (100, 5000, 50000),$$

and the rest of the variables are specified as $T_u = 115600, T_l = 116, K_w = 12045$.

For the original ALHDs, 4 by 4-run LHDs are used as initial designs and for the sequential ALHDs, $T = 3$ is specified. For both designs, simulations are performed using different quantiles of the response as the criterion for adaptive selection and they all lead to similar conclusions; therefore, we only demonstrate one particular criterion here to save space. The criterion is specified as $y \geq 160$, where 160 is approximately the upper 0.2 quantile. Based on 10000 iterations, the simulation results are summarized in Table 2.12 and Table 2.13. In both tables, the last row indicates the efficiency of the proposed estimator with respect to the simple random sample estimator based on the sample size. If $\hat{\mu}_.$ denotes one of the proposed estimators, then the efficiency of such an estimator is calculated by $\text{var}(\hat{\mu}_{\text{SRS}})/\text{var}(\hat{\mu}_.)$. The results show that both types of ALHDs have significant improvement on the estimation efficiency over simple random sampling. The average sample size is 76.26 for the original ALHDs, and 58.81 for the sequential ALHDs. Comparing the results in Tables 2.12 and 2.13, it shows that with a 23% (= (76.26 − 58.81)/76.26) smaller sample size, the average variance can be reduced by 15% (= (2464.47 − 2097.21)/2464.47) using the sequential ALHDs and they can be further reduced

Table 2.2: Toy example: Sequential Adaptive Latin Hypercube Designs

	Design points	P	$\hat{\mu}_M$	$\hat{\mu}_M^{RB}$	$\hat{\mu}_M^*$	$\widehat{\text{var}}(\hat{\mu}_M)$	$\widehat{\text{var}}(\hat{\mu}_M^{RB})$	$\widehat{\text{var}}(\hat{\mu}_M^*)$
1	(1,3)(2,2);	$\frac{1}{36}$	3.5	3.5	3.5	-0.03	-0.03	-0.03
2	(2,2)(1,3);	$\frac{1}{36}$	3.5	3.5	3.5	-0.03	-0.03	-0.03
3	(1,3)(3,1);	$\frac{1}{36}$	3.5	3.5	3.5	-0.03	-0.03	-0.03
4	(3,1)(1,3);	$\frac{1}{36}$	3.5	3.5	3.5	-0.03	-0.03	-0.03
5	(2,3)(3,1);	$\frac{1}{36}$	5	5	5	-0.11	-0.11	-0.11
6	(3,1)(2,3));	$\frac{1}{36}$	5	5	5	-0.11	-0.11	-0.11
7	(2,2)(3,1);	$\frac{1}{36}$	4	4	4	0	0	0
8	(3,1)(2,2);	$\frac{1}{36}$	4	4	4	0	0	0
9	(2,3)(3,2);(3,1)(2,2)(3,3)	$\frac{1}{36}$	7.25	6.75	6.75	1.22	2.33	2.33
10	(3,2)(2,3);(3,1)(2,2)(3,3)	$\frac{1}{36}$	7.25	6.75	6.75	1.22	2.33	2.33
11	(2,2)(3,3);(2,3)(3,1)(3,2)	$\frac{1}{36}$	6.25	6.75	6.75	3.94	2.33	2.33
12	(3,3)(2,2);(2,3)(3,1)(3,2)	$\frac{1}{36}$	6.25	6.75	6.75	3.94	2.33	2.33
13	(1,3)(3,2);(2,2)(2,3)(3,1)(3,3)	$\frac{1}{36}$	5.75	5.75	5.75	5.88	5.88	5.88
14	(3,2)(1,3);(2,2)(2,3)(3,1)(3,3)	$\frac{1}{36}$	5.75	5.75	5.75	5.88	5.88	5.88
15	(1,3)(2,1);(1,1)(1,2)(2,2)(3,1)	$\frac{1}{36}$	7	7.15	7.21	16	14.17	13.83
16	(2,1)(1,3);(1,1)(1,2)(2,2)(3,1)	$\frac{1}{27}$	7	7.15	7.21	16	14.17	13.83
17	(1,2)(3,1);(1,1)(1,3)(2,1)(2,2)	$\frac{1}{27}$	7.43	7.15	7.21	11.76	14.17	13.83
18	(3,1)(1,2);(1,1)(1,3)(2,1)(2,2)	$\frac{1}{36}$	7.43	7.15	7.21	11.76	14.17	13.83
19	(1,1)(2,2);(1,2)(1,3)(2,1)(3,1)	$\frac{1}{36}$	7	7.15	7	15	14.17	15
20	(2,2)(1,1);(1,2)(1,3)(2,1)(3,1)	$\frac{1}{36}$	7	7.15	7	15	14.17	15
21	(1,2)(2,3);(1,1)(1,3)(2,1)(2,2)(3,1)	$\frac{1}{27}$	8.15	8.15	8.15	0.95	0.95	0.95
22	(2,3)(1,2);(1,1)(1,3)(2,1)(2,2)(3,1)	$\frac{1}{36}$	8.15	8.15	8.15	0.95	0.95	0.95
23	(1,1)(2,3);(1,1)(1,3)(2,1)(2,2)(3,1)	$\frac{1}{36}$	8.15	8.15	8.15	0.95	0.95	0.95
24	(2,3)(1,1);(1,1)(1,3)(2,1)(2,2)(3,1)	$\frac{1}{36}$	8.15	8.15	8.15	0.95	0.95	0.95
25	(1,2)(3,3);all	$\frac{1}{27}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
26	(3,3)(1,2);all	$\frac{1}{36}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
27	(1,1)(3,3);all	$\frac{1}{36}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
28	(3,3)(1,1);all	$\frac{1}{36}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
29	(1,1)(3,2);all	$\frac{1}{36}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
30	(3,2)(1,1);all	$\frac{1}{36}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
31	(2,1)(3,3);all	$\frac{1}{27}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
32	(3,3)(2,1);all	$\frac{1}{36}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
33	(2,1)(3,2);all	$\frac{1}{27}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
34	(3,2)(2,1);all	$\frac{1}{36}$	9.32	9.32	9.32	-0.012	-0.012	-0.012
	mean		7.11	7.11	7.11	3.36	3.33	3.33

