Submitted to *Operations Research*
manuscript (Please, provide the manuscript number!)

Rare-Event Simulation for Distribution Networks

Jose Blanchet
Department of Industrial Engineering and Operations Research, Department of Statistics, Columbia University, New York, NY 10027, jose.blanchet@columbia.edu

Juan Li
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, jl3035@columbia.edu

Marvin K. Nakayama
Department of Computer Science, New Jersey Institute of Technology, University Heights Newark, NJ 07102, marvin@njit.edu

We model equilibrium allocations in a distribution network as the solution of a linear program (LP) which minimizes the cost of unserved demands across nodes in the network. The constraints in the LP dictate that once a given node's supply is exhausted, its unserved demand is distributed among neighboring nodes. All nodes do the same and the resulting solution is the equilibrium allocation. Assuming that the demands are random (following a jointly Gaussian law), our goal is to study the probability that the optimal cost (i.e. the cost of unserved demands in equilibrium) exceeds a large threshold, which is a rare event. Our contribution is the development of importance sampling and conditional Monte Carlo algorithms for estimating this probability. We establish the asymptotic efficiency of our algorithms and also present numerical results which illustrate strong performance of our procedures.

Key words: distribution network; linear program; rare event simulation; importance sampling; conditional Monte Carlo

1. Introduction

Consider the following model of a distribution network. We assume that there is a commodity to be distributed among various nodes in a network. Each node is endowed with a given supply of the commodity and at the same time it experiences a random demand. We assume that the commodity is infinitely divisible. If the demand at a given node exceeds its supply, then the excess demand is distributed according to some proportions to each of its neighbors, which in turn do the same. In order to obtain the distribution amounts in equilibrium, we solve a linear program (LP), where the objective function to minimize is the sum across nodes of the unserved demands.

One possible practical example where such a problem might arise is an electric power grid. Here, the commodity is electricity, and each node represents a geographic region. Each region has generators, which provide the region's supply of electricity. Also, each region has a random load (i.e., demand for electricity). Regions are connected by transmission lines, and if a region's load exceeds its supply, then the network tries to serve a node's excess load by sending it to neighboring regions. If the total amount of load not served at their originating regions exceeds a threshold k, then we consider the network to have failed.

Another application involves load distribution for internet services, such as web servers, cloud-computing services, and domain name servers (DNS). A company may have a number of fixed-capacity servers situated in different geographic regions. As the requests to servers (i.e. the demand) arrive, a specific server tries to fulfill its own local requests, but if the demand exceeds its capacity, then the server may offload its excess to a neighboring server. Since this shifting may incur additional delays for the user, we want to minimize the amount of distributed demand. This is similar to load balancing; e.g., see Kopparapu (2002).

Let $\alpha(k)$ be the probability that the sum of unserved demands, in equilibrium, exceeds threshold k. Our goal is to estimate the probability $\alpha(k)$, especially in the case in which k is large. Assuming jointly distributed multivariate Gaussian demands, we provide asymptotically optimal estimators, together with numerical experiments showing their performance, and associated large deviations results. We recall that an unbiased estimator for $\alpha(k)$ is asymptotically optimal if the logarithm of its second moment is asymptotically equivalent to the logarithm of $\alpha(k)^2$ (see Asmussen and Glynn 2007, for notions of efficiency in rare-event simulation).

As far as we know, this paper provides the first type of large deviations analysis and efficient Monte Carlo for solutions of linear programs with random input. More precisely, our contributions are as follows:

1) For our model formulation, we show that our optimal allocation is invariant if one replaces the objective function by any other criterion that is increasing as a function of the unserved demands (see Theorem 3).

2) We establish large deviations estimates for our class of linear programs with random input (see Theorem 4).

3) We develop an importance sampling (IS) algorithm for estimating $\alpha(k)$, and we show that the algorithm is asymptotically optimal as the threshold $k \to \infty$ (see Section 5.2).

4) We develop a conditional Monte Carlo (CMC) algorithm for the evaluation of $\alpha(k)$, and we prove the asymptotic optimality of this procedure as $k \to \infty$ (see Section 5.3).

5) We provide several numerical examples in Section 6 that validate the performance of our algorithm.

Some of the results regarding CMC previously appeared in a conference version of this paper (Blanchet, Li and Nakayama 2011). Our conference paper restricted the LP's objective function to be the sum of the unserved demands, and we now prove its invariance, as described in contribution 1), which greatly expands the applicability of our approach. Regarding contribution 2), we study the asymptotic behaviors of this network which is not discussed in the conference version. Regarding contribution 3), we develop an importance sampling algorithm which is not studied in the conference version, and we provide a proof of optimality and algorithm implementation. As for contribution 4), although in the conference version, we have studied the CMC algorithm and its implementation (see Blanchet, Li and Nakayama 2011, Section 4.3), no mathematical proof is provided regarding the optimality of this algorithm. Here, in the journal version, we prove it rigorously. Finally, regarding contribution 5), instead of only comparing the naive simulation and CMC, we compare IS as well. In addition, to make the numerical experiments more complete, we add two more examples in which the threshold is fixed while the rarity parameter changes.

We now explain how our paper relates to prior work. First, regarding 1), we note that similar results, with different types of networks and other applications, have been obtained in the literature (see Eisenberg and Noe 2001). We only learned about these applications after we obtained our model formulation, but we believe the connections are relevant. For the IS algorithm (contribution 2), we introduce a probability measure that is obtained by connecting the event of interest (i.e. total unserved demands in equilibrium exceeding a threshold) with a simple union event involving the demands. Then we use an IS distribution inspired by an approach developed by Adler, Blanchet and Liu (2012). Regarding the CMC estimator, we express the Gaussian demands in polar coordinates. Given the angle, the conditional probability of the LP's optimal objective function value exceeding k can be expressed as the probability of the radial component of the Gaussian lying in an interval or union of intervals, and this conditional probability can be computed analytically. We prove the asymptotic optimality of these two methods using the theory of excursions of Gaussian random fields (Adler and Taylor 2007). The use of polar transformations for CMC and rare event simulation has been used in the past, see for example, Asmussen et al. (2011). Asmussen and Glynn (2007), Chapters V and VI, provide additional background material on importance sampling and conditional Monte Carlo.

Our work also has other potential applications, in particular to cascading failures. For example, Watts (2002) studies cascades in a sparse, random network of interacting agents whose decisions are determined by the actions of their neighbors according to a simple threshold rule. Dobson et al. (2007) consider a branching process model of cascading failures in an electric power grid. Iyer, Nakayama and Gerbessiotis (2009) analyze a continuous-time Markov chain of a dependability model with cascading failures.

We would like to point out that although we assume multivariate Gaussian demands in this paper, the CMC algorithm can be applied to the case when the demands follow elliptical distribution (see McNeil, Frey and Embrechts 2005). Further more, while elliptical copula exhibits symmetric tail dependence, the well known Archimedean copula allows asymmetric tail dependence (Brechmann, Hendrich and Czado 2013). Making use the results in McNeil and Neslehova (2009), we can see that CMC algorithm is also applicable to Archimedean copula, which makes this algorithm very powerful in solving a wide range of problems; additional details on the application to Archimedean copulas are given in our last section on final comments.

The rest of the paper develops as follows. Section 2 presents the model of the distribution network, and it also defines the LP problem and its dual. We establish some properties of the primal and dual LPs in Section 3. The asymptotic behavior of the model is discussed in Section 4. We describe the asymptotic optimality and implementations of importance sampling and conditional Monte Carlo methods for estimating $\alpha(k)$ in Section 5. Section 6 contains the experimental results from some examples, and we give some final comments in Section 7.

2. Model Description

As we introduce our model and discuss its properties we will follow closely the discussion in Blanchet, Li and Nakayama (2011). Suppose there is a directed graph $G=(V,E)$, where $V=\{1,2,\ldots,d\}$ is the set of vertices and $E=\{(i,j):\exists \text{ directed edge from vertex } i \text{ to vertex } j\}$ is the set of edges. The incidence matrix of the graph is denoted by $H=(H(i,j):i,j\in V)$, where $H(i,j)=1$ if $(i,j)\in E$, and $H(i,j)=0$ otherwise, and we assume $H(i,i)=0$ for any $i\in V$. The network model we consider is induced by this graph, and we also assume the following:

1 The network is irreducible in the sense that the matrix H is irreducible.

2 Each node i has a given fixed supply s_i.

3 Each node i is subjected to a random demand D_i. The demand vector $\boldsymbol{D}=(D_1,D_2\ldots,D_d)'$ is jointly Gaussian $N(\boldsymbol{\mu},\Sigma)$, where prime denotes transpose.

4 The expectation of D_i is less than or equal to s_i for each node i.

Each node tries to serve its realized demand. However, if a given nodes supply is exhausted, it distributes the unserved demand to its neighbors, which, in turn, do the same with their respective neighbors. Nevertheless, there is a cost associated with transferring unserved demands which should be minimized. We construct a linear program to describe this problem. The demands achieve an equilibrium point at each feasible solution, and the objective function is to minimizing the sum of the excess demands across the nodes. Let $\boldsymbol{s}=(s_1,s_2,\ldots,s_d)'$, and the LP is:

$$\min \quad \sum_{i=1}^{d} x_i^+$$

$$\text{s.t.}\quad D_i - s_i + \sum_{j:(j,i)\in E} x_j^+ a_{ji} = x_i^+ - x_i^-, \forall i$$

$$x_i^+ \geq 0, x_i^- \geq 0, \forall i. \tag{1}$$

The quantity $x_i^+ \geq 0$ represents the shedded demand from node i in equilibrium, which is distributed among its neighbors using a fixed distribution scheme, which we describe shortly. The quantity $x_i^- \geq 0$ represents the unused supply at node i in equilibrium. Therefore, in equilibrium, if $x_i^+ - x_i^- > 0$, then node i sheds demand; if $x_i^+ - x_i^- < 0$, then node i has unused supply. When node j has excess demand, a_{ji} denotes the proportion of unserved demand at node j distributed to node i. We assume that if $H(i,j) = 0$, then $a_{ij} = 0$; if $H(i,j) = 1$, then $a_{ij} > 0$. In addition, $\sum_{j=1}^{d} a_{ij} = 1, \forall i = 1, 2, \ldots, d$. The solution moves around excess demands and supplies to neighbors but does so in such a way that the sum of x_i^+'s, which are the equilibrium shedded demands, is minimized. The problem can be expressed in matrix notation as follows. Define $A(i,j) = a_{ij}$ (note that $A(i,i) = 0$). Let $\mathbf{1} = (1, 1, \ldots, 1)'$ denote the d-dimensional column vector with all components equal to 1. Then the previous linear programming problem (1) can be written as:

$$\begin{aligned} \min \quad & \mathbf{1}'\boldsymbol{x}^+ + \mathbf{0}'\boldsymbol{x}^- \\ \text{s.t.} \quad & (A' - I)\boldsymbol{x}^+ + I\boldsymbol{x}^- = \boldsymbol{s} - \boldsymbol{D} \\ & \boldsymbol{x}^+ \geq \mathbf{0}, \boldsymbol{x}^- \geq \mathbf{0}, \end{aligned} \tag{2}$$

where $\mathbf{0} = (0, 0, \ldots, 0)'$ is the d-dimensional column vector with all components equal to 0, $A = (A(i,j) : i, j \in V)$, I is the $d \times d$ identity matrix, $\boldsymbol{x}^+ = (x_1^+, x_2^+, \ldots, x_d^+)'$, and $\boldsymbol{x}^- = (x_1^-, x_2^-, \ldots, x_d^-)'$. The goal is that the sum of shedded demands is as small as possible because, e.g., the cost of distributing demands is high. If the cost is too high, for example, larger than a given number, say k, or the LP is infeasible, we consider the network to have failed.

