

# Path-sampling for State-dependent Importance Sampling

Blanchet, J. and Liu, J. C.

June, 2007

## Abstract

State-dependent importance sampling (SDIS) has proved to be particularly useful in simulation (specially in rare event analysis of stochastic systems). One approach for designing SDIS is to mimic the zero-variance change-of-measure by using a likelihood ratio that is proportional to an asymptotic approximation that may be available for the problem at hand. Using such approximation poses the problem of computing the corresponding normalizing constants at each step. In this paper, we propose the use of path-sampling, which allows to estimate such normalizing constants in terms of one dimensional integrals. We apply path-sampling to the tail of the delay in a G/G/1 queue with regularly varying input and argue such procedure, combined with a change-of-measure proposed by Blanchet and Glynn (2007a), yields an estimator with good relative accuracy with quadratic computational complexity (as a function of the tail parameter).

## 1 Introduction

Let  $Y = (Y_n : n \geq 0)$  be a Markov chain living on a space  $\mathcal{X}$  with a  $\sigma$ -field  $\mathcal{B}$ . Assume that  $Y$  has transition kernel  $(K(y, A) : y \in \mathcal{X}, A \in \mathcal{B})$  and consider, as a stylized problem, estimating

$$u(y) = P_y(T_A < T_B, T_A < \infty), \quad (1)$$

where  $P_y(\cdot)$  is the law of the process  $Y$  given  $Y_0 = y$  and  $T_A = \inf\{n > 0 : Y_n \in A\}$  (an analogous definition applies to  $T_B$  and we assume that

$A \cap B = \emptyset$ ). It is well known that applying importance sampling with the kernel  $(K_*(y, A) : y \in \mathcal{X}, A \in \mathcal{B})$  defined via

$$K_*(y, A) = \int_A K(y, dz) \frac{u(z)}{u(y)} dz,$$

yields an unbiased estimator with zero-variance for  $u(y)$  (see, for instance, Blanchet and Glynn (2007a)). However, since clearly applying  $K_*(\cdot)$  is not a feasible choice for simulation, a natural alternative consists in using a kernel  $K_v(\cdot)$  of the form

$$K_v(y, A) = \int_A K(y, dz) \frac{v(z)}{w(y)} dz, \quad (2)$$

where  $v(\cdot)$  is (an integrable) non-negative function and  $w(\cdot)$  represents the normalization that makes  $K_v(\cdot)$  as well defined Markov transition kernel. The importance sampling estimator for estimating  $u(y)$  takes the form (after simulating transitions under  $K_v(\cdot)$  with  $Y_0 = y$ )

$$Z = \prod_{k=0}^{T_A-1} \frac{w(Y_k)}{v(Y_{k+1})} I(T_A < T_B, T_A < \infty). \quad (3)$$

In some cases, the simulator has access to a function  $v(\cdot)$  that approximates  $u(\cdot)$  – examples are given in Section 2. Our focus on this paper is on developing a methodology for estimating the normalizing function  $(w(y) : y \in \mathcal{X})$  using a Monte Carlo strategy based on sample paths generated under  $K_v(\cdot)$ . Our motivation stems from the fact that, in some cases, there are known approximating functions that are developed under meaningful asymptotic regimes. For instance, in rare-event simulation, one often considers a family of problem instances indexed by some parameter that goes to infinity and that captures a situation of practical interest – such as the fact that  $u(y)$  could be small because  $y$  is closer to  $B$  than to  $A$  in some sense. An example of this situation is given in Blanchet and Glynn (2007a), where an approximation for the tail of the maximum of a random walk (rw) was used to develop the first strongly efficient algorithm for the delay of a heavy-tailed  $G/G/1$  queue (a quantity that admits the representation (1) for appropriate choices of  $W$ ,  $A$  and  $B$ ). The paper of Blanchet and Glynn (2007a) shows an important example in which asymptotic approximations are enhanced by efficient simulation algorithms, enabling the possibility of obtaining fast and

accurate estimates when the asymptotics may not be accurate enough. The technique developed in Blanchet and Glynn (2007a) has been subsequently applied to other situations, such as estimating large deviation probabilities for heavy-tailed rw's (Blanchet and Liu (2006)), a strongly efficient algorithm for light-tailed rw's (Blanchet and Glynn (2007)) and to counting problems involving complex combinatorial structures (such as graphs), see for instance Blanchet (2007).

As we indicated before, if an approximation  $v(\cdot)$  to  $u(\cdot)$  is available, it is desirable to use state-dependent importance sampling based on  $K_v(\cdot)$ . Path generation according to  $K_v(\cdot)$  may introduce additional complications that vary from one application to another. Hence, we consider that such issue should be studied depending on the situation at hand. In this paper, we assume that an algorithm for generating paths according to  $K_v(\cdot)$  is available but computing the corresponding normalizing function  $w(\cdot)$  along the generated trajectory could be more demanding. In Section 2 we discuss a couple of examples in the context of rare-event simulation that illustrate this type of situation. Such examples are taken from Blanchet and Glynn (2007a) and Blanchet and Liu (2006) in the context of first passage time probabilities for heavy-tailed random walks. In such examples, generating increments for the random walk according to  $K_v(\cdot)$  can be done efficiently using a well-constructed acceptance / rejection scheme. In situations where exact (unbiased) generation of transition steps according to  $K_v(\cdot)$  is not possible, then one could resort to the use of Markov chain Monte Carlo (MCMC) (see, for instance Liu (2001)). However, it is important to keep in mind that the use of MCMC introduces bias and in order to control the contribution of such bias one needs to develop rates of convergence to stationarity for Markov chains.

