Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Efficient Rare-Event Simulation for Multiple Jump Events in Regularly Varying Random Walks and Compound Poisson Processes

Bohan Chen, Jose Blanchet, Chang-Han Rhee, Bert Zwart

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 44, No. 3, August 2019, pp. 919–942
ISSN 0364-765X (print), ISSN 1526-5471 (online)
http://pubsonline.informs.org/journal/moor/

Efficient Rare-Event Simulation for Multiple Jump Events in Regularly Varying Random Walks and Compound Poisson Processes

Bohan Chen,[a] **Jose Blanchet,**[b] **Chang-Han Rhee,**[c] **Bert Zwart**[a,d]

[a] Stochastics Group, Centrum Wiskunde & Informatica, 1098 XG Amsterdam, Netherlands; [b] Department of Management Science and Engineering, Stanford University, Stanford, California 94306; [c] Department of Engineering and Management Sciences, Northwestern University, Evanston, IL 60208; [d] Centrum Wiskunde & Informatica, Amsterdam 1098 XG, Netherlands

Contact: bohan.chen@cwi.nl, http://orcid.org/0000-0002-0586-3855 (BC); jose.blanchet@stanford.edu, http://orcid.org/0000-0001-5895-0912 (JB); chang-han.rhee@northwestern.edu, http://orcid.org/0000-0002-1651-4677 (C-HR); bert.zwart@cwi.nl, http://orcid.org/0000-0001-9336-0096 (BZ)

Received: June 14, 2017
Revised: November 13, 2017
Accepted: January 7, 2018
Published Online in Articles in Advance: May 23, 2019

MSC2000 Subject Classification: Primary: 65C05; secondary: 60F10, 60G51
OR/MS Subject Classification: Primary: simulation, efficiency; secondary: algorithms, limit theorems, stochastic model applications, random walk

https://doi.org/10.1287/moor.2018.0950

Abstract. We propose a class of strongly efficient rare-event simulation estimators for random walks and compound Poisson processes with a regularly varying increment/jump-size distribution in a general large deviations regime. Our estimator is based on an importance sampling strategy that hinges on a recently established heavy-tailed sample-path large deviations result. The new estimators are straightforward to implement and can be used to systematically evaluate the probability of a wide range of rare events with bounded relative error. They are "universal" in the sense that a single importance sampling scheme applies to a very general class of rare events that arise in heavy-tailed systems. In particular, our estimators can deal with rare events that are caused by multiple big jumps (therefore, beyond the usual principle of a single big jump) as well as multidimensional processes such as the buffer content process of a queueing network. We illustrate the versatility of our approach with several applications that arise in the context of mathematical finance, actuarial science, and queueing theory.

Funding: The work of B. Chen, C-H. Rhee, and B. Zwart was supported by the Netherlands Organisation for Scientific Research [Vici Grant 639.033.413]. The work of J. Blanchet was supported by the National Science Foundation [Grants 132055, 1538217, 1820942] and the Defense Advanced Research Projects Agency [Grant N660011824028].

Keywords: rare-event simulation • strong efficiency • regularly varying distribution • principle of multiple big jumps • random walks • compound Poisson processes • large deviations results

1. Introduction

In this paper, we develop a strongly efficient importance sampling scheme for computing rare-event probabilities involving path functionals of heavy-tailed random walks and compound Poisson processes in a general large deviations regime. Heavy-tailed distributions play an important role in many man-made stochastic systems. For example, they accurately model inputs to computer and communications networks (see, e.g., Foss et al. [20]), and they are an essential component of the description of many financial risk processes (see, e.g., Embrechts et al. [17]).

We focus on stochastic processes with a regularly varying increment/jump-size distribution. The estimators produced with our sampling scheme are straightforward to implement and can be used to estimate the likelihood of a wide range of rare events with bounded relative error. In particular, such a single sampling scheme applies to a very general class of rare events whose occurrence is caused by one or several components in the system that exhibit extreme behavior, whereas the rest of the system is operating in "normal" circumstances (therefore, beyond the usual principle of a single big jump). In particular, our results apply to a large class of continuous functionals of multiple random walks and compound Poisson processes.

Our estimators are based on importance sampling, a Monte Carlo technique that consists of biasing the nominal distribution of the underlying process to induce the rare event of interest. The estimator is obtained by weighting each sample by the corresponding likelihood ratio to obtain unbiased estimators. Our goal is to find biasing techniques leading to estimators that have a bounded coefficient of variation uniformly, as the probability of the event of interest tends to zero in a suitable large deviations regime. A brief review of importance sampling and the notion of strong efficiency will be given later in this paper; for a more in-depth discussion, see Asmussen and Glynn [1].

The construction of our sampling scheme is driven by recently developed large deviations results in Rhee et al. [27] for regularly varying Lévy processes. Specifically, let $X(t)$, $t \geq 0$, be a one-dimensional compensated compound Poisson process with unit arrival rate and a positive jump distribution W that is regularly varying at infinity (see Definition 1). Define $\bar{X}_n = \{\bar{X}_n(t)\}_{t\in[0,1]}$ with $\bar{X}_n(t) = X(nt)/n$. For a measurable set $A \subseteq \mathbb{D}$ satisfying a specific topological property, the large deviations results derived in Rhee et al. [27] establish that $\mathbb{P}(\bar{X}_n \in A) = \Theta((n\mathbb{P}(W \geq n))^{l^*})$, where precise details can be found in Section 2. In practice, exact estimates are often demanded. Hence, we design a sampling scheme for rare events that take the form $\mathbb{P}(\bar{X}_n \in A)$. We illustrate our approach with several applications that arise in mathematical finance, actuarial science, and queueing theory.

To contextualize our contribution, let us provide a review of the theory and methods that are standard in rare-event simulation settings similar to those studied in this paper.

In the context of stochastic processes with light-tailed characteristics, such as random walks with increments possessing a finite moment generating function in a neighborhood of the origin, large deviations theory can be used to bias the process of interest in order to induce the occurrence of the rare event in question. In fact, it is well known that for a conventional type of proof of the asymptotic lower bound in large deviations analysis, one can extract an exponential change of measure that can sometimes be proved to be efficient (for counterexamples, see, e.g., Glasserman and Kou [21] and Glasserman and Wang [22]). By connecting the design of efficient importance sampling estimators with a game-theoretic formulation, Dupuis et al. [16] and Dupuis and Wang [12, 13] provide the foundations for the use of large deviations theory in the construction and analysis of provably efficient rare-event simulation estimators. Moreover, a weakly efficient "universal" sampler has been proposed by Dupuis and Wang [14] for a general class of hitting sets in arbitrary Jackson network topologies. Examples of additional recent papers are Boxma et al. [10] and Torrisi [31].

The setting of stochastic processes with heavy-tailed increments brings up additional challenges compared with its light-tailed counterpart discussed in the previous paragraph. These challenges were exposed in Asmussen et al. [2]. First of all, typically, the asymptotic conditional distribution of any particular increment given the rare event of interest converges to the underlying nominal distribution. Intuitively, if a rare event is caused by a large jump that may occur in a single "unlucky" increment out of many possible alternatives, then the chance that any specific increment is precisely the unlucky one is, naturally, small. So any particular increment is likely to behave "normally," and therefore, in contrast to the light-tailed setting, there is no direct way in which one might attempt to bias a particular increment in order to stir the process toward the rare event of interest.

Moreover, as pointed out in Asmussen et al. [2], the asymptotic description of the most likely way in which a rare event may occur, for example, as a result of the presence of a single large jump, does not lead to a valid change of measure for importance sampling because it is possible that several large jumps (or no large jump at all) might actually produce the event of interest under the nominal dynamics of the system. In other words, the natural biasing mechanism induced by directly approximating the zero-variance importance sampling distribution in the heavy-tailed setting assigns zero probability to events that are possible under the nominal dynamics leading to an ill-defined likelihood ratio.

The use of state-dependent importance sampling provides a way to deal with these difficulties. In Blanchet and Glynn [4], the authors explain how approximating Doob's h-transform can lead to a feasible change of measure that produces a strongly efficient importance sampling estimator in the setting of first-passage-time probabilities for one-dimensional random walks. A Lyapunov technique was introduced for the analysis of state-dependent importance sampling estimators. But the direct approximation of Doob's h-transform might be difficult to implement in higher dimensions because of both sampling implementation challenges and the evaluation of normalizing constants.

In the setting of one-dimensional compound sums of independent and identically distributed (i.i.d.) regularly varying random variables, Dupuis et al. [15] produced a state-dependent change of measure whose normalizing constant is straightforward to implement. Their idea can be described as follows: each increment is sampled by either the original measure or—with small probability, which is a design parameter—a different measure, which is essentially the original measure conditional on exhibiting a large jump. The advantage of the mixture samplers is that implementation challenges and the evaluation of normalizing constants can often be addressed by choosing a suitable set of parameters.

Under the setting where the time horizon is growing in large and moderate deviation schemes, Blanchet and Liu [5] show how to use Lyapunov inequalities to address the parameter selection while enforcing a bounded relative error. A key step in the methodology is the construction of a suitable Lyapunov function (for an example of the technique in multidimensional settings, see Blanchet and Liu [6]). Blanchet and Liu [5] suggest using the type of fluid analysis that is prevalent in the large deviations literature of heavy-tailed stochastic processes (see, e.g., Foss and Korshunov [18, 19]). However, the construction of the Lyapunov function and the verification of the Lyapunov inequality becomes highly nontrivial in settings involving multiple jumps and the presence of boundaries that are common in queueing systems; for an example of the types of complications which arise in queueing settings, see Blanchet et al. [7].

The idea of using mixtures (for literature on simulation of heavy-tailed random walks from other perspectives, such as Markov chain Monte Carlo (MCMC) and cross-entropy, see, e.g., Gudmundsson and Hult [23], Botev et al. [9], and the references therein), suggested in Dupuis et al. [15], is also used here. But whereas Dupuis et al. [15] treats a particular one-dimensional setting involving a rare event that is caused by a single big jump during a bounded time horizon, our setting is more general. We allow for a wide range of events, which might be caused by multiple jumps during a growing time horizon in a large deviations scaling.

Recall that we are interested in estimating $\mathbb{P}(\bar{X}_n \in A)$. The concept behind our sampling scheme can be described as follows. On the basis of the large deviations results derived in Rhee et al. [27], we construct first an auxiliary set B^γ that is closely related to the optimization problem given by (1). Then, given a fixed mixture probability parameter $w \in (0,1)$, we generate the sample path of $\bar{X}_n$ under the nominal measure. And with probability $1-w$, we generate the sample path of $\bar{X}_n$ under the measure $\mathbb{Q}^\gamma(\,\cdot\,) \triangleq \mathbb{P}(\cdot \mid \bar{X}_n \in B^\gamma)$. Finally, as a consequence of applying the importance sampling technique, we scale our samplers with a suitable likelihood ratio given as in (5). It should be noted that the set A can be as general as in the setting of Rhee et al. [27]. Therefore, our methodological contribution in this paper addresses precisely those types of difficulties mentioned in the previous paragraphs, such as dealing with events that are caused by multiple jumps, working with time scales of order $\mathcal{O}(n)$, avoiding the evaluation of normalizing constants, and bypassing the verification of Lyapunov inequalities. The advantages of our sampling scheme are that the new estimators are strongly efficient and straightforward to implement. Moreover, they are "universal" in the sense that a single importance sampling scheme applies to a very general class of rare events involving multiple jumps that arise in heavy-tailed systems. As a final remark, it should be mentioned that constructing the auxiliary set B^γ requires choosing a set of suitable parameters γ whose existence is guaranteed by the topological property we impose on A. Hence, one of the main challenges is to select the set of parameters specifically for each application.

Our mathematical contributions in this paper can be summarized as follows.

- We propose a simulation algorithm for estimating the rare-event probability of $\bar{X}_n \in A$, together with a sampling scheme for $\bar{X}_n \in \cdot$ given $\bar{X}_n \in B^\gamma$, which is based on a rejection sampling with an unconditional acceptance probability bounded away from zero as $n \to \infty$.
- By showing the existence of the parameter γ, we prove the strong efficiency of our sampling scheme under a very general setting (see Assumption 2).
- We showcase the versatility of the algorithm by illustrating the implementation of the proposed sampling scheme to the rare events that arise in finance, actuarial science, and queueing theory.
- In the application to queueing networks in particular (see Section 6), we show that the tail index of the rare-event probability—which usually exhibit a complex boundary behavior as a result of the nonlinear nature of the associated Skorokhod mapping—can be determined by solving knapsack problem with nonlinear constraints.

The rest of the paper is organized as follows. Section 2 deals with basic background and notations required to state our contributions. Section 3 introduces our estimators and describes the main result. Applications and numerical implementations are discussed in Sections 4–6. All the proofs of results presented in this paper are given in Section 7.

2. Notations and Preliminaries

This section is split into two parts. The first discusses general notions that will be employed in this paper. The second reviews recently developed results involving large deviations for regularly varying Lévy processes and random walks.

2.1. Notations

We start with a summary of notations that will be employed in this paper. Let $\mathbb{Z}_+$ denote the set of nonnegative integers, and let $\mathbb{R}_+$ denote the set of nonnegative real numbers. Let A° and A^- denote the interior and the closure of A, respectively. Let $(\mathbb{D}_{[0,1],\mathbb{R}}, d)$ be the metric space of real-valued right-continuous with left limits functions on $[0,1]$, denoted by $\mathbb{D} = \mathbb{D}_{[0,1],\mathbb{R}}$, equipped with the Skorokhod J_1 metric on $\mathbb{D}$ that is defined by $d(x,y) = \inf_{\lambda \in \Lambda} \|\lambda - id\|_\infty \vee \|x \circ \lambda - y\|_\infty$, $x, y \in \mathbb{D}$, where id denotes the identity mapping, $\|\cdot\|_\infty$ denotes the uniform metric (i.e., $\|x\|_\infty \triangleq \sup_{t \in [0,1]} |x(t)|$), and Λ denotes the set of all strictly increasing, continuous bijections from $[0,1]$ to itself. Let $\mathbb{D}^k$ denote the k-fold product space of $\mathbb{D}$. Let $\mathbb{D}^k_\uparrow$ denote the subset of functions in $\mathbb{D}^k$ that are nonnegative and nondecreasing in each coordinate. When it comes to the tail indices of a regularly varying distribution, we use β (or β_i in the multidimensional case) for the right tail and α for the left tail. Let $\mathbb{D}_l$ denote the subspace of $\mathbb{D}$ consisting of nondecreasing step functions vanishing at time 0 with l jumps, and let $\mathbb{D}_{<l^*}$ denote the subspace of $\mathbb{D}$ consisting of nondecreasing step functions vanishing at 0 with at most $l-1$ jumps (i.e., $\mathbb{D}_{<l^*} = \bigcup_{l \le l^*-1} \mathbb{D}_l$). Define $\mathbb{D}_{<(l_1^*,\ldots,l_d^*)} \triangleq \bigcup_{(l_1,\ldots,l_d) \in I_{<(l_1^*,\ldots,l_d^*)}} \prod_{i=1}^d \mathbb{D}_{l_i}$, where $I_{<(l_1^*,\ldots,l_d^*)} \triangleq \{(l_1,\ldots,l_d) \in \mathbb{Z}_+^d \setminus \{(l_1^*,\ldots,l_d^*)\} \mid$

$\mathcal{I}(l_1,\ldots,l_d)\le\mathcal{I}(l_1^*,\ldots,l_d^*)\}$, and $\mathcal{I}(l_1,\ldots,l_d)\triangleq(\beta_1-1)l_1+\ldots+(\beta_d-1)l_d$. Define a partial order $\prec$ on $\mathbb{Z}_+^d$ such that $(l_1,\ldots,l_d)\prec(m_1,\ldots,m_d)$ if and only if $\mathbb{C}_{(l_1,\ldots,l_d)}\subsetneq\mathbb{C}_{(m_1,\ldots,m_d)}$, where $\mathbb{C}_{(l_1,\ldots,l_d)}\triangleq\bigcup_{i=1}^d\mathbb{D}^{i-1}\times\mathbb{D}_{<l_i}\times\mathbb{D}^{d-i}$. Define $J_{(j_1,\ldots,j_d)}\triangleq\{(l_1,\ldots,l_d)\in\mathbb{Z}_+^d\setminus I_{<(j_1,\ldots,j_d)}\,|\,(m_1,\ldots,m_d)\prec(l_1,\ldots,l_d)$ implies $(m_1,\ldots,m_d)\in I_{<(j_1,\ldots,j_d)}\}$. To get familiar with the notation, an illustration of $I_{<(l_1*,\ldots,l_d*)}$, $J_{(l_1*,\ldots,l_d*)}$, and the partial order $\prec$ is given in Figure 1. Let $\mathbb{D}_{l_-;l_+}$ denote the subspace of the Skorokhod space consisting of step functions vanishing at the origin with exactly l_- downward jumps and l_+ upward jumps, and define

$$\mathbb{D}_{<l_-^*;l_+^*}\triangleq\bigcup_{(l_-,l_+)\in I_{<l_-^*;l_+^*}}\mathbb{D}_{l_-;l_+},$$

where $I_{<l_-^*;l_+^*}\triangleq\{(l_-,l_+)\in\mathbb{Z}_+^2\setminus\{(l_-^*,l_+^*)\}\,|\,(\alpha-1)l_-+(\beta-1)l_+\le(\alpha-1)l_-^*+(\beta-1)l_+^*\}$.

Given nonnegative sequences of real numbers x_n and y_n, we write $x_n=\mathcal{O}(y_n)$, $x_n=o(y_n)$ and $x_n=\Theta(y_n)$ if $\limsup_{n\to\infty}x_n/y_n<\infty$, $\lim_{n\to\infty}x_n/y_n=0$, and $0<\liminf_{n\to\infty}x_n/y_n\le\limsup_{n\to\infty}x_n/y_n<\infty$, respectively. Given two $\mathbb{R}$-valued functions f and g, we write $f\propto g$ if there exists $c\in\mathbb{R}$ such that $f=cg$. For $x=(x_1,\ldots,x_k)$, $y=(y_1,\ldots,y_k)\in\mathbb{R}^k$, we write $x\le y$ if $x_i\le y_i$ for all $i\in\{1,\ldots,k\}$. Let the cardinality of S be denoted by $|S|$ or $\#S$. Finally, let $\mathscr{C}(S,k)$ and $\mathscr{P}(S,k)$ denote the set of all k-combinations and k-permutations of a set S, respectively. Note that $|\mathscr{C}(S,k)|=\binom{|S|}{k}$ and $|\mathscr{P}(S,k)|=|\mathscr{C}(S,k)|\cdot k!$.

To describe the efficiency of a rare-event simulation algorithm, we adopt a widely applied criterion, which requires that the relative mean squared error of the associated estimator is appropriately controlled. To be more precise, suppose that we are interested in a sequence of rare events A_n, which becomes more and more rare as $n\to\infty$. For each $n\in\mathbb{Z}_+$, let L_n be an unbiased estimator of the rare-event probability $\eta_n=\mathbb{P}(A_n)$. Note that L_n is said to be strongly efficient if $\mathbb{E}L_n^2=\mathcal{O}(\eta_n^2)$. In particular, strong efficiency implies that the number of simulation runs required to estimate the target probability to a given relative accuracy is bounded with respect to (w.r.t.) n.

