Proceedings of 2007 IEEE International Conference on Grey Systems
and Intelligent Services, November 18-20, 2007, Nanjing, China

An Importance Sampling Method with Applications to Rare Event Probability

Yue QIU, Hong ZHOU, and Yue-qin WU , *Fellow, IEEE*

***Abstract*—It usually takes long time to simulate rare event using traditional Monte Carlo method, while importance sampling techniques can effectively reduce the simulation time and improve simulation efficiency. A new implementation for importance sampling method to estimate rare event probability in simulation models is proposed in this paper, in which the classical exponential change of measure is adopted to construct the family of importance sampling distributions, and the optimal importance sampling distribution is obtained by minimizing the variance of importance sampling estimator. Numerical experiment has been conducted and the result indicates that the method can effectively estimate the rare event probability.**

I. Introduction

RARE event simulation has attracted extensive attentions since the concept was first proposed, but there haven't been many research results yet [1], [14]. Performances of modern systems, such as coherent reliability systems, inventory systems, insurance risk, storage systems, computer networks and telecommunications networks, are sometimes characterized by probabilities of rare events and usually studied by simulation [9],[13],[15],[16],[19],[20],[22]. Straightforward simulation for rare events requires a large number of trials and hard to implement because the occurrence of rare events are extremely little in a standard simulation [2], [4], [5], hence new methods are needed to be investigated and developed. One way to overcome this difficulty is to use importance sampling (IS). The main idea of IS is to make the occurrence of rare events more frequent by carrying out the simulation with a different probability distribution – what is the so-called change of measure (CM) – and to estimate the probability of interest via a corresponding likelihood ratio (LR) estimator[5],[14].

It is commonly known that the greatest difficulty in obtaining an efficient IS method is to find out the optimal importance sampling distribution function. It has been shown that, in many cases, the computation and experiments required for achieving an efficient IS method may be even more complicated than the original problem. Thus, this implies that we have to find simple and efficient IS algorithms. There are two main ways to solve this problem. One is to utilize Variance Minimization (VM) technique[21],[23]; the other is to use minimizing Cross Entropy (CE) technique[18],[19]. It is well known that there theoretically exists a CM that yields a zero-variance likelihood ratio estimator. However, in practice such an optimal CM cannot be computed since it depends on the underlying quantity being estimated. However, investigations have shown that the reduced variance method can be employed to approach the optimal solution in the sense of minimizing the variance [7], [10], [16]. In other words, minimizing estimator's variance with original density function is equivalent to minimizing the expected likelihood ratio conditioned on the rare event happening; and minimizing cross entropy is equivalent to minimizing the expected logarithm of likelihood ratio conditioned on the rare event happening. Therefore, minimizing cross entropy in a sense is close to, but definitely is different from minimizing estimator's variance. In terms of estimator's variance, cross entropy method does not seek for the optimal solution.

The aim is to select a CM that minimizes the variance of the likelihood ratio estimator in variance minimization. Prominent among the CMs is the exponential change of measure (ECM). Here, instead of the original PDF $f(x)$, the simulation is carried out with an "exponentially twisted" PDF $f_\theta(x) = c\exp(\theta x) f(x)$, where θ is called the twisting or tilting parameter and c is a normalizing constant. ECM often yields efficient and sometimes "optimal" IS estimates [11].

To reduce the number of simulation and improve the efficiency of sampling, a novel method of estimating the probability of the rare event is presented in this paper. We estimate the rare event probability via importance sampling, using the classical exponential change of measure, in which the importance sampling distribution is chosen from the same parametric family as the transformed distribution. And the optimal parameter vector of the importance sampling distribution is estimated by minimizing the variance of importance sampling estimator. The obtained numerical results suggest that the method can effectively estimates rare event probability.

II. Importance Sampling

IS is based on the idea to make the occurrence of rare events more frequent, or in other words, to speed up the simulation. Technically, IS aims to select a probability distribution that minimizes the variance of the IS estimate. The efficiency of IS depends on obtaining a good change of

Manuscript received May 18, 2007. This work was supported by the National Natural Science Foundation of China (Project no. 70531010, 70521001), and Program for New Century Excellent Talents in University of China

Yue QIU is with the School of Economics & Management, Beihang University Beijing 100083, P.R China (corresponding author to provide phone: +86-131-4129-7743; e-mail:qiuyueqq@sem.buaa.edu.cn.)

