Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

NONPARAMETRIC INPUT-OUTPUT UNCERTAINTY COMPARISONS

Jaime Gonzalez[1], Johannes Milz[1], and Eunhye Song[1]

[1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology,
Atlanta, GA, USA

ABSTRACT

We consider the problem of inferring the system with the best simulation output mean among k systems when the simulation model is subject to input uncertainty caused by estimated common input models from finite data. The Input-Output Uncertainty Comparisons (IOU-C) procedure is designed to return a set of solutions that contains the best solution with an asymptotic probability guarantee when parametric input models are adopted. We extend this framework to nonparametric IOU-C (NIOU-C) when empirical distributions of the data are adopted as input models. Representing the simulation output mean of each system as a functional of the common empirical distributions via the functional Taylor series expansion, we propose two methods that rely on the nonparametric delta method and an ambiguity set formulation, respectively. We provide numerical examples to test the performance of our methods and show that they outperform the IOU-C.

1 INTRODUCTION

Simulation is a common tool for solving a decision-making problem defined for a complex real-world stochastic system. These simulation models are built to approximate the stochastic behaviors in the target systems by generating random variates from probability distributions and feeding them through the simulation logic to produce outputs. Typically, the input probability distributions are unknown and must be estimated from data observed from the target system. Inevitably, when the data size is finite, there is estimation error in the postulated input distribution functions. The uncertainty in the simulation output caused by the estimation error in the input models is referred to as *input uncertainty* (IU) and has been actively studied in the literature; see Barton et al. (2022) for a recent review.

When a simulation model has IU, any statistical inference made from its outputs is also subject to error induced by IU. When we make decisions based on the simulation outputs, there is risk of making a suboptimal decision due to the discrepancy between estimated input models and true input distributions in the system. This is known as input model risk.

In this work, we discuss robustifying optimization via simulation against input model risk from the statistical inference point of view. In particular, we focus on the problems known as ranking and selection (R&S), where the solutions in consideration are finite and categorical. Traditional R&S does not consider the input model risk; the focus is on controlling the uncertainty from the stochastic error inherent to simulation to find the optimal solution. In recent years, several methods have been proposed to incorporate IU in R&S. Gao et al. (2017) reformulates the traditional R&S problem to find a robust solution under input model risk and devise finite-budget algorithms to maximize the probability of correct selection. Pearce and Branke (2017) propose a Bayesian Optimization framework that incorporates IU. Corlu and Biller (2013) and Corlu and Biller (2015) extend the traditional subset selection procedures to account for IU. Song and Nelson (2019) propose the Input-Output Uncertainty Comparisons (IOU-C) procedure that extends the multiple comparisons with the best (MCB) framework (Chang and Hsu 1992) to incorporate IU. However, majority of these work either assume that the input distributions have known family of distributions, or the true input distribution belongs in a finite set of candidate distributions. Both assumptions are limiting.

In this work, we propose two Nonparametric IOU-C (NIOU-C) procedures to account for IU in the R&S problem. We do not impose any assumptions on the distribution families of the inputs. Instead, we adopt the empirical distribution function (edf) to approximate the true distributions. We introduce the influence function, first defined by Hampel (1974), to represent the simulation output mean as a functional of the edf using the functional Taylor series expansion. In the first procedure, we construct CI widths from stochastic error and IU separately, each of which relies on a Central Limit Theorem (CLT). In the second procedure, we exploit the empirical likelihood (EL) theory (Owen 2001) to construct an ambiguity set for distributions that lets us infer the true performance of each system with a target level of confidence, asymptotically. Then, the CI widths are computed by solving a set of convex optimization problems while constraining the true input distribution to belong in the ambiguity set.

Both approaches are benchmarked against IOU-C in our empirical study assuming that incorrect parametric family assumptions are made for IOU-C. The results show that our first approach selects a subset of systems that includes the best one with very high probability in all cases, but the number of systems in the subset could be potentially high, which represents a conservative method. In contrast, our second approach selects a subset of at most 3 systems, but at the price of having less probability compared to the first procedure. This method represents an aggressive one. Nevertheless, both approaches shows better results than the IOU-C methods for the chosen example.

The rest of the paper is organized as follows. In Section 2, we describe the problem, review the MCB framework, and introduce the nonparametric delta method. In Sections 3 and 4, we provide the NIOU-C procedures based on a CLT (NIOU-C:CLT) and the ambiguity set formulation (NIOU-C:AS), respectively. Empirical evaluations are provided in Section 5.

2 PROBLEM DEFINITION

Consider k systems where stochasticity in each system is characterized by the same set of input distributions, $\mathbf{F}^c$, where the c stands for "correct". We assume that $\mathbf{F}^c$ is a collection of m independent input probability distribution functions $F_1^c,\ldots,F_m^c$. We also assume that there exists a stochastic simulator that perfectly emulates the system output given $\mathbf{F}^c$. The ith system's simulation output given $\mathbf{F}^c$ is denoted by $Y_i(\mathbf{F}^c)$ and its performance measure is defined as the expected value of the simulation output, $\mathrm{E}[Y_i(\mathbf{F}^c)]$. The goal is to find the system with the largest performance measure, i.e. $i^c \triangleq \arg\max_{1\le i\le k}\mathrm{E}[Y_i(\mathbf{F}^c)]$. We assume i^c to be unique in this work. Consequently, $\mathrm{E}[Y_{i^c}(\mathbf{F}^c)] \neq \mathrm{E}[Y_i(\mathbf{F}^c)], \forall i \neq i^c$.

In practice, $\mathbf{F}^c$ is unknown and thus must be estimated from finite data observed from the real-world system. For each $p = 1,\ldots,m$, let $X_{p1},\ldots,X_{pn_p}$ denote the independent and identically distributed (i.i.d.) observations from F_p^c. We do not make any parametric assumptions on the input distributions. Instead, we adopt an edf $\hat{\mathbf{F}}$ constructed from the observations as an estimator for $\mathbf{F}^c$:

$$\hat{\mathbf{F}}(x_1,\ldots,x_m) \triangleq \prod_{p=1}^m n_p^{-1}\sum_{j=1}^{n_p}\mathbf{1}\{X_{pj}=x_p\}.$$

We emphasize that $\hat{\mathbf{F}}$ is an empirical probability distribution function, not cumulative distribution function.

Let $\eta_i(\mathbf{F}) = \mathrm{E}[Y_i(\mathbf{F})|\mathbf{F}]$, for any generic $\mathbf{F}$. Then, the R&S problem aims to find $i^c = \arg\max_i \eta_i(\mathbf{F}^c)$. When the simulation is run with $\hat{\mathbf{F}}$, the sample average of the simulation outputs is an unbiased estimator for the conditional output mean, $\eta_i(\hat{\mathbf{F}}) = \mathrm{E}[Y_i(\hat{\mathbf{F}})|\hat{\mathbf{F}}]$, not $\eta_i(\mathbf{F}^c)$. Simply applying a R&S procedure with estimated $\hat{\mathbf{F}}$ would let us select the best solution given $\hat{\mathbf{F}}$, not i^c. Representing the simulation output at each solution with a mixed effects model, Song et al. (2015) show that under IU, one may not provide a desired level ($1-\alpha$ for some arbitrary α) of statistical guarantee for selecting i^c even if the sampling distribution of $\eta_i(\hat{\mathbf{F}})$ for each i can be characterized. Namely, if the uncertainty about $\hat{\mathbf{F}}$ is too large compared to the performance difference between $\eta_{i^c}(\mathbf{F}^c)$ and $\eta_l(\mathbf{F}^c)$ for all $l \neq i^c$, then any R&S procedure would not be able to separate i^c from the rest of the systems with high confidence.

