

## DISTRIBUTIONAL INPUT UNCERTAINTY

Motong Chen  
Zhenyuan Liu  
Henry Lam

Department of Industrial Engineering and  
Operations Research  
Columbia University  
500 West 120th Street  
New York, NY 10027, USA

### ABSTRACT

The vast majority of the simulation input uncertainty literature focuses on estimating target output quantities that are real-valued. However, outputs of simulation models are random and real-valued targets essentially serve only as summary statistics. In this paper, we study the input uncertainty problem from a distributional view, namely we construct simultaneous confidence bands for the entire output distribution function. Our approach utilizes a novel test statistic that consists of the supremum of the sum of a Brownian bridge and a mean-zero Gaussian process whose covariance function is characterized by the influence function of the true output distribution function, which generalizes the Kolmogorov-Smirnov statistic to account for input uncertainty. We demonstrate how subsampling helps estimate the covariance function of the Gaussian process, thereby leading to an implementable estimation of the quantile of the test statistic and a statistically valid confidence band. Finally, we present some supporting numerical experiments.

### 1 INTRODUCTION

Stochastic simulation is widely used for operational decision-making and scientific modeling (Law et al. 2007; Banks et al. 2014). It consists of the generation of random input variates, fed into the system logic, to produce outputs that are subsequently analyzed for downstream tasks. In many real-world settings, the distributions that govern the input variates are unknown and need to be calibrated from external data. In this case, proper output analysis needs to account for both the Monte Carlo noises in running the simulation and also the input data noises that propagate to the outputs. This issue, which is beyond the handling of only the Monte Carlo noises in traditional output analysis, is known commonly as the input uncertainty problem in the simulation literature. For an overview on this topic, see, e.g., Henderson (2003), Barton (2012), Song et al. (2014), Lam (2016), Corlu et al. (2020), Barton et al. (2022), and Nelson (2013) Chapter 7.

More concretely, the goal of handling input uncertainty can often be cast as the construction of confidence intervals that are statistically valid under both sources of noises from Monte Carlo and input data. A closely related task is variance estimation, or the estimation of the different components of the variance that attribute to each source of noise. There are several major methods in the literature. The first is bootstrap resampling, which includes the variance bootstrap (Cheng and Holloand 1997; Song and Nelson 2015), quantile-based bootstrap (Barton and Schruben 1993; Barton and Schruben 2001), and enhancements such as the metamodel-assisted bootstrap (Barton et al. 2014) and subsampling (Lam and Qian 2021). The second line of methods uses the delta method directly which involves estimating gradient information of the model with respect to the inputs. This includes the two-point method (Cheng and Holland 1998; Cheng and Holland 2004), regression approach (Lin et al. 2015; Song and Nelson 2019), and subsampled

infinitesimal jackknife (Lam and Qian 2019; Lam et al. 2022). Other recent approaches include the use of robust optimization (Ghosh and Lam 2019) which relates to the empirical likelihood method (Lam and Qian 2016; Lam and Qian 2017). Finally, there is also a stream of work on Bayesian approaches that rely on posterior sampling (Chick 2001; Zouaoui and Wilson 2003; Zouaoui and Wilson 2004; Xie et al. 2014).

Despite the growing literature in input uncertainty, most, if not all, of the works have been focusing on the estimation of target quantities that are real-valued. However, in stochastic simulation the outputs are random (hence the Monte Carlo noise) and the real-valued target quantities often only serve as a summary statistic. For instance, suppose the simulation model represents a queuing system and we are interested in the waiting time or queue length at a certain time horizon. Then, in the input uncertainty literature, existing approaches would consider a summary target such as the expected waiting time. Such type of summary statistic is arguably convenient but, depending on our downstream decision-making tasks, could hide valuable information (in this example this could be the quantiles or tails of the waiting time distribution). Having an approach to understand the interplay of Monte Carlo and input uncertainty at the *distributional* level would therefore address this issue. In a sense, doing so would capture both the *epistemic* uncertainty, i.e., the uncertainty coming from the data noise, and *aleatory* uncertainty, i.e., the intrinsic uncertainty from the randomness in the stochastic model.

Our main contribution of this paper is to propose an approach to construct a *confidence band* for the entire output distribution that accounts for both Monte Carlo and input data noises. Our approach relies on a novel statistic that comprises of the supremum of a sum of a Brownian bridge and a mean-zero Gaussian process with a covariance structure governed by the so-called influence function of the target output distribution. In this statistic, the Brownian bridge signifies the Monte Carlo noises and resembles the limit of the Kolmogorov-Smirnov (KS) statistic. The mean-zero Gaussian process, on the other hand, signifies the input data noises, where the influence function that controls the covariance structure can be regarded as the derivative of the target output distribution with respect to the input parameters and carry sensitivity information like in the standard delta method. Our approach can be viewed as a generalization of the KS statistic, which can be used to construct confidence band under a single source of randomness from data, to the simulation input uncertainty setting where both epistemic and aleatory uncertainties need to be accounted for. In fact, we will illustrate that when there is input uncertainty, using the standard KS machinery to construct confidence bands for the output distribution would lead to a systematic under-coverage. The additional Gaussian process serves to inflate the variability to lift up the coverage by the right amount.

With the test statistic, we propose a subsampling approach to efficiently estimate the covariance function of the Gaussian process. Note that the influence function described above, while conceptually understandable, is difficult to use in practice because it requires computing essentially infinite-dimensional derivatives. Thus the bootstrap is used in lieu of the direct estimation of the influence function. However, standard variance bootstrap is inefficient because of the entanglement of the Monte Carlo and input data noises, which necessitates washing away the Monte Carlo noises via a large amount of simulation runs. This phenomenon has been pointed out in Lam and Qian (2021) even when only real-valued target quantities are considered. To address this computational challenge, we generalize the idea in Lam and Qian (2021) to devise a subsampling method where we draw a resample of smaller size than the original input data size, at the outer layer of a nested simulation scheme used to bootstrap estimate the covariance function.

The rest of this paper is as follows. Section 2 introduces the input uncertainty problem for the stochastic simulation output distribution. Section 3 briefly discusses the theory of the input uncertainty quantification for the distribution function. Section 4 then explains and presents the algorithms for estimating the covariance function of the Gaussian process and constructing the confidence bands. Finally Section 5 displays some supporting numerical results.

