SOME LARGE DEVIATIONS RESULTS FOR LATIN HYPERCUBE SAMPLING

Shane S. Drew
Tito Homem-de-Mello

Department of Industrial Engineering
and Management Sciences
Northwestern University
Evanston, IL 60208-3119, U.S.A.

ABSTRACT

Large deviations theory is a well-studied area which has shown to have numerous applications. The typical results, however, assume that the underlying random variables are either i.i.d. or exhibit some form of Markovian dependence. Our interest in this paper is to study the validity of large deviations results in the context of estimators built with *Latin Hypercube sampling*, a well-known sampling technique for variance reduction. We show that a large deviation principle holds for Latin Hypercube sampling for functions in one dimension and for separable multi-dimensional functions. Moreover, the upper bound of the probability of a large deviation in these cases is no higher under Latin Hypercube sampling than it is under Monte Carlo sampling. We extend the latter property to functions that preserve negative dependence (such as functions that are monotone in each argument). Numerical experiments illustrate the theoretical results presented in the paper.

1 INTRODUCTION

Suppose we wish to calculate $\mathbb{E}_P[g(X)]$ where $X = [X^1, \ldots, X^d]$ is a random vector in $\mathbb{R}^d$ with probability distribution P and $g(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}$ is a measurable function. Further, suppose that the expected value is finite and cannot be written in closed form or be easily calculated, but that $g(X)$ can be easily computed for a given value of X. Let $\mathbb{E}_P[g(X)] = \mu \in (-\infty, \infty)$. To estimate the expected value, we can use the sample average approximation:

$$\frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i(\omega)) \qquad (1)$$

where the $X_i(\omega)$ are random realizations of X. When the $X_i(\omega)$ are i.i.d. (i.e. Monte Carlo sampling), by the law of large numbers the sample average approximation should approach the true mean μ (with probability one) as

the number of samples n becomes large. Large deviations theory ensures that the probability that the sample average approximation deviates from μ by a fixed amount $\delta > 0$ approaches zero exponentially fast as n goes to infinity. Formally, this is expressed as

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\left|\frac{1}{n} S_n - \mu\right| > \delta\right) = -\beta_\delta,$$

where β_δ is a positive constant.

The above description, of course, is a small fraction of a much more general theory, but conveys a basic concept — that one obtains *exponential convergence* of estimators under certain conditions. This idea has found applications in numerous areas, from simulation to telecommunications to optimization; we refer to classical books in the area such as (Shwartz and Weiss 1995), (Dembo and Zeitouni 1998) and (Bucklew 2004) for further discussions.

Despite the exponential convergence results mentioned above, it is well known that Monte Carlo methods have some drawbacks, particularly when one wants to calculate the *errors* corresponding to given estimates. Although the theory behind such calculations — notably the Central Limit Theorem — is solid, in practice the error may be large even for large sample sizes. That has led to the development of many variance reduction techniques as well as alternative sampling methods (see, e.g., Law and Kelton 2000 for a general discussion of this topic).

One alternative approach for sampling the $X_i(\omega)$ is called *Latin Hypercube sampling* (LHS, for short), introduced by McKay, Beckman, and Conover (1979). Broadly speaking, the method calls for splitting each dimension into n strata (yielding n^d hypercubes) and, for every dimension, sampling all n strata exactly once. This technique has been extensively used in practice, not only because of simplicity of implementation but also because of its nice properties. Indeed, McKay, Beckman, and Conover (1979) show that if $g(X)$ is monotone in all of its arguments, then the variance of the estimator obtained with LHS (call it Var_{LHS})

is no higher than the sample variance from Monte Carlo sampling (Var$_{MC}$). Hoshino and Takemura (2000) extend this result to the case where $g(\cdot)$ is monotone in all but one of its arguments. Stein (1987) splits the function $g(\cdot)$ into a constant component μ plus first order terms plus a residual component (this is called the ANOVA decomposition of g), i.e.,

$$g(X) = \mu + g_1(X^1) + \cdots + g_d(X^d) + g_{resid}(X). \quad (2)$$

Stein shows that, asymptotically, the sample variance from Latin Hypercube sampling is just equal to the variance of the residual term and is no worse than the variance from Monte Carlo sampling. Loh (1996) extends this result to the multivariate case where $g : \mathbb{R}^d \mapsto \mathbb{R}^m$. Owen (1997) shows that for any n and any function g, Var$_{LHS} \leq \frac{n}{n-1}Var_{MC}$. Also, Owen (1992) shows that LHS satisfies a Central Limit Theorem with the variance equal to the variance of the residual term.

The above discussion shows that the LHS method has been well studied and possesses many nice properties. However, to the best of our knowledge there have been no studies on the exponential convergence of estimators obtained with LHS. Thus, it is of interest to know whether large deviations results hold under Latin Hypercube sampling. This is by no means a trivial question — since the $X_i(\omega)$ are no longer i.i.d. under LHS, Cramér's Theorem (which is the basic pillar of the results for i.i.d. sampling) can no longer be applied.

In this paper, we study the above problem. We derive conditions under which large deviations results hold under Latin Hypercube sampling. More specifically, our results apply when the integrand function is either one-dimensional, multi-dimensional but separable (i.e. functions with no residual term), or where the LH samples are negatively dependent. One special case of negative dependence is the case when the underlying function is monotone in all of its arguments. Further, in these situations, we show that the upper bound for the large deviations probability is lower under LHS than under Monte Carlo sampling. Jin, Fu, and Xiong (2003) show this property holds when negatively dependent sampling is used to estimate a probability quantile, whereas we prove it for the situations mentioned above.

