Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

A SEQUENTIAL EXPERIMENT DESIGN FOR INPUT UNCERTAINTY QUANTIFICATION IN STOCHASTIC SIMULATION

Yuan Yi
Wei Xie

Industrial and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180, USA

Enlu Zhou

Industrial and Systems Engeering
Georgia Institute of Technology
Atlanta, GA 30324, USA

ABSTRACT

When we use simulations to estimate the performance of a stochastic system, simulations are often driven by input distributions that are estimated from real-world data. There is both input and simulation uncertainty in the performance estimates. Non-parametric sampling approaches, e.g., the bootstrap, could be used to generate samples of input distributions quantifying both input model and parameter uncertainty. In this paper, a sequential experiment design is proposed to efficiently propagate the input uncertainty to output mean and deliver a percentile confidence interval to quantify the impact of input uncertainty on the system performance. Compared to the classical equal allocation, it could assign more computational budget to samples of input distributions that contribute most to the percentile confidence interval estimation. Our approach is supported by rigorous theoretical and empirical study.

1 INTRODUCTION

Stochastic simulation could be used to evaluate the performance of complex stochastic systems. The input models that are used to drive the simulation are often estimated from real-world data. Therefore, there are two sources of uncertainty in the system performance estimate: input estimation error and simulation uncertainty. In this paper, the input estimation error is also called *input uncertainty*. Since the input uncertainty could dominate the simulation uncertainty, ignoring it may lead to unfounded confidence in the simulation assessment of system performance (Barton and Schruben 2001). In our study, we want to efficiently estimate the impact of input uncertainty and control the simulation estimation error.

The approaches to quantify input uncertainty could be divided into parametric, semi- and non-parametric approaches. When the parametric families of input distributions are known, the input uncertainty could be characterized by the parameter estimation error (Xie, Nelson, and Barton 2015). The semi-parametric, e.g., Bayesian model average (BMA) (Chick 2001) and nonparametric approach, e.g., the bootstrap (Barton and Schruben 2001), were proposed to quantify both input model and parameter uncertainty. Notice that BMA delivers posteriors of all data coming from each of candidate distributions and uses them to quantify the input uncertainty. Without any prior information, it could be hard to specify an appropriate set of candidate distributions.

Approaches to quantify the input uncertainty could be divided into frequentist and Bayesian approaches. If a *percentile* confidence or credible interval is desired to measure the impact of input uncertainty on the system mean performance estimate, it is typically recommended to have a few thousands of samples of input distributions to quantify the input uncertainty. These samples could be generated by sampling approaches, including the bootstrapping and Bayesian sampling approaches. Since each simulation run could be computationally expensive, it is important to develop approaches that efficiently propagate the input uncertainty to outputs. When input models could be specified by finite "parameters", a metamodel

of the system mean response could be constructed to efficiently propagate the input uncertainty to output mean (Xie, Nelson, and Barton 2015). When nonparametric approaches, such as the bootstrap, are used to quantify the input uncertainty, the direct simulation with equal computational allocation at all samples of input distributions is commonly used to propagate the input uncertainty to outputs (Barton and Schruben 2001).

For illustration purposes, in this paper, we focus on using the non-parametric bootstrapping to quantify the input uncertainty and building a percentile confidence interval (CI) to quantify the impact of input uncertainty. Our approach is also applicable to the situation where a Bayesian approach is used to quantify the input uncertainty and a percentile credible interval is desired for quantifying the impact of input uncertainty. Notice that it is easy to extend our approach to construct a percentile CI quantifying both input and simulation estimation uncertainty; see Xie, Nelson, and Barton (2015).

When we propagate the input uncertainty to outputs, since samples of input distributions do not equally contribute to estimating the percentile CI quantifying the impact of input uncertainty, the equal allocation could waste simulation resources to precisely estimate the mean response for those unimportant samples. This leads to the desire of computational efficient and precise algorithms. *In this paper, we propose a sequential experiment design that could gradually find the important samples of input distributions and assign more computational resource there. Therefore, we could efficiently use the computational budget to deliver a percentile CI quantifying the impact of input uncertainty.*

The next section provides the problem statement. In Section 3, we propose a sequential experiment design to efficiently estimate the percentile CI quantifying the impact of input uncertainty. This is followed by theoretical support. In Section 4, we report a numerical study of an $M/M/1/50$ queue, and we conclude the paper in Section 5.

2 PROBLEM STATEMENT

Simulation is driven by input models, denoted by F. For the jth replication, the simulation output Y_j is

$$Y_j(F) = \mu(F) + \varepsilon_j(F)$$

where, $\mu(F)$ denotes the unknown output mean and $\varepsilon_j(F)$ represents the simulation error with $\varepsilon_j(F) \sim N(0, \sigma^2(F))$. Notice that the simulation outputs depend on the choice of input models. F could be composed of univariate and multivariate joint distributions. For the notation simplification, we only consider one univariate input distribution.

We denote the underlying true input model by F^c. The unknown F^c is estimated by a random sample of m real-world data, denoted by $\{X_1, X_2, \ldots, X_m\}$, with $X_i \overset{i.i.d.}{\sim} F^c$ and $i = 1, 2, \ldots, m$. Let $\widehat{F}$ represent the point estimator of F^c. We are interested in quantifying the impact of input uncertainty on the system mean performance estimate. Specifically, we want to find the $(1-\alpha^*)100\%$ two-sided equal probability percentile CI, denoted by $\text{CI}^* \equiv [Q_L, Q_U]$, such that

$$\Pr(\mu(F^c) \in [Q_L, Q_U]) = 1 - \alpha^*. \tag{1}$$

If $\mu(\cdot)$ is known, we have $Q_L = \inf\{q : \Pr(\mu(\widehat{F}) \le q) \ge \alpha^*/2\}$ and $Q_U = \inf\{q : \Pr(\mu(\widehat{F}) \le q) \ge 1 - \alpha^*/2\}$.

It could be difficult to get the sampling distribution for $\widehat{F}$. In this paper, the bootstrapping is used to quantify the input model and parameter uncertainty. That means we draw with replacement to generate m bootstrapped samples, denoted by $\{X_1^{(1)}, X_2^{(1)}, \ldots, X_m^{(1)}\}$. Based on them, we could construct a bootstrapped empirical distribution, denoted by F_1. By repeating this procedure B times, we could generate B bootstrapped empirical distributions, denoted by $F_1, F_2, \ldots, F_B$, to quantify the input uncertainty. In this paper, we ignore the sampling uncertainty for finite B bootstrapped samples. Accounting for this source of uncertainty is our on-going research.