## 2 DISTRIBUTIONAL INPUT UNCERTAINTY PROBLEM

Consider a stochastic simulation model where each run outputs a target quantity, say  $Y$ , that follows a distribution  $Q$ . This simulation model can depend on  $m$  input distributions  $\underline{P} := (P_1, \dots, P_m)$ , so we may write

$Q = Q(\underline{P})$  which highlights that the output distribution depends on the input distributions. For example, think of a queuing system with two input distributions where  $P_1$  denotes the inter-arrival time distribution and  $P_2$  denotes the service time distribution. The quantity  $Y$  can be the average waiting time of the first 10 customers in a queue, which follows some distribution  $Q(\underline{P})$ .

Suppose we face input uncertainty for the stochastic simulation model, i.e., the true input distributions are unknown. Instead, we have a finite number of data points for each distribution, say  $\{X_{i,j} : i = 1, \dots, m, j = 1, \dots, n_i\}$  where  $X_{i,j}, j = 1, \dots, n_i$  are i.i.d. random variables from  $P_i$  and they are independent for different  $i$ 's. In order to obtain an approximate distribution of  $Y$ , we need to drive the simulation model using an estimate of  $\underline{P}$ . A natural choice is the empirical distribution  $\hat{\underline{P}} := (\hat{P}_1, \dots, \hat{P}_m)$  constructed from the data

$$\hat{P}_i(\cdot) = \frac{1}{n_i} \sum_{j=1}^{n_i} \delta_{X_{i,j}}(\cdot),$$

where  $\delta_x$  denotes the degenerate Dirac measure at  $x$ . Then, pretending that the empirical distributions are accurate enough, we use them to drive  $R_e$  simulation runs to obtain  $R_e$  output samples such as the average waiting time.

The output estimator we just mentioned contains statistical noises coming from both the external data used to calibrate the input distributions, and the Monte Carlo noise in running the simulation model. To conduct statistically valid inference for  $Y$ , we need to quantify both sources of noises. Existing literature mainly focuses on inferring a real-value performance measure, e.g., the expected value  $\mathbb{E}[Y]$ . In this paper, we are interested in the inference of the entire output distribution which, depending on the downstream decision-making tasks, could give more complete and useful information about the aleatory uncertainty of the output.

More precisely, our goal is to construct simultaneous confidence bands for the output distribution function. For convenience, let  $Q(t, \underline{P}) = \mathbb{P}_{\underline{P}}(Y \leq t)$  denote the true output distribution function where the subscript  $\underline{P}$  makes clear that  $Y$  is the output generated using the true unknown input distributions  $\underline{P}$ . We would like to find two functions  $L(t)$  and  $U(t)$  such that

$$\mathbb{P}(L(t) \leq Q(t, \underline{P}) \leq U(t) \ \forall t \in \mathbb{R}) \rightarrow 1 - \alpha, \tag{1}$$

where  $1 - \alpha$  is a pre-specified confidence level, e.g., 95%, where the limit is taken as  $n_i$  and  $R_e$  get large.

To get a sense of how one can construct these  $L(t)$  and  $U(t)$ , note that when we do not have noises in calibrating the input distribution (i.e., suppose we know the input distributions completely), the only uncertainty is from the Monte Carlo runs, and in this case we can use, for instance, the KS statistic. Suppose the true distribution function  $Q(t, \underline{P})$  is continuous in  $t$ . Let  $\hat{Q}(t, \underline{P})$  be the output empirical distribution constructed from  $R_e$  simulation runs driven by the input distribution  $\underline{P}$ , i.e.,

$$\hat{Q}(t, \underline{P}) = \frac{1}{R_e} \sum_{r=1}^{R_e} I(Y_r \leq t),$$

where  $Y_r, r = 1, \dots, R_e$  denote i.i.d. Monte Carlo outputs from the simulation model. Then the confidence band (1) is accomplished with

$$L(t) = \hat{Q}(t, \underline{P}) - \frac{q_{1-\alpha}}{\sqrt{R_e}}, \quad U(t) = \hat{Q}(t, \underline{P}) + \frac{q_{1-\alpha}}{\sqrt{R_e}},$$

where  $q_{1-\alpha}$  is the  $(1 - \alpha)$ -quantile of the asymptotic of the KS statistic, namely  $\sup_{s \in [0,1]} |BB(s)|$  where  $BB(\cdot)$  is a standard Brownian bridge. The above  $L$  and  $U$  are piece-wise constant and can be computed readily. Thus, when there is no input uncertainty, the problem straightforwardly reduces to the classical goodness-of-fit problem.

However, when the input distributions need to be calibrated from external data, the above  $L$  and  $U$  would *under-cover* the true output distribution  $Q(t, \underline{P})$ , because they have ignored this additional source of input noise. To remedy this, we need to properly inflate the variability of the considered statistic. This is discussed next.

### 3 A STATISTIC TO QUANTIFY DISTRIBUTIONAL INPUT UNCERTAINTY

We describe our procedure to construct confidence band for the entire output distribution. Along the way, we explain the statistical underpinnings and briefly discuss the mathematical theory.

Let  $n = (1/m)\sum_{i=1}^m n_i$  be the average data size. We assume the data sizes are balanced, i.e.,

$$\lim_{n \rightarrow \infty} n_i/n = \beta_i > 0, \forall i = 1, \dots, m. \tag{2}$$

Then we have the following approximation results when  $n$  and  $R_e$  are large:

$$\hat{Q}(t, \hat{P}) - Q(t, \underline{P}) = [Q(t, \hat{P}) - Q(t, \underline{P})] + [\hat{Q}(t, \hat{P}) - Q(t, \hat{P})] \stackrel{d}{\approx} \frac{\mathbb{G}(t)}{\sqrt{n}} + \frac{BB(Q(t, \underline{P}))}{\sqrt{R_e}}, \tag{3}$$

and consequently

$$\sup_{t \in \mathbb{R}} |\hat{Q}(t, \hat{P}) - Q(t, \underline{P})| \stackrel{d}{\approx} \sup_{t \in \mathbb{R}} \left| \frac{\mathbb{G}(t)}{\sqrt{n}} + \frac{BB(Q(t, \underline{P}))}{\sqrt{R_e}} \right|, \tag{4}$$

where  $\mathbb{G}(t)/\sqrt{n}$  is an approximation of  $Q(t, \hat{P}) - Q(t, \underline{P})$  and  $BB(Q(t, \underline{P}))/\sqrt{R_e}$  is an approximation of  $\hat{Q}(t, \hat{P}) - Q(t, \hat{P})$ . Here,  $BB(Q(t, \underline{P}))$  is a Brownian bridge with changed time scale  $Q(t, \underline{P}), t \in \mathbb{R}$ , and  $\mathbb{G}(t)$  is a Gaussian process independent of  $BB(Q(t, \underline{P}))$  with domain on  $\mathbb{R}$ , mean zero, and covariance structure given by

$$\text{Cov}(\mathbb{G}(t), \mathbb{G}(s)) = \sum_{i=1}^m \frac{1}{\beta_i} \text{Cov}(IF_i(t, X_i; \underline{P}), IF_i(s, X_i; \underline{P})),$$

where  $X_i$  follows the distribution  $P_i$  and  $IF_i(t, x; \underline{P})$  is the so-called influence function of  $Q(t, P)$  for the  $i$ -th input distribution defined as

$$IF_i(t, x; \underline{P}) = \lim_{\varepsilon \downarrow 0} \frac{Q(t, P_1, \dots, P_{i-1}, (1-\varepsilon)P_i + \varepsilon\delta_x, P_{i+1}, \dots, P_m) - Q(t, \underline{P})}{\varepsilon}.$$