The remaining of the paper is organized as follows. In Section 2, we give some background on large deviations theory and Latin Hypercube sampling. In Section 3, we show our results for functions in one-dimension. In Section 4, we extend the one-dimensional results to separable functions with multi-dimensional domain and functions where negatively dependent sampling is used. In Section 5 we show some examples of our results and in Section 6 we present concluding remarks.

Due to space constraints, we omit most of the proofs; those can be found — along with more detailed discussions — in (Drew and Homem-de-Mello 2005).

2 BACKGROUND

2.1 Large Deviations

We begin with an overview of some of the basic results from large deviations theory. All of the results in this section can be found in any large deviations textbook such as (Dembo and Zeitouni 1998) or (den Hollander 2000).

Suppose Y is a real-valued variable with mean $\mu = \mathbb{E}[Y]$ (possibly infinite) and let $S_n = \sum_{i=1}^{n} Y_i$, where $Y_1, \ldots, Y_n$ are (not necessarily i.i.d.) *unbiased* samples of Y, i.e., $\mathbb{E}[Y_i] = \mu$.

Define the function

$$\phi_n(\theta) := \frac{1}{n} \log \mathbb{E}[\exp(\theta S_n)], \quad (3)$$

or equivalently,

$$\exp(n\phi_n(\theta)) = \mathbb{E}[\exp(\theta S_n)],$$

and let

$$\phi(\theta) := \lim_{n \to \infty} \phi_n(\theta) \text{ when the limit exists.} \quad (4)$$

In the case where the Y_i are i.i.d., we have

$$\begin{aligned}
\phi_n(\theta) &= \frac{1}{n} \log(\mathbb{E}[\exp(\theta S_n)]) \\
&= \frac{1}{n} \log(\{\mathbb{E}[\exp(\theta Y_1)]\}^n) \\
&= \log(\mathbb{E}[\exp(\theta Y_1)]) \\
&= \log M_{Y_1}(\theta)
\end{aligned}$$

where $M_{Y_1}(\theta)$ is the moment generating function of Y_1 evaluated at θ. Notice that the above quantity does not depend on n, so $\phi(\theta) = \phi_n(\theta)$.

The *large deviations rate function* of Y at a given point x is defined as

$$I(x) := \sup_{\theta} [\theta x - \phi(\theta)]. \quad (5)$$

It can be shown that both $\phi(\cdot)$ and $I(\cdot)$ are convex.

The estimator $\frac{1}{n} S_n$ is said to satisfy a *large deviation principle (LDP)* with rate function $I(\cdot)$ if the following conditions hold:

1. $I(\cdot)$ is lower semi-continuous, i.e., it has closed level sets;

2.   For every closed subset $F \in \mathbb{R}$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} S_n \in F\right) \leq -\inf_{x \in F} I(x)$$

3.   For every open subset $G \in \mathbb{R}$,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} S_n \in G\right) \geq -\inf_{x \in G} I(x).$$

$I(\cdot)$ is said to be a *good rate function* if it has compact level sets. Note that this implies that there exists some point x such that $I(x) = 0$.

Next, define

$$I(n, x) = \sup_{\theta \in \mathbb{R}}[\theta x - \phi_n(\theta)]. \tag{6}$$

When using the asymptotic $\phi(\theta)$ instead of $\phi_n(\theta)$, we denote this function $I(x)$. In the i.i.d. case, $I(x) = I(n, x)$.

Let (a, b) be an interval on the real line containing μ. We wish to calculate the probability of a large deviation, i.e.

$$\mathbb{P}\left(\frac{1}{n} S_n \notin (a, b)\right) = \mathbb{P}\left(\frac{1}{n} S_n \leq a\right) + \mathbb{P}\left(\frac{1}{n} S_n \geq b\right).$$

For all $\theta > 0$, it holds that $\mathbb{P}(\frac{1}{n} S_n \geq b) = \mathbb{P}(S_n \geq bn) = \mathbb{P}(\exp(\theta S_n) \geq \exp(\theta bn))$. Using Chebyshev's inequality, $\mathbb{P}(\frac{1}{n} S_n \geq b) \leq \exp(-\theta bn)\mathbb{E}[\exp(\theta S_n)]$. Thus,

$$\frac{1}{n} \log\left[\mathbb{P}\left(\frac{1}{n} S_n \geq b\right)\right] \leq -\left(\theta b - \frac{1}{n} \log \mathbb{E}[\exp(\theta S_n)]\right)$$
$$= -[\theta b - \phi_n(\theta)].$$

Note that this inequality holds regardless of any assumptions on the Y_is. Moreover, since the above inequality is true for all $\theta \geq 0$ it follows that

$$\frac{1}{n} \log\left[\mathbb{P}\left(\frac{1}{n} S_n \geq b\right)\right] \leq \inf_{\theta \geq 0} -[\theta b - \phi_n(\theta)]. \tag{7}$$

By Jensen's inequality, we have $\mathbb{E}[\exp(\theta S_n)] \geq \exp(\theta \mathbb{E}[S_n]) = \exp(\theta n \mu)$ for any $\theta \in \mathbb{R}$ and hence

$$\phi_n(\theta) \geq \theta \mu \qquad \text{for all } \theta \in \mathbb{R}. \tag{8}$$

It follows that $\theta b - \phi_n(\theta) \leq \theta(b - \mu)$. Since $b \geq \mu$, we can take the infimum in (7) over $\theta \in \mathbb{R}$ — which is then equal to $-I(n, b)$ — and, moreover, if $b = \mu$ the right-hand side of (7) is zero.

