Submitted to *Operations Research*
manuscript (OPRE-2014-08-508)

Rare-Event Simulation for Distribution Networks

Jose Blanchet
Department of Management Science and Engineering, Stanford University, Stanford, CA 94305,
jblanche@stanford.edu
http://orcid.org/0000-0001-5895-0912

Juan Li
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027,
jl3035@columbia.edu
https://orcid.org/0000-0002-1560-0469

Marvin K. Nakayama
Department of Computer Science, New Jersey Institute of Technology, University Heights Newark, NJ 07102,
marvin@njit.edu
https://orcid.org/0000-0003-1503-373X

We model optimal 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 optimal allocation. Assuming that the demands are random (following a jointly Gaussian law), our goal is to study the probability that the optimal cost of unserved demands 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 optimal distribution amounts, we solve a linear program (LP), where the objective function to minimize is the sum across nodes of the unserved demands.

An objective function in the form of the sum of unserved demands is frequently used in the model formulation of network problems involving supply and demand. Zeineldin et al. (2006) aim at minimizing the total generation cost of an electricity distribution system, which includes the sum of unserved energy. Brown et al. (2006) study the optimal defense for multiple infrastructure sectors, e.g., banking and finance, energy, food, public health, transportation, etc., where the sum of economic losses resulting from unmet demands is considered as cost to be minimized. Lisser et al (1999) represent the penalty costs as the sum of unserved demands weighted by unit penalty cost coefficients, and explore the optimal capacity assignment in telecommunications networks with minimum costs.

To illustrate a possible practical setting for which our framework may apply, consider an electric power grid. Each node represents a geographic region, and there is an edge between two nodes if transmission lines directly connect them. Each region has generators, which provide the region's supply of electricity. Also, each region has a random load (i.e., demand for electricity) from consumers. 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. One of the most important issues in operating a power grid is to keep the stability of the network and make sure demands can be satisfied. If the total amount of load not served at their originating regions exceeds a threshold k, then we consider the network to have failed. To better operate this power transmission system, it is essential to estimate the probability that this network fails.

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, at an optimal solution, the sum of unserved demands exceeds threshold k. We also want to study this failure probability when supplies are large, with possibly the threshold large as well, which permits for modeling flexibility. To define an asymptotic regime for analyzing this setting, we introduce a "rarity parameter" n, which increases to infinity, and assume that supplies grow linearly in n. We further assume that the threshold $k = k_n$ is a constant or grows strictly slower than linearly in n. As this framework has supplies (and possibly also the threshold) that depend on n, we now write the failure probability as $\alpha_n(k_n)$, which we want to estimate. 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 $\alpha_n(k_n)$ is *asymptotically optimal* when n goes to infinity if the logarithm of its second moment is asymptotically equivalent to the logarithm of $\alpha_n^2(k_n)$ (see Chapter VI of 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 analysis for our class of linear programs with random input (see Theorem 4).

3) We develop an importance sampling (IS) algorithm for estimating $\alpha_n(k_n)$, and we show that the algorithm is asymptotically optimal as the supply gets large with n, and the threshold k_n is a constant or increases with n as well but at a slower rate (see Section 5.2).

4) We develop a conditional Monte Carlo (CMC) algorithm for the evaluation of $\alpha_n(k_n)$, and we prove the asymptotic optimality of this procedure as the supply gets large with n, and the threshold k_n is a constant or increases with n as well but at a slower rate (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 asymptotic 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 asymptotic 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 show the asymptotic optimality of our algorithms, we include numerical examples in which the rarity parameter changes.

We now explain how our paper relates to prior work in the literature. 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 at an optimal solution 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). IS algorithms have also been used in Wadman,

Crommelin and Frank (2013), and Perninge, Lindskog and Söder (2012) to solve a network operation problem with random inputs. While those two papers focus on the assessment of electrical constraints violation, we make use of IS to assess the optimal solution of an LP. 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. 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, which has been an interesting and important research topic. 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. Bienstock, Blanchet and Li (2016) study control algorithms to limit the probability that cascading failures happen due to overheating. It has been well recognized that efficient simulation algorithms are powerful for rare event analysis associated with cascading blackouts. Wang et al. (2015) apply splitting simulation to evaluate the probabilities associated with cascading blackouts in an electric grid. Belmudes, Ernst, and Wehenkel (2008) make use of cross-entropy method to identify rare events that may endanger power systems.

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 an elliptical distribution (see McNeil, Frey and Embrechts 2005). Furthermore, while an elliptical copula exhibits symmetric tail dependence, the well known Archimedean copula allows asymmetric tail dependence (Brechmann, Hendrich and Czado 2013). Making use of 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.

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 $\alpha_n(k_n)$ 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 $E = \{(i, j) : \exists$ directed edge from vertex i 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, $\boldsymbol{\mu}$ is the mean vector, and Σ is the covariance matrix.

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 node's 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

optimal point at each feasible solution, and the objective function is to minimize the sum of the excess demands across the nodes. Let $\boldsymbol{s} = (s_1, s_2, \ldots, s_d)'$, and the LP is:

$$\begin{aligned} \min \quad & \sum_{i=1}^{d} x_i^+ \\ \text{s.t.} \quad & D_i - s_i + \sum_{j:(j,i)\in E} x_j^+ a_{ji} = x_i^+ - x_i^-, \forall i \in V \\ & x_i^+ \geq 0, x_i^- \geq 0, \forall i \in V. \end{aligned} \tag{1}$$

The quantity $x_i^+ \geq 0$ represents the shedded demand from node i at an optimal solution, 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 at an optimal solution. Therefore, 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 optimal shedded demands, is minimized. The problem can be expressed in matrix notation as follows. Define a matrix $A = (A(i,j) : i, j \in V)$ with $A(i,j) = a_{ij}$ (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, 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.

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.

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 $\boldsymbol{r} = \boldsymbol{D} - \boldsymbol{s}$.

We are interested in computing the probability that the network fails. Let $\alpha(k)$ represent this failure probability, and $L(\boldsymbol{D})$ denote the optimal value of the dual when the demand vector is $\boldsymbol{D}$. 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.

Since the discussion in Section 3 is valid for all k and $\boldsymbol{s}$, we do not assume them as functions of the rarity parameter n until Section 4 .

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 $\sum_{i=1}^{d} D_i \leq \sum_{i=1}^{d} s_i$.

3.2. Uniqueness and Positivity of the Solution to the Primal

THEOREM 2. *When the primal problem (2) is feasible, it 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,* $\forall 1 \leq k \leq d$.

To emphasize the main results of the paper, we postpone the formal proof to Appendix A, and only give a brief explanation here. For (a), assuming there are two optimal solutions and making use of LP duality theory, we can prove that these two solutions are the same. Part (b) can be proved by contradiction.

