Stochastic Networks: Modeling, Simulation Design and Risk Control

Juan Li

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

Stochastic Networks: Modeling, Simulation Design and Risk Control

Juan Li

This dissertation studies stochastic network problems that arise in various areas with important industrial applications. Due to uncertainty of both external and internal variables, these networks are exposed to the risk of failure with certain probability, which, in many cases, is very small. It is thus desirable to develop efficient simulation algorithms to study the stability of these networks and provide guidance for risk control.

Chapter 2 models 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. Assuming that the demands are random (following a jointly Gaussian law), we study the probability that the optimal cost 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 that demonstrate the strong performance of our algorithms.

Chapter 3 studies an insurance-reinsurance network model that deals with default contagion risks with a particular aim of capturing cascading effects at the time of defaults. We capture these effects by finding an equilibrium allocation of settlements that can be found as the unique optimal solution of an optimization problem. We are able to obtain an asymptotic description of the most likely ways in which the default of a specific group

of participants can occur, by solving a multidimensional Knapsack integer programming problem. We also propose a class of strongly efficient Monte Carlo estimators for computing the expected loss of the network conditioned on the failure of a specific set of companies.

Chapter 4 discusses control schemes for maintaining low failure probability of a transmission system power line. We construct a stochastic differential equation to describe the temperature evolution in a line subject to stochastic exogenous factors such as ambient temperature, and present a solution to the resulting stochastic heat equation. A number of control algorithms designed to limit the probability that a line exceeds its critical temperature are provided.

Contents

List of Figures

List of Tables

Acknowledgements

I have a mixed feeling. This dissertation gives me an opportunity to summarize what I have been doing in the past few years, but also marks the end of my PhD life at the IEOR Department of Columbia University, where I learned knowledge, made good friends, experienced happiness and sadness, gained confidence and encourage.

I wish to express my sincere thanks to Professor Jose Blanchet, my advisor, collaborator, and friend. I am indebted to him for invaluable guidance and encouragement extended to me. He is one of the smartest people I have ever met, and his passion for research is contagious. I always enjoy and benefit from discussing with him, whether it was about any questions regarding our research, or an interesting trip he had just taken. Besides his vast knowledge, what also impresses me is his deep love for his family, which tells me how to keep a good balance between work and life. In the past year, I have been keeping doing research while working full-time, and I really appreciate his sacrifice of rest time meeting with me on weekends. Without his guidance and persistent help this dissertation would not have been possible.

I am very grateful to my dissertation committee, Daniel Bienstock, Marvin K. Nakayama, Tim Leung, and Agostino Capponi, for taking time reading this dissertation and providing constructive feedbacks. I would like to dearly thank Daniel and Marvin who collaborated with me along with Jose in the work of Chapter 2 and 4, and they have shown excellent examples as successful researcher and professor. Special thanks go to David Yao and Donald Goldfarb, who encouraged me to pursue a doctoral degree when I was a master

student, and who are always willing to help me and give their best suggestions on my personal development for the past few years, from which I would benefit lifelong. I also want to express my gratitude to all the professors who taught me at Columbia University: Daniel Bienstock, Jose Blanchet, Donald Goldfarb, Soulaymane Kachani, Mariana Olvera-Cravioto, Karl Sigman, David Yao, Assaf Zeevi (CBS), Mark Broadie (CBS), Ciamac C. Moallemi (CBS). I thank all staff from IEOR Department who create a pleasant atmosphere, and are always so helpful and friendly.

My time at Columbia University was made enjoyable in large part due to the many friends I have made. I feel so lucky to have all of them with me: Xinyun Chen, Yujiao Chen, Jing Dong, Anran Li, Yina Lu, Bo Ren, Xingbo Xu...

Lastly, I would like to dedicate this dissertation to my parents, Kewei Li and Fengrong Guo, for their unconditional love, encouragement and support. Without them, I can hardly imagine getting through some tough times and getting this far. I love them so much.

Juan Li

May 18, 2015

Chapter 1

Introduction

1.1 Overview

This dissertation is devoted to algorithms for risk control on stochastic networks. Networks are very useful in modeling real-life problems. A network is identified by the characteristics of each vertex and the relationships across vertices, and it usually operates in some normal way. However, because of the uncertainty of both external and internal conditions, the characteristics of some individual vertex may change, therefore the status of the whole network may be affected as well, which sometimes may result in the failure of this network. Some examples are given in [35], [29] and [51]. It is thus desirable to develop efficient algorithms to study stochastic networks. This dissertation contributes in three themes:

- Modeling: Given a complex network, our model is able to capture how the un-

certainty impacts each vertex, and how the change of each vertex impacts other correlated vertices in this network.

- Simulation Design: The failure of the network is usually a rare event and the failure probability is difficult to estimate through crude Monte Carlo. Asymptotic behaviors are studied, and efficient simulation algorithms are developed and implemented which have proven to be both fast and accurate.

- Risk Control: Some guidance on risk management is provided which ensures that the failure probability is limited to a certain level.

Chapter 2 presents a distribution network in which each vertex is subjected to jointly Gaussian distributed demands. The equilibrium is determined by solving a liner program to minimize the total unserved demands across all vertices. This networks fails if the optimal objective function exceeds a large threshold, which is a rare event. Efficient Importance Sampling and Conditional Monte Carlo algorithms are developed to estimate the failure probability.

Chapter 3 discusses an insurance-reinsurance network in which stochastic claims arrive as the weighted sum of Pareto distributed common factors and individual factors. Reinsurance companies absorb some proportion of claims, but return spillover losses if they default. The equilibrium is obtained by solving a linear program. This network fails if all companies in a particular set default. When initial reserve and premium are large, the failure probability is very small. An efficient Importance Sampling algorithm is developed to estimate this probability and the spillover losses.

Chapter 4 studies an electrical power transmission network. The temperature evolution of each power line is determined by random ambient temperature through a stochastic differential equation. A line fails if the temperature is too high. Several control schemes are provided to limit the failure probability.

1.2 Preliminaries

1.2.1 Long-tailed, Heavy-tailed, and Light-tailed Distributions

Here we introduce some definitions of distributions that are used in this dissertation.

Definition 1.2.1. *The distribution of a random variable X with distribution function F is* long-tailed *if for all $c > 0$,*

$$\lim_{x\to\infty} P(X > x+c|X > x) = 1,$$

or equivalently

$$\lim_{x\to\infty} \frac{\bar{F}(x+c)}{\bar{F}(x)} = 1,$$

where $\bar{F}(\cdot)$ represents tail distribution function.

Definition 1.2.2. *The distribution of a random variable X with distribution function F is* heavy-tailed *if for all $\lambda > 0$,*

$$\lim_{x\to\infty} e^{\lambda x} P(X > x) = \infty.$$

In other words, for heavy-tailed distributions, tails are not exponentially bounded. All long-tailed distributions are heavy-tailed, while the converse is false.

Definition 1.2.3. *The distribution of a random variable X with distribution function F is* light-tailed *if there exists some $\lambda > 0$,*

$$\lim_{x\to\infty} e^{\lambda x} P(X > x) < \infty.$$

1.2.2 Notations for Asymptoticity

When we design efficient Monte Carlo algorithms, asymptotic behaviors of the probability of interest play an important role. Here we introduce some related definitions that are used throughout this dissertation.

Definition 1.2.4. *Let f and g be two functions defined on some subset of the real numbers. Then*

1. *$g(n) = O(f(n))$, if there exists a positive real number M and a real number n_0 such that $|g(n)| \leq M|f(n)|$ for all $n \geq n_0$.*

2. *$g(n) = \Omega(f(n))$, if there exists a positive real number M and a real number n_0 such that $|g(n)| \geq M|f(n)|$ for all $n \geq n_0$.*

3. *$g(n) = \Theta(f(n))$, if $g(n) = O(f(n))$, and $g(n) = \Omega(f(n))$.*

Definition 1.2.5. *Let $\{X_n\}$ and $\{a_n\}$ be a sequence of random variables and a sequence*

of deterministic positive constants, respectively. Suppose there exists a probability space $\{X_n'\}$ *such that* $X_n' \stackrel{d}{=} X_n$*, where* $\stackrel{d}{=}$ *represents equal in distribution. We denote by*

1. $X_n = O_p(a_n)$ *if there exists a random variable* M_1*, non-negative and finite almost surely, such that*

$$P\left(|X_n'| \leq a_n M_1\right) = 1.$$

2. $X_n = \Omega_p(a_n)$ *if there exists a random variable* M_2*, non-negative and finite almost surely, such that*

$$P\left(|X_n'| \geq a_n M_2\right) = 1.$$

3. $X_n = \Theta_p(a_n)$ *if* $X_n = O_p(a_n)$ *and* $X_n = \Omega_p(a_n)$.

1.2.3 Importance Sampling (IS)

Importance Sampling is one of the variance reduction techniques when doing Monte Carlo simulation, which is very powerful for rare event probability estimation.

Suppose we are interested in estimating: $\alpha = P(X \in A_b)$. Instead of obtaining samples from the original probability measure $P(\cdot)$, an alternative approach is to use another probability measure $P_g(\cdot)$, as long as the Radon-Nikodym derivative between these two measures is well defined on the event of interest. Then the IS estimator is

$$Y = 1_{\{\omega \in A_b\}} \frac{dP}{dP_g}(\omega),$$

where ω is the random outcome generated under P_g. Note that Y is unbiased since

$$E_g[Y] = \int_{A_b} \frac{dP}{dP_g}(\omega) dP_g(\omega) = \int_{A_b} dP(\omega) = P(X \in A_b).$$

An interesting probability measure is when $P_g(\cdot) = P(\cdot|X \in A_b) = P(\cdot, X \in A_b)/\alpha$, which depends on the unknown α. This measure is called zero-variance change-of-measure which is virtually impossible to implement because A_b is involved. However, studying the asymptotic behavior of this measure as $b \to \infty$ can provide useful guidance on developing an efficient IS algorithm.

The technique that we will use is inspired by this spirit and can be traced back to [28]. Generally, a natural way to develop an efficient IS algorithm is to bias the sampling distribution to be compatible with large deviation behaviors (see [16]).

1.2.4 Conditional Monte Carlo (CMC)

Another variance reduction technique is Conditional Monte Carlo. Suppose we are interested in estimating α, and U is an unbiased estimator. According to the conditional variance formula: $Var(U) = E[Var(U|Y)] + Var(E[U|Y])$, we have $Var(U) \geq Var(E[U|Y])$. Therefore, using $E[U|Y]$ as an estimator may help to reduce variance.

1.2.5 Notions of Efficiency

For a given Monte Carlo simulation algorithm, the following definitions provide a way to evaluate the efficiency.

Definition 1.2.6. *A simulation estimator Z_b for $\rho(b)$ is said to be* strongly efficient *if $E[Z_b] = \rho(b)$ and if*

$$\sup_{b>0} \frac{E(Z_b^2)}{\rho^2(b)} < \infty,$$

Definition 1.2.7. *A simulation estimator Z_b for $\rho(b)$ is said to be* weakly efficient *or* asymptotically optimal *if $E[Z_b] = \rho(b)$ and if*

$$\sup_{b>0} \frac{E(Z_b^2)}{\rho^{2-\varepsilon}(b)} < \infty, \ \ \forall \varepsilon > 0.$$

Asymptotic optimality also amounts to showing that

$$\frac{\log E(Z_b^2)}{2\log(\rho(b))} \to 1, \ \ b \to \infty.$$

These two definitions are motivated by the number of replications required to achieve a certain level of error. To see this, let us define the coefficient of variation of Z_b as

$$CV\,(Z_b) \stackrel{\Delta}{=} \left[\frac{Var\,(Z_b)}{\rho^2(b)}\right]^{1/2}.$$

Using Chebyshev's inequality we obtain

$$P\left\{\frac{|Z_b - \rho(b)|}{\rho(b)} > \varepsilon\right\} \leq \frac{CV(Z_b)^2}{\varepsilon^2}.$$

For crude Monte Carlo estimator Z_b^c, based on n independently and identically dis-

tributed (i.i.d) replications, the variance of Z_b^c is $\rho(b)(1-\rho(b))/n$, so

$$CV\left(Z_b^c\right) = \frac{(1-\rho(b))^{1/2}}{\rho^{1/2}(b)n^{1/2}}.$$

Therefore, in order to achieve the relative error with probability at least $1-\delta$, one needs $\Omega(1/\rho(b))$ replications through crude Monte Carlo simulation. On the other hand, efficient simulation algorithm only needs $O(\varepsilon^{-2}\delta^{-1})$ replications as $\rho(b) \to 0$.

Since we only focus on unbiased estimators, this requirement essentially reduces to controlled second moment.

1.2.6 Other Definitions

Definition 1.2.8. *A function $L(\cdot)$ is* slowly varying *if for all $x > 0$,*

$$\lim_{t\to\infty} \frac{L(xt)}{L(t)} = 1.$$

Definition 1.2.9. *A random variable X is said to be* regularly varying *with index $\alpha > 0$ (which we shall write $X \in \mathcal{RV}(\alpha)$) if for any $x > 0$,*

$$\lim_{t\to\infty} \frac{\overline{F}(xt)}{\overline{F}(t)} = x^{-\alpha}.$$

Equivalently, $X \in \mathcal{RV}(\alpha)$ if $\overline{F}(t) = t^{-\alpha}L(t)$ for some slowly varying function $L(\cdot)$.

Chapter 2

Distribution Networks

2.1 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. 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 trans-

mission 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 this network is recognized 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, see [35].

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 $\alpha^2(k)$ (see [9], for notions of efficiency in rare-event simulation).

As far as we know, this chapter 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 2.3.3).

2) We establish large deviations estimates for our class of linear programs with random input (see Theorem 2.4.2).

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 2.5.1).

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 2.5.2).

5) We provide several numerical examples in Section 2.6 that validate the performance of our algorithm.

We now explain how our research 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 [29]). 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 [1]. 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 proba-

bility 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 ([3]). The use of polar transformations for CMC and rare event simulation has been used in the past, see for example, [8]. [9], 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, [54] 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. [26] considers a branching process model of cascading failures in an electric power grid. [34] analyzes 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 chapter, the CMC algorithm can be applied to the case when the demands follow elliptical distribution (see [40]). Furthermore, while elliptical copula exhibits symmetric tail dependence, the well known Archimedean copula allows asymmetric tail dependence ([21]). Making use the results in [41], 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 this chapter develops as follows. Section 2.2 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 2.3. The asymptotic behavior of the model is discussed in Section 2.4. We describe the asymptotic optimality and implementations of importance sampling and conditional Monte Carlo methods for estimating $\alpha(k)$ in Section 2.5. Section 2.6 contains the experimental results from some examples, and we give some final comments in Section 2.7.

2.2 Model Description

We consider a network model which is induced by 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 the transpose of a matrix or vector.

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:

$$\begin{aligned} \min \quad & \sum_{i=1}^{d} x_i^+ \\ \text{s.t.} \quad & D_i - s_i + \sum_{j:(j,i)\in E} x_j^+ a_{ji} = x_i^+ - x_i^-, \forall i \\ & x_i^+ \geq 0, x_i^- \geq 0, \forall i. \end{aligned} \tag{2.2.1}$$

The quantity $x_i^+ \geq 0$ represents the demand shed from node i in equilibrium that 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 demands shed, is minimized. The problem can be expressed in matrix notation as follows. Define $A(i,j) = a_{ij}$ (note that $A(i,i) = 0$). Therefore, A is also inreducible. 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 (2.2.1) can be written as:

$$\begin{aligned} \min \quad & \mathbf{1}'\boldsymbol{x}^+ + \mathbf{0}'\boldsymbol{x}^- \\ \text{s.t.} \quad & (A' - I)\boldsymbol{x}^+ + I\boldsymbol{x}^- = \boldsymbol{s} - \boldsymbol{D} \\ & \boldsymbol{x}^+ \geq \mathbf{0}, \boldsymbol{x}^- \geq \mathbf{0}, \end{aligned} \tag{2.2.2}$$

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 demands shed 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:

$$\begin{aligned} \max \quad & \boldsymbol{y}'\boldsymbol{r} \\ \text{s.t.} \quad & M\boldsymbol{y} \leq \mathbf{1} \\ & \boldsymbol{y} \geq \mathbf{0}, \end{aligned} \tag{2.2.3}$$

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

We are interested in computing the probability that the network fails, for different values of k. Let $\alpha(k)$ represent this failure probability. We consider the network to have failed in two cases:

1. The primal is infeasible, i.e., the total demand exceeds the total supply.

2. The primal is feasible, but the cost of distributing demand is too high, i.e., $F(\boldsymbol{D})$ is larger than k.

Let $Q(\boldsymbol{D}) = 1$ if the primal is feasible, and $Q(\boldsymbol{D}) = 0$ otherwise. Let $\beta_0 = P\{Q(\boldsymbol{D}) = 0\}$, which is the probability of the first kind of failure, and $\beta_1(k) = P\{Q(\boldsymbol{D}) = 1, F(\boldsymbol{D}) \geq k\}$, which is the probability of the second kind of failure. If β_0 or $\beta_1(k)$ is large, the network is at a high risk of failure. Then we are interested in computing $\alpha(k) = \beta_0 + \beta_1(k)$, which is the probability that the networks fails.

Let $L(\boldsymbol{D})$ denote the optimal value of the dual when the demand vector is $\boldsymbol{D}$. Note that

$$\begin{aligned}
\alpha(k) &= \beta_0 + \beta_1(k) \\
&= P\{Q(\boldsymbol{D}) = 0\} + P\{Q(\boldsymbol{D}) = 1, F(\boldsymbol{D}) \geq k\} \\
&= P\{Q(\boldsymbol{D}) = 0\} + P\{Q(\boldsymbol{D}) = 1\} P\{F(\boldsymbol{D}) \geq k | Q(\boldsymbol{D}) = 1\} \\
&= P\{Q(\boldsymbol{D}) = 0\} P\{L(\boldsymbol{D}) \geq k | Q(\boldsymbol{D}) = 0\} + P\{Q(\boldsymbol{D}) = 1\} P\{L(\boldsymbol{D}) \geq k | Q(\boldsymbol{D}) = 1\} \\
&= P\{L(\boldsymbol{D}) \geq k\}, \qquad (2.2.4)
\end{aligned}$$

where the fourth equality follows from strong duality, and the fact that when $Q(\boldsymbol{D}) = 0$, i.e., the primal is infeasible, we have $L(\boldsymbol{D}) = +\infty$. Therefore, we want to compute the probability that the optimal value of the dual $L(\boldsymbol{D})$ is larger than k.

Notice that while β_0 is independent of k, $\beta_1(k)$ is decreasing in k. Thus, if β_0 is large compared to $\beta_1(k)$, then $\alpha(k) \approx \beta_0$ as k increases, which is a constant. On the other hand, if β_0 is small compared to $\beta_1(k)$ even for large k, then $\alpha(k) \approx \beta_1(k)$ is decreasing in k. In our experiments (Section 2.6), we consider examples in which $\beta_0 \ll \beta_1(k)$ for all k, so the latter situation occurs.

2.3 Properties of Our Primal and Dual Linear Programs

2.3.1 Feasibility

Theorem 2.3.1. *The primal and the dual problems have the following properties:*

(a) *The dual problem (2.2.3) is always feasible.*

(b) *The primal problem (2.2.2) is feasible if and only if* $\sum_{i=1}^{d} D_i \leq \sum_{i=1}^{d} s_i$.

Proof. For part (a), take $\boldsymbol{y} = \mathbf{0}$, which is clearly feasible.

For part (b), note that if $\sum_{i=1}^{d} D_i > \sum_{i=1}^{d} s_i$, i.e., $\mathbf{1}'\boldsymbol{r} > 0$, then the dual is unbounded and therefore the primal is infeasible. To see this, note that $-M$ can be interpreted as the rate matrix of a continuous-time Markov chain, so $M\mathbf{1} = \mathbf{0}$, and $\boldsymbol{y} = m\mathbf{1}$ is a feasible solution to the dual, for all $m > 0$. Therefore, if $\mathbf{1}'\boldsymbol{r} > 0$, we clearly have that the dual is unbounded.

If $\sum_{i=1}^{d} D_i \leq \sum_{i=1}^{d} s_i$, i.e., $\mathbf{1}'\boldsymbol{r} \leq 0$, then we claim the dual is bounded and therefore the primal has an optimal feasible solution. To see this, we argue by contradiction. Suppose that the dual is unbounded. Then there exists a vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_d)' \geq \mathbf{0}$ such that $\boldsymbol{v}'\boldsymbol{r} > 0$, and $M\boldsymbol{v} \leq \mathbf{0}$. Let $v_{i_0} = \max_{j=1,2,\ldots,d} v_j$. Without loss of generality, we assume that $i_0 = 1$. Since A is irreducible, there exists some $h_2 \geq 1$ such that $a_{12}^{(h_2)} > 0$, where $a_{ij}^{(h_2)}$ denotes the (i, j) element of A^{h_2}. Because A is the transition matrix of a discrete-time Markov chain, $A\mathbf{1} = \mathbf{1}$ and $A^{h_2}\mathbf{1} = \mathbf{1}$. Thus,

$$
\begin{aligned}
(I - A^{h_2})\boldsymbol{v} &= \begin{pmatrix} 1 - a_{11}^{(h_2)} & -a_{12}^{(h_2)} & \ldots & -a_{1d}^{(h_2)} \\ -a_{21}^{(h_2)} & 1 - a_{22}^{(h_2)} & \ldots & -a_{2d}^{(h_2)} \\ \vdots & \vdots & & \vdots \\ -a_{d1}^{(h_2)} & -a_{d2}^{(h_2)} & \ldots & 1 - a_{dd}^{(h_2)} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} \\
&= \begin{pmatrix} a_{12}^{(h_2)}(v_1 - v_2) + a_{13}^{(h_2)}(v_1 - v_3) + \cdots + a_{1d}^{(h_2)}(v_1 - v_d) \\ a_{21}^{(h_2)}(v_2 - v_1) + a_{23}^{(h_2)}(v_2 - v_3) + \cdots + a_{2d}^{(h_2)}(v_2 - v_d) \\ \vdots \\ a_{d1}^{(h_2)}(v_d - v_1) + a_{d2}^{(h_2)}(v_d - v_2) + \cdots + a_{d,d-1}^{(h_2)}(v_d - v_{d-1}) \end{pmatrix},
\end{aligned}
$$

where the second equality follows from $A^{h_2}\mathbf{1} = \mathbf{1}$. Also we have

$$
(I - A^{h_2})\boldsymbol{v} = (I + A + A^2 + \cdots + A^{h_2 - 1})(I - A)\boldsymbol{v} = (I + A + A^2 + \cdots + A^{h_2 - 1})M\boldsymbol{v} \leq \mathbf{0},
$$

where the inequality follows from $M\boldsymbol{v} \leq \mathbf{0}$, and the fact that every entry of the matrix

$I + A + A^2 + \cdots + A^{h_2 - 1}$ is non-negative. Therefore,

$$(I - A^{h_2})\boldsymbol{v} = \begin{pmatrix} a_{12}^{(h_2)}(v_1 - v_2) + a_{13}^{(h_2)}(v_1 - v_3) + \cdots + a_{1d}^{(h_2)}(v_1 - v_d) \\ a_{21}^{(h_2)}(v_2 - v_1) + a_{23}^{(h_2)}(v_2 - v_3) + \cdots + a_{2d}^{(h_2)}(v_2 - v_d) \\ \vdots \\ a_{d1}^{(h_2)}(v_d - v_1) + a_{d2}^{(h_2)}(v_d - v_2) + \cdots + a_{d,d-1}^{(h_2)}(v_d - v_{d-1}) \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Consider the first element of $(I - A^{h_2})\boldsymbol{v}$. Because $a_{1j}^{(h_2)} \geq 0$ and $v_1 - v_j \geq 0$ for all $j \neq 1$, we have $\sum_{j \neq 1} a_{1j}^{(h_2)}(v_1 - v_j) = 0$. Therefore, $a_{1j}^{(h_2)}(v_1 - v_j) = 0$ for all $j \neq 1$. Because $a_{12}^{(h_2)} > 0$, we must have $v_1 = v_2$. Similarly, by irreducibility, there exists some $h_3 \geq 1$ such that $a_{13}^{(h_3)} > 0$. Again we conclude that $v_1 = v_3$. Therefore, by iteration, we have that $v_1 = v_2 = \cdots = v_d > 0$. Then $\mathbf{1}'\boldsymbol{r} = \boldsymbol{v}'\boldsymbol{r}/v_1 > 0$, which contradicts our assumption that $\mathbf{1}'\boldsymbol{r} \leq 0$. □

2.3.2 Uniqueness and Positivity

Theorem 2.3.2. *The primal problem (2.2.2) 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,*

$\forall 1 \leq k \leq d$.

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:

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

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

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

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:

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

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 (2.2.2). 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 (2.2.2). 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, $\forall 1 \leq k \leq d$. □

2.3.3 Insensitivity of the Solution to the Primal

Theorem 2.3.3. *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*

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

Proof. Consider the problem

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

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.

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

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, $x_i^{*+} > 0, \forall 1 \leq i \leq d$. Note that the conclusion of Theorem 2.3.2(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 (2.2.2) 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 $f = f_2$. □

Note that although Theorem 2.3.3 established the insensitivity of the optimal solution to a large class of nonlinear objective functions, for the rest of the chapter, our discussion is based on the primal problem (2.2.2) and the dual problem (2.2.3) with linear objective functions.