Similarly, for all $\theta < 0$ it holds that $\mathbb{P}(\frac{1}{n} S_n \leq a) = \mathbb{P}(S_n \leq an) = \mathbb{P}(\exp(\theta S_n) \geq \exp(\theta an))$. By repeating the

argument in the above paragraphs we conclude that

$$\frac{1}{n} \log\left[\mathbb{P}\left(\frac{1}{n} S_n \geq b\right)\right] \leq -I(n, b) \tag{9a}$$

$$\frac{1}{n} \log\left[\mathbb{P}\left(\frac{1}{n} S_n \leq a\right)\right] \leq -I(n, a). \tag{9b}$$

In the i.i.d. case, since $\phi(\theta) = \phi_n(\theta)$ for all n we have that (9) holds with the terms $-I(b)$ and $-I(a)$ on the right-hand side, thus yielding Chernoff's upper bound, i.e.,

$$\mathbb{P}\left(\frac{1}{n} S_n \notin (a, b)\right) \leq e^{-nI(a)} + e^{-nI(b)}. \tag{10}$$

The upper bounds in (9) hold for any positive integer n. In the i.i.d. case, it is possible to show that (9) holds with the reverse inequalities in the limit as n approaches infinity (note however that the lower bounds do not necessarily hold for a given n). Together, the above statements constitute Cramér's Theorem, which states that, under i.i.d. sampling, $\frac{1}{n} S_n$ satisfies a large deviation principle with good rate function $I(\cdot)$.

The above results can be extended to the non-i.i.d. case, which is our main interest as Latin Hypercube samples are dependent. The main tool for that case is the Gartner-Ellis Theorem, which we describe below. Roughly speaking, the theorem asserts that, under proper conditions, a large deviation principle holds for the estimator $\frac{1}{n} S_n$, with the rate defined in terms of the limiting $\phi(\theta)$ defined in (4).

The assumptions of the theorem are the following:

Assumption 1 : *For each $\theta \in \mathbb{R}$, the function $\phi(\theta)$ defined in (4) exists as an extended real number.*

Assumption 2 : *0 belongs to the interior of D_ϕ where $D_\phi = \{\theta \in \mathbb{R} : \phi(\theta) < \infty\}$.*

Assumption 3 : *$\phi(\theta)$ is essentially smooth, i.e. the following three conditions hold:*

i) The interior of D_ϕ is nonempty

ii) $\phi(\theta)$ is differentiable on the interior of D_ϕ

iii) Either $D_\phi = \mathbb{R}$ or $\phi(\theta)$ is steep, i.e. for $\theta \in D_\phi$ as θ approaches the boundary of D_ϕ, $|\phi'(\theta)| = \infty$ (where $\phi'(\theta)$ is the derivative of ϕ with respect to θ).

Assumption 4 : *$\phi(\theta)$ is lower semi-continuous.*

Theorem 1 (Gartner-Ellis Theorem)*:*

1.   *If Assumptions 1 and 2 hold, then for every closed subset F of $\mathbb{R}$,*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} S_n \in F\right) \leq -\inf_{x \in F} I(x).$$

2.   *If Assumptions 1-3 hold, then for every open subset G of $\mathbb{R}$,*

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} S_n \in G\right) \geq -\inf_{x \in G} I(x).$$

3. *If Assumptions 1-4 all hold, then a large deviation principle holds with the good rate function $I(\cdot)$.*

That rate function is $I(x) = \sup_\theta [\theta x - \phi(\theta)]$.

Our main goal is to derive conditions under which the Gartner-Ellis Theorem can be applied under Latin Hypercube sampling. We do that in Sections 3 and 4. Before that, we review in detail the basic ideas of LHS.

2.2 Latin Hypercube Sampling

Let $X = [X^1, X^2, \ldots, X^d]$ be the vector of the d input variables of a simulation and let $Y = g(X) = g(X^1, X^2, \ldots, X^d)$ be the output of the simulation. Assume that all of the dimensions are independent. Let $F_j(\cdot)$ be the marginal cumulative distribution function for X^j. Suppose the quantity of interest is $\mathbb{E}[Y]$.

One possible sampling method to estimate $\mathbb{E}[Y]$ is to randomly sample n points in the sample space (Monte Carlo sampling). For each replication i from 1 to n, Uniform(0,1) random numbers $U_i = [U_i^1, \ldots, U_i^d]$ are generated (one per dimension) which, assuming we can use the inverse transform method, yield the input random vector $X_i = [F_1^{-1}(U_i^1), \ldots, F_d^{-1}(U_i^d)]$ and the output $Y_i = g(X_i)$.

One problem with random sampling is that there is no guarantee that all sections of the sample space are equally represented. Input points could be clustered in one particular region. An improvement on this is Latin Hypercube sampling, first proposed by McKay, Beckman, and Conover (1979). Each dimension of the sample space is split into n sections (or strata) each with probability $\frac{1}{n}$, and one observation is randomly sampled from each stratum. The algorithm is comprised of three steps — it generates some uniform random numbers, then some random permutations and finally these elements are put together to yield the samples. The detailed algorithm is the following:

1. *Generate uniform random numbers:*
   (a) Generate a $n \times d$ matrix U of Uniform(0,1) random numbers. Let U_i^j be the $(i, j)^{th}$ entry of this matrix.
   (b) Create another $n \times d$ matrix $V(U)$ with $(i, j)^{th}$ entry $V_i^j(U) = \frac{i-1+U_i^j}{n}$. Thus each $V_i^j(U)$ is uniform on the interval $[\frac{i-1}{n}, \frac{i}{n}]$.