Table 2.3: Borehole $4 * 3$ example: ALHDs

	$\hat{\mu}$	$\hat{\mu}^{RB}$	$\hat{\mu}^{*}$	$\hat{\mu}_{\mathrm{SRS}}$
mean	91.99	92.22	92.04	91.71
variance	2464.47	2425.61	2444.54	5048.80
efficiency	2.05	2.08	2.07	1

Table 2.4: Borehole $4 * 3$ example: Sequential ALHDs

	$\hat{\mu}_M$	$\hat{\mu}_M^{RB}$	$\hat{\mu}_M^{*}$	$\hat{\mu}_{\mathrm{SRS}}$
mean	126.58	125.96	124.67	92.37
variance	2097.21	1808.83	1965.59	5133.68
efficiency	2.45	2.84	2.61	1

by more than 25% ($= (2425.61 - 1808.83)/2425.61$) with the Rao-Blackwell estimators.

Another example is 3 by 3 experiment. We use levels $(170, 270, 400)$ to present $H_u - H_l$, $(0.05, 0.1, 0.15)$ to present r_w, and $(1120, 1440, 1680)$ for L. For other five variables, the settings are $r = 100, T_u = 115600, T_l = 116, K_w = 12045$. We run this 27-run experimental design with criteria defined as $y \geq y's$ upper 0.25 quantile. 10000 iterations are tried and the results are summarized in Table 2.5 and Table 2.6. Mean sample size is 13.24 in adaptive sampling and 9.32 in sequential sampling.

Table 2.5: Borehole $3 * 3$ example: ALHDs

	$\hat{\mu}$	$\hat{\mu}_{RB}$	$\hat{\mu}^{*}$	$\hat{\mu}_{\mathrm{SRS}}$
mean	89.23	89.31	89.22	90.08
variance	1986.16	1783.55	1831.24	6005.65
efficiency	0.33	0.30	0.30	1

2.5.2 Data center thermal management study

The proposed design is demonstrated using computational fluid dynamic (CFD) simulations conducted at IBM for a data center thermal management study. A data center is an integrated facility housing multipe-unit servers, providing application services or management for data

Table 2.6: Borehole $3*3$ example: Sequential ALHDs

	$\hat{\mu}$	$\hat{\mu}_{RB}$	$\hat{\mu}^*$	$\hat{\mu}_{\text{SRS}}$
mean	89.39	89.22	89.41	89.44
variance	1165.57	1117.94	1162.46	6009.82
efficiency	0.19	0.19	0.19	1

Table 2.7: Factors in a data center thermal management study

Factor	levels
Computer room air conditioning (CRAC) unit 1 flow rate (cfm)	0, 7750, 10750, 13000
Computer room air conditioning (CRAC) unit 2 flow rate (cfm)	0, 2500, 4000, 5500
Percentage of tile open area	20%, 40%, 60%, 75%

processing. Data center Facilities constantly generate large amounts of heat to the room, 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, the goal of the study is to design a data center with an efficient heat-removal mechanism (Schmidt, Cruz, and Iyengar 2005; Hamann 2008).

To achieve the foregoing goal, three factors at four levels are considered in this study. Details regarding these factors and their levels are given in Table 2.7. Using CFD simulations, we are able to obtain the maximum room temperature (MRT) under each data center setting and the objective is to find the optimal setting so that MRT can be minimized. A sequential ALHD was conducted with $T = 3$ and it proceed by adding design points adaptively if the observed MRT is lower than or equal to $80°F$. By doing this, more design points are sequentially conducted in the neighborhood of the existing design that have smaller MRT, because they are more likely to be the optimal heat-removal mechanism. The final sample size of the sequential ALHD is 28. The unbiased estimators and their estimated variances are given by

$$\hat{\mu}_M = 88.50, \hat{\mu}_M^{\text{RB}} = 87.35, \hat{\mu}_M^* = 87.05, \widehat{\text{var}}(\hat{\mu}_M) = 35.73, \widehat{\text{var}}(\hat{\mu}_M^{RB}) = 7.01, \widehat{\text{var}}(\hat{\mu}_M^*) = 14.54.$$

For adaptive sampling, the unbiased estimator $\hat{\mu}$ based on adaptive designs has efficiency 0.33 compared to simple random sampling with the same sample size. With smaller sample size than adaptive sampling, sequential sampling has smaller variances 1165.57 compared to 1986.16, and smaller efficiencies 0.19. Consistent with previous proof, $var(\hat{\mu}_{RB}) \leq var(\hat{\mu}^*) \leq var(\hat{\mu})$ for both adaptive sampling and sequential sampling.