2.4 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 2.4.1. *If a random variable $X \sim N(\bar{\mu}, \bar{\sigma}^2)$, where $\bar{\sigma} > 0$, then for all $\alpha > \bar{\mu}$,*

$$P\{X > \alpha\} \geq \frac{1}{\sqrt{2\pi}} \frac{\bar{\sigma}}{\alpha - \bar{\mu}} \exp\{-\frac{(\alpha - \bar{\mu})^2}{2\bar{\sigma}^2}\}. \tag{2.4.1}$$

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 $\alpha > \bar{\mu}$,

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

□

Theorem 2.4.2. *Let $L_n(\boldsymbol{D})$ denote the optimal value of the dual (2.2.3), when the demand*

vector is $\boldsymbol{D}$ *and the rarity parameter is n. Then for all* $k \geq 0$*,*

$$\lim_{n\to\infty} n^{-2\beta} \log P\{L_n(\boldsymbol{D}) > k\} = \lim_{n\to\infty} n^{-2\beta} \log P\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > k\} \tag{2.4.2}$$

$$= \lim_{n\to\infty} n^{-2\beta} \log P\{\sup_{t\in T} D(t) - s(t) > k\} \tag{2.4.3}$$

$$= -\frac{\gamma^2(t^*)}{2\sigma^2(t^*)}, \tag{2.4.4}$$

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

$$\{L_n(\boldsymbol{D}) > k\} \subseteq \{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\}. \tag{2.4.5}$$

To see this, if we assume $\max_{i=1,\ldots,d} D(t_i) - s(t_i) \leq 0$, then $D(t_i) \leq s(t_i), \forall i = 1,2,\ldots,d$. According to Theorem 2.3.1(b), the primal problem (2.2.2) is feasible, and it is easy to see that $x_i^+ = 0, x_i^- = s(t_i) - D(t_i) \geq 0, \forall i = 1,2,\ldots,d$, is an optimal solution to the primal problem. In this case $L_n(\boldsymbol{D}) = 0$. Thus $\{\max_{i=1,\ldots,d} D(t_i) - s(t_i) > 0\}^c \subseteq \{L_n(\boldsymbol{D}) > k\}^c$, where

"c" represents the complement of a set. Therefore,

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

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

$$\begin{aligned} P\{L_n(\boldsymbol{D}) > k\} &\leq P\{\sup_{t\in T} h(t) > \frac{n^{\beta}\gamma(t^*)}{\sigma(t^*)} - \frac{\mu(\hat{t})}{\sigma(\hat{t})}\} \\ &= P\{\|h\| > \frac{n^{\beta}\gamma(t^*)}{\sigma(t^*)} - \frac{\mu(\hat{t})}{\sigma(\hat{t})}\} \\ &\leq \exp\{-\frac{1}{2}(\frac{n^{\beta}\gamma(t^*)}{\sigma(t^*)} - \frac{\mu(\hat{t})}{\sigma(\hat{t})} - \lambda)^2\}, \end{aligned} \tag{2.4.6}$$

where the last inequality is obtained by Borell-TIS inequality ([3], 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* ([3], 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

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

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 $k \geq 0$ is some constant. Then $g(t)$ is continuous in T. Therefore, for a given $0 < \delta < 1$, there exists a ball $B(t^*,\varepsilon)$ centered at t^* such that $|g(t) - g(t^*)| \leq \delta g(t^*)$, for all $t \in B(t^*,\varepsilon)$. We now claim that

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

To see this, note that if $\max_{i=1,\ldots,d} D(t_i) - s(t_i) > k$, then there exists some $1 \leq i_0 \leq d$ 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 (2.2.3) and $\boldsymbol{y}'(\boldsymbol{D} - \boldsymbol{s}) = D(t_{i_0}) - s(t_{i_0}) > k$. Therefore, $L_n(\boldsymbol{D}) > k$. Then,

$$\begin{aligned}
P\{L_n(\boldsymbol{D}) > k\} &\geq P\{\max_{i=1,\ldots,n} D(t_i) - s(t_i) > k\} \qquad (2.4.7)\\
&\geq P\{D(t_1) - s(t_1) > k\}\\
&= \int_T P\{D(t) - s(t) > k\} g_T(t)\, dt\\
&\geq \int_{B(t^*,\varepsilon)} P\{D(t) - s(t) > k\} g_T(t)\, dt\\
&\geq \int_{B(t^*,\varepsilon)} \frac{1}{\sqrt{2\pi}} \frac{\sigma(t)}{s(t)-\mu(t)+k} \exp\{-\frac{(s(t)-\mu(t)+k)^2}{2\sigma^2(t)}\} g_T(t)\, dt\\
&\geq g(t^*)(1-\delta)C, \qquad (2.4.8)
\end{aligned}$$

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

Therefore, (2.4.6) and (2.4.8) imply for n sufficiently large,

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

Taking logarithms, we have

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

Because

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

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 (2.4.4).

We now establish (2.4.2) and (2.4.3). By (2.4.7) and (2.4.8), it follows that

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

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

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

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

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

□

2.5 Efficient Algorithms: Importance Sampling and Conditional Monte Carlo

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).

2.5.1 Importance Sampling

Algorithm Description

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

$$Q\{\boldsymbol{D} \in B\} = \sum_{i=1}^{d} p(i) P\{\boldsymbol{D} \in B | D(t_i) - s(t_i) > 0\}, \tag{2.5.1}$$

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

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

Asymptotic Optimality

Theorem 2.5.1.

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

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 (2.4.5), we have

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

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, $\sum_{j=1}^d I\{D(t_j) - s(t_j) > 0\} \geq 1$,

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

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,

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

where the last equation follows from Theorem 2.4.2. □

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 (2.5.1). To do this, we use the algorithm described in [46] 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.

2.5.2 Conditional Monte Carlo

Algorithm Description

Note that the multivariate-normal random demand has polar-coordinate representation (see [40]):

$$\boldsymbol{D} = \boldsymbol{\mu} + RW\boldsymbol{\Psi}, \tag{2.5.2}$$

where the radius R satisfies $R^2 \sim \Gamma(d/2, 1/2)$, i.e., its density function

$$g(x) = x^{d/2-1}e^{-x/2}(1/2)^{d/2}/\Gamma(d/2),$$

where $\Gamma(\cdot)$ is the gamma function, $WW' = \Sigma$, 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.

In the set $\{(R, \boldsymbol{\Psi}) : \text{primal with } (R, \boldsymbol{\Psi}) \text{ is feasible}\}$, let $F(\boldsymbol{D})$ denote the optimal value of the primal when the demand vector is $\boldsymbol{D} = \boldsymbol{\mu} + RW\boldsymbol{\Psi}$, and let $F_{\boldsymbol{\Psi}}(R)$ denote $F(\boldsymbol{D})$ when $\boldsymbol{\Psi}$ is fixed. Then we have the following theorem regarding the shape of the optimal value function.

Theorem 2.5.2. *For fixed* $\boldsymbol{\Psi}$*,*

(a) $F_{\boldsymbol{\Psi}}(R)$ *is a piecewise linear function of* R *in the set*

$$J_{\boldsymbol{\Psi}} = \{R \geq 0 : \textit{primal with } (R, \boldsymbol{\Psi}) \textit{ is feasible}\}; \tag{2.5.3}$$

(b) *$F_{\boldsymbol{\Psi}}(R)$ is a convex function of R in the set $J_{\boldsymbol{\Psi}}$.*

Proof. We can write the primal as

$$\begin{aligned} \min \quad & \boldsymbol{g}'\boldsymbol{x} \\ \text{s.t.} \quad & \Theta\boldsymbol{x} = \boldsymbol{s} - \boldsymbol{\mu} - RW\boldsymbol{\Psi} \\ & \boldsymbol{x} \geq \mathbf{0}, \end{aligned}$$

where $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix}$, $\boldsymbol{g} = (1,1,\ldots,1,0,0,\ldots,0)'$, and $\Theta = (A' - I, I)$. Obviously, the rows of Θ are linearly independent.

The dual can be written as

$$\begin{aligned} \max \quad & \boldsymbol{y}'(\boldsymbol{\mu} + RW\boldsymbol{\Psi} - \boldsymbol{s}) \\ \text{s.t.} \quad & \boldsymbol{y}'(-\Theta) \leq \boldsymbol{g}'. \end{aligned}$$

Because the dual is always feasible, for any $R \in J_{\boldsymbol{\Psi}}$, $F_{\boldsymbol{\Psi}}(R)$ is finite and by strong duality (e.g., Theorem 4.4 of [13]), is equal to the optimal value of the dual objective.

Let us fix a particular element $R^* \in J_{\boldsymbol{\Psi}}$. Then there exists a primal optimal basic feasible solution. Let B be the corresponding optimal basis matrix. The vector $\boldsymbol{x}_B$ of basic variables at that optimal solution is given by $\boldsymbol{x}_B = B^{-1}(\boldsymbol{s} - \boldsymbol{\mu} - R^*W\boldsymbol{\Psi})$, which is strictly positive with probability 1 because R^{*2} follows the continuous distribution $\Gamma(d/2, 1/2)$. In addition, the vector of reduced costs is nonnegative. If we change R^* to R and if the

difference $R - R^*$ is sufficiently small, $B^{-1}(\boldsymbol{s} - \boldsymbol{\mu} - RW\boldsymbol{\Psi})$ remains positive and we still have a basic feasible solution. The reduced costs are not affected by the change from R^* to R and remain nonnegative. Therefore, B is an optimal basis for the new problem with R as well. Then $F_{\boldsymbol{\Psi}}(R)$ for the new problem is given by

$$F_{\boldsymbol{\Psi}}(R) = \boldsymbol{g}_B' B^{-1}(\boldsymbol{s} - \boldsymbol{\mu} - RW\boldsymbol{\Psi}) \text{ for } R \text{ close to } R^*,$$

where $\boldsymbol{g}_B$ is the cost vector corresponding to the basis. This establishes that in the vicinity of R^*, $F_{\boldsymbol{\Psi}}(R)$ is a linear function of R.

To show part (b), let $R^{(1)}$ and $R^{(2)}$ be two elements of $J_{\boldsymbol{\Psi}}$. For $i = 1, 2$, let $\boldsymbol{x}^{(i)}$ be an optimal solution to the primal corresponding to $R^{(i)}$. Thus, $F_{\boldsymbol{\Psi}}(R^{(1)}) = \boldsymbol{g}'\boldsymbol{x}^{(1)}$ and $F_{\boldsymbol{\Psi}}(R^{(2)}) = \boldsymbol{g}'\boldsymbol{x}^{(2)}$. Fix a scalar $\gamma \in [0,1]$, and note that the vector $\boldsymbol{x} = \gamma\boldsymbol{x}^{(1)} + (1-\gamma)\boldsymbol{x}^{(2)}$ is a feasible solution to the primal, so $\gamma R^{(1)} + (1-\gamma)R^{(2)} \in J_{\boldsymbol{\Psi}}$. Therefore,

$$F_{\boldsymbol{\Psi}}(\gamma R^{(1)} + (1-\gamma)R^{(2)}) \leq g'x = \gamma\boldsymbol{g}'\boldsymbol{x}^{(1)} + (1-\gamma)\boldsymbol{g}'\boldsymbol{x}^{(2)} = \gamma F_{\boldsymbol{\Psi}}(R^{(1)}) + (1-\gamma)F_{\boldsymbol{\Psi}}(R^{(2)}),$$

establishing the convexity of $F_{\boldsymbol{\Psi}}(R)$. □

Making use of the polar-coordinate representation and the above the theorem, we now develop a conditional Monte Carlo approach for estimating $\alpha(k)$.

Note that $\alpha(k)$ can be written as $\alpha(k) = E[E[U|\boldsymbol{\Psi}]]$, so if we can compute $E[U|\boldsymbol{\Psi}] = P\{L(\boldsymbol{D}) \geq k|\boldsymbol{\Psi}\}$, then sampling $E[U|\boldsymbol{\Psi}]$ rather than U as in naive simulation reduces

variance. Below is an algorithm for computing an estimator of $\alpha(k)$ with conditional Monte Carlo.

Asymptotic Optimality

Recall that we defined in Section 2.4 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 2.5.3. *There exist constants $n_0 > 0$, $c_3 > 0$, $s^* > 0$, $\eta_1 \in \mathbb{R}$, such that when $n > n_0$,*

$$T_n(\boldsymbol{\Psi}) \triangleq P\{L_n(\boldsymbol{D}) > k|\boldsymbol{\Psi}\} \leq P\{R > n^{\beta}s^* + \eta_1\}, \ \ \forall \|\boldsymbol{\Psi}\| = 1, \tag{2.5.4}$$

$$P\{L_n(\boldsymbol{D}) > k\} \geq c_3 P\{R > n^{\beta}s^* + O(1)\}n^{-(d-1)\beta}. \tag{2.5.5}$$

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

Proof. We first prove (2.5.4). Let $\Omega = \{\boldsymbol{y} : M\boldsymbol{y} \leq \mathbf{1}, \boldsymbol{y} \geq \mathbf{0}\}$ denote the feasible region of the dual problem (2.2.3). 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 2.4. We are interested in the failure probability, which includes two cases as we noted previously in Section 2.2. One case is that the primal problem is infeasible, which, according to Theorem 2.3.1(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 $\{\tilde{\mathbf{y}}_i : i = 1,2,\ldots,m\}$ are the extreme points of Ω, excluding $\mathbf{0}$, and we have

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

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

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

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

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

Define

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

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

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

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

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

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

$$P\{L_n(\boldsymbol{D}) > k\} = P\{R > H(\boldsymbol{\Psi}, n)\}. \tag{2.5.6}$$

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 (2.5.4) 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

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

$$\inf_{\boldsymbol{\Psi}\in\Gamma_a} \min_{i\in M_{\boldsymbol{\Psi}}} \frac{-\langle \tilde{\boldsymbol{y}}_i, \boldsymbol{\mu}\rangle + k_i}{\langle \tilde{\boldsymbol{y}}_i, W\boldsymbol{\Psi}\rangle} = \eta_1. \tag{2.5.7}$$

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 (2.5.4) is established.

Now we consider the inequality (2.5.5). We claim that for any $\boldsymbol{\Psi}$ in Γ_a, there exists $n_2(\boldsymbol{\Psi}) > 0$ such that when $n > n_2(\boldsymbol{\Psi})$,

$$H(\boldsymbol{\Psi}, n) = n^{\beta} S(\boldsymbol{\Psi}) + \frac{k_{\boldsymbol{\Psi}} - \langle \tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}}, \boldsymbol{\mu}\rangle}{\langle \tilde{\boldsymbol{y}}_{\boldsymbol{\Psi}}, W\boldsymbol{\Psi}\rangle}, \tag{2.5.8}$$

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 $i \in M_{\boldsymbol{\Psi}}$,

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

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

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

Choose

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

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

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

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

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

$$H(\mathbf{\Psi}^*, n) = n^{\beta} S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\mathbf{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\mathbf{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle}. \tag{2.5.9}$$

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 $H(\mathbf{\Psi}, n) - H(\mathbf{\Psi}^*, n) \leq (n^{\beta} c_1 + c_2)\|\mathbf{\Psi} - \mathbf{\Psi}^*\|$ on Γ_a. To see this, for any $\delta > 0$ and $\boldsymbol{\theta} \in \Gamma_a$, define $B(\boldsymbol{\theta}, \delta) = \{\mathbf{\Psi} \in \Gamma_a : \|\mathbf{\Psi} - \boldsymbol{\theta}\| \leq \delta\}$. Note that there exists $\delta_1 > 0$, such that when $0 < \delta \leq \delta_1$, 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 $\tilde{\mathbf{y}}_{\mathbf{\Psi}^*}$ is in $M_{\mathbf{\Psi}}$, and

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

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

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

where c_0 is some positive constant.

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

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

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

$$H(\mathbf{\Psi}, n) \leq H(\mathbf{\Psi}^*, n) + (n^{\beta} c_1 + c_2)\delta. \tag{2.5.10}$$

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

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

Let $\delta = n^{-\beta}$. By equations (2.5.6) and (2.5.10), it follows that

$$\begin{aligned}
P\{L_n(\boldsymbol{D}) > k\} &= P\{R > H(\mathbf{\Psi}, n)\} \\
&\geq P\{R > H(\mathbf{\Psi}^*, n) + (n^{\beta} c_1 + c_2)\delta, \| \mathbf{\Psi} - \mathbf{\Psi}^* \| \leq \delta\} \\
&\geq c_3 P\{R > H(\mathbf{\Psi}^*, n) + (n^{\beta} c_1 + c_2)\delta\}\delta^{(d-1)} \\
&= c_3 P\{R > n^{\beta} S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\mathbf{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\mathbf{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta})\} n^{-(d-1)\beta} \\
&= c_3 P\{R > n^{\beta} S(\mathbf{\Psi}^*) + O(1)\} n^{-(d-1)\beta}.
\end{aligned} \tag{2.5.11}$$

Hence, we have proven (2.5.5).

We now establish the last part of the theorem. By (2.5.4) and (2.5.11), we have

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

$$= \frac{2\log P\{R > n^{\beta} S(\mathbf{\Psi}^*) + \eta_1\}}{\log c_3 + \log P\{R > n^{\beta} S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\mathbf{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\mathbf{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta})\} - (d-1)\beta \log n}$$

$$= 2\Big(\frac{\log P\{R > n^{\beta} S(\mathbf{\Psi}^*) + \frac{k_{\mathbf{\Psi}^*} - \langle \tilde{\mathbf{y}}_{\mathbf{\Psi}^*}, \boldsymbol{\mu} \rangle}{\langle \tilde{\mathbf{y}}_{\mathbf{\Psi}^*}, W\mathbf{\Psi}^* \rangle} + (c_1 + c_2 n^{-\beta})\}}{\log P\{R > n^{\beta} S(\mathbf{\Psi}^*) + \eta_1\}} + \frac{\log c_3 - (d-1)\beta \log n}{\log P\{R > n^{\beta} S(\mathbf{\Psi}^*) + \eta_1\}} \Big)^{-1}. \tag{2.5.12}$$

Recall that $n^{\beta} c_1 + c_2 > 0$ when $n > \max\{n_1, n_2(\mathbf{\Psi}^*), n_3\}$, so (2.5.7) implies

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

Therefore,

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

and

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

When $n > n_4 = e^{\log b / \beta(d-1)}$, the second term inside the parentheses in (2.5.12) is non-negative. Then when $n > n_0 = \max\{n_1, n_2(\mathbf{\Psi}^*), n_3, n_4\}$, it follows that (2.5.12) is bounded above by 2, thereby concluding the result. □

Algorithm Implementation

We now explain how to implement the Conditional Monte Carlo algorithm.

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

2. Generate $\mathbf{\Psi}_i = z_i/\|z_i\|$, where $z_i \sim N(0,I)$.

3. Search for the root R_i^* of the equation $F_{\mathbf{\Psi}_i}(R) = k$ if it exists, and calculate $E[U_i|\mathbf{\Psi}_i]$.

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

5. Compute $\widetilde{\alpha}_n(k) = (\sum_{i=1}^n E[U_i|\mathbf{\Psi}_i])/n$ as our conditional Monte Carlo estimator of $\alpha(k)$, and a $100(1-\delta)\%$ confidence interval for $\alpha(k)$ is $(\widetilde{\alpha}_n(k) \pm \Phi^{-1}(1-\delta/2)\widetilde{S}_n/\sqrt{n}))$, where $\widetilde{S}_n^2 = (\sum_{i=1}^n (E[U_i|\mathbf{\Psi}_i] - \widetilde{\alpha}_n(k))^2)/(n-1)$.

Now we give more details about how we calculate $E[U_i|\mathbf{\Psi}_i]$ and search for the root R_i^* of the equation $F_{\mathbf{\Psi}_i}(R) = k$. Note that

$$\begin{aligned} E[U|\mathbf{\Psi}] &= P\{L(D) \geq k|\mathbf{\Psi}\} \\ &= P\{F(D) \geq k, Q(D) = 1|\mathbf{\Psi}\} + P\{Q(D) = 0|\mathbf{\Psi}\} \\ &= P\{F_{\mathbf{\Psi}}(R) \geq k, Q(D) = 1|\mathbf{\Psi}\} + P\{Q(D) = 0|\mathbf{\Psi}\} \\ &= P\{R \in K_{\mathbf{\Psi}}|\mathbf{\Psi}\} + P\{R \in J_{\mathbf{\Psi}}^c|\mathbf{\Psi}\} \\ &= P\{R \in K_{\mathbf{\Psi}} \cup J_{\mathbf{\Psi}}^c|\mathbf{\Psi}\}, \end{aligned} \tag{2.5.13}$$

where $K_{\mathbf{\Psi}} = \{R \in J_{\mathbf{\Psi}} : F_{\mathbf{\Psi}}(R) \geq k\}$, $J_{\mathbf{\Psi}}$ is defined in (2.5.3), and $J^c_{\mathbf{\Psi}} = \{R \geq 0 : \text{primal with } (R, \mathbf{\Psi}) \text{ is infeasible}\}$ is the complement of $J_{\mathbf{\Psi}}$. The last equality follows from the fact that $K_{\mathbf{\Psi}} \cap J^c_{\mathbf{\Psi}} = \emptyset$.

By Theorem 2.5.2, for fixed $\mathbf{\Psi}$, $F_{\mathbf{\Psi}}(R)$ is piecewise linear and convex on $J_{\mathbf{\Psi}}$. Therefore, $K_{\mathbf{\Psi}}$ is an interval or union of intervals if not empty. By Theorem 2.3.1, for fixed $\mathbf{\Psi}$, $J_{\mathbf{\Psi}} = \{R \geq 0 : R\mathbf{1}'W\mathbf{\Psi} \leq \mathbf{1}'s - \mathbf{1}'\mu\}$. By Assumption 4 in Section 2.2, we have $\mathbf{1}'s > \mathbf{1}'\mu$. If $\mathbf{1}'W\mathbf{\Psi} > 0$, then $J_\Psi = [0, (\mathbf{1}'s - \mathbf{1}'\mu)/\mathbf{1}'W\mathbf{\Psi}]$; $J_{\mathbf{\Psi}} = [0, +\infty)$ otherwise. In both cases, $J_{\mathbf{\Psi}}$ is an interval. Thus, $K_{\mathbf{\Psi}} \cup J^c_\Psi$ is an interval or union of intervals if not empty. Because $R^2 \sim \Gamma(d/2, 1/2)$, $E[U|\mathbf{\Psi}]$ can be calculated analytically using (2.5.13).

As noted above, $J^c_{\mathbf{\Psi}}$ is easy to find, and calculating $K_{\mathbf{\Psi}}$ only requires finding $R^* \in J_{\mathbf{\Psi}}$ such that

$$F_{\mathbf{\Psi}}(R^*) = k. \tag{2.5.14}$$

There are three cases for the root, as we now explain.

- There is no root in the set $J_{\mathbf{\Psi}}$.

  If $F_{\mathbf{\Psi}}(R) < k$ for all $R \in J_{\mathbf{\Psi}}$, then $K_{\mathbf{\Psi}} = \emptyset$ and $E[U|\mathbf{\Psi}] = P\{R \in J^c_{\mathbf{\Psi}}\}$. If $F_{\mathbf{\Psi}}(R) > k$ for all $R \in J_{\mathbf{\Psi}}$, then $K_{\mathbf{\Psi}} = J_{\mathbf{\Psi}}$ and $E[U|\mathbf{\Psi}] = 1$.

- There is only one root R^* in the set $J_{\mathbf{\Psi}}$.

  If the slope of $F_{\mathbf{\Psi}}(R)$ at R^* is positive, then $K_{\mathbf{\Psi}} = [R^*, (\mathbf{1}'s - \mathbf{1}'\mu)/\mathbf{1}'W\mathbf{\Psi}]$ when $\mathbf{1}'W\mathbf{\Psi} > 0$, and $K_{\mathbf{\Psi}} = [R^*, +\infty)$ when $\mathbf{1}'W\mathbf{\Psi} \leq 0$; in both cases $E[U|\mathbf{\Psi}] = P\{R \in [R^*, +\infty)\}$. If the slope of $F_{\mathbf{\Psi}}(R)$ at R^* is negative, then $K_{\mathbf{\Psi}} = [0, R^*]$ and $E[U|\mathbf{\Psi}] = P\{R \in [0, R^*] \cup J^c_{\mathbf{\Psi}}\}$.

- There are two different roots $R^{*(1)} < R^{*(2)}$ in the set $J_{\mathbf{\Psi}}$.

  Because $F_{\mathbf{\Psi}}(R)$ is convex on $J_{\mathbf{\Psi}}$, it must be that the slope at $R^{*(1)}$ is negative and the slope at $R^{*(2)}$ is positive. Then $K_{\mathbf{\Psi}} = [0, R^{*(1)}] \cup [R^{*(2)}, (\mathbf{1}'s - \mathbf{1}'\mu)/\mathbf{1}'W\mathbf{\Psi}]$ when $\mathbf{1}'W\mathbf{\Psi} > 0$, and $K_{\mathbf{\Psi}} = [0, R^{*(1)}] \cup [R^{*(2)}, +\infty)$ when $\mathbf{1}'W\mathbf{\Psi} \leq 0$; in both cases $E[U|\mathbf{\Psi}] = P\{R \in [0, R^{*(1)}] \cup [R^{*(2)}, +\infty)\}$.

Now we explain how we search for the root R^* of (2.5.14) for a fixed $\mathbf{\Psi}$ using Newton's method. Recall that if $\mathbf{1}'W\mathbf{\Psi} > 0$, $J_{\mathbf{\Psi}}$ is a finite interval; $J_{\mathbf{\Psi}}$ is an infinite interval otherwise. We hope to find a uniform way to search in both cases to simplify coding. Meanwhile, we want to control the computational error. If $\mathbf{1}'W\mathbf{\Psi} > 0$, set $b = G_\Gamma((\mathbf{1}'s - \mathbf{1}'\mu)/\mathbf{1}'W\mathbf{\Psi})^2$, where G_Γ is the distribution function of $\Gamma(d/2, 1/2)$; set $b = 1$ otherwise. Then we let $b_l = \sqrt{G_\Gamma^{-1}(0.5c)}$ and $b_u = \sqrt{G_\Gamma^{-1}(b - 0.5c)}$, where c is a small constant such that $P\{R \in [b_l, b_u]\} > 0$. In our algorithm, b_l and b_u are lower and upper bounds for the root R^*, and we search for $R^* \in [b_l, b_u]$, which is a finite interval in both cases. Let $\widetilde{K}_{\mathbf{\Psi}} = \{R \in [b_l, b_u] : F_{\mathbf{\Psi}}(R) \geq k\}$. Then our algorithm calculates $\widetilde{K}_{\mathbf{\Psi}}$ instead of $K_{\mathbf{\Psi}}$. Because $P\{R \in J_{\mathbf{\Psi}}\} - P\{R \in [b_l, b_u]\} = c$, we have $0 \leq \{R \in K_{\mathbf{\Psi}}\} - P\{R \in \widetilde{K}_{\mathbf{\Psi}}\} \leq c$, and the error between $E[U|\mathbf{\Psi}]$ and our computed value of it will not exceed c. We first start from an initial value $R_0 = b_l$, and by strong duality, we solve the dual LP with demand $\boldsymbol{D} = \boldsymbol{\mu} + R_0 W\mathbf{\Psi}$ using the simplex algorithm to obtain the corresponding $F_{\mathbf{\Psi}}(R_0)$ and slope l_0 of $F_{\mathbf{\Psi}}(R)$ at R_0.

- If $l_0 > 0$, there is at most one root.

  - If $F_{\mathbf{\Psi}}(R_0) > k$ or $F_{\mathbf{\Psi}}(b_u) < k$, then there is no root.

- Otherwise, there is one root. We update $R_1 = \min(R_0 + (k - F_{\boldsymbol{\Psi}}(R_0))/l_0, b_u)$. Solve the dual LP with demand $\boldsymbol{D} = \boldsymbol{\mu} + R_1 W \boldsymbol{\Psi}$ to obtain the corresponding $F_{\boldsymbol{\Psi}}(R_1)$ and slope l_1. We keep updating until we find the root R^*.

- If $l_0 < 0$, there are at most two different roots.

  - If $F_{\boldsymbol{\Psi}}(R_0) < k$ and $F_{\boldsymbol{\Psi}}(b_u) < k$, then there is no root.

  - If $F_{\boldsymbol{\Psi}}(R_0) \geq k$ and $F_{\boldsymbol{\Psi}}(b_u) < k$, then there is one root. We update $R_1 = R_0 + (k - F_{\boldsymbol{\Psi}}(R_0))/l_0$. Solve the dual LP with demand $\boldsymbol{D} = \boldsymbol{\mu} + R_1 W \boldsymbol{\Psi}$ to obtain the corresponding $F_{\boldsymbol{\Psi}}(R_1)$ and slope l_1. We keep updating until we find the root R^*.