2. *Generate random permutations:*
   (a) Let $P(n)$ be the set of column vectors of permutations of the numbers $(1, 2, \ldots, n)$. There are $n!$ possible permutations, each equally likely. Let $\mathcal{P}$ be the set of $n \times d$ matrices where each column (representing an input variable) is a random permutation in $P(n)$ with all columns mutually independent. There are $(n!)^d$ elements in $\mathcal{P}$, each equally likely. Index these

with $k = 1, \ldots, (n!)^d$ and let K be a random index.

   (b) Randomly select $\Pi(K) \in \mathcal{P}$ (i.e. the Kth element of $\mathcal{P}$). Let $\pi_i^j(K)$ be the $(i, j)^{th}$ entry of this matrix. Note that the permutation matrix $\Pi(K)$ is independent of the random number matrix $V(U)$.

   (c) In Latin Hypercube sampling, only n of the n^d strata are sampled. The rows of the $\Pi(K)$ matrix determine which hypercubes get sampled. Let $\pi_i(K) = [\pi_i^1(K), \ldots, \pi_i^d(K)]$ be the ith row of $\Pi(K)$. This corresponds to the hypercube that covers the $\pi_i^1(K)^{th}$ stratum of X^1, the $\pi_i^2(K)^{th}$ stratum of X^2, ..., and the $\pi_i^d(K)^{th}$ stratum of X^d.

3. *Determine the randomly sampled point within each hypercube.*
   (a) Create matrix $Z(\omega) = Z(V, K)$ with $(i, j)^{th}$ entry $Z_i^j(\omega) = V_{\pi_i^j(K)}^j(U)$. In other words, the $(i, j)^{th}$ entry of $Z(\omega)$ corresponds to the $(\pi_i^j(K), j)^{th}$ entry of $V(U)$ based on the permutation matrix. Thus the jth column $V^j(U)$ of the random number matrix $V(U)$ is permuted according to the jth column of the permutation matrix $\Pi(K)$.

   (b) Let $X_i^j(\omega) = F_j^{-1}[Z_i^j(\omega)]$. Then $X_i(\omega) = [X_i^1(\omega), \ldots, X_i^d(\omega)]$ and $Y_i(\omega) = g(X_i(\omega))$.

The above algorithm generates n random vectors $Z_i(\omega) = [Z_i^1(\omega), \ldots, Z_i^d(\omega)]$, each of which is uniformly distributed on $[0, 1]^d$. Unlike standard Monte Carlo, of course, the vectors $Z_1, \ldots, Z_n$ are *not* independent. These vectors are mapped via inverse transform into vectors $X_1, \ldots, X_n$, which then are used to generate the samples $Y_1, \ldots, Y_n$. It is well known that each Y_i generated by the LHS method is an unbiased estimate of $\mathbb{E}[Y]$ (see, e.g., the appendix in McKay, Beckman, and Conover 1979).

More formally, let $f : [0, 1]^d \mapsto \mathbb{R}^d$ be the function that converts the uniform random vector $Z_i(\omega)$ into the random vector $X_i(\omega)$, and let $h := g \circ f$. Then we have $Y_i(\omega) = g(X_i(\omega)) = g(f(Z_i(\omega))) = h(Z_i(\omega))$. Thus, without loss of generality we will assume that the outputs Y_i are functions of random vectors that are uniformly distributed on $[0, 1]^d$.

3 THE ONE-DIMENSIONAL CASE

We study now large deviations properties of the estimators generated by LHS. In order to facilitate the analysis, we start by considering the one-dimensional case.

Let $h : [0, 1] \mapsto \mathbb{R}$ be a real-valued function in one variable. It is known that a large deviations principle is satisfied when we use Monte Carlo sampling to estimate $\mathbb{E}[h(Z)]$ via the sample average approximation. We wish to show that a large deviations principle is also satisfied when we use Latin Hypercube sampling to estimate $\mathbb{E}[h(Z)]$ and that the upper bound on the probability of a large deviation is lower for Latin Hypercube sampling than for Monte Carlo sampling.

In standard Monte Carlo sampling, the $Z_i(\omega)$ are all independent Uniform(0,1) random variables. We can calculate

$$\exp(n\phi_n^{MC}(\theta)) = \mathbb{E}[\exp(\theta S_n)]$$
$$= \mathbb{E}[\exp(\theta \sum_{i=1}^n h(Z_i))]$$
$$= \mathbb{E}[\prod_{i=1}^n \exp(\theta h(Z_i))]$$
$$= \prod_{i=1}^n \mathbb{E}[\exp(\theta h(Z_i))]$$
$$= \{\mathbb{E}[\exp(\theta h(Z_1))]\}^n$$

and so

$$\phi_n^{MC}(\theta) = \log(\mathbb{E}[\exp(\theta h(Z_1))])$$
$$= \log\left[\int_0^1 \exp(\theta h(z))dz\right],$$

which is independent of n.

Thus

$$\phi^{MC}(\theta) = \log\left[\int_0^1 \exp(\theta h(z))dz\right]. \qquad (11)$$

In LHS, when the interval $[0, 1]$ is split into n strata of equal probability $\frac{1}{n}$, the intervals are all of the form $[\frac{j-1}{n}, \frac{j}{n}]$ and each random variable $Z_i(\omega)$ is now uniform on some interval of length $\frac{1}{n}$. Further, independence no longer holds.

We make the following assumptions about the function $h(z) : [0, 1] \mapsto \mathbb{R}$:

Assumption 5

1. $h(z)$ is an integrable function (i.e. $|\int_0^1 h(z)dz| < \infty$).
2. $h(z)$ has at most a finite number of singularities.
3. $h(z)$ has a finite moment generating function (i.e. $\int_0^1 \exp(\theta h(z))dz < \infty$ for all $\theta \in \mathbb{R}$).