Now, we also introduce the dual linear program:

$$\begin{aligned} \max \quad & \boldsymbol{y}'\boldsymbol{r} \\ \text{s.t.} \quad & M\boldsymbol{y} \leq \mathbf{1} \\ & \boldsymbol{y} \geq \mathbf{0}, \end{aligned} \tag{3}$$

where $M = I - A$ and $r = \boldsymbol{D} - \boldsymbol{s}$.

Note that while in Blanchet, Li and Nakayama (2011), we assume that the unserved demands are equally distributed to neighbors, here we make a small but important extension. We allow the proportions to be any non-negative numbers.

We are interested in computing the probability that the network fails, for different values of k. Let $\alpha(k)$ represent this failure probability. As discussed in Blanchet, Li and Nakayama (2011),

$$\alpha(k) = \beta_0 + \beta_1(k) = P\{L(\boldsymbol{D}) > k\}, \tag{4}$$

where β_0 is the probability that the primal is infeasible, and $\beta_1(k)$ is the probability that the primal is feasible, but the cost is larger than k. $L(\boldsymbol{D})$ denotes the optimal value of the dual when the demand vector is $\boldsymbol{D}$.

3. Properties of Our Primal and Dual Linear Programs

3.1. Feasibility of the Solutions to the Primal and Dual

Our previous conference paper proves two theorems on properties of the primal and dual LPs for the special case when $A(i,j) = H(i,j)/\sum_{l=1}^{d} H(i.l)$. We claim that both theorems are still valid for our more general $A(i,j)$, and the proofs are exactly the same. Here we only list the property regarding feasibility which will be used later, but omit the proof.

Theorem 1.

(a) The dual problem (3) is always feasible.

(b) The primal problem (2) is feasible if and only if $\sum_{i=1}^{d} D_i \le \sum_{i=1}^{d} s_i$.

3.2. Uniqueness and Positivity of the Solution to the Primal

Theorem 2. *The primal problem (2) has the following properties:*

(a) It has a unique optimal solution.

(b) At the optimal solution, at most one element in the pair (x_k^+, x_k^-) *is strictly positive,* $\forall 1 \le k \le d$.

Proof: Suppose both $\boldsymbol{x}_1 = \begin{pmatrix} \boldsymbol{x}_1^+ \\ \boldsymbol{x}_1^- \end{pmatrix}$ and $\boldsymbol{x}_2 = \begin{pmatrix} \boldsymbol{x}_2^+ \\ \boldsymbol{x}_2^- \end{pmatrix}$ are optimal solutions. Let $\boldsymbol{d}^* = \boldsymbol{x}_1 - \boldsymbol{x}_2 = \begin{pmatrix} \boldsymbol{x}_1^+ - \boldsymbol{x}_2^+ \\ \boldsymbol{x}_1^- - \boldsymbol{x}_2^- \end{pmatrix} = \begin{pmatrix} \boldsymbol{d}^{*+} \\ \boldsymbol{d}^{*-} \end{pmatrix}$, which is of dimension $2d$. We want to prove that $\boldsymbol{d}^* = 0$. Consider the following linear program:

$$
\begin{aligned}
(P) \qquad \min \quad & \mathbf{0}'\boldsymbol{d} \\
\text{s.t.} \quad & \mathbf{1}'\boldsymbol{d}^+ = 0 \\
& (A' - I)\boldsymbol{d}^+ + I\boldsymbol{d}^- = \mathbf{0} \\
& \boldsymbol{d} \geq \boldsymbol{e}_j,
\end{aligned}
$$

where $\boldsymbol{e}_j$ is a $2d$-dimensional vector with the jth element equal to 1 and other elements equal to 0. Equivalently, we write the LP (P) as

$$
\begin{aligned}
\min \quad & \mathbf{0}'\boldsymbol{d} \\
\text{s.t.} \quad & B\boldsymbol{d} = 0 \qquad (\alpha) \\
& \boldsymbol{d} \geq \boldsymbol{e}_j, \qquad (\beta)
\end{aligned}
$$

where $B = \begin{pmatrix} \mathbf{1}' & \mathbf{0}' \\ A' - I & I \end{pmatrix}$. Then we only need to prove the above LP is infeasible for all $1 \leq j \leq 2d$. Consider the corresponding dual problem:

$$
\begin{aligned}
(D) \qquad \max \quad & \boldsymbol{\beta}'\boldsymbol{e}_j \\
\text{s.t.} \quad & B'\boldsymbol{\alpha} + \boldsymbol{\beta} = 0 \\
& \boldsymbol{\beta} \geq 0.
\end{aligned}
$$

Then, for all $m > 0$, $\boldsymbol{\alpha} = \begin{pmatrix} -m \\ -m\mathbf{1} \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} m\mathbf{1} \\ m\mathbf{1} \end{pmatrix}$ is a feasible solution to (D) since $(I - A)\mathbf{1} = 0$. The value of the objective function is m. Due to the arbitrariness of m, we see that the optimal value of the dual is unbounded. Therefore, for all $1 \leq j \leq 2d$, the primal is infeasible. Hence, each element of d must be 0, which means that $\boldsymbol{x}_1 = \boldsymbol{x}_2$, proving part (a). To establish (b), suppose $(\boldsymbol{x}^+, \boldsymbol{x}^-)$ is the optimal solution of the primal (2). Suppose for some $1 \leq k \leq d$, both x_k^+ and x_k^- are strictly positive, i.e., $x_k^+ > \delta$ and $x_k^- > \delta$ for some $\delta > 0$. Let $\hat{x}_k^+ = x_k^+ - \delta$, $\hat{x}_k^- = x_k^- - \delta$, and define a new vector $(\bar{\boldsymbol{x}}^+, \bar{\boldsymbol{x}}^-)$ as follows:

$$\begin{cases} \bar{x}_i^+ = \hat{x}_k^+, \bar{x}_i^- = \hat{x}_k^-, & \text{if } \ i = k; \\ \bar{x}_i^+ = x_i^+, \bar{x}_i^- = \bar{x}_i{}^+ - (D_i - s_i + \sum_{j:(j,i)\in E} \bar{x}_j^+ a_{ji}), & \text{otherwise.} \end{cases}$$

Then it is not hard to show that $\bar{\boldsymbol{x}} = \begin{pmatrix} \bar{\boldsymbol{x}}^+ \\ \bar{\boldsymbol{x}}^- \end{pmatrix}$ is a feasible solution to the problem (2). In addition, the value of the objective function at $\bar{\boldsymbol{x}}$ is strictly less than the value at $\boldsymbol{x}$, which conflicts with the optimality of $\boldsymbol{x}$. Therefore, at least one element in the pair (x_k^+, x_k^-) is zero, $\forall 1 \leq k \leq d$. □

3.3. Insensitivity of the Solution to the Primal

THEOREM 3. *Suppose* $\boldsymbol{x}^* = \begin{pmatrix} \boldsymbol{x}^{*+} \\ \boldsymbol{x}^{*-} \end{pmatrix}$ *is the optimal solution to the problem*

$$\begin{aligned} \min \quad & f_1(\boldsymbol{x}^+) \\ s.t. \quad & (A' - I)\boldsymbol{x}^+ + I\boldsymbol{x}^- = \boldsymbol{s} - \boldsymbol{D} \\ & \boldsymbol{x}^+ \geq \boldsymbol{0}, \boldsymbol{x}^- \geq \boldsymbol{0}, \end{aligned}$$

where $f_1(\boldsymbol{x}^+)$ *is differentiable and increasing with respect to* x^+. *Let* $f_2(\boldsymbol{x}^+)$ *be another differentiable and increasing function. Then* $\boldsymbol{x}^*$ *is also the optimal solution to the problem*

$$\begin{aligned} \min \quad & f_2(\boldsymbol{x}^+) \\ s.t. \quad & (A' - I)\boldsymbol{x}^+ + I\boldsymbol{x}^- = \boldsymbol{s} - \boldsymbol{D} \\ & \boldsymbol{x}^+ \geq \boldsymbol{0}, \boldsymbol{x}^- \geq \boldsymbol{0}. \end{aligned}$$

Proof: Consider the problem

$$
\begin{aligned}
(P') \qquad & \min \quad f_1(\boldsymbol{x}^+) \\
& \text{s.t.} \quad (A'-I)\boldsymbol{x}^+ + I\boldsymbol{x}^- = \boldsymbol{s} - \boldsymbol{D} && (\boldsymbol{\alpha}) \\
& \qquad \boldsymbol{x}^+ \geq \boldsymbol{0} && (\boldsymbol{\mu}) \\
& \qquad \boldsymbol{x}^- \geq \boldsymbol{0}, && (\boldsymbol{\lambda})
\end{aligned}
$$

Suppose $\boldsymbol{x}^* = \begin{pmatrix} \boldsymbol{x}^{*+} \\ \boldsymbol{x}^{*-} \end{pmatrix}$ is the optimal solutions to (P'), and the Lagrange function is

$$L(\boldsymbol{x}^*, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\boldsymbol{x}^{*+}) + \boldsymbol{\alpha}'[(A'-I)\boldsymbol{x}^{*+} + I\boldsymbol{x}^{*-} - \boldsymbol{s} + \boldsymbol{D}] - \boldsymbol{\mu}'\boldsymbol{x}^{*+} - \boldsymbol{\lambda}'\boldsymbol{x}^{*-}.$$

Then $(\boldsymbol{x}^{*+}, \boldsymbol{x}^{*-})$ and $(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\lambda})$ satisfy the *Karush-Kuhn-Tucher (KKT)* conditions when $f = f_1$, i.e.

