On the Inefficiency of State-independent Importance Sampling in the Presence of Heavy Tails

Achal Bassamboo*
Northwestern University

Sandeep Juneja†
Tata Institute of Fundamental Research

Assaf Zeevi‡
Columbia University

Submitted: August 2004, Revised: July 2005, January 2006

Abstract

We consider importance sampling simulation for estimating rare event probabilities in the presence of heavy-tailed distributions that have polynomial-like tails. In particular, we prove the following negative result: there does not exist an asymptotically optimal *state-independent* change-of-measure for estimating the probability that a random walk (respectively, queue length for a single server queue) exceeds a "high" threshold before going below zero (respectively, becoming empty). Furthermore, we derive explicit bounds on the best asymptotic variance reduction achieved by state-independent importance sampling relative to naïve simulation. We illustrate through a simple numerical example that a "good" *state-dependent* change-of-measure may be developed based on an approximation of the zero-variance measure.

1 Introduction

Importance sampling (IS) simulation has proven to be an extremely successful method in efficiently estimating certain rare events associated with *light-tailed* random variables; see, e.g., [15] and [11] for queueing applications, and [10] for applications in financial engineering. (Roughly speaking, a random variable is said to be light-tailed if the tail of the distribution decays at least exponentially fast.) The main idea of IS algorithms is to perform a *change-of-measure*, then estimate the rare event in question by generating independent identically distributed (iid) copies of the underlying random variables (rv's) according to this new distribution. Roughly speaking, a "good" IS distribution should assign high probability to realizations of the rv's that give rise to the rare event of interest (while simultaneously not reducing by too much the probability of more likely events).

*Kellogg School of Management. e-mail: a-bassamboo@northwestern.edu
†School of Technology and Computer Science. email: juneja@tifr.res.in
‡Graduate School of Business. e-mail: assaf@gsb.columbia.edu

Recently, *heavy-tailed* distributions have become increasingly important in explaining rare event related phenomena in many fields including data networks and teletraffic models (see, e.g., [14]), and insurance and risk management (cf. [9]). Unlike the light-tailed case, designing efficient IS simulation techniques in the presence of heavy-tailed random variables has proven to be quite challenging. This is mainly due to the fact that the manner in which rare events occur is quite different than that encountered in the light-tailed context (see, [2] for further discussion).

In this paper we highlight a fundamental difficulty in applying IS techniques in the presence of heavy-tailed random variables. For a broad class of such distributions having polynomial-like tails, we prove that if the constituent random variables are independent under an IS change-of-measure then this measure cannot achieve *asymptotic optimality*. (Roughly speaking, a change-of-measure is said to be asymptotically optimal, or *efficient*, if it asymptotically achieves zero variance on a logarithmic scale; a precise definition is given in Section 2.) In particular, we give explicit asymptotic bounds on the level of improvement that state-independent IS can achieve vis-a-vis naïve simulation. These results are derived for the following two rare events.

i.) A negative drift random walk (RW) $S_n = \sum_{i=1}^{n} X_i$ exceeding a large threshold before taking on a negative value (see Theorem 1).

ii.) A stable GI/GI/1 queue exceeding a large threshold within a busy cycle (see Theorem 2). This analysis builds on asymptotes for the maximum of the queue length process (see Proposition 1).

The above probabilities are particularly important in estimating steady-state performance measures related to waiting times and queue lengths in single-server queues, when the regenerative ratio representations is exploited for estimation (see, e.g., [11]).

Our negative results motivate the development of state-dependent IS techniques (see, e.g., [13]). In particular, for the probabilities that we consider the zero variance measure has a straightforward "state-dependent" representation. In the random walk setting this involves generating each increment X_i using a distribution that depends on the position of the RW prior to that, *i.e.*, the distribution of X_i depends on $S_{i-1} = \sum_{j=1}^{i-1} X_j$. For a simple example involving an $M/G/1$ queue, we illustrate numerically how one can exploit approximations to the zero-variance measure (see Proposition 2) to develop state-dependent IS schemes that perform reasonably well.

Related literature. The first algorithm for efficient simulation in the heavy-tailed context was given in [3] using conditional Monte Carlo. Both [4] and [12] develop successful IS techniques to estimate level crossing probabilities of the form $P(\max_n S_n > u)$, for random walks with heavy tails, by relying on an alternative ladder height based representation of this probability. (Our negative results do not apply in such cases since the distribution that is being "twisted" is not that of the increment but rather that of the ladder height.) It is important to note that the ladder height

representation is only useful for a restricted class of random walks where each X_i is a difference of a heavy tailed random variable and an exponentially distributed random variable. The work in [7] also considers the level crossing problem and obtains positive results for IS simulation in the presence of Weibull-tails. They avoid the inevitable variance build-up by truncating the generated paths. However, even with truncation they observe poor results when the associated random variables have polynomial tails.

Recently, in [5] it was shown that performing a change in parameters within the family of Weibull or Pareto distributions does not result in an asymptotically optimal IS scheme in the random-walk or in the single server queue example. In [5] the authors also advocate the use of cross-entropy methods for selecting the "best" change-of-measure for IS purposes. Our paper provides further evidence that any state-independent change-of-measure (not restricted to just parameter changes in the original distribution) will not lead to efficient IS simulation algorithms. We also explicitly bound the loss of efficiency that results from restricting use to iid IS distributions.

The remainder of this paper. In Section 2, we briefly describe IS and the notion of asymptotic optimality. Section 3 describes the main results of the paper. In Section 4 we illustrate numerically the performance of a state-dependent approximate zero variance change-of-measure for a simple discrete time queue. Proofs of the main results (Theorems 1 and 2) are given in Appendix A. For space considerations we omit the proof of secondary results (Proposition 1 and Proposition 2), the details of which can be found in [6].

