

A CONDITIONAL MONTE CARLO METHOD FOR ESTIMATING THE FAILURE PROBABILITY OF A DISTRIBUTION NETWORK WITH RANDOM DEMANDS

Jose Blanchet
Juan Li

Industrial Engineering and Operations Research Department
Columbia University
New York, NY 10027, USA

Marvin K. Nakayama

Computer Science Department
New Jersey Institute of Technology
Newark, NJ 07102, USA

ABSTRACT

We consider a model of an irreducible network in which each node is subjected to a random demand, where the demands are jointly normally distributed. Each node has a given supply that it uses to try to meet its demand; if it cannot, the node distributes its unserved demand equally to its neighbors, which in turn do the same. The equilibrium is determined by solving a linear program (LP) to minimize the sum of the unserved demands across the nodes in the network. One possible application of the model might be the distribution of electricity in an electric power grid. This paper considers estimating the probability that the optimal objective function value of the LP exceeds a large threshold, which is a rare event. We develop a conditional Monte Carlo algorithm for estimating this probability, and we provide simulation results indicating that our method can significantly improve statistical efficiency.

1 INTRODUCTION

This paper considers a model of a distribution network, in which there is a single commodity to be distributed and each node is subject to a random demand of the item. Each node has a fixed supply of the commodity, which it uses to try to meet its demand. If the demand at a node exceeds its supply, then the excess demand is distributed equally to each of its neighbors, which in turn do the same. A linear program (LP) determines the equilibrium, where the objective is to minimize the sum of the unserved demands across all nodes. Our goal is to estimate the probability $\alpha(k)$ that the optimal objective function value exceeds a threshold k , which we consider to be a failure. We are particularly interested in the case when k is large.

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

For our model, we consider estimating $\alpha(k)$ via simulation. When the threshold k is large, naive simulation (i.e., without the use of any variance reduction) is inefficient. We thus develop a method based on conditional Monte Carlo (e.g., see Section 8.3 of Ross 2006) for the case when the demands have a

joint Gaussian distribution. To do this, we express the Gaussian demands in polar coordinates. Given the angle, the conditional probability of the LP's optimal objective function value exceeding k can be expressed as the probability of the radial component of the Gaussian lying in an interval or union of intervals, and this conditional probability can be computed analytically. We ran experiments comparing our conditional Monte Carlo approach with naive simulation on some simple networks, and we find that our method can significantly improve efficiency when the threshold k is large.

Our study has some connection to cascading failures, on which there has been previous work. For example, Watts (2002) studies cascades in a sparse, random network of interacting agents whose decisions are determined by the actions of their neighbors according to a simple threshold rule. Dobson, Carreras, Lynch, and Newman (2007) consider a branching process model of cascading failures in an electric power grid. Iyer, Nakayama, and Gerbessiotis (2009) analyze a continuous-time Markov chain of a dependability model with cascading failures.

The rest of the paper is organized as follows. Section 2 presents the model of the distribution network, and it also defines the LP problem and its dual. We establish some properties of the primal and dual LPs in Section 3. We describe the naive simulation and conditional Monte Carlo method for estimating $\alpha(k)$ in Section 4. Section 5 contains the experimental results from some examples, and we give some concluding remarks in Section 6.

2 MODEL

We consider a network model which is induced by a directed graph $G = (V, E)$, where $V = \{1, 2, \dots, d\}$ 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 $i \in V$. We also make the following assumptions:

Assumption 1 The graph is irreducible in the sense that the matrix H is irreducible.

Assumption 2 Each node i has a given supply s_i , and let $s = (s_1, s_2, \dots, s_d)'$, where prime denotes transpose.

Assumption 3 Each node i is subjected to a random demand D_i , and let $D = (D_1, D_2, \dots, D_d)'$. A discussion on the distribution of the vector D will be given momentarily.

Assumption 4 The total supply exceeds the mean total demand.