The influence function  $IF_i(t, x; \underline{P})$  signifies the functional gradient or the Gateaux derivative of  $Q(t, P)$  with respect to the  $i$ -th input distribution  $P_i$  (see Wasserman (2006) Section 2.3 and Serfling (2009) Chapter 6). It has mean zero under  $P_i$ , i.e.,  $\mathbb{E}_{P_i}[IF_i(t, X_i; \underline{P})] = 0$ . Now we briefly explain why the approximation in (3) holds. For the Brownian bridge part, when  $n$  is large, we have  $\hat{P} \approx \underline{P}$ . Therefore,  $\hat{Q}(t, \hat{P}) - Q(t, \hat{P}) \approx \hat{Q}(t, \underline{P}) - Q(t, \underline{P}) \stackrel{d}{\approx} BB(Q(t, \underline{P}))/\sqrt{R_e}$  where the second approximation is due to the Donsker Theorem (see, e.g., Van der Vaart (2000) Theorem 19.3). Then we explain how to obtain the Gaussian process  $\mathbb{G}(t)$ , especially how we obtain the form of its covariance function. Since the influence function can serve as the first order Taylor expansion of the statistical functional, we can obtain

$$Q(t, \hat{P}) - Q(t, \underline{P}) \approx \sum_{i=1}^m \int IF_i(t, x; \underline{P}) d(\hat{P}_i(x) - P_i(x)) = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} IF_i(t, X_{i,j}; \underline{P}),$$

where the equality follows from the mean zero property of  $IF_i(t, x; \underline{P})$  under  $P_i$ . Now using the independence among  $X_{i,j}$ 's, balanced data assumption in (2) and multivariate central limit theorem, we can easily show that

$$\sqrt{n} \left( \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} IF_i(t, X_{i,j}; \underline{P}) \right) \stackrel{f.d.}{\rightarrow} \mathbb{G}(t), t \in \mathbb{R},$$

where  $\stackrel{f.d.}{\rightarrow}$  means any finite dimensional distribution of the left hand side (LHS) process converges in distribution to the corresponding finite dimensional distribution of the right hand side process. The

argument is rigorized via proper checking of tightness, control of remainder terms in the linearization, and asymptotic independence between the components elicited by the Monte Carlo and input data noises.

Based on the approximation in (4), if we can estimate the  $(1 - \alpha)$ -quantile of

$$\sup_{t \in \mathbb{R}} \left| \frac{\mathbb{G}(t)}{\sqrt{n}} + \frac{BB(Q(t, \underline{P}))}{\sqrt{R_e}} \right|$$

say  $\tilde{q}_{1-\alpha}$ , then a statistically valid pair of confidence bands is given by

$$L(t) = \hat{Q}(t, \hat{P}) - \tilde{q}_{1-\alpha}, \quad U(t) = \hat{Q}(t, \hat{P}) + \tilde{q}_{1-\alpha}, \tag{5}$$

which are readily computable since they are piece-wise constant.

#### 4 CONFIDENCE BAND CONSTRUCTION

In view of the approximation (4), in order to construct a  $(1 - \alpha)$ -level confidence band for  $Q(t, \underline{P})$ , the key is to estimate the  $(1 - \alpha)$ -quantile of the random variable

$$\sup_{t \in \mathbb{R}} \left| \frac{\mathbb{G}(t)}{\sqrt{n}} + \frac{BB(Q(t, \underline{P}))}{\sqrt{R_e}} \right|$$

Exact simulation is hard to implement since it contains the supremum over the whole real line. To this end, we discretize  $t$  on a grid over the real line, say  $t_1, \dots, t_k$  and replace the supremum over  $t \in \mathbb{R}$  with the maximum over  $t \in \{t_1, \dots, t_k\}$ , i.e.,

$$\max_{i=1, \dots, k} \left| \frac{\mathbb{G}(t_i)}{\sqrt{n}} + \frac{BB(Q(t_i, \underline{P}))}{\sqrt{R_e}} \right| \stackrel{d}{\approx} \sup_{t \in \mathbb{R}} \left| \frac{\mathbb{G}(t)}{\sqrt{n}} + \frac{BB(Q(t, \underline{P}))}{\sqrt{R_e}} \right|. \tag{6}$$

If we simulate the quantile of the LHS random variable in (6) and construct the confidence band like (5), we actually end up giving a statistically valid confidence band only at these discrete values. But when the discretization scale is refined enough, this discretized confidence band becomes an accurate approximation and its natural extrapolation should cover the true distribution function in  $\mathbb{R}$  with probability close to the nominal level. Therefore, we next aim at estimating the quantile of the LHS random variable in (6).  $BB(Q(t_i, \underline{P}))$  can be easily approximated by  $BB(\hat{Q}(t_i, \hat{P}))$ . But approximating  $\mathbb{G}(t_i)$  or equivalently  $\mathbb{G}(t_i)/\sqrt{n}$  is more complicated. Note that  $\mathbb{G}(t)$  is a mean-zero Gaussian process. In order to simulate  $\mathbb{G}(t)$ , it suffices to know its covariance function. However, the covariance function is hard to obtain since it involves the influence function which is not easy to compute in general. In this paper, we resort to the bootstrap to estimate the covariance function. In particular we will use subsampling to relieve the computation burden.

