ISSN No. 0976-5697
Volume 1, No. 4, Nov-Dec 2010
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

Some Advantages on Monte Carlo Integration using Variance Reduction Procedures

Behrouz Fathi Vajargah*
Department of Mathematics, Faculty of Science
University of Guilan
Iran
fathi@guilan.ac.ir

Farshid Mehrdoust
Department of Mathematics, Faculty of Science
University of Guilan
Iran
fmehrdoust@guilan.ac.ir

Ahmad Pourdarvish
Department of Mathematics, Faculty of Science
University of Mazandaran
Iran
apourdarvish@umz.ac.ir

Forough Norouzi Saziroud
Department of Mathematics, Faculty of Science
University of Mazandaran
Iran
foroughnorouzi@yahoo.com

Abstract: This paper presents the variance reduction methods for increasing the accuracy of Monte Carlo integration. Also, based on robust linear regression and multiple control variates method an alternative method is proposed.

Keywords: Antithetic variates(AV); Multiple control variates (MCV); Multivariate numerical integration; Naive Monte Carlo(NMC); Single control variates (SCV)

I. INTRODUCTION

The Monte Carlo method generally is an applicable way to find the solution of problems by stochastic method. The problem of evaluating of high dimension integrals is very important since it appears in many applications of control theory, statistical physics and mathematical economics. However, it is well known fact that the number of realization to be generated and deterministic problems to be solved often becomes very large and therefore the CPU time requirements may increase. In order to improve the efficiency of MC based on methods a variety of variance reduction techniques have been developed during the last decades [1, 3, 5]. Finding ways of constructing estimators with smaller variance can often lead to an improvement in the efficiency as well. The efficiency is a quality measure for an estimator that takes into account both their variance and computation time.

Now, consider the problem of estimating the multiply integral (1) over the region $D=[0,1]^d \subset R^d$
We may represent the integral

$$\mathrm{I}=\int_D f(\mathrm{X})P(\mathrm{X})d\mathrm{X} \tag{1}$$

As an expectation $\mathrm{E}[f(\mathrm{X})]$, such that $\mathrm{X}=(\mathrm{X}_1,...,\mathrm{X}_d)$ is a random vector with probability density function $P(\mathrm{X})$. Suppose that we have mechanism for drawing points $\{\mathrm{X}_i\}_{i=1,...,N}$ independently and sampled with density $P(\mathrm{X})$ from the region D. Evaluating the function f at N of these random points and averaging the results produces the MC estimate,

$$\mathrm{I}_N=\frac{1}{N}\sum_{i=1}^{N} f(\mathrm{X}_i) \tag{2}$$

By strong law of large numbers, as $N\to\infty$, we have

$$\mathrm{I}_N \to \mathrm{I}, (\mathrm{Pr}\,ob.1) \tag{3}$$

The quantity I_N is an estimator for I, therefore we can write

$$\mathrm{I}_N=\mathrm{I}+error \tag{4}$$

If the variance, σ_f^2, exists, the error appearing in the last statement is a random variable whose mean is zero and width is characterized for large N by

$$|error| \cong \frac{\sigma_f}{\sqrt{N}} \tag{5}$$

where

$$\sigma_f^2=\int_D f^2(\mathrm{X})P(\mathrm{X})d\mathrm{X}-\mathrm{I}^2 \tag{6}$$

It is shown that the error of the numerical integration depends on the smoothness of the integrand and for large dimensions d the convergence of the super convergent adaptive MC method goes asymptotically to $O(N^{-\frac{1}{2}})$, which corresponds to the convergence to the simple adaptive method [2]. The last equation shows that one way of improving the MC integration error is to try for reducing the variance σ^2 of the integrand f . The goal is to find another function Φ whose integral is equal to the integral of f but whose variance is smaller than of f . Methods that achieve this are called variance reduction techniques. Finding ways of constructing estimators with smaller variance can often lead to an improvement in the efficiency as well. The efficiency is a quality measure for estimators that takes into account both their variance and computation time. Before we begin our investigation of the most commonly used variance reduction techniques, we first briefly discuss the concept of efficiency.

