RARE EVENT SIMULATION TECHNIQUES

Jose Blanchet   Henry Lam

Columbia University   Boston University
500 W 120th Street   111 Cummington Street
New York, NY 10027, USA   Boston, MA 02215, USA

ABSTRACT

We discuss rare event simulation techniques based on state-dependent importance sampling. Classical examples and counter-examples are shown to illustrate the reach and limitations of the state-independent approach. State-dependent techniques are helpful to deal with these limitations. These techniques can be applied to both light and heavy tailed systems and often are based on subsolutions to an associated Isaacs equation and on Lyapunov bounds.

1   INTRODUCTION

We study recent techniques that have been developed for the design of provably efficient rare event simulation estimators. We concentrate mostly on importance sampling techniques. Our discussion here is based on the survey paper Blanchet and Lam (2011).

We mention that besides importance sampling, another powerful set of tools for variance reduction in rare event problems are based on splitting. Interested readers can consult L'Ecuyer et al. (2009) and Chapter VI Section 9 of Asmussen and Glynn (2007) for basic notions on splitting.

The types of techniques that we consider in this paper focus on the interplay between large deviations results and their applications to provably efficient importance sampling estimators. Early references in this spirit include Siegmund (1976) and Sadowsky (1991). Applications of these techniques in settings such as insurance and queueing include Asmussen (1985), Collamore (2002), Asmussen and Rubinstein (1995), and Parekh and Walrand (1989). We refer the reader to Juneja and Shahabuddin (2006), Bucklew (1990) and Chapter VI of Asmussen and Glynn (2007) for a wider scope on importance sampling for rare event simulation.

A series of examples and counter-examples (Glasserman and Kou 1995, Glasserman and Wang 1997) related to the use of large deviations principles in the design of efficient importance sampling estimators for light-tailed systems has motivated research on state-dependent techniques. In addition, as explained in Bassamboo et al. (2006), efficient importance sampling estimators in the setting of heavy tails often must be state-dependent (see also Asmussen et al. 2000 for more on the challenges that arise in the context of efficient rare-event simulation for heavy-tailed systems).

In order to systematically address the construction of provably efficient state-dependent importance sampling estimators for light-tailed systems P. Dupuis, H. Wang and their students have developed a method based on control theory and the use of subsolutions of an associated Isaacs equation for constructing and testing importance sampling estimators (Dupuis and Wang 2004; Dupuis and Wang 2007; Dupuis et al. 2007; Dupuis et al. 2007). A related approach, based on the construction of Lyapunov inequalities has been also developed (Blanchet and Glynn 2008; Blanchet and Liu 2008; Blanchet et al. 2007) for the construction and analysis of state-dependent importance sampling estimators for heavy-tailed systems.

In this paper we review the development of state-dependent importance sampling estimators and illustrate their application. We start by discussing standard notions of statistical efficiency that are used in order to rigorously test the optimality of a simulation estimator. This will be done in Section 2. The introduction of

these notions of efficiency leads to the question of how to construct importance sampling estimators that can be rigorously shown to be efficient. The zero-variance change-of-measure, while virtually impossible to implement in most situations of interest, provides a good guidance in order to construct such estimators.

The zero-variance change-of-measure turns out to be the conditional distribution of the underlying process given the rare event of interest. So, it is natural to leverage-off the available asymptotic theory for rare events in order to understand such a conditional distribution. This is basically the starting point in the design of efficient importance sampling estimators for rare events. The idea is to take advantage of the available large deviations theory, which provides insight into the occurrence of rare events. In the light-tailed case, the theory of large deviations is comprehensively documented in Dembo and Zeitouni (1998). References in connection to large deviations for heavy-tailed problems (mainly related to random walk settings) can be found in Embrechts et al. (1997) and Chapter IX of Asmussen (2000).

The successful application of large deviations theory in the design of efficient importance sampling estimators is not straightforward and this feature turned out to be the norm rather than the exception, especially in applications such as the analysis of queueing networks (see Glasserman and Kou 1995). To avoid such poor variance behavior, Dupuis and Wang formulate the design of state-dependent importance sampling estimators in terms of a stochastic control problem in which the value function is the second moment of the estimator. Taking an asymptotic limit, as we shall see in Section 3, the stochastic control problem becomes a deterministic control problem. The application of Dupuis and Wang's method is simplified thanks to the fact that only a subsolution to the value function is needed for the design of an efficient sampling scheme. A trivial supersolution is given, thanks to Jensen's inequality, by twice the underlying large deviations rate function. The whole point of Dupuis and Wang's approach is to construct a subsolution that matches the trivial supersolution at one point, namely, the initial condition of the system under consideration. In many problems of interest it turns out that such a subsolution is piecewise affine and therefore relatively easy to manipulate.

We will then move to the analysis of heavy-tailed systems, which will be introduced in our last section, namely Section 4. These types of systems arise in the analysis of finance and insurance models. There are fundamental qualitative differences between light and heavy tailed systems when it comes to the analysis of rare events. This is why it is incorrect to use a model with light-tailed components to study large deviations properties of systems that are built from heavy-tailed building blocks. Light-tailed large deviations occur gradually, while heavy-tailed large deviations occur suddenly and this feature manifests itself in the design of efficient importance sampling estimators for heavy-tailed systems.

We emphasize that this paper aims to introduce readers the key concepts and ideas in rare-event problems. Throughout the exposition we will drop many of the mathematical details and highlight only the central arguments, pointing readers to other references. Finally, we briefly mention the existence of some empirical approaches aimed at selecting good importance sampling distributions. We do not discuss these methods in detail but rather refer the reader to some related references including: adaptive Monte Carlo (Kollman, Baggerly, Cox, and Picard 1999) and the cross-entropy method (Rubinstein and Kroese 2004). In this vein, we mention the recent work of Chan and Lai (2010) who considers an interesting importance sampling approach for Markov-modulated problems that avoids the need for computing the associated eigenfunctions.

2 NOTIONS ON EFFICIENCY AND IMPORTANCE SAMPLING

2.1 Notions of Efficiency

In order to design algorithms that achieve high accuracy in relative terms and that perform well in rare event settings it is useful to embed the question of interest in a suitable asymptotic regime, often motivated by an applied standpoint. For example, in the setting of bankruptcy of an insurance company, it is often the case that the capital or reserve of the company is large relative to individual claim sizes, and this is basically why one might expect the probability of bankruptcy, $P(A)$, to be small. So, one might introduce

a parameter, say n, for instance the initial capital in the insurance example, and consider $\alpha_n = P(A_n)$ as a function of the parameter n which is often referred to as the "rarity parameter". From now on we will assume that underlying the rare event there is a rarity parameter n such that $\alpha_n \longrightarrow 0$ as $n \longrightarrow \infty$.