$$
\begin{cases}
\nabla_{\boldsymbol{x}^+} f + (A-I)\boldsymbol{\alpha} - \boldsymbol{\mu} = 0 \\
\boldsymbol{\alpha} - \boldsymbol{\lambda} = 0 \\
x_i^{*+}\mu_i = 0, \forall i \\
x_i^{*-}\lambda_i = 0, \forall i \\
(A'-I)\boldsymbol{x}^{*+} + I\boldsymbol{x}^{*-} = \boldsymbol{s} - \boldsymbol{D} \\
\boldsymbol{x}^{*+} \geq 0, \boldsymbol{x}^{*-} \geq 0, \boldsymbol{\mu} \geq 0, \boldsymbol{\lambda} \geq 0,
\end{cases}
$$

where $\nabla_{\boldsymbol{x}^+} f$ represents the gradient of f with respect to $\boldsymbol{x}^+$. Now we would like to construct the dual solution vector $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\lambda}})$, such that when $f = f_2$, $(\boldsymbol{x}^{*+}, \boldsymbol{x}^{*-})$ and $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\lambda}})$ satisfy the above KKT conditions. Then we can claim that $(\boldsymbol{x}^{*+}, \boldsymbol{x}^{*-})$ is also the optimal solution when $f = f_2$. Define $\mathcal{H} = \{1 \leq i \leq d : x_i^{*+} > 0\}$, and $\bar{\mathcal{H}} = \{1, 2, \ldots, d\} \backslash \mathcal{H}$. For each $i \in \mathcal{H}$, set $\hat{\mu}_i = 0$; and for each $i \in \bar{\mathcal{H}}$, set $\hat{\lambda}_i = 0$. Without loss of generality we assume that $\mathcal{H} = \{1, 2, \ldots, |\mathcal{H}|\}$. Let $\boldsymbol{\mu}_{\bar{\mathcal{H}}} = \{\mu_{|\mathcal{H}|+1}, \mu_{|\mathcal{H}|+2}, \ldots, \mu_d\}$, $\boldsymbol{\lambda}_{\mathcal{H}} = \{\lambda_1, \lambda_2, \ldots, \lambda_{|\mathcal{H}|}\}$, and $\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\lambda}_{\mathcal{H}} \\ \boldsymbol{\mu}_{\bar{\mathcal{H}}} \end{pmatrix}$. Let Q be a $d \times d$ diagonal matrix with the first $|\mathcal{H}|$ diagonal elements equal to 1 and the remaining elements equal to 0. Considering the second KKT condition, the first KKT condition becomes

$$\nabla_{\boldsymbol{x}^+} f + (A-I)\boldsymbol{\alpha} - \boldsymbol{\mu} = \nabla_{\boldsymbol{x}^+} f + (A-I)\boldsymbol{\lambda} - \boldsymbol{\mu} = \nabla_{\boldsymbol{x}^+} f + (A-I)Q\boldsymbol{\xi} - (I-Q)\boldsymbol{\xi} = 0$$

$$\Rightarrow [(I-Q)-(A-I)Q]\boldsymbol{\xi} = \nabla_{\boldsymbol{x}^+} f$$

$$\Rightarrow (I-AQ)\boldsymbol{\xi} = \nabla_{\boldsymbol{x}^+} f.$$

Notice that the matrix A is irreducible and stochastic. Also we claim that Q cannot be the identity matrix with probability 1. To see this, suppose Q is the identity matrix, in other words, $x_i^{*+} > 0, \forall 1 \le i \le d$. Note that the conclusion of Theorem 2(b) is still valid when the objective function is f, and the proof is exactly the same. Then $x_i^{*-} = 0, \forall 1 \le i \le d$. Adding all constraints in the primal problem (2) gives us $\sum_{i=1}^d D_i = \sum_{i=1}^d s_i$. But this equality holds with probability 0. Therefore, $(I - AQ)$ is invertible with probability 1, and $\boldsymbol{\xi} = (I-AQ)^{-1}\nabla_{\boldsymbol{x}^+} f$. Because f is increasing in $\boldsymbol{x}^+$ and $(I-AQ)^{-1} \ge 0$, we have that $\boldsymbol{\xi} \ge 0$. It is obvious that $(\boldsymbol{x}^{*+}, \boldsymbol{x}^{*-})$ and $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\lambda}}) = (Q\boldsymbol{\xi}, (I-Q)\boldsymbol{\xi}, Q\boldsymbol{\xi})$ satisfy the above KKT conditions when $f = f_2$. □

Note that although Theorem 3 established the insensitivity of the optimal solution to a large class of nonlinear objective functions, for the rest of the paper, our discussion is based on the primal problem (2) and the dual problem (3) with linear objective functions.

4. Asymptotic Behavior

Now we would like to discuss the asymptotic behavior of the failure probability of this distribution network, which is useful when we develop efficient simulation algorithms for estimating the failure probability in the next section. We will now assume a specific geometric layout for the network. In particular, the vertices in the network lie in the plane $T = [0,1]^2$, and their locations are i.i.d. We next specify the vertices' supplies and the distribution for the demands.

Let $t \in T$ represent a location in this network, where we note that T is a compact set. Suppose we have positive continuous functions $\gamma(t), \mu(t), \sigma(t)$ on T, and $\sigma^2(s,t)$ on $T \times T$. Let $g_T(t)$ be the density function of $t \in T$, which is positive and continuous. We independently generate this random graph with d nodes at locations $\{t_1, t_2, \ldots, t_d\}$ in T. For each node i with location $t_i \in T$, there is a deterministic supply $s_n(t_i) \triangleq s(t_i) = n^\beta \gamma(t_i)$, where $\beta > 0$, n is a rarity parameter, and a random demand $D(t_i) \sim N(\mu(t_i), \sigma^2(t_i))$, where the covariance between the demands at two vertices with locations t_i and t_j is $cov[(D(t_i), D(t_j)] = \sigma^2(t_i, t_j)$. The demands correspond to a Gaussian random field with parameter space T. Since T is a compact set, we have that $\gamma(t), \mu(t), \sigma(t)$ are bounded and $\inf_{t\in T} \sigma(t) > 0$. Also note that only the supply function $s(t)$ involves n, not the demand function. Let Σ be the covariance matrix of $(D(t_1), D(t_2), \ldots, D(t_d))$, which we require to be symmetric positive definite.

We now establish a theorem that describes the asymptotic behavior of this network. More specifically, it tells what is the most likely way in which this network fails. This result is crucial in designing an efficient importance sampling algorithm. To prove the theorem, we will use the following result.

LEMMA 1. *If a random variable $X \sim N(\bar{\mu}, \bar{\sigma}^2)$, where $\bar{\sigma} > 0$, then for all $\alpha > \bar{\mu}$,*

$$P\{X > \alpha\} \geq \frac{1}{\sqrt{2\pi}} \frac{\bar{\sigma}}{\alpha - \bar{\mu}} \exp\{-\frac{(\alpha - \bar{\mu})^2}{2\bar{\sigma}^2}\}. \tag{5}$$

Proof: Let $g(x)$ be the density function of X, so $g(x) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2}} \exp\{-\frac{(x-\bar{\mu})^2}{2\bar{\sigma}^2}\}$. Note that $g(x) = -\frac{\bar{\sigma}^2}{x-\bar{\mu}} g'(x)$, where $g'(x)$ represents the derivative of $g(x)$, and for all $\alpha > \bar{\mu}$,

$$P\{X > \alpha\} = \int_\alpha^\infty g(x)\,dx = \int_\alpha^\infty -\frac{\bar{\sigma}^2}{x - \bar{\mu}} g'(x)\,dx \geq -\frac{\bar{\sigma}^2}{\alpha - \bar{\mu}} \int_\alpha^\infty g'(x)\,dx = \frac{\bar{\sigma}^2}{\alpha - \bar{\mu}} g(\alpha). \quad \square$$

THEOREM 4. *Let $L_n(\boldsymbol{D})$ denote the optimal value of the dual (3), when the demand vector is $\boldsymbol{D}$ and the rarity parameter is n. Then for all $k \geq 0$,*

$$\lim_{n\to\infty} n^{-2\beta} \log P\{L_n(\boldsymbol{D}) > k\} = \lim_{n\to\infty} n^{-2\beta} \log P\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > k\} \tag{6}$$

$$= \lim_{n\to\infty} n^{-2\beta} \log P\{\sup_{t\in T} D(t) - s(t) > k\} \tag{7}$$

$$= -\frac{\gamma^2(t^*)}{2\sigma^2(t^*)}, \tag{8}$$

where $t^* = \arg\inf_{t\in T} \frac{\gamma(t)}{\sigma(t)}$.

Proof: We will prove this result by establishing upper and lower bounds on $P\{L_n(\boldsymbol{D})>k\}$. We start with deriving an upper bound. Note that $h(t)\triangleq \frac{D(t)-\mu(t)}{\sigma(t)}$ is a centered Gaussian process, almost surely (a.s.) bounded on T. Then $\sigma_T^2 \triangleq \sup_{t\in T} E[h^2(t)]=1$, $\lambda \triangleq E[\|h\|]<\infty$, where $\|h\|=\sup_{t\in T}h(t)$. We first claim that

$$\{L_n(\boldsymbol{D})>k\}\subseteq\{\max_{i=1,\ldots,d} D(t_i)-s(t_i)>0\}. \tag{9}$$

To see this, if we assume $\max_{i=1,\ldots,d} D(t_i)-s(t_i)\le 0$, then $D(t_i)\le s(t_i), \forall i=1,2,\ldots,d$. According to Theorem 1(b), the primal problem (2) is feasible, and it is easy to see that $x_i^+=0, x_i^-=s(t_i)-D(t_i)\ge 0, \forall i=1,2,\ldots,d$, is an optimal solution to the primal problem. In this case $L_n(\boldsymbol{D})=0$. Thus $\{\max_{i=1,\ldots,d} D(t_i)-s(t_i)>0\}^c\subseteq\{L_n(\boldsymbol{D})>k\}^c$, where "$c$" represents the complement of a set. Therefore,

$$\begin{aligned} P\{L_n(\boldsymbol{D})>k\} &\le P\{\max_{i=1,\ldots,d}\frac{D(t_i)-s(t_i)}{\sigma(t_i)}>0\}\\ &\le P\{\sup_{t\in T}\frac{D(t)-s(t)}{\sigma(t)}>0\}\\ &= P\{\sup_{t\in T}(h(t)-\frac{s(t)-\mu(t)}{\sigma(t)})>0\}. \end{aligned}$$

Set $\hat{t}=\arg\sup_{t\in T}\frac{\mu(t)}{\sigma(t)}$. Note that when n is large enough, $\frac{n^\beta\gamma(t^*)}{\sigma(t^*)}-\frac{\mu(\hat{t})}{\sigma(\hat{t})}>0$. Then

$$\begin{aligned} P\{L_n(\boldsymbol{D})>k\} &\le P\{\sup_{t\in T}h(t)>\frac{n^\beta\gamma(t^*)}{\sigma(t^*)}-\frac{\mu(\hat{t})}{\sigma(\hat{t})}\}\\ &= P\{\|h\|>\frac{n^\beta\gamma(t^*)}{\sigma(t^*)}-\frac{\mu(\hat{t})}{\sigma(\hat{t})}\}\\ &\le \exp\{-\frac{1}{2}(\frac{n^\beta\gamma(t^*)}{\sigma(t^*)}-\frac{\mu(\hat{t})}{\sigma(\hat{t})}-\lambda)^2\}, \end{aligned} \tag{10}$$

where the last inequality is obtained by Borell-TIS inequality (Adler and Taylor 2007, p. 50). This establishes the desired upper bound on $P\{L_n(\boldsymbol{D})>k\}$.