Let $q_1 = \alpha^*/2$ and $q_2 = 1 - \alpha^*/2$. We have $Q_L = \mu_{[q_1 B]}$ and $Q_U = \mu_{[q_2 B]}$, where $\mu_b \equiv \mu(F_b)$ and the permutation, denoted by $[\cdot]$, is defined by $\mu_{[1]} \le \mu_{[2]} \le \cdots \le \mu_{[B]}$. Suppose $q_1 B$ and $q_2 B$ are integers for

simplicity. That means there exists samples of input distributions with the index, denoted by $[q_1 B]$ and $[q_2 B]$, in the set $\{F_1, F_2, \ldots, F_B\}$ corresponding to the true percentiles. *When $\mu(\cdot)$ is known, there is no additional simulation estimation error introduced when we propagate the input uncertainty to output mean. Thus, $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$ define the faithful CI we want to estimate.*

At any F, the system true mean response $\mu(F)$ is unknown and it could be estimated by running simulations. If we know the true input model F^c, we could directly run simulations at F^c and estimate $\mu(F^c)$ by using the sample mean of simulation outputs. However, F^c is unknown and estimated by m real-world data. The input uncertainty is quantified by bootstrapped samples, $\{F_1, F_2, \ldots, F_B\}$. The straightforward approach to propagate the input uncertainty to output means is using *the direct bootstrap* that equally allocates, say n replications, at each bootstrapped sample of input distributions. Then, we could estimate $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$ by $\bar{Y}_{(q_1 B)}$ and $\bar{Y}_{(q_2 B)}$, where $\bar{Y}(F_b) = \sum_{j=1}^{n} Y_j(F_b)/n$ and $\bar{Y}_b \equiv \bar{Y}(F_b)$. The permutation, denoted by $(\cdot)$, is defined by $\bar{Y}_{(1)} \leq \bar{Y}_{(2)} \leq \cdots \leq \bar{Y}_{(B)}$. *There exist two sources of error in the point estimator $\bar{Y}_{(\gamma B)}$ with $\gamma = q_1$ and q_2: the selection error and simulation estimation error.* To estimate the percentiles $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$ well, we need to control both errors. Since we do not need to estimate μ_b precisely for the sample F_b with mean response far from the percentiles, the equal allocation does not efficiently use the computational resource.

Our objective is to find a good computational allocation to bootstrapped input distributions $F_1, F_2, \ldots, F_B$, so that we could precisely estimate percentiles $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$. We propose a sequential experiment design that could efficiently allocate computational resources to important samples of input distributions that contribute most to the percentile estimation. By doing so, we achieve the goal of computational efficiency, and our algorithm delivers an accurate percentile estimate along with a specified accuracy.

3 SEQUENTIAL EXPERIMENT DESIGN FOR INPUT UNCERTAINTY QUANTIFICATION

In this section, we first describe a sequential procedure in Section 3.1 that allocates the computational resource to $F_1, F_2, \ldots, F_B$ so that we could efficiently estimate percentile values $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$. Then, we provide the theoretical support for our algorithm in Section 3.2.

Motivated by Lesnevski, Nelson, and Staum (2008), we propose a sequential procedure that could efficiently estimate the percentiles $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$, while simultaneously controlling the estimation accuracy. Specifically, we develop two-sided screening supported by Theorem 1 in Section 3.2 that gradually screens out samples that are statistically impossible to be $[q_1 B]$ and $[q_2 B]$, assign more replications to surviving samples and run simulations. We iteratively repeat this screening procedure until the percentile estimation reaches to the desired accuracy. Since the percentile estimates are based on order statistics $\bar{Y}_{(q_1 B)}$ and $\bar{Y}_{(q_2 B)}$, screening over large samples could cause selection bias; see Lesnevski, Nelson, and Staum (2007). We divide the whole screening procedure into two phases. The main purpose of Phase I is to screen out most samples with mean responses far from $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$. Then, we use the technique of *restart* to reduce selection bias caused by screening in Phase I. The main focus in Phase II is the percentile estimation. The number of samples screened out in Phase II is typically small and the corresponding selection bias is negligible.

Based on simulation outputs of samples in surviving sets, we could construct $(1-\alpha)100\%$ CIs, denoted by CI_1 and CI_2, for the lower and upper percentile estimation by Equation (2). The percentile estimation accuracy is measured by the widths of these CIs. Theorem 2 proves that CI_1 and CI_2 cover $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$ with probability $(1-\alpha)$. By the Bonferroni inequality, the total significant level α could be divided into the significant levels for screening and simulation estimation error, α_I and α_0. Since most screening supposes to finish in Phase I, we typically assign larger proportion of screening significant level to Phase I, denoted by α_{I1}, than that to Phase II, denoted by α_{I2}. α_{I1} and α_{I2} are tunable parameters. In the empirical study, we set $\alpha_{I1} = 0.8\alpha_I$.

Since we want to quantify the impact of input uncertainty on the system mean performance, $\{\mu(F_1), \mu(F_2), \ldots, \mu(F_B)\}$ are the mean responses of the same system with different input distributions. When we employ common random number (CRN) to generate X from F_b and $F_{b'}$, simulation outputs

$Y(F_b)$ and $Y(F_{b'})$ could have stronger correlation if F_b and $F_{b'}$ are close to each other, where $b \neq b'$ and $b, b' = 1, 2, \ldots, B$. Therefore, CRN is implemented in our procedure to efficiently screen out samples with mean responses very similar.

Notice that there exist three types of CIs in our procedure. $[\mu_{[q_1 B]}, \mu_{[q_2 B]}]$ represents the percentile CI quantifying the impact of input uncertainty on system mean performance. The estimation accuracy for percentiles $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$ is characterized by CI_1 and CI_2 in Equation (2). The CI widths, $|\text{CI}_k|$ with $k = 1, 2$, are used as the stopping criteria for our sequential experiment design. Let L_k denote the desired width for CI_k. For a given input distribution, say F_b, the third type of CI characterizes the simulation estimation uncertainty for $\mu(F_b)$ with $b = 1, 2, \ldots, B$. Since $|\text{CI}_k|$ with $k = 1, 2$ depends on the simulation estimation accuracy of all samples in the surviving sets, it is hard to use it to specify the maximum number of iterations required. Therefore, the third type of CI is used to specify the maximum number of iterations.

