

EFFICIENT SPLITTING-BASED RARE EVENT SIMULATION ALGORITHMS FOR HEAVY-TAILED SUMS

Jose Blanchet
Yixi Shi

Columbia University
New York, NY 10027, USA

ABSTRACT

Rare events in heavy-tailed systems are challenging to analyze using splitting algorithms because large deviations occur suddenly. So, every path prior to the rare event is viable and there is no clear mechanism for rewarding and splitting paths that are moving towards the rare event of interest. We propose and analyze a splitting algorithm for the tail distribution of a heavy-tailed random walk. We prove that our estimator achieves the best possible performance in terms of the growth rate of the relative mean squared error, while controlling the population size of the particles.

1 INTRODUCTION

Most of the probably efficient rare event simulation estimators for heavy-tailed systems have been designed using importance sampling, see for example (Dupuis et al. 2006, Blanchet et al. 2007), and (Blanchet and Glynn 2008, Blanchet and Liu 2008, Blanchet and Li 2011, Chan and Lai 2012). Conditional Monte Carlo, is also a very effective technique, and one of the first to yield provably efficient estimators in heavy-tailed settings, see (Binswanger and Hojgaard 1997, Asmussen and Kroese 2006). In contrast, in the setting of light-tailed systems (where the inputs have exponentially decaying tails), both importance sampling and splitting procedures are popular approaches applied in the construction of provably efficient rare event simulation algorithms, see (Asmussen and Glynn 2008). Apart from the one that we report in this paper, we do not know of any splitting algorithm for large deviations in heavy tailed random walks which has been proved to be efficient. Efficiency is understood in the sense of controlling the optimal rate of decay of the amount of work required to produce an estimate of the probability of interest with a given degree of relative precision, see (Asmussen and Glynn 2008), and also Theorem 1 below.

In order to explain why splitting is difficult to apply in rare event simulation for heavy-tailed random walks let us recall how splitting works at a high level. In traditional splitting applications, the simulator constructs a sequence of nested “milestone” events (with the last milestone event corresponding to the rare event of interest). The milestone events, whose sequential occurrence is not rare, are judiciously placed. Particles are then evolved according to the system’s nominal dynamics and kept splitting whenever a new milestone is reached. Attached with each particle is a weight defined by the total number of times the particle has split so that the final estimator is unbiased. We refer readers to (Glasserman et al. 1999) and (Villén-Altamirano and Villén-Altamirano 2002) for a review on earlier developments in the splitting method. In summary, the milestone events and the splitting procedure at the time of reaching milestone events, are designed to increase the population of particles that behave in a way that is consistent with producing the event of interest.

The issue that arises in the context of large deviations for heavy-tailed random walks, for instance, is that just prior to the occurrence of a large deviations event, the random walk evolves basically according to its nominal (unconditional) dynamics. So, there is no obvious way to place the milestone events in order to encourage particles that behave in a way that is consistent with producing large deviations. Simply put,

all typical paths prior to the event of interest behave in a way that is consistent with the large deviations event.

We are interested in designing and studying splitting estimators for heavy-tailed systems because, often, provably efficient splitting estimators are often easier to design than importance sampling estimators; at least there is evidence that supports this feature of splitting estimators in light-tailed systems. Indeed, in Dupuis et al. (2007) and Dupuis and Wang (2009) it is shown that in a conventional large deviations environment for light-tailed systems, asymptotically optimal importance sampling strategies can be constructed out of classical sense sub-solutions of a certain control problem. Meanwhile, according to Dean and Dupuis (2009), in the same large deviations environment, asymptotically optimal splitting based algorithms can be constructed using only weak-sense sub-solutions to the same equation. Weak-sense sub-solutions are often easier to construct than classical sense sub-solutions.

In this paper we take a step to explore rare event simulation via splitting based estimators for heavy-tailed stochastic systems. A natural class of problems to start with is the efficient estimation of the tail probability

$$p_b = \mathbb{P}(S_n > b), \quad (1)$$

where $S_n = X_1 + X_2 + \dots + X_n$ and the X_i 's are i.i.d. random variables, with a suitable heavy-tailed structure. This problem has served as a starting point in the analysis of virtually any importance sampling strategy that has been considered for heavy-tailed systems. This is partly why this model problem appears to be a natural starting point to start with our investigations in the setting of splitting.

We must be clear that there are a number of very efficient importance sampling based algorithms for (1) as $b \rightarrow \infty$ (see, for example, Dupuis et al. 2006, Blanchet and Liu 2012, Blanchet and Li 2011, Juneja and Shahabuddin 2002, Blanchet and Lam 2012). The goal of this paper is not trying to develop an algorithm that is superior in efficiency to some of the existing algorithms in this setting; but rather we contribute by giving a first attempt to explore the idea of crafting splitting procedures in rare event simulation for heavy-tailed systems that are provably efficient. Our motivation is to see if, as in the light-tailed case, splitting algorithms might have a hope of being easier to set up while still maintaining provable efficiency (in the form of weak efficiency or logarithmic efficiency). As we shall see, we conclude that, in some sense, there seems to be some evidence that this may well be the case. While some of the importance sampling strategies that are available (see Blanchet and Liu 2012) provide a stronger control on the second moment of the estimator, those importance sampling strategies require different setup depending on specific tail properties of the increments (see, for example, the counterexample Shi 2013). The splitting based algorithms considered here benefit from an easier setup as the same treatment is given to all applicable random variables (including Weibull and Pareto). This flexibility, we believe, is similar in spirit to our prior discussion involving the difficulty in applying importance sampling vs splitting in light-tailed settings.