2 Importance Sampling and Asymptotic Optimality

2.1 Two rare events

Random walk. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random walk $S_n = \sum_{m=1}^{n} X_m$, $S_0 = 0$ where $X_1, X_2, ...$ are iid copies of X. We assume that $\mathbb{E}X < 0$, and we denote the cumulative distribution function of X by F. Define τ to be the time at which the random walk first goes below zero, i.e.,

$$\tau = \inf\{n \geq 1 : S_n < 0\}.$$

Let $\zeta = \mathbb{E}\tau$, and $M_n = \max_{0 \leq m \leq n} S_m$. The probability of interest is $\gamma_u = \mathbb{P}(M_\tau > u)$. To estimate this probability by naïve simulation, we generate m iid samples of the function $\mathbb{I}_{\{M_\tau > u\}}$ and average over them to get an unbiased estimate $\hat{\gamma}_u^m$. The relative error of this estimator (defined as the ratio of standard deviation and mean) is given by $\sqrt{\frac{(1-\gamma_u)}{m\gamma_u}}$. Since $\gamma_u \to 0$ as $u \to \infty$, the number of simulation runs must increase without bound in order to have fixed small relative error as u becomes large.

Consider another probability distribution $\widetilde{\mathbb{P}}$ on the same sample space such that the sequence

$\{X_1, X_2, ...\}$ is iid under $\widetilde{\mathbb{P}}$ with marginal distribution $\widetilde{F}$, and F is absolutely continuous w.r.t. $\widetilde{F}$. Let $T_u = \inf\{n : S_n \geq u\}$. Define

$$Z_u = L_u \mathbb{I}_{\{M_\tau > u\}}, \tag{1}$$
$$\text{where} \qquad L_u = \prod_{i=1}^{\min\{\tau, T_u\}} \frac{dF(X_i)}{d\widetilde{F}(X_i)},$$

and let $\widetilde{\mathbb{E}}[\cdot]$ be the expectation operator under $\widetilde{\mathbb{P}}$. Then, using Wald's likelihood ratio identity (see [16, Proposition 2.24]), we have that Z_u under measure $\widetilde{\mathbb{P}}$ is an unbiased estimator of the probability $\mathbb{P}(M_\tau > u)$. Thus, we can generate iid samples of Z_u under the measure $\widetilde{\mathbb{P}}$, the average of these would be an unbiased estimate of γ_u. We refer to $\widetilde{\mathbb{P}}$ as the IS change-of-measure and L_u as the likelihood ratio. In many cases, by choosing the IS change-of-measure appropriately, we can substantially reduce the variance of this estimator.

Note that a similar analysis can be carried out to get an estimator when the sequence $\{X_1, X_2, ...\}$ is not iid under $\widetilde{\mathbb{P}}$. The likelihood ratio L_u in that case can be expressed as the Radon-Nikodyn derivative of the original measure w.r.t. the IS measure restricted to the appropriate stopping time.

Queue length process. The second rare event studied in this paper is that of buffer overflow during a busy cycle. Consider a $GI/GI/1$ queue, and let the inter-arrival and service times have finite means λ^{-1} and μ^{-1}, respectively. Let $Q(t)$ represent the number of customers in the system (in queue and in service) at time t under FCFS (first come first serve) service discipline. Assume that the busy cycle starts at time $t = 0$, *i.e.*, $Q(0^-) = 0$ and $Q(0) = 1$, and let τ denote the end of the busy cycle, namely

$$\tau = \inf\{t \geq 0 : \ Q(t) = 0\}.$$

Let the cumulative distribution of inter-arrival times and service times be F and G, respectively. Let S_i be the service time of the i^{th} customer and A_i be the inter-arrival time for the $(i+1)^{th}$ customer. The probability of interest is $\gamma_u = \mathbb{P}(\max_{0 \leq t \leq \tau} Q(t) \geq u)$. (We assume that $u > 1$ is integer-valued for simplicity.) Again we note that $\gamma_u \to 0$ as $u \to \infty$; to estimate this probability efficiently we can use IS.

Let the number of arrivals until the queue length exceeds (or is equal to) level $u - 1$ be

$$M = \inf\left\{ n \geq 1 : \sum_{i=1}^{n} A_i < \sum_{i=1}^{n-u+2} S_i \right\}.$$

Let $N(t)$ represent the number of arrivals up until time t. Then $N(\tau)$ is the number of customers arriving during a busy period. Let $\widetilde{F}$ and $\widetilde{G}$ be the cumulative IS distributions of inter-arrival and service times, respectively. Then, again using Wald's likelihood ratio identity, Z_u under the

measure $\widetilde{\mathbb{P}}$ is an unbiased estimator for the probability $\mathbb{P}(\max_{0\le t\le\tau} Q(t) > u)$, where

$$Z_u = L_u \mathbb{I}_{\{M\le N(\tau)\}}, \tag{2}$$

$$\text{and} \qquad L_u = \prod_{i=1}^{M} \frac{dF(A_i)}{d\widetilde{F}(A_i)} \prod_{j=1}^{M-u+2} \frac{dG(S_j)}{d\widetilde{G}(S_j)}.$$

2.2 Asymptotic optimality

Consider a sequence of rare-events indexed by a parameter u. Let $\mathbb{I}_u$ be the indicator of this rare event, and suppose $\mathbb{E}[\mathbb{I}_u] \to 0$ as $u \to \infty$ (e.g., for the first rare event defined above, $\mathbb{I}_u = \mathbb{I}_{\{M_\tau > u\}}$). Let $\widetilde{\mathbb{P}}$ be an IS distribution and L be the corresponding likelihood ratio. Put $Z_u = L\mathbb{I}_u$.

Definition 1 (asymptotic optimality [11]) *A sequence of IS estimators is said to asymptotically optimal if*

$$\frac{\log \widetilde{\mathbb{E}}[Z_u^2]}{\log \widetilde{\mathbb{E}}[Z_u]} \to 2 \quad as \quad u \to \infty. \tag{3}$$

Note that $\widetilde{\mathbb{E}}[Z_u^2] \ge (\widetilde{\mathbb{E}}[Z_u])^2$ and $\log \widetilde{\mathbb{E}}[Z_u] < 0$, therefore for any sequence of IS estimators we have

$$\limsup_{u\to\infty} \frac{\log \widetilde{\mathbb{E}}[Z_u^2]}{\log \widetilde{\mathbb{E}}[Z_u]} \le 2.$$

Thus, loosely speaking, asymptotic optimality implies minimal variance on logarithmic scale. Bucklew [8] uses the term *efficiency* to describe this property of a simulation-based estimator.

