State-dependent Importance Sampling and Large Deviations

Jose Blanchet
Statistics Department
Harvard University
Cambridge, MA. USA
blanchet@stat.harvard.edu

Jingchen Liu
Statistics Department
Harvard University
Cambridge, MA. USA
jcliu@stat.harvard.edu

Peter Glynn
Management Science and Engineering
Stanford University
Stanford, CA. USA
glynn@stanford.edu

ABSTRACT

Large deviations analysis for light-tailed systems provides an asymptotic description of the optimal importance sampler in the scaling of the Law of Large Numbers. As we will show by means of a simple example related to computational finance, such asymptotic description can be interpreted in different ways suggesting several importance sampling algorithms, some of them state-dependent. In turn, the performance of the suggested algorithms can be substantially different.

Keywords

Importance sampling, Random Walks, Rare-event Simulation, Large Deviations.

1. INTRODUCTION

Large deviations techniques are often used in designing efficient importance sampling estimators for rare-event simulation. Recall that an unbiased estimator is said to be efficient (or, more precisely, logarithmically efficient) if the ratio of the logarithm of its second moment to the logarithm of the square of its mean is bounded (see, for instance, Bucklew (2004)).

Our purpose is to illustrate how large deviations analysis describes the most likely path in a macroscopic scaling (often called "fluid scaling") that can hide some of the fine structure that is present in the conditional behavior, given the rare event, at the microscopic scale. In particular, there are frequently several different changes of measure that are consistent with the macroscopic desciption of the large deviations path associated with the rare event, but that are very different at the microscopic scale. Not surprisingly, the specific change of measure that is simulated at a microscopic scale can have a big impact on the variance and efficiency of the corresponding importance sampling estimator. Our exposition here focuses on a model problem related to the pricing of digital knock-in options in the context of computational finance. This problem exhibits enough complexity to illustrate the performance of different importance sampling algorithms that, as we indicated, are each consistent with a large deviations analysis (in fluid scale) of the associated problem.

To be precise, let $(X_k : k \geq 1)$ be a sequence of i.i.d. r.v.'s (independent and identically distributed random variables) and consider the random walk (r.w.) process $(S_n : n \geq 0)$ defined via $S_n = X_1 + ... + X_n$. Given $a < 0 < b$ and $S_0 = 0$ we are interested in efficient estimation via simulation of the probability

$$u(0, n) = P\left(\min_{0 \leq k \leq n} S_k < an, S_n > nb\right). \quad (1)$$

Such probability can be interpreted as the price of a digital knock-in option in the absence of interest rates. That is, the price of an option that upon exercise pays one unit at time n only if both $S_n > b$ and in addition the r.w. hit a level lower than a before time n. (See Glasserman (2003) for more applications of importance sampling to option pricing). For mathematical convenience in our exposition, we shall assume that the X_k's follow a standard Gaussian distribution.

Large deviations theory tells us that $u(0, n)$ decreases to zero at an exponentially fast rate in n that can be explicitly computed (see, for instance, Dembo and Zeitouni (1998)). It is also possible to obtain a precise description in the fluid scale proportional to n (i.e. the Law of Large Numbers (LLN's) scale) of the most likely path that the random walk follows conditional on the exercise of the option. This path is also known as the optimal path in large deviations analysis, because it can be computed by solving a variational problem.

It is not hard to see that the optimal path is piecewise linear. In particular, it has slope $s_1 < 0$ from time zero up to some deterministic time $k_0 < n$ at which time it reaches level a, and then it has positive slope $s_2 > 0$ from time k_0 up until time n at which time it reaches b. Because of the connection between exponential tilting and large deviations, this description of the optimal path suggests the first algorithm that we study. The exponential tilting algorithm (or **ET** algorithm) proceeds by performing i.i.d. exponential tilting for the first k_0 increments in order to match the slope s_1. Regardless of the state of the random walk at time k_0, the remaining $n - k_0$ increments are then exponentially tilted to induce drift s_2. Note that the **ET** is state-independent.