Instead of selecting i^c, an alternative approach is to provide some statistical inference on the identity of i^c. Song and Nelson (2019) take this approach and focus on providing the simultaneous confidence intervals (CIs) for the mean differences, $\eta_i(\mathbf{F}^c) - \max_{l\neq i}\eta_l(\mathbf{F}^c)$, for all $1 \leq i \leq k$. These CIs are referred to as MCB CIs. The following theorem by Chang and Hsu (1992) stipulates how to derive the MCB CIs when $\mathbf{F}^c$ is known:

Theorem 1 (Chang and Hsu 1992) Let $\hat{\eta}_i(\mathbf{F}^c)$ be an estimator of $\eta_i(\mathbf{F}^c)$, $x^+ = \max\{0,x\}$ and $x^- = \max\{0,-x\}$. If, for each $i \in \{1,\ldots,k\}$

$$\Pr\{\hat{\eta}_i(\mathbf{F}^c) - \hat{\eta}_l(\mathbf{F}^c) - (\eta_i(\mathbf{F}^c) - \eta_l(\mathbf{F}^c)) \geq -q_{il}, \forall l \neq i\} \geq 1-\alpha, \tag{1}$$

then the following statement holds:

$$\Pr\left\{\eta_i(\mathbf{F}^c) - \max_{l\neq i}\eta_l(\mathbf{F}^c) \in [D_i^-, D_i^+], \forall i\right\} \geq 1-\alpha,$$

where $D_i^+ = (\min_{l\neq i}\{\hat{\eta}_i(\mathbf{F}^c) - \hat{\eta}_l(\mathbf{F}^c) + q_{il}\})^+$, $I = \{i : D_i^+ > 0\}$ and

$$D_i^- = \begin{cases} 0 & \text{if } I = \{i\} \\ -(\min_{l\in I: l\neq i}\{\hat{\eta}_i(\mathbf{F}^c) - \hat{\eta}_l(\mathbf{F}^c) - q_{li}\})^- & \text{otherwise.} \end{cases}$$

Theorem 1 allows us to construct the MCB CIs from the k sets of simultaneous CIs comparing each solution's mean to the rest of the solutions'. Moreover, it is guaranteed that $\Pr\{i^c \in I\} \geq 1-\alpha$. The MCB framework is powerful since it does not impose any distributional assumption on the simulation outputs and the CIs are valid as long as the set of $\{q_{il}\}_{i\neq l}$ satisfying (1) exists.

However, with unknown $\mathbf{F}^c$, finding such $\{q_{il}\}_{i\neq l}$ proves to be challenging. Song and Nelson (2019) study this problem for the case when the parametric family of $\mathbf{F}^c$ is assumed known. Extending the MCB CIs in Theorem 1, they propose the IOU-C procedure that accounts for both IU and simulation error. In particular, for the former, they coin the term common-input-data (CID) effect highlighting that the IU is caused by $\hat{\mathbf{F}}$ commonly adopted by all k systems.

Under the parametric assumption, estimating $\mathbf{F}^c$ boils down to estimating its unknown parameter vector. While this makes the problem simpler, it may be unrealistic in many practical problems to know or even assume that the input distribution function belongs to a certain parametric family.

In this paper, we do not impose a parametric assumption on $\mathbf{F}^c$. Instead, we suppose that $\mathbf{F}^c$ is estimated by an edf $\hat{\mathbf{F}}$. Let $b_i(\hat{\mathbf{F}},\mathbf{F}^c) \triangleq \eta_i(\hat{\mathbf{F}}) - \eta_i(\mathbf{F}^c)$. Note that $b_i(\hat{\mathbf{F}},\mathbf{F}^c)$ denotes the CID effect of $\hat{\mathbf{F}}$ on system i. Then, the simulation output of system i run with $\hat{\mathbf{F}}$ is

$$Y_i(\hat{\mathbf{F}}) = \eta_i(\hat{\mathbf{F}}) + \varepsilon_i(\hat{\mathbf{F}}) = \eta_i(\mathbf{F}^c) + b_i(\hat{\mathbf{F}},\mathbf{F}^c) + \varepsilon_i(\hat{\mathbf{F}}),$$

where the simulation error, $\varepsilon_i(\hat{\mathbf{F}})$, has zero mean and a finite variance given $\hat{\mathbf{F}}$. Observe that $b_i(\hat{\mathbf{F}},\mathbf{F}^c)$ accounts for IU and $\varepsilon_i(\hat{\mathbf{F}})$ represents the stochastic error in simulation.

For each system i, let

$$\hat{\eta}_i(\hat{\mathbf{F}}) = \bar{Y}_i(\hat{\mathbf{F}}) = R^{-1}\sum_{j=1}^{R} Y_{ij}(\hat{\mathbf{F}}),$$

where $Y_{ij}(\hat{\mathbf{F}})$ is the simulation output of the jth replication using $\hat{\mathbf{F}}$ as input distribution. Then $\bar{Y}_i(\hat{\mathbf{F}})$ is the sample average of R replications.

In Sections 3 and 4, we propose nonparametric extensions of the IOU-C procedure. Both extensions rely on the functional Taylor series expansion of $\eta_i(\mathbf{F})$ with respect to $\hat{\mathbf{F}}$, which is discussed below. To set it up, we explicitly denote the simulation output of solution i as $Y_i(\hat{\mathbf{F}}) = h_i(\mathbf{X}_{i1},\ldots,\mathbf{X}_{im})$ where $\mathbf{X}_{ip} = (X_{ip}(1),\ldots,X_{ip}(T_{ip}))$ are the random variates drawn from the pth edf to run the ith solution's simulation. This notation highlights the dependency of the simulation output with the random variates.

Let $\mathrm{IF}_{ip}(x)$ denote the influence function of the performance measure of the system i, $\eta_i(\mathbf{F})$, where $F_p \in \mathbf{F}$, when there is a perturbation of the input distribution F_p in the direction of x on the support of F_p. In other words, $\mathrm{IF}_{ip}(x)$ represents the change in the expected value of the simulation output when the corresponding input distribution is infinitesimally perturbed in the direction of x, where x is a support point of F_p. Mathematically, we have:

$$\mathrm{IF}_{ip}(x) = \lim_{\varepsilon \to 0} \frac{\eta(F_1,\ldots,F_{i-1},(1-\varepsilon)F_i+\varepsilon\delta_x,F_{i+1},\ldots,F_m)-\eta(F_1,\ldots,F_m)}{\varepsilon}, \tag{2}$$

where δ_x puts the unit probability mass on x. Lam and Qian (2019) rewrite (2) as

$$\mathrm{IF}_{ip}(x) = \textstyle\sum_{t=1}^{T_{ip}} \mathbb{E}_{\mathbf{F}}[h_i(\mathbf{X}_{i1},\ldots,\mathbf{X}_{im})|X_{ip}(t)=x] - T_{ip}\eta_i(\mathbf{F}). \tag{3}$$

Note that $\mathbb{E}_{\mathbf{F}}[\mathrm{IF}_{ip}(x)] = 0$, where $\mathbb{E}_{\mathbf{F}}$ denotes the expectation taken with respect to $\mathbf{X} \sim \mathbf{F}$.