3 Main Results

3.1 Random walk

Consider the random walk defined in Section 2.1. We assume that the distribution of X satisfies

$$\frac{\log \mathbb{P}(X > x)}{\log x} \to -\alpha \quad \text{and} \quad \frac{\log \mathbb{P}(X < -x)}{\log x} \to -\beta, \quad \text{as } x \to \infty, \tag{4}$$

where $\alpha \in (1, \infty)$ and $\beta \in (1, \infty]$. Further, we assume that $\mathbb{P}(X > x) \sim 1 - B(x)$ as $x \to \infty$, for some distribution B on $(0, \infty)$ which is *subexponential*, that is, it satisfies

$$\limsup_{x\to\infty} \frac{1 - (B * B)(x)}{1 - B(x)} \le 2,$$

(cf. [9]). We write $f(u) \sim g(u)$ as $u \to \infty$ if $\frac{f(u)}{g(u)} \to 1$ as $u \to \infty$. Thus, distributions with regularly varying tails are a subset of the class of distributions satisfying our assumptions. (Regularly

varying distributions have $1-F(x) = \mathcal{L}(x)/x^{\alpha}$, where $\alpha > 1$ and $\mathcal{L}(x)$ is slowly varying; for further discussion see [9, Appendix A.3].) Note that (4) allows the tail behavior on the negative side to be lighter than polynomial as $\beta = \infty$ is permitted. We denote the cumulative distribution function of X by F. From [2] it follows that

$$\mathbb{P}(M_\tau > u) \sim \zeta \mathbb{P}(X > u) \text{ as } u \to \infty, \tag{5}$$

where ζ is the expected time at which the random walk goes below zero. Consider the IS probability distribution $\widetilde{\mathbb{P}}$ such that the sequence $\{X_1, X_2, ...\}$ is iid under $\widetilde{\mathbb{P}}$ with marginal distribution $\widetilde{F}$, and F is absolutely continuous w.r.t. $\widetilde{F}$. Let $\mathcal{P}$ be the collection of all such probability distributions on the sample space $(\Omega, \mathcal{F})$. Let Z_u be an unbiased estimator of $\mathbb{P}(M_\tau > u)$ defined in (1). We then have the following result.

Theorem 1 *For any* $\widetilde{\mathbb{P}} \in \mathcal{P}$

$$\limsup_{u\to\infty} \frac{\log \widetilde{\mathbb{E}}[Z_u^2]}{-\alpha \log u} \leq 2 - \frac{\min(\alpha, \beta)}{\alpha(1 + \min(\alpha, \beta))},$$

where α *and* β *are defined in (4).*

Intuition and proof sketch. The proof follows by contradiction. We consider two disjoint subsets $\mathcal{B}$ and $\mathcal{C}$ of the "rare set" $\mathcal{A} = \{\omega : M_\tau > u\}$ and use the fact that $\widetilde{\mathbb{E}}[L_u^2 \mathbb{I}_{\{\mathcal{A}\}}] \geq \widetilde{\mathbb{E}}[L_u^2 \mathbb{I}_{\{\mathcal{B}\}}] + \widetilde{\mathbb{E}}[L_u^2 \mathbb{I}_{\{\mathcal{C}\}}]$. The first subset, $\mathcal{B}$, consists of sample paths where the first random variable is "large" and causes the random walk to immediately exceed level u. Assuming that the limit in the above equation exceeds $2 - \frac{\min(\alpha,\beta)}{\alpha(1+\min(\alpha,\beta))}$, we obtain a lower bound on the probability that X exceeds u under the IS distribution $\widetilde{F}$. The above, in turn, restricts the mass that can be allocated below level u. We then consider the subset $\mathcal{C}$ which consists of sample paths where the X_i's are of order u^γ for $i = 2, \ldots, \lfloor u^{1-\gamma} \rfloor$ followed by one "big" jump. By suitably selecting the parameter γ and the value of X_1, we can show that $\widetilde{\mathbb{E}}[L_u^2 \mathbb{I}_{\{\mathcal{C}\}}]$ is infinite, leading to the desired contradiction.

3.2 Queue length process

Consider a GI/GI/1 queue described in Section 2.1 with service times being iid copies of S and inter-arrival times being iid copies of A. Put $\Lambda(x) := -\log \mathbb{P}(S > x) = -\log(1 - G(x))$. Assume that

$$\frac{\Lambda(x)}{\log x} \to \alpha \text{ as } x \to \infty, \tag{6}$$

where $\alpha \in (1, \infty)$, and $(S - A)$ has a subexponential distribution. We then have the following logarithmic asymptotics for the buffer overflow probability in a busy cycle.

Proposition 1 *Let assumption (6) hold. Then,*

$$\lim_{u\to\infty} \frac{\log \mathbb{P}(\max_{0\le t\le\tau} Q(t) > u)}{\log u} = -\alpha.$$

Recall that $\widetilde{F}$ and $\widetilde{G}$ are the cumulative IS distribution of inter-arrival and service times, respectively, and an unbiased estimator for the probability $\mathbb{P}(\max_{0\le t\le\tau} Q(t) > u)$ is Z_u defined in (2). Let $\widetilde{\mathbb{P}}$ be the product measure generated by $(\widetilde{F}, \widetilde{G})$, and let $\mathcal{D}$ be the collection of all such measures.

Theorem 2 *For any* $\widetilde{\mathbb{P}} \in \mathcal{D}$

$$\limsup_{u\to\infty} \frac{\log \widetilde{\mathbb{E}}[Z_u^2]}{-\alpha \log u} \le 2 - \frac{1}{1+\alpha}.$$

The basic idea of the proof is similar to that sketched immediately following Theorem 1.

3.3 Discussion

1. Theorems 1 and 2 imply that for our class of heavy-tailed distributions no state-independent change-of-measure can be asymptotically optimal, since by Definition 1 such a distribution must satisfy

$$\liminf_{u\to\infty} \frac{\log \widetilde{\mathbb{E}}[Z_u^2]}{-\alpha \log u} \ge 2.$$

Note that Theorems 1 and 2 hold even when the IS distribution is allowed to depend on u, and Theorem 2 continues to hold when the inter-arrival time distribution is changed in a state-dependent manner. (The proof is a straightforward modification of the one given in Appendix A.)

2. The bounds given in Theorems 1 and 2 indicate that the asymptotic inefficiency of the "best" state-independent IS distribution is more severe the heavier the tails of the underlying distributions are. As these tails become lighter, a state-independent IS distribution may potentially achieve near-optimal asymptotic variance reduction.