The idea is then to design a simulation estimator whose error is controlled in relative terms as $n \nearrow \infty$ compared to naive (or crude) Monte Carlo, which provides an obvious alternative and, therefore, a general benchmark. Note that naive Monte Carlo involves simulating N (independent and identically distributed) iid Bernoullies, $I_n(1), ..., I_n(N)$, with success parameter α_n, thereby producing an estimate of the form $\widehat{\alpha}_n^c(N) = \sum_{j=1}^N I_n(j)/N$. The super-index "$c$" that appears in $\widehat{\alpha}_n^c(N)$ corresponds to crude Monte Carlo. Since $\widehat{\alpha}_n^c(N)$ is an unbiased estimator of α_n, its mean squared error equals $Var(\widehat{\alpha}_n^c(N))^{1/2} = \alpha_n^{1/2}(1-\alpha_n)^{1/2}/N^{1/2}$. In turn, we obtain a relative mean squared error equal to the coefficient of variation of $\widehat{\alpha}_n^c(N)$, namely, $CV(\widehat{\alpha}_n^c(N)) = (1-\alpha_n)^{1/2}/(\alpha_n^{1/2}N^{1/2})$. In order to control the relative error of $\widehat{\alpha}_n^c(N)$ one needs $N = \Omega(1/\alpha_n)$ Bernoullies.

The problem with the estimator $\widehat{\alpha}_n^c(N)$ is that each of the underlying replications, the $I_n(j)$'s, have a very big variance in relative terms. The number of replications N must grow at the same rate as $CV(I_n(j)) = (1-\alpha_n)^{1/2}/\alpha_n^{1/2}$. The same phenomenon occurs in any estimator obtained out of averaging iid replications, as in the definition of $\widehat{\alpha}_n^c(N)$. The objective is then to design an estimator R_n with a controlled mean squared error. We concentrate only on unbiased estimators, so controlling the behavior of the mean squared error boils down to controlling the behavior of the second moment. The overall estimate is then obtained by averaging N iid copies of R_n. Motivated by these considerations we have the following definitions.

Definition 1 An estimator R_n is said to be *strongly efficient* if $ER_n^2 = O(\alpha_n^2)$ as $n \nearrow \infty$.

Definition 2 R_n is said to be *weakly efficient* or *asymptotically optimal* if for each $\varepsilon > 0$ we have that $ER_n^2 = O(\alpha_n^{2-\varepsilon})$ as $n \nearrow \infty$.

In most cases the analysis of importance sampling estimators (especially in light-tailed cases) concludes only weak efficiency of the estimators.

In order to see the gain in efficiency, suppose that an estimator R_n is strongly efficient and consider $\widetilde{\alpha}_n(N) = \sum_{i=1}^N R_n(i)/N$ where the $R_n(i)$'s are independent replications of R_n. It follows from Chebyshev's inequality that at most $N = O(\varepsilon^{-2}\delta^{-1})$ (uniformly in n) replications are required to conclude that $\widetilde{\alpha}_n(N)$ is ε-close to α_n in relative terms with at least $1 - \delta$ confidence.

Besides variance control, it is worth pointing out the other measures of good importance sampling estimators, and we refer interested readers to L'Ecuyer et al. (2010) for these considerations (including work-normalized measures to account for cost-per replication).

2.2 Basic Properties of Importance Sampling Estimators

Importance sampling is a variance reduction technique that is often applied to design an estimator, R_n, with efficiency characteristics such as those described by Definitions 1 and 2. We shall review basic properties of importance sampling next; for a more detailed discussion see, for example, Rubinstein and Kroese (2009) and Asmussen and Glynn (2007). The basic idea is to introduce a probability measure $\widetilde{P}(\cdot)$ such that the likelihood ratio or Radon-Nikodym derivative between the nominal (original) probability measure, $P(\cdot)$, and $\widetilde{P}(\cdot)$ is well defined on the event A_n of interest. Then we can simply let

$$R_n = \frac{dP}{d\widetilde{P}}(\omega) I(\omega \in A_n),$$

as usual, we use ω to denote the underlying random outcome which is simulated according to the probability measure $\widetilde{P}(\cdot)$. We use $\widetilde{E}(\cdot)$ to denote the expectation operator associated to the probability measure $\widetilde{P}(\cdot)$.

Note that

$$\widetilde{E} R_n = \int_{A_n} \frac{dP}{d\widetilde{P}}(\omega) \, d\widetilde{P}(\omega) = P(A_n) = \alpha_n,$$

so R_n is an unbiased estimator, and

$$\widetilde{E}(R_n^2) = \int_{A_n} \left(\frac{dP}{d\widetilde{P}}(\omega) \right)^2 d\widetilde{P}(\omega) = \int_{A_n} \frac{dP}{d\widetilde{P}}(\omega) \, dP(\omega) = E(R_n). \tag{1}$$

In principle one can design an importance sampling estimator with zero variance. Indeed, if we let $\widetilde{P}(\cdot) = P_n^*(\cdot) := P(\cdot \mid A_n)$, then

$$R_n^* := \frac{dP}{dP_n^*}(\omega) \, I(\omega \in A_n) = P(A_n),$$

which is clearly an exact estimate of $\alpha_n = P(A_n)$. Of course, assuming that one has access to $P^*(\cdot)$ defeats the point of using simulation to estimate $P(A_n)$. However, the underlying lesson behind the characterization of the zero-variance change-of-measure as the conditional distribution given the rare event of interest is that one can use asymptotic theory to describe $P_n^*(\cdot)$ as $n \nearrow \infty$.

3   STATE-DEPENDENT TECHNIQUES FOR LIGHT-TAILED SYSTEMS

3.1 Solutions to a Deterministic Control Problem and Efficient Importance Sampling

As we indicated in the Introduction, Dupuis and Wang introduced a technique to guarantee efficiency (in the sense of Definition 2) of importance sampling estimators based on control theoretic considerations.

We will formally characterize the solution to a deterministic control problem proposed by Dupuis and Wang (2004). In this section we assume that $(X_k : k \geq 1)$ is a sequence of iid rv's taking values in R^d and such that $\psi(\theta) = \log E \exp(\langle X_k, \theta \rangle)$ is finite everywhere, where $\langle \cdot, \cdot \rangle$ is defined as the inner product under Euclidean norm. We then let $S_k = X_1 + ... + X_k$ and consider, for some $y \in R^d$,

$$Y_n(t) = y + S_{\lfloor nt \rfloor}/n. \tag{2}$$

Let B be a closed set with non-empty interior and define $\tau_B(n) = \inf\{t > 0 : Y_n(t) \in B\}$. We assume that B is attainable in the sense that $P(\tau_B(n) < \infty | Y_n(0) = y) > 0$ for each $y \in R^d$.

Consider two disjoint sets B and C. We are interested in computing

$$\alpha_n(y) = P_y(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty) := P(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty | Y_n(0) = y).$$

We assume that B and C are regular in the sense that $\lim_{n \to \infty} n^{-1} \log \alpha_n(y) = -I_{B,C}(y)$, where $I_{B,C}(y)$ is the associated large deviations rate function (which we do not need to specify but can be consulted in Dembo and Zeitouni 1998, p. 176).