3.1 Sequential Procedure

We define I^k with $k = 1, 2$ as the sets of samples of input distributions that survive screening for estimating percentiles $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$ accordingly. Let I^{L_k} and $I^{U_k}, k = 1, 2$ denote the sets of eliminated samples that have mean responses statistically smaller or larger than the desired one.

3.1.1 Phase I

In this section, we describe the procedure in Phase I. We first allocate n_0 replications to each F_b with $b = 1, 2, \ldots, B$ in Step 2. Generally, n_0 is small, in most cases n_0 can be set no greater than 30. We run simulations, obtain outputs $\{Y_1(F_b), Y_2(F_b), \ldots, Y_{n_0}(F_b)\}$ and calculate the sample mean and variance, denoted by $\bar{Y}_b$ and S_b^2, with $b = 1, 2, \ldots, B$. Then, we iteratively screen out samples with mean responses far from the percentiles. The maximum number of iterations for Phase I is $M = \max(M_1, M_2)$, where M_k denotes the number of iterations required to have the CI width of $\mu_{[q_k B]}$ less than L_k with $k = 1, 2$; see Step 3. Notice that it is based on the third type of CI. In Phase I, we update I^1 and I^2 simultaneously in Step 4, which could save the computational resources. The accumulated replications for samples in surviving sets increases by a growth factor R. At iteration ℓ, each surviving sample has the accumulated replications $N(\ell) = \lceil n_0 R^\ell \rceil$. Screening can stop early if early stopping criteria are satisfied: a) only one sample remains in I^k, or b) $|\text{CI}_k|$ is no greater than L_k; see Step 5. Screening for lower and upper percentiles could stop at different iterations. We denote the procedure stops screening for $I^k, k = 1, 2$ at iterations J_k respectively.

Phase I mainly includes following steps.

1. Specify q_1, q_2, α_{I1}, α_{I2}, α_0. Define $I_0^k \leftarrow \{1, 2, \ldots, B\}$ and $I_0^{L_k} = I_0^{U_k} = \emptyset$ for $k = 1, 2$.
2. Initial allocation. Assign n_0 replications to F_b with $b = 1, 2, \ldots, B$. Simulate and sort sample means of simulation outputs: $\bar{Y}_{(1)} \leq \bar{Y}_{(2)} \leq \cdots \leq \bar{Y}_{(B)}$. Set $L_k = \beta \cdot \bar{Y}_{(q_k B)}$ for $k = 1, 2$, where β is a constant, such as 0.01.
3. Estimate the maximum number of iterations for Phase I. Let $M = \max(M_1, M_2)$, where for $k = 1, 2$

$$N_{\max} = \left(2 \frac{S_{(q_k B)} \cdot t_{n_0 - 1, 1 - \frac{\alpha_0}{2}}}{L_k} \right)^2 \text{ and } M_k = \max\left(\lceil \log_R \frac{N_{max}}{n_0} \rceil, 0 \right) + 1$$

   Define $\alpha_{I1}^{kl} = \frac{\alpha_{I1}}{M_k}$, which is the screening error allowance for each iteration during Phase I.
4. Screening. For stage $\ell = 1, 2, \ldots, M$:

$$I_\ell^k = \{ \forall i \in I_{\ell-1}^k : \sum_{\substack{j \neq i \\ \forall j \in I_{\ell-1}^k}} \mathbf{1}(\bar{Y}_j \leq \bar{Y}_i + W_{ij}) \geq q_k B - 1 - |I_{\ell-1}^{L_k}|$$

$$\textit{and} \quad \sum_{\substack{j \neq i \\ \forall j \in I_{\ell-1}^k}} \mathbf{1}(\bar{Y}_j \geq \bar{Y}_i - W_{ij}) \geq B - q_k B - |I_{\ell-1}^{U_k}| \}$$

$$I_\ell^{L_k} = I_{\ell-1}^{L_k} \bigcup \{\forall i \in I_{\ell-1}^k : \sum_{\substack{j\neq i \\ \forall j \in I_{\ell-1}^k}} \mathbf{1}(\bar{Y}_j \leq \bar{Y}_i + W_{ij}) < q_k B - 1 - |I_{\ell-1}^{L_k}|\}$$

$$I_\ell^{U_k} = I_{\ell-1}^{U_k} \bigcup \{\forall i \in I_{\ell-1}^k : \sum_{\substack{j\neq i \\ \forall j \in I_{\ell-1}^k}} \mathbf{1}(\bar{Y}_j \geq \bar{Y}_i - W_{ij}) < B - q^k B - |I_{l-1}^{U_k}|\}$$

Where, $W_{ij} = t_{N(\ell)-1,1-\frac{\alpha_{I1}^{k\prime}}{|I_{\ell-1}^k|-1}} \cdot \frac{S_{ij}}{\sqrt{N(\ell)}}$ and $S_{ij}^2 = \frac{1}{N(\ell)-1}\sum_{h=1}^{N(\ell)}(Y_{ih} - Y_{jh} - (\bar{Y}_i - \bar{Y}_j))^2$ and $\mathbf{1}(x)$ is the indicator function.

5. Check stopping criteria. If $|I_\ell^k| = 1$ or $\ell = M_k$ or $|\mathrm{CI}_k| \leq L_k$, then stops updating I^k for $k = 1,2$, where CI_k is the CI for percentile estimation

$$\mathrm{CI}_k \equiv \left[\min_{i\in I_\ell^k}\left(\bar{Y}_i - t_{N(\ell)-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{N(\ell)-1}}\right), \max_{i\in I_\ell^k}\left(\bar{Y}_i + t_{N(\ell)-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{N(\ell)-1}}\right)\right]. \quad (2)$$

If stopping criteria for both lower and upper percentiles hold, the procedure moves to Phase II. Otherwise assign additional $n_0 R^l(R-1)$ replications to each sample in the surviving set I_ℓ^k, and run simulations, let $\ell \leftarrow \ell + 1$. Loop back to Step 4.