4 A State Dependent Change-of-Measure

In this section we briefly describe the Markovian structure of the "state dependent" zero variance measure in settings that include the probabilities that we have considered. To keep the analysis simple we focus on a discrete state process. As an illustrative example, we consider the probability that the queue length in an $M/G/1$ queue exceeds a large threshold u in a busy cycle. Here we develop an asymptotic approximation for the zero variance measure and empirically test the performance of the corresponding IS estimator. (Such approximations of the zero variance measure can be developed more generally, see [6].)

Preliminaries and the proposed approach. Consider a Markov process $(S_n : n \geq 0)$ taking integer values. Let $\{p_{xy}\}$ denote the associated transition matrix. Let A and R be two disjoint sets in the state space, e.g., A may denote a set of non-positive integers and R may denote the set $\{u, u+1, \ldots\}$ for a large integer u. For any set B, let $\tau_B = \inf\{n \geq 1 : S_n \in B\}$. Let $J_x = \mathbb{P}(\tau_R < \tau_A | S_0 = x)$ for an integer x. In this set up, the zero variance measure for estimating the probability J_{s_0}, for $s_0 \geq 1$, admits a Markovian structure with transition matrix

$$p^*_{xy} = \frac{p_{xy} J_y}{J_x},$$

for each x, y (see, e.g., [1]). The fact that the transition matrix is stochastic follows from first step analysis. Also note that $\{\tau_R < \tau_A\}$ occurs with probability one under the p^* distribution, and the likelihood ratio, due to cancellation of terms, equals J_{s_0} along each generated path. Thus, if good approximations can be developed for J_x for each x, then the associated approximation of the zero variance measure may effectively estimate J_{s_0}.

Description of the numerical example. For the purpose of our numerical study, we consider the M/G/1 queue observed at times of customer departures. The arrival stream is Poisson with rate λ and service times are iid copies of S having Pareto distribution with parameter $\alpha \in (1, \infty)$, *i.e.*,

$$\mathbb{P}(S \geq x) = \left\{ \begin{array}{ll} x^{-\alpha} & \text{if } x \geq 1 \\ 1 & \text{otherwise.} \end{array} \right\}$$

Let Y_n denote the number of arrivals during the service of the nth customer. Note that $\{Y_n\}$ are iid, and conditioned on the first service time taking a value s, Y_1 is distributed as a Poisson random variable with mean λs. We assume that $\mathbb{E}[Y_1] < 1$ to ensure that the system is stable. Let $X_n = Y_n - 1$, $S_n = \sum_{i=1}^n X_i$ with $S_0 = x$, and $\tau_0 = \inf\{n \geq 1 : S_n \leq 0\}$. Then τ_0 denotes the length of a busy cycle that commences with x customers in the system, and for $1 \leq n \leq \tau_0$, S_n denotes the number in the system at the departure of the nth customer. Let $\tau_u = \inf\{n \geq 1 : S_n \geq u\}$. The event of interest is that the number of customers in the system exceeds level u during the first busy cycle, conditioned on $S_0 = x$. We denote this probability by $J_x(u) = \mathbb{P}(\tau_u < \tau_0 | S_0 = x)$. The following proposition provides an asymptotic for $J_x(u)$. Here, F denotes the distribution function of X_1.

Proposition 2 *For all* $\beta \in (0, 1)$

$$J_{\lfloor \beta u \rfloor}(u) \sim \mathbb{E}[\tau_0] \left[\int_{x=(1-\beta)u}^{u} (1 - F(x)) dx \right] \text{ as } u \to \infty. \tag{7}$$

This suggests that $J_x(u) \approx E[\tau_0]\, g(x)$, where $g(x) = \sum_{z=u-x}^{u} \mathbb{P}(X_1 \geq z)$. Therefore a reasonable approximation for the zero variance measure would have transition probabilities

$$\tilde{p}_{xy} = \frac{\mathbb{P}(X_1 = y - x) g(y)}{\sum_{z=x-1}^{\infty} \mathbb{P}(X_1 = z - x) g(z)}$$

for $x \geq 1$ and $y \geq x - 1$. These probabilities are easy to compute in this simple setting.

Note that in the existing literature no successful methodology exists for estimating $J_1(u)$ by simulating $(X_1, X_2, \ldots)$ under an IS distribution. As mentioned in the Introduction, [4] and [12] estimate the related level crossing probabilities by exploiting an alternative ladder height based representation.

The simulation experiment. We estimate the exceedence of level u in a busy cycle for the following cases: $u = 100$ and 1000; tail parameter values $\alpha = 2, 9$ and 19; and traffic intensities $\rho = 0.3, 0.5$ and 0.8. (The traffic intensity ρ equals $\lambda\alpha/(\alpha - 1)$.) The number of simulation runs in all cases is taken to be $500,000$. To test the precision of the simulation results, we also calculate the probabilities of interest, $J_1(u)$, using first step analysis. Results in Table 1 illustrate the following points. First, the precision of the proposed IS method decreases as the traffic intensity increases, and/or the tail becomes "lighter." Second, accuracy for the problem involving buffer level 1000 is better than the case of buffer level 100, in accordance with the fact that we are using a "large buffer" asymptotic approximation to the zero variance measure. Finally, the relative error on logarithmic scale is quite close to the best possible value of 2, hence we anticipate that our proposed IS scheme might be asymptotically optimal. (In [6] performance of an adaptive version of this algorithm is shown to give further improvement in the performance of the IS estimator.) The rigorous derivation of such results and their generalizations to continuous state space is left for future work.

