Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, eds.

EFFICIENT IMPORTANCE SAMPLING UNDER PARTIAL INFORMATION

Henry Lam

Boston University
111 Cummington Street
Boston, MA 02215, USA

ABSTRACT

Importance sampling is widely perceived as an indispensable tool in Monte Carlo estimation for rare-event problems. It is also known, however, that constructing efficient importance sampling scheme requires in many cases a precise knowledge of the underlying stochastic structure. This paper considers the simplest problem in which part of the system is not directly known. Namely, we consider the tail probability of a monotone function of sum of independent and identically distributed (i.i.d.) random variables, where the function is only accessible through black-box simulation. A simple two-stage procedure is proposed whereby the function is learned in the first stage before importance sampling is applied. We discuss some sufficient conditions for the procedure to retain asymptotic optimality in well-defined sense, and discuss the optimal computational allocation. Simple analysis shows that the procedure is more beneficial than a single-stage mixture-based importance sampler when the computational cost of learning is relatively light.

1 INTRODUCTION

Importance sampling has been well documented as one of the most useful tools in Monte Carlo estimation related to rare events. In many settings, the importance sampler is constructed so as to mimic the zero-variance estimator, typically constructed by analyzing the sample path conditional on the occurrence of the rare event of interest. Plenty of literature is devoted to developing the techniques in various systems; see, for example, Bucklew (2004), Juneja and Shahabuddin (2006), Heidelberger (1995), and Blanchet and Lam (2012) for general overview. The main analytical tools for such development include large deviations theory (some recent work for tackling general problems under this framework includes Dupuis and Wang 2007 and Blanchet and Glynn 2008), and adaptive approximation and h-transform (see, for example, L'Ecuyer and Tuffin 2008 and Kollman et. al. 1999). Generally speaking, successful application of importance sampling typically requires a sufficient level of knowledge on the problem structure. This also means, on the negative side, that a poor judgement can substantially increase or even blow up the variance (Glasserman and Kou 1995). Such a high sensitivity poses even greater emphasis on enough knowledge of the system structure.

As a classical example, consider the large deviations probability of sum of one-dimensional i.i.d. random variables. Under light-tail assumption, this large deviations is exponentially decaying, with the exponential constant or so-called rate function dependent on the distribution of the summands and the cross-level of the probability. To construct an efficient importance sampler, these information for determining the rate function are needed in precision. If, for example, the target cross-level used by the importance sampler is lower than the actual level, the variance of the resulting estimator can in fact grow exponentially in the rarity parameter.

Our goal in this paper is to mathematically analyze the scenario when part of the system is not analytically known, and as a result the parameter in the importance sampler cannot be accurately specified. We assume, however, that knowledge on this missing part is accessible through data collection or simulation. The goal as stated is of course very broad, so as a first attempt we shall confine our discussion on a concrete problem regarding sum of i.i.d. random variables; namely, we consider the large deviations probability for

978-1-4673-4781-5/12/$31.00 ©2012 IEEE

an unknown function of this i.i.d. sum. To be more precise, let us consider

$$P(\mu(S_n) > y_n) \tag{1}$$

where $S_n = \sum_{i=1}^{n} U_i$ is a sum of zero-mean i.i.d. random variables, $\mu(\cdot)$ is a well-behaved function, and y_n is a sequence of cross-level, with $y_n = cn$ for some constant $c > 0$. The quantity n acts as the "rarity" parameter such that the probability (1) goes to zero as $n \to \infty$.

We assume complete knowledge about U_i i.e., we know the analytical form of the distribution function of U_i and also how to simulate U_i efficiently. The interesting feature about problem (1) lies in the function $\mu(\cdot)$, which is assumed to lack analytical closed-form expression, but nevertheless can be evaluated by inputting arbitrary value to get, via a "black box", the mapped function value subject to random noise. In other words, we can simulate (or "collect data" from) a sequence of

$$Y_t = \mu(X_t) + Z_t(X_t)$$

where X_t is the input at each time, and $Z_t(\cdot)$ is random noise depending on the input. To get an example for practical motivation, one can think about nested simulation in financial risk management. The S_n in (1) can be regarded as the random walk modeling an economic factor or stock movement, and we are interested in estimating the tail risk of say an option portfolio. In this case $\mu(S_n)$ is the sum of market values of a set of options at underlying price S_n. If the options are non-standard and have no closed-form expression, Monte Carlo method is necessary in estimating their values, and hence the value of $\mu(S_n)$. In this case $\mu(\cdot)$ is not directly known in closed-form but can be simulated. For convenience, let us call the function $\mu(\cdot)$ the *mean function* throughout this manuscript.

To further our discussion, we introduce some assumptions for this mean function $\mu(\cdot)$. It is worth pointing out that the model and assumptions considered in this paper are by no means practical in real applications, but serve to facilitate a transparent illustration of our analysis. First, we assume that $\mu(\cdot)$ is smooth enough and monotonic:

Assumption 1 The mean function $\mu(\cdot)$ is differentiable with bounded positive derivative i.e., $0 < \gamma_1 \leq \mu'(x) \leq \gamma_2$ for all $x \in \mathbb{R}$. The bounds γ_1 and γ_2 are known.

Let us denote Σ as the class of functions that satisfy Assumption 1, which implies in particular that $\mu(\cdot)$ is strictly increasing and grows steadily. This assumption, though restrictive, is natural for carrying out nonparametric procedure. In fact, a shifted version of $\mu(\cdot)$ can be shown to be quasi-linear, which leads to tractable large deviations behavior (Woodroofe 1972). Moreover, note that when $\mu(\cdot)$ is a known function, the problem can be reduced to $P(S_n > \xi_n)$ where ξ_n is the root of $\mu(\xi_n) = y_n$, which is the standard large deviations problem for i.i.d. sum.

Our focus in this paper is on the amount of information we need to know about $\mu(\cdot)$, quantified via the number of simulation trials or "data", such that one can still retain efficiency of the importance sampler. Learning the function takes up computation or other effort, so there is an intrinsic tradeoff on resource allocation to learning versus carrying out the importance sampler. We shall propose and provide analysis on this tradeoff for a simple two-stage procedure: learn about the function $\mu(\cdot)$ in the first stage, and use the gathered information to devise the importance sampler in the second stage. Throughout the paper, we shall use a minimax framework on the class Σ to address the above question.

