Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

JOINT RESOURCE ALLOCATION FOR INPUT DATA COLLECTION AND SIMULATION

Jingxu Xu
Zeyu Zheng

Department of Industrial Engineering
and Operations Research
University of California, Berkeley
Berkeley, CA 94720, USA

Peter W. Glynn

Department of Management Science
and Engineering
Stanford University
Stanford, CA 94305, USA

ABSTRACT

Simulation is often used to evaluate and compare performances of stochastic systems, where the underlying stochastic models are estimated from real-world input data. Collecting more input data can derive closer-to-reality stochastic models while generating more simulation replications can reduce stochastic errors. With the objective of selecting the system with the best performance, we propose a general framework to analyze the joint resource allocation problem for collecting input data and generating simulation replications. Two commonly arised features, correlation in input data and common random numbers in simulation, are jointly exploited to save cost and enhance efficiency. In presence of both features, closed-form joint resource allocation solutions are given for the comparison of two systems.

1 INTRODUCTION

The need to compare the expected performances of two or multiple stochastic systems naturally arise in the areas of healthcare, supply chain, logistics, production, queueing systems, portfolio management, among others. The stochastic systems under consideration typically involve uncertainties that are captured by a set of probability models. The associated probability distributions are either specified by domain experts or estimated from data. These probability models, serving as inputs to the stochastic system, can be the customer arrival and service processes in a service system or the daily demands and lead times in a supply chain system. After the input probability distributions are specified or estimated, simulation is often used to evaluate the expected system performance, particularly in cases where analytical solutions are not available.

This paper presumes that there is independent and identically distributed data available or that can be collected that faithfully represents the true input distributions. When the input probability distributions are estimated from data, there exists a statistical estimation error due to the finite amount of data collected. The statistical estimation error from the input distributions may propagate and cause uncertainties in the evaluation of system performances. The resulting uncertainty in the performance evaluation caused by using an incorrectly specified input distribution (due to estimation error) is called *input uncertainty*. The input uncertainty cannot be eliminated by increasing the number of simulation replications used to evaluate the system performances, but can potentially be reduced by collecting more data. In many applications, the data to be collected is multi-dimensional and is usually generated sequentially from real operations. Therefore typically the available data that can be collected in one operational period is corrected. Consider a site selection problem where a manager chooses one out of m sites to run a new branch store. The manager may need to run simulation to evaluate revenues for each site with different store designs. One of the key input distributions is the daily traffic flows. On day i, the full set of data that may be collected is $\mathscr{A}_i = (A_{i,1}, A_{i,2}, \ldots, A_{i,m})$ in which $A_{i,j}$ denotes the traffic flow at the j-th site on day i. Due to the nature of the data generation process, $(A_{i,1}, A_{i,2}, \ldots, A_{i,m})$ are typically correlated, possibly due to common

unobserved features from the same day and other observed dependencies such as the adjacency between sites. Ideally, collecting this full set of data $\mathscr{A}_i$ is preferred and provides the best possible information, but this can be costly or even inaccessible. On the other hand, when the correlation structure within the full set of input data generated in the same period is exploited, it may not be necessary to collect the full set of data. Therefore this correlation structure may be used to save input data collection cost.

In this paper, we propose a general framework to study the joint resource allocation problem for input data collection and simulation. The objective is appropriately allocating resource to maximize the probability of correctly selecting the system with the best performance. Two commonly arised features, correlation in input data and common random numbers in simulation, are jointly exploited to save costs. For input data collection, in presence of the correlation structure among different sources, options are available to either jointly collect data simultaneously from the different sources, or to collect data solely from a particular source. For simulation, one has the option to either use common random numbers to simultaneously evaluate performances for different systems or to evaluate independently a single system. We provide closed-form optimal resource allocation solutions that maximize the asymptotic probability of correct selection, when the resource budget is large. Our results explicitly show that how the correlation structure is exploited to save costs and improve performances, and how the optimal resource allocation strategy depends on the correlation structure.

Two scenarios are considered in our framework. First, we consider scenarios (in Section 3) where the "monetary" cost of generating a simulation replication is much smaller than the cost of collecting a sample of input data. For example, these situations happen when the scale and structure of the problem permits the generation of simulation replications efficiently even on personal computers, while the input data needs to be purchased from a data vendor at a significant price or needs to be collected by multiple staff throughout a number of days. One may then assume a simplification that once the input distributions are estimated, the expected performance evaluation via simulation is immediately available at no cost. Therefore in these situations, the resource allocation problem focuses entirely on the input data collection part. Second, we consider scenarios (in Section 5) where the simulation cost is not negligible compared with the input data collection cost. These scenarios arise in performance evaluation for complicated systems, in which high performance computing resources are needed for simulation. Otherwise if not using designated high performance computing resources, the simulation may take too long a time. The simulation costs therefore may be evaluated by monetary costs for purchasing computing resources or by the opportunity costs for long simulation time. In these scenarios, the optimal resource allocation strikes a balance between the input data collection costs and simulation costs, to jointly control input uncertainty and simulation error.