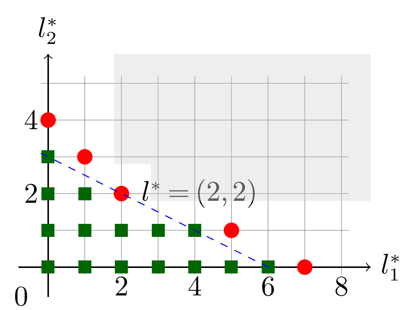
2.2. Preliminaries

As we will see, the simulation algorithm that we propose in this paper is constructed based on the asymptotic behavior of rare-event probabilities; therefore we review some recently developed large deviations results for scaled Lévy processes with heavy-tailed Lévy measures, introduced in Rhee et al. [27]. To begin with, we give the definition of regularly varying functions/random variables.

Definition 1. A positive function f is called regularly varying (at infinity) with index β if $f(x)=x^\beta L(x)$, where L is a slowly varying function (i.e., $\lim_{x\to\infty}L(cx)/L(x)=1$ for all $c>0$). Moreover, a random variable X is said to be regularly varying at infinity and minus infinity with index β if $\mathbb{P}(X\ge x)$ and $\mathbb{P}(X\le -x)$ are regularly varying with index β, respectively.

Now, let X be a Lévy process with Lévy measure ν, where ν is spectrally positive and regularly varying (at infinity) with index $-\beta<-1$. Let $\bar{X}_n\triangleq\{X(nt)/n\}_{t\in[0,1]}$ denote the associated scaled process. Let ν_β^l denote the restriction of l-fold product measure of ν_β to $\{x\in\mathbb{R}_+^l: x_1\ge x_2\ge\ldots\ge x_l\}$, where $\nu_\beta(x,\infty)\triangleq x^{-\beta}$. For $l\ge1$, define a (Borel) measure $C_l(\,\cdot\,)\triangleq\mathbb{E}[\nu_\beta^l\{y\in(0,\infty)^l\mid\sum_{i=1}^l y_i\mathbb{1}_{[U_i,1]}\in\cdot\}]$, where U_i, $i\ge1$ are i.i.d. uniformly distributed on $[0,1]$. Note that C_l is concentrated on $\mathbb{D}_l$ (i.e., $C_l(\mathbb{D}_l)=1$). Moreover, we make the convention that C_0 is the Dirac measure concentrated on the zero function. The following result is useful in designing an efficient algorithm for rare events involving one-dimensional scaled processes. Throughout the rest of this paper, all measurable sets are understood to be Borel measurable.

Figure 1. (Color online) An example of important notations introduced in Section 2.1. For $(\beta_1-1)/(\beta_2-1)=2$ and $(l_1^*,l_2^*)=(2,2)$, we mark the elements in $I_{<(l_1*,l_2*)}$ and $J_{(l_1^*,l_2^*)}$ with squares and circles, respectively. Moreover, the shaded area contains all those points (l_1,l_2) such that $(l_1^*,l_2^*)\prec(l_1,l_2)$.



Result 1 (Theorem 3.1 of Rhee et al. [27]). *Suppose that A is a measurable set. If A is bounded away from $\mathbb{D}_{<l^*}$—that is, $d(A, \mathbb{D}_{<l^*}) > 0$, where $l^* \triangleq \min\{l \in \mathbb{Z}_+ \mid \mathbb{D}_l \cap A \neq \emptyset\} < \infty$—then we have that*

$$C_{l^*}(A^\circ) \le \liminf_{n\to\infty} \frac{\mathbb{P}(\bar{X}_n \in A)}{(n\nu[n,\infty))^{l^*}} \le \limsup_{n\to\infty} \frac{\mathbb{P}(\bar{X}_n \in A)}{(n\nu[n,\infty))^{l^*}} \le C_{l^*}(A^-).$$

As one can see in Section 6, some applications can be interpreted as sample-path rare events in a multidimensional setting. Therefore, it is particularly interesting to consider large deviations results for multidimensional processes. Let $X^{(1)}, \ldots, X^{(d)}$ be independent centered one-dimensional Lévy processes with spectrally positive Lévy measures $\nu_1(\cdot), \ldots, \nu_d(\cdot)$, respectively, where each ν_i is regularly varying with index $-\beta_i < -1$ at infinity. Moreover, for the finite product of metric spaces, we use the maximum metric; that is, we use $d_{\mathbb{S}_1\times\cdots\times\mathbb{S}_d}((x_1,\ldots,x_d),(y_1,\ldots,y_d)) \triangleq \max_{i=1,\ldots,d} d_{\mathbb{S}_i}(x_i,y_i)$ for the product $\mathbb{S}_1\times\cdots\times\mathbb{S}_d$ of metric spaces $(\mathbb{S}_i, d_{\mathbb{S}_i})$. Finally, for $(l_1,\ldots,l_d)\in\mathbb{Z}_+^d$, we define $C_{l_1}\times\cdots\times C_{l_d}(\cdot)$ (which is concentrated on $\prod_{i=1}^d \mathbb{D}_{l_i}$) as the product measure of $C_{l_i}(\cdot) \triangleq \mathbb{E}[\nu_{\beta_i}^{l_i}\{y \in (0,\infty)^{l_i} \mid \sum_{j=1}^{l_i} y_j \mathbb{1}_{[U_j,1]} \in \cdot\}]$. Result 2 states a large deviations result for d-dimensional process $\bar{X}_n(t) \triangleq (X^{(1)}(nt)/n, \ldots, X^{(d)}(nt)/n)$ for $t\in[0,1]$.

Result 2 (Theorem 3.6 of Rhee et al. [27]). *Suppose that A is measurable. If A is bounded away from $\mathbb{D}_{<(l_1^*,\ldots,l_d^*)}$, where*

$$(l_1^*,\ldots,l_d^*) = \operatorname*{arg\,min}_{(l_1,\ldots,l_d)\in\mathbb{Z}_+^d,\ \prod_{i=1}^d \mathbb{D}_{l_i}\cap A\neq\emptyset} \mathcal{I}(l_1,\ldots,l_d) \tag{1}$$

and $\mathcal{I}(l_1,\ldots,l_d) = (\beta_1-1)l_1 + \ldots + (\beta_d-1)l_d$, then we have that

$$C_{l_1^*}\times\cdots\times C_{l_d^*}(A^\circ) \le \liminf_{n\to\infty} \frac{\mathbb{P}(\bar{X}_n\in A)}{\prod_{i=1}^d (n\nu_i[n,\infty))^{l_i^*}} \le \limsup_{n\to\infty} \frac{\mathbb{P}(\bar{X}_n\in A)}{\prod_{i=1}^d (n\nu_i[n,\infty))^{l_i^*}} \le C_{l_1^*}\times\cdots\times C_{l_d^*}(A^-).$$

Note that the assumption that A is bounded away from $\mathbb{D}_{<(l_1^*,\ldots,l_d^*)}$ guarantees the uniqueness of $(l_1^*,\ldots,l_d^*)$. Finally, we conclude this section with an extension of Result 2, which will be useful in constructing an efficient simulation algorithm for heavy-tailed random walks. Let S_k, $k\ge 0$, be a random walk; set $\bar{S}_n(t) = S_{\lfloor nt\rfloor}/n$, $t\ge 0$; and define $\bar{S}_n = \{\bar{S}_n(t), t\in[0,1]\}$. Let ν_β^l be as defined above. Similarly, let ν_α^m denote the restriction of m-fold product measure of ν_α to $\{x\in\mathbb{R}_+^m : x_1\ge x_2\ge\ldots\ge x_m\}$, where $\nu_\alpha(x,\infty)\triangleq x^{-\alpha}$. Let $C_{0,0}(\cdot)\triangleq\delta_{\mathbf{0}}(\cdot)$ be the Dirac measure concentrated on the zero function. For each $(l_-,l_+)\in\mathbb{Z}_+^2\backslash\{(0,0)\}$, define a measure (which is concentrated on $\mathbb{D}_{l_-;l_+}$) $C_{l_-;l_+}(\cdot)\triangleq\mathbb{E}[\nu_\alpha^{l_-}\times\nu_\beta^{l_+}\{(x,y)\in (0,\infty)^{l_-}\times(0,\infty)^{l_+} : \sum_{i=1}^{l_+} y_i\mathbb{1}_{[V_i,1]} - \sum_{i=1}^{l_-} x_i\mathbb{1}_{[U_i,1]}\in\cdot\}]$, where U_i's and V_i's are i.i.d. uniform on $[0,1]$.

Result 3. *Suppose that $\mathbb{P}(S_1\le -x)$ is regularly varying with index $-\alpha$ and $\mathbb{P}(S_1\ge x)$ is regularly varying with index $-\beta$. Let A be a measurable set bounded away from $\mathbb{D}_{<l_-^*;l_+^*}$, where*

$$(l_-^*, l_+^*) = \operatorname*{arg\,min}_{(l_-,l_+)\in\mathbb{Z}_+^2,\ \mathbb{D}_{l_-;l_+}\cap A\neq\emptyset} (\alpha-1)l_- + (\beta-1)l_+. \tag{2}$$

Then,

$$\begin{aligned} C_{l_-^*;l_+^*}(A^\circ) &\le \liminf_{n\to\infty} \frac{\mathbb{P}(\bar{S}_n\in A)}{(n\mathbb{P}(S_1\le -n))^{l_-^*}(n\mathbb{P}(S_1\ge n)))^{l_+^*}} \\ &\le \limsup_{n\to\infty} \frac{\mathbb{P}(\bar{S}_n\in A)}{(n\mathbb{P}(S_1\le -n))^{l_-^*}(n\mathbb{P}(S_1\ge n)))^{l_+^*}} \le C_{l_-^*;l_+^*}(A^-). \end{aligned}$$

3. Main Results

In this section, we present our main results. Although the large deviations results reviewed in Section 2 are stated for Lévy processes, we focus on compensated compound Poisson process for simulation purposes. Let X denote a d-dimensional compensated compound Poisson process, and recall that $\bar{X}_n$ is the scaled process with $\bar{X}_n(t) = X(nt)/n$, $t\in[0,1]$. For a measurable set $A\in\mathbb{D}^d$, we are interested in estimating the probability of the event $A_n\triangleq\{\bar{X}_n\in A\}$, when n is large. Note that, in view of the law of large numbers, one can expect that $\mathbb{P}(\bar{X}_n\in A)\to 0$ for A's that are bounded away from the zero function, and hence, A_n's are rare events for large n's. In Section 3.1, we first illustrate the idea of our algorithm in the special case for $d=1$, where the notations are simpler. In Section 3.2, we extend this result to general d.

3.1. The One-Dimensional Case

Let $\{X(t)\}_{t\geq 0}$ be a one-dimensional compensated compound Poisson process with i.i.d. jump sizes $\{W(k)\}_{k\geq 1}$. That is, $X(t)=\sum_{k=1}^{N(t)}W(k)-\lambda t\mathbb{E}W(1)$, where $\{N(t)\}_{t\geq 0}$ is a Poisson process with arrival rate λ, and let $\bar{X}_n\triangleq\{X(nt)/n\}_{t\in[0,1]}$ denote the associated scaled process. Moreover, let $\mathbb{P}(W(1)>x)$ be regularly varying of index $-\beta<-1$. The following assumption is essential for analyzing the asymptotic behavior of the rare-event probability and, hence, deriving the strong efficiency of our estimator.

Assumption 1. *Let A be a measurable set in $\mathbb{D}$. We assume that A is bounded away from $\mathbb{D}_{<l^*}$, where $l^*=\min\{l\in\mathbb{Z}_+\mid\mathbb{D}_l\cap A\neq\emptyset\}$ denotes the minimal number of upward jumps of a step function in A. Moreover, assume that $C_{l^*}(A^\circ)>0$.*

Remark 1. As one can see in Sections 4–6, one of the typical settings that arises in applications is that the set A can be written as a finite combination of unions and intersections of $F_1^{-1}(A_1),\ldots,F_m^{-1}(A_m)$, where each $F_i:\mathbb{D}\to\mathbb{S}_i$ is a continuous function, and all sets A_i are subsets of a general topological space $\mathbb{S}_i$. If we denote this operation of taking unions and intersections by Ψ (i.e., $A=\Psi(F_1^{-1}(A_1),\ldots,F_m^{-1}(A_m))$), then it holds that $\Psi(F_1^{-1}(A_1^\circ),\ldots,F_m^{-1}(A_m^\circ))\subseteq A^\circ\subseteq A\subseteq A^-\subseteq\Psi(F_1^{-1}(A_1^-),\ldots,F_m^{-1}(A_m^-))$. Hence, $C_{l^*}(A^\circ)>0$ holds if $\hat{T}_{l^*}^{-1}(\Psi(F_1^{-1}(A_1^\circ),\ldots,F_m^{-1}(A_m^\circ)))$ has a positive Lebesque measure, where $\hat{T}_j:\hat{S}_j\to\mathbb{D}_j$ is defined by $\hat{T}_j(x,u)\triangleq\sum_{i=1}^j x_i\mathbb{1}_{[u_i,1]}$ for $j\in\mathbb{Z}_+$, and $\hat{S}_j\triangleq\{(x,u)\in\mathbb{R}_+^j\times[0,1]^j\mid x_1\geq\cdots\geq x_j,0,1,u_1,\ldots,u_j\text{ are distinct}\}$. Analogously, one can derive a sufficient condition for $C_{l_1^*}\times\cdots\times C_{l_d^*}(A^\circ)>0$ (see Assumption 2).

Remark 2. There are several examples that satisfy Assumption 1. For instance, considering $A=\{\xi\in\mathbb{D}_{[0,1]}:\xi(1)\geq a\}$ corresponds to estimating the rare-event probability $\mathbb{P}(X(n)\geq an)$. Another application that fits into this framework is the ruin probability of an insurance company, where reinsurance policy is taken into account. For details of this application, we refer to Section 4. Finally, for one of many examples of A in the multidimensional setting, we refer to Section 6, where the workload in a queueing network is considered.

We design a simulation algorithm that estimates the probability of $A_n\triangleq\{\bar{X}_n\in A\}$ efficiently, based on an importance sampling strategy. To construct an importance distribution, we introduce a constant $\gamma>0$ and define $B_n^\gamma\triangleq\{\bar{X}_n\in B^\gamma\}$, where $B^\gamma\triangleq\{\xi\mid\#\{t\mid\xi(t)-\xi(t^-)>\gamma\}\geq l^*\}$. In the construction of our rare-event simulation algorithm, we will take advantage of the fact that one can always choose γ so that B_n^γ is sufficiently "close" to A_n. The specific choice of γ will be further discussed later in Sections 4–6 for concrete examples. Let $\mathbb{Q}_\gamma(\,\cdot\,)\triangleq\mathbb{P}(\,\cdot\mid B_n^\gamma)$ denote the conditional distribution given $\bar{X}_n\in B^\gamma$. One should notice that $d\mathbb{Q}_\gamma/d\mathbb{P}=\mathbb{P}(B_n^\gamma)^{-1}\mathbb{1}_{B_n^\gamma}$. Moreover, by the Fubini–Tonelli theorem, a closed-form expression for $\mathbb{P}(B_n^\gamma)$ is given by

$$\mathbb{P}(B_n^\gamma)=1-\exp\Big\{-\lambda n\mathbb{P}(W(1)>n\gamma)\Big\}\sum_{j=0}^{l^*-1}\frac{(\lambda n)^j}{j!}\mathbb{P}(W(1)>n\gamma)^j. \tag{3}$$

From (3), one should recognize that B_n^γ can be interpreted as the event of a Poisson distributed random variable with rate $\lambda n\mathbb{P}(W(1)>\gamma n)$ crossing the level l^*. Now, let $w\in(0,1)$ be arbitrary but fixed. We propose an importance distribution $\mathbb{Q}_{\gamma,w}$ that is absolutely continuous w.r.t. $\mathbb{P}$ and is given by

$$\mathbb{Q}_{\gamma,w}(\,\cdot\,)\triangleq w\mathbb{P}(\,\cdot\,)+(1-w)\mathbb{Q}_\gamma(\,\cdot\,). \tag{4}$$

We give here an algorithm for generating the sample path of $\bar{X}_n$ under the probability measure $\mathbb{Q}_\gamma(\,\cdot\,)$. Because $\{\bar{X}_n\in B^\gamma\}\subseteq\{N(n)\geq l^*\}$, we observe that $\mathbb{Q}_\gamma(\bar{X}_n\in\,\cdot\,)=\mathbb{P}(B_n^\gamma)^{-1}\mathbb{P}(\bar{X}_n\in\,\cdot\,,B_n^\gamma)=\sum_{m=l^*}^\infty h_m\,\mathbb{P}(\bar{X}_n\in\,\cdot\mid B_n^\gamma,N(n)=m)$, where $h_m=h_m(n)\triangleq\mathbb{P}(B_n^\gamma,N(n)=m)/\mathbb{P}(B_n^\gamma)$ satisfies $\sum_{m\geq l^*}h_m=1$. Hence, it remains to discuss sampling from $\mathbb{P}(\bar{X}_n\in\,\cdot\mid B_n^\gamma,N(n)=m)$. It turns out that we can proceed a rejection sampling, where drawing from the proposal distribution can be achieved as follows: first, sample $\{b_k\}_{k\leq l^*}$ uniformly from $\mathscr{C}(\{1,\ldots,m\},l^*)$; then, sample each $W(b_k)$, $k\leq l^*$, conditional on $W(b_k)>n\gamma$; and finally, sample $W(m')$, $m'\leq m$, $m'\notin\{b_k\}_{k\leq l^*}$, under the nominal measure. Note that the target density $f_{\text{target};m}$, defined by

$$f_{\text{target};m}(w_1,\ldots,w_m)dw_1\cdots dw_m\triangleq\mathbb{P}(W(1)\in w_1+dw_1,\ldots,W(m)\in w_m+dw_m\mid B_n^\gamma,N(n)=m),$$

can be bounded by $M_m f_{\text{proposal};m}(w_1,\ldots,w_m)$, where

$$f_{\text{target};m}(w_1,\ldots,w_m)\propto\frac{1}{\mathbb{P}(B_n^\gamma\mid N(n)=m)}\prod_{j=1}^m\frac{d}{dw_j}\mathbb{P}(W(j)\leq w_j)\mathbb{1}_{B_n^\gamma}(w_1,\ldots,w_m),$$

$$f_{\text{proposal};m}(w_1,\ldots,w_m)=\frac{1}{\binom{m}{l^*}\mathbb{P}(W(1)>n\gamma)^{l^*}}\prod_{j=1}^m\frac{d}{dw_j}\mathbb{P}(W(j)\leq w_j)\sum_{\substack{(b_1,\ldots,b_{l^*})\in\\ \mathscr{C}(\{1,\ldots,m\},l^*)}}\mathbb{1}_{\{W(b_k)>n\gamma,\forall k\leq l^*\}},$$

and hence, $M_m = M_m(n) \triangleq \binom{m}{l^*} \mathbb{P}(W(1) > n\gamma)^{l^*} \mathbb{P}(B_n^\gamma \,|\, N(n) = m)^{-1}$. Now, it is natural to accept $(W(1), \ldots, W(m))$ with probability $a(W(1), \ldots, W(m)) = \binom{\#\{i \in \{1,\ldots,m\} \,|\, W(i) > n\gamma\}}{l^*}^{-1}$. Finally, we are able to formulate the pseudocode for generating $\bar{X}_n$ under $\mathbb{Q}_\gamma$ in Algorithm 1. Moreover, we show in Proposition 1 that the expected running time of Algorithm 1 is uniformly bounded from above w.r.t. n.