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

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

To prove it, we construct the solution of the dual problem and make use of Karush-Kuhn-Tucker (KKT) conditions. See Bertsimas and Tsitsiklis (1997) for more information about KKT conditions. A detailed proof appears in Appendix B.

Although Theorem 3 establishes 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 discuss the asymptotic behavior of the failure probability of this distribution network, which will be useful when we develop efficient simulation algorithms for estimating the failure probability in the next section. For our asymptotic analysis, we introduce a rarity parameter n that grows large, and let the vertices supplies and thresholds be functions of n, but the number d of vertices remains fixed. We next give the details.

We consider the distribution network to lie in $\mathbb{R}^m$ for some positive integer m, and let $t_i \in \mathbb{R}^m$, $i = 1, 2, \ldots, d$, represent the locations of d vertices in the network, and $T = \{t_1, t_2, \ldots, t_d\}$. Suppose we have positive functions $\gamma(t), \mu(t), \sigma(t)$ on T, and $\sigma^2(t, u)$ on $T \times T$. For each node i with location $t_i \in T$, there is a deterministic supply $s_n(t_i) \triangleq s(t_i) = n\gamma(t_i)$, where n is a rarity parameter, and a random demand $D(t_i) \sim N(\mu(t_i), \sigma^2(t_i))$. The covariance between the demands at two vertices with locations t_i and t_j is $cov[D(t_i), D(t_j)] = \sigma^2(t_i, t_j)$. 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 first introduce the little o notion, which is used in the theorem that will be discussed momentarily.

Definition 1. Let f and g: $\mathbb{R}^+ \to \mathbb{R}^+$ be positive functions. Then $f(x) = o(g(x))$ as $x \to \infty$ if for every constant $\epsilon > 0$, there exists a constant number N such that for all $x \geq N$, we have $f(x) \leq \epsilon g(x)$. If $g(x) > 0$ for all $x \in \mathbb{R}^+$, $f(x) = o(g(x))$ as $x \to \infty$ is equivalent to $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

We now establish a theorem that describes the asymptotic behavior of this network when supply grows faster than the threshold $k = k_n$, which results in a rare event as n gets large. 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.

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 $k = k_n \geq 0$ with $k_n = o(n)$,*

$$\lim_{n\to\infty} n^{-2} \log P\{L_n(\boldsymbol{D}) > k_n\} = \lim_{n\to\infty} n^{-2} \log P\{\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) > k_n\} \tag{5}$$

$$= -\frac{\gamma^2(t^*)}{2\sigma^2(t^*)}, \tag{6}$$

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

To prove this result, we derive upper and lower bounds with the same limit $-\frac{\gamma^2(t^*)}{2\sigma^2(t^*)}$. The details appear in Appendix C.

5. Efficient Algorithms: Importance Sampling and Conditional Monte Carlo

5.1. Asymptotic Optimality

Recall that t_i, $i = 1, 2, \ldots, d$, are the locations of the d vertices. 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 2. 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:

$$Q\{\boldsymbol{D} \in B\} = \sum_{i=1}^{d} p(i) P\{\boldsymbol{D} \in B | D(t_i) - s_n(t_i) > 0\}, \tag{7}$$

where $B \subset \mathbb{R}^d$ is a Borel set, and $p(i) = \frac{P\{D(t_i)-s_n(t_i)>0\}}{\sum_{j=1}^d P\{D(t_j)-s_n(t_j)>0\}}$.

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. In other words, we force the demand to be larger than the supply for at least one node, which ensures that the network fails more often under the new measure. Since

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

it is easy to see that

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

5.2.1. Asymptotic Optimality We next establish the asymptotic optimality of the IS approach based on Q.

THEOREM 5.

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

is an asymptotically optimal estimator for $\alpha_n(k_n) \triangleq P\{L_n(\boldsymbol{D}) > k_n\}$, *where* $k_n = o(n)$.

To prove this result, we find an upper bound of $\frac{\log E_Q[Z_n^2(\boldsymbol{D})]}{\log P\{L_n(\boldsymbol{D})>k_n\}}$ with limit 2, and make use of Theorem 4. The proof appears in Appendix D.

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 $\boldsymbol{D}^{(i)} = (D^{(i)}(t_1), D^{(i)}(t_2), \ldots, D^{(i)}(t_d))$ from distribution Q as in (7). To do this, we choose a node i with probability $p(i)$, and begin by generating untruncated normal variables and reject those if the demand of node i does not exceed its supply. If the acceptance rate becomes too small after some iterations with escalating sample sizes, we switch to use a Gibbs sampler algorithm described in Robert (1995) to sample truncated normal variables.

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

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

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

5.3. Conditional Monte Carlo

We first briefly introduce the Conditional Monte Carlo (CMC) approach, which is a variance-reduction technique. Suppose we are interested in estimating α, and U is an unbiased estimator. Let Y be an auxiliary random variable generated along with U. Then iterated expectations implies that $E[U|Y]$ is an unbiased estimator of α. Moreover, since $Var[U] = E[Var[U|Y]] + Var[E[U|Y]] \geq Var[E[U|Y]]$, using $E[U|Y]$ as an estimator may help to reduce variance.

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

$$\boldsymbol{D} = \boldsymbol{\mu} + RW\boldsymbol{\Psi}, \tag{8}$$

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 and conditioning on $\boldsymbol{\Psi}$, Blanchet, Li and Nakayama (2011) developed a conditional Monte Carlo approach for estimating $\alpha_n(k_n)$, along with algorithmic details on how to implement the method. However, we did not discuss the optimality of the CMC algorithm in the conference paper. We now provide such an analysis.

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

THEOREM 6. *For $k_n = o(n)$, there exist $n_0 > 0$, $c_3 > 0$, $s^* > 0$, $\eta_1 = O(n)$, such that when $n > n_0$, the conditional Monte Carlo estimator*

$$T_n(\mathbf{\Psi}) \triangleq P\{L_n(\boldsymbol{D}) > k_n|\mathbf{\Psi}\} \leq P\{R > ns^* + \eta_1\}, \quad \forall \|\mathbf{\Psi}\| = 1, \tag{9}$$

$$P\{L_n(\boldsymbol{D}) > k_n\} \geq c_3 P\{R > ns^* + o(n)\} n^{-(d-1)}. \tag{10}$$

Also, $T_n(\mathbf{\Psi})$ is asymptotically optimal.

To prove (9), since the dual problem is an LP, we only need to consider the extreme points of the feasible region. Making use of the polar-coordinate representation of the random demand, we show that $P\{L_n(\boldsymbol{D}) > k_n|\mathbf{\Psi}\}$ is equal to the conditional probability that the radius R is larger than a function of $\mathbf{\Psi}$, which has minimum value $ns^* + \eta_1$ when n is large enough.

