Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

IMPORTANCE SAMPLING OF COMPOUNDING PROCESSES

Jose Blanchet

Department of Statistics
Harvard University
Cambridge, MA 02138, U.S.A.

Bert Zwart

Georgia Institute of Technology
H. Milton Stewart School of
Industrial and Systems Engineering
765 Ferst Drive
Atlanta, GA 30332, USA

ABSTRACT

Compounding processes, also known as perpetuities, play an important role in many applications; in particular, in time series analysis and mathematical finance. Apart from some special cases, the distribution of a perpetuity is hard to compute, and large deviations estimates sometimes involve complicated constants which depend on the complete distribution. Motivated by this, we propose provably efficient importance sampling algorithms which apply to qualitatively different cases, leading to light and heavy tails. Both algorithms have the non-standard feature of being state-dependent. In addition, in order to verify the efficiency, we apply recently developed techniques based on Lyapunov inequalities.

1 INTRODUCTION

This paper develops efficient simulation methods for estimating tail probabilities for random variables called perpetuities (also known as infinite horizon discounted rewards). An example of a perpetuity is a random variable D given by

$$D = \sum_{n=1}^{\infty} B_n e^{S_n}, \tag{1}$$

with $S_n = X_1 + \ldots + X_n$ a random walk and $\{B_k : k \geq 1\}$ an i.i.d. sequence of non-negative random variables which are independent of $\{S_n : n \geq 1\}$. The reason that D is called a perpetuity comes from the following interpretation. Consider a bond, generating a reward of B_n dollars at time n. The total present value of the reward at time n is not B_n, but a random variable depending on interest rate in the first n time units. Let $-X_n$ be the interest rate at time n. Then the present value of the reward at time n is $B_n e^{S_n}$. Thus, D is the present value of the bond at time 0. Extensions of this example lead to cases where time may be continuous (leading to integrals rather than sums), where the rewards may be negative, or both.

Perpetuities appear in many different applications. An obvious area is mathematical finance, but there are also applications in physics, communication networks, sorting algorithms, number theory, and more. For examples, see Vervaat (1979), Embrechts and Goldie (1994), Goldie and Grübel (1996). In particular, perpetuities naturally appear in time series analysis. For example, if we consider the generalized ARCH(1) sequence

$$D_{n+1} = A_n(D_n + B_n),$$

with (A_n, B_n) an i.i.d. sequence, the choice $A_n = e^{X_n}$ leads to (1), with D the weak limit of D_n as $n \to \infty$ (assuming $EX_n < 0$ and $E\log(|B_n|+1) < \infty$).

An explicit analysis of perpetuities is surprisingly hard. Even for the special case considered in this paper, it is hard to come up with exact results for the distribution of Z. Although some particular cases exist that allow for an explicit analysis (see e.g. Vervaat 1979, Gjessing and Paulsen 1997) it is clear that there is a need for different methods. Monte Carlo simulation arises as a natural approach to deal with analysis of perpetuities. Our focus here is on rare-event simulation methodology for efficient tail estimation of perpetuities, namely, $\beta(x) = P(D > x)$, for large values of x. The theoretical analysis of rare-event simulation algorithms involve measures of efficiency such as bounded relative error (or strong efficiency) and logarithmic efficiency, also known as asymptotic optimality or weak efficiency. Recall that an unbiased estimator $Z(x)$ for $\beta(x)$ possesses logarithmic efficiency if

$$\sup_{x \geq 1} \frac{E(Z(x))^2}{\beta(x)^{2-\varepsilon}} < \infty \tag{2}$$

for every $\varepsilon > 0$. If (2) holds with $\varepsilon = 0$, we say that $Z(x)$ possesses bounded relative error. Because of the infinite horizon nature of perpetuities (which forces an estimator that must be implemented in finite time at the expense of introducing some bias) it is relevant to study efficiency in the case of biased estimators and show that the relative bias can also be handled appropriately. In order to simplify the exposition in this paper, we shall just study the theoretical properties of the variance (in terms of (2)) for infinite horizon estimators. Dealing with the bias is typically not a problem in the current setting because the exponential discount allows to truncate the time horizon at a time that grows as a low degree polynomial as a function of the tail parameter x while controlling the bias.

There is a substantial amount of literature on this topic; a non-exhaustive list is Blanchet and Glynn (2005), Goldie (1991), Goldie and Grübel (1996), Konstantinides and Mikosch (2005) and Maulik and Zwart (2006). Depending on the distribution of X_1 and B_1, there can be many different types of tail behavior, ranging from extremely heavy (for example $1/\log x$), to extremely light even inducing bounding support. We focus on three qualitatively different examples in this paper.