To obtain a lower bound on the probability, we first define a metric d on T, known as the *canonical metric* (Adler and Taylor 2007, p. 12), by setting

$$d(s,t)\triangleq\{E[(h(s)-h(t))^2]\}^{1/2},$$

where h is a centered Gaussian process, a.s. bounded on T. We denote the ball of radius ϵ in the canonical metric, centered at a point $t \in T$, by

$$B(t,\epsilon) \triangleq \{s \in T : d(s,t) \le \epsilon\}.$$

Define $g(t) \triangleq \frac{1}{\sqrt{2\pi}} \frac{\sigma(t)}{s(t)-\mu(t)+k} \exp\{-\frac{(s(t)-\mu(t)+k)^2}{2\sigma^2(t)}\}, t \in T$, where $k \ge 0$ is some constant. Then $g(t)$ is continuous in T. Therefore, for a given $0 < \delta < 1$, there exists a ball $B(t^*,\epsilon)$ centered at t^* such that $|g(t) - g(t^*)| \le \delta g(t^*)$, for all $t \in B(t^*,\epsilon)$. We now claim that

$$P\{L_n(\boldsymbol{D}) > k\} \ge P\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > k\}.$$

To see this, note that if $\max_{i=1,\ldots,d} D(t_i) - s(t_i) > k$, then there exists some $1 \le i_0 \le d$ such that $D(t_{i_0}) - s(t_{i_0}) > k$. Let $\boldsymbol{y}$ be the vector with the i_0-th element equal to 1 and the rest of the elements equal to 0. It is easy to see that $\boldsymbol{y}$ is a feasible solution to the dual problem (3) and $\boldsymbol{y}'(\boldsymbol{D} - \boldsymbol{s}) = D(t_{i_0}) - s(t_{i_0}) > k$. Therefore, $L_n(\boldsymbol{D}) > k$. Then,

$$\begin{aligned} P\{L_n(\boldsymbol{D}) > k\} &\ge P\{\max_{i=1,\ldots,n} D(t_i) - s(t_i) > k\} \qquad (11)\\ &\ge P\{D(t_1) - s(t_1) > k\}\\ &= \int_T P\{D(t) - s(t) > k\} g_T(t)\,dt\\ &\ge \int_{B(t^*,\epsilon)} P\{D(t) - s(t) > k\} g_T(t)\,dt\\ &\ge \int_{B(t^*,\epsilon)} \frac{1}{\sqrt{2\pi}} \frac{\sigma(t)}{s(t)-\mu(t)+k} \exp\{-\frac{(s(t)-\mu(t)+k)^2}{2\sigma^2(t)}\} g_T(t)\,dt\\ &\ge g(t^*)(1-\delta)C, \qquad (12) \end{aligned}$$

where $C = \int_{B(t^*,\epsilon)} g_T(t)\,dt$, and the second-to-last step applied Lemma 1, and the last step follows from the fact that for all $t \in B(t^*,\epsilon)$, $g(t) \le (1-\delta)g(t^*)$, giving us the desired lower bound on $P\{L_n(\boldsymbol{D}) > k\}$.

Therefore, (10) and (12) imply for n sufficiently large,

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \frac{\sigma(t^*)}{n^\beta \gamma(t^*) - \mu(t^*) + k} \exp\{-\frac{(n^\beta \gamma(t^*) - \mu(t^*) + k)^2}{2\sigma^2(t^*)}\}(1-\delta)C\\ &\le P\{L_n(\boldsymbol{D}) > k\} \le \exp\{-\frac{1}{2}(\frac{n^\beta \gamma(t^*)}{\sigma(t^*)} - \frac{\mu(\hat{t})}{\sigma(\hat{t})} - \lambda)^2\}. \end{aligned}$$

Taking logarithms, we have

$$\log[\frac{1}{\sqrt{2\pi}}\frac{\sigma(t^*)}{n^\beta\gamma(t^*)-\mu(t^*)+k}]-\frac{(n^\beta\gamma(t^*)-\mu(t^*)+k)^2}{2\sigma^2(t^*)}+\log[(1-\delta)C]$$
$$\leq \log P\{L_n(\boldsymbol{D})>k\}\leq -\frac{1}{2}(\frac{n^\beta\gamma(t^*)}{\sigma(t^*)}-\frac{\mu(\hat{t})}{\sigma(\hat{t})}-\lambda)^2.$$

Because

$$\lim_{n\to\infty}\frac{1}{n^{2\beta}}\big(\log[\frac{1}{\sqrt{2\pi}}\frac{\sigma(t^*)}{n^\beta\gamma(t^*)-\mu(t^*)+k}]-\frac{(n^\beta\gamma(t^*)-\mu(t^*)+k)^2}{2\sigma^2(t^*)}+\log[(1-\delta)C]\big)$$
$$=\lim_{n\to\infty}-\frac{1}{n^{2\beta}}\frac{1}{2}\frac{(n^\beta\gamma(t^*)-\mu(t^*)+k)^2}{\sigma^2(t^*)}=-\frac{\gamma^2(t^*)}{2\sigma^2(t^*)},$$

it follows that

$$\lim_{n\to\infty}n^{-2\beta}\log P\{L_n(\boldsymbol{D})>k\}=-\frac{\gamma^2(t^*)}{2\sigma^2(t^*)},$$

thereby verifying (8).

We now establish (6) and (7). By (11) and (12), it follows that

$$\frac{1}{\sqrt{2\pi}}\frac{\sigma(t^*)}{n^\beta\gamma(t^*)-\mu(t^*)+k}\exp\{-\frac{(n^\beta\gamma(t^*)-\mu(t^*)+k)^2}{2\sigma^2(t^*)}\}(1-\delta)C\leq P\{\max_{i=1,\ldots,d}D(t_i)-s(t_i)>k\}.$$

In addition, similar to what we did for $P\{L_n(D)>k\}$, to obtain an upper bound for the right side of the previous display, we can use the Borell-TIS inequality to conclude that

$$P\{\max_{i=1,\ldots,d}D(t_i)-s(t_i)>k\}\leq P\{\sup_{t\in T}D(t)-s(t)>k\}\leq\exp\{-\frac{1}{2}(\frac{n^\beta\gamma(t^*)}{\sigma(t^*)}-\frac{\mu(\hat{t})}{\sigma(\hat{t})}+\frac{k}{\sigma(\bar{t})}-\lambda)^2\},$$

where $\bar{t}=\arg\inf_{t\in T}\frac{1}{\sigma(t)}$. Then by taking logarithms across these two sets of inequalities, we see that

$$\lim_{n\to\infty}n^{-2\beta}\log P\{\max_{i=1,\ldots,d}D(t_i)-s(t_i)>k\}=\lim_{n\to\infty}n^{-2\beta}\log P\{\sup_{t\in T}D(t)-s(t)>k\}=-\frac{\gamma^2(t^*)}{2\sigma^2(t^*)}.\quad\square$$

5. Efficient Algorithms: Importance Sampling and Conditional Monte Carlo

5.1. Asymptotic Optimality

Suppose the locations of the vertices $t_i, i=1,2,\ldots,d$, have been generated. When n is large, the failure of this network is a rare event. To estimate this failure probability, we develop two efficient simulation algorithms: one based on importance sampling (IS) and the other using conditional Monte Carlo (CMC). To evaluate the efficiency of these two algorithms, we need to introduce a definition.

DEFINITION 1. A collection $(Z_n : n \geq 0)$ of estimators for $\rho(n)$ is said to be asymptotically optimal if $E[Z_n] = \rho(n)$ and if

$$\sup_{n>0} \frac{E(Z_n^2)}{\rho(n)^{2-\epsilon}} < \infty, \forall \epsilon > 0.$$

Asymptotic optimality also amounts to showing that

$$\frac{\log E(Z_n^2)}{2\log(\rho(n))} \to 1, \quad n \to \infty.$$

5.2. Importance Sampling

We now develop an IS estimator making use of a new probability measure Q:

$$Q\{\boldsymbol{D} \in B\} = \sum_{i=1}^{d} p(i) P\{\boldsymbol{D} \in B | D(t_i) - s(t_i) > 0\}, \tag{13}$$

where $B \subset \mathbb{R}^d$ is a Borel set, and

$$p(i) = \frac{P\{D(t_i) - s(t_i) > 0\}}{\sum_{j=1}^{d} P\{D(t_j) - s(t_j) > 0\}}.$$

Note that Q is a mixture of d measures, where the i-th measure in the mixture is the conditional distribution given that the i-th node's demand exceeds its supply. Since

$$Q\{\boldsymbol{D} \in B\} = \frac{1}{\sum_{j=1}^{d} P\{D(t_j) - s(t_j) > 0\}} \sum_{i=1}^{d} P\{\boldsymbol{D} \in B, D(t_i) - s(t_i) > 0\},$$

it is easy to see that

$$\frac{dP}{dQ} = \frac{\sum_{j=1}^{d} P\{D(t_j) - s(t_j) > 0\}}{\sum_{j=1}^{d} I\{D(t_j) - s(t_j) > 0\}}.$$

5.2.1. Asymptotic Optimality

THEOREM 5.

$$Z_n(\boldsymbol{D}) = \frac{dP}{dQ} I\{L_n(\boldsymbol{D}) > k\} = \frac{\sum_{j=1}^{d} P\{D(t_j) - s(t_j) > 0\}}{\sum_{j=1}^{d} I\{D(t_j) - s(t_j) > 0\}} I\{L_n(\boldsymbol{D}) > k\}$$

is an asymptotically optimal estimator for $\alpha_n(k) \triangleq P\{L_n(\boldsymbol{D}) > k\}$.

