TECHNOMETRICS, MAY 1987, VOL. 29, NO. 2

Large Sample Properties of Simulations Using Latin Hypercube Sampling

Michael Stein

Department of Statistics
University of Chicago
Chicago, IL 60637

Latin hypercube sampling (McKay, Conover, and Beckman 1979) is a method of sampling that can be used to produce input values for estimation of expectations of functions of output variables. The asymptotic variance of such an estimate is obtained. The estimate is also shown to be asymptotically normal. Asymptotically, the variance is less than that obtained using simple random sampling, with the degree of variance reduction depending on the degree of additivity in the function being integrated. A method for producing Latin hypercube samples when the components of the input variables are statistically dependent is also described. These techniques are applied to a simulation of the performance of a printer actuator.

KEY WORDS: Variance reduction; Sampling with dependent random variables; Rank procedure; Exchangeability.

1 INTRODUCTION

Suppose we have some device or process the behavior of which depends on a random vector $\mathbf{X} = (X_1, \ldots, X_K)$ of fixed length K. For example, consider an electrical circuit the performance of which depends on a number of quantities (capacitances, resistances) that vary from circuit to circuit in some random fashion. A mathematical model for the device is developed (e.g., a set of differential equations) from which we can simulate the behavior of the device on a computer. Very often, we want to estimate the expected value of some measure of performance of the device, given by the function $h(\mathbf{X})$. Thus we have the problem of approximating the expected value of some function. If K, the number of variables, is large, then deterministic methods are difficult to use. Hence, Monte Carlo methods are usually used for high-dimensional problems. That is, N values of the input random vector, $\mathbf{X}_1, \ldots, \mathbf{X}_N$, are generated in some fashion such that the expected value $Eh(\mathbf{X})$ can be estimated by

$$\bar{h} = N^{-1} \sum_{j=1}^{N} h(\mathbf{X}_j). \tag{1}$$

Since $h(\mathbf{x})$ may be difficult to compute for each new value of $\mathbf{x}$ [we may have to solve numerically a large system of differential equations to obtain $h(\mathbf{x})$], it is important to pick a sampling scheme that allows us to estimate $Eh(\mathbf{X})$ well while keeping N, the number of simulations, to a minimum. Many methods for choosing $\mathbf{X}_1, \ldots, \mathbf{X}_N$ exist. The simplest is to generate N iid random vectors with the distribution of $\mathbf{X}$, a method that I shall refer to as *simple random sampling*. McKay, Conover, and Beckman (1979) suggested an alternative method of generating $\mathbf{X}_1, \ldots, \mathbf{X}_N$ that they call Latin hypercube sampling. In Section 2, I describe this procedure. In Section 3, I derive the asymptotic variance of $\bar{h}$, the estimator of $Eh(\mathbf{X})$, based on a Latin hypercube sample. I find that as long as N, the number of simulations, is large compared with K, the number of variables, Latin hypercube sampling gives an estimator with lower variance than simple random sampling for any function $h(\cdot)$ having finite second moment. Moreover, the closer $h(\mathbf{X})$ is to additive in the components of $\mathbf{X}$, the more Latin hypercube sampling helps relative to simple random sampling. I also show that $\bar{h}$ based on a Latin hypercube sample is asymptotically normal as N increases. In Section 4, I briefly consider estimating the variance of $\bar{h}$ when using Latin hypercube sampling. In Section 5, I give a method for producing Latin hypercube samples when the components of $\mathbf{X}$ are statistically dependent such that $\mathbf{X}_1, \ldots, \mathbf{X}_N$ have approximately the correct joint distribution for their components. In Section 6, I apply these methods to a model for a printer actuator.

2. LATIN HYPERCUBE SAMPLING

Suppose that the joint distribution of the random vector of parameters $\mathbf{X}$ is given by F. Throughout

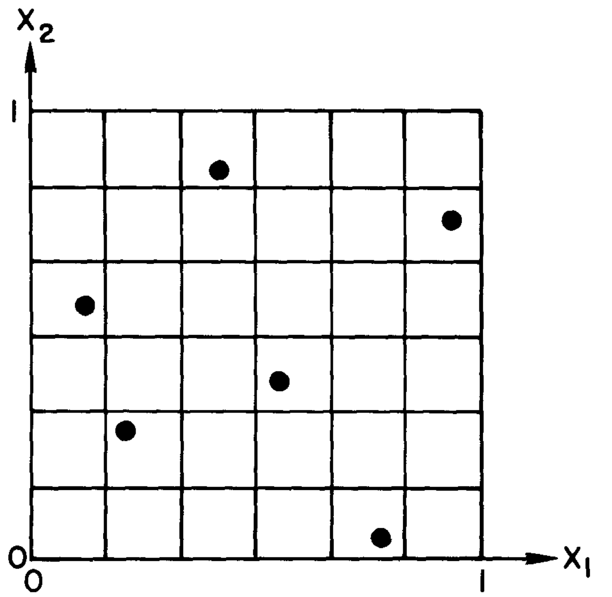

Figure 1. A Latin Hypercube Sample With $N = 6$, $K = 2$ for **X** *Distributed Uniformly on the Unit Square.*

this work, we will assume that F is specified. For now, we will also assume that the components of $\mathbf{X}$ are independent. Denote by F_k the cumulative distribution function of X_k, and let X_{jk} be the kth component of $\mathbf{X}_j$, the jth simulated value.

I now describe the procedure for producing a Latin hypercube sample of size N as given by McKay et al. (1979). Define $P = (p_{jk})$ to be an $N \times K$ matrix, where each column of P is an independent random permutation of $\{1, 2, \ldots, N\}$. Moreover, let ξ_{jk} ($j = 1, \ldots, N$; $k = 1, \ldots, K$) be NK iid $U[0, 1]$ (uniformly distributed on $[0, 1]$) random variables independent of P. Then X_{jk} is defined by

$$X_{jk} = F_k^{-1}(N^{-1}(p_{jk} - 1 + \xi_{jk})). \tag{2}$$

An example of a Latin hypercube sample when $\mathbf{X}$ is uniformly distributed on the unit square is shown in Figure 1. We see that $p_{j1}, \ldots, p_{jK}$ determine in which "cell" $\mathbf{X}_j$ is located, and $\xi_{j1}, \ldots, \xi_{jK}$ determine where in the cell $\mathbf{X}_j$ is located. Note that there is exactly one observation in each row in Figure 1. Roughly speaking, Latin hypercube sampling stratifies each marginal distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_N$ as much as possible but otherwise picks the $\mathbf{X}_i$'s randomly.