3.1.2 Phase II

In this section, we describe the procedure for Phase II. To reduce the selection bias from the screening procedure in Phase I, our procedure "restarts" in Step 1 (Lan, Nelson, and Staum 2010; Boesel, Nelson, and Kim 2003; Nelson and Goldsman 2001). All simulations from Phase I are discarded, and for every sample that survives Phase I, do simulations with replications $\lceil n_0 R^{J_k+1} \rceil$ with $k = 1,2$.

We estimate the maximum number of iterations in Step 2. If we can ensure that every sample in I^k has an individual CI width no greater than $L_k/|I^k|$, then $|\mathrm{CI}_k| \leq L_k$ for $k = 1,2$ because all samples in each surviving set are statistically indifferent. Therefore, the maximum number of iterations for Phase II is determined by the number of iterations required for every sample in the surviving set I^k having CI width less than $L_k/|I^k|$ with $k = 1,2$. Notice that it is based on the third type of CI.

To fully utilize the error allowance on screening, two updates are made. First, our procedure "recycles" the unused portions of α_{I1}, $\left(1 - J_k/M_k\right)\alpha_{I1}, k = 1,2$, since Phase I may exit early. Therefore, during Phase II, the allowable significant level on screening is $\alpha_{I2}^k = \alpha_I - J_k\alpha_{I1}/M_k$. Second, the number of iteations required in Phase II, denoted by P, is updated at each iteration. Unlike Phase I, where α_{I1} is equally distributed among all iterations, α_{I2} may be unevenly allocated to all iterations. The reason is that P is determined in a conservative manner. Therefore, constantly updating the number of stages P_ℓ required at iteration ℓ would give a tighter upper bound on P, which comes with higher utilization of screening error allowance. For the theoretical support for this adaptively updating the number of iterations, please see Theorem 3 in Section 3.2.

For q_k with $k = 1,2$:

1. Restart. Allocate $N(J_k + 1) = n_0 R^{J_k+1}$ replications to each sample in the surviving set $I_{J_k}^k$, run simulations and calculate sample variance of simulation outputs S_i^2 for $i \in I_{J_k}^k$.
2. Estimate the number of iterations needed. Set

$$N'_{\max} = \left(\frac{2\max_{i\in I_{J_k}^k}(S_i) \cdot t_{N(J_k+1)-1,1-\frac{\alpha_0}{2}}}{L_k/|I_{J_k}^k|}\right)^2$$

Denoted the maximum number of iteration needed for Phase II by $P^k_{J_k} = \max\left(\lceil \log_R \frac{N'_{\max}}{N(J_k+1)} \rceil, 0\right) + 1$. Let $\alpha^k_{I2} = \alpha_I - \frac{J_k}{M_k}\alpha_{I1}$. Define $\alpha'_{I2,J_k} = \frac{\alpha^k_{I2}}{P^k_{J_k}}$ as the error allowance for each iteration.

3. Screening. For stage $\ell = J_k+1, J_k+2, \ldots, J_k+P^k_\ell$:

$$I^k_\ell = \{\forall i \in I^k_{\ell-1} : \sum_{\substack{j\neq i \\ \forall j \in I^k_{\ell-1}}} \mathbf{1}(\bar{Y}_j \leq \bar{Y}_i + W_{ij}) \geq q_k B - 1 - |I^{L_k}_{\ell-1}|$$
$$\textit{and} \quad \sum_{\substack{j\neq i \\ \forall j \in I^k_{l-1}}} \mathbf{1}(\bar{Y}_j \geq \bar{Y}_i - W_{ij}) \geq B - q_k B - |I^{U_k}_{\ell-1}|\}$$

$$I^{L_k}_\ell = I^{L_k}_{\ell-1} \bigcup \{\forall i \in I^k_{\ell-1} : \sum_{\substack{j\neq i \\ \forall j \in I^k_{\ell-1}}} \mathbf{1}(\bar{Y}_j \leq \bar{Y}_i + W_{ij}) < q_k B - 1 - |I^{L_k}_{\ell-1}|\}$$

$$I^{U_k}_\ell = I^{U_k}_{\ell-1} \bigcup \{\forall i \in I^k_{\ell-1} : \sum_{\substack{j\neq i \\ \forall j \in I^k_{\ell-1}}} \mathbf{1}(\bar{Y}_j \geq \bar{Y}_i - W_{ij}) < B - q_k B - |I^{U_k}_{\ell-1}|\}$$

Where, $W_{ij} = t_{N(\ell)-1,1-\frac{\alpha'_{I2,\ell-1}}{|I^k_{\ell-1}|-1}} \cdot \frac{S_{ij}}{\sqrt{N(\ell)}}$ and $S^2_{ij} = \frac{1}{N(\ell)-1}\sum_{h=1}^{N(\ell)}(Y_{ih} - Y_{jh} - (\bar{Y}_i - \bar{Y}_j))^2$.

4. Report the point estimate $\bar{Y}_{(q_k B)}$ and CI for percentile $\mu_{[q_k B]}$,

$$\mathrm{CI}_k = \left[\min_{i\in I^k_\ell}\left(\bar{Y}_i - t_{N(\ell)-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{N(\ell)-1}}\right), \quad \max_{i\in I^k_\ell}\left(\bar{Y}_i + t_{N(\ell)-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{N(\ell)-1}}\right)\right]$$

If $|\mathrm{CI}_k| \leq L_k$, stop. Otherwise, continue.

5. Updating the number of iterations and the remaining screening significant level. Let

$$N'_{\max} \leftarrow \left(\frac{2\max_{i\in I^k_\ell}(S_i)\cdot t_{N(\ell)-1,1-\frac{\alpha_0}{2}}}{L_k/|I^k_\ell|}\right)^2$$
$$P^k_\ell = \max\left(\lceil \log_R \frac{N'_{\max}}{N(J_k+1)} \rceil, 0\right) + 1.$$

Update the screening error allowance,

$$\alpha^k_{I2} = \alpha^k_{I2} - \frac{\alpha^k_{I2}}{P^k_{\ell-1} + J_k - (\ell-1)} \text{ and } \alpha'_{I2,\ell} = \frac{\alpha^k_{I2}}{P^k_\ell + J_k - \ell} \tag{3}$$

6. Simulation. Assign additional $n_0 R^l(R-1)$ replications to samples in the surviving set I^k_ℓ, and run simulations, let $\ell \leftarrow \ell+1$. Loop back to Step 3.