Finally, we comment on the form of our splitting algorithm. The different nature of how large deviations occur in a heavy-tailed system forces us to abandon the idea of splitting in the original state space. Instead of splitting in the original state space, we embed a splitting procedure in the hazard function space, and then transform back to the original space to obtain the sampled increments.

We close our discussion by noting that there is a literature on particle methods in general (of which splitting algorithms are just a particular case). This important body of work has been fueled by the connection to genealogical particle systems and Feynman-Kac flows, see for instance (del Moral 2004), and in a related class of problems (C erou et al. 2006). Likely, the algorithms that we propose here can be embedded, in form, in a Feynman-Kac flow representation by enlarging the state-space.

The rest of the paper is organized as follows. Section 2 we list our assumptions, then provide the form of the splitting estimator that we analyze and then state our main result, which summarizes the asymptotic optimality of the estimator as $b \rightarrow \infty$. The proof of our main result is given in Section 3. We include some numerical examples in Section 4.

2 ASSUMPTIONS, ESTIMATOR AND MAIN RESULT

2.1 Assumptions

Let $\{X_j : 1 \leq j \leq n\}$ be independent and identically distributed (i.i.d.) random variables (r.v.'s) with distribution $F(\cdot)$, with support $[0, \infty)$. The integrated hazard function, $\Lambda(\cdot)$, of X is defined as $\Lambda(x) = -\log \bar{F}(x)$.

Assumption 1 The spectrum of distributions we consider satisfy the following conditions.

- A1 $\Lambda(x)$ is strictly increasing in x .
- A2 The random variables X_i are subexponential in the sense that

$$P(X_1 + \dots + X_n > b) = n \exp(-\Lambda(b)) (1 + o(1))$$

as $b \rightarrow \infty$.

- A3 The hazard rate $\lambda(t) = d\Lambda(t)/dt$ exists is non increasing.

Assumption A3) can be replaced by requiring $\lambda(\cdot)$ to be eventually non increasing. These assumptions are satisfied by most heavy-tailed random variables that arise in practice (such as Pareto and Weibull distributions, see Embrechts et al. 1997).

2.2 The Splitting Estimator and Main Result

Sampling of a single increment X is conducted in two phases. In the first phase we use a splitting based procedure to sample $\Lambda(X)$, which is exponentially distributed with unit mean. In the second phase, we transform it back to the original space with the inverse function $\Lambda^{-1}(\cdot)$. We explain how to sample the component X_1 and then we will explain how to proceed with the rest.

The splitting based procedure is perhaps best described in terms of a tree construction. Let Δ be a pre-determined positive number and define $m(b)$, the total number of Δ -sized levels, via

$$m = m(b) = \min\{k \leq 1 : k\Delta \geq \Lambda(b)\} = \lceil \Lambda(b)/\Delta \rceil.$$

We introduce an integer $r \geq 2$ which will correspond to the number of children that are obtained at every single splitting of a particle.

We now explain precisely how to construct a tree required to sample the first increment X_1 . Let us write Π_0 to denote a trivial tree containing only a single element which we call root; we write $\Pi_0 = \{0\}$. Let $\mathcal{Z}_0 = \{0\}$ be the set of elements (nodes) in Π_0 that are associated to active particles; initially we have only one active particle corresponding to the root. The set of elements in Π_0 that are inactive is denoted by $\mathcal{D}_0 = \Pi_0 \setminus \mathcal{Z}_0 = \emptyset$.

Now, we introduce the following notation to enlarge any element of a tree. Given $s = (s_1, \dots, s_k)$ and $s_{k+1} \in \{0, 1, \dots, r\}$ we have that $((s_1, \dots, s_k), s_{k+1}) = (s_1, \dots, s_k, s_{k+1})$.

We say that $t = (t_1, \dots, t_k)$ is an *enlargement* of $s = (s_1, \dots, s_l)$, denoted via $t \succ s$, if $l \leq k$ and $s_1 = t_1, \dots, s_l = t_l$. We say that t is a *strict enlargement* of s , denoted via $t \succeq s$, if $k < l$ and $s_1 = t_1, \dots, s_l = t_l$.

Using this notation, the trees Π_1, \dots, Π_m , along with the set of active nodes $\mathcal{Z}_1, \dots, \mathcal{Z}_m$ and inactive nodes $\mathcal{D}_1, \dots, \mathcal{D}_m$ are constructed according to the following recursive procedure. Define $\mathcal{I}_{k-1} = \{\mathbf{I}(u) = I(A(u) > \Delta) : u \in \mathcal{Z}_{k-1}\}$, where $A(u)$'s form a collection of i.i.d. exponentially distributed r.v.'s with unit mean (if $\mathcal{Z}_{k-1} = \emptyset$, then $\mathcal{I}_{k-1} = \emptyset$). So, in particular $E(\mathbf{I}(u)) = \exp(-\Delta)$. Then,

$$\mathcal{Z}_k = \cup_{u \in \mathcal{Z}_{k-1} : \mathbf{I}(u)=1} \{(u, 1), \dots, (u, r)\}, \tag{2}$$

$$\Pi_k = \Pi_{k-1} \cup_{u \in \mathcal{Z}_k} \{(u, 1), \dots, (u, r)\}, \tag{3}$$

$$\mathcal{D}_k = \{u \in \mathcal{Z}_{k-1} : \mathbf{I}(u) = 0\}. \tag{4}$$

for $1 \leq k \leq m$. Finally, define $\mathcal{D}_{m+1} = \mathcal{L}_m$. Note that if $\mathcal{L}_k = \emptyset$, then $\Pi_k = \Pi_{k-1}$, and $\mathcal{D}_k = \mathcal{L}_{k-1}$.

