Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

EFFICIENT SIMULATION FOR LINEAR PROGRAMMING UNDER UNCERTAINTY

Dohyun Ahn
Lewen Zheng

Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong
Shatin, New Territories, HONG KONG

ABSTRACT

We consider a problem of estimating the probability that the optimal value of a stochastic linear program exceeds a large threshold. Inspired by the classical theory of linear programming, we partition the sample space of random components so that the optimal value can be generated without solving a linear program for each sample. This enables us to develop an efficient importance sampling scheme for computing the said probability when the random components are jointly normal. We prove its asymptotic efficiency under the regime where the threshold increases. Our numerical experiments reveal that the proposed method significantly outperforms the existing simulation techniques in the literature.

1 INTRODUCTION

Over the past few decades, the development of optimization models has greatly facilitated operational decision making and risk management. In many related cases, we solve optimization problems designed with prescribed components built on data obtained from various channels. However, some of the underlying data could be highly uncertain, and the estimation of the associated components may not be accurate enough. This is a critical issue since uncertain components often distort the optimization outcomes such as optimal solutions and values, amplifying the uncertainty in those outcomes. To address such an uncertainty issue in optimization, different approaches have been introduced in the literature: relevant examples include stochastic programming (Shapiro et al. 2014), simulation optimization (Fu 2015), and robust optimization (Gabrel et al. 2014). Nevertheless, most of those approaches focus on finding sensible decisions in stochastic environments rather than considering the likelihood of obtaining extreme-yet-possible outcomes.

In this paper, we particularly consider a linear programming problem with random components and develop an efficient simulation scheme for computing the probability that its optimal value exceeds a large threshold. As we shall discuss in Section 4, one representative example of such a problem is the estimation of network instability. In this example, the optimal value of the corresponding linear program represents a measure of network losses, and thus, its tail probability indicates how likely extreme network losses occur. This is applicable to, but not limited to, interbank networks (Eisenberg and Noe 2001; Glasserman and Young 2015) and supply chain networks (Blanchet et al. 2019).

Our problem setting shares a common theme with the literature of stochastic linear programming. Many works in this literature focus on the expectation of the optimal value rather than its distribution or tail probability. For such expected optimal values, upper and lower bounds are derived (Madansky 1960; Avriel and Williams 1970; Morton and Wood 1999), and correlation effects are investigated (Soyster et al. 1984). Simulation methods for the expected optimal values are also developed; see, for example, Parpas et al. (2015) and references therein. Only a few studies work on the distribution of the optimal value by deriving its explicit formulas; see, for example, Bereanu (1963) and Ewbank et al. (1974). However, it is impractical to apply those formulas to the numerical calculation of the tail probabilities due to their computational complexity in high-dimensional circumstances.

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

In terms of considering the interface between simulation and optimization, our work could be compared with the literature of simulation optimization (Fu 2015). However, our work is distinct from that literature: simulation optimization mainly focuses on optimizing simulation outcomes, whereas in this paper, we are interested in estimating the quantity associated with optimization outcomes. It is also worth noting that the aforementioned uncertainty of optimization components differs from the issue of input uncertainty in simulation optimization, which typically means the uncertainty in the input distribution (Zhou and Xie 2015; Lam 2016). This work concentrates on reducing simulation errors when the distribution of the components is a priori known.

In the literature of rare event simulation, estimating the probability that a random quantity exceeds a large threshold has been widely studied. Examples of the random quantity include portfolio losses (Glasserman et al. 2000; Glasserman et al. 2002), the sum of random variables (Juneja and Shahabuddin 2002; Asmussen and Kroese 2006; Chan and Kroese 2011), and the maximum of a random walk (Juneja et al. 2007). Nevertheless, the optimal value of optimization problems with random components has been rarely considered as such a random quantity. To the best of our knowledge, only two papers share the same target probability as our work: Blanchet et al. (2019) and Ahn and Kim (2018). The first paper develops two variance reduction schemes: importance sampling and conditional Monte Carlo. However, the importance sampling scheme is useful only when the associated linear program has a specific form, and both methods require solving at least one linear program for each iteration, leading to much heavier computational burden than ours. The second work converts the target event into a union of a finite number of rare events and proposes an efficient conditional Monte Carlo method, but such a conversion is not always feasible.

The main contribution of this paper is to propose an efficient importance sampling scheme for estimating the probability that the optimal value of a standard linear program with a random right hand side exceeds a certain threshold. This general method can be applied to all specific situations discussed in Blanchet et al. (2019) and Ahn and Kim (2018). We derive a large deviation result on the target probability as the threshold increases and find the relationship between its decay rate and the bases of the linear program. Based on this result, we prove the asymptotic efficiency of our method under the large threshold regime. Numerical results show that our scheme is extremely fast and significantly outperforms those existing methods, especially when the target probability is small.

Before we proceed to the next section, let us introduce the basic notation used throughout the paper. The d-dimensional Euclidean space is denoted by $\mathbb{R}^d$. For any two functions f and g, $f(v) \sim g(v)$ and $f(v) = o(g(v))$ imply $f(v)/g(v) \to 1$ and $f(v)/g(v) \to 0$, respectively, as $v \to \infty$. For any two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$, $\mathbf{v} \geq \mathbf{w}$ means entry-wise inequality. For any $\mathbf{v} \in \mathbb{R}^d$ and a set $A \subset \mathbb{R}^d$, $\mathbf{1}_A(\mathbf{v})$ yields 1 if $\mathbf{v} \in A$ and 0 otherwise. We denote by $\mathbf{0}$ and $\mathbf{1}$ the vectors of zeros and ones, respectively, in a suitable dimension. The matrix $\mathbf{I}$ represents the identity matrix in a suitable dimension.

The rest of the paper is organized as follows. In Section 2, we formalize the main problem of this paper and introduce the classical theory of linear programming from which we decompose the target probability. Section 3 describes our importance sampling scheme based on the probability decomposition, discusses the optimal allocation of the number of simulation trials, and shows the asymptotic efficiency of our scheme. In Section 4, we conduct numerical experiments to demonstrate the effectiveness of the proposed method. The proofs of our theoretical results can be found in the appendix.

2 PROBLEM FORMULATION

2.1 The Main Problem

We consider the following linear program (LP) in standard form:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{1}$$

where an $m \times n$ matrix $\mathbf{A}$ and a vector $\mathbf{c} \in \mathbb{R}^n$ are deterministic and known, $\mathbf{b} \in \mathbb{R}^m$ is a random vector, and $m \leq n$. We assume that the matrix $\mathbf{A}$ has full row rank, and $\mathbf{b}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ which is positive definite. Note that every linear program can be converted into the standard form (Bertsimas and Tsitsiklis 1997). We are interested in the estimation of the probability $p(v)$ that the optimal value of (1) exceeds a large threshold $v > 0$, i.e.,

$$p(v) = \mathsf{P}(\mathbf{c}^\top \mathbf{x}(\mathbf{b}) \geq v),$$

where $\mathbf{x}(\mathbf{b})$ is the optimal solution to the problem (1). We further assume that at least one component in $\mathbf{c}$ is positive; otherwise, we have $p(v) = 0$. Also, the probability that the problem (1) is infeasible is assumed to be negligible so that $p(v)$ approaches 0 as v increases. Although the probability of infeasibility could be large in some cases, this paper concerns more interesting cases where it is significantly dominated by the probability of feasibility even when v is sufficiently large, motivated by practical applications (Ahn and Kim 2018; Blanchet et al. 2019).