Proposition 1. *Let $T_{\text{alg1}}(n)$ denote the expected running time of Algorithm 1. Under the condition that $W(1)$ is regularly varying of index $-\beta < -1$, we have that $T_{\text{alg1}}(n) = \sum_{m \geq l^*} h_m(n) M_m(n)$ is uniformly bounded from above w.r.t. n (i.e., $\max_{n \geq 0} T_{\text{alg1}}(n) < \infty$).*

In view of the observations we made so far, we propose an estimator Z_n for $\mathbb{P}(A_n)$ that is given by

$$Z_n = \mathbb{1}_{A_n} \frac{d\mathbb{P}}{d\mathbb{Q}_{\gamma,w}} = \frac{\mathbb{1}_{A_n}}{w + \frac{1-w}{\mathbb{P}(B_n^\gamma)} \mathbb{1}_{B_n^\gamma}}. \tag{5}$$

Intuitively, an importance sampling technique is used to get more samples from the interesting region by sampling from a distribution that overweights the important region. Based on this, the choice of B_n^γ can be "justified" because B_n^γ is mimicking the asymptotic behavior of the probability of interest. However, as one can see in the proof of strong efficiency (see Theorem 2), we should analyze the second moment of our estimator to avoid "backfire," yielding an estimator with larger or even infinite variance. It turns out that this intuition can be made rigorous by applying Result 1. We end this section with a theorem regarding to the strong efficiency of our estimator.

Algorithm 1 (Generating the Sample Path of $\bar{X}_n$ Under $\mathbb{Q}_\gamma$)

1: sample $m \sim h_m$ ▷ $m = m'$ with probability $h_{m'} = \mathbb{P}(N(n) = m' \,|\, B_n^\gamma)$
2: $R \leftarrow$ **true**
3: **while** $R =$ **true do**
4: sample $\{b_k\}_{k \leq l^*} \sim \text{unif}(\mathscr{C}(\{1, \ldots, m\}, k))$ ▷ uniform distribution on $\mathscr{C}(\{1, \ldots, m\}, k)$
5: **for** $i \in \{b_k\}_{k \leq l^*}$ **do**
6: sample $W(i) \sim W(1) \,|\, W(1) > n\gamma$
7: **for** $i \notin \{b_k\}_{k \leq l^*}$ **do**
8: sample $W(i) \sim W(1)$
9: $c \leftarrow \#\{j \in \{1, \ldots, m\} \,|\, W(j) > n\gamma\}$; $a \leftarrow \binom{c}{l^*}^{-1}$; sample $u \sim \text{uniform}[0,1]$; $R \leftarrow$ **true**
10: **if** $u < a$ **then**
11: $R \leftarrow$ **false**
return $\bar{X}_n$

Theorem 1. *Under Assumption 1, there exists a $\gamma > 0$ such that the estimator constructed in (5) is strongly efficient for estimating $\mathbb{P}(A_n)$.*

Remark 3. Note that, under Assumption 1, there exists $r > 0$ such that $d(A, \mathbb{D}_{<l^*}) > r$. One sufficient way to make Z_n in (5) strongly efficient is to choose γ such that $\mathbb{Z} \not\ni \lceil r/\gamma \rceil \geq l^* + 1$. Sometimes, finding r can be application specific, though generally r is the smallest size a big jump needs to take to make the rare event happen, and physical intuition—which can be obtained from solving the large-deviations problem—is helpful in making an educational guess on r. For more details about finding r as well as choosing γ we refer to Sections 4–6.

3.2. Extension to General *d*

In this section, we extend the results in Section 3.1 to the case with general d. To be precise, let $X \triangleq (X^{(1)}, \ldots, X^{(d)})$ be a superposition of d independent compensated compound Poisson processes with upward jumps, where $\{N^{(i)}(t)\}$ is a Poisson process with arrival rate λ_i, and $X^{(i)}(t) = \sum_{k=1}^{N^{(i)}(t)} W^{(i)}(k) - \lambda_i t \mathbb{E} W^{(i)}(1)$. Moreover, let $\mathbb{P}(X^{(i)}(1) > x)$ be regularly varying of index $-\beta_i < -1$ at infinity. Finally, let $\bar{X}_n$ denote the corresponding scaled process. As we can see in Result 2, the large deviations results for $\mathbb{P}(\bar{X}_n \in A)$ depend heavily on the value of $\mathcal{I}(l_1^*, \ldots, l_d^*)$, where $(l_1^*, \ldots, l_d^*)$ is as defined in (1). However, for $c \in \mathbb{R}$, the grid $(l_1, \ldots, l_d) \in \mathbb{Z}_+^d$ satisfying $\mathcal{I}(l_1, \ldots, l_d) = c$ is not unique, in general. Therefore, assuming A being bounded away from $\prod_{i=1}^d \mathbb{D}_{<l_i}$ is not sufficient for our purposes. The following assumption, which is slightly different from Assumption 1, corresponds to the extension of Result 1 to Result 2.

Assumption 2. *Let A be a measurable set. Assume that A is bounded away from $\mathbb{D}_{<(l_1^*,\ldots,l_d^*)}$, where $(l_1^*,\ldots,l_d^*)$ is the unique solution of the minimization problem given by (1). Moreover, assume that $C_{l_1^*}\times\cdots\times C_{l_d^*}(A^\circ)>0$.*

If the solution to (1) is not unique, we may partition A. As in Section 3.1, we focus now on constructing the auxiliary set B^γ for the importance distribution. Define $A_n \triangleq \{\bar{X}_n\in A\}$ and $B_n^\gamma \triangleq \{\bar{X}_n\in B^\gamma\}$. As one can see in the proof of Theorem 2, the ability to control the probability of $A_n\cap(B_n^\gamma)^c$ should be taken into account in choosing the auxiliary set B^γ. In the one-dimensional case, letting B^γ mimic the optimal path leading to the rare event makes us capable of controlling the relative error of our estimator. By "mimicking the optimal path," we mean that the minimal number of jumps l^* that are needed for $\mathbb{D}_{l^*}\cap A\neq\emptyset$ is used as parameter in the construction of B^γ. However, the same strategy would fail in the multidimensional case, because the rare event can be reached through other feasible but not necessarily optimal paths. Thus, we require a more complicated construction of B^γ.

Definition 2. Let A be a measurable set in $\mathbb{D}^d$, and let $(l_1^*,\ldots,l_d^*)$ denote the unique solution to (1). Let $\gamma\in\mathbb{R}^d$ with $\gamma_i>0$ for all $i\in\{1,\ldots,d\}$, and define

$$B^\gamma \triangleq \bigcup_{(l_1,\ldots,l_d)\in J_{(l_1^*,\ldots,l_d^*)}} B^{\gamma;l}, \tag{6}$$

where $B^{\gamma;l}$ is the set of càdlàg functions on $\mathbb{R}^d$ that have at least l_i number of jumps with sizes larger than γ_i in its ith coordinate; that is, $B^{\gamma;l}\triangleq\{(\xi^{(1)},\ldots,\xi^{(d)})\in\mathbb{D}^d \mid \#\{t\mid \xi^{(i)}(t)-\xi^{(i)}(t^-)>\gamma_i\}\geq l_i,\ \forall i\in\{1,\ldots,d\}\}$.

Remark 4. Note that the cardinality of $J_{(l_1^*,\ldots,l_d^*)}$ is finite. To design a strongly efficient simulation algorithm for estimating $\mathbb{P}(A_n)$, we will take advantages of an important property of $J_{(l_1^*,\ldots,l_d^*)}$. That is, for all $\xi\in A$ with A being bounded away from $\mathbb{D}_{<(l_1^*,\ldots,l_d^*)}$, there exists $(l_1,\ldots,l_d)\in J_{(l_1^*,\ldots,l_d^*)}$ such that the path of ξ in its ith coordinate is bounded away from $\mathbb{D}_{<l_i^*}$ for every $i\in\{1,\ldots,d\}$.

Let $\mathbb{Q}_\gamma(\,\cdot\,)\triangleq\mathbb{P}(\,\cdot\mid B_n^\gamma)$ and let $\mathbb{Q}_{\gamma,w}$ be as defined in (4); following the same strategy as in Section 3.1, we propose an estimator that takes the same form as in (5). Before turning to the efficiency analysis of our estimator, we summarize the findings above in Algorithm 2.

Algorithm 2 (Efficient Sampling of $\mathbb{P}(\bar{X}_n\in A)$)
1: sample $u\sim$ uniform$[0,1]$ $\quad\triangleright$ uniform distribution on $[0,1]$
2: sample $\bar{X}_n\sim\mathbb{P}(\bar{X}_n\in\cdot\mid\bar{X}_n\in B^\gamma)$
3: **if** $u<w$ **then**
4: $\quad$ sample $\bar{X}_n\sim\mathbb{P}(\bar{X}_n\in\cdot)$
5: **if** $\bar{X}_n\in A$ **then**
6: $\quad L\leftarrow[w+(1-w)\mathbb{I}_{B_n^\gamma}/\mathbb{P}(\mathrm{B}_\mathrm{n}^\gamma)]^{-1}$
7: **else**
8: $\quad L\leftarrow 0$
$\quad$ **return** L

To complete our algorithm, we need to discuss the computation of $\mathbb{P}(B_n^\gamma)$ as well as the strategy of sampling from the conditional distribution $\mathbb{Q}_\gamma(\,\cdot\,)$. Because B^γ constructed in Definition 2 is the union of $B^{\gamma;l}$ with $l=(l_1,\ldots,l_d)\in J_{(l_1^*,\ldots,l_d^*)}$, by the inclusion-exclusion principle, it is sufficient to discuss computing the probability of sets of the form $\bigcap_{(l_1,\ldots,l_d)\in I}B^{\gamma;l}$, where I is a finite collection of elements in $\mathbb{Z}_+^d$. It turns out that the probability of such a set can be computed similarly as in Section 3.1. From this observation, we give the following proposition.

Proposition 2. *The probability of B_n^γ is equal to*

$$\sum_{k=1}^{|J_{(l_1^*,\ldots,l_d^*)}|}(-1)^{k-1}\sum_{\substack{|I|=k\\ I\subseteq J_{(l_1^*,\ldots,l_d^*)}}}\prod_{i=1}^{d}\left(1-\exp\left\{-\lambda_i n\mathbb{P}(W^{(i)}(1)>n\gamma_i)\right\}\sum_{j=0}^{\hat{l}_{i;I}-1}\frac{(\lambda_i n)^j}{j!}\mathbb{P}(W^{(i)}(1)>n\gamma_i)^j\right),$$

where $\hat{l}_{i;I}\triangleq\max_{(l_1,\ldots,l_d)\in I}l_i$.

Remark 5. It should be mentioned that the complexity of computing $\mathbb{P}(B_n^\gamma)$ can be reduced rapidly in the case, where, for example, one is able to take a smaller (in the sense of cardinality) set than $J_{(l_1^*,\ldots,l_d^*)}$ (see, e.g., Corollary 1 and Sections 5 and 6).

As in Section 3.1, we now discuss generating the sample path of $\bar{X}_n$ under $\mathbb{Q}_\gamma$ in the next step. To begin with, we need the following lemma, which shows that B^γ can be decomposed into finitely many disjoint sets.

Lemma 1. *Let* $B^{\gamma;l}(i,j) \triangleq \{\xi \in \mathbb{D}^d \mid \#\{t \mid \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \geq (l(j))_i\}$. *Let the elements in* $J_{(l_1^*,\ldots,l_d^*)}$, *denoted by* $l(1),\ldots,l(|J_{(l_1^*,\ldots,l_d^*)}|)$, *be ordered so that* $(l(1))_d \leq \cdots \leq (l(|J_{(l_1^*,\ldots,l_d^*)}|))_d$. *Define*

$$\Delta B^{\gamma;l}(i,j) \triangleq B^{\gamma;l}(i,j) \backslash \left(\bigcup_{m=1}^{j-1} B^{\gamma;l}(i,m) \right), \quad i \in \{1,\ldots,d-1\}. \tag{7}$$

Then we have that

$$B^\gamma = \bigcup_{m_1=1}^{|J_{(l_1^*,\ldots,l_d^*)}|} \bigcup_{m_2=1}^{m_1} \cdots \bigcup_{m_{d-1}=1}^{m_{d-2}} \left(\left(\bigcap_{i=1}^{d-1} \Delta B^{\gamma;l}(i,m_i) \right) \cap B^{\gamma;l}(d,1) \right).$$

Lemma 1 shows that B^γ can be decomposed into finitely many disjoint sets. This implies that

$$\mathbb{Q}_\gamma(\bar{X}_n \in \cdot\,) = \sum_{m_1=1}^{|J_{(l_1^*,\ldots,l_d^*)}|} \sum_{m_2=1}^{m_1} \cdots \sum_{m_{d-1}=1}^{m_{d-2}} h_{1;m_1,\ldots,m_{d-1}} \mathbb{P}(\bar{X}_n \in \cdot \mid \bar{X}_n \in B^\gamma(m_1,\ldots,m_{d-1})),$$

where

$$B^\gamma(m_1,\ldots,m_{d-1}) \triangleq \left(\bigcap_{i=1}^{d-1} \Delta B^{\gamma;l}(i,m_i) \right) \cap B^{\gamma;l}(d,1),$$

and $h_{1;m_1,\ldots,m_{d-1}} \triangleq \mathbb{P}(\bar{X}_n \in B^\gamma(m_1,\ldots,m_{d-1}))\mathbb{P}(\bar{X}_n \in B^\gamma)^{-1}$ satisfying

$$\sum_{m_1=1}^{|J_{(l_1^*,\ldots,l_d^*)}|} \sum_{m_2}^{m_1} \cdots \sum_{m_{d-1}}^{m_{d-2}} h_{1;m_1,\ldots,m_{d-1}} = 1.$$

Hence, it remains to design a sampling scheme for generating the sample path of $\bar{X}_n$ under $\mathbb{P}(\,\cdot \mid \bar{X}_n \in B^\gamma(m_1,\ldots,m_{d-1}))$ (for details about generating multidimensional discrete random numbers, see, e.g., Hu and Cui [25]). Because of the special structure of $B^\gamma(m_1,\ldots,m_{d-1})$, we are able to generate $\bar{X}_n^{(1)},\ldots,\bar{X}_n^{(d)}$ independently under $\mathbb{P}(\,\cdot \mid \bar{X}_n \in B^\gamma(m_1,\ldots,m_{d-1}))$. To see this, first note that sampling $\bar{X}_n^{(d)}$ is trivial because of the discussion in Section 3.1. Define $\check{l}(m_i;i) \triangleq \min_{\xi \in \Delta B^{\gamma;l}(i,m_i)} \#\{t \mid \xi(t) - \xi(t^-) > \gamma_i)\}$ and $\hat{l}(m_i;i) \triangleq \max_{\xi \in \Delta B^{\gamma;l}(i,m_i)} \#\{t \mid \xi(t) - \xi(t^-) > \gamma_i)\}$ for $i \in \{1,\ldots,d-1\}$. By (7), we have that

$$\mathbb{P}(\bar{X}_n^{(i)} \in \cdot \mid \bar{X}_n \in B^\gamma(m_1,\ldots,m_{d-1})) = \sum_{q_i=\check{l}(m_i;i)}^{\infty} h_{2;q_i} \mathbb{P}(\bar{X}_n^{(i)} \in \cdot \mid \Delta B^{\gamma;l}(i,m_i), N^{(i)}(n) = q_i),$$

where $h_{2;q_i} \triangleq \mathbb{P}(\Delta B^{\gamma;l}(i,m_i), N^{(i)}(n) = q_i)/\mathbb{P}(\Delta B^{\gamma;l}(i,m_i))$ satisfies $\sum_{q_i \geq \check{l}(m_i;i)} h_{2;q_i} = 1$. Note that

$$\mathbb{P}(\Delta B^{\gamma;l}(i,m_i), N^{(i)}(n) = q_i) = e^{-\lambda_i n} \frac{(\lambda n)^{q_i}}{q_i!} \left(\sum_{i=\check{l}(m_i;i)}^{\hat{l}(m_i;i) \wedge q_i} \binom{q_i}{i} \mathbb{P}(W^{(i)}(1) > n\gamma)^i \mathbb{P}(W^{(i)}(1) \leq n\gamma)^{q_i - i} \right).$$

Therefore, it suffices to consider sampling $\bar{X}_n^{(i)}$ under $\mathbb{P}(\,\cdot \mid \Delta B^{\gamma;l}(i,m_i), N^{(i)}(n) = q_i)$. Again, we can proceed using a similar approach as in Section 3.1: sample $\{b_k\}_{k \leq l}$ uniformly from $\mathscr{C}(\{1,\ldots,q_i\}, \check{l}(m_i;i))$; sample each $W^{(i)}(b_k)$, $k \leq q_i$, conditional on $W^{(i)}(b_k) > n\gamma_i$; sample $W^{(i)}(q_i')$, $q_i' \leq q_i$, and $q_i' \notin \{b_k\}_{k \leq l^*}$, under the nominal measure; and accept $(W^{(i)}(1),\ldots,W^{(i)}(q_i))$ with probability

$$a(W^{(i)}(1),\ldots,W^{(i)}(q_i)) = \binom{\#\{j \in \{1,\ldots,q_i\} \mid W^{(i)}(j) > n\gamma_i\}}{\check{l}(m_i;i)}^{-1} \mathbb{1}_{\{\#\{j \in \{1,\ldots,q_i\} \mid W^{(i)}(j) > n\gamma_i\} \leq \hat{l}(m_i;i)\}}.$$

Finally, we are able to give the pseudocode of this sampling scheme in Algorithm 3. For its expected running time, an analogous result to Proposition 1 is formulated in Proposition 3.

Proposition 3. *Let $T_{alg3}(n)$ denote the expected running time of Algorithm 3. Under the assumption that $W^{(i)}(1)$ is regularly varying of index $-\beta_i < -1$ for all $i \in \{1, \ldots, d\}$, we have that $T_{alg3}(n)$ is uniformly bounded from above w.r.t. n (i.e., $\max_{n\geq 0} T_{alg3}(n) < \infty$).*

The discussion above shows that sampling from the conditional distribution $\mathbb{Q}_\gamma(\,\cdot\,)$ is tractable. As we mentioned in the introduction, our estimator is straightforward to implement. Moreover, its strong efficiency, which is formulated in Theorem 2, can be proved based on Lemma 2. Moreover, we state in Theorem 2 that our estimator is strongly efficient. Without introducing any new notations, we formulate a corollary to address a special case where it is sufficient to consider a smaller (in the sense of cardinality) set than $J_{(l_1^*,\ldots,l_d^*)}$, as in Definition 2. Note that Corollary 1 can be shown by following similar arguments as in the proofs of Lemma 2 and Theorem 2; thus the proof is omitted.

