Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

NONPARAMETRIC ADAPTIVE IMPORTANCE SAMPLING FOR RARE EVENT SIMULATION

Yun Bae Kim
Deok Seon Roh

Sung Kyun Kwan University
School of Systems and Management Engineering
Suwon, Korea

Myeong Yong Lee

Telecommunications Network Laboratory
Korea Telecom R&D Group
Taejeon, Korea

ABSTRACT

Simulating rare events in telecommunication networks such as estimation for cell loss probability in Asynchronous Transfer Mode (ATM) networks requires a major simulation effort due to the slight chance of buffer overflow. Importance Sampling (IS) is applied to accelerate the occurrence of rare events. Importance Sampling depends on a biasing scheme to make the estimator from IS unbiased. Adaptive Importance Sampling (AIS) employs an estimated sampling distribution of IS to the system of interest during the course of simulation. In this study, we propose a Nonparametric Adaptive Importance Sampling (NAIS) technique, a non-parametrically modified version of AIS, and estimate the probability of rare event occurrence in an M/M/1 queueing model. Compared with classical Monte Carlo simulation and AIS, the computational efficiency and variance reductions gained via NAIS are reasonable. A possible extension of NAIS with regards to random number generation is also discussed.

1 INTRODUCTION

Simulation is a powerful tool that is used to assess the performance evaluation of ATM networks. The desired cell loss probability for ATM networks is within the range of 10^{-6} to 10^{-12} which makes it relatively costly to use classical simulation techniques to estimate the loss probability. This limitation has been reported in several previous simulation studies of ATM networks. The problems caused by this limitation can be classified in two ways. First, the random number generator may exhaust its cycle. Second, the excessive simulation time required to generate even a few cell losses may tempt the analyst to settle for estimators with large variance.

There are however, several fast simulation techniques that can be used to remedy these limitations, including: Importance Sampling, Parallel Simulation, Regenerative Method, and Hybrid Simulation. Smith (1997) and Glynn and Iglehart (1989), provide thorough surveys on these fast simulation techniques.

In rare-event simulation, the simulation software may not be able to provide a large enough sequence of random numbers without degeneration occurring (i.e., the same random numbers are repeated within a single simulation run).

IS is a powerful technique that is used for rare event simulation and the success of IS has been reported in several papers. The fundamental concept of IS is to modify the probabilities for rare event occurrences that govern the outcomes of the simulation in a way that allows for low-probability events to occur more frequently. To estimate the probability of rare events, we simulate with a biased sampling distribution that makes rare events more likely. The sample values from a biased sampling distribution are then adjusted to make the final estimates unbiased. However, selecting an optimal sampling distribution in order to make the events occur more frequently is not easy; how to make this happen is a very important issue. For example, an arbitrarily selected sample distribution can produce estimators with inflated variance. Hence the most crucial problem in IS is the selection of an optimal sampling distribution (often system specific) that guarantees variance reduction. How to select an optimal sampling distribution properly is still unresolved.

The main idea behind AIS is to recognizing that the distribution of the samples of error events is identical to the optimal sampling distribution of IS. Therefore, the distribution of the samples of error events may be used to estimate the properties of an optimal sampling distribution in an iterative way in order to close the gaps between the actual optimal sampling distribution and the estimate of the optimal sampling distribution.

Most of the works in IS focus on the calculation of an estimate for a sampling distribution using IS in a parametric way: see for instance Glynn and Iglehart (1989), Oh and Berger (1992, 1993), and West (1992, 1993). The nonparametric way has been studied by Givens and Raftery

(1996), and it can provide a significant improvement in the selection of an optimal sampling distribution.

Zhang (1996) proposes a nonparametric method to estimate an IS sampling distribution for any given system that uses an estimated sampling distribution to generate random variates rather than simply estimating the parameters of the optimal sampling distribution. He extends Nonparametric Importance Sampling (NIS) to Nonparametric Adaptive Importance Sampling (NAIS), which is just an iteration of NIS that requires more computation. Our NAIS is based on AIS, uses the initial sampling distribution under the conditions that the samples of rare enents occurred during the initial simulation run and adopts Zhang's nonparametric approach to estimate the optimal sampling distribution.