2 LITERATURE REVIEW

Comparing the expected performances of two or multiple systems via simulation is a fundamental component in the problems of Ranking and Selection (R&S) and Discrete Optimization via Simulation (DOvS). When one knows explicitly the input distributions, or has the ability to generate simulation replications from the true input distributions, the focus of these problems is then on developing efficient simulation procedures to select the best system. Procedures developed in the literature typically adopt a frequentist or a Bayesian view. See Kim and Nelson (2007) and Chen et al. (2015) for an overview. Our paper follows a frequentist perspective.

When the input distributions are not explicitly known, or when one does not have the ability to generate samples from the true input distributions, a series of works discuss the quantification of input uncertainty and its impact on the comparison of system performances. See Cheng and Holloand (1997), Chick (2001), Barton et al. (2014), Xie et al. (2014), Song et al. (2015), Corlu and Biller (2013), Wu and Zhou (2017), Wu and Zhou (2018) among others. The closest to our work are Wu and Zhou (2017) and Song and Nelson (2019). Song and Nelson (2019) exploits the effect of common input distribution to reduce the uncertainty in system performances comparison, when the input data set is given. They construct valid confidence intervals for system comparison that incorporate input uncertainty, the common input distribution effect,

and simulation uncertainty. Wu and Zhou (2017) allows the collection of additional input data from multiple independent sources and discusses the optimal resource allocation for input data collection and simulation. Our paper is different to the literature in three folds: (1) we exploit the correlation structure in the input data to reduce the input data collection cost, and show how the correlation structure impacts the optimal resource allocation strategy. (2) we propose a general framework that integrates input data collection and simulation in which the data collection and simulation costs themselves can be random; (3) we investigate the joint optimal resource allocation when both the correlation in the input data and the use of common random numbers in the simulation procedures are exploited.

When the performance of each system is evaluated independently, the best system selection problem shares the same formulation as the *best arm identification* problem; see Karnin et al. (2013), Kaufmann and Kalyanakrishnan (2013), Glynn and Juneja (2015), Ryzhov (2016), Russo (2020) among others. In fact, if we separate the input data collection problem and the best system selection problem, each problem shares a very similar formulation with a best arm identification problem, provided that the observation of each dimension in the data is independent and that the performance evaluation of each system is independent. Differences emerge when correlation is present either in the input data collection or among simulation evaluations.

The use of common random numbers (CRN) in the simulation procedures for R& S and DOvS have been widely discussed. See Yang and Nelson (1991), Glasserman and Vakili (1994), Nelson and Matejcik (1995), Dai and Chen (1997), Kim and Nelson (2001) among others. Specifically, Fu et al. (2007) discusses the optimal allocation of simulation replications on each system when the CRN technique is used. Their work did not discuss errors created from input data. We propose an alternative framework that allows the simulation costs to be random, and consider the joint resource allocation problem for both input data collection and simulation.

3 A GENERAL FRAMEWORK

We introduce a basic and general framework that allows us to integrate simulation for performance evaluation and input data collection for input distribution estimation. In this framework, the costs for simulation and input data collection can be random. We first focus on describing the input data collection and input distribution estimation in the framework.

Consider a set of systems labeled by index set $[m] = \{1, 2, \ldots, m\}$. When $m = 2$, for example, there are two systems to compare. For $i \in [m]$, the expected performance for system i is given by $\alpha_i(\theta_i^*)$, where $\alpha_i : \mathbb{R}^{d_i} \to \mathbb{R}$ is a continuously differentiable function, and $\theta_i^* \in \mathbb{R}^{d_i}$ is the true input distribution parameter associated with system i (e.g., arrival and service rates, lead time expectation, shape parameters, etc.). We first consider scenarios where the simulation cost is negligible compared to input data collection cost, so that the expected performance function $\alpha_i(\cdot)$'s are viewed to be available at no cost whenever the input distribution is specified. We recognize that the input distribution parameters $\theta_1^*, \theta_2^*, \ldots, \theta_m^*$ need to be estimated. This can be based on common observations $(\tilde{X}_{ij} : i \in [m], j \geq 1)$ or based on independently gathered observations $(X_{ij} : j \geq 1)$ for $i \in [m]$. We assume that the cost of collecting the j-th copy of a set of common observations $(\tilde{X}_{ij} : i \in [m])$ collected simultaneously is given by $\tilde{\tau}_j$, while the cost of collecting individual X_{ij} is given by τ_{ij}. These data collection costs can be random variables themselves. Then, with a budget c in hand, we can either collect

