İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl: 5 Sayı:10 Güz 2006/2 s.35-41

VARIANCE REDUCTION VIA IMPORTANCE SAMPLING

Semih YÖN*, Dave GOLDSMAN**

ABSTRACT
Variability always occurs to be the most frightening phenomena in implementation of various kinds of experiments. We desire to control variability and decrease the variance of experiments in order to be aware of the accuracy of the constructed models and consequently supply reliable results. Importance Sampling, also called Biased Sampling is one of the variance reduction techniques especially used in Monte Carlo Methods. This study includes a research to gather the appropriate importance sampling density which gives the lowest variance. We illustrate the importance sampling method on an M/M/1 queuing problem involving a limited waiting capacity of 50 of buffer size and solve it with an efficient C coded simulation program. We first execute naïve simulation, afterwards we carried out importance sampling method and supplied meaningful decrease in the estimated variance of the case which queue length ever exceeds buffer size. By this way, one can calculate any expectation that cannot be calculated by analytically. Numerical results indicate that longer tailed proposal distributions provide much more meaningful decrease.

Keywords: Variance Reduction, Importance Sampling, Monte Carlo Simulation, M/M/1 Queue.

TARAFLI ÖRNEKLEME YÖNTEMİ İLE VARYANS AZALTMA

ÖZET
Değişkenlik veya rassal sayılara bağlı hata çeşitli deneylerde ortaya çıkan en korkutucu problemlerdendir. Gerçeğe uygun modeller kurup bunlardan güvenilir sonuçlar elde etmek istenir. Bunun için Monte Carlo uygulamalarında tahmini varyansı azaltan Taraflı Örnekleme (Importance Sampling) yöntemi kullanılabilir. Bu çalışmada en az varyansı veren dağılımlar bulunmaya çalışılmıştır. Bunun için basit bir M/M/1 kuyruk sistemi benzetim modellemesi ile analiz edilmiş ve 50 birimlik bir ön tamponun dolup aşılma olasılığı bulunmaya çalışılmıştır. Önce basit Monte Carlo benzetim modeli daha sonar Taraflı Örnekleme benzetim modeli kullanılarak sonuçlar alınmıştır ve sayısal sonuçlar daha uzun kuyruğa sahip dağılımların daha olumlu sonuç verdiğini göstermiştir.

Anahtar Kelimeler: Varyans Azaltma, Taraflı Örnekleme, Monte Carlo Benzetim, M/M/1 Kuyruk Modeli.

** Dept. of Industrial Engineering, Boğaziçi University,E-mail : semih.yon@boun.edu.tr*
***Dept. of Industrial Engineering, Boğaziçi University, E-mail : dave.goldsman@boun.edu.tr*

1. INTRODUCTION

Simulations driven by random inputs will produce random outputs. If we can somehow reduce the variance of an output random variable of interest without disturbing its expectation, we should have greater precision in means of smaller confidence intervals. The classical Monte Carlo method is a random number based approach to estimate physical quantities that are hard to compute exactly. Monte Carlo methods are used for numerically approximating integrals and have many application areas such as global optimization, nuclear shielding, and computational chemistry. Many problems in these areas can be formulated as integrals over a single model distribution or highly multi-modal distributions in the result of expectations which can be shown as

$$\theta = E[q(X)] = \int_{R^d} q(x) f(x) dx \tag{1.1}$$

where X= $(X_1, ..., X_d)$ denote a vector of iid random variables in R^d, having a joint density function (or joint mass function in the discrete case) of $f(x) = f(x_1, ..., x_d)$ and *q(x)* is an arbitrary real valued function in the sampled region. The accuracy of this estimation depends strongly on quality of sampling which can be improved in two ways: increasing the cardinality of sampling or introducing some kind of selection rules that make it more representative, (Dupuis, 2005).

To make Monte Carlo calculations faster and improve the accuracy, a biased sampling with weight coefficients, so called importance sampling, is used. The contribution of importance sampling is to introduce definite selection rules to generate the most likely samples or configurations and hence to obtain more accurate values of statistical averages, (Touzig et al., 2003). The basic idea is to compute a correction factor to the importance sampling estimates, based on sample weights accumulated during sampling. With proper weights the correction factor compensates for statistical fluctuations and lead to a lower variance. Very commonly used variance reduction techniques are importance sampling, stratification, common random numbers, antithetic variates and control variates. First two methods reflect the idea of using weighted sampling based on a priori qualitative or quantitative information in an attempt to reduce variance whereas others concentrate on introducing correlation to reduce variance.

In this article we address the problem of selection of high computational importance sampling density (so called proposal density). The article is organized in four main sections. Section 2 gives the method of importance sampling in detail. Section 3 gives two different applications of importance sampling on an M/M/1 queuing system. And finally conclusions are set in section 4.