3.2 Theoretical Support

In this section, we provide the theoretical support for the sequential procedure in Section 3.1. For notation simplification, we drop the index for the lower and upper percentiles, k. Theorem 1 shows our screening rule could guarantee the surviving set I includes the sample $[qB]$ with probability $(1-\alpha_I)$, where $q = q_1$ or q_2.

Theorem 1 Define the surviving set

$$I = \{i : \sum_{j \neq i} \mathbf{1}(\bar{Y}_j \leq \bar{Y}_i + W_{ij}) \geq qB - 1 \text{ and } \sum_{j \neq i} \mathbf{1}(\bar{Y}_j \geq \bar{Y}_i - W_{ij}) \geq B - qB\}$$

where, $W_{ij} = t_{n-1,1-\frac{\alpha_I}{B-1}} \cdot \frac{S_{ij}}{\sqrt{n}}$ and $S_{ij}^2 = \frac{1}{n-1}\sum_{h=1}^{n}(Y_{ih} - Y_{jh} - (\bar{Y}_i - \bar{Y}_j))^2$. Then, $\mathrm{P}([qB] \in I) \geq 1 - \alpha_I$.

Proof: Define

$$\begin{aligned} A_1 = \{&all \quad j \quad with \quad \mu_j < \mu_{[qB]} : \bar{Y}_j \leq \bar{Y}_{[qB]} + W_{[qB]j} \\ &and \quad all \quad j \quad with \quad \mu_j > \mu_{[qB]} : \bar{Y}_j \geq \bar{Y}_{[qB]} - W_{[qB]j}\} \end{aligned}$$

Since $A_1 \subseteq \{[qB] \in I\}$, we only need to show $P(A_1) \geq 1 - \alpha_I$. Define

$$\begin{aligned} A_2 &= \{all \quad j \quad with \quad \mu_j < \mu_{[qB]} : \bar{Y}_j \leq \bar{Y}_{[qB]} + W_{[qB]j}\} \\ A_3 &= \{all \quad j \quad with \quad \mu_j > \mu_{[qB]} : \bar{Y}_j \geq \bar{Y}_{[qB]} - W_{[qB]j}\}. \end{aligned}$$

We have

$$\begin{aligned} P({A_2}^c) &= P(\exists j \quad with \quad \mu_j < \mu_{[qB]} : \bar{Y}_j > \bar{Y}_{[qB]} + W_{[qB]j}) \\ &= P\Big(\bigcup_{\substack{j \neq [qB] \\ \mu_j < \mu_{[qB]}}} \{\bar{Y}_j > \bar{Y}_{[qB]} + W_{[qB]j}\} \Big) \\ &\leq \sum_{\substack{j \neq [qB] \\ \mu_j < \mu_{[qB]}}} P(\bar{Y}_j > \bar{Y}_{[qB]} + W_{[qB]j}) \\ &= \sum_{\substack{j \neq [qB] \\ \mu_j < \mu_{[qB]}}} P\left(\frac{\bar{Y}_j - \bar{Y}_{[qB]} - (\mu_j - \mu_{[qB]})}{S_{[qB]j}/\sqrt{n}} > \frac{W_{[qB]j} + (\mu_{[qB]} - \mu_j)}{S_{[qB]j}/\sqrt{n}} \right) \\ &= \sum_{\substack{j \neq [qB] \\ \mu_j < \mu_{[qB]}}} P\left(\frac{\bar{Y}_j - \bar{Y}_{[qB]} - (\mu_j - \mu_{[qB]})}{S_{[qB]j}/\sqrt{n}} > t_{n-1,1-\frac{\alpha_I}{B-1}} \right) \\ &= (qB-1) \cdot \frac{\alpha_I}{B-1}. \end{aligned}$$

Step 3 follows by the Bonferroni inequality again. Similarly, we could have $P(A_3^c) = (B - qB) \cdot \alpha_I/(B-1)$. Therefore, by Bonferroni inequality,

$$\begin{aligned} P([qB] \in I) \geq P(A_1) &\geq 1 - P({A_2}^c) - P({A_3}^c) \\ &= 1 - (qB-1)\frac{\alpha_I}{B-1} - (B - qB)\frac{\alpha_I}{B-1} = 1 - \alpha_I. \square \end{aligned}$$

After screening and estimation, we use $\bar{Y}_{(qB)}$ as the point estimate for $\mu_{[qB]}$. Then, we contruct a CI for the percentile estimation, denoted by

$$\mathrm{CI}_q = \left[\min_{i \in I} \left(\bar{Y}_i - t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}} \right), \max_{i \in I} \left(\bar{Y}_i + t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}} \right) \right].$$

Notice that to account for $(qB) \neq [qB]$, this CI does not center around $\bar{Y}_{(qB)}$. Theorem 2 shows that CI_k delivered by our sequential procedure in Section 3.1 could cover the true quantile $\mu_{[q_k B]}$ with probability $(1-\alpha)$ with $k = 1, 2$.

Theorem 2 Suppose $\alpha = 2\alpha_I + \alpha_0$. Then, $P(\mu_{[qB]} \in \mathrm{CI}_q) \geq 1-\alpha$.

Proof:

$$
\begin{aligned}
&P\Big(\mu_{[qB]} \geq \min_{i\in I}(\bar{Y}_i - t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}})\Big) \\
&\geq P\Big([qB] \in I, \quad \mu_{[qB]} \geq \min_{i\in I}(\bar{Y}_i - t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}})\Big) \\
&\geq P\Big([qB] \in I, \quad \mu_{[qB]} \geq \bar{Y}_{[qB]} - t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}}\Big) \\
&\geq 1 - P\Big([qB] \notin I\Big) - P\Big(\mu_{[qB]} < \bar{Y}_{[qB]} - t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}}\Big) \quad \geq 1-\alpha_I - \frac{\alpha_0}{2}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&P\Big(\mu_{[qB]} \leq \max_{i\in I}(\bar{Y}_i + t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}})\Big) \\
&\geq P\Big([qB] \in I, \quad \mu_{[qB]} \leq \max_{i\in I}(\bar{Y}_i + t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}})\Big) \\
&\geq P\Big([qB] \in I, \quad \mu_{[qB]} \leq \bar{Y}_{[qB]} + t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}}\Big) \\
&\geq 1 - P\Big([qB] \notin I\Big) - P\Big(\mu_{[qB]} > \bar{Y}_{[qB]} + t_{n-1,1-\frac{\alpha_0}{2}} \cdot \frac{S_i}{\sqrt{n-1}}\Big) \quad \geq 1-\alpha_I - \frac{\alpha_0}{2}.
\end{aligned}
$$

Therefore, we have $\mathrm{P}(\mu_{[qB]} \in \mathrm{CI}_q) \geq 1-\alpha_I - \frac{\alpha_0}{2} - \alpha_I - \frac{\alpha_0}{2} = 1-\alpha$. □

Theorem 3 supports the strategy to adaptively update the number of iterations P required in Phase II. By the Bonferroni inequality, we only need to show that the accumulated screening significance level consumed in Phase II is α_{I2}.