Algorithm 3 (Generating the Sample Path of $\bar{X}_n^{(1)}, \ldots, \bar{X}_n^{(d)}$ Under $\mathbb{Q}_\gamma$)

sample $(m_1, \ldots, m_{d-1}) \sim h_{1;m_1,\ldots,m_{d-1}}$
for $i = 1$ to d **do**
  sample $q_i \sim h_{2;q_i}$; $R \leftarrow$ **true**
  while $R =$ **true do**
    sample $\{b_k\}_{k\leq \check{l}(m_i;i)} \sim \mathrm{unif}(\mathscr{C}(\{1,\ldots,q_i\}, \check{l}(m_i;i)))$
    for $j \in \{b_k\}_{k\leq \check{l}(m_i;i)}$ **do**
      sample $W^{(i)}(j) \sim W^{(i)}(1) \mid W^{(i)}(1) > n\gamma_i$
    for $j \notin \{b_k\}_{k\leq \check{l}(m_i;i)}$ **do**
      sample $W^{(i)}(j) \sim W^{(i)}(1)$
    $c \leftarrow \#\{j \in \{1,\ldots,q_i\} \mid W^{(i)}(j) > n\gamma_i\}$; $a \leftarrow 0$
    if $c < \hat{l}(m_i;i)$ **then**
      $a \leftarrow \binom{c}{\check{l}(m_i;i)}^{-1}$
    sample $u \sim \mathrm{uniform}[0,1]$; $R \leftarrow$ **true**
    if $u < a$ **then**
      $R \leftarrow$ **false**
  return $\bar{X}_n^{(1)}, \ldots, \bar{X}_n^{(d)}$

Theorem 2. *Let $B_n^\gamma \triangleq \{\bar{X}_n \in B^\gamma\}$, where B^γ is as defined in (6). Under Assumption 2, there exists γ such that the estimator given by (5) is strongly efficient for estimating $\mathbb{P}(A_n)$.*

Corollary 1. *Along with Assumption 2, we assume additionally that there exists an index set $I \subseteq J_{(l_1^*,\ldots,l_d^*)}$ and $r > 0$ such that for every $\xi \in A$, there exists $(l_1,\ldots,l_d) \in I$ satisfying $d(\xi, \mathbb{C}_{(l_1,\ldots,l_d)}) \geq r$. Set $\tilde{J}_{(l_1^*,\ldots,l_d^*)} = I \backslash \Delta I$, where $(l_1,\ldots,l_d) \in \Delta I$ if and only if*

- *$(l_1,\ldots,l_d) \in I$ satisfies that $\mathcal{I}(l_1,\ldots,l_d) > 2\mathcal{I}(l_1^*,\ldots,l_d^*)$, and*
- *for every $(l'_1,\ldots,l'_d) \in I \backslash \{(l_1,\ldots,l_d)\}$, we have that $\mathcal{I}(l_1,\ldots,l_d) \neq \mathcal{I}(l'_1,\ldots,l'_d)$.*

Setting $B_n^\gamma = \{\bar{X}_n \in B^\gamma\}$ with $B^\gamma \triangleq \cup_{(l_1,\ldots,l_d)\in \tilde{J}_{(l_1^*,\ldots,l_d^*)}} B^{\gamma;l}$, there exists γ such that the estimator given by (5) is strongly efficient for estimating $\mathbb{P}(A_n)$.

Remark 6. Even though our simulation algorithm is constructed in the context of Poisson processes with positive jump distributions, it can be easily generalized to the case where the jump distributions are regularly varying at both $-\infty$ and ∞ (for details, see the proof of theorem 3.5 in Rhee et al. [27] and the references therein).

Remark 7. We end this section with a final remark to point out the connection between the one-dimensional case and the multidimensional case. That is, if we set $d = 1$, then Assumption 2 coincides with Assumption 1, and no additional conditions are imposed on the set A. Moreover, the auxiliary sets B^γ in both cases are essentially the same. Thus, Theorem 2 can be considered as a special case of Theorem 1.

3.3. Extension to Random Walks

Let S_k, $k \geq 0$ be a centered random walk with increments $\{Y_k\}_{k\geq 1}$. Let $\mathbb{P}(Y_1 \leq -x)$ be regularly varying with index $-\alpha$, and let $\mathbb{P}(Y_1 \geq x)$ be regularly varying with index $-\beta$. Define $\bar{S}_n(t) \triangleq S_{\lfloor nt \rfloor}/n$, $t \geq 0$. In this subsection, we want to design an efficient simulation algorithm for estimating the probability of $\bar{S}_n \in A$. As in Sections 3.1 and 3.2, we make the following assumption for the set A.

Assumption 3. *Assume that A is a measurable set bounded away from $\mathbb{D}_{<l_-^*;l_+^*}$, where (l_-^*, l_+^*) is the unique solution of (2). Moreover, assume that $C_{l_-^*,l_+^*}(A^\circ) > 0$.*

Then, we construct the auxiliary set B^γ as follows.

Definition 3. Let (l_-^*, l_+^*) denote the unique solution to (2), and let

$$J_{l_-^*;l_+^*} \triangleq \left\{(l_-, l_+) \in \mathbb{Z}_+^2 \setminus I_{<l_-^*;l_+^*} \,\middle|\, (m_-, m_+) \prec (l_-, l_+) \text{ implies } (m_-, m_+) \in I_{<l_-^*;l_+^*}\right\},$$

where $I_{<l_-^*;l_+^*} \triangleq \{(l_-, l_+) \in \mathbb{Z}_+^2 \setminus \{(l_-^*, l_+^*)\} \mid (\alpha-1)l_- + (\beta-1)l_+ \le (\alpha-1)l_-^* + (\beta-1)l_+^*\}$. For $\gamma_- > 0$ and $\gamma_+ > 0$, define

$$B^\gamma \triangleq \bigcup_{(l_-, l_+) \in J_{l_-^*;l_+^*}} B^{\gamma; l_-^*; l_+^*}, \tag{8}$$

where $B^{\gamma; l_-^*; l_+^*} \triangleq \{\xi \in \mathbb{D} \mid \#\{t \mid \xi(t^-) - \xi(t) > \gamma_-\} \ge l_-^*, \#\{t \mid \xi(t) - \xi(t^-) > \gamma_+\} \ge l_+^*\}$.

Defining $A_n \triangleq \{\bar{S}_n \in A\}$ and $B_n^\gamma \triangleq \{\bar{S}_n \in B^\gamma\}$, we propose an estimator for $\mathbb{P}(\bar{S}_n \in A)$ that is given by (5). Note that, computing $\mathbb{P}(\bar{S}_n \in B^\gamma)$, as well as generating the sample path $\bar{S}_n$ under $\mathbb{Q}^\gamma$, can be achieved by following similar strategies as in Sections 3.1 and 3.2. Hence, the details are omitted (for examples, see Sections 4 and Section 5). We state the strong efficiency of our estimator in the following theorem without giving the proof, because it is analogous to the proof of Theorem 2.

Theorem 3. *Let $B_n^\gamma \triangleq \{\bar{X}_n \in B^\gamma\}$, where B^γ is as defined in (8). Under Assumption 3, there exists γ_- and γ_+ such that the estimator given by (5) is strongly efficient for estimating $\mathbb{P}(A_n)$.*

With the results presented in this section, we are able to apply our general simulation algorithm to three examples in the next sections. These examples can be found in the applications of mathematical finance, actuarial science, and queueing networks.

4. An Application to Finite-Time Ruin Probabilities

4.1. Problem Settings

Let S_k, $k \ge 0$, be a centered random walk with increments $\{Y_k\}_{k\ge 1}$. Moreover, let $\mathbb{P}(Y_1 > x)$ be regularly varying at infinity with index $-\beta$. For $c \ge 0$, define $A_n \triangleq \{\max_{0\le k\le n} Y_k \le nb, \max_{0\le k\le n} S_k - ck \ge na\}$. Additionally, we make a technical assumption that $a/b \notin \mathbb{Z}$. We are interested in computing $\mathbb{P}(A_n)$. This probability is particularly interesting because it is related to, for example, insurance, where huge claims may be reinsured and therefore are irrelevant in the sense of estimating the finite-time ruin probability of an insurance company.

4.2. Large Deviations Results

The rare-event probability can be estimated efficiently using the technique introduced in Section 3. To see this, define $A \triangleq \{\xi \in \mathbb{D}\colon \sup_{t\in[0,1]}[\xi(t) - ct] \ge a; \sup_{t\in[0,1]}[\xi(t) - \xi(t^-)] \le b\}$ and $\bar{S}_n \triangleq \{\bar{S}_n(t)\}_{t\in[0,1]}$, where $\bar{S}_n(t) = S_{\lfloor nt\rfloor}/n$ for $t \ge 0$. Note that $\mathbb{P}(A_n) = \mathbb{P}(\bar{S}_n \in A)$. Set $l^* = \lceil a/b \rceil$. Intuitively, l^* should be the key parameter, as it takes at least l^* jumps of size b to cross level a. This intuition has been made rigorous by Rhee et al. [27, section 5.1], where the authors show that A is bounded away from $\mathbb{D}_{<l^*}$, and hence, $\mathbb{P}(A_n) = \Theta(n^{l^*}\mathbb{P}(S_1 \ge n)^{l^*})$.

4.3. Construction of B^γ

We set $B^\gamma = \{\xi \in \mathbb{D} \mid \#\{t \mid \xi(t) - \xi(t^-) > \gamma\} \ge l^*\}$ and $B_n^\gamma = \{\bar{S}_n \in B^\gamma\} = \{\#\{k \in \{1, \ldots, n\} \mid Y_k > n\gamma\} \ge l^*\}$, where γ is the parameter that needs to be tuned. For the completeness of our algorithm, we give a closed-form expression for $\mathbb{P}(B_n^\gamma)$. Let p denote the probability of $\mathbb{P}(Y_1 > \gamma n)$; then we have that

$$\mathbb{P}(B_n^\gamma) = \sum_{i=l^*}^{n} \binom{n}{i} p^i (1-p)^{n-i} = 1 - \sum_{i=0}^{l^*-1} \binom{n}{i} p^i (1-p)^{n-i}, \tag{9}$$

where the latter representation in (9) is for numerical purposes.

4.4. Choice of γ

As we mentioned in Remark 3, a strategy of choosing the parameters γ needs to be discussed in the next step. From the proof of Theorem 2, it is sufficient to select γ such that $\mathbb{P}\left(A_n \cap (B_n^\gamma)^c\right) = o\left(\mathbb{P}(A_n)^2\right)$. We propose to select γ such that $(a - (l^*-1)b)/\gamma \notin \mathbb{Z}_+$ and that

$$\left\lceil \frac{a - (l^*-1)b}{\gamma} \right\rceil > l^* + 1. \tag{10}$$

In view of Theorem 3, it is sufficient to show that $A \cap (B^\gamma)^c$ is bounded away from $\mathbb{D}_{<2l^*+1}$ with γ satisfying (10). To see this, choose θ with $d(\theta, \mathbb{D}_{<2l^*+1}) < r$. This implies that there exists $\xi \in \mathbb{D}_{<2l^*+1}$ satisfying $d(\theta, \xi) < r$ and $\xi(t) = \sum_{j=1}^{2l^*} x_j \mathbb{1}_{[u_j,1]}(t)$. In particular, there exists a homeomorphism $\lambda : [0,1] \to [0,1]$ satisfying $\| \lambda - id \|_\infty \vee \| \xi \circ \lambda - \theta \|_\infty < r$. Hence, for $\theta \in A$, using the identity $\xi \circ \lambda = \theta + (\xi \circ \lambda - \theta)$, we conclude that the following holds:

1. $x_j < b + 2r$ for every $j \in \{1, \ldots, 2l^*\}$, and
2. there exists t' such that $\sum_{u_j \le 1} x_j \ge \sum_{u_j \le \lambda(t')} x_j > a - 2r$.

To see this, note that $\xi \circ \lambda(t) = \sum_{j=1}^{2l^*} x_j \mathbb{1}_{[u_j,1]}(\lambda(t)) = \theta + (\xi \circ \lambda - \theta)$ and $\|\xi \circ \lambda - \theta\|_\infty < r$. By the fact that $\sup_{t \in [0,1]} \cdot [\theta(t) - \theta(t^-)] \le b$, conclusion 1 follows. Moreover, by the fact that $\sup_{t\in[0,1]}[\theta(t) - ct] \ge a$, there exists t' such that $\xi \circ \lambda(t') = \sum_{u_j \le \lambda(t')} x_j > a - 2r$, and hence, conclusion 2 is obtained. This implies that $\sum_{j \ge l^*} x_j > a - 2r - (l^* - 1)(b + 2r)$. Moreover, for $\theta \in (B^\gamma)^c$, every jump of ξ should be bounded by $\gamma + 2r$ after having $l^* - 1$ jumps with sizes bigger than b. Because γ satisfies (10) and a is not a multiple of b, we obtain the result by choosing r small.

4.5. Sampling from Q_γ

Summarizing the discussion from previous paragraphs, we are able to propose a strongly efficient estimator for $\mathbb{P}(A_n)$ that is given by (5). As the last ingredient of our simulation algorithm, a strategy of sampling from $\mathbb{Q}_\gamma(\,\cdot\,)$ $(= \mathbb{P}(\,\cdot\,|B_n^\gamma))$ needs to be discussed. We use a similar strategy as in Algorithm 3 and formulate the pseudocode in Algorithm 4.

Algorithm 4

```
R ← true
while R = true do
    sample (i_1, ..., i_{l*}) uniformly from 𝒞({1, ..., n}, l*)
    for j ∈ {i_1, ..., i_{l*}} do
        sample Y_j ~ Y_1 | γn < Y_1 ≤ bn
    for j ∉ {i_1, ..., i_{l*}} do
        sample Y_j ~ Y_1
    sample u ~ uniform[0, 1]; c ← #{m ∈ {1, ..., n} | γn < Y_1 ≤ bn}; a ← (c choose l*); R ← true
    if u < a^{-1} then
        R ← false
    return (Y_1, ..., Y_n)
```

4.5. Numerical Results

Finally, we investigate our algorithm numerically based on a concrete example. Let $Y_1 = Y_1' - \mathbb{E}Y_1'$, where $\mathbb{P}(Y_1' > t) = (1/t)^\beta$. In Table 1, we select $c = 0.05$, $w = 0.05$ (for a heuristic of the choice of w and its impact on the empirical performance, see Section 5) and summarize the estimated probability and the level of precision (ratio between the radius of the 95% confidence interval and the estimated value) for different combinations of n, β, a, b, and c (based on 10^6 samples). We observe that, for different values of β, a, and b, the precision stays roughly constant as n grows. This confirms our theoretical results.

5. An Application in Barrier Option Pricing

In this section, we consider an application that arises in the context of financial mathematics; in particular, we consider a down-in barrier option (see section 11.3 in Tankov and Cont [30]).

5.1. Problem Settings

Let S_k, $k \ge 0$, be a centered random walk with increments $\{Y_k\}_{k\ge1}$. Let $\mathbb{P}(Y_1 \le -x)$ be regularly varying with index $-\alpha$, and let $\mathbb{P}(Y_1 \ge x)$ be regularly varying with index $-\beta$. Let a, b, and c be positive real numbers. We provide a strongly efficient estimator for the probability of $A_n \triangleq \{S_n \ge bn, \min_{0\le k\le n} S_k + ck \le -an\}$, which can be interpreted as the chance of exercising a down-in barrier option. This application is interesting because, as we will see, the large deviations behavior of $\mathbb{P}(A_n)$ is caused by two large jumps.

5.2. Large Deviations Results

Define $A \triangleq \{\xi \in \mathbb{D} : \xi(1) \ge b, \ \inf_{0 \le t \le 1} \xi(t) + ct \le -a\}$. Obviously, we have that $(l_-^*, l_+^*) = (1, 1)$, where (l_-^*, l_+^*) denotes the solution to (2). To verify the topological property of A, we define $m, \pi_1 : \mathbb{D} \to \mathbb{R}$ by $m(\xi) = \inf_{0\le t\le 1}\{\xi(t) + ct\}$, and $\pi_1(\xi) = \xi(1)$. Note that F, π_1, and m are continuous; therefore $F^{-1}(A) = m^{-1}(-\infty, -a] \cap \pi_1^{-1}[b, \infty)$ is a closed

Table 1. Estimated rare-event probability and level of precision for the application as described in Section 4.

EstPr	$n=80$		$n=100$		$n=200$	
Pr	$\beta=1.5$	$\beta=2.0$	$\beta=1.5$	$\beta=2.0$	$\beta=1.5$	$\beta=2.0$
$a=2$, $b=1.2$	1.171×10^{-3}	3.904×10^{-5}	1.043×10^{-3}	2.361×10^{-5}	6.316×10^{-4}	5.167×10^{-6}
$(l^*=2)$	2.053×10^{-2}	3.133×10^{-2}	2.057×10^{-2}	3.376×10^{-2}	2.130×10^{-2}	3.975×10^{-2}
$a=4$, $b=1.2$	5.099×10^{-7}	3.778×10^{-10}	3.860×10^{-7}	1.592×10^{-10}	1.326×10^{-7}	8.911×10^{-12}
$(l^*=4)$	1.799×10^{-2}	2.278×10^{-2}	1.761×10^{-2}	2.366×10^{-2}	1.717×10^{-2}	2.780×10^{-2}
$a=2$, $b=0.3$	1.635×10^{-10}	1.147×10^{-12}	1.795×10^{-10}	3.983×10^{-13}	1.202×10^{-10}	6.775×10^{-15}
$(l^*=7)$	6.441×10^{-2}	1.662×10^{-2}	5.456×10^{-2}	1.635×10^{-2}	3.535×10^{-2}	1.826×10^{-2}

set. By adapting the results in Rhee et al. 27, section 5.2, it can be shown that, for any arbitrary $i\ge 0$, $\mathbb{D}_{i;0}$ and $\mathbb{D}_{0;i}$ are bounded away from $m^{-1}(-\infty,-a]$ and $\pi_1^{-1}[b,\infty)$, respectively. Hence, A is bounded away from $\mathbb{D}_{<1;1}$. Applying Result 3, we obtain that $\mathbb{P}(\bar{X}_n\in A)=\Theta(n^2\mathbb{P}(S_1\ge n)\mathbb{P}(S_1\le n))$.