3. THE ASYMPTOTIC DISTRIBUTION BASED ON LATIN HYPERCUBE SAMPLING

I now compare the variance of $\bar{h}$, our estimator of $Eh(\mathbf{X})$ [Eq. (1)], depending on whether simple random sampling or Latin hypercube sampling is used. If simple random sampling is used to produce $\mathbf{X}_1, \ldots, \mathbf{X}_N$, then the estimator is unbiased and

$$\operatorname{var}(\bar{h}) = N^{-1} \operatorname{var}(h(\mathbf{X})).$$

If Latin hypercube sampling as described in Section 2 is used, then $\bar{h}$ is still unbiased, and

$$\operatorname{var}(\bar{h}) = N^{-1} \operatorname{var}(h(\mathbf{X})) + N^{-1}(N - 1)\operatorname{cov}(h(\mathbf{X}_1), h(\mathbf{X}_2)) \tag{3}$$

(McKay et al. 1979, p. 245). Thus Latin hypercube sampling lowers the variance if and only if $\operatorname{cov}(h(\mathbf{X}_1), h(\mathbf{X}_2)) < 0$. Iman and Conover (1980) gave various exact expressions for the variance given in (3), but these expressions are difficult to apply in practice. McKay et al. (1979) showed that this covariance is negative whenever $h(\mathbf{x})$ is monotonic in each of its components. In many problems, monotonicity will not hold. For example, if $h(\mathbf{x})$ is an indicator of failure in an electrical circuit, then since failures will tend to occur when one or more of the parameters take on very high or low values, $h(\mathbf{x})$ will usually not be monotonic. I will show, however, that as $N \to \infty$, the covariance term is asymptotically nonpositive. Define

$$g_k(x_k) = \int h(\mathbf{x}) \prod_{\substack{i=1 \\ i \neq k}}^{K} dF_i(x_i).$$

Moreover, let $\{\mathbf{X}_{jN}\}$ ($j = 1, \ldots, N$; $N = 1, 2, \ldots$) be an infinite triangular array of random vectors such that $\mathbf{X}_{1N}, \ldots, \mathbf{X}_{NN}$ is a Latin hypercube sample as defined in Section 2. Then we have the following:

Theorem 1. If $Eh^2 < \infty$, then as $N \to \infty$,

$$\operatorname{cov}(h(\mathbf{X}_{1N}), h(\mathbf{X}_{2N})) = KN^{-1}(Eh)^2 - N^{-1} \sum_{k=1}^{K} \int g_k(x)^2 \, dF_k(x) + o(N^{-1}). \tag{4}$$

The proof is in Appendix A. By Jensen's inequality,

$$\int g_k(x)^2 \, dF_k(x) \geq \left(\int g_k(x) \, dF_k(x) \right)^2 = (Eh)^2;$$

thus the highest-order term in the expansion of $\operatorname{cov}(h(\mathbf{X}_{1N}), h(\mathbf{X}_{2N}))$ is nonpositive. That is,

$$\lim_{N \to \infty} N \operatorname{cov}(h(\mathbf{X}_{1N}), h(\mathbf{X}_{2N})) \leq 0.$$

We see that, for any square integrable $h(\mathbf{X})$, Latin hypercube sampling does at least as well asymptotically as simple random sampling. We can write Theorem 1 in a more interpretable form. Define

$$h_a(\mathbf{x}) = \sum_{k=1}^{K} g_k(x_k) - (K - 1)Eh \tag{5}$$

and

$$r(\mathbf{x}) = h(\mathbf{x}) - h_a(\mathbf{x}). \tag{6}$$

The function $h_a(\mathbf{x})$ is the best additive fit to $h(\mathbf{x})$; that is,

$$\int r^2(\mathbf{x})\, dF(\mathbf{x}) \le \int \Big(h(\mathbf{x}) - \sum_{k=1}^{K} h_k(x_k)\Big)^2\, dF(\mathbf{x})$$

for any set of univariate functions $h_1, \ldots, h_K$. A proof is given in Appendix B. Then we have the following:

Corollary 1. Define $\mathbf{X}_{jN}$ as in Theorem 1. As $N \to \infty$,

$$\operatorname{var}\Big(N^{-1} \sum_{j=1}^{N} h(\mathbf{X}_{jN})\Big) = N^{-1} \int r(\mathbf{x})^2\, dF(\mathbf{x}) + o(N^{-1}). \tag{7}$$

By using Latin hypercube sampling, we essentially filter out the additive component of $h(\mathbf{x})$. Furthermore, for large N, we do no worse with the nonadditive part of $h(\mathbf{x})$ [the function $r(\mathbf{x})$] than we would with simple random sampling. Thus we see that the closer $h(\mathbf{x})$ is to additive, the more Latin hypercube sampling will help. An important aspect of this theorem is that it is relevant whenever N is much larger than K. Appendix A shows that, in a reasonable sense, the error in Corollary 1 is small relative to K/N. This appendix also shows that under additional conditions, the error in Corollary 1 is, in a reasonable sense, $O((K/N)^2)$.

In Appendix A, under the additional condition that $h(\mathbf{X})$ has a finite fourth moment, I show that $\bar{h}$ is asymptotically normal as $N \to \infty$. That is, we have the following:

Theorem 2. If $Eh(\mathbf{X})^4 < \infty$, then as $N \to \infty$

$$N^{1/2}\Big(N^{-1} \sum_{j=1}^{N} h(\mathbf{X}_{jN}) - Eh(\mathbf{X})\Big) \xrightarrow{\mathscr{L}} N\Big(0, \int r(\mathbf{x})^2\, dF(\mathbf{x})\Big),$$

where $\xrightarrow{\mathscr{L}}$ means converges in distribution.