Next, define $\mathcal{L}(\Pi_{k-1})$ to be the set of leaves at the top of Π_{k-1} . In precise terms, $s \in \mathcal{L}(\Pi_{k-1})$ if s does not have any enlargement in Π_{k-1} . Observe the decomposition

$$\mathcal{L}(\Pi_m) = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \dots \cup \mathcal{D}_m \cup \mathcal{L}_m, \quad (5)$$

where $\mathcal{D}_1, \dots, \mathcal{D}_m, \mathcal{L}_m$ are disjoint sets.

Next, for any $l \geq 0$, let us write $\Pi_l(s)$ to denote an i.i.d. copy of the tree Π_l , but rooted at $(s, 0)$; so, $\Pi_0(s) = \{(s, 0)\}$ and $\mathcal{L}(\Pi_0(s)) = \Pi_0(s)$. The definition of \mathcal{L}_k and \mathcal{D}_k is adapted in an obvious way to their counterparts $\mathcal{L}_k(s)$ and $\mathcal{D}_k(s)$, constructed when evolving the recursions rooted at s . In particular, given $\mathcal{I}_{k-1}(s) = \{\mathbf{I}(u) = I(A(u) \leq \Delta) : u \in \mathcal{L}_{k-1}(s)\}$, we write

$$\begin{aligned} \mathcal{L}_k(s) &= \cup_{u \in \mathcal{L}_{k-1}(s) : \mathbf{I}(u)=1} \{(u, 1), \dots, (u, r)\}, \\ \Pi_k(s) &= \Pi_{k-1}(s) \cup_{u \in \mathcal{L}_k(s)} \{(u, 1), \dots, (u, r)\}, \\ \mathcal{D}_k(s) &= \{u \in \mathcal{L}_{k-1}(s) : \mathbf{I}(u) = 1\}, \end{aligned} \quad (6)$$

for $1 \leq k \leq m$ and define $\mathcal{D}_{m+1}(s) = \mathcal{L}_m(s)$.

We now are ready to provide a recursive construction of the overall splitting procedure. Define $\bar{\Pi}_{0,l} = \Pi_l$ for $0 \leq l \leq m$, and set

$$\bar{\Pi}_{i,l} = \bar{\Pi}_{i-1,m} \cup \{\Pi_l(s) : s \in \mathcal{L}(\bar{\Pi}_{i-1,m})\}, \quad (7)$$

for $i = 1, 2, \dots, n-1$. The splitting procedure is completed after constructing $\bar{\Pi}_{n-1,m}$.

In order to extract the increments from $\bar{\Pi}_{n-1,m}$ we need some final definitions. For every $s \in \mathcal{L}(\bar{\Pi}_{j-1,l})$ for $1 \leq j \leq n$ and $0 \leq l \leq m-1$ define the *steam* of s via $(h_i(s) : 1 \leq i \leq j-1)$ where $h_{i-1}(s) \in \mathcal{L}(\bar{\Pi}_{i-1,m})$ and $s \succeq h_{i-1}(s)$ and we set $h_0(s) = 0$. Further, for every $s \in \mathcal{L}(\bar{\Pi}_{j-1,l})$ for $1 \leq j \leq n$ and $0 \leq l \leq m$ define

$$L(s) = \dim(s) - \dim(h_{j-1}(s)) - 1,$$

where $\dim(t)$ is the dimension (the number of components) of t .

Finally, for each $t \in \mathcal{L}(\bar{\Pi}_{i-1,m})$ with $1 \leq i \leq n$ define

$$V_i(t) = L(t)\Delta + (A(t) - \Delta)I(L(t) = m) + A(t)I(L(t) \leq m-1). \quad (8)$$

It follows, by the memory less property, that $V(t)$ is exponentially distributed with mean one. Moreover, for each $s \in \mathcal{L}(\bar{\Pi}_{n-1,m})$ we have that

$$\{V(h_1(s)), V(h_2(s)), \dots, V(h_{n-1}(s)), V(h_n(s))\},$$

where $h_n(s) = s$, forms a set of n i.i.d. r.v.'s exponentially distributed with unit mean. Consequently, for each $X_j(s) = \Lambda^{-1}(V(h_j(s)))$ is distributed according to $F(\cdot)$.

The splitting estimator, \hat{p}_b , can be computed as follows: Select $r = \exp(\Delta)$, and compute

$$\hat{p}_b = \sum_{s \in \mathcal{L}(\bar{\Pi}_n)} r^{-L(h_1(s))} r^{-L(h_2(s))} \dots r^{-L(h_n(s))} \mathbf{J}_b(s), \quad (9)$$

where $J_b(s) = I(X_1(s) + \dots + X_n(s) > b)$.

For each $s \in \mathcal{L}(\bar{\Pi}_{n-1,m})$, the factor $r^{-L(h_1(s))} r^{-L(h_2(s))} \dots r^{-L(h_n(s))}$ is called the weight of particle s .

It is straightforward to check that \hat{p}_b is unbiased, that is, $E(\hat{p}_b) = p_b$. We now state our main result which estimates both $E(\hat{p}_b^2)$ and the cost required to generate a single replication of \hat{p}_b , which is computed in terms

$$\mathcal{W}(b) = E(|\mathcal{L}(\bar{\Pi}_{n-1,m})|),$$

where $|\mathcal{A}|$ denotes the cardinality of any set \mathcal{A}

Theorem 1 Under assumptions A1) to A3) we have that the work normalized achieves asymptotic optimality. More precisely, we have that

$$E(\widehat{p}_b^2) \mathscr{W}(b) = O(p_b^{2-\varepsilon}), \tag{10}$$

for any $\varepsilon > 0$.