Other than a simple two-stage procedure, another method is a uniformly efficient mixture-based importance sampler, analyzed in Glasserman and Juneja (2008). The purpose of their algorithm was to tackle the problem of simultaneous estimation for rare-event probabilities over a range of cross-levels, also under a minimax framework. They consider a mixture of exponential tilting schemes with different tilting parameters, and choose the best mixture probabilities and tilting parameters i.e., with minimax risk. Although the initial motivation was different, their algorithm can be readily applied to our scenario where the range of cross-levels is now inferred from the class Σ of mean functions. We will provide a comparison of this mixture algorithm with the simple two-stage approach we propose, and demonstrate

under what situation our algorithm works better. Intuitively, the two-stage procedure is preferable when the computational cost for learning is relatively light; in such case, the mixture algorithm, which is designed to work well under the whole class Σ, sacrifices the efficiency that can otherwise be recovered by learning more about $\mu(\cdot)$ with little computational cost. Lastly, we should mention that adaptive approaches, such as the cross-entropy method (Rubinstein and Kroese(2004)), can plausibly be applied to further improve the performance; this will be a subject of further research.

We also note the paper by Zhang et. al. (2007), which points out the issue of complex system description that hinders the design of optimal variance reduction algorithms for rare-event problems. There they focus on the first hitting probabilities of Markov processes that are analytically complex, and propose algorithms that perform favorably relative to crude Monte Carlo yet are unlikely to be optimal. Olvera-Cravioto (2007) studies the relation on estimation error between the truncation levels of heavy-tailed service times and the large deviations thresholds for single-server queues. Finally, we also mention the paper of L'Ecuyer et. al. (2010) that studies robustness properties of various rare-event estimators regarding their higher moments.

We state our assumptions on the i.i.d. summands and noise process as follows:

Assumptions on i.i.d. random variables. We assume $S_n = \sum_{i=1}^n U_i$, where $U_i, i = 1, \ldots, n$ are i.i.d. zero-mean random variables on $\mathbb{R}$ with light tail i.e., $\psi(\theta) := \log Ee^{\theta U_i} < \infty$ for θ in a neighborhood of 0. Let $\text{Dom}(\psi) = \{\theta : \psi(\theta) < \infty\}$ be the domain of ψ. We assume that

Assumption 2 (Light Tail) The logarithmic moment generating function $\psi(\cdot)$ is twice continuously differentiable and is steep on its positive domain i.e., $\psi'(\theta) \to \infty$ as θ goes to the positive boundary of $\text{Dom}(\psi)$.

Assumption 3 (Smoothness) We have $\int_{-\infty}^{\infty} |\phi_U(\theta)| d\theta < \infty$ where $\phi_U(\cdot) = Ee^{i\theta U_j}$ is the characteristic function of U_j. Hence U_j has density on the real axis.

Assumptions on noise process. The noises $Z_t(\cdot)$ are assumed to satisfy:

Assumption 4 The noise process $Z_t(\cdot)$ is a sequence of i.i.d. random functions. Denote $F(z; x)$ as the distribution function of $Z_t(x)$ given x, with

$$\psi(\theta; x) := \int e^{\theta z} F(dz; x) \leq c_1$$

for some constant $c_1 > 0$, for $-2h_1 \leq \theta \leq 2h_1$ for some $h_1 > 0$.

In other words, the moment generating function of the noises is uniformly bounded over $x \in \mathbb{R}$ for θ close to 0. The constants c_1 and h_1 are not necessarily known. However, knowledge about the bounds γ_1 and γ_2 in Assumption 1 is required.

The rest of the paper is as follows. In Section 2 we describe our two-stage procedure, followed by our main result in Section 3. Section 4 is devoted to a brief discussion. All proofs are left to the appendix.

2 TWO-STAGE PROCEDURE AND EFFICIENCY CRITERION

As mentioned before, since $\mu(\cdot)$ is strictly increasing, the probability $P(\mu(S_n) > y_n)$ can be rewritten as $P(S_n > \xi_n)$ where ξ_n is the solution to $\mu(\xi_n) = y_n$. Since the mean function $\mu(\cdot)$ is unknown, one would have to estimate the root ξ_n. Consequently, we use a natural two-stage approach. In Stage 1, we convert the problem into the form $P(S_n > \hat{\xi}_n^{(t)})$ by solving the equation $\mu(\xi_n) = y_n$ using standard Robbins-Monro procedure. Here $\hat{\xi}_n^{(t)}$ is an estimate of ξ_n using t iterations (we sometimes abbreviate as $\hat{\xi}_n = \hat{\xi}_n^{(t)}$ when no confusion arises). In Stage 2, we take the value $\hat{\xi}_n$ as the cross-level and carry out standard state-independent exponential tilting according to $\hat{\xi}_n$.

As the computational resources are not fully devoted to Monte Carlo simulation of the probability, we define logarithmic efficiency, or asymptotic optimality, by a criterion that involves the overall computational allocation:

Definition 1 A sequence of scheme, parametrized by n, that is used to generate an estimator Z_n for the probability of interest is called asymptotically optimal if given any $\epsilon > 0$, a subexponential (in n) expected amount of computational resources is enough to guarantee that the worst-case relative mean square error

$$\sup_{\mu \in \Sigma} \frac{MSE(Z_n)}{P(\mu(S_n) > y_n)^2} < \epsilon \tag{2}$$

where $MSE(Z_n)$ denotes the mean square error $E(Z_n - P(\mu(S_n) > y_n))^2$.

By Markov inequality, criterion (2) also ensures that the ratio of $Z_n - P(\mu(S_n) > y_n)$ to $P(\mu(S_n) > y_n)$ deviates from a given constant with controlled probability.