5.3. Construction of B^γ

Now we are in the framework of Theorem 3. Note that by Definition 3 we have $J_{1;1}=\{(1,1),(l,0),(0,m)\}$, where $l=\min\{l'\in\mathbb{Z}_+\,|\,(l'-1)(\beta-1)>(\alpha-1)\}$ and $m=\min\{m'\in\mathbb{Z}_+\,|\,(m'-1)(\alpha-1)>(\beta-1)\}$. However, adapting the idea behind Corollary 1 together with the fact that A is bounded away from both $\mathbb{D}_{i;0}$ and $\mathbb{D}_{0;i}$, it is sufficient to consider $\tilde{J}_{1;1}=\{(1,1)\}$. Hence, we can set $B^\gamma=\left\{\xi\in\mathbb{D}\,\middle|\,\#\{t\,|\,\xi(t^-)-\xi(t)>\gamma_-\}\ge 1,\ \#\{t\,|\,\xi(t)-\xi(t^-)>\gamma_+\}\ge 1\right\}$. As we mentioned in the introduction, it is possible that estimators may be crafted specifically for the events of interest to obtain (up to constant factors) better performance. Because at least one downward jump should happen *before* upward jumps, without introducing new notations, we can modify B^γ such that $B^\gamma=\left\{\left\{\xi\in\mathbb{D}\,\middle|\,\exists\, t_1<t_2:\xi(t_1^-)-\xi(t_1)>\gamma_-,\ \xi(t_2)-\xi(t_2^-)>\gamma_+\right\}\right\}$. This implies that $B_n^\gamma=\left\{\left\{\exists\, i<j: Y_i<-\gamma_- n,\ Y_j>\gamma_+ n\right\}\right\}$. By straightforward computation, we obtain that $\mathbb{P}(B_n^\gamma)=1-p_2(1-p_1)^n/(p_2-p_1)+p_1(1-p_2)^n/(p_2-p_1)$, where $p_1\triangleq\mathbb{P}(Y_1>\gamma_+ n)$ and $p_2\triangleq\mathbb{P}(Y_1<-\gamma_- n)$.

5.4. Choice of γ_- and γ_+

We discuss here the strategy of choosing the parameters γ_- and γ_+. From the proof of Theorem 2, it is sufficient to select γ_-, γ_+ such that $\mathbb{P}(A_n\cap(B_n^\gamma)^c)=o(\mathbb{P}(A_n)^2)$. Hence, we propose to choose γ_- and γ_+ such that $((a+b)/\gamma_+,a/\gamma_-)\notin\mathbb{Z}_+^2$ and that

$$\min\left\{(\alpha-1)+\left\lceil\frac{a+b}{\gamma_+}\right\rceil(\beta-1),\left\lceil\frac{a}{\gamma_-}\right\rceil(\alpha-1)+(\beta-1)\right\}>2(\alpha+\beta-2).\tag{11}$$

Without loss of generality, we assume that $\lceil a/\gamma_-\rceil(\alpha-1)+(\beta-1)$ is the unique minimum of (11). It suffices to prove that $A\cap(B^\gamma)^c$ is bounded away from $\mathbb{D}_{<\lceil a/\gamma_2\rceil;1}$. To show that $\cup_{(l_-,l_+)}\mathbb{D}_{<l_-;l_+}$ with $l_-\le\lceil a/\gamma_2\rceil-1$ is bounded away from $A\cap(B^\gamma)^c$, choose θ with $d(\theta,\cup_{(l_-,l_+)}\mathbb{D}_{<l_-;l_+})<r$. This implies that there exists $\xi\in\cup_{(l_-,l_+)}\mathbb{D}_{<l_-;l_+}$ satisfying $d(\theta,\xi)<r$, where $\xi=\sum_{k=1}^{l_+}x_k\mathbb{1}_{[u_k,1]}(t)-\sum_{k=1}^{l_-}y_k\mathbb{1}_{[v_k,1]}(t)$. In particular, there exists homeomorphism $\lambda:[0,1]\to[0,1]$ satisfying

$$||\,\lambda-id\,||_\infty\vee||\,\xi\circ\lambda-\theta\,||_\infty<r.\tag{12}$$

Using (12) and the identity $\xi\circ\lambda=\theta+(\xi\circ\lambda-\theta)$, we conclude that, for $\theta\in(B^\gamma)^c$ and $t\in[0,1]$, at least one of the following holds:

- $x_k\le\gamma_++2r$ for every $u_k\ge t$; or
- $y_k\le\gamma_-+2r$ for every $v_k<t$.

For $\theta\in m^{-1}(-\infty,-a]$, by (12), there exists t' such that

$$\sum_{u_j\le\lambda(t')}x_j-\sum_{v_j\le\lambda(t')}y_j\ <-a+3r.\tag{13}$$

Moreover, we can assume that $y_j\le\gamma_-+2r$ for j satisfying $v_j\le\lambda(t')$. Otherwise, x_j is bounded by γ_++2r for j satisfying $v_j>\lambda(t')$. By choosing r sufficiently small, this leads to a contradiction: that $\theta\in\pi_1^{-1}[b,\infty)$, and $\lceil a/\gamma_-\rceil(\alpha-1)+(\beta-1)$ is the minimum of (11). Hence, (13) implies that $(\lceil a/\gamma_-\rceil-1)(\gamma_-+2r)>a-3r$. Because $(\lceil a/\gamma_-\rceil-1)\gamma_-<a$, choosing r sufficiently small, we obtain the result. Similarly, it can be can shown that $A\cap(B^\gamma)^c$ is bounded away from $\cup_{(l_-,l_+)}\mathbb{D}_{<l_-;l_+}$ for $l_+\le\lceil(a+b)/\gamma_+\rceil-1$.

5.5. Sampling from Q_γ

As in Section 4, a strategy of sampling from $\mathbb{Q}_\gamma(\,\cdot\,)$ needs to be discussed. Even though B^γ is modified to obtain smaller relative error, a similar strategy as in Algorithm 3 can be used here. Hence we omit the details.

5.6. Numerical Results

We end this section with some numerical investigations. First let $Y_1 = Y_1' - \mathbb{E}Y_1'$, where Y_1' is a random variable with density function f_Y that is given by $f_Y = \frac{1}{3}\left(\frac{1}{y}\right)^\beta \mathbb{1}_{(1,\infty)}(y) + \frac{1}{3}\left(-\frac{1}{y}\right)^\alpha \mathbb{1}_{(-\infty,-1)}(y) + \frac{1}{6}\mathbb{1}_{[-1,1]}(y)$. We apply our algorithm to estimate $\mathbb{P}(S_n \geq bn, \min_{0\leq k\leq n} S_k \leq -an)$ with $a = 2$ and $b = 1.5$. In Figure 2, we plot the precision of the estimated probability against the parameter w for different values of n. We observe that the estimated probabilities become more precise as w decreases. This heuristic suggests the upper bound we derive in (28), where the latter term in (28) is the order of $o(\mathbb{P}(A_n)^2)$, as long as w is strictly positive. From this observation, we choose $w = 0.05$ for all numerical investigation presented in this paper. In Table 2, we compare the estimated rare-event probability and precision w.r.t. different values of n, α and β. We observe that the precision stays roughly constant as n increases for different combinations of α and β, which suggests the strong efficiency of our estimator.

Next, we make a comparison between the algorithms developed in this paper and in Gudmundsson and Hult [23], where a simulation algorithm is designed for estimating $\mathbb{P}(S_n \geq bn)$ from an MCMC perspective. First note that, instead of unbiased estimators, MCMC algorithms give us only consistent estimators. Furthermore, note that the event $\{S_n \geq bn\}$ is a special case of the event studied in this section with $a = 0$. Here, we consider Y_1 with density function $f_Y(y) = 2(x+1)^{-3}$. In Table 3, we present the estimated rare-event probability, the level of precision, the computational time (in seconds), and the normalized workload—that is, the (estimated) standard deviation multiplied by the computational time divided by the sample mean—produced by the two algorithms, based on 10^6 samples. Note that our algorithm typically outperforms the MCMC algorithm in terms of *computational time*—especially as n increases—while producing slightly larger *coefficient variation* compared with the MCMC algorithm. Overall, our algorithm seems to be more efficient for larger values of n, whereas MCMC seems to be more efficient for small values of n in terms of *normalized workload*. This can be explained by the fact that our estimator is state independent (i.e., the increments of the random walk S_n can be updated simultaneously). On the other hand, in the MCMC case, the algorithm needs a burn-in period to converge, and the increments have to be simulated following a specific order. Moreover, updating the value of, say, Y_k relies on the values of $Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n$ to run an MCMC algorithm. It may be noted that the range of probabilities examined in Table 3 is smaller than the typical range of practical interest in applications. For example, insurance companies (according to the Solvency II Directive) are suggested to have a capital

Figure 2. (Color online) A plot of the precision of the estimators discussed in Section 5 w.r.t. w for different values of n.

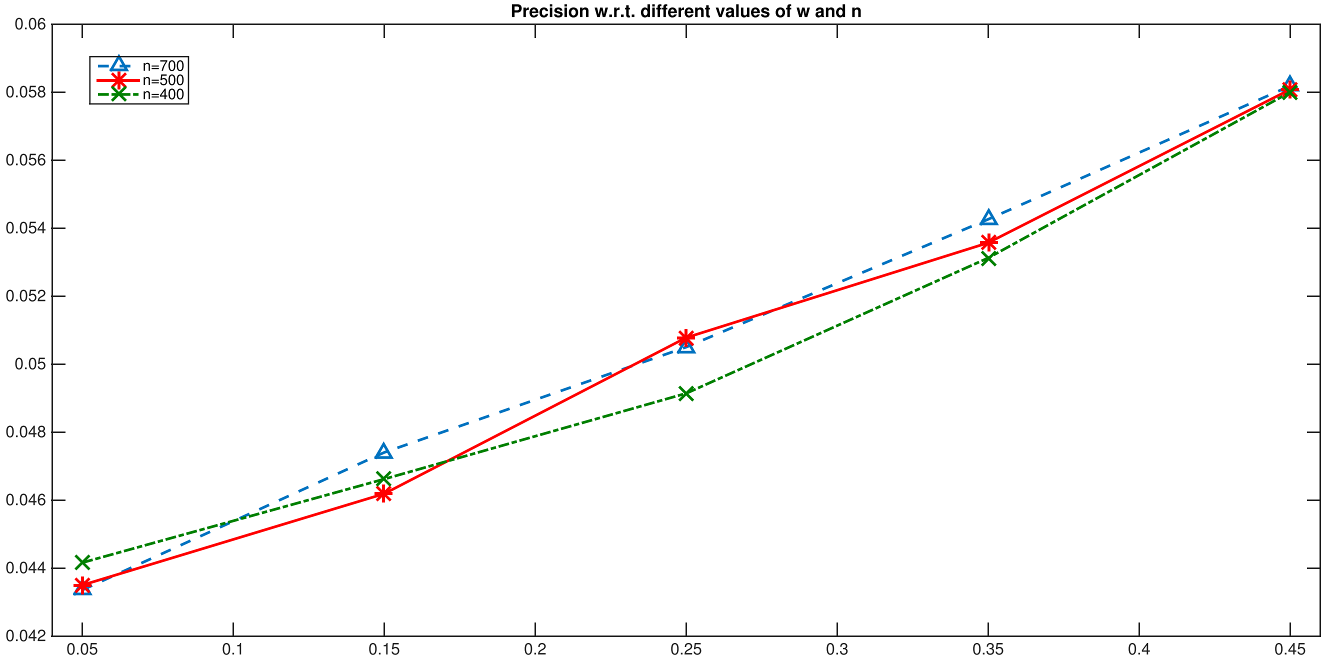
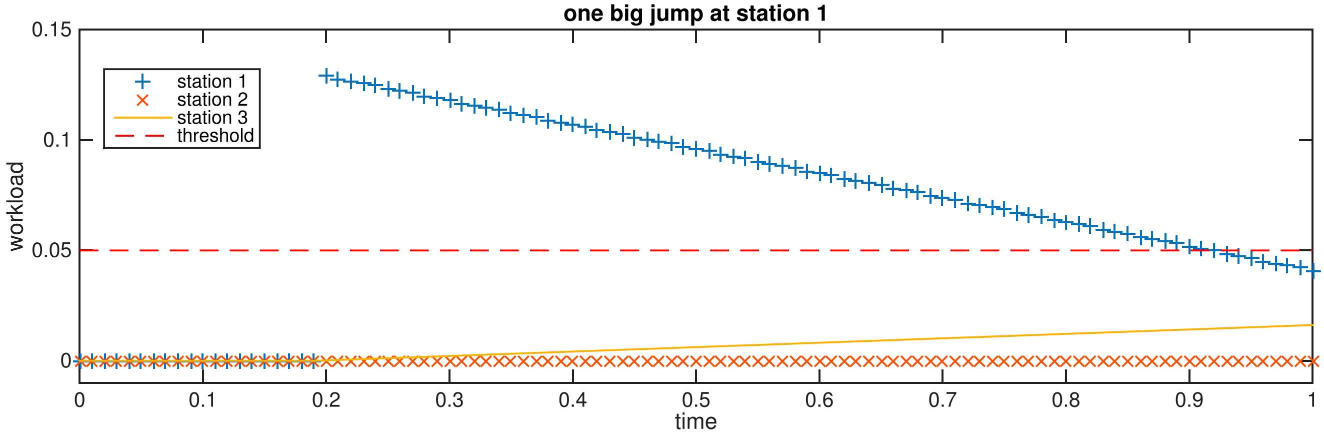
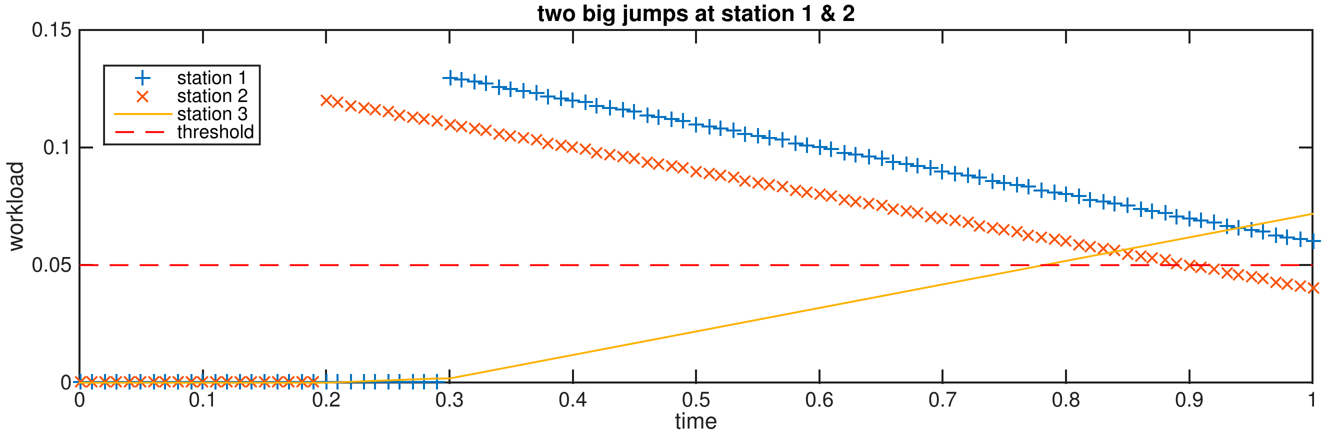


Figure 3. (Color online) An illustration of two different sample paths of workload processes (under the setting of Example 1), whose associated input processes have the same form as in (18).





reserve corresponding to bankruptcy events in the order of a 0.5% likelihood (on an annual basis). Nevertheless, when model uncertainty is taken into account, considering bankrupcy probability of the putative (assumed) parametric model in the range of likelihood that is considerably smaller than nominal values suggested by regulation may be necessary. For example, calibrating any distribution of claims with a degree of precision corresponding to a bankrupt probability of 0.5% in a nonparametric way is practically impossible because one would need (as a result of the central limit theorem) millions of observations on an annual basis. In such a case, an additional safety margin should be added to the capital requirement to account for model error, as discussed in Blanchet et al. [8] (see section 9.2.1). That is, one may have to consider the range of likelihood in the order of $10^{-4} \sim 10^{-5}$ for the putative parametric model to ensure the 0.005 likelihood for the true claims. Note that our importance sampling algorithm seems to be comparable or preferable to the MCMC algorithm in this range.

Table 2. Estimated rare-event probability and level of precision for the application as described in Section 5.

Est Pr	$n = 250$	$n = 500$	$n = 750$	$n = 1,000$	$n = 1,250$	$n = 1,500$
$\alpha = 2, \beta = 1.5$	3.913×10^{-7} 0.043	1.370×10^{-7} 0.043	6.992×10^{-8} 0.044	4.539×10^{-8} 0.044	3.305×10^{-8} 0.044	2.471×10^{-8} 0.044
$\alpha = 1.8, \beta = 1.7$	3.322×10^{-7} 0.037	1.154×10^{-7} 0.037	6.040×10^{-8} 0.038	3.840×10^{-8} 0.038	2.870×10^{-8} 0.038	2.225×10^{-8} 0.037
$\alpha = 2.3, \beta = 2$	1.923×10^{-9} 0.053	4.004×10^{-10} 0.053	1.491×10^{-10} 0.054	7.601×10^{-11} 0.054	4.632×10^{-11} 0.054	3.072×10^{-11} 0.054
$\alpha = 2.7, \beta = 1.8$	6.838×10^{-10} 0.068	1.121×10^{-10} 0.070	4.092×10^{-11} 0.070	2.079×10^{-11} 0.069	1.105×10^{-11} 0.071	6.896×10^{-12} 0.071

6. An Application to Queueing Networks

In this section, an application to queueing networks is considered. More specifically, the probability of the number of customers in a subset of the system crossing a high level is estimated. Although some particular cases exist that allow for an explicit analysis (see, e.g., section 13 in Dębicki and Mandjes [11]), it is hard to come up with exact results for the distribution of the workload process in general. Hence, implementing our algorithm in such a context is particularly interesting.

6.1. Model Description and Preliminaries

To be specific, we consider a d-dimensional stochastic fluid model. Suppose that jobs arrive to the ith station in the network according to a Poisson process with unit rate, which is denoted by $\{N^{(i)}(t)\}_{t\geq 0}$ and independent of $\{N^{(j)}\}$ for $j \neq i$. Moreover, the kth arrival of the ith station brings a job of size $W^{(i)}(k)$. We are assuming that $\{W(k) \triangleq (W^{(1)}(k), \ldots, W^{(d)}(k))^T\}_{k\geq 1}$ is a sequence of i.i.d. positive random vectors and that $\{W(k)\}_{k\geq 1}$ is independent of $\{N(t)\}_{t\geq 0}$. Therefore, the total amount of external work that arrives to the ith station up to time t is given by $J^{(i)}(t) = \sum_{k=1}^{N^{(i)}(t)} W^{(i)}(k)$. Now, assume that the workload at the ith station is processed as a fluid by the server at a rate r_i and that a proportion $Q_{ij} \geq 0$ of the fluid processed by the ith station is routed to the jth server. Moreover, we assume that Q is a substochastic matrix with $Q_{ii} = 0$ and that $Q^n \to 0$ as $n \to \infty$. The dynamics of the model are expressed formally by the so-called Skorokhod map (for details, see Skorokhod [28, 29]; see also, e.g., Harrison and Reiman [24]), that is defined in terms of a pair of processes (Z, Y) satisfying a stochastic differential equation that we shall describe now. Let $R = (I - Q)^T$, $r = (r_1, \ldots, r_d)^T$, $X(t) \triangleq J(t) - Rrt$, and $Z^{(i)}(t)$ denote the workload of the ith station at time t. For a given $Z^{(i)}(0)$, we have that

$$dZ(t) = dX(t) + RdY(t), \tag{14}$$

where $Y(\cdot)$ encodes the minimal amount of pushing required to keep $Z(\cdot)$ nonnegative. To describe how to characterize the solution (Z, Y) to (14), we need to introduce some notations. Let $\psi : \mathbb{D}^d \to \mathbb{D}^d_\uparrow$ with $\psi(x) \triangleq \inf\{w \in \mathbb{D}^d_\uparrow \mid x + Rw \geq 0\}$; that is, $\psi^{(i)}(x)(t) \triangleq \inf\{w^{(i)}(t) \in \mathbb{R} \mid w \in \mathbb{D}^d_\uparrow,\ x + Rw \geq 0\}$ for all i and t, and $\phi : \mathbb{D}^d \to \mathbb{D}^d$ with $\phi(x) \triangleq x + R\psi(x)$. The following results summarize useful properties and characterizations of the Skorokhod mappings ψ and ϕ, as well as the workload process $Z(t)$.