Definition 1.

The efficiency of an estimator $\hat{\mu}$ for a quantity μ is given by

$$\mathrm{E}ff(\hat{\mu}) = [MSE(\hat{\mu}) \times C(\hat{\mu})]^{-1}$$

where $MSE(\hat{\mu}) = Var(\hat{\mu}) + [\mathrm{E}(\hat{\mu}) - \mu]^2$ is the mean-square error of $\hat{\mu}$ and $C(\hat{\mu})$ is the expected computation time for $\hat{\mu}$.

This definition implies that if we have two unbiased $\hat{\mu}_1$ and $\hat{\mu}_2$ that require the same computation time, then if $Var(\hat{\mu}_1) < Var(\hat{\mu}_2)$, we prefer $\hat{\mu}_1$ over $\hat{\mu}_2$.

II. VARIANCE REDUCTION METHODS

Generally, variance reduction technique may increase the accuracy of the estimator by a decreased sample standard deviation, instead of larger samples. Now, we study three methods for variance reduction. Here, each of the variance reduction techniques will mostly be discussing how and why they reduce the variance, but we also use numerical examples to compare the efficiency of the corresponding estimators with the MC method.

A. Single and multiple control variates

Suppose that we want to estimate $\theta = \mathrm{E}(\mathrm{Y})$ where $\mathrm{Y} = h(\mathrm{X})$ is the output it of a simulation experiment. Suppose that Z is also an output of the simulation or that we can easily output it if we wish. Then we can construct many unbiased estimators of θ such that

$$\hat{\theta}_c = \mathrm{Y} + c(\mathrm{Z} - \mathrm{E}[\mathrm{Z}]) \tag{7}$$

where c is some real number. It is clear that $\mathrm{E}[\hat{\theta}_c] = \theta$. We should choose the value of c such that to minimize $Var(\hat{\theta}_c)$. Using some calculation implies that the optimal value of c . For this purpose let

$$g(c) = \mathrm{var}(\hat{\theta}_c) = \mathrm{var}(\mathrm{Y}) + c^2\,\mathrm{var}(\mathrm{Z}) + 2c\,\mathrm{cov}(\mathrm{Y},\mathrm{Z}) \tag{8}$$

We have

$$\frac{dg(c)}{dc} = 2c\,\mathrm{var}(\mathrm{Z}) + 2\,\mathrm{cov}(\mathrm{Y},\mathrm{Z}) \tag{9}$$

By setting (9) to zero, we see that $c_0 = -\frac{\mathrm{cov}(\mathrm{Y},\mathrm{Z})}{\mathrm{var}(\mathrm{Z})}$, is critical point of g.

Since $\frac{d^2 g(c_0)}{d^2 c_0} > 0$, then c_0 is a local minimum of $g(c)$. On the other hand, c_0 is just a single extremum of g in its domain. Then c_0 is absolute minimum of g. Thus, with selecting c_0 the variance $\mathrm{var}(\hat{\theta}_c)$ minimizes, that is

$$\mathrm{var}(\hat{\theta}_{c_0}) = \mathrm{var}(\hat{\theta}) - \frac{\mathrm{cov}(\mathrm{Y},\mathrm{Z})^2}{\mathrm{var}(\mathrm{Z})} \le \mathrm{var}(\hat{\theta}). \tag{10}$$

We note that for achieving a variance reduction it is only necessary that $\mathrm{cov}(\mathrm{Y},\mathrm{Z}) \neq 0$. (We note that in practice $\mathrm{cov}(\mathrm{Y},\mathrm{Z})$ never is known and we thus to simulate it). In this case, the random variable Z is called a control variates for Y and also this method is known as single control variates (SCV) method.

Now, suppose that for $\mathrm{Z}_i, i = 1,\ldots,m$ is an output or that we can easily output it and we assume that for each i the value $\mathrm{E}[\mathrm{Z}_i]$ is known. Consider the following unbiased estimator of θ

$$\hat{\theta} = \mathrm{Y} + c_1(\mathrm{Z}_1 - \mathrm{E}[\mathrm{Z}_1]) + \ldots + c_m(\mathrm{Z}_m - \mathrm{E}[\mathrm{Z}_m]) \tag{11}$$

where each $c_i, i = 1,\ldots,n$ are real numbers. The $\hat{\theta}$ in (11) which is an unbiased estimator and an extended equation of SCV is called multiple control variates (MCV) estimator.