Proof: Let E_Q denote the expectation under Q, so by (9), we have

$$\log E_Q[Z_n^2(\boldsymbol{D})] = \log E_Q[(\frac{dP}{dQ} I\{L_n(\boldsymbol{D}) > k\})^2] \\ \leq \log E_Q[\big(\frac{dP}{dQ} I\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\}\big)^2].$$

Since $I\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\} = 1$ implies $\sum_{j=1}^d I\{D(t_j) - s(t_j) > 0\} \geq 1$, and under measure Q, $\sum_{j=1}^d I\{D(t_j) - s(t_j) > 0\} \geq 1$,

$$\frac{dP}{dQ} I\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\} = \frac{\sum_{j=1}^d P\{D(t_j) - s(t_j) > 0\}}{\sum_{j=1}^d I\{D(t_j) - s(t_j) > 0\}} I\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\} \\ \leq \sum_{j=1}^d P\{D(t_j) - s(t_j) > 0\}.$$

Thus

$$\log E_Q[Z_n^2(\boldsymbol{D})] \leq \log \big(\sum_{j=1}^d P\{D(t_j) - s(t_j) > 0\}\big)^2 = 2 \log \sum_{j=1}^d P\{D(t_j) - s(t_j) > 0\}.$$

Since

$$P\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\} \leq \sum_{j=1}^d P\{D(t_j) - s(t_j) > 0\} \leq d \times P\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\},$$

we have

$$\lim_{n\to\infty} \frac{\log \sum_{j=1}^d P\{D(t_j) - s(t_j) > 0\}}{\log P\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\}} = 1.$$

Therefore,

$$\lim_{n\to\infty} \frac{\log E_Q[Z_n^2(\boldsymbol{D})]}{\log P\{L_n(\boldsymbol{D}) > k\}} \leq \frac{2 \log \sum_{j=1}^d P\{D(t_j) - s(t_j) > 0\}}{\log P\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\}} \frac{\log P\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\}}{\log P\{L_n(\boldsymbol{D}) > k\}} = 2,$$

where the last equation follows from Theorem 4. □

5.2.2. Algorithm Implementation We now explain how to implement the IS algorithm.

1. Set $i = 1$ and let N be the total number of replications to simulate.

2. Generate demand vector $\boldsymbol{D}^{(i)}$ from distribution Q as in (13). To do this, we use the algorithm described in Robert (1995) to sample truncated normal variables.

3. Calculate $Z_n(\boldsymbol{D}^{(i)}) = \frac{\sum_{j=1}^d P\{D(t_j) - s(t_j) > 0\}}{\sum_{j=1}^d I\{D(t_j) - s(t_j) > 0\}} I\{L_n(\boldsymbol{D}^{(i)}) > k\}$.

4. If $i < N$, set $i = i + 1$ and go to step 2; otherwise, go to step 5.

5. Compute $\widehat{\alpha}_n(k) = (\sum_{i=1}^N Z_n(\boldsymbol{D}^{(i)}))/N$ as our importance-sampling estimator of $\alpha_n(k) = P\{L_n(\boldsymbol{D}) > k\}$, and a $100(1-\delta)\%$ confidence interval for $\alpha_n(k)$ is $(\widehat{\alpha}_n(k) \pm \Phi^{-1}(1-\delta/2)\widehat{S}/\sqrt{N}))$, where $\widehat{S}^2 = \left(\sum_{i=1}^N (Z_n(\boldsymbol{D}^{(i)}) - \widehat{\alpha}_n(k))^2\right)/(N-1)$, and $\Phi(\cdot)$ is the distribution function of a standard normal.

5.3. Conditional Monte Carlo

Note that the multivariate-normal random demand has polar-coordinate representation (see McNeil, Frey and Embrechts 2005):

$$\boldsymbol{D} = \boldsymbol{\mu} + RW\boldsymbol{\Psi}, \tag{14}$$

where the radius R satisfies $R^2 \sim \Gamma(d/2, 1/2)$, i.e., its density function $g(x) = x^{d/2-1}e^{-x/2}(1/2)^{d/2}/\Gamma(d/2)$, $\Gamma(\cdot)$ is the gamma function, $WW^T = \Sigma$, the angle $\boldsymbol{\Psi} = \boldsymbol{z}/\|\boldsymbol{z}\|$, is uniformly distributed over the unit sphere, $\boldsymbol{z} = (z_1, z_2, \ldots, z_d)' \sim N(0, I)$, and $\|\boldsymbol{z}\| = \sqrt{z_1^2 + z_2^2 + \cdots + z_d^2}$. In addition, the radius R and angle $\boldsymbol{\Psi}$ are independent.

Making use of this representation, Blanchet, Li and Nakayama (2011) developed a conditional Monte Carlo approach for estimating $\alpha(k)$, along with algorithmic details on how to implement the method. However, we did not discuss the optimality of the CMC algorithm in the conference version. Here, we would like to solve this problem.

5.3.1. Asymptotic Optimality Recall that we defined in Section 4 the deterministic supply of node i at location t_i as $s_n(t_i) = n^\beta \gamma(t_i)$, where $\beta > 0$ is a constant, n is the rarity parameter, and $\gamma(\cdot)$ is a fixed positive function.

THEOREM 6. *There exist constants $n_0 > 0$, $c_3 > 0$, $s^* > 0$, $\eta_1 \in \mathbb{R}$, such that when $n > n_0$,*

$$T_n(\boldsymbol{\Psi}) \triangleq P\{L_n(\boldsymbol{D}) > k|\boldsymbol{\Psi}\} \leq P\{R > n^\beta s^* + \eta_1\}, \quad \forall \|\boldsymbol{\Psi}\| = 1, \tag{15}$$

$$P\{L_n(\boldsymbol{D}) > k\} \geq c_3 P\{R > n^\beta s^* + O(1)\} n^{-(d-1)\beta}. \tag{16}$$

Also, the conditional Monte Carlo estimator $T_n(\boldsymbol{\Psi})$ is asymptotically optimal.

Proof: We first prove (15). Let $\Omega = \{\boldsymbol{y} : M\boldsymbol{y} \leq \mathbf{1}, \boldsymbol{y} \geq \mathbf{0}\}$ denote the feasible region of the dual problem (3). Then $L_n(\boldsymbol{D}) = \max \boldsymbol{y}'(\boldsymbol{\mu} + RW\boldsymbol{\Psi} - n^\beta \boldsymbol{\gamma}), \boldsymbol{y} \in \Omega$, where $\boldsymbol{\gamma} = (\gamma(t_1), \gamma(t_2), \ldots, \gamma(t_d))'$ as defined in Section 4. We are interested in the failure probability, which includes two cases as we noted previously in Section 2. One case is that the primal problem is infeasible, which, according to Theorem 1(b), occurs if and only if when $\mathbf{1}'(\boldsymbol{\mu} + RW\boldsymbol{\Psi} - n^\beta \boldsymbol{\gamma}) > 0$. The other case is that the primal problem is feasible but the optimal value is greater than k. Since the dual problem is an LP, for the second case, we can focus on the extreme points of the feasible region Ω. Since $k > 0$, when $\boldsymbol{y} = \mathbf{0}$, the optimal value is 0, so we do not have a failure. Therefore, we do not need to consider the solution $\mathbf{0}$ when calculating the failure probability.

Suppose $\{\tilde{\boldsymbol{y}}_i : i = 1, 2, \ldots, m\}$ are the extreme points of Ω, excluding $\mathbf{0}$, and we have

$$\begin{aligned}\{L_n(\boldsymbol{D}) > k\} &= \{\mathbf{1}'(\boldsymbol{\mu} + RW\boldsymbol{\Psi} - n^\beta \boldsymbol{\gamma}) > 0\} \bigcup \Big[\bigcup_{i=1}^m \{\tilde{\boldsymbol{y}}_i'(\boldsymbol{\mu} + RW\boldsymbol{\Psi} - n^\beta \boldsymbol{\gamma}) > k\}\Big] \\ &= \bigcup_{i=0}^m \{\tilde{\boldsymbol{y}}_i'(\boldsymbol{\mu} + RW\boldsymbol{\Psi} - n^\beta \boldsymbol{\gamma}) > k_i\},\end{aligned}$$

where $\tilde{\boldsymbol{y}}_0 = \mathbf{1}$, and

$$k_i = \begin{cases} 0, & i = 0; \\ k, & i = 1, 2, \ldots, m. \end{cases}$$

Let $n_1 = \max\{0, \max_{i=0,1,\ldots,m} \frac{\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle - k_i}{\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\gamma}\rangle}\}^{1/\beta}$, where $\langle \cdot, \cdot \rangle$ denotes inner product. Then when $n > n_1$, we have $n^\beta \langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\gamma}\rangle - \langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle + k_i > 0$. Recall that R is a positive random variable, so

$$\tilde{\boldsymbol{y}}_i'(\boldsymbol{\mu} + RW\boldsymbol{\Psi} - n^\beta \boldsymbol{\gamma}) > k_i \quad \Rightarrow \quad \begin{cases} R > \frac{n^\beta \langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\gamma}\rangle - \langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle}, & \text{if } \langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle > 0; \\ R \in \emptyset, & \text{if } \langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle \leq 0. \end{cases}$$

Define

$$\Gamma_0 = \{\boldsymbol{\Psi} : \|\boldsymbol{\Psi}\| = 1, \max_{i=0,1,\ldots,m} \langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle > 0\},$$

$$M_{\boldsymbol{\Psi}} = \{i = 0, 1, \ldots, m : \langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle > 0\}.$$

For $\boldsymbol{\Psi} \in \Gamma_0$, define

$$H(\boldsymbol{\Psi}, n) = \min_{i \in M_{\boldsymbol{\Psi}}} \frac{n^\beta \langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\gamma}\rangle - \langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle},$$

$$S(\boldsymbol{\Psi}) = \min_{i\in M_{\boldsymbol{\Psi}}} \frac{\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\gamma}\rangle}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle}, \quad i_{\boldsymbol{\Psi}} \in \arg\min_{i\in M_{\boldsymbol{\Psi}}} \frac{\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\gamma}\rangle}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle}, \qquad \tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}} = \tilde{\boldsymbol{y}}_{i_{\boldsymbol{\Psi}}}.$$

It is easy to see that when $n > n_1$,

$$P\{L_n(\boldsymbol{D}) > k\} = P\{R > H(\boldsymbol{\Psi}, n)\}. \tag{17}$$

In the non-trivial case when $\Gamma_0 \neq \emptyset$, there exists some $\boldsymbol{\Psi}_0 \in \Gamma_0$. Let $a = \max_{i=0,1,\ldots,m} \langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}_0\rangle > 0$. Define

$$\Gamma_a = \{\boldsymbol{\Psi} : \|\boldsymbol{\Psi}\| = 1, \max_{i=0,1,\ldots,m} \langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle \geq a\}.$$