We are interested in choosing a suitable importance sampling estimator with optimal performance in the sense of minimizing its second moment. The class of policies that we consider are based on exponential tilting. In particular, the increment X_k is ultimately simulated according to the distribution

$$P_{\theta_k}(X_k \in dx) = \exp(\langle \theta_k, X_k \rangle - \psi(\theta_k)) P(X_k \in dx). \tag{3}$$

If X_k has a density with respect to the Lebesgue measure, say $f(\cdot)$, then $P(X_k \in dx) = f(x) \, dx$. The notation adopted in (3) is simply a mechanism that allows us to consider general distributions (including discrete or mixtures of discrete and continuous distributions). We will use $E_{\theta_k}(\cdot)$ to denote the expectation operator associated to $P_{\theta_k}(\cdot)$. The parameter θ_k might depend on $(X_1, ..., X_{k-1})$ but it is not allowed to depend on future observations. We say that θ_k is non-anticipating. As examples of exponentially tilted

distributions consider the following: If X_k is a one dimensional standard Gaussian under $P(\cdot)$, then, under $P_{\theta_k}(\cdot)$, X_k is normal with mean θ_k and unit variance; on the other hand, if under X_k is exponentially distributed with unit mean under $P(\cdot)$, then under $P_{\theta_k}(\cdot)$ (for $\theta_k < 1$), X_k is exponentially distributed with mean $1/(1 - \theta_k)$.

One reason that motivates the use of exponentially tilted distributions is that one can sometimes show that the distribution of $(X_1, ..., X_k)$ given a large deviations event of interest converges to a suitable optimal exponential tilting as $n \to \infty$ (see, for instance, Asmussen and Glynn 2007, Section VI.5).

Using exponential tilting the estimator ultimately takes the form

$$R_n = \exp(-\sum_{k=1}^{n\tau_B(n)} \langle \theta_k, X_k \rangle + \sum_{k=1}^{n\tau_B(n)} \psi(\theta_k))I(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty). \tag{4}$$

It follows immediately that the HJB (Hamilton-Jacobi-Bellman) equation corresponding to finding the optimal policy (i.e., the optimal sequence of θ_k's) to minimize the second moment of R_n takes the form

$$V_n(y) = \inf_\theta E_\theta[\exp(-2\langle \theta, X \rangle + 2\psi(\theta))V_n(y + X/n)]$$

$$= \inf_\theta E[\exp(-\langle \theta, X \rangle + \psi(\theta))V_n(y + X/n)] \tag{5}$$

for $y \notin B \cup C$ and subject to the boundary condition that $V_n(y) = 1$ for $y \in B$ and $V_n(y) = 0$ if $y \in C$. If there is an optimal policy $(\theta_k^* : k \geq 1)$ generated by the previous HJB equation, then generally it would be the case that

$$V_n(y) = E\exp(-\sum_{k=1}^{n\tau_B(n)} \langle \theta_k^*, X_k \rangle + \sum_{k=1}^{n\tau_B(n)} \psi(\theta_k^*))I(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty)$$

$$\leq E\exp(-\sum_{k=1}^{n\tau_B(n)} \langle \theta_k, X_k \rangle + \sum_{k=1}^{n\tau_B(n)} \psi(\theta_k))I(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty),$$

for any non-anticipating policy $(\theta_k : k \geq 1)$.

The large deviations scaling suggests writing $V_n(y) = \exp(-nH_n(y))$ and thus we should expect $H_n(y) \to H(y)$ as $n \to \infty$ for some function $H(y)$. If we proceed using this postulated limit in the previous HJB equation, after taking logarithms of (5), formally and without being careful about underlying smoothness assumptions and errors incurred, we arrive at the approximation

$$-nH(y) \approx \min_\theta \log E[\exp(-\langle \theta, X \rangle + \psi(\theta) - nH(y + X/n))] \tag{6}$$

$$\approx \min_\theta \log E[\exp(-\langle \theta, X \rangle + \psi(\theta) - nH(y) - \langle \nabla H(y), X \rangle)].$$

Equivalently, we have that

$$0 \approx \min_\theta \log E[\exp(-\langle \theta, X \rangle + \psi(\theta) - \langle \nabla H(y), X \rangle)]$$

$$= \min_\theta \log \exp(\psi(\theta) + \psi(-\nabla H(y) - \theta)) = \min_\theta[\psi(\theta) + \psi(-\nabla H(y) - \theta)]. \tag{7}$$

First order optimality conditions imply that at the optimal value $\theta^*(y)$ one has $\nabla \psi(\theta^*(y)) = \nabla \psi(-\nabla H(y) - \theta^*(y))$, which yields $\theta^*(y) = -\nabla H(y)/2$ and therefore we conclude that equation (7) can be expressed as

$$2\psi(-\nabla H(y)/2) = 0, \tag{8}$$

subject to the boundary conditions (inherited from (5)) equal to $H(y) = 0$ for $y \in B$ and $H(y) = \infty$ for $y \in C$. Equation (8) corresponds to a so-called Isaacs equation. The name originates from a connection that can be established to a solution of a two-player differential game, see Dupuis and Wang (2004).

On the other hand, we have assumed $\alpha_n(y) = \exp(-nI_{B,C}(y) + o(n))$ as $n \to \infty$ and, by conditioning on the first increment of the random walk, we also have that

$$\alpha_n(y) = E\alpha_n(y + X/n) \tag{9}$$

subject to the constraints that $\alpha_n(y) = 1$ for $y \in B$ and $\alpha_n(y) = 0$ if $y \in C$. Proceeding to analyze equality (9) formally as we did for the discussion leading to (8), we conclude that

$$\exp(-nI_{B,C}(y) + o(n)) = E\exp(-nI_{B,C}(y + X/n)) \approx E\exp(-nI_{B,C}(y) - \langle \nabla I_{B,C}(y), X \rangle).$$

Equivalently, taking logarithms we arrive at $\psi(-\nabla I_{B,C}(y)) = 0$, subject to the boundary conditions implied by $\alpha_n(y)$, namely, $I_{B,C}(y) = 0$ for $y \in B$ and $I_{B,C}(y) = \infty$ if $y \in C$. These considerations, together with our analysis leading to (8) yield that $H(y) = 2I_{B,C}(y)$ and therefore, under suitable regularity conditions we obtain that applying importance sampling with exponential tilting given by the tilting parameter $\theta^*(y) = -\nabla I_{B,C}(y)$ allows us to obtain an asymptotically optimal estimator (recall Definition 2). Sufficient conditions required to rigorously substantiate this result are given in Dupuis and Wang (2004), at least in the case of large deviations problems for S_n/n.