For example, Williams et al. (2000) proposed a sequential procedure according to the expected improvement calculated based on fitted Gaussian process models. Gramacy and Lee (2009) introduced a class of sequential designs based on treed Gaussian process models.

2.5.3 Piezoelectric Energy Harvester Example

Energy harvesting (EH) is the mechanism which converts energy from ambient waste energy sources, like waste heat, solar energy, wind, vibration, human energy on the iphone, or key typing motion energy. Among varied energy harvesting processes, piezoelectric devices attracts attentions because they directly convert mechanical energy into electrical power. This gives a possible solution to making the portable electronic devices like PDAs, cellular phones, and laptops more sustainable. One interesting idea is to insert the piezoelectric device into the keyboard mechanism so that the typing motion energy may be used or stored for sustainable use. Therefore, how to increase the energy conversion efficiency, or maximize output energy under static or dynamic loading receives great attention.

The example here aims to explore the induced voltage performance in an unimorph piezoelectric cantilever device under static force load, and how geometric parameters affect the output power. This unimorph cantilever device (Figure 2.4) consists of two layers, with upper layer is piezoelectric material and lower layer is non-piezoelectric material. A concentrated force is imposed on the free end. The surface of the cantilever is a rectangular shape with length L, width b_1 for the upper layer and b_2 for the lower layer. The two layers are bonded by a conductive adhesive. Force F is continuously pushed along z direction. Nickel electrodes are applied on both sides of piezoelectric layer for voltage measurement. The thicknesses of non-piezoelectric layer and piezoelectric layer are t_s and t_p respectively. Here Lead Zirconate Titanate (PZT) is the piezoelectric material, which is commonly used in energy harvesting, and stainless steel(SS) is the material in non-piezoelectric layer.

The following function describes the relationship between the induced voltage and other parameters:

$$V_{in,ave} = \frac{1}{4}\frac{d_{31}}{\varepsilon_0\varepsilon_r}\frac{FE_p}{D}(z_2^2 - z_1^2)L \tag{2.23}$$

where ε_r and ε_0 are the dielectric constant of the piezoelectric layer and the permittivity of free space respectively.d_{31} is the piezoelectric coefficient of $PZT.E_s$ and E_p are the Yang's modulus of stainless steel and PZT. And the distance from the interface between the stainless steel and PZT layers to neutral line is:

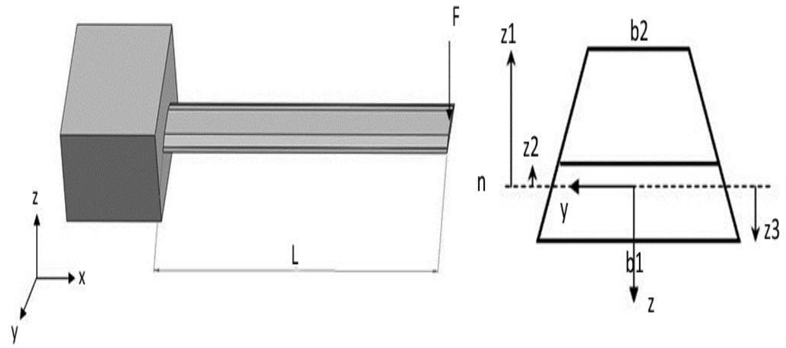

Figure 2.4: Piezoelectric Energy Harvester Example

$$a = \frac{1}{3}\frac{-3E_s t_s^2 b_2 t_p - E_s t_s^3 b_1 - 2E_s t_s^3 b_2 + E_p t_p^3 b_2 + 3E_p t_p^2 b_1 t_s + 2E_p t_p^3 b_1}{2E_s t_s b_2 t_p + E_s t_s^2 b_1 + E_s t_s^2 b_2 + E_p t_p^2 b_2 + 2E_p t_p b_1 t_s + E_p t_p^2 b_1} \tag{2.24}$$

Then $z_1 = -(t_p + a), z_2 = -a, z_3 = t_s - a$ are corresponding to the distances from the neutral axis to the top of the SS layer, neutral line, and bottom of the PZT layer, and bending modulus

$$D = E_s[\frac{A}{3}(z_3^3 - z_2^3) + \frac{B}{4}(z_3^4 - z_2^4)] + E_p[\frac{A}{3}(z_2^3 - z_1^3) + \frac{B}{4}(z_2^4 - z_1^4)] \tag{2.25}$$

with $A = b_2 + \frac{b_1 - b_2}{h}(t_s + a); B = \frac{b_1 - b_2}{h}$, and $h = z_1 + z_3$ is the distance from upper layer top to lower layer bottom.