We shall call the second algorithm that we analyze the sequential exponential tilting algorithm (or **SET** algorithm), following the terminology adopted by Sadowsky (1996) used in describing a similar class of methods. **SET** is a small vari-

Valuetools'06 October 11-13, Pisa, Italy

ation of **ET.** Here, we apply exponential tilting to induce slope s_1 until the random walk hits level a,and from that time onwards (up until time n) we apply the tilting that induces mean $s_2 > 0$. **SET** is a simple state-dependent algorithm, but still there are only two tiltings that are applied during the course of the simulation.

Finally, the third algorithm, which we call **OSDET** for optimal state-dependent exponential tilting, corresponds to applying the large deviations analysis at each step in the course of the simulation. That is, we re-compute the optimal path depending on the current position and the remaining time. We then apply the corresponding exponential tilting to the next increment only and continue repeating this procedure until time n. **OSDET** is fully state-dependent in that typically one would use n different randomly determined tilting parameters over the course of the simulation.

The state-independent algorithm, **ET**, provides an estimator that exhibits exponential complexity. In particular, the estimator is not logarithmically efficient and this can be seen by its poor performance in our numerical experiments. In contrast to **ET**, the second algorithm, namely **SET**, turns out to be logarithmically efficient. This may be somewhat surprising given the similarities between these two algorithms. However, a little thought reveals a crucial difference between **SET** and **ET**. Indeed, the reflection principle yields that the probability of exercise coincides (up to quantities of order $O\left(n^{-1/2}\right)$) with $P\{S_n \leq -(b-2a)\,n\}$. Applying importance sampling to induce slope $s_1$up to the first passage time of level $a < 0$ and then applying tilting with mean $s_2 > 0$ (which by symmetry turns out to be $-s_1$) is essentially equivalent to estimating the probability of $\{S_n \leq -(b-2a)\,n\}$ using the corresponding optimal exponential tilting (which induces mean $-s_1$). Since this tilting is well known to be efficient for estimating such tail sum large deviations probabilities, it follows that **SET** is efficient. The third algorithm, **OSDET**, is also efficient. In fact, a very similar algorithm applied to related random walk problems has been shown to be strongly efficient (in that the coefficient of variation of the estimator remains bounded as $n \nearrow \infty$) (see Blanchet and Glynn (2006)). The connection between **OSDET**-type algorithms and their performance in terms of strong efficiency will be studied elsewhere.

The second and third algorithms described above can be related to a recently developed class of algorithms. In particular, the second algorithm corresponds to a subsolution to the associated Issacs equation introduced by Dupuis and Wang (2005). Interestingly, the third algorithm corresponds to the solution to the Issacs equation (Dupuis and Wang (2004)). It turns out that in great generality the **OSDET** importance sampling strategy, in which one re-computes the fluid scale optimal path at each step and exponentially tilts as required, corresponds to the strategy suggested by the solution to the differential game described by the Issacs equation of Dupuis and Wang. We shall not pursue the details of this connection here but defer those details for a future work.

2. ALGORITHMS AND PERFORMANCE

Let us first provide an explicit description of the three procedures. We shall start with the most complex of them, namely **OSDET.**

We need to evaluate for $t \in (0,1)$ and $x > a$ the asymptotic optimal path (in fluid scale) that the random walk follows when the knock-in option is exercised, namely if the event

$$\{\min_{nt<l\leq n} S_l \leq an, S_n \geq nb\}$$

occurs and it is further given that $S_{\lfloor nt \rfloor} = xn$.