The problem with (8) is that it is rarely applicable in most situations of interest, especially because it is often the case that there is no classical (smooth) solution to the Isaacs equation (8). (It is apparent from our heuristic derivation that smoothness appears to be an important feature.) The function $H(y) = 2I_{B,C}(y)$ typically provides a solution only in a weak sense. Fortunately, as we shall see next, only a subsolution is required, as long as it coincides with the value function at the initial condition.

3.2 Subsolutions to a Deterministic Control Problem and Efficient Importance Sampling

We take as starting point the HJB equation introduced in (5). The next lemma provides an inequality whose solution gives an upper bound for the value function in (5); for a proof see Blanchet and Lam (2011).

Lemma 1 Suppose that one finds a non-negative function $U_n(\cdot)$ such that

$$U_n(y) \geq E[\exp(-\langle \theta, X \rangle + \psi(\theta))U_n(y + X/n)] \geq \inf_\theta E[\exp(-\langle \theta, X \rangle + \psi(\theta))U_n(y + X/n)],$$

where $y \notin B \cup C$ and subject to the boundary condition that $U_n(y) \geq 1$ for $y \in B$. Then,

$$U_n(y) \geq E[\exp(-\sum_{j=1}^{\tau_B}\langle \theta_j, X_j \rangle + \sum_{j=1}^{\tau_B}\psi(\theta_j))I(\tau_B < \infty, \tau_B < \tau_C)]$$

for any non-anticipating policy $(\theta_j : j \geq 1)$.

Using a similar heuristic argument as the one introduced in the previous section in (6) we write $U_n(y) = \exp(-nG_n(y))$ and postulate $G_n(y) \to G(y)$ as $n \to \infty$ for some function $G(y)$. If we proceed using this postulated limit in the inequality obtained in the previous lemma, after taking logarithms, we arrive at

$$-nG(y) \gtrsim \min_\theta \log E[\exp(-\langle \theta, X \rangle + \psi(\theta) - nG(y + X/n))]$$
$$\approx \min_\theta \log E[\exp(-\langle \theta, X \rangle + \psi(\theta) - nG(y) - \langle \nabla G(y), X \rangle)].$$

We then conclude, as in the analysis leading to (7) that

$$0 \geq \min_{\theta}[\psi(\theta) + \psi(-\nabla G(y) - \theta)]. \tag{10}$$

First order optimality conditions imply that at the optimal value $\theta^*(y)$ one has $\nabla\psi(\theta^*(y)) = \nabla\psi(-\nabla G(y) - \theta^*(y))$, which yields $\theta^*(y) = -\nabla G(y)/2$ and therefore we conclude that inequality (10) can be expressed as

$$0 \geq 2\psi(-\nabla G(y)/2), \tag{11}$$

for $y \notin B \cup C$ and subject to the boundary conditions (inherited from Lemma 1) equal to $G(y) \leq 0$ for $y \in B$. In order to conclude asymptotic optimality for $\alpha_n(y_0)$ we must have $G(y_0) \geq 2I_{B,C}(y_0)$. Sufficient conditions required to rigorously substantiate this subsolution equation are given in Theorem 8.1 of Dupuis and Wang (2007).

We now apply the previous result to an example, which is a generalization of a one dimensional ruin problem.

Example 1 (multidimensional first passage time problems) Assume that $d = 2$ and let us write $X_i^{(j)}$ to denote the j-th component of the i-th increment; in particular, we have that $j = 1, 2$. Similarly, we use $Y_n^{(i)}(t)$ to denote the i-th component of the process $Y_n(\cdot)$ at time t, recall equation 2. Assume that X_i follows a bivariate Gaussian distribution with mean $EX_i = (-1, -1)^T$ and variance $Var(X_i) = I$ (that is, the components of X_i are independent). Therefore, we have that $\psi(\theta) = ||\theta||^2/2$.

Define the set $B(i) = \{y = (y^{(1)}, y^{(2)})^T : y^{(i)} \geq 1\}$ and let $B = B(1) \cup B(2)$. Consider the first passage time probability

$$\alpha_n(y) = P(\tau_B(n) < \infty|Y_n(0) = y),$$

where $\tau_B(n) = \inf\{t \geq 0 : Y_n(t) \in B\}$. We will construct an efficient estimator for $\alpha_n(y)$ based on (11).

Similarly, we let $\tau_{B(i)}(n) = \inf\{t > 1 : Y_n^{(i)}(t) \geq 1\}$ for each $i = 1, 2$, and note that one has the elementary inequalities

$$\max_{i=1,2} P(\tau_{B(i)}(n) < \infty|Y_n(0) = y) \leq \alpha_n(y) \leq \sum_{i=1}^{2} P(\tau_{B(i)}(n) < \infty|Y_n(0) = y). \tag{12}$$

Evidently, $P(\tau_{B(i)}(n) < \infty|Y_n(0) = y)$ is equivalent to a one dimensional first passage time probability. It is well known (see Asmussen 2003 p. 352) that

$$P(\tau_{B(i)}(n) < \infty|Y_n(0) = y) = \exp[-n\theta_i^*(1 - y^{(i)}) + o(n)], \tag{13}$$

where $\theta_i^* = 2$. We then obtain from (12) that $\lim_{n\to\infty} n^{-1}\log\alpha_n(y) = -\min_{i=1,2}\{\theta_i^*(1 - y^{(i)})\}$. The value of θ_i^* corresponds to the conditional asymptotic mean of the $i - th$ component if the random walk hits B through $B(i)$.

To apply the subsolution approach we need to find a function $G(\cdot)$ satisfying the inequality (11) together with the corresponding boundary condition that $G(y) \leq 0$ for $y \in B$. Moreover, since in the end we are after an asymptotic upper bound for the second moment of the importance sampling estimator and we wish to achieve the same decay rate as $\alpha_n(y)^2$, it makes sense to use the union bound (12) and (13) to postulate

$$\exp(-nG(y)) = (\sum_{i=1}^{2}\exp[-n\theta_i^*(1 - y^{(i)})])^2.$$

We now verify that our selection is indeed a valid subsolution. First, note that $\nabla G(y) = -2\sum_{i=1}^{2}\theta_j^* a_j p_j(y)$, where $a_1 = (1, 0)$, $a_2 = (0, 1)$ and

$$p_j(y, n) = \frac{\exp[-n\theta_j^*(1 - y^{(j)})])}{\sum_{i=1}^{2}\exp[-n\theta_i^*(1 - y^{(i)})]}.$$

Note that for each y, $(p_1(y, n), p_2(y, n))$ can be interpreted as a probability vector. So, using Jensen's inequality and convexity of $\psi(\cdot)$ we conclude that

$$\psi(-\nabla G(y)/2) = \psi(\sum_{j=1}^{2} \theta_j^* a_j p_j(y, n)) \leq \sum_{j=1}^{2} p_j(y, n) \psi(\theta_j^* a_j) = 0.$$

In the last equality we used that $\psi(\theta_j^* a_j) = 0$ for $j = 1, 2$. Now, if $y \in B$, then there exists i such that $1 - y^{(i)} \leq 0$ and therefore evidently $\exp(-nG(y)) \geq 0$ or, equivalently, $G(y) \leq 0$. We then conclude that the importance sampling scheme generated by using as exponential tilting

$$\theta_k(S_{k-1}/n) = \sum_{j=1}^{2} \theta_j^* a_j p_j(S_{k-1}/n, n) \tag{14}$$

for $k/n < \tau_B(n)$ is asymptotically optimal; actually in this case the estimator can be shown to be strongly efficient in the sense of Definition 1.