The rest of this paper is organized as follows. In Section 2, we introduce the basic idea of the IS method and the AIS method. Section 3 is devoted to the NAIS method. In Section 4, we test NAIS in an M/M/1 queueing model. Conclusions and ideas for future research are discussed in Section 5.

2 IMPORTANCE SAMPLING AND ADAPTIVE IMPORTANT SAMPLING

2.1 Importance Sampling (IS)

Let a random variable X be defined on the probability space (Ω, Γ, P), where Ω, Γ, and P are sample space, event space, and probability measure, respectively. The occurrence of a rare event, E, is defined as $E\in\Gamma$. The indicator function $\phi(x)$ can be defined as follows:

$$\phi(x)=\begin{cases}1, & \text{if } X\in E\\ 0, & \text{if otherwise.}\end{cases}$$

Consider the problem of estimating the probability of the rare event E :

$$E_P[\phi(x)]=\mu_{\phi(x)}, \tag{1}$$

where P is a measure with respect to the expectation is taken.

In classical simulation, (1) can be estimated with N independent samples as follows:

$$\hat{\mu}_{\phi(x)}=\frac{1}{N}\sum_{i=1}^{N}\phi(x_i).$$

According to the Strong Law of Large Numbers, $\hat{\mu}_{\phi(x)}$ converges to $\mu_{\phi(x)}$ as N increases. For $E_P[\phi(x)^2]<\infty$, a confidence interval of $\hat{\mu}_{\phi(x)}$ can be constructed using the Central Limit Theorem(CLT) as

$\hat{\mu}_{\phi(x)}-z_{\alpha/2}\sqrt{var_P[\phi(x)/N]}$,

$\hat{\mu}_{\phi(x)}+z_{\alpha/2}\sqrt{var_P[\phi(x)/N]}$), where $z_{\alpha/2}$ is the 100(1-α/2)% quantile for a standard normal distribution. Since the variance $var_P[\phi(x)]$ is not known beforehand, it must be replaced with the sample variance.

Let X be defined for the probability space (Ω, Γ, P_I); where P_I is the IS probability measure, and $dP_I(x)$ is absolutely continuous with respect to $dP(x)$, and is different from $dP(x)$ (the rare event occurs more frequently). This condition implies that if the old probability measure is positive, then the new probability measure will also positive. Then

$$\begin{aligned}\mu_{\phi(x)}&=E_P[\phi(x)]=\int\phi(x)dP(x)\\&=\int\phi(x)\frac{dP(x)}{dP_I(x)}dP_I(x)\\&=\int\phi(x)L(x)dP_I(x)\\&=E_{P_I}[\phi(x)L(x)],\end{aligned} \tag{2}$$

where $L(x)=dP(x)/dP_I(x)$ is the likelihood ratio. Using the samples $\{(\phi(x_1),L(x_1)),\cdots,(\phi(x_N),L(x_N))\}$ generated from P_I, an unbiased estimator ($\hat{\mu}_I$) is given by

$$\hat{\mu}_I=\frac{1}{N}\sum_{i=1}^{N}\phi(x_i)L(x_i). \tag{3}$$

To obtain the variance reduction in IS, selecting $dP_I(x)$ as the likelihood ratio $L(x)<1$ when $\phi(x)=1$ is important

$$\begin{aligned}E_{P_I}[\phi(x)^2\cdot L(x)^2]&<E_{P_I}[\phi(x)^2\cdot L(x)]\\&=E_P[\phi(x)^2].\end{aligned}$$