Result 4 (Theorems 14.2.1, 14.2.5, and 14.2.7 of Whitt [32]). *The mappings ψ and ϕ are well defined for all $x \in \mathbb{D}^d$. Moreover, ψ and ϕ are Lipschitz continuous w.r.t. both the uniform metric and the Skorokhod J_1 metric. If $Y(t) \triangleq \psi(X)(t)$ and $Z(t) \triangleq \phi(X)(t)$, then $(Y(t), Z(t))$ solve the Skorokhod problem given by (14).*

Result 5 (Lemma 14.3.3, Corollary 14.3.4, and Corollary 14.3.5 of Whitt [32]). *Let $x \in \mathbb{D}^d$. For the discontinuity points of $\psi(x)$ (denoted by $\mathrm{Disc}(\psi(x))$) and $\phi(x)$, we have that $\mathrm{Disc}(\psi(x)) \cup \mathrm{Disc}(\phi(x)) = \mathrm{Disc}(x)$. Moreover, if x has only positive jumps, then $\psi(x)$ is continuous, and $\phi(x)(t) - \phi(x)(t^-) = x(t) - x(t^-)$.*

Table 3. Estimated rare-event probability, level of precision, and computational time for estimating $\mathbb{P}(S_n \geq bn)$ using the algorithms introduced in this paper and in Gudmundsson and Hult [23].

Est Pr Time (s)	$n = 5$		$n = 20$		$n = 200$		$n = 1{,}000$	
NW	MCMC	IS	MCMC	IS	MCMC	IS	MCMC	IS
$b = 20$	5.340×10^{-4}	5.286×10^{-4}	1.375×10^{-4}	1.369×10^{-4}	1.384×10^{-5}	1.384×10^{-5}	2.770×10^{-6}	2.769×10^{-6}
	0.587×10^{-3}	1.020×10^{-3}	0.636×10^{-3}	1.060×10^{-3}	0.645×10^{-3}	1.071×10^{-3}	0.644×10^{-3}	1.073×10^{-3}
	25.5	19.6	67.9	21.5	561.0	44.4	3,686.8	145.0
	7.633	10.196	22.046	11.624	184.488	24.267	1,211.854	79.370
$b = 150$	8.962×10^{-6}	8.958×10^{-6}	2.250×10^{-6}	2.250×10^{-6}	2.252×10^{-7}	2.252×10^{-7}	4.505×10^{-8}	4.503×10^{-8}
	2.052×10^{-4}	8.287×10^{-4}	2.239×10^{-4}	8.331×10^{-4}	2.289×10^{-4}	8.353×10^{-4}	2.289×10^{-4}	8.370×10^{-4}
	24.8	17.1	63.1	19.9	545.7	43.7	3687.0	146.4
	2.594	7.227	7.204	8.468	63.712	18.633	430.399	62.543
$b = 1{,}000$	2.002×10^{-7}	2.001×10^{-7}	5.008×10^{-8}	5.005×10^{-8}	5.009×10^{-9}	5.011×10^{-9}	1.002×10^{-9}	1.002×10^{-9}
	0.796×10^{-5}	8.051×10^{-4}	0.848×10^{-5}	8.068×10^{-4}	0.851×10^{-5}	8.042×10^{-4}	0.881×10^{-5}	8.052×10^{-4}
	27.8	17.5	65.9	20.4	577.4	45.5	4,011.9	149.7
	1.130	7.206	2.854	8.384	25.070	18.684	180.317	61.482

Result 6 (Theorem 14.2.2 of Whitt [32]). *The regulator map $y = \psi(x)$ can be characterized as the unique fixed point of the map $\pi_{x,Q}: \mathbb{D}^d_\uparrow \to \mathbb{D}^d_\uparrow$, where $\pi_{x,Q}(w)(t) \triangleq \max\{0, \sup_{s\in[0,t]} Q^T w(s) - x(s)\}$.*

Result 7 (Consequence of Theorem 4.1 of Ramasubramanian [26]). *Let $\Delta \in \mathbb{D}^d$ be a nondecreasing function such that $\Delta(0) \geq 0$. Then, for $x \in \mathbb{D}^d$, we have that $\psi(x) \geq \psi(x+\Delta)$, $\phi(x) \leq \phi(x+\Delta)$, and $\phi(x)(t_2) - \phi(x)(t_1) \leq \phi(x+\Delta)(t_1) - \phi(x+\Delta)(t_2)$ for any $0 \leq t_1 \leq t_2 \leq 1$.*

Finally, we assume that the right tail of $W^{(i)}(1)$ is regularly varying with index $-\beta_i$ and that the stability condition holds (i.e., $R^{-1}\rho < r$, where $\rho \triangleq \mathbb{E}J(1)$). Let $\bar{Z}_n(t) \triangleq Z(nt)/n$ and $\bar{X}_n(t) \triangleq X(nt)/n$. Let $c \in \{0,1\}^d$ be a binary vector, and let $\mathcal{J}_c$ denote the index set encoded by c (i.e., $j \in \mathcal{J}_c$ if $c_j = 1$). Set $\bar{Z}_n(t) \triangleq Z(nt)/n$ and $\bar{X}_n(t) \triangleq X(nt)/n$. Define $l_c : \mathbb{R}^d \to \mathbb{R}$ by $l_c(x) = c^T x$, and define $\pi_1 : \mathbb{D}^d \to \mathbb{R}^d$ by $\pi_1(\xi) = \xi(1)$. Moreover, let $F \triangleq l_c \circ \pi_1 \circ \phi$. We are interested in estimating the probability of $\mathbb{P}(c^T \bar{Z}_n(1) \geq a)$. By theorem 14.2.6(iii) of Whitt [32], we have that $\bar{Z}_n = \phi(\bar{X}_n)$, and hence it holds that, for $a > 0$ and $A \triangleq \{\xi \in \mathbb{D} : F(\xi) \geq a\}$,

$$\mathbb{P}\big(c^T \bar{Z}_n(1) \geq a\big) = \mathbb{P}\big(F(\bar{X}_n) \geq a\big) = \mathbb{P}\big(\bar{X}_n \in A\big). \tag{15}$$

6.2. Large Deviations Results

To obtain the large deviations asymptotics for the rare-event probability as in (15), we proceed as follows.

- To determine the tail index of the rare-event probability, we study first the optimization problem given by (1) and transform it into a (nonstandard) knapsack problem with nonlinear constraints (see (21) and Proposition 4).
- Under a certain assumption (see Assumption 4), we show that A, as defined in (15), is bounded away from $\mathbb{D}_{<(l_1^*,\ldots,l_d^*)}$, where $l_1^*, \ldots, l_d^*$ is the optimal solution to the knapsack problem derived in the first step.
- Finally, we derive a large deviations result for $\mathbb{P}(\bar{X}_n \in A)$ by applying Result 2.

We start with the optimization problem given by (1). Because $X(t)$ is, in general, not a compensated compound Poisson process but one with certain drift, it is convenient to consider a slightly different problem, which is given by

$$\operatorname*{arg\,min}_{(l_1,\ldots,l_d)\in\mathbb{Z}_+^d,\ \prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)\cap A\neq\emptyset} \mathcal{I}(l_1,\ldots,l_d), \tag{16}$$

where $\mu \triangleq \mathbb{E}X(1) = \rho - Rr$, $r' = r - R^{-1}\rho > 0$, because of the stability condition, and $\mathbb{L}_{l_i}(\mu_i) \triangleq \{\xi \mid \exists \xi' \in \mathbb{D}_{l_i} : \xi(t) = \xi'(t) + \mu_i t = \xi'(t) - (Rr')_i t\}$. Define $E_0 \triangleq \big\{(l_1,\ldots,l_d) \in \mathbb{Z}_+^d \,\big|\, l_i = 0,\ \forall i \in \mathcal{J}_c\big\}$ and $E_1 \triangleq \{e_i \mid i \in \mathcal{J}_c\}$, where e_i denotes the unit vector with entries 0 except for the ith coordinate. By Result 5, instead of (16), we can solve two separate problems that are given by

$$\operatorname*{arg\,min}_{(l_1,\ldots,l_d)\in E_0,\ \prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)\cap A\neq\emptyset} \mathcal{I}(l_1,\ldots,l_d), \quad \text{and} \quad \operatorname*{arg\,min}_{(l_1,\ldots,l_d)\in E_1,\ \prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)\cap A\neq\emptyset} \mathcal{I}(l_1,\ldots,l_d). \tag{17}$$

Note that the latter problem in (17) can be solved easily by considering $\min_{i\in\mathcal{J}_c} \beta_i - 1$; therefore we focus on the first problem in (17). Let $\mathcal{J}$ be a subset of $(\mathcal{J}_c)^c$. Moreover, let $\theta \in \mathbb{D}_1$, and let $\xi \in \mathbb{D}^d$ be such that

$$\xi^{(i)}(t) = \begin{cases} -(Rr')_i t,\ t\in[0,1] & \text{for } i \notin \mathcal{J}, \\ \theta^{(i)} - (Rr')_i t,\ t \in [0,1] & \text{for } i \in \mathcal{J}. \end{cases} \tag{18}$$

A necessary and sufficient condition for the existence of $\xi \in A$ is given in the following proposition.

Proposition 4. *Let $\mathcal{J} \subseteq (\mathcal{J}_c)^c$. Moreover, let $\{r_i^*\}_{i\notin\mathcal{J}}$ be such that*

$$r_i^* = \max\left\{ r_i' - \sum_{j\neq i} Q_{ji} r_j' + \sum_{\substack{j\neq i\\ j\notin\mathcal{J}}} Q_{ji} r_j^*, 0 \right\} \qquad \textit{for } i \notin \mathcal{J}. \tag{19}$$

Define

$$\partial_z(\mathcal{J}) \triangleq \sum_{i\in\mathcal{J}_c} \left(r_i^* - r_i' + \sum_{j\neq i} Q_{ji} r_j' - \sum_{\substack{j\neq i,\\ j\notin\mathcal{J}}} Q_{ji} r_j^* \right). \tag{20}$$

If $\partial_z(\mathcal{J}) \neq a$, then there exists ξ satisfying (18) and $c^T\phi(\xi)(1) \geq a$ if and only if $\partial_z(\mathcal{J}) > a$. Additionally, if $\mathcal{J}_1 \subseteq \mathcal{J}_2 \subseteq (\mathcal{J}_c)^c$, then we have that $\partial_z(\mathcal{J}_1) \leq \partial_z(\mathcal{J}_2)$.

We give a sketch of the proof and refer to Section 7 for details. Note that $\partial_z(\mathcal{J})$ given by (20) is the increasing rate of the subset $\mathcal{J}_c$ of the workload process, whose associated input process does not have any jumps but starts with sufficiently large initial value. Based on this observation, a ξ can be constructed for the "if" part of the first statement. For the "only if" part, suppose that there is a ξ satisfying $c^T\phi(\xi)(1) \geq a$. By Result 7, enlarging the size of jumps in ξ will preserve the fact that $c^T\phi(\xi)(1) \geq a$. Hence, we can construct a new ξ, such that

- the associated workload process $\phi(\xi)$ is piecewise linear between two neighboring discontinuity points, and
- the increasing rate of $c^T\phi(\xi)$ is always smaller than or equal to $\partial_z(\mathcal{J})$ given by (20).

Remark 8. Note that (19) can be written in a matrix notation that is given by $r^* = \max\{((I-Q^T)r')_{\notin\mathcal{J}} + (Q_{\notin\mathcal{J}})^T r^*, 0\} = \max\{(Rr-\rho)_{\notin\mathcal{J}} + (Q_{\notin\mathcal{J}})^T r^*, 0\}$, where $(Rr-\rho)_{\notin\mathcal{J}}$ and $Q_{\notin\mathcal{J}}$ denote the vector and matrix, respectively, with the ith row and column being removed for all $i \in \mathcal{J}$. Using the Banach fixed point theorem, we obtain that $r^* = \lim_{n\to\infty}\underline{\pi}^n(0)$, where $\underline{\pi}^n \triangleq \underline{\pi} \circ \underline{\pi}^{n-1}$ and $\underline{\pi}(x) \triangleq \max\{(Rr-\rho)_{\notin\mathcal{J}} + (Q_{\notin\mathcal{J}})^T x, 0\}$.

Define $E_{\mathcal{J}} \triangleq \{(l_1,\ldots,l_d) \in E_0' \mid \partial_z(\mathcal{J}_{(l_1,\ldots,l_d)}) > a\}$, where $E_0' \triangleq E_0 \cap \{(l_1,\ldots,l_d) \in \mathbb{Z}_+^d \mid l_i \in \{0,1\}, \forall i \notin \mathcal{J}_c\}$, and $\partial_z(\mathcal{J}_{(l_1,\ldots,l_d)})$ is as defined in (20), with $\mathcal{J}_{(l_1,\ldots,l_d)}$ denoting the index set encoded by $(l_1,\ldots,l_d) \in E_0'$. By Proposition 4, we conclude that the first problem in (17) is equivalent to

$$\underset{(l_1,\ldots,l_d)\in E_{\mathcal{J}}}{\arg\min}\ \mathcal{I}(l_1,\ldots,l_d). \tag{21}$$

Thanks to the last statement of Proposition 4, it is unnecessary to check every $(l_1,\ldots,l_d) \in E_{\mathcal{J}}$ for solving (21). However, the optimization problem formulated in (21) is a nonstandard knapsack problem with nonlinear constraints. In the following example, we consider a specific fluid network and illustrate how to solve (21) using Proposition 4.

Example 1. Consider the fluid network given by $\rho = (0.8\ 0.8\ 1)^T$, $r = (1\ 1\ 2.5)^T$, and

$$Q = \begin{bmatrix} 0 & 0.1 & 0.8 \\ 0.1 & 0 & 0.8 \\ 0 & 0 & 0 \end{bmatrix}.$$

We are interested in the probability of the rare event that the third station crosses the level na at time n for large n (i.e., $\mathcal{J}_c = \{3\}$). It is easy to check that the stability condition holds. By an easy computation, we obtain that $\partial_z(\{1,2\}) = 0.1$ and $\partial_z(\{1\}) = \partial_z(\{2\}) = 0.02$. For $a = 0.05$, the optimal solution to (21) is given by $(1,1,0)$.

Suppose that we have solved (21). To obtain the large deviations results, the following technical assumption needs to be made.

Assumption 4. *Assume that* (21) *satisfies the conditions as follows.*
(a) *The optimization problem given by* (21) *has a unique solution.*
(b) *For every* $\mathcal{J} \subseteq (\mathcal{J}_c)^c$, *it holds that* $\partial_z(\mathcal{J}) \neq a$.
(c) *Let* $(l_1^*,\ldots,l_d^*)$ *denote the optimal solution to* (21). *We assume that* $\mathcal{I}(l_1^*,\ldots,l_d^*) < \min_{i\in\mathcal{J}_c}\beta_i - 1$.

By Result 2, Assumption 4(c) implies that the objective value of the first problem in (17) is strictly less than the objective value of the latter one in (17), and hence the optimal solution $(l_1^*,\ldots,l_d^*)$ to (21) solves (16). In view of this observation, the rare event is caused by multiple large jumps. Throughout the rest of this section, we assume that Assumption 4 holds. We end this subsection with a large deviations result for $\mathbb{P}(\bar{X}_n \in A) = \mathbb{P}(c^T\bar{Z}_n(1) \geq a)$, which is formulated in the following proposition.

Proposition 5. *Suppose that Assumption 4 holds. Let F be as defined in* (15). *Then* $A = F^{-1}[a,\infty)$ *is bounded away from* $\bigcup_{(l_1,\ldots,l_d)\in I_{<(l_1^*,\ldots,l_d^*)}}\prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)$, *where* $(l_1^*,\ldots,l_d^*)$ *denotes the unique optimal solution of* (21). *Moreover, we have that*

$$\begin{aligned} C_{l_1^*}\times\cdots\times C_{l_d^*}((F^{-1}[a,\infty))^\circ) &\leq \liminf_{n\to\infty}\frac{\mathbb{P}(\bar{X}_n\in A)}{\prod_{i=1}^d(n\nu_i[n,\infty))^{l_i^*}} \\ &\leq \limsup_{n\to\infty}\frac{\mathbb{P}(\bar{X}_n\in A)}{\prod_{i=1}^d(n\nu_i[n,\infty))^{l_i^*}} \leq C_{l_1^*}\times\cdots\times C_{l_d^*}(F^{-1}[a,\infty)). \end{aligned}$$

6.3. Simulation

Again, we are in the setting of Theorem 2. To be able to discuss the choice of $J_{(l_1^*,\ldots,l_d^*)}$ and the parameter γ in a more precise context, let us consider the stochastic fluid network introduced in Example 1.