4 STATE-DEPENDENT TECHNIQUES FOR HEAVY-TAILED SYSTEMS

As mentioned earlier, various challenges of rare-event simulation with heavy tails are pointed out in Asmussen et al. (2000) and Bassamboo et al. (2006). The former focuses on the singularity of the asymptotic conditional distribution with respect to the original measure and the difficulty of finding a suitable importance sampler, while suggesting provably efficient conditional Monte Carlo scheme. The latter suggests the non-existence of provably efficient state-independent importance samplers for some first passage problems. Conditional Monte Carlo is also studied in Asmussen and Binswanger (1997) in the context of insurance, and further improved in Asmussen and Kroese (2006). A technique called hazard rate twisting is studied in Juneja et al. (1999) and Juneja and Shahabuddin (2002).

Our goal here is to illustrate a technique, based on Lyapunov inequalities and appropriate mixture samplers to estimate rare event probabilities. The Lyapunov approach was proposed in Blanchet and Glynn (2008) and it has subsequently applied to various settings (see, for instance, Blanchet et al. 2007; Blanchet and Liu 2008; Blanchet and Liu 2010b). The mixture distributions that we shall use were proposed in Dupuis et al. (2006) in the context of power-law decaying tails; see also Blanchet and Liu (2010a) for extensions beyond power-law decaying tails.

We first start with some basic results on heavy-tailed large deviations.

4.1 Large Deviations for Heavy-tailed Sums

For simplicity we concentrate on a special case of heavy-tailed random variables, namely, regularly varying random variables, which posses power-law type tails; see Embrechts et al. (1997) for more information on more general heavy-tailed models.

Definition 3 A random variable X with tail distribution $\overline{F}(\cdot) = P(X > \cdot) = 1 - F(\cdot)$ has a regularly varying right tail with index $\alpha > 0$ if $\overline{F}(\beta x)/\overline{F}(x) \longrightarrow \beta^{-\alpha}$ as $x \nearrow \infty$ for all $\beta > 0$. Similarly, X has a regularly varying left tail with index $\alpha > 0$ if $-X$ has a regularly varying right tail with index $\alpha > 0$.

A function $L(\cdot)$ is slowly varying at infinity if $L(\beta x)/L(x) \longrightarrow 1$ as $x \to \infty$ for each $\beta > 0$. Therefore, if X has regularly varying right tail with index $\alpha > 0$ we often write $\overline{F}(t) = t^{-\alpha} L(t)$.

The following result provides insight on how large deviations tend to occur in heavy-tailed models.

Proposition 1 Suppose that $S_m = X_1 + ... + X_m$ where the X_i's are iid non-negative regularly varying rv's. Then $P(\max_{1 \leq j \leq m} X_j > n | S_m > n) \longrightarrow 1$ as $n \nearrow \infty$.

The previous proposition illustrates the so-called "catastrophe principle" behind extreme behavior of heavy-tailed systems, which postulates that large deviations are caused by extremes in one or few components. In this case, the sum is large because one single component, namely the largest one, is large.

In the heavy-tailed case there is a powerful heuristic that allows to "guess" a plausible form for large deviations results and the associated "most likely paths" to the rare event. This heuristic, which is known as the "fluid heuristic", is based on elementary Law of Large Numbers analysis and basic facts of regularly varying distributions. In order to guide our discussion we shall revisit a one dimensional analogue of the example studied earlier in the light tailed case.

Example 2 (first passage time probabilities for heavy-tailed random walk)   Suppose that the X_j's are regularly varying with index $\alpha > 1$. Assume that $EX_k = -\mu < 0$ and let $S_0 = y$. Define $\tau_n = \inf\{k \geq 0 : S_k > n\}$ and consider the problem of approximating the ruin probability

$$\alpha_n(y) = P\left(\tau_n < \infty | S_0 = y\right)$$

as $n \nearrow \infty$.

The fluid heuristic consists in substituting the random walk by its fluid or Law of Large Numbers behavior prior to the big jump that causes ruin. The associated fluid path in this case is $S_{\lfloor nt \rfloor}/n \approx y(t) = y - \mu t$ and therefore the fluid heuristic suggests

$$\alpha_n(y) \sim \sum_{k=0}^{\infty} P(X_{k+1} + y(k) > n) \sim \int_0^{\infty} P\left(X > n + \mu t - y\right) dt$$

as $n \nearrow \infty$, thereby neglecting the contribution of more than one jump and fluctuations beyond the Law of Large Numbers. Letting $u = n + \mu t - y$ we obtain

$$\alpha_n(y) \sim \frac{1}{\mu} \int_{n-y}^{\infty} P\left(X > u\right) dy.$$

as $n \nearrow \infty$. This approximation turns out to be correct even beyond the assumption of regularly varying service times (see Pakes 1975; Veraverbeke 1977; Embrechts and Veraverbeke 1982).

4.2 Conditional Distributions and Implications For Simulation

As we discussed in Section 2, it is useful to understand the conditional distribution of the random walk given the occurrence of the rare event of interest because such conditional distribution provides the optimal change-of-measure (in terms of variance minimization). A natural starting point would be to obtain a conditional limit theorem for the X_k's given the rare event of interest. The asymptotic conditional distribution would then be a natural candidate for an importance sampling distribution. One can pursue this idea in the context of Example 2. In such case, it is not difficult to verify that for any $-\infty < x_i^- < x_i^+ < \infty$, with $i = 1, ..., k$ we have that

$$\lim_{n \to \infty} P\left(\cap_{i=1}^k \{X_i \in (x_i^-, x_i^+)\} | \tau_n < \infty\right) = P\left(\cap_{i=1}^k \{X_i \in (x_i^-, x_i^+)\}\right).$$

Thus, concluding that the asymptotic conditional distribution of $(X_1, ..., X_k)$ given that $\tau_n < \infty$ remains unchanged as $n \to \infty$. This is one of the main reasons that makes the design of efficient importance sampling estimators for heavy-tailed systems a challenging problem.