4. ESTIMATION OF VARIANCE

Along with the estimator $\bar{h}$, it is usually important to produce some estimate of error of $\bar{h}$. If simple random sampling is used, then a consistent estimate of the var $\bar{h} = N^{-1}$ var $h(\mathbf{X})$ is obtained by using the sample variance divided by N. If we use this estimate with a Latin hypercube sample, we will essentially still estimate N^{-1} var $h(\mathbf{X})$, and not the variance of the estimator. We can use the fact that for N large,

$$\operatorname{var}(\bar{h}) \approx N^{-1} \int r(\mathbf{x})^2\, dF(\mathbf{x})$$

for a Latin hypercube sample to produce estimates of the variance. One possibility is to approximate $h_a(\mathbf{x})$, the best additive fit to $h(\mathbf{x})$ [see (5)] by a regression equation. Some preliminary investigations into such estimators have been done (details available from me). Unfortunately, the performance of these estimators is uncertain, so it is difficult to assess their usefulness.

A simple way to get an estimate of the variance of $\bar{h}$ is to produce several independent Latin hypercube samples and then estimate the variance using the sample variance between samples, a method called *replicated Latin hypercube sampling* by Iman and Conover (1980). More specifically, if we are planning to do N simulations, instead of producing one Latin hypercube sample of size N, we can produce α independent Latin hypercube samples of size M, where $M\alpha = N$. Define $\mathbf{X}_1^i, \ldots, \mathbf{X}_M^i$ to be the ith sample so that $(\mathbf{X}_1^1, \ldots, \mathbf{X}_M^1), \ldots, (\mathbf{X}_1^\alpha, \ldots, \mathbf{X}_M^\alpha)$ are α independent Latin hypercube samples of size M, and let

$$\bar{h}_i = M^{-1} \sum_{j=1}^{M} h(\mathbf{X}_j^i).$$

Then

$$[\alpha(\alpha - 1)]^{-1}\Big[\sum_{i=1}^{\alpha} \bar{h}_i^2 - \Big(\alpha^{-1} \sum_{i=1}^{\alpha} \bar{h}_i\Big)^2\Big] \tag{8}$$

is an unbiased estimator for the variance of $\alpha^{-1} \sum_{i=1}^{\alpha} \bar{h}_i$, our estimator of Eh. Of course, if α is small, then this estimator of the variance will be imprecise. By considering the second term in the asymptotic expansion of var $\bar{h}$ (App. A), we see that increasing α (and keeping N fixed) will increase the variance of our estimator of Eh. As long as MK^{-1} is large, however, the increase in variance will tend to be small. Thus if we had 20 variables and were planning to do 1,000 simulations, we might consider using 5 independent Latin hypercube samples of size 200, for which $MK^{-1} = 10$.

5. LATIN HYPERCUBE SAMPLING WITH DEPENDENT VARIABLES

The results of the previous sections all depend on the assumption that the components of $\mathbf{X}$ are independent. In many applications, strong dependencies among the components may exist. In this section, I introduce a procedure for producing Latin hypercube samples such that each sample vector has approximately the correct joint distribution when the sample size is large. In Section 6, another application of such a sampling scheme appears, one in which transforming variables to increase the effectiveness of Latin hypercube sampling is considered. In this case, the transformed variables are dependent even though the original variables are independent, so I need to produce Latin hypercube samples with dependent variables to implement this procedure.

Iman and Conover (1982) described a method for producing a Latin hypercube sample with rank correlation matrix of the sample approximately equal to a specified value. Although this procedure will tend to reproduce joint distributions more accurately than assuming the components are independent, the points produced under this scheme do not necessarily have even approximately the correct joint distribution for large N. For example, if $\mathbf{X} = (X_1 X_2)'$ and the conditional expectation of X_1 given X_2 is not monotonic in X_2, then using rank correlations to describe the joint distributions of the components would be highly inappropriate.

I now describe a procedure for producing a Latin hypercube sample of size N such that each sample vector has approximately the correct joint distribution when N is large. Assume that we are able to produce an iid sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ such that each $\mathbf{Y}_i$ has the correct joint distribution F. Now, make $\mathbf{Y}_i$ the ith row of an $N \times K$ matrix $\mathscr{Y}$; that is, let

$$\mathscr{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{pmatrix} . \tag{9}$$

For $k = 1, \ldots, K$, replace each element of the kth column of $\mathscr{Y}$ by its rank in the column (assume that each component of $\mathbf{Y}$ is continuous so that there are no ties), producing an $N \times K$ matrix of ranks R whose jkth element is denoted by r_{jk}. Obtain a Latin hypercube sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ by defining the kth component of $\mathbf{Z}_j$ by

$$Z_{jk} = F_k^{-1}(N^{-1}(r_{jk} + \xi_{jk} - 1)). \tag{10}$$

Although this method is based on ranks like the procedure by Iman and Conover (1982), it uses more than just the rank correlation structure and thus is able to reproduce the joint distribution more faithfully (see point 6 following). This procedure has the following properties:

1. It is easy to implement. The only problem might be in the last step, in which we need to know F_k^{-1}, the inverse cumulative distribution function of Y_k. When we cannot obtain a good analytic approximation to F_k^{-1}, however, we may be able to simulate it. When the cost of computing $h(\cdot)$ is much greater than the cost of computing $\mathbf{Y}$, then, with a relatively small increase in total computing time, we can obtain a good approximation of the cumulative distribution functions of the Y_k's by simulating a large number (much greater than N) of $\mathbf{Y}$'s and using the empirical marginal distributions to estimate the actual marginal distributions.

2. Each Z_{jk} has the correct distribution F_k.

3. If the components of $\mathbf{Y}$ are independent, then $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ will have the correct joint distribution; that is, they have the same joint distribution as a Latin hypercube sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ produced using the procedure given in Section 2.

4. By construction, the marginals of $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ are stratified. Thus if $h(T^{-1}(\mathbf{Y}))$ is exactly additive in the components of $\mathbf{Y}$, then

$$\operatorname{var}\left(N^{-1} \sum_{j=1}^{N} h(T^{-1}(\mathbf{Z}_j))\right) = o(N^{-1})$$

for any square integrable $h(\cdot)$.