Large deviations theory tells us that the limit

$$\begin{aligned} & I_1(x,t) \\ = & \lim_{n\to\infty} \frac{1}{n} \log P\left(\min_{nt<l\leq n} S_l \leq an, S_n \geq nb \middle| S_{nt} = xn \right) \end{aligned}$$

exists, as does

$$I_2(x,t) = \lim_{n\to\infty} \frac{1}{n} \log P\left(S_n \geq nb \middle| S_{nt} = xn \right).$$

for $x < b$.The function $I_1(\cdot)$ can be computed as

$$I_1(x,t) = \inf_{y(\cdot)\in A(x)} \frac{1}{2} \int_0^{1-t} \dot{y}(s)^2\, ds, \qquad (2)$$

where

$$\begin{aligned} A(x) &= \{y(\cdot) \in AC[0,1-t] : y(0) = x, \\ \min_{0\leq u\leq 1-t} y(u) &\leq a, y(1-t) \geq b\} \end{aligned}$$

and $AC[0,t]$ represents the set of real valued absolutely continuous function on the interval $[0,t]$. The solution $y_1^* \in A(x)$ attaining the infimum in (2) is called the optimal path and it can be computed explicitly as

$$\begin{aligned} y_1^*(u) &= [x + s_1 u] I(u \leq t_0) \\ &\quad + [s_2(u - t_0) - a] I(t_0 < u \leq 1-t), \end{aligned}$$

where

$$\begin{aligned} s_1 &= (a - x)/t_0, \\ s_2 &= (b-a)/(1-t-t_0), \\ t_0 &= (1-t)(x-a)/(b-2a+x). \end{aligned}$$

Now, let

$$\tau_n(a) = \inf\{m \geq 0 : S_m \leq -na\},$$

and suppose that $k < \tau_n(a)$ and that $S_k = nx$. Set $t = k/n$. Then, the description of the optimal path suggests applying exponential tilting to the $((k+1)$-*st* increment in order to induce mean s_1. Because the increments are standard Gaussian, this suggests sampling a Gaussian r.v. with mean $\theta_0^*(S_k/n, k/n) = s_1$ and unit variance.

If $k \geq \tau_n(a)$ then the next increment's mean matches that of the optimal path induced by $I_2(\cdot)$ (using a similar characterization as (2)). The optimal path is

$$y_2^*(u) = x + \frac{(b-x)}{1-t} u,$$

which yields that if $\tau_n(a) \leq k = nt$ and $x = S_k/n < b$ then the $(k+1)$-*st* increment follows a Gaussian distribution with mean $\theta_0^*(S_k/n, k/n) = (b-x)/(1-t)$.

In summary, at time k, given S_k, the $(k+1)$-*th* increment is generated according to a unit variance Gaussian distribution with mean

$$\begin{aligned} &\theta_0^* \left(S_k/n, k/n\right) \\ = \quad & -\frac{(b-2a+S_k/n)}{(1-k/n)} I\left(\tau_n\left(a\right) > k\right) \\ & + \frac{(b-S_k/n)}{1-k/n} I\left(\tau_n\left(a\right) \leq k, S_k/n < b\right). \end{aligned}$$

The corresponding importance sampling estimator is

$$\begin{aligned} L_0\left(n\right) &= \exp\left(-\sum_{k=0}^{n-1} \theta_0^* \left(S_k/n, k/n\right) X_{k+1}\right) \\ & \exp\left(\sum_{k=0}^{n-1} \theta_0^* \left(S_k/n, k/n\right)^2/2\right) \quad (3) \\ & \cdot I\left(\min_{0\leq k\leq n} S_k \leq an, S_n > bn\right). \end{aligned}$$

Now, we shall explain the ideas behind **ET.** The precise dynamics of this procedure can be easily explained using the notation introduced previously for **OSDET**. The suggested tilting is completely path-independent and is given by

$$\begin{aligned} \theta_E^*\left(k/n\right) &= -\left(b-2a\right) I\left(-na/(b-2a) > k\right) \\ & + \frac{(b-a)}{1-\left(-a\right)/\left(b-2a\right)} \\ & \cdot I\left(-na/(b-2a) \leq k\right) \\ &= -\left(b-2a\right) I\left(-na/(b-2a) > k\right) \\ & + \left(b-2a\right) I\left(-na/(b-2a) \leq k\right). \end{aligned}$$