Theorem 3 The screening significant level spent in each iteration is $\alpha'_{I2,\ell}$ with $\ell = J_k, J_k+1, \ldots, J_k+P-1$ and it is updated by eq (3). The accumulated screening significance level consumed in Phase II is α_{I2}: $\sum_{\ell=J_k}^{J_k+P-1} \alpha'_{I2,\ell} = \alpha_{I2}$, where P denotes the number of iterations in Phase II.

Proof: We prove it by induction. Let r denote the number of times we update P. First consider the case with $r=1$. In Phase II, P is updated once to P_ℓ after the ℓth iteration. That means P and α'_{I2} remain constant until the ℓth iteration. All screening allowance consumed until the ℓth iteration is $\frac{\ell-J}{P}\alpha_{I2}$, and the remaining allowance is $\frac{P-\ell+J}{P}\alpha_{I2}$.

Two cases are discussed. Case 1 is: $P_\ell + J \leq \ell$. By our definition of P, that means every sample in I, say F_b, has an individual CI from $\bar{Y}_b$ with width less than $L/|I|$, the percentile CI_q must have the width no greater than L, the stopping criteria is satisfied and the procedure just stops. Case 2 is: $\ell < P_\ell + J < P + J$. In such case, we update α'_{I2} by eq (3): $\alpha'_{I2} = ((P+J-\ell)\alpha_{I2}/P)/(P_\ell + J - \ell)$. Hence, the total screening significance level spent in Stage II is: $(\ell - J)\alpha_{I2}/P + \alpha'_{I2} \cdot (P_\ell + J - \ell) = (\ell - J)\alpha_{I2}/P + (P+J-\ell)\alpha_{I2}/P = \alpha_{I2}$.

Assume the guarantee that the accumulated screening significant level spent in Phase II equals to α_{I2} holds when $r = r'$, where $r' \geq 1$ is an integer. We want to show this guarantee also holds when $r = r'+1$. When the maximum number of iterations P is updated $r'+1$ times, we can choose only update P at first r' chances, leaving the last update point constant. Then the problem is simply the base case with only one update point, which we have proved. Therefore, for $r = r'+1$, the guarantee also holds. □

4 EMPIRICAL STUDY

In this section we use an $M/M/1/50$ queue to illustrate the performance of our approach. Customer arrival rate is $\lambda^c = 6$ and service rate is $\mu^c = 10$. We are interested in the expected number of customers in the system.

To evaluate our approach, we pretend that the distributions for inter-arrival and service times are unknown. They are estimated from m inter-arrival times $A_1, A_2, \ldots, A_m$ and service times $S_1, S_2, \ldots, S_m$ with $A_i \overset{i.i.d.}{\sim} \exp(\lambda^c)$ and $S_i \overset{i.i.d.}{\sim} \exp(\mu^c)$. In our experiment, $m = 100$. To quantify the input uncertainty, we generate B bootstrapped empirical input distributions. We want to find a 99% CI quantifying the impact of input uncertainty on system mean response. Therefore, we want to efficiently estimate $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$ with $q_1 = 0.005$ and $q_2 = 0.995$.

The simulation of the queueing system starts with an empty system. The warmup length is 150 time units. Ideally, we would compare the percentile estimates obtained by our algorithm to the true values $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$. Since the true percentiles are not known, we use very long simulation runs to estimate them. To find an adequate run-length to estimate true values of $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$, we did a side experiment. Specifically, we have 10 macro-replications. In each macro-replication, we draw $m = 100$ independent real-world observations of interarrival and service times. Then, we generate 10 bootstrapped empirical distributions and use them to drive the simulations. We run simulatons with run lengths equal to 10^3, 2×10^3, 5×10^4, 10^5, 5×10^5 and 10^6 time units to estimate the average number of customers in the system. We chose 10^6 time units as the benchmark. The relative error defined as the maximum relative difference of the results obtained by each runlength compared to those obtained using 10^6 time units is recorded in Table 1. The runlength with 5×10^5 achieves maximum relative error 0.02. Balancing precision and computational cost, we choose 5×10^5 time units to estimate true percentiles $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$.

Table 1: The maximum absolute relative difference relative to the results by using the runlength equal to 10^6 time units.

run length	10^3	2×10^3	5×10^4	10^5	5×10^5
relative error	0.18	0.18	0.11	0.10	0.02

We compare the performance of our approach with the equal allocation under different settings. We run 100 macro-replications for each setting. In each macro-replication, we first obtain $m = 100$ data from $\exp(\mu^c)$ and $\exp(\lambda^c)$. Then, we generate $B = 500, 1000$ samples of bootstrapped input distributions to quantify input uncertainty. At each bootstrapped sample, we run simualtions. We set the runlength to be RL $= 10, 100$ time units. We set up the significance level $\alpha = 0.05$ with $\alpha_I = 0.015$ and $\alpha_0 = 0.02$. The number of initial replications is $n_0 = 30$. Let the desired width for percentile estimation CIs to be $L_k = 0.05\bar{Y}_{(q_k B)}$, where $\bar{Y}_{(q_k B)}$ with $k = 1, 2$ is the point estimates of percentiles based on the initial allocation.