5. $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ have the same ordering for every component as $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$, the iid sample on which the $\mathbf{Z}_j$'s are based. Thus, in particular, the rank correlations of the $\mathbf{Z}_j$'s are identical to the rank correlations of the $\mathbf{Y}_j$'s.

6. $\| \mathbf{Z}_j - \mathbf{Y}_j \| = o_p(1)$, where $\| \cdot \|$ indicates Euclidean distance and $o_p(\cdot)$ means of smaller order in probability (Cox and Hinkley 1974, p. 282). A proof is given in Appendix C. Thus $\mathbf{Z}_j$ must have approximately the correct joint distribution F, since $\mathbf{Y}_j \sim F$.

This last property is not nearly strong enough to obtain an asymptotic expansion for $\operatorname{var}(N^{-1} \sum_{j=1}^{N} h(T^{-1}(\mathbf{Z}_j)))$. Thus it is possible that in certain situations this estimator will have a substantially higher mean squared error (MSE) than an estimator based on a simple random sample of size N. In particular, if N is not large enough, then the joint distribution of $\mathbf{Z}_j$ may be substantially distorted, which could lead to bias problems. For example, if the range of $\mathbf{Y}$ is not rectangular (or, more precisely, a product space), then the $\mathbf{Z}_j$'s do not necessarily fall in the range of $\mathbf{Y}$.

6. APPLICATION TO A PRINTER ACTUATOR

In this section, I apply the techniques discussed in the previous sections to a model for an impact printer actuator. The actuator is made up of a permanent magnet, an armature coil, an electromagnet, and a spring. The motion of the armature of the printer actuator is modeled by the following set of differ-

Table 1. Distributions of Variables

Variable	Mean	Standard deviation
R (ohms)	61.3	2
L (millihenrys)	2.92	.2
m (grams)	.403	.04
B (gauss)	9.67	.6
K (newton/meter)	117.5	12
l (meters)	6.67	.3

Table 2. Comparison of Simple Random Sampling to Latin Hypercube Sampling

	Variance based on sample size of 100		
Function[a]	Simple random sampling	Latin hypercube sampling[b]	Mean
h^1_{IE}	.00183	.00143	.2318
h^2_{IE}	.00959	.00527	.6940
h^1_{CT}	.00129	.000447	.1425
h^2_{CT}	2.51×10^{-11}	7.91×10^{-12}	3.164×10^{-5}

[a] Function whose integral is being estimated. Functions are defined in Equation (19).

[b] Estimated by 100 independent replications of Latin hypercube samples of size 100.

ential equations:

$$m(d^2x/dt^2) = Bli - Kx - F_{\text{paper}}(x)$$
$$V = iR + L(di/dt) + Bl(dx/dt), \quad (11)$$

where x gives the displacement of the armature, t is time, m is the mass of the armature, B is the magnetic flux generated by the permanent magnet, l is the length of each conductor in the coil, i is the current, K is the spring constant, V is the external voltage applied to the actuator, R is the resistance, L is the inductance, and F_{paper} is the external force applied to the armature when it is in contact with the paper. Unlike the other terms, F_{paper} is highly nonlinear in x. This model was described in detail by Hendriks (1983). The two performance functions that we will consider are impact energy (in kiloergs) when the armature hits the paper and the length of time (in milliseconds) the armature is in contact with the paper. Both functions are functionals of the solution to the differential equation in (11), so it takes no more effort to compute the contact time and impact energy than to compute just the contact time. There are six random quantities, and I take them to be independent and normally distributed with means and standard deviations given in Table 1. The mean values are taken from Chen, Wang, and Zug (1984, fig. 6.4) and the standard deviations are hypothetical.

I am interested in estimating how much impact energy and contact time deviate from some desired value or how often they fall outside some desired range. Thus we will consider estimating the expected values of the following four functions:

$$\begin{aligned} h^1_{IE}(\mathbf{x}) &= 0 \quad \text{if } 12 \le \text{impact energy} \le 14 \\ &= 1 \quad \text{otherwise} \\ h^2_{IE}(\mathbf{x}) &= (\text{impact energy} - 13)^2 \\ h^1_{CT}(\mathbf{x}) &= 0 \quad \text{if } .155 \le \text{contact time} \le .17 \\ &= 1 \quad \text{otherwise} \\ h^2_{CT}(\mathbf{x}) &= (\text{contact time} - .162)^2. \end{aligned} \quad (12)$$

Comparing simple random sampling to Latin hypercube sampling, both with sample sizes of 100, we see that in all four cases, Latin hypercube sampling produces considerable reductions in variance (from 22% to 69%; see Table 2).

Impact energy and, to a lesser extent, contact time can be fit quite well by a linear combination of the six variables. Using this fact, I now describe a procedure for transforming the variables to make Latin hypercube sampling more effective. The idea is to transform the variables so that the functions whose expectations we are estimating are more nearly additive in the transformed variables than in the original variables. Ten iid random vectors, $\mathbf{X}_1, \ldots, \mathbf{X}_{10}$, were generated using a normal distribution with independent components and means and standard deviations given in Table 1. Impact energy and contact time were computed for each of these sample values, and these outputs were regressed on the six inputs, leading to the following least squares fits: impact energy $(\mathbf{X}) \approx a_{IE} + \mathbf{b}'_{IE}\mathbf{X}$; contact time $(\mathbf{X}) \approx a_{CT} + \mathbf{b}'_{CT}\mathbf{X}$.

Table 3. Comparison of Latin Hypercube Sampling Based on Original and Transformed Variables

	Original variables[b]		Transformed variables[c]	
Function[a]	Mean	Variance	Mean	MSE[d]
h^1_{IE}	.2318	.00143	.2298	.000302
h^2_{IE}	.6940	.00527	.7062	.00121
h^1_{CT}	.1425	.000447	.1496	.000510
h^2_{CT}	3.164×10^{-5}	7.91×10^{-12}	3.283×10^{-5}	1.06×10^{-11}

[a] Function whose integral is being estimated, defined as in (12).

[b] Sample means and variances based on 100 Latin hypercube samples of size 100.

[c] Defined as in (13). Fifty Latin hypercube samples of size 100 were obtained using the procedure described in Section 5 for producing Latin hypercube samples with dependent parameters.