Hong ZHOU is with the School of Economics & Management, Beihang University (e-mail: h_zhou@buaa.edu.cn).

Yue-qin WU is with the School of Economics & Management, Beihang University (e-mail: wuyueqin01@sina.com).

measure. Large deviation results have proven their applicability for small systems with medium complexity [2], [6], [12].

Let $x=(X_1,\ldots,X_n)$ be a random vector taking values in some space Ω and $f(x)$ be a probability density function on Ω. Let s be some real-valued function on Ω. Suppose we are interested in the probability that $s(x)$ is greater than or equal to some real number r, this probability can be expressed as $l=P\{s(x)\geq r\}=EI_{(s(x)\geq r)}$. If this probability is very small, say smaller than 10^{-5}, we call $\{s(x)\geq r\}$ a rare event.

A straightforward way to estimate $l=P\{s(x)\geq r\}=EI_{(s(x)\geq r)}$ is to use crude Monte-Carlo simulation: Draw a random sample $X_1,\ldots,X_N$ from $f(x)$; then

$$\hat{l}=\frac{1}{N}\sum_{i=1}^{N} I_{\{s(X_i)\}}=\frac{1}{N}\sum_{i=1}^{N} I_{\{s(X_i)\geq r\}}. \tag{1}$$

is an unbiased estimator of l. However this poses serious problems when $\{s(x)\geq r\}$ is a rare event. In that case a large simulation effort is required in order to estimate l accurately, i.e., with small relative error or a narrow confidence interval.

An alternative is based on importance sampling: take a random sample $X_1,\ldots,X_N$ from an importance sampling density $g(x)$ on Ω, then

$$\begin{aligned} l &= \int_{\{s(x)\geq r\}} f(x)dx \\ &= \int_{\{s(x)\geq r\}} L(x)g(x)dx \end{aligned} \tag{2}$$

where $L(x)=\dfrac{f(x)}{g(x)}$ is called the likelihood ratio. It is well known that the best way to estimate l is to use the change of measure with density

$$g_{opt}(x)=\frac{I_{(s(x)\geq r)}f(x)}{r}. \tag{3}$$

The estimator (3) has zero variance, and we need to produce only N = 1 sample. The obvious difficulty is of course that this $g^*(x)$ depends on the unknown parameter, so it cannot be computed in practice. In estimate rare event probability: Draw a random sample $X_1,\ldots,X_N$ from $g(x)$, and then get estimator according to (2)

$$\hat{l}_{IS}=\sum_{i=1}^{R} I_{\{s(X_i)\geq r\}}L(X_i). \tag{4}$$

We will call the estimator in (4) the *importance sampling estimator.*

To obtain the optimal – or at least a good – change of measure from the original sampling distribution $f(x)$ to a new distribution $g(x)$, is the main challenge with respect to making importance sampling efficient and robust. From equation (2), the only restriction to the probability density $g(x)$ is that $g(x)>0$ for all samples x. This is a necessary condition which serves as a guideline for choosing $g(x)$. The efficiency of the importance sampling estimate in (4) is dependent on an appropriate choice of the new distribution. An unfortunate choice of $g(x)$ may cause the variance of (4) to be larger than the variance of the estimates obtained by direct simulation. For some $g(x)$, the variance may be infinite, and, hence, it is crucial to find a good $g(x)$.

III. IMPORTANCE SAMPLING METHOD BASED ON THE CROSS-ENTROPY

We briefly review some basics concepts about importance sampling method based on the cross-entropy. As discuses above, the obvious difficulty is of course that this $g_{opt}(x)$ depends on the unknown parameter l. Moreover, it is often convenient to choose a $g(x)$ in the family of densities $\{f(.;v)\}$. The idea now is to choose the parameter vector, called the reference parameter v such that the distance between the densities $g_{opt}(x)$ and $f(.;v)$ is minimal. A particular convenient measure of distance between two densities g and h is the Kullback-Leibler distance, which is also termed the cross-entropy between g and h. The Kullback-Leibler distance is defined as:

$$\begin{aligned} &D(g,h) \\ &= E_g \ln\frac{g(x)}{h(x)} \\ &= \int g(x)\ln g(x)dx-\int g(x)\ln h(x)dx \end{aligned}$$

We note that $D(g,h)$ is not a “distance” in the formal sense; for example, it is not symmetric.