We must choose the c_i's so that $\hat{\theta}_{\hat{c}}$ has a lower variance than $\hat{\theta}$. From (11) we achieve that

$$g(c_1,c_2,\ldots,c_m) = \mathrm{var}(\hat{\theta}_{\hat{c}}) = \mathrm{var}(\mathrm{Y}) + 2\sum_{i=1}^{m} c_i\,\mathrm{cov}(\mathrm{Y},\mathrm{Z}_j) + \sum_{i=1}^{m}\sum_{j=1}^{m} c_i c_j\,\mathrm{cov}(\mathrm{Z}_i,\mathrm{Z}_j) \tag{12}$$

and so

$$\begin{pmatrix} \frac{\partial g}{\partial c_1} \\ \frac{\partial g}{\partial c_2} \\ \vdots \\ \frac{\partial g}{\partial c_m} \end{pmatrix} = \begin{pmatrix} \mathrm{var}(Z_1) & \mathrm{cov}(Z_1,Z_2) & \dots & \mathrm{cov}(Z_1,Z_m) \\ \mathrm{cov}(Z_1,Z_2) & \mathrm{var}(Z_2) & \dots & \mathrm{cov}(Z_2,Z_m) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{cov}(Z_1,Z_m) & \mathrm{cov}(Z_2,Z_m) & \dots & \mathrm{var}(Z_m) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} + \begin{pmatrix} \mathrm{cov}(Y,Z_1) \\ \mathrm{cov}(Y,Z_2) \\ \vdots \\ \mathrm{cov}(Y,Z_m) \end{pmatrix}$$

which can be written in matrix format as

$$\nabla g = \sum C + \mathrm{Y} \tag{13}$$

Now, we set

$$\nabla g = 0 \tag{14}$$

A way for estimating the c_i's is to observe that $\hat{c}_i = -\beta_i$, for each $i = 1,...,m$ and the β_i's are the solution to the following linear regression

$$Y = \beta_0 + \beta_1 Z_1 + ... + \beta_m Z_m + \varepsilon \tag{15}$$

where $\varepsilon \sim N(0,\sigma^2)$ is an error term.

B. Robust MCV

Here, we use the robust linear regression instead of usual linear regression for evaluating the coefficients in (15). It is well known that the least squares method is a simple computational technique for estimating a linear regression model, but its estimation may behave badly, when the distribution of ε does not follow normal distribution. Indeed, one single outlier can have an arbitrarily large effect on the least squares estimation. Robust linear regression is not sensitive respect to outlier and specially when the realization follow normal distribution, the efficiency of robust linear regression respect to least squares estimate is 90-95 percent. For example, suppose that ε has exponential distribution,

$$f(\varepsilon_i) = \frac{1}{2\sigma} e^{-\frac{|\varepsilon_i|}{\sigma}}, \qquad \varepsilon_i \in R \tag{16}$$

Based on MLE method we have

$$L(\beta_1,...,\beta_m) = \frac{1}{(2\sigma)^n} e^{-\frac{\sum|\varepsilon_i|}{\sigma}} \tag{17}$$

Maximizing the above equation is equivalent to minimizing $\sum|\varepsilon_i|$. Because, the distribution of errors, ε_i's have heavy tail the least squares method does not work very well. In this paper, we use the robust linear regression based on Huber estimators [2].