Example 1 (Continued). Recall that, for $a = 0.05$, the optimal solution of (16) is given by $\beta_1 + \beta_2 - 2$ if we assume that $\beta_1 + \beta_2 - 2 < \beta_3 - 1$. Moreover, it can be easily shown that A is bounded away from both $\mathbb{D}_{<i} \times \mathbb{D}_0 \times \mathbb{D}_0$ and $\mathbb{D}_0 \times \mathbb{D}_{<j} \times \mathbb{D}_0$. Combining this with $\mathcal{I}(1,1,1) > 2\mathcal{I}(1,1,0)$, as well as Corollary 1, it is sufficient to take $\tilde{J}_{(l_1^*,\dots,l_d^*)} = \{(1,1,0),(0,0,1)\}$. This implies that

$$B_n^\gamma = \{\#\{k \,|\, W^{(i)}(k) > n\gamma_i, k \le N^{(i)}(n)\} \ge 1, i \in \{1,2\}\} \cup \{\#\{k \,|\, W^{(3)}(k) > n\gamma_3, k \le N^{(3)}(n)\} \ge 1\},$$

and hence $(B_n^\gamma)^c = \{\exists i \in \{1,2\} : W^{(i)}(k) \le n\gamma_i,\ \forall k \le N^{(i)}(n)\} \cap \{W^{(3)}(k) \le n\gamma_3,\ \forall k \le N^{(3)}(n)\}$. We choose γ such that $\mathbb{P}(A_n \cap (B_n^\gamma)^c) = o(\mathbb{P}(A_n)^2)$. To begin with, we assume, without loss of generality, that $\beta_3 - 1 \le 2(\beta_1 + \beta_2 - 2)$; otherwise, we can simply set $J_{(l_1^*,\dots,l_d^*)} = \{(1,1,0)\}$, because $\mathcal{I}(0,0,1) > 2\mathcal{I}(1,1,0)$. Now the parameter γ_3 can be chosen such that $\lceil \frac{1}{20}/\gamma_3 \rceil (\beta_3 - 1) > 2(\beta_1 + \beta_2 - 2)$. For the choice of γ_1, we observe that the job arriving at the second station can have an arbitrarily large size. Hence, it is sufficient to consider the inequality $\partial_z(\{1,2\})t' + \partial_z(\{2\})(1 - t') > a$, where $\partial_z(\{1,2\}) = 0.1$ and $\partial_z(\{2\}) = 0.02$. Solving the inequality, we obtain that $t' < 3/8$. This simply means that the workload process of the third station cannot exceed the level a at time 1 if we keep both the first and the second stations overloaded less than 3/8 of the time. Because the workload process of the first station decays at rate 1/10, one can choose γ_1 such that $\lceil \frac{3}{80}/\gamma_1 \rceil (\beta_1 - 1) + (\beta_2 - 1) > 2(\beta_1 + \beta_2 - 2)$. Analogously, it is sufficient to set γ_2 such that $(\beta_1 - 1) + \lceil \frac{3}{80}/\gamma_2 \rceil (\beta_2 - 1) > 2(\beta_1 + \beta_2 - 2)$.

We give a closed-form expression for $\mathbb{P}(B_n^\gamma)$. We assume that $\{W^{(i)}(k)\}_{1 \le i \le d}$ are mutually independent; therefore we have that

$$\begin{aligned}\mathbb{P}((B_n^\gamma)^c) &= \mathbb{P}(\exists i \in \{1,2\} : W^{(i)}(k) \le n\gamma_i,\ \forall k \le N^{(i)}(n))\, \mathbb{P}(W^{(3)}(k) \le n\gamma_3,\ \forall k \le N^{(3)}(n)) \\ &= \left[1 - \prod_{i=1}^{2}(1 - \mathbb{P}(W^{(i)}(k) \le n\gamma_i,\ \forall k \le N^{(i)}(n)))\right] \mathbb{P}(W^{(3)}(k) \le n\gamma_3,\ \forall k \le N^{(3)}(n)).\end{aligned}$$

Conditional on $N^{(i)}(n)$, we obtain $\mathbb{P}(W^{(i)}(k) \le n\gamma_i,\ \forall k \le N^{(i)}(n)) = \exp\{-n(1 - \mathbb{P}(W^{(i)}(1) \le n\gamma_i))\}$. Summarizing the findings from above, we are able to propose a strongly efficient estimator for $\mathbb{P}(A_n)$ that is given by (5). Moreover, Algorithm 3 can be used to sample from $\mathbb{Q}^\gamma$. To see this, we decompose B_n^γ into two disjoint sets, $B_n^\gamma(1)$ and $B_n^\gamma(2)$, that are given by $B_n^\gamma(1) \triangleq \{\#\{k \,|\, W^{(3)}(k) > n\gamma_3,\ k \le N^{(3)}(n)\} \ge 1\}$ and

$$B_n^\gamma(2) \triangleq \{\#\{W^{(i)}(k) > n\gamma_i,\ k \le N^{(i)}(n)\} \ge 1, \forall i \in \{1,2\}\} \cap \{W^{(3)}(k) \le n\gamma_3,\ \forall 1 \le k \le N^{(3)}(n)\},$$

respectively. Using Algorithm 3, the sample path of $\bar{X}_n^{(1)}, \bar{X}_n^{(2)}, \bar{X}_n^{(3)}$ can be simulated independently on both $B_n^\gamma(1)$ and $B_n^\gamma(2)$. We present the numerical results based on 20,000 samples in Table 4. We choose $W^{(i)}(1)$ such that $\mathbb{P}(W^{(i)}(1) > t) = (t_{r,i}/t)^{\beta_i}$ and $t_{r,i} = \rho_i(\beta_i - 1)/\beta_i$ for $i \in \{1,2,3\}$. As one can see, the numerical results suggest again what our theory predicts.

7. Proofs

In this section, we provide proofs of the results presented in this paper.

Proof of Proposition 1. Recall that the expected running time of the rejection sampling (see Algorithm 1), which is used to generate the jumps of $\bar{X}_n$, is given by $M_l = \binom{l}{l^*}\mathbb{P}(W(1) > n\gamma)^{l^*}\mathbb{P}(B_n^\gamma | N(n) = l)^{-1}$. Hence, for the expected running time of Algorithm 1, denoted by $T_{\text{alg1}}(n)$, we have that

$$\begin{aligned}T_{\text{alg1}}(n) &= \sum_{l \ge l^*} h_l M_l = \mathbb{P}(B_n^\gamma)^{-1} \sum_{l \ge l^*} \mathbb{P}(B_n^\gamma \,|\, N(n) = l)\mathbb{P}(N(n) = l) M_l \\ &= \mathbb{P}(B_n^\gamma)^{-1} \sum_{l \ge l^*} \mathbb{P}(N(n) = l) \binom{l}{l^*} \mathbb{P}(W(1) > n\gamma)^{l^*} \\ &= \frac{n^{l^*}(\lambda \mathbb{P}(W(1) > n\gamma))^{l^*}}{\mathbb{P}(B_n^\gamma)} e^{-\lambda n} \sum_{l \ge l^*} \frac{(\lambda n)^{l - l^*}}{(l - l^*)!} = \frac{n^{l^*}(\lambda \mathbb{P}(W(1) > n\gamma))^{l^*}}{\mathbb{P}(B_n^\gamma)}.\end{aligned}$$

Recall that $B_n^\gamma \triangleq \{\bar{X}_n \in B^\gamma\}$, where $B^\gamma \triangleq \{\xi \,|\, \#\{t \,|\, \xi(t) - \xi(t^-) > \gamma\} \ge l^*\}$. Noting that B_n^γ is bounded away from $\mathbb{D}_{<l^*}$ and $l^* = \min\{l \in \mathbb{Z}_+ | \mathbb{D}_l \cap B_n^\gamma\}$, by Result 1, we obtain that $\limsup_{n \to \infty} T_{\text{alg1}}(n) \le\ n^{l^*}(\lambda \mathbb{P}(W(1) > n\gamma))^{l^*}\mathbb{P}(B_n^\gamma)^{-1} \le C_{l^*}((B^\gamma)^\circ)^{-1} < \infty$. ◻

Table 4. Estimated rare-event probability and level of precision for the application as described in Section 6.3.

Est Pr	$n = 1{,}200$	$n = 1{,}600$	$n = 2{,}000$	$n = 2{,}400$
$\beta_1 = 1.5, \beta_2 = 1.5, \beta_3 = 2.2$	7.719×10^{-2} 0.045	6.228×10^{-2} 0.058	4.541×10^{-2} 0.057	3.973×10^{-2} 0.057
Est Pr	$n = 800$	$n = 1{,}200$	$n = 1{,}600$	$n = 2{,}000$
$\beta_1 = 2.5, \beta_2 = 2.3, \beta_3 = 4$	2.894×10^{-2} 0.325	1.686×10^{-2} 0.404	6.153×10^{-3} 0.445	2.023×10^{-3} 0.448
Est Pr	$n = 600$	$n = 1{,}000$	$n = 1{,}400$	$n = 1{,}800$
$\beta_1 = 2.2, \beta_2 = 2.9, \beta_3 = 4.5$	5.139×10^{-2} 0.249	1.858×10^{-2} 0.347	9.987×10^{-3} 0.351	1.028×10^{-3} 0.377

Proof of Proposition 2. Let B^γ be as in (6), and let $I \subseteq J_{(l_1^*,\dots,l_d^*)}$. Define $B_I^{\gamma;l} \triangleq \bigcap_{(l_1,\dots,l_d)\in I} B^{\gamma;l}$. By the inclusion–exclusion principle, we have that

$$\mathbb{P}(\bar{X}_n \in B^\gamma) = \sum_{k=1}^{|J_{(l_1^*,\dots,l_d^*)}|} (-1)^{k-1} \sum_{|I|=k,\, I\subseteq J_{(l_1^*,\dots,l_d^*)}} \mathbb{P}(\bar{X}_n \in B_I^{\gamma;l}). \tag{22}$$

Moreover, for any finite collection I of elements in $\mathbb{Z}_+^d$ with $I \subseteq J_{(l_1^*,\dots,l_d^*)}$, we have that

$$\begin{aligned} B_I^{\gamma;l} &= \bigcap_{i=1}^d \bigcap_{(l_1,\dots,l_d)\in I} \{(\xi^{(1)},\dots,\xi^{(d)}) \,|\, \#\{t \,|\, \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \geq l_i\} \\ &= \bigcap_{i=1}^d \left\{(\xi^{(1)},\dots,\xi^{(d)}) \,|\, \#\{t \,|\, \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \geq \hat{l}_{i;I}\right\}, \end{aligned} \tag{23}$$

where $\hat{l}_{i;I} \triangleq \max_{(l_1,\dots,l_d)\in I} l_i$. Because $\bar{X}_n^{(1)},\dots,\bar{X}_n^{(d)}$ are independent processes, we obtain that

$$\mathbb{P}(B_I^{\gamma;l}) = \prod_{i=1}^d \left(1 - \exp\left\{-\lambda_i n \mathbb{P}(W^{(i)}(1) > n\gamma_i)\right\} \sum_{j=0}^{\hat{l}_{i;I}-1} \frac{(\lambda_i n)^j}{j!} \mathbb{P}(W^{(i)}(1) > n\gamma_i)^j\right). \quad \square$$

Proof of Lemma 1. Recall that $B^{\gamma;l}(i,j) \triangleq \{\xi \in \mathbb{D}^d \,|\, \#\{t \,|\, \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \geq (l(j))_i\}$. Hence, we have that

$$B^\gamma = \bigcup_{j=1}^{|J_{(l_1^*,\dots,l_d^*)}|} \bigcap_{i=1}^d B^{\gamma;l}(i,j) = \bigcup_{j=1}^{|J_{(l_1^*,\dots,l_d^*)}|} \left(B^{\gamma;l}(1,j) \cap \bigcap_{i=2}^d B^{\gamma;l}(i,j)\right). \tag{24}$$

By definition, $\Delta B^{\gamma;l}(i,j) \triangleq B^{\gamma;l}(i,j) \backslash \left(\bigcup_{m=1}^{j-1} B^{\gamma;l}(i,m)\right)$. Therefore, we have that

$$B^{\gamma;l}(i,j) = \bigcup_{m_i=1}^{j} \Delta B^{\gamma;l}(i,j). \tag{25}$$

Plugging (25) into (24), we obtain that

$$\begin{aligned} B^\gamma &= \bigcup_{j=1}^{|J_{(l_1^*,\dots,l_d^*)}|} \left(\left(\bigcup_{m_1=1}^{j} \Delta B^{\gamma;l}(1,m_1)\right) \cap \bigcap_{i=2}^d B^{\gamma;l}(i,j)\right) \\ &= \bigcup_{m_1=1}^{|J_{(l_1^*,\dots,l_d^*)}|} \left(\bigcup_{j=m_1}^{|J_{(l_1^*,\dots,l_d^*)}|} \left(\Delta B^{\gamma;l}(1,m_1) \cap \bigcap_{i=2}^d B^{\gamma;l}(i,j)\right)\right) \\ &= \bigcup_{m_1=1}^{|J_{(l_1^*,\dots,l_d^*)}|} \left(\Delta B^{\gamma;l}(1,m_1) \cap \left(\bigcup_{j=1}^{m_1} \bigcap_{i=2}^d B^{\gamma;l}(i,j)\right)\right). \end{aligned}$$

Applying the same procedure to $\bigcup_{j=1}^{m_1}\bigcap_{i=2}^{d} B^{\gamma;l}(i,j)$, we obtain that

$$B^{\gamma} = \bigcup_{m_1=1}^{|J_{(l_1^*,\ldots,l_d^*)}|} \bigcup_{m_2=1}^{m_1} \left(\Delta B^{\gamma;l}(1,m_1) \cap \Delta B^{\gamma;l}(2,m_2) \cap \left(\bigcup_{j=1}^{m_2}\bigcap_{i=3}^{d} B^{\gamma;l}(i,j)\right)\right).$$

Iterating the same procedure $d-1$ times, we obtain that

$$B^{\gamma} = \bigcup_{m_1=1}^{|J_{(l_1^*,\ldots,l_d^*)}|} \bigcup_{m_2=1}^{m_1} \cdots \bigcup_{m_{d-1}=1}^{m_{d-2}} \left(\left(\bigcap_{i=1}^{d-1} \Delta B^{\gamma;l}(i,m_i)\right) \cap \left(\bigcup_{j=1}^{m_{d-1}} B^{\gamma;l}(d,j)\right)\right). \tag{26}$$

Because $l(1),\ldots,l(|J_{(l_1^*,\ldots,l_d^*)}|)$ are ordered such that $(l(1))_d \le \cdots \le (l(|J_{(l_1^*,\ldots,l_d^*)}|))_d$, we obtain that

$$\bigcup_{j=1}^{m_{d-1}} B^{\gamma;l}(d,j) = B^{\gamma;l}(d,1). \tag{27}$$

Plugging (27) into (26), we obtain that

$$B^{\gamma} = \bigcup_{m_1=1}^{|J_{(l_1^*,\ldots,l_d^*)}|} \bigcup_{m_2=1}^{m_1} \cdots \bigcup_{m_{d-1}=1}^{m_{d-2}} \left(\left(\bigcap_{i=1}^{d-1} \Delta B^{\gamma;l}(i,m_i)\right) \cap B^{\gamma;l}(d,1)\right). \quad \square$$

Proof of Theorem 2. For the second moment of Z (under the change of measure), we have that

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}_{\gamma,w}}[Z_n^2] = \mathbb{E}[Z_n] = \mathbb{E}[Z_n \mathbb{1}_{B_n^{\gamma}}] + \mathbb{E}[Z_n \mathbb{1}_{(B_n^{\gamma})^c}] &\le \frac{1}{1-w}\mathbb{P}(A_n \cap B_n^{\gamma})\mathbb{P}(B_n^{\gamma}) + \frac{1}{w}\mathbb{P}(A_n \cap (B_n^{\gamma})^c) \\ &\le \frac{1}{1-w}\mathbb{P}(A_n)\mathbb{P}(B_n^{\gamma}) + \frac{1}{w}\mathbb{P}(A_n \cap (B_n^{\gamma})^c).\end{aligned} \tag{28}$$

Combining this with Lemma 2, we obtain the strong efficiency of our estimator. $\square$

Lemma 2. *Let B^{γ} be as defined in (6). Under Assumption 2, we have that $\mathbb{P}(\bar{X}_n \in B^{\gamma}) = \mathcal{O}(\mathbb{P}(\bar{X}_n \in A))$. Moreover, there exists γ such that $\mathbb{P}(\bar{X}_n \in A \cap (B^{\gamma})^c) = o(\mathbb{P}(\bar{X}_n \in A)^2)$.*

Proof of Lemma 2. First, note that $\mathbb{P}(\bar{X}_n \in B^{\gamma}) = \mathcal{O}(\mathbb{P}(\bar{X}_n \in A))$ follows immediately from Result 2. It remains to show the existence of γ such that $\mathbb{P}(\bar{X}_n \in A \cap (B^{\gamma})^c) = o(\mathbb{P}(\bar{X}_n \in A)^2)$. Because A is bounded away from $\mathbb{D}_{<(l_1^*,\ldots,l_d^*)}$ by assumption, there exists r such that $d\left(A, \mathbb{D}_{<(l_1^*,\ldots,l_d^*)}\right) \ge r$. On the one hand, from Rhee et al. [27], we have that

$$A \subseteq \left\{(\xi^{(1)},\ldots,\xi^{(d)}) \,\middle|\, \exists (l_1,\ldots,l_d) \in J_{(l_1^*,\ldots,l_d^*)} : d(\xi^{(i)}, \mathbb{D}_{<l_i}) \ge r,\ \forall i \in \{1,\ldots,d\}\right\}. \tag{29}$$

On the other hand, we have that

$$(B^{\gamma})^c = \left\{(\xi^{(1)},\ldots,\xi^{(d)}) \,\middle|\, \forall (l_1,\ldots,l_d) \in J_{(l_1^*,\ldots,l_d^*)} : \exists i : \#\{t \,|\, \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \le l_i - 1\right\}. \tag{30}$$