$$\tilde{N}(c) = \max\{n \geq 0 : \tilde{\tau}_1 + \ldots + \tilde{\tau}_n \leq c\}$$

copies of "common observations" or

$$N_i(c) = \max\{n \geq 0 : \tau_{i1} + \ldots + \tau_{in} \leq c\}$$

copies of observations solely from system i. In this framework, we assume that

1. $(\tilde{\tau}_j, (\tilde{X}_{ij} : i \in [m])) : j \geq 1)$, $(\tau_{1j}, X_{1j} : j \geq 1)$, …, $(\tau_{mj}, X_{mj} : j \geq 1)$ are independent sequences.

2. $(\tau_i, (\tilde{X}_{ij} : i \in [m]) : j \geq 1)$ is iid in j.
3. For each $i \in [m]$, $((\tau_{ij}, X_{ij}) : j \geq 1)$ is iid in j.
4. $\tilde{X}_{i1} \stackrel{\mathscr{D}}{=} X_{i1}$ for $i \in [m]$. $\mathrm{Var}(X_{i1}) < \infty$ for $i \in [m]$.
5. $\mathbb{E}\,\tilde{\tau}_1 < \infty$, $\mathbb{E}\,\tau_{i1} < \infty$, for $i \in [m]$.

In practice, it is often the case that $\mathbb{E}\,\tau_{i1} < \mathbb{E}\,\tilde{\tau}_1$ for $i \in [m]$ and $\mathbb{E}\,\tilde{\tau}_1 \leq \sum_{i=1}^{m} \mathbb{E}\,\tau_{i1}$. Since $\tilde{N}(\cdot)$ and $N_i(\cdot)$'s are renewal counting processes, it is known that

$$\frac{1}{c}\tilde{N}(c) \stackrel{a.s.}{\to} \lambda \triangleq \frac{1}{\mathbb{E}\,\tilde{\tau}_1}$$

and

$$\frac{1}{c}N_i(c) \stackrel{a.s.}{\to} \lambda_i \triangleq \frac{1}{\mathbb{E}\,\tau_{i1}}.$$

Given an overall budget c, suppose we allocate a fraction p to collecting common observations, and a fraction p_i to collecting independent observations from system i, where

$$p + p_1 + p_2 + \ldots + p_m = 1,$$

with $p \geq 0, p_i \geq 0, i \in [m]$.

We now assume that one estimates θ_i^* via an maximum likelihood estimator (MLE) $\hat{\theta}_i$, where $\hat{\theta}_i$ maximizes the likelihood

$$\prod_{j=1}^{\tilde{N}(pc)} f_i(\theta, \tilde{X}_{ij}) \prod_{j=1}^{N_i(p_i c)} f_i(\theta, X_{ij}),$$

where $f_i(\theta, \cdot)$ is the marginal probability density function of the input data for the i'th system. By taking the logarithm of the likelihood function

$$\tilde{L}_{ij}(\theta) = \log f_i(\theta, \tilde{X}_{ij}),$$
$$L_{ij}(\theta) = \log f_i(\theta, X_{ij}),$$

maximizing the likelihood is equivalent to maximizing the log-likelihood, given by

$$\sum_{j=1}^{\tilde{N}(pc)} \tilde{L}_{ij}(\theta) + \sum_{j=1}^{N_i(p_i c)} L_{ij}(\theta).$$

Under appropriate technical conditions, the maximum likelihood estimator $\hat{\theta}_i$ satisfies

$$\sum_{j=1}^{\tilde{N}(pc)} \nabla \tilde{L}_{ij}(\hat{\theta}_i) + \sum_{j=1}^{N_i(p_i c)} \nabla L_{ij}(\hat{\theta}_i) = 0$$

with $\hat{\theta}_i \to \theta_i^*$ almost surely as $c \to \infty$. Note that

$$\sum_{j=1}^{\tilde{N}(pc)} \left[\nabla \tilde{L}_{ij}(\hat{\theta}_i) - \nabla \tilde{L}_{ij}(\theta_i^*)\right] + \sum_{j=1}^{N_i(p_i c)} \left[\nabla L_{ij}(\hat{\theta}_i) - \nabla L_{ij}(\theta_i^*)\right] = -\sum_{j=1}^{\tilde{N}(pc)} \nabla \tilde{L}_{ij}(\theta_i^*) - \sum_{j=1}^{N_i(p_i c)} \nabla L_{ij}(\theta_i^*).$$