2. IMPORTANCE SAMPLING

Importance sampling (IS) is used for numerically approximating integrals besides viewed as a variance reduction technique. The idea behind IS is that certain values of the input random variables in a simulation have more impact on the parameter being estimated than others. If these important values are emphasized by sampling more frequently, then the estimator variance can be reduced. Hence, the basic methodology in IS is to choose a distribution which encourages the important values. This use of a biased distribution will result in a biased estimator. However, the simulation outputs are weighted to correct for the use of the biased distribution, and this ensures that the new IS estimator is unbiased. We can demonstrate the application of IS by using equation (1.1) with a mathematical approach as

$$\theta = E_g[q(X)\frac{f(X)}{g(X)}] = \int_{R^d} q(x)\frac{f(x)}{g(x)}g(x)dx \tag{2.1}$$

where *g(x)* is the proposal density such that *g(x)*=0 whenever *f(x)*=0 and always be of one sign. E_g emphasizes that random vector X has joint density *g(x)*. Technique of importance sampling is applicable for only rare events in wide sampling space. If the volume to be sampled is large, but can be characterized by small probabilities over most parts, IS can be carried out by approximating the probability distribution, *f(x)*, by proposal density, *g(x)*, and generating randomly x (forms a vector of iid variables) according to *g(x)*, then weighting each result at the same time by *w(x)=f(x)/g(x)*. The basic idea is to compute the correction factor, *w(x)*, to the IS estimates, based on sample weights accumulated during sampling. With proper weights the correction factor will compensate for statistical fluctuations and lead to a lower variance, (Bekaert et al.,2000). To supply this it is obvious that w(x) should be approximately constant, (Hörmann et al., 2005). Although the likelihood ratio *f(X)/g(X)* will usually be small in comparison to 1, average weight is obviously **1**. That is because q(x) is an indicator function that has the value of 1 when the target condition (X>k, k∈ R) is supplied and zero otherwise, (Sminchisescu et al., 2002).

$$\overline{w}(X) = E_g[\frac{f(X)}{g(X)}] = \int_{R^d}\frac{f(x)}{g(x)}g(x)dx = \int_{R^d} f(x)dx = 1.0 \tag{2.2}$$

2.1. Tilted Densities

Tilted densities are so useful in selecting a proposal density at the very beginning of the search. A density function of the form

$$f_t(x) = \frac{e^{tx}f(x)}{M(t)} \tag{2.3}$$

is a tilted density of f, where $-\infty < t < \infty$ and

$$M(t) = E_f[e^{tx}] = \int e^{tx} f(x)dx \tag{2.4}$$

is the moment generating function of one dimensional density *f*. A random variable with density f_t tends to be larger than *f* when t>0, and tends to be smaller when t<0, (Ross, 2002).

2.2. Conventional Biasing Methods

There are many kinds of biasing methods; following two methods are most widely used in the applications of importance sampling.

i) Scaling : Shifting probability mass into the event region X>k by positive scaling of the random variable X with a number greater than unity has the effect of increasing the variance (mean also) of the density function. This results in a heavier tail of the density, leading to an increase in the event probability. For random variable aX, a>1 we can have $g(x) = \frac{1}{a} f(\frac{x}{a})$ by transformation. While scaling shifts probability mass into the desired event region, it also pushes mass into the complementary region which is undesirable. If X is a sum of n random variables, the spreading of mass takes place in an n dimensional space. The consequence of this is a decreasing importance sampling gain for increasing n, and is referred to dimensionality effect.

ii) Translation : This technique employs translation of the density function (and hence random variable) to place much of its probability mass in the rare event region. Translation does not suffer from a dimensionality effect and has been successfully used in several applications. It often provides better simulation gains than scaling. In biasing by translation, the simulation density is given by $g(x) = f(x-b), b > 0$ where *b* is the amount of shift and is to be chosen to minimize the variance of the importance sampling estimator.

3. APPLICATION OF IMPORTANCE SAMPLING ON AN M/M/1 QUEUING SYSTEM

We implement a simulation of M/M/1 queue in the C program and had one million replications for each of a total of 15 long-run experiments. We investigate the behavior of the single server system that customers come into a poison process in accordance and served during an exponentially distributed time in FCFS queue discipline. The buffer size is fixed to 50 for 1000 limited arrivals. The goal is that to reduce the estimated variance of the case which buffer size ever exceeds 50 via IS method or in other words we desire to estimate θ in equation $\theta = E[q(X > 50)]$

more accurately by using equation (2.1) instead of equation (1.1). In the first subtitle below we deal with a tilted exponential proposal density for exponential inter arrival times and a translated Pareto proposal density for exponential service times.

Table 1. Comparison of Naive Simulation & Importance Sampling for Tilted Exponential Proposal Density.