  - If $F_{\boldsymbol{\Psi}}(R_0) < k$ and $F_{\boldsymbol{\Psi}}(b_u) \geq k$, then there exists one root. We first increase R from R_0 to R_1 in $J_{\boldsymbol{\Psi}}$ using a fixed step way such that $l_1 > 0$. We try $R_1 = \min(R_0 + t, b_u)$, where t is the fixed step length. If $l_1 \leq 0$, we try $R_1 = \min(R_0 + 2t, b_u)$. We increase until we find a point R_1 with positive slope. Then we update $R_2 = \min(R_1 + (k - F_{\boldsymbol{\Psi}}(R_1))/l_1, b_u)$. Solve the dual LP with demand $\boldsymbol{D} = \boldsymbol{\mu} + R_2 W \boldsymbol{\Psi}$ to obtain the corresponding $F_{\boldsymbol{\Psi}}(R_2)$ and slope l_2. We keep updating until we find the root R^*.

  - If $F_{\boldsymbol{\Psi}}(R_0) \geq k$ and $F_{\boldsymbol{\Psi}}(b_u) \geq k$, then there exists no root or there exist two roots. We first test if there exists a root. We update $R_1 = R_0 + (k - F_{\boldsymbol{\Psi}}(R_0))/l_0$. Solve the dual LP with demand $\boldsymbol{D} = \boldsymbol{\mu} + R_1 W \boldsymbol{\Psi}$ to obtain the corresponding $F_{\boldsymbol{\Psi}}(R_1)$ and slope l_1. We keep updating until we find a root $R^{*(1)}$, or we reach some point with non-negative slope before we find a root. If the latter case happens,

there is no root. If the former case happens, there exist two roots. We increase R from $R^{*(1)}$ to R'_1 in $J_{\mathbf{\Psi}}$ using the fixed step way such that $l'_1 > 0$, and then update $R'_2 = \min(R'_1 + (k - F_{\mathbf{\Psi}}(R'_1))/l'_1, b_u)$. Solve the dual LP with demand $\boldsymbol{D} = \boldsymbol{\mu} + R'_2 W \mathbf{\Psi}$ to obtain the corresponding $F_{\mathbf{\Psi}}(R'_2)$ and slope l'_2. We keep updating until we find the second root $R^{*(2)}$.

2.6 Numerical Examples

Here we use the same basis for comparing the estimators using different simulation algorithms as in [18]. 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 [30], 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.

2.6.1 Example 1: $d = 3$, fixed $\boldsymbol{s}$

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 2.1: 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}

2.6.2 Example 2: $d = 10$, fixed $\boldsymbol{s}$

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 2.2: 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}

2.6.3 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 2.3: 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}

2.6.4 Example 4: $d = 10$, fixed k

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

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

Table 2.4: 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}

2.6.5 Discussion of Results and Comparison 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 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 asymptotically 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 2.5.1) 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 [18]) 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.

2.7 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 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 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, $\mathbf{\Psi}$. Therefore the conditional Monte Carlo algorithm applies as long as the demand vector $\boldsymbol{D}$ 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 [41]). 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 [41] indicates

that the following equality in distribution holds, namely,

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

where $\mathbf{\Psi}$ is uniformly distributed in the l_1 ball, and R is independent of $\mathbf{\Psi}$, with a distribution depending on $\varphi(\cdot)$ to be discussed shortly. We can simulate $\mathbf{\Psi}$ 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 chapter, 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.

Chapter 3

Insurance-Reinsurance Networks

3.1 Introduction

We develop both modeling and simulation methodology for a class of networks consisting of multiple institutions with inter-correlated exposures. Our modeling framework is introduced in light of an insurance setting, which involves a market with two types of participants: insurance and reinsurance companies. The dynamics of the network are governed by well-defined multidimensional stochastic processes subject to funding and contractual constraints. An equilibrium approach is taken to arbitrate settlements when company defaults occur. The types of models we consider are closely related to and, in fact, inspired by network models that have been analyzed in the literature in recent years, for example [23], [24], [5], [29] and [36], to name a few.

Some stylized features of insurance markets that make our framework reasonable include: 1) risk factors display fat tails and independent structures in time, e.g., annual

hurricane risks, and 2) the contractual relationships, summarized in a contractual graph to be defined later, either remain constant, when there is no default, or the associated graph reduces to a subgraph, if defaults occur.

We say a *market or system dislocation* occurs when a specific group of participants fails. Using our results and simulation procedures we aim at characterizing the features that dictate a significant change in the nature of the system's exposures given market dislocation. For instance, if a specific set of market participants is not sufficiently capitalized to fulfill their obligations, what is the most likely responsible externalities for such a scenario, a systemic shock in the market or a sequence of specific idiosyncratic events pertaining to the specific set of participants? (See Theorem 3.3.1.) We believe that our results can aid policy makers by shedding light on regulatory questions such as determining the appropriate capital requirements. Our goal is to provide a framework for systemically studying and gauging risks in such stochastic insurance-reinsurance networks. Our contributions spanning modeling, asymptotic analysis and computations are summarized as follows:

a) We propose a network model that allows to deal with default contagion risks. Cascading default losses are captured by means of the solution of a *linear programming problem* that can be interpreted in terms of an equilibrium. This formulation allows us to define the evolution of reserve processes in the network throughout time, see Proposition 3.2.4.

b) The linear programming formulation and therefore the associated equilibrium of

settlements at the time of default recognizes: 1) the correlations among the risk factors, which are assumed to follow a linear factor model, 2) the contractual obligations among the companies, which are assumed to follow popular contracts in the insurance industry (such as proportional reinsurance retrocession), and 3) the interconnectedness of the network. The equilibrium approach we adopted (see (3.2.8)) turns out to be closely related to the market clearing framework established in [29], see Remark 3.2.1. Our approach, however, permits companies to net against each other's losses in the wake of default.

c) Our model allows to obtain asymptotic results and a description of the asymptotic most likely way market dislocation can occur. This description is fleshed out explicitly by means of an *integer programming problem* (a multidimensional Knapsack problem). Such a description emphasizes the impact on the systemic risk landscape of the interactions between the severity of the exogenous claims, their dependence structure, and the interconnectedness of the companies, see Theorem 3.3.1 and Theorem 3.3.2.

d) Our model provides a method to quantify the higher economic role that companies with large endowment play in the insurance market. This is achieved in terms of an integer programming problem, which is the result of a large deviations asymptotic limit, and reduces to the Knapsack problem mentioned in c) when companies have comparable economic power. Theorem 3.3.2 thus puts forward a regulatory suggestion, pointing out that companies might play only a limited role in mitigating

systemic risks unless adequate capital requirements are in place (these should be an order of magnitude higher than those held by regular companies).

e) On the computation side, we propose a class of strongly efficient estimators for computing the expected loss of the network at the time of dislocation conditioned on the event that a specific set of market participants fails to meet their obligations. In addition, these estimators allow to compute associated conditional distributions of the network exposures given the dislocation of a specific set of companies. The estimation of these conditional distributions is performed with a computational cost (as measured by the number of simulation replications) that remains bounded even if the event of interest becomes increasingly rare, see Theorem 3.4.2. The design of strongly efficient estimators in this setting presents some challenges as the most likely paths occur through several jumps, (according to the solution to the combinatorial problem in c)), instead of only one jump, as it is customary in heavy tailed models.

We are aware of only a limited amount of research that provides a risk analytical framework in an integrated insurance-reinsurance market with heavy-tailed risks. The work of [47] considers a simple two-node insurance-reinsurance network involving light-tailed claims. Our work, however, puts into consideration a more complex and general network that captures more stylized features of the insurance market in practice. We have formulated our results in terms of regularly varying distributions for simplicity. Deriving logarithmic asymptotics with basically the same qualitative conclusions under other types

of tail distributions, such as intermediate regularly varying or lognormal distributions, is straightforward; Weibullian distributions present additional complications. Our asymptotic results are obtained with the intention of gaining qualitative insight in the form of logarithmic asymptotics. The role of the simulation algorithms, then, is to endow these asymptotic approximations with a computational device that allows one to efficiently obtain quantitatively accurate results. Thus, the entire approach we use, namely analysis and efficient computation, must be thought as a coherent contribution.

Now, as the connections in the network increase, one must account for all possibilities in which failure can occur. We have aimed at laying out a program which is adequate for a fixed network architecture, as the probability of a failure event becomes more rare. The question of performing rare-event analysis in large networks is certainly important and we plan to investigate this avenue in future research.

The rest of this chapter is organized as follows. Section 3.2 describes in detail our network model and discuss the associated linear programming formulation for the evolution of contract settlements in the event of company failures. The asymptotic analysis of the model is given in Section 3.3. Section 3.4 proposes a simulation scheme that balances practicality and efficiency, accompanied by a rigorous efficiency analysis at the end of the section. Numerical experiments are given in Section 3.5 on a test network under various configurations and target sets.

3.2 The Network Model and Its Properties

3.2.1 The Eisenberg and Noe Framework

Since our model is closely related to the framework developed in [29], we explain it first.

Suppose there are n nodes in a network. Each node i has an initial cash flow e_i, and is subject to liability $\bar{p}_i$. L_{is} represents the liability of node i to node s, $L_{ii} = 0$, and $\bar{p}_i = \sum_{s=1}^{n} L_{is}$. The relative liability matrix Π is defined as

$$\Pi_{is} = \begin{cases} \frac{L_{is}}{\bar{p}_i}, & \text{if } \bar{p}_i > 0 \\ 0, & \text{otherwise} \end{cases}. \tag{3.2.1}$$

Let p_i represent the payment made by node i to other nodes, and $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)'$. Then a clearing vector can be decided by the following linear program:

$$\begin{aligned} [EN1]: \quad \max \quad & \mathbf{1}'\boldsymbol{p} \\ \text{s.t.} \quad & \boldsymbol{p} \leq \Pi'\boldsymbol{p} + \boldsymbol{e} \\ & \boldsymbol{p} \leq \bar{\boldsymbol{p}} \\ & \boldsymbol{p} \geq \mathbf{0}, \end{aligned} \tag{3.2.2}$$

where $\mathbf{1}$ is the n-dimensional column vector with all components equal to 1, $\mathbf{0}$ is the n-dimensional column vector with all components equal to 0, $\boldsymbol{e} = (e_1, e_2, \ldots, e_n)'$, and $\bar{\boldsymbol{p}} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_n)'$.

We would like to point out that, actually, constraint (3.2.2) is redundant, i.e. $[EN1]$ is equivalent to:

$$
\begin{aligned}
[EN2]: \quad \max \quad & \mathbf{1}'\boldsymbol{p} \\
\text{s.t.} \quad & \boldsymbol{p} \leq \Pi'\boldsymbol{p} + \boldsymbol{e} \\
& \boldsymbol{p} \leq \bar{\boldsymbol{p}}
\end{aligned}
$$

To see this, we prove by contradiction. Suppose we ignore constraint (3.2.2), and at some optimal solution $\boldsymbol{p}^*$ to the new optimization problem $[EN2]$, there exist some elements that are negative. We can construct a new vector $\tilde{\boldsymbol{p}}^*$ by replacing all negative elements of $\boldsymbol{p}^*$ with 0. Then the new vector is also feasible to $[EN1]$ but have a larger value of objective function, which contracts with the optimality of $\boldsymbol{p}^*$. Therefore, we can remove constraint (3.2.2) from now on.

3.2.2 An Equivalent Form

To find a clearing vector, while the Eisenberg and Noe framework maximizes the total payments, a more intuitive way is to minimize the total unserved demands. We introduce

slack variables $\boldsymbol{x}^+$ and $\boldsymbol{x}^-$, and get the equivalent form:

$$\begin{aligned} \max \quad & \mathbf{1}'\boldsymbol{p} \\ \text{s.t.} \quad & \boldsymbol{p}+\boldsymbol{x}^+ = \Pi'\boldsymbol{p}+\boldsymbol{e} \\ & \boldsymbol{p}+\boldsymbol{x}^- = \bar{\boldsymbol{p}} \\ & \boldsymbol{x}^+, \boldsymbol{x}^- \geq \mathbf{0}. \end{aligned}$$

Replacing $\boldsymbol{p}$ with $\bar{\boldsymbol{p}} - \boldsymbol{x}^-$ gives the following equivalent form

$$\begin{aligned} \min \quad & \mathbf{1}'\boldsymbol{x}^- \\ \text{s.t.} \quad & \boldsymbol{x}^+ - \boldsymbol{x}^- = \boldsymbol{e} - (I-\Pi')\bar{\boldsymbol{p}} - \Pi'\boldsymbol{x}^- \\ & \boldsymbol{p}+\boldsymbol{x}^- = \bar{\boldsymbol{p}} \\ & \boldsymbol{x}^+, \boldsymbol{x}^- \geq \mathbf{0}, \end{aligned} \tag{3.2.3}$$

where I is the identity matrix. Note that $\bar{\boldsymbol{p}}$ is constant and has no impact on the optimal solution, so it has been removed from the objective function. Also since there is no constraint on $\boldsymbol{p}$, equation (3.2.3) can be removed, and we obtain:

$$\begin{aligned} \min \quad & \mathbf{1}'\boldsymbol{x}^- \\ \text{s.t.} \quad & \boldsymbol{x}^+ - \boldsymbol{x}^- = \boldsymbol{e} - (I-\Pi')\bar{\boldsymbol{p}} - \Pi'\boldsymbol{x}^- \\ & \boldsymbol{x}^+, \boldsymbol{x}^- \geq \mathbf{0}, \end{aligned} \tag{3.2.4}$$

Our model is constructed by adding randomness features based on (3.2.4) and settlement mechanism developed in [29].

3.2.3 The Network Model

In this section we provide a precise description of our model in light of the insurance setting. The architecture we build reflects a stylized insurance market structure. [1] Specifically, we consider an insurance market with two types of roles:

- *Insurance companies or Insurers* whose core business involves underwriting insurance policies and thereby providing protection to policy-holders.
- *Reinsurance companies or Reinsurers*, acting as "insurers of insurers", sell reinsurance contracts to insurance companies, in exchange for collections of reinsurance premiums to get funded.

In practice, there are insurance companies that act as reinsurance providers, so we don't differentiate between insurance companies and reinsurance companies unless otherwise stated. In order to cover typical features of an insurance market with these two sets of participants, the model is set up to allow reasonable generalities regarding

1) *contractual specifications*, popular in the insurance literature;

2) *network topology/architecture*, specifying the dependence among the participants;

[1] See p. 51 of "Systemic risk in insurance: An analysis of insurance and financial stability, Special Report of The Geneva Association Systemic Risk Working Group," 2010.

3) *settlement/clearing mechanisms*, allowing us to deal with the settlements of cascading defaults.

Contractual Specifications and Network Topology

Let us denote by $\mathcal{I} = \{1, 2, \ldots, K\}$ the set of companies in the market, including both insurance and reinsurance companies. We then endow the following claim structure to this insurance network.

Claim arrival and heavy-tailed claim structure

We consider a single-period setting. Claims arrive at each company I_i, $i = 1, \ldots, K$ exogenously. We add up all claims and consider them as one single claim. The claim size is assumed to adopt a linear factor model with heavy-tailed structure. Let V_i be the size of the claim that I_i incurs, and we have

$$V_i = \sum_{h=1}^{d} \gamma_{i,h} Z_h + \beta_i Y_i. \tag{3.2.5}$$

Here $\{Z_h\}_{h \leq d}$ is a series of common factors. For instance, if $d = 2$, Z_1, Z_2 might represent exposures to hurricanes and earthquakes, respectively. Non-negative constant $\gamma_{i,h}$ is the associated factor loading for I_i to Z_h. If $\gamma_{i,h}$ is positive, I_i is exposed to Z_h. The set of common factors $\{Z_h\}$ quantifies the "sectoral risk" that is shared by a subset of companies. On the other hand, Y_i is the idiosyncratic factor corresponding to I_i and is independent of all the common factors Z_h, $h \leq d$. Non-negative constant β_i is the factor loading for I_i associated with Y_i. In general $\gamma_{i,h}$'s and β_i's can be taken as random, as long

as $E\left[\exp\left(\theta\gamma_{i,h}\right)\right] < \infty$ and $E\left[\exp\left(\theta\beta_i\right)\right] < \infty$ for any $\theta > 0$. For simplicity, we assume the factor loadings are deterministic. We also assume that for each $\sum_{h=1}^{d} \gamma_{i,h} + \beta_i > 0, \forall i \in \mathcal{I}$, which is reasonable since otherwise that company has no exposure at all. The factors are assumed to be independent in time.

Factors are assumed to have heavy tails. In particular, they belong to the class of regularly varying distributions. Specifically, we assume $Z_h \in \mathcal{RV}(\alpha_h^Z)$, $Y_i \in \mathcal{RV}(\alpha_i)$. The regularly varying class requires the random variable to basically possess polynomial decaying tails. It clearly includes the well-known Pareto distributions. Recall that a random variable X is said to have Pareto distribution, $X \sim Pareto\left(\theta, \alpha\right)$, if

$$P\left\{X > x\right\} = \left(\frac{\theta}{\theta + x}\right)^{\alpha}, \; x > 0.$$

Sources of Funding and Risks

Companies are funded by: 1) an initial reserve and 2) net premiums, with the latter defined as

$$\text{net premiums} = \text{premiums received} - \text{premiums paid}.$$

For each $i \in \mathcal{I}$, we denote by u_i the initial reserve for I_i; and denote by C_i the total net premiums for I_i. On the other hand, companies are faced with two sources of risks:

1) *Claim risks*, which is the primary risk incurred in the main insurance/reinsurance business. This type of risk is captured in the *effective claims*, which is defined as

$$\text{effective claims} = \text{total liability} - \text{liability reinsured}. \tag{3.2.6}$$

We shall detail the mathematical form of the definition later in Subsection 3.2.3.

2) *Reinsurance counterparty risks*, which is the risk associated with the event of reinsurance counterparty defaults and the subsequent dishonor of reinsurance payments.

As a result, one can define the end reserve as follows

end reserve = initial reserve + net premiums − effective claims − losses from counterparty defaults.

Naturally, company I_i is *insolvent* if the available funds is insufficient to cover the losses from the realized risks, that is, if the end reserve level, denoted by u_i' for $i \in \mathcal{I}$, is negative. Precise definitions of $\{u_i'\}_{i\in\mathcal{I}}$ will be given in (3.2.9) later in Subsection 3.2.3. If a company is insolvent, a *default* process is triggered, which has direct impacts on the solvency of other companies. A detailed rule must be specified to ensure that such a process be executed in an orderly fashion, and that all the counterparties (and counterparties of any subsequently defaulted companies) get their deserved shares of default payments. We assume that defaults only occur at the end of the period.

While recognizing that in reality, a company is faced with more sophisticated sources of risks than the ones specified above (for example, market risks from trading financial derivatives such as credit default swaps), we restrict the risks to claim risks and reinsurance counterparty risks in order to single out and gauge the risks primarily from the insurance business. The insurance/reinsurance companies might be inter-connected with respect to those other risks that are not directly modeled here. But from a regulatory perspective, any such risks can be thought of *incremental* to the claim risks and reinsur-

ance counterparty risks, and can be segregated and studied in separate frameworks. Note that this "divide and conquer" approach is directly applied in the *Basel* 2.5 regulatory framework for banks' trading book capital, in which relatively long-term credit migration/default losses are monitored within the *incremental risk charge* framework, whereas short-term market risks are scrutinized using the *VaR* and *Stressed VaR* measures.

Contractual Links and Network Topology

Naturally, the effective claims received by the companies are contingent on the survival of their counterparties, which in turn is influenced by how the participants deal with each other in the network. It is therefore crucial to first set the rules that govern the contractual connectivity of the network. Once such rules are enforced, we are able to visualize the business interactions among the participants in a directed graph, as depicted in Figure 3.1. We now state the rules in the following assumption.

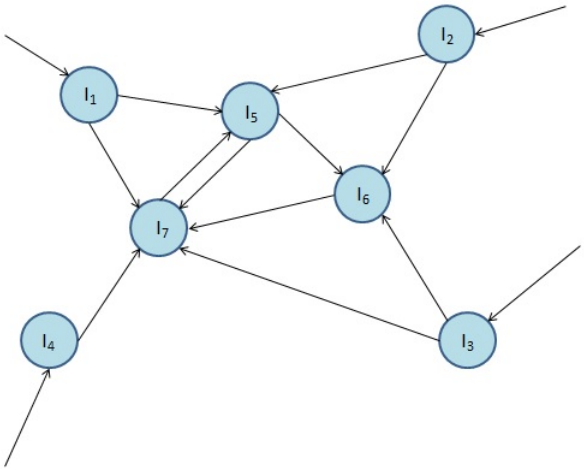


Figure 3.1: Directed contractual graph of a simple example network.

Assumption 3.2.1 (Contractual Links and Network Topology)**.**

i) ***Liability Allocation:*** *Each company I_i enters into contracts with a set of other companies. The exposure taken by another company I_s, is a proportion, $\omega_{i,s} \geq 0$, of each claim. $\omega_{i,i}$ represents the retained proportion by I_i. We assume that $\sum_{s\in\mathcal{I}} \omega_{i,s} = 1$, $\forall i \in \mathcal{I}$. If $\omega_{i,s} > 0, i \neq s$, there is a directed edge from I_i to I_s in the graph representing a contractual presence in the network.*

ii) ***Reinsuring:*** *Each company $I_s, s \in \mathcal{I}$, cannot reinsure the exposure transferred from one company $I_{s_1}, s_1 \neq s$ to some other company $I_{s_2}, s_2 \neq s_1, s$ (i.e. there are only two "hops" in the reinsurance sequence).*

iii) ***Network Connectivity:*** *For any two companies $i, s \in \mathcal{I}$, we assume that there is a path connecting them.*

iv) ***Extra Node*** *There is an extra node $I_{|\mathcal{I}|+1}$ in this network representing the customers served by these companies. Let $\mathcal{I}^+$ represent the union of this extra node and $\mathcal{I}$.*

Assumption 3.2.1-ii) forbids a company to cede coverage back to the companies which initially seek protection on that particular coverage. The stipulation of no more than two 'hops' in the retrocession sequence is imposed merely for the sake of expositional simplicity. In fact, as long as the contract ends up with a party other than the one that buys protection at the first place, or equivalently if the "hops" do not create a "loop", the framework introduced in this chapter works. Note also that Assumption 3.2.1-iii) is a very mild one. If the network is not connected, we can study each connected part separately.

Matrix Π and Effective Claims

Now we are in a good position to construct the matrix Π defined in (3.2.1). To do this, we only need to know $L_{is}, i \neq s$. Let V_i denote the size of the claim received by i, then $L_{is} = \omega_{si} V_s$. Let $L_{|\mathcal{I}|+1,i} = 0, i \in \mathcal{I}$. So Π can be computed as soon as all the claims to the network system within a given period have been collected.

The effective claim for I_i is

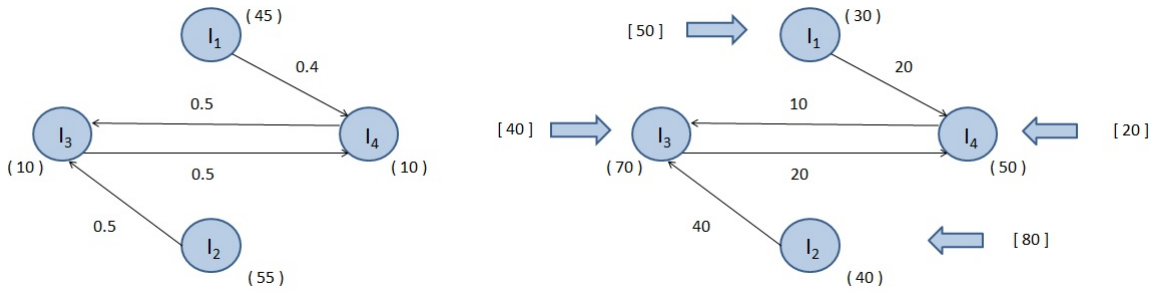
$$\begin{aligned} H_i &= \text{total liability} - \text{liability reinsured} \\ &= \bar{p}_i - \sum_{s \in \mathcal{I}} \Pi_{si} \bar{p}_s \\ &= \sum_{s \in \mathcal{I}} V_s \omega_{si}. \end{aligned} \tag{3.2.7}$$

Settlement Mechanism

At the end of a period, each existing company in the network is faced with *the settlement of the claims* collected during the period. In the event of a default, the state of the system is defined according to a settlement mechanism. In order to describe the mechanism, we first specify the following assumption on the rules governing the allocation of spillover losses in the network system.

Assumption 3.2.2 (Rules for Spillover Loss Allocation)**.** *Upon the incident of I_s defaulting during a period, $I_i, i \neq s$ gets partially settled by an amount proportional to its unsettled exposure to I_s, if any. This proportion is Π, i.e., Π_{si} the proportion of spillover loss that I_i gets if I_s fails, $i, s \in \mathcal{I}$, $i \neq s$.*

Nevertheless, having Assumption 3.2.2 alone turns out to be inadequate to secure a well-defined settlement mechanism in the event of a cascade of failures. Let us take a closer look using the following example.



(a) Initial Configurations (b) Before Write-offs

Figure 3.2: A network example

a) For each company, the initial reserve plus premium is given in the parentheses next to it. Transfer ratios ω are given next to the arrow representing the flow of contracts.

b) State of the network after all claims have been collected, before the write-offs. Bracketed numbers are the sizes of the claims V (defined in (3.2.5)). Numbers in parentheses are effective claims to the companies. And the rest is the transferred amount.

Consider the simple network illustrated in Figure 3.2. Let I_5 represent the extra node, then the Liability matrix L, matrix Π, vector $\bar{p}$, and the effective claims vector H are as follows:

$$L = \begin{pmatrix} 0 & 0 & 0 & 0 & 50 \\ 0 & 0 & 0 & 0 & 80 \\ 0 & 40 & 0 & 10 & 40 \\ 20 & 0 & 20 & 0 & 20 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Pi = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & \frac{4}{9} & 0 & \frac{1}{9} & \frac{4}{9} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \bar{p} = \begin{pmatrix} 50 \\ 80 \\ 90 \\ 60 \\ 0 \end{pmatrix}, \quad H = \begin{pmatrix} 30 \\ 40 \\ 70 \\ 50 \\ -190 \end{pmatrix}.$$

Right after the claims have been collected, I_3 does not have sufficient reserve base to buffer the size of the claims arrived at that period. A write-off procedure is therefore triggered. According to Assumption 3.2.2, I_4 will get an amount of the spillover loss from I_3 equal to $(10-70)\times(1/9) = -20/3$. With this allocation of contagion loss, I_4 is subsequently forced to fail because $10-50-20/3 < 0$. But we immediately run into a dilemma if the recurrent spillover loss from I_4 is to be allocated to I_1 and I_3: should I_3, a bankrupt company, take on the spillover loss from I_4? If we allow this process to iterate by arguing that any failure/bankruptcy shall not be declared until all the subsequent cascading write-offs are settled, then a more precise write-off mechanism is called for to ensure a unique network state after all the contagion losses have been settled and received.

Network Equilibrium

In order to address the afore-mentioned issue, we take the *equilibrium approach* as stated earlier. In particular, we require that, in addition to the principle stipulated in Assumption 3.2.2, the companies work out the spillover loss allocation at the end of each period

according to the single-period linear optimization problem (3.2.3), which gives an equilibrium state of the network after all companies mark write-offs and make settlements.