As mentioned previously, we are interested in developing an algorithm that allows to evaluate the normalizing function  $w(\cdot)$  along a sample path generated under  $K_v(\cdot)$  (and output an estimator of the form of (3)). Such sample paths often require the generation of many transitions under  $K_v(\cdot)$  and therefore, for each such path, one requires estimating  $w(\cdot)$  at many different points. Standard numerical integration rules are available in low dimensions, however, as is well known, the computational burden of such rules degrades rapidly as the dimension of the integral increases. An alternative in those cases, is to use Monte Carlo methods at each transition in order to estimate  $w(\cdot)$ . Such methods could be effective, but may require a properly designed strategy – specially for problems that require good precision (as in

rare-event simulation). Our procedure here can be seen as a refinement of the Monte Carlo alternative that is well suited for rare-event simulation problems. In particular, we propose to take advantage of the samples generated by the proposed importance sampler (which is designed to mimic the zero-variance change-of-measure) in order to estimate the normalizing constants. Since the importance sampler already incorporates the bias towards the regions that are relevant, the potential use of the generated sample paths to guide the estimation of the normalizing constants is particularly appealing.

Let us summarize our strategy. Suppose that one can accurately estimate  $w(y)$  (i.e. the normalizing function at the initial condition of the chain only). We deal also with this issue in Section 4. The more interesting part involves estimating the remaining normalizing constants. We use so-called path-sampling to estimate  $\log(w(z)/w(y))$  for values of  $z$  that correspond to the points visited by the generated sample paths. The key fact behind path-sampling is that the previous log-ratio can be expressed as a line integral (an integral over a path that joins  $y$  and  $z$ ) and such integral in turn can be expressed as an expectation. The log-ratio then can be estimated via Monte Carlo and this estimation procedure is called path-sampling. We have many degrees of freedom with regard to the path selection, but a sensitive choice is naturally guided by the proposed importance sampler (involving  $K_v(\cdot)$ ). The use of path sampling seems to be particularly advantageous in several dimensions because it allows to estimate the log-ratios through one dimensional integrals (along path trajectories). In addition, in many situations that arise in the context of rare-event simulation, the paths tend to be highly concentrated along deterministic trajectories that correspond to “fluid limits” based on conditional Laws of Large Numbers given that the rare event of interest occurs. Since the proposed importance sampler is designed to mimic the zero-variance change-of-measure (which corresponds to the conditional law mentioned previously), the sample paths will tend to be concentrated along such fluid trajectories, which in turn will naturally guide the construction of the path samplers.

The estimators for the log-ratios can be easily combined to obtain an estimator of the form of (3). As an illustration of our strategy, we apply our ideas to the problem of estimating the tail of the maximum of a random walk with regularly varying increments (a problem that was studied in Blanchet and Glynn (2007a)) and argue that the computational cost required to produce a path-sampling estimator that has a good relative accuracy is quadratic in the tail parameter of the maximum.

It is important to mention that path-sampling has been proved to be useful in various settings in which Monte Carlo methods are widely used. A comprehensive reference in the context of Statistical analysis is Gelman and Meng (1998), who also discuss certain optimality properties of various path-sampling strategies. In the Chemistry and Physics literature, path-sampling is known as thermodynamic integration or Ogata’s method (see Bennett (1976) and Ogata (1989)). Johansen, Del Moral and Doucet (2006) also discuss applications of path-sampling to estimating normalizing constants in rare-event simulation. Their discussion on the use of path-sampling, however, focuses only on a single constant and not on all the constants required along the generate trajectories.

The rest of the paper is organized as follows. In Section 2 we discuss few examples involving that illustrate the types of problems that motivate our development. Section 3 discusses basic notions on path-sampling. Section 3 provides a precise description of the path-sampling strategy. Finally, in Section 4 we provide error estimates for the path-sampling procedure applied to the tail of the delay in a heavy-tailed G/G/1 queue. Numerical experiments are also given at the end of Section 4.

## 2 Examples of Approximating State-dependent Samplers

We shall adopt the notation described at the beginning of the Introduction. Our discussion throughout the rest of the paper is applicable to importance sampling estimators for a general class of expectations, such as the expected accumulated reward up to the first hitting time to a set  $A$ , namely,

$$\tilde{u}(y) = E_y \left( g(Y_{T_A}) + \sum_{k=0}^{T_A-1} h(k, Y_k); T_A < \infty \right).$$

Awad, Glynn and Rubinstein (2007) discuss properties of the zero-variance change-of-measure these expectations assuming that  $g(\cdot)$  and  $h(\cdot)$  are non-negative. Such zero-variance change-of-measure can be expressed in terms of  $\tilde{u}(\cdot)$ . So, as we explained in the Introduction, if one can an approximation to  $\tilde{u}(\cdot)$  (in some sense), a natural approach is to build an importance sampler based on such approximation – just as we suggested in equation (2) in the

context of first-passage time probabilities. Assuming that path generation under the proposed importance sampler can be done efficiently, the problem then arises when trying to compute the normalizing constants in order to output the estimator (as the one displayed in (3)). Given the discussion in the Introduction, the issues that motivate our current development apply in substantial generality. However, to simplify the exposition, we shall focus on the first-passage time formulation proposed at the beginning of the paper.