##### 4.1 Subsampling to Estimate Covariance Function

Recall that we have  $n_i$  data points  $X_{i,j}, j = 1, \dots, n_i$  for the  $i$ -th input distribution  $P_i$ . We specify a subsample ratio  $\theta \in (0, 1]$ , and compute  $s_i = \lfloor \theta n_i \rfloor, \forall i = 1, \dots, m$  as the subsample sizes. Based on the approximation (3), we have

$$nCov(Q(t, \hat{P}), Q(s, \hat{P})) \approx Cov(\mathbb{G}(t), \mathbb{G}(s)). \tag{7}$$

At this point, we could consider using the bootstrap to approximate the LHS. However, note that the bootstrap requires a nested simulation where in the outer layer, we resample the input distributions and, in the inner layer, we use the resample distributions to drive simulation runs to obtain the output distributions. The issue with a naive use of the bootstrap in this context is that the total required simulation effort can be huge (see Lam and Qian (2021) Section 2.3), essentially an amount that can wash away the Monte Carlo noises. More concretely, as shown in Lam and Qian (2021), this amount is of larger order than the data

size itself. To remedy this computational demand, we utilize a generalization of Lam and Qian (2021) to devise a subsampling procedure. This procedure aims to exploit the form of the covariance in (7) that scales with the data size  $n$  and turns out to reduce the computational demand from a larger order than  $n$  to an order that is independent of  $n$ .

To explain our procedure, let  $X_{i,1}^*, \dots, X_{i,s_i}^*$  be random resamples uniformly drawn with replacement from  $X_{i,1}, \dots, X_{i,n_i}$  and let  $\hat{P}_{i,s_i}^* = (1/s_i) \sum_{j=1}^{s_i} \delta_{X_{i,j}^*}$  be the corresponding subsample empirical distribution. Denote  $\hat{P}_\theta^* = (\hat{P}_{1,s_1}^*, \dots, \hat{P}_{m,s_m}^*)$ . Using the rationale of the bootstrap, we can approximate the LHS of (7) as follows

$$\theta n \text{Cov}_*(Q(t, \hat{P}_\theta^*), Q(s, \hat{P}_\theta^*)) \approx n \text{Cov}(Q(t, \hat{P}), Q(s, \hat{P})) \approx \text{Cov}(\mathbb{G}(t), \mathbb{G}(s)),$$

and consequently we obtain

$$\theta \text{Cov}_*(Q(t, \hat{P}_\theta^*), Q(s, \hat{P}_\theta^*)) \approx \text{Cov}\left(\frac{\mathbb{G}(t)}{\sqrt{n}}, \frac{\mathbb{G}(s)}{\sqrt{n}}\right), \quad (8)$$

where  $\text{Cov}_*$  denotes the covariance conditional on the input data. Note that we rescale the covariance in the LHS of (8) by  $\theta$  because we are estimating a quantity of order  $1/n$ . When we use the smaller average data size  $s = (1/m) \sum_{i=1}^m s_i$  in our subsampling, the estimated quantity we get is of order  $1/s = 1/\theta n$ . The rescaling factor  $\theta$  thus helps us retain correct estimation for the original quantity of interest.

To implement the approximation in (8), for  $b = 1, \dots, B$ , we uniformly sample with replacement from  $X_{i,1}, \dots, X_{i,n_i}$   $s_i$  times to obtain  $X_{i,1}^{*b}, \dots, X_{i,s_i}^{*b}$ . Then we construct the subsample empirical distribution  $\hat{P}_{i,s_i}^{*b} = (1/s_i) \sum_{j=1}^{s_i} \delta_{X_{i,j}^{*b}}$  and let  $\hat{P}_\theta^{*b} = (\hat{P}_{1,s_1}^{*b}, \dots, \hat{P}_{m,s_m}^{*b})$ . We use  $\hat{P}_\theta^{*b}$  to drive  $R$  independent simulation runs to obtain  $R$  outputs  $Y_r^{*b}, r = 1, \dots, R$  and then construct the empirical output distribution as follows

$$\hat{Q}(t, \hat{P}_\theta^{*b}) = \frac{1}{R} \sum_{r=1}^R I(Y_r^{*b} \leq t).$$

Then we use the ANOVA-based estimator to estimate the covariance in (8):

$$\theta \left( \frac{1}{B-1} \sum_{b=1}^B (\hat{Q}(t, \hat{P}_\theta^{*b}) - \bar{Q}(t)) (\hat{Q}(s, \hat{P}_\theta^{*b}) - \bar{Q}(s)) - \frac{1}{BR(R-1)} \sum_{b=1}^B \sum_{r=1}^R (I(Y_r^{*b} \leq t) - \hat{Q}(t, \hat{P}_\theta^{*b})) (I(Y_r^{*b} \leq s) - \hat{Q}(s, \hat{P}_\theta^{*b})) \right),$$

where  $\bar{Q}(\cdot) = (1/B) \sum_{b=1}^B \hat{Q}(\cdot, \hat{P}_\theta^{*b})$  is the sample mean of  $\hat{Q}(\cdot, \hat{P}_\theta^{*b})$ . We can show this estimator is unbiased for estimating the LHS of (8) conditional on the input data. Algorithm 1 summarizes the procedure to estimate the covariance matrix of  $\mathbb{G}(t)/\sqrt{n}$  for  $t = t_1, \dots, t_k$ . Importantly, this estimator allows to use less simulation effort than the standard bootstrap by choosing  $s$  to be much smaller than  $n$ , while still retaining meaningful statistical consistency.

## 4.2 Confidence Band Construction for the Output Distribution Function

Once we obtain the estimated covariance matrix  $V$  returned by Algorithm 1, we can simulate a  $k$ -dimensional Gaussian vector  $(Z_1, \dots, Z_k) \sim N(0, V)$  as the estimator for  $(\mathbb{G}(t_1)/\sqrt{n}, \dots, \mathbb{G}(t_k)/\sqrt{n})$ . We can also simulate  $(W_1, \dots, W_k) \stackrel{d}{=} (BB(\hat{Q}(t_1, \hat{P})), \dots, BB(\hat{Q}(t_k, \hat{P})))$  as the estimator for  $(BB(Q(t_1, P)), \dots, BB(Q(t_k, P)))$ . Since the processes  $\mathbb{G}(t)$  and  $BB(Q(t, P))$  are independent, the simulation procedures of  $(Z_1, \dots, Z_k)$  and  $(W_1, \dots, W_k)$  should also be independent, conditional on the input data. Now in view of the approximation (6), we can deduce that

$$\max_{j=1, \dots, k} \left| Z_j + \frac{W_j}{\sqrt{R_e}} \right| \stackrel{d}{\approx} \sup_{t \in \mathbb{R}} \left| \frac{\mathbb{G}(t)}{\sqrt{n}} + \frac{BB(Q(t, P))}{\sqrt{R_e}} \right|. \quad (9)$$

---

**Algorithm 1** Subsampling to Estimate Covariance Matrix

---

**Inputs:**  $B, R$ , observed inputs  $\{X_{i,j} : i = 1, \dots, m, j = 1, \dots, n_i\}$ , subsample ratio  $\theta$ , a grid  $\{t_l : l = 1, \dots, k\}$