A simple situation where the above assumptions are satisfied is when $h(\cdot)$ is a bounded function; however, we do allow $h(\cdot)$ to be unbounded. Also, it can be shown that the third part of this assumption is equivalent to assuming that $D_\phi = \mathbb{R}$.

To show that LHS satisfies a large deviation principle, we will show that it satisfies the assumptions of the Gartner-Ellis Theorem. The key element of the proof is the lemma below, which shows that $\{\phi_n(\theta)\}$ converges to a *linear* function in θ.

Lemma 1 *Suppose $h(z) : [0, 1] \mapsto \mathbb{R}$ and that Assumption 5 holds. Let Z be a Uniform(0,1) random variable. If Latin Hypercube sampling is used to estimate $\mathbb{E}[h(Z)]$, then $\lim_{n\to\infty} \phi_n(\theta) = \theta \int_0^1 h(z)dz$.*

The main result of this section is the following:

Theorem 2 *Suppose $h(z) : [0, 1] \mapsto \mathbb{R}$ and that Assumption 5 holds. Let Z be a Uniform(0,1) random variable and define $\mu_1 := \mathbb{E}[h(Z)] = \int_0^1 h(z)dz$. Then, the LHS estimator of μ_1 satisfies a large deviation principle with good rate function*

$$I_{LHS}(x) = \begin{cases} \infty, & \text{if } x \neq \mu_1 \\ 0, & \text{if } x = \mu_1. \end{cases}$$

Proof From Lemma 1, Assumption 1 holds for the 1-dimensional case. Let $\phi(\theta)$ denote the linear function $\theta \int_0^1 h(z)dz = \theta\mu_1$.

Since by assumption $h(z)$ is integrable, μ_1 is finite and $D_\phi = \mathbb{R}$. Thus the interior of D_ϕ is also $\mathbb{R}$, meaning that Assumption 2 holds. Also, since $\phi(\theta)$ is a linear function of θ, it is differentiable everywhere and lower semi-continuous and thus Assumptions 3 and 4 also hold and the Gartner-Ellis Theorem can be applied. The resulting rate function is

$$I_{\text{LHS}}(x) = \sup_\theta[\theta x - \phi(\theta)]$$
$$= \sup_\theta[\theta(x - \mu_1)]$$
$$= \begin{cases} \infty, & \text{if } x \neq \mu_1 \\ 0, & \text{if } x = \mu_1 \end{cases}$$

which is a good rate function since $\{x : I_{\text{LHS}}(x) \leq \alpha\} = \{0\}$ for any $\alpha \geq 0$. $\square$

Theorem 2 implies that, for any closed subset F of $\mathbb{R}$, as long as $\mu_1 \notin F$ we have that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} S_n \in F\right) \leq -\inf_{x\in F} I(x) = -\infty.$$

That is, we have an *infinite decay rate*, as opposed to the exponential rate obtained with standard Monte Carlo. This shows that, asymptotically, LHS is much more precise than Monte Carlo.

The next result suggests that superiority of LHS (in the context of deviation probabilities) in fact holds for any finite n.

Theorem 3 *Consider the setting of Theorem 2. Let $I_{MC}(x)$ and $I_{LHS}(n, x)$ denote the (non-asymptotic) functions defined in (6) respectively for Monte Carlo and for LHS. Then, for* any *sample size n and all x we have that $I_{LHS}(n, x) \geq I_{MC}(x)$.*

In particular, from (9) we see that the upper bound for the probability of a large deviation is smaller under Latin Hypercube sampling than under Monte Carlo sampling for any sample size n.

Although the above result compares only the upper bounds corresponding to LHS and Monte Carlo, the importance of Theorem 3 lies in the fact that the Monte Carlo upper bound is *tight* asymptotically. This suggests that even for small sample sizes the deviation probabilities under LHS may be smaller than under Monte Carlo — a fact that is corroborated in the examples of Section 5.

4 THE MULTI-DIMENSIONAL CASE

We consider now the multi-dimensional case $h : [0, 1]^d \mapsto \mathbb{R}$. Assume that the dimensions are all independent and define $Z_i(\omega) = [Z_i^1(\omega), \ldots, Z_i^d(\omega)]$.

For Monte Carlo sampling, a large deviation principle holds, and we can show using a similar calculation to the one-dimensional case that the function ϕ_n defined in (3) is equal to

$$\phi^{MC}(\theta) = \log\left[\int_{[0,1]^d} \exp(\theta h(z))dz\right]. \quad (12)$$

Again, we would like to show that a large deviation principle holds for Latin Hypercube sampling in the multi-dimensional case and that the upper bound for the probability of a large deviation under LHS is lower than it is for Monte Carlo sampling. While the latter assertion is in general not true for multidimensional functions, we will focus on two special cases: (1) $h(\cdot)$ is a *separable* function, and (2) the samples are *negatively dependent*.