In order to provide a more concrete idea of the types of issues that we address let us discuss a couple of rare-event simulation examples. In these examples, strongly efficient estimators are constructed using a transition  $K_v(\cdot)$  as in (2). Recall that a strongly efficient (unbiased) estimator,  $Z_b$ , for a probability  $u_b = P\{C_b\}$ , computed for a family of sets  $\{C_b : b \geq 1\}$  such that  $u_b \searrow 0$  as  $b \nearrow \infty$ , satisfies

$$\sup_{b \geq 1} E(Z_b)^2 / u_b^2 < \infty.$$

(It is important to emphasize – specially when using importance sampling estimators – that the expectation  $E Z_b^2$  is computed under the law under which  $Z_b$  is simulated).

**Example 1 (Blanchet and Glynn (2007a))** Suppose that

$$Y_{n+1} = Y_n + X_{n+1},$$

where the  $X_n$ 's are iid rv's with  $EX_n = -\mu < 0$ . For simplicity, let us assume that the  $X_k$ 's possess regularly varying tails (the discussion that follows holds in greater generality, basically under subexponential assumptions, which include Weibull and lognormal rv's). Regular variation means that if  $P(X > \cdot) = \bar{F}(\cdot) = 1 - F(\cdot)$ , then

$$\bar{F}(t) = t^{-\alpha} L(t),$$

where  $L(\cdot)$  is a slowly varying function (i.e.  $L(t\lambda)/L(t) \rightarrow 1$  as  $t \nearrow \infty$  for each  $\lambda > 0$ ). Now, set  $A_b = [b, \infty)$  and put  $B = \{-\infty\}$ . Then, the Pakes-Veraverbeke Theorem (see, for instance, Asmussen (2003)) yields that (for each  $a \in R$ )

$$u_b(y) = P_y(T_{A_b} < \infty) \sim P(Z > b - w + a) \tag{4}$$

as  $b \nearrow \infty$ , where  $Z$  is a rv with tail distribution given by

$$P(Z > t) = \min \left( \frac{1}{\mu} \int_t^\infty P(X_1 > s) ds, 1 \right).$$

As a consequence, (4) suggests using, for some  $a \in R$  (which can be selected by the simulator in order to further reduce variance) the function  $v_b(y) = P(Z > b - y + a)$  for the construction of a good importance sampling transition kernel  $K_v(\cdot)$  of the form given by expression (2). Blanchet and Glynn (2007a) proved that such sampling scheme gives rise to a strongly efficient simulation procedure. It follows that  $w_b(y) = P(Z + X > b - y + a)$  and therefore we obtain

$$K_v(y, A) = \frac{E(I(y + X \in A) \cdot P(Z + X > b - y + a | X))}{P(Z + X > b - y + a)}.$$

Sampling an increment distribution,  $X$ , according to  $K_v(\cdot)$  given that the current position of the random walk is  $y$  involves sampling from the conditional distribution law of  $X$  given the event  $\{X + Z > b - y + a\}$ . So, for instance, if  $X$  has density  $f_X(\cdot)$ , generating transitions of the  $Y$  under  $K_v(y, \cdot)$  (given current position  $y$ ) involves sampling increments from the conditional density (letting  $\beta = b - y + a_*$ )  $f_{X|X+Z>\beta}(\cdot)$  defined via

$$f_{X|X+Z>\beta}(x) = \frac{f_X(x) P(Z + x > \beta)}{P(Z + X > \beta)}. \quad (5)$$

Blanchet and Glynn (2007a) propose an acceptance / rejection procedure that allows to sample from the previous conditional density with an acceptance probability that is uniformly bounded away from zero for all values of  $\beta$ . As a consequence, sampling transitions from  $K_v(\cdot)$  can be done efficiently as  $b \nearrow \infty$  at any point visited by  $Y$ . The procedure that we propose here addresses the question of computing  $(w(Y_k) : 0 \leq k \leq T_{A_b} - 1)$ .

**Example 2 (Blanchet and Liu (2006))** Let  $(N(t) : t > 0)$  be a Poisson process with intensity 1. Define

$$S_t = \sum_{j=1}^{N(t)} X_j + S_0,$$

where the  $X_n$ 's are iid rv's (independent of  $N(\cdot)$ ) with  $EX_n = 0$ ,  $Var(X_n) = 1$  and possess regularly varying tails. Let  $Y_t = (S_t, t)$  for  $t \geq 0$ , define

$A_b = [\eta b, \infty) \times [b, \infty)$  for some  $\eta > 0$  and put  $B_b = R \times [b, \infty)$ . Then, a result of Rozovskii (1989) in particular implies the asymptotic approximation

$$\begin{aligned} u_b(0, 0) &= P_{(0,0)}(T_{A_b} < T_{B_b}) \\ &= P(S_t \geq \eta b) \sim bP(X_1 > \eta b) \end{aligned}$$

as  $b \nearrow \infty$ . Blanchet and Liu (2006) proved (in an analogous discrete time formulation) that strong efficiency of an asymptotic scheme that combines the previous approximation with the Central Limit Theorem through a kernel of the form  $K_v(\cdot)$  given in (2). The current continuous formulation is convenient in order to apply path sampling (as we shall see in the next section). As in Example 1, sampling exactly from efficiently  $K_v(\cdot)$  is feasible and the methods that we develop in this paper can be used to estimate the corresponding normalizing constants.