If $E_{P_I}[\phi(x)^2L(x)^2]<\infty$, then a new confidence interval can be calculated as $(\hat{\mu}_I-z_{\alpha/2}\sqrt{var_{P_I}[\phi(x)\cdot L(x)]/N}$, $\hat{\mu}_I+z_{\alpha/2}\sqrt{\mathrm{var}_{P_I}[\phi(x)\cdot L(x)]/N})$. Shahabudin (1994) reports that there exists probability measure (P_I), that gives a variance of 0, but it requires a knowledge of the probability of the rare event. The most important task in IS is to find an easily tractable measure that guarantees variance reduction. Therefore, it is necessary to select a sampling distribution that reflects the rare event (E) well. The theoretical optimal sampling distribution of IS can be given by

$$dP_I(x) = \phi(x) \cdot dP(x) / \mu_{\phi(x)} . \quad (4)$$

Using equation (4), the original estimator can be calculated as follows:

$$\begin{aligned} \mu_{\phi(x)} &= \phi(x_i)L(x_i) \\ &= \phi(x_i)\frac{dP(x)}{\phi(x_i) \cdot dP(x) / \mu_{\phi(x)}} . \end{aligned} \quad (5)$$

Since the optimal sampling distribution is dependent on the unknown estimator, $\mu_{\phi(x)}$, the random variates, x_i's cannot be generated directly from the theoretical optimal sampling distribution formula used in (4). The incorrect selection of a sampling distribution for IS done in a parametric way may provide an imprecise estimator. Much of the research on IS is focused on choosing a reasonable approximation of optimal sampling distribution. In classical IS approach, one assumes that sampling distribution belongs to a parametric family dP_I (•, θ), $\theta \subset \Theta$. Choosing a parameter value θ that satisfies certain optimality criterion is problematic. Sigmund (1976) assumed the sampling distribution belongs to an exponential family, whereas Oh and Berger (1993) assumed that the sampling distribution is a mixture distribution. Adaptive Important Sampling (AIS) was thus developed to overcome this problem.

2.2 Adaptive Important Sampling (AIS)

It is obvious that if we use the optimal sampling distribution then the variance of the estimator becomes zero as Shahabudin (1994) has noted. This implies that a perfect estimate of $\mu_{\phi(x)}$ can be obtained in a single simulation run. AIS basically assumes that the distribution of the samples of observations conditional on the rare event occurring area and the optimal sampling distribution for IS are the same. This can be expressed as,

$$dP(x \mid \mathrm{X} \in E) = \phi(x) \cdot dP(x) / \mu_{\phi(x)} = dP_I(x) . \quad (6)$$

AIS uses the simulation results to estimate the parameters of the unknown optimal sampling distribution. AIS can minimize computational efforts by using the probability density function (pdf) of the simulation output to estimate the parameters of the unknown optimal sampling distribution and the probability of rare event occurrence simultaneously. Several short simulation runs are performed in an AIS algorithm. For each run, $\mu_{\phi(x)}$ and $dP_I(x)$ are estimated. Then the sampling distribution of IS is modified such that its properties match the estimated properties of optimal sampling distribution to be used in the subsequent simulation runs. In this way, the sampling distribution of IS becomes more like the optimal sampling distribution and the estimate of $\mu_{\phi(x)}$ becomes more accurate as the simulation runs are performed successively (For a more detailed algorithm of AIS, see Stadler and Roy (1993)). AIS is a simulation technique that estiamtes every unknown quantities during the course of simulation. AIS has advantage over calssical IS for estimating both θ^* and $\mu_{\phi(x)}$ using the same set of samples, hence reduce the computation time.

3 NON-PARAMETRIC ADAPTIVE IMPORTANCE SAMPLING (NAIS)

An improperly selected IS distribution may bring variance inflation instead of reduction. The proper selection of an initial distribution remains problematic in AIS. If prior information about the sampling distribution of IS is not available for a given system, a nonparametric approach may be more helpful.

Based on equation (6) of AIS, our NAIS method uses the initial sampling distribution based on the samples of rare events that occurred during the initial run, but it also and uses Zhang's nonparametric idea to estimate the optimal sampling distribution. NAIS is estimating the optimal sampling distribution itself for a given system rather than estimating the parameter of the optimal sampling distribution. We then use this estimated optimal sampling distribution to generate random variates directly. NAIS use the samples conditioned on the rare event occurring area, which requies less samples compared to NIS.