Let us now take a more direct approach to approximate the increment X_{k+1} given the current position $S_k = s$ of the random walk at time k and given that $k < \tau_n < \infty$. We note (adopting the notation introduced in (3)) that

$$P\left(X_{k+1} \in dy | \tau_n \in [k+1, \infty), S_k = s\right)$$
$$= P\left(X_{k+1} \in dy, \tau_n = k+1 | \tau_n \in [k+1, \infty), S_k = s\right)$$
$$+ P\left(X_{k+1} \in dy, \tau_n > k+1 | \tau_n \in [k+1, \infty), S_k = s\right).$$

Now,

$$P\left(X_{k+1} \in dy, \tau_n = k+1 | \tau_n \in (k, \infty), S_k = s\right) = p_n^*(s) \frac{P\left(X_{k+1} \in dy\right) I(y > n - s)}{P\left(X_{k+1} > n - s | s\right)}, \quad (15)$$

where $p_n^*(s) = P\left(X_{k+1} > n - s\right) / \alpha_n(s)$. Using the approximation in Example 2, we have that

$$p_n^*(s) \sim p(n - s) := \mu \frac{P\left(X_{k+1} > n - s\right)}{\int_{n-s}^\infty P\left(X_{k+1} > t\right) dt} \sim \frac{\mu(\alpha - 1)}{(n - s)}$$

as $n - s \to \infty$. On the other hand, we have that

$$P\left(X_{k+1} \in dy, \tau_n > k+1 | \tau_n \in [k+1, \infty), S_k = s\right)$$
$$= \frac{P\left(X_{k+1} \in dy, \tau_n \in [k+2, \infty) | S_k = s\right) I(y \le n - s)}{\alpha_n(s)}$$
$$= \frac{P\left(X_{k+1} \in dy\right) I(y \le n - s)}{P\left(X_{k+1} \le n - s\right)} \times P\left(X_{k+1} \le n - s\right) \frac{\alpha_n(s + y)}{\alpha_n(s)}. \quad (16)$$

If we apply the fluid heuristic locally, that is by replacing X_{k+1} by its mean, namely $E X_{k+1} = -\mu$, we arrive at

$$\frac{\alpha_n(s + X_{k+1})}{\alpha_n(s)} \approx \frac{\alpha_n(s - \mu)}{\alpha_n(s)} \approx \frac{(n - s + \mu)^{-\alpha+1}}{(n - s)^{-\alpha+1}} \approx [1 + \mu/(n - s)]^{-\alpha+1} = 1 - \frac{\mu(1 - \alpha)}{n - s} + o(\frac{1}{n - s}).$$

Therefore, applying the previous local version of the fluid heuristic into (16) and combining this with (15) we arrive at the non-rigorous approximation

$$P\left(X_{k+1} \in dy | \tau_n \in [k+1, \infty), S_k = s\right) \approx \frac{P\left(X_{k+1} \in dy\right) I(y > n - s)}{P(X_{k+1} \le n - s)} p(n - s)$$
$$+ \frac{P\left(X_{k+1} \in dy\right) I(y \le n - s)}{P\left(X_{k+1} \le n - s\right)} [1 - p(n - s)]. \quad (17)$$

It turns out that this approximation is indeed valid in total variation (Blanchet and Liu 2010a) as $n \to \infty$.

4.3 State-dependent Importance Sampling Techniques for Heavy-tailed Random Walks

We now discuss Lyapunov inequalities, which were introduced in Blanchet and Glynn (2008) in the context of importance sampling, to show the efficiency of the family of samplers suggested by (17). This technique is closely related and in some sense parallel to the subsolution approach explained earlier in the light-tailed setting.

Lemma 2 Suppose that there exists $r\left(x, y\right) > 0$ satisfying

$$\int r^{-1}(x, y) P(S_j \in dy | S_{j-1} = x) = 1$$

for all x and $j \in \{1, 2, ...\}$. Then, we can define a Markov transition kernel via $K(x, dy) = r^{-1}\left(x, y\right) P(S_j \in dy | S_{j-1} = x)$ and an importance sampling estimator of the form

$$R = \prod_{j=1}^{T_A \wedge T_B} r(S_{j-1}, S_j) I(T_A < T_B).$$

Suppose that there exists a non-negative function $G\left(\cdot\right)$ and a constant $\rho \in (0, \infty)$ such that

$$E[r(x, S_1) G(S_1) | S_0 = x] \le G(x) \quad (18)$$

for $x \notin A \cup B$ and $G(x) \ge \rho$ for $x \in A$. Then, $E\left[R | S_0 = x\right] \le G(x)/\rho$.

The proof of the previous result is completely analogous to Lemma 1 (see Blanchet and Lam 2011). We now explain how to apply the previous result, combined with the fluid heuristics to design efficient importance sampling estimators in the context of Example 2.

Example 3 (Lyapunov inequalities and state-dependent importance sampling) We shall apply importance sampling following the analysis in Section 4.2. We distinguish two situations; first, if the current position s is sufficiently far away from the boundary n, say if $n - s > \Delta$ for some $\Delta > 0$, then the intuition developed in Section 4.2 is accurate and then we apply importance sampling using the mixture family.

However, in the second situation, if s is sufficiently close to n we do not need to apply importance sampling. Also, if we use the previous family when s is sufficiently close to n, we might create variance because the mixture sampling distribution is derived using the large deviations regime, so turning off the sampling when s is sufficiently close to n is a sensible defense mechanism. We introduce a boundary layer $0 \le n - s \le \Delta$ for some $\Delta > 0$ and if s lies in this layer no importance sampling is applied.

Our family, then, takes the form

$$\widetilde{P}(X_{k+1} \in dy | S_k = s) = \frac{P\left(X_{k+1} \in dy\right) I(y > a(n-s))}{P(X_{k+1} > a(n-s))} p(n-s) I(n - s > \Delta)$$

$$+ \frac{P\left(X_{k+1} \in dy\right) I(y \le a(n-s))}{P\left(X_{k+1} \le a(n-s)\right)} [1 - p(n-s)] I(n - s > \Delta)$$

$$+ P\left(X_{k+1} \in dy\right) I(n - s \le \Delta) .$$

The parameter $a \in (0, 1)$ is selected in order to enforce uniform integrability conditions, as we shall see. Following the intuition developed earlier we propose

$$p(n - s) = \theta P(X > a(n-s)) / \int_{n-s}^{\infty} P(X > t) dt,$$

for some $\theta > 0$. In fact, we expect $\theta \approx \mu$ to be optimal in the sense of variance minimization.

For notational convenience define $H(n - s) = \int_{n-s}^{\infty} P(X > t) dt$. We now need to construct a valid Lyapunov function. We wish to prove strong efficiency, so it is natural to suggest

$$G(n - s) = \min(\kappa H^2(n - s), 1) = O(\alpha_n(s)^2)$$

for some $\kappa > 0$. The strategy is then to select (given $a \in (0, 1)$) θ, κ and Δ in order to satisfy the Lyapunov inequality. If we can show that there exists a selection of these parameters that satisfies (18), then, given that the corresponding boundary condition in Lemma 2, namely

$$G(n - s) \ge \rho := \min(\kappa \int_0^{\infty} P(X > t) dt, 1) = \min(\kappa EX^+, 1) > 0$$

is satisfied for $s > n$, we would conclude strong efficiency of the estimator.