### 3 Ogata's Method and Path Sampling

Ogata (1989) proposed a method for numerical integration in several dimensions. In order to describe Ogata's method let us focus on a single dimension. Suppose that we have a family of positive functions,  $f(\cdot, \theta)$ , indexed by a parameter  $\theta \in R$  and integrable with respect to some underlying measure  $\mu(\cdot)$ . We are interested in computing

$$\tilde{w}(\theta_1) = \int_{-\infty}^{\infty} f(x, \theta) \mu(dx)$$

and assume that we know  $\tilde{w}(\theta_0)$  for some  $\theta_0 < \theta_1$ . First, we observe that

$$\log(\tilde{w}(\theta_1) / \tilde{w}(\theta_0)) = \int_{\theta_0}^{\theta_1} \partial_u \log \tilde{w}(u) du.$$

Now, assuming that it is possible to exchange derivatives and integrals (which is verified with standard techniques, see for instance, Durrett (2004), Appendix A9), we can write

$$\begin{aligned} \partial_\theta \log \tilde{w}(\theta) &= \int_{-\infty}^{\infty} \frac{\partial_\theta f(x, \theta)}{f(x, \theta)} \frac{f(x, \theta)}{\tilde{w}(\theta)} \mu(dx) \\ &= E_\theta R(X, \theta), \end{aligned}$$

where  $X$  is a random variable with law  $P_\theta(\cdot)$  given by  $P_\theta(X \in A) = \int_A f(x, \theta) / \tilde{w}(\theta) \mu(dx)$ , and  $R(x, \theta) = \partial_\theta \log f(x, \theta)$ . Combining the previous two displays we obtain

$$\begin{aligned} \log(\tilde{w}(\theta_1) / \tilde{w}(\theta_0)) &= \int_{\theta_0}^{\theta_1} E_u R(X, u) du \\ &= (\theta_1 - \theta_0) E(R(X, U)), \end{aligned} \tag{6}$$

where  $U$  is uniformly distributed over  $(\theta_0, \theta_1)$  and the pair  $(X, U)$  has law given by

$$P(X \in A, U \in B) = \frac{1}{\theta_1 - \theta_0} \int_B P_u(X \in A) du.$$

As a consequence, in order to estimate  $\tilde{w}(\theta_1)$  one can sample according to the previous law  $n$  iid replicates  $((X_k, U_k) : 1 \leq k \leq n)$  and consider

$$\begin{aligned} \tilde{R}_n &= \frac{1}{n} \sum_{j=1}^n (\theta_1 - \theta_0) R(X_j, U_j) \\ &\approx (\theta_1 - \theta_0) E(R(X, U)) = \log(\tilde{w}(\theta_1) / \tilde{w}(\theta_0)). \end{aligned} \tag{7}$$

Therefore, an estimator  $\tilde{w}(\theta_1)$  (which is biased but obviously consistent) is  $\exp(\tilde{Z}_n) \tilde{w}(\theta_0)$ . Note that one can take advantage of the representation in terms of an expectation given in (6) to introduce standard variance reductions techniques. For our purposes it suffices to consider the plain version that we just described. Now, also note that a very similar development applies if  $\theta$  takes values in higher dimensions. In such case the whole development can be done along a parameterized smooth trajectory that connects  $\theta_0$  and  $\theta_1$ . We prefer to simplify the exposition and focus on the one dimensional case here.

Ogata's method applies directly in our context, we just identify its elements with our formulation described in previous sections. In particular, we have  $\theta = y$ ,

$$\begin{aligned} f(x, \theta) \mu(dx) &= K(y, dx) v(x), \\ \tilde{w}(\theta) &= w(y). \end{aligned}$$

Note that it is required to have some degree of smoothness in the parameter  $\theta$ , for Example 1 this is satisfied in because  $v(\cdot)$  is differentiable. In Example 2, the Poissonization trick (i.e. basically, we have let the random walk

run on a Poisson clock) is convenient to previous ideas. Other alternatives, which we do not explore here, is to embed the discrete problem in a continuous formulation or use smoothing – Poissonization can be seen as a form of smoothing.

In fact, Ogata’s method can be traced back to the statistical physics literature where the procedure is known as thermodynamic integration. The term “path sampling” was coined by Gelman and Meng (1998), which studied these techniques in the context of Bayesian computations. Gelman and Meng (1998) provide a more detailed analysis of path sampling and certain optimality properties and several important references in settings such as Chemistry, Physics and Statistics.

## 4 Path Sampling in Action

We shall illustrate the path sampling technique through a detailed example. In particular, we shall work in the context of Example 1. Suppose that the increments of the random walk (i.e. the  $X_k$ ’s) possess a density  $f_X(\cdot)$  with respect to the Lebesgue measure. As we indicated in the description of Example 1,

$$w(y) = \int_{-\infty}^{\infty} f_X(y-z)v(z)dz = P(X+Z > b+a-y).$$

Recall that our estimator is

$$Z_b = \prod_{i=0}^{T_{A_b}-1} \frac{w(Y_k)}{v(Y_{k+1})}. \tag{8}$$

Under the change-of-measure proposed by Blanchet and Glynn (2007a) (indicated in Example 1) we have that  $P_y^{(v)}\{T_{A_b} < \infty\} = 1$  (where  $P_y^{(v)}(\cdot)$  is the law of the process  $Y$  induced by the transition kernel  $K_v(\cdot)$  – described in Example 1 – given that  $Y_0 = y$ ). So, we omit the indicator  $I\{T_{A_b} < \infty\}$  in expression (8).