We first study the screening performance of our approach. Figure 1 shows the number of samples in the surviving sets I^k with $k = 1, 2$ obtained from one macro-replication when $B = 500$ and RL=10. The horizontal axis gives the iteration index and the vertical axis gives the number of samples in surviving sets. We could observe that the majority of samples are screened out in a few iterations. It takes 13 iterations to precisely estimate $\mu_{[q_1 B]}$ and 18 iterations for $\mu_{[q_2 B]}$. Figure 2 illustrates the average of mean respones of surviving samples $\sum_{b \in I^k} \bar{Y}_b / |I^k|$ for $k = 1, 2$. Green horizontal lines represent the true percentile values. Clearly, the average mean response of surviving samples converges to the true percentiles $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$. Even though different macro-replications could require different number of iterations to get a precise percentile estimation, these plots represent the typical screening behavior for our sequential approach.

Figure 3 shows the histogram of the relative percentile estimation errors obtained by our approach and equal allocation under the same computation cost. Here, $B = 1000$ and RL=10 time units. The dark grey denotes the result from our approach and light grey denotes the result from the equal allocation. Figure 3 indicates that our approach provides more accurate estimation for $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$. In general, by gradually screening out samples with mean responses far from $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$, our sequential procedure can effectively use computational budget and precisely estimate the percentiles.

The first part of Table 2 show the mean, standard deviation (SD) and maximum of absolute relative error of percentile estimates obtained by our approach and equal allocation under the same simulation budget when $q_1 = 0.005$. The bottom of Table 2 shows the results for $q_2 = 0.995$. In general, our algorithm

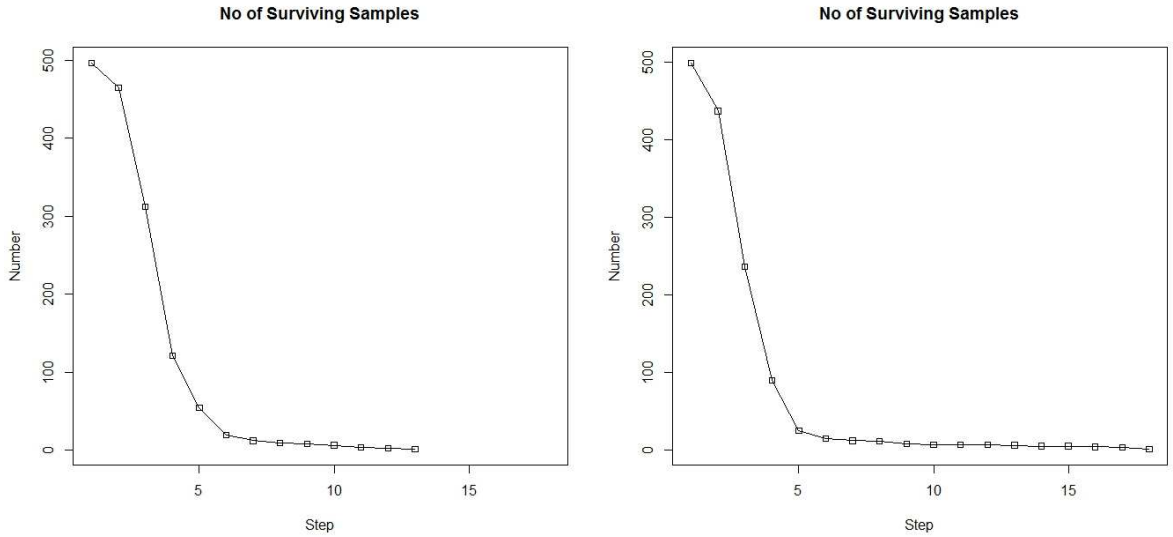

Figure 1: The number of surviving samples in I^k with $k = 1,2$ when $B = 500$.

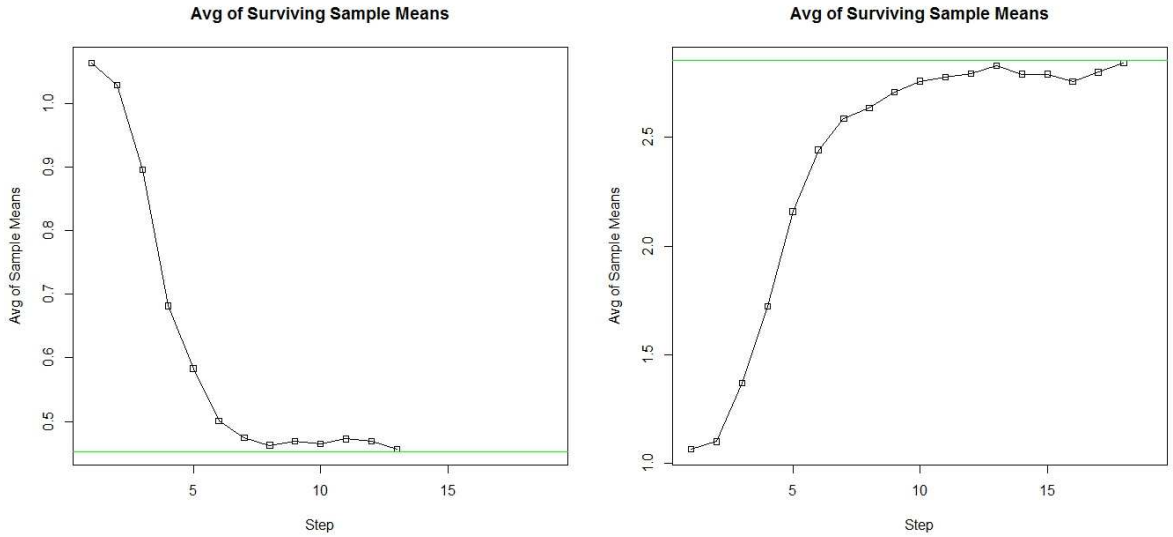

Figure 2: Average of mean responses for surviving samples in I^k with $k = 1,2$ when $B = 500$.

outperforms the equal allocation. It produces more accurate and stable percentile estimation. In some experiments, we observe that our procedure reports a point estimate extremely far away from the true one. This might be due to insufficient initial sample size (Lesnevski et al. 2008). One may increase n_0 in this case.