Minimizing the Kullback-Leibler distance between $g_{opt}(x)$ and $f(.;v)$ is equivalent to choosing v such that $-\int g_{opt}(x)\ln f(x;v)dx$ is minimized, which is equivalent to solving the maximization problem

$$\max_v \int g_{opt}(x) \ln f(x;v) dx \quad (5)$$

Substituting $g_{opt}(x)$ from (3) into (5) we obtain the maximization program

$$\max_v \int \frac{I_{\{s(x)\geq r\}} f(x;u)}{l} \ln f(x;v) dx$$

which is equivalent to the program

$$\begin{aligned} &\max_v D(v) \\ &= \max_v E_u I_{\{s(x)\geq r\}} \ln f(x;v) \end{aligned} \quad (6)$$

Using again importance sampling, we can rewrite (6) as

$$\begin{aligned} &\max_v D(v) \\ &= \max_v E_w I_{\{s(x)\geq r\}} W(x;u,w) \ln f(x;v) \end{aligned} \quad (7)$$

where, $W(x;u,w) = \frac{f(x;u)}{f(x;w)}$ is the likelihood ratio. The optimal solution of (7) can be written as

$$\begin{aligned} &v^* \\ &= \arg\max_v E_w I_{\{s(x)\geq r\}} W(x;u,w) \ln f(x;v) \end{aligned} \quad (8)$$

We may estimate v^* by solving the following stochastic program

$$\begin{aligned} &\max_v \hat{D}(v) \\ &= \max_v \frac{1}{N} \sum_{i=1}^{N} I_{\{s(X_i)\geq r\}} W(X_i;u,w) \ln f(X_i;v) \end{aligned} \quad (9)$$

where $X_1, \ldots, X_N$ is a random sample from $f(x;w)$. In typical applications the function in (9) is convex and differentiable with respect to v, the solution of (9) may be readily obtained by solving (with respect to v) the following system of equations:

$$\frac{1}{N} \sum_{i=1}^{N} I_{\{s(X_i)\geq r\}} W(X_i;u,w) \nabla \ln f(X_i;v) = 0.$$

IV. Minimizing the Importance Sampling Techniques Based on the VariancE Minimization

The existence important sampling methods are mainly based on minimizing the cross-entropy techniques to estimate the probability of rare events [4], [8], [10]. The problem of these methods is that they cannot be approach the optimal solution in the sense of minimizing the estimator's variance. To overcome this difficulty, this paper proposes an IS method based on variance minimization. The well-known problem of the IS method is that it is hard to derive the optimal change: the choice of new process to simulate which will minimize the variance in the estimator. A non-optimal choice could make the behavior of the importance sampling estimator even worse than the MC estimator. Because the efficiency of IS is observed to be rather sensitive to the change of measure, a great deal of work is being done to obtain optimal – or at least good – simulation parameters[24],[25].

Since the effectiveness of importance sampling estimator totally depend on the choice of importance sampling distribution function, we often select importance sampling distribution function from a family of importance sampling distributions, and evaluate its effectiveness by computing the variance of the estimator. In this paper, we construct the family of importance sampling distribution using the exponential change of measure.

The variance of $\hat{l}_{IS}$ in (4):

$$Var[\hat{l}_{IS}] = \frac{1}{R} Var_{g(x)}[I_{\{s(x)\geq r\}} L(x)]$$

Let f_θ denote the family of distributions, that is $f_\theta(x) = c\exp(\theta x) f(x)$, we would like to minimize the variance of $\hat{l}_{IS}$ with respect to the family of distributions f_θ, that is, we want to solve

$$\min_\theta Var_{f_\theta(x)}[I_{\{s(x)\geq r\}} L_\theta(x)]. \quad (10)$$

where $L_\theta(x) = \frac{f(x)}{f_\theta(x)}$.

