1998年度日本オペレーションズ・リサーチ学会
春季研究発表会

The Latin Hypercube Importance Sampling Method for Monte Carlo Simulation

01503140 Railway Technical Research Institute Hiroshi FUKUOKA

1. Introduction

Recently, a lot of efforts are devoted to the methods of Monte Carlo simulation, for example, Markov Chain Monte Calro methods for Baysian estimation.

In the risk (safety) assessment, the Laten Hypercube Sampling (LHS) Monte Carlo method is defaultly necessary. Because its important task, the uncertainty analysis, requires the estimation of the propagation of uncertainty, which is usually not analytically tractable, and even an ordinal Monte Carlo method can not be applied due to the high level of dimensionality. Only the LHS Monte Carlo methods can overcome the difficulty with its capability to reduce the variance of the output of the Monte Carlo simulation.

The LHS method, developed by McKay, Conover and Beckman[1], is an extension of stratifying method, and shows good performance in variance reduction, but that requires the stratification of the marginals with equal probabilities (equiprobability property), which means the necessity of the easily tractable inverse function of the marginal distribution function. Then especially, it is not directly applicable to the problem with correlated (dependent) variables.

To handle the dependency in LHS Monte Carlo simulation, the author introduce the idea of importance sampling method developed by Rubin [2]. Practically, this new method do not require the intractable integration of the density of multi-variate distribution function, so be applied to some large system problem.

2. Dependent Variables

The dependency in variables often appears in the Monte Carlo simulation. For example, in the uncertainty analysis, some uncertainty has the same origin for the specified group members, then some correlations within the same variables of uncertainty of such members are to be assumed. On the other hand, in some large system model of structual reliablity problem, one can find the dependency in stress variables and/or fragility variables of some adjacent components.

For the control of the correlations, Iman and Conover proposed dependence induction algorithm (ranked Cholesky (RC) method) [3] and the ranked Gram-Schmidt (RGS) algorithm was introduced by Owen[4]. But these methods can handle only low value of correlations. That is essentially due to the Cartesian stratification framework of LHS method itself.

In this paper, the author proposes the new methods which can introduce non-Cartesian stratifications and then can handle the high correlation problem.

3. Latin Hypercube Sampling(LHS) Method

We are interested in the class of estimators of the form

$$T(X_1,\cdots,X_n)=(1/N)\sum_{i=1}^{n} g(X_i) \qquad (1)$$

where $g(\cdot)$ = arbitrary function.

The estimator T represents the sample mean, which is an estimator of $E(X)$, for $g(X)=X$, r^{th} sample moment for $g(X)=X^r$ and empirical distribution function if $g(X)=1$ for $X \le x$, =0 otherwise.

For Latin Hypercube Sampling,

$$X_i^j = F_j^{-1}((\pi_j(i)-U_{ij})/n)$$
$$i=1,\ldots,n \quad j=1,\ldots,p \qquad (2)$$

where

$\pi_j(1),\ldots,\pi_j(n)$: random permutation of $i=1,\ldots,n$

and all $n!$ outcomes are equally probable,

U_{ij} : U (0,1) random variable

(these n p uniform variates are mutually independent) .

In this paper, we discuss the lattice version , in which U_{ij} in (2) is replaced by the constant value, 0.5 (Owen [4]).

4. Latin Hypercube Importance Sampling (LHIS) Method

The idea of importance-sampling method is introduced as follows:

Let the sample space S of $\mathbb{X}$ be partitioned into I disjoint strata S_i. Let Y be the indicator of the stara S_i, that is S_y for $Y=y$. Then we introduce the state space $(\mathbb{X},Y)$, then

$$f(\mathbb{X},Y)=f(\mathbb{X}),\ \text{if } \mathbb{X}\in S_y, Y=y$$
$$=0,\quad \text{otherwise.} \qquad (3)$$

Next, choose an importance-sampling distribution for Y, where its density

$f_S(Y)$ is an approximation of $f(Y)$, and $f_S(Y)$ has positive support wherever $f(Y)$ does.

The key is the selection of sample distribution. We have three variations:

(a) Cartesian(LHIS-C)

$$f_S(Y) = \prod_{j=1}^{p} f^j(Y_j), \quad (4)$$

in the case of little correlations between X_l's. To get Y, we can draw Y_1 from $f^1(Y_1)$, and next draw Y_2 from $f^2(Y_2)$, ..., Y_n from $f^n(Y_n)$ in the manner of the lattice version of (2). If $f(Y)$ can be represented as the same form as (4), that is an ordinary LHS, but one can control (reduce) the correlations by the importance formulation.