Denote by u_i the levels of reserves at the beginning of the period for I_i, C_i the net premium, $i \in \mathcal{I}$. Recall that e_i represents initial cash flow, together with equation (3.2.7), (3.2.3) can be rewritten as:

$$\begin{aligned}
\left[P^+\right]: \quad \min \quad & \sum_{i\in\mathcal{I}^+} x_i^- \\
\text{s.t.} \quad & x_i^+ - x_i^- = e_i - H_i - \sum_{s\in\mathcal{I}^+} x_s^- \cdot \Pi_{si}, \qquad \forall i \in \mathcal{I}^+ \\
& x_i^+, x_i^- \geq 0, \qquad \forall i \in \mathcal{I}^+,
\end{aligned}$$

where $e_i = u_i + C_i$.

Suppose I_e represent the extra node. It is easy to see that the optimal value $x_e^{-*} = 0$, so we only need to consider $\mathcal{I}$.

Netting

Note that $L_{is} \neq L_{si}$. We can introduce a netting parameter $\kappa \in [0,1]$, which controls the degree of *netting agreement* between two companies. Define the netted liability as

$$\tilde{L}_{is} = L_{is} - \kappa \cdot \min(L_{is}, L_{si}).$$

When $\kappa = 0$, none of the contracts between two companies are netted, which is the case as we discussed above and the case in [29] setting. And $\kappa = 1$ corresponds to a

fully netted scenario, for example, when all the contracts between two nodes are fungible/exchangeable. Of course the netting parameter κ can be made arc dependent, but for simplicity we consider the situation where κ is identical throughout the network.

In fact, enforcing netting can have a substantial impact on defaults.

Theorem 3.2.3. *Netting is beneficial (in terms of smaller optimal value) when there is at most one company defaults.*

Proof. Recall $[EN2]$:

$$\begin{aligned} [EN2]: \quad \max \ & \mathbf{1}'\boldsymbol{p} \\ \text{s.t.} \ & \boldsymbol{p} \leq \Pi'\boldsymbol{p} + \boldsymbol{e} \\ & \boldsymbol{p} \leq \bar{\boldsymbol{p}}. \end{aligned}$$

Let $M_{ij} = \min(L_{ij}, L_{ji}), r_i = \sum_{s=1}^{n} M_{is}, \bar{\boldsymbol{r}} = (r_1, r_2, \ldots, r_n)'$. Define $v_i \overset{\Delta}{=} \frac{p_i}{\bar{p}_i - \kappa \bar{p}_i}$. In other words, v_i represents the proprotion of liability that has been paid for I_i. Let $\bar{\boldsymbol{v}} = (v_1, v_2, \ldots, v_n)'$. Taking κ into account, we can rewrite $[EN2]$ as

$$\begin{aligned} [EN3]: \quad \max \ & (\bar{\boldsymbol{p}} - \kappa \bar{\boldsymbol{r}})' \bar{\boldsymbol{v}} & \\ \text{s.t.} \ & W\bar{\boldsymbol{v}} \leq \boldsymbol{e} & (\boldsymbol{\alpha}) \\ & \bar{\boldsymbol{v}} \leq \mathbf{1} & (\boldsymbol{\beta}) \end{aligned}$$

where,

$$W_{ij} = \begin{cases} \bar{p}_i - \kappa \bar{r}_i, & \text{if } i = j \\ -(L_{ji} - \kappa M_{ji}), & \text{otherwise.} \end{cases}$$

Its dual problem is:

$$\begin{aligned} [D]: \quad \min \quad & \boldsymbol{e}'\boldsymbol{\alpha} + \mathbf{1}'\boldsymbol{\beta} \\ \text{s.t.} \quad & W'\boldsymbol{\alpha} + \boldsymbol{\beta} = \bar{\boldsymbol{p}} - \kappa \bar{\boldsymbol{r}} \\ & \boldsymbol{\alpha} \geq \mathbf{0}, \boldsymbol{\beta} \geq \mathbf{0}. \end{aligned}$$

Suppose $\bar{\boldsymbol{v}}^*$ and $\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix}$ are the optimal solutions to the primal and dual problems respectively. Then they satisfy the *Karush-Kuhn-Tucher (KKT)* conditions, i.e.

$$\begin{cases} \alpha_i (W\bar{\boldsymbol{v}} - \boldsymbol{e})_i = 0, \forall i \\ \beta_i (\bar{\boldsymbol{v}} - \mathbf{1})_i = 0, \forall i \\ W\bar{\boldsymbol{v}} \leq \boldsymbol{e} \\ \bar{\boldsymbol{v}} \leq \mathbf{1} \\ W'\boldsymbol{\alpha} + \boldsymbol{\beta} = \bar{\boldsymbol{p}} - \kappa \bar{\boldsymbol{r}} \\ \boldsymbol{\alpha} \geq \mathbf{0}, \boldsymbol{\beta} \geq \mathbf{0}, \end{cases}$$

where $(\cdot)_i$ represents the i-th element of a vector.

- When there is no company defaults.

  It is easy to see that $\bar{v}^* = \mathbf{1}, \boldsymbol{\alpha} = \mathbf{0}, \boldsymbol{\beta} = \bar{\boldsymbol{p}} - \kappa\bar{\boldsymbol{p}}$. When the netting parameter is set to $1 \geq \kappa' > \kappa$. Then $\hat{v}^* = \mathbf{1}, \hat{\boldsymbol{\alpha}} = \mathbf{0}, \hat{\boldsymbol{\beta}} = \bar{\boldsymbol{p}} - \kappa'\bar{\boldsymbol{p}}$ also satisfy the (KKT) conditions, thus are the optimal solutions. Since $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$, $\hat{\boldsymbol{\beta}} \leq \boldsymbol{\beta}$, doing more netting is beneficial.

- When there is one company i_0 defaults.

  Then the optimal solutions are

$$\bar{v}_i* = \begin{cases} \frac{e_{i_0} + \sum_{j \neq i_0} (L_{j,i_0} - \kappa M_{j,i_0})}{\bar{p}_{i_0} - \kappa \bar{r}_{i_0}}, & \text{if } i = i_0 \\ 1, & \text{otherwise.} \end{cases}$$

$$\alpha_i = \begin{cases} 1, & \text{if } i = i_0 \\ 0, & \text{otherwise.} \end{cases}, \quad \beta_i = \begin{cases} 0, & \text{if } i = i_0 \\ (\bar{p}_i - \kappa \bar{r}_i) + (L_{i,i_0} - \kappa M_{i,i_0}), & \text{otherwise.} \end{cases}$$

  When the netting parameter is set to $1 \geq \kappa' > \kappa$, and $\kappa' - \kappa$ is small enough, the new optimal solutions are

$$\hat{v}_i* = \begin{cases} \frac{e_{i_0} + \sum_{j \neq i_0} (L_{j,i_0} - \kappa' M_{j,i_0})}{\bar{p}_{i_0} - \kappa' \bar{r}_{i_0}}, & \text{if } i = i_0 \\ 1, & \text{otherwise.} \end{cases}$$

$$\hat{\alpha}_i = \begin{cases} 1, & \text{if } i = i_0 \\ 0, & \text{otherwise.} \end{cases} \qquad \hat{\beta}_i = \begin{cases} 0, & \text{if } i = i_0 \\ (\bar{p}_i - \kappa' \bar{r}_i) + (L_{i,i_0} - \kappa' M_{i,i_0}), & \text{otherwise.} \end{cases}$$

Since $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$, $\hat{\boldsymbol{\beta}} \leq \boldsymbol{\beta}$, doing more netting is beneficial.

□

Using netted liability to calculate Π matrix, together with the fact that we only need to focus on $\mathcal{I}$, now the new linear program becomes:

$$\begin{aligned} [P]: \quad & \min \quad \sum_{i \in \mathcal{I}} x_i^- && (3.2.8) \\ & \text{s.t.} \quad x_i^+ - x_i^- = e_i - H_i - \sum_{s \in \mathcal{I}} x_s^- \cdot \Pi_{si}, \qquad \forall i \in \mathcal{I} && (I) \\ & \qquad x_i^+, x_i^- \geq 0, \qquad \forall i \in \mathcal{I}. \end{aligned}$$

We shall interpret the linear program shortly after we state the following results, which indicates desirable "stability" properties of the equilibrium state of the network underscored by the preceding linear program. Proposition 3.2.4 and 3.2.5 have been explained in [29]. Since our model is equivalent to Eisenberg and Noe model, these two results are also valid.

Proposition 3.2.4. *The linear program $[P]$, given in (3.2.8), admits a unique optimal solution.*

Moreover, it is easy to see that at this optimal solution, at least one element in each pair, $\left(x_i^+, x_i^-\right)$, is equal to zero, for each $i \in \mathcal{I}$.

Proposition 3.2.5. *Let* $\boldsymbol{x}^- = \left(\ldots, x_i^-, \ldots\right)$, $i \in \mathcal{I}$. *Let* $f\left(\boldsymbol{x}^-\right)$ *be a function that is differentiable and non-decreasing with respect to its variables. And define* $[P_f]$ *be the set of optimization problems with objective function* $f\left(\boldsymbol{x}^-\right)$ *and with constraints identical to the ones in* $[P]$. *Then the* $[P]$*-optimal solution is also* $[P_f]$*-optimal.*

The property of stable optimality suggests that, the equilibrium state found by solving $[P]$ is the best settlement solution to the system, as long as the companies in the network negotiate to minimize any sensible measure, f, of the incremental system loss.

Let us denote the optimal solution pairs to $[P]$ by $\{\check{x}_i^+, \check{x}_i^-\}_{i\in\mathcal{I}}$. At optimality, if $\check{x}_s^- > 0$ and $\check{x}_s^+ = 0$, it means that I_s has failed, and constraint (I) in $[P]$ ensures that I_i receives the contagion loss of amount equal to $\check{x}_s^- \cdot \Pi_{si}$. If the capital base of I_i is solid enough to weather the total spillover loss from other companies (which is represented by the amount $\sum_{s\in\mathcal{I}} \check{x}_s^- \Pi_{si}$), i.e., $e_i > H_i + \sum_{s\in\mathcal{I}} \check{x}_s^- \Pi_{si}$, then I_i will remain solvent, in which case $\check{x}_i^+ > 0 = \check{x}_i^-$. If otherwise, then I_i fails, in which case $\check{\pi}_i^+ = 0$. As a result, $\{\check{\pi}_i^-\}_{i\in\mathcal{I}}$ represent the loss at default for I_i at the equilibrium state of the network. The preceding optimization problem would yield the same optimal solution if we impose the additional constraint that $x_i^+ \times x_i^- = 0, \forall i \in \mathcal{I}^+$. Therefore, we can interpret the state associated with the optimal solution vector to $[P]$ as the equilibrium state of the network.

Remark 3.2.1. Our model is amenable to a similar type of analysis that has been done for the Eisenberg and Noe framework [37].

In addition to allowing default netting, the LP formulation developed in this chapter is convenient for our purpose because

a) The output of the linear optimization problem $[P]$ are the end-of-the-period reserve levels. [29]'s approach (in the case when $\kappa = 0$), however, requires an extra step of calculation in order to transform the payment vector to the vector of reserve levels, which is somewhat involved. See [48].

b) Our LP formulation, as we shall see in the discussion preceding Theorem 3.3.1, helps build natural intuitions on the large deviations description of the system.

Reserve Processes

At the end of period, the system reaches the equilibrium state associated with the unique optimal solution to $[P]$. And the end-of-period reserves are determined by the unique optimal solution vectors $\{\check{x}_i^+, \check{x}_i^-\}_{i\in\mathcal{I}}$ via

$$u_i' = \check{x}_i^+ - \check{x}_i^-, i \in \mathcal{I}. \tag{3.2.9}$$

Conditional Spillover Loss at System Dislocation

Usually, people are interested in the failure of a particular set of companies in the network, for example, companies in the same locations, with similar sizes or business lines. Motivated by the insurance applications discussed in the previous section, we shall study the performance measure *Conditional Spillover Loss at System Dislocation* which is in the form of a conditional expectation. In simple words, it is the expected loss to primary risk originators conditioned on the failure of a sub set of the network constituents.

We assume that A is a subset of $\mathcal{I}$, and define F_A as the event that all companies in set

A fail at the end of period. Define

$$D_i = -\min\{u_i', 0\},$$

the *lost reserve at system dislocation* at the end of period. We can therefore introduce the following formal definition of **Conditional Spillover Loss at System Dislocation**:

Definition 3.2.6. *The Conditional Spillover Loss at System Dislocation for the subset $A \subseteq \mathcal{I}$ is defined as*

$$\mathcal{D}(A) = E\left[\sum_{i\in\mathcal{I}} D_i \Big| F_A\right]. \tag{3.2.10}$$

The idea of such a measure is motivated by the so-called *Systemic Risk Index* or *Contagion Index*, following the terminology in [11], and studied in, for example [23].

3.3 Asymptotic Description of the Network

Having fixed the architecture of the network, we now embark on providing a qualitative characterization of the large deviations behavior of the system given F_A, i.e., the event of system dislocation caused by the set A. As we shall reveal momentarily, this characterization can be identified *by solving another optimization problem.* In the analysis that follows let us scale the initial cash reserves by $b > 0$. Assume that $e_i = r_i b$ is the initial reserve for I_i, $i \in \mathcal{I}$, where r_i is a fixed positive constant. Depending on the tail structure of the claim size distributions, the failure of the designated set arises from different numbers of extreme shocks in the claims. This particular feature of the system inspires us to tailor

a sequential algorithm for evaluating $\mathcal{D}(A)$, for any given set A, which we shall describe in details in the next section.

3.3.1 Large Deviations Description via An Integer Program

It is interesting to realize that useful implications about the asymptotic behavior of the system can be obtained from the linear program $[P]$ given in (3.2.8). To see this, recall that constraints (I) in (3.2.8) require, for each $i \in \mathcal{I}$ that,

$$x_i^+ - x_i^- = e_i - H_i - \sum_{s \in \mathcal{I}} x_s^- \cdot \Pi_{si}.$$

Since $e_i - H_i = \Theta_p(b)$ (recall that stochastic order notations have been explained in Sec 1.2.2), the intuition is that, $P(\check{x}_i^- > 0) = \Theta(1)$ if and only if there exists $s \in \mathcal{I}$, $s \neq i$, such that *both* of the following are satisfied:

i) $\check{x}_s^- = \Theta_p(b)$,

ii) $\Pi_{si} = \Theta_p(1)$.

In other words, both the default loss for I_s and the contractual link between I_i and I_s need to be sufficiently large in order for I_i to fail with $\Theta(1)$ probability. This can occur due to either of the following two possible cases:

a) $Z_h = \Theta_p(b)$, for some $1 \leq h \leq d$ such that $\gamma_{i,h} > 0$,

b) $Y_i = \Theta_p(b)$.

The intuitions above are certainly helpful, for now we are able to restrict the enumeration of possible paths (leading to the event F_A) down to a much smaller subset. In fact, as we shall see shortly, the combinatorial task of singling out the cheapest route to the target event boils down to solving a *Knapsack problem with multiple constraints*.

Now it is not difficult to see that the intercorrelation structure of the network can be summarized in a directed graph. An example is shown in Figure 3.3 below. The left part of the graph (a factor-company bipartite sub-graph) summarizes the exogenous risks underlying the participants, while the part of the graph on the right (company-company) reflects the endogenous default contagion risks, and corresponds to the directed contractual graph in Figure 3.1.

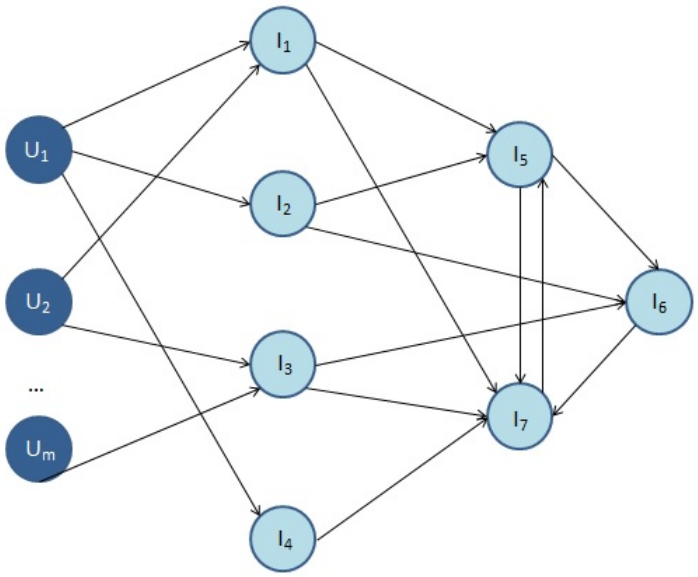


Figure 3.3: Directed graph combining risk factor exposures and contractual links.

Let us denote by Ξ the *factor exposure matrix* in the network, which is an $|\mathcal{I}| \times m$ matrix, where $m = (d + |\mathcal{I}|)$. Each column corresponds to a specific factor. We align the

factors in such a way that the first d factors are the common factors, and the remaining $|\mathcal{I}|$ factors are the idiosyncratic factors for companies in $\mathcal{I}$. On the other hand, each row of Ξ represents a particular market participant, and we align the companies such that the first $|A|$ rows correspond to set A and the last $|\mathcal{I}| - |A|$ rows correspond to set $A^c = \mathcal{I} - A$. The entries of Ξ are binary: $\Xi_{ij} = 1$ if and only if there exists a directed path from factor U_j, common or idiosyncratic, to the node representing company i. For example, in the network represented by Figure 3.3, $\Xi_{1,1} = 1$ and $\Xi_{3,1} = 0$. In other words, $\Xi_{ij} = 1$ if and only if company i is exposed to factor U_j, either in a direct or an indirect way.

Let Ξ_j^c be the j-th column of Ξ. In what follows we shall denote by U_j the factor corresponding to Ξ_j^c. Let us further define $\widetilde{\alpha}_j$ to be the regularly varying index of U_j, i.e., $\widetilde{\alpha}_j = \alpha_j^Z$ if $j \leq d$, and $\widetilde{\alpha}_j = \alpha_i$ if $j = i + d$, $i \in \mathcal{I}$. The following result shows that the large deviation description of the system is simply obtained by solving an *integer programming problem*, which is easily identified as a Knapsack problem with multiple knapsacks. We shall postpone the proof of the theorem to the end of §3.4. We mention that a one dimensional Knapsack formulation has also been proposed by [56] in the setting of heavy-tailed large deviations.

We first introduce some definitions. Define

$$\mathcal{R}(i) = \{s \in \mathcal{I} : \Pi_{si} > 0\} = \{s \in \mathcal{I}, s \neq i : \omega_{is} > 0\},$$

for $i \in \mathcal{I}$. In other words, $\mathcal{R}(i)$ is the set of reinsurance counterparties of I_i.

$$H(i) = \{s \in \mathcal{I} : \Pi_{is} > 0\} = \{s \in \mathcal{I}, s \neq i : \omega_{si} > 0\},$$

or $i \in \mathcal{I}$. In other words, $H(i)$ is the set of companies that I_i reinsures for. Let

$$\mathcal{R}(A) = \bigcup_{i \in A} \mathcal{R}(i), \quad H(A) = \bigcup_{i \in A} H(i).$$

Theorem 3.3.1. *As $b \to \infty$, we have*

$$\frac{\log P(F_A(b))}{\log b} \longrightarrow -\zeta(A), \tag{3.3.1}$$

where $\zeta(A)$ is the optimal cost to the following integer programming problem:

$$\begin{aligned} [IP]: \quad \min \quad & \sum_{j=1}^{m} \widetilde{\alpha}_j y_j \\ s.t. \quad & \sum_{j=1}^{m} y_j \Xi_{i,j} \geq 1, \ \forall i \in A \\ & y_j \in \{0,1\}, \ 1 \leq j \leq m \end{aligned} \tag{3.3.2}$$

Remark 3.3.1. For any $[IP]$-optimal solution $\boldsymbol{y}^* = (y_1^*, \ldots, y_m^*)'$, y_j^* is interpreted as the "indicator of activation" which dictates the occurrence of a large factor U_j. In particular, if $y_{j+d}^* = 1$, then $Y_j = \Theta(b)$ in the large deviations description of the system; if $y_h^* = 1$, for some $h \leq d$, then $Z_h = \Theta(b)$ in the large deviations description of the system. In general, knapsack problems are known to be *NP*-hard. However, since the exposure matrix Ξ has

binary entries, this particular knapsack problem is known to be solvable in polynomial time, see [43].

Remark 3.3.2. An interesting feature of this characterization is that the large deviations behavior of the network is dictated only by a set of tail indices. Depending on the choice of A, the description of the most likely way leading to F_A may change domains. For instance, the event F_{A_1}, where $A_1 = \{\text{AIG, Prudential}\}$, could most likely result from the occurrence of a few large common factors, while F_{A_2}, where $A_2 = \{\text{Lincoln Benefit, Northwestern Mutual}\}$, might occur most likely due to multiple phenomenal idiosyncrasies, or a mixture of extremal idiosyncratic and common shocks.

3.3.2 The Role of Companies in $\mathcal{R}(A)$

An important implication of Theorem 3.3.1 is that, the existence of the companies in $R = \mathcal{R}(A) - A$ does not alter the large deviations description of the network system if they have roughly the same economic power as companies in A. That is, the same Knapsack problem arises if no R companies exist in the system. This observation appears to be consistent with various empirical studies, which argue that reinsurance failure may not be a substantial source of systemic risk for the insurance industry, see for example [52]. [2] [3]

We could further strengthen the roles of the companies in R by enforcing a more

[2] See also p. 50 of "Systemic risk in insurance: An analysis of insurance and financial stability, Special Report of The Geneva Association Systemic Risk Working Group," 2010.

[3] See also "Reinsurance - a systemic risk? Sigma, Swiss Re," 2003.

stringent capital requirement for them. In order to see this, let us assume that

$$e_s = \Theta\left(b^{\rho}\right), \rho > 1,$$

for all $s \in R$, thereby demanding each company in R to pledge more capital than the companies in A (recall that $e_i = \Theta(b)$ for $i \in A$). The following result indicates that asymptotic description for the system with this modified assumption can be identified by solving a different integer programming problem.

Theorem 3.3.2. *As $b \to \infty$,*

$$\frac{\log P\left(F_A(b)\right)}{\log b} \longrightarrow -\widetilde{\zeta}(\rho, A),$$

where $\widetilde{\zeta}(\rho, A)$ is the optimal cost to the following integer programming problem:

$$\begin{aligned} \left[\widetilde{IP}^{(\rho)}\right]: \quad \min \quad & \sum_{j=1}^{m} \rho\widetilde{\alpha}_j z_j + \sum_{j=1}^{m} \widetilde{\alpha}_j y_j \\ s.t. \quad & \sum_{j=1}^{m} \Xi_{i,j} z_j \geq 1, \ \forall i \in R \\ & \sum_{j=1}^{m} \Xi_{l,j}\left(z_j + y_j\right) \geq 1, \ \forall l \in A \\ & z_j, y_j \in \{0,1\}, \ 1 \leq j \leq m \end{aligned} \tag{3.3.3}$$

Proof. This theorem can be proved using the similar way as the proof of Theorem 3.3.1. The idea is that, since $e_s = \Theta\left(b^{\rho}\right), \forall s \in R$, in order for the set A to fail, the cheapest way

is to bring down the reinsurance counterparty set of A first. Factors of size $\Omega(b^{\rho})$ are required for each failure in R. If factor j has magnitude $\Omega(b^{\rho})$, the overshoot from such failures, again of sizes $\Omega(b^{\rho})$ (see [44]), will lead to the default of any company k such that $\Xi_{kj} = 1$, with $\Omega(1)$ probability. A few more factors of size $\Omega(b)$ are needed to fail the surviving companies in A after R has failed. The optimization problem $\left[\widetilde{IP}^{(\rho)}\right]$ returns the routes with the largest probability (in logarithmic scale). □

Remark 3.3.3. In any $\left[\widetilde{IP}^{(\rho)}\right]$-optimal solution $(\boldsymbol{z}^*, \boldsymbol{y}^*)$, z_j^* and y_j^* are interpreted as the "strong" and "weak" activation indicators, respectively. If $z_j^* = 1$, then the corresponding factor U_j is among the factors that most likely lead to the failure of the counterparty set R, i.e., $U_j = \Theta(b^{\rho})$; if $y_j^* = 1$, then U_j is among the factors that result in the failure of some companies in A after they lost protections from their reinsurance counterparties, and in that case, $U_j = \Theta(b)$. The structure of the program ensures that in any optimal solution, $z_j^* y_j^* = 0$. This is because if there exists an optimal solution $(\boldsymbol{z}^*, \boldsymbol{y}^*)$ in which $z_k = y_k = 1$ for some k, then the solution, $\left(\widetilde{\boldsymbol{z}}^*, \widetilde{\boldsymbol{y}}^*\right)$ which is the same as $(\boldsymbol{z}^*, \boldsymbol{y}^*)$ except that the l-th pair is given by $(\widetilde{z}_l^*, \widetilde{y}_l^*) = (1,0)$, is also $\left[\widetilde{IP}^{(\rho)}\right]$-feasible, with smaller objectives, contradicting the optimality of $(\boldsymbol{z}^*, \boldsymbol{y}^*)$.

The characterization given by Theorem 3.3.2 offers a vehicle for guiding policy-level decisions from a regulatory stance. Questions such as "*what is the minimum level of capital required by the companies in R in order to control the probability of a specific set A?*" can be formulated and studied in a structured way.

3.4 Design of Efficient Simulation Algorithms

The asymptotic analysis in the preceding section is useful in obtaining a qualitative description of the systemic risk landscape of the entire network. The resulting asymptotic description is coarse. In this section we aim to achieve a more precise quantitative assessment by means of efficient computations. We resort to Monte Carlo methods, and our goal is to propose an efficient simulation algorithm to evaluate the *conditional system dislocation* (3.2.10), with the aid of the large deviations analysis presented in the previous section. We do this by designing an algorithm for the probability

$$q(b) = P\{F_A(b)\}$$

instead. An estimator for (3.2.10) is a natural consequence.

3.4.1 Importance Sampling

As explained in Sec 1.2.3, to develop an efficient Importance Sampling algorithm, we first study the asymptotic probability as $b \to \infty$.

Proposition 3.4.1. *Given the network $\mathcal{N}_e$ defined in Section 3.2, define*

$$\delta \stackrel{\Delta}{=} \min_{i \in A} \frac{r_i}{2\sum_{s \in i \cup H(i)} \left(\sum_{h=1}^{d} \gamma_{s,h} + \beta_s\right)}.$$

Let $\mathcal{X}$ be the set of feasible solutions to the IP given in (3.3.2). And define

$$\mathcal{A}_\delta(b) \stackrel{\Delta}{=} \bigcup_{\boldsymbol{y}\in\mathcal{X}} \left\{ \bigcap_{i\in A} \left[\left(\bigcup_{\substack{1\le h\le d \\ \gamma_{i,h}y_h>0}} \{Z_h \ge \delta b\} \right) \bigcup \left(\bigcup_{\beta_i y_{i+d}>0} \{Y_i \ge \delta b\} \right) \bigcup \left(\bigcup_{\substack{j\in H(i) \\ y_{j+d}=1}} \{Y_j \ge \delta b\} \right) \right] \right\}$$

Then we have

i) $\mathcal{A}_\delta(b)$ is a superset of $F_A(b)$, i.e.,

$$\mathcal{A}_\delta(b) \supseteq F_A(b). \tag{3.4.1}$$

ii) We have, as $b \to \infty$,

$$\frac{\log P\{\mathcal{A}_\delta(b)\}}{\log b} \longrightarrow -\zeta,$$

where ζ is the optimal cost to $[IP]$ in (3.3.2).