To prove (10), we show that $P\{L_n(\boldsymbol{D}) > k_n\}$ is equal to the probability that radius R is larger than a function of $\mathbf{\Psi}$. We then find a lower bound by considering a small ball when n is large enough.

The asymptotic optimality follows since we have found an upper bound of $\log\left(E[T_n^2(\mathbf{\Psi})]\right)$, and a lower bound of $\log\left(P\{L_n(\boldsymbol{D}) > k_n\}\right)$ with their ratio less than or equal to 2 when n is large enough. The complete proof appears in Appendix E.

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 $RSE^2 \times CT$. For each example, assume d

locations $t_1, t_2, \cdots, t_d$ have been chosen, we give incidence matrix H, supply parameter $\boldsymbol{\gamma} = (\gamma(t_1), \gamma(t_2), \ldots, \gamma(t_d))'$, and demand parameters $\boldsymbol{\mu}, \Sigma$. We have proven the asymptotic optimality of the IS and CMC estimators when the threshold $k = k_n$ is a constant or increases with the rarity parameter n. Examples 1 and 2 show how failure probability changes with n for constant k_n. Example 3 shows how failure probability changes when k_n is a function of n, with $k = k_n = 20 \times n^{0.5}$. We set the sample size $N = 10^5$ for all of the three examples.

We choose parameters based on the following considerations:

- Network size d: we did three experiments with networks of three different sizes $d = 3$, 10, and 30. We believe that a network with 30 nodes represents a sufficiently large example for actual applications. In addition, these experiments are used to compare the relative efficiency among different simulation algorithms. While larger networks take more time to simulate, we expect that the results across the methods would be similar.
- Incidence matrix H: it was chosen so that the network is irreducible.
- Supply and demand related parameters $\boldsymbol{\gamma}, \boldsymbol{\mu}, \Sigma$: it is not easy to obtain this information from real-life examples, so we constructed them so that failure rarely happens.
- Rarity parameter n and threshold k: they were chosen so that failure probability $\alpha_n(k_n)$ exhibits different orders of magnitude. Although our results establish asymptotic optimality of the IS and CMC estimators, the experiments consider a range of parameter values to study when $\alpha_n(k_n)$ is not too small so we can assess the performance.

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

The first example, whose simulation results are in Table 1, is a 3-dimensional network with the following parameters:

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

Table 1 Results for Example 1 using Naive Simulation, IS, and CMC for $d=3$, fixed k_n.

	Naive Simulation		Importance Sampling		Conditional MC	
n	$\alpha_n(k_n)$	$RSE^2\times CT$	$\alpha_n(k_n)$	$RSE^2\times CT$	$\alpha_n(k_n)$	$RSE^2\times CT$
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}

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

The second example, whose simulation results are in Table 2, is a 10-dimensional network with the following parameters:

$$H=\begin{pmatrix}
0&1&1&0&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&1&0&0\\
0&0&0&0&1&0&1&0&0&0\\
0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&0&0&0&1\\
1&0&0&0&0&0&0&0&0&0
\end{pmatrix},\boldsymbol{\gamma}=\begin{pmatrix}3\\5\\3\\3\\3\\3\\3\\3\\3\\15\end{pmatrix},\boldsymbol{\mu}\begin{pmatrix}1\\5\\1\\1\\1\\1\\1\\1\\1\\1\end{pmatrix},\Sigma=\begin{pmatrix}
0.5&0.3&0.3&0.25&0.2&0.15&0.2&0.25&0.2&0.15\\
0.3&0.5&0.25&0.2&0.15&0.1&0.15&0.2&0.15&0.1\\
0.3&0.25&0.5&0.3&0.25&0.2&0.25&0.3&0.25&0.2\\
0.25&0.2&0.3&0.5&0.3&0.25&0.3&0.25&0.2&0.15\\
0.2&0.15&0.25&0.3&0.5&0.3&0.25&0.2&0.15&0.1\\
0.15&0.1&0.2&0.25&0.3&0.5&0.3&0.25&0.2&0.15\\
0.2&0.15&0.25&0.3&0.25&0.3&0.5&0.3&0.25&0.2\\
0.25&0.2&0.3&0.25&0.2&0.25&0.3&0.5&0.3&0.25\\
0.2&0.15&0.25&0.2&0.15&0.2&0.25&0.3&0.5&0.3\\
0.15&0.1&0.2&0.15&0.1&0.15&0.2&0.25&0.3&0.5
\end{pmatrix},k_n=2.$$

6.3. Example 3: $d=30$, k_n changes with n

The third example, whose simulation results are in Table 3, is a 30-dimensional network with the following parameters:

$H(i,i+1)=1, i=1,2,\ldots,29.$ $H(30,1)=1$. All other elements of H are equal to 0.

Table 2 Results for Example 2 using Naive Simulation, IS, and CMC for $d = 10$, fixed k_n.

	Naive Simulation		Importance Sampling		Conditional MC	
n	$\alpha_n(k_n)$	$RSE^2 \times CT$	$\alpha_n(k_n)$	$RSE^2 \times CT$	$\alpha_n(k_n)$	$RSE^2 \times CT$
1.0	3.64×10^{-2}	1.21×10^{-1}	3.67×10^{-2}	9.57×10^{-2}	3.66×10^{-2}	2.00×10^{-1}
1.3	3.05×10^{-3}	1.39×10^{0}	3.38×10^{-3}	2.09×10^{-1}	3.38×10^{-3}	6.85×10^{-1}
1.5	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}
1.8	0	NaN	7.34×10^{-7}	5.24×10^{1}	7.26×10^{-7}	6.87×10^{0}

$\gamma(t_i) = 2, \mu(t_i) = 1, i = 1, 2, \ldots, 30.$ $k_n = 20 \times n^{0.5}$.

$\Sigma(i,i) = \sigma^2(t_i, t_i) = 1, i = 1, 2, \ldots, 30$. All other elements of Σ are equal to 0.4.

Table 3 Results for Example 3 using Naive Simulation, IS, and CMC for $d = 30$, k_n increases with n.