The estimation of $p(v)$ requires numerical integral procedures over a high-dimensional region with a large number of grid points (Ewbank et al. 1974), which are very challenging to conduct in practice. Alternatively, one might use the naive Monte Carlo method in which for each sample of $\mathbf{b}$, we solve (1) to obtain the optimal value and check if the optimal value is greater than v. However, two major issues arise when using the naive Monte Carlo method in our problem. Firstly, large computational costs are inevitable due to the fact that a linear program needs to be solved each time a sample of $\mathbf{b}$ is generated. Solving a linear program often results in significant computational burden unless the problem size is very limited. Secondly, inefficiency and low accuracy are aggravated as the threshold v increases. Since the optimal value rarely exceeds large v, a huge number of simulation trials are required to estimate $p(v)$, and the standard deviation of the naive Monte Carlo estimator is often much greater than the estimate, which is common in rare event simulations; see, e.g., Rubino and Tuffin (2009) for a good review. In this paper, we propose a new Monte Carlo scheme for $p(v)$ that addresses both of the issues raised above, demonstrating its asymptotic efficiency and effectiveness via theoretical and numerical results, respectively.

Remark 1 Although our distributional assumption on $\mathbf{b}$ is seemingly restrictive, as pointed out by Bucklew (2004), the normality assumption has been commonly used in the literature of stochastic simulation due to its practical applicability. Furthermore, our method is applicable to the cases where $\mathbf{b}$ has a normal mean–variance mixture distribution by conditioning on a mixing variable. Note that a wide family of generalized hyperbolic distributions including multivariate t, variance-gamma, and normal inverse Gaussian distributions belongs to the class of normal mean–variance mixtures (McNeil et al. 2015).

2.2 Theoretical Motivation from the Classical Linear Programming Theory

Given a linear program in standard form (1), let B be a subset of $\{1,\ldots,n\}$ with m elements, and denote by $\mathbf{A}_B$ and $\mathbf{c}_B$ the matrix made of m columns of $\mathbf{A}$ indexed by B and the vector composed of $\{c_i\}_{i\in B}$, respectively. We call B a basis of (1) if $\mathbf{A}_B^{-1}$ exists. We say that a basis B is optimal if $\mathbf{A}_B^{-1}\mathbf{b} \geq \mathbf{0}$ (primal feasibility) and $\mathbf{c}^\top \geq \mathbf{c}_B^\top \mathbf{A}_B^{-1}\mathbf{A}$ (dual feasibility). Note that the dual feasibility of a basis is independent of $\mathbf{b}$. If such an optimal basis B is found, the optimal solution $\mathbf{x}(\mathbf{b})$ exists and is written in closed form as

$$\mathbf{x}_B(\mathbf{b}) = \mathbf{A}_B^{-1}\mathbf{b},\ \mathbf{x}_{B^{\mathrm{c}}}(\mathbf{b}) = \mathbf{0}, \tag{2}$$

where $\mathbf{x}_B(\mathbf{b})$ and $\mathbf{x}_{B^{\mathrm{c}}}(\mathbf{b})$ are the vectors obtained by restricting the entries of $\mathbf{x}(\mathbf{b})$ to the index sets B and $B^{\mathrm{c}} = \{1,\cdots,n\} \setminus B$, respectively. See Bertsimas and Tsitsiklis (1997) for more details. For each basis B, we define its critical region by $\{\mathbf{b} \in \mathbb{R}^m : \mathbf{A}_B^{-1}\mathbf{b} \geq \mathbf{0}\}$. Then, according to Gal and Nedoma (1972) and Chapter 15.2 of Prékopa (1995), there exists a finite collection $\mathscr{B}$ of dual feasible bases such that the space of feasible $\mathbf{b}$ is partitioned by the corresponding critical regions. This implies that if $\mathbf{b}$ follows a continuous distribution, we have the following decomposition of $p(v)$ based on (2):

$$p(v) = \sum_{B\in\mathscr{B}} p_B(v), \tag{3}$$

Algorithm 1 Finding Partitioning Bases

Set $\mathscr{B} = \emptyset$ and let Q be a queue consisting of an initial dual feasible basis B_0.
while $|\mathsf{Q}| > 0$ **do**
  Pop out the first basis B in Q and set $\mathscr{B} = \mathscr{B} \cup \{B\}$, $\overline{\mathbf{A}} = \mathbf{A}_B^{-1}\mathbf{A}$, and $\overline{\mathbf{c}}^\top = \mathbf{c}^\top - \mathbf{c}_B^\top \overline{\mathbf{A}}$.
  for $l = 1, \cdots, m$ **do**
    Set $\mathbf{u} = \overline{\mathbf{A}}^l$, the l-th row of $\overline{\mathbf{A}}$.
    if $u_i < 0$ for some $i \in \{1, \cdots, n\}$ **then**
      Set $\overline{B} = B$, $j = \arg\min_{i:u_i<0} \frac{\overline{c}_i}{|u_i|}$, and $\overline{B}(l) = j$.
      if $\overline{B} \notin \mathscr{B} \cup \mathsf{Q}$ **then**
        Place $\overline{B}$ at the end of the queue Q.
      end if
    end if
  end for
end while
return $\mathscr{B}$.

where $p_B(v) = \mathsf{P}(\mathbf{c}_B^\top \mathbf{A}_B^{-1}\mathbf{b} \geq v, \mathbf{A}_B^{-1}\mathbf{b} \geq \mathbf{0})$. An important observation from (3) is that $p(v)$ can be estimated without solving LPs for each realization of $\mathbf{b}$, which motivates us to focus on variance reduction in estimating $p_B(v)$ for each basis $B \in \mathscr{B}$ in the next section. We shall also consider an optimal allocation of the total number of simulation trials to the bases so that the overall estimation error could be further reduced.

Remark 2 Gal and Nedoma (1972) show that starting from one dual feasible basis B_0, by a dual simplex pivot, one can get another dual feasible basis B_1 of which corresponding critical region is disjoint from that of B_0. By repeating this procedure, we can obtain all bases that form the collection $\mathscr{B}$. The initial dual feasible basis B_0 could be identified by a standard technique; see Chapter 4 of Luenberger and Ye (2016). The complete procedure is given in Algorithm 1. In the worst case, the running time of Algorithm 1 could grow exponentially in m or n; see Section 5.2 of Luenberger and Ye (2016). However, since $\mathscr{B}$ is independent of the threshold v, once it is identified via Algorithm 1, we can flexibly use the collection for estimating $p(v)$ for different v. In Section 4, we numerically observe that the computational cost of Algorithm 1 is negligible when $m = 10$ and $n = 20$. Note that it is beyond the scope of this paper to develop an efficient scheme of finding dual feasible bases that significantly affect the probability for large-scale linear programs.

3 EFFICIENT IMPORTANCE SAMPLING

In this section, we first develop an effective and tractable importance sampling (IS) method for the estimation of $p_B(v)$ that minimizes the asymptotic second moment of the estimator under the regime where v goes to infinity. We then show that our proposed estimator $Z^*(v)$ for $p(v)$ is logarithmically efficient, i.e.,

$$\liminf_{v\to\infty} \frac{\log \mathsf{Var}(Z^*(v))}{\log p(v)^2} \geq 1;$$

see Asmussen and Glynn (2007) for more details on the efficiency notions. The logarithmic efficiency implies that the decay rate of the estimate with respect to v corresponds to that of the standard deviation of the estimator. This allows us to expect significant variance reduction in estimating $p(v)$ via the proposed method compared to the Monte Carlo method when v is large enough.