When the demands are realized, each node serves its demand if it is less than its supply. Otherwise, it equally distributes the unserved demand to its neighbors, which, in turn, do the same with their respective neighbors. Thus, in the end, the demands achieve an equilibrium point given by the following linear program, whose objective is to minimizing the sum of the excess demands across the nodes:

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

The quantity $n(j) = \sum_{i \in V} H(j, i)$ is the number of neighbors of j . Since H is irreducible we have that $n(j) \geq 1$. The quantity $x_i^+ \geq 0$ represents the shedded demand from node i in equilibrium, which is distributed equally among its neighbors. 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. The solution moves around excess demands and supplies to neighbors but does so in such a way that the sum of x_i^+ 's, which are the equilibrium shedded demands, is minimized. The

problem can be expressed in matrix notation as follows. Define $n = H\mathbf{1}$, and let $A(i, j) = H(i, j)/n(i)$ (note that $A(i, i) = 0$), where $\mathbf{1} = (1, 1, \dots, 1)'$ is the d -dimensional column vector with all components equal to 1. Then the previous linear programming problem can be written as:

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

where $\mathbf{0} = (0, 0, \dots, 0)'$ 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, $x^+ = (x_1^+, x_2^+, \dots, x_d^+)'$, and $x^- = (x_1^-, x_2^-, \dots, x_d^-)'$. The goal is that the sum of shedded demands is as small as possible because, e.g., the cost of distributing demands is high. If the cost is too high, for example, larger than a given number, say k , or the LP is infeasible, we consider the network to have failed. We are interested in computing the probability that the network fails, for different values of k .

The dual linear program is the following:

$$\begin{aligned} \max \quad & \lambda'(s - D) \\ \text{s.t.} \quad & (A - I)\lambda \leq \mathbf{1} \\ & \lambda \leq \mathbf{0}. \end{aligned} \tag{3}$$

Equivalently, we can write the dual as

$$\begin{aligned} \max \quad & y'r \\ \text{s.t.} \quad & My \leq \mathbf{1} \\ & y \geq \mathbf{0}, \end{aligned} \tag{4}$$

where $M = I - A$ and $r = D - s$.

3 PROPERTIES OF OUR PRIMAL AND DUAL LINEAR PROGRAMS

We now establish some properties of our primal and dual LPs.

3.1 Feasibility

Theorem 1

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

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

For part (b), note that if $\sum_{i=1}^d D_i > \sum_{i=1}^d s_i$, i.e., $\mathbf{1}'r > 0$, 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 $y = m\mathbf{1}$ is a feasible solution to the dual, for all $m > 0$. Therefore, if $\mathbf{1}'r > 0$, we clearly have that the dual is unbounded.

If $\sum_{i=1}^d D_i \leq \sum_{i=1}^d s_i$, i.e., $\mathbf{1}'r \leq 0$, 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 $v = (v_1, v_2, \dots, v_d)'$ $\geq \mathbf{0}$ such that $v'r > 0$, and $Mv \leq \mathbf{0}$. Let $v_{i_0} = \max_{j=1,2,\dots,d} v_j$.

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,

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

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

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

where the inequality follows from $Mv \leq \mathbf{0}$, and the fact that every entry of the matrix $I + A + A^2 + \cdots + A^{h_2-1}$ is non-negative. Therefore,

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

Consider the first element of $(I - A^{h_2})v$. 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 $\mathbf{1}'r = v'r/v_1 > 0$, which contradicts our assumption that $\mathbf{1}'r \leq 0$. \square

3.2 Shape of the Objective Function of the Dual

Assume that $D \sim N(\mu, \Sigma)$ is jointly Gaussian, i.e., its density function $f(x) = (2\pi)^{-d/2} |\Sigma|^{-1/2} e^{-(x-\mu)'\Sigma^{-1}(x-\mu)/2}$, where $|\Sigma|$ is the determinant of matrix Σ . We can write the polar-coordinate representation: $D = \mu + RW\Psi$, where $R^2 \sim \Gamma(d/2, 1/2)$, i.e., its density function $f(x) = x^{d/2-1} e^{-x/2} (1/2)^{d/2} / \Gamma(d/2)$, $\Gamma(\cdot)$ is the gamma function, $WW^T = \Sigma$, $\Psi = z/\|z\|$, $z = (z_1, z_2, \dots, z_d)' \sim N(0, I)$, $\|z\| = \sqrt{z_1^2 + z_2^2 + \cdots + z_d^2}$, and R and Ψ are independent. Note that D_i could be either positive or negative, for $i = 1, 2, \dots, d$. If D_i is negative for some i , we assume that node i receives an extra supply.