Suppose η_i is a smooth function of $\hat{\mathbf{F}}$. From the nonparametric delta method, we have

$$\eta_i(\hat{\mathbf{F}}) \approx \eta_i(\mathbf{F}^c) + \sum_{p=1}^{m} \int \mathrm{IF}_{ip}(x)d\hat{\mathbf{F}}_p(x). \tag{4}$$

Then, we can write

$$b_i(\hat{\mathbf{F}},\mathbf{F}^c) - b_l(\hat{\mathbf{F}},\mathbf{F}^c) \approx \sum_{p=1}^{m} \int (\mathrm{IF}_{ip}(x) - \mathrm{IF}_{lp}(x))d\hat{\mathbf{F}}_p(x). \tag{5}$$

Observe that the left hand side of (5) represents the difference in the two systems' CID effects.

3 NONPARAMETRIC IOU-C: ASYMPTOTIC NORMALITY APPROACH

In this section, we describe the NIOU-C:CLT procedure. If $\mathbf{F}^c$ is known, a natural choice of $\hat{\eta}_i(\mathbf{F}^c)$ is $\bar{Y}_i(\mathbf{F}^c)$ in which case $\{q_{il}\}_{i\neq l}$ accounts only for stochastic error. When input distributions are estimated and $\hat{\eta}_i(\mathbf{F}^c) = \bar{Y}_i(\hat{\mathbf{F}})$ is adopted as an estimator for each i, $\{q_{il}\}_{i\neq l}$ depend on both stochastic error and input uncertainty. Let $q_{il} = q_{il}^{(1)} + q_{il}^{(2)}$ with $q_{il}^{(1)}, q_{il}^{(2)} > 0$. Then, Song and Nelson (2019) show that

$$\begin{aligned}
&\Pr\{\hat{\eta}_i(\mathbf{F}^c) - \hat{\eta}_l(\mathbf{F}^c) - (\eta_i(\mathbf{F}^c) - \eta_l(\mathbf{F}^c)) \geq -q_{il}, \forall l \neq i\} \\
&\geq \mathrm{E}\left[\mathbf{1}\{b_i(\hat{\mathbf{F}},\mathbf{F}^c) - b_l(\hat{\mathbf{F}},\mathbf{F}^c) \geq -q_{il}^{(1)}, \forall l \neq i\} \Pr\{\bar{\varepsilon}_i(\hat{\mathbf{F}}) - \bar{\varepsilon}_l(\hat{\mathbf{F}}) \geq -q_{il}^{(2)}, \forall l \neq i | \hat{\mathbf{F}}\}\right]
\end{aligned}$$

Suppose we choose $q_{il}^{(2)}$ such that $\Pr\{\bar{\varepsilon}_i(\hat{\mathbf{F}}) - \bar{\varepsilon}_l(\hat{\mathbf{F}}) \geq -q_{il}^{(2)}, \forall l \neq i | \hat{\mathbf{F}}\} = 1 - \alpha_2$, then the last expression becomes $\mathrm{E}[\mathbf{1}\{b_i(\hat{\mathbf{F}},\mathbf{F}^c) - b_l(\hat{\mathbf{F}},\mathbf{F}^c) \geq -q_{il}^{(1)}, \forall l \neq i\}(1-\alpha_2)] = \Pr\{b_i(\hat{\mathbf{F}},\mathbf{F}^c) - b_l(\hat{\mathbf{F}},\mathbf{F}^c) \geq -q_{il}^{(1)}, \forall l \neq i\}(1-\alpha_2)$. Then, we can find $q_{il}^{(1)}$ such that $\Pr\{b_i(\hat{\mathbf{F}},\mathbf{F}^c) - b_l(\hat{\mathbf{F}},\mathbf{F}^c) \geq -q_{il}^{(1)}, \forall l \neq i\} = 1 - \alpha_1$ so $\Pr\{\hat{\eta}_i(\mathbf{F}^c) - \hat{\eta}_l(\mathbf{F}^c) - (\eta_i(\mathbf{F}^c) - \eta_l(\mathbf{F}^c)) \geq -q_{il}, \forall l \neq i\} \geq (1-\alpha_1)(1-\alpha_2)$.

In the next sections, we discuss different methods to find $\{q_{il}^{(1)}\}_{i\neq l}$ and $\{q_{il}^{(2)}\}_{i\neq l}$ satisfying the conditions.

3.1 CID Effects

Let $n = m^{-1}\sum_{p=1}^{m} n_p$ denote the average number of observations from all data sources. We assume $\frac{n_p}{n} \to \beta_p > 0$ when n increases to infinity. In other words, the proportion of observations from each input distribution converges to a positive number. Then, under some regularity conditions, the following CLT can be derived from (4):

$$\sqrt{n}\left(b_i(\hat{\mathbf{F}},\mathbf{F}^c) - b_l(\hat{\mathbf{F}},\mathbf{F}^c)\right) \xrightarrow{D} \mathrm{N}\left(0, \sum_{p=1}^{m} \frac{1}{\beta_p} \mathrm{E}\left[\left(\mathrm{IF}_{ip}(x) - \mathrm{IF}_{lp}(x)\right)^2\right]\right), \quad \text{as } n \to \infty. \tag{6}$$

Here, $\xrightarrow{D}$ means convergence in distribution. Moreover, (6) can be extended to obtain the joint asymptotic distribution of $\{\sqrt{n}(b_i(\hat{\mathbf{F}},\mathbf{F}^c) - b_l(\hat{\mathbf{F}},\mathbf{F}^c))\}_{l=1;l\neq i}^{k}$, denoted by $\mathrm{N}(\mathbf{0}_{k-1}, \Sigma^i)$, where $\mathbf{0}_{k-1}$ is the zero vector of dimension $k-1$ and Σ^i is the variance-covariance matrix. Observe that, for each i, Σ^i is a $(k-1)\times(k-1)$ matrix whose (l,l')th element is $\Sigma^i_{ll'} = \lim_{n\to\infty} n\mathrm{Cov}(b_i(\hat{\mathbf{F}},\mathbf{F}^c)-b_l(\hat{\mathbf{F}},\mathbf{F}^c), b_i(\hat{\mathbf{F}},\mathbf{F}^c)-b_{l'}(\hat{\mathbf{F}},\mathbf{F}^c))$. From (6), we have

$$\Sigma^i_{ll} = \sum_{p=1}^{m}\frac{1}{\beta_p}\mathrm{E}\left[\left(\mathrm{IF}_{ip}(x)-\mathrm{IF}_{lp}(x)\right)^2\right], \text{ and } \Sigma^i_{ll'} = \sum_{p=1}^{m}\frac{1}{\beta_p}\mathrm{E}\left[\left(\mathrm{IF}_{ip}(x)-\mathrm{IF}_{lp}(x)\right)\left(\mathrm{IF}_{ip}(x)-\mathrm{IF}_{l'p}(x)\right)\right].$$

Therefore, we can set $\{q^{(1)}_{il}\}_{i\neq l}$ to be the $(1-\alpha_1)$-quantile vector of $\mathrm{N}(\mathbf{0}_{k-1},\Sigma^i/n)$.

For general simulation output function h_i, however, $\mathrm{IF}_{ip}(\cdot)$ does not have an analytical expression and needs to be estimated via simulation. We adopt the estimators proposed by Lam and Qian (2019), which can be computed by running simulations at $\hat{\mathbf{F}}$ only.