Finally, we describe the tilting strategy for **SET.** Using the notation introduced for **OSDET,** we can easily describe the corresponding tilting strategy as

$$\begin{aligned} \theta_S^*\left(S_k/n, k/n\right) &= -\left(b-2a\right) I\left(\tau_n\left(a\right) > k\right) \\ & + \left(b-2a\right) I\left(\tau_n\left(a\right) \leq k, x < b\right). \end{aligned}$$

For both **ET** and **SET,** the estimators are completely analogous to **OSDET** (displayed in (3)). For instance,

$$\begin{aligned} L_E\left(n\right) &= \exp\left(-[\theta_E^*\left(0\right) S_{k_0} - k_0 \theta_E^*\left(0\right)^2/2]\right) \\ & \cdot \exp\left(-\theta_E^*\left(1\right)\left(S_n - S_{k_0}\right)\right) \\ & \cdot \exp\left((n-k_0)\theta_E^*\left(1\right)^2/2\right) \\ & I\left(\min_{0\leq k\leq n} S_k \leq an, S_n > bn\right), \end{aligned}$$

where $k_0 = \lfloor -an/\left(b-2a\right)\rfloor$.

We report the output for our numerical experiments in the table at the end of the paper. The sample size (i.e. number of replications of each replication) is 10^5. We assume $a = -1$ and $b = .8$ ($S_0 = 0$) and estimate $u\left(0, n\right)$ for $n = 10$, 20 and 30 respectively. The first row indicates the corresponding estimate for $u\left(0, n\right)$ (which corresponds to the empirical mean of each of the estimators) while the second row is the estimated standard deviation for each estimator and the third row is the estimated coefficient of variation (i.e. the ratio of the estimated standard deviation to the estimated mean).

The performance of **OSDET** is extremely good. This is perhaps not surprising given that the simulated random walk is able to recover and closely follow the optimal path even if random perturbations push the process temporarily off this path. The performance of **ET** is very poor, which may suggest that a state-dependent algorithm is required for efficient estimation via simulation of $u\left(0, n\right)$ for large values of n. In fact, as we make clear in the theorem below, the estimator given by **ET** has exponential complexity. However, the minor variation in the algorithm obtained by moving **ET** to **SET** provides an estimator having a performance that seems to be very good, as can be seen by comparing the form of $\theta_E^*\left(\cdot\right)$ and $\theta_S^*\left(\cdot\right)$ above. In fact, it turns out that the algorithm **SET** can be proved to be efficient.

The next result summarizes the efficiency properties of the estimators. Recall that an importance sampling estimator is said to be asymptotically (logarithmically) efficient if

$$\underline{\lim}_{n\to\infty} \log \widetilde{E}[L\left(n\right)^2]/\log u\left(0,n\right)^2 \geq 1, \quad (4)$$

where $\widetilde{E}\left(\cdot\right)$ represents the expectation under the change-of-measure. For our estimators above, we have $\widetilde{E}[L\left(n\right)^2] = E[L\left(n\right)]$ ($E\left(\cdot\right)$ is the underlying probability under which the increments are i.i.d. standard Gaussian).

THEOREM 1. *Let $L_E\left(n\right)$, $L_S\left(n\right)$ and $L_0\left(n\right)$ be the estimators provided by **ET**, **SET**, and **OSDET** respectively. Then,*

$$\overline{\lim}_{n\to\infty} \log E[L_E\left(n\right)]/\log u\left(0,n\right)^2 \leq 1-\delta$$

for some $\delta > 0$. On the other hand, both $L_S\left(n\right)$ and $L_0\left(n\right)$ are asymptotically efficient.