We shall replace  $w(Y_k)$  in (8) by an estimator  $\hat{w}(Y_k)$  obtaining

$$\begin{aligned}\hat{Z}_b &= \prod_{i=0}^{T_{A_b}-1} \frac{\hat{w}(Y_k)}{v(Y_{k+1})} = Z_b \prod_{i=0}^{T_{A_b}-1} \frac{\hat{w}(Y_k)}{w(Y_k)} \\ &= Z_b \exp \left( \sum_{k=0}^{T_{A_b}-1} (\log(\hat{w}(Y_k)) - \log w(Y_k)) \right).\end{aligned}$$

Now, using the estimator  $\hat{Z}_b$  requires estimating  $w(\cdot)$  at points  $\mathcal{S} = \{y_1, \dots, y_m\}$  – that is, we must construct estimators  $(\hat{w}(y_1), \dots, \hat{w}(y_l))$  (we shall explain what  $l > 0$  is momentarily). Let us first assume that we have a good estimator  $\hat{w}(y_0)$  for some  $y_0$  (we shall also come back to this issue later). We will refer to the point  $y_0$  as the *reference point*. Usually  $y_0$  is chosen to be the starting point of the process (in the random walk problem described in Example 1,  $y_0 = 0$ ). The set  $\mathcal{S}$  correspond to the states visited (prior to the hitting the set  $A_b$ ) for the whole set of iid replications that are used to construct the final estimator for the first-passage time probability. Let us describe  $\mathcal{S}$  in more precise terms. We generate  $n$  iid replications of the process  $Y$  up to time  $T_{A_b}$  under  $K_v(\cdot)$ ; namely, we obtain  $\{Y_k^{(j)} : 0 \leq k \leq T_{A_b}^{(j)}, 1 \leq j \leq n\}$  (where the super-index  $j$  denotes the  $j$ -th replication). We denote the observed values via  $\tilde{y}_k^{(j)}$  (i.e.  $\tilde{y}_k^{(j)} = Y_k^{(j)}$ ). We then put

$$\mathcal{S} = \{y_1, \dots, y_m\} = \{\tilde{y}_k^{(j)} : 0 \leq k \leq T_{A_b}^{(j)}, 1 \leq j \leq n\}.$$

Since we are assuming that  $X_k$ 's possess a density, then the  $\tilde{y}_k^{(j)}$ 's are all distinct and  $l = \sum_{j=1}^n T_{A_b}^{(j)}$  – however, having different  $\tilde{y}_k^{(j)}$ 's does not play an important role in general.

The strategy is to apply Ogata's method to estimate the  $\hat{w}(y_j)$ 's. First, we construct a grid  $t_1 < \dots < t_{k^*}$  such that  $t_1 < \min_j y_j < \max_j y_j < t_{k^*}$ . Then, for each  $1 \leq j < k^*$ , we can estimate  $\log w(t_{i+1})/w(t_i)$  through (7). For instance, let us write  $(X^{(t_i)}, U^{(t_i, t_{i+1})})$  to denote a random variable with law

$$\begin{aligned}P(X^{(t_i)} \in A, U^{(t_i, t_{i+1})} \in B) &= \frac{1}{t_{i+1} - t_i} \int_{B \cap [t_i, t_{i+1}]} \int_A K_v(u, u+x) dx du \\ &= \frac{1}{t_{i+1} - t_i} \int_{B \cap [t_i, t_{i+1}]} P(X \in A | X + Z > b + a - u) du.\end{aligned}\tag{9}$$

We can sample  $n_i$  iid copies  $((X_j^{(t_i)}, U_j^{(t_i, t_{i+1})}) : 1 \leq j \leq n)$  of  $(X^{(t_i)}, U^{(t_i, t_{i+1})})$  and consider the estimator

$$\tilde{R}_{m_i}^{(t_i, t_{i+1})} = \frac{t_{i+1} - t_i}{m_i} \sum_{j=1}^{m_i} \frac{\partial \log v \left( U_j^{(t_i, t_{i+1})} + X_j^{(t_i)} \right)}{\partial y}. \quad (10)$$

As we saw in the previous section,

$$\tilde{R}_{m_i}^{(t_i, t_{i+1})} \longrightarrow \log(w(t_{i+1}) / w(t_i))$$

as  $m_i \nearrow \infty$ . (Note that taking derivatives under the integral sign here is valid almost everywhere if the increments have a density.) For simplicity, we shall assume that the  $m_i$ 's are constant equal to  $m$ .

In the context of Example 1,  $X_j^{(t_i)}$  is sampled according to the density (5). Now, to estimate  $w(y_k)$  for  $y_k \in \mathcal{S}$ , we shall take advantage of (10). Without loss of generality let us assume that  $y_k > y_0$  (otherwise, swap the roles of  $y_k$  and  $y_0$  in the discussion that follows). Let  $r = \min \{i : t_i \geq y_0\}$ ,  $s = \max \{i : t_i < y_k\}$ . Then,

$$\tilde{R}_m^{(y_0, y_k)} = \tilde{R}_m^{(y_0, t_r)} + \sum_{i=r}^s \tilde{R}_m^{(t_i, t_{i+1})} + \tilde{R}_m^{(t_s, y_k)}. \quad (11)$$

Therefore  $\tilde{R}_m^{(y_0, y_k)}$  provides an unbiased estimator for  $\log w(y_k) / w(y_0)$  and we can form the estimator

$$\hat{w}(y_k) = \hat{w}(y_0) \exp \left( \tilde{R}_m^{(y_0, y_k)} \right).$$

Usually, the  $t_j$ 's are chosen to be equally spaced or one can take advantage of the precise form of  $v(\cdot)$  on the region of interest – which is explored *after the simulations have been performed*. With regard to the number of samples  $m_i$  required for  $\tilde{R}_{m_i}^{(t_i, t_{i+1})}$  (or the rest of the log-ratio estimators), it is typically possible to choose  $m_i$  relatively low because the cardinality of the set  $\mathcal{S}$  will typically tend to be large. We can also choose  $m_i$  based on the output observed by the importance sampler, for instance, choosing  $m_i$  to be proportional to the number points landed in  $[t_i, t_{i+1}]$ . That is the higher the importance of the region (as dictated by the importance sampling outcome) the larger the accuracy of the corresponding log-ratio estimator.