Note that the inequality (18) holds trivially if $G(n - s) = 1$ is equivalent to $n - s \le \Delta$. Indeed, in this case, since $G(x) \le 1$ for every $x \in R$, inequality (18) takes the form $EG(n - s - X) \le G(n - s) = 1$.

Now observe that $G(n - s) < 1$ holds if and only if $H(n - s) < 1/\kappa^{1/2}$, which in turn holds if and only if $n - s > H^{-1}(1/\kappa^{1/2})$. Therefore we can simply choose $\Delta = H^{-1}(1/\kappa^{1/2})$ and simply select θ and κ. In other words, we shall apply importance sampling using the mixture family if and only if our current position s is such that $G(n - s) < 1$. In order to proceed with the verification of inequality (18) we henceforth assume that $G(n - s) < 1$. In this case inequality (18) takes the form

$$J_1 + J_2 \le 1, \tag{19}$$

where

$$J_1 = \frac{E[G(n-s-X); X > a(n-s)]P(X > a(n-s))}{\kappa H^2(n-s)p(n-s)},$$

$$J_2 = \frac{E[G(n-s-X); X \le a(n-s)]P(X \le a(n-s))}{\kappa H^2(n-s)[1-p(n-s)]}.$$

We first bound the term J_1. Observe, noting that $G(\cdot) \le 1$, and using the form of $p(n-s)$, we obtain that

$$J_1 \le \frac{P(X > a(n-s))}{\theta \kappa H(n-s)} = \frac{p(n-s)}{\theta^2 \kappa}.$$

Then we analyze the term J_2. Note that

$$J_2 \le \frac{E[H^2(n-s-X); X \le a(n-s)]}{H^2(n-s)[1-p(n-s)]}.$$

Using the a Taylor expansion with remainder, expressed using elementary calculus with $t = x + yu$, we obtain

$$H^2(x+y) - H^2(x) = \int_x^{x+y} 2H(t)\dot{H}(t)\, dt = 2E[yH(x+yU)\dot{H}(x+yU)],$$

where U is uniformly distributed in $[0,1]$. Therefore, we have, letting $x = (n-s)$ and $y = -X$,

$$\frac{E[H^2(n-s-X)I(X \le a(n-s))]}{H^2(n-s)}$$

$$= P(X \le a(n-s)) - 2\frac{E[X\dot{H}(n-s-XU)H(n-s-XU); X \le a(n-s)]}{H^2(n-s)},$$

where X and U are independent. Now,

$$-\frac{E[X\dot{H}(n-s-XU)H(n-s-XU); X \le a(n-s)]}{H^2(n-s)}$$

$$= \frac{E[X\overline{F}(n-s-XU)H(n-s-XU); X \le a(n-s)]}{H^2(n-s)}.$$

We are assuming that $G(n-s) < 1$, or equivalently, that $n-s > H^{-1}(1/\kappa^{1/2}) = \Delta$. On the other hand, we have that

$$X\frac{\overline{F}(n-s-XU)H(n-s-XU)I(X \le a(n-s))}{H^2(n-s)} \sim X\frac{\alpha-1}{(n-s)}$$

almost surely as $n-s \nearrow \infty$ and also we have, because of regular variation and Karamata's theorem (see for example Resnick 2004 p. 17) that there exists a constant $K \in (0,\infty)$ such that

$$|X|\frac{\overline{F}(n-s-XU)H(n-s-XU)I(X \le a(n-s))}{H^2(n-s)}$$

$$\le |X|\frac{\overline{F}((n-s)(1-a))H((n-s)(1-a))}{H^2(n-s)} \le \frac{K|X|}{n-s}.$$

Consequently, if κ (or equivalently $\Delta = H^{-1}(1/\kappa^{1/2})$) is chosen sufficiently large, we conclude, applying the dominated convergence theorem (note here that $a \in (0,1)$ has been crucial) that for each $\varepsilon > 0$

$$J_2 \leq 1 + 2EX\frac{(\alpha - 1)}{n - s}(1 - \varepsilon)$$

as long as $G(n - s) < 1$. Combining our estimates for J_1 and J_2 and choosing κ sufficiently large so that

$$p(n - s) \leq \theta a^{-\alpha}\frac{(\alpha - 1)(1 + \varepsilon)}{n - s}, \quad \frac{1}{1 - p(n - s)} \geq 1 + \theta a^{-\alpha}\frac{(\alpha - 1)(1 - \varepsilon)}{n - s},$$

we arrive at

$$J_1 + J_2 \leq \frac{(\alpha - 1)(1 + \varepsilon)}{\theta\kappa(n - s)}a^{-\alpha} + (1 - 2\mu\frac{(\alpha - 1)}{n - s}(1 - \varepsilon))(1 + \theta a^{-\alpha}\frac{(\alpha - 1)(1 - \varepsilon)}{n - s})$$
$$\leq a^{-\alpha}\frac{(\alpha - 1)(1 + \varepsilon)}{\theta\kappa(n - s)} + (1 - 2\mu\frac{(\alpha - 1)}{n - s}(1 - \varepsilon) + \theta a^{-\alpha}\frac{(\alpha - 1)(1 - \varepsilon)}{n - s}).$$

In order to enforce (19) we then need to ensure that

$$a^{-\alpha}\frac{1 + \varepsilon}{\theta\kappa} - 2\mu(1 - \varepsilon) + a^{-\alpha}\theta(1 - \varepsilon) < 0,$$

which clearly feasible given that $\mu > 0$. Note that it is crucial that $-\mu = EX < 0$, otherwise it would be impossible to select the parameters appropriately. This makes sense given that when $EX \geq 0$, $\alpha_n(y) = 1$ and it would be therefore impossible to bound the second moment of the estimator by a function that decreases to zero as $n \to \infty$, as we are proposing. See Blanchet and Liu (2010a) for further discussion on the associated running time of the sampler.

ACKNOWLEDGEMENT

REFERENCES

Asmussen, S. 1985. "Conjugate processes and the simulation of ruin problems". *Stochastic Processes and their Applications* 20:213–229.

Asmussen, S. 2000. *Ruin Probabilities*. London, UK: World Scientific.

Asmussen, S. 2003. *Applied Probability and Queues*. New York, NY, USA: Springer-Verlag.

Asmussen, S., and K. Binswanger. 1997. "Simulation of Ruin probabilities for Subexponential Claims". *ASTIN Bulletin* 27:297–318.

Asmussen, S., K. Binswanger, and B. Hojgaard. 2000. "Rare events simulation for heavy-tailed distributions". *Bernoulli* 6:303–322.

Asmussen, S., and P. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. New York, NY, USA: Springer-Verlag.