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

Theorem 2 For fixed Ψ ,

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

$$J_\Psi = \{R \geq 0 : \text{primal with } (R, \Psi) \text{ is feasible}\}; \quad (5)$$

(b) $F_\Psi(R)$ is a convex function of R in the set J_Ψ .

Proof. We can write the primal as

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

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

The dual can be written as

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

Because the dual is always feasible, for any $R \in J_\Psi$, $F_\Psi(R)$ is finite and by strong duality (e.g., Theorem 4.4 of Bertsimas and Tsitsiklis 1997), is equal to the optimal value of the dual objective.

Let us fix a particular element $R^* \in J_\Psi$. Then there exists a primal optimal basic feasible solution. Let B be the corresponding optimal basis matrix. The vector x_B of basic variables at that optimal solution is given by $x_B = B^{-1}(s - \mu - R^*W\Psi)$, 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, $B^{-1}(s - \mu - RW\Psi)$ 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 $F_\Psi(R)$ for the new problem is given by

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

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

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

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

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

4 FAILURE PROBABILITY AND ESTIMATION ALGORITHMS

4.1 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., $F(D)$ is larger than k .

Let $Q(D) = 1$ if the primal is feasible, and $Q(D) = 0$ otherwise. Let $\beta_0 = P(Q(D) = 0)$, which is the probability of the first kind of failure, and $\beta_1(k) = P(Q(D) = 1, F(D) \geq k)$, 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 $L(D)$ denote the optimal value of the dual when the demand vector is D . Note that

$$\begin{aligned}
 \alpha(k) &= \beta_0 + \beta_1(k) \\
 &= P(Q(D) = 0) + P(Q(D) = 1, F(D) \geq k) \\
 &= P(Q(D) = 0) + P(Q(D) = 1)P(F(D) \geq k|Q(D) = 1) \\
 &= P(Q(D) = 0)P(L(D) \geq k|Q(D) = 0) + P(Q(D) = 1)P(L(D) \geq k|Q(D) = 1) \\
 &= P(L(D) \geq k),
 \end{aligned} \tag{6}$$

where the fourth equality follows from strong duality, and the fact that when $Q(D) = 0$, i.e., the primal is infeasible, we have $L(D) = +\infty$. Therefore, we want to compute the probability that the optimal value of the dual $L(D)$ 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 5), we consider examples in which $\beta_0 \ll \beta_1(k)$ for all k , so the latter situation occurs.

4.2 Naive Simulation

We first describe how to use naive simulation (i.e., with no variance reduction) to estimate $\alpha(k)$. By (6), we express $\alpha(k) = E[U]$, where $U = I\{L(D) \geq k\}$. Below is an algorithm for computing an estimator of $\alpha(k)$ with naive simulation.

1. Set $i = 1$ and let n be the total number of replications to simulate.
2. Generate $\Psi_i = z_i/\|z_i\|$, where $z_i \sim N(0, I)$. Generate $R_i^2 \sim \Gamma(d/2, 1/2)$, independent of Ψ_i . Let $D^{(i)} = \mu + R_i W \Psi_i$.
3. Solve the dual problem with demand $D^{(i)}$. If the optimal value $L(D^{(i)}) \geq k$, set $U_i = 1$; otherwise, set $U_i = 0$.
4. If $i < n$, then set $i = i + 1$ and go to step 2; otherwise, go to step 5.
5. Compute $\hat{\alpha}_n(k) = (\sum_{i=1}^n U_i)/n$ as our naive simulation estimator of $\alpha(k)$, and a $100(1 - \delta)\%$ confidence interval for $\alpha(k)$ is $(\hat{\alpha}_n(k) \pm \Phi^{-1}(1 - \delta/2)S_n/\sqrt{n})$, where $S_n^2 = (\sum_{i=1}^n (U_i - \hat{\alpha}_n(k))^2)/(n - 1)$, and Φ is the distribution function of a standard normal.