First, Lam and Qian (2019) propose to approximate $\mathrm{IF}_{ip}(x)$ with the influence function of $\eta_i(\cdot)$ evaluated at the empirical distribution functions, $\hat{\mathbf{F}}$:

$$\hat{\mathrm{IF}}_{ip}(x) = \textstyle\sum_{t=1}^{T_{ip}} \mathbb{E}_{\hat{\mathbf{F}}}[h_i(\mathbf{X}_{i1},\ldots,\mathbf{X}_{im})|X_{ip}(t)=x] - T_{ip}\eta_i(\hat{\mathbf{F}}), \quad \text{for } x\in\{X_{p1},\ldots,X_{pn_p}\}. \tag{7}$$

Unlike in (3), the expectation in (7) is taken with respect to $\hat{\mathbf{F}}$. Since we have $\hat{\mathbf{F}}$, we can obtain an estimator of (7) via simulations. Namely, the estimator of $\hat{\mathrm{IF}}_{ip}(\cdot)$ at X_{pj} can be computed as

$$\hat{\hat{\mathrm{IF}}}_{ip}(X_{pj}) = \frac{1}{R}\sum_{r=1}^{R}\left(Y_{ir}(\hat{\mathbf{F}})-\bar{Y}_i\right)\left(n\sum_{t=1}^{T_{ip}}\mathbf{1}\{X^r_{ip}(t)=X_{pj}\}-T_{ip}\right), \tag{8}$$

where $X^r_{ip}(1),\ldots,X^r_{ip}(T_{ip})$ are the random variates drawn from the pth input model to run the rth replication simulation of system i. The variation we make in this work is to allow adoption of the common random numbers (CRNs) across all systems. That is, for each r and p, $X^r_{ip}(1),\ldots,X^r_{ip}(T_{ip})$ are reused across all systems as much as possible. If T_{ip} remains the same for all systems, then all T_{ip} inputs are reduced.

From (8), we can estimate the (l,l')th element of Σ^i as

$$\hat{\Sigma}^i_{ll'} = \sum_{p=1}^{m}\frac{n}{n_p}\sum_{j=1}^{n_p}\frac{1}{n_p}(\hat{\hat{\mathrm{IF}}}_{ip}(X_{pj})-\hat{\hat{\mathrm{IF}}}_{lp}(X_{pj}))(\hat{\hat{\mathrm{IF}}}_{ip}(X_{pj})-\hat{\hat{\mathrm{IF}}}_{l'p}(X_{pj})), \tag{9}$$

for each i and $1\le l,l'\le k$ such that $l,l'\neq i$. Let $\hat{\Sigma}^i$ be the resulting estimator of Σ^i. Now, $\{q^{(1)}_{il}\}$ can be estimated by a $(1-\alpha_1)$ quantile of $\mathrm{N}(\mathbf{0}_{k-1},\hat{\Sigma}^i/n)$. Note that there are infinitely many candidates for the quantile. In our empirical study in section 5, we adopt the quantile estimation procedure in Appendix EC.3 of Song and Nelson (2019).

3.2 Simulation error

To complete the MCB CIs, our goal is to find $q^{(2)}_{il}$ such that for any $\hat{\mathbf{F}}$, we have $\Pr\{\bar{\varepsilon}_i(\hat{\mathbf{F}})-\bar{\varepsilon}_l(\hat{\mathbf{F}})\ge -q^{(2)}_{il}, \forall l\neq i|\hat{\mathbf{F}}\} = 1-\alpha_2$. We have $\{\sqrt{R}(\bar{\varepsilon}_i(\hat{\mathbf{F}})-\bar{\varepsilon}_l(\hat{\mathbf{F}}))\}^k_{l=1,l\neq i}$ converge in distribution from a CLT to $N(0,V_i)$ as R increases, where V_i is the variance-covariance matrix of $\{Y_{ij}(\hat{\mathbf{F}})-Y_{lj}(\hat{\mathbf{F}})\}^k_{l=1,l\neq i}$. Although V_i is unknown, we can estimate its sample version $\hat{V}_i$ from the R replications already made to estimate the influence functions. Then, $\{q^{(2)}_{il}\}^k_{l=1,l\neq i}$ can be computed as a $(1-\alpha_2)$ quantile of $\mathrm{N}(\mathbf{0}_{k-1},\hat{V}_i/R)$.

3.3 Algorithm

We present the NIOU-C procedure that constructs the MCB CIs from the CLTs discussed in sections 3.1 and 3.2:

1. Choose $0<\alpha_1,\alpha_2<1/2$ such that $(1-\alpha_1)(1-\alpha_2)=1-\alpha$.

2. From observations $X_{p1},...,X_{pn}$, compute the edf, $\hat{F}_p$, as the estimator of F_p^c for each $p = 1,...,m$.
3. For each $i \in \{1,...,k\}$:
   (a) Run R simulation replications to obtain $Y_{i1}(\hat{\mathbf{F}}),...,Y_{iR}(\hat{\mathbf{F}})$.
   (b) Compute $\bar{Y}_i = \frac{1}{R}\sum_{r=1}^{R} Y_{ir}(\hat{\mathbf{F}})$.
   (c) Compute $\hat{\hat{\mathrm{IF}}}_{ip}(X_{pj})$ in (8) for $p = 1,...,m$ and $j = 1,...,n_p$.
4. (CID effects) For each $i \in \{1,...,k\}$:
   (a) For $1 \le l,l' \le k$ and each i, with $l,l' \ne i$, compute $\hat{\Sigma}^i_{ll'}$ as in (9).
   (b) Compute the $(k-1)$-dimensional $(1-\alpha_1)$ quantile $q_{il}^{(1)}$ of the distribution $\mathrm{N}(0,\hat{\Sigma}^i/n)$.
5. (Stochastic error) For each system i:
   (a) Compute the sample var-covariance matrix $\hat{V}_i$ of $\{\varepsilon_i - \varepsilon_l\}_{l=1,l\ne i}^{k}$.
   (b) Compute the $(k-1)$-dimensional $(1-\alpha_2)$ quantile $q_{il}^{(2)}$ of the distribution $\mathrm{N}(0,\hat{V}_i/R)$.
6. For each pair $(i,l), i \ne l$, set $q_{il} = q_{il}^{(1)} + q_{il}^{(2)}$. Use Theorem 1 to derive $1-\alpha$ simultaneous comparisons CIs.

The NIOU-C:CLT procedure runs R replications at each system, which costs kR replications in total. Note that the algorithm may be run with or without the CRNs. If CRNs are adopted, then in step 3(a), the same random number stream should be repeated for each i.

4 NONPARAMETRIC IOU-C: AMBIGUITY SET APPROACH

In this section, we extend the IOU-C framework by introducing an ambiguity set to help us infer each system's performance under $\mathbf{F}^c$ with high probability, where the probability guarantee is supported by the empirical likelihood theory. Below, we first discuss how the MCB CI framework can be modified to exploit the ambiguity set formulation.