Finally, we must indicate how to estimate  $w(y_0)$ . We propose as estimator simply

$$\hat{w}(y_0) = \left( \frac{1}{m} \sum_{j=1}^m v(Y_1^{(j)}) \right)^{-1}.$$

Here we write  $(Y_1^{(j)} : 1 \leq j \leq m)$  to denote  $m$  iid copies of the first transition of  $Y$  under  $K_v(y, \cdot)$ . In the case of Example 1, using the notation introduced in (9), we have

$$\hat{w}(y_0) = \left( \frac{1}{m} \sum_{j=1}^m v(y_0 + X_j^{(y)})^{-1} \right)^{-1}. \quad (12)$$

It follows easily that  $\hat{w}(y_0)$  is a consistent estimator (as  $m \nearrow \infty$ ). Although  $\hat{w}(y_0)$  is biased, it is appealing because it takes explicit advantage of the proposed importance sampler  $K_v(y, \cdot)$  – which is assumed to be good in principle. In order to have a better understanding of why taking advantage of  $K_v(y, \cdot)$  is appealing, it is useful to consider the situation in Example 1. Note that if  $y_0 = 0$ , then

$$w(y_0) = P(Z + X > b + a),$$

which is a small probability if  $b$  is large. The use of naive Monte Carlo is not appropriate because the sample size required to estimate  $w(y_0)$  with good relative precision would tend to be high. An algorithm that is useful to estimate  $w(y_0)$  efficiently in this context is easy to derive; however, since our intention is to provide a procedure that can be used more widely, we propose using (12). It is worth pointing out that one may wish to compute  $w(\cdot)$  at multiple positions in order to reduce the number of grid points (the  $t_j$ 's) required to produce the estimators for the log-ratios.

## 5 Quantifying the Error of Path-Sampling

It is important to analyze the size of the error introduced by Ogata's method and this analysis has to be done case-by-case. We are interested in finding an estimate on the number of variate generations (in addition to the generation

of transitions under  $K_v(\cdot)$  required to obtain an estimator  $\hat{Z}_b$  that is close to  $Z_b$  in relative terms, that is,

$$P\left(\left|\log\left(\frac{\hat{Z}_b}{Z_b}\right) - 1\right| \geq \varepsilon\right) \leq \delta$$

(for  $\varepsilon$  and  $\delta$  fixed). We provide an illustration of this analysis in a specific case corresponding to Example 1. Let  $X_i = V_i - \tau$ , where  $V_i$ 's are positive iid rv's with density

$$f_V(x) \sim \frac{L(x)}{(1+x)^{1+\alpha}},$$

where  $L(\cdot)$  is a slowly varying function. For simplicity we shall assume that  $\tau$  is a deterministic quantity such that  $EX_i < 0$  – in the context of the G/G/1 queue application that motivates Example 1, this assumption deterministic arrival times.

We apply a change-of-measure suggested by Blanchet and Glynn (2007a) with  $v(y) = P(Z > b - y + a)$  and

$$P(Z > t) = 1 \wedge \frac{1}{-EX_1} \int_t^\infty P(X > x) dx = \frac{1}{(1+t)^{\alpha-1}} \wedge 1.$$

The following lemma summarizes important properties of  $w(y)$ .

**Lemma 1** *There exists  $\tilde{\kappa} > 0$  such that for  $\tau + b - y + a > -1$ ,*

$$\begin{aligned} 0 &< E\left(\frac{\partial \log v(y+X)}{\partial y} \frac{v(x+X)}{w(X)}\right) \\ &= \frac{\partial \log w(y)}{\partial y} = \frac{\partial \log P(V+Z > \tau + b - y + a)}{\partial y} \\ &\leq \frac{\tilde{\kappa}}{1 + \tau + b - y + a}. \end{aligned} \tag{13}$$

and

$$E\left(\left(\frac{\partial \log v(y+X)}{\partial y}\right)^2 \frac{v(y+X)}{w(X)}\right) \leq \frac{\tilde{\kappa}}{(1 + \tau + b - y + a)^2}. \tag{14}$$

**Proof.** Equation (13) from the Karamata's theorem (see Embrechts et al (1997)) since

$$\frac{\partial \log w(y)}{\partial y} = \frac{\partial_y P(X + Z > b - y + a)}{P(X + Z > b - y + a)}.$$

Similarly, note that there exists  $\tilde{\kappa}_1, \tilde{\kappa}_2 > 0$  such that (for  $\tau + b - y + a > -1$ )

$$\frac{\partial \log v(y)}{\partial y} \leq \rho(y) = \min \left( \frac{\tilde{\kappa}_1}{(1 + \tau + b - y + a)^+}, \tilde{\kappa}_2 \right).$$

From the previous estimate we obtain,

$$\begin{aligned} & E \left( \left( \frac{\partial \log v(y + X)}{\partial y} \right)^2 \frac{v(X + y)}{w(y)} \right) \\ & \leq E \left( \frac{v(y + X) \rho(y + X)^2}{w(y)} \right). \end{aligned}$$