Proof. i) Suppose there exists $i' \in A$, such that 1) $Z_h < \delta b$ for all $h \le d$, $\gamma_{i',h}y_h > 0$, and 2) $Y_{i'} < \delta b$, $\beta_{i'}y_{i'+d} > 0$, and 3) $Y_j < \delta b$, for all $j \in H(i')$, $y_{j+d} = 1$, then we

have,

$$\begin{aligned} u'_{i'} &= r_{i'}b - \sum_{s\in\mathcal{I}} V_s\omega_{si'} - \sum_{s\in\mathcal{I}} \check{x}_s^- \cdot \Pi_{si'} \\ &= r_{i'}b - \sum_{s\in i'\cup H(i')} \left(\sum_{h=1}^{d} \gamma_{s,h}Z_h + \beta_s Y_s\right)\omega_{si'} - \sum_{s\in\mathcal{I}} \check{x}_s^- \cdot \Pi_{si'} \\ &\geq r_{i'}b - \delta b \cdot \sum_{s\in i'\cup H(i')} \left(\sum_{h=1}^{d} \gamma_{s,h} + \beta_s\right) - \sum_{s\in\mathcal{I}} \check{x}_s^- \cdot \Pi_{si'} \\ &\geq r_{i'}b/2 - \sum_{s\in\mathcal{I}} \check{x}_s^- \cdot \Pi_{si'}, \end{aligned}$$

where $\check{x}_s^-$ is the optimal solution x_s^-, $s \in \mathcal{I}$ of the linear program $[P]$. Furthermore, the model setup ensures that at any point in time, each company *cannot* receive an allocation of the spillover losses from all of its reinsurance counterparties of an aggregate amount larger than the total amount it reinsures. Actually,

$$x_s^- \cdot \Pi_{si} = x_s^- \frac{L_{si}}{\bar{p}_s} \leq L_{si}, \quad \forall s, i \in \mathcal{I}.$$

In what follows, we shall refer to this observation as *limited spillover impact*. Therefore, we have

$$\sum_{s\in\mathcal{I}} \check{x}_s^- \cdot \Pi_{si'} \leq \sum_{s\in i'\cup H(i')} \left(\sum_{h=1}^{d} \gamma_{s,h}Z_h + \beta_s Y_s\right) \leq r_{i'}b/2.$$

Consequently $u'_{i'} \geq 0$, and this implies that $F_A^c(b)$, where c represents complement set. We have thus established (3.4.1).

ii) An equivalent expression for $\mathcal{A}_\delta(b)$ is given by

$$\mathcal{A}_\delta(b) = \bigcup_{\boldsymbol{y}\in\mathcal{X}} \left\{ \bigcap_{i\in A} \left(\bigcup_{1\le j\le m, \Xi_{ij}y_j\ge 1} \left\{ U_j \ge \delta b \right\} \right) \right\},$$

where Ξ is the factor exposure matrix defined shortly before Theorem 3.3.1, and $m = d + |\mathcal{I}|$ is the number of columns of Ξ. Recall that $U_j = Z_h$ if $1 \le j \le d$, and $U_j = Y_i$ if $j = d+i$, $i \in \mathcal{I}$. For a feasible solution $\boldsymbol{y}$ associated with $[IP]$, let us further define

$$\mathcal{S}(\boldsymbol{y}) = \left\{ j = d+i : i \in \mathcal{I}, y_j = 1 \right\} \cup \left\{ h \le d : y_h = 1 \right\}, \tag{3.4.2}$$

i.e., $\mathcal{S}(\boldsymbol{y})$ is the index set of active factors associated with $[IP]$-feasible solution $\boldsymbol{y}$.

For the lower bound, we note that

$$\begin{aligned} P\{\mathcal{A}_\delta(b)\} &\ge P\left\{ \bigcap_{i\in A} \left(\bigcup_{1\le j\le m, \Xi_{ij}y^*_j\ge 1} \left\{ U_j \ge \delta b \right\} \right) \right\} \\ &= \prod_{j\in\mathcal{S}(\boldsymbol{y}^*)} P\left\{ U_j \ge \delta b \right\} \\ &\ge \kappa_1 b^{-\widetilde{\boldsymbol{\alpha}}'\boldsymbol{y}^*}, \end{aligned}$$

for some positive constant κ_1, where $\boldsymbol{y}^*$ is an $[IP]$-optimal solution.

And for the other direction, we utilize a union bound instead. In particular,

$$P\{\mathcal{A}_\delta(b)\} \le \sum_{\boldsymbol{y}\in\mathcal{X}} P\left\{ \bigcap_{i\in A} \left(\bigcup_{1\le j\le m, \Xi_{ij}y_j\ge 1} \left\{ U_j \ge \delta b \right\} \right) \right\} \le \kappa_2 b^{-\widetilde{\boldsymbol{\alpha}}'\boldsymbol{y}^*}, \tag{3.4.3}$$

for some positive constant κ_2, where $\boldsymbol{y}^*$ is again an optimal solution to $[IP]$. The result follows immediately after taking log for both the lower and upper bounds.

□

An immediate implication of the previous results is a sampling scheme that induces the occurrence of adequately large (of size at least δ) common or individual factors might be sufficient to guarantee bounded relative error of the estimator. From the simulation perspective, the order of occurrence of the factors during each period deems irrelevant. Our strategy is therefore to view the factors as if they arrive sequentially. We can consider the random sums of the factors, as random walks themselves, thereby creating this "internal" layer of random walks. From this point on we can borrow apparatus from established *state-dependent* rare event simulation algorithms to aid the design of our importance sampling estimator.

The key ingredient is a mixture based importance sampling distribution for the increments: with some probability p, the increment is sampled conditioning on it being "large", and with probability $1-p$, it's sampled as if it's a "normal" shock. Let X be the increment of the system, and without loss of generality suppose its density is given by $f(x)$, then the nth increment is drawn from the importance density $g(\cdot)$, defined as

$$g(x) = \left[p\frac{I\big(x \in A(b)\big)}{P\big\{X \in A(b)\big\}} + (1-p)\frac{I\left(x \in \overline{A(b)}\right)}{P\left\{X \in \overline{A(b)}\right\}} \right] f(x), \tag{3.4.4}$$

where $A(b)$ specifies the region in which the increment is qualified to be a large shock. Note that the part in (3.4.4) corresponding to the "normal" jumps is necessary in order

to conciliate the sensitivities of large deviations probabilities to the likelihood ratio of those paths that have more than one jump of order $\Omega(b)$, a crucial observation pointed out by [15], and [19].

$A(b)$ is typically chosen to be proportional to the "distance to go" for the current position of the random walk, and can be derived from some "auxiliary" or "steering" processes other than the targeting process. A convenient choice of such an auxiliary process in our setting is obtained by "eliminating" the participants in $\mathcal{R}(A)$ a priori and allocating the reserve process e_s, $s \in \mathcal{R}(A)$ proportionally to each e_i, $i \in A$. Equivalently, we pretend that the I_i's, $i \in A$ absorb full sized claims without reaching out to $\mathcal{R}(A)$ to hedge risks. In other words, by removing reinsurance we construct a convenient relaxation yielding a system with strictly small reserves for insurance companies.

At the beginning of the period, we first sample the common factors in order to strip off the first layer of dependence among the claims; and then sequentially sample the remaining individual factors. The mixture sampling density (3.4.4) is used to sample each factor, with the "distance to go" $A(b)$ properly defined. We shall detail this choice in the next subsection. The resulting sampling scheme is easy to carry out, self-adjusting in nature, and saves the user the trouble of setting up the algorithm differently according to different network structures. Formally we have the following efficiency result, the proof of which is postponed after we have detailed the algorithm in the next subsection.

Theorem 3.4.2. *The adaptive importance sampling estimator $\widehat{q}_{Z,Y}$ (to be defined in (3.4.9)) is strongly efficient for estimating $q(b)$. If, in addition, $\alpha_i > 2$, for all $i \in \mathcal{I}$, and $\alpha_h^Z > 2$,*

for all $1 \leq h \leq d$*, then the estimator*

$$\widehat{h}_{Z,Y} \stackrel{\Delta}{=} \sum_{i \in \mathcal{J}} \widehat{q}_{Z,Y} D_i(A)$$

is also strongly efficient for estimating $\mathcal{D}_0(A) = E\left[\left(\sum_{i\in\mathcal{J}} D_i(A)\right) I\left(F_A\right)\right]$.

3.4.2 The Algorithm

We are now ready to explain the state-dependent importance sampling idea in detail. Let

$$e_{i,j} = \begin{cases} 1, \text{if } \ i = j \\ \\ \omega_{i,j}, \text{otherwise.} \end{cases}$$

Define the auxiliary process via

$$\begin{aligned} S_i^{(0)} &= \sum_{k \in i \cup H(i)} e_{k,i} \sum_{h=1}^{d} \gamma_{k,h} Z_h, \\ S_i^{(1)} &= S_i^{(0)} + \sum_{k \in i \cup H(i)} e_{k,i} \beta_k Y_k = \sum_{h=1}^{d} \gamma_{i,h} Z_h + \beta_i Y_i + \sum_{k \in H(i)} \omega_{k,i} \left(\sum_{h=1}^{d} \gamma_{k,h} Z_h + \beta_k Y_k \right), \end{aligned} \tag{3.4.5}$$

for each $i \in A$. And we introduce the following notations:

$$\mathcal{S} = \bigcup_{\boldsymbol{y} \in \mathcal{X}} \mathcal{S}(\boldsymbol{y}), \ \mathcal{S}^* = \bigcap_{\boldsymbol{y} \in \mathcal{X}} \mathcal{S}(\boldsymbol{y}), \tag{3.4.6}$$

where $\mathcal{S}(\boldsymbol{y})$ is defined in (3.4.2). In other words, $l \in \mathcal{S}$ if the l-th factor is active in *some*

$[IP]$-feasible solutions, and $l \in \mathcal{S}^*$ if the l-th factor is active in *all* $[IP]$-optimal solutions. We then summarize the details of our general SDIS (Stochastic Dynamic Importance Sampling) algorithm as follows. Recall that $\mathcal{X}$ is the set of feasible solutions to $[IP]$.

Description of The SDIS Algorithm

Step 1) Solve the integer program, $[IP]$, given in (3.3.2).

Step 2) For each $1 \leq h \leq d$, let $f_{Z_h}(\cdot)$ be the density for the common factor Z_h. For $h \in \mathcal{S}$, sample Z_h from the mixture density $g_h(z) = [\eta_h(z|b)]^{-1} f_{Z_h}(z)$, where

$$[\eta_h(z|b)]^{-1} = p_{Z_h} \frac{I\big(z \geq a\rho^Z(b)\big)}{P\big\{Z_h \geq a\rho^Z(b)\big\}} + \big(1 - p_{Z_h}\big) \frac{I\big(z < a\rho^Z(b)\big)}{P\big\{Z_h < a\rho^Z(b)\big\}}, \tag{3.4.7}$$

defines the *local likelihood ratio* between the importance sampling and the nominal measures. Here, $a \in (0,1)$ is a constant, the mixing probability $p_{Z_h} \in (0,1)$ is positive, and the "distance to go" $\rho^Z(b)$ is defined as

$$\rho^Z(b) = \min_{j \in A, i \in j \cup H(j), \gamma_{i,h} > 0, h \in \mathcal{S}} \frac{r_j b}{v_j d e_{i,j} \gamma_{i,h}},$$

where $v_i = |i \cup H(i)|$. For $h \notin \mathcal{S}$, sample Z_h from its original density. Then, we can obtain $S_i^{(0)}, \forall i \in A \cup H(A)$.

Step 3) For each $i \in A \cup H(A)$, if $d + i \in \mathcal{S}$, given $s = S_i^{(0)}$, sample Y_i from the mixture density given by $g_i(y|s) = [\eta_i(y|b,s)]^{-1} f_{Y_i}(y)$, where

$$[\eta_i(y|b,s)]^{-1} = p_i \frac{I\big(y > a\rho_i(b,s)\big)}{P\big\{Y_i > a\rho_i(b,s)\big\}} + \big(1 - p_i\big) \frac{I\big(y \leq a\rho_i(b,s)\big)}{P\big\{Y_i \leq a\rho_i(b,s)\big\}}, \tag{3.4.8}$$

defines the local likelihood ratio for Y_i between the importance sampling and the original measures. Here $p_i \in (0,1)$ is a positive mixing probability, and the "distance to go" is defined via

$$\rho_i(b,s) = \begin{cases} \min_{\substack{j\in(\mathcal{R}(i)\cup i)\cap A \\ r_j b > S_j^{(0)}}} \frac{r_j b - S_j^{(0)}}{v_j e_{i,j} \beta_i}, & \text{if } \max_{j\in(\mathcal{R}(i)\cup i)\cap A} \frac{r_j b - S_j^{(0)}}{v_j e_{i,j}\beta_i} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

And if $d+i \notin \mathcal{S}$, sample Y_i from its original density. It is understood that $p_i = 0$, if $\rho_i(b,s) = 0$, i.e., importance sampling is *switched off* when the auxiliary process hits the corresponding initial reserve level.

Step 4) Given $Z_h, h \le d$ and $Y_i, i \in A$ sampled in Step 2) and 3), set matrix Π.

Step 5) Solve the single-period linear program $[P]$ given in (3.2.8), Update the true reserve processes according to (3.2.9), i.e., $u_i' = \check{x}_i^+ - \check{x}_i^-$ for each $i \in \mathcal{I}$.

Remark 3.4.1. Ideally, in Step 1) one should pick uniformly at random among the set of optimal solutions. Also, we can further guide the choices of the mixing probabilities p_{Z_h} and p_i by setting $p_{Z_h} = \theta$, if $h \in \mathcal{S}^*$, and setting $p_i = \theta'$, if $d+i \in \mathcal{S}^*$, and θ, θ' are some positive constants independent of b.

It is, however, necessary to assign a small (bounded away from zero) probability to the mixing probability for which the associated factor is active in some but not all $[IP]$-optimal solutions. This is because paths in which these factors are large create a non-negligible contribution to the variance of the estimator. Therefore, if $h \in \mathcal{S}\backslash\mathcal{S}^*$, we set

$p_{Z_h} = \varepsilon_Z \ll \theta$; and if $d + i \in \mathcal{S} \backslash \mathcal{S}^*$, we set $p_i = \varepsilon_Y \ll \theta'$, where both ε_Z and ε_Y are small positive constants.

The choices are consistent with the asymptotic behaviors of the system in the sense that they reflect the large deviations description of the system, as specified by Theorem 3.3.1 (i.e., we endow a large value to the mixing probability if the associated factor is active in all $[IP]$-feasible solutions, and hence must be active in all $[IP]$-optimal solutions).

Remark 3.4.2. It is necessary to simulate all the claims within a period for I_i even if some intermediate claim causes its reserve to go below zero. This is because claims are assumed to be aggregated at the end of each period. However, the SDIS scheme should be switched off as soon as that insurer fails, and one shall continue with crude Monte Carlo towards the end of that period.

The Estimator

The estimator for the probability $q(b)$ is given by

$$\begin{aligned} & \widehat{q}_{Z,Y} \\ = & \left[\prod_{h \in \mathcal{S}} \frac{f_{Z_h}(Z_h)}{g_h(Z_h|b)} \prod_{i \in A \cup H(A), d+i \in \mathcal{S}} \left(\frac{f_{Y_i}(Y_i)}{g_i\left(Y_i|b, S_i^{(0)}\right)} \right) \right] I(F_A) \\ = & \left[\prod_{h \in \mathcal{S}} \eta_h(Z_h|b) \prod_{i \in A \cup H(A), d+i \in \mathcal{S}} \left(\eta_i\left(Y_i|b, S_i^{(0)}\right) \right) \right] I(F_A), \end{aligned} \tag{3.4.9}$$

where $\eta_h(z|b)$ and $\eta_i(y|b,s)$ are defined in (3.4.7), and (3.4.8), respectively.

3.4.3 Proof of Theorem 3.3.1 and 3.4.2.

We first prove Theorem 3.4.2, which concludes our efficiency analysis of the algorithm, and then we finish the proof of Theorem 3.3.1 given in Section 3.3.

Proof. Proof of Theorem 3.4.2. Let $\widetilde{P}(\cdot)$ be the probability measure induced by the proposed importance sampling distribution, and $\widetilde{E}(\cdot)$ the associated expectation operator. Note that along a sample path generated under $\widetilde{P}$ that eventually leads to the ruin of the set A, at least one of the following cases occurs:

$$\begin{aligned} &1.\ Z_h > a\rho^Z(b), \text{ for some } h \le d, \\ &2.\ Y_j > a\rho_j\left(b, S_j^{(0)}\right), \text{for some } j \in i \cup H(i) \end{aligned} \tag{3.4.10}$$

for all $i \in A$. Otherwise, we would obtain, for some $i_0 \in A$,

$$S_{i_0}^{(0)} \le a(r_{i_0} b),$$

$$S_{i_0}^{(1)} \le r_{i_0} b\,(a + a(1-a)) < r_{i_0} b,$$

which implies F_A^c.

Therefore, when F_A happens, for each $i \in A \cup H(A)$, either (1) in 3.4.10 occurs, or (1)

doesn't occur but (2) does. In the former case, we have

$$Z_h > a\rho^Z(b), \text{ for some } h \leq d. \tag{3.4.11}$$

In the latter case, we have that for some $j \in i \cup H(i)$,

$$\begin{aligned} Y_j > a\rho_j(b, S_j^{(0)}) &= \min_{\substack{k \in (\mathcal{R}(j) \cup j) \cap A \\ r_k b > S_k^{(0)}}} \frac{a(r_k b - S_k^{(0)})}{v_k e_{j,k} \beta_j} \\ &\geq \min_{k \in (\mathcal{R}(j) \cup j) \cap A} \frac{a(1-a) r_k b}{v_k e_{j,k} \beta_j}. \end{aligned} \tag{3.4.12}$$

Now, let $\widetilde{\Omega}(\mathcal{X})$ be the subset of all the sample paths generated under $\widetilde{P}(\cdot)$ that contains large common factors or large idiosyncratic factors (in the sense of (3.4.10)) matching the active factors corresponding to *any* $[IP]$-feasible solution in $\mathcal{X}$. It follows from (**??**) and (3.4.12) that those paths must be included on the event F_A. Let the indicator $I\left((Z,Y) \in \widetilde{\Omega}(\mathcal{X})\right)$ be equal to one if the sample path encoded by the vector (Z,Y) belongs to $\widetilde{\Omega}(\mathcal{X})$, and zero otherwise. Further define

$$c = \min \left[\min_{j \in A, i \in j \cup H(j), \gamma_{i,h} > 0, h \in \mathcal{S}} \left(\frac{a r_j}{v_j d e_{i,j} \gamma_{i,h}} \right), \min_{i \in A} \min_{j \in (\mathcal{R}(i) \cup i) \cap A} \frac{a(1-a) r_j}{v_j e_{i,j} \beta_i} \right] > 0, \tag{3.4.13}$$

and let the set $\mathcal{A}_{c,\boldsymbol{y}}(b)$ be defined as

$$\mathcal{A}_{c,\boldsymbol{y}}(b) = \left\{ \bigcap_{i \in A} \left(\bigcup_{1 \leq j \leq m, \Xi_{ij} y_j \geq 1} \left\{ U_j \geq cb \right\} \right) \right\}, \tag{3.4.14}$$

for $\boldsymbol{y} \in \mathcal{X}$, where we have used the unified factor representation U introduced in §3.3 (see the paragraph before Theorem 3.3.1). Let

$$\phi\left(\underline{\boldsymbol{p}}\right) = \prod_{h \in \mathcal{S}} \frac{1}{\min\left(p_{Z_h}, 1 - p_{Z_h}\right)} \prod_{i: i \in A \cup H(A), d+i \in \mathcal{S}} \left(\frac{1}{\min\left(p_i, 1 - p_i\right)} \right).$$

Then, we have

$$\widehat{q}_{Z,Y} I\left((Z,Y) \in \widetilde{\Omega}\left(\mathcal{X}\right)\right) \leq \max_{\boldsymbol{y} \in \mathcal{X}} P\Big\{\mathcal{A}_{c,\boldsymbol{y}}(b)\Big\} \phi\left(\underline{\boldsymbol{p}}\right). \tag{3.4.15}$$

For any $\boldsymbol{y} \in \mathcal{X}$,

$$\begin{aligned} P\Big\{\mathcal{A}_{c,\boldsymbol{y}}(b)\Big\} &= P\left\{ \bigcap_{i \in A} \left(\bigcup_{1 \leq j \leq m, \Xi_{ij} y_j \geq 1} \Big\{U_j \geq cb\Big\} \right) \right\} \\ &= \left[\prod_{i: i \in A \cup H(A), d+i \in \mathcal{S}(\boldsymbol{y})} P\left\{Y_i \geq cb\right\} \prod_{h \in \mathcal{S}(\boldsymbol{y})} P\left\{Z_h \geq cb\right\} \right] \\ &\leq \widetilde{K}_1 \prod_{i: i \in A \cup H(A), d+i \in \mathcal{S}(\boldsymbol{y})} P\left\{Y_i \geq b\right\} \prod_{h \in \mathcal{S}(\boldsymbol{y})} P\left\{Z_h \geq b\right\} \end{aligned} \tag{3.4.16}$$

for some positive constant $\widetilde{K}_1$ independent of b, where $\widetilde{\alpha}$ is defined in the paragraph before Theorem 3.3.1, and $\mathcal{S}(\boldsymbol{y})$ is defined in (3.4.2).

Meanwhile, on defining

$$\underline{c}(\boldsymbol{y}) = \min_{i \in A} \left(\min_{l \in \mathcal{S}(\boldsymbol{y}), \Xi_{il} y_l \geq 1} r_i \Xi_{il} \right) > 0,$$

we have the following lower bound for $q(b)$,

$$\begin{aligned} q(b) &\geq \max_{\boldsymbol{y}\in\mathcal{X}} P\left\{\bigcap_{i\in A}\bigcup_{1\leq j\leq m,\Xi_{ij}y_j\geq 1}\left\{U_j \geq \underline{c}(\boldsymbol{y})b\right\}\right\} \\ &\geq \max_{\boldsymbol{y}\in\mathcal{X}}\left[\widetilde{K}_2 \prod_{i:i\in A\cup H(A),d+i\in\mathcal{S}(\boldsymbol{y})} P\{Y_i\geq b\}\prod_{h\in\mathcal{S}(\boldsymbol{y})} P\{Z_h\geq b\}\right], \quad (3.4.17) \end{aligned}$$

for some positive constant $\widetilde{K}_2$ independent of b.

The way we choose the mixing probabilities (see Step 2 and Step 3 in the description of the algorithm in the previous subsection) leads us to the following bound for $\phi\left(\underline{\boldsymbol{p}}\right)$,

$$0 < \phi\left(\underline{\boldsymbol{p}}\right) \leq (1/p_*)^m, \quad (3.4.18)$$

where

$$p_* \stackrel{\Delta}{=} \min\left(\min_{h\leq d}\left(p_{Z_h}, 1-p_{Z_h}\right), \min_{i\in A\cup H(A)}\left(p_i(k), 1-p_i(k)\right)\right) > 0.$$

Now combining (3.4.16), (3.4.17) and (3.4.18) we conclude that

$$\frac{\widehat{q}_{Z,Y}}{q(b)} \leq 2C,$$

where the positive constant C is defined as

$$C = \widetilde{K}_1\left(1/p_*\right)^m/\widetilde{K}_2 > 0, \quad (3.4.19)$$

therefore,

$$\widetilde{E}\left[\widehat{q}^2_{Z,Y}\right] \leq 4C^2q^2(b) = O\left[q^2(b)\right].$$

And the result follows. □

Proof. Proof of Theorem 3.3.1. From the proof of Proposition 3.4.1 we know that $\mathcal{A}_{\delta_N}(b) \supseteq \{F_A(b)\}$. And from (3.4.3), we have

$$P\{F_A(b)\} \leq P\left\{\mathcal{A}_{\delta_N}(b)\right\} \leq \kappa_2 b^{-\widetilde{\alpha}'\boldsymbol{y}^*},$$

where κ_2 is some positive constant independent of b, and $\boldsymbol{y}^*$ is an optimal solution to $[IP]$ given in (3.3.2). On the other hand, from the lower bound in (3.4.17), it's immediate that

$$P\{F_A(b)\} \geq \widetilde{K}_2 b^{-\widetilde{\alpha}'\boldsymbol{y}^*}.$$

Consequently the result follows. □

3.5 Numerical Examples

In this section we illustrate how to apply the simulation strategy described in the previous section on a network described in Figure 3.2, which consists of four companies. We assume the factors follow Pareto distributions. In particular,

$$P\{Z > z\} = \left(\frac{\theta}{\theta + z}\right)^{\alpha}, \text{ and } P\{Y_i > y\} = \left(\frac{\theta_i}{\theta_i + y}\right)^{\alpha_i}, i = 1,2,3.$$

Table 3.1: Values of model parameters.

	I_1	I_2	I_3	I_4	Z
γ	0.3	0.2	0.4	0.5	
β	1.0	1.0	1.0	1.0	
θ	100	100	100	100	100
α	2.5	3.5	4.5	2.2	2.2
p	0.9	0.9	0.9	0.9	0.9
r	4.5	5.5	1	1	

- Scenario 1: $A = \{I_4\}$
- Scenario 2: $A = \{I_3, I_4\}$

The simulation results are demonstrated in Table 3.2 and Table 3.3 below. Each estimate is based on 10^5 replications of the procedure described in the previous section. We report the mean estimate of the probability $q(b)$, standard error as a percentage of the probability estimate, as well as the estimate of the Conditional Spillover Loss at System Dislocation of the set A, standard error as a percentage of the Conditional Spillover Loss, $\mathcal{D}(A)$. For moderate values of b we compare our estimates against crude Monte Carlo in order to verify that our implementations are correct. The cost per replication of our importance sampling estimator and that of crude Monte Carlo are very comparable.

Table 3.2: Numerical results with scenarios 1.

Scenario # 1	$b = 10^4$	$b = 1.5 \times 10^4$	$b = 4 \times 10^4$	$b = 10^5$
$\hat{q}$	1.15×10^{-5}	4.66×10^{-6}	5.38×10^{-7}	7.36×10^{-8}
$s.e./\hat{q}$	1.17%	1.19%	1.20%	1.19%
$\widehat{\mathcal{D}}_0(A)$	2.60×10^{-1}	1.51×10^{-1}	4.87×10^{-2}	1.62×10^{-2}
$s.e./\widehat{\mathcal{D}}_0(A)$	3.25%	2.74%	3.11%	2.85%
$\widehat{\mathcal{D}}(A)$	2.27×10^4	3.25×10^4	9.05×10^4	2.21×10^5

Table 3.3: Numerical results with scenarios 2.