u	α	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.8$
100	2	$3.31 \times 10^{-6} \pm 0.019\%$ [1.97]	$2.43 \times 10^{-5} \pm 0.050\%$ [1.86]	$1.00 \times 10^{-4} \pm 0.837\%$ [1.26]
	9	$1.57 \times 10^{-23} \pm 0.051\%$ [1.97]	$1.52 \times 10^{-20} \pm 0.119\%$ [1.93]	$5.12 \times 10^{-19} \pm 2.409\%$ [1.79]†
	19	$4.70 \times 10^{-48} \pm 0.080\%$ [1.98]	$5.30 \times 10^{-42} \pm 0.543\%$ [1.94]†	$4.58 \times 10^{-39} \pm 4.182\%$ [1.89]†
1000	2	$3.19 \times 10^{-8} \pm 0.006\%$ [1.99]	$2.25 \times 10^{-7} \pm 0.015\%$ [1.98]	$8.16 \times 10^{-7} \pm 0.079\%$ [1.84]
	9	$1.02 \times 10^{-32} \pm 0.007\%$ [2.00]	$9.21 \times 10^{-30} \pm 0.032\%$ [1.99]	$2.49 \times 10^{-28} \pm 0.103\%$ [1.96]
	19	$7.22 \times 10^{-68} \pm 0.022\%$ [2.00]	$6.72 \times 10^{-62} \pm 0.041\%$ [2.00]	$3.30 \times 10^{-59} \pm 0.403\%$ [1.96]

Table 1: Performance of the state-dependent IS estimator for the probability of exceeding level u in a busy cycle: Simulation results are for $u = 100$ and 1000, using $500,000$ simulation runs. The $\pm X$ represents 95% confidence interval, and $[Y]$ represents the ratio defined in (3) with 2 being the asymptotically optimal value. † The actual probability, which is calculated using first step analysis, lies outside the 95% confidence interval.

A Proofs of the main results

We use the following simple lemma repeatedly in the proof of the main results.

Lemma 1 *Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two probability distribution on the same sample space $(\Omega, \mathcal{F})$. Then for any set $A \in \mathcal{F}$ if $\mathbb{P}$ is absolutely continuous w.r.t. $\widetilde{\mathbb{P}}$ on the set A, then*

$$\mathbb{P}^2(A) \leq \widetilde{\mathbb{P}}(A) \int_A \frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}} d\mathbb{P}. \tag{8}$$

Proof. Fix a set $A \in \mathcal{F}$. Using the Cauchy-Schwartz inequality, we have

$$\left(\int_A \frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}} d\widetilde{\mathbb{P}} \right)^2 \leq \widetilde{\mathbb{P}}(A) \int_A \left(\frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}} \right)^2 d\widetilde{\mathbb{P}}.$$

Rearranging terms we get (8). ■

Proof of Theorem 1. Let

$$\Lambda_+(x) := -\log \mathbb{P}(X > x), \tag{9}$$

$$\Lambda_-(x) := -\log \mathbb{P}(X < -x). \tag{10}$$

Assumption (4) on the distribution of X states that

$$\frac{\Lambda_+(x)}{\log x} \to \alpha, \ \frac{\Lambda_-(x)}{\log x} \to \beta, \text{ as } x \to \infty. \tag{11}$$

and $\alpha \in (1, \infty)$ and $\beta \in (1, \infty]$. Fix $\widetilde{\mathbb{P}} \in \mathcal{P}$. Let c be defined as follows

$$c = \limsup_{u \to \infty} \frac{\log \widetilde{\mathbb{E}}[Z_u^2]}{-\alpha \log u}, \tag{12}$$

where Z_u is defined in (1). Then, there exists a subsequence $\{u_n : n = 1, 2, \ldots\}$ over which c is achieved; for brevity, simply assume that the limit holds on the original sequence. Appealing to (5), we have

$$\widetilde{\mathbb{E}}(Z_u^2) \geq (\widetilde{\mathbb{E}}(Z_u))^2 = \mathbb{P}^2(M_\tau > u) \sim \zeta^2 e^{-2\Lambda_+(u)},$$

hence c cannot be greater than 2. Put

$$\kappa = \frac{\min(\alpha, \beta)}{\alpha(1 + \min(\alpha, \beta))} < 1,$$

and assume towards a contradiction that in (12), $c \in (2 - \kappa, 2]$. Let $\mathcal{B} = \{\omega : X_1(\omega) > u\}$. Then by Lemma 1 we have

$$\widetilde{\mathbb{P}}(\mathcal{B}) \geq \frac{\mathbb{P}^2(\mathcal{B})}{\widetilde{\mathbb{E}}[Z_u^2 \mathbb{I}\{\mathcal{B}\}]}.$$

Now, using (11) and (12) to lower bound the right-hand-side above we get that for sufficiently large u, $1 - \widetilde{F}(u) \geq e^{-(2-c')\Lambda_+(u)}$ where $c' \in (2 - \kappa, c)$. Here we use the fact that $\mathbb{P}(\mathcal{B}) = 1 - F(u)$ and $\widetilde{\mathbb{P}}(\mathcal{B}) = 1 - \widetilde{F}(u)$. Thus, for u sufficiently large,

$$\widetilde{F}(u) \leq 1 - e^{-(2-c')\Lambda_+(u)}. \tag{13}$$

Appealing to Lemma 1, and using the definitions of Λ_+ and Λ_-, and (13) for any $0 < a < b < u$ and $-a' < 0 < b' < u$, we have

$$\begin{aligned}\int_a^b \frac{dF(s)}{d\widetilde{F}(s)} dF(s) &\geq \frac{(F(b)-F(a))^2}{\widetilde{F}(u)} \\ &\geq (e^{-\Lambda_+(a)} - e^{-\Lambda_+(b)})^2, \end{aligned} \tag{14}$$

and

$$\begin{aligned}\int_{-a'}^{b'} \frac{dF(s)}{d\widetilde{F}(s)} dF(s) &\geq \frac{(F(b')-F(-a'))^2}{\widetilde{F}(u)} \\ &\geq \frac{(1-e^{-\Lambda_+(b')} - e^{-\Lambda_-(a')})^2}{(1-e^{-(2-c')\Lambda_+(u)})} \\ &\geq (1-2e^{-\Lambda_+(b')} - 2e^{-\Lambda_-(a')})(1+e^{-(2-c')\Lambda_+(u)}). \end{aligned} \tag{15}$$