The first example leads to heavy tails: If $X_n > 0$ with positive probability, it can happen that e^{S_n} can become very large: only $O(\log x)$ consecutive positive realizations of the X_i suffice for D to be of the order x. Therefore, D is typically heavy tailed. Under certain regularity conditions, it can be shown that

$$P(D > x) \sim C_{HT} x^{-\theta^*}, \tag{3}$$

with $\theta^* > 0$ the strictly positive solution of $E[e^{\theta^* X_1}] = 1$, a key reference is Goldie (1991). (For two functions $f(x)$ and $g(x)$ we use the notational convention $f(x) \sim g(x)$ to denote $\lim_{x\to\infty} f(x)/g(x) = 1$.) The constant C_{HT} given in Goldie (1991) depends on the entire distribution of D. An alternative, but still rather complicated expression has been given by Siegmund (2001). Thus, putting aside how good this asymptotic estimate performs for moderate values of x, this large deviations estimate is simply not explicit enough to use as approximation.

In the second example we consider the case in which the rate of return X_n at time n is negative with probability one. If, in addition, $P(B_n > x) \sim \gamma(x)e^{-\mu x}$ (with $\gamma(x)$ equal to a constant, or a power of x), one can show that

$$P(D > x) \sim C_{EXP} P(B_1 > x), \tag{4}$$

see Maulik and Zwart (2006). Again, the lack of explicitness of the pre-factor C_{EXP} is an issue. If this prefactor is not finite, it has to be replaced by a more complicated, subexponential function, which is not known in general.

Finally, if the rewards B_n are constant and $X_n < 0$ a.s., the tail behavior of Z is substantially lighter, leading to Poisson tails, i.e. $-\log P(D > x) = O(x \log x)$. To the best of our knowledge, the behavior of $P(D > x)$ itself is not known in this case, although logarithmic asymptotics can be found in Goldie and Grübel (1996) and de Bruijn (1951).

These results motivate the exploration of alternative methods (to large deviations) for evaluating $P(D > x)$. In the present work, we present importance sampling algorithms that can be used to obtain reliable estimates for $P(D > x)$ in all three cases. The algorithms we present are based on exponential twisting, but have the non-standard features that they are *time and/or state dependent*. Although state-dependent importance sampling schemes can be traced back early in the history of Monte Carlo methods (see Hammersley and Morton 1954), the theoretical analysis of such schemes is relatively new. In the presence of large deviation principles for light-tailed systems, Dupuis and Wang (2004) and Dupuis and Wang (2005) developed methodology for the efficient design and analysis of state-dependent importance samplers. Such methodology rests on the construction of solutions to non-linear partial differential equations and inequalities – in the latter case these inequalities correspond to subsolutions of an associated Isaacs equation. In situations where the stochastic system of interest is heavy-tailed (in particular, does not admit a traditional large deviations principle) or is purely combinatorial or discrete in nature, Blanchet and Glynn (2007) and Blanchet (2007) develop a methodology for efficient design and analysis of state-dependent importance samplers for rare-event simulation. In these situations, the analysis depends on stability bounds and Lyapunov inequalities that control the behavior of the second moment of the estimator. These Lyapunov inequalities basically correspond to subsolutions of the associated Isaacs equation when applied to traditional large deviations settings. We apply techniques based on Lyapunov inequalities in Section 4 in order to quantify the efficiency of the proposed scheme.

This paper is organized as follows. The Poisson case is investigated in Section 2. Our algorithm for exponential-tailed perpetuities is presented in Section 3. The heavy tailed case is investigated in Section 4.

2 POISSON TAILS

The first case we consider is an example leading to Poisson tails for D. Specifically, let $(-X_k : k \geq 1)$ be a sequence of exponentially distributed random variables with mean λ. We are interested in developing an efficient state-dependent importance sampling algorithm for efficient evaluation of $P(D > x)$ as $x \nearrow \infty$, where

$$\begin{aligned} D &= \sum_{k=1}^{\infty} \exp(S_k) = \int_0^{\infty} \exp(-u)\, dN(u) \\ &= -\int_0^{\infty} N(u)\, d\exp(-u) = \int_0^{\infty} \exp(-u) N(u)\, du. \end{aligned}$$

Here $N(\cdot)$ represents the Poisson process generated by the sequence $(-X_k : k \geq 1)$.