Before we describe our IS scheme, we present two large deviations results on the probabilities $p_B(v)$ and $p(v)$, respectively. To facilitate later discussions, we let $\Theta_B^v = \{\mathbf{z} \in \mathbb{R}^m | \mathbf{c}_B^\top \mathbf{A}_B^{-1}\mathbf{z} \geq v, \mathbf{A}_B^{-1}\mathbf{z} \geq \mathbf{0}\}$ so that $p_B(v) = \mathsf{P}(\mathbf{b} \in \Theta_B^v)$. One can easily see that $p_B(v) = 0$ if and only if $\Theta_B^1 = \emptyset$. We also define $Q(\mathbf{u}) = \mathbf{u}^\top \mathbf{\Sigma}^{-1}\mathbf{u}/2$ for $\mathbf{u} \in \mathbb{R}^m$ and denote by q_B the optimal value of the following quadratic program

$$\min_{\mathbf{z}\in\Theta_B^1} Q(\mathbf{z}), \tag{4}$$

where by convention we set $q_B = \infty$ if $\Theta_B^1 = \emptyset$.

Theorem 1 For each $B \in \mathscr{B}$, $\lim_{v\to\infty} v^{-2} \log p_B(v) = -q_B$, if $\Theta_B^1 \neq \emptyset$.

One implication from Theorem 1 is that for any two bases $B_1, B_2 \in \mathscr{B}$ satisfying $q_{B_1} > q_{B_2}$ and $\Theta_{B_1}^1, \Theta_{B_2}^1 \neq \emptyset$, the ratio of $p_{B_1}(v)$ to $p_{B_2}(v)$ can be approximated as $p_{B_1}(v)/p_{B_2}(v) \approx \exp(-v^2(q_{B_1} - q_{B_2}))$, which exponentially decays to 0 as $v \to \infty$, meaning that $p_{B_1}(v)$ is significantly smaller than $p_{B_2}(v)$ for large v. Therefore, when v is large, our target probability $p(v)$ is highly dominated by the probabilities $p_B(v)$ associated with minimal q_B, which is formalized in the next theorem.

Theorem 2 $\lim_{v\to\infty} v^{-2} \log p(v) = -\min_{B\in\mathscr{B}} q_B$.

3.1 Efficient IS for Estimating $p_B(v)$

In this subsection, as alluded to earlier, we devise an efficient simulation scheme for estimating $p_B(v)$ for fixed $B \in \mathscr{B}$. We basically consider changing the mean $\boldsymbol{\mu}$ in order to generate more samples in Θ_B^v, which is referred to as mean shifting; see, for example, Asmussen and Glynn (2007) for more discussions on this scheme. Under the mean shifting from $\boldsymbol{\mu}$ to a new point $\widetilde{\boldsymbol{\mu}}$, the unbiased IS estimator $Z_B(\widetilde{\boldsymbol{\mu}}, v)$ for $p_B(v)$ can be written as

$$Z_B(\widetilde{\boldsymbol{\mu}}, v) = \exp\left(-Q(\mathbf{Z} - \boldsymbol{\mu}) + Q(\mathbf{Z} - \widetilde{\boldsymbol{\mu}})\right) \cdot \mathbf{1}_{\Theta_B^v}(\mathbf{Z}), \tag{5}$$

where $\mathbf{Z}$ is a normal random vector with mean $\widetilde{\boldsymbol{\mu}}$ and covariance matrix $\boldsymbol{\Sigma}$.

Using the Laplace principle in large deviations theory, one can observe that the second moment of the IS estimator (5), denoted by $M_2(\widetilde{\boldsymbol{\mu}}, v)$, has the following asymptotic relationship:

$$\log M_2(\widetilde{\boldsymbol{\mu}}, v) - \kappa = \log \int_{\Theta_B^v} \exp(-2Q(\mathbf{z} - \boldsymbol{\mu}) + Q(\mathbf{z} - \widetilde{\boldsymbol{\mu}}))\mathrm{d}\mathbf{z} \sim \max_{\mathbf{z}\in\Theta_B^v} \left\{-2Q(\mathbf{z} - \boldsymbol{\mu}) + Q(\mathbf{z} - \widetilde{\boldsymbol{\mu}})\right\},$$

for some constant κ. Following a standard approach (e.g., Glasserman et al. (2000)), we find the optimal mean shift that asymptotically minimizes the second moment of (5), i.e., we solve

$$\min_{\widetilde{\boldsymbol{\mu}}\in\mathbb{R}^m} \max_{\mathbf{z}\in\Theta_B^v} \left\{-2Q(\mathbf{z} - \boldsymbol{\mu}) + Q(\mathbf{z} - \widetilde{\boldsymbol{\mu}})\right\}. \tag{6}$$

From the minimax theorem (e.g., Du and Pardalos (1995)), for any T sufficiently large, we have

$$\min_{\widetilde{\boldsymbol{\mu}}\in\Omega_T} \max_{\mathbf{z}\in\Theta_B^v\cap\Omega_T} \left\{-2Q(\mathbf{z} - \boldsymbol{\mu}) + Q(\mathbf{z} - \widetilde{\boldsymbol{\mu}})\right\} = \max_{\mathbf{z}\in\Theta_B^v\cap\Omega_T} \min_{\widetilde{\boldsymbol{\mu}}\in\Omega_T} \left\{-2Q(\mathbf{z} - \boldsymbol{\mu}) + Q(\mathbf{z} - \widetilde{\boldsymbol{\mu}})\right\}, \tag{7}$$

where $\Omega_T = [-T, T]^m$. The inner minimization in the right hand side of (7) is achieved at $\widetilde{\boldsymbol{\mu}} = \mathbf{z}$. Thus, by letting $T \to \infty$, one can easily see that solving (6) is equivalent to solving

$$q_B^*(v) = \min_{\mathbf{z}\in\Theta_B^v} Q(\mathbf{z} - \boldsymbol{\mu}), \tag{8}$$

and we denote its optimal solution by $\boldsymbol{\mu}_B^*$ to which we shift the mean. We write $Z_B^*(v) = Z_B(\boldsymbol{\mu}_B^*, v)$ for ease of exposition. The asymptotically optimal mean shifting can be interpreted as changing the mean to the point with the greatest density in Θ_B^v so that the event Θ_B^v could occur more frequently under the new distribution. Note that since $\boldsymbol{\mu} \notin \Theta_B^v$ for large v, it is likely that the point $\boldsymbol{\mu}_B^*$ is at the boundary of Θ_B^v.

Theorem 3 For each $B \in \mathscr{B}$, $\lim_{v\to\infty} v^{-2} \log M_2(\boldsymbol{\mu}_B^*, v) = -2q_B$ if $\Theta_B^1 \neq \emptyset$.

The above theorem provides a large deviations result on the second moment $M_2(\boldsymbol{\mu}_B^*, v)$ of the IS estimator under the asymptotically optimal mean shifting. According to this result, we observe that the decay rate of the second moment $M_2(\boldsymbol{\mu}_B^*, v)$ is twice as large as that of the probability $p(v)$. This implies that the second moment decays at the maximum achievable rate, which is straightforward from Jensen's inequality. In other words, $Z_B^*(v)$ is a logarithmically efficient estimator for $p_B(v)$.

3.2 The Aggregate IS Estimator

In this subsection, we first discuss the problem of optimally allocating the total number of simulation trials, denoted by N, to the bases in $\mathscr{B}$ so that for each $B \in \mathscr{B}$, $p_B(v)$ can be estimated using N_B simulation trials allocated to B. We then provide a logarithmically efficient estimator for our target probability $p(v)$ based on the optimal allocation of the number of simulation trials and the results in Section 3.1. Given any unbiased estimator $Z_B(v)$ for $p_B(v)$, we can construct the estimator $Z(v)$ for $p(v)$ as $Z(v) = \sum_{B\in\mathscr{B}} N_B^{-1} \sum_{i=1}^{N_B} Z_{B,i}(v)$, where $\{Z_{B,i}(v)\}_{i=1,\cdots,N_B}$ are i.i.d. realizations of $Z_B(v)$ for each $B \in \mathscr{B}$. This estimator is unbiased if $N_B \geq 1$ for all $B \in \mathscr{B}$, and its variance is given by $\mathsf{Var}(Z(v)) = \sum_{B\in\mathscr{B}} N_B^{-1} \mathsf{Var}(Z_B(v))$. Then, using $N = \sum_{B\in\mathscr{B}} N_B$ and the Cauchy-Schwarz inequality, we observe that the following allocation of the total number of simulation trials minimizes $\mathsf{Var}(Z(v))$:

$$N_B = N \cdot \frac{\sqrt{\mathsf{Var}(Z_B(v))}}{\sum_{B\in\mathscr{B}} \sqrt{\mathsf{Var}(Z_B(v))}}. \tag{9}$$

Note that we ignore minor technicalities associated with N_B not being a positive integer.