Now consider the following set $\mathcal{C}$ of sample paths, $X_1 \in [2u^{1-\epsilon}, 3u^{1-\epsilon'}]$, and $X_i \in [-u^{\gamma-\epsilon}, u^\gamma]$ for $i = 2, ..., \lfloor 0.5u^{1-\gamma} \rfloor$ and $X_{\lfloor 0.5u^{1-\gamma}\rfloor+1} \in [u, \infty)$, where

$$\epsilon = \begin{cases} \frac{\beta-\alpha}{(1+\alpha)\beta} & \text{if } \alpha < \beta \\ 1 - \frac{(2-c')(1+\beta)\alpha}{\beta} & \text{otherwise,} \end{cases} \tag{16}$$

$$\epsilon' \in (0, \epsilon), \tag{17}$$

$$\gamma = \begin{cases} \frac{1}{1+\alpha} & \text{if } \alpha < \beta \\ \frac{1+\epsilon\beta}{1+\beta} & \text{otherwise.} \end{cases} \tag{18}$$

Note that on the set of sample paths $\mathcal{C}$, we have $0 < S_n < u$ for all $n = 1, ..., \lfloor 0.5u^{1-\gamma} \rfloor$ and $S_{\lfloor 0.5u^{1-\gamma}\rfloor+1} > u$. Hence, $\mathbb{I}_{\{M_\tau > u\}} = 1$ on this set. Then,

$$\begin{aligned}\widetilde{\mathbb{E}}(Z_u^2) &\geq \int_{\mathcal{C}} \left(\frac{dF(x_1)}{d\widetilde{F}(x_1)} \cdots \frac{dF(x_{\lfloor 0.5u^{1-\gamma}\rfloor+1})}{d\widetilde{F}(x_{\lfloor 0.5u^{1-\gamma}\rfloor+1})} \right)^2 d\widetilde{F}(x_1) \ldots d\widetilde{F}(x_{\lfloor 0.5u^{1-\gamma}\rfloor+1}) \\ &\overset{(a)}{=} \int_{x_1 \in [2u^{1-\epsilon}, 3u^{1-\epsilon'}]} \frac{dF(x_1)}{d\widetilde{F}(x_1)} dF(x_1) \left(\prod_{i=2}^{\lfloor 0.5u^{1-\gamma}\rfloor} \int_{x_i \in [-u^{\gamma-\epsilon}, u^\gamma]} \frac{dF(x_i)}{d\widetilde{F}(x_i)} dF(x_i) \right) \\ &\quad \times \int_{x_{\lfloor 0.5u^{1-\gamma}\rfloor+1} \in [u,\infty)} \frac{dF(x_{\lfloor 0.5u^{1-\gamma}\rfloor+1})}{d\widetilde{F}(x_{\lfloor 0.5u^{1-\gamma}\rfloor+1})} dF(x_{\lfloor 0.5u^{1-\gamma}\rfloor+1}) \\ &\overset{(b)}{\geq} (e^{-\Lambda_+(2u^{1-\epsilon})} - e^{-\Lambda_+(3u^{1-\epsilon'})})^2 \left[(1-2e^{-\Lambda_+(u^\gamma)} - 2e^{-\Lambda_-(u^{\gamma-\epsilon})})(1+e^{-(2-c')\Lambda_+(u)}) \right]^{\lfloor 0.5u^{1-\gamma}\rfloor} \\ &\quad \times e^{-2\Lambda_+(u)} \quad \text{for sufficiently large } u, \end{aligned}$$

where (a) follows due to independence of the X_i's, and (b) follows from the inequalities (14) and (15) and the following application of Lemma 1

$$\int_u^\infty \frac{dF}{d\widetilde{F}} dF \geq \frac{\mathbb{P}^2(X_1 \geq u)}{\widetilde{\mathbb{P}}(X_1 > u)} \geq \mathbb{P}^2(X_1 \geq u).$$

Taking the logarithm of both sides, we have for sufficiently large u

$$\begin{aligned}\log(\widetilde{\mathbb{E}}(Z_u^2)) &\geq 2\log(e^{-\Lambda_+(2u^{1-\epsilon})} - e^{-\Lambda_+(3u^{1-\epsilon'})}) \\ &\quad +\lfloor 0.5u^{1-\gamma}\rfloor \left[\log(1-2e^{-\Lambda_+(u^\gamma)} - 2e^{-\Lambda_-(u^{\gamma-\epsilon})}) + \log(1+e^{-(2-c')\Lambda_+(u)})\right] \\ &\quad -2\Lambda_+(u) \\ &=: I_1(u) + \lfloor 0.5u^{1-\gamma}\rfloor I_2(u) + I_3(u).\end{aligned}$$

Now, using (11) we have

$$\begin{aligned}\frac{(I_1(u)+I_3(u))}{-\log u} &= \frac{2\Lambda_+(2u^{1-\epsilon}) - 2\log(1-e^{\Lambda_+(2u^{1-\epsilon})-\Lambda_+(3u^{1-\epsilon'})}) + 2\Lambda_+(u)}{\log u} \\ &\to 2\alpha(2-\epsilon) \text{ as } u\to\infty.\end{aligned}$$

Using the choice of ϵ, ϵ' and γ in (16)-(18), a straight forward calculation shows that $-u^{1-\gamma}I_2(u)/\log u \to \infty$, as $u\to\infty$. Thus, we have, in contradiction, that

$$\lim_{u\to\infty}\frac{\log(\widetilde{\mathbb{E}}(Z_u^2))}{-\log u} = \infty.$$

This completes the proof. ■

Proof of Theorem 2. Suppose, towards a contradiction, that there exist $(\widetilde{F},\widetilde{G})\in\mathcal{D}$ such that the corresponding IS change-of-measure for the probability $\mathbb{P}\left(\max_{0\leq t\leq\tau}Q(t)>u\right)$ satisfies

$$\limsup_{u\to\infty}\frac{\log\widetilde{\mathbb{E}}[Z_u^2]}{-\alpha\log u} = c,$$

and $c\in(2-\frac{1}{1+\alpha},2]$. Then, there exists a subsequence $\{u_n : n=1,2,\ldots\}$ over which c is achieved; for brevity, assume the limit holds for the original sequence. (Note that c cannot be greater than 2, using the same reasoning as in the proof of Theorem 1.) Thus, there exists $c'\in(2-\frac{1}{1+\alpha},c)$ such that for sufficiently large u, $\widetilde{\mathbb{E}}[Z_u^2]\leq e^{-c'\Lambda(u)}$. In what follows, assume without loss of generality that u is integer valued. We represent the original probability distribution by $\mathbb{P}$ and the IS probability distribution by $\widetilde{\mathbb{P}}$.