To facilitate our discussion, we assume that each iteration of the Robbins-Monro procedure in Stage 1 consumes $\gamma(n)$ computational resource, and each sample of the importance sampler in Stage 2 consumes $\lambda(n)$ computational resource. The quantity $\gamma(n)$ and $\lambda(n)$ are taken to be subexponential in n. In fact, regarding each arithmetic operation and generation of random variable as one unit of computation, then typically $\gamma(n) = O(1)$ and $\lambda(n) = O(n)$. Note that when $\mu(\cdot)$ is known, Stage 1 is not necessary and the number of simulation trials in Stage 2 needed to sustain (2) is $O(\sqrt{n})$ (by the exact asymptotic for i.i.d. sum; see Bucklew 2004), which gives an expected computational cost of $O(\lambda(n)\sqrt{n})$.

To get some intuition about $MSE(Z_n)$, we write

$$\begin{aligned} MSE(Z_n) &= Var(Z_n) + (EZ_n - P(\mu(S_n) > y_n))^2 \\ &= E[Var(Z_n|\hat{\xi}_n)] + Var(E[Z_n|\hat{\xi}_n]) + (E[E[Z_n|\hat{\xi}_n]] - P(S_n > \xi_n))^2 \\ &= E[Var(Z_n|\hat{\xi}_n)] + MSE(E[Z_n|\hat{\xi}_n]) \\ &= E[Var(Z_n|\hat{\xi}_n)] + MSE(P(S_n > \hat{\xi}_n|\hat{\xi}_n)). \end{aligned} \tag{3}$$

The expression (3) can be interpreted as the decomposition of mean square error into Stages 1 and 2. The first term captures the contribution to the mean square error due to estimation by importance sampling at Stage 2, taken as its variance averaged over the value of $\hat{\xi}_n$ obtained at Stage 1. The second term of (3) is the mean square error contributed by Stage 1, which can be interpreted as the expected squared bias due to Stage 1.

To close this section, we list our assumptions on the Robbins-Monro procedure in Stage 1 for estimating the solution of $\mu(\xi_n) = y_n$, where $y_n = cn$. Recall that we can observe $Y_t = \mu(X_t) + Z_t(X_t)$. We use the updating rule $X_{t+1} = X_t - a_t[Y_t - y_n]$. Also, note that the assumption $\gamma_1 \leq \mu'(\cdot) \leq \gamma_2$ implies that $m_1 n \leq \xi_n \leq m_2 n$ for $m_1 = c/\gamma_2$ and $m_2 = c/\gamma_1$. Therefore, we truncate the estimate of $\hat{\xi}_n$ at the end of the Robbins-Monro procedure to the closest point in $[m_1 n, m_2 n]$. This truncation procedure can only improve our estimate of $\hat{\xi}_n$ from a risk-theoretic perspective.

We assume the step size a_t and the initial value X_0 satisfy the following assumptions:

Assumption 5 $a_t = a/t$ where the constant a satisfies $a\gamma_1 \geq 1$, and γ_1 is the constant in Assumption 1.

Assumption 6 $Ee^{\theta X_0} < C(n)$ for θ in a neighborhood of 0, and $C(n)$ grows at most exponentially in n.

Assumption 6 is natural. For example, X_0 can be taken to be merely 0, or $c_0 n$, where c_0 is a constant hoping to match up the growth of y_n. The choice of X_0 indeed does not affect qualitatively the large deviations behavior of $\hat{\xi}_n$, and hence the overall mean square error, as shown by Theorem 2 in the sequel.

3 MAIN RESULTS

In this section we report and discuss a performance bound on our two-stage procedure.

3.1 Bound on Relative Mean Square Error

We denote t as the number of iterations in Stage 1 and m as the number of simulation using exponential tilting in Stage 2. Our main result is the following:

Theorem 1 With Assumptions 1–6, and suppose the number of Stage 1 iterations $t = \omega(n)$ and $t = o(e^{cn})$ for any $c > 0$ i.e., t is superlinear but subexponential. The relative mean square error satisfies

$$\sup_{\mu \in \Sigma} \frac{MSE(Z_n)}{P(\mu(S_n) > y_n)^2} = O\left(\frac{\sqrt{n}}{m} + \frac{n}{t}\right). \tag{4}$$

This in particular shows that the two-stage scheme, by choosing m to be $\omega(\sqrt{n})$ and $o(e^{cn})$ and t chosen as in the theorem, is asymptotically optimal according to Definition 1.

3.2 Optimal Allocation

Now suppose we are given a computational budget H (possibly dependent on n). The bound (4) gives a way to check if an allocation scheme achieves asymptotic optimality as in Definition 1. It also allows one to minimize the mean square error via suitable resource allocation. Recall that the expected computational cost per Stage 1 iteration is $\gamma(n)$ and that for each Stage 2 simulation is $\lambda(n)$. We then consider the minimization of $C_1\sqrt{n}/m + C_2 n/t$ under the constraint $t\gamma(n) + m\lambda(n) \le H$, where C_1 and C_2 are some positive constants. Applying Lagrange multiplier α gives $C_1\sqrt{n}/m^2 = \alpha\lambda(n)$ and $C_2 n/t^2 = \alpha\gamma(n)$. Upon solving, we get

$$m = \frac{\sqrt{C_1} H n^{1/4}}{(\sqrt{C_1\lambda(n)}n^{1/4} + \sqrt{C_2\gamma(n)n})\sqrt{\lambda(n)}} \quad \text{and} \quad t = \frac{\sqrt{C_2} H \sqrt{n}}{(\sqrt{C_1\lambda(n)}n^{1/4} + \sqrt{C_2\gamma(n)n})\sqrt{\gamma(n)}}.$$

This gives the minimal relative mean square error

$$\begin{aligned} & O\left(\frac{n^{1/4}\sqrt{\lambda(n)}(\sqrt{C_1\lambda(n)}n^{1/4} + \sqrt{C_2\gamma(n)n})}{H} + \frac{\sqrt{n}\sqrt{\gamma(n)}(\sqrt{C_1\lambda(n)}n^{1/4} + \sqrt{C_2\gamma(n)n})}{H}\right) \\ = \; & O\left(\frac{(\sqrt{\lambda(n)}n^{1/4} + \sqrt{\gamma(n)n})^2}{H}\right). \end{aligned} \tag{5}$$

In the scenario where $\gamma(n) = 1$ and $\lambda(n) = n$, we have $m = O(H/n)$ and $t = O(H/n^{1/4})$, which gives a minimal relative mean square error of $O(n^{3/2}/H)$.