C. Antithetic variates

The method of antithetic variates attempts to reduce variance by introducing negative dependence between pairs of replication. In this technique, we generate two sample Y_1 and Y_2 and then let the unbiased estimator for θ as

$$\hat{\theta} = \frac{\mathrm{Y}_1 + \mathrm{Y}_2}{2}$$

Obviously, we see

$$\mathrm{var}(\hat{\theta}) = \frac{\mathrm{var}(\mathrm{Y}_1) + \mathrm{var}(\mathrm{Y}_2) + 2\,\mathrm{cov}(\mathrm{Y}_1,\mathrm{Y}_2)}{4} \tag{18}$$

Now, we could construct random variables $\mathrm{Y}_1, \mathrm{Y}_2$ such that $\mathrm{cov}(Y_1,Y_2) < 0$. Then we reduce the variance of our new unbiased estimator $\hat{\theta}$. For this aim, suppose that $U_1,U_2,...,U_n$ are a random sample of uniform distribution $U(0,1)$, and let

$$\mathrm{Y}_i = h(U_i) \qquad , \quad \hat{\mathrm{Y}}_i = h(1-U_i)$$

For monotone function (sufficient condition) h it con easily be shows that the covariance of Y_i and $\hat{\mathrm{Y}}_i$ is negative. Therefore the estimator, $\mathrm{Z}_i = \frac{\mathrm{Y}_i + \hat{\mathrm{Y}}_i}{2}$ is better then the usual estimator.

Theorem 1.

Let $f : [0,1) \times ... \times [0,1) \to R$ be a bounded and monotone function in each of arguments. Suppose also that f is not constant in the interior of its domain. If let $U = (U_1,...,U_n) \sim U([0,1) \times ... \times [0,1))$ and so that $\tilde{U} = (1-U_1,...,1-U_n) \sim U([0,1) \times ... \times [0,1))$ then we have $\mathrm{cov}(f(U), f(\hat{U})) < 0$.

Proof. See [4].

III. COMPUTATIONAL RESULTS

Here, we apply the variance reduction methods for calculating two examples. We can see in the following tables the deference of methods. Figures 4-5 and 1-3 show comparison of variance reduction method for θ_1 and θ_1.

Table 1: comparison between NMC method and reduced variance MC methods for calculating

$$\theta_1 = \int_0^1\int_0^1 \log|0.5-x||0.5-y|\,dxdy \ (\theta_1 = -3.3862, N = 10000)$$

Table I

Variance Reduction procedure	Solution	Rel. Error	Std. Dev.	CPU(Sec.)
NMC	-3.4076	0.0063	1.51	0.03
AV	−3.3681	0.0054	1.43	0.004
SCV	-3.3868	1.83e-4	0.85	0.04
MCV	-3.3880	5.25e-4	0.82	0.4
Robust MCV	−3.3865	7.37e-5	0.76	0.4

Table 2: comparison between NMC method and reduced variance MC methods for calculating

$$\theta_1 = \int_0^1\int_0^1 e^{-x^2-y^2}\,dxdy\,(\theta_1 = 0.5577, N = 1000)$$

Table II

Variance Reduction **procedure**	Solution	**Rel. Error**	**Std. Dev.**	**CPU(Sec.)**
NMC	0.5617	0.071	0.2145	0.04
AV	0.5580	3.955e-4	0.0499	0.0008
SCV	0.5570	0.0013	0.0487	0.03
MCV	0.5579	2.6665e- 4	0.0483	0.12
Robust MCV	0.5576	2.0678e-4	0.0476	0.12

IV. CONCLUDING REMARKS

In this paper, three different approaches to reduce NMC method error have been presented. We conclude that the variance reduction methods improve the accuracy of desired integral. As we can see from Fig. 1, Fig. 2 and Fig. 3 and comparing the standard deviation of usual MCV and AV together, the robust MCV has minimum standard deviation of robust MCV. This was expected to happen, since the condition, $\mathrm{cov}(f(U), f(\hat{U})) < 0$, in theorem 1 is not valid here.