We finally define our *aggregate IS estimator* $Z^*(v)$ as follows:

$$Z^*(v) = \sum_{B\in\mathscr{B}} \frac{1}{N_B^*} \sum_{i=1}^{N_B^*} Z_{B,i}^*(v), \tag{10}$$

where N_B^* is the number of simulation trials (9) for $B \in \mathscr{B}$ calculated by using the estimator $Z_B^*(v)$. Our estimator thus reduces estimation error in two ways: importance sampling for $p_B(v)$ and the optimal allocation of the number of simulation trials. The asymptotic efficiency of the estimator is demonstrated in the following theorem.

Theorem 4 The aggregate IS estimator $Z^*(v)$ is logarithmically efficient.

In practice, it is challenging to compute N_B^* since $\{\mathsf{Var}(Z_B^*(v))\}_{B\in\mathscr{B}}$ are unknown. Alternatively, as mentioned in Chapter V.7 of Asmussen and Glynn (2007), one may adaptively allocate samples based on the sample variance updated at each step or estimate them in advance using a pilot run. However, both of them require additional computational costs. To avoid such an issue, in this paper, we propose to use the approximation $\mathsf{Var}(Z_B^*(v)) \approx \exp(-2q_B^*(v))$, which results from Theorems 1 and 3 and Lemma 1 in the appendix. This approximation leads to the the number of simulation trials N_B^* proportional to $N\exp(-q_B^*(v))$ by (9). Algorithm 2 describes our proposed method given the collection $\mathscr{B}$ identified by Algorithm 1.

In Figure 1, we numerically observe that the said approximation works well for the example in Section 4. For $v = 1,2,3$, we compute $q_B^*(v)$ and use 10^5 simulation trials to estimate $\mathsf{Var}(Z_B^*(v))$ for each $B \in \mathscr{B}$. The blue dots represent the points $(2q_B^*(v), -\log\mathsf{Var}(Z_B^*(v)))$ for different B, while the red line is the identity line. In all cases we consider, the blue dots are mostly parallel to the red line, meaning that $-\log\mathsf{Var}(Z_B^*(v)) \approx 2q_B^*(v) + \beta(v)$, where for fixed v, $\beta(v)$ is a constant independent of B. However, such a constant does not affect the optimal allocation since it cancels out in the ratio (9).

4 NUMERICAL EXPERIMENTS

Linear programming models have been widely used to describe networked systems that address the distribution of resources, e.g., electricity and water, and the spread of negative impacts, e.g., economic shocks and epidemics. In this section, we particularly consider the following linear program that has been applied to equilibrium modeling in various networks:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^m} \quad & \mathbf{1}^\top \mathbf{x} \\ \text{s.t.} \quad & (\mathbf{I} - \mathbf{\Pi})^\top \mathbf{x} \geq \boldsymbol{\eta},\ \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{11}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)^\top$ represents a vector composed of excess resources (or impacts) generated at each node, $\mathbf{\Pi}$ is an $m \times m$ matrix consisting of $\{\pi_{ij}\}_{i,j=1,\ldots,m}$, and π_{ij} represents the proportion of excess resources

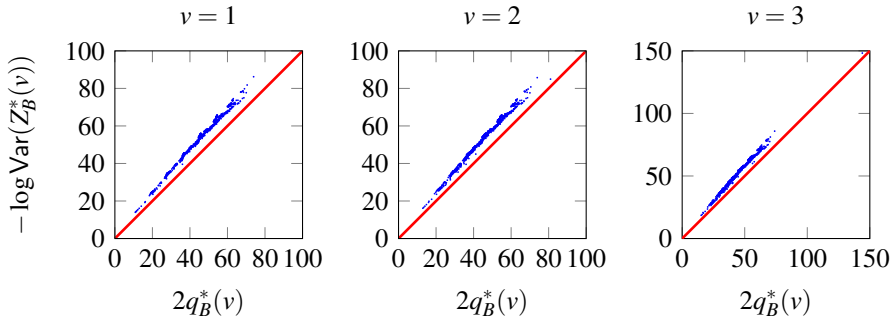

Figure 1: The quality of the approximation $\mathsf{Var}(Z_B^*(v)) \approx \exp(-2q_B^*(v))$.

at node i transferred to node j. An important underlying assumption of the above model is that the resources are distributed according to the pro rata allocation rule. The associated applications include risk allocation in financial networks (Eisenberg and Noe 2001; Glasserman and Young 2015) and commodity allocation in distribution networks (Blanchet et al. 2019). In the former, the goal is to minimize the total shortfall in interbank payments, whereas the latter aims to minimize the total excess demand in the network.

The linear program formulation (11) can be rewritten in standard form by introducing a redundant vector $\mathbf{y}$, and we denote its optimal value by $h(\boldsymbol{\eta})$, i.e., $h(\boldsymbol{\eta}) = \min\{\mathbf{1}^\top \mathbf{x} \mid (\mathbf{I} - \boldsymbol{\Pi})^\top \mathbf{x} - \mathbf{y} = \boldsymbol{\eta},\ \mathbf{x}, \mathbf{y} \geq \mathbf{0}\}$. Assuming that $\boldsymbol{\eta}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$, we are concerned with the estimation of the probability that $h(\boldsymbol{\eta})$ exceeds a certain threshold v, i.e., $\mathsf{P}(h(\boldsymbol{\eta}) \geq v)$. This probability can be interpreted as systemic risk in the case of financial networks and a failure probability in the case of distribution networks. Recent papers including Ahn and Kim (2018) and Blanchet et al. (2019) also address this issue. In particular, assuming that the random vector has a multivariate elliptical distribution, Ahn and Kim (2018) propose a conditional Monte Carlo (CMC) method by geometrically interpreting the target event $\{h(\boldsymbol{\eta}) \geq v\}$ as a union of half-spaces identified by the extreme points of the feasible set. Blanchet et al. (2019) provide importance sampling and conditional Monte Carlo schemes under the assumption that the random vector is jointly normal. The former shows the asymptotic efficiency of their method under a regime in which the size of the random vector diminishes to zero, whereas the latter considers a regime where its mean vector decreases to negative infinity. Although those asymptotic regimes are different from ours, the three regimes share the same spirit in that all of them make the target event rarer.

Table 1 compares our proposed method with five different schemes including the naive Monte Carlo, the aggregate IS with the uniform allocation of the number of simulation trials, and those developed in Ahn

Algorithm 2 Aggregate Importance Sampling

for $B \in \mathscr{B}$ **do**
  Compute $\boldsymbol{\mu}_B^*$ and $q_B^*(v)$ by solving (8).
end for
for $B \in \mathscr{B}$ **do**
  Set $N_B^* = N \exp(-q_B^*(v)) \cdot (\sum_{B \in \mathscr{B}} \exp(-q_B^*(v)))^{-1}$.
  for $i = 1, \ldots, N_B^*$ **do**
    Sample $\mathbf{Z} = \mathbf{z}$ from the multivariate normal distribution with mean $\boldsymbol{\mu}_B^*$ and covariance matrix $\boldsymbol{\Sigma}$.
    Set $Z_{B,i}^*(v) = \exp\left(-Q(\mathbf{z} - \boldsymbol{\mu}) + Q(\mathbf{z} - \boldsymbol{\mu}_B^*)\right) \cdot \mathbf{1}_{\Theta_B^v}(\mathbf{z})$.
  end for
end for
Compute $Z^*(v)$ by (10).
return $Z^*(v)$.