[d] Mean squared error. Estimated by using the sample means based on the original parameterization (second column of this table) as the true mean.

We can then use the transformation

$$\mathbf{Y} = T(\mathbf{X}) = (\mathbf{b}'_{IE}\mathbf{X}\ \mathbf{b}'_{CT}\mathbf{X}\ X_1\ X_2\ X_3\ X_4). \quad (13)$$

If the preceding approximations for impact energy and contact time are good, we would have that $h^1_{IE}(\cdot)$ and $h^2_{IE}(\cdot)$ are largely determined by the first component of $\mathbf{Y}$ and $h^1_{CT}(\cdot)$ and $h^2_{CT}(\cdot)$ by the second component of $\mathbf{Y}$. Thus, by taking this transformation, we try to make the four functions nearly additive in the components of $\mathbf{Y}$. The first two components of $\mathbf{Y}$ are not independent, so we use the procedure for producing Latin hypercube samples with dependent variables described in Section 5. The marginals of $\mathbf{Y}$ are all normal, so the last step of this procedure, Equation (10), is easy to implement. Based on 50 samples of size 100, the bias and variance of the estimates of the four functions in (11) were obtained; the results are given in Table 3. There does not appear to be any substantial bias in the estimates of any of the four integrals, although there is evidence of slight upward bias in the estimates of Eh^1_{CT} and Eh^2_{CT}. We see that by using this sampling procedure, we obtained quite large reductions in MSE relative to using Latin hypercube sampling with the marginals of the original variables stratified for estimating Eh^1_{IE} and Eh^2_{IE}. There was a slight increase in MSE for estimating Eh^1_{CT} and Eh^2_{CT}, however, although the MSE was still less than that which is obtained by using simple random samples of size 100. These results are not surprising, because impact energy is very well fit by a linear model and contact time is fit less well. By taking the linear transformation of the variables, the additive components of $h^1_{IE}(\cdot)$ and $h^2_{IE}(\cdot)$ were apparently greatly increased, but the additive components of $h^1_{CT}(\cdot)$ and $h^2_{CT}(\cdot)$ were apparently not increased, or at least not increased enough to compensate for any errors introduced by the sampling scheme. Thus, based on a very small preliminary sample (sample size 10), I was able to obtain a transformation that substantially improved estimation of two of the four expected values of interest. Of course, these comparisons do not take into account the 10 initial simulations used to obtain the transformation. More important, there are no general guidelines as to when transforming will yield an improved estimator. As the gains can be dramatic, however, as in the cases of estimating Eh^1_{IE} and Eh^2_{IE}, this procedure deserves further study.

7. CONCLUSIONS

We have considered the use of Latin hypercube sampling for variance reduction in simulations of high dimensional integrals. By computing the asymptotic variance of an estimator based on Latin hypercube sampling, I showed that Latin hypercube sampling does reduce the variance relative to simple random sampling in a relevant asymptotic sense (Sec. 3 and App. A). The amount of variance reduction increases with the degree of additivity in the random quantities on which the function we are simulating depends. I have also given a method for producing Latin hypercube samples when the parameters are dependent that, for large sample sizes, gives approximately the correct joint distribution for each sample point (Sec. 5). These procedures were tested using a model for a printer actuator, with reductions in MSE relative to simple random sampling ranging from 22% to 87% (Tables 2 and 3).

ACKNOWLEDGMENT

The example of the printer actuator model and the program for solving the differential equation of its motion were provided by Peter Zug. Most of this work was done while the author was at the IBM T. J. Watson Research Center.

APPENDIX A: PROOFS OF THE THEOREMS

I outline the derivation of the asymptotic variance of $\bar{h}_N = N^{-1}\sum_{j=1}^{N} h(\mathbf{X}_{jN})$ as $N \to \infty$, where $\mathbf{X}_{1N}, \ldots, \mathbf{X}_{NN}$ is a Latin hypercube sample of size N defined as in Section 2. I also prove that $\bar{h}_N$ is asymptotically normal as $N \to \infty$. An additional result, Theorem 3, giving the next order term in the asymptotic expansion of the variance, is stated without proof. Details are available from me. Assume that the components of $\mathbf{X}$ are independent. Then we have the following results:

Theorem 1. If $Eh(\mathbf{X})^2 < \infty$, then as $N \to \infty$,

$$\operatorname{cov}(h(\mathbf{X}_{1N}), h(\mathbf{X}_{2N})) = KN^{-1}(Eh)^2 - N^{-1}\sum_{k=1}^{K}\int g_k^2(x)\, dF_k(x) + o(N^{-1}),$$

where F_k is the distribution of the kth component of $\mathbf{X}$ and

$$g_k(x_k) = \int h(\mathbf{x}) \prod_{\substack{i=1\\ i\neq k}}^{K} dF_i(x_i).$$

Corollary 1. Under the same conditions as Theorem 1,

$$\operatorname{var}(\bar{h}_N) = N^{-1}\int r(\mathbf{x})^2\, dF(\mathbf{x}) + o(N^{-1}),$$

where $r(\mathbf{x}) = h(\mathbf{x}) - \sum_{k=1}^{K} g_k(x_k) + (K-1)Eh$.

Under the additional condition that $h(\mathbf{X})$ has a finite fourth moment, $\bar{h}_N$ is asymptotically normal.

Theorem 2. If $Eh(\mathbf{X})^4 < \infty$, then as $N \to \infty$,

$$N^{1/2}(\bar{h}_N - Eh(\mathbf{X})) \xrightarrow{\mathscr{L}} N\left(0, \int r(\mathbf{x})^2\, dF(\mathbf{x})\right),$$

where $\xrightarrow{\mathscr{L}}$ means converges in distribution.