4.3 Conditional Monte Carlo

We now develop a conditional Monte Carlo approach for estimating $\alpha(k)$. Note that $\alpha(k)$ can be written as $\alpha(k) = E[E[U|\Psi]]$, so if we can compute $E[U|\Psi] = P(L(D) \geq k|\Psi)$, then sampling $E[U|\Psi]$ rather than U as in naive simulation reduces variance since $Var[U] = E[Var[U|\Psi]] + Var[E[U|\Psi]]$. Below is an algorithm for computing an estimator of $\alpha(k)$ with conditional Monte Carlo.

1. Set $i = 1$ and let n be the total number of replications to simulate.
2. Generate $\Psi_i = z_i/\|z_i\|$, where $z_i \sim N(0, I)$.
3. Search for the root R_i^* of the equation $F_{\Psi_i}(R) = k$ if it exists, and calculate $E[U_i|\Psi_i]$.
4. If $i < n$, then set $i = i + 1$ and go to step 2; otherwise, go to step 5.

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

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

$$\begin{aligned} E[U|\Psi] &= P(L(D) \geq k|\Psi) \\ &= P(F(D) \geq k, Q(D) = 1|\Psi) + P(Q(D) = 0|\Psi) \\ &= P(F_{\Psi}(R) \geq k, Q(D) = 1|\Psi) + P(Q(D) = 0|\Psi) \\ &= P(R \in K_{\Psi}|\Psi) + P(R \in J_{\Psi}^c|\Psi) \\ &= P(R \in K_{\Psi} \cup J_{\Psi}^c|\Psi), \end{aligned} \tag{7}$$

where $K_{\Psi} = \{R \in J_{\Psi} : F_{\Psi}(R) \geq k\}$, J_{Ψ} is defined in (5), and $J_{\Psi}^c = \{R \geq 0 : \text{primal with } (R, \Psi) \text{ is infeasible}\}$ is the complement of J_{Ψ} . The last equality follows from the fact that $K_{\Psi} \cap J_{\Psi}^c = \emptyset$.

By Theorem 2, for fixed Ψ , $F_{\Psi}(R)$ is piecewise linear and convex on J_{Ψ} . Therefore, K_{Ψ} is an interval or union of intervals if not empty. By Theorem 1, for fixed Ψ , $J_{\Psi} = \{R \geq 0 : R\mathbf{1}'W\Psi \leq \mathbf{1}'s - \mathbf{1}'\mu\}$. By Assumption 4 in Section 2, we have $\mathbf{1}'s > \mathbf{1}'\mu$. If $\mathbf{1}'W\Psi > 0$, then $J_{\Psi} = [0, (\mathbf{1}'s - \mathbf{1}'\mu)/\mathbf{1}'W\Psi]$; $J_{\Psi} = [0, +\infty)$ otherwise. In both cases, J_{Ψ} is an interval. Thus, $K_{\Psi} \cup J_{\Psi}^c$ is an interval or union of intervals if not empty. Because $R^2 \sim \Gamma(d/2, 1/2)$, $E[U|\Psi]$ can be calculated analytically using (7).