Compute  $s_i = \lfloor \theta n_i \rfloor, i = 1, \dots, m$

**for**  $b = 1, \dots, B$  **do**

For each  $i = 1, \dots, m$  draw a subsample  $X_i^{*b} = \{X_{i,1}^{*b}, \dots, X_{i,s_i}^{*b}\}$  uniformly with replacement from the input  $\{X_{i,1}, \dots, X_{i,n_i}\}$ . Form resample empirical distribution  $\hat{P}_{i,s_i}^{*b}$

Let  $\hat{P}_\theta^{*b} = (\hat{P}_{1,s_1}^{*b}, \dots, \hat{P}_{m,s_m}^{*b})$

**for**  $r = 1 \dots R$  **do**

Simulate  $\hat{Y}_r^{*b} = f(\hat{P}_\theta^{*b})$  from the stochastic simulation model  $f$

**end for**

Form empirical distribution for  $\hat{Q}(\cdot, \hat{P}_\theta^{*b})$  at points  $\{t_l : l = 1, \dots, k\}$ :

$\hat{Q}(t_l, \hat{P}_\theta^{*b}) = \frac{1}{R} \sum_{r=1}^R I(Y_r^{*b} \leq t_l), l = 1, \dots, k$

**end for**

$\bar{Q}(t_l) = \frac{1}{B} \sum_{b=1}^B \hat{Q}(t_l, \hat{P}_\theta^{*b}),$  where  $l = 1, \dots, k$

Compute the estimated covariance matrix  $V \in \mathbb{R}^{k \times k}$  at points  $\{t_l : l = 1, \dots, k\}$  from  $B$  empirical distributions constructed from the subsampled runs:

$$V_{i,j} = \theta \left( \frac{1}{B-1} \sum_{b=1}^B (\hat{Q}(t_i, \hat{P}_\theta^{*b}) - \bar{Q}(t_i)) (\hat{Q}(t_j, \hat{P}_\theta^{*b}) - \bar{Q}(t_j)) - \frac{1}{BR(R-1)} \sum_{b=1}^B \sum_{r=1}^R (I(Y_r^{*b} \leq t_i) - \hat{Q}(t_i, \hat{P}_\theta^{*b})) (I(Y_r^{*b} \leq t_j) - \hat{Q}(t_j, \hat{P}_\theta^{*b})) \right)$$

**return**  $V$

---

Thus, the quantile of the LHS distribution in (9) can be taken as the approximation of the quantile of the RHS distribution in (9).

To estimate the quantile of the LHS distribution in (9), we independently repeat the above  $R_q$  times and output the  $(1 - \alpha)$ -quantile of

$$\max_{j=1, \dots, k} \left| Z_j^i + \frac{W_j^i}{\sqrt{R_e}} \right|, i = 1, \dots, R_q$$

as  $\tilde{q}_{1-\alpha}$ . Based on (4), a statistically valid pair of confidence bands on the discretized time grid is given by

$$L(t_l) = \hat{Q}(t_l, \hat{P}) - \tilde{q}_{1-\alpha}, \quad U(t_l) = \hat{Q}(t_l, \hat{P}) + \tilde{q}_{1-\alpha}, \quad \forall l = 1, \dots, k,$$

where  $\hat{Q}(t_l, \hat{P}) = (1/R_e) \sum_{r=1}^{R_e} I(Y_r \leq t_l)$  and  $Y_r, r = 1, \dots, R_e$  are i.i.d. outputs of the simulation model driven by the empirical input distribution  $\hat{P}$ . We summarize our confidence band construction procedure in Algorithm 2.

## 5 NUMERICAL EXPERIMENTS

We conduct experiments on an M/M/1 queue and report our experimental findings. This system has 2 input distributions  $\underline{P} = (P_1, P_2)$  where  $P_1$  denotes the inter-arrival time distribution and  $P_2$  denotes the service time distribution, for which we have  $n_1$  and  $n_2$  i.i.d. data points available respectively. We consider two ground truths, one with arrival rate 0.5 and service rate 1, and another with arrival rate 0.9 and service

---

**Algorithm 2** Construction of Confidence Band for the Output Distribution Function

---

**Inputs:**  $R_e, R_q, \alpha$ , observed inputs  $\{X_{i,j} : i = 1, \dots, m, j = 1, \dots, n_i\}$ , a grid  $\{t_l : l = 1, \dots, k\}$ , estimated covariance matrix  $V$

**for**  $r = 1, \dots, R_e$  **do**

$Y_r = f(\hat{P})$  from the stochastic simulation model  $f$

**end for**

Form empirical distribution for  $\hat{Q}(\cdot, \hat{P})$  at points  $\{t_l : l = 1, \dots, k\}$ :

$\hat{Q}(t_l, \hat{P}) = \frac{1}{R_e} \sum_{r=1}^{R_e} I(Y_r \leq t_l), l = 1, \dots, k$

**for**  $i = 1, \dots, R_q$  **do**

Simulate  $Z^i = (Z_1^i, \dots, Z_k^i)$  from the  $k$ -dimensional Gaussian random vector  $N(0, V)$

Simulate  $W^i = (W_1^i, \dots, W_k^i)$  as a time transformed discretized brownian bridge  $BB(\hat{Q}(t_l, \hat{P})), l = 1, \dots, k$

$e_i = \max_{j=1, \dots, k} \left| Z_j^i + \frac{W_j^i}{\sqrt{R_e}} \right|$

**end for**

Find  $\tilde{q}_{1-\alpha} = (1 - \alpha)$ -quantile of  $e_1, \dots, e_{R_q}$

**return**  $L(\cdot) = \hat{Q}(\cdot, \hat{P}) - \tilde{q}_{1-\alpha}$  and  $U(\cdot) = \hat{Q}(\cdot, \hat{P}) + \tilde{q}_{1-\alpha}$

---

rate 1. To satisfy the balanced data condition (2), we choose  $n_2 = 2n_1$  in both settings and we only report  $\min_i n_i$  for convenience. Our goal is to estimate the entire distribution of the average waiting time of the first 10 customers entering the system, more precisely the 95%-level confidence band for the distribution of  $\bar{W}_{10}$ , the average waiting time of the first 10 customers. i.e.,  $Q(t, \underline{P}) = \mathbb{E}_{\underline{P}}[\mathbf{1}\{\bar{W}_{10} \leq t\}]$ .