This random variable appears in several applications, particularly in analytic number theory and sorting algorithms, see de Bruijn (1951), Vervaat (1972) and Goldie and Grübel (1996). These papers all treat the behavior of $P(D > x)$ as $x \to \infty$. It is not difficult to see from the previous expression involving an exponential integral of the Poisson process that D has Poisson-type tails. Specifically, it holds that

$$\log P(D > x) \sim -x \log x. \tag{5}$$

Surprisingly, this estimate does not depend on the rate of the exponential distribution. This rate does play a role if one considers more refined estimates; see Vervaat (1972) for a higher order expansion of $\log P(D > x)$. To obtain estimates for $P(D > x)$, it is therefore reasonable to develop an efficient importance sampling algorithm. This is the goal of the current section. For convenience, we assume that the rate λ of the Poisson process equals 1.

Note first that

$$\begin{aligned} E \exp(\theta D) &= E \exp\left(\theta \int_0^\infty \exp(-u)\, dN(u)\right) \\ &= \exp\left(\int_0^\infty \psi(\theta \exp(-u))\, du\right) \\ &= \exp\left(\int_0^\theta \frac{\psi(y)}{y} dy\right). \end{aligned}$$

In order to construct the algorithm, select θ_x such that

$$\psi_D'(\theta_x) = x,$$

where

$$\psi_D(\theta) = \log E \exp(\theta D) = \int_0^\theta \frac{\psi(y)}{y} dy.$$

In the Poisson case that we consider here we have that

$$\psi(\beta) = \log E \exp(\beta N(u)) = \exp(\beta) - 1$$

and therefore $\psi_D'(\theta_x) = x$ yields

$$\psi(\theta_x) = \exp(\theta_x) - 1 = x\theta_x.$$

It follows that

$$\begin{aligned} \theta_x &= \log(1 + x\theta_x + \delta\theta_x) \\ &= \log(x\theta_x) + \log\left(1 + \frac{1}{x\theta_x}\right) \\ &= \log(x) + \log(\theta_x) + \frac{1}{x\theta_x} + O\left(\left(\frac{1}{x\theta_x}\right)^2\right), \end{aligned}$$

which in particular yields that

$$\theta_x = \log(x) + \log\log(x) + o(\log\log(x))$$

as $x \to \infty$.

We apply importance sampling according by tilting the Poisson process $N(\cdot)$ dynamically in time. In order to describe the dynamics of the proposed importance sampling scheme, let us denote by $\widetilde{P}(\cdot)$ the probability law induced by the proposed change-of-measure. The dynamics under the importance sampler can be described as follows. Let $M(\cdot)$ be a Poisson process with unit rate under $\widetilde{P}(\cdot)$. Next, define

$$\Lambda(t) = \int_0^t \exp(\theta_x \exp(-s))\, ds.$$

Then, under $\widetilde{P}(\cdot)$ we have that $N(\cdot)$ has the same distribution as $M(\Lambda(\cdot))$. In other words, we have

$$\begin{aligned} &\widetilde{P}(N(t+h) - N(t) = 0) = 1 - \exp(\theta_x \exp(-t))\, h + o(h) \\ &\widetilde{P}(N(t+h) - N(t) = 1) = \exp(\theta_x \exp(-t))\, h + o(h) \\ &\widetilde{P}(N(t+h) - N(t) = 1) = o(h) \end{aligned}$$

as $h \searrow 0$. Note that representing $N(\cdot)$ in terms of $M(\Lambda(\cdot))$ under $\widetilde{P}(\cdot)$ is particularly useful in terms of implementing the algorithm by means of a thinning procedure. For brevity reasons we shall not provide the complete details here.

The estimator for $P(D > x)$ then becomes

$$L = \exp\left(\int_0^\infty \psi(\theta_x \exp(-u))\, du - \theta_x D\right) I(D > x). \tag{6}$$

The second moment of the estimator is

$$\begin{aligned} &\widetilde{E}L^2 = \exp(-2x\theta_x) \times \\ &\widetilde{E}\left(\exp\left(\int_0^\infty 2\psi(\theta_x \exp(-u))\, du - 2\theta_x Z\right) I(Z > 0)\right) \\ &\leq \exp(-2x\theta_x) \exp\left(\int_0^\infty 2\psi(\theta_x \exp(-u))\, du\right), \end{aligned}$$

where $Z = D - x$. Note that

$$\int_0^\infty 2\psi(\theta_x \exp(-u))\, du = 2\int_0^{\theta_x} \frac{\exp(y) - 1}{y} = O(x)$$

(in fact, this can be further expanded). We can now invoke (5) to conclude that our algorithm is logarithmically efficient. We record this result in the form of a proposition.