Scenario # 2	$b = 10^4$	$b = 1.5 \times 10^4$	$b = 4 \times 10^4$	$b = 10^5$
$\hat{q}$	1.09×10^{-5}	4.42×10^{-6}	5.08×10^{-7}	6.95×10^{-8}
$s.e./\hat{q}$	1.21%	1.22%	1.24%	1.22%
$\widehat{\mathcal{D}}_0(A)$	2.73×10^{-1}	1.61×10^{-1}	5.14×10^{-2}	1.75×10^{-2}
$s.e./\widehat{\mathcal{D}}_0(A)$	4.19%	3.43%	3.89%	4.70%
$\widehat{\mathcal{D}}(A)$	2.52×10^4	3.64×10^4	1.01×10^5	2.51×10^5

From the resulting tables we have a few noteworthy remarks. First of all, the relative stable ratio between the standard error and the mean of the estimates is in line with the strong efficiency of the algorithm. In other words, as b increases, it's not necessary to increase the number of replications in order to achieve the same relative accuracy. On the other hand, there is some discernible performance differential across various system configurations, for example, the relative error experiences a deterioration moving from Scenario 1 to Scenario 2. This relates to Remark 3.4.1. A quick and simple solution is to weight the factors corresponding to the "trouble-makers" substantially more than the rest of the other factors. The asymptotically optimal waiting requires explicitly computing the asymptotic conditional distributions of each factor's contribution to the rare event. This becomes even more difficult if multiple periods are considered. Another approach to find optimal mixture probabilities could be to use cross-entropy or another adaptive technique as illustrated in [19].

Chapter 4

Electrical Power Transmission Networks

4.1 Introduction

Electrical power transmission systems play a significant role in the nation's economy, and a critical aspect in the secure operation of electrical power transmission systems concerns how to keep power lines at safe temperature levels. When a power line overheats it becomes exposed to a number of risk factors. If the overheating is severe the physical/mechanical attributes of the line may be compromised, rendering it unusable. Under less severe overheating the line may sag, thus bringing it into proximity with other objects, and thereby potentially causing a contact or arc which will trip the line. If overheating is determined, the line may be protectively tripped (be taken out of service). In any of these cases the line will become unavailable, and the power flow on that line will instead

become rerouted, in a nontrivial fashion that obeys the laws of physics. This rerouting may possibly cause other lines to become overloaded. In a failure scenario of a transmission system, this sequence of events may result in a *cascade* resulting in a large-scale blackout. The Northeast U.S.-Canada blackout of 2003 produced precisely this type of event, see [51].

The temperature of a power line primarily depends on the amount of (active, or real) power flowing on that line (we refer the reader to [6], [12] or [25] for background on power engineering). However, high-voltage power lines are uninsulated and exposed to numerous exogenous factors, such as in particular wind and ambient temperature, among many. All of these factors can and do influence power line temperature. IEEE Standard 738 [33] amounts to a deterministic codification of the impact of a very large number of such factors, starting from a base model that relies on the classical heat equation. Even though this is a very thorough approach, an examination of the standard highlights the potential for mis-estimation due to erroneous, missing or variable data. The previously mentioned report [51] describes instances during the 2003 cascade where incorrect calibration of a power line leads to unexpected tripping which contributed to system instability. A somewhat more nuanced analysis of power line temperature based on the heat equation is given in [7].

In this chapter, we study a model for the temperature of a line which is based on a diffusion equation for heat conduction with an additional stochastic component used to model variations in ambient temperature. Based on the solution to this diffusion equation,

we suggest a set of guidelines for control of a power line subject to risk constraints. More specifically, we consider risk control under two failure criteria.

The first criterion is the maximum temperature along an electrical power line during a certain period of time. A guideline for the choice of inspection frequency and ways to control current (such control may result in intermittent load shedding) is provided. Firstly, one can use our results to choose an appropriate time horizon that is long enough so that it does not require impractical frequent monitoring but short enough so that risk can be effectively managed (by selecting a time horizon under which empirical statistics of ambient temperature data support the light-tailed assumption). Secondly, in the case of light-tailed temperature distributions, we explain how the theory of Gaussian random fields allows us to easily and practically estimate the risk of tripping, thereby providing a computable upper bound that can be used in optimization routines as a side constraint.

The second criterion is the average temperature along an electrical power line during a certain period of time. We assume that randomness is primarily of a spatial nature and ignore the time component. This is a reasonable model given the nature of current control practice, with various levels of control applied in the context of OPF, or Optimal Power Flow, a computation performed every fifteen minutes (or even more frequently) so as to set generator output levels and average power flows in the next time window. From this perspective, and given the length of typical lines in a transmission system (short lines may measure 50 miles, and long lines much more), local and random variations in geographically-dependent exogenous factors (such as direction and strength of wind) leading to heating or cooling of lines can and do manifest themselves, and may well be

persistent across multiple OPF intervals. We suggest several control mechanisms relying on so-called "chance constraints" to maximize delivered power while maintaining an acceptable level of risk.

Various exogenous factors are in fact the considered in IEEE 738, albeit in a simplified, deterministic manner. A comprehensive model that accounts for short-term (time dependent) exogenous variability over a large spatial domain may prove challenging; though in future work we plan to address this point.

Our particular model for stochastics of line temperature and our specific analysis are motivated by the application. The study of stochastic variants of the heat equation is a classical subject. See e.g. [49], [50] and citations therein. In terms of specific focus to power line temperature, a model different than ours is given in [20].

4.2 Formulation

We now focus on a particular power line on the time domain $[0,\tau]$. The line is modeled as a one-dimensional object parameterized by x, $0 \leq x \leq L$, used to model the spatial dependence of temperature on an exogenous stochastic factor. Let

- $I = I(t)$ denote the current of that line at time t, with the dependence on t highlighted so as to allow for control actions.

- $T(x,t)$ the temperature at x at time t.

The *heat equation* states:

$$\frac{\partial T(x,t)}{\partial t} = \kappa \frac{\partial^2 T(x,t)}{\partial x^2} + \alpha I^2(t) - \nu(T(x,t) - T^{ext}(x,t)), \tag{4.2.1}$$

where $\kappa \geq 0$, $\alpha \geq 0$ and $\nu \geq 0$ are (line dependent) constants, and $T^{ext}(x,t)$ is the ambient temperature at (x,t). In order to account for stochasticity, we model $T^{ext}(x,t) = G(h(x))$ where $h(x)$ denotes a random variable at x, with distribution that is either known or can be estimated (in what follows boldface will be used to denote uncertain quantities) and $G(\cdot) > 0$. We thus obtain

$$\frac{\partial T(x,t)}{\partial t} = \kappa \frac{\partial^2 T(x,t)}{\partial x^2} + \alpha I^2(t) - \nu(T(x,t) - G(h(x)). \tag{4.2.2}$$

We will further assume $\kappa = 0$. This is consistent with the use of the heat equation in [33], [7]; it is justified by noting that propagation in the time domain is much faster than in the spatial domain. [We will acount for the randomness of exogenous conditions in the spatial domain in an average, or aggregated manner made precise below]. We therefore obtain:

$$\frac{\partial T(x,t)}{\partial t} = \alpha I^2 - \nu(T(x,t) - G(h(x)). \tag{4.2.3}$$

4.3 Criterion 1: Maximum Temperature

If we know the boundary condition $T(x,0) = l(x)$, where $l(x)$ is some continuous function, then for any $\kappa \geq 0$, making use of Feynman-Kac formula gives us

$$\begin{aligned} T(x,t) = & e^{-\nu t} \int_{-\infty}^{\infty} l(x)\phi(y;x,2\kappa t)\,dx \\ & + \frac{\alpha I^2}{\nu}\left(1 - e^{-\nu t}\right) \\ & + \int_0^t \nu e^{-\nu s} \int_{-\infty}^{\infty} G(h(x))\phi(y;x,2\kappa s)\,dx\,ds, \end{aligned} \tag{4.3.1}$$

where

$$\phi(y;x,2\kappa t) = \frac{\exp(-(y-x)^2/(4\kappa t))}{(4\pi\kappa t)^{1/2}}.$$

For the simplified version when $\kappa = 0$, we have that

$$T(x,t) = (1 - e^{-\nu t})(G(h(x)) + \frac{\alpha I^2}{\nu}) + e^{-\nu t} l(x). \tag{4.3.2}$$

Note that in both solutions (4.3.1) and (4.3.2) we have that x is not bounded. However, we might only interested in x inside some compact interval $[0,L]$, so we look only at $x \in [0,L]$, for some $L > 0$. While in our technical development in this chapter we will only consider the simplified case in which $\kappa = 0$, some of our insights can be easily seen to also apply to the case $\kappa > 0$. In particular, as we shall briefly discuss the fact that the tail of the temperature distribution is relatively insensitive to the current in the case of long-tailed temperature distributions is a feature that remains valid in the case of $\kappa > 0$.

If the internal temperature of a line is too high, it (or its connectors) might suffer physical damage. In a more likely scenario, the line may sag and contact a foreign object such as a tree. In either case the line will trip and be taken out of operation. Such an event (or set of events) could trigger a cascading failure of the underlying grid. Therefore, we want to limit the tripping probability which, in order to capture the effects described previously, we define as follows:

$$P\{\max_{x\in[0,L],t\in[0,\tau]} T(x,t) > k\}, \tag{4.3.3}$$

where k is some critical temperature. In order to obtain robust insights in the level of generality that we are aiming at, we shall perform an asymptotic analysis in the context of k large. We believe that this asymptotic environment is meaningful in the applications considered because, presumably, failure events happen mostly when temperatures reach a threshold.

4.3.1 Current Control

Notice that $T(x,t)$ is a function of the current, I. To limit the tripping probability (4.3.3) to a certain level, I should satisfy some constraint, which gives us a guidance how we should control the current in the line.

We set the length of each time window equals to τ. At the beginning of each time window, operators inspect the line and input updated κ, α, ν and $l(x)$ at that time, and

modify I such that the tripping probability in this window is limited to a certain level to guarantee the line is safe before the next inspection.

4.3.2 Study of Tripping Probability

Now we want to learn when and where the line temperature reaches its maximum. To do this, we further assume that G is a monotone strictly increasing and continuous function which we will discuss later, $h(x) \sim N(\mu(x), \sigma^2(x))$ follows Gaussian distribution with mean and variance depending on x. For $x \neq y$ we have that $h(x)$ and $h(y)$ are jointly distributed Gaussian random variables with some correlation structure which satisfies mild assumptions required to make the $h(\cdot)$ a Hölder continuous function with probability one (i.e. every realization of the function $h(\cdot)$ satisfies $|h(x) - h(x)| \leq K|x-y|^{\rho}$ for a deterministic constant $\rho \in (0,1)$ and some constant $K > 0$).

Combining (4.2.3) and (4.3.2) we get

$$\frac{\partial T(x,t)}{\partial t} = e^{-\nu t}[\nu G(h(x)) + \alpha I^2 - \nu l(x)]. \tag{4.3.4}$$

Therefore,

$$\max_{x\in[0,L],t\in[0,\tau]} T(x,t) = \begin{cases} \max\limits_{x\in[0,L]} T(x,\tau), & \text{if } q(x,I) > 0 \\ \max\limits_{x\in[0,L]} T(x,0), & \text{if } q(x,I) \leq 0 \end{cases}$$

where $q(x,I) = \nu G(h(x)) + \alpha I^2 - \nu l(x)$.

If $q(x,I) \leq 0$, then $\max_{x\in[0,L],t\in[0,\tau]} T(x,t) = \max_{x\in[0,L]} l(x)$. The line trips if $\max_{x\in[0,L]} l(x) > k$, which means that this line is already tripped when operators inspect it. We believe that this case is not of practical interest for our purposes since we aim to find controls that prevent potential future tripping. We also might be able to modify I such that $q(x,I) > 0$ with large probability. Thus when the critical temperature k is large, tripping is most likely to happen at time τ. Therefore, we are more interested in the probability:

$$P\{\max_{x\in[0,L]} T(x,\tau) > k\}. \tag{4.3.5}$$

For the rest of chapter, when we say tripping probability, we refer to probability (4.3.5) by default unless otherwise stated.

We conclude that the tripping probability behaves very differently in the case when $\max_{x\in[0,L]} G(h(x))$ is long-tailed from the case when it is light-tailed. The definitions of long-tailed and light-tailed distributions have been given in Section 1.2.1. We claim that:

Theorem 4.3.1. *When $k \to \infty$, if $\max_{x\in[0,L]} G(h(x))$ is long-tailed, then the tripping probability is independent of I; the tripping probability depends on I otherwise.*

Proof.

$$\begin{aligned} &P\{\max_{x\in[0,L]} T(x,\tau) > k\} \\ &= P\{\max_{x\in[0,L]} (1-e^{-\nu\tau})(G(h(x)) + \frac{\alpha I^2}{\nu}) + e^{-\nu\tau} l(x) > k\} \qquad (4.3.6) \\ &\sim P\{\max_{x\in[0,L]} (1-e^{-\nu\tau}) G(h(x)) > k\}, \text{ as } k \to \infty, \qquad (4.3.7) \end{aligned}$$

where, in this chapter, "$\sim$" represents that those two probabilities are of the same order. The last asymptotic equivalence is obtained by the definition of long-tailed distribution and the boundedness of $l(x)$. Notice that (4.3.7) doesn't involve I, so the tripping probability is (asymptotically) independent of current. On the other hand, (4.3.6) does involve I. □

We believe that this simple result provides useful and important guidance. If the variations of the maximum ambient temperature are long-tailed, the capability to control tripping probability through controlling current is limited. To make sure our control is effective, we need to make sure the variations of maximum ambient temperature are light-tailed, which required that the length of time window τ should not be too large. Given historical data of ambient temperature, we could choose a proper τ. Although we have used the simplified model with $\kappa = 0$, note that if $\max_{x\in[0,L]} G(h(x))$ is long-tailed, then the tail behavior of $a + \max_{x\in[0,L]} G(h(x))$ coincides with that of $\max_{x\in[0,L]} G(h(x))$. In equation (4.3.2), I^2 only appears as in a term added to $G(h(x))$, which is the only random term, so we conclude that controlling I^2 in the long-tailed setting might not be effective to substantially reduce risk (measured in terms of tail behavior). By the same token, in equation (4.3.1), I^2 also only appears as in a term added to the only term that involves randomness, which is a more involved expression of $G(h(x))$. Thus the same insights indicated in the theorem prevail in the case when $\kappa > 0$ although the mathematical proof will be somewhat more complicated.

Define

$$g(x) = h(x) - \mu(x),$$

$$\bar{\mu} = \max_{x\in[0,L]} \mu(x),$$

$$\bar{\sigma}^2 = \max_{x\in[0,L]} E[g^2(x)] = \max_{x\in[0,L]} E[h^2(x)],$$

$$\bar{\lambda} = E[\max_{x\in[0,L]} g(x)],$$

$$\bar{l} = \max_{x\in[0,L]} l(x),$$

$$\tilde{l} = \min_{x\in[0,L]} l(x).$$

We assume that $\mu(x)$, $\sigma(x)$ and $l(x)$ are continuous, then $\bar{\mu} < \infty$, $\bar{\sigma}^2 < \infty$ and $\bar{l} < \infty$. If we further assume that the Gaussian random fields $h(x)$ is almost surely continuous on $[0,L]$, then $\bar{\lambda} < \infty$. The next result allows us to obtain asymptotic approximations (at least in logarithmic scale). These approximations, although somewhat coarse, allow to quantify the tripping probability in great generality (including both light-tailed and long-tailed distributions).

Theorem 4.3.2. *If $G^{-1}(\cdot)$ satisfies $G^{-1}(x) \to \infty$ as $x \to \infty$ and that for any $c > 0$,*

$$\lim_{k\to\infty} \frac{G^{-1}(k-c)}{G^{-1}(k)} = 1, \tag{4.3.8}$$

then as $k \to \infty$,

$$\log P\{\max_{x\in[0,L]} T(x,\tau) > k\} \sim -\frac{(G^{-1}(\frac{k}{1-e^{-\nu\tau}}))^2}{2\sigma^2(x^*)}, \tag{4.3.9}$$

where $x^* = \arg\max_{x\in[0,L]} \sigma(x) \in [0,L]$.

Proof. We start with deriving an upper bound.

$$\begin{aligned}
&P\{\max_{x\in[0,L]} T(x,\tau) > k\} \\
&= P\{\max_{x\in[0,L]} (1-e^{-\nu\tau})(G(h(x)) + \frac{\alpha I^2}{\nu}) + e^{-\nu\tau} l(x) > k\} \\
&\le P\{\max_{x\in[0,L]} (1-e^{-\nu\tau})(G(g(x)+\bar{\mu}) + \frac{\alpha I^2}{\nu}) + e^{-\nu\tau}\bar{l} > k\} \\
&= P\{\max_{x\in[0,L]} (G(g(x)+\bar{\mu}) > \frac{k-e^{-\nu\tau}\bar{l}}{1-e^{-\nu\tau}} - \frac{\alpha I^2}{\nu}\} \\
&= P\Big\{\max_{x\in[0,L]} g(x) > G^{-1}(\frac{k-e^{-\nu\tau}\bar{l}}{1-e^{-\nu\tau}} - \frac{\alpha I^2}{\nu}) - \bar{\mu}\Big\} \\
&\le e^{-\frac{(y_u-\bar{\mu}-\bar{\lambda})^2}{2\sigma^2(x^*)}},
\end{aligned} \tag{4.3.10}$$

where $y_u = G^{-1}(\frac{k-e^{-\nu\tau}\bar{l}}{1-e^{-\nu\tau}} - \frac{\alpha I^2}{\nu})$ and the last inequality is obtained by Borell-TIS inequality, see [4], page 50.

Now we consider the lower bound.

$$
\begin{aligned}
&P\{\max_{x\in[0,L]} T(x,\tau) > k\} \\
&\geq P\Big\{(1-e^{-\nu\tau})(G(g(x^*)+\mu(x^*))+\frac{\alpha I^2}{\nu})+e^{-\nu\tau}\tilde{l} > k\Big\} \\
&= P\Big\{g(x^*) > G^{-1}(\frac{k-e^{-\nu\tau}\tilde{l}}{1-e^{-\nu\tau}}-\frac{\alpha I^2}{\nu})-\mu(x^*)\Big\} \\
&\geq (\frac{\sigma(x^*)}{\sqrt{2\pi}(y_l-\mu(x^*))}-\frac{\sigma^3(x^*)}{\sqrt{2\pi}(y_l-\mu(x^*))^3})e^{-\frac{(y_l-\mu(x^*))^2}{2\sigma^2(x^*)}}, \qquad (4.3.11)
\end{aligned}
$$

where $y_l = G^{-1}(\frac{k-e^{-\nu\tau}\tilde{l}}{1-e^{-\nu\tau}}-\frac{\alpha I^2}{\nu})$.

Combining (4.3.10) and (4.3.11), and taking logarithm, we have

$$
\begin{aligned}
&-\frac{(y_u-\bar{\mu}-\bar{\lambda})^2}{2\sigma^2(x^*)} \\
&\geq \log P\{\max_{x\in[0,L]} T(x,\tau) > k\} \\
&\geq \log(\frac{\sigma(x^*)}{\sqrt{2\pi}(y_l-\mu(x^*))}-\frac{\sigma^3(x^*)}{\sqrt{2\pi}(y_l-\mu(x^*))^3}) \\
&-\frac{(y_l-\mu(x^*))^2}{2\sigma^2(x^*)}.
\end{aligned}
$$

Since $G^{-1}(\cdot)$ satisfies (4.3.8), then

$$
\begin{aligned}
&\lim_{k\to\infty}\frac{-\frac{(y_u-\bar{\mu}-\bar{\lambda})^2}{2\sigma^2(x^*)}}{-(G^{-1}(\frac{k}{1-e^{-\nu\tau}}))^2} \\
&= \lim_{k\to\infty}\frac{\log(\frac{\sigma(x^*)}{\sqrt{2\pi}(y_l-\mu(x^*))}-\frac{\sigma^3(x^*)}{\sqrt{2\pi}(y_l-\mu(x^*))^3})-\frac{(y_l-\mu(x^*))^2}{2\sigma^2(x^*)}}{-(G^{-1}(\frac{k}{1-e^{-\nu\tau}}))^2} \\
&= \frac{1}{2\sigma^2(x^*)}.
\end{aligned}
$$

Therefore, as $k \to \infty$,

$$\log P\{\max_{x\in[0,L]} T(x,\tau) > k\} \sim -\frac{(G^{-1}(\frac{k}{1-e^{-\nu\tau}}))^2}{2\sigma^2(x^*)}.$$

□

After the discussion regarding to the general function $G(\cdot)$, next we will discuss two specific functions: $G(x) = e^x$ and $G(x) = x$. In the former case, $\max_{x\in[0,L]} G(h(x))$ is long-tailed, and in the latter case, it is light-tailed.

4.3.3 Examples

We would like to discuss a long-tailed example and a light-tailed example. Before that, we need to introduce some definitions and a lemma on the maximum of Gaussian non-centered field, see [10], pages 190-193.

Definition 4.3.3. *Let the collection $\alpha_1,\ldots,\alpha_k$ of positive number is be given, as well as the collection $l_1,\ldots,l_k$ of positive integers such that $\sum_{i=1}^k l_i = n$. We set $l_0 = 0$. These two collections will be called a* structure. *For any vector $\boldsymbol{x} = (x_1,\ldots,x_n)^T$ its* structural module *is defined by*

$$|\boldsymbol{x}|_\alpha = \sum_{i=1}^{k} \Big(\sum_{j=E(i-1)+1}^{E(i)} x_j^2 \Big)^{\frac{\alpha_i}{2}},$$

where $E(i) = \sum_{j=0}^{i} l_j, j = 1,\ldots,k$.

Definition 4.3.4. *Let an $\alpha-$structure be given on R^n. We say that $h(\boldsymbol{x})$, $\boldsymbol{x} \in A \subset R^n$, has a local $(\alpha, D_{\boldsymbol{x}})$-stationary structure, or $h(\boldsymbol{x})$ is locally $(\alpha, D_{\boldsymbol{x}})$-stationary, if for any $\varepsilon > 0$*

there exists a positive $\delta(\varepsilon)$ *such that for any* $\boldsymbol{s} \in \mathrm{A}$ *one can find a non-degenerate matrix* $D_{\boldsymbol{s}}$ *such that the covariance function* $r(\boldsymbol{x}_1,\boldsymbol{x}_2)$ *of* $h(\boldsymbol{x})$ *satisfies*

$$1-(1+\varepsilon)|D_{\boldsymbol{s}}(\boldsymbol{x}_1,\boldsymbol{x}_2)|_\alpha \le r(\boldsymbol{x}_1,\boldsymbol{x}_2) \le 1-(1-\varepsilon)|D_{\boldsymbol{s}}(\boldsymbol{x}_1,\boldsymbol{x}_2)|_\alpha,$$

provided $\| \boldsymbol{x}_1-\boldsymbol{s} \| < \delta(\varepsilon)$ *and* $\| \boldsymbol{x}_2-\boldsymbol{s} \| < \delta(\varepsilon)$.

Lemma 4.3.5. *Let* $h(\boldsymbol{x})$, $\boldsymbol{x} \in R^n$, *be a Gaussian locally* $(\alpha, D_{\boldsymbol{x}})$*-stationary field, with some* $\alpha > 0$ *and continuous matrix function* $D_{\boldsymbol{x}}$). *Let* $\mathcal{M} \subset R^n$ *be a smooth p-dimensional compact set,* $0 < p \le n$. *Let the expectation* $m(\boldsymbol{x}) = \boldsymbol{E}h(\boldsymbol{x})$ *be continuous on* $\mathcal{M}$ *and attains its maximum on* $\mathcal{M}$ *at the only point* $\boldsymbol{x}_0$, *with*

$$m(\boldsymbol{x}) = m(\boldsymbol{x}_0) - (\boldsymbol{x}-\boldsymbol{x}_0)B(\boldsymbol{x}-\boldsymbol{x}_0)' + \mathrm{O}(\| \boldsymbol{x}-\boldsymbol{x}_0 \|^{2+\beta}),$$

as $\boldsymbol{x} \to \boldsymbol{x}_0$, *for some* $\beta > 0$ *and positive matrix B. Then,*

$$P\{\sup_{\boldsymbol{x}\in\mathcal{M}} h(\boldsymbol{x}) > u\} = bu^{\theta}\Psi(u-m(\boldsymbol{x}_0))(1+o(1)), \ \ as \ \ u \to \infty, \tag{4.3.12}$$

where b and θ *are constants, and*

$$\Psi(u) = \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-x^2/2} dx.$$

For the rest of this section, we assume that $h(x)$ satisfies the conditions required in the above theorem.

Long-tailed Example: $G(h(x)) = e^{h(x)}$

1) $\max_{x\in[0,L]} e^{h(x)}$ is long-tailed

Proof. Making use the above lemma, we have

$$\begin{aligned} & P\{\max_{x\in[0,L]} e^{h(x)} > k+c\} \\ & = P\{\max_{x\in[0,L]} h(x) > \log(k+c)\} \\ & \sim b[\log(k+c)]^{\theta}\Psi(\log(k+c)-m(x_0)), \end{aligned}$$

where $m(x_0)$ is defined as the same as in the theorem.

Since

$$\begin{aligned} & \Psi(\log(k+c)-m(x_0)) \\ & \sim \frac{1}{\log(k+c)-m(x_0)}\frac{1}{\sqrt{2\pi}}e^{-(\log(k+c)-m(x_0))^2/2}, \end{aligned}$$

as $k\to\infty$, it is easy to see that

$$\begin{aligned} & P\{\max_{x\in[0,L]} e^{h(x)} > k+c\} \\ & \sim b(\log k)^{\theta}\Psi(\log k - m(x_0)) \\ & \sim P\{\max_{x\in[0,L]} e^{h(x)} > k\}. \end{aligned}$$

□

2) Asymptotic behavior of tripping probability

In this case, $G^{-1}(x) = \log x$. Therefore,

$$\log P\{\max_{x\in[0,L]} T(x,\tau) > k\} \sim -\frac{(\log(\frac{k}{1-e^{-\nu\tau}}))^2}{2\sigma^2(x^*)}, \text{ as } k \to \infty.$$

Light-tailed Example: $G(h(x)) = h(x)$

1) $\max_{x\in[0,L]} h(x)$ is light-tailed

Proof. Since

$$P\{\max_{x\in[0,L]} h(x) > k\}$$
$$\sim ck^{\theta-1}\frac{1}{\sqrt{2\pi}}e^{-k^2/2},$$

for any $\lambda > 0$,

$$e^{\lambda k}P\{\max_{x\in[0,L]} h(x) > k\} < \infty, \text{ as } k \to \infty.$$

□

2) Asymptotic behavior of tripping probability

In this case, $G^{-1}(x) = x$. Therefore,

$$\log P\{\max_{x\in[0,L]} T(x,\tau) > k\} \sim -\frac{(\frac{k}{1-e^{-\nu\tau}})^2}{2\sigma^2(x^*)}, \text{ as } k \to \infty.$$

3) A useful upper bound for current control

As we said earlier, in this case, we could control the risk effectively through current control. The reason is we can find an upper bound of the tripping probability.

Define

$$\hat{\mu} = \max_{x\in[0,L]} E[h(x)] < \infty,$$

$$\hat{\sigma}^2 = \max_{x\in[0,L]} Var[h(x)] < \infty,$$

According to the following Borell-Sudakov-Tsirelson inequality in [10], page 192, we can find an upper bound of the tripping probability.

Lemma 4.3.6. *If there exists some a such that*

$$P\{\max_{x\in[0,L]} h(x) - E[h(x)] \geq a\} \leq \frac{1}{2}$$

Then, for all b,

$$P\{\max_{x\in[0,L]} h(x) > b\} \leq 2\Psi(\frac{b-\hat{\mu}-a}{\hat{\sigma}}). \tag{4.3.13}$$

In fact, it is not difficult to prove that such a exists using Borell-TIS inequality because we are assuming that the underlying Gaussian process is continuous. In practice, such a could be estimated through Monte Carlo simulation.