Now define

$$\Delta_k^N(y) = g_k(F_k^{-1}(y)) - N \int_{I_N(y)} g_k(F_k^{-1}(y))\, dy, \quad \text{(A.1)}$$

where $I_N(y)$ is the interval

$$I_N(y) = [(\alpha - 1)N^{-1}, \alpha N^{-1})$$

for the positive integer α satisfying $y \in [(\alpha - 1)N^{-1}, \alpha N^{-1})$. If the kth component of $\mathbf{X}$ has a continuous distribution, $\Delta_k^N(y)$ is equivalent to

$$\tilde{\Delta}_k^N(x) = g_k(x) - N \int_{I_N(F_k(x))} g_k(x)\, dF_k(x), \quad \text{(A.2)}$$

where $y = F_k(x)$. Then we have the following extension of Corollary 1:

Theorem 3. Under the same conditions as in Theorem 1,

$$\operatorname{var} \bar{h}_N = N^{-1} \int r(\mathbf{x})^2\, dF(\mathbf{x}) + N^{-1} \sum_{k=1}^{K} \int_0^1 (\Delta_k^N(y))^2\, dy$$

$$+ N^{-2} \sum_{k=2}^{K} \sum_{l=1}^{k-1} \int [g_{kl}(x_k, x_l) - g_k(x_k) - g_l(x_l) + Eh]^2$$

$$\times\, dF_k(x_k)\, dF_l(x_l) + o(N^{-2}),$$

where

$$g_{kl}(x_k, x_l) = \int h(\mathbf{x}) \prod_{\substack{i=1\\ i \neq k,l}}^{K} dF_i(x_i).$$

If the kth component of $\mathbf{X}$ has a continuous distribution, then we have by (A.2),

$$\int_0^1 (\Delta_k^N(y))^2\, dy = \int (\tilde{\Delta}_k^N(x))^2\, dF_k(x).$$

Corollary 1 essentially says that $\int (\Delta_k^N(y))^2\, dy = o(1)$, but it may in fact be as small as $O(N^{-2})$, as we shall see later.

To obtain Theorem 1, I first prove it in the case of $\mathbf{X}$ distributed uniformly on the K-dimensional unit hypercube and then extend it to the case of an arbitrary distribution with independent components. Define for $0 \leq z_1, z_2 < 1$,

$$r_N(z_1, z_2) = 1 \quad \text{if } [Nz_1] = [Nz_2]$$
$$= 0 \quad \text{otherwise,}$$

where $[x]$ is the greatest integer less than or equal to x. When $\mathbf{X}$ is uniformly distributed, the joint density of $(\mathbf{X}_{1N}, \mathbf{X}_{2N}) = (\mathbf{x}_1, \mathbf{x}_2)$ can be easily derived, and is given by

$$p(\mathbf{x}_1, \mathbf{x}_2) = [N/(N-1)]^K \prod_{k=1}^{K} (1 - r_N(x_{k1}, x_{k2}))$$

(see McKay et al. 1979). Then

$$\operatorname{cov}(h(\mathbf{X}_{1N}), h(\mathbf{X}_{2N}))$$

$$= \int h(\mathbf{x}_1)h(\mathbf{x}_2)[N/(N-1)]^K$$

$$\times \prod_{k=1}^{K} (1 - r_N(x_{k1}, x_{k2}))\, d\mathbf{x}_1\, d\mathbf{x}_2 - (Eh)^2$$

$$= [N/(N-1)]^K \int h(\mathbf{x}_1)h(\mathbf{x}_2)\, d\mathbf{x}_1\, d\mathbf{x}_2$$

$$- [N/(N-1)]^K \int h(\mathbf{x}_1)h(\mathbf{x}_2) r_N(x_{k1}, x_{k2})\, d\mathbf{x}_1\, d\mathbf{x}_2$$

$$- (Eh)^2 + O(N^{-2}),$$

where the remainder includes all lower-order terms in the expansion of the product; this equals

$$[KN^{-1} + O(N^{-2})](Eh)^2 - [1 + O(N^{-1})]$$

$$\times \sum_{k=1}^{K} \int h(\mathbf{x}_1)h(\mathbf{x}_2) r_N(x_{k1}, x_{k2})\, d\mathbf{x}_1\, d\mathbf{x}_2 + O(N^{-2}). \quad \text{(A.3)}$$

Define $I_{jN} = [(j-1)N^{-1}, jN^{-1})$. Then

$$\int h(\mathbf{x}_1)h(\mathbf{x}_2) r_N(x_{k1}, x_{k2})\, d\mathbf{x}_1\, d\mathbf{x}_2$$

$$= \int g_k(x_1) g_k(x_2) r_N(x_1, x_2)\, dx_1\, dx_2$$

$$= \sum_{j=1}^{N} \left(\int_{I_{jN}} g_k(x)\, dx\right)^2.$$

If g_k^2 is Riemann integrable, we immediately have

$$N \sum_{j=1}^{N} \left(\int_{I_{jN}} g_k(x)\, dx\right)^2 \to \int g_k(x)^2\, dx. \quad \text{(A.4)}$$

Substituting (A.4) into (A.3), we obtain Theorem 1 when each of the g_k's are square Riemann integrable. Corollary 1 is obtained by applying Theorem 1 to Equation (3). When g_k^2 is not Riemann integrable, Equation (A.4), and hence Theorem 1 and Corollary 1 are still valid as long as $Eh^2 < \infty$, which implies $\int g_k^2 < \infty$, for all k (details are available from me).

To obtain Theorem 1 for $\mathbf{X}$ with independent components but arbitrary marginal distributions, we merely replace $h(\mathbf{X})$ by $h^*(\mathbf{Y}) = h(F^{-1}(\mathbf{Y}))$, where $\mathbf{Y}$ is uniformly distributed; note that we can apply Theorem 1 to $h^*(\mathbf{Y})$, since $Eh(\mathbf{X})^2 < \infty$ implies that $Eh^*(\mathbf{Y})^2 < \infty$. Rewriting Theorem 1 in terms of $\mathbf{X}$

and $h(\cdot)$ instead of in terms of $\mathbf{Y}$ and $h^*(\cdot)$, we obtain the general case of Theorem 1.