3.3 Comparison to Mixture-Based Algorithm

As discussed, the problem $P(\mu(S_n) > y_n)$ can be reformulated as $P(S_n > \xi_n)$ with ξ_n taking some value on $[m_1 n, m_2 n]$. Consequently, another approach is to ignore learning the mean function $\mu(\cdot)$ and carry out an importance sampling scheme that is uniformly asymptotically efficient on the interval $[m_1 n, m_2 n]$. Glasserman and Juneja (2008) investigates the efficiency of a mixture scheme that randomizes the target cross-level between m_1 and m_2, followed by the optimal exponential tilting with the realized target level. Suppose that the mixture is discrete on k points on $[m_1, m_2]$, call them $q_i, i = 1, \ldots, k$. Moreover, let the discrete probability be $1/k$, which is shown to be nearly optimal in well-defined sense (Glasserman and Juneja 2008, Remark 3.1). We choose to avoid discussion on continuous mixture distribution as it requires numerical integration and complicates the comparison.

The likelihood ratio can then be written as

$$L = \frac{1}{\sum_{i=1}^{k} (1/k) e^{\theta_i S_n - n\psi(\theta_i)}} \tag{6}$$

where each θ_i satisfies $\psi'(\theta_i) = q_i$. The resulting relative second moment per simulation run is

$$\frac{\tilde{E}[L^2; \mu(S_n) > y_n]}{P(\mu(S_n) > y_n)^2} = O(\sqrt{n} k e^{nc/k^2}) \tag{7}$$

for some constant c. In fact, from Glasserman and Juneja (2008) Proposition 3.3 (it appears that the constant c in Equation (16) in their paper can be refined in terms of k) and further analysis (Lam 2012), the expression on the right hand side of (7) can be shown to be also a minimax lower bound (with a different constant in the exponent). Hence (7) is a good benchmark for comparison with our two-stage procedure. It is easy to see that picking $k = \Theta(\sqrt{n})$ would minimize (7) to attain an order n. So let us insist on using $k = \Theta(\sqrt{n})$ in our comparison.

Note that the calculation of likelihood ratio (6) would add an order k arithmetic operations to the cost per simulation, and this is taken as insignificant (to be conservative in our comparison and also because typically $\lambda(n) = n$). Given a budget capacity H, and choosing $k = \Theta(\sqrt{n})$, the expected relative second moment for the mixture algorithm is of order $n\lambda(n)/H$. Comparing this to (5), we see that it is more beneficial to use our two-stage procedure if

$$(\sqrt{C_1\lambda(n)}n^{1/4} + \sqrt{C_2\gamma(n)n})^2 \le n\lambda(n)$$

which is implied by $\gamma(n)/\lambda(n) = o(1)$. Therefore, it is better to pursue two-stage procedure if the computational cost for stochastic root finding is small relative to the cost of simulation, namely when $\gamma(n) = o(\lambda(n))$.

4 DISCUSSION: UNCERTAIN INPUT PARAMETERS

Although this paper deals with the scenario involving an unknown function in the probability, similar methodology can be utilized to analyze efficient importance sampling strategy when certain parameters in the stochastic object are uncertain and has to be estimated through other data. For example, consider again the large deviations probability of i.i.d. sum, where the distribution of the summands are known up to a shift of the mean i.e., the summands are $U_i + \mu$ for known distribution U_i but μ is uncertain. If μ lies in an interval $[m_1 n, m_2 n]$, then the probability is $P(\sum_{i=1}^n U_i > y_n - \mu n) = P(\sum_{i=1}^n U_i > \xi_n)$ where now ξ_n is again some value in a bounded interval $[m_1' n, m_2' n]$. The problem then quickly falls into the framework discussed in this paper. Further work along this line includes other types of estimation for the U_i and the establishment of information-theoretic lower bounds to identify the optimal procedures in these contexts.

5 PROOFS

5.1 Main Proof

In this section we provide the main arguments for proving Theorem 1. The following uniform large deviations estimate of Robbins-Monro procedure is needed:

Theorem 2 Suppose Assumptions 1, 4, 5 and 6 are in place. Then X_t satisfies the following large deviations result

$$P(|X_{t+1} - \xi_n| \ge x) \le 2(t + c_4 e^{c_5 n}) \begin{cases} \exp\left\{-\frac{1}{2d}tx^2\right\} & \text{for } 0 \le x \le dh \\ \exp\left\{-\frac{1}{2}thx\right\} & \text{for } x \ge dh \end{cases}$$

uniformly over n, for some constants d, h, c_4, c_5.

The proof uses heavily a result in Woodroofe (1972), and is left to the next subsection.

Next, we also have the following bound on the relative bias of $P(S_n > \hat{\xi}_n)$ for $\hat{\xi}_n$ close to ξ_n, uniformly over $\xi_n \in [m_1 n, m_2 n]$:

Theorem 3 Suppose Assumptions 1–6 hold. For $x_n \le \epsilon n$ for some small $\epsilon > 0$, we have

$$\sup_{\xi_n \in [m_1 n, m_2 n]} \sup_{|\hat{\xi}_n - \xi_n| < x_n} \frac{|P(S_n > \hat{\xi}_n) - P(S_n > \xi_n)|}{P(S_n > \xi_n)} \le C_1 e^{C_2 x_n} \tag{8}$$

for some constants $C_1, C_2 > 0$. Moreover, when $x_n \le \epsilon$ for some small $\epsilon > 0$, we have

$$\sup_{\xi_n \in [m_1 n, m_2 n]} \sup_{|\hat{\xi}_n - \xi_n| < x_n} \frac{|P(S_n > \hat{\xi}_n) - P(S_n > \xi_n)|}{P(S_n > \xi_n)} \le C_3 x_n \tag{9}$$

for some constant $C_3 > 0$.