Let $\xi = (\xi^{(1)},\ldots,\xi^{(d)}) \in A \cap (B^{\gamma})^c$ be a step function in the set $\prod_{i=1}^{d}\mathbb{D}_{l_i'}$. By (28), there exists $(l_1,\ldots,l_d) \in J_{(l_1^*,\ldots,l_d^*)}$ such that $\xi^{(i)} = \sum_{j=1}^{l_i+m_i} c_j^{(i)} \mathbb{1}_{[t_j^{(i)},1]}$, $m_i \in \mathbb{Z}_+$, and $d(\xi^{(i)}, \mathbb{D}_{<l_i}) \ge r$ for all $i \in \{1,\ldots,d\}$ with $l_i \ne 0$. Combining $d\left(\xi^{(i)}, \mathbb{D}_{<l_i'}\right) \ge r$ with the fact that $\xi^{(i)} = \sum_{j=1}^{l_i-1} c_j^{(i)} \mathbb{1}_{[t_j^{(i)},1]} \in \mathbb{D}_{<l_i}$, we conclude that

$$\sum_{j=l_i}^{l_i+m_i} c_j^{(i)} \ge d\left(\sum_{j=1}^{l_i+m_i} c_j^{(k)} \mathbb{1}_{[t_j^{(i)},1]}, \sum_{j=1}^{l_i-1} c_j^{(i)} \mathbb{1}_{[t_j^{(i)},1]}\right) \ge r, \tag{31}$$

or in other words, the sum of the m_i+1-smallest jump is bounded from below by r for each $\xi^{(i)}$ of $\{\xi^{(i)}\}_{i\in\{1,\ldots,d\}}$ satisfying $l_i \ne 0$. Combining (30) with (31), as well as choosing γ_k sufficiently small, there exists at least one $k \in \{1,\ldots,d\}$ such that the smallest jump of $\xi^{(k)}$ is bounded from below by $r' > 0$ for an arbitrary but fixed m_k. Repeating the same procedure as described above, we can construct $(m_1,\ldots,m_d)$ for every $(l_1,\ldots,l_d) \in J_{(l_1^*,\ldots,l_d^*)}$ such that the optimization problem given by

$$\underset{(l_1,\ldots,l_d)\in\mathbb{Z}_+^d,\ \prod_{i=1}^{d}\mathbb{D}_{l_i}\cap A\cap(B^{\gamma})^c\ne\emptyset}{\arg\min} \mathcal{I}(l_1,\ldots,l_d) \tag{32}$$

has a unique solution $(l_1^{**}, \dots, l_d^{**})$ satisfying $\mathcal{I}(l_1^{**}, \dots, l_d^{**}) > 2\mathcal{I}(l_1^*, \dots, l_d^*)$. We denote this specific choice of $(m_1, \dots, m_d)$ for every $(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}$ by $\{m^{(l_1, \dots, l_d)}\}_{(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}}$. It should be noted that the existence and the uniqueness of $(l_1^{**}, \dots, l_d^{**})$ can be guaranteed by enlarging the set A (because we are looking for an upper bound for $\mathbb{P}(\bar{X}_n \in A \cap (B^\gamma)^c)$), together with choosing the corresponding γ_i sufficiently small. Therefore, it remains to show that, under the chosen γ, the set $A \cap (B^\gamma)^c$ is bounded away from $\mathbb{D}_{<(l_1^{**}, \dots, l_d^{**})}$. Select ξ satisfying $d(\xi, \mathbb{D}_{<(l_1^{**}, \dots, l_d^{**})}) < \delta$, and hence, there exists $\theta \in \mathbb{D}_{<(l_1^{**}, \dots, l_d^{**})}$ such that $d(\xi, \theta) < \delta$. On the one hand, combining $d(\xi, \theta) < \delta$ with (29), there exists $(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}$ such that $d(\theta^{(i)}, \mathbb{D}_{<l_i}) > r - \delta$ for all $i \in \{1, \dots, d\}$. Hence, we have that $\theta^{(i)} = \sum_{j=1}^{l_i + m_i} c_j^{(i)} \mathbb{1}_{[t_j^{(i)}, 1]}$, $m_i \in \mathbb{Z}_+$, satisfying

$$\sum_{j=l_i}^{l_i+m_i} c_j^{(i)} \geq d\left(\sum_{j=1}^{l_i+m_i} c_j^{(k)} \mathbb{1}_{[t_j^{(i)},1]}, \sum_{j=1}^{l_i-1} c_j^{(i)} \mathbb{1}_{[t_j^{(i)},1]}\right) \geq r - \delta \tag{33}$$

for all $i \in \{1, \dots, d\}$ with $l_i \neq 0$. On the other hand, there exist homeomorphisms $\{\lambda_i\}_{i \in \{1, \dots, d\}}$ such that

$$\| \lambda_i - id \|_\infty \vee \| \theta^{(i)} \circ \lambda_i - \xi^{(i)} \|_\infty < \delta \qquad \text{for all } i \in \{1, \dots, d\}. \tag{34}$$

Combining (34) with (30), we conclude the existence of at least one $i \in \{1, \dots, d\}$ such that

$$\#\{t \mid \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i - \delta\} \leq l_i - 1. \tag{35}$$

Because $\theta \in \mathbb{D}_{<(l_1^{**}, \dots, l_d^{**})}$, we have that

$$m_i \leq (m^{(l_1, \dots, l_d)})_i - 1 \tag{36}$$

for all $i \in \{1, \dots, d\}$ with $l_i \neq 0$. Finally, by (33), (35) and the choice of γ, we conclude that choosing δ sufficiently small leads to a contradiction of (36). □

Proof of Proposition 4. We derive a necessary and sufficient condition for $c^T \phi(\xi)(1) \geq a$ with ξ (18).

For the "only if" part, suppose that $\partial_z(\mathcal{J}) > a$. Let $(v_1, \dots, v_d) \in \mathbb{R}_+^d$, $\delta \in (0, 1)$, and ξ be such that

$$\xi^{(i)}(t) = \begin{cases} -(Rr')_i t, \ t \in [0,1] & \text{for } i \notin \mathcal{J}, \\ v_i \mathbb{1}_{[\delta,1]}(t) - (Rr')_i t, \ t \in [0,1] & \text{for } i \in \mathcal{J}. \end{cases}$$

Obviously, ξ satisfies (18). For $t \in [0, \delta)$, by Result 6, the regulator process $y_\xi \triangleq \psi(\xi)$ should satisfy the fixed point equation that is given by $y_\xi^{(i)}(t) = \max\left\{0, \sup_{s \in [0,t]} \sum_{j \neq i} Q_{ji} y_\xi^{(j)}(s) + (Rr')_i s\right\}$ for all $i \in \{1, \dots, d\}$. Using the fact that $r' > 0$, we obtain that $y_\xi(t) = r't$ for $t \in [0, \delta)$. For $t \in [\delta, 1]$, again by Result 6, it holds that

$$y_\xi^{(i)}(t) = \max\left\{0, -v_i + \sup_{s \in [0,t]} r_i' s + \sum_{j \neq i} Q_{ji} (y_\xi^{(j)}(s) - r_j' s)\right\}, \quad \text{for all } i \in \mathcal{J}, \tag{37}$$

and

$$y_\xi^{(i)}(t) = \max\left\{0, \sup_{s \in [0,t]} r_i' s + \sum_{j \in \mathcal{J}} Q_{ji} (y_\xi^{(j)}(s) - r_j' s) + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}}} Q_{ji} (y_\xi^{(j)}(s) - r_j' s)\right\}, \quad \text{for all } i \notin \mathcal{J}. \tag{38}$$

Because $\{v_i\}_{i \in \mathcal{J}}$ are nonnegative, by Result 5 we conclude that $y_\xi(s)$ and $r_i' s + \sum_{j \neq i} Q_{ji} (y_\xi^{(j)}(s) - r_j' s)$ are continuous in s on $[0, 1]$. Using the Bolzano–Weierstrass theorem, there exists a set of sufficiently large $\{v_i\}_{i \in \mathcal{J}}$ (depending on y_ξ) such that $y_\xi^{(i)}(t) = y_\xi^{(i)}(\delta) = r_i' \delta$ for $i \in \mathcal{J}$. Plugging this into (37) along with setting $y_\xi^{(i)}(t) = r_i' \delta + r_i^*(t - \delta)$ for $i \notin \mathcal{J}$, $t \in [\delta, 1]$, we obtain that

$$\begin{aligned} r_i' \delta + r_i^*(t - \delta) &= \max\left\{0, \sup_{s \in [0,t]} r_i' s + \sum_{j \in \mathcal{J}} Q_{ji} (y_\xi^{(j)}(s) - r_j' s) + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}}} Q_{ji} (y_\xi^{(j)}(s) - r_j' s)\right\} \\ &= \max\left\{r_i' \delta, r_i' \delta + \max_{s \in [\delta,t]} r_i'(s - \delta) - \sum_{j \neq i} Q_{ji} r_j' (s - \delta) + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}}} Q_{ji} r_j^* (s - \delta)\right\} \quad \text{for } i \in \mathcal{J}. \end{aligned} \tag{39}$$

Note that (39) is solved by r_i^* satisfying (19). Moreover, by a straightforward computation, for the workload process $z_\xi \triangleq \phi(\xi)$, we obtain that $c^T z_\xi(1) = \partial_z(\mathcal{J})(1-\delta)$. Because by assumption $\partial_z(\mathcal{J}) > a$, we can choose δ such that $c^T z_\xi(1) \geq a$.

For the other direction of the proof, suppose that $c^T\phi(\xi)(1) \geq a$ for some ξ satisfying (18). Let the jump sizes and the associated jump times of ξ be denoted by $\{u_i\}_{i\in\mathcal{J}}$ and $\{t_i\}_{i\in\mathcal{J}}$, respectively. First we should mention that, by Result 7, enlarging $\{u_i\}_{i\in\mathcal{J}}$ will preserve the fact that $c^T\phi(\xi)(1)\geq a$. Moreover, let $d_1 < \cdots < d_m$ denote the discontinuity points of ξ with $m \leq |\mathcal{J}|$ and define $\mathcal{J}_i \triangleq \{k \mid t_k \leq d_i\}$ for every $i \in \{1,\ldots,m\}$. Now observe that $y_\xi(t) = r't$, $t\in[0,d_1)$. Hence, we have that $z'_\xi(t) = 0 \leq \partial_z(\mathcal{J})$ for $t \in [0,d_1)$. For $y_\xi(t)$, $t\in[d_1,d_2)$, we can easily check that

$$y_\xi^{(i)}(t) = \begin{cases} r'_i d_1, & \text{for all } i \in \mathcal{J}_1, \\ r'_i d_1 + r_i^{*,1}(t-d_1), & \text{for all } i \notin \mathcal{J}_1, \end{cases}$$

by taking sufficiently large $\{u_i\}_{i\in\mathcal{J}_1}$, where $r_i^{*,1} = \max\{r'_i - \sum_{j\neq i} Q_{ji}r'_j + \sum_{\substack{j\neq i\\ j\notin\mathcal{J}_1}} Q_{ji} r_j^{*,1}, 0\}$ for $i \notin \mathcal{J}_1$. Because $\mathcal{J}_1 \subseteq \mathcal{J}$, by Result 7 and (19), we conclude that $z'_\xi(t) \leq \partial_z(\mathcal{J})$ for $t\in[d_1,d_2)$. Defining $\mathcal{J}_1{}' \triangleq \mathcal{J}_1 \cup \{k \mid r_k^{*,1} = 0\}$, we consider $y_\xi(t)$ for $t \in [d_2, d_3)$. Following a similar argument as above, we claim that

$$y_\xi^{(i)}(t) = \begin{cases} r'_i d_1 & \text{for all } i \in \mathcal{J}'_1, \\ r'_i d_1 + r_i^{*,1}(d_2-d_1) & \text{for all } i \in \mathcal{J}_2 \backslash \mathcal{J}'_1, \\ r'_i d_1 + r_i^{*,1}(d_2-d_1) + r_i^{*,2}(t-d_2) & \text{for all } i \notin \mathcal{J}'_1 \cup \mathcal{J}_2, \end{cases}$$

for sufficiently large $\{u_i\}_{i\in\mathcal{J}_1\cup\mathcal{J}_2}$, where $r_i^{*,2} = \max\{r'_i - \sum_{j\neq i} Q_{ji} r'_j + \sum_{j\neq i, j\notin\mathcal{J}'_1\cup\mathcal{J}_2} Q_{ji} r_j^{*,2}, 0\}$ for $i \notin \mathcal{J}'_1 \cup \mathcal{J}_2$. Consider the fixed point equation that is given by

$$\tilde{r}_i^{*,2} = \max\left\{ r'_i - \sum_{j\neq i} Q_{ji} r'_j + \sum_{\substack{j\neq i\\ j\notin\mathcal{J}_1\cup\mathcal{J}_2}} Q_{ji}\tilde{r}_j^{*,2}, 0 \right\}, \quad \text{for } i \notin \mathcal{J}_1 \cup \mathcal{J}_2. \tag{40}$$

Because $\mathcal{J}_1 \subseteq \mathcal{J}_1 \cup \mathcal{J}_2$, by Result 7, we obtain that $\tilde{r}_k^{*,2} = 0$ for every $k \in \mathcal{J}'_1 \backslash \mathcal{J}_1$. By making the convention that $r_k^{*,2} = 0$ for $k \in \mathcal{J}'_1\backslash\mathcal{J}_1$, we claim that $r_i^{*,2} = \tilde{r}_i^{*,2}$ for $i \notin \mathcal{J}_1 \cup \mathcal{J}_2$. Because $\mathcal{J}_1 \cup \mathcal{J}_2 \subseteq \mathcal{J}$, by Result 7, (40), and (19), we conclude that $z'_\xi(t) \leq \partial_z(\mathcal{J})$ for $t \in [d_2, d_3)$. Iterating the same procedure m more times, we can construct a ξ (by taking $\{u_i\}_{i\in\mathcal{J}}$ sufficiently large) such that z_ξ is piecewise linear between neighboring discontinuity points. Moreover, the increasing rate of z_ξ is less than $\partial_z(\mathcal{J})$ (i.e., $z'_\xi(t) \leq \partial_z(\mathcal{J})$ for $t \in [0,1]$). Therefore, we obtain that $\partial_z(\mathcal{J}) > a$. The last statement of Proposition 4 is a consequence of Result 7. □

Proof of Proposition 5. Let the unique optimal solution of (21) be denoted by $(l_1^*, \ldots, l_d^*)$. To prove that A is bounded away from $\bigcup_{(l_1,\ldots,l_d)\in\mathcal{I}_{<l_1^*,\ldots,l_d^*}} \prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)$, it is sufficient to show that $A = F^{-1}[a,\infty)$ is bounded away from $\prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)$ for all $(l_1,\ldots,l_d) \in \mathcal{I}_{<l_1^*,\ldots,l_d^*}$. To begin with, let $(l_1,\ldots,l_d)\in\mathcal{I}_{<l_1^*,\ldots,l_d^*}$. Under Assumption 4, we have that $\partial_z(\mathcal{I}) < a$, where $j \in \mathcal{I}$ if and only if $l_j \neq 0$. Applying a similar approach as in the proof of Proposition 4, it can be shown that $F\left(\prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)\right) \subseteq (-\infty, \partial_z(\mathcal{I})]$. This implies that there exists $\delta > 0$ satisfying

$$d\left(F\left(\prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)\right), [a,\infty)\right) > \delta. \tag{41}$$

Moreover, by Result 4, we conclude that the mapping F as composition of Lipschitz continuous mappings (for continuity of π_1, see, e.g., theorem 12.5 in Billingsley [3]) is again Lipschitz continuous. Let K_F denote the Lipschitz constant of F. Combining this with (41), we conclude that $d\left(\prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i), F^{-1}([a,\infty))\right) > \delta/K_F$; hence the second statement is obtained by applying Result 2. □

References

[1] Asmussen S, Glynn P (2007) *Stochastic Simulation: Algorithms and Analysis*, 1st ed., Stochastic Modelling and Applied Probability, vol. 57. (Springer-Verlag, New York).

[2] Asmussen S, Schmidli H, Schmidt V (1999) Tail probabilities for non-standard risk and queueing processes with subexponential jumps. *Adv. Appl. Probab.* 31(2):422–447.

[3] Billingsley P (2013) *Convergence of Probability Measures*, 2nd ed. (John Wiley & Sons, New York).

[4] Blanchet J, Glynn P (2008) Efficient rare-event simulation for the maximum of heavy-tailed random walks. *Ann. Appl. Probab.* 18(4): 1351–1378.

[5] Blanchet J, Liu J (2008) State-dependent importance sampling for regularly varying random walks. *Adv. Appl. Probab.* 40(4):1104–1128.
[6] Blanchet J, Liu J (2010) Efficient importance sampling in ruin problems for multidimensional regularly varying random walks. *J. Appl. Probab.* 47(2):301–322.
[7] Blanchet J, Glynn P, Liu J (2007) Efficient rare event simulation for heavy-tailed multiserver queues. Technical report, Columbia University, New York.
[8] Blanchet J, Lam H, Tang Q, Yuan Z (2017) Applied robust performance analysis for actuarial applications. Technical report, Columbia University, New York.
[9] Botev ZI, Ridder A, Rojas-Nandayapa L (2016) Semiparametric cross entropy for rare-event simulation. *J. Appl. Probab.* 53(3):633–649.
[10] Boxma OJ, Cahen EJ, Koops D, Mandjes M (2018) Linear networks: Rare-event simulation and Markov modulation. *Methodology Comput. Appl. Probab.*, ePub ahead of print June 4, https://doi.org/10.1007/s11009-018-9644-1.
[11] Dębicki K, Mandjes M (2015) Queues and Lévy fluctuation theory, 1st ed., *Universitext* (Springer International Publishing, Cham, Switzerland).
[12] Dupuis P, Wang H (2004) Importance sampling, large deviations, and differential games. *Stochastics Stochastic Rep.* 76(6):481–508.
[13] Dupuis P, Wang H (2005) On the convergence from discrete to continuous time in an optimal stopping problem. *Ann. Appl. Probab.* 15(2): 1339–1366.
[14] Dupuis P, Wang H (2009) Importance sampling for Jackson networks. *Queueing Systems* 62(1):113–157.
[15] Dupuis P, Leder K, Wang H (2007) Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. Model. Comput. Simulation* 17(3):Article 14.
[16] Dupuis P, Sezer AD, Wang H (2007) Dynamic importance sampling for queueing networks. *Ann. Appl. Probab.* 17(4):1306–1346.
[17] Embrechts P, Klüppelberg C, Mikosch T (1997) *Modelling Extremal Events for Insurance and Finance*, 1st ed., Stochastic Modelling and Applied Probability, vol. 33 (Springer-Verlag, Berlin).
[18] Foss S, Korshunov D (2006) Heavy tails in multi-server queue. *Queueing Systems Theory Appl.* 52(1):31–48.
[19] Foss S, Korshunov D (2012) On large delays in multi-server queues with heavy tails. *Math. Oper. Res.* 37(2):201–218.
[20] Foss S, Korshunov D, Zachary S (2013) *An Introduction to Heavy-Tailed and Subexponential Distributions*, 2nd ed., Springer Series in Operations Research and Financial Engineering, vol. 38 (Springer-Verlag, New York).
[21] Glasserman P, Kou SG (1995) Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Model. Comput. Simulation* 5(1): 22–42.
[22] Glasserman P, Wang Y (1997) Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.* 7(3):731–746.
[23] Gudmundsson T, Hult H (2014) Markov chain Monte Carlo for computing rare-event probabilities for a heavy-tailed random walk. *J. Appl. Probab.* 51(2):359–376.
[24] Harrison JM, Reiman MI (1981) Reflected Brownian motion on an orthant. *Ann. Probab.* 9(2):302–308.
[25] Hu X, Cui H (2010) Generating multi-dimensional discrete distribution random number. *Proc. 2010 6th Internat. Conf. Natl. Comput.*, vol. 3 (IEEE, Piscataway, NJ), 1102–1104.
[26] Ramasubramanian S (2000) A subsidy-surplus model and the Skorokhod problem in an orthant. *Math. Oper. Res.* 25(3):509–538.
[27] Rhee C-H, Blanchet J, Zwart B (2016) Sample path large deviations for heavy-tailed Lévy processes and random walks. Arxiv preprint arXiv:1606.02795v3.
[28] Skorokhod AV (1961) Stochastic equations for diffusion processes in a bounded region. *Theory Probab. Appl.* 6(3):264–274.
[29] Skorokhod AV (1962) Stochastic equations for diffusion processes in a bounded region, II. *Theory Prob. Appl.* 7(1):3–23.
[30] Tankov P, Cont R (2015) *Financial Modelling with Jump Processes*, 2nd ed., Chapman and Hall/CRC Financial Mathematics Series (Taylor & Francis, Boca Raton, FL).
[31] Torrisi G (2004) Simulating the ruin probability of risk processes with delay in claim settlement. *Stochastic Processes Appl.* 112(2):225–244.
[32] Whitt W (2002) *Stochastic-Process Limits*, 1st ed., Springer Series in Operations Research and Financial Engineering (Springer-Verlag, New York).