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

$$F_{\Psi}(R^*) = k. \tag{8}$$

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

- There is no root in the set J_{Ψ} .
If $F_{\Psi}(R) < k$ for all $R \in J_{\Psi}$, then $K_{\Psi} = \emptyset$ and $E[U|\Psi] = P(R \in J_{\Psi}^c)$. If $F_{\Psi}(R) > k$ for all $R \in J_{\Psi}$, then $K_{\Psi} = J_{\Psi}$ and $E[U|\Psi] = 1$.
- There is only one root R^* in the set J_{Ψ} .
If the slope of $F_{\Psi}(R)$ at R^* is positive, then $K_{\Psi} = [R^*, (\mathbf{1}'s - \mathbf{1}'\mu)/\mathbf{1}'W\Psi]$ when $\mathbf{1}'W\Psi > 0$, and $K_{\Psi} = [R^*, +\infty)$ when $\mathbf{1}'W\Psi \leq 0$; in both cases $E[U|\Psi] = P(R \in [R^*, +\infty))$. If the slope of $F_{\Psi}(R)$ at R^* is negative, then $K_{\Psi} = [0, R^*]$ and $E[U|\Psi] = P(R \in [0, R^*] \cup J_{\Psi}^c)$.
- There are two different roots $R^{*(1)} < R^{*(2)}$ in the set J_{Ψ} .
Because $F_{\Psi}(R)$ is convex on J_{Ψ} , it must be that the slope at $R^{*(1)}$ is negative and the slope at $R^{*(2)}$ is positive. Then $K_{\Psi} = [0, R^{*(1)}] \cup [R^{*(2)}, (\mathbf{1}'s - \mathbf{1}'\mu)/\mathbf{1}'W\Psi]$ when $\mathbf{1}'W\Psi > 0$, and $K_{\Psi} = [0, R^{*(1)}] \cup [R^{*(2)}, +\infty)$ when $\mathbf{1}'W\Psi \leq 0$; in both cases $E[U|\Psi] = P(R \in [0, R^{*(1)}] \cup [R^{*(2)}, +\infty))$.

Now we explain how we search for the root R^* of (8) for a fixed Ψ using Newton's method. Recall that if $\mathbf{1}'W\Psi > 0$, J_{Ψ} is a finite interval; J_{Ψ} is an infinite interval otherwise. We hope to find a uniform way to search in both cases to simplify coding. In the meanwhile, we want to control the computational error. If $\mathbf{1}'W\Psi > 0$, set $b = G_{\Gamma}((\mathbf{1}'s - \mathbf{1}'\mu)/\mathbf{1}'W\Psi)^2$, 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 $\tilde{K}_{\Psi} = \{R \in [b_l, b_u] : F_{\Psi}(R) \geq k\}$. Then our algorithm calculates \tilde{K}_{Ψ} instead of K_{Ψ} . Because $P(R \in J_{\Psi}) - P(R \in [b_l, b_u]) = c$, we have $0 \leq P(R \in K_{\Psi}) - P(R \in \tilde{K}_{\Psi}) \leq c$, and the error between $E[U|\Psi]$ 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 $D = \mu + R_0 W\Psi$ using the simplex algorithm to obtain the corresponding $F_\Psi(R_0)$ and slope l_0 of $F_\Psi(R)$ at R_0 .