Proposition 1. *The estimator (6) obtained using the law $\widetilde{P}(\cdot)$ described before is logarithmically efficient as $x \nearrow \infty$.*

3 EXPONENTIAL TAILS

We assume now that the rewards B_i are exponentially distributed with rate λ, and that the interest rates are non-negative. We also sum from 0 in (1). Although the assumption on B_i is not critical, this case is of intrinsic interest, since in this case, D can be expressed as

$$D = \int_0^\infty e^{-X(s)} ds, \tag{7}$$

with $\{X(s)\}$ a compound Poisson process. Functionals this type are natural continuous-time analogues of perpetuities and appear in many applications. See Bertoin and Yor (2005) for a survey.

The logarithmic asymptotics of $P(D > x)$ are the same as those of $P(B_1 > x)$, but the exact asymptotics may require a polynomial correction term, see Maulik and Zwart (2006). This suggests that a large value of D is caused by a single or several large values of B, which is similar to what happens in exceedance probabilities for heavy-tailed random walks.

To explain our algorithm, note that

$$E[e^{\theta D}] = E\left[\prod_{k=0}^{\infty} \frac{\lambda}{\lambda - \theta e^{S_k}}\right].$$

We propose to perform a state-dependent simulation algorithm where the distribution of the random walk $\{S_k\}$ is unperturbed, but where the B_k's are sampled from an exponential distribution with rate $\lambda - \theta_x e^{S_k}$, with $\theta_x = \lambda - c/x$ for some $c \in (0,1)$. This yields the estimator

$$L = I(\sum_{k=0}^{\infty} e^{S_k} B_k > x) \prod_{k=0}^{\infty} \frac{\lambda}{\lambda - \theta_x e^{S_k}} \exp\{-\theta_x \sum_{k=0}^{\infty} e^{S_k B_k}\}. \tag{8}$$

We now prove that this estimator is efficient. Under the original measure, the second moment of this estimator is equal to

$$E\left[I(\sum_{k=0}^{\infty} e^{S_k} B_k > x) \prod_{k=0}^{\infty} \frac{\lambda}{\lambda - \theta_x e^{S_k}} \exp\{-\theta_x \sum_{k=0}^{\infty} e^{S_k B_k}\}\right],$$

which is smaller than

$$e^{-\theta_x x} E\left[I(\sum_{k=0}^{\infty} e^{S_k} B_k > x) \prod_{k=0}^{\infty} \frac{\lambda}{\lambda - \theta_x e^{S_k}}\right].$$

To complete the proof, we make the simplifying assumption that there exists an $\varepsilon > 0$ such that $X_1 \leq -\varepsilon$ a.s.; the general case will be treated in the extended version of this paper. Under this assumption, we can upper bound the second moment further by

$$e^{-\theta_x x} P(\sum_{k=0}^{\infty} e^{S_k} B_k > x) \prod_{k=0}^{\infty} \frac{\lambda}{\lambda - \theta_x \delta^k},$$

with $\delta = e^{-\varepsilon} < 1$. This regularity condition also guarantees that $C_{EXP} < \infty$. Thus, the first term behaves like $P(D > x)$ up to a constant, while the second term equals $P(D > x)$. To establish logarithmic efficiency, not that the first term in the infinite product behaves like $O(x)$, and that the remaining (from $k = 1$ on) infinite product remains bounded as $x \to \infty$. To see this, write

$$\prod_{k=1}^{\infty} \frac{\lambda}{\lambda - \theta_x \delta^k} = -\sum_{k=1}^{\infty} \log(1 - \delta^k(1 - c/(\lambda x))).$$

Let x be larger than c/λ and let M be large enough such that $\delta^M < 1/2$. Then $\log(1 - \delta^k(1 - c/(\lambda x))) \geq -2\delta^k(1 - c/(\lambda x))$ for $k \geq M$. From this, the desired boundedness easily follows.

We summarize our findings as follows:

Proposition 2. *If $X_1 \leq -\varepsilon$ a.s., the estimator (8) obtained using the law $\widetilde{P}(\cdot)$ described before has logarithmic efficiency as $x \nearrow \infty$.*

To control the bias when implementing this algorithm, note that the infinite sum needs to be truncated at some point M. If one let M depend on x in such a way that $M = M(x) \to \infty$ when $x \to \infty$, then one can show that the relative bias vanishes, while preserving optimality.

To illustrate the applicability of the algorithm we consider a case that is not covered by our proposition: assume that the negative discount rates, $-X_k$, are exponentially distributed with mean 10%. The rewards, B_k's, are exponentially distributed with mean 1. In this case, D is gamma distributed with both mean and variance equal to 11 (cf. Bertoin and Yor 2005). The following table was obtained using 10,000 replications (100 periods per replication were simulated). The column IS displays the importance sampling estimator. $\widehat{\sigma}_{IS}$ shows the estimated standard deviation of the importance sampling estimator and $\widehat{\sigma}_{MCM}$ corresponds to the sample standard deviation for the crude Monte Carlo estimator (we did not include the values of the CMC estimator because they are far away from the exact values).