Let us consider inequality (15) first. We have

$$T_n(\boldsymbol{\Psi}) = P\{R > H(\boldsymbol{\Psi}, n) | \boldsymbol{\Psi}\} \leq P\{R > \inf_{\boldsymbol{\Psi}\in\Gamma_0} H(\boldsymbol{\Psi}, n)\} = P\{R > \inf_{\boldsymbol{\Psi}\in\Gamma_a} H(\boldsymbol{\Psi}, n)\},$$

and

$$\begin{aligned}
\inf_{\boldsymbol{\Psi}\in\Gamma_a} H(\boldsymbol{\Psi}, n) &= \inf_{\boldsymbol{\Psi}\in\Gamma_a} \min_{i\in M_{\boldsymbol{\Psi}}} \frac{n^{\beta}\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\gamma}\rangle - \langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle} \\
&\geq \inf_{\boldsymbol{\Psi}\in\Gamma_a} \min_{i\in M_{\boldsymbol{\Psi}}} \frac{n^{\beta}\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\gamma}\rangle}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle} + \inf_{\boldsymbol{\Psi}\in\Gamma_a} \min_{i\in M_{\boldsymbol{\Psi}}} \frac{-\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle} \\
&= n^{\beta} \inf_{\boldsymbol{\Psi}\in\Gamma_a} S(\boldsymbol{\Psi}) + \inf_{\boldsymbol{\Psi}\in\Gamma_a} \min_{i\in M_{\boldsymbol{\Psi}}} \frac{-\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle}.
\end{aligned}$$

Note that both $S(\boldsymbol{\Psi})$ and $\min_{i\in M_{\boldsymbol{\Psi}}} \frac{-\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle}$ are continuous with respect to $\boldsymbol{\Psi}$ on the compact set Γ_a. Then there exist $\boldsymbol{\Psi}^* \in \Gamma_a$ and $\eta_1 \in \mathbb{R}$ such that

$$\begin{aligned}
\inf_{\boldsymbol{\Psi}\in\Gamma_a} S(\boldsymbol{\Psi}) = S(\boldsymbol{\Psi}^*) = \frac{\langle \tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}^*}, \boldsymbol{\gamma}\rangle}{\langle \tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}^*}, W\boldsymbol{\Psi}^*\rangle}, \\
\inf_{\boldsymbol{\Psi}\in\Gamma_a} \min_{i\in M_{\boldsymbol{\Psi}}} \frac{-\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle} = \eta_1.
\end{aligned} \tag{18}$$

Therefore,

$$\inf_{\boldsymbol{\Psi}\in\Gamma_a} H(\boldsymbol{\Psi}, n) \geq n^{\beta} S(\boldsymbol{\Psi}^*) + \eta_1.$$

Then we have

$$T_n(\boldsymbol{\Psi}) \leq P\{R > n^{\beta} S(\boldsymbol{\Psi}^*) + \eta_1\}.$$

Let $s^* \triangleq S(\boldsymbol{\Psi}^*)$, then (15) is established.

Now we consider the inequality (16). We claim that for any $\mathbf{\Psi}$ in Γ_a, there exists $n_2(\mathbf{\Psi})>0$ such that when $n>n_2(\mathbf{\Psi})$,

$$H(\mathbf{\Psi},n)=n^{\beta}S(\mathbf{\Psi})+\frac{k_{\mathbf{\Psi}}-\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},W\mathbf{\Psi}\rangle}, \tag{19}$$

where $k_{\mathbf{\Psi}}$ is the k_i corresponding to $\tilde{\boldsymbol{y}}_{\mathbf{\Psi}}$. To see why this is true, observe that for any $i\in M_{\mathbf{\Psi}}$,

$$\begin{aligned}\lambda_i \triangleq & n^{\beta}S(\mathbf{\Psi})+\frac{k_{\mathbf{\Psi}}-\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},W\mathbf{\Psi}\rangle}-\frac{n^{\beta}\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\gamma}\rangle-\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\mu}\rangle+k_i}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}\\ =& n^{\beta}\big(S(\mathbf{\Psi})-\frac{\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\gamma}\rangle}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}\big)+\big(\frac{k_{\mathbf{\Psi}}-\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},W\mathbf{\Psi}\rangle}-\frac{k_i-\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}\big).\end{aligned}$$

We know that $S(\mathbf{\Psi})-\frac{\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\gamma}\rangle}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}\le 0$. Define

$$\mathcal{I}_{\mathbf{\Psi}}=\{i\in M_{\mathbf{\Psi}}:S(\mathbf{\Psi})-\frac{\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\gamma}\rangle}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}=0\},\quad \mathcal{I}_{\mathbf{\Psi}}^{-}=\{i\in M_{\mathbf{\Psi}}:S(\mathbf{\Psi})-\frac{\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\gamma}\rangle}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}<0\}.$$

Choose

$$i_{\mathbf{\Psi}}\in\arg\min_{i\in\mathcal{I}_{\mathbf{\Psi}}}\frac{k_{\mathbf{\Psi}}-\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},W\mathbf{\Psi}\rangle},$$

then $\lambda_i\le 0,\forall i\in\mathcal{I}_{\mathbf{\Psi}}$. For $i\in\mathcal{I}_{\mathbf{\Psi}}^{-}$, note that both $S(\mathbf{\Psi})-\frac{\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\gamma}\rangle}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}$ and $\frac{k_{\mathbf{\Psi}}-\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},W\mathbf{\Psi}\rangle}-\frac{k_i-\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}$ are bounded on Γ_a. Then there exist $\eta_2(\mathbf{\Psi}),\eta_3(\mathbf{\Psi})>0$, such that

$$S(\mathbf{\Psi})-\frac{\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\gamma}\rangle}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}\le-\eta_2(\mathbf{\Psi}),$$

$$-\eta_3(\mathbf{\Psi})\le\frac{k_{\mathbf{\Psi}}-\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}},W\mathbf{\Psi}\rangle}-\frac{k_i-\langle\tilde{\boldsymbol{y}}_i,\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_i,W\mathbf{\Psi}\rangle}\le\eta_3(\mathbf{\Psi}).$$

Then when $n>n_2(\mathbf{\Psi})=(\eta_3(\mathbf{\Psi})/\eta_2(\mathbf{\Psi}))^{1/\beta}$, we have that $\lambda_i\le 0,\forall i\in\mathcal{I}_{\mathbf{\Psi}}^{-}$. Therefore, when $n>\max\{n_1,n_2(\mathbf{\Psi}^*)\}$, it follows that $\lambda_i\le 0,\forall i\in M_{\mathbf{\Psi}^*}$, so

$$H(\mathbf{\Psi}^*,n)=n^{\beta}S(\mathbf{\Psi}^*)+\frac{k_{\mathbf{\Psi}^*}-\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*},\boldsymbol{\mu}\rangle}{\langle\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*},W\mathbf{\Psi}^*\rangle}. \tag{20}$$

We also claim that there exist $c_1>0$, $c_2\in\mathbb{R}$, such that if $n>\max\{n_1,n_2(\mathbf{\Psi}^*)\}$, then $H(\mathbf{\Psi},n)-H(\mathbf{\Psi}^*,n)\le(n^{\beta}c_1+c_2)\|\mathbf{\Psi}-\mathbf{\Psi}^*\|$ on Γ_a. To see this, for any $\delta>0$ and $\boldsymbol{\theta}\in\Gamma_a$, define $B(\boldsymbol{\theta},\delta)=\{\mathbf{\Psi}\in\Gamma_a:\|\mathbf{\Psi}-\boldsymbol{\theta}\|\le\delta\}$. Note that there exists $\delta_1>0$, such that when $0<\delta\le\delta_1$, and $n>$

$\max\{n_1, n_2(\mathbf{\Psi}^*)\}$, for any $\mathbf{\Psi} \in B(\mathbf{\Psi}^*, \delta)$, we have that the index corresponding to $\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}$ is in $M_{\mathbf{\Psi}}$, and

$$\begin{aligned} H(\mathbf{\Psi}, n) - H(\mathbf{\Psi}^*, n) &= \min_{i \in M_{\mathbf{\Psi}}} \frac{n^\beta \langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\gamma} \rangle - \langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu} \rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W\mathbf{\Psi} \rangle} - \frac{n^\beta \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\gamma} \rangle - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle + k_{\mathbf{\Psi}^*}}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle} \\ &\le \frac{n^\beta \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\gamma} \rangle - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle + k_{\mathbf{\Psi}^*}}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi} \rangle} - \frac{n^\beta \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\gamma} \rangle - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle + k_{\mathbf{\Psi}^*}}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle} \\ &= (n^\beta \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \gamma \rangle - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle + k_{\mathbf{\Psi}^*}) \frac{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi} \rangle \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle} \\ &= (n^\beta \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \gamma \rangle - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle + k_{\mathbf{\Psi}^*}) \frac{\langle W' \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \mathbf{\Psi}^* - \mathbf{\Psi} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi} \rangle \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle}. \end{aligned}$$

Since $\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi} \rangle \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle$ is continuous on $B(\mathbf{\Psi}^*, \delta)$, there exists $\delta_2 \ge 0$ such that when $0 < \delta \le \min\{\delta_1, \delta_2\}$, we have

$$\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi} \rangle \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle \ge \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle^2 - c_0 > 0,$$

where c_0 is some positive constant.

Define $c_1 = \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\gamma} \rangle \frac{\|W' \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}\|}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle^2 - c_0} > 0$, $c_2 = (k_{\mathbf{\Psi}^*} - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle) \frac{\|W' \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}\|}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle^2 - c_0}$, $n_3 = (\frac{\max\{0, -c_2\}}{c_1})^{1/\beta}$. When $n > \max\{n_1, n_2(\mathbf{\Psi}^*), n_3\}$, we have $n^\beta c_1 + c_2 > 0$. Therefore,

$$\begin{aligned} H(\mathbf{\Psi}, n) - H(\mathbf{\Psi}^*, n) &\le (n^\beta \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \gamma \rangle - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle + k_{\mathbf{\Psi}^*}) \frac{\langle W' \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \mathbf{\Psi}^* - \mathbf{\Psi} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi} \rangle \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle} \\ &\le (n^\beta \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \gamma \rangle - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle + k_{\mathbf{\Psi}^*}) \frac{\|W' \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}\| \|\mathbf{\Psi}^* - \mathbf{\Psi}\|}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle^2 - c_0} \\ &= (n^\beta c_1 + c_2) \|\mathbf{\Psi}^* - \mathbf{\Psi}\|. \end{aligned}$$

So for any $\mathbf{\Psi} \in B(\mathbf{\Psi}^*, \delta)$,

$$H(\mathbf{\Psi}, n) \le H(\mathbf{\Psi}^*, n) + (n^\beta c_1 + c_2)\delta. \tag{21}$$

Since $\mathbf{\Psi}$ is uniformly distributed over the unit sphere, which is a $(d-1)$-dimensional manifold, there exists some constant $c_3 > 0$ such that

$$P\{\| \mathbf{\Psi} - \mathbf{\Psi}^* \| \le \delta\} \ge c_3 \delta^{(d-1)}.$$

Let $\delta = n^{-\beta}$. By equations (17) and (21), it follows that

$$P\{L_n(\boldsymbol{D}) > k\} = P\{R > H(\mathbf{\Psi}, n)\}$$

$$
\begin{aligned}
&\geq P\{R > H(\mathbf{\Psi}^*, n) + (n^\beta c_1 + c_2)\delta, \| \mathbf{\Psi} - \mathbf{\Psi}^* \| \leq \delta\} \\
&\geq c_3 P\{R > H(\mathbf{\Psi}^*, n) + (n^\beta c_1 + c_2)\delta\} \delta^{(d-1)} \\
&= c_3 P\{R > n^\beta S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W \mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta})\} n^{-(d-1)\beta} \qquad (22) \\
&= c_3 P\{R > n^\beta S(\mathbf{\Psi}^*) + O(1)\} n^{-(d-1)\beta}.
\end{aligned}
$$

Hence, we have proven (16).