Recall that the MCB CIs are constructed from the joint CIs in (1) constructed for all $1 \le i \le k$. Defining $U_{il} \triangleq \hat{\eta}_i(\mathbf{F}^c) - \hat{\eta}_l(\mathbf{F}^c) + q_{il}$, (1) can be rewritten as

$$\Pr\{U_{il} \ge \eta_i(\mathbf{F}^c) - \eta_l(\mathbf{F}^c), \forall l \ne i\} \ge 1-\alpha. \tag{10}$$

The MCB procedure in Theorem 1 computes D_i^+ and D_i^- from $\hat{\eta}_i(\mathbf{F}^c) - \hat{\eta}_l(\mathbf{F}^c) + q_{il} = U_{il}$ and $\hat{\eta}_i(\mathbf{F}^c) - \hat{\eta}_l(\mathbf{F}^c) - q_{li} = -\{\hat{\eta}_l(\mathbf{F}^c) - \hat{\eta}_i(\mathbf{F}^c) + \hat{\eta}_l(\mathbf{F}^c) - q_{li}\} = -U_{li}$ for each i. Therefore, it suffices to find $\{U_{il}\}_{l\ne i}$ satisfying (10) for $1 \le i \le k$. Suppose we have an ambiguity set $\mathscr{U}_\alpha$ such that $\Pr\{\mathbf{F}^c \in \mathscr{U}_\alpha\} \ge 1-\alpha$. Then, $U_{il}, l \ne i$ can be found by solving

$$\max_{\mathbf{F}} \quad \eta_i(\mathbf{F}) - \eta_l(\mathbf{F}) \qquad \text{subject to} \quad \mathbf{F} \in \mathscr{U}_\alpha. \tag{11}$$

for all $l \ne i$ for each $1 \le i \le k$. Clearly, one would prefer a choice of $\mathscr{U}_\alpha$ such that (11) can be easily solved. Below, we discuss how to construct such $\mathscr{U}_\alpha$ based on the empirical likelihood theory.

To set up the discussion, consider some d-dimensional random vector $\mathbf{Z} \sim G_0$ with mean $\mathrm{E}[\mathbf{Z}] = \mu_0$ and full-rank variance-covariance matrix $\mathrm{V}[\mathbf{Z}] = V_0$. Let $\mathbf{z}_1,\ldots,\mathbf{z}_n$ be the observed size-n i.i.d. sample from G_0. Given the sample, the empirical likelihood, L, of distribution function G defined on the support of $\mathbf{Z}$ is (Owen 2001)

$$L(G) = \prod_{j=1}^{n} G(\mathbf{z}_j). \tag{12}$$

Thus, $L(G)$ is maximized when $G = \hat{G}$, where $\hat{G}$ is the edf constructed from the size-n sample. Note that (12) is a nonparametric parallel to the parametric likelihood function. Thus, a natural estimator of G_0 is the G that maximizes (12).

Indeed, it is easy to see that if G puts no probability mass on any of the sampled observations, then $L(G) = 0$. Moreover, if G assigns nonzero probability to a set of points other than the sample, then one can also reassign the probability to the sample and increase the empirical likelihood.

From (12), the empirical likelihood ratio of G to the $\hat{G}$ is defined as

$$\mathsf{R}(G) \triangleq L(G)/L(\hat{G}).$$

If the goal is to make inference on the mean of $\mathbf{Z}$, μ_0, then we can further define the profile likelihood function given $\mu \in \mathbb{R}^d$ as

$$\mathsf{R}(\mu) \triangleq \sup\{\mathsf{R}(G)\colon \mathrm{E}[\mathbf{Z}] = \mu,\ \text{where } \mathbf{Z} \sim G\}.$$

In words, $\mathsf{R}(\mu)$ returns the largest likelihood ratio of G among those that have mean μ.

We further focus on the class of G that assigns nonzero probability weights to the sampled observations, $\mathbf{z}_1, \ldots, \mathbf{z}_n$, i.e., G is represented by probability simplex vector $\mathbf{w} = (w_1, \ldots, w_n)$ such that $w_i \geq 0$ and $\sum_{i=1}^n w_i = 1$. Therefore, $\mathsf{R}(\mu)$ can be redefined as

$$\mathsf{R}(\mu) \triangleq \max\left\{\prod_{j=1}^n n w_j\colon\ \sum_{j=1}^n w_j \mathbf{z}_j = \mu; \sum_{j=1}^n w_j = 1; w_j \geq 0 \text{ for } 1 \leq j \leq n\right\}.$$

where $L(\mathbf{w}) = \prod_{j=1}^n w_j$ and $\prod_{j=1}^n n w_j$ is the empirical likelihood ratio between the probability simplex, $\mathbf{w}$, and the edf, which assigns $1/n$ weight to each sampled observation. The following theorem restated from Theorem 3.2 in Owen (2001) stipulates the asymptotic distribution of $-2\log(\mathsf{R}(\mu_0))$ computed from a random sample of size n.

Theorem 2 Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be i.i.d. observations of G_0 defined earlier. Then, $-2\log(\mathsf{R}(\mu_0))$ converges in distribution to a χ^2_d random variable with d degrees of freedom as $n \to \infty$.

Theorem 2 justifies adopting the following ambiguity set for G_0, if estimating μ is the goal of the inference:

$$\left\{\mathbf{w}\colon\ -2\sum_{j=1}^n \log(n w_j) \leq \chi^2_{d,1-\alpha}; \sum_{j=1}^n w_j = 1; w_j \geq 0 \text{ for } 1 \leq j \leq n\right\}, \tag{13}$$

where $\chi^2_{d,1-\alpha}$ represents the $1-\alpha$ quantile of a χ^2_d random variable. That is, as n increases, some distribution G that has the correct mean, μ_0, is included in (13) with probability $1-\alpha$.

To utilize (13) to construct the ambiguity set, $\mathscr{U}_\alpha$ in Problem (11), we first define n_p-dimensional probability simplex $\mathbf{w}_p$ on $X_{p1}, \ldots, X_{pn_p}$ for $1 \leq p \leq m$. Then, $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ represents the m input models defined by assigning probability masses to the observed data. Suppose we define a joint distribution function $\mathbf{F} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$. Then, the nonparametric delta method in (4) can be rewritten as

$$\eta_i(\mathbf{F}) \approx \eta_i(\mathbf{F}^c) + \sum_{p=1}^m \sum_{j=1}^{n_p} \mathrm{IF}_{ip}(X_{pj}) w_{pj}.$$

Notice that the right-hand side can be regarded as the sum of m expectations, where weight w_{pj} is assigned to $\mathrm{IF}_{ip}(X_{pj})$ for $1 \leq j \leq n_p$ for $1 \leq p \leq m$.

In our framework, the goal is to make simultaneous inference on $\{\eta_i(\mathbf{F}^c) - \eta_l(\mathbf{F}^c), l \neq i\}$ for each i, which is a $(k-1)$-dimensional vector, where each entry is a sum of m expectations under the nonparametric delta effect approximation. Therefore, we conjecture that the corresponding $\mathsf{R}(\mu)$ and μ defined for the $(k-1)$-dimensional vector admits the following asymptotic convergence: $-2\log(\mathsf{R}(\mu_0)) \xrightarrow{D} \chi^2_{k-1}$ for appropriately defined μ_0. From this result, we adopt the following ambiguity set

$$\mathscr{U}_\alpha = \left\{(\mathbf{w}_1, ..., \mathbf{w}_m) : -2\sum_{p=1}^m \sum_{j=1}^{n_p} \log n_p w_{pj} \leq \chi^2_{k-1,1-\alpha}; \sum_{j=1}^{n_p} w_{pj} = 1, \forall p = 1, ..., m; w_{pj} \geq 0, \forall p, j\right\}.$$

This conjecture is based on the following theoretical result.