4 HEAVY TAILS

Suppose that $(X_n : n \geq 1)$ is a sequence of i.i.d. rv's with logarithmic moment generating function $\psi(\theta) = \log E \exp(\theta X_n)$ and there exists $\theta^* > 0$ such that $\psi(\theta^*) = 0$.

x	$P(D>x)$	IS	$\widehat{\sigma}_{IS}$	$\widehat{\sigma}_{CMC}$
15	$1.18e-01$	$1.15e-01$	$1.54e-01$	$4.53e-01$
20	$1.08e-02$	$1.01e-02$	$1.96e-02$	$3.18e-01$
25	$5.86e-04$	$5.39e-04$	$1.34e-03$	$2.03e-01$
30	$2.23e-05$	$2.08e-05$	$6.13e-05$	$1.18e-01$
35	$6.61e-07$	$5.97e-07$	$2.02e-06$	$6.32e-02$

Figure 1: Illustration of the algorithm.

We want to estimate $P(D>x)$ for x large, where we take the rewards $B_n = 1$, so that

$$D = \sum_{k=1}^{\infty} \exp(S_k).$$

Note that by convexity of $\psi(\cdot)$ we have that $\psi'(0) = EX_1 < 0$ and therefore D is a well defined random variable. Let us write $x = b/\Delta$ for some $b > 0$ and $\Delta > 0$. We shall let $x \nearrow \infty$ by sending $\Delta \searrow 0$ while holding $b > 0$ fixed. Introducing the parameter Δ is slightly more convenient in order to describe a scaled process for which the tail probability of interest coincides with a first-passage time computation. More precisely, let us define the process $W = ((D_n, S_n) : n \geq 0)$ via

$$(D_{n+1}, S_{n+1}) = (D_n + \Delta \exp(S_n + X_{n+1}), S_n + X_{n+1})$$

and put $T_\Delta = \inf\{n \geq 0 : D_n > b\}$. Let us write $E_{d,s}(\cdot)$ to denote the corresponding expectation operator. In the next steps we shall drop the super-script Δ on D_n. In other words, we wish to estimate $P_{(0,0)}(T_\Delta < \infty)$.

Our goal is to design a state-dependent importance sampling strategy based on exponential tilting. The strategy is to select the exponential tilting according to the state of the chain W. Our choice of tilting is guided by the selection of a parametric family of Lyapunov functions that control the behavior of the second moment of our estimator. We tune both the parameters of the Lyapunov function and the exponential tilting in order to satisfy a linear inequality which implies the bound on the second moment.

A closely related strategy has been used by Dupuis and Wang (2005) in the context of systems in which standard large deviations scaling is in force – in which case, the Lyapunov inequalities is transformed, after taking a limit under the large deviations scaling, into differential inequalities. Blanchet, Glynn, and Liu (2007) use a similar strategy in the context of heavy-tailed multi-server queues. In such case, the proposed family of changes-of-measure is, obviously, not parameterized in terms of exponential changes-of-measure. However, a completely analogous program to the one described next can be shown to give rise to an efficient algorithm by selecting the family of changes-of-measure based on mixtures.

Our estimator takes the form

$$L = \exp\left(-\sum_{k=0}^{T_\Delta - 1} \theta_k X_{k+1} + \sum_{k=0}^{T_\Delta - 1} \psi(\theta_k)\right) I(T_\Delta < \infty),$$

where θ_k depends on the D_k and S_k (one could even allow dependence on the whole history up to time k, but this is not necessary). If we denote by $\widetilde{P}_{d,s}(\cdot)$ the distribution of the chain W under the change-of-measure generated by the exponential tiltings (i.e. the θ_k's) given that $(D_0, S_0) = (d, s)$, then we have

$$\begin{aligned} &\widetilde{P}_{d,s}(X_{k+1} \in dx | (D_1, S_1), \ldots, (D_k, S_k)) \\ &= P_{d,s}(X_{k+1} \in dx) \exp(\theta_k x - \psi(\theta_k)). \end{aligned} \tag{9}$$

The estimator L is obtained by generating the X_k's according to (9).

Note that the second moment of the estimator L can be expressed via a function $r(\cdot)$ according to the equality

$$\begin{aligned} &r(d,s) = \\ &\widetilde{E}_{d,s}\left(\exp\left(-\sum_{k=0}^{T_\Delta-1} 2\theta_k X_{k+1} + \sum_{k=0}^{T_\Delta-1} 2\psi(\theta_k)\right) I(T_\Delta < \infty)\right) \\ &= E_{d,s}\left(\exp\left(-\sum_{k=0}^{T_\Delta-1} \theta_k X_{k+1} + \sum_{k=0}^{T_\Delta-1} \psi(\theta_k)\right) I(T_\Delta < \infty)\right); \end{aligned}$$

in particular, $EL^2 = r(0,0)$.