In the experiments, we validate the performance of our procedure and investigate the impact of procedural parameters: subsample ratio  $\theta$ , subsampling split between outer loop  $B$  and inner loop  $R$  when estimating the covariance matrix by Algorithm 1, and provide practical guidelines for choosing these parameters. Moreover, we investigate the budget split between Algorithms 1 and 2. Note that Algorithm 1 consists of  $R_v = BR$  simulation model runs for the estimation of the covariance matrix while Algorithm 2 consists of  $R_e$  simulation model runs for the estimation of  $\hat{Q}(\cdot, \hat{P})$ . Suppose we are given a total budget of  $N$  simulation runs (the total number of times we can run the stochastic simulation model) and we allocate  $R_v$  to estimate the covariance matrix  $V$  and  $R_e = N - R_v$  to compute the CDF estimator  $\hat{Q}(\cdot, \hat{P})$ . We investigate the optimal split between  $R_v$  and  $R_e$ . Here we do not take into account the  $R_q$  runs in Algorithm 2 because it does not require running the simulation model, but rather just the generation of normal vectors and thus the runtime is relatively negligible. The performances of different configurations are measured by the coverage probability as well as the width of the confidence band.

### 5.1 Performance Validation and Guidelines for Algorithmic Configuration

We validate and examine the performance of a wide range of selections of  $\theta, B, R$ . For each example and for input data size ranging from 300 to 2000, we test different combinations of  $\theta, B, R$ . The subsample size  $\theta \min_i n_i \in \{5, 30, 60, 90\}$  and the budge allocation  $(B, R) \in \{(10, 100), (12, 80), (16, 60), (25, 40), (40, 25), (200, 5)\}$  when  $R_v = 1000$  and  $B = \lfloor R_v/R \rfloor$ . We also consider the case when we have a smaller budget  $R_v = 500$  and test the corresponding budget allocation  $(B, R) \in \{(5, 100), (6, 80), (8, 60), (12, 40), (20, 25), (100, 5)\}$ .

To calculate the coverage probability of the estimated confidence band, we first run the simulation model with the true input distributions (exponential distributions in the M/M/1 case) 100000 times to approximate the true output distribution. When the number of simulation runs is large enough, the empirical output distribution we constructed with the true input distributions is a good enough approximation of the true output distribution. We then perform 200 independent runs of the confidence band construction procedure and report the percentage of runs in which the constructed confidence band covers the true output distribution.



We first establish a guideline for choosing the optimal allocation between  $B$  and  $R$  given a subsample size  $\theta \min_i n_i$ . As shown in Figures 1-3, for different  $\theta \min_i n_i$ , the optimal  $R$  is generally between  $(2/3)\theta \min_i n_i$  and  $\theta \min_i n_i$  since larger  $R$  results in undercoverage while smaller  $R$  produces a much wider confidence band which is less informative and leads to overcoverage. As observed from Figures 1-3, when  $R$  is below  $(2/3)\theta \min_i n_i$ , most of the experiments have 100% coverage probability and thus these confidence bands are too conservative. To verify that this selection of  $R$  is independent of the simulation budget  $R_v$ , we also take a look at  $R_v = 500$  in Figure 4 and observe a similar optimal choice of  $R$ . We therefore conclude that  $R \in [(2/3)\theta \min_i n_i, \theta \min_i n_i]$  is recommended as a practical choice in budget allocation.

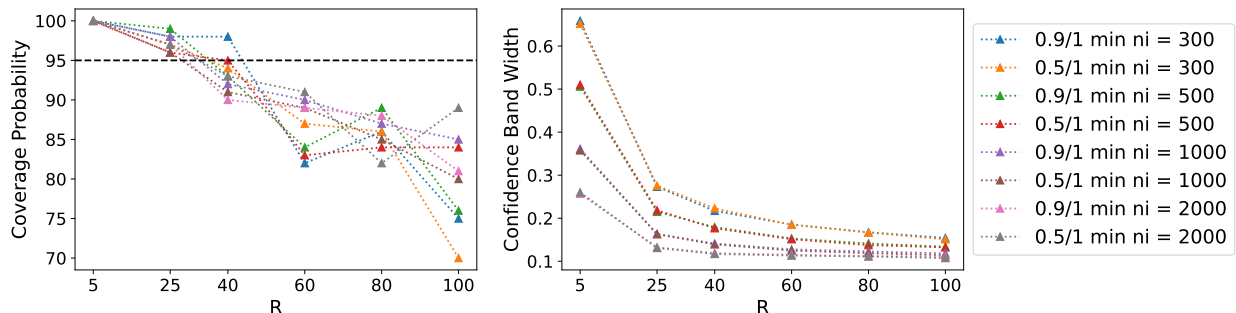


Figure 1: Coverage probability and confidence band width for different  $R$  with  $R_v = 1000$  and  $\theta \min_i n_i = 30$ .

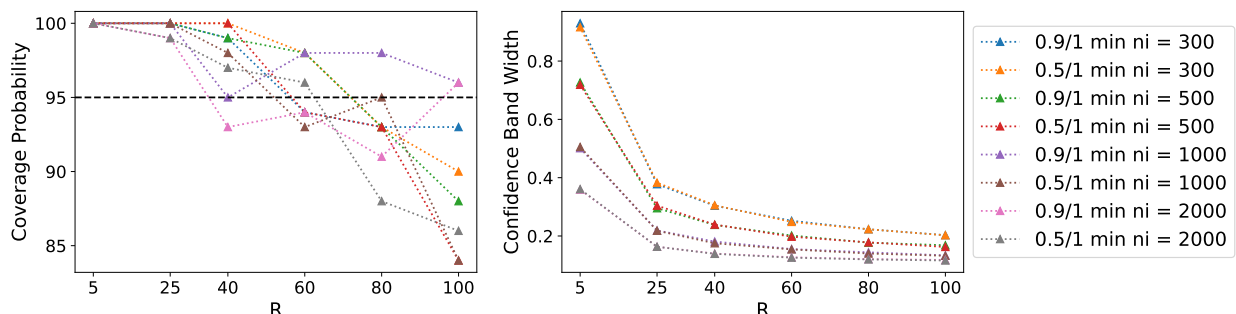


Figure 2: Coverage probability and confidence band width for different  $R$  with  $R_v = 1000$  and  $\theta \min_i n_i = 60$ .