	Naive Simulation		Importance Sampling		Conditional MC	
n	$\alpha_n(k_n)$	$RSE^2 \times CT$	$\alpha_n(k_n)$	$RSE^2 \times CT$	$\alpha_n(k_n)$	$RSE^2 \times CT$
1.20	3.29×10^{-2}	2.09×10^{-1}	3.22×10^{-2}	2.94×10^{-1}	3.23×10^{-2}	5.96×10^{-1}
1.50	2.72×10^{-3}	2.16×10^{0}	2.58×10^{-3}	1.06×10^{0}	2.61×10^{-3}	2.96×10^{0}
1.70	2.80×10^{-4}	2.03×10^{1}	3.03×10^{-4}	3.33×10^{0}	3.03×10^{-4}	1.20×10^{1}
1.95	1.00×10^{-5}	5.78×10^{2}	1.18×10^{-5}	2.34×10^{1}	1.17×10^{-5}	4.47×10^{1}
2.05	0	NaN	2.92×10^{-6}	6.39×10^{1}	3.02×10^{-6}	9.92×10^{1}
2.16	0	NaN	3.83×10^{-7}	3.07×10^{2}	3.84×10^{-7}	2.15×10^{2}

6.4. Discussion of Results and Comparisons Between Algorithms

1. When 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, 2, and 3, when n is very large, we do not get even one observation of the event $\{L_n(\boldsymbol{D}) \geq k_n\}$. However,

although the performance of CMC becomes worse as well, it does not deteriorate as quickly as the other two. No matter how large n is, we can obtain a non-zero estimate of $\alpha_n(k_n)$.

2. Although both IS and CMC are asymptotically optimal, when 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. Also, generating demand vector $\boldsymbol{D}^{(i)}$ from distribution Q is likely to give a non-zero $Z_n(\boldsymbol{D}^{(i)})$, which makes RSE relatively small. In contrast, our conditional Monte Carlo method needs to solve several optimization problems to find the roots R_i^* that equate the optimal value of the primal and the threshold k_n for a fixed angle $\boldsymbol{\Psi}$ (see equation (8) in Blanchet, Li and Nakayama 2011) in each replication i, which incurs larger CT. Thus, the added computational effort required by CMC can lead to it performing worse than IS. However, as n increases, conditional Monte Carlo method works much better. The larger n is, the bigger the advantage CMC has compared to naive simulation. The advantage arises from the fact that although IS can still save some CT since it requires a smaller number of optimizations, generating demand vector $\boldsymbol{D}^{(i)}$ from distribution Q is likely to give a zero $Z_n(\boldsymbol{D}^{(i)})$, which results in a large RSE. The significant variance reduction obtained by CMC for large n overwhelms the additional computational effort. In conclusion, for a given network, IS performs best when n is small, and CMC is better when n is large.

3. We have established the asymptotic optimality of our methods as the rarity parameter $n \to \infty$. But as with any technique for which an asymptotic property has been proven, the performance for fixed n when the asymptotics are not yet in effect may differ from that for large n, and may not outperform naive simulation. For instance, this is the case for $n = 1.20$ in Table 3. We explore this by varying n in our experiments.

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 optimal point is determined by a linear program whose objective is to minimize 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 the rarity parameter n is large.

We would like to further discuss the following extensions to our current results, as well as some related topics.

- Cost Structure: We assume unit cost associated with pushing unit demand from one node to another. In other words, let c_{ij} be the cost by distributing unit demand from node i to node j. Currently, $c_{ij} = 1$ for all $(i,j) \in E$. We can generalize this setting by using a path dependent cost structure and allowing c_{ij} to differ for different (i,j). At the same time, the objective function of the primal problem (2) now becomes: min $\sum_{i=1}^{d} \sum_{j=1}^{d} x_i^+ a_{ij} c_{ij}$.

Here, we claim that, all theorems in the paper are still valid for the generalized structure as long as $c_{ij} > 0$ for all $(i,j) \in E$. To see this, Theorem 3 has generalized the cost structure for Theorems 1 and 2. We can also prove Theorems 4, 5, and 6 with straightforward modifications.

- Objective Function: In this paper, the objective function is set as the sum of the excess demands across all nodes, i.e. min $\sum_{i=1}^{d} x_i^+$. We would like to discuss two extensions that make our IS and CMC algorithms applicable to a wide range of problems.

1) min $\sum_{i,j:(i,j)\in E} x_{ij}^+ c_{ij}$, where x_{ij}^+ represents the excess demand from node i to node j, and c_{ij} has been defined in the last bullet point. In other words, we can make the decision variables have a more granular structure. Instead of considering the total excess demands of each node, we can let the LP decide how those demands are allocated from each node to its neighbors. Correspondingly, the constraints should be rewritten as:

$$\text{s.t. } D_i - s_i + \sum_{j:(j,i)\in E} x_{ji}^+ = \sum_{j:(i,j)\in E} x_{ij}^+ - \sum_{j:(i,j)\in E} x_{ij}^-, \forall i \in V$$

$$x_{ij}^+ \geq 0, x_{ij}^- \geq 0, \forall (i,j) \in E,$$

where x_{ij}^- represents the unused supply that can be distributed from node i to node j. The asymptotic optimality of both IS and CMC can be proved in a similar way to Theorems 5 and 6.

2) min $\sum_{i=1}^{d} f(\boldsymbol{x}^+)$, where $f(\cdot)$ is differentiable and increasing with respect to $\boldsymbol{x}^+$, lower bounded by a linear and increasing function of $\boldsymbol{x}^+$, and $f(\mathbf{0}) = 0$. In other words, $f(\cdot)$ doesn't need to be a linear function, and as long as the above conditions are satisfied, the asymptotic optimality of both IS and CMC can be established. The former can be proved in a similar way to Theorem 5. The latter can be proved by making use of (11) and (13), and the proof outline appears in Appendix F.

- Elliptical Copula: 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 calculate the root for the optimal value of the primal as a function of R conditional on the angular part, Ψ. Therefore the conditional Monte Carlo algorithm applies as long as the demand vector D is an elliptical copula. In order to prove the asymptotic optimality, some regularity properties are needed, which are similar to what are stated in Theorem 1 in Blanchet and Rojas-Nandayapa (2011).

- Growing Number of Nodes: In this paper, all of our discussion focuses on a given graph with a fixed number of nodes. We can also consider the asymptotic behavior of a graph when the number of nodes grows large. Similar properties and simulation algorithms can be developed by embedding the Gaussian vector of demands in a continuous Gaussian random field, so that Borell-TIS inequality (Adler and Taylor 2007, p. 50) can be applied in the proof of Theorem 4.

- IS vs. CMC as the Dimension (d) Changes: In this paper, we presented two asymptotically efficient estimators for fixed d, and we advocate using both because they have advantages and disadvantages. For example, CMC is always guaranteed to reduce variance. However, as the dimension of the problem increases, the angular component of the demand distribution, which is the variable on which we condition to apply the CMC, plays a more important role in the occurrence of the rare event. Because the contribution of the angular component to the rare event of interest is not fully accounted for in the CMC estimator (which only integrates the contribution of the radial component), then we expect the relative mean squared error of CMC to degrade as the dimension increases. On the other hand, the importance sampling estimator is closely related to estimators

which are known to work well in continuous Gaussian random fields (see Adler, Blanchet and Liu (2012)), and may perform better than CMC in the context of very small probabilities. Due to these trade-offs, we suggest the users to consider using both estimators.