The following proposition can be proved as in Blanchet and Glynn (2007), the details are given in an extended version of this paper, namely Blanchet and Zwart (2007).

Proposition 3. *If there exists a positive function $h(\cdot)$ such that*

$$\begin{aligned} h_\Delta(d,s) \quad \geq \quad & E_{d,s}(\exp(-\theta X_1 + \psi(\theta)) \times \\ & h_\Delta(d + \Delta \exp(s + X_1), s + X_1)) \end{aligned}$$

subject to $h_\Delta(d,s) \geq \delta$ if $d > b$, then

$$\begin{aligned} &\frac{h_\Delta(d,s)}{\delta} \geq \\ &E_{d,s}\left(\exp\left(-\sum_{k=0}^{T_\Delta-1} \theta_k X_{k+1} + \sum_{k=0}^{T_\Delta-1} \psi(\theta_k)\right) I(T_\Delta < \infty)\right). \end{aligned}$$

The previous result allows to measure the efficiency of our estimator in terms of its second moment. In particular, we note that

$$\frac{\widetilde{E}_{0,0}L^2}{P_{0,0}(T_\Delta < \infty)^2} \leq \frac{h_\Delta(0,0)}{\delta P_{0,0}(T_\Delta < \infty)^2}.$$

If $h_\Delta(d,s)$ behaves similarly as $P_{d,s}(T_\Delta < \infty)$, then we will be in good shape to prove efficiency. It is known (Goldie 1991) that if $d < b$

$$P_{d,s}(T_\Delta < \infty) \sim C\left(\frac{b-d}{\Delta}\right)^{-\theta^*} \exp(s\theta^*)$$

for some constant $C \in (0,\infty)$ (independent of d and s) as $\Delta \searrow 0$. Motivated by this observation, we choose $h_\Delta(d,s)$ to equal

$$\min\left(c_\Delta \Delta^{2\theta^*-\rho_\Delta}(b-d)_+^{-2\theta^*+\rho_\Delta}\exp((2\theta^*-\rho_\Delta)s),1\right), \tag{10}$$

where $(b-d)_+ \triangleq \max(b-d,0)$ and the constants c_Δ and ρ_Δ (possibly depending on Δ) will be selected in order to satisfy the Lyapunov inequality. At the end we shall send $\rho_\Delta \searrow 0$ as $\Delta \searrow 0$ at a suitable rate so that logarithmic efficiency is maintained. In particular, we have the following result.

Proposition 4. *Let $c_\Delta = \log(1/\Delta)^{2\theta^*}$. Then, one can compute $\lambda, \Delta_0 > 0$ such that if $\rho_\Delta = \lambda c_\Delta^{-1/(2\theta^*)}$ and $\Delta \leq \Delta_0$, inequality (10) holds.*

In the remainder of this section we shall prove the previous proposition and at the same time provide the elements of the algorithm. If $h_\Delta(d,s) < 1$ then

$$\begin{aligned}
&\frac{E_{d,s}[\exp(-\theta X + \psi(\theta))h_\Delta(d+\Delta e^{s+X}, s+X)]}{h_\Delta(d,s)} \\
&= E_{d,s}\left(g_\Delta(s) \wedge h_\Delta(d,s)^{-1}\right),
\end{aligned} \tag{11}$$

with $g_\Delta(s)$ equal to

$$\exp(-(\theta-2\theta^*+\rho_\Delta)X+\psi(\theta)) \times \left(1-\Delta\frac{\exp(s+X)}{(b-d)}\right)_+^{-2\theta^*+\rho_\Delta}.$$

Now fix $a \in (0,1)$ and consider the events

$$A_1 = \left\{\Delta\frac{\exp(s+X)}{(b-d)} < a\right\},$$
$$A_2 = \left\{\Delta\frac{\exp(s+X)}{(b-d)} \geq a\right\}.$$