- If $l_0 > 0$, there is at most one root.
 - If $F_\Psi(R_0) > k$ or $F_\Psi(b_u) < k$, then there is no root.
 - Otherwise, there is one root. We update $R_1 = \min(R_0 + (k - F_\Psi(R_0))/l_0, b_u)$. Solve the dual LP with demand $D = \mu + R_1 W\Psi$ to obtain the corresponding $F_\Psi(R_1)$ 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 $F_\Psi(R_0) < k$ and $F_\Psi(b_u) < k$, then there is no root.
 - If $F_\Psi(R_0) \geq k$ and $F_\Psi(b_u) < k$, then there is one root. We update $R_1 = R_0 + (k - F_\Psi(R_0))/l_0$. Solve the dual LP with demand $D = \mu + R_1 W\Psi$ to obtain the corresponding $F_\Psi(R_1)$ and slope l_1 . We keep updating until we find the root R^* .
 - If $F_\Psi(R_0) < k$ and $F_\Psi(b_u) \geq k$, then there exists one root. We first increase R from R_0 to R_1 in J_Ψ 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 $R_2 = \min(R_1 + (k - F_\Psi(R_1))/l_1, b_u)$. Solve the dual LP with demand $D = \mu + R_2 W\Psi$ to obtain the corresponding $F_\Psi(R_2)$ and slope l_2 . We keep updating until we find the root R^* .
 - If $F_\Psi(R_0) \geq k$ and $F_\Psi(b_u) \geq k$, then there exists no root or there exist two roots. We first test if there exists a root. We update $R_1 = R_0 + (k - F_\Psi(R_0))/l_0$. Solve the dual LP with demand $D = \mu + R_1 W\Psi$ to obtain the corresponding $F_\Psi(R_1)$ 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_Ψ using the fixed step way such that $l'_1 > 0$, and then update $R'_2 = \min(R'_1 + (k - F_\Psi(R'_1))/l'_1, b_u)$. Solve the dual LP with demand $D = \mu + R'_2 W\Psi$ to obtain the corresponding $F_\Psi(R'_2)$ and slope l'_2 . We keep updating until we find the second root $R^{*(2)}$.

5 EXAMPLES

5.1 Efficiency of Simulation Estimators

We now discuss a basis for comparing the estimators $\hat{\alpha}_n(k)$ for naive simulation and $\tilde{\alpha}_n(k)$ for conditional Monte Carlo.

Definition 1 Suppose we want to estimate $\alpha = E[X]$, and X_1, X_2, \dots, X_n are independent replications of X . Then $\hat{\alpha}_n = (\sum_{i=1}^n X_i)/n$ is an unbiased estimator of α , and $S_n^2 = (\sum_{i=1}^n (X_i - \hat{\alpha}_n)^2)/(n-1)$ is an unbiased estimator of $\text{Var}[X] = \sigma^2$, which we assume is finite. By the central limit theorem, the error $\hat{\alpha}_n - \alpha$ is approximately normally distributed with mean 0, and standard deviation σ/\sqrt{n} , which can be estimated by S_n/\sqrt{n} . We then define the *RSE (relative standard error)* as $S_n/(\sqrt{n}\hat{\alpha}_n)$.

The standard deviation σ/\sqrt{n} is a measure of the absolute error $\hat{\alpha}_n - \alpha$. Other things being equal, when comparing two unbiased estimator of the same quantity, we prefer the one with smaller standard deviation. However, what if the estimator with smaller standard deviation takes longer computing time (*CT*)? Glynn and Whitt (1992) suggest a criterion: in comparing alternative estimators, each of which is the average of unbiased independent replications, we prefer the one with smaller (variance per replication) \times (expected computing time per replication), which can be estimated by $S_n^2/n \times CT$. We also would like to compare the estimators for different thresholds k . Therefore, we use the relative measure $RSE^2 \times CT$ as the criterion.

In our experiments we apply naive simulation method and conditional Monte Carlo method to different networks, and compare $RSE^2 \times CT$.

5.2 Design of Networks

We designed the networks in our examples so that the correlation of the demands between two nodes is related to the topology of the network, with correlation decreasing the “further apart” the nodes are. For each $i, j \in V$ with $i < j$, we define a distance $d(i, j)$ to measure the closeness. Because we assume that the network is irreducible, there exists some $m \geq 1$ such that there is a directed path $i = i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_m = j$. Let $d(i, j)$ denote the minimum of such m 's. Then if $d(i, j)$ is small, we assign a “large” correlation between the two nodes. On the contrary, if $d(i, j)$ is large, we assign a “small” correlation. Considering the correlation matrix is symmetric, we set $d(j, i) = d(i, j)$ for each $i, j \in V$ with $i < j$. We also assign a variance to D_i for each $i \in V$. Moreover, because we need to calculate the Cholesky decomposition of covariance matrix of Σ , we choose Σ so that it is positive definite.