The criterion (10) is equivalent to requiring the total number of random variables necessary to obtain one single estimate has to grow at least at the same rate as the work-normalized squared coefficient of variation. When considering the splitting based estimator, this notion of efficiency is by far the most commonly used (see (Dean and Dupuis 2009)). The proof of Theorem 1 is given in Section 3. In particular, Lemma 1 provides a bound for $\mathscr{W}(b)$ and Lemma 2 provides estimates for $E(\widehat{p}_b^2)$. These results combined in particular imply Theorem 1.

It is important to mention, as we shall see, that the estimates that we shall obtain appear to deteriorate significantly as a function of n . No effort has been made in order to improve the estimates for large n . Our emphasis here is on b large and moderate n . We believe that improved bounds for n and b growing simultaneously are possible, but the construction of the subtrees corresponding to the increments must change somewhat, depending on which increment one is simulating. Progress on these types of constructions will be reported in future work.

3 PROOF OF THEOREM 1

The proof of Theorem 1 is divided in several parts. First, we study an estimate for $\mathscr{W}(b)$. The second part involves an estimate for $E(\widehat{p}_b^2)$.

3.1 Expected Number of Particles

The following result provides an estimate for $\mathscr{W}(b)$.

Lemma 1

$$\mathscr{W}(b) = [(r \exp(-\Delta))^m + (1 - \exp(-\Delta)) \sum_{k=1}^m (r \exp(-\Delta))^{k-1}]^n.$$

In particular, if $r \exp(-\Delta) = 1$, we have that $\mathscr{W}(b) = O(m^n) = O(\Lambda(b)^n)$ as $b \rightarrow \infty$.

Proof. From the (2)-(4), we have that if $Z_k = |\mathscr{Z}_k|$ and $D_k = |\mathscr{D}_k|$ we have the recursion

$$Z_k = \sum_{s \in \mathscr{Z}_{k-1}} r \mathbf{I}(s), \text{ and } D_k = \sum_{s \in \mathscr{Z}_{k-1}} (1 - \mathbf{I}(s)) = Z_{k-1} - Z_k/r,$$

with $Z_0 = 1$. In particular, Z_k follows a Galton-Watson process (see Harris 1963). We then have that

$$E(Z_k) = (r \exp(-\Delta))^k, \text{ and } E(D_k) = (r \exp(-\Delta))^{k-1} (1 - \exp(-\Delta)).$$

The decomposition (5) implies

$$E|\mathscr{L}(\Pi_m)| = E(Z_m) + \sum_{k=1}^m E(D_k) = (r \exp(-\Delta))^m + (1 - \exp(-\Delta)) \sum_{k=1}^m (r \exp(-\Delta))^{k-1}.$$

In total, we have by construction that $\mathscr{W}(b) = (E|\mathscr{L}(\Pi_m)|)^n$, and thus we obtain the result. □

3.2 The Second Moment of the Splitting Estimator

In order to estimate $E(\widehat{p}_b^2)$, following the ideas in (Dean and Dupuis 2009), it is useful to introduce the so-called ‘‘Fully Branching Representation’’ of \widehat{p}_b . To this end we introduce a sequence of full trees whose construction is completely parallel to that of $\bar{\Pi}_l$ and $\bar{\Pi}_i$. In particular, Π'_l is a tree with root at 0, with l generations, and each node - except for the leaves at the top - has exactly r children. So, in total $\mathcal{L}(\Pi'_l)$ has r^l elements and $\Pi'_0 = \{0\}$. Now, define $\bar{\Pi}'_{0,l} = \Pi'_l$ and recursively set for $i \geq 1$ and $0 \leq l \leq m$,

$$\bar{\Pi}'_{i,l} = \bar{\Pi}'_{i-1,m} \cup \{\Pi'_l(s) : s \in \mathcal{L}(\bar{\Pi}'_{i-1,m})\},$$

where $\Pi'_l(s)$ is a full tree (each node except the leaves has r children) with l generations rooted at $(s, 0)$. Observe that $|\mathcal{L}(\bar{\Pi}'_{i,l})| = r^{im+l}$.

Now, if $s \in \mathcal{L}(\bar{\Pi}_n)$ we have the steam information $h_1(s), h_2(s), \dots, h_n(s)$, with $h_0(s) = 0$ and $h_n(s) = s$. We also have that

$$h_1(s) \in \mathcal{D}_{L(h_1(s))+1}(h_0(s)), h_2(s) \in \mathcal{D}_{L(h_2(s))+1}(h_1(s)), \dots, h_n(s) \in \mathcal{D}_{L(h_n(s))+1}(h_{n-1}(s)).$$

So, each particle $s \in \mathcal{L}(\bar{\Pi}_n)$ is replaced by $r^{m-L(h_1(s))} r^{m-L(h_2(s))} \dots r^{m-L(h_n(s))}$ particles each with weight r^{-nm} , and these particles are all put together in a set denoted by $\mathcal{K}_n(s)$. The sets $\{\mathcal{K}_n(s) : s \in \mathcal{L}(\bar{\Pi}_n)\}$ are disjoint and their union contains r^{nm} elements, and so it can be put into one to one correspondence with the set $\mathcal{L}(\bar{\Pi}'_{n-1,m})$. Thus, we will write $\mathcal{L}(\bar{\Pi}'_{n-1,m}) = \cup\{\mathcal{K}_n(s) : s \in \mathcal{L}(\bar{\Pi}_n)\}$. Now to each $y \in \mathcal{L}(\bar{\Pi}'_{n-1,m})$ we can associate a unique element $\theta_y = s \in \mathcal{L}(\bar{\Pi}_n)$; in simple words, θ_y is the unique element in $\mathcal{L}(\bar{\Pi}_n)$ such that $y \in \mathcal{K}_n(\theta_y)$.