Theorem 3 Let m, $t \in \mathbb{N}$, and let $\mathbf{Z}_p \in \mathbb{R}^t$, $p = 1,\ldots,m$, be integrable, independent random vectors such that each $\mathbf{Z}_p$ has a full-rank variance-covariance matrix. Let $\mathbf{Y}_{pj} \in \mathbb{R}^t$, $j = 1,\ldots,n_p$, $p = 1,\ldots,m$, be independent such that for each $p \in \{1,\ldots,m\}$, $\mathbf{Y}_{pj}$, $j = 1,\ldots,n_p$, have the same distribution as that of $\mathbf{Z}_p$. We define R on $\mathbb{R}^t$ by

$$R(\mu) \triangleq \max\left\{ \prod_{p=1}^{m}\prod_{j=1}^{n_p} n_p w_{pj} \colon \sum_{p=1}^{m}\sum_{j=1}^{n_p} \mathbf{Y}_{pj} w_{pj} = \mu; \sum_{j=1}^{n_p} w_{pj} = 1, \forall p = 1,...,m; w_{pj} \geq 0, \forall p,j \right\}.$$

Then, $-2\log(R(\sum_{p=1}^{m}\mathrm{E}[\mathbf{Z}_p])) \xrightarrow{D} \chi_t^2$ as $N \triangleq \sum_{p=1}^{m} n_p \to \infty$ and $\liminf_{N\to\infty} \min n_p / \max n_p > 0$.

For $m = 1$, Theorem 3 recovers Theorem 2. For $t = 1$, Theorem 3 corresponds to Theorem 1 in Lam and Qian (2016).

We provide a sketch of proof to support Theorem 3 following section 11.4 in Owen (2001). The setting of Theorem 3 can be formulated as a multi-distribution multivariate estimation problem, $\mathbb{E}[\psi(\mathbf{Z}_1,\ldots,\mathbf{Z}_m,\mu)] = 0$ with unknown $\mu \in \mathbb{R}^t$, where $\psi(\mathbf{z}_1,\ldots,\mathbf{z}_m,\mu) \triangleq \sum_{i=1}^{m}\mathbf{z}_i - \mu$. Here $\psi \colon \mathbb{R}^t \times \cdots \times \mathbb{R}^t \to \mathbb{R}^t$ with the Cartesian product taken $(t+1)$-times. For multi-distribution multivariate estimation, the profile empirical likelihood ratio function is the maximum value of $\prod_{p=1}^{m}\prod_{j=1}^{n_p} n_p w_{pj}$ subject to $\sum_{j=1}^{n_p} w_{pj} = 1, \forall p = 1,\ldots,m$, $w_{pj} \geq 0, \forall p,j$, and $\sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2}\cdots\sum_{j_m=1}^{n_m} w_{1j_1}w_{2j_2}\cdots w_{m,j_m}\psi(\mathbf{Y}_{1j_1},\ldots,\mathbf{Y}_{mj_m},\mu) = 0$. Since $\sum_{j=1}^{n_p} w_{pj} = 1, \forall p = 1,\ldots,m$, the latter constraint equals $\sum_{p=1}^{m}\sum_{j=1}^{n_p}\mathbf{Y}_{pj}w_{pj} - \mu = 0$, which is the constraint in the maximization problem in Theorem 3.

4.1 Estimation of critical values

Observe that if we replace the objective function of (11) with $\eta_i(\mathbf{F}) - \eta_l(\mathbf{F}) - (\eta_i(\mathbf{F}^c) - \eta_l(\mathbf{F}^c))$, the optimal solution remains the same. From (5), we can approximate the new objective function as:

$$\eta_i(\mathbf{F}) - \eta_l(\mathbf{F}) - (\eta_i(\mathbf{F}^c) - \eta_l(\mathbf{F}^c)) \approx \sum_{p=1}^{m}\int (\mathrm{IF}_{ip}(x) - \mathrm{IF}_{lp}(x)) d\mathbf{F}_p(x).$$

The unknown true influence function is replaced with its estimator in (8), which results in the following optimization problem:

$$\max_{\mathbf{w}} \quad \sum_{p=1}^{m}\sum_{j=1}^{n_p} (\hat{\hat{\mathrm{IF}}}_{ip}(X_{pj}) - \hat{\hat{\mathrm{IF}}}_{lp}(X_{pj}))w_{pj} \qquad \text{s.t.} \quad \mathbf{w} \in \mathscr{U}_\alpha. \tag{14}$$

Because $\mathscr{U}_\alpha$ is a convex set of $\mathbf{w}$ and the objective function of (14) is linear in $\mathbf{w}$, (14) is a convex optimization problem. Therefore, (14) can be solved by a convex optimization algorithm. Suppose we obtain the optimal solution $\mathbf{w}_{il}^{max} = (\mathbf{w}_{il1}^{max},...,\mathbf{w}_{ilm}^{max})$, where $\mathbf{w}_{ilp}^{max} \in \mathbb{R}^{n_p}$ for $p = 1,...,m$. However, plugging $\mathbf{w}_{il}^{max}$ back into the objective function of (14) does not provide U_{il} since $\eta_i(\mathbf{F}^c) - \eta_l(\mathbf{F}^c)$ is unknown. Instead, we compute the following estimator of U_{il} by simulating R_2 replications with $\mathbf{w}_{il}^{max}$:

$$\hat{U}_{il} = \frac{1}{R_2}\sum_{r=1}^{R_2}(Y_{ir}(\mathbf{w}_{il}^{max}) - Y_{lr}(\mathbf{w}_{il}^{max})),$$

where $Y_{ir}(\mathbf{w}_{il}^{max})$ refers to system i's simulation output from the rth replication when the random variates are drawn from a distribution with the same support as $\hat{\mathbf{F}}$ with weights $\mathbf{w}_{il}^{max}$.

4.2 Algorithm

The algorithm to obtain the MCB CIs following the NIOU-C:AS method is:

1. From observations $X_{p1},...,X_{pn}$, compute the edf, $\hat{F}_p$, as the estimator of F_p^c for each $p = 1,...,m$.
2. For each $i \in \{1,...,k\}$:
   (a) Run R_1 replications of the simulation to get $Y_{i1}(\hat{\mathbf{F}}),...,Y_{iR_1}(\hat{\mathbf{F}})$.
   (b) Compute $\bar{Y}_i = \frac{1}{R_1}\sum_{r=1}^{R_1} Y_{ir}(\hat{\mathbf{F}})$.
   (c) Compute $\hat{\hat{\mathrm{IF}}}_{ip}(X_{pj})$ in (8) for each $p = 1,...,m$ and $j = 1,...,n_p$.
3. For each pair $i \neq l$, compute the optimal solution $\mathbf{w}_{il}^{max}$ of (14).
4. For each pair $i \neq l$, compute $\hat{U}_{il} = \frac{1}{R_2}\sum_{r=1}^{R_2}\left(Y_{ir}(\mathbf{w}_{il}^{max}) - Y_{lr}(\mathbf{w}_{il}^{max})\right)$.
5. For each i, compute $D_i^+ = (\min_{l\neq i}\{\hat{U}_{il}\})^+$ and $I = \{i : D_i^+ > 0\}$. If $I = \{i\}$, $D_i^- = 0$; otherwise, $D_i^- = -(\min_{l\in I: l\neq i}\{-\hat{U}_{li}\})^-$.
6. For each i, we have the CI $[D_i^-, D_i^+]$.