And also

$$\begin{aligned} &Var_{f_\theta(x)}[I_{\{s(x)\geq r\}} L_\theta(x)] \\ &= E_{f_\theta}[I_{\{s(x)\geq r\}} L_\theta^2(x)] - (E_{f_\theta}[I_{\{s(x)\geq r\}} L_\theta(x)])^2 \\ &= E_{f_\theta}[I_{\{s(x)\geq r\}} L_\theta^2(x)] - l^2 \end{aligned}$$

So (10) can be rewrite as:

$$\min_\theta E_{f_\theta}[I_{\{s(x)\geq r\}} L_\theta^2(x)]. \quad (11)$$

The above rationale results in the following algorithm:

step1. Define θ_0;

step2. Generate a sample $X_1^0, \ldots, X_N^0$ from importance sampling probability density function and solve the problem (11), update the parameter θ_i;

step3. Generate a new sample $X_1^i, \ldots, X_N^i$ from the importance sampling probability density function that updates, and solve the problem (11);

step4. If $\|\theta_i - \theta_{i-1}\| < \varepsilon$, get the solution θ^*, then stop; otherwise, go to step3;

step5. Estimate the rare event probability using (4).

An Armijo-type step selection rule is selected in the above algorithm [17]. Moreover, a reasonable choice for initial value is a key problem in algorithm. First, a recursive algorithm estimating conditional mode was employed to generate rare event samples [3]. And then, the subsequent iterative simulation procedures were taken with initial importance sampling probability density function whose parameters were estimated by using this set of samples.

V. SIMULATING ANALYSIS

To illustrate the ideas set forth in the previous sections, we present now numerical results obtained for project management problem. There are 14 key activities on key routes in a project. Let x denote the man-hour of all activities, and let $s(x)$ denote the total completion time of this project. Our goal is to estimate the probability that the total completion time of this project will be more than a given time r. That is, we want to estimate $l = P(s(x) > r)$. If $l < 10^{-5}$, and this is a rare event. We consider the case where all man-hours of all activities have exponential distribution. For simplicity, we assume that the times of all activities are i.i.d. with mean $\mu = 25$. To show the efficiency of the method, we used the Monte Carlo method to estimate $l = P(s(x) > r)$ at the same time. The above procedure was replicated 100 times, and the average and a simultaneous 90% confidence interval were built from those 100 independent estimates, both for the two methods. We consider three values for r, namely, $r = 1000$ $r = 1250$ $r = 2500$ .The following tables displays the estimation results. In the table, $\hat{l}$ is the estimate for $l = P(s(x) > r)$, "90% H.W." denotes the half-width of a 90% confidence interval and N is the sample size.

TABLE I.
SIMULATION RESULT BASED ON MONTE CARLO

r	$\hat{l}$	*90%H.W.*	N
1000	5.219e-007	4.4606e-009	200000
1250	—	—	5000000
2500	—	—	5000000

TABLE II.
SIMULATION RESULT BASED ON VARIANCE MINIMIZATION

r	$\hat{l}_{IS}$	*90%H.W.*	N
1000	6.153e-007	3.3905e-009	1000
1250	4.984e-010	7.3494e-012	2000
2500	2.185e-032	8.7000e-035	500000

Note: "—"in table 1 denote rare event did not happen after 5000000 samples and could not establish effective interval estimation.

The above results indicate high efficiency of the VM method for estimating rare events probabilities, where the naive Monte Carlo method fails. For events that are not very rare, the VM method may still help in terms of providing estimates with smaller variance.

VI. CONCLUSION

In this paper, an important sampling method based on variance minimization procedure is presented. The key concept is to construct the family of importance sampling distributions through selecting the classical exponential change of measure, and seeks the optimal importance sampling distribution function by minimizing variance of the importance sampling estimator, then generates samples from optimal importance sampling distribution function, get the probability of the rare event at the end. We present simulation result which supports the efficiency of the method. Rare event simulation research is still at the primary stage, there is a broad space for development both in theory and in application [1], [8], [14]. Currently, we are extending this method for other types of applications, such as economic crisis and the early-warning of financial risk.

REFERENCES

[1] P.W. Glynn,"Efficiency Improvement Techniques," *Annals of Operation Research*.1994,vol 53,pp. 175-197.

[2] J.S.Sadowsky, "Large Deviations Theory and Efficient Simulation of Excessivebacklogs in a GI/GI/m Queue,"*IEEE Transactions on Automatic Control*, 1991,pp.1383-1394.

[3] McComb W.D., Johnston C.,"Conditional mode elimination and scale-invariant dissipation in isotropic turbulence," *Physica A: Statistical Mechanics and its Applications,* 2001,pp.346-382.

[4] M Hsieh, "Adaptive Importance Sampling for Rare Event Simulation of Queuing Networks," Ph. D. thesis, Stanford University, December, 1997.