PROOF. Define

$$A = \left\{S_n \geq bn, \min_{0\leq k\leq n} S_k \leq an\right\},$$

$\theta = (b-2a)$, $\theta_- = -\theta$ and $k_0 = \lfloor -an/\left(b-2a\right)\rfloor$.

$$L_E\left(n\right) = \exp\left(\sum_{i=1}^{k_0}\left(\theta_-^2/2 - \theta_- X_i\right) + \sum_{i=k_0+1}^{n}\left(\theta^2/2 - \theta X_i\right)\right) I_A.$$

Now we will analyze the second moment of $L_E\left(n\right)$

$$\widetilde{E}[L_E\left(n\right)^2] = \exp\left(-n\theta^2\right)\widetilde{E}\left[\exp(2\theta\left[S_{k_0} - \left(S_n - S_{k_0}\right) + n\theta\right]); A\right].$$

Let $B_\varepsilon = A \cap \{S_{k_0} + n\theta \geq n\varepsilon\}$ and

$$C_\varepsilon = B_\varepsilon \cap \{S_{k_0} - \left(S_n - S_{k_0}\right) + n\theta \geq n\varepsilon\}$$

then,

$$\begin{aligned} \widetilde{E}[L_E\left(n\right)^2] &\geq \widetilde{E}[\exp(2\theta\left[S_{k_0} - \left(S_n - S_{k_0}\right) + n\theta\right]); B_\varepsilon] \\ & \cdot \exp\left(-n\theta^2\right) \\ &\geq \exp\left(-n\theta^2\right)\exp\left(4\theta\varepsilon n\right)\widetilde{P}\left(C_\varepsilon\right) \end{aligned}$$

It follows that,

$$\lim_{n\to\infty}\frac{\log\widetilde{P}\left(C_\varepsilon\right)}{n} = -2\theta\varepsilon + O\left(\varepsilon^2\right)$$

So,

$$\log\widetilde{E}\left(L_E\left(n\right)^2\right) \geq -\theta^2 + 2\theta\varepsilon + O\left(\varepsilon^2\right),$$

and one can pick ε small enough such that $2\theta\varepsilon + O\left(\varepsilon^2\right) \geq \delta > 0$. That $L_S\left(n\right)$ is logarithmically efficient is straightforward using the reflection principle and the fact that θ_- is the optimal tilting for computing $P\left(S_n \leq -n\left(2b-a\right)\right)$. Finally, as we have indicated before, $L_0\left(n\right)$ corresponds to the estimator suggested by the solution to the Issacs equation proposed by Dupuis and Wang (2004) and thus, in particular, is logarithmically efficient. □

	OSDET	ET	SET
n=10	1.669e-05	1.601e-05	1.653e-05
	2.051e-07	8.955e-07	1.858e-07
	3.885e+00	1.769e+01	3.553e+00
n=20	9.556e-09	8.662e-09	9.824e-09
	1.120e-10	1.514e-09	1.229e-10
	3.708e+00	5.529e+01	3.956e+00
n=30	5.921e-12	6.089e-12	6.051e-12
	6.324e-14	9.685e-13	8.167e-14
	3.377e+00	5.030e+01	4.267e+00

Table 1: Numerical Results

3. REFERENCES

[1] Blanchet, J. and Glynn, P. (2006) Efficient simulation of light-tailed sums: an old-folk song sung to a faster new tune. Preprint.

[2] Bucklew, J. (2004) *Introduction to Rare-event Simulation.* Springer, New York.

[3] Dembo, A. and Zeitouni O. (1998) *Large Deviations Techniques and Applications*: Second Edition, Springer-Verlag, New-York.

[4] Dupuis, P. and Wang, H. (2004) Importance sampling, large deviations, and differential games. *Stochastics and Stochastics Reports* 76, 481-508.

[5] Dupuis, P. and Wang, H. (2005) Subsolutions of an Isaacs equation and efficient schemes for importance sampling: Convergence analysis. Preprint.

[6] Glasserman, P. (2003) *Monte Carlo Methods in Financial Engineering,* Springer Verlag, New York.

[7] Glasserman, P. and Wang, Y. (1997) Counter examples in importance sampling for large deviations probabilities. *Ann. Appl. Prob.*, 7:731–746, 1997.

[8] Sadowsky, J.S. (1996) On Monte Carlo estimation of large deviations probabilities. *Ann. Appl. Prob.*, 6:399-422.