In the multi-dimensional case, each Latin Hypercube permutation is equally likely with probability $\mathbb{P}[K = k] = \frac{1}{(n!)^d}$ (recall that the permutation matrices are indexed by k, and that K is a random index). As in the one-dimensional case, given a particular permutation $\Pi(k)$, the point sampled from each strata is independent of the point sampled from any other strata, so the product and the expectation can be

switched. Thus, we can write

$$\exp(n\phi_n^{LHS}(\theta)) = \mathbb{E}\left[\prod_{i=1}^{n}\exp(\theta h(Z_i(V, K)))\right]$$

$$= \sum_{k=1}^{(n!)^d}\mathbb{E}\left[\prod_{i=1}^{n}\exp(\theta h(Z_i(V, K))) \mid K = k\right]\mathbb{P}[K = k]$$

$$= \frac{1}{(n!)^d}\sum_{k=1}^{(n!)^d}\prod_{i=1}^{n}\mathbb{E}\left[\exp(\theta h(Z_i(V, K))) \mid K = k\right]. \quad (13)$$

Also, given a particular permutation index $K = k$, for each sample i we have that

$$Z_i(V, k) \in \left[\frac{\pi_i^1(k) - 1}{n}, \frac{\pi_i^1(k)}{n}\right] \times \cdots \times \left[\frac{\pi_i^d(k) - 1}{n}, \frac{\pi_i^d(k)}{n}\right].$$

For notational convenience, define $a_i^j(k) := \frac{\pi_i^j(k)-1}{n}$ and $b_i^j(k) := \frac{\pi_i^j(k)}{n}$. Also, let $z := [z^1, \ldots, z^d]$ and $dz := dz^1 \cdots dz^d$. Note that $Z_i^j(V, k)$ is uniformly distributed on $(a_i^j(k), b_i^j(k))$. Then, (13) becomes

$$\exp(n\phi_n^{LHS}(\theta))$$

$$= \frac{1}{(n!)^d}\sum_{k=1}^{(n!)^d}\prod_{i=1}^{n}n^d\int_{a_i^d(k)}^{b_i^d(k)}\cdots\int_{a_i^1(k)}^{b_i^1(k)}\exp(\theta h(z))dz. \quad (14)$$

We now specialize the calculations for the two cases mentioned above.

4.1 Case 1: The Separable Function Case
Definition 1 *A function $h(\cdot)$ defined on a d-dimensional space is said to be* separable *if there exist one-dimensional functions $h^1, \ldots, h^d$ such that*

$$h(z^1, \ldots, z^d) = h^1(z^1) + \ldots + h^d(z^d).$$

Note that this is equivalent to saying that the ANOVA decomposition of h (cf. (2)) has residual part equal to zero. Also, when a function is separable,

$$\int_{[0,1]^d} h(z)dz = \int_0^1 h^1(z^1)dz^1 + \cdots + \int_0^1 h^d(z^d)dz^d.$$

Since a separable multidimensional function can be decomposed into a sum of one dimensional functions, it is intuitive that our results from the one-dimensional case can be extended to this case. The theorem below states precisely that:

Theorem 4   *Suppose* $h(z) : [0, 1]^d \mapsto \mathbb{R}$ *is a separable function and that each component* h^j *of* h *satisfies Assumption 5. Let* Z *be a random vector uniformly distributed on* $[0, 1]^d$, *and define* $\mu_d := \mathbb{E}[h(Z)] = \int_{[0,1]^d} h(z)dz$. *Then, the LHS estimator of* μ_d *satisfies a large deviation principle with good rate function*

$$I(x) = \begin{cases} \infty, & \text{if } x \neq \mu_d \\ 0, & \text{if } x = \mu_d. \end{cases}$$

Proof   As in the proof of Theorem 2, the basic idea is to show that the functions $\{\phi_n^{LHS}(\theta)\}$ converge to the linear function $\phi^{LHS}(\theta) := \theta\mu_d$. Then, applying the Gartner-Ellis Theorem yields the desired result. Details of the proof can be found in (Drew and Homem-de-Mello 2005). □

As before, for any closed subset F of $\mathbb{R}$, as long as $\mu_d \notin F$ we have a decay with infinite rate, i.e.,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}S_n \in F\right) \leq -\inf_{x\in F} I(x) = -\infty.$$

Moreover, as in the one-dimensional case, in the case where $h(z)$ is separable the Chernoff upper bound for the large deviations rate for LHS is better than that of Monte Carlo sampling for any number of samples, i.e., we have an extension of Theorem 3:

Theorem 5   *Consider the setting of Theorem 4. Let* $I_{MC}(x)$ *and* $I_{LHS}(n, x)$ *denote the (non-asymptotic) functions defined in* (6) *respectively for Monte Carlo and for LHS. Then, for any sample size n and all x we have that* $I_{LHS}(n, x) \geq I_{MC}(x)$.

In particular, from (9) *we see that the upper bound for the probability of a large deviation is smaller under Latin Hypercube sampling than under Monte Carlo sampling for any sample size n.*

4.2 Case 2: The Negative Dependent Case

We move now to the case of negative dependence. The relevance of this case is due to the fact that Latin Hypercube samples are negatively dependent (as shown in Jin, Fu, and Xiong 2003) and so a function preserving that property will inherit some of the behavior of the samples.

We start by defining negative dependence.

Definition 2   *Random variables* $Y_i, i = i \ldots n$ *are called* negatively dependent *if*

$$\mathbb{P}[Y_1 \leq y_1, \ldots, Y_n \leq y_n] \leq \mathbb{P}[Y_1 \leq y_1] \cdots \mathbb{P}[Y_n \leq y_n].$$

In our context, we are interested in the case where $Y_i = h(Z_i^1, \ldots, Z_i^d)$, where the vectors Z_i, $i = 1, \ldots, n$ are LH samples of a Uniform$([0, 1]^d)$ random vector Z. In that setting, one situation where the Y_i are negatively dependent is when the function h is monotone in each

argument when the other arguments are held fixed (we refer again to Jin, Fu, and Xiong 2003).

Unfortunately, in this case it is not clear whether the Gartner-Ellis Theorem can be applied — the reason being that we do not know if negative dependence suffices to ensure convergence of the functions $\{\phi_n^{LHS}(\theta)\}$. We must note, however, that the Gartner-Ellis Theorem only provides *sufficient* (but not necessary) conditions for the validity of a large deviation principle; that is, it is possible that a large deviation principle holds in the present case even if the assumptions of the theorem are violated. A definite answer to that question is still an open problem.