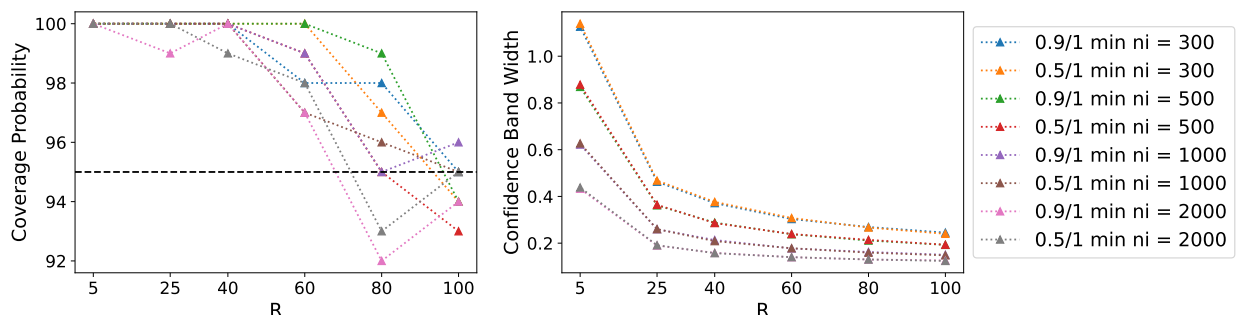


Figure 3: Coverage probability and confidence band width for different  $R$  with  $R_v = 1000$  and  $\theta \min_i n_i = 90$ .

We then investigate the optimal subsample size. We compare the performance as  $\theta \min_i n_i$  varies with optimally tuned  $R = \theta \min_i n_i$ . In Figure 5, we see that almost all cases have a good coverage probability close to the nominal level 95%. A small subsample size such as 5 gives us a slightly wider confidence band

and lower coverage probability in most cases compared to moderately larger subsample size. Choosing a moderate subsample size such as 30 gives us a fairly accurate coverage with narrower confidence band. Increasing the subsample size from 30 results in a wider confidence band. We therefore recommend choosing subsample size  $\theta \min_i n_i \approx 30$ .

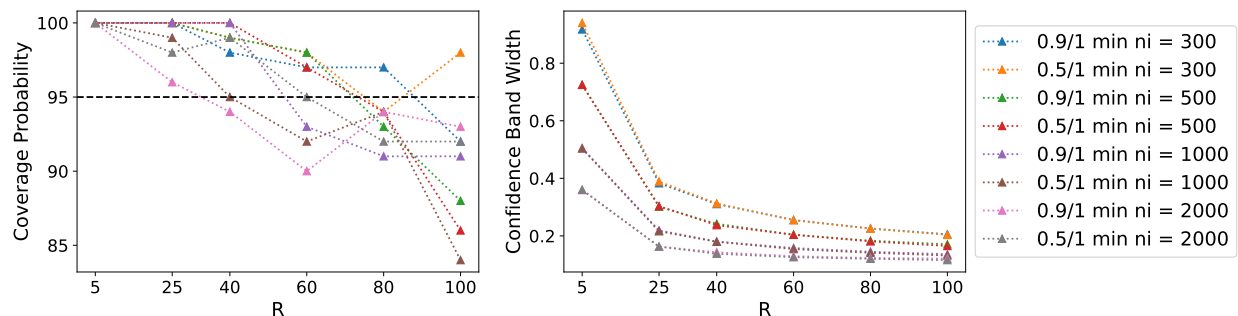


Figure 4: Coverage probability and confidence band width for different  $R$  with  $R_v = 500$  and  $\theta \min_i n_i = 60$ .

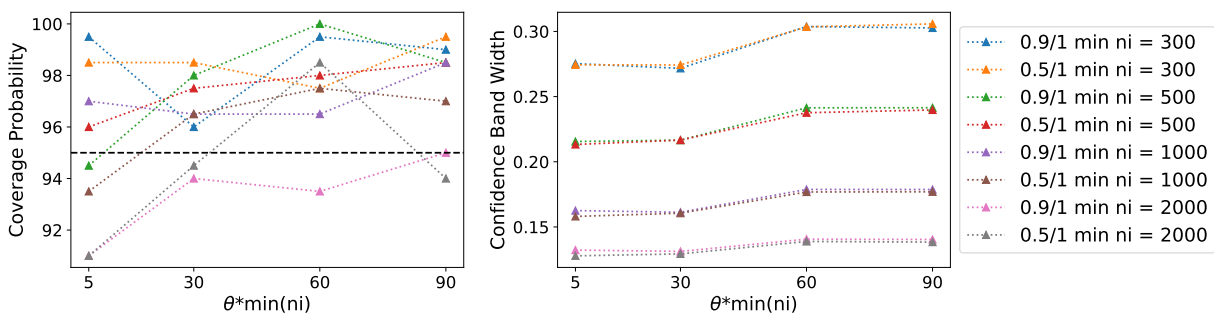


Figure 5: Coverage probability and confidence band width for different  $\theta \min_i n_i$  with optimally tuned  $R$ .

## 5.2 Budget Split between Covariance Matrix Estimation and Distribution Function Construction

We also investigate the optimal split between  $R_v$  and  $R_e$  given a total budget  $N$ . Assume that we have total budget  $N = 1000$  to be allocated to  $R_v$  and  $R_e$ . Based on the practical guidelines for choosing  $R$  and  $\theta \min_i n_i$  in Section 5.1, we fix the optimal subsample size  $\theta \min_i n_i = 30$  and optimal  $R = \theta \min_i n_i = 30$  in all cases and only vary  $R_v$ . We try six different splits  $R_v = 90, 180, 300, 510, 690, 900$  (accordingly,  $R_e = 1000 - R_v$ ). From Figure 6, we see that  $R_v = 180$  or  $R_v = 300$  is generally a good choice of budget assigned to covariance matrix estimation in all cases. Such an allocation gives us the tightest confidence band while having the proper coverage probability. Therefore, we suggest that given a fixed total budget  $N$ , we assign 20%-30% to the estimation of covariance matrix and allocate the rest to estimate the output distribution function. This practical guideline is also verified when the total budget is increased to  $N = 1500$ , from which we observe that the optimal  $R_v$  is around 300-500 (we do not present a figure for  $N = 1500$  due to page limit).

## ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Science Foundation under grants CAREER CMMI-1834710 and IIS-1849280.

## REFERENCES

Banks, J., J. Carson, B. Nelson, and D. Nicol. 2014. *Discrete Event System Simulation*. Fifth ed. Harlow: Pearson.

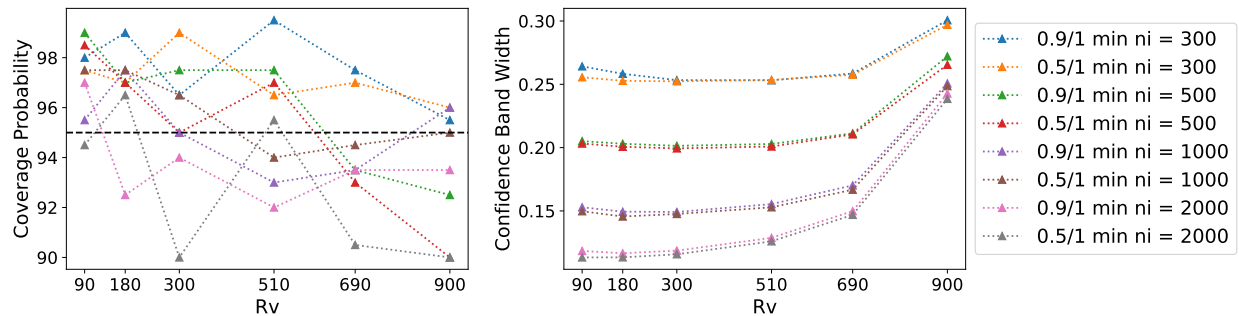


Figure 6: Coverage probability and confidence band width for different splits with  $N = 1000$  and optimally tuned  $R$  and  $\theta \min_i n_i$ .