The simulation is conduced based on three factors at four levels with following settings:

$$t_s = (10^{-6}, 10^{-4}, 10^{-3}, 10^{-2}), t_p = (10^{-6}, 10^{-4}, 10^{-3}, 10^{-2}), b_2 = (0.01, 0.02, 0.03, 0.04) \tag{2.26}$$

$E_s = 20*10^{10}, E_p = 6.2*10^{10}, d_{31} = -274*10^{-12}, \varepsilon_r = 3130, \varepsilon_0 = 8.854*10^{-12}, F = 1, L = 12*10^{-3}, b_1 = 2*10^{-3}$. Induced voltage are calculated using classic equation (Wang 2003). 50000 iterations of adaptive and sequential sampling are proceeded, and results are as follows.

2.6 Discussion

The computation of $P(s_0)$ can be intensive when T is large because it requires the consideration of $T!$ permutations. This is not surprising given the same disadvantage with Murthy's estimator. However, this approach is mainly recommended for the experiments in which observations of interest are usually clustered in a few locations. Therefore, the number of network with size greater than one in the sample is few and this reduces the computation. We are currently

Table 2.8: EH simulation results based on ALHDs:$V \geq quantile(V, 0.75)$, average final sample size=25.32

	$\hat{\mu}$	$\hat{\mu}^{RB}$	$\hat{\mu}^{*}$	$\hat{\mu}_{\text{SRS}}$	$\hat{\mu}_{\text{LHD}}$
mean	6.89	6.83	6.86	7.18	1.92
variance	318.28	290.39	311.83	383.33	352.20
efficiency	1.2	1.32	1.23	1	1.09

Table 2.9: EH simulation results based on sequential ALHDs:$V \geq quantile(V, 0.75)$,average final sample size=15.48

	$\hat{\mu}$	$\hat{\mu}^{RB}$	$\hat{\mu}^{*}$	$\hat{\mu}_{\text{SRS}}$	
mean	6.86	7.10	6.86	6.93	19.50
variance	478.62	359.63	478.12	1492.63	1713.37
efficiency	3.12	4.15	3.12	1	0.87

Table 2.10: EH simulation results based on ALHDs:$V \geq quantile(V, 0.8)$, average final sample size = 21.88

	$\hat{\mu}$	$\hat{\mu}^{RB}$	$\hat{\mu}^{*}$	$\hat{\mu}_{\text{SRS}}$	$\hat{\mu}_{\text{LHD}}$
mean	6.99	6.93	6.97	6.64	2.60
variance	373.88	355.4	371.63	468.46	485.86
efficiency	1.25	1.32	1.26	1	0.96

Table 2.11: EH simulation results based on sequential ALHDs:$V \geq quantile(V, 0.8)$, average final sample size = 12.40

	$\hat{\mu}$	$\hat{\mu}^{RB}$	$\hat{\mu}^{*}$	$\hat{\mu}_{\text{SRS}}$	$\hat{\mu}_{LHD}$
mean	6.65	6.90	6.65	6.96	22.46724
variance	570.39	338.92	522.38	1660.39	1904.717
efficiency	2.91	4.90	3.18	1	0.87

Table 2.12: EH simulation results based on ALHDs:$V \geq quantile(V, 0.9)$, average final sample size = 11.54

	$\hat{\mu}$	$\hat{\mu}^{RB}$	$\hat{\mu}^{*}$	$\hat{\mu}_{\text{SRS}}$	$\hat{\mu}_{\text{LHD}}$
mean	7.37	7.50	7.50	7.36	3.83
variance	568.60	557.36	560.48	828.34	771.38
efficiency	1.46	1.49	1.48	1	1.07

Table 2.13: EH simulation results based on Sequential ALHDs:$V \geq quantile(V, 0.9)$, average final sample size = 6.02

	$\hat{\mu}_M$	$\hat{\mu}_M^{RB}$	$\hat{\mu}_M^{*}$	$\hat{\mu}_{\text{SRS}LHD}$	
mean	6.98	7.07	6.98	6.76	27.76
variance	772.89	643.44	771.48	2061.18	2191.18
efficiency	2.67	3.2	2.67	0.94	

developing algorithms to efficiently evaluate $P(s_0)$ under a general setting and the result will be reported elsewhere.

Another interesting extension is the study of adaptive designs with optimized initial designs. This is because the initial designs in both types of ALHDs are not unique. Careful construction of the initial designs with some desirable properties, such as maximizing the minimum inter-site distances, is expected to be useful in improving the estimation efficiency. Design criteria, search algorithms, and theoretical properties are the issues that warrants further investigation. Moreover, further research on generalization of the proposed sequential idea to other classes of space-filling designs as the initial designs would be interesting.