Nevertheless, we can still provide results that are analogous to Theorems 3 and 5, i.e., comparing the upper bounds on deviation probabilities under LHS and under Monte Carlo. Jin, Fu, and Xiong (2003) show that, when the quantity to be estimated is a *quantile*, the upper bound on a deviation probability with negatively dependent sampling is less than that from Monte Carlo sampling. Here we show a similar result but in the context of estimation of the mean.

Theorem 6   *Suppose* $h(z) : [0, 1]^d \mapsto \mathbb{R}$ *is a function that preserves negative dependence, and let* $\mu_d := \int_{[0,1]^d} h(z)dz$.

Let $I_{MC}(x)$ *and* $I_{LHS}(n, x)$ *denote the (non-asymptotic) functions defined in* (6) *respectively for Monte Carlo and for LHS. Then, for any sample size n and all x we have that* $I_{LHS}(n, x) \geq I_{MC}(x)$.

In particular, from (9) *we see that the upper bound for the probability of a large deviation is smaller under Latin Hypercube sampling than under Monte Carlo sampling for any sample size n.*

5   EXAMPLES

We now show examples comparing the probability of a large deviation under Latin Hypercube sampling and Monte Carlo sampling on four different functions. For each function, we generated both Monte Carlo and Latin Hypercube samples for various sample sizes n ($n = 50, 100, 500, 1000, 5000, 10000$). For each sampling method and each n, we calculated the numerical integral (i.e. the sample mean). We then repeated this for 1000 replications. A large deviation was determined to be any value of the sample mean that was not within 0.1% of the true mean. The estimate of the probability of a large deviation is just the number of large deviations divided by 1000.

In each graph below, the x-axis represents the different sample sizes while the y-axis shows the estimated large deviations probabilities for each sample size. Estimates for both Latin Hypercube and Monte Carlo sampling are graphed as well as the upper and lower 95% confidence intervals for each estimate (represented by the dashed lines).

Example 1:

$$h(z) = \log(\frac{1}{\sqrt{z_1}})$$

This is a one-dimensional function with a singularity at $z_1 = 0$. Its integral on $[0, 1]$ is equal to $\frac{1}{2}$. Latin Hypercube sampling considerably outperforms Monte Carlo sampling with a large deviation probability of essentially zero by the time $n = 5000$. Meanwhile the probability of a large deviation is still roughly 90% for Monte Carlo sampling with $n = 10000$. This is shown in Figure 1.
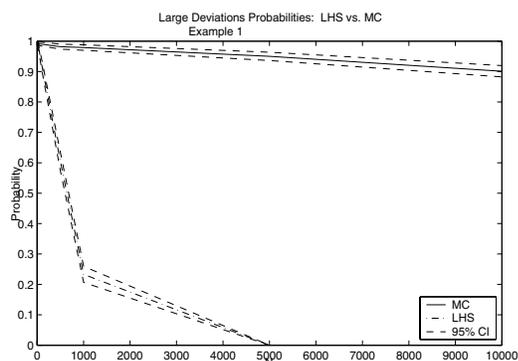
Figure 1: $h(z) = \log(\frac{1}{\sqrt{z_1}})$

Example 2:

$$h(z) = \log(z_1 z_2 z_3 z_4 z_5)$$

This function is separable, so by Theorem 4 we expect the large deviation probability to be essentially zero under Latin Hypercube sampling with large n. The integral of the function is -5. Again the Latin Hypercube sampling dominates the Monte Carlo sampling which has a large deviations probability of nearly 80% at $n = 10000$. This is shown in Figure 2.
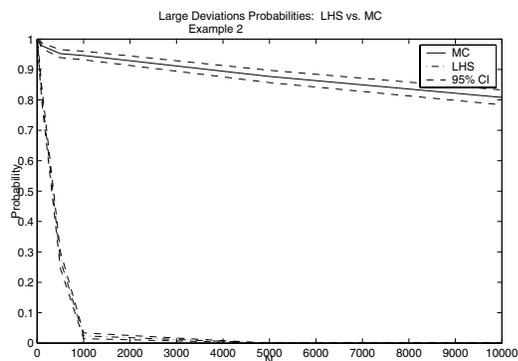
Figure 2: $h(z) = \log(z_1 z_2 z_3 z_4 z_5)$

Example 3:

$$h(z) = \log(\frac{1}{\sqrt{z_1}} + \frac{1}{\sqrt{z_2}})$$

While not separable, this function is negative dependent because it is monotone in both z_1 and z_2. Its integral is $\frac{5}{4}$. From Theorem 6, we know that the upper bound for the large deviations probability is guaranteed to be smaller under Latin Hypercube sampling than under Monte Carlo sampling for each value of n, and indeed we see that Latin Hypercube sampling again dominates Monte Carlo sampling. This is shown in Figure 3.
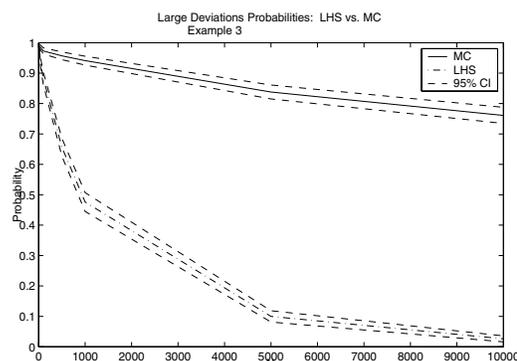
Figure 3: $h(z) = \log(\frac{1}{\sqrt{z_1}} + \frac{1}{\sqrt{z_2}})$