Theorem 2 is proved using a central limit theorem for exchangeable random variables (Weber 1980). We have

$$\bar{h}_N = N^{-1}\sum_{j=1}^{N} h_a(\mathbf{X}_{jN}) + N^{-1}\sum_{j=1}^{N} r(\mathbf{X}_{jN}). \quad \text{(A.5)}$$

Applying Corollary 1 to $h_a(\mathbf{x})$, we obtain

$$N^{1/2}\left(N^{-1}\sum_{j=1}^{N} h_a(\mathbf{X}_{jN}) - Eh(\mathbf{X})\right) \xrightarrow{\mathscr{L}} 0, \quad \text{(A.6)}$$

since $h_a(\mathbf{x})$ is additive. By Theorem 1,

$$\operatorname{cov}(r(\mathbf{X}_{1N}), r(\mathbf{X}_{2N})) = o(N^{-1}). \quad \text{(A.7)}$$

Applying Theorem 1 to $r(\mathbf{x})^2$ and using $Eh(\mathbf{X})^4 < \infty$, we have

$$\operatorname{cov}(r(\mathbf{X}_{1N})^2, r(\mathbf{X}_{2N})^2) = O(N^{-1}).$$

Thus, as $N \to \infty$,

$$\begin{aligned} &E[r(\mathbf{X}_{1N})^2 r(\mathbf{X}_{2N})^2] \\ &\quad = \operatorname{cov}(r(\mathbf{X}_{1N})^2, r(\mathbf{X}_{2N})^2) + [Er(\mathbf{X}_{1N})^2]^2 \\ &\quad \to \left[\int r(\mathbf{x})^2\, dF(\mathbf{x})\right]^2. \end{aligned} \quad \text{(A.8)}$$

We also have that $r(\mathbf{X}_{1N}), \ldots, r(\mathbf{X}_{NN})$ are exchangeable (Weber 1980). By (A.7) and (A.8), the conditions of corollary 2 of Weber (1980) for the asymptotic normality of sums of exchangeable random variables are met, so we can conclude that

$$N^{-1/2}\sum_{j=1}^{N} r(\mathbf{X}_{jN}) \xrightarrow{\mathscr{L}} N\left(0, \int r(\mathbf{x})^2\, dF(\mathbf{x})\right). \quad \text{(A.9)}$$

Theorem 2 follows from (A.5), (A.6), and (A.9).

Equation (A.4) is equivalent to $\int_0^1 (\Delta_k^N(y))^2\, dy \to 0$ as $N \to \infty$. Comparing Theorem 3 with Corollary 1, we may conclude that Corollary 1 gives a good approximation to the variance whenever N is much larger than K. In fact, $\int_0^1 (\Delta_k^N(y))^2\, dy$ will, under additional conditions, converge to 0 at a faster rate. If $g_k(F_k^{-1}(y))$ has a bounded first derivative, then

$$\begin{aligned} \int_0^1 (\Delta_k^N(y))^2\, dy &= \sum_{j=1}^{N} \int_{I_{jN}} (\Delta_k^N(y))^2\, dy \\ &\le \sum_{j=1}^{N} \int_{I_{jN}} (y - (j + \tfrac{1}{2})N^{-1})^2 C^2\, dy \\ &= C^2/12N^2, \end{aligned}$$

where C is the bound on the absolute value of $(d/dy)g_k(F_k^{-1}(y))$. If F_k has a density f_k, then $(d/dy)g_k(F_k^{-1}(y))$ being bounded is equivalent to $((d/dx)g_k(x))/f_k(x)$ being bounded. Thus we can obtain the following corollary to Theorem 3:

Corollary 2. If $Eh(\mathbf{X})^2 < \infty$ and

$$((d/dx)g_k(x))/f_k(x) \text{ is bounded for all } k, \quad \text{(A.10)}$$

then

$$\begin{aligned} \operatorname{var} \bar{h}_N = N^{-1} &\int r(\mathbf{x})^2\, dF(\mathbf{x}) + N^{-2} \\ &\times \sum_{k=2}^{K}\sum_{l=1}^{k-1} \int [g_{kl}(x_k, x_l) - g_k(x_k) - g_l(x_l) + Eh]^2 \\ &\qquad \times dF_k(x_k)\, dF_l(x_l) + o(N^{-2}). \end{aligned}$$

I can show that the second-order term is less than $\frac{1}{2}N^{-2}K(K-1)\operatorname{var} h(\mathbf{X})$. Thus, in a reasonable sense the error in Corollary 1 is $O(K^2N^{-2})$ if (A.10) is satisfied. Note that this extra condition can be very strong if f_k has thin tails, such as with a normal distribution. If $(d/dx)g_k(x)$ is 0 outside some bounded interval, however, then the tail behavior of $f_k(x)$ will not cause (A.5) not to be satisfied. For example, suppose a device does not work at all if the kth component of $\mathbf{X}$ is outside some interval, no matter what values the other components of $\mathbf{X}$ take. If $h(\mathbf{X})$ is some constant for all values of $\mathbf{X}$ for which the device does not work at all, then $(d/dx)\, g_k(x)$ will be 0 for x outside this interval.

Based on Theorem 3, we can assess the effect on the variance of our estimator of Eh of using α samples of size M, where $M\alpha = N$, as suggested in Section 4. Considering α fixed and letting $M \to \infty$, we get the following (see Sec. 4 for notation):

$$\begin{aligned} &\operatorname{var}\left(N^{-1}\sum_{i=1}^{\alpha}\sum_{j=1}^{M} h(\mathbf{X}_j^i)\right) \\ &= \alpha^{-1}\Bigg[M^{-1}\int r(\mathbf{x})^2\, dF(\mathbf{x}) + M^{-1}\sum_{k=1}^{K}\int_0^1 (\Delta_k^M(y))^2\, dy \\ &\quad + M^{-2}\sum_{k=2}^{K}\sum_{l=1}^{k-1}\int [g_{kl}(x_k, x_l) - g_k(x_k) \\ &\quad - g_l(x_l) + Eh]^2\, dF_k(x_k)\, dF_l(x_l)\Bigg] + o(N^{-2}) \\ &= N^{-1}\int r(\mathbf{x})^2\, dF(\mathbf{x}) + N^{-1}\sum_{k=1}^{K}\int_0^1 (\Delta_k^M(y))^2\, dy \\ &\quad + (NM)^{-1}\sum_{k=2}^{K}\sum_{l=1}^{k-1}\int [g_{kl}(x_k, x_l) - g_k(x_k) \\ &\quad - g_l(x_l) + Eh]^2\, dF_k(x_k)\, dF_l(x_l) + o(N^{-2}). \end{aligned} \quad \text{(A.11)}$$

Now, since N is a multiple of M, I can show that

$$\int (\Delta_k^M(y))^2\, dy \ge \int (\Delta_k^N(y))^2\, dy.$$

Comparing (A.11) with Theorem 3, the highest-order term in the variance is unchanged by using $\alpha > 1$, but the two terms that are $o(N^{-1})$ and potentially not $o(N^{-2})$ are both increased by using $\alpha > 1$. Thus we lose some precision by choosing $\alpha > 1$, but if M is sufficiently large, the loss will be negligible.