An analogous result holds for the expected variance in the first term of (3):

Theorem 4 Suppose Assumptions 1–6 hold, and Z_n is generated from one simulation sample. For $x_n \le \epsilon n$ for some small ϵ, we have

$$\sup_{\xi_n \in [m_1 n, m_2 n]} \sup_{|\hat{\xi}_n - \xi_n| < x_n} \frac{E[Z_n^2 | \hat{\xi}_n]}{P(S_n > \xi_n)^2} \le C_1 e^{C_2 x_n} \sqrt{n} \tag{10}$$

for some constants $C_1, C_2 > 0$. Moreover, for $x_n \le \epsilon$ for some small ϵ, we have

$$\sup_{\xi_n \in [m_1 n, m_2 n]} \sup_{|\hat{\xi}_n - \xi_n| < x_n} \frac{E[Z^2 | \hat{\xi}_n]}{P(S_n > \xi_n)^2} \le C_3 \sqrt{n} \tag{11}$$

for some constant $C_3 > 0$.

The proof of Theorem 3 is left to the end of this section, while the proof of Theorem 4 is similar and is skipped. With these estimates, we are ready to prove Theorem 1:

Proof of Theorem 1. Let us analyze the two terms in (3) one by one, starting with the second term. Note that since $\xi_n, \hat{\xi}_n \in [m_1 n, m_2 n]$, $|P(S_n > \hat{\xi}_n) - P(S_n > \xi_n)|$ is bounded by $O(e^{cn})$ for some constant c uniformly over $\xi_n, \hat{\xi}_n \in [m_1 n, m_2 n]$. Therefore,

$$\begin{aligned} MSE(P(S_n > \hat{\xi}_n)) &= E(P(S_n > \hat{\xi}_n) - P(S_n > \xi_n))^2 \\ &= E[(P(S_n > \hat{\xi}_n) - P(S_n > \xi_n))^2; |\hat{\xi}_n - \xi_n| < x_n] + P(|\hat{\xi}_n - \xi_n| \ge x_n) C e^{cn} \\ &\le \sup_{|\hat{\xi}_n - \xi_n| < x_n} |P(S_n > \hat{\xi}_n) - P(S_n > \xi_n)|^2 + P(|\hat{\xi}_n - \xi_n| \ge x_n) C e^{cn} \end{aligned} \tag{12}$$

where C is a constant. By Theorems 2 and 3, we have, from (12), that

$$\sup_{\xi_n \in [m_1 n, m_2 n]} \frac{MSE(P(S_n > \hat{\xi}_n))}{P(S_n > \xi_n)^2} \le C_1' e^{C_2' x_n} + 2(t + c_4 e^{c_5 n}) e^{-c_6 t x_n} C e^{cn}$$

for some constants $C_1', C_2', C, c, c_6 > 0$ when $x_n = \Omega(1)$, which can be seen to be a suboptimal bound. On the other hand, in the case that $x_n = o(1)$, we have

$$\sup_{\xi_n \in [m_1 n, m_2 n]} \frac{MSE(P(S_n > \hat{\xi}_n))}{P(S_n > \xi_n)^2} \le C_3' x_n^2 + 2(t + c_4 e^{c_5 n}) e^{-c_7 t x_n^2} C e^{cn} \tag{13}$$

for some constant $C_3', c_7 > 0$. For a given t, to minimize the right hand side of (13), we set

$$-2 \log x_n = c_7 t x_n^2 - (c_5 + c) n$$

which gives $x_n = \eta \sqrt{n/t}$ for large enough $\eta > 0$. Then the right hand side of (13) is of order $O(n/t)$.

Next we analyze the first term in (3). Suppose first that Z_n is generated from one Stage 2 simulation sample. For an estimated value of ξ_n, namely $\hat{\xi}_n$, we use exponential tilting with parameter $\hat{\theta}_n$, the solution

to $\psi'(\hat{\theta}_n) = \hat{\xi}_n/n$. Since the estimation procedure ensures that $\hat{\xi}_n$ is bounded in $[m_1 n, m_2 n]$, the variance of Z_n, given any $\hat{\xi}_n$, is uniformly bounded by Ce^{cn} for some constants $C, c > 0$. Hence

$$\begin{aligned} E[Var(Z_n|\hat{\xi}_n)] &\le E[Var(Z|\hat{\xi}_n); |\hat{\xi}_n - \xi_n| < x_n] + P(|\xi_n - \hat{\xi}_n| \ge x_n) Ce^{cn} \\ &= E[E[Z^2|\hat{\xi}_n] - (E[Z|\hat{\xi}_n])^2; |\hat{\xi}_n - \xi_n| < x_n] + P(|\xi_n - \hat{\xi}_n| \ge x_n) Ce^{cn} \\ &\le E[E[Z^2|\hat{\xi}_n]; |\hat{\xi}_n - \xi_n| < x_n] + P(|\xi_n - \hat{\xi}_n| \ge x_n) Ce^{cn} \quad (14) \\ &\le \sup_{|\hat{\xi}_n - \xi_n| < x_n} E[Z^2|\hat{\xi}_n] + P(|\xi_n - \hat{\xi}_n| \ge x_n) Ce^{cn}. \quad (15) \end{aligned}$$

From Theorem 2 and 4, for $x_n \le \epsilon n$,

$$\sup_{\xi_n \in [m_1 n, m_2 n]} \frac{E[Var(Z_n|\hat{\xi}_n)]}{P(S_n > \xi_n)^2} \le C_1' e^{C_2' x_n} \sqrt{n} + 2(t + c_4 e^{c_5 n}) e^{-c_6 t x_n} Ce^{cn}$$

for some constants $C_1', C_2', C, c, c_6 > 0$. On the other hand, for $x_n \le \epsilon$, we have

$$\sup_{\xi_n \in [m_1 n, m_2 n]} \frac{E[Var(Z_n|\hat{\xi}_n)]}{P(S_n > \xi_n)^2} \le C_3' \sqrt{n} + (t + c_4 e^{c_5 n}) e^{-c_7 t x_n^2} Ce^{cn}$$

for some constant $C_3', c_7 > 0$. Recall the assumption that $t = \omega(n)$. Hence, to minimize the above expressions, we choose $x_n \to 0$ arbitrarily slow to get

$$\sup_{\xi_n \in [m_1 n, m_2 n]} \frac{E[Var(Z_n|\hat{\xi}_n)]}{P(S_n > \xi_n)^2} \le C' \sqrt{n}$$

for some constant $C' > 0$. Lastly, for a sample mean of m trials of Z_n, the relative expected variance is $O(\sqrt{n}/m)$. This concludes the theorem. □