Example 4:

$$h(z) = \log\left[2 + \sin(2\pi z_1)\cos(2\pi z_2^2)\right]$$

This function is neither separable nor negative dependent — in fact, it is highly non-separable. We have no guarantee that Latin Hypercube sampling will produce a lower probability of a large deviation than Monte Carlo sampling. This function has integral 0.653212638. From the run, we see that it is possible for Monte Carlo sampling to have a lower probability of large deviation than Latin Hypercube sampling, even at $n = 10000$. In fact, the two sampling methods give similar results for this function. This is shown in Figure 4.
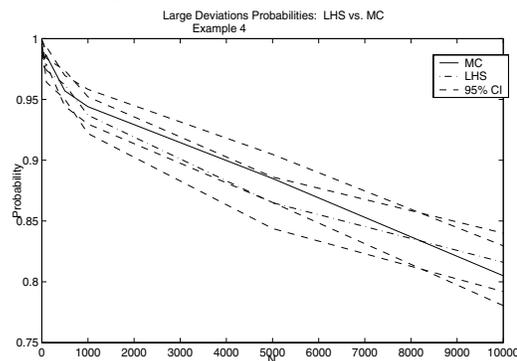
Figure 4: $h(z) = \log\left[2 + \sin(2\pi z_1)\cos(2\pi z_2^2)\right]$

6 CONCLUSIONS

In this paper, we have shown that Latin Hypercube sampling satisfies a large deviation principle for real-valued functions of one variable and for separable real-valued functions in multiple variables. We have also shown that the upper bound of the probability of a large deviation is smaller under LHS than it is for Monte Carlo sampling in these cases regardless of the sample size. This is analogous to the result that Latin Hypercube sampling gives a smaller variance than Monte Carlo sampling in these same cases since Var_{LHS} approaches the variance of the residual term, which in these cases is nonexistent. Further, as the number of samples becomes large, the probability of a large deviation from the true mean is essentially zero under Latin Hypercube sampling.

We have also shown that, if the underlying function preserves negative dependence (e.g., a function which is monotone in each component), then the upper bound for the large deviation probability is again less than that of Monte Carlo sampling regardless of the sample size. Again, this is analogous to the fact that the variance from LHS is no greater than that of Monte Carlo sampling when the function is monotone in all arguments. Unfortunately we do not know whether the large deviations rate is infinite, as it is in the separable case.

Large deviations results for LHS for general functions still remain to be shown, though the Latin Hypercube variance results found in the literature seem to provide a good direction. In general, the variance of a Latin Hypercube estimate may not be smaller than that of a Monte Carlo estimate (recall the bound $\text{Var}_{\text{LHS}} \leq \frac{n}{n-1}\text{Var}_{\text{MC}}$ proven by Owen (1997)); however, asymptotically it is no worse. This might also be the case for the upper bound of the large deviations probability. Also, Stein (1987) has shown that asymptotically, Var_{LHS} is equal to just the variance of the residual term. In the separable function case, the upper bound for the large deviation probability is zero, which is also the variance of the residual term. This suggests that the rate of convergence of large deviations probabilities for LHS may depend only on the residual terms. Further research on this topic is underway.

REFERENCES

Bucklew, J. A. 2004. *Introduction to rare event simulation*. New York: Springer-Verlag.

Dembo, A., and O. Zeitouni. 1998. *Large deviations techniques and applications*. 2nd. ed. New York, NY: Springer-Verlag.

den Hollander, F. 2000. *Large deviations*. Number 14 in Fields Institute Monographs. Providence, RI: American Mathematical Society.

Drew, S. S., and T. Homem-de-Mello. 2005. Some large deviations results for latin hypercube sampling. Manuscript, Department of Industrial Engineering and Management Sciences, Northwestern University.

Hoshino, N., and A. Takemura. 2000. On reduction of finite sample variance by extended latin hypercube sampling. *Bernoulli* 6 (6): 1035–1050.

Jin, X., M. C. Fu, and X. Xiong. 2003. Probabilistic error bounds for simulation quantile estimators. *Management Science* 49 (2): 230–246.

Law, A. M., and W. D. Kelton. 2000. *Simulation modeling and analysis*. 3rd. ed. New York, NY: McGraw-Hill.

Loh, W. 1996. On latin hypercube sampling. *The Annals of Statistics* 24 (5): 2058–2080.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21:239–245.

Owen, A. B. 1992. A central limit theorem for latin hypercube sampling. *J. Roy. Statist. Soc. Ser. B* 54:541–551.

Owen, A. B. 1997. Monte Carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.* 34 (5): 1884–1910.

Shwartz, A., and A. Weiss. 1995. *Large deviations for performance analysis*. Stochastic Modeling Series. London: Chapman & Hall. Queues, communications, and computing, With an appendix by Robert J. Vanderbei.

Stein, M. L. 1987. Large sample properties of simulations using latin hypercube sampling. *Technometrics* 29:143–151.

AUTHOR BIOGRAPHIES

SHANE S. DREW is a graduate student in the Department of Industrial Engineering and Management Sciences at Northwestern University. His e-mail address is <s-drew@northwestern.edu>.

TITO HOMEM-DE-MELLO is an Associate Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He received his PhD degree in Operations Research from the Georgia Institute of Technology in 1998. His research interests include stochastic optimization, simulation analysis methodology, and stochastic models for revenue management. His e-mail and web addresses are <tito@northwestern.edu> and <http://users.iems.northwestern.edu/~tito/>.