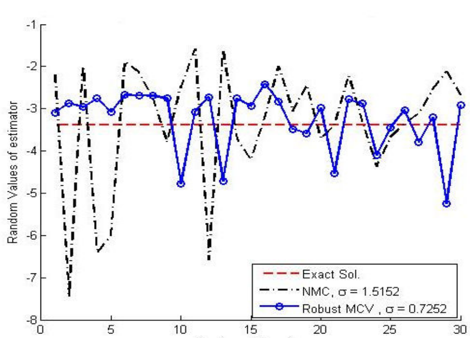

Figure 1 Comparison between NMC method and robust MCV method

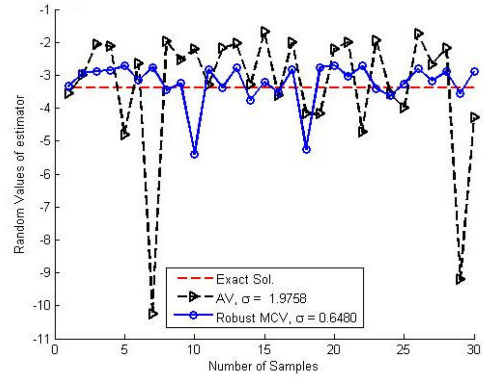

Figure 2 Comparison between AV method and robust MCV method

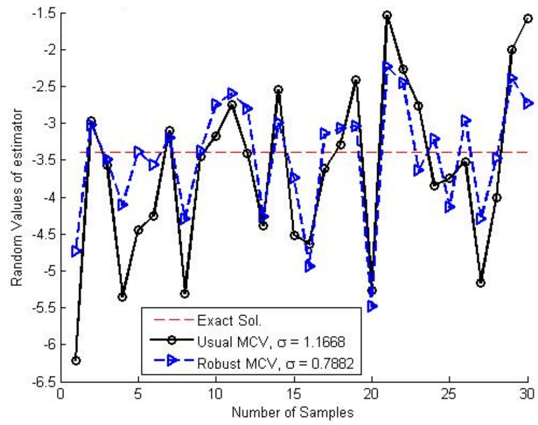

Figure 3 Comparison between usual MCV method and robust MCV method

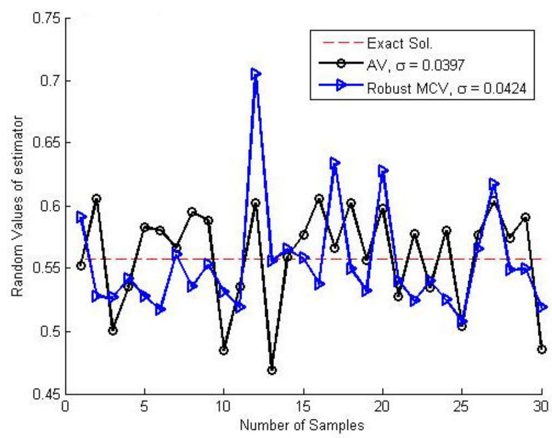

Figure 4 Comparison between AV method and robust MCV method

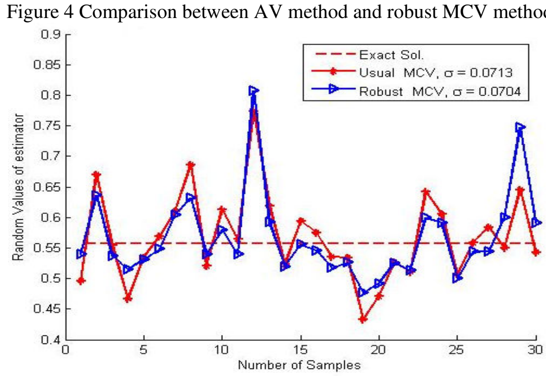

Figure 5 Comparison between usual MCV method and robust MCV method

V. REFERENCES

[1] A. Karaivanova, I. Dimov, Error analysis of an adaptiv Monte Carlo method for numerical integration, Mathematics and Coputers in Simulation, 1998, 47, 201-213.

[2] E. L. Lehmann, Some concepts of dependence. Annals of Mathematical Statistics, 1966,47,1137-1153.

[3] J. M. Hammersly, and D. C. Handscomb.Monte Carlo methods, Landon Methuen and Company, 1964.

[4] P. J. Huber, Robust statistics, John Wiley & Sons, 1981.

[5] R. Y. Rubinstein, Simulation and the Monte Carlo method, John Wiley & Sons, New York, 1981.