5.2 Proof of Theorem 2

We need the following result from Woodroofe (1972):

Theorem 5 (adapted from Woodroofe 1972) Consider a sequence of X_t, with $Y_t = \mu(X_t) + Z_t(X_t)$ as the value of the mean function $\mu(\cdot)$ evaluated at X_t, corrupted with independent noise $Z_t(X_t)$. The sequence X_t is generated by the Robbins-Monro process $X_{t+1} = X_t - a_t Y_t$. We assume the following:

- 1' $Z_t(\cdot)$ is a sequence of i.i.d. random functions. Denote $F(z; x)$ as the distribution function of $Z_t(x)$ given x, with $\psi(\theta; x) := \int e^{\theta z} F(dz; x) \le c_1$ for some constant $c_1 > 0$, and for $-2h_1 \le \theta \le 2h_1$ for some $h_1 > 0$. This in particular implies $\int z^2 F(dz; x) \le c_3$ for some constant $c_3 > 0$.
- 2' $\mu(0) = 0$, and $\gamma_1 \le \mu(x)/x \le \gamma_2$ for any $x \in \mathbb{R} \setminus \{0\}$, for some $\gamma_1, \gamma_2 > 0$ i.e., $\mu(\cdot)$ is quasi-linear.
- 3' $a_t = a/t$, where the constant a satisfies $a\gamma_1 \ge 1$, with γ_1 as defined in Assumption 2' above.
- 4' $Ee^{\theta X_r} \le c_2$ for $-h_2 \le \theta \le h_2$ for some $h_2, c_2 > 0$, and for some $r \ge 5a\gamma_2/2$.

Then we have the following inequality

$$P(|X_{t+1}| \ge x) \le 2(t + c_2) \begin{cases} \exp\left\{-\frac{1}{2d} t x^2\right\} & \text{for } 0 \le x \le dh \\ \exp\left\{-\frac{1}{2} t h x\right\} & \text{for } x \ge dh \end{cases}. \quad (16)$$

Here d is a constant depending on $a, \gamma_1, h_1, c_1, c_3$, and h depends on h_1, h_2, a and r.

A few remarks are in place:

Remark 1 In fact, from Woodroofe (1972), the constants d and h can be explicitly written as follows:

1. Let $b = c_3 + 4c_1/h_1^2$. Then d can be taken as $b\tau$, where τ is an upper bound for $\sum_{j=k}^{n} n a_j^2 \prod_{m=j+1}^{n}(1-\gamma_1 a_m) < \infty$.
2. h can be taken as $\min(h_1/(2a), h_2/r)$.

Remark 2 The formulation in Woodroofe (1972) focuses on the one-sided probability $P(X_{t+1} \leq -x)$ (with the corresponding one-sided assumptions). The upper-tail result can be easily extended by taking negative on both sides of the relation $X_{t+1} = X_t - a_t Y_t$ to get $-X_{t+1} = -X_t - a_t(-Y_t)$, with $-Y_t$ now becoming $-\mu(-(-X_t)) - Z_t(-(-X_t))$ with mean function $-\mu(-\cdot)$ and noise $-Z_t(-\cdot)$.

Proof of Theorem 2. Note that $\mu(x) = y_n$ can be rewritten as $\mu(x) - y_n = 0$, and we use the recursion $X_{t+1} = X_t - a_t[Y_t - y_n]$. Consider the recursive process shifted by ξ_n i.e., let $W_t = X_t - \xi_n$, and write $W_{t+1} = W_t - a_t[\mu(W_t + \xi_n) - y_n]$. We will show that the recursion W_t satisfies Assumptions 1', 2', 3' and 4'. Assumption 1' is obvious. So we focus on the remaining ones.

Note that by Assumption 1 the expression $\mu(w + \xi_n) - y_n$ satisfies the inequality $0 < \gamma_1 \leq \mu(w + \xi_n) - y_n \leq \gamma_2$ by a straightforward invocation of the mean value theorem, where γ_1 and γ_2 are the constants defined in Assumption 1. Moreover, obviously $\mu(0 + \xi_n) - y_n = 0$. Hence Assumption 2' is satisfied, and Assumption 3' follows.

It remains to show Assumption 4'. Observe that $Ee^{\theta W_0} = Ee^{\theta(X_0 - \xi_n)} \leq C(n) Ee^{-\theta\xi_n}$. By Assumption 1 $\mu(x)$ satisfies $\gamma_1' \leq \mu(x)/x \leq \gamma_2'$ for some γ_1', γ_2', and hence $\xi_n = \Theta(n)$. This implies $Ee^{-\theta\xi_n} \leq e^{cn}$ for some constant c. In overall, we then have $Ee^{\theta W_0} \leq c_5 e^{c_6 n}$ for some constants c_5 and c_6.

For convenience, write $\tilde{\mu}(w) = \mu(w + \xi_n)$ and $\tilde{Z}_t(w) = Z_t(w + \xi_n)$. Following the argument in Woodroofe (1972) p.338, we write

$$Ee^{W_t} = Ee^{\theta(W_{t-1} - a_t\tilde{\mu}(W_{t-1}) - a_t\tilde{Z}_{t-1}(W_{t-1}))} \leq c_2 Ee^{\theta(W_{t-1} - a_t\tilde{\mu}(W_{t-1}))} \leq c_2(Ee^{\theta W_{t-1}} + Ee^{-a_t\gamma_2\theta W_{t-1}})$$

for θ in a small enough neighborhood of 0. Recursing the relation above leads to $Ee^{\theta W_r} \leq c(r, c_3)\sum_i Ee^{\theta_i W_0}$ where $C(r, c_3)$ is a constant depending on r and c_3 and the summation is finite with each θ_i satisfying $|\theta_i| \leq C(r, a, \gamma_2, \theta)$. Therefore $Ee^{\theta W_r} \leq c_4 e^{c_5 n}$ for some constants c_4, c_5, and Assumption 4' is satisfied. The conclusion then follows directly. Finally, note that the inequality is also satisfied for an estimator that projects the last iteration onto the interval $[m_1 n, m_2 n]$. □