Chapter 3

Appendix

$$\begin{aligned} P(n_k > 0) &= \textstyle\sum_{j=1}^{\min(\psi_k,n)} (-1)^{j+1} P(\text{ at least } j \text{ points from network } k \text{ appear in LHD}) \\ &= \textstyle\sum_{j=1}^{\min(\psi_k,n)} (-1)^{j+1} q_{k,j} P(j \text{ points appear in LHD}) \\ &= \textstyle\sum_{j=1}^{\min(\psi_k,n)} (-1)^{j+1} q_{k,j} [(n-j)!]^{p-1} (n!)^{1-p} \\ &= \textstyle\sum_{j=1}^{\min(\psi_k,n)} (-1)^{j+1} q_{k,j} \Big[n(n-1)\cdots(n-j+1) \Big]^{1-p} \end{aligned} \tag{3.1}$$

To proof the unbiasedness of $\hat{\mu}$, we rewrite $\hat{\mu}$ as follows

$$\hat{\mu} = \frac{1}{N} \sum_{k}^{K} \frac{(\sum_{j \in \Psi_k} y_j) I(n_k > 0)}{P(n_k > 0)} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i I(n_k > 0)}{P(n_k > 0)} \tag{3.2}$$

and thus, the result follows.

APPENDIX A: PROOF OF LEMMA 1

Because all the networks have size one, we have $\psi_k = 1$ and $q_{k,1} = 1$ for all $k = 1, \cdots, K$. $K = n^p$, $q_{(k,1),(h,1)} = 1$ only if y_k and y_h have no coordinates in common.

APPENDIX B: DERIVATION OF (2.4) AND (2.5)

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

$$\hat{\mu}^{\mathrm{RB}} = E(\mu | M = m) = \sum_{s_0' \in S} \hat{\mu} P(S_0 = s_0' | M = m) \tag{3.3}$$

and the conditional probability can be written as

$$P(S_0 = s_0' | M = m) = \frac{I(g(s_0') = m)}{\sum_{s_0' \in S} I(g(s_0') = m)}. \tag{3.4}$$

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

$$\begin{aligned} \hat{\mu} &= \tfrac{1}{N} \textstyle\sum_{k=1}^{K} \tfrac{y_k^* J_k (1-e_k^*)}{P(n_k>0)} + \tfrac{1}{N} \textstyle\sum_{k=1}^{K} \tfrac{y_k^* J_k e_k}{P(n_k>0)} \\ &= \tfrac{1}{N} \textstyle\sum_{k=1}^{K} \tfrac{y_k^* J_k (1-e_k^*)}{P(n_k>0)} + \tfrac{1}{N} \textstyle\sum_{i \in s_0, e_i = 1} y_i n^{p-1}. \end{aligned} \tag{3.5}$$

The first term on the right hand side of (3.5) 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 $P(n_k > 0) = P(n_k = 1) = n^{1-p}$. Therefore, we have

$$\begin{aligned} \hat{\mu}^{\mathrm{RB}} = & \textstyle\sum_{s_0' \in S} \hat{\mu} \frac{I(g(s_0')=m)}{\sum_{s_0' \in S} I(g(s_0')=m)} \\ = & \textstyle\frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* J_k (1-e_k^*)}{P(n_k>0)} + \frac{1}{NL} \sum_{s_0' \in S} \{ I(g(s_0') = m) [\sum_{i \in s_0', e_i=1} y_i n^{p-1}] \}, \\ = & \textstyle\frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* J_k (1-e_k^*)}{P(n_k>0)} + \frac{1}{nL} \sum_{s_0' \in S} \{ I(g(s_0') = m) [\sum_{i \in s_0', e_i=1} y_i] \}, \end{aligned} \tag{3.6}$$

where $L = \sum_{s_0' \in S} I(g(s_0') = m)$. The variance of $\hat{\mu}^{\mathrm{RB}}$ can be calculated by $var(\hat{\mu}^{\mathrm{RB}}) = var(\hat{\mu}) - E[(\hat{\mu} - \hat{\mu}^{\mathrm{RB}})^2]$ and the detail derivation is omitted.

APPENDIX C: DERIVATION OF (1.5) AND (1.2.3)

Define S_0 as the set of initial design.

$$\hat{\mu}^* = E(\mu | M^* = m^*) = \sum_{s_0' \in S} \hat{\mu} P(S_0 = s_0' | M^* = m^*) \tag{3.7}$$

Because the conditional probability can be calculated by

$$P(S_0 = s_0' | M^* = m^*) = \frac{I(g(s_0') = m^*)(n!)^{1-p}}{\sum_{s_0' \in S} I(g(s_0') = m^*)(n!)^{1-p}} = \frac{I(g(s_0') = m^*)}{\sum_{s_0' \in S} I(g(s_0') = m^*)}. \tag{3.8}$$

Similarly, we decompose $\hat{\mu}$ into two parts. We have

$$\begin{aligned} \hat{\mu} = & \textstyle\frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* I(n_k>0)(1-e_k^*)}{P(n_k>0)} + \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* I(n_k>0) e_k}{P(n_k>0)} \\ = & \textstyle\frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* I(n_k>0)(1-e_k)}{P(n_k>0)} + \frac{1}{N} \sum_{i \in s_0, e_i=1} y_i n^{p-1}. \end{aligned} \tag{3.9}$$