Next we describe another representation of the likelihood ratio which will be used in the proof. Consider any set of sample path $\mathcal{B}$, which can be decomposed into the sample paths for inter-arrival times, $\mathcal{B}_a$, and service times, $\mathcal{B}_s$, such that, $\mathcal{B}=\mathcal{B}_a\cap\mathcal{B}_s$. Since under the IS distribution the services times and inter-arrival times are iid random variables, we have

$$\int_{\mathcal{B}}\frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}}d\mathbb{P} = \int_{\mathcal{B}_a}\frac{d\mathbb{P}^a}{d\widetilde{\mathbb{P}}^a}d\mathbb{P}^a\int_{\mathcal{B}_s}\frac{d\mathbb{P}^s}{d\widetilde{\mathbb{P}}^s}d\mathbb{P}^s. \tag{19}$$

In the above equality, $\mathbb{P}^a$ and $\mathbb{P}^s$ represent the probability measure associated with the inter-arrival and service time process and services respectively. Hence, $\mathbb{P}=\mathbb{P}^a\times\mathbb{P}^s$. We define $\widetilde{\mathbb{P}}$, $\widetilde{\mathbb{P}}^a$ and $\widetilde{\mathbb{P}}^s$ in a similar manner.

Consider the set of sample paths on which $\{S_1(\omega) > 2u\lambda^{-1}\}$ and $\{\sum_{i=1}^{u} A_i(\omega) < 2u\lambda^{-1}\}$. Put $\mathcal{B}_a = \{\omega : \sum_{i=1}^{u} A_i(\omega) < 2u\lambda^{-1}\}$ and $\mathcal{B}_s = \{\omega : S_1(\omega) > 2u\lambda^{-1}\}$. Using Markov's inequality we have $\mathbb{P}^a(\mathcal{B}_a) \geq 1/2$. Using Lemma 1, we have

$$\int_{\mathcal{B}_a} \frac{d\mathbb{P}^a}{d\widetilde{\mathbb{P}}^a} d\mathbb{P}^a \geq \frac{\left(\mathbb{P}^a(\mathcal{B}_a)\right)^2}{\widetilde{\mathbb{P}}^a(\mathcal{B}_a)} \geq \left(\mathbb{P}^a(\mathcal{B}_a)\right)^2 \geq \left(\frac{1}{2}\right)^2.$$

Since the buffer overflows on the set $\mathcal{B}_a \cap \mathcal{B}_s$ we have

$$\begin{aligned}
\widetilde{\mathbb{E}}[Z_u^2] &\geq \int_{\mathcal{B}_a} \frac{d\mathbb{P}^a}{d\widetilde{\mathbb{P}}^a} d\mathbb{P}^a \int_{\mathcal{B}_s} \frac{d\mathbb{P}^s}{d\widetilde{\mathbb{P}}^s} d\mathbb{P}^s \\
&\geq \left(\frac{1}{2}\right)^2 \int_{2u\lambda^{-1}}^{\infty} \frac{dG(x)}{d\widetilde{G}(x)} dG(x).
\end{aligned}$$

Thus, we have for large enough u that

$$e^{-c'\Lambda(u)} \geq \frac{1}{4} \int_{2u\lambda^{-1}}^{\infty} \frac{dG(x)}{d\widetilde{G}(x)} dG(x).$$

Appealing to Lemma 1 and (6) we have for u sufficiently large

$$1 - \widetilde{G}(2u\lambda^{-1}) \geq \frac{e^{-2\Lambda(2u\lambda^{-1}) + c'\Lambda(u)}}{4}.$$

Thus, for sufficiently large u we have $\widetilde{G}(u^\gamma) \leq 1 - \frac{e^{-2\Lambda(2u\lambda^{-1}) + c'\Lambda(u)}}{4}$ where $\gamma < 1$. Using Lemma 1, this in turn implies that

$$\begin{aligned}
\int_0^{0.5u^\gamma\lambda^{-1}} \frac{dG(x)}{d\widetilde{G}(x)} dG(x) &\geq \frac{(1 - e^{-\Lambda(0.5u^\gamma\lambda^{-1})})^2}{1 - \frac{e^{-2\Lambda(2u\lambda^{-1}) + c'\Lambda(u)}}{4}} \qquad (20) \\
&\geq \left(1 + \frac{e^{-2\Lambda(2u\lambda^{-1}) + c'\Lambda(u)}}{4}\right) \left(1 - 2e^{-\Lambda(0.5u^\gamma\lambda^{-1})}\right). \qquad (21)
\end{aligned}$$

Now, consider the following set of sample paths represented by $\mathcal{B}'$.

1. The first service time $S_1 \in [2u^{1-\gamma}\lambda^{-1}, 3u^{1-\gamma}\lambda^{-1}]$.
2. The sum of the first $\lfloor u^{1-\gamma} \rfloor$ inter-arrival times is less than $2u^{1-\gamma}\lambda^{-1}$, i.e., $\sum_{i=1}^{\lfloor u^{1-\gamma} \rfloor} A_i \leq 2u^{1-\gamma}\lambda^{-1}$. This ensures that by the end of service of the first customer at least $\lfloor u^{1-\gamma} \rfloor$ customers are in the queue.
3. The next $\lfloor u^{1-\gamma} \rfloor - 1$ services lie in the interval $[0, 0.5u^\gamma\lambda^{-1}]$. This ensures that at most $0.5u\lambda^{-1}$ time has elapsed before the beginning of service of customer $\lfloor u^{1-\gamma} \rfloor$.
4. The service time for customer $\lfloor u^{1-\gamma} \rfloor$ exceeds $2u\lambda^{-1}$.
5. The next $\lfloor 0.6u \rfloor$ arrivals are such that $0.5u\lambda^{-1} \leq \sum_{i=\lfloor u^{1-\gamma} \rfloor + 1}^{\lfloor u^{1-\gamma} \rfloor + \lfloor 0.6u \rfloor} A_i \leq 0.75u\lambda^{-1}$. This ensures that the buffer does not overflow before the beginning of service of customer $\lfloor u^{1-\gamma} \rfloor$.

6. The next $\lfloor 0.4u \rfloor$ arrivals are such that $0.3u\lambda^{-1} \leq \sum_{i=\lfloor u^{1-\gamma} \rfloor + \lfloor 0.6u \rfloor}^{\lfloor u^{1-\gamma} \rfloor + \lfloor u \rfloor} A_i \leq 0.75u\lambda^{-1}$. This ensures that the buffer overflows during the service of customer $\lfloor u^{1-\gamma} \rfloor$.