It is clear then that following equivalent representation for \widehat{p}_b holds,

$$\widehat{p}_b = \sum_{y \in \mathcal{L}(\bar{\Pi}'_n)} r^{-mn} \mathbf{J}_b(\theta_y). \tag{11}$$

Of course, representation (11) is only useful for theoretical purposes, not for computational ones. For computational purposes we we should keep in mind (9).

Representation (11) then implies

$$\widehat{p}_b^2 = \sum_{y \in \mathcal{L}(\bar{\Pi}'_{n-1,m})} r^{-2mn} \mathbf{J}_b(\theta_y) + \sum_{y,z \in \mathcal{L}(\bar{\Pi}'_{n-1,m}); y \neq z} r^{-2mn} \mathbf{J}_b(\theta_y) \mathbf{J}_b(\theta_z). \tag{12}$$

It will be useful to decompose the second summation in (12) in terms of the last common ancestor of y and z (i.e. $x \in \bar{\Pi}'_{n-1,m}$ such that $y, z \succeq x$ and there is no $v \in \bar{\Pi}'_{n-1,m}$, with $y, z \succeq v \succ x$). Just as we partitioned $\mathcal{L}(\bar{\Pi}'_{n-1,m})$ into as many equivalence classes as members of the set $\mathcal{L}(\bar{\Pi}_{n-1,m})$ we can also partition $\mathcal{L}(\bar{\Pi}'_{i-1,l})$ for $1 \leq i \leq n$ and $0 \leq l \leq m$, into equivalence classes $\{\mathcal{K}_{i-1,l}(s) : s \in \mathcal{L}(\bar{\Pi}_{i-1,l})\}$. So, if $z \in \mathcal{L}(\bar{\Pi}'_{i-1,l})$ we can define $\theta_z = s \in \mathcal{L}(\bar{\Pi}_{i-1,l})$ such that $z \in \mathcal{K}_{i-1,l}(s)$.

Now, if $x \in \bar{\Pi}'_{i-1,l}$ with $1 \leq i \leq n$ and $0 \leq l \leq m-1$ we define *the set of children of x* , denoted by $\mathcal{C}(x)$, as the immediate descendants of x , namely $\mathcal{C}(x) = \{(x, 1), \dots, (x, r)\}$. We then can write

$$\begin{aligned} & \sum_{y,z \in \bar{\Pi}'_{n-1,m}; y \neq z} r^{-2mn} \mathbf{J}_b(\theta(z)) \mathbf{J}_b(\theta(y)) \\ &= \sum_{i=1}^n \sum_{l=0}^{m-1} \sum_{x \in \mathcal{L}(\bar{\Pi}'_{i-1,l})} \sum_{y', z' \in \mathcal{C}(x); y' \neq z'} r^{-2(i-1)m} r^{-2l} \left(\sum_{y: y \succeq y'} r^{-mn+(i-1)m+l} \mathbf{J}_b(\theta_y) \right) \left(\sum_{z: z \succeq z'} r^{-mn+(i-1)m+l} \mathbf{J}_b(\theta_z) \right). \end{aligned} \tag{13}$$

Now we are ready to state a result that described the behavior of $E(\widehat{p}_b^2)$.

Lemma 2 If $r = \exp(\Delta)$, under Assumptions A1) to A3) there exists a (computable) constant $\kappa_n > 0$ such that

$$E(\widehat{p}_b^2) \leq \kappa_n p_b^2.$$

Proof. Define

$$G_1(b) = \sum_{y \in \mathcal{L}(\overline{\Pi}'_{n-1,m})} r^{-2mn} E[\mathbf{J}_b(\theta_y)], \quad G_2(b) = \sum_{y,z \in \mathcal{L}(\overline{\Pi}'_{n-1,m}): y \neq z} r^{-2mn} E[\mathbf{J}_b(\theta_y) \mathbf{J}_b(\theta_z)].$$

Using (12) we have that

$$E(\widehat{p}_b^2) = G_1(b) + G_2(b).$$

Using A2) we have that

$$G_1(b) = \sum_{y \in \mathcal{L}(\overline{\Pi}'_{n-1,m})} r^{-2mn} p(b) = r^{-mn} p(b) \leq r p(b)^{n+1} C_n, \quad (14)$$

for some $C_n \in (0, \infty)$. In order to analyze $G_2(b)$ using representation (13), define for $x \in \mathcal{L}(\overline{\Pi}'_{i-1,l})$ with $1 \leq i \leq n$ and $0 \leq l \leq m-1$, the σ -field

$$\mathcal{F}_x = \sigma(\{\mathbf{I}(s) : \theta_x \succeq s\}).$$

Note that conditional on \mathcal{F}_x , we have we know that $L_j = l_j$ for $j = 1, \dots, i$, where the random variables L_j are defined via

$$L_1 = L(h_1(\theta_x)), \dots, L_{i-1} = L(h_{i-1}(\theta_x)), L_i = L(\theta_x) \wedge l.$$

Now define i.i.d. r.v.'s $(V_j : 1 \leq j \leq n)$ exponentially distributed with unit mean. Note that we can write always write

$$V_j = I(V_j < m\Delta) (\Delta \lfloor V_j/\Delta \rfloor + (V_j - \Delta \lfloor V_j/\Delta \rfloor)) + V_j I(V_j \geq m\Delta);$$

this decomposition corresponds to (8), in the sense that we can easily check that $(V(t), \min(\lfloor V(t)/\Delta \rfloor, m)) =_d (V(t), L(t))$, where the symbol “ $=_d$ ” denotes equality in distribution. Given \mathcal{F}_x , the random variables $\sum_{y: y \succeq y'} r^{-mn+im+l} \mathbf{J}_b(\theta(y))$ and $\sum_{z: z \succeq z'} r^{-mn+im+l} \mathbf{J}_b(\theta(z))$ are independent and identically distributed and