	Mean	**Variance**	**Weight**
ρ =0.80			
naive simulation	0.000554	0.000554	
is λ =0.85 tilted rate	0.000448	0.000532	0.996387
is λ=0.82 tilted rate	0.000500	0.000305	0.989788
ρ =0.85			
naive simulation	0.005517	0.005487	
is λ =0.90 tilted rate	0.005543	0.012338	1.000641
is λ =0.87 tilted rate	0.005576	0.003921	0.999768
ρ =0.90			
naive simulation	0.041661	0.039925	
is λ=0.95 tilted rate	0.041761	0.108903	0.997965
is λ=0.92 tilted rate	0.041604	0.030433	1.002034
ρ =0.95			
naive simulation	0.194171	0.156469	
is λ =0.97 tilted rate	0.193781	0.139327	0.999893

3.1. Tilted Exponential Proposal Density

We illustrate IS method on a queuing system and simulate it for different traffic intensity values (ρ=.80, .85, .90, .95). First we executed naïve simulation and calculated mean and variance for binomial random variable. Then we carried out IS method for two tilted exponential densities with 2% and 5% rate increments in order to have longer tails. For constant service rate μ=1, we change the arrival rate λ's for different ρ values (as ρ=λ/μ). The related parameters are shown below for both original and proposal densities. The results for estimated mean, variance and average weight are shown in Table 1.

$f(x) = \lambda_f e^{-\lambda_f x}$, $g(x) = \lambda_g e^{-\lambda_g x}$ where $\lambda_g = 1.02 * \lambda_f$, and $\lambda_g = 1.05 * \lambda_f$

3.2 Translated Pareto Proposal Density

In order to increase the probability of buffer overload or in other words to make it more visualize, we can also increase the service times instead of decreasing the inter arrival times. We applied translated Pareto proposal density as

$g(x) = \frac{\alpha}{\beta}(\frac{\beta}{x+\beta})^{\alpha+1}, \alpha > 0, \beta > 0, x > 0$ α and β are shape and scale parameters respectively. When $x = 0, \mu = \alpha/\beta$. So we selected the parameters likewise to fit exponential service density accurately. We have μ =1, so α=β. The

results can be found in Table 2. When we compare Table 1 and Table 2 we can conclude that it is better to select longer tailed distributions in advance.

Table 2. Comparison of Naive Simulation & Importance Sampling for Translated Pareto Proposal Density.

	Mean	Variance	Weight
ρ =0.80			
naive simulation	0.000554	0.000554	
is α=20 β=20	0.000346	0.000305	0.980622
ρ =0.85			
naive simulation	0.005517	0.005487	
is α=20 β=20	0.002276	0.001523	1.004464
ρ =0.90			
naive simulation	0.041661	0.039925	
is α=20 β=20	0.013295	0.010563	0.996148
ρ =0.95			
naive simulation	0.194171	0.156469	
Is α=20 β=20	0.170268	0.119723	0.983216

4. CONCLUSIONS

Fundamental idea is that the sampling process is distorted in order to take into account the weighting of the underlying distribution. A key issue in order to achieve small errors on the obtained result is a suitable strategy of sampling the available one dimensional or multidimensional space. The term "importance sampling" also refers to choosing the proposal density so that the sampled values lie in the region that is important for the value of the integral. Numerical results indicate that it is better to select longer tailed distributions that fit the original density accurately. For complex problems finding samplable approximating distributions occurs to be a hard job, so it is then useful to look at sequential samplers based on distributions derived from the original density. The rewards for a good distribution can be huge run-time savings; the penalty for a bad distribution can be longer run times than for a general Monte Carlo simulation without any special techniques.

5. REFERENCES

Bekaert. P. Sbert. M. and Willems. Y.D.(2000) "Weighted Importance Sampling Technique for Monte Carlo Radiosity" 11th Eurographics Workshop on Rendering. Brno. Czech Republic.

Dupuis. P. (2005) "Dynamic Importance Sampling for Uniformly Recurrent Markov Chains" Inst. of Math.

Hesterberg. T. (1995). "Weighted Average Importance Sampling and Defensive Mixture Distributions." Technometrics 37, 2, 185-194

Hörmann. W. and Leydold. J. (2005)"Monte Carlo Integration Using Importance Sampling and Gibbs Sampling."

Law. A.M. and Kelton. W.D. (2000) "Simulation Modelling and Analysis." Third Edition. International Editions

Ross. S. (2002) "Stochastic Process Applications." Fifth Edition.

Sminchisescu. C. and Triggs. B. "Hyperdynamics Importance Sampling" in ECCV 2002.

Touzig. A. Hermann. H. (2003) "General Purpose Software for Monte Carlo Simulations", Elsevier.

Teşekkür: Bu yayının hazırlanması sürecinde Tübitak Bilim Adamı Destekleme Dairesi'nden (2210) aldığım burs herşeyi kolaylaştırdı.