The services in condition 3 are assigned sufficiently less probability under the new measure so that the second moment of the estimator builds up along such realizations. The remaining conditions ensures that the buffer overflows for the paths in this set. Note that the set $\mathcal{B}'$ can be decomposed into sample paths for arrivals (satisfying 2, 5 and 6 above), represented by $\mathcal{B}'_a$ and sample paths for service times (satisfying 1, 3 and 4 above), represented by $\mathcal{B}'_s$.

First we focus on the contribution of the arrival paths in $\mathcal{B}'_a$ to the likelihood ratio. Using Markov's inequality we have

$$\mathbb{P}\left(\sum_{i=1}^{\lfloor u^{1-\gamma} \rfloor} A_i \geq 2u^{1-\gamma}\lambda^{-1} \right) \leq \frac{1}{2}.$$

By the strong law of large numbers we get for sufficiently large u

$$\mathbb{P}\left(0.5u\lambda^{-1} \leq \sum_{i=\lfloor u^{1-\gamma} \rfloor + 1}^{\lfloor u^{1-\gamma} \rfloor + \lfloor 0.6u \rfloor} A_i \leq 0.75u\lambda^{-1} \right) \geq \frac{1}{2},$$
$$\mathbb{P}\left(0.3u\lambda^{-1} \leq \sum_{i=\lfloor u^{1-\gamma} \rfloor + \lfloor 0.6u \rfloor}^{\lfloor u^{1-\gamma} \rfloor + \lfloor u \rfloor} A_i \leq 0.75u\lambda^{-1} \right) \geq \frac{1}{2}.$$

Thus, we have $\mathbb{P}^a(\mathcal{B}'_a) \geq \frac{1}{8}$, which using Lemma 1 implies that

$$\int_{\mathcal{B}'_a} \frac{d\mathbb{P}^a}{d\widetilde{\mathbb{P}}^a} d\mathbb{P}^a \geq \frac{1}{8^2}. \tag{22}$$

We now focus on the contribution of the service time paths in $\mathcal{B}'_s$ to the likelihood ratio. In particular, using (20) we have

$$\begin{aligned} \int_{\mathcal{B}'_s} \frac{d\mathbb{P}^s}{d\widetilde{\mathbb{P}}^s} d\mathbb{P}^s \quad \geq \quad & \left[e^{-\Lambda(2u\lambda^{-1})} - e^{-\Lambda(3u\lambda^{-1})} \right]^2 \cdot \\ & \times \left[\left(1 + \frac{e^{-2\Lambda(2u\lambda^{-1}) + c'\Lambda(u)}}{4} \right) \left(1 - 2e^{-\Lambda(0.5u^\gamma \lambda^{-1})} \right) \right]^{\lfloor u^{1-\gamma} \rfloor - 1} \left[e^{-2\Lambda(2u\lambda^{-1})} \right] \end{aligned} \tag{23}$$

Combining (22) and (23) with (19), and taking a logarithm on both sides we have

$$\begin{aligned} \log \widetilde{\mathbb{E}}[Z_u^2] \quad \geq \quad & 2\log(e^{-\Lambda(2u\lambda^{-1})} - e^{-\Lambda(3u\lambda^{-1})}) \\ & + \left[u^{1-\gamma} - 1 \right] \left[\log(1 + \frac{e^{-2\Lambda(2u\lambda^{-1}) + c'\Lambda(u)}}{4}) + \log(1 - 2e^{-\Lambda(0.5u^\gamma \lambda^{-1})}) \right] \\ & - 2\Lambda(2u\lambda^{-1}) - \log 64. \end{aligned}$$

Repeating the arguments in the proof of Theorem 1 with the specific choice of c and γ, we get the desired contradiction

$$\lim_{u\to\infty} \frac{\log(\widetilde{\mathbb{E}}[Z_u^2])}{-\log u} = \infty,$$

which completes the proof. ■

References

[1] I. Ahamed, V. S. Borkar, and S. Juneja. Adaptive importance sampling for Markov chains using stochastic approximation. 2005. To appear in *Oper. Res.*

[2] S. Asmussen. Subexponential asymptotics for stochastic processes: extreme behavior, stationary distributions and first passage probabilities. *Ann. Appl. Probab.*, 8:354–374, 1998.

[3] S. Asmussen and K. Binswanger. Ruin probability simulation for subexponential claims. *ASTIN Bull*, 27:297–318, 1997.

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

[5] S. Asmussen, D. P. Kroese, and R. Y. Rubinstein. Heavy tails, importance sampling and cross-entropy. 2004. Preprint.

[6] A. Bassamboo, S. Juneja, and A. Zeevi. Importance sampling simulation in the presence of heavy tails. *Proceedings of Winter Simulation Conference*, 2005.

[7] N.K. Boots and P. Shahabuddin. Simulating tail probabilities in GI/GI/1 queues and insurance risk processes with sub-exponential distributions. 2001. Preprint.

[8] J. A. Bucklew. *Large deviation techniques in decision, simulation, and estimation.* Wiley, New York, 1990.

[9] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin, 1997.

[10] P. Glasserman. *Monte Carlo Methods in Financial Engineeing.* Springer-Verlag, 2003.

[11] P. Heidelberger. Fast simulation of rare events in queueing and reliability models. *ACM TOMACS*, 5:43–85, 1995.

[12] S. Juneja and P. Shahabuddin. Simulating heavy tailed processes using delayed hazard rate twisting. *ACM TOMACS*, 12,2:94–118, 2002.

[13] C. Kollman, K. Baggerly, D. Cox, and R. Picard. Adaptive importance sampling on discrete markov chains. *Ann. Appl. Probab.*, 9:391–412, 1999.

[14] S. Resnick. Heavy tail modelling and teletraffic data. *Ann. Statis.*, 25:1805–1869, 1997.

[15] J.S. Sadowsky. Large deviations and efficient simulation of excessive backlogs in a GI/G/m queue. *IEEE Trans. on Aut. Control*, 36:1383–1394, 1991.

[16] D. Siegmund. *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag, New York, 1985.