[5] P.W.Glynn and D.L.Iglehart, "Importance Sampling for Stochastic Simulations," *Management Science*, 1989,vol 35,pp.1367-1392.

[6] D.Siegmund, "Importance Sampling in the Monte Carlo Study of Sequential Tests," *The Annals of Statistics*. 1976,vol 4,pp. 673-684.

[7] D.P.Krose and R.Y.Rubinstein, "The Transform Likelihood Ratio Method for Rare Event Simulation with Heavy Tails," *Queuing Systems*. 2004,vol 46,pp. 317-351.

[8] PT de Boer, DP Kroese, S Mannor, RY Rubinstein, "A Tutorial on the Cross-Entropy Method," *Annals of Operations Research*,2005,pp.19-67.

[9] Al-Qaq WA, Devetsiklotis M, Townsend J K, "Stochastic Gradient Optimization of Importance Sampling for the Efficient Simulation of Digital Communication Systems," *IEEE Trans. Communications*.1995,vol 43,pp.2975-2985.

[10] Tito Homem-de-Mello and R.Y.Rubinstein, "Estimation of Rare Event Probabilities Using Cross-Entropy," *Proceedings of the 2002 Winter simulation Conference,*2002,pp. 310-319.

[11] J.S.Sadowsky, "On the Optimality and Stability of Exponential Twisting in Monte Carlo Simulation,"*IEEE Trans. In Theory* 1993,vol 39,pp.119-128.

[12] A Bohdanowicz, JH Weber, "Simulations of Communication Systems via Integrated Variance Reduction Techniques," *Vehicular Technology Conference* 2003,pp.2314-2318.

[13] Devetsikiotis M, Townsend J K, "an Algorithmic Approach to the Optimization of Importance Sampling Parameters in Digital Communication System Simulation," *IEEE Trans. Communications.* 1993,vol 41 pp.1464-1473.

[14] JA Bucklew, "Introduction to Rare Event Simulation" Springer 2004.

[15] JK Townsend, Z Haraszti, JA Freebersyser, M. Devetsikiotis, "Simulation of rare events in communications networks," *IEEE Communications Magazine,*1998,pp.36-41.

[16] M. Villén-Altamirano, A. Martínez-Marrón, J. Gamo, F. Fernández-Cuesta, "Enhancement of the Accelerated Simulation Method RESTART by Considering Multiple Thresholds," *In Proceedings 14th International Teletraffic Conference*,1994,pp.797–810.

[17] Jorge Nocedal and Stephen J. Wright, "Numerical Optimization," Springer,1999.

[18] Krishna Chepuri, Tito Homem-De-Mello, "Solving the Vehicle Routing Problem with Stochastic Demands using the Cross-Entropy Method," *Annals of Operations Research,* 2005,vol 134,pp. 153–181

[19] G. Alon, D.P. Kroese, T. Raviv, R.Y. Rubinstein, "Application of the Cross-Entropy Method to the Buffer Allocation Problem in a Simulation-Based Environment," *Annals of Operations Research,* 2005, vol 134, pp.137–151

[20] Josselin Garnier, Pierre Del Moral, "Simulations of Rare Events in Fiber Optics by Interacting Particle Systems," *Optics Communications,* 2006, vol 267,pp. 205–214

[21] Jeffrey P. Astheimer, Wayne C. Pilkington, Robert C. Waag, "Reduction of Variance in Spectral Estimates for Correction of Ultrasonic Aberration," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 2006, vol 53(1),pp.79-89

[22] K.P. Hui, N. Bean, M. Kraetzi, Dirk P. Kroese, "The Cross-Entropy Method for Network Reliability." *Annals of Operations Research*,2005,vol 134,pp. 101–118

[23] Xingbin Yu, Chanan Singh, "Expected Power Loss Calculation Including Protection Failures using Importance Sampling and SOM," *Power Engineering Society General Meeting,* 2004. IEEE, Vol.1, pp. 206- 211

[24] Reuven Rubinstein, "The Cross-Entropy Method for Combinatorial and Continuous Optimization." *Methodology and Computing in Applied Probability,* 1999, vol 1pp.127-190

[25] Qiuming Zhu, "Minimum Cross-Entropy Approximation for Modeling of Highly Intertwining Data Sets at Subclass Levels," *Journal of Intelligent Information Systems,* 1998, vol 11pp. 139–152