Now, we set  $\beta = 1 + \tau + b - y + a$  and pick  $\delta \in (0, 1)$  such that

$$\begin{aligned} & E \left( \frac{v(X + y) \rho(y + X)^2}{w(y)}; X < \delta\beta \right) \\ & \leq \tilde{\kappa}_1^2 E \left( \frac{1}{(\beta - X)^2} \frac{v(X + y)}{w(y)}; X < \delta\beta \right). \end{aligned}$$

On the other hand, we have (using  $f_X(\cdot)$  to denote the density of  $X$ )

$$\begin{aligned} & E \left( \frac{v(X + y) \rho(y + X)^2}{w(y)}; X \geq \delta\beta \right) \\ & = \int_{\delta\beta}^{\beta} \min \left( \frac{\tilde{\kappa}_1^2}{(\beta - x)^2}, \tilde{\kappa}_2^2 \right) \frac{v(y + x)}{w(y)} f_X(x) dx \\ & \leq \tilde{\kappa}_3 \int_{\delta\beta}^{\beta} \min \left( \frac{\tilde{\kappa}_1^2}{(\beta - x)^2}, \tilde{\kappa}_2^2 \right) \frac{\beta^{\alpha-1} L(\beta - x) L(x)}{L(\beta) (\beta - x + 1)^{\alpha-1} x^{\alpha+1}} dx \\ & \leq \frac{\tilde{\kappa}}{\delta^{\alpha+1} \beta^2} \int_{\delta\beta}^{\beta} \min \left( \frac{\tilde{\kappa}_1^2}{(\beta - x)^2}, \tilde{\kappa}_2^2 \right) \frac{L(\beta - x)}{(\beta - x + 1)^{\alpha-1}} dx, \end{aligned}$$

for some constants  $\tilde{\kappa}$  and  $\tilde{\kappa}_3$ . This completes the proof of the lemma. ■

Now, we analyze

$$\log \left( \frac{\hat{Z}_b}{Z_b} \right) = \sum_{i=0}^{T_{A_b}-1} \log \frac{\hat{w}(Y_i)}{w(Y_i)}.$$

For simplicity, since our main goal is to focus on the error due to path-sampling, we shall assume that  $w(y_0)$  is known. We have

$$E_{y_0}^{(v)} \left( \sum_{i=0}^{T_{A_b}-1} \log \frac{\hat{w}(Y_i)}{w(Y_i)} \right) = 0,$$

where  $E_{y_0}^{(v)}(\cdot)$  is the expectation operator associated to the kernel  $K_v(\cdot)$  given that  $Y_0 = y_0$ . On the other hand,

$$\begin{aligned} \sum_{i=0}^{T_{A_b}-1} \log \frac{\hat{w}(Y_i)}{w(Y_i)} &= \sum_{i=0}^{T_{A_b}-1} \left( \log \frac{\hat{w}(Y_i)}{w(y_0)} + \log \frac{w(y_0)}{w(Y_i)} \right) \\ &= \sum_{i=0}^{T_{A_b}-1} \log \frac{w(y_0)}{w(Y_i)} + \sum_{i=0}^{T_{A_b}-1} \tilde{R}_m^{(y_0, Y_i)}, \end{aligned}$$

where we put  $\hat{w}(y_0) = 1$  and  $\tilde{R}_m^{(y_0, y_0)} = 0$ . We construct the grid  $t_1 < \dots < t_{k^*}$  in such a way that the  $Y_j$ 's included in the grid (i.e.  $Y_i = t_{j_i}$  for  $i \in \{0, \dots, T_{A_b} - 1\}$ ). We sort the  $t_{j_i}$ 's and denote the sorted values by  $t_0^* < t_1^* < \dots < t_{T_{A_b}-1}^*$ . Now, note that

$$\begin{aligned} \sum_{i=0}^{T_{A_b}-1} \tilde{R}_m^{(y_0, Y_i)} &= \sum_{i=0}^{T_{A_b}-1} \tilde{R}_m^{(t_0^*, t_i^*)} = \sum_{i=0}^{T_{A_b}-1} \sum_{k=0}^{i-1} \tilde{R}_m^{(t_k^*, t_{k+1}^*)} \\ &= \sum_{i=0}^{T_{A_b}-1} (T_{A_b} - i - 1) \tilde{R}_m^{(t_i^*, t_{i+1}^*)}. \end{aligned}$$

Let  $\mathcal{F}_{T_{A_b}}$  be the stopped  $\sigma$ -field generated by  $Y$  up to time  $T_{A_b}$ . Then,

$$\begin{aligned}
& \text{Var}^{(v)} \left( \sum_{i=0}^{T_{A_b}-1} \log \frac{\hat{w}(Y_i)}{w(Y_i)} \middle| \mathcal{F}_{T_{A_b}} \right) \\
&= \text{Var} \left( \sum_{i=0}^{T_{A_b}-1} (T_{A_b} - i - 1) \tilde{R}_m^{(t_i^*, t_{i+1}^*)} \middle| \mathcal{F}_{T_{A_b}} \right) \\
&= \sum_{i=0}^{T_{A_b}-1} (T_{A_b} - i - 1)^2 \text{Var} \left( \tilde{R}_m^{(t_i^*, t_{i+1}^*)} \middle| \mathcal{F}_{T_{A_b}} \right).
\end{aligned} \tag{15}$$

We now can select  $k^* = c_1 T_{A_b}^2$  and  $m > 0$ . Lemma 1 can be used to conclude that for some constant  $\kappa > 0$ ,

$$\text{Var}^{(v)} \left( \tilde{R}_m^{(t_i^*, t_{i+1}^*)} \middle| \mathcal{F}_{T_{A_b}} \right) \leq \frac{\kappa}{m T_{A_b}^3}.$$