If we select $\theta = \theta^*$, then (11) is less or equal than

$$\begin{aligned}
&E_{d,s}\Bigg(\exp((\theta^*-\rho_\Delta)X) \times \\
&\left(1+\Delta(2\theta^*-\rho_\Delta)\eta\frac{\exp(s+X)}{(b-d)}\right); A_1\Bigg) \\
&+ h_\Delta(d,s)^{-1}E_{d,s}[\exp((\theta^*-\rho_\Delta)X); A_2],
\end{aligned}$$

where

$$\eta = \frac{(1-a)^{-2\theta^*+\rho_\Delta}-1}{2\theta^*-\rho_\Delta}.$$

Suppose that if $Z = e^X$ then there exists $c \in (0,\infty)$ such that $P(Z > t) \leq c(1+t)^{-3\theta^*-1}$. This holds if $E[e^{(3\theta^*+1)X}] < \infty$. Since $c_\Delta = o(\Delta^{-\varepsilon})$ for each $\varepsilon > 0$, then

$$\begin{aligned}
&\frac{E_{d,s}[\exp((\theta^*-\rho_\Delta)X); e^X > a\Delta^{-1}e^{-s}(b-d)]\exp(\psi(\theta_\Delta))}{h_\Delta(d,s)} \\
&\leq \frac{h_\Delta(d,s)^{-1}E_{d,s}[Z^{\theta^*-\rho_\Delta}; Z > a\Delta^{-1}e^{-s}(b-d)]}{h_\Delta(d,s)} \\
&\leq \widetilde{c}c_\Delta^{-1}\left(\frac{b-d}{\Delta}\right)^{-1}e^s,
\end{aligned}$$

for some constant $\widetilde{c} > 0$. Combining these estimates we obtain that expectation (11) is less or equal to

$$\begin{aligned}
&E_{d,s}\left(\exp((\theta^*-\rho_\Delta)X)\left(1+\Delta 2\theta^*\eta\frac{\exp(s+X)}{(b-d)}\right)\right) \\
&+\widetilde{c}c_\Delta^{-1}\frac{\Delta\exp(s)}{(b-d)} \\
&\leq \exp(\psi(\theta^*-\rho_\Delta)) + 2\theta^*\eta\frac{\exp(s)\Delta}{b-d}\exp(\psi(\theta^*+1)) \\
&+\widetilde{c}c_\Delta^{-1}\frac{\Delta\exp(s)}{(b-d)}.
\end{aligned} \tag{12}$$

There exists $\delta_1 > 0$ such that if $\rho_\Delta \in (0,\delta_1)$, then

$$\exp(\psi(\theta^*-\rho_\Delta)) \leq 1-\psi'(\theta^*)\rho_\Delta/2.$$

On the other hand, we are assuming that $h_\Delta(d,s) < 1$, which implies

$$\frac{\Delta\exp(s)}{b-d} < c_\Delta^{-1/(2\theta^*-\rho_\Delta)}.$$

Hence, if $\rho_\Delta \in (0, \delta_1)$ we obtain that the right hand side of the inequality in display (12) is upper bounded by

$$1-\psi'(\theta^*)\rho_\Delta/2 \\ +2\theta^*\eta c_\Delta^{-1/(2\theta^*-\rho_\Delta)}\exp(\psi(\theta^*+1)) \\ +\widetilde{c}c_\Delta^{-1}c_\Delta^{-1/(2\theta^*-\rho_\Delta)}.$$

In order to satisfy the Lyapunov bound on the region $h_\Delta(d,s) < 1$ it suffices to choose ρ_Δ such that

$$\psi'(\theta^*)\rho_\Delta/2 \geq 2\theta^*\eta c_\Delta^{-1/(2\theta^*-\rho_\Delta)}\exp(\psi(\theta^*+1)) \\ +\widetilde{c}c_\Delta^{-1}c_\Delta^{-1/(2\theta^*-\rho_\Delta)}.$$

This implies that is feasible to choose $\rho_\Delta = O\left(c_\Delta^{-1/(2\theta^*)}\right)$ as $\Delta \searrow 0$.

Now, observe that when $h_\Delta(d,s) = 1$, satisfying the Lyapunov bound is an easy task. In such case, just pick $\theta_k = 0$ and the bound is guaranteed to hold because $h_\Delta(\cdot)$ is less than one throughout its domain.

We now provide an explicit description of the algorithm:

Algorithm
1. Set ρ_Δ, c_Δ and $h(\cdot)$ according to (10).
2. Let $L \longleftarrow 1$, $D \longleftarrow 0$ and $S \longleftarrow 0$
3. Repeat until $D > b/\Delta$
   If $h(d,s) = 1$, then sample X according to the nominal (original) distribution
   Else, sample X with distribution

$$P_{\theta^*}(X \in dx) = \exp(\theta^* x)P(X \in dx)$$

   and put $L \longleftarrow \exp(-\theta^* X)\cdot L$.
   Endif
   Update $S \longleftarrow S + X$, $D \longleftarrow D + \exp(S)$
End
Output L

The next result summarizes the efficiency properties of the estimator.