Theorem 4.3.7. *There exists a, such that*

$$P\{\max_{x\in[0,L]} T(x,\tau) > k\} \leq 2\Psi(\frac{\hat{k}-\hat{\mu}-a}{\hat{\sigma}}), \tag{4.3.14}$$

where $\hat{k} = \frac{k-(1-e^{-\nu\tau})\frac{\alpha I^2}{\nu}-e^{-\nu\tau}\bar{l}}{1-e^{-\nu\tau}}$.

Proof. Since

$$T(x,\tau) = (1-e^{-\nu\tau})(h(x)+\frac{\alpha I^2}{\nu})+e^{-\nu\tau}l(x).$$

Then

$$\begin{aligned}
&P\{\max_{x\in[0,L]} T(x,\tau) > k\} \\
&= P\{\max_{x\in[0,L]} (1-e^{-\nu\tau})(h(x)+\frac{\alpha I^2}{\nu})+e^{-\nu\tau}l(x) > k\} \\
&\leq P\Big\{(1-e^{-\nu\tau})\max_{x\in[0,L]} h(x)+(1-e^{-\nu\tau})\frac{\alpha I^2}{\nu}+e^{-\nu\tau}\bar{l} > k\Big\} \\
&= P\Big\{\max_{x\in[0,L]} h(x) > \frac{k-(1-e^{-\nu\tau})\frac{\alpha I^2}{\nu}-e^{-\nu\tau}\bar{l}}{1-e^{-\nu\tau}}\Big\} \\
&\leq 2\Psi(\frac{\hat{k}-\hat{\mu}-a}{\hat{\sigma}}).
\end{aligned}$$

□

Note that the upper bound (4.3.14) is increasing in I. Suppose we need to keep the risk of tripping at a certain level η, or in other words, we require that $P(\max_{x\in[0,L]} T(x,\tau) > k) \leq \eta$. We only need to let $2\Psi(\frac{\hat{k}-\hat{\mu}-a}{\hat{\sigma}}) \leq \eta$, which gives us an upper bound of current. This provides a mechanism for periodic control of the line. Note that current magnitude

is directly proportional to (active, or real) power flow magnitude; in other words the risk exposure that we compute as inputed by a given current level can be used as a guide for load shedding in an emergency situation.

Suppose that in the presence of bound (4.3.14), we are interested in maximizing an increasing function of I, subject to a constraint that guarantees that the tripping probability is less than η. We then can easily compute an optimal value of I_* satisfying

$$I_*^2 = \frac{\nu}{\alpha(1-e^{-\nu\tau})}\Big(k - e^{-\nu\tau}\bar{l} - (1-e^{-\nu\tau})(\hat{\sigma}\Psi^{-1}(\eta/2) + \hat{\mu} + a)\Big). \tag{4.3.15}$$

Next, we would look into a numerical example based on this conclusion.

4.3.4 Numerical Light-Tailed Example: $G(h(x)) = h(x)$

Basically, we need to do three things.

1. Define $h(x), x \in [0, L]$ such that it is both Hölder continuous and locally stationary.
2. Estimate a and calculate I_*^2.
3. Set the current equal to I_*, simulate $h(x)$ up to time τ, estimate the tripping probability.

Definition of $h(x), x \in [0, L]$

We assume that $h(x)$ follows an Ornstein-Uhlenbeck process, i.e.

$$dh(x) = (c_0 - h(x))dx + \sigma B(x), \tag{4.3.16}$$

where c_0 and $\sigma > 0$, and $B(x)$ denotes the standard Brownian motion. Then $E[h(x)] = c_0, \forall x \in [0,L]$, and $cov(h(x),h(y)) = e^{-|x-y|}/2, \forall x,y \in [0,L]$.

It is known that the Ornstein-Uhlenbeck process has almost surely Hölder continuous sample path, and it is obviously that $h(x)$ is also locally stationary. We have

$$\hat{\mu} = c_0, \quad \hat{\sigma}^2 = \sigma^2/2.$$

Note that $E[h(x)] = c_0$ does not satisfy the conditions in Lemma 4.3.5. However, $h(x)$ can be considered as a centered field. Therefore, we can make use of Theorem 7.1 in [45], page 108, to find a similar equation as (4.3.12).

Estimation of a and I_*^2

We want to find a, such that

$$P\{\max_{x\in[0,L]} h(x) - c_0 \geq a\} \leq \frac{1}{2}$$

Since the solution to the stochastic differential equation (4.3.16) is

$$h(x) = \frac{\sigma e^{-x}}{\sqrt{2}} B(e^{2x} - 1) + e^{-x}h(0) + (1 - e^{-x})c_0,$$

we could estimate a using Monte Carlo simulation. Then I_* could be calculated using formula (4.3.15).

Estimation of the Tripping Probability by Simulating $h(x)$ up to Time τ

For simplicity, we assume $l(x) = \bar{l}, \forall x \in [0,L]$, then the tripping probability

$$\begin{aligned}
&P\{\max_{x\in[0,L]} T(x,\tau) > k\} \\
&= P\{\max_{x\in[0,L]} (1-e^{-\nu\tau})(h(x)+\frac{\alpha I_*^2}{\nu}) + e^{-\nu\tau}\bar{l} > k\} \\
&= P\Big\{\max_{x\in[0,L]} h(x) > \hat{k}\Big\},
\end{aligned}$$

where $\hat{k} = \frac{k-(1-e^{-\nu\tau})\frac{\alpha I_*^2}{\nu} - e^{-\nu\tau}\bar{l}}{1-e^{-\nu\tau}}$.

We want to efficiently compute the tail probability for the suprema of Gaussian random fields. We apply algorithm 7.3 in [1]. Before we explain this algorithm, let us first introduce a definition.

Definition 4.3.8. *We call* $\tilde{X} = \{x_1, x_2, \ldots, x_M\} \subset [0,L]$ *a* θ-regular discretization *of* $[0,L]$ *if,*

$$\min_{i\neq j} |x_i - x_j| \geq \theta, \quad \sup_{x\in[0,L]} \min_i |x_i - x| \leq 2\theta.$$

To decide the number of replications n, we could choose $n = O(\varepsilon^{-2}\delta^{-1})$ to achieve ε relative error with probability at least $1-\delta$.

Let $H = (h(x_1), h(x_2), \ldots, h(x_M))$, we define distribution Q as

$$Q(H \in B) = \sum_{i=1}^{M} \frac{1}{M} P\{H \in B | h(x_i) > \hat{k} - \frac{1}{\hat{k}}\}. \tag{4.3.17}$$

Given a number of replications n and an $\varepsilon/\hat{k}$-regular discretization, the simulation algorithm is as follows:

Step 1: Sample $H^{(1)}, H^{(2)}, \ldots, H^{(n)}$ i.i.d copies of H with distribution Q given by (4.3.17).

Step 2: Compute and output

$$\hat{L}_n = \frac{1}{n}\sum_{i=1}^{n} \tilde{L}_{\hat{k}}^{(i)},$$

where

$$\tilde{L}_{\hat{k}}^{(i)} = \frac{M \times P\{h(x_1) > \hat{k} - 1/\hat{k}\}}{\sum_{j=1}^{M} \mathbf{1}(h(x_j)^{(i)} > \hat{k} - 1/\hat{k})} \mathbf{1}(\max_{1 \le i \le M} h(x_j)^{(i)} > \hat{k}).$$

Results

We set parameters as follows:

- Stochastic differential equation (4.2.3)

$$\alpha = 1, \nu = 1, L = 1, \tau = 1/4, \bar{l} = 70.$$

- $h(x)$

$$c_0 = 70, \sigma = 10.$$

- Simulation

$$k = 100, \eta = 0.05, \varepsilon = 0.05, \delta = 0.05, n = 8000.$$

In other words, We want the tripping probability

$$P\{\max_{x\in[0,1]} T(x,1/4) > 100\} \leq 0.05.$$

The simulation result is:

$$I_*^2 = 99.97, P\{\max_{x\in[0,1]} T(x,1/4) > 100\} = 7.4\times 10^{-6} < 0.05.$$

Therefore, if we set the risk level $\eta = 0.05$, and let the current $I_*^2 = 99.97$, we could make sure that the tripping probability is below this level. The simulation results confirm the validity of the current control methodology we develop.

Analysis of Results

Assuming we use the current control methodology in (4.3.15), then one can verify that

$$P\{\max_{x\in[0,L]} T(x,\tau) > k\} \leq \sqrt{\sigma^2/2}\Psi^{-1}(\eta/2) + c_0 + a.$$

Note that the upper bound above is independent of k. Thus, we can keep the tripping probability bounded by a certain level no matter what the value of k is.

Another observation is that the tripping probability is much smaller than the risk level. In terms of risk control, this result shows that our control is very effective. But if we would like to simultaneously maximize the current, a tighter upper bound than (4.3.14) should be developed.

4.4 Criterion 2: Average Temperature

4.4.1 Line limits in the context of Optimal Power Flow (OPF)

Here we briefly review how OPF is used in current power engineering practice and how our work here ultimately relates to guiding certain upper bound parameters in the evaluation of OPF equations. For a review see e.g. [39].

The OPF (or "economic dispatch") computation is now performed as often as every fifteen minutes, and is used to set the output of generators (in the next fifteen-minute window) so as to minimize cost while meeting estimates of demands in a safe manner, that is to say, without exceeding flow limits on the transmission lines in a systematic manner. Real-time, random fluctuations in demands are handled through automatic adjustment of generator output. Such automatic, short-term actions may cause power lines to exceed their flow limits in moderate form and for brief periods of time; however such overages are tolerated unless sustained over longer periods of time. See [6], [12].

Using the linearized "DC" model of power flows, which is common in the case of OPF,

the optimization problem that is solved in power engineering practice is of the form:

$$\min \sum_{i \in G} w_i(P_i) \tag{4.4.1a}$$

$$\text{s.t. } B\theta - P = 0 \tag{4.4.1b}$$

$$|y_{ij}(\theta_i - \theta_j)| \leq u_{ij} \tag{4.4.1c}$$

$$P_i = \text{-(load at } i\text{)}, \text{ if } i \notin G \tag{4.4.1d}$$

$$\theta \in \mathbb{R}^n. \tag{4.4.1e}$$

In this formulation we consider a transmission system with n buses in total. We denote by G the set of buses attached to generators. The vector $P \in \mathbb{R}^n$ indicates net output at each bus. Thus when $i \in G$, P_i is the output for the generator at i. When bus i is a load bus (i.e., $i \notin G$) the quantity P_i is fixed in equation (4.4.1d) to minus the load at i, possibly equal to zero. Matrix B is the so-called network susceptance matrix and for a bus i, θ_i is the phase angle at bus i, a variable in the computation. Equation (4.4.1b) represents the (linearized) relationship between net power output levels and phase angles; under the DC model the power flow from i to j on line ij equals

$$y_{ij}(\theta_i - \theta_j)$$

where $y_{ij} > 0$ is the susceptance of line ij, a physical parameter. Finally, the quantity u_{ij} is the power transmission limit for line ij. Thus, equation (4.4.1c) states that the flow on ij, in absolute value does not exceed the limit on that line. Finally, the objective to

be minimized is the overall cost of generation; at generator bus i the function $w_i(\cdot)$ is a convex quadratic on the output P_i. Other constraints may be present, e.g. upper and lower bounds on the generator outputs P_i.

An important part of formulation (4.4.1), of interest to us in the context of this section, is the determination of the line limits u_{ij}. If these are set to large values then lines may become overheated and at risk of tripping; on the other hand if the u_{ij} are set too low the cost of operation (i.e. the objective value (4.4.1a) at optimality) may become too large or we may even obtain an infeasible optimization problem. In current power engineering practice a simple procedure is employed:

(a) Before each OPF period, if a line u_{ij} is deemed to have become overheated (in the previous time period) or at risk of becoming overheated in the next time period, the u_{ij} quantity is decreased to a "safe" value.

(b) The adjustment in (a) is typically done by reducing u_{ij} to one of a (very) small set of precomputed values.

We stress that (a) and (b) are performed one line at a time. Effectively, one is setting the quantity u_{ij} to act as a proxy for the maximum power flow that line ij can safely sustain in the next time window; but of course, the computation in (4.4.1) may set the actual power flow to a lower value than u_{ij}.

At this point we note that there are several procedures that one could employ in order to incorporate into OPF the stochasticity of line temperatures. Let us denote by $B_{ij} = B_{ij}(y_{ij}(\theta_i - \theta_j))$ the (random) event that in the next time window line ij will reach a

critical temperature given its flow value $y_{ij}(\theta_i - \theta_j)$. This event depends on exogenous events; we consider a specific model below. Using $\mathcal{P}$ to denote probability, one can then consider the following variations of (4.4.1):

$$\min \sum_{i \in G} w_i(P_i) \tag{4.4.2a}$$

$$\text{s.t.constraints (4.4.1b) - (4.4.1e)}, \tag{4.4.2b}$$

$$\mathcal{P}\left(\bigcup_{ij} B_{ij}\right) \leq \varepsilon, \tag{4.4.2c}$$

and

$$\min \sum_{i \in G} w_i(P_i) \tag{4.4.3a}$$

$$\text{s.t.constraints (4.4.1b) - (4.4.1e)}, \tag{4.4.3b}$$

$$\mathcal{P}\left(B_{ij}\right) \leq \varepsilon, \quad \forall\, ij. \tag{4.4.3c}$$

Either (4.4.2) or (4.4.3) are risk-aware variants of OPF; with (4.4.2) likely much more conservative. Either one of these optimization problems is, likely, quite challenging due to the nature of the stochasticity (i.e. the "chance constraint" (4.4.3c) may be quite non-trivial). And in fact either model amounts to a significant departure from current practice.

In this section we follow a different approach that adjusts process (a)-(b). Our goal will simply be to refine the current practice of resetting the u_{ij} with an intelligent compu-

tation that is risk-aware. In other words the OPF problem to be solved will still be of the form (4.4.1), albeit with carefully constructed quantities u_{ij}.

Our approach will involve the computation of a general solution to a stochastic heat equation, with an explicit spatial dependence on stochastic effects, and suggest several control mechanisms relying on so-called "chance constraints" to maximize delivered power while maintaining an acceptable level of risk.

Our particular model for stochastics of line temperature and our specific analysis are motivated by the OPF application. The study of stochastic variants of the heat equation is a classical subject. See e.g. [49], [50] and citations therein.

4.4.2 Formulation

Integrating on both sides of (4.2.3) with respect to x, and dividing by L, we have

$$\begin{aligned}\frac{1}{L}\int_0^L \frac{\partial T(x,t)}{\partial t}dx \;&=\; \alpha I^2(t) - \frac{\nu}{L}\int_0^L T(x,t)dx \\ &+\frac{\nu}{L}\int_0^L G(h(x))dx. \end{aligned} \tag{4.4.4}$$

Denoting by $H(t)$ average internal temperature along the line at time t, by R the average ambient temperature along the line, i.e.,

$$H(t) \triangleq \frac{1}{L}\int_0^L T(x,t)dx, \quad R \triangleq \frac{1}{L}\int_0^L G(h(x))dx,$$

we therefore have

$$\frac{dH(t)}{dt} = \frac{d}{dt}\frac{1}{L}\int_0^L T(x,t)dx = \frac{1}{L}\int_0^L \frac{\partial T(x,t)}{\partial t}dx.$$

Then (4.4.4) becomes:

$$\frac{dH(t)}{dt} = \alpha I^2(t) - \nu H(t) + \nu R, \tag{4.4.5}$$

with solution

$$\begin{aligned} H(t) &= \int_0^t e^{-\nu(t-s)}(\alpha I^2(s) + \nu R)ds + Ce^{-\nu t} \\ &= \int_0^t e^{-\nu(t-s)}\alpha I^2(s)ds + R(1-e^{-\nu t}) + Ce^{-\nu t}, \end{aligned} \tag{4.4.6}$$

where

$$C = H(0) = \frac{1}{L}\int_0^L T(x,0)dx.$$

We note that the quantity R is not observed – however we can assume that its distribution can be estimated. We are interested in control schemes that vary $I(t)$ in response to observed conditions. As criterion for stability, we will enforce the chance-constraint [22], [42]

$$P\left\{\max_{t\in[0,\tau]} H(t) > k\right\} \leq \varepsilon, \tag{4.4.7}$$

where $k > 0$ is a given limit and $\varepsilon > 0$ is small.

4.4.3 Constant $I(t)$, $t \in [0,\tau]$

The case where $I(t)$ is constant in the time window of interest is of special interest because of its simplicity. We are interested in computing those values $\breve{I}$ such that setting $I(t) = \breve{I}$ for $0 \leq t \leq \tau$ satisfies (4.4.7). From the closed-form solution above we obtain

$$H(t) = ((\alpha/\nu)\breve{I}^2 + R)(1 - e^{-\nu t}) + Ce^{-\nu t}. \tag{4.4.8}$$

Now, let us assume that $G(x) \geq 0$, which implies that $R \geq 0$.

It follows from (4.4.8) together with (4.4.5) that

$$\begin{aligned} \frac{dH(t)}{dt} &= \alpha I^2(t) - \nu((\alpha/\nu)\breve{I}^2 + R)(1 - e^{-\nu t}) + vCe^{-\nu t} + \nu R, \\ &= \alpha\breve{I}^2 e^{-\nu t} + vRe^{-\nu t} + vCe^{-\nu t}. \end{aligned}$$

In other words, $H'(t) > 0$ if one assumes that $\breve{I}^2 + C \geq 0$, which in turn will always hold if $H(0) = C \geq 0$. Consequently, if one assumes that $C = H(0) < k$, it follows that (4.4.7) is equivalent to

$$P\{H(\tau) > k\} \leq \varepsilon.$$

Using (4.4.8) this implies

$$\breve{I}^2 \leq \frac{\nu}{\alpha}\frac{k - Ce^{-\nu\tau} - r_\varepsilon(1 - e^{-\nu\tau})}{1 - e^{-\nu\tau}}, \tag{4.4.9}$$

where r_ε is the ε-quantile of R (i.e. r_ε is the smallest x such that $P\{R > x\} \leq \varepsilon$). As

discussed above we assume that the distribution of R is known, and consequently the bound in (4.4.9) is computable.

For future use, we use $L(\tau, k)$ to denote the right-hand side of expression (4.4.9).

Now consider an entire grid where several lines are assumed to be thermally stressed. We can then compute the upper bound on the current for each such line that is implied by computation in (4.4.9). These values can then be used the OPF computation (generator dispatch) at time $t = 0$.

Numerical examples

We next perform some numerical experiments designed to understand the behavior of the threshold expression (4.4.9) in a range of numerical values which are informed by actual IEEE Standards.

- Calibration of parameters

  Expression (4.4.9) contains two line-dependent parameters, α and ν, which need to be calibrated. Also the constant $C = H(0)$ needs to be estimated. For the purpose of computing the line-dependent parameters, we use some formulas provided in the IEEE Standard 738 [33] applied to an ACSR (aluminum conductor steel reinforced) line. We include these parameters here because they give a sense of the physical measurements involved in ultimately evaluating the various parameters in the underlying model, and because we ultimately wish to show that the numerical values that we obtain make practical sense.

The formulas for α and v require that we introduce some notation and parameters arising in the standard:

ρ_f: Density of air $(1.029\ kg/m^3)$.

V_w: Speed of air stream at conductor $(0.61\ m/s)$.

μ_f: Dynamic viscosity of air $(0.0000204\ Pa \cdot s)$.

k_f: Thermal conductivity of air $(0.0295\ W/(m - {}^oC))$.

K_{angle}: Wind direction factor (90^o).

mC_{ACSR}: Total heat capacity of conductor (ACSR in our case).

q_c: Convected heat loss rate per unit length per unit temp.

q_r: Radiated heat loss rate per unit length per unit temp.

q_s: Heat gain rate from the sun per unit temp.

R_a: AC resistance of conductor at average temperature $(5.34 \times 10^{-4}\ \Omega/m)$.

mC_{ACSR}: Total heat capacity of conductor (ACSR in our case).

D: Conductor diameter $(.01\ m)$.

T^{ext}: External temperature.

Thus (as per the standard) we have that

$$\text{low wind speed per temp.} : q_{c1} = \left(1.01 + 0.0372\left(\frac{D\rho_f V_w}{\mu_f}\right)^{0.52}\right) k_f K_{angle},$$

$$\text{high wind speed per temp.} : q_{c2} = 0.0119\left(\frac{D\rho_f V_w}{\mu_f}\right)^{0.6} k_f K_{angle},$$

$$q_c = \max(q_{c1}, q_{c_2}).$$

The IEEE Standard 738 [33] provides equations for q_r and q_s, but these two quantities typically approximately offset each other (CITATION) and hence $q_r \approx q_s$, thereby obtaining that

$$\alpha = \frac{R_a}{mC_{ACSR}}, \tag{4.4.10}$$

and

$$\nu = \frac{q_r + q_c - q_s}{mC_{ACSR}} \approx \frac{\max(q_{c1}, q_{c_2})}{mC_{ACSR}}. \tag{4.4.11}$$

We use equations (4.4.10) and (4.4.11) to estimate $\alpha = 3.99 \times 10^{-6}$ and $\nu = 2.96 \times 10^{-4}$.

In addition, we assume that the initial average external temperature is $C := H(0)$ = $70^o C$, we perform our analysis in a time window of τ = $900s$ (approximately 15 minutes). Finally, we shall set the line temperature limit to $k = 110^o C$.

Figure 4.1 indicates how $\sqrt{L(\tau, k)}$ (namely, the square root of the right hand side in (4.4.9)) changes with r_ε.

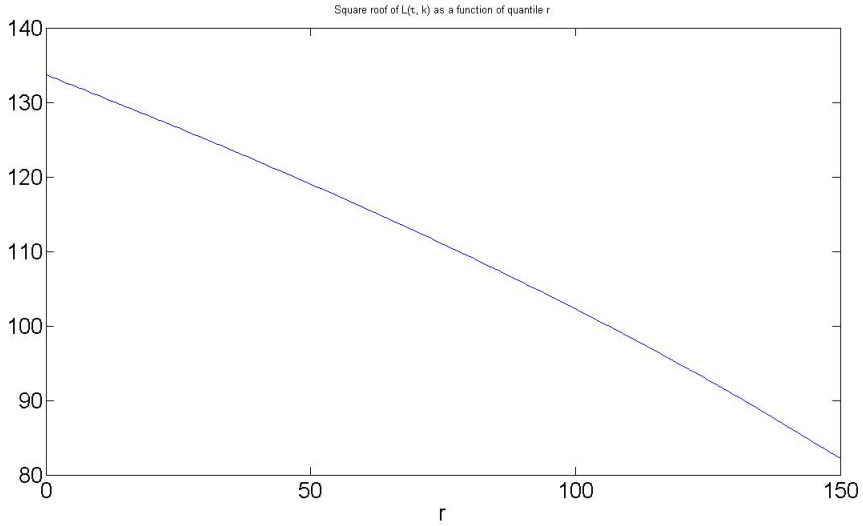


Figure 4.1: Upper Bound - $\sqrt{L(\tau,k)}$ as a function of r_ε

- Applying control to a power grid

  Next we describe experiments where the methodology presented above for setting upper bounds on power flows is used in a system-wide fashion for OPF calculations. In the computations described in this section we will rely on formulation (4.4.1) given above.

  In the tests performed below, we use varying choices of the parameter of r_ε to set line limits as indicated above. More precisely, we assume that $u_{i,j}$ in (4.4.1) is set equal to

$$u_{ij} = \sqrt{L(\tau,k)}.$$

  Of course, for line ij, the evaluation of $\sqrt{L(\tau,k)}$ depends on a wide range of line-dependent characteristics as indicated in Section 4.4.3. For the purpose of this

exercise we assume that all the lines have the same characteristics and therefore u_{ij} is constant across lines.

We shall study the resulting total "carrying capacity" for the network as measured by the optimal value of the objective function in (4.4.1), and we will choose below two different objective functions (linear and quadratic). Moreover, a question of central importance for any procedure that sets line limits in protective fashion is precisely how much such procedure may constrain (or over-constrain) operations. So, an important part of the outcome of our study is to understand when the problem (4.4.1) becomes unfeasible as we change a parameter which reflects underlying stochastic fluctuations, and we will choose r_ε to perform this parametric study. As we will see, we might obtain a critical threshold r_ε^* such that if $r_\varepsilon > r_\varepsilon^*$ then (4.4.1) becomes unfeasible, whereas if $r_\varepsilon \leq r_\varepsilon^*$, (4.4.1) is feasible.

For our testing we used the 9-bus system in [55], with three buses attached to generators (buses 1, 2, and 3). We used the following procedure:

- For $r_\varepsilon \leq r_\varepsilon^*$ (that is, if (4.4.1) is feasible) we study how the optimal value of (4.4.1) changes as we vary r_ε.
- For $r_\varepsilon > r_\varepsilon^*$ (that is, if (4.4.1) is unfeasible), then at least one of the constrains represented in (4.4.1c) is violated. In this case, we introduce for each line, ij, a deficit variable $z_{ij} \geq 0$, and replace each of the constrains (4.4.1c) by the constrains

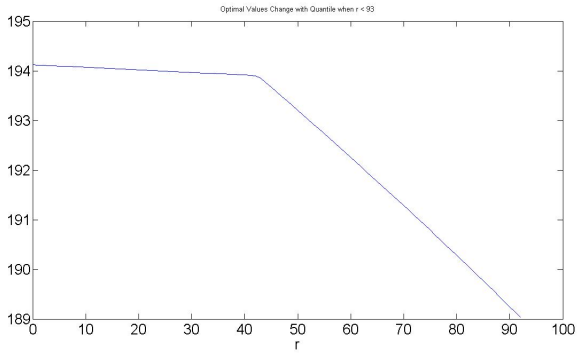
$$-u_{ij} - z_{ij} \leq y_{ij}(\theta_i - \theta_j) \leq u_{ij} + z_{ij},$$

we also replace the objective function (4.4.1) by minimizing $\sum_{ij} z_{ij}$. We then study how the optimal sum of deficits changes as we vary $r_\varepsilon > r_\varepsilon^*$.

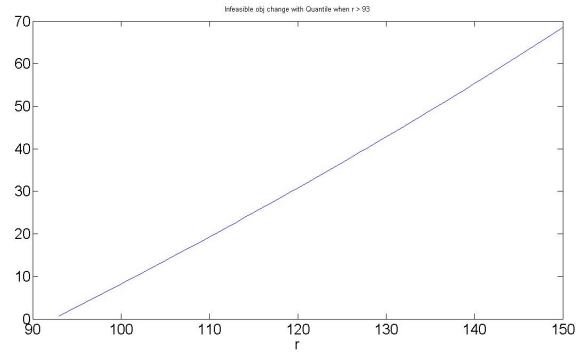
We performed two tests with two different objective functions:

1. Linear function: $\max 0.5P_1 + 0.6P_2 + 0.7P_3$.

   The goal of this experiment is to study how a weighted sum of generation adapts to a change in risk aversion.