- Potential Application in the Area of Cascading Failures: Our IS strategy (discussed in Section 5.2) can be modified to estimate the probability of observing a cascading failure under the policy advocated by Bienstock, Blanchet and Li (2016). We can, e.g., change the objective function to minimize instead the worst case temperature over the lines in the network. Non-linear constraints stated in that paper need to be modified or approximated by linear constraints.

Acknowledgments

Jose Blanchet acknowledges support from NSF grants CMMI-1069064, CMMI-1436700, DMS-1820942 and DARPA grant N660011824028.

The work of Marvin K. Nakayama has been supported in part by the NSF under Grants No. CMMI-0926949, and CMMI-1200065. 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.

Appendix A: Proof of Theorem 2

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

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

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 $1 \le j \le 2d$. Consider the corresponding dual problem:

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

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 $1 \le j \le 2d$, the primal is infeasible. Hence, each element of d must be 0, which means that $\boldsymbol{x}_1 = \boldsymbol{x}_2$, proving part (a). Note that the objective function of the LP can be of multiple forms since we only aim to prove the infeasibility, and different choice of the objective function only leads to different construction of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

To establish (b), suppose $(\boldsymbol{x}^+, \boldsymbol{x}^-)$ is the optimal solution of the primal (2). Suppose for some $1 \le k \le d$, 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, $\forall 1 \le k \le d$. $\square$

Appendix B: Proof of Theorem 3

Proof: Consider the problem

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

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

$$\begin{aligned} &\nabla_{\boldsymbol{x}^+} f + (A - I)\boldsymbol{\alpha} - \boldsymbol{\mu} = \nabla_{\boldsymbol{x}^+} f + (A - I)\boldsymbol{\lambda} - \boldsymbol{\mu} = \nabla_{\boldsymbol{x}^+} f + (A - I)Q\boldsymbol{\xi} - (I - Q)\boldsymbol{\xi} = 0 \\ &\Rightarrow [(I - Q) - (A - I)Q]\boldsymbol{\xi} = \nabla_{\boldsymbol{x}^+} f \Rightarrow (I - AQ)\boldsymbol{\xi} = \nabla_{\boldsymbol{x}^+} f. \end{aligned}$$

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, $x_i^{*+} > 0, \forall 1 \leq i \leq d$. 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 $x_i^{*-} = 0, \forall 1 \leq i \leq d$. 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 $(I - AQ)^{-1} \geq 0$, we have that $\boldsymbol{\xi} \geq 0$. 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 $f = f_2$. □

Appendix C: Proof of Theorem 4

Proof: We will prove this result by establishing upper and lower bounds on $P\{L_n(\boldsymbol{D}) > k_n\}$. We start with deriving an upper bound. Note that $h(t) \triangleq \frac{D(t) - \mu(t)}{\sigma(t)}$ follows standard Gaussian distribution. We first claim that

$$\{L_n(\boldsymbol{D}) > k_n\} \subseteq \{\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) > 0\}. \tag{11}$$

To see this, if we assume $\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) \leq 0$, then $D(t_i) \leq s(t_i), \forall i = 1, 2, \ldots, d$. According to Theorem 1(b), the primal problem (2) is feasible, and it is easy to see that $x_i^+ = 0, x_i^- = s_n(t_i) -$

$D(t_i) \geq 0, \forall i = 1, 2, \ldots, d$, is an optimal solution to the primal problem. In this case $L_n(\boldsymbol{D}) = 0$. Thus $\{\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) > 0\}^c \subseteq \{L_n(\boldsymbol{D}) > k_n\}^c$, where "$c$" represents the complement of a set, and Equation (11) is valid. Therefore,

$$P\{L_n(\boldsymbol{D}) > k_n\} \leq P\{\max_{i=1,\ldots,d} \frac{D(t_i) - s_n(t_i)}{\sigma(t_i)} > 0\} = P\{\max_{t\in T}(h(t) - \frac{s_n(t) - \mu(t)}{\sigma(t)}) > 0\}.$$

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

$$\begin{aligned} P\{L_n(\boldsymbol{D}) > k_n\} &\leq P\{\max_{t\in T} h(t) > \frac{n\gamma(t^*)}{\sigma(t^*)} - \frac{\mu(\hat{t})}{\sigma(\hat{t})}\} \\ &\leq \bar{C} \exp\{-\frac{1}{2}(\frac{n\gamma(t^*)}{\sigma(t^*)} - \frac{\mu(\hat{t})}{\sigma(\hat{t})})^2\}, \end{aligned} \tag{12}$$

where $\bar{C}$ is some positive constant, and the last step makes use of the fact that if a random variable X follows standard Gaussian distribution, then for any $x > 0$, $P\{X > x\} \leq \frac{\exp\{-x^2/2\}}{x\sqrt{2\pi}}$. This establishes the desired upper bound on $P\{L_n(\boldsymbol{D}) > k_n\}$.

To obtain a lower bound on the probability, define $g(t) \triangleq \frac{1}{\sqrt{2\pi}} \frac{\sigma(t)}{s_n(t) - \mu(t) + k_n} \exp\{-\frac{(s_n(t) - \mu(t) + k_n)^2}{2\sigma^2(t)}\}$, $t \in T$, where $k_n \geq 0$ is some constant. We now claim that

$$P\{L_n(\boldsymbol{D}) > k_n\} \geq P\{\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) > k_n\}. \tag{13}$$

To see this, note that if $\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) > k$, then there exists some $1 \leq i_0 \leq d$ such that $D(t_{i_0}) - s_n(t_{i_0}) > k_n$. 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 $\boldsymbol{y}'(\boldsymbol{D} - \boldsymbol{s}) = D(t_{i_0}) - s_n(t_{i_0}) > k$. Therefore, $L_n(\boldsymbol{D}) > k_n$. Then,

$$\begin{aligned} P\{L_n(\boldsymbol{D}) > k_n\} &\geq P\{\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) > k_n\} \qquad (14) \\ &\geq P\{D(t^*) - s_n(t^*) > k_n\} \\ &\geq \frac{1}{\sqrt{2\pi}} \frac{\sigma(t^*)}{s_n(t^*) - \mu(t^*) + k_n} \exp\{-\frac{(s_n(t^*) - \mu(t^*) + k_n)^2}{2\sigma^2(t^*)}\} \\ &= g(t^*)C, \qquad (15) \end{aligned}$$

where C is some positive constant, and the second-to-last step applied the fact that if a random variable $X \sim N(\bar{\mu}, \bar{\sigma}^2)$, where $\bar{\sigma} > 0$, then for all $\alpha > \bar{\mu}$,

$$P\{X > \alpha\} \geq \frac{1}{\sqrt{2\pi}} \frac{\bar{\sigma}}{\alpha - \bar{\mu}} \exp\{-\frac{(\alpha - \bar{\mu})^2}{2\bar{\sigma}^2}\}, \tag{16}$$

giving us the desired lower bound on $P\{L_n(\boldsymbol{D}) > k_n\}$.