Asmussen, S., and D. Kroese. 2006. "Improved algorithms for rare event simulation with heavy tails". *Advances in Applied Probability* 38:545–558.

Asmussen, S., and R. Rubinstein. 1995. "Steady state rare events simulation in queueing models and its complexity properties". In *Advances in Queueing*, edited by J. Dshalalow, 429–461. CRC Press.

Bassamboo, A., S. Juneja, and A. Zeevi. 2006. "On the inefficiency of state-independent importance sampling in the presence of heavy-tails". *Operations Research Letters* 34:521–531.

Blanchet, J., and P. Glynn. 2008. "Efficient rare-event simulation for the maximum of a heavy-tailed random walk.". *Ann. of Appl. Probab.* 18:1351–1378.

Blanchet, J., P. Glynn, and J. C. Liu. 2007. "Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed $G/G/1$ queue". *QUESTA* 57:99–113.

Blanchet, J., and H. Lam. 2011. "State-dependent importance sampling for rare-event simulation: An overview and recent advances". *under review in Surveys in Operations Research and Management Sciences*.

Blanchet, J., and J. C. Liu. 2008. "State-dependent importance sampling for regularly varying random walks". *Adv. in Appl. Probab.* 40:1104–1128.

Blanchet, J., and J. C. Liu. 2010a. "Efficient simulation and conditional functional limit theorems for ruinous heavy-tailed random walks". *submitted, http://www.columbia.edu/ jb2814/papers/ConcaveHazard9f.pdf*.

Blanchet, J., and J. C. Liu. 2010b. "Importance sampling in ruin problems for multidimensional regularly varying random walks". *Journal of Applied Probability* 47:301–322.

Bucklew, J. 1990. *Large Deviation Techniques in Decision, Simulation and Estimation*. New York, NY, USA: Wiley.

Chan, H. P., and T. L. Lai. 2010. "A sequential Monte Carlo approach to computing tail probabilities in stochastic models". *Ann. of Appl. Prob., forthcoming*.

Collamore, J. 2002. "Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random vectors". *Ann. Appl. Probab.* 12:382–421.

Dembo, A., and O. Zeitouni. 1998. *Large deviations techniques and applications, 2nd edition*. New York, NY, USA: Springer.

Dupuis, P., K. Leder, and H. Wang. 2006. "Importance sampling for sums of random variables with regularly varying tails.". *ACM TOMACS* 17.

Dupuis, P., K. Leder, and H. Wang. 2007. "Large deviations and importance sampling for a tandem network with slowdown.". *QUESTA* 57:71–83.

Dupuis, P., A. Sezer, and H. Wang. 2007. "Dynamic importance sampling for queueing networks.". *Ann. Appl. Probab.* 17:1306–1346.

Dupuis, P., and H. Wang. 2004. "Importance sampling, large deviations, and differential games". *Stoch. Stoch. Reports* 76:481–508.

Dupuis, P., and H. Wang. 2007. "Subsolutions of an Isaacs equation and efficient schemes of importance sampling". *Mathematics of Operations Research* 32:723–757.

Embrechts, P., C. Klppelberg, and T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. New York: Springer-Verlag.

Embrechts, P., and N. Veraverbeke. 1982. "Estimates for the probability of ruin with special emphasis on the possibility of large claims". *Insurance: Math. Economics* 1:55–72.

Glasserman, P., and S. Kou. 1995. "Analysis of an importance sampling estimator for tandem queues". *ACM Trans. Mod. Comp. Sim.* 5:22–42.

Glasserman, P., and Y. Wang. 1997. "Counter examples in importance sampling for large deviations probabilities". *Annals of Applied Probability* 7:731–746.

Juneja, S., and P. Shahabuddin. 2002. "Simulating heavy tailed processes using delayed hazard rate twisting". *ACM Trans. Model. Comput. Simul.* 12 (2): 94–118.

Juneja, S., and P. Shahabuddin. 2006. "Rare event simulation techniques: An introduction and recent advances". *Handbook on Simulation*:291–350.

Juneja, S., P. Shahabuddin, and A. Chandra. 1999, December. "Simulating heavy tailed processes using delayed hazard rate twisting". In *Proceedings of the 1999 Winter Simulation Conference*, edited by P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. Evans, 420–427. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kollman, C., K. Baggerly, D. Cox, and R. Picard. 1999. "Adaptive importance sampling on discrete Markov chains". *Annals of Applied Probability* 9:391–412.

L'Ecuyer, P., J. Blanchet, B. Tuffin, and P. Glynn. 2010. "Asymptotic robustness of estimators in rare-event simulation". *ACM TOMACS* 20:6:1–6:41.

L'Ecuyer, P., F. L. Gland, P. Lezaud, and B. Tuffin. 2009. "Splitting techniques". In *Rare Event Simulation Using Monte Carlo Methods*, edited by G. Rubino and B. Tuffin, 39–62. West Sussex, United Kingdom: Wiley. Chapter 3.

Pakes, A. G. 1975. "On the tails of waiting-time distribution". *J. Appl. Probab.* 12:555–564.

Parekh, S., and J. Walrand. 1989. "Quick simulation of rare events in networks". *IEEE Trans. Automat. Contr.* 34:54–66.

Resnick, S. I. 2004. *Extreme Values, Regular Variation, and Point Processes*. New York, USA: Springer.

Rubinstein, R., and D. Kroese. 2004. *The Cross-Entropy Method*. New York, NY: Springer.

Rubinstein, R. Y., and D. P. Kroese. 2009. *Simulation and the Monte Carlo Method, 2nd Edition*. John Wiley.

Sadowsky, J. S. 1991. "Large deviations theory and efficient simulation of excessive backlogs in a $GI/GI/m$ queue". *IEEE Trans. Autom. Control* 36:1383–1394.

Siegmund, D. 1976. "Importance sampling in the Monte Carlo study of sequential tests". *Ann. Stat.* 4:673–684.

Veraverbeke, N. 1977. "Asymptotic behavior of Wiener-Hopf factors of a random walk". *Stochastic Process. Appl.* 5:27–31.

AUTHOR BIOGRAPHIES

JOSE BLANCHET is a faculty member of the IEOR Department at Columbia University. Jose holds a Ph.D. in Management Science and Engineering from Stanford University. Prior to joining Columbia he was a faculty member in the Statistics Department at Harvard University. Jose is a recipient of the 2009 Best Publication Award given by the INFORMS Applied Probability Society and of the 2010 Erlang Prize. He also received a PECASE award sponsored by NSF in 2010. He worked as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability and Monte Carlo methods. He serves in the editorial board of Advances in Applied Probability, Journal of Applied Probability, Mathematics of Operations Research and QUESTA.

HENRY LAM is an Assistant Professor in the Department of Mathematics and Statistics at Boston University. He recently graduated from Harvard University with a Ph.D. degree in statistics. His research interests lie in applied probability and Monte Carlo methods with applications in queueing, operations management and insurance modeling.