Because we are interested in the case when $\alpha(k)$ is decreasing in k , we choose parameters s, u , and Σ such that β_0 is very small. We tested two examples with $n = 10^5$. We set $c = 10^{-10}$ in our search algorithm for the root R^* of (8); this value was selected as it is significantly smaller than $\tilde{\alpha}_n(k)$ in all of our experiments. The step length t should not be too small, and we set $t = \sqrt{d}$ because the mean of R^2 is d . Simulation results are as follows.

5.3 Example 1: $d = 3$

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

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

Table 1: Results of Naive Simulation and Conditional Monte Carlo for $d = 3$.

k	Naive Simulation		Conditional MC	
	$\hat{\alpha}_n(k)$	$RSE^2 \times CT$	$\tilde{\alpha}_n(k)$	$RSE^2 \times CT$
1	1.82×10^{-1}	1.45×10^{-2}	1.81×10^{-1}	1.61×10^{-2}
5	2.16×10^{-2}	1.44×10^{-1}	2.11×10^{-2}	4.78×10^{-2}
10	2.34×10^{-3}	1.36×10^0	2.47×10^{-3}	7.45×10^{-2}
13	5.30×10^{-4}	6.01×10^0	5.20×10^{-4}	9.31×10^{-2}
16	1.30×10^{-4}	2.45×10^1	8.89×10^{-5}	1.13×10^{-1}
20	3.00×10^{-5}	1.06×10^2	6.13×10^{-6}	1.41×10^{-1}
24	0	NA	2.93×10^{-7}	1.69×10^{-1}
26	0	NA	5.49×10^{-8}	1.78×10^{-1}

The estimate of β_0 is 3.30×10^{-9} , which is smaller than $\tilde{\alpha}_n(k)$ for all the above values of k .

5.4 Example 2: $d = 10$

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

$$H = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad s = \begin{pmatrix} 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \end{pmatrix}, \quad u = \begin{pmatrix} 1 \\ 5 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

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

Table 2: Results of Naive Simulation and Conditional Monte Carlo for $d = 10$.

k	Naive Simulation		Conditional MC	
	$\hat{\alpha}_n(k)$	$RSE^2 \times CT$	$\tilde{\alpha}_n(k)$	$RSE^2 \times CT$
1	5.03×10^{-1}	4.20×10^{-3}	5.02×10^{-1}	7.23×10^{-3}
5	3.20×10^{-2}	1.28×10^{-1}	3.05×10^{-2}	2.12×10^{-1}
10	3.88×10^{-3}	1.09×10^0	3.88×10^{-3}	6.38×10^{-1}
15	4.50×10^{-4}	9.40×10^0	4.42×10^{-4}	1.62×10^0
20	3.00×10^{-5}	1.41×10^2	4.07×10^{-5}	3.93×10^0
23	0	NA	8.76×10^{-6}	5.28×10^0
30	0	NA	3.95×10^{-7}	1.18×10^1
34	0	NA	4.58×10^{-8}	1.62×10^1

The estimate of β_0 is 9.85×10^{-10} , which is smaller than $\tilde{\alpha}_n(k)$ for all the above values of k .

5.5 Discussion of Results

1. As we expect, if we choose parameters s, u , and Σ such that β_0 is very small, $\alpha(k)$ is decreasing in k .
2. For naive simulation, the performance deteriorates very quickly as k increases. Because we fix the number n of simulations, when k is very large (i.e., $\alpha(k) < 1/n$), we do not get even one observation of the event $\{L(D) \geq k\}$.

3. For conditional Monte Carlo, the performance deteriorates as well, but not as quickly as naive simulation. No matter how large k is, we can obtain a non-zero estimate of $\alpha(k)$ even though we fix n .
4. For non-rare events (i.e., when k is small), naive simulation method works fine, sometimes even better than our conditional Monte Carlo algorithm, as we now explain. The naive simulation method only needs to solve a single optimization problem to determine U_i in each replication i . In contrast, our conditional Monte Carlo method needs to solve several optimization problems to find R_i^* in each replication i . Thus, the added computational effort required by conditional Monte Carlo can lead to it performing worse than naive simulation. However, as k increases, conditional Monte Carlo method works much better. The larger k is, the bigger the advantage conditional Monte Carlo has compared to naive simulation. The advantage arises because of the significant variance reduction obtained for large k overwhelms the additional computational effort.
5. As the dimension of the network increases, neither method gives good results when k is large. However, conditional Monte Carlo offers much better estimates.