5 CONCLUSION

When we use nonparametric approaches to estimate both input model and parameter uncertainty, the input uncertainty is quantified by many samples of input distributions. Since each simulation run could be computationally expensive, it is desired to develop an approach that could efficiently propagate the input

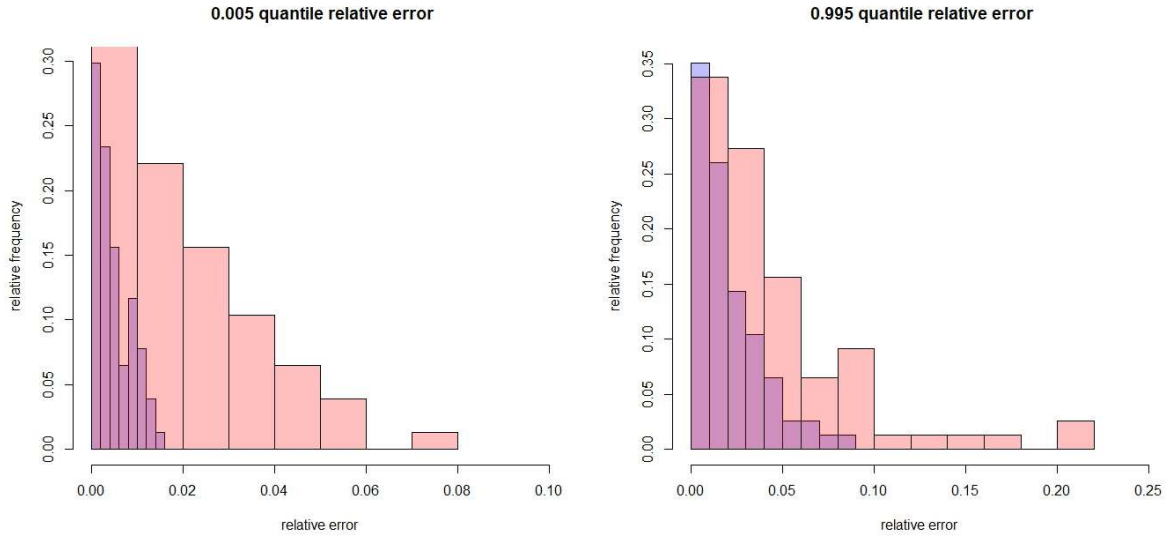

Figure 3: Histogram of relative error of percentile estimates when $B = 1000$.

Table 2: Relative estimation error of percentiles $\mu_{[q_1 B]}$ and $\mu_{[q_2 B]}$ obtained from the sequential experiment design and the equal allocation.

$\lambda^c = 6$, $q_1 = 0.005$	our procedure			direct equal allocation		
	mean	SD	maximum	mean	SD	maximum
$B = 500$, RL=10	0.71%	0.50%	2.08%	2.04%	1.63%	6.53%
$B = 500$, RL=100	0.94%	1.74%	12.44%	1.42%	1.14%	5.56%
$B = 1000$, RL=10	0.47%	0.39%	1.42%	1.86%	1.65%	7.90%
$B = 1000$, RL=100	0.70%	0.59%	3.20%	2.21%	1.56%	8.53%
$\lambda^c = 6$, $q_2 = 0.995$	our procedure			direct equal allocation		
	mean	SD	maximum	mean	SD	maximum
$B = 500$, RL=10	2.26%	2.02%	12.10%	5.21%	4.38%	15.54%
$B = 500$, RL=100	3.2%	7.94%	54.01%	3.84%	4.20%	18.22%
$B = 1000$, RL=10	2.13%	1.80%	8.34%	4.47%	4.30%	20.60%
$B = 1000$, RL=100	2.19%	2.19%	12.30%	6.27%	5.35%	27.40%

uncertainty to output mean. In this paper, we propose a sequential experiment design. It could gradually screen out samples of input distributions with mean responses far from the true percentiles and assign more computational budget to the important samples that contribute most to the percentile estimation. Our approach could efficiently estimate a percentile CI quantifying the impact of input uncertainty on the system mean performance. It is supported with theoretical analysis and an empirical study demonstrates our approach performs better than the classical equal allocation approach.

ACKNOWLEDGMENTS

The third author is grateful to the support by the National Science Foundation under Grant CMMI-1413790 and Grant CAREER CMMI-1453934, and Air Force Office of Scientific Research under Grant YIP FA-9550-14-1-0059.

REFERENCES

Barton, R. R., and L. W. Schruben. 2001. "Resampling Methods for Input Modeling". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 372–378: Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Boesel, J., B. L. Nelson, and S.-H. Kim. 2003. "Using Ranking and Selection to 'Clean Up' after Simulation Optimization". *Operations Research* 51 (5): 814–825.

Chick, S. E. 2001. "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty". *Operations Research* 49:744–758.

Lan, H., B. L. Nelson, and J. Staum. 2010. "A Confidence Interval Procedure for Expected Shortfall Risk Measurement via Two-Level Simulation". *Operations Research* 58 (5): 1481–1490.

Lesnevski, V., B. L. Nelson, and J. Staum. 2007. "Simulation of Coherent Risk Measures Based on Generalized Scenarios". *Management Science* 53 (11): 1756–1769.

Lesnevski, V., B. L. Nelson, and J. Staum. 2008. "An Adaptive Procedure for Estimating Coherent Risk Measures Based on Generalized Scenarios". *Journal of Computational Finance* 11 (4): 1–31.

Nelson, B. L., and D. Goldsman. 2001. "Comparisons with a Standard in Simulation Experiments". *Management Science* 47 (3): 449–463.

Xie, W., B. L. Nelson, and R. R. Barton. 2015. "Statistical Uncertainty Analysis for Stochastic Simulation". Working Paper, Department of Industrial and Systems Engineering, Rensselaer Polytechnic Institute, Evanston, NY.

AUTHOR BIOGRAPHIES

YUAN YI is a Ph.D. student in the Department of Industrial and Systems Engineering at the Rensselaer Polytechnic Institute. His research interests are in the interplay of simulation and supply chain, and risk management. His email address is yiy2@rpi.edu.

WEI XIE is an assistant professor in the Department of Industrial and Systems Engineering at Rensselaer Polytechnic Institute. Her research interests are in computer simulation, risk management and data analytics. Her email address is xiew3@rpi.edu and her web page is http://homepages.rpi.edu/~xiew3/.

ENLU ZHOU is an Assistant Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. She received the B.S. degree with highest honors in electrical engineering from Zhejiang University, China, in 2004, and received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2009. She is an recipient of AFOSR Young Investigator Award and NSF CAREER Award. Her research interests include stochastic control and simulation optimization, with applications towards financial engineering. Her email address is enlu.zhou@isye.gatech.edu and her web page is http://enluzhou.gatech.edu/.