Therefore, (12), (14), and (15) imply for n sufficiently large,

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}}\frac{\sigma(t^*)}{n\gamma(t^*)-\mu(t^*)+k_n}\exp\{-\frac{(n\gamma(t^*)-\mu(t^*)+k_n)^2}{2\sigma^2(t^*)}\}C\\
&\le P\{\max_{i=1,\ldots,d} D(t_i)-s_n(t_i)>k_n\}\\
&\le P\{L_n(\boldsymbol{D})>k_n\}\le\exp\{-\frac{1}{2}(\frac{n\gamma(t^*)}{\sigma(t^*)}-\frac{\mu(\hat{t})}{\sigma(\hat{t})})^2\}\bar{C}.
\end{aligned}$$

Taking logarithms, we have

$$\begin{aligned}
&\log[\frac{1}{\sqrt{2\pi}}\frac{\sigma(t^*)}{n\gamma(t^*)-\mu(t^*)+k_n}]-\frac{(n\gamma(t^*)-\mu(t^*)+k_n)^2}{2\sigma^2(t^*)}+\log C]\\
&\le\log P\{\max_{i=1,\ldots,d} D(t_i)-s_n(t_i)>k_n\}\\
&\le\log P\{L_n(\boldsymbol{D})>k_n\}\le-\frac{1}{2}(\frac{n\gamma(t^*)}{\sigma(t^*)}-\frac{\mu(\hat{t})}{\sigma(\hat{t})})^2+\log\bar{C}.
\end{aligned}$$

Because

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n^2}\big(\log[\frac{1}{\sqrt{2\pi}}\frac{\sigma(t^*)}{n\gamma(t^*)-\mu(t^*)+k_n}]-\frac{(n\gamma(t^*)-\mu(t^*)+k_n)^2}{2\sigma^2(t^*)}+\log C\big)\\
&=\lim_{n\to\infty}-\frac{1}{2n^2}\frac{(n\gamma(t^*)-\mu(t^*)+k_n)^2}{\sigma^2(t^*)}=-\frac{\gamma^2(t^*)}{2\sigma^2(t^*)},
\end{aligned}$$

it follows that

$$\lim_{n\to\infty}\frac{1}{n^2}\log P\{L_n(\boldsymbol{D})>k_n\}=\lim_{n\to\infty}\frac{1}{n^2}\log P\{\max_{i=1,\ldots,d} D(t_i)-s_n(t_i)>k_n\}=-\frac{\gamma^2(t^*)}{2\sigma^2(t^*)},$$

thereby verifying (5) and (6).

Appendix D: Proof of Theorem 5

Proof: Let E_Q denote the expectation under Q, so by (11), we have

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

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

$$\begin{aligned}
\frac{dP}{dQ}I\{\max_{i=1,\ldots,d} D(t_i)-s_n(t_i)>0\}&=\frac{\sum_{j=1}^d P\{D(t_j)-s_n(t_j)>0\}}{\sum_{j=1}^d I\{D(t_j)-s_n(t_j)>0\}}I\{\max_{i=1,\ldots,d} D(t_i)-s_n(t_i)>0\}\\
&\le\sum_{j=1}^d P\{D(t_j)-s_n(t_j)>0\}.
\end{aligned}$$

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

Since $P\{\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) > 0\} \le \sum_{j=1}^d P\{D(t_j) - s_n(t_j) > 0\} \le d \times P\{\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) > 0\}$, we have

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

Therefore,

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

where the last equation follows from Theorem 4. □

Appendix E: Proof of Theorem 6

Proof: We first prove (9). Let $\Omega = \{\boldsymbol{y} : M\boldsymbol{y} \le \mathbf{1}, \boldsymbol{y} \ge \mathbf{0}\}$ denote the feasible region of the dual problem (3). Then $L_n(\boldsymbol{D}) = \max \boldsymbol{y}'(\boldsymbol{\mu} + RW\boldsymbol{\Psi} - n\boldsymbol{\gamma}), \boldsymbol{y} \in \Omega$, 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 $\mathbf{1}'(\boldsymbol{\mu} + RW\boldsymbol{\Psi} - n\boldsymbol{\gamma}) > 0$. The other case is that the primal problem is feasible but the optimal value is greater than k_n. Since the dual problem is an LP, for the second case, we can focus on the extreme points of the feasible region Ω. Since $k_n \ge 0$, 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

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

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

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

Let $n_1 = \max\{0, \max_{i=0,1,\ldots,m} \frac{\tilde{\boldsymbol{y}}_i'\boldsymbol{\mu} - k_i}{\tilde{\boldsymbol{y}}_i'\boldsymbol{\gamma}}\}$.Then when $n > n_1$, we have $n\tilde{\boldsymbol{y}}_i'\boldsymbol{\gamma} - \tilde{\boldsymbol{y}}_i'\boldsymbol{\mu} + k_i > 0$. Recall that R is a positive random variable, so

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

Define

$$\Gamma_0 = \{\boldsymbol{\Psi} : \|\boldsymbol{\Psi}\| = 1, \max_{i=0,1,\ldots,m} \tilde{\boldsymbol{y}}_i'W\boldsymbol{\Psi} > 0\}, \quad M_{\boldsymbol{\Psi}} = \{i = 0, 1, \ldots, m : \tilde{\boldsymbol{y}}_i'W\boldsymbol{\Psi} > 0\}.$$

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

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

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

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

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

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

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

Let us consider inequality (9) 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

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

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

$$\inf_{\boldsymbol{\Psi} \in \Gamma_a} S(\boldsymbol{\Psi}) = S(\boldsymbol{\Psi}^*) = \frac{\tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}^*}'\boldsymbol{\gamma}}{\tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}^*}' W \boldsymbol{\Psi}^*},$$

$$\inf_{\boldsymbol{\Psi} \in \Gamma_a} \min_{i \in M_{\boldsymbol{\Psi}}} \frac{-\tilde{\boldsymbol{y}}_i'\boldsymbol{\mu} + k_i}{\tilde{\boldsymbol{y}}_i' W \boldsymbol{\Psi}} = \eta_1. \tag{18}$$

Therefore, $\inf_{\boldsymbol{\Psi} \in \Gamma_a} H(\boldsymbol{\Psi}, n) \geq nS(\boldsymbol{\Psi}^*) + \eta_1$. Then we have $T_n(\boldsymbol{\Psi}) \leq P\{R > nS(\boldsymbol{\Psi}^*) + \eta_1\}$.