$$E\left(\sum_{y: y \succeq y'} r^{-mn+im+l} \mathbf{J}_b(\theta(y)) \mid \mathcal{F}_x\right) = p_{n-i}(b \mid L_1, \dots, L_i),$$

where

$$\begin{aligned} p_{n-i}(b \mid l_1, \dots, l_i) &= P(\Lambda^{-1}(V_1) + \dots + \Lambda^{-1}(V_n) > b \mid \lfloor V_k/\Delta \rfloor \wedge m = l_k, k \leq i-1, \lfloor V_i/\Delta \rfloor \wedge l = l_i) \\ &\leq P(\Lambda^{-1}((V_i - \Delta)^+ + (l_i + 1)\Delta) + \Lambda^{-1}(V_{i+1}) + \dots + \Lambda^{-1}(V_n) > b - l_1^- - \dots - l_{i-1}^-), \end{aligned}$$

Therefore,

$$E\left(\left[\sum_{y: y \succeq y'} r^{-mn+im+l} \mathbf{J}_b(\theta(y))\right] \times \left[\sum_{z: z \succeq z'} r^{-mn+im+l} \mathbf{J}_b(\theta(z))\right] \mid \mathcal{F}_x\right) = p_{n-i}^2(b \mid l_1, \dots, l_i), \quad (15)$$

where

$$\begin{aligned} &p_{n-i}(b \mid l_1, \dots, l_i) \\ &= P(\Lambda^{-1}(V_1) + \dots + \Lambda^{-1}(V_n) > b \mid \lfloor V_k/\Delta \rfloor \wedge m = l_k, k \leq i-1, \lfloor V_i/\Delta \rfloor \wedge l = l_i). \end{aligned}$$

Note that if $\lfloor V_k/\Delta \rfloor \wedge m = l_k$, we have that

$$\Lambda^{-1}(V_k) \leq \Lambda^{-1}((V_k - \Delta m)^+ + (l_k + 1)\Delta).$$

Moreover, conditional on $V_k > \Delta m$, we have that

$$\Lambda^{-1}((V_k - \Delta m)^+ + (l_k + 1)\Delta) =_d \Lambda^{-1}(V'_k + \Delta),$$

where V'_k is independent of V_k . Of course, conditional on $V_k \leq \Delta m$ we simply have

$$\Lambda^{-1}((V_k - \Delta m)^+ + (l_k + 1)\Delta) = \Lambda^{-1}((l_k + 1)\Delta).$$

Putting these considerations together we see that

$$\begin{aligned} & p_{n-i}(b|l_1, \dots, l_i) \\ & \leq P(\Lambda^{-1}(V'_1 + \Delta) + \dots + \Lambda^{-1}(V'_i + \Delta) + \Lambda^{-1}(V_{i+1}) + \dots + \Lambda^{-1}(V_n) > b - l_1^- - \dots - l_i^-), \end{aligned} \quad (16)$$

where $l_j^- = \Lambda^{-1}((l_j + 1)\Delta)$ for $j = 1, \dots, i$. It then follows easily from the subexponential property (A2) and (16) that there exists $C'_n \in (0, \infty)$ such that

$$p_{n-i}(b|l_1, \dots, l_i) \leq C'_n \exp(-\Lambda(b - l_1^- - \dots - l_i^-)). \quad (17)$$

Now, since $\lambda(t)$ is non increasing (by A3) we have that for any $a_1, a_2 > 0$

$$\Lambda(a_1 + a_2) \leq \Lambda(a_1) + \int_{a_1}^{a_1+a_2} \lambda(s) ds \leq \Lambda(a_1) + \int_0^{a_2} \lambda(s) ds = \Lambda(a_1) + \Lambda(a_2).$$

Therefore, applying the previous inequality i times we conclude that

$$\Lambda(b) \leq \Lambda(b - l_1^- - \dots - l_i^-) + \Lambda(l_1^-) + \dots + \Lambda(l_i^-),$$

which yields that

$$\exp(-\Lambda(b - l_1^- - \dots - l_i^-)) \leq \exp(-\Lambda(b) + \Delta l_1 + \dots + \Delta l_i + i\Delta). \quad (18)$$

Similarly

$$\exp(-\Lambda(b - l_1^- - \dots - l_{i-1}^-)) P(V_i > \Delta l) \leq \exp(-\Lambda(b) + \Delta l_1 + \dots + \Delta l_{i-1} + (i-1)\Delta - \Delta l). \quad (19)$$

We then conclude that there is $C_n \in (0, \infty)$ such that

$$p_{n-i}(b|l_1, \dots, l_i) \leq \bar{p}_{n-i}(b|l_1, \dots, l_i) := C_n \exp(-\Lambda(b) + \Delta l_1 + \dots + \Delta l_i).$$

Plugging in this bound into (15) and using (13) we conclude that $G_2(b)$ is bounded by

$$\begin{aligned} & \sum_{i=1}^n \sum_{l=0}^{m-1} \sum_{x \in \mathcal{L}(\bar{\Pi}'_{i-1,l})} \sum_{y', z' \in \mathcal{C}(x): y' \neq z'} r^{-2(i-1)m} r^{-2l} E(\bar{p}_{n-i}^2(b|L_1, \dots, L_i)) \\ & = C_n \exp(-2\Lambda(b)) \sum_{i=1}^n \sum_{l=0}^{m-1} \sum_{x \in \mathcal{L}(\bar{\Pi}'_{i-1,l})} \sum_{y', z' \in \mathcal{C}(x): y' \neq z'} \exp(-2(i-1)m\Delta) \exp(-2l\Delta) E \exp(2\Delta L_1 + \dots + 2\Delta L_i). \end{aligned} \quad (20)$$