Based on (3.7) and (3.9) and the fact that the first term on the right hand side of (3.9) is fixed given $M^* = m^*$, it follows that

$$\begin{aligned} \hat{\mu}^* = & \textstyle\sum_{s_0' \in S} \hat{\mu} \frac{I(g(s_0')=m^*)}{\sum_{s_0' \in S} I(g(s_0')=m^*)} \\ = & \textstyle\frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* I(n_k>0)(1-e_k^*)}{P(n_k>0)} + (nL)^{-1} \sum_{s_0' \in S} I(g(s_0') = m^*) [\sum_{i \in s_0', e_i=1} y_i] \end{aligned} \tag{3.10}$$

Because only a subset of the edge points can lead to the same sufficient statistics, we further define the subset of the edge points by $\Omega_{s_c} = \{w \in s_{\bar{c}} : g(s_0) = m^*, s_0 = (s_c, w)\}$, where s_c is a subset of s containing the networks with size larger than one. Each edge point in Ω_{s_c} has an equal probability of being selected in the initial Latin hypercube design, therefore we have

$$\begin{aligned} & \textstyle\sum_{s_0' \in S} I(g(s_0') = m^*) \left(\sum_{i \in s_0', e_i=1} y_i \right) \\ = & \textstyle\frac{L}{G} \begin{pmatrix} e_s - 1 \\ e_{s_0} - 1 \end{pmatrix} (\sum_{i \in s} e_i y_i t_i) \end{aligned} \tag{3.11}$$

where $t_i = 1$ if $i \in \Omega_{s_c}$ and $t_i = 0$ otherwise, $e_{s_0} = \sum_{i \in s_0} e_i$ is the total number of edge points in the initial sample, $e_s = \sum_{i \in s} e_i t_i$ is the total number of edge points in the final sample that can lead to the same m^*, and

$$G = \begin{pmatrix} e_s \\ e_{s_0} \end{pmatrix}$$

Thus,

$$\begin{aligned} \hat{\mu}^* = & \tfrac{1}{N} \textstyle\sum_{k=1}^{K} \tfrac{y_k^* I(n_k > 0)(1 - e_k^*)}{P(n_k > 0)} + \tfrac{e_{s_0}}{n e_s} (\textstyle\sum_{i \in s} e_i t_i y_i) \\ = & \tfrac{1}{N} \textstyle\sum_{k=1}^{K} \tfrac{y_k^* I(n_k > 0)(1 - e_k^*)}{P(n_k > 0)} + \tfrac{1}{n} e_{s_0} \bar{y}_e \\ = & \tfrac{1}{N} \textstyle\sum_{k=1}^{K} \tfrac{y_k^* I(n_k > 0)(1 - e_k^*)}{P(n_k > 0)} + \tfrac{1}{n} \textstyle\sum_{i \in s_0, e_i = 1} \bar{y}_e \end{aligned} \tag{3.12}$$

where $\bar{y}_e = \frac{1}{e_s} \sum_{i \in s} e_i t_i y_i$ So (1.5) holds.

For the variance, we have

$$\begin{aligned} var(\hat{\mu}^*) = & var(\hat{\mu}) - \textstyle\sum_{m^* \in M^*} \tfrac{P(m^*)}{L} \textstyle\sum_{s_0' \in S} I(g(s_0') = m^*)(\hat{\mu} - \hat{\mu}^*)^2 \\ = & var(\hat{\mu}) - \textstyle\sum_{m^* \in M^*} \tfrac{P(m^*)}{L} \textstyle\sum_{s_0' \in S} \Big\{ I(g(s_0') = m^*) \\ & \Big[\tfrac{1}{n} \textstyle\sum_{i \in s_0, e_i = 1} (y_i - \bar{y}_e) \Big]^2 \Big\}. \end{aligned} \tag{3.13}$$

APPENDIX D: DERIVATION OF (2.19) AND (2.22)

The easy-to-compute unbiased estimator is given by

$$\hat{\mu}_M^* = E(\hat{\mu}_M \mid D^* = d^*) = \sum_{s_0' \in S} \hat{\mu}_M P(S_0 = s_0' \mid D^* = d^*). \tag{3.14}$$

We decompose $\hat{\mu}_M$ into two parts

$$N^{-1} \sum_{i=1}^{T} \frac{P(s_0|i)}{P(s_0)} y_i^* = N^{-1} \left[\sum_{i=1}^{T} \frac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) + \sum_{i=1}^{T} \frac{P(s_0|i)}{P(s_0)} y_i^* e_i \right] \tag{3.15}$$