In conclusion, naive simulation can do well when k is small, but conditional Monte Carlo method outperforms naive simulation when k is large.

6 CONCLUDING REMARKS

We presented a model of network with a goal of fulfilling random demands at the nodes given fixed supplies. Excess demand at a node is distributed equally to its neighbors and the equilibrium is determined by solving a linear program. We are interested in estimating the probability that the optimal objective function value of the LP exceeds a threshold k . We developed a conditional Monte Carlo approach to efficiently simulate the problem when k is large. Empirical results show significant improvement over naive simulation for large k .

There are several extensions we are now pursuing. Our model currently assumes that nodes equally distribute unserved demands to neighbors, but we can modify our (primal) LP to determine the optimal allocation of unserved demands to neighbors. The current objective function of our (primal) LP is the sum of the unserved demands across the nodes; alternatively, we can consider weighted sums, which can take into account differences in costs of distributing demands over different edges. We are also investigating applying importance sampling (e.g., Section 8.6 of Ross 2006) to estimate $\alpha(k)$ for large k .

ACKNOWLEDGMENTS

This work has been supported in part by the National Science Foundation under Grant No. DMS-0902075, DMS-0806145, DMS-0846816, DMS-1069064, CMMI-0926949, CMMI-0846816 and CMMI-1069064.

REFERENCES

- Bertsimas, D., and N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Massachusetts: Athena Scientific.
- Dobson, I., B. A. Carreras, V. E. Lynch, and D. E. Newman. 2007. "Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization". *Chaos* 17:article 026103 (13 pages).
- Glynn, P. W., and W. Whitt. 1992. "The asymptotic efficiency of simulation estimators". *Operations Research* 40:505–520.
- Iyer, S. M., M. K. Nakayama, and A. V. Gerbessiotis. 2009. "A Markovian dependability model with cascading failures". *IEEE Transactions on Computers* 139:1238–1249.
- Ross, S. M. 2006. *Simulation*. 4 ed. San Diego, CA: Academic Press.
- Watts, D. J. 2002. "A simple model of global cascades on random networks". *Proceedings of the National Academy of Sciences USA* 99:5766–5771.

AUTHOR BIOGRAPHIES

JOSE BLANCHET is a faculty member of the IEOR Department at Columbia University. Jose holds a Ph.D. in Management Science and Engineering from Stanford University. Prior to joining Columbia he was a faculty member in the Statistics Department at Harvard University. Jose is a recipient of the 2009 Best Publication Award given by the INFORMS Applied Probability Society and of the 2010 Erlang Prize. He also received a PECASE award given by NSF in 2010. He worked as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability and Monte Carlo methods. He serves in the editorial board of *Advances in Applied Probability*, *Journal of Applied Probability*, *Mathematics of Operations Research* and *QUESTA*. His email is jose.blanchet@gmail.com.

JUAN LI is a Ph.D. student in the IEOR Department at Columbia University, and a member of INFORMS. She received her B.S. in Mathematics from Peking University. Her research interests are in application of simulation and financial engineering. Her email is li.juan611@gmail.com.

MARVIN K. NAKAYAMA is a professor in the Department of Computer Science at the New Jersey Institute of Technology. He received his Ph.D. in operations research from Stanford University. He won second prize in the 1992 George E. Nicholson Student Paper Competition sponsored by INFORMS and is a recipient of a CAREER Award from the National Science Foundation. He is the Simulation Analysis and Stochastic Modeling Area Editor for *ACM Transactions on Modeling and Computer Simulation* and the Simulation Area Editor for *INFORMS Journal on Computing*. His email is marvin@njit.edu.