Silverman (1986) introduces four nonparametric methods to estimate a density function*:* histogram, kernel estimation, nearest neighbor, and variable kernel. We have focused on the kernel estimation method for the ease of use and accuracy.

We propose the NAIS algorithm as follows:

Step 1: Initialize a simulation run to collect the rare events occurring samples y_i i=1,…., *p*., y_i is the *i*th sample (inter-arrival time or service time)that caused an error

Step 2: Esitmate the optimal sampling distribution $f^*_{opt}(x)$ using the kernel function estimation method.

$$\hat{f}^*_{opt}(x) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{h}K(\frac{x - y_i}{h}) , \quad (6)$$

where h is a smoothing parameter and $K(\bullet)$ is a simple rectangular kernel function such as

$$K(x)=\begin{cases}\frac{1}{2}, & if\ |x|<1\\ 0, & if\ otherwise.\end{cases}$$

Step 3: Run a simulation where $\hat{f}^*_{opt}(x)$ is the optimal sampling distribution of IS, and calculate $\hat{\mu}_{\phi(x)}$ as follows:

$$\hat{\mu}_{\phi(x)}=\frac{1}{n}\sum_{i=1}^{n}\frac{f(x_i)}{\hat{f}^*_{opt}(x_i)}\phi(x_i),$$

where n is the number of replications or the number of regeneration cycles.

4 NUMERICAL RESULTS

We tested the proposed NAIS algorithm in an M/M/1 model. Let λ and μ be the mean arrival rate and the mean service rate in an M/M/1 queueing model, respectively. Consider the problem of estimating the probability, $\mu_{\phi(x)}$, that the number of customers reaches a certain queue level (A) during a busy period. Assume that the sample path (ω) reaches queue level (A) during a busy period and that this occurrs at time τ before the system becomes empty. If the number of departures is m, then there are A+m-1 arrivals during any given busy period.

Now the NAIS algorithm can be described as follows:

Step 1: Initialize a short simulation run to collect the samples of inter-arrival and service times when the number of customers reaches the queue level A during a busy period. Collect $t_i\ i=1,.,A+m-1$ and $s_j, j=1,..,m$.

Step 2: Use the kernel function estimation method with the samples collected from Step 1 to estimate the optimal sampling distributions of inte-rarrival and service times to be used later in the simulation. Estimate $f^*(x)$ and $g^*(x)$ using equation (8).

Step 3: Proceed in the simulation with the estimated optimal sampling distributions, $f^*(x)$ and $g^*(x)$ obtained from Step 2.

The sampling distributions of IS can be determined by modifying the samples' inter-arrival and service times during a busy period. It follows that for any busy period, if a sample path (ω) represents the arrivals and departures from the queue, then the likelihood function $L(\omega)$ can be defined as follows:

$$L(\omega)=\prod_{i=1}^{A+m-1}\frac{f(t_i)}{f^*(t_i)}\bullet\prod_{j=1}^{m}\frac{g(s_j)}{g^*(s_j)}, \tag{7}$$

where

t_i : inter-arrival time for the i^{th} customer

s_j : service time for the j^{th} customer

$f(x)$: pdf for inter-arrival times

$f^*(x)$: sampling pdf of NAIS for inter-arrival times

$g(x)$: pdf for service times

$g^*(x)$: sampling pdf of NAIS for service times.

For N busy periods (it can be obtained through either N independent replications or N regenerative periods), the probability $\mu_{\phi(x)}$ can be calculated as follows :

$$\mu_{\phi(x)}=\frac{1}{N}\sum_{i=1}^{N}L(\omega_i)\bullet\phi(\omega_i),\quad i=1,....N, \tag{8}$$

where

$$\phi(\omega)=\begin{cases}1, & if\ the\ number\ of\ customers\ reaches\\ & the\ queue\ level\ (A)\\ 0, & if\ otherwise.\end{cases}$$

We then run the Monte Carlo, AIS, and NAIS simulation varying the queue level (A) for 10,000 busy periods at 10 different times.