(b) step transition(LHIS-ST)

$$f_S(Y) = f^1(Y_1)\prod_{j=2}^{p} f^j(Y_j|Y_{j-1}), \quad (5)$$

in the case of low value of correlations between adjacent X_l's and little correlations for the others'. Y_j's are drawn in the same manner as (a). If Y_j's have identical distribution then we can use the same $f^j(Y_j|Y_{j-1})$.

(c) full transition(LHIS-FT)

$$f_S(Y) = f^1(Y_1)\prod_{j=2}^{p} f^j(Y_j|Y_{j-1},\cdots,Y_1), \quad (6)$$

for the highly correlated case.

Then $f(X|Y)*f_S(Y)$ provides an importance-sampling distribution for (X,Y). Therefore, calculate

$$r_l = \frac{f(X_l,Y_l)}{f(X_l|Y_l)*f_S(Y_l)} (l = 1,\ldots,n) \quad (7)$$

and estimate the marginal density for $f(X)$ by

$$f(\hat{X}) = \sum_{l=1}^{n} \frac{f(X|Y_l)\, r_l}{\sum_{i=1}^{n} r_i}. \quad (8)$$

5. Performance

Owen[4] examined the effectiveness of two algorithms, (1) ranked Cholesky (RC) method and (2) ranked Gram-Schmidt (RGS) algorithm for reducing off-diagonal correlations. He adopted the root mean square correlation (RMSC) among columns of $\mathbf{X}$ as a performance measure. We use the root mean square error of correlation (RMSEC) among columns of $\mathbf{X}$. If RMSEC is used to the uncorrelated case, then it reduced to RMSC.

Sample problems are (1) uncorrelated case for LHIS-C, (each marginal dist. $\sim N(0,1)$, $\rho_{lm} = 0$, for $l \neq m$, where ρ_{lm} : correlation between column i and column j of $\mathbf{X}$) (2) cases of low value correlation between adjacent components for LHIS-ST ($\rho_{lm} = c$:constant, for $m = l-1$ or $m = l+1$, $\rho_{lm} = 0$ otherwise) , (3) correlations are set to each components for LHIS-FT ($\rho_{lm} = c$: constant, for $l \neq m$).

Figure 1 shows RMSEC (RMSC) versus n for p = n - 1. The new methods shows rather good performance.

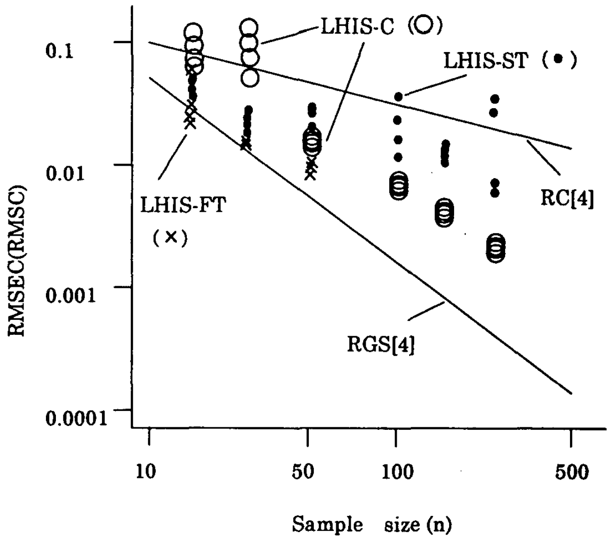

Fig. 1 RMSEC (RMSC) versus n for p = n - 1

(the values for RC and RGS algorithm from [4] are also shown : two lines).

6. Concluding Remarks

New methods for LHS with dependent variables are proposed. These methods allows LHS Monte Carlo simulations to be performed with high values of correlation.

[References] [1] McKey, M. D. et. al., Technometrics, 21, 239-245 (1979)[2] Rubin, D. B., JASA, 82, 543-546 (1987)[3] Iman ,R. L., et. al., Communication In Statistics Part B, Simulations and Computation, 11, 311-334 [4] Owen, A. B., JASA, 89, 1517-1522(1994)