The NIOU-C:AS procedure runs R_1 replications at each system to estimate the influence functions. Additionally, in Step 4, a total of $2k(k-1)R_2$ replications are made to compute $\{\hat{U}_{il}\}_{i\neq l}$ for $1 \leq i \leq k$. In summary, the total simulation cost of NIOU-C:AS is $kR_1 + 2k(k-1)R_2$. In addition, the procedure solves $k(k-1)$ convex optimization problems in Step 3. We use the `cvxpy` package implemented in Python to solve Problem (14) in the empirical study presented in Section 5.

5 EMPIRICAL STUDY

In this section we compare the performances of NIOU-C:CLT and NIOU-C:AS procedures to those of IOU-C: All-in and IOU-C: Plug-in procedures, by Song and Nelson (2019). We first describe our example below.

Let us consider a tandem queueing system with three servers, where each server has a FIFO service rule. The arrival process to the system is a Poisson process with mean $\lambda^{-1} = 0.15$ and is independent from all service times. For each $s \in \{1,2,3\}$, let S_s represent the service time of the sth server. We choose the following bimodal distribution for S_s:

$$S_s = \mathbf{1}\{Z=1\}\cdot a_s^1 \cdot \mathrm{Beta}(a_s^2, a_s^3) + \mathbf{1}\{Z=0\}\cdot a_s^4 \cdot \mathrm{Beta}(a_s^5, a_s^6),$$

where Z is a Bernoulli random variable with success probability γ_s. The parameters take the following values: $\mathbf{a}^1 = (1,1,1), \mathbf{a}^2 = (2,2,2), \mathbf{a}^3 = (6,6,6), \mathbf{a}^4 = (3,2.3,1), \mathbf{a}^5 = (10,6,12), \mathbf{a}^6 = (2,2,2)$ and $\gamma = (0.785, 0.7, 0.1)$. The mean of the three services times are approximately $\mu_1 = 0.73, \mu_2 = 0.7$ and $\mu_3 = 0.8$, respectively. To examine the effect of incorrectly specifying the parametric families, we run IOU-C procedures with the assumption that the service times are exponentially distributed.

We assume the first server has an infinite-capacity queue so that every customer is accepted into the system. However, the second and third servers have queue capacities of 2 and 3, respectively, which may cause blocking. For instance, if server 2 has no available resources and two customers are in the queue of server 2 (i.e. full capacity), then a customer finishing its service at server 1 will be blocked from joining the second queue preventing server 1 to release the resource.

The base capacity of each server is 4. We consider adding extra server capacities to the system. Let $\mathbf{c} = (2,5,6)$ be the vector of the cost per extra server at each station. With budget 9, the set of all possible systems is $\{(d_1,d_2,d_3) \in \mathbb{Z}_+ : \sum_{s=1}^{3} c_s d_s \leq 9\}$, which results in $k = 9$ systems. We define the average total waiting time of the first 100 customers as our performance measure. Table 1 summarizes the systems and their estimated performance measures from 1000 Monte Carlo simulations; the smaller, the better.

Table 1: List of feasible systems with their expected simulation outputs (waiting times).

System	1	2	3	4	5	6	7	8	9
Added Capacities	(2, 1, 0)	(1, 0, 1)	(1, 1, 0)	(3, 0, 0)	(2, 0, 0)	(1, 0, 0)	(0, 1, 0)	(0, 0, 1)	(0, 0, 0)
Expected Wait Time	2.36	2.40	2.55	2.77	2.86	3.03	3.22	3.24	3.73

We assume the three service time distributions must be estimated from data. At each macrorun we generate n_p, $1 \leq p \leq 3$, i.i.d. data from the true service time distribution. We set $n_p = n$ for each p, where $n = 100$ or $n = 400$. For IOU-C procedures, we estimate the maximum likelihood estimators (MLEs) of the exponential service time distributions from the data. We set the confidence level $1-\alpha$ to 0.9 and adopt CRNs to run the simulations. We generate exactly 100 random variates from each input distribution, because every customer enters the system and goes through all three servers. That is, $T_{ip} = 100$ for $1 \leq i \leq k$ and $1 \leq p \leq 3$.

To make fair comparisons, we assign the same total simulation budget to each method. For the IOU-C methods, Song and Nelson (2019) suggest to run, for each system i, n replications at $\hat{\mathbf{F}}$ and to sample $B = n^{1.1}$ design points from the asymptotic sampling distribution of the MLEs and to make one replication at each to fit the linear regression for each system. Let R be the total number of simulations to run at each system $i \in \{1,...,k\}$. Then, we have $R = B + n$, which accounts to the total budget of kR. We allocate the same budget for the NIOU-C methods. For NIOU-C:CLT, we run R replications at $\hat{\mathbf{F}}$ for each system, which are used to calculate both IU and stochastic error CI widths. Finally, for the NIOU-C:AS method we allocate $R/2$ to run simulations at $\hat{\mathbf{F}}$ to estimate the influence function for each system. The other half is spent for estimating the bounds $\{\hat{U}_{il}\}_{i \neq l}$.

Table 2 shows the summary results of 1000 macro runs of all methods when the number of observations for each service time distribution is $n = 100$ with two settings of simulation budget, $R = 259$ and $R = 1000$. For each method we estimate the probability that each system is in the set, I, as well as the joint coverage probability of the MCB CIs and the average size of I.

Table 2: Probability of $i \in I$ for each system i, MCB Coverage and Average Size Set when $n = 100$ for NIOU-C and IOU-C procedures using $R = 259$ and $R = 1000$. Results are computed from 1000 macroruns.

System	$R = 259$				$R = 1000$			
	NIOU-C		IOU-C		NIOU-C		IOU-C	
	CLT	AS	All-in	Plug-in	CLT	AS	All-in	Plug-in
1	1	0.923	0.993	0.701	0.988	0.99	0.987	0.671
2	1	0.92	0.999	0.979	0.998	0.995	1	0.976
3	0.999	0	0.964	0.609	0.97	0	0.917	0.567
4	0.998	0.088	0.761	0.134	0.918	0.203	0.684	0.123
5	0.996	0	0.676	0.01	0.886	0	0.599	0.007
6	0.976	0	0.382	0.005	0.479	0	0.331	0.005
7	0.973	0	0.592	0.233	0.716	0	0.527	0.22
8	0.989	0	0.681	0.668	0.732	0	0.642	0.617
9	0.541	0	0.026	0.001	0.001	0	0.019	0
MCB Coverage	1	0.842	0.969	0.631	0.984	0.986	0.968	0.592
Avg set size	8.5	1.9	6.1	3.3	6.7	2.2	5.7	3.2

Table 2 shows that, when $R = 259$, the NIOU-C:CLT method is overly conservative; the average size of I is 8.5 and the joint probability coverage is 1. In contrast, the NIOUC:AS tends to be aggressive; it selects system 1 only with probability 0.923. System 2, which is very close to the optimal, is selected with probability 0.92. The average size set is 1.9 and the coverage is 0.842. IOU-C:All-in is conservative as its I contains 6.1 systems on average with a joint coverage of 0.969. System 1 and 2 are in I with probability higher than 0.99. The IOU-C:Plug-in procedure is more aggressive, however, observe that the probability of $i = 1$ being in I is 0.701, close to that of system 8, which is the second worst. Whereas, system 2 is included in I with a significantly higher frequency. This is likely caused by the incorrect parametric assumptions.