Now, recall that $L_i =_d \lfloor V_i/\Delta \rfloor \wedge m$ and therefore

$$E \exp(2\Delta L_1) = \sum_{l=0}^{m-1} \exp(2\Delta l) \exp(-\Delta l) (1 - \exp(-\Delta)) + \exp(\Delta m) \leq 2 \exp(\Delta m).$$

We conclude that

$$E \exp(2\Delta L_1 + \dots + 2\Delta L_i) \leq 2^i \exp(\Delta m(i-1)) \exp(\Delta l).$$

Consequently, plugging this estimate into (20) we have that $G_2(b)$ is bounded by

$$\begin{aligned} C_n 2^n \exp(-2\Lambda(b)) & \sum_{i=1}^n \sum_{l=0}^{m-1} \sum_{x \in \mathcal{L}(\bar{\Pi}_{i-1,l})} \sum_{y', z' \in \mathcal{C}(x): y' \neq z'} \exp(-(i-1)m\Delta) \exp(-l\Delta) \\ & = C_n 2^n \exp(-2\Lambda(b)) \sum_{i=1}^n \sum_{l=0}^{m-1} r^{(i-1)m+1} \exp(-(i-1)m\Delta) \exp(-l\Delta) \\ & = C_n 2^n \exp(-2\Lambda(b)) r \sum_{i=1}^n \sum_{l=0}^{m-1} \exp(-l\Delta) \leq \frac{n C_n 2^n \exp(-2\Lambda(b)) r}{1 - \exp(-\Delta)} = O(p_b^2). \end{aligned} \tag{21}$$

Combining bound (14) and (21) together we conclude the proof of the lemma. □

4 NUMERICAL EXAMPLES

In this section, we implement and test the proposed splitting algorithm on the following examples, for various choices of b :

- (i) $p_1 = P(X_1 + \dots + X_4 > b)$, where X_j 's are Pareto with index $\alpha = 1.5$, i.e., $P(X > x) = 1/(1+x)^\alpha$.
- (ii) $p_2 = P(Y_1 + \dots + Y_4 > b)$, where Y_j 's are Weibull, with parameter $\gamma = 0.5$ such that $P(Y > y) = \exp(-2(y+1)^\gamma)$.

The performance of the splitting algorithm (SPLT) is compared against 1) crude Monte Carlo, 2) the state-dependent importance sampling (SDIS) in (Dupuis et al. 2006) and (Blanchet and Liu 2012) (for the Pareto and Weibull cases respectively), and 3) hazard-rate twisting (JS-IS) in (Juneja and Shahabuddin 2002). Note that importance sampling requires a different set-up depending on the type of tail behavior (Weibull or Pareto). In contrast, splitting preserves their basic structure (i.e. do not need a special adaptation) both for Pareto and Weibull tails and in that sense is robust. As we indicated in the introduction, importance sampling is expected to work better because it takes into account more information on the structure of the tails, where as splitting basically only uses the (eventual) non-increasing property of the hazard rate in an essential way. We included the JS-IS strategy because this one is state-independent. Our splitting strategy is also state-independent so the comparison here, we believe is more meaningful with regard to the main objective of this experiment, namely, to have a sense of how much does one lose when using splitting vs importance sampling. We conclude that the efficiency loss could be significant specially when compared to state-dependent strategies. However, splitting still appears to be a good alternative against crude Monte Carlo (running from around six to twenty times faster). We believe that our results call for additional investigation on splitting strategies in which the tree construction varies slightly depending on the corresponding increment corresponding to the subtree in question. We intend to pursue these investigations in the future.

The results are demonstrated in Tables 1 - 2 below. For each algorithm, we report the following quantities:

- 1) *Estimate*. The SDIS, JS-IS and SPLT algorithms are all run $N = 10^6$ times. For crude Monte Carlo, we produce $N = 10^8$ replications for each example.

- 2) *Work-normalized relative error.* For each algorithm, this is calculated as the equivalent relative error of the estimate as if the algorithm is run for the same length of time as the benchmark crude Monte Carlo. In particular, let T, T_c be the running time for the current algorithm and the crude Monte Carlo, respectively, then the work-normalized relative error for this splitting algorithm is calculated as

$$\overline{RE}^{normalized} = \left(\frac{Var(\hat{p}) T}{N\hat{p} T_c} \right)^{1/2},$$

where \hat{p} is a single replication of the associated estimator under consideration.

- 3) *Variance reduction factor,* which is calculated as $\overline{RE}^{crudeMC} / \overline{RE}^{normalized}$, where $\overline{RE}^{crudeMC}$ is the relative error of the crude Monte Carlo estimator.

Table 1: Numerical results for p_1 , i.e., sums of Pareto with $\alpha = 1.5$.

$b = 5 \times 10^4$	Crude MC	SDIS	JS-IS	SPLT
Estimate	3.80×10^{-7}	3.58×10^{-7}	3.54×10^{-7}	3.51×10^{-7}
Work-normalized rel. err.	16.22%	0.02%	0.24%	1.89%
Var. reduction factor	1.00	811.00	67.74	8.60
$b = 10^5$	Crude MC	SDIS	JS-IS	SPLT
Estimate	1.10×10^{-7}	1.26×10^{-7}	1.26×10^{-7}	1.25×10^{-7}
Work-normalized rel. err.	30.15%	0.02%	0.27%	1.25%
Var. reduction factor	1.00	1918.64	111.00	24.15

Table 2: Numerical results for p_2 , i.e., sums of Weibull with $\beta = 0.5$.