Let $s^* \triangleq S(\boldsymbol{\Psi}^*)$, then (9) is established.

Now we consider the inequality (10). We claim that for any $\boldsymbol{\Psi}$ in Γ_a, there exists $n_2(\boldsymbol{\Psi}) > 0$ such that when $n > n_2(\boldsymbol{\Psi})$,

$$H(\boldsymbol{\Psi}, n) = nS(\boldsymbol{\Psi}) + \frac{k_{\boldsymbol{\Psi}} - \tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}}'\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}}' W \boldsymbol{\Psi}}, \tag{19}$$

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

$$\lambda_i \triangleq nS(\boldsymbol{\Psi}) + \frac{k_{\boldsymbol{\Psi}} - \tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}}'\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}}' W \boldsymbol{\Psi}} - \frac{n\tilde{\boldsymbol{y}}_i'\boldsymbol{\gamma} - \tilde{\boldsymbol{y}}_i'\boldsymbol{\mu} + k_i}{\tilde{\boldsymbol{y}}_i' W \boldsymbol{\Psi}} = n\big(S(\boldsymbol{\Psi}) - \frac{\tilde{\boldsymbol{y}}_i'\boldsymbol{\gamma}}{\tilde{\boldsymbol{y}}_i' W \boldsymbol{\Psi}}\big) + \big(\frac{k_{\boldsymbol{\Psi}} - \tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}}'\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}}' W \boldsymbol{\Psi}} - \frac{k_i - \tilde{\boldsymbol{y}}_i'\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}_i' W \boldsymbol{\Psi}}\big).$$

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

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

Choose

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

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

$$S(\mathbf{\Psi}) - \frac{\tilde{\boldsymbol{y}}_i'\boldsymbol{\gamma}}{\tilde{\boldsymbol{y}}_i'W\mathbf{\Psi}} \le -\eta_2(\mathbf{\Psi}), \quad -\eta_3(\mathbf{\Psi}) \le \frac{k_{\mathbf{\Psi}} - \tilde{\boldsymbol{y}}_{\mathbf{\Psi}}'\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}_{\mathbf{\Psi}}'W\mathbf{\Psi}} - \frac{k_i - \tilde{\boldsymbol{y}}_i'\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}_i'W\mathbf{\Psi}} \le \eta_3(\mathbf{\Psi}).$$

Since $k_n = o(n)$, there exists $n_2(\mathbf{\Psi}) > 0$, such that when $n > n_2(\mathbf{\Psi})$, $\lambda_i \le 0, \forall i \in \mathcal{I}_{\mathbf{\Psi}}^-$. Therefore, when $n > \max\{n_1, n_2(\mathbf{\Psi}^*)\}$, it follows that $\lambda_i \le 0, \forall i \in M_{\mathbf{\Psi}^*}$, so

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

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 $H(\mathbf{\Psi}, n) - H(\mathbf{\Psi}^*, n) \le (nc_1 + c_2)\|\mathbf{\Psi} - \mathbf{\Psi}^*\|$ 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

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

Since $\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}^*$ 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 $\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}^* \ge (\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}^*)^2 - c_0 > 0$, where c_0 is some positive constant.

Define $c_1 = \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'\boldsymbol{\gamma}\frac{\|W'\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}\|}{(\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}^*)^2 - c_0} > 0$, $c_2 = (k_{\mathbf{\Psi}^*} - \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'\boldsymbol{\mu})\frac{\|W'\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}\|}{(\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}^*)^2 - c_0}$. Since $k_n = o(n)$, there exists $n_3(\mathbf{\Psi}^*) > 0$, such that when $n > \max\{n_1, n_2(\mathbf{\Psi}^*), n_3(\mathbf{\Psi}^*)\}$, we have $nc_1 + c_2 > 0$. Therefore,

$$\begin{aligned}
H(\mathbf{\Psi}, n) - H(\mathbf{\Psi}^*, n) &\le (n\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'\boldsymbol{\gamma} - \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'\boldsymbol{\mu} + k_{\mathbf{\Psi}^*})\frac{\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W(\mathbf{\Psi}^* - \mathbf{\Psi})}{\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}^*} \\
&\le (n\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'\boldsymbol{\gamma} - \tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'\boldsymbol{\mu} + k_{\mathbf{\Psi}^*})\frac{\|W'\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}\|\|\mathbf{\Psi}^* - \mathbf{\Psi}\|}{(\tilde{\boldsymbol{y}}_{\mathbf{\Psi}^*}'W\mathbf{\Psi}^*)^2 - c_0} \\
&= (nc_1 + c_2)\|\mathbf{\Psi}^* - \mathbf{\Psi}\|.
\end{aligned}$$

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

$$H(\mathbf{\Psi}, n) \le H(\mathbf{\Psi}^*, n) + (nc_1 + c_2)\delta. \tag{21}$$

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 $\delta = n^{-1}$. By equations (17) and (21), it follows that

$$\begin{aligned} P\{L_n(\boldsymbol{D}) > k_n\} &= P\{R > H(\mathbf{\Psi}, n)\} \\ &\ge P\{R > H(\mathbf{\Psi}^*, n) + (nc_1 + c_2)\delta, \| \mathbf{\Psi} - \mathbf{\Psi}^* \| \le \delta\} \\ &\ge c_3 P\{R > H(\mathbf{\Psi}^*, n) + (nc_1 + c_2)\delta\}\delta^{(d-1)} \\ &= c_3 P\{R > nS(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*}\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*} W \mathbf{\Psi}^*} + (c_1 + c_2 n)\} n^{-(d-1)} \\ &= c_3 P\{R > nS(\mathbf{\Psi}^*) + o(n)\} n^{-(d-1)}. \end{aligned} \tag{22}$$

Hence, we have proven (10).