Table 1 shows the results of the AIS, NAIS, and the Monte Carlo simulation: first column, $\hat{\mu}_{\phi(x)}$, is the probability that the number of customers reaches the queue level (A), the second column is the standard deviation (SD) of $\hat{\mu}_{\phi(x)}$, and the third column is the half-width of a 90% confidence interval for $\hat{\mu}_{\phi(x)}$.

Table 1: NAIS, AIS and Monte Carlo Results of an M/M/1 Queueing Model (with 10,000 Busy Periods)

Queue level (A)	$\hat{\mu}_{\phi(x)}$			SD			Half-width		
	Monte Carlo	AIS	NAIS	Monte Carlo	AIS	NAIS	Monte Carlo	AIS	NAIS
10	2.53E-03	2.43E-03	2.36E-03	2.96E-04	3.86E-04	1.02E-04	3.36E-04	2.24E-04	1.16E-04
15	1.40E-04	1.90E-04	1.25E-04	9.66E-05	6.66E-06	4.88E-06	1.10E-04	3.86E-06	5.54E-06
20	3.00E-05	1.45E-05	7.37E-06	4.83E-05	2.20E-07	2.25E-07	5.49E-05	1.28E-07	2.56E-07
25		1.15E-06	6.61E-07		2.37E-08	3.10E-08		1.37E-08	3.52E-08
30		8.69E-08	6.41E-08		4.40E-09	4.82E-09		2.55E-09	5.48E-09
35		6.90E-09	3.27E-09		2.54E-10	1.46E-10		1.47E-10	1.66E-10
40		4.94E-10	5.38E-11		6.99E-11	4.78E-11		4.05E-11	5.23E-11

($\lambda = 0.3$, $\mu = 0.5$)

In Table 1, the probability, $\hat{\mu}_{\phi(x)}$, decreases as the queue level (A) increases. The probability below 10^{-5} cannot be estimated for the Monte Carlo simulation since the num ber of rare events is not sufficient. The curves illustrated in Figure 1 are the AIS, NAIS and Monte Carlo (MC) simulation results of $\hat{\mu}_{\phi(x)}$.

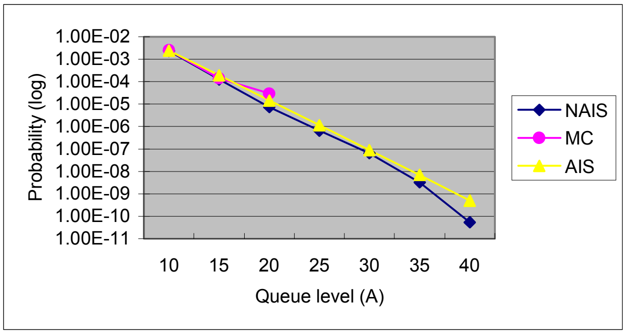

Figure 1: Probability of Customer Reaches the Level A During Busy Period

Figure 2 shows the half-widths of the 90% confidence intervals of the MC, AIS, and NAIS simulations. We can see that a variance reduction is obtained with the AIS and NAIS. Since variance in theses cases is reduced, the confidence intervals become tighter, which implies these the estimates are more accurate. In an M/M/1 queueing model, the asymptotical optimal sampling distribution of AIS is found by swapping λ and μ. This explains the medicore performance of NAIS. However, if the optimal sampling distribution is not known (such as in an ATM network simulation), the performance of NAIS will be better than AIS.