Table 1: Estimates of the probability $\mathsf{P}(h(\boldsymbol{\eta}) \geq v)$ with $v = 1,2,3$ based on six different simulation methods.

v	Method	Estimate	Standard Error	Time (sec)	VR	ER
1	Naive Monte Carlo	1.91×10^{-3}	1.38×10^{-4}	516	-	-
	Aggregate IS	2.10×10^{-3}	1.39×10^{-5}	0.7	99	73,804
	Aggregate IS (unif. alloc.)	2.09×10^{-3}	1.64×10^{-4}	0.8	0.7	449
	CMC (Ahn and Kim 2018)	2.06×10^{-3}	3.34×10^{-5}	4.8	17	1,856
	IS (Blanchet et al. 2019)	2.10×10^{-3}	1.54×10^{-5}	527	81	79
	CMC (Blanchet et al. 2019)	2.13×10^{-3}	3.39×10^{-5}	2,291	17	3.7
2	Naive Monte Carlo	5.40×10^{-4}	7.35×10^{-5}	520	-	-
	Aggregate IS	4.81×10^{-4}	3.49×10^{-6}	0.8	444	302,919
	Aggregate IS (unif. alloc.)	4.88×10^{-4}	4.58×10^{-5}	0.8	2.6	1,778
	CMC (Ahn and Kim 2018)	4.78×10^{-4}	1.12×10^{-5}	4.7	43	4,686
	IS (Blanchet et al. 2019)	4.80×10^{-4}	7.85×10^{-6}	527	88	86
	CMC (Blanchet et al. 2019)	4.97×10^{-4}	1.18×10^{-5}	2,132	39	10
3	Naive Monte Carlo	1.30×10^{-4}	3.61×10^{-5}	506	-	-
	Aggregate IS	1.17×10^{-4}	9.48×10^{-7}	0.7	1,447	1,037,690
	Aggregate IS (unif. alloc.)	1.03×10^{-4}	1.08×10^{-5}	0.7	11	7,932
	CMC (Ahn and Kim 2018)	1.11×10^{-4}	3.69×10^{-6}	4.7	95	10,347
	IS (Blanchet et al. 2019)	1.16×10^{-4}	3.91×10^{-6}	523	85	82
	CMC (Blanchet et al. 2019)	1.21×10^{-4}	4.02×10^{-6}	1,990	80	20

and Kim (2018) and Blanchet et al. (2019). For each algorithm, we use 10^5 simulation trials to compute $\mathsf{P}(h(\boldsymbol{\eta}) \geq v)$ for different values of v. We set $m = 10$ and $\boldsymbol{\Pi} = 0.1(\mathbf{e}_m, \mathbf{e}_1, \cdots, \mathbf{e}_{m-1}) + 0.9(\mathbf{J} - \mathbf{I})/(m-1)$, where $\mathbf{e}_i$ is the i-th column of $\mathbf{I}$, and $\mathbf{J}$ is the $m \times m$ matrix of ones. We also assume that $\eta_1, \ldots, \eta_m$ are independent, and for each i, η_i has mean $\mu_i = -i$ and variance $\sigma_{ii} = (i/3)^2$. Using Algorithm 1, we identify 1,023 dual feasible bases in 0.3 seconds. The extreme points required to implement the CMC method of Ahn and Kim (2018) can simply be obtained from those dual feasible bases. Note that these computation times are found to be negligible and thus are not included in the results in Table 1. For each result, we use two efficiency measures: variance ratio (VR) and efficiency ratio (ER). VR is the ratio of the variance of the naive Monte Carlo estimator to that of the new estimator, which captures the effect of the new estimator on variance reduction. ER is obtained by multiplying VR by the ratio of the running time of the naive Monte Carlo scheme to that of the new scheme, which describes the practical effectiveness of the new scheme compared to the naive Monte Carlo method.

According to the VR and ER estimates in the table, we observe that our proposed method thoroughly dominates the other methods in all instances, and its performance improves as v increases, i.e., the target event becomes rarer. To be more specific, the aggregate IS provides significant variance reduction compared to the other methods. This might result from the asymptotically optimal mean shifting that minimizes the approximate second moment of the estimator, which the schemes of Ahn and Kim (2018) and Blanchet et al. (2019) do not consider. Also, by comparing the aggregate IS schemes with optimal allocation and uniform allocation, we find that the optimal allocation greatly contributes to variance reduction. Furthermore, the proposed method substantially reduces computational costs. Recall that our algorithm does not require solving LPs and its computational time is not affected by the number of dual feasible bases. Although the method of Ahn and Kim (2018) can also be implemented without solving LPs, it is slightly slower than ours because its computational time depends on the number of extreme points as well as the number of simulation trials. In the naive Monte Carlo and the schemes of Blanchet et al. (2019), at least one linear program needs to be solved for each iteration, which results in a huge computational burden.

5 CONCLUDING REMARKS

We found this article not without any limitations. We hence provide two other interesting topics that could be investigated for further research. Firstly, it would be challenging but practically important to devise an efficient simulation method for more complicated cases where the matrix $\mathbf{A}$ is stochastic. Our current work does not cover such cases, but may help initiate the related discussion. Secondly, the distributional assumption could be relaxed. In addition to applying the current approach to normal mean–variance mixtures (Remark 1), one might consider an extension to the case where the distribution of the random components is not a priori known and thus could be misspecified. Despite those limitations, we hope that this work sheds new light on the combination of rare–event simulation and optimization theory.

ACKNOWLEDGEMENTS

The authors would like to thank the three anonymous referees for their thoughtful comments and valuable suggestions which helped them improve the manuscript. This work was supported by the Hong Kong Research Grant Council Early Career Scheme (24210420) and CUHK Direct Grants (4055101 and 4055134).

A PROOFS OF THEORETICAL RESULTS

Proof of Theorem 1. Let $\mathbf{z}^*$ denote the optimal solution of (4). We note that the Karush–Kuhn–Tucker conditions for (4) can be written as follows:

$$\mathbf{z}^* = \mathbf{\Sigma}\mathbf{A}_B^{-\top}(\boldsymbol{\lambda} + \gamma \mathbf{c}_B), \;\; (\boldsymbol{\lambda} + \gamma \mathbf{c}_B)^\top \mathbf{A}_B^{-1}\mathbf{z}^* = \gamma, \tag{12}$$

for some $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\gamma \geq 0$ in $\mathbb{R}^m$ and $\mathbb{R}$, respectively. Let $s_{\max}$ denote the largest eigenvalue of $\mathbf{\Sigma}$. Then, for any $\mathbf{z} \in \Theta_B^1$, we have

$$Q(\mathbf{z}) - Q(\mathbf{z}^*) = Q(\mathbf{z}-\mathbf{z}^*) + \boldsymbol{\lambda}^\top \mathbf{A}_B^{-1}\mathbf{z} + \gamma(\mathbf{c}_B^\top \mathbf{A}_B^{-1}\mathbf{z} - 1) \geq Q(\mathbf{z}-\mathbf{z}^*) \geq \frac{||\mathbf{z}-\mathbf{z}^*||^2}{2s_{\max}}, \tag{13}$$

where the equality is from (12), the first inequality holds since $\mathbf{z} \in \Theta_B^1$, $\boldsymbol{\lambda} \geq \mathbf{0}$, and $\gamma \geq 0$, and the second inequality results from the Courant–Fischer theorem; see, e.g., Theorem 4.2.6 of Horn and Johnson (2013). Fix $B \in \mathscr{B}$, by a change of variable, $p_B(v)$ can be recast in the following integral form:

$$p_B(v) = \frac{v^m \exp(-v^2 q_B)}{\sqrt{(2\pi)^m |\det(\mathbf{\Sigma})|}} \int_{\Theta_B^1} \exp\left(-v^2 \left(Q(\mathbf{z} - \boldsymbol{\mu}/v) - q_B\right)\right) \mathrm{d}\mathbf{z}. \tag{14}$$