- Barton, R., H. Lam, and E. Song. 2022. "Input Uncertainty in Stochastic Simulation". *To appear in the Palgrave Handbook of Operations Research, Chapter 17*.
- Barton, R. R. 2012. "Tutorial: Input Uncertainty in Output Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Barton, R. R., B. L. Nelson, and W. Xie. 2014. "Quantifying Input Uncertainty via Simulation Confidence Intervals". *INFORMS Journal on Computing* 26(1):74–87.
- Barton, R. R., and L. W. Schruben. 1993. "Uniform and Bootstrap Resampling of Empirical Distributions". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans, M. Mollaghasemi, E. Russell, and W. Biles, 503–508. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- 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.
- Cheng, R. C., and W. Holland. 1998. "Two-Point Methods for Assessing Variability in Simulation Output". *Journal of Statistical Computation Simulation* 60(3):183–205.
- Cheng, R. C., and W. Holland. 2004. "Calculation of Confidence Intervals for Simulation Output". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 14(4):344–362.
- Cheng, R. C., and W. Hollond. 1997. "Sensitivity of Computer Simulation Experiments to Errors in Input Data". *Journal of Statistical Computation and Simulation* 57(1-4):219–241.
- Chick, S. E. 2001. "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty". *Operations Research* 49(5):744–758.
- Corlu, C. G., A. Akcay, and W. Xie. 2020. "Stochastic Simulation Under Input Uncertainty: A Review". *Operations Research Perspectives* 7:100162.
- Ghosh, S., and H. Lam. 2019. "Robust Analysis in Stochastic Simulation: Computation and Performance Guarantees". *Operations Research* 67(1):232–249.
- Henderson, S. G. 2003. "Input Modeling: Input Model Uncertainty: Why Do We Care and What Should We Do About It?". In *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 90–100. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Lam, H. 2016. "Advanced Tutorial: Input Uncertainty and Robust Analysis in Stochastic Simulation". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 178–192. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Lam, H., Z. Liu, and H. Qian. 2022. "Efficient Input Uncertainty Quantification Without Nested Simulation: A Subsampled Infinitesimal Jackknife Approach". *Preprint*.
- Lam, H., and H. Qian. 2016. "The Empirical Likelihood Approach to Simulation Input Uncertainty". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 791–802. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Lam, H., and H. Qian. 2017. "Optimization-Based Quantification of Simulation Input Uncertainty via Empirical Likelihood". *arXiv preprint arXiv:1707.05917*. <https://arxiv.org/abs/1707.05917>, accessed 13<sup>th</sup> May 2022.
- Lam, H., and H. Qian. 2019. "Random Perturbation and Bagging to Quantify Input Uncertainty". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, and C. Szabo, 320–331. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Lam, H., and H. Qian. 2021. "Subsampling to Enhance Efficiency in Input Uncertainty Quantification". *Operations Research*.

- Law, A. M., W. D. Kelton, and W. D. Kelton. 2007. *Simulation Modeling and Analysis*, Volume 3. New York: Mcgraw-hill.
- Lin, Y., E. Song, and B. L. Nelson. 2015. "Single-Experiment Input Uncertainty". *Journal of Simulation* 9(3):249–259.
- Nelson, B. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. New York: Springer Science & Business Media.
- Serfling, R. J. 2009. *Approximation Theorems of Mathematical Statistics*, Volume 162. New York: John Wiley & Sons.
- Song, E., and B. L. Nelson. 2015. "Quickly Assessing Contributions to Input Uncertainty". *IIE Transactions* 47(9):893–909.
- Song, E., and B. L. Nelson. 2019. "Input-Output Uncertainty Comparisons for Discrete Optimization via Simulation". *Operations Research* 67(2):562–576.
- Song, E., B. L. Nelson, and C. D. Pegden. 2014. "Advanced Tutorial: Input Uncertainty Quantification". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Diallo, I. Ryzhov, L. Yilmaz, S. Buckley, and J. Miller, 162–176. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Van der Vaart, A. W. 2000. *Asymptotic Statistics*, Volume 3. Cambridge: Cambridge University Press.
- Wasserman, L. 2006. *All of Nonparametric Statistics*. New York: Springer Science & Business Media.
- Xie, W., B. L. Nelson, and R. R. Barton. 2014. "A Bayesian Framework for Quantifying Uncertainty in Stochastic Simulation". *Operations Research* 62(6):1439–1452.
- Zouaoui, F., and J. R. Wilson. 2003. "Accounting for Parameter Uncertainty in Simulation Input Modeling". *Iie Transactions* 35(9):781–792.
- Zouaoui, F., and J. R. Wilson. 2004. "Accounting for Input-Model and Input-Parameter Uncertainties in Simulation". *IIE Transactions* 36(11):1135–1151.

## AUTHOR BIOGRAPHIES

**MOTONG CHEN** is a Master of Financial Engineering student in the Department of Industrial Engineering and Operations Research at Columbia University. She received her bachelor's degree in Computer Science and Philosophy from Columbia University, and was a rate options trader at Bank of America. Her research interests include Monte Carlo simulation and uncertainty quantification. Her email address is [mc3917@columbia.edu](mailto:mc3917@columbia.edu).

**ZHENYUAN LIU** is a Ph.D. student in the Department of Industrial Engineering and Operations Research at Columbia University. He received his bachelor's degree from the Department of Financial Mathematics, School of Mathematical Sciences, Peking University. His research focuses on data-driven stochastic optimization, uncertainty quantification, rare-event estimation and applied probability. His email address is [zl2817@columbia.edu](mailto:zl2817@columbia.edu).

**HENRY LAM** is an Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. He received his Ph.D. degree in statistics from Harvard University in 2011, and was on the faculty of Boston University and the University of Michigan before joining Columbia in 2017. His research focuses on Monte Carlo simulation, uncertainty quantification, risk analysis, and stochastic and robust optimization. His email address is [henry.lam@columbia.edu](mailto:henry.lam@columbia.edu).