(a) Linear Obj Function: Feasible, When $r < 93$



(b) Linear Obj Function: Infeasible, When $r > 93$

Figure 4.2: Adding upper bound to a power grid DC problem - linear objective function

Figure 4.2 shows that when $r_\varepsilon > r_\varepsilon^* = 93$, the LP problem becomes infeasible (as a result of power flow limits), and thereafter the infeasibility increases monotonically because u_{ij} is monotone in r_ε.

2. Quadratic function: $\min 0.2P_1^2 + 0.3P_2^2 + 0.5P_3^2$.

   This objective function is of the typical type used in an OPF computation. See Figure 4.3.

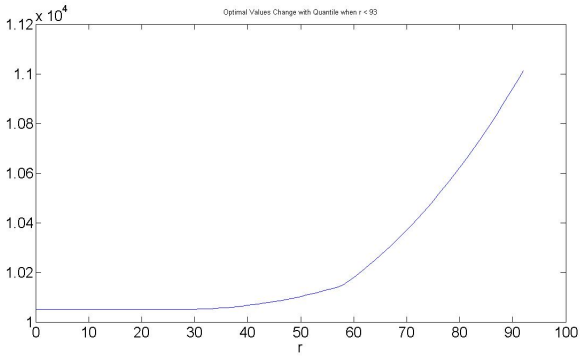


Figure 4.3: Quadratic Obj function: Feasible, When $r < 93$

4.4.4 Adaptive Control

As an extension to the above analysis above, we consider a setup where it is known at time $t = 0$ that uncertainty will be resolved at later point in the time window $[0, \tau]$. We wish to make use of this fact by designing a better line limit than that obtained through e.g. (4.4.9); this improvement will come about by incorporating recourse actions that will be deployed when uncertainty is resolved. We present our specific approach below; first we will motivate our approach through some examples. These examples deviate from current power engineering practice, but only to a small degree.

We also note that the problem studied in this section falls under the broad category of stochastic programming with recourse, which has received much attention, with many elegant and powerful results that are progressively finding more applications in power

engineering. More to the point, we will be solving optimization problems subject to chance constraints. See, e.g. [27], [31], [32].

Consider, first, the OPF setting, with generators output levels computed so as to cover the time interval $[0,\tau]$. As discussed above, if some line is known to be thermally stressed at time $t = 0$, special attention should (and will) be devoted to setting its limit in the OPF computation. Using (4.4.9) provides a risk-aware methodology for doing so.

Suppose, however, that at time $t = 0$ it is known that the numerical value of W will be known at $t = \tau/2$, and, further, that a second generator dispatch (possibly in limited form) can be performed at $t = \tau/2$. Then additional efficiencies might be attained by resetting the line limit at $t = \tau/2$. We thus have a two-stage problem: set a line limit to be applied over $[0,\tau/2]$ and another limit over $[\tau/2,\tau]$ contingent on the value of W observed at $\tau/2$. The combined action must be done in risk-aware fashion (i.e. guaranteeing that line temperature will exceed its critical value with probability $\leq \varepsilon$) while maximizing some weighted average of the two line limits. In addition, we would also want to guarantee that with probability 1, some higher critical temperature is not exceeded.

A second justification refers to a method already found, to some degree, in OPF practice. This is the idea that a given OPF computation should be performed in such a way to also incorporate information from the time window after the current one, i.e. the time window $[\tau,2\tau]$. In our setting, suppose that at time $t = 0$ the line is thermally stressed, with uncertainty, but that the random variable W will become known at time $t = \tau$. The computation used to set the line limit over $[0,\tau]$ should be performed aware of the fact that the limit should be set again over $[\tau,2\tau]$ because there is hysteresis in the thermal process.

The computation should also be done so as to take advantage of the fact that uncertainty will disappear at time $t = \tau$.

A third justification involves the Unit Commitment problem, which covers a much larger span of time (e.g. 24 hours). However, if thermal stresses are expected over this longer time period, line limits will have to be accordingly reduced. Again we have a setting where the resolution of uncertainty at some intermediate point permits a better line limit setting (while relying on a second generator output computation).

For simplicity, in what follows we will assume that uncertainty is resolved at time $t = \tau/2$. The specific details of our model are as follows: Define the random variable

$$W \triangleq R \cdot (1 - e^{-\nu\tau}),$$

and to simplify the analysis we assume that W is discretely distributed, $P\{W = w_i\} = p_i, i = 1, 2, \ldots, m$ with a known distribution (i.e., the distribution of R is known). We are interested in a control scheme with the following characteristics

(1) At time $t = 0$, we compute values I_1, and $I_{2,i}$ for $i = 1, 2, \ldots, m$. These values are used as follows:

(2) In the time window $[0, \tau/2]$ the upper bound on current is set to the value I_1.

(3) At time $\tau/2$, we observe the value of R and thus of W. Assuming $W = w_i$ then the upper bound on current is set, in the interval $[\tau/2, \tau]$, to the value $I_{2,i}$.

The values I_1, and $I_{2,i}$, $1 \leq i \leq m$ are computed according to the following criteria

(a) $P\{H(\tau) > k\} < \varepsilon$.

(b) $I_1 \leq L(\tau/2, k)$.

(c) Let k be the critical line temperature. For a given value $k^* > k$ we guarantee, with probability 1, that line temperature will not exceed k^*.

(d) Where $F : \mathbb{R}_+^2 \to \mathbb{R}_+$ is coordinate-wise monotonic increasing, we want to maximize, in expectation

$$\sum_{i=1}^{m} F(I_1, I_{2,i}) p_i$$

We now discuss these modeling features. First, our control scheme broadly falls within the class of "two-stage stochastic optimization" problems. In the initial time window we use a (computed) current value I_1. At the midway point we observe prevailing conditions and we adjust our policy as per criterion (3). Note that as per the analysis in the previous section, for any choice of the values I_1, $I_{2,i}$ we will generally have that $H(t)$ is monotonely increasing or decreasing for $t \in [0, \tau/2]$ and also for $t \in [\tau/2, \tau]$.

Thus, our control scheme may possibly result in a realization where e.g. $H(\tau/2) > k$. However, (a) guarantees that even if this were to be the case, the line temperature will have reached a safe value by time τ. And of course (b) helps ensure that the probability of this event is small.

Regarding item (d), many examples are reasonable, keeping in mind that a central goal is to attain "high" values for I_1 or I_2 without incurring risk. As discussed at the end of Section 4.4.3 and in the Introduction, a choice of safe value for the current parameter

I imputes an effective value for the line limit and thus is a central parameter to be used in computing generator dispatch (OPF). For example, we could set $F(I_1, I_2) = \pi_1 I_1 + \pi_2 I_2$ where $\pi_1, \pi_2 \geq 0$. Or, $F(I_1, I_2) = \pi_1 I_1^2 + \pi_2 I_2^2$. Below we discuss several cases.

We will now cast the choice of the values I_1, $I_{2,i}$ as an optimization problem. To do so, suppose that $W = w_i$. Then, using (4.4.6),

$$H(\tau) = v_1 I_1^2 \; + \; v_2 I_2^2(i) \; + \; w_i \; + \; Ce^{-\nu\tau}, \tag{4.4.12}$$

where

$$v_1 \; \triangleq \; \int_0^{\tau/2} e^{-\nu(\tau - s)} \alpha \, ds \tag{4.4.13}$$

and

$$v_2 \; \triangleq \; \int_{\tau/2}^{\tau} e^{-\nu(\tau - s)} \alpha \, ds. \tag{4.4.14}$$

Define

$$z_1 \triangleq v_1 I_1^2, \quad z_2(i) \triangleq v_2 I_2^2(i), \;\; 1 \leq i \leq n, \tag{4.4.15}$$

and, for $1 \leq i \leq n$

$$\bar{k}_i \triangleq k - Ce^{-\nu\tau} - w_i,$$

Using this notation we have that, when $W = w_i$

$$H(\tau) > k \quad \text{is equivalent to:} \quad z_1 + z_2(i) > \bar{k}_i.$$

Likewise, if we define $u_i \triangleq k^* - Ce^{-\nu\tau} - w_i$, if $W = w_i$, then $H(\tau) > k^*$ is equivalent to

$z_1 + z_2(i) > u_i$. Finally, let us define

$$\tilde{F}(z_1, z_2) \triangleq F(\sqrt{z_1/v_1}, \sqrt{z_2(i)/v_2}), \tag{4.4.16}$$

which is simply recasting function F in terms of the z variables. It follows that we can write our optimal control problem as

$$\begin{aligned}
\mathcal{P}_1: \quad & \max_{z_1, z_2} \sum_{i=1}^{m} \tilde{F}(z_1, z_2(i)) p_i, \\
& \text{s.t.} \sum_{i=1}^{m} \mathcal{I}\{z_1 + z_2(i) > \bar{k}_i\} p_i \ \leq \ \varepsilon && (4.4.17) \\
& z_1 \ \leq \ v_1 L(\tau/2, k) && (4.4.18) \\
& z_1 + z_2(i) \ \leq \ u_i, \quad \forall i, && (4.4.19) \\
& z_1 \geq 0, \quad z_2(i) \geq 0, \forall i. && (4.4.20)
\end{aligned}$$

In (4.4.17), $\mathcal{I}$ is the indicator function; the square roots in the objective, and the bound (4.4.18), arise from our definition of the z variables. Constraint (4.4.19) guarantees that the line temperature at time τ not exceed k^*, with probability 1. Of course, this constraint may render the problem above infeasible – however assuming that $H(0)$ is "safe" the problem will be feasible (if necessary by setting $z_1 = z_2(i) = 0$ for all i) assuming realistic R. As per the above discussions we have that $u_i > \bar{k}_i$.

Remark 4.4.1. Let z_1^*, $z_2^*(i)$ $(1 \leq i \leq m)$ be an optimal solution to problem $\mathcal{P}_1$. Then, for

each $1 \leq i \leq m$

$$z_1^* + z_2^*(i) \;=\; \bar{k}_i \text{ or } u_i.$$

Proof. Suppose that for some i, $z_1^* + z_2^*(i) < \bar{k}_i$. Then increasing $z_2^*(i)$ maintains feasibility, and the monotonicity assumption on F implies that the objective improves. The same reasoning applies if $u_i > z_1^* + z_2^*(i) > \bar{k}_i$. □

Using this observation we can simplify the optimization problem. Define, for $1 \leq i \leq m$ a binary variable y_i such that

$$y_i = \begin{cases} 0 \text{when } z_1 + z_2(i) = \bar{k}_i \\ 1 \text{when } z_1 + z_2(i) = u_i \end{cases} \tag{4.4.21}$$

Then the above optimization problem can be recast as:

$$\begin{aligned} \mathcal{P}_2: \quad & \max_{z_1, y} \sum_{i=1}^{m} \tilde{F}(z_1, \bar{k}_i - z_1) p_i (1 - y_i) + \tilde{F}(z_1, u_i - z_i) p_i y_i \\ & \text{s.t.} \sum_{i=1}^{m} p_i y_i \;\leq\; \varepsilon \\ & 0 \;\leq\; z_1 \;\leq\; \min\{v_1 L(\tau/2, k)\,, \min_i\{u_i\}\} \\ & y_i = 0 \text{ or } 1, \text{ all } i. \end{aligned} \tag{4.4.22}$$

Problem $\mathcal{P}_2$ is a nonlinear, binary optimization problem. We are interested in methodologies and special cases where a near optimal solution can be obtained in practicable form. Here we will present two results:

- We will describe a general approximation method that should prove effective when the parameter m is moderately large, say $m < 10^4$.

- We also describe a provably good approximation algorithm for the case where m is very large, which however only applies in a special case of the function F.

As an aside, the issue of the magnitude of m concerns several practical questions, primarily with regards with how accurate a representation of the random variable R can be constructed in "real time". Adequate sensorization of power lines should help in this regard, however there is a larger issue of how data uncertainty can arise in this context (e.g., the spatial distribution of exogenous temperatures over a small time window).

General approach for moderately large m

Suppose that in problem $\mathcal{P}_2$ we were to fix the variable z_1 to a value $\hat{z}_1$ satisfying (4.4.22). The remaining part of problem $\mathcal{P}_2$ has the following general structure:

$$\begin{aligned} \mathcal{P}_2(\hat{z}_1): \quad & \max \sum_{i=1}^{m} \tilde{f}_i(\hat{z}_1) y_i \\ & \text{s.t.} \sum_{i=1}^{m} p_i y_i \ \leq \ \varepsilon, \\ & y_i = 0 \text{ or } 1, \text{ all } i. \end{aligned}$$

Problem $\mathcal{P}_2(\hat{z}_1)$ is a linear (binary) knapsack problem. Knapsack problems are NP-hard – however in this case we are dealing with values of m that are not very large. In fact, it is fair to say that knapsack problems are the easiest of the NP-hard problems, and, more

to the point, commercial mixed-integer program solvers can handle such problems with ease even for large m.

These observations suggest the following (grid- or mesh-parameterization) approach:

(1) Enumerate equally spaced values of $\hat{z}_1$ between the two bounds in (4.4.22).

(2) For each enumerated value, solve $\mathcal{P}_2(\hat{z}_1)$.

While suffering from an enumerative component, this approach does have the attribute of handling any objective function F in the definition of our problem.

A separate issue regarding the small n case concerns the robustness of the computed answers with respect to e.g. the (necessarily estimated) parameters p_i and w_i. Using a small n has the effect of accumulating more probability mass into fewer values, with a resulting increase in numerical sensitivity (to the choices for the p_i and w_i).

Very large m

Suppose now that m is very large. As stated above we expect that even in this case a good mixed-integer programming solver should be able to solve the problems $\mathcal{P}_2(\hat{z}_1)$. Nevertheless, we would like to discuss a case where a solution methodology with sound theoretical foundations is available.

Recall the formula (4.4.12) for $H(\tau)$ as well as (4.4.15), and note that $z_1 + z_2(i)$ appears in $H(\tau)$ and thus, in the chance constraint (4.4.17). Consider the special case where

$$F(I_1, I_2) \;=\; v_1^2 I_1^2 \;+\; v_2^2 I_2^2 \tag{4.4.23}$$

for all I_1, I_2. Below we will discuss the implications of this selection of $F(I_1, I_2)$. Using (4.4.23), it will follow that using (4.4.16),

$$\tilde{F}(z_1, z_2) = z_1 + z_2, \quad \text{for all } z_1, z_2.$$

Thus, problem $\mathcal{P}_2$ can be equivalently restated as:

$$\begin{aligned}
\mathcal{P}_3 : &\max \sum_{i=1}^{m} \bar{k}_i p_i (1 - y_i) + u_i p_i y_i \\
&\text{s.t.} \sum_{i=1}^{m} p_i y_i \;\leq\; \varepsilon \\
&\quad 0 \;\leq\; z_1 \;\leq\; \min\{v_1 L(\tau/2, k)\,, \min_i\{u_i\}\} \qquad (4.4.24) \\
&\quad y_i = 0 \text{ or } 1, \text{ all } i.
\end{aligned}$$

We can see that constraint (4.4.24) is not needed. In short, $\mathcal{P}_3$ can be rewritten in the form:

$$\begin{aligned}
\mathcal{P}_3' : &\max \sum_{i=1}^{m} f_i y_i \\
&\text{s.t.} \sum_{i=1}^{m} p_i y_i \;\leq\; \varepsilon \\
&\quad y_i = 0 \text{ or } 1, \text{ all } i,
\end{aligned}$$

for appropriate quantities $f_i > 0$ and $q_i > 0$; this problem constitutes a standard, linear $0-1$ knapsack problem.

In comparing this approach to that used for the small n case, we can see that we have simplified the problem (no enumeration over the $\hat{z}_1$ values). Of course, we do have to solve the possibly large knapsack problem $\mathcal{P}_3$. As we discussed before, this should prove routine (and very fast) even for n in the thousands. However, even though the knapsack problem is NP-hard, there is a very large literature regarding rigorous algorithms for obtaining nearly-optimal solutions to a knapsack problem, within arbitrary precision, in an efficient manner. For a broad survey, see [53]. Also see [14] and references therein. An alternative procedure would rely on branch-and-cut to directly tackle the underlying chance-constrained problem; see [38]. Another alternative would rely on affine controls, see e.g. [2].

Finally we comment on (4.4.23). We would argue that this is a "reasonable" functional form for $F(I_1, I_2)$ in that it amounts to a weighted sum. Of course the weights are not flexibly chosen. Nevertheless, note (see (4.4.13), (4.4.14)) that $v_2 > v_1$. Thus, (4.4.23) places more emphasis on what happens in the time interval $[\tau/2, \tau]$. We would argue that this is a reasonable approach, in the sense that we focus in the later time interval, where, coincidentally, we are able to make more precise decisions since randomness has been resolved.

Numerical example

We implement a numerical example for moderately large $m = 50$ using the methodology in Section 4.4.4. In the test,

- The distribution of W is shown in Figure 4.4.
- $k^* = k + 10 = 120^o C$.
- The number of enumerated values of $\hat{z}_1$ is 1000.
- The remaining parameters are as in Section 4.4.3. This implies that $L(0,\tau) \approx 110.80$.

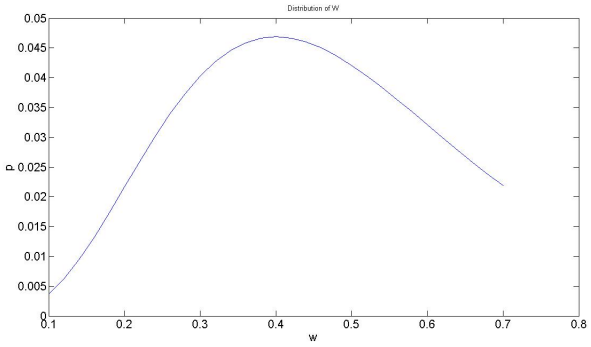


Figure 4.4: Distribution of W

In the experiment we performed, $F(I_1, I_2) = I_1 + \sum_{i=1}^{m} p_i I_{2,i}$, i.e. we seek to maximize the average line limit (averaged over first and second stages and across all realizations). Solution of the resulting ensemble of knapsack problems yields $I_1 = 120.98$, and the values of $I_{2,i}, i = 1,2,\ldots,m$, as shown in Figure 4.5. The entire computation (1000 knapsacks) required approximately one CPU second on a current workstation.

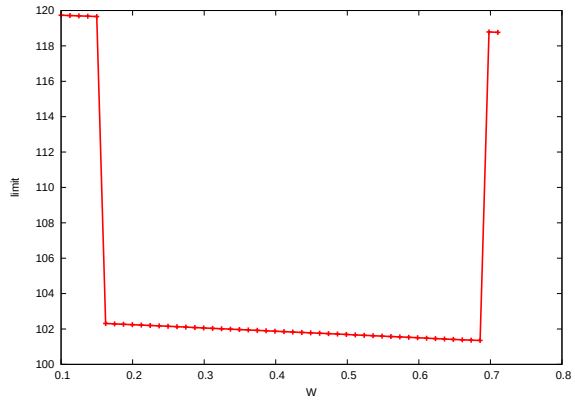


Figure 4.5: Values of I_2 for first 2-stage experiment

The figure displays a compelling behavior: when W is large then I_2 is small (because it should be) and when W is small then I_2 is large (because it can be). Moreover, this intelligent second-stage behavior is leveraged in the first stage by setting I_1 quite large; in fact large compared to $L(0,\tau)$. Hence the main benefit of the two-stage approach is that during the first stage the line limit is approximately 9% higher; moreover the average of first- and second-stage limits is approximately 111.83, again larger than $L(0,\tau)$, and all of this under the much more conservative additional constraint that guarantees that line temperature will not exceed the value k^* with probability one.

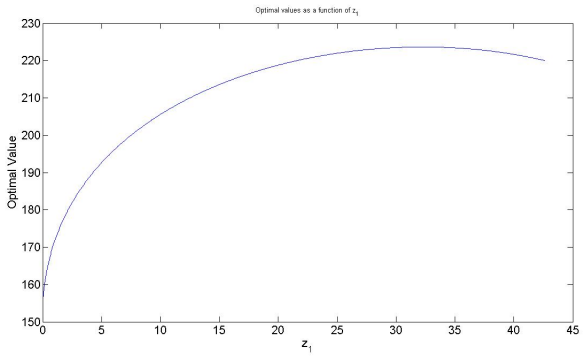


Figure 4.6: Optimal Values change with z_1

4.5 Conclusion

In this chapter we have presented a rigorous analysis of a stochastic variant of the heat equation which leads to control guidelines under different failure criteria. If we are interested in the probability that the maximum temperature over a fixed amount of time along a single line exceeds a critical threshold, then a guidance for the choice of frequency and the way to modify current so that the failure probability is kept at a safe level is provided. If we are interested in the average temperature, then accurate chance-constrained policies for setting line limits, obtained as the (approximate) solution to an appropriate optimization problem are provided. In this case, we also consider a two-stage stochastic optimization problem that leads to an adaptive control mechanism, where observations made at an intermediate point in a time window provide a means to adjust line limits.

Bibliography

[1] 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.

[2] Chen, X., Sim, M., Sun, P., and Zhang, J. 2008. A Linear Decision-Based Approximation Approach to Stochastic Programming, Operations Research 56 344-357.

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

[4] Adler, R. J., J. E. Taylor. 2010. Random Fields and Geometry. New York: Springer Monographs in Mathematics.

[5] Amini, H., R. Cont, A.Minca. 2011. Stress testing the resilience of financial networks. International Journal of Theoretical and Applied Finance 14.

[6] Andersson, G. 2004. Modelling and Analysis of Electric Power Systems. Power Systems Laboratory, ETH Zürich.

[7] Anghel, M., K.A. Werley, and A.E. Motter. 2007. Stochastic model for power grid dynamics.Proceedings of the 40^{th} Hawaii Int. Conf. on System Sciences.

[8] 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.

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

[10] Balakrishnan, N., I.A. Ibragimov and V.B. Nevzorov. 2001. Asymptotic Methods in Probability and Statistics with Applications. Birkhauser Boston.

[11] Bandt, O.D., P. Hartmann. 2000. Systemic risk: A survey. European Central Bank, Frankfurt, Germany.

[12] Bergen, A.R., V. Vittal. 1999. Power Systems Analysis. Prentice-Hall.

[13] Bertsimas, D., J. N. Tsitsiklis. 1997. Introduction to Linear Optimization. Athena Scientific, Belmont, Massachusetts.

[14] Bienstock, D., and B. McClosky. 2012. Tightening simple mixed-integer sets with guaranteed bounds. Math. Programming 133 337-363.

[15] Binswanger, S. Asmussen K., B. Hojgaard. 1997. Rare events simulation for heavy-tailed distributions. Bernoulli 6 303-322.

[16] Blanchet, J., H. Lam. 2011. State-dependent importance sampling for rare-event simulation: An overview and recent advances. Submitted to Surveys in Operations Research and Management Sciences.

[17] Blanchet, J., J. C. 2008. State-dependent importance sampling for regularly varying random walks. Advances in Applied Probability 40 1104-1128.

[18] 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.

[19] Blanchet, J., Y. Shi. 2011. Efficient Rare Event Simulation for Heavy-tailed Systems via Cross Entropy. S. Jain and R. R. Creasey and J. Himmelspach and K. P. White and M. Fu, eds., Proceedings of the 2011 Winter Simulation Conference. IEEE Press.

[20] Bosman, J., J. Nair, B. Zwart. 2014. On the probability of current and temperature overloading in power grids: a large deviations approach. ACM Sigmetrics Performance Evaluation Review 42 33-85.

[21] 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.

[22] Charnes, A., W. Cooper, G. Symonds. 1958. Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. Management Science 4 235-263.

[23] Cont, R., A. Moussa. 2009. Too interconnected to fail: contagion and systemic risk in financial networks. Financial Engineering Report 2009-04, Columbia University.

[24] Cont, R., Q.Moussa, Edson Bastos e Santos. 2010. The Brazilian financial system: network structure and systemic risk analysis. Working Paper.

[25] D., J., GLover, M.S. Sarma, T. J. Overbye. 2012. Power System Analysis and Design. CENGAGE Learning.

[26] 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.

[27] Dentcheva, D. 2009. Optimization Models with Probabilistic Constraints. Lectures on Stochastic Optimization, Shapiro, A., Dentcheva, D., and Ruszczynski, A. eds., MPS-SIAM Seres in Optimization. SIAM, Phil., 9 , Ch. 4.

[28] Dupuis, P., H. Wang. 2007. Subsolutions of an Isaacs equation and efficient schemes of importance sampling. Mathematics of Operations Research 32 723-757.

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

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

[31] Henrion, R. 2004. Introduction to chance constraint programming. Tutorial paper for the Stochastic Programming Community Home Page.

[32] Henrion, R. 2013. A critical note on empirical (sample average, Monte Carlo) approximation of solutions to chance constrained programs, in: D. HÃ¶mberg and F. Tröltzsch (eds.) CSMO 2011, IFIP AICT 391. Springer, Berlin 25-37.

[33] IEEE Std. 738-2006. 2006. IEEE standard for calculating the current-temperature of bare overhead conductors 1-59.

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

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

[36] L. C. G.,L. A. M. Veraat. 2011. Failure and Rescue in an Interbank Network. Working Paper.

[37] Liu, M., J.Staum. 2010. Sensitivity analysis of the Eisenberg-Noe model of contagion. Operations Research Letters 38 5 489-491.

[38] Luedtke, J. 2014. A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. Mathematical Programming 146 219-244.

[39] Huneault, M., Galiana, F.D. 1991. A Survey of the Optimal Power Flow Literature. IEEE Transactions on Power Systems, 6 762-770.

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

[41] 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.

[42] Miller L., H.Wagner. 1965. Chance-constrained programming with joint constraints. Operations Research 13 930-945.

[43] Papadimitriou, C. H., K. Steiglitz. 1998. Combinatorial Optimization: Algorithms and Complexity. Dover Publications, Mineola, New York.

[44] Resnick, S.I. 2006. Heavy Tail Phenomena: Probabilistic and Statistical Modeling. New York.

[45] Piterbarg, V. I. 1996. Asymptotic Methods in the Theory of Gaussian Processes and Fields. Amer Mathematical Society.

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

[47] Sezer, A.D. 2010. Modeling of an insurance system and its large deviations analysis. Journal of Computational and Applied Mathematics 235 3 535-546.

[48] Shi, Y. 2012. Rare Events in Stochastic Systems: Modeling, Simulation Design and Algorithmic Analysis. Ph.D Dissertation, Columbia University.

[49] Swanson, J. 2007. Variations of the Solution to a Stochastic Heat Equation. The Annals of Probability 35 2122-2150.

[50] Tudor, C.A. 2013. Analysis of Variations for Self-similar Processes. Series on Probability and Its Applications, Springer.

[51] U.S.-Canada Power System Outage Task Force. 2004. Report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations. `https://reports.energy.gov`.

[52] Van Lelyveld, I., F. Liedorp, M. Kampman. 2011. An Empirical Assessment of Reinsurance Risk. Journal of Financial Stability 7 4 191-203.

[53] Vazirani, V. 2001. Approximation Algorithms. Springer.

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

[55] Zimmerman, Ray D., Carlos E. Murillo-Sanchez. 2011. MATPOWER 4.1 User's Manual. Power Systems Engineering Research Center (PSERC).

[56] Zwart, B., S. Borst, M. Mandjes. 2004. Exact Asymptotics for Fluid Queues Fed by Multiple Heavy-Tailed On-Off Flows. The Annals of Applied Probability 14 903-957.