5.3 Proof of Theorem 3

Proof of Theorem 3. Consider first the asymptotic for $P(S_n > \xi_n)$, where $\xi_n \in [m_1 n, m_2 n]$. Let θ_n be the solution to $\psi'(\theta_n) = \xi_n/n$, and let $\tilde{E}_n[\cdot]$ be the expectation under the exponential change of measure for each U_i with parameter θ_n. We have

$$\begin{aligned} P(S_n > \xi_n) &= \tilde{E}_n[e^{-\theta_n S_n + n\psi(\theta_n)}; S_n > \xi_n] = e^{-nI(\xi_n/n)}\tilde{E}_n[e^{-\theta_n(S_n - \xi_n)}; S_n > \xi_n] \\ &= e^{-nI(\xi_n/n)}\tilde{E}_n[e^{-\theta_n\sqrt{n\psi''(\theta_n)}Z_n}; Z_n > 0] \end{aligned} \tag{17}$$

where $Z_n = (S_n - \xi_n)/\sqrt{n\psi''(\theta_n)}$ is a sum of i.i.d. variables with mean zero and unit variance under $\tilde{E}_n[\cdot]$. Denoting $F_n(\cdot)$ as the distribution function of Z_n, (17) can be written as

$$e^{-nI(\xi_n/n)}\int_0^\infty e^{-\theta_n\sqrt{n\psi''(\theta_n)}z} dF_n(z) = e^{-nI(\xi_n/n)}\int_0^\infty e^{-z} dF_n\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n}\right)$$

by a change of variable. Then integration by parts gives

$$e^{-nI(\xi_n/n)}\int_0^\infty \left[F_n\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n}\right) - F_n(0)\right] e^{-z} dz. \tag{18}$$

We now consider $\sup_{|\hat{\xi}_n-\xi_n|<x_n}|P(S_n>\hat{\xi}_n)-P(S_n>\xi_n)|\le P(S_n>\xi_n-x_n)-P(S_n>\xi_n+x_n)$. Suppose $x_n\le \epsilon n$ for some small ϵ. By the same argument leading to (18), we have

$$P(S_n>\xi_n\pm x_n)=e^{-nI(\xi_n/n\pm x_n/n)}\int_0^\infty\left[F_n\left(\frac{z}{\sqrt{n\psi''(\theta_n^\pm)}\theta_n^\pm}\right)-F_n(0)\right]e^{-z}dz. \tag{19}$$

Here $\theta_n^\pm$ is the solution to the equations $\psi'(\theta_n^\pm)=\xi_n/n\pm x_n/n$. Since $\psi(\cdot)$ is twice continuously differentiable, $\psi'(\cdot)$ is continuously differentiable, and so is $(\psi')^{-1}(\cdot)$ (defined as the positive root). Moreover, by continuity $((\psi')^{-1})'(s)$ is bounded over any bounded set of s. So

$$\theta_n^\pm=(\psi')^{-1}\left(\frac{\xi_n}{n}\pm\frac{x_n}{n}\right)=(\psi')^{-1}\left(\frac{\xi_n}{n}\right)\pm\frac{x_n}{n}((\psi')^{-1})'(\zeta)$$

for some ζ between ξ_n/n and $\xi_n/n\pm x_n/n$. Since ξ_n/n and $\xi_n/n\pm x_n/n$ are uniformly bounded over n,ξ_n, we have

$$\theta_n^\pm=\theta_n\pm\frac{x_n}{n}O(1) \tag{20}$$

uniformly in n,ξ_n. In particular, we have

$$\psi''(\theta_n^\pm)\sim\psi''(\theta_n) \tag{21}$$

by continuity of $\psi''(\cdot)$.

Now

$$\begin{aligned}I\left(\frac{\xi_n}{n}\pm\frac{x_n}{n}\right)&=\theta_n^\pm\left(\frac{\xi_n}{n}\pm\frac{x_n}{n}\right)-\psi(\theta_n^\pm)=\left(\theta_n\pm\frac{x_n}{n}O(1)\right)\left(\frac{\xi_n}{n}\pm\frac{x_n}{n}\right)-\psi\left(\theta_n\pm\frac{x_n}{n}O(1)\right)\\&=\left(\theta_n\frac{\xi_n}{n}-\psi(\theta_n)\right)\pm\theta_n\frac{x_n}{n}\pm\frac{x_n}{n}O(1)\left(\frac{\xi_n}{n}\pm\frac{x_n}{n}\right)\mp\frac{x_n}{n}O(1)\psi(\zeta)\end{aligned} \tag{22}$$

for some ζ between θ_n and $\theta_n\pm(x_n/n)O(1)$. Since θ_n is uniformly bounded, by the uniform bound on ξ_n/n and the continuity of $(\psi')^{-1}(\cdot)$, (22) is written as

$$I(\theta_n)\pm\frac{x_n}{n}O(1) \tag{23}$$

uniformly over n,ξ_n. Then from (20), (21) and (23), the probability in (19) can be written as

$$\begin{aligned}&e^{-nI(\xi_n/n)\pm x_nO(1)}\left\{\int_0^\infty\left[F_n\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n}\right)-F_n(0)\right]e^{-z}dz\right.\\&\left.\pm\frac{x_n}{n}\int_0^\infty f_n\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n}\right)\frac{z}{\sqrt{n}}\frac{d}{dx}\frac{1}{\sqrt{\psi''(\theta(x))}\theta(x)}e^{-z}dz\right\}\end{aligned} \tag{24}$$

for some x in a compact interval, uniformly over n,ξ_n. Here $\theta(x)$ is the function denoting the solution to $\psi'(\theta)=x$, and $f_n(\cdot)$ denotes the density function of Z_n. Since $\int_{-\infty}^\infty|\phi_U(\theta)|d\theta<\infty$, Edgeworth expansion gives $|f_n(x)-\phi(x)|\le C/\sqrt{n}$ uniformly over $x\in\mathbb{R}$, where $\phi(\cdot)$ denotes the density function of standard normal variable. This implies that $f_n(x)$ is uniformly bounded by some number M as $n\to\infty$. Therefore, (24) can be rewritten as

$$e^{-nI(\xi_n/n)\pm x_nO(1)}\left[\int_0^\infty\left[F_n\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n}\right)-F_n(0)\right]e^{-z}dz\pm\frac{x_n}{n^{3/2}}O(1)\right]$$

uniformly over n, ξ_n.