The first term on the right side of (3.15) remains the same given $D^* = d^*$, therefore we have

$$\begin{aligned} & \textstyle\sum_{s_0' \in S} \left(\textstyle\sum_{i=1}^{T} \tfrac{P(s_0|i)}{P(s_0)} y_i^* \right) \tfrac{I(g(s_0') = d^*) P(s_0')}{\sum_{s_0' \in S} I(g(s_0') = d^*) P(s_0')} \\ = & \textstyle\sum_{s_0' \in S} \left(\textstyle\sum_{i=1}^{T} \tfrac{P(s_0|i)}{P(s_0)} y_i^* \right) \tfrac{I(g(s_0') = d^*)}{\sum_{s_0' \in S} I(g(s_0') = d^*)} \\ = & \textstyle\sum_{i=1}^{T} \tfrac{P(s_0|i)}{P(s_0)} y_i^* (1 - e_i^*) + (1/L^*) \textstyle\sum_{s_0' \in S} \Big\{ I(g(s_0') = d^*) \Big[\textstyle\sum_{i \in s_0',\ e_i = 1} \tfrac{P(s_0'|i)}{P(s_0')} y_i \Big] \Big\} \end{aligned} \tag{3.16}$$

Because $\frac{P(s_0'|i)}{P(s_0')}$ is the same given i being an edge point and $D^* = d^*$,we can further denote this value by $\frac{P(s_0|e)}{P(s_0)}$. Hence, the second term on the right of (3.16) can be written as

$$\frac{P(s_0|e)}{L^* P(s_0)} \left[\sum_{s_0' \in S} I(g(s_0') = d^*) \Big(\sum_{i \in s_0',\ e_i = 1} y_i \Big) \right]. \tag{3.17}$$

Similar to ...., we define the subset of edge points that can lead to the same d^* by $\Delta_{s_c} = \{w \in s_{\bar{c}} : g(s_0) = d^*, s_0 = (s_c, w)\}$. Based a similar argument in 3.16, we have

$$\begin{aligned} \textstyle\sum_{s_0' \in S} I(g(s_0') = d^*)\left[\sum_{i \in s_0',\ e_i=1} y_i\right] &= \tfrac{L^* e_{s_0}}{e_s^*} \textstyle\sum_{i \in s,\ e_i=1} y_i \delta_i \\ &= \tfrac{L^* e_{s_0}}{e_s^*} \textstyle\sum_{i \in s} e_i y_i \delta_i, \end{aligned} \tag{3.18}$$

where $\delta_i = 1$ if $i \in \Delta_{s_c}$ and $\delta_i = 0$ otherwise and $e_s^* = \sum_{i \in s} e_i \delta_i$ is the total number of edge points in the final sample that can lead to the same d^*. Therefore, we have

$$\mu_M^* = N^{-1}\left\{ \sum_{i=1}^{T} \frac{P(s_0|i)}{P(s_0)} y_i^*(1 - e_i^*) + \frac{P(s_0|e)e_{s_0}}{P(s_0)e_s^*} \sum_{i \in s} e_i y_i \delta_i \right\}. \tag{3.19}$$

Therefore, (2.19) holds.

For the variance, we have

$$\begin{aligned} \mathrm{var}(\hat{\mu}_M^*) = &\ \mathrm{var}(\hat{\mu}_M) - E[(\hat{\mu}_M - \hat{\mu}_M^*)^2] \\ = &\ \mathrm{var}(\hat{\mu}_M) - \textstyle\sum_{d^* \in D^*} \frac{P(d^*)}{L^*} \sum_{s_0' \in S} I(g(s_0') = d^*)(\hat{\mu}_M - \hat{\mu}_M^*)^2 \\ = &\ \mathrm{var}(\hat{\mu}_M) - \textstyle\sum_{d^* \in D^*} \frac{P(d^*)}{L^* N^2} \sum_{s_0' \in S} \Big\{ I(g(s_0') = d^*) \\ &\ \left[\left(\textstyle\sum_{i \in s_0',\ e_i=1} \frac{P(s_0'|i)}{P(s_0')} y_i\right) - \left(\frac{P(s_0'|e)e_{s_0'}}{P(s_0')e_s^*} \textstyle\sum_{i \in s} e_i y_i \delta_i\right)\right] \Big\}^2. \end{aligned} \tag{3.20}$$

Bibliography

Chaudhuri, P. and Mykland, P. A. (1993). Nonlinear experiments: Optimal design and inference based on likelihood.

Chernoff, H. (1959). Sequential design of experiments. *The Annals of Mathematical Statistics*, pages 755–770.

Cochran, W. C. (1977). *Sampling techniques*. Wiley, New York.

Constantine, G. M. (1981). Some e-optimal block designs. *The Annals of Statistics*, 9(4):886–892.

DraguljiÄ, D., Santner, T. J., and Dean, A. M. (2012). Noncollapsing space-filling designs for bounded nonrectangular regions. *Technometrics*, 54(2):169–178.

Dryver, A. L. and Thompson, S. K. (2005). Improved unbiased estimators in adaptive cluster sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):157–166.

Fang, K.-T., Li, R., and Sudjianto, A. (2006). *Design and modeling for computer experiments*. CRC Press.

Fang, K.-T., Ma, C.-X., and Winker, P. (2002). Centered -discrepancy of random sampling and latin hypercube design, and construction of uniform designs. *Mathematics of Computation*, 71(237):275–296.