Since $q_B = Q(\mathbf{z}^*)$, by (13), the exponent in (14) can be rewritten as

$$\begin{aligned} -v^2\left(Q(\mathbf{z} - \boldsymbol{\mu}/v) - q_B\right) &= -v^2\big(Q(\mathbf{z}-\mathbf{z}^*) + \boldsymbol{\lambda}^\top \mathbf{A}_B^{-1}\mathbf{z} + \gamma(\mathbf{c}_B^\top \mathbf{A}_B^{-1}\mathbf{z} - 1)\big) - Q(\boldsymbol{\mu}) + v\mathbf{z}^\top \mathbf{\Sigma}^{-1}\boldsymbol{\mu} \qquad (15) \\ &\leq -v^2 ||\mathbf{z}-\mathbf{z}^*||^2/(2s_{\max}) + v\mathbf{z}^\top \mathbf{\Sigma}^{-1}\boldsymbol{\mu}. \qquad (16) \end{aligned}$$

Using (16) and changing the integral region to $\mathbb{R}^m$, we have $p_B(v) \leq \exp(-v^2 q_B + o(v^2))$, and hence, $\limsup_{v\to\infty} v^{-2}\log p_B(v) \leq -q_B$.

We next claim that $\liminf_{v\to\infty} v^{-2}\log p_B(v) \geq -q_B$. By letting $\mathbf{y} = \mathbf{A}_B^{-1}\mathbf{z}$ and $\mathbf{y}^* = \mathbf{A}_B^{-1}\mathbf{z}^*$, we obtain

$$p_B(v) = \widetilde{\kappa} v^m \exp(-v^2 q_B) \int_{\widetilde{\Theta}_B^1} \exp\left(-v^2\left(Q(\mathbf{A}_B(\mathbf{y}-\mathbf{y}^*)) + \boldsymbol{\lambda}^\top \mathbf{y} + \gamma(\mathbf{c}_B^\top \mathbf{y} - 1)\right) + v\boldsymbol{\mu}^\top \mathbf{\Sigma}^{-1}\mathbf{A}_B\mathbf{y}\right) \mathrm{d}\mathbf{y},$$

where $\widetilde{\kappa} = |\det(\mathbf{A}_B)|\exp(-Q(\boldsymbol{\mu}))(2\pi)^{-m/2}|\det(\mathbf{\Sigma})|^{-1/2}$ is a constant, and $\widetilde{\Theta}_B^1 = \{\mathbf{y} \in \mathbb{R}^m | \mathbf{c}_B^\top \mathbf{y} \geq 1, \mathbf{y} \geq \mathbf{0}\}$. Let $I = \{i \in B \mid c_i > 0\}$, $J = \{i \in B \mid c_i < 0\}$, and $K = \{i \in B \mid c_i = 0\}$. Note that $I \neq \emptyset$ since $\Theta_B^1 \neq \emptyset$. Define $\delta = (\sum_{i\in I} c_i)/(2\sum_{j\in J}|c_j| + \varepsilon)$ for some constant $\varepsilon > 0$. Then, we have

$$\left\{\mathbf{y} \in \mathbb{R}^m \middle| y_i - y_i^* \geq v^{-1/2}, \delta v^{-1/2} \leq y_j - y_j^* \leq 2\delta v^{-1/2} \;\text{ for all } i \in I \cup K, j \in J\right\} \subset \widetilde{\Theta}_B^1, \tag{17}$$

where y_i is the i-th component of a vector $\mathbf{y}$. By Theorem 4.2.6 of Horn and Johnson (2013), we have $2Q(\mathbf{A}_B(\mathbf{y}-\mathbf{y}^*)) \le \alpha_{\max}||\mathbf{y}-\mathbf{y}^*||^2$, where $\alpha_{\max} > 0$ is the largest eigenvalue of $\mathbf{A}_B^\top \boldsymbol{\Sigma}^{-1}\mathbf{A}_B$. Note that the existence and positiveness of $\alpha_{\max}$ are guaranteed since $\mathbf{A}_B^\top \boldsymbol{\Sigma}^{-1}\mathbf{A}_B$ is positive definite. Define the following probability density functions

$$f_i(y) = \frac{\exp(-v^2\alpha_{\max}(y-y_i^*)^2/2)I\{y \ge y_i^* + v^{-1/2}\}}{\int_{y_i^*+v^{-1/2}}^{\infty}\exp(-v^2\alpha_{\max}(y-y_i^*)^2/2)\mathrm{d}y} \text{ for } i \in I\cup K, \text{ and}$$

$$f_j(y) = \frac{\exp(-v^2\alpha_{\max}(y-y_j^*)^2/2)I\{y_j^* + \delta v^{-1/2} \le y \le y_j^* + 2\delta v^{-1/2}\}}{\int_{y_j^*+\delta v^{-1/2}}^{y_j^*+2\delta v^{-1/2}}\exp(-v^2\alpha_{\max}(y-y_j^*)^2/2)\mathrm{d}y} \text{ for all } j \in J.$$

By (17) and Jensen's inequality, for large v, we thus obtain the following asymptotic relationship:

$$\begin{aligned}
\widetilde{\kappa}^{-1}v^{-m}\exp(v^2 q_B)p_B(v) \ge & \prod_{i\in I\cup K}\int_{y_i^*+v^{-1/2}}^{\infty} e^{-v^2\alpha_{\max}(y_i-y_i^*)^2/2}\mathrm{d}y_i \cdot \prod_{j\in J}\int_{y_j^*+\delta v^{-1/2}}^{y_j^*+2\delta v^{-1/2}} e^{-v^2\alpha_{\max}(y_j-y_j^*)^2/2}\mathrm{d}y_j \\
& \cdot e^{\gamma v^2}\prod_{i\in B}\int_{-\infty}^{\infty} e^{-v^2(\lambda_i y_i + \gamma c_i y_i) + v u_i y_i} f_i(y_i)\mathrm{d}y_i \\
\ge & \left(\frac{2\pi}{v^2\alpha_{\max}}\right)^{m/2}\overline{\Phi}(\sqrt{v\alpha_{\max}})^{|I\cup K|}\left(\overline{\Phi}(\delta\sqrt{v\alpha_{\max}}) - \overline{\Phi}(2\delta\sqrt{v\alpha_{\max}})\right)^{|J|} \\
& \cdot \exp\left(-v^2\left(\boldsymbol{\lambda}^\top\overline{\mathbf{y}} + \gamma(\mathbf{c}_B^\top\overline{\mathbf{y}} - 1)\right) + v\mathbf{u}^\top\overline{\mathbf{y}}\right) \\
\sim & \frac{\exp\left(-v^2\left(\boldsymbol{\lambda}^\top\overline{\mathbf{y}} + \gamma(\mathbf{c}_B^\top\overline{\mathbf{y}} - 1)\right) - v\left(\alpha_{\max}(\delta^2|J| + |I\cup K|)/2 - \mathbf{u}^\top\overline{\mathbf{y}}\right)\right)}{\delta^{|J|}v^{3m/2}\alpha_{\max}^m}, \qquad (18)
\end{aligned}$$