APPENDIX B: BEST ADDITIVE FITS

I prove that $h_a(\mathbf{x})$, as defined in (5), provides the best additive fit to $h(\mathbf{x})$. For any set of functions $h_1, \ldots, h_K$,

$$\begin{aligned}
&\int \left(h(\mathbf{x}) - \sum_{k=1}^{K} h_k(x_k) \right)^2 dF(\mathbf{x}) \\
&= \int \left[(h(\mathbf{x}) - h_a(\mathbf{x})) + \left(h_a(\mathbf{x}) - \sum_{k=1}^{K} h_k(x_k) \right) \right]^2 dF(\mathbf{x}) \\
&= \int (h(\mathbf{x}) - h_a(\mathbf{x}))^2 \, dF(\mathbf{x}) \\
&\quad + \int \left(h_a(\mathbf{x}) - \sum_{k=1}^{K} h_k(x_k) \right)^2 dF(\mathbf{x}) \\
&\geq \int (h(\mathbf{x}) - h_a(\mathbf{x}))^2 \, dF(\mathbf{x}),
\end{aligned}$$

as required, where the second equality follows by noting that

$$\begin{aligned}
&\int \left(h_a(\mathbf{x}) - \sum_{k=1}^{K} h_k(x_k) \right)(h(\mathbf{x}) - h_a(\mathbf{x})) \, dF(\mathbf{x}) \\
&= \sum_{k=1}^{K} \int \left(g_k(x_k) - \frac{K-1}{K} Eh - h_k(x_k) \right) \\
&\quad \times \left(h(\mathbf{x}) - \sum_{j=1}^{K} g_j(x_j) + (K-1)Eh \right) dF(\mathbf{x}) \\
&= \sum_{k=1}^{K} \int \left(g_k(x_k) - \frac{K-1}{K} Eh - h_k(x_k) \right) \\
&\quad \times \left[\int \left(h(\mathbf{x}) - \sum_{j=1}^{K} g_j(x_j) + (K-1)Eh \right) \right. \\
&\quad \left. \times \prod_{i \neq k} dF_i(x_i) \right] dF_k(x_k) \\
&= \sum_{k=1}^{K} \int \left(g_k(x_k) - \frac{K-1}{K} Eh - h_k(x_k) \right) \\
&\quad \times [g_k(x_k) - g_k(x_k) - (K-1)Eh + (K-1)Eh] \, dF_k(x_k) \\
&= 0.
\end{aligned}$$

APPENDIX C: A PROPERTY OF THE DEPENDENT LATIN HYPERCUBE SAMPLING SCHEME

I now prove point 6 in Section 5. Using the definitions from that section,

$$\| \mathbf{Z}_j - \mathbf{Y}_j \|^2 = \sum_{k=1}^{K} [F_k^{-1}(N^{-1}(r_{jk} + \xi_{jk} - 1)) - Y_{jk}]^2. \tag{C.1}$$

For F continuous, conditional on Y_{jk}, $r_{jk} - 1$ has a binomial distribution with parameters $N - 1$ and $F_k(Y_{jk})$. It follows that, conditional on Y_{jk},

$$N^{-1}(r_{jk} - 1) - F_k(Y_{jk}) = O_p(N^{-1/2});$$

hence

$$N^{-1}(r_{jk} + \xi_{jk} - 1) - F_k(Y_{jk}) = O_p(N^{-1/2}).$$

Now, if F_k^{-1} is continuously differentiable, for any $\varepsilon > 0$, F_k^{-1} has a bounded first derivative on $[\varepsilon, 1 - \varepsilon]$. Thus, conditional on $\varepsilon \leq F_k(Y_{jk}) \leq 1 - \varepsilon$,

$$F_k^{-1}(N^{-1}(r_{jk} + \xi_{jk} - 1)) - Y_{jk} = O_p(N^{-1/2}).$$

Since ε is arbitrary, we have, unconditionally,

$$F_k^{-1}(N^{-1}(r_{jk} + \xi_{jk} - 1)) - Y_{jk} = o_p(1). \tag{C.2}$$

Point 6 follows from (C.1) and (C.2). Under stronger conditions on F(F_k^{-1} is boundedly differentiable for each k, for example), we would have $\| \mathbf{Z}_j - \mathbf{Y}_j \| = O_p(N^{-1/2})$.

[*Received June 1985. Revised October 1986.*]

REFERENCES

Chen, J. S-C., Wang, S., and Zug, P. (1984), "Optimal Tolerance Design of a Printer Actuator," *Transactions of the ASME—Journal of Mechanisms, Transmissions, and Automation in Design*, 106, 510–517.

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London: Chapman & Hall.

Hendriks, R. (1983), "Bounce and Chaotic Motion in Impact Printer Hammers," *IBM Journal of Research and Development*, 27, 273–280.

Iman, R. L., and Conover, W. J. (1980), "Small Sample Sensitivity Analysis Techniques for Computer Models, With an Application to Risk Assessment" (with discussion), *Communications in Statistics, Part A—Theory and Methods*, 9, 1749–1874.

——— (1982), "A Distribution Free Approach to Inducing Rank Correlations Among Input Variables," *Communications in Statistics, Part B—Simulation and Computation*, 11, 311–334.

McKay, M. D., Conover, W. J., and Beckman, R. J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code," *Technometrics*, 21, 239–245.

Weber, N. C. (1980), "A Martingale Approach to Central Limit Theorems for Exchangeable Random Variables," *Journal of Applied Probability*, 17, 662–673.