When R is increased to 1000, NIOU-C:CLT becomes less conservative: the average size set and MCB coverage decrease to 6.7 and 0.984 respectively. Each of the best two systems are in I with a probability higher than 0.98. The NIOU-C:AS, on the other hand, becomes less aggressive: the average size set increases to 2.2 and the MCB coverage increases to 0.986. Each of the two best systems are in I with probabilities of at least 0.99, increased from when $R = 259$. These results suggest that increasing R improves the performance of each NIOU-C procedure as CI bounds are computed more precisely.

Table 2 also shows that IOU-C:All-in and IOU-C:Plug-in methods select the best system with probabilities 0.987 and 0.671, respectively. Although the former exceeds the target coverage 0.9, this is likely caused by its inherent conservatism. We expect the performance would worsen when R increases because All-in and Plug-in methods start behaving similarly for large R.

Table 3 shows the average of 1000 macro runs of all methods when $n = 400$. When $R = 1129$, the NIOU-C:CLT method has a joint coverage probability of 1 and the average size set is 5.4, both are smaller than when $n = 100$. Such reduction comes from reduced IU thanks to higher n. The NIOU-C:AS procedure, in contrast, has the joint coverage probability of 0.822 and average size set of 1.9. Moreover, it only selects systems 1 and 2 to be in I; no other system is considered to be the true best under the unknown true service distributions. On the contrary, system $i = 1$ is ruled out 2.2% of the times by the IOU-C:All-in procedure and 46.9% of the times by the IOU-C:Plug-in procedure. The MCB coverage probabilities for these methods are 0.949 and 0.412, respectively, and the average sizes set are 2.9 and 1.9, respectively. The probability of selecting the best system in IOU-C:Plug-in is far below the target indicating the model risk caused by wrong parametric assumptions.

Table 3: Probability of $i \in I$ for each system i, MCB Coverage and Average Size Set when $n = 400$ for NIOU-C and IOU-C procedures using $R = 1129$ and $R = 5000$. Results are computed from 1000 macroruns.

System	$R = 1129$				$R = 5000$			
	NIOU-C		IOU-C		NIOU-C		IOU-C	
	CLT	AS	All-in	Plug-in	CLT	AS	All-in	Plug-in
1	1	0.912	0.978	0.531	0.999	0.993	0.977	0.539
2	1	0.943	1	0.993	0.998	0.995	1	0.993
3	1	0	0.663	0.287	0.632	0	0.64	0.279
4	0.942	0	0.078	0.001	0.09	0.001	0.079	0.001
5	0.771	0	0.019	0	0	0	0.021	0
6	0	0	0	0	0	0	0	0
7	0.521	0	0.065	0.003	0	0	0.059	0.001
8	0.125	0	0.098	0.058	0	0	0.087	0.061
9	0	0	0	0	0	0	0	0
MCB Coverage	1	0.822	0.949	0.412	0.991	0.983	0.94	0.399
Avg set size	5.4	1.9	2.9	1.9	2.7	2.0	2.9	1.9

Finally, when R increases to 5000, we can observe that systems 1 and 2 are chosen over 99% of times by the NIOU-C methods. NIOU-C:CLT also selects system 3 and 4 with probability 0.632 and 0.09, which makes the average size set to be 2.7. The MCB coverage probability is 0.991 for NIOU-C:CLT and 0.983 for NIOU-C:AS. In the IOU-C procedures, the results are similar to the case when $R = 1129$.

Results in Table 3 suggest that increasing R makes the NIOU-C perform better as it should: for the NIOU-C:CLT procedure, it reduces the average size set allowing us to make a better decision while maintaining a high MCB coverage; for the NIOU-C:AS procedure, it increases the probability of selecting the best system up to 99%. The IOU-C:Plug-in method does not perform well due to the wrong parametric assumptions. IOU-C:All-in procedure shows good MCB coverage, thanks to its conservatism, but it is not expected to perform well if R increases.

In this queuing example, the NIOU-C procedures show good performances. The NIOU-C:AS method returns the smallest I among all methods and system 1 is included in I with probability over $1-\alpha$. For the NIOU-C:CLT method, the probability of system 1 being in I is always close to 1 due to its conservatism.

ACKNOWLEDGMENTS

This work is partially funded by the National Science Foundation Grant CMMI-2045400.

REFERENCES

Barton, R. R., H. Lam, and E. Song. 2022. "Input Uncertainty in Stochastic Simulation". In *The Palgrave Handbook of Operations Research*, edited by S. Salhi and J. Boylan, 573–620. Cham: Springer International Publishing.

Chang, J. Y. and J. C. Hsu. 1992. "Optimal Designs for Multiple Comparisons with the Best". *Journal of Statistical Planning and Inference* 30(1):45–62.

Corlu, C. G. and B. Biller. 2013. "A Subset Selection Procedure Under Input Parameter Uncertainty". In *2013 Winter Simulations Conference (WSC)*, 463–473 https://doi.org/10.1109/WSC.2013.6721442.

Corlu, C. G. and B. Biller. 2015. "Subset Selection for Simulations Accounting for Input Uncertainty". In *2015 Winter Simulation Conference (WSC)*, 437–446 https://doi.org/10.1109/WSC.2015.7408185.

Gao, S., H. Xiao, E. Zhou, and W. Chen. 2017. "Robust Ranking and Selection with Optimal Computing Budget Allocation". *Automatica* 81:30–36.

Hampel, F. R. 1974. "The Influence Curve and its Role in Robust Estimation". *Journal of the american statistical association* 69(346):383–393.

Lam, H. and H. Qian. 2016. "The Empirical Likelihood Approach to Simulation Input Uncertainty". In *2016 Winter Simulation Conference (WSC)*, 791–802 https://doi.org/10.1109/WSC.2016.7822142.

Lam, H. and Qian, H. 2019. "Optimization-based Quantification of Simulation Input Uncertainty via Empirical Likelihood". *arXiv preprint arXiv: 1707.05917v2*.

Owen, A. B. 2001. *Empirical Likelihood*. 1st ed. New York: Chapman and Hall/CRC.

Pearce, M. and J. Branke. 2017. "Bayesian Simulation Optimization with Input Uncertainty". In *2017 Winter Simulation Conference (WSC)*, 2268–2278 https://doi.org/10.1109/WSC.2017.8247958.

Song, E. and B. L. Nelson. 2019. "Input-Output Uncertainty Comparisons for Discrete Optimization via Simulation". *Oper. Res.* 67(2):562–576.

Song, E., B. L. Nelson, and L. J. Hong. 2015. "Input Uncertainty and Indifference-zone Ranking & Selection". In *2015 Winter Simulation Conference (WSC)*, 414–424 https://doi.org/10.1109/WSC.2015.7408183.

AUTHOR BIOGRAPHIES

JAIME GONZALEZ is a Ph.D. student in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. His research interests are input uncertainty quantification and simulation optimization. His e-mail address is jhodar3@gatech.edu.

JOHANNES MILZ is an Assistant Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. His research interests include optimization under uncertainty and optimal control of uncertain systems. He received his Ph.D. degree in Applied Mathematics from the Technical University of Munich. His email address is johannes.milz@isye.gatech.edu and his website is https://milzj.github.io.

EUNHYE SONG is a Coca-Cola Foundation Early Career Professor and Assistant Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. She earned her PhD degree in Industrial Engineering and Management Sciences at Northwestern University. Her research interests include simulation design of experiments, uncertainty and risk quantification, and simulation optimization. Her email address is eunhye.song@isye.gatech.edu. Her website is http://eunhyesong.info.