where $\mathbf{u}^\top = \boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{A}_B$, $\overline{y}_i = \int y f_i(y)\mathrm{d}y$ for each $i \in B$, $\overline{\Phi}(\cdot)$ is the tail distribution of the standard normal distribution, and (18) follows from the so-called Mills ratio (Feller 1968, Chapter 7): $\overline{\Phi}(t) \sim (t\sqrt{2\pi})^{-1}e^{-t^2/2}$ as $t \to \infty$. Using $\overline{y}_i - y_i^* = \int (y-y_i^*)f_i(y)\mathrm{d}y$ for $i \in B$, the Mills ratio also implies that

$$\begin{aligned}
\overline{y}_i - y_i^* &= \frac{\exp(-v\alpha_{\max}/2)}{\overline{\Phi}(\sqrt{v\alpha_{\max}})\sqrt{2\pi v^2\alpha_{\max}}} \sim v^{-1/2} \text{ for } i \in I\cup K, \text{ and} \\
\overline{y}_j - y_j^* &= \frac{\exp(-\delta^2 v\alpha_{\max}/2) - \exp(-2\delta^2 v\alpha_{\max})}{\left(\overline{\Phi}(\delta\sqrt{v\alpha_{\max}}) - \overline{\Phi}(2\delta\sqrt{v\alpha_{\max}})\right)\sqrt{2\pi v^2\alpha_{\max}}} \sim \delta v^{-1/2} \text{ for } j \in J. \qquad (19)
\end{aligned}$$

From (12), we know $(\boldsymbol{\lambda} + \gamma\mathbf{c}_B)^\top\mathbf{y}^* = \gamma$, and hence, according to (19), the logarithm of (18) is equal to

$$-v^2(\boldsymbol{\lambda} + \gamma\mathbf{c}_B)^\top(\overline{\mathbf{y}} - \mathbf{y}^*) - v\left(\frac{\alpha_{\max}(\delta^2|J| + |I\cup K|)}{2} - \mathbf{u}^\top\overline{\mathbf{y}}\right) + \log(\delta^{|J|}v^{3m/2}\alpha_{\max}^m) = o(v^2).$$

This completes the proof. □

Proof of Theorem 2. By Theorem 1, for each $B \in \mathscr{B}$, $p_B(v) = \exp(-q_B v^2 + o(v^2))$ as $v \to \infty$ if $\Theta_B^1 \neq \emptyset$. Let $\mathscr{B}^* = \mathrm{argmin}_{B\in\mathscr{B}} q_B$ and $q^* = \min_{B\in\mathscr{B}} q_B$. Then, $q^* = q_B$ for all $B \in \mathscr{B}^*$, and we observe that as $v \to \infty$,

$$\begin{aligned}
p(v) &= \sum_{B\in\mathscr{B}^*}\exp(-q^*v^2 + o(v^2)) + \sum_{B\notin\mathscr{B}^*, \Theta_B^1\neq\emptyset}\exp(-q_B v^2 + o(v^2)) \\
&\le |\mathscr{B}^*|\exp(-q^*v^2 + o(v^2)) + \sum_{B\notin\mathscr{B}^*, \Theta_B^1\neq\emptyset}\exp(-q_B v^2 + o(v^2)) \sim |\mathscr{B}^*|\exp(-q^*v^2),
\end{aligned}$$

where $|\mathscr{B}^*|$ denotes the number of bases in $\mathscr{B}^*$. Hence, $\limsup_{v\to\infty} v^{-2}\log p(v) \le -q^*$ since $f(v) \sim g(v)$ and $g(v) \to 0$ implies $\log f(v) \sim \log g(v)$. Also, we have $p(v) \ge p_B(v) = \exp(-q^*v^2 + o(v^2))$ for any $B \in \mathscr{B}^*$. □

Lemma 1 Let $\mathbf{z}^*$ denote the optimal solution of (4). If $\mathbf{z}_v^* = \text{argmin}_{\mathbf{z}\in\Theta_B^1} Q(\mathbf{z}-\boldsymbol{\mu}/v)$ for all $v>0$, then $\mathbf{z}_v^* \to \mathbf{z}^*$ as $v\to\infty$.

Proof. Consider the difference function $D(\mathbf{z}) = Q(\mathbf{z}-\boldsymbol{\mu}/v) - Q(\mathbf{z})$. Then, by the Cauchy-Schwarz inequality, $|D(\mathbf{z})-D(\mathbf{z}')| = v^{-1}|(\mathbf{z}-\mathbf{z}')^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}| \le v^{-1}||\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}||\cdot||\mathbf{z}-\mathbf{z}'||$, for any $\mathbf{z},\mathbf{z}'\in\Theta_B^1$. Combining with (13), according to Proposition 4.32 of Bonnans and Shapiro (2000), $||\mathbf{z}_v^*-\mathbf{z}^*|| \le 2v^{-1}s_{\max}||\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}||$. □

Proof of Theorem 3. Fix $B\in\mathscr{B}$ such that Θ_B^1 is nonempty. We note that $M_2(\boldsymbol{\mu}_B^*,v)\ge p_B(v)^2$ implies $\liminf_{v\to\infty} v^{-2}\log M_2(\boldsymbol{\mu}_B^*,v) \ge -2q_B$. Thus, it is enough to show that $\limsup_{v\to\infty} v^{-2}\log M_2(\boldsymbol{\mu}_B^*,v)\le -2q_B$. Recall Lemma 1 and observe that $\boldsymbol{\mu}_B^* = v\mathbf{z}_v^*$. By a change of variable, $M_2(\boldsymbol{\mu}_B^*,v)$ can be recast as

$$M_2(\boldsymbol{\mu}_B^*,v) = \frac{v^m}{\sqrt{(2\pi)^m|\det(\boldsymbol{\Sigma})|}}\int_{\Theta_B^1}\exp\left(-2v^2Q(\mathbf{z}-\boldsymbol{\mu}/v)+v^2Q(\mathbf{z}-\mathbf{z}_v^*)\right)\mathrm{d}\mathbf{z}. \tag{20}$$

Using (15) and (13), the exponent of the integrand in (20) has an upper bound:

$$\begin{aligned}
&-2v^2Q(\mathbf{z}-\boldsymbol{\mu}/v)+v^2Q(\mathbf{z}-\mathbf{z}_v^*)\\
&= -v^2\left(2q_B+2Q(\mathbf{z}-\mathbf{z}^*)+2\boldsymbol{\lambda}^\top\mathbf{A}_B^{-1}\mathbf{z}+2\gamma(\mathbf{c}_B^{-1}\mathbf{A}_B^{-1}\mathbf{z}-1)+2Q(\boldsymbol{\mu}/v)-2\mathbf{z}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/v-Q(\mathbf{z}-\mathbf{z}_v^*)\right)\\
&\le -v^2(2q_B+2Q(\mathbf{z}-\mathbf{z}^*)-Q(\mathbf{z}-\mathbf{z}_v^*))+2v\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{z}\\
&\le -v^2(3q_B-Q(\mathbf{z}_v^*)+||\mathbf{z}-\mathbf{z}^*||^2/(2s_{\max}))+v(2\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{z}+\mathbf{z}^\top\boldsymbol{\Sigma}^{-1}\mathbf{s}_v),
\end{aligned}$$

where $\mathbf{s}_v = v(\mathbf{z}^*-\mathbf{z}_v^*)$ is a bounded vector since $||\mathbf{z}_v^*-\mathbf{z}^*||\le 2v^{-1}s_{\max}||\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}||$ from the proof of Lemma 1. Therefore, by relaxing the integral region and noticing $Q(\mathbf{z}_v^*)\to q_B$ as $v\to\infty$ by Lemma 1, we have

$$M_2(\boldsymbol{\mu}_B^*,v) \le \frac{v^m e^{-v^2(3q_B-Q(\mathbf{z}_v^*))}}{\sqrt{(2\pi)^m|\det(\boldsymbol{\Sigma})|}}\int e^{-v^2||\mathbf{z}-\mathbf{z}^*||^2/(2s_{\max})+v(2\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{z}+\mathbf{z}^\top\boldsymbol{\Sigma}^{-1}\mathbf{s}_v)}\mathrm{d}\mathbf{z} = \exp(-2v^2q_B+o(v^2)).$$