Theorem 5. *Suppose that $E[e^{(3\theta^*+1)X}] < \infty$. The estimator L obtained from the previous algorithm is logarithmically efficient, that is*

$$\lim_{\Delta\searrow 0}\frac{\log \widetilde{E}_{0,0}L^2}{2\log P(D > b/\Delta)} = 1.$$

ACKNOWLEDGMENT

We are grateful to Peter Glynn for valuable discussions related to Section 3. This research was partially supported by NSF grant DMS 0595595.

REFERENCES

Bertoin, J., and M. Yor. 2005. Exponential functionals of Lévy processes. *Probab. Surv.* 2:191–212 (electronic).

Blanchet, J. 2007. Efficient importance sampling for counting. *Submitted for Publication*.

Blanchet, J., and P. Glynn. 2005. Large deviations and sharp asymptotics for perpetuities with small discount rates. *Submitted for Publication*.

Blanchet, J., and P. Glynn. 2007. Efficient rare-event simulation for the maximum of heavy-tailed random walks. *Submitted for Publication*.

Blanchet, J., P. Glynn, and J. Liu. 2007. Fluid heuristics, lyapunov bounds and efficient importance sampling for a heavy-tailed g/g/1 queue. *Submitted for Publication*.

Blanchet, J., and B. Zwart. 2007. Efficient rare event simulation for perpetuities. *In preparation*.

de Bruijn, N. G. 1951. The asymptotic behaviour of a function occurring in the theory of primes. *J. Indian Math. Soc. (N.S.)* 15:25–32.

Dupuis, P., and H. Wang. 2004. Importance sampling, large deviations, and differential games. *Stoch. Stoch. Rep.* 76 (6): 481–508.

Dupuis, P., and H. Wang. 2005. Dynamic importance sampling for uniformly recurrent Markov chains. *Ann. Appl. Probab.* 15 (1A): 1–38.

Embrechts, P., and C. M. Goldie. 1994. Perpetuities and random equations. In *Asymptotic statistics (Prague, 1993)*, Contrib. Statist., 75–86. Heidelberg: Physica.

Gjessing, H. K., and J. Paulsen. 1997. Present value distributions with applications to ruin theory and stochastic equations. *Stochastic Process. Appl.* 71 (1): 123–144.

Goldie, C. M. 1991. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* 1 (1): 126–166.

Goldie, C. M., and R. Grübel. 1996. Perpetuities with thin tails. *Adv. in Appl. Probab.* 28 (2): 463–480.

Hammersley, J. M., and K. W. Morton. 1954. Poor man's Monte Carlo. *J. Roy. Statist. Soc. Ser. B.* 16:23–38; discussion 61–75.

Konstantinides, D. G., and T. Mikosch. 2005. Large deviations and ruin probabilities for solutions to stochastic recurrence equations with heavy-tailed innovations. *Ann. Probab.* 33 (5): 1992–2035.

Maulik, K., and B. Zwart. 2006. Tail asymptotics for exponential functionals of Lévy processes. *Stochastic Process. Appl.* 116 (2): 156–177.

Siegmund, D. 2001. Note on a stochastic recursion. In *State of the art in probability and statistics (Leiden, 1999)*, Volume 36 of *IMS Lecture Notes Monogr. Ser.*, 547–554. Beachwood, OH: Inst. Math. Statist.

Vervaat, W. 1972. *Success epochs in Bernoulli trials (with applications in number theory)*. Amsterdam: Mathematisch Centrum. Mathematical Centre Tracts, No. 42.

Vervaat, W. 1979. On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab.* 11 (4): 750–783.

AUTHOR BIOGRAPHIES

JOSE BLANCHET is Assistant Professor of Statistics at Harvard University. Jose holds a M.Sc. in Engineering-Economic Systems and Operations Research and a Ph.D. in Management Science and Engineering, both from Stanford University. He also holds two B.Sc. degrees: one in Actuarial Science and another one in Applied Mathematics from ITAM (Mexico). Jose worked for two years as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability, computational finance, performance engineering, queueing theory, risk management, rare-event analysis, statistical inference, stochastic modeling, and simulation. He serves on the editorial boards of *ACM TOMACS* and *Statistica Sinica.*

BERT ZWART is Coca Cola Associate Professor of Industrial and Systems Engineering at Georgia Tech. Bert holds an M.A. in Econometrics from the Free University, Amsterdam, and a Ph.D in Applied Mathematics from Eindhoven University of Technology. His research is concerned with modeling, analysis and simulation of stochastic systems arising in actuarial and financial mathematics, computer and communication systems, manufacturing systems, and service systems. He serves on the editorial boards of *Operations Research*, *Mathematics of Operations Research*, and *Mathematical Methods of Operations Research*.