We now establish the last part of the theorem. By (15) and (22), we have

$$
\begin{aligned}
&\frac{\log E[T_n(\mathbf{\Psi})^2]}{\log P\{L(\boldsymbol{D}) > k\}} \leq \frac{\log P\{R > n^\beta S(\mathbf{\Psi}^*) + \eta_1\}^2}{\log c_3 P\{R > n^\beta S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W \mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta})\} n^{-(d-1)\beta}} \\
&= \frac{2 \log P\{R > n^\beta S(\mathbf{\Psi}^*) + \eta_1\}}{\log c_3 + \log P\{R > n^\beta S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W \mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta})\} - (d-1)\beta \log n} \\
&= 2 \Big(\frac{\log P\{R > n^\beta S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W \mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta})\}}{\log P\{R > n^\beta S(\mathbf{\Psi}^*) + \eta_1\}} + \frac{\log c_3 - (d-1)\beta \log n}{\log P\{R > n^\beta S(\mathbf{\Psi}^*) + \eta_1\}} \Big)^{-1}. \qquad (23)
\end{aligned}
$$

Recall that $n^\beta c_1 + c_2 > 0$ when $n > \max\{n_1, n_2(\mathbf{\Psi}^*), n_3\}$, so (18) implies

$$
\eta_1 = \inf_{\mathbf{\Psi} \in \Gamma_a} \inf_{i \in M_{\mathbf{\Psi}}} \frac{-\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu} \rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W \mathbf{\Psi} \rangle} \leq \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W \mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta}).
$$

Therefore,

$$
P\{R > n^\beta S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W \mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta})\} \leq P\{R > n^\beta S(\mathbf{\Psi}^*) + \eta_1\},
$$

and

$$
\frac{\log P\{R > n^\beta S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}, W \mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta})\}}{\log P\{R > n^\beta S(\mathbf{\Psi}^*) + \eta_1\}} \geq \frac{\log P\{R > n^\beta S(\mathbf{\Psi}^*) + \eta_1\}}{\log P\{R > n^\beta S(\mathbf{\Psi}^*) + \eta_1\}} = 1.
$$

When $n > n_4 = e^{\log b / \beta(d-1)}$, the second term inside the parentheses in (23) is non-negative. Then when $n > n_0 = \max\{n_1, n_2(\mathbf{\Psi}^*), n_3, n_4\}$, it follows that (23) is bounded above by 2, thereby concluding the result. □

6. Numerical Examples

Here we use the same basis for comparing the estimators using different simulation algorithms as in Blanchet, Li and Nakayama (2011). Suppose we want to estimate $\alpha = E[X]$, and $X_1, X_2, \ldots, X_N$ are independent replications of X. Then $\widehat{\alpha} = (\sum_{i=1}^{N} X_i)/N$ is an unbiased estimator of α, and $S^2 = (\sum_{i=1}^{N}(X_i - \widehat{\alpha})^2)/(N-1)$ is an unbiased estimator of $Var[X] = \sigma^2$, which we assume is finite. We then define the RSE *(relative standard error)* as $S/(\sqrt{N}\widehat{\alpha})$. To consider both the accuracy and computational efficiency when comparing different unbiased estimators, as suggested in Glynn and Whitt (1992), we use the relative measure $RSE^2 \times CT$(Computing Time) as the criterion.

In our experiments we apply naive simulation, importance sampling, and conditional Monte Carlo methods to different networks, and compare $RSE^2 \times CT$. Examples 1 and 2 show how failure probability changes with the threshold k for fixed supply, and Examples 3 and 4 show how failure probability changes with the rarity parameter n for fixed k. We set $N = 10^5$ for all of the four examples.

6.1. Example 1: $d = 3$, fixed s

The first example is a 3-dimensional network with the following parameters:

$$H = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{s} = \begin{pmatrix} 3 \\ 1 \\ 13 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 & 0.1 \\ 0.5 & 1 & 0.5 \\ 0.1 & 0.5 & 1 \end{pmatrix}.$$

Table 1 Results of Naive Simulation, IS, and CMC for $d=3$, fixed $\boldsymbol{s}$.

	Naive Simulation		importance sampling		Conditional MC	
k	$\alpha(k)$	$RSE^2 \times CT$	$\alpha(k)$	$RSE^2 \times CT$	$\alpha(k)$	$RSE^2 \times CT$
1	1.73×10^{-1}	2.07×10^{-2}	1.73×10^{-1}	8.40×10^{-3}	1.73×10^{-1}	2.95×10^{-2}
5	1.63×10^{-2}	2.71×10^{-1}	1.62×10^{-2}	9.22×10^{-2}	1.64×10^{-2}	6.23×10^{-2}
10	1.72×10^{-3}	2.93×10^{0}	1.61×10^{-3}	7.60×10^{-1}	1.55×10^{-3}	9.80×10^{-2}
12	4.80×10^{-4}	8.79×10^{0}	5.46×10^{-4}	2.17×10^{0}	5.09×10^{-4}	1.17×10^{-1}
16	2.00×10^{-5}	1.92×10^{2}	6.01×10^{-5}	1.99×10^{1}	3.96×10^{-5}	1.51×10^{-1}
20	0	NaN	0	NaN	2.08×10^{-6}	1.85×10^{-1}
22	0	NaN	0	NaN	4.23×10^{-7}	1.96×10^{-1}
25	0	NaN	0	NaN	3.20×10^{-8}	1.86×10^{-1}

6.2. Example 2: $d=10$, fixed s

The second example is a 10-dimensional network with the following parameters:

$$H=\begin{pmatrix} 0&1&1&0&0&0&0&0&0&0\\ 1&0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&1&0&0\\ 0&0&0&0&1&0&1&0&0&0\\ 0&0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&0&1\\ 1&0&0&0&0&0&0&0&0&0 \end{pmatrix}, \quad \boldsymbol{s}=\begin{pmatrix}3\\5\\3\\3\\3\\3\\3\\3\\3\\15\end{pmatrix}, \quad \boldsymbol{\mu}=\begin{pmatrix}1\\5\\1\\1\\1\\1\\1\\1\\1\\1\end{pmatrix},$$

$$\Sigma = \begin{pmatrix} 0.5 & 0.3 & 0.3 & 0.25 & 0.2 & 0.15 & 0.2 & 0.25 & 0.2 & 0.15 \\ 0.3 & 0.5 & 0.25 & 0.2 & 0.15 & 0.1 & 0.15 & 0.2 & 0.15 & 0.1 \\ 0.3 & 0.25 & 0.5 & 0.3 & 0.25 & 0.2 & 0.25 & 0.3 & 0.25 & 0.2 \\ 0.25 & 0.2 & 0.3 & 0.5 & 0.3 & 0.25 & 0.3 & 0.25 & 0.2 & 0.15 \\ 0.2 & 0.15 & 0.25 & 0.3 & 0.5 & 0.3 & 0.25 & 0.2 & 0.15 & 0.1 \\ 0.15 & 0.1 & 0.2 & 0.25 & 0.3 & 0.5 & 0.3 & 0.25 & 0.2 & 0.15 \\ 0.2 & 0.15 & 0.25 & 0.3 & 0.25 & 0.3 & 0.5 & 0.3 & 0.25 & 0.2 \\ 0.25 & 0.2 & 0.3 & 0.25 & 0.2 & 0.25 & 0.3 & 0.5 & 0.3 & 0.25 \\ 0.2 & 0.15 & 0.25 & 0.2 & 0.15 & 0.2 & 0.25 & 0.3 & 0.5 & 0.3 \\ 0.15 & 0.1 & 0.2 & 0.15 & 0.1 & 0.15 & 0.2 & 0.25 & 0.3 & 0.5 \end{pmatrix}.$$

Table 2 Results of Naive Simulation, IS, and CMC for $d = 10$, fixed $\boldsymbol{s}$.

	Naive Simulation		importance sampling		Conditional MC	
k	$\alpha(k)$	$RSE^2 \times CT$	$\alpha(k)$	$RSE^2 \times CT$	$\alpha(k)$	$RSE^2 \times CT$
2	3.64×10^{-2}	1.21×10^{-1}	3.67×10^{-2}	9.57×10^{-2}	3.66×10^{-2}	2.00×10^{-1}
7	3.53×10^{-3}	1.29×10^{0}	3.30×10^{-3}	8.27×10^{-1}	3.30×10^{-3}	7.23×10^{-1}
12	4.70×10^{-4}	9.67×10^{0}	4.08×10^{-4}	3.98×10^{0}	4.01×10^{-4}	1.44×10^{0}
20	3.00×10^{-5}	1.52×10^{2}	4.44×10^{-5}	1.83×10^{1}	3.87×10^{-5}	3.32×10^{0}
30	0	NaN	3.00×10^{-6}	2.04×10^{2}	3.86×10^{-6}	8.17×10^{0}
40	0	NaN	0	NaN	3.47×10^{-7}	1.38×10^{1}
47	0	NaN	0	NaN	4.88×10^{-8}	2.95×10^{1}

6.3. Example 3: $d = 3$, fixed k

The third example is a 3-dimensional network with the following parameters:

$$H = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} 3 \\ 1 \\ 13 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 & 0.1 \\ 0.5 & 1 & 0.5 \\ 0.1 & 0.5 & 1 \end{pmatrix}, \quad k = 1.$$

Table 3 Results of Naive Simulation, IS, and CMC for $d = 3$, fixed k.