As a result, for $x_n \leq \epsilon n$ for small enough ϵ, we have

$$P(S_n > \xi_n - x_n) - P(S_n > \xi_n + x_n) \leq e^{-nI(\xi_n/n)+x_nO(1)} \int_0^\infty \left[F_n\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \right) - F_n(0) \right] e^{-z} dz \tag{25}$$

whereas for $x_n \leq \epsilon$ for small enough ϵ, we have

$$P(S_n > \xi_n - x_n) - P(S_n > \xi_n + x_n) \leq x_nO(1)e^{-nI(\xi_n/n)} \int_0^\infty \left[F_n\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \right) - F_n(0) \right] e^{-z} dz. \tag{26}$$

Let us now analyze the expression

$$\int_0^\infty \left[F_n\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \right) - F_n(0) \right] e^{-z} dz. \tag{27}$$

Following Dembo and Zeitouni (1998), Berry-Essen Theorem (or Edgeworth expansion up to the first order) states that

$$F_n(z) = \Phi(z) + \frac{(1-z^2)}{6\sqrt{n}}\phi(z) + o\left(\frac{1}{\sqrt{n}}\right) \tag{28}$$

uniformly over $z \in \mathbb{R}$, where $\Phi(z)$ and $\phi(z)$ are the distribution and density function of the standard normal variable. Then (27) becomes

$$\begin{aligned} &\int_0^\infty \left[\Phi\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \right) - \Phi(0) \right] e^{-z} dz \\ &+ \int_0^\infty \frac{1}{6\sqrt{n}} \left[\left(1 - \left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \right)^2 \right) \phi\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \right) - \phi(0) \right] e^{-z} dz + o\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \tag{29}$$

uniformly over ξ_n. Note that, for fixed z,

$$\frac{\Phi(z/(\sqrt{n\psi''(\theta_n)}\theta_n)) - \Phi(0)}{\phi(0)/(\sqrt{n\psi''(\theta_n)}\theta_n)} \to z \tag{30}$$

since

$$\Phi\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \right) - \Phi(0) = \phi(0)\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} + \phi'(\zeta)\frac{z^2}{2n\psi''(\theta_n)\theta_n^2}$$

for some ζ between 0 and $z/(\sqrt{n\psi''(\theta_n)}\theta_n)$. Also note that

$$\left(1 - \left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \right)^2 \right) \phi\left(\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \right) - \phi(0) = \phi'(\zeta)\frac{z}{\sqrt{n\psi''(\theta_n)}\theta_n} \leq \frac{Cz}{\sqrt{n}} \tag{31}$$

uniformly over z. From (30) and (31), and by dominated convergence, (29) becomes

$$\frac{1}{\sqrt{2\pi n\psi''(\theta_n)}\theta_n} \int_0^\infty ze^{-z} dz + o\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi n\psi''(\theta_n)}\theta_n} + o\left(\frac{1}{\sqrt{n}}\right) \tag{32}$$

uniformly over ξ_n.

As a result, (25) gives (8) for $x_n \leq \epsilon n$ and (26) gives (9) for $x_n \leq \epsilon$. This concludes the theorem.

□

ACKNOWLEDGMENTS

The author would like to thank the anonymous referees for their fruitful comments that help improve the presentation of this paper, and for pointing out some of the references.

REFERENCES

Blanchet, J., and H. Lam. 2012. "State-dependent importance sampling for rare-event simulation: An overview and recent advances". *Surveys in Operations Research and Management Science* 17:38-59.

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

Bucklew, J. 2004. *Introduction to Rare-Event Simulation*. New York: Springer-Verlag.

Dembo, A., and O. Zeitouni. 1998. *Large Deviations Techniques and Applications*. 2nd ed. New York: Springer-Verlag.

Dupuis, P., and H. Wang. 2007. "Subsolutions of an Isaacs equation and efficient schemes of importance sampling". *Math. OR* 32:723-757.

Glasserman, P., and S. Juneja. 2008. "Uniformly efficient importance sampling for the tail distribution of sums of random variables". *Math. OR* 33:36-50.

Glasserman, P., and S. Kou. 1995. "Analysis of an importance sampling estimator for tandem queues". *ACM Trans. Modeling Comp. Simulation* 4:22-42.

Heidelberger, P. 1995. "Fast simulation of rare events in queueing and reliability models". *ACM Trans. Modeling Comp. Simulation* 5:43-85.

Juneja, S., and P. Shahabuddin. 2006. "Rare event simulation techniques: An introduction and recent advances". *Handbook in OR & MS* 13:291-346.

Kollman, C., K. Baggerly, D. Cox, and R. Picard. 1999. "Adaptive importance sampling on discrete Markov chains". *Ann. Appl. Probab.* 9:391-412.

Lam, H. 2012. "Efficient rare-event estimation under model uncertainty". *Working paper*.

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

L'Ecuyer, P. and B. Tuffin. 2008. "Approximate zero-variance simulation". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Monch, O. Rose, T. Jefferson, and J. W. Fowler, 170-181. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Olvera-Cravioto, M. 2007. "The single-server queue with heavy tails". *Ph.D. dissertation, Stanford University*.

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

Woodroofe, M. 1972. "Normal approximation and large deviations for the Robbins-Monro process". *Probability Theory and Related Fields* 21:329-338.

Zhang, X., P. W. Glynn, and J. Blanchet. 2007. "Efficient sub-optimal rare-event simulation". In *Proceedings of the 2007 Winter Simulation Conference*, edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 389-394. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

HENRY LAM is an Assistant Professor in the Department of Mathematics and Statistics at Boston University. He graduated from Harvard University with a Ph.D. degree in statistics in 2011. His research interests lie in applied probability and Monte Carlo methods with applications in queueing, operations management and insurance modeling.