We now establish the last part of the theorem. By (9) and (22), we have

$$\begin{aligned} &\frac{\log\left(E[T_n^2(\mathbf{\Psi})]\right)}{\log\left(P\{L(\boldsymbol{D}) > k_n\}\right)} \le \frac{\log\left(P^2\{R > nS(\mathbf{\Psi}^*) + \eta_1\}\right)}{\log\Big(c_3 P\{R > nS(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*}\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*} W \mathbf{\Psi}^*} + (c_1 + c_2 n^{-1})\} n^{-(d-1)}\Big)} \\ &= \frac{2\log\left(P\{R > nS(\mathbf{\Psi}^*) + \eta_1\}\right)}{\log c_3 + \log\Big(P\{R > nS(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*}\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*} W \mathbf{\Psi}^*} + (c_1 + c_2 n^{-1})\}\Big) - (d-1)\log n} \\ &= 2\Big(\frac{\log\Big(P\{R > nS(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*}\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*} W \mathbf{\Psi}^*} + (c_1 + c_2 n^{-1})\}\Big)}{\log\left(P\{R > nS(\mathbf{\Psi}^*) + \eta_1\}\right)} + \frac{\log c_3 - (d-1)\log n}{\log\left(P\{R > nS(\mathbf{\Psi}^*) + \eta_1\}\right)}\Big)^{-1}. \end{aligned} \tag{23}$$

Recall that $nc_1 + c_2 > 0$ when $n > \max\{n_1, n_2(\mathbf{\Psi}^*), n_3(\mathbf{\Psi}^*)\}$, so (18) implies

$$\eta_1 = \inf_{\mathbf{\Psi} \in \Gamma_a} \inf_{i \in M_{\mathbf{\Psi}}} \frac{-\tilde{\boldsymbol{y}}'_i \boldsymbol{\mu} + k_i}{\tilde{\boldsymbol{y}}'_i W \mathbf{\Psi}} \le \frac{k_{\mathbf{\Psi}^*} - \tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*}\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*} W \mathbf{\Psi}^*} + (c_1 + c_2 n^{-1}).$$

Therefore,

$$P\{R > nS(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*}\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*} W \mathbf{\Psi}^*} + (c_1 + c_2 n^{-1})\} \le P\{R > nS(\mathbf{\Psi}^*) + \eta_1\},$$

and

$$\frac{\log P\{R > nS(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*}\boldsymbol{\mu}}{\tilde{\boldsymbol{y}}'_{\mathbf{\Psi}^*} W \mathbf{\Psi}^*} + (c_1 + c_2 n^{-1})\}}{\log P\{R > nS(\mathbf{\Psi}^*) + \eta_1\}} \ge \frac{\log P\{R > nS(\mathbf{\Psi}^*) + \eta_1\}}{\log P\{R > nS(\mathbf{\Psi}^*) + \eta_1\}} = 1.$$

When $n > n_4 = e^{\log c_3/(d-1)}$, the second term inside the parentheses in (23) is non-negative. Then when $n > n_0 = \max\{n_1, n_2(\mathbf{\Psi}^*), n_3(\mathbf{\Psi}^*), n_4\}$, it follows that (23) is bounded above by 2, thereby concluding the result. $\square$

Appendix F: Proof Outline of the Asymptotic Optimality of CMC in the Second Extension

Proof: First recall our definition of $T_n(\Psi)$ in (9). Then as in Theorem 6, the asymptotic optimality will hold if we can find an upper bound of $\log\left(E[T_n^2(\boldsymbol{\Psi})]\right)$, and a lower bound of $\log\left(P\{L_n(\boldsymbol{D}) > k_n\}\right)$ with their ratio less than or equal to 2 when n is large enough.

To obtain an upper bound, we make use of (11) to get

$$P\{L_n(\boldsymbol{D}) > k_n\} \leq P\{\max_{i=1,\ldots,d} D(t_i) - s_n(t_i) > 0\}.$$

It is not difficult to see that for $k_n = o(n)$, when n is large enough, there exist $s^* > 0$, and $\eta_1 > 0$ such that the conditional Monte Carlo estimator satisfies

$$T_n(\boldsymbol{\Psi}) \leq P\{R > ns^* + \eta_1\}, \quad \forall \|\boldsymbol{\Psi}\| = 1. \tag{24}$$

To obtain a lower bound, first recall (13). It can be proved that when n is large enough, there exists $c_3 > 0$, such that

$$P\{L_n(\boldsymbol{D}) > k_n\} \geq c_3 P\{R > ns^* + o(n)\} n^{-(d-1)}. \tag{25}$$

By (24) and (25), following steps similar to those applied in the proof of Theorem 6 after (22) leads to the desired result. □

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.

Belmudes, F., D. Ernst, L. Wehenkel. 2008. Cross-entropy based rare-Event simulation for the identification of dangerous events in power systems. Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS 2008), 1-7.

Bertsimas, D. and N. Tsitsiklis. 1997. Introduction to Linear Optimization. Athena Scientific, Massachusetts.

Bienstock, D., J. Blanchet, J. Li. 2016. Stochastic models and control for electrical power line temperature. Energy Systems 7 1 173-192.

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. Proceedings of the 2011 Winter Simulation Conference (WSC), 3832-3843.

Blanchet, J. and L. Rojas-Nandayapa. 2011. Efficient simulation of tail probabilities of sums of dependent random variables. J. Appl. Probab. 48A, 147-164.

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.

Brown, G., M. Carlyle, J. Salmerón, K. Wood. 2006. Defending critical infrastructure. Interfaces, 36 6 530-544.

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.

Lisser, A., A. Ouorou, J. Vial, J. Gondzio. 1999. Capacity planning under uncertain demand in telecommunications networks. https://archive-ouverte.unige.ch/unige:5871.

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.

Perninge, M., F. Lindskog, L. Söder. 2012. Importance sampling of injected powers for electric power system security analysis. IEEE Transactions on Power Systems 27 1 3-11.

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

Wadman, W.S., D.T. Crommelin, J.E. Frank. 2013. Applying a splitting technique to estimate electrical grid reliability. Proceedings of the 2013 Winter Simulation Conference, 577-588.

Wang, S.P., A. Chen, C.W. Liu, C.H. Chen, J. Shortle, J.Y. Wu. 2015. Efficient splitting simulation for blackout analysis. IEEE Transactions on Power Systems, 30 4 1775-1783.

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

Zeineldin, H.H., K. Bhattacharya, E.F. El-Saadany, M.M.A. Salama. 2006. Impact of intentional islanding of distributed generation on electricity market prices. IEE Proceedings - Generation, Transmission and Distribution, 153 2 147-154.

Jose Blanchet holds a Ph.D. in Management Science and Engineering from Stanford University. Prior to joining MS&E he was a faculty member of Columbia and Harvard University. Jose is a recipient of the 2009 Best Publication Award given by the INFORMS Applied Probability Society and of the 2010 Erlang Prize. He also received a PECASE award given by NSF in 2010. He worked as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability and Monte Carlo methods. He serves on the editorial board of Advances in Applied Probability, Journal of Applied Probability, Mathematics of Operations Research, QUESTA, Stochastic Models, and Stochastic Systems.

Juan Li is a vice president at AQR Capital Management. Prior to joining AQR, she was a vice president at Goldman Sachs. Juan is a recipient of the 2011 INFORMS Simulation Society CUMW Awards. She obtained her Ph.D. and M.S. in operations research from Columbia University and her B.S. in mathematics from Peking University.

Marvin K. Nakayama is a professor of computer science at the New Jersey Institute of Technology. His research interests include simulation, modeling, applied probability, statistics, risk analysis, and energy.