This leads to $\limsup_{v\to\infty} v^{-2}\log M_2(\boldsymbol{\mu}_B^*,v)\le -2q_B$. The proof thus completes. □

Proof of Theorem 4. From (9), we observe the following relationship for $\mathsf{Var}(Z^*(v))^{1/2}$:

$$\mathsf{Var}(Z^*(v))^{1/2} = N^{-1/2}\sum_{B\in\mathscr{B}}\mathsf{Var}(Z_B^*(v))^{1/2} \le \sum_{B\in\mathscr{B}^*}M_2(\boldsymbol{\mu}_B^*,v)^{1/2}+\sum_{B\notin\mathscr{B}^*,\Theta_B^1\neq\emptyset}M_2(\boldsymbol{\mu}_B^*,v)^{1/2}.$$

By Theorem 3, $M_2(\boldsymbol{\mu}_B^*,v)^{1/2} = \exp(-q_Bv^2+o(v^2))$ if $\Theta_B^1\neq\emptyset$. Then, the result follows using the same approach in the proof of Theorem 2. □

REFERENCES

Ahn, D., and K.-K. Kim. 2018. "Efficient Simulation for Expectations over the Union of Half-Spaces". *ACM Transactions on Modeling and Computer Simulation* 28(3):Article 23.

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. New York: Springer-Verlag.

Asmussen, S., and D. P. Kroese. 2006. "Improved Algorithms For Rare Event Simulation With Heavy Tails". *Advances in Applied Probability* 38(2):545–558.

Avriel, M., and A. C. Williams. 1970. "The Value of Information and Stochastic Programming". *Operations Research* 18(5):947–954.

Bereanu, B. 1963. "On Stochastic Linear Programming I. Distribution Problems: A Single Random Variable". *Revue Roumaine De Mathématiques Pures et Appliquées* 8(4):683–697.

Bertsimas, D., and J. N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Belmont, MA: Athena Scientific.

Blanchet, J., J. Li, and M. K. Nakayama. 2019. "Rare-Event Simulation for Distribution Networks". *Operations Research* 67(5):1383–1396.

Bonnans, J. F., and A. Shapiro. 2000. *Perturbation Analysis of Optimization Problems*. New York: Springer.
Bucklew, J. A. 2004. *Introduction to Rare Event Simulation*. New York: Springer-Verlag.
Chan, J. C. C., and D. P. Kroese. 2011. "Rare-Event Probability Estimation with Conditional Monte Carlo". *Annals of Operations Research* 189:43–61.
Du, D.-Z., and P. M. Pardalos. 1995. *Minimax and Applications*. Dordrecht: Kluwer Academic Publishers.
Eisenberg, L., and T. H. Noe. 2001. "Systemic Risk in Financial Systems". *Management Science* 47(2):236–249.
Ewbank, J. B., B. L. Foote, and H. J. Kumin. 1974. "A Method for the Solution of the Distribution Problem of Stochastic Linear Programming". *SIAM Journal on Applied Mathematics* 26(2):225–238.
Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*. 3rd ed, Volume I. New York: John Wiley & Sons, Inc.
Fu, M. C. 2015. *Handbook of Simulation Optimization*. New York: Springer.
Gabrel, V., C. Murat, and A. Thiele. 2014. "Recent Advances in Robust Optimization: An Overview". *European Journal of Operational Research* 235(3):471–483.
Gal, T., and J. Nedoma. 1972. "Multiparametric Linear Programming". *Management Science* 18(7):406–422.
Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000. "Variance Reduction Techniques for Estimating Value-at-Risk". *Management Science* 46(10):1349–1364.
Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2002. "Portfolio Value-at-Risk with Heavy-Tailed Risk Factors". *Mathematical Finance* 12(3):239–269.
Glasserman, P., and H. P. Young. 2015. "How Likely is Contagion in Financial Networks?". *Journal of Banking & Finance* 50:383–399.
Horn, R. A., and C. R. Johnson. 2013. *Matrix Analysis*. 2nd ed. New York: Cambridge University Press.
Juneja, S., R. L. Karandikar, and P. Shahabuddin. 2007. "Asymptotics and Fast Simulation for Tail Probabilities of Maximum of Sums of Few Random Variables". *ACM Transactions on Modeling and Computer Simulation* 17(2):Article 7.
Juneja, S., and P. Shahabuddin. 2002. "Simulating Heavy Tailed Processes Using Delayed Hazard Rate Twisting". *ACM Transactions on Modeling and Computer Simulation* 12(2):94–118.
Lam, H. 2016. "Advanced tutorial: input uncertainty and robust analysis in stochastic simulation". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, , and S. E. Chick, 178–192. Arlington, Virginia: Institute of Electrical and Electronics Engineers, Inc.
Luenberger, D. G., and Y. Ye. 2016. *Linear and Nonlinear Programming*. 4th ed. Cham: Springer.
Madansky, A. 1960. "Inequalities for Stochastic Linear Programming Problems". *Management Science* 6(2):197–204.
McNeil, A. J., R. Frey, and P. Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised ed. Princeton and Oxford: Princeton University Press.
Morton, D. P., and R. K. Wood. 1999. "Restricted-Recourse Bounds for Stochastic Linear Programming". *Operations Research* 47(6):943–956.
Parpas, P., B. Ustun, M. Webster, and Q. K. Tran. 2015. "Importance Sampling in Stochastic Programming: A Markov Chain Monte Carlo Approach". *INFORMS Journal on Computing* 27(2):358–377.
Prékopa, A. 1995. *Stochastic Programming*. Netherlands: Springer.
Rubino, G., and B. Tuffin. 2009. *Rare Event Simulation using Monte Carlo Methods*. Chichester: John Wiley & Sons, Ltd.
Shapiro, A., D. Dentcheva, and A. Ruszczyński. 2014. *Lectures on Stochastic Programming: Modeling and Theory*. 2nd ed. Philadelphia: SIAM.
Soyster, A. L., R. D. Foley, and F. H. Murphy. 1984. "Technical Note–A Class of Stochastic Mathematical Programs with Correlated Scale Parameters in the Objective and Right-Hand Side". *Operations Research* 32(4):945–951.
Zhou, E., and W. Xie. 2015. "Simulation Optimization when facing Input Uncertainty". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 3714–3724. Huntington Beach, CA: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

DOHYUN AHN is an assistant professor in the Department of Systems Engineering and Engineering Management at The Chinese University of Hong Kong. He received his B.S., M.S., and Ph.D. degrees in Industrial & Systems Engineering, all from KAIST. His research interests include, but are not limited to, quantitative risk management, stochastic simulation, decision making via uncertainty quantification, and network analysis in finance and operations. His email address is dohyun.ahn@cuhk.edu.hk.

LEWEN ZHENG is a Ph.D. candidate in the Department of Systems Engineering and Engineering Management at The Chinese University of Hong Kong. He received his B.S. degree from the School of Mathematics at Sun Yat-sen University in Guangzhou, China. His current research interest is in stochastic simulation; in particular rare event simulation and the applications. His email address is lwzheng@se.cuhk.edu.hk.