Fedorov, V. (1972). *Theory Of Optimal Experiments*. ACADEMIC PRESS, INC.

Ghosh, P. K. S. B. (1991). *Handbook of Sequential Analysis*. Marcel Dekker, Inc.

Glover, F. (1990). Tabu search: A tutorial. *Interfaces*, 20(4):74–94.

Gramacy, R. B. and Lee, H. K. H. (2009). Adaptive design and analysis of supercomputer experiments. *Technometrics*, 51(2):130–145.

Hamann, H. F. (2008). A measurement-based method for improving data center energy efficiency. In *SUTC*, pages 312–313.

Hancock, J. T., Thom-Santelli, J., and Ritchie, T. (2004). Deception and design: The impact of communication technology on lying behavior. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–134. ACM.

Handcock, M. S. (1994). *Measuring the uncertainty in kriging*, pages 436–447. Geostatistics for the Next Century. Springer.

Handcock, M. S. and Stein, M. L. (1993). A bayesian analysis of kriging. *Technometrics*, 35(4):403–410.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Hung, Y. (2011). Adaptive probability-based latin hypercube designs. *Journal of the American Statistical Association*, 106(493):213–219.

Hung, Y., Amemiya, Y., and Wu, C.-F. J. (2010). Probability-based latin hypercube designs for slid-rectangular regions. *Biometrika*, 97(4):961–968.

Iman, R. L. (2008). *Latin hypercube sampling*. Wiley Online Library.

Iman, R. L. and Conover, W. (1982a). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics-Simulation and Computation*, 11(3):311–334.

Iman, R. L. and Conover, W. (1982b). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics-Simulation and Computation*, 11(3):311–334.

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2):131–148.

Joseph, V. R. and Hung, Y. (2008). Orthogonal-maximin latin hypercube designs. *Statistica Sinica*, 18(1):171.

Jr, D. L. S. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, 99(465):262–278.

Karr, J. R. (1981). Assessment of biotic integrity using fish communities. *Fisheries*, 6(6):21–27.

Matérn, B. (1986). Spatial variation, vol. 36 of. *Lecture Notes in Statistics*, 2.

Mentré, F., Mallet, A., and Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika*, 84(2):429–442.

Morris, M. D. and Mitchell, T. J. (1995). Exploratory designs for computational experiments. *Journal of statistical planning and inference*, 43(3):381–402.

Murty, V. N. (1971). Minimax designs. *Journal of the American Statistical Association*, 66(334):319–320.

Ouwens, M. J. N. M., Tan, F. E. S., and Berger, M. P. F. (2002). Maximin d-optimal designs for longitudinal mixed effects models. *Biometrics*, 58(4):735–741.

Palacios, M. B. and Steel, M. F. J. (2006). Non-gaussian bayesian geostatistical modeling. *Journal of the American Statistical Association*, 101(474):604–618.

Raj, D. (1956). Some estimators in sampling with varying probabilities without replacement. *Journal of the American Statistical Association*, 51(274):269–284.

Rankin, E. T. and Ohio, E. (1989). *The qualitative habitat evaluation index [QHEI]: Rationale, methods, and application*. State of Ohio Environmental Protection Agecny.

Roesch, F. A. (1993). Adaptive cluster sampling for forest inventories. *Forest Science*, 39(4):655–669.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423.

Salehi, M. and Seber, G. A. (1997). Two-stage adaptive cluster sampling. *Biometrics*, pages 959–970.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The design and analysis of computer experiments*. Springer.

Schmidt, R. R., Cruz, E., and Iyengar, M. (2005). Challenges of data center thermal management. *IBM Journal of Research and Development*, 49(4.5):709–723.

Seber, G. and Thompson, S. (1994). 6 environmental adaptive sampling. *Handbook of statistics*, 12:201–220.

Smith, D. R., Conroy, M. J., and Brakhage, D. H. (1995). Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, pages 777–788.

Stehman, S. V. and Overton, W. S. (1994). 9 environmental sampling and monitoring. *Handbook of statistics*, 12:263–306.

Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, 85(412):1050–1059.

Thompson, S. K. (1991). Stratified adaptive cluster sampling. *Biometrika*, 78(2):389–397.

Thompson, S. K., Ramsey, F. L., and Seber, G. A. (1992). An adaptive procedure for sampling animal populations. *Biometrics*, pages 1195–1199.

Thompson, S. K. and Seber, G. A. F. (1996). *Adaptive sampling*. Wiley, New York.

Tsay, J.-Y. (1976). On the sequential construction of d-optimal designs. *Journal of the American Statistical Association*, 71(355):671–674.

Wang, L.-P., Jr, R. A. W., Wang, Y., Deng, K. K., Zou, L., Davis, R. J., and Trolier-McKinstry, S. (2003). Design, fabrication, and measurement of high-sensitivity piezoelectric microelectromechanical systems accelerometers. *Microelectromechanical Systems, Journal of*, 12(4):433–439.

Williams, B. J., Santner, T. J., and Notz, W. I. (2000). Sequential design of computer experiments to minimize integrated response functions. *Statistica Sinica*, 10(4):1133–1152.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.