Therefore, we conclude that for some  $\kappa > 0$

$$\text{Var}^{(v)} \left( \sum_{i=0}^{T_{A_b}-1} \log \frac{\hat{w}(Y_i)}{w(Y_i)} \middle| \mathcal{F}_{T_{A_b}} \right) \leq \frac{\kappa}{m}.$$

Blanchet and Glynn (2007a) proved that  $E_{y_0}^{(v)} T_{A_b} \leq cb$  for some  $c > 0$ . Therefore, we conclude that we require roughly  $O(b^2 m)$  variate generations in order to obtain an estimator  $\hat{Z}_b$  so that

$$\begin{aligned}
& P \left( \left| \log \left( \frac{\hat{Z}_b}{Z_b} \right) - 1 \right| \geq \varepsilon \right) \\
&= E \left( P \left( \left| \log \left( \frac{\hat{Z}_b}{Z_b} \right) - 1 \right| \geq \varepsilon \middle| \mathcal{F}_{T_{A_b}} \right) \right) \leq \frac{\kappa}{m}.
\end{aligned}$$

We conclude this section with a numerical experiment. Assume that the  $V_i$ 's have tail distributions  $P(V_i > x) = (1+x)^{-5}$  and suppose that  $\tau = 1/2$ . In this case,  $EX = -.25$ . We select  $a = 1$  in this case. Our simulation results are presented in the next table for  $b = 20, 50, \text{ and } 100$ . We simulate 1,000 independent  $Z_b$ 's (which correspond to 1,000 sample paths) and

we denote their sample mean by  $\bar{Z}_b$  – we computed  $w(\cdot)$  with an efficient one dimensional numerical integral in order to produce  $\bar{Z}_b$ . The sample path generation was done according to the acceptance/rejection scheme proposed by Blanchet and Glynn (2007a). We then implemented path-sampling by selecting setting  $k^* = 250b^2$  (ranging the  $t_i$ 's from the minimum value observed in the simulations to  $b$ ) and set  $m = 10$ . We computed the average using the path-sampling estimates and denoted this quantity by  $\bar{Z}'_b$ . We use the notation  $\hat{\sigma}_{R|Y}$  to for the estimated standard error of the ratio  $\log(\bar{Z}'_b/\bar{Z}_b)$  given the outcome of the (1,000) generated sample paths (i.e. this standard error just accounts for the variability introduced in the path-sampling procedure only).  $\hat{\sigma}_{R|Y}$  is computed by applying the path-sampling procedure 10 independent times given the (1,000 generated previously). Finally, the column  $\hat{u}(b)$  is our benchmark, based on 200,000 Monte Carlo simulations using an importance sampling procedure developed by Blanchet, Glynn and Liu (2007a), which can be shown to be strongly efficient.

$b$	$\hat{u}(b)$	$\bar{Z}_b$	$\log(\bar{Z}'_b/\bar{Z}_b)$	$\hat{\sigma}_{R Y}$
20	$5.04e - 06$	$5.33e - 06$	0.0072	0.07
50	$1.47e - 07$	$1.456e - 07$	0.0539	0.149
100	$9.571e - 09$	$1.05e - 08$	-0.002	0.127

**Acknowledgement 2** *This research was partially supported by NSF grant DMS 0595595*

## References

- [1] Asmussen, S. (2003) *Applied Probability and Queues*. Springer-Verlag. New York.
- [2] Awad, H., Glynn, P., and Rubinstein, R. (2007) Importance sampling for Markov process expectations. Preprint.
- [3] Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* No. 22, p. 245-268.
- [4] Blanchet, J. (2007) Efficient Importance Sampling for Counting. Submitted.

- [5] Blanchet, J. and Glynn, P. (2007a) Efficient rare event simulation for the maximum of heavy-tailed random walks. Submitted.
- [6] Blanchet, J. and Glynn, P. (2007b) Efficient simulation of light-tailed sums: an old-folk song sung to a faster new tune. Preprint.
- [7] Blanchet, J., Glynn, P., and Liu, J. C. (2007a) Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue. Submitted.
- [8] Dupuis, P., Leder, K., and Wang, H. (2006) Notes on importance sampling for random variables with regularly varying tails. Preprint.
- [9] Durrett, R. (2004) *Probability: Theory and Examples*. Duxbury. Belmont, California.
- [10] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag. New York.
- [11] Gelman, A. and Meng, X. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statist. Sci.* No. 13 p. 163–185.
- [12] Johansen, A., Del Moral, P. and Doucet, A. (2006) Sequential Monte Carlo samplers for rare events. Proceedings of the 6th International Workshop on Rare Event Simulation.
- [13] Liu, J. (2001) *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York.
- [14] Ogata, Y. (1989). A Monte Carlo method for high dimensional integration. *Numer. Math.* No. 55, p. 137-157.

## AUTHOR BIOGRAPHY

**JOSE BLANCHET** is Assistant Professor of Statistics at Harvard University. Jose holds a M.Sc. in Engineering-Economic Systems and Operations Research and a Ph.D. in Management Science and Engineering, both from

Stanford University. He also holds two B.Sc. degrees: one in Actuarial Science and another one in Applied Mathematics from ITAM (Mexico). Jose worked for two years as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability, computational finance, performance engineering, queueing theory, risk management, rare-event analysis, statistical inference, stochastic modeling, and simulation. He has served in the editorial board of *ACM TOMACS*.