	Naive Simulation		importance sampling		Conditional MC	
n	$\alpha(k)$	$RSE^2 \times CT$	$\alpha(k)$	$RSE^2 \times CT$	$\alpha(k)$	$RSE^2 \times CT$
1	1.73×10^{-1}	2.07×10^{-2}	1.73×10^{-1}	8.40×10^{-3}	1.73×10^{-1}	2.95×10^{-2}
1.5	6.77×10^{-2}	5.04×10^{-2}	6.76×10^{-2}	1.59×10^{-2}	6.69×10^{-2}	4.35×10^{-2}
2.5	6.44×10^{-3}	5.34×10^{-1}	6.19×10^{-3}	4.40×10^{-2}	6.21×10^{-3}	7.74×10^{-2}
3.2	6.10×10^{-4}	5.63×10^{0}	6.92×10^{-4}	8.82×10^{-2}	6.88×10^{-4}	1.14×10^{-1}
3.9	8.00×10^{-5}	4.27×10^{1}	4.82×10^{-5}	4.68×10^{-1}	4.83×10^{-5}	1.43×10^{-1}
4.5	0	NaN	3.39×10^{-6}	1.62×10^{0}	3.30×10^{-6}	1.84×10^{-1}
4.9	0	NaN	4.80×10^{-7}	7.08×10^{0}	4.89×10^{-7}	2.03×10^{-1}

6.4. Example 4: $d = 10$, fixed k

The fourth example is a 10-dimensional network with the following parameters:

$$H = \begin{pmatrix} 0&1&1&0&0&0&0&0&0&0\\ 1&0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&1&0&0\\ 0&0&0&0&1&0&1&0&0&0\\ 0&0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&0&1\\ 1&0&0&0&0&0&0&0&0&0 \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} 3\\5\\3\\3\\3\\3\\3\\3\\3\\15 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} 1\\5\\1\\1\\1\\1\\1\\1\\1\\1 \end{pmatrix}, \quad k = 2.$$

$$\Sigma = \begin{pmatrix} 0.5&0.3&0.3&0.25&0.2&0.15&0.2&0.25&0.2&0.15\\ 0.3&0.5&0.25&0.2&0.15&0.1&0.15&0.2&0.15&0.1\\ 0.3&0.25&0.5&0.3&0.25&0.2&0.25&0.3&0.25&0.2\\ 0.25&0.2&0.3&0.5&0.3&0.25&0.3&0.25&0.2&0.15\\ 0.2&0.15&0.25&0.3&0.5&0.3&0.25&0.2&0.15&0.1\\ 0.15&0.1&0.2&0.25&0.3&0.5&0.3&0.25&0.2&0.15\\ 0.2&0.15&0.25&0.3&0.25&0.3&0.5&0.3&0.25&0.2\\ 0.25&0.2&0.3&0.25&0.2&0.25&0.3&0.5&0.3&0.25\\ 0.2&0.15&0.25&0.2&0.15&0.2&0.25&0.3&0.5&0.3\\ 0.15&0.1&0.2&0.15&0.1&0.15&0.2&0.25&0.3&0.5 \end{pmatrix}.$$

6.5. Discussion of Results and Comparisons Between Algorithms

1. When k or n increases, the performance of both the naive simulation and IS deteriorates quickly in terms of $RSE^2 \times CT$. Because we fix the number of simulations N, as in Example 1 and

Table 4 Results of Naive Simulation, IS, and CMC for $d=10$, fixed k.

	Naive Simulation		importance sampling		Conditional MC	
n	$\alpha(k)$	$RSE^2 \times CT$	$\alpha(k)$	$RSE^2 \times CT$	$\alpha(k)$	$RSE^2 \times CT$
1	3.64×10^{-2}	1.21×10^{-1}	3.67×10^{-2}	9.57×10^{-2}	3.66×10^{-2}	2.00×10^{-1}
1.32	3.05×10^{-3}	1.39×10^{0}	3.38×10^{-3}	2.09×10^{-1}	3.38×10^{-3}	6.85×10^{-1}
1.48	2.10×10^{-4}	2.00×10^{1}	2.70×10^{-4}	6.14×10^{-1}	2.73×10^{-4}	2.28×10^{0}
1.6	4.00×10^{-5}	1.04×10^{2}	3.20×10^{-5}	2.19×10^{0}	3.23×10^{-5}	3.79×10^{0}
1.7	0	NaN	4.13×10^{-6}	1.09×10^{1}	4.02×10^{-6}	6.07×10^{0}
1.78	0	NaN	7.34×10^{-7}	5.24×10^{1}	7.26×10^{-7}	6.87×10^{0}

2, when k is very large, we do not get even one observation of the event $\{L_n(\boldsymbol{D}) \geq k\}$. However, although the performance of CMC becomes worse as well, it does not deteriorate as quickly as the other two. No matter how large k is, we can obtain a non-zero estimate of $\alpha(k)$.

2. Although both IS and CMC are asymptotic optimal, when k or n is small, IS performs better than CMC, as we now explain. The IS method only needs to solve a single optimization problem to determine $Z_n(\boldsymbol{D})$ (see Section 5.2.2) in each replication i. In contrast, our conditional Monte Carlo method needs to solve several optimization problems to find the roots R_i^* which equate the optimal value of the primal and the threshold k for a fixed angle $\boldsymbol{\Psi}$ (see equation (8) in Blanchet, Li and Nakayama 2011) in each replication i. Thus, the added computational effort required by CMC can lead to it performing worse than IS. However, as k or n increases, conditional Monte Carlo method works much better. The larger k or n is, the bigger the advantage CMC has compared to naive simulation. The advantage arises because of the significant variance reduction obtained for large k or n overwhelms the additional computational effort. In conclusion, for a given network, IS performs best when k or n is small, and CMC is better when k or n is large.

7. Final Comments

We discuss a distribution network model with each node subjected to given fixed supply and Gaussian random demand. The unserved demand at a node is distributed proportionally to its neighbors.

The equilibrium point is determined by a linear program whose objective is to minimizing the sum of excess demands across all nodes in this network. We developed IS and CMC approaches to efficiently estimate the failure probability. Numerical results show that these two algorithms greatly outperform naive simulation, especially when the threshold k is large.

For CMC algorithm, note that the algorithm requires that the radial component, R, is a positive continuous random variable and that we are able to simulate the radial component conditional on the angular part, Ψ. Therefore the conditional Monte Carlo algorithm applies as long as the demand vector D is an elliptical copula. As for the case of Archimedean copula, recall that if $(U_1,...,U_d)$ follows an Archimedean copula then $C(u_1,...,u_d) := P(U_1 \leq u_1,...,U_d \leq u_d)$ satisfies

$$C(u_1,...,u_d) = \varphi\left(\varphi^{-1}(u_1) + ... + \varphi^{-1}(u_d)\right),$$

for $u_i \in (0,1)$ and $\varphi(\cdot)$ is the so-called generator of $C(\cdot)$ and is $d-$monotonic (see McNeil and Neslehova 2009). A sufficient condition for d-monotonicity is that $\varphi(\cdot)$ is the Laplace transform of some non-negative random variable, for example $\varphi(x) = 1/(1+\theta x)_+^{1/\theta}$, for $\theta > 0$, and x_+ denotes $\max(0,x)$, which gives rise to the so-called Clayton family. Theorem 3.1 in McNeil and Neslehova (2009) indicates that the following equality in distribution holds, namely,

$$\left(\varphi^{-1}(U_1),...,\varphi^{-1}(U_d)\right) = R\Psi,$$

where Ψ is uniformly distributed in the l_1 ball, and R is independent of Ψ, with a distribution depending on $\varphi(\cdot)$ to be discussed shortly. We can simulate Ψ by sampling i.i.d. exponential random variables $(\tau_1,...,\tau_d)$ with unit mean and letting $\Psi_i = \tau_i/(\tau_1 + ... + \tau_d)$. Moreover, we have that the distribution function of R, $F_R(\cdot)$, satisfies

$$F_R(r) = 1 - \sum_{k=0}^{d-2} \frac{(-1)^k r^k \varphi^{(k)}(r)}{k!} - \frac{(-1)^{d-1} r^{d-1} \varphi_+^{(d-1)}(r)}{(d-1)!},$$

for $r > 0$. The conditional Monte Carlo algorithm that we propose is then readily applicable also in the setting of Archimedean copulas. The proof of asymptotic optimality of the estimator would follow similar lines of reasoning, assuming, in addition, that D_i is sufficiently heavy tailed, for example regularly varying tails suffice.

We can make several extensions. In this paper, all of our discussion focuses on a given graph. We can also consider the asymptotic behavior of a graph when the number of nodes grows. For the case when the demand is fixed and supply is jointly Gaussian, which is probably more appropriate for modeling the electric power grid, similar properties and simulation algorithms can be developed.

Acknowledgments

Support from NSF grants CMMI-1069064 and CMMI-1436700 is gratefully acknowledged by the first author.

The work of the third author has been supported in part by the NSF under Grants No. CMMI-0926949, CMMI-1200065, and DMS-1331010. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

Adler, R.J., J. H. Blanchet, J. Liu. 2012. Efficient monte carlo for high excursions of gaussian random fields. Annals of Applied Probability 22 1167-1214.

Adler, R.J., J. E. Taylor. 2007. Random Fields and Geometry. Springer, New York.

Asmussen, S., J. Blanchent, S. Juneja, L. Rojas-Nandayapa. 2011. Efficient simulation of tail probabilities of sums of correlated lognormals. Annals of Operations Research 189 5-23.

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

Brechmann, E. C., Hendrich, K., and Czado, C. 2013. Conditional copula simulation for systemic risk stress testing. Insurance: Mathematics and Economics 53 3 722-732.

Blanchet, J., J. Li, M.K. Nakayama. 2011. A conditional monte carlo method for estimating the failure probability of a distribution network with random demands. Simulation Conference (WSC), Proceedings of the 2011 Winter 3832-3843.

Dobson, I., B. A. Carreras, V. E. Lynch, D. E. Newman. 2007. Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization. Chaos 17 article 026103.

Eisenberg, L., T.H. Noe. 2001. Systemic risk in financial systems. Management Science 47 226-249.

Glynn, P. W., W. Whitt. 1992. The asymptotic efficiency of simulation estimators. Operations Research 40 505-520.

Iyer, S. M., M.K.Nakayama, A. V. Gerbessiotis. 2009. A Markovian dependability model with cascading failures. IEEE Transactions on Computers 139 1238-1249.

Kopparapu, C. 2002. Load Balancing Servers, Firewalls, and Caches. John Wiley & Sons.

McNeil, A. J., R. Frey, P. Embrechts. 2005. Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, New Jersey.

McNeil, A.J., Neslehova, J. 2009. Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. The Annals of Statistics 37 5B 3059-3097.

Robert, C. P. 1995. Simulation of truncated normal variables. Statistics and Computing 5 121-125.

Watts, D. J. 2002. A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences USA 99 5766-5771.