5 CONCLUSION

This paper proposed a modified fast simulation technique, NAIS, and demonstrated its performance in the estimation

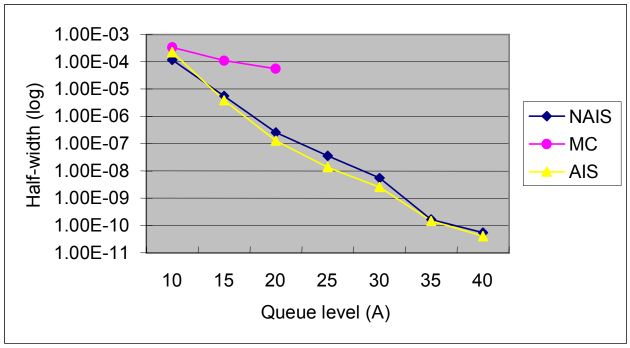

Figure 2: Half-widths of 90% Confidence Intervals for $\hat{\mu}_{\phi(x)}$

of rare event probabilities (when the queue level exceeds a certain level) in an M/M/1 queueing model. The experiments done using NAIS show substantial gains regarding computational efficiency and the variance reduction when compared to classical Monte Carlo simulation and AIS. The difficulty of choosing a proper optimal sampling distribution for IS and AIS can be eased by applying NAIS, since we can use data that is collected during the simulation run directly to modify the optimal sampling distribution estimate regardless of the characteristics of the system. To improve efficiency when modifying the estimate for an optimal sample distribution a more complicated kernel function, such as Gaussian density, is worth of investigation. We noted that the time taken for random number generation could be decreased if a more efficient random variate generation technique is used. A method that guarantees an optimal sampling distribution, which is invertible regardless of the density estimation methods also needs to be developed. Therefore, NAIS should be used for the simulation of highly reliable systems whose general characteristics are not known beforehand. We will applied NAIS in estimating the cell loss probability in ATM networks.

REFERENCES

Givens, G. H. and A. Raftery. 1996. Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. *Journal of American Statistical Association*, 91: 132-141.

Glynn, P., and D. Iglehart. 1989. Importance sampling for stochastic simulations. *Management Science,* 35(4): 1367-1392

Oh, M. S., and J. Berger, 1992. Adaptive importance sampling in Monte Carlo integration. *Journal of Statistical Computation and Simulation*, 41: 143-168.

Oh, M. S., and J. Berger, 1993. Integration of multimodal functions by Monte Carlo importance sampling. *Journal of American Statistical Association*, 88: 450-456.

Shahabuddin, P., 1994. Importance sampling for the simulation of highly reliable Markovian systems. *Management Science*, 40(3): 333-352.

Sigmund, D. 1976. Importance sampling in Monte Carlo study of sequential tests. The Annals of Statistics, 4: 673-684.

Silverman, B., 1986. Density Estimation. London: Chap man and Hall.

Stadler, J. S. and S. Roy, 1993. Adaptive importance sampling," *IEEE J. on Selected Areas in Communications*, 11(3): 309-316.

West, M., 1992. Modeling with mixtures in *Bayesian Statistics 4*, Eds. J. M. Bernardo, 503-524. University Press Oxford, Oxford, United .Kingdom.

Zhang, P. 1996. Non-parametric importance sampling . *Journal of American Statistical Association,* 91(435): 1245-1253.

AUTHOR BIOGRAPHIES

YUN BAE KIM is an Associate Professor in the School of Systems Management Engineering at Sungkyunkwan University. He received his B.S.I.E from Sungkyunkwan University, M.S and Ph.D. from Rensselaer Polytechnic Institute. He is a member of INFORMS, ASA, KORMS, KSC and IEEE. His interests include simulation output analysis, bootstrap methods, fast simulation techniques, and telecommunication network performance. His email and web addresses are <kimyb@yurim.skku.ac.kr> and <bootstrap.skku.ac.kr>.

DEOK SUN ROH is an officer of Korean Navy. He received his B.S.I.E., and M.S.I.E from Sungkyunkwan University. His interests include simulation output analysis and telecommunication network performance.

MYEONG YONG LEE is a senior member of technical staff at Korea Telecom. He received his B.S.I.E., M.S.I.E., and Ph.D. from Seoul National University. His interests include traffic engineering of high speed networks. His email addresses is <yonglee@kt.co.kr >.