$b = 50$	Crude MC	SDIS	JS-IS	SPLT
Estimate	2.28×10^{-6}	2.14×10^{-6}	2.11×10^{-6}	2.14×10^{-6}
Work-normalized rel. err.	6.62%	0.18%	0.16%	1.13%
Var. reduction factor	1.00	360.53	228.55	5.85
$b = 75$	Crude MC	SDIS	JS-IS	SPLT
Estimate	1.00×10^{-7}	9.12×10^{-8}	9.17×10^{-8}	9.15×10^{-8}
Work-normalized rel. err.	35.36%	0.20%	0.26%	2.24%
Var. reduction factor	1.00	173.66	138.37	15.81

5 CONCLUSIONS

We have discussed a provably efficient splitting estimator for tail probabilities of random walks. The estimator is particularly suited for heavy-tailed increments. The results are interesting because standard splitting estimators require the underlying process to gradually move towards the rare-event region. Heavy-tailed random walks, however, will typically reach a rare-event region by a sudden jump, so it is not clear how to apply splitting (i.e. it is not clear when to split particles and thus reward "good" behavior). However, heavy-tailed increments typically have a decreasing hazard rate. Thus, it is natural to perform splitting by looking both at the state of the random walk and by simulating the next increment in (conditional) segments that become increasingly large and perform splitting accordingly. This is the strategy that this paper studies and we show its asymptotic efficiency. The resulting estimator is also robust in the sense of requiring a uniform setup to produce probably efficient estimators - contrary to state-dependent importance sampling, which requires different setup depending on the asymptotic behavior of the hazard rate. It is an interesting (future) research project to consider random walks in large deviations scaling (i.e. large number of increments) and other types of processes.

ACKNOWLEDGMENTS

We thank the referee for his/her comments. We gratefully acknowledge NSF support through grants CMMI-0846816 and CMMI-1069064.

REFERENCES

- Asmussen, S., and P. Glynn. 2008. *Stochastic Simulation: Algorithms and Analysis*. New York, NY, USA: Springer-Verlag.
- Asmussen, S., and D. Kroese. 2006. “Improved algorithms for rare event simulation with heavy tails.”. *Advances in Applied Probability* 38:545–558.
- Binswanger, S. A. K., and B. Hojgaard. 1997. “Rare events simulation for heavy-tailed distributions.”. *Bernoulli* 6:303–322.
- Blanchet, J., and P. Glynn. 2008. “Efficient rare-event simulation for the maximum of a heavy-tailed random walk.”. *Ann. of Appl. Probab.* 18:1351–1378.
- Blanchet, J., P. Glynn, and J. C. Liu.. 2007. “Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed $G/G/1$ queue.”. *QUESTA* 57:99–113.
- Blanchet, J., and H. Lam. 2012. “State-dependent importance sampling for rare-event simulation: An overview and recent advances”. *Surveys in Operations Research and Management Science* 17 (1): 38–59.
- Blanchet, J., and C. Li. 2011. “Efficient rare-event simulation for heavy-tailed compound sums”. *ACM Transactions on Modeling and Computer Simulation - TOMACS* 21:1–10.
- Blanchet, J., and J. Liu. 2012. “Efficient Simulation and Conditional Functional Limit Theorems for Ruinous Heavy-tailed Random Walks”. *Stochastic Processes and Their Applications* 122:2994–3031.
- Blanchet, J., and J. C. Liu. 2008. “State-dependent importance sampling for regularly varying random walks”. *Adv. in Appl. Probab.* 40:1104–1128.
- Cérou, F., P. D. Moral, F. L. Gland, and P. Lezaud. 2006. “Genetic genealogical models in rare event analysis”. *Latin American Journal of Probability and Mathematical Statistics*.
- Chan, H. P., and T. L. Lai. 2012. “Rare-event simulation of heavy-tailed random walks by sequential importance sampling and resampling.”. *Advances in Applied Probability* 44:1173–1196.
- Dean, T., and P. Dupuis. 2009. “Splitting for Rare Event Simulation: A Large Deviation Approach to Design and Analysis”. *Stochastic Processes and Its Applications* 119:562–587.
- del Moral, P. 2004. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York, USA: Springer-Verlag.
- Dupuis, P., K. Leder, and H. Wang. 2006. “Importance sampling for sums of random variables with regularly varying tails.”. *ACM TOMACS* 17:Article 14.
- Dupuis, P., A. Sezer, and H. Wang.. 2007. “Dynamic importance sampling for queueing networks.”. *Ann. Appl. Probab.* 17:1306–1346.
- Dupuis, P., and H. Wang. 2009, June. “Importance sampling for Jackson networks”. *Queueing Syst. Theory Appl.* 62 (1-2): 113–157.
- Embrechts, P., C. Klppelberg, and T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. New York: Springer-Verlag.
- Glasserman, P., P. Heidelberger, P. Shahabuddin, and T. Zajic. 1999. “Multilevel Splitting for Estimating Rare Event Probabilities”. *Operations Research* 47:585–600.
- Harris, T. 1963. *The Theory of Branching Processes*. New York: Springer-Verlag.
- Juneja, S., and P. Shahabuddin. 2002. “Simulating heavy-tailed processes using delayed hazard rate twisting”. *ACM TOMACS* 12:94–118.
- Shi, Y. 2013. *Rare Events in Stochastic Systems*. Ph. D. thesis, Columbia University.
- Villén-Altamirano, M., and J. Villén-Altamirano. 2002. “Analysis of restart simulation: Theoretical basis and sensitivity study”. *European Transactions on Telecommunications* 13:373–385.

AUTHOR BIOGRAPHIES

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

YIXI SHI is a PhD candidate in Department of Industrial Engineering and Operations Research, School of Engineering and Applied Science at Columbia University. He holds a B.Sc. in Actuarial Science from University of Hong Kong. His email address is ys2347@columbia.edu.