We adopt the convention that the gradient is a row vector. If $L_{ij}(\cdot)$ is appropriately smooth, then the mean value theorem implies that

$$\sqrt{c}(\hat{\theta}_i - \theta_i^*)\Big(\frac{\tilde{N}(pc)}{c} H_i + \frac{N_i(p_i c)}{c} H_i + o_p(1)\Big) = -\frac{\sum_{j=1}^{\tilde{N}(pc)} \nabla \tilde{L}_{ij}(\theta_i^*)}{\sqrt{c}} - \frac{\sum_{j=1}^{N_i(p_i c)} \nabla L_{ij}(\theta_i^*)}{\sqrt{c}},$$

where the notion $o_P(1)$ indicates a small random quantity that weakly converges to zero as $c \to \infty$. The Hessian matrix H_i is given by

$$H_i = \big(\mathbb{E}\frac{\partial^2}{\partial\theta_k\partial\theta_l}L_{i1}(\theta^*) : 1 \le k \le l \le d_i\big).$$

Assume that H_i is negative definite so that $\det H_i$ is non-singular. Then, when c is large,

$$c^{\frac{1}{2}}(\hat{\theta}_i - \theta_i^*) = -\frac{1}{\lambda p + \lambda_i p_i}\Big(\sum_{j=1}^{\tilde{N}(pc)}\frac{\nabla\tilde{L}_{ij}(\theta^*)H_i^{-1}}{\sqrt{c}} + \sum_{j=1}^{N_i(p_ic)}\frac{\nabla L_{ij}(\theta^*)H_i^{-1}}{\sqrt{c}}\Big) + o_p(1).$$

Hence, we have the following result.

Theorem 1 Assume that for $i \in [m]$, there exists an open subset w_i that is a subset of the feasible parameter region where the true parameter $\theta_i^* \in w_i$, and that all third-order partial derivatives of the log-likelihood functions with respect to input parameters are uniformly bounded for $\theta_i \in w_i$. When $c \to \infty$,

$$c^{\frac{1}{2}}((\hat{\theta}_i - \theta_i^*) : i \in [m]) \Rightarrow \Big(-\frac{1}{\lambda p + \lambda_i p_i}\big(\sqrt{\lambda p}\,\tilde{G}_i + \sqrt{\lambda_i p_i}\,G_i\big) : i \in [m]\Big),$$

where:

- $(\tilde{G}_1, \ldots, \tilde{G}_m)$ is jointly Gaussian with mean 0.
- $\tilde{G}_i \stackrel{\mathcal{D}}{=} G_i$, where the covariance matrix of G_i is given by $H_i^{-1}\mathbb{E}\nabla L_{i1}(\theta^*)^\top \nabla L_{i1}(\theta^*)H_i^{-1}$.
- The random variables (rv's) $G_1, G_2, \ldots, G_m$ are independent and independent of $(\tilde{G}_1, \ldots, \tilde{G}_m)$.

Then, with Theorem 1 in hand,

$$\begin{aligned} c^{\frac{1}{2}}(\alpha_i(\hat{\theta}_i) - \alpha_i(\theta_i^*)) &= (\hat{\theta}_i - \theta_i^*)\nabla\alpha_i(\theta^*)^\top + o_p(1) \\ &= -\frac{1}{\lambda_p + \lambda_i p_i}(\sqrt{\lambda p}\tilde{G}_i + \sqrt{\lambda_i p_i}G_i)\nabla\alpha_i(\theta^*)^\top + o_p(1) \\ &= -\frac{1}{\lambda_p + \lambda_i p_i}(\sqrt{\lambda p}\tilde{\mathscr{G}}_i + \sqrt{\lambda_i p_i}\mathscr{G}_i) + o_p(1) \end{aligned}$$

as $c \to \infty$, where $\tilde{\mathscr{G}}_i = \tilde{G}_i\nabla\alpha_i(\theta^*)^\top, \mathscr{G}_i = G_i\nabla\alpha_i(\theta^*)^\top, i \in [m]$. Note that $\tilde{\mathscr{G}}_i \stackrel{\mathcal{D}}{=} \mathscr{G}_i, i \in [m]$. Hence, when $c \to \infty$,

$$c^{\frac{1}{2}}(\alpha_i(\hat{\theta}_i) - \alpha_j(\hat{\theta}_j) - (\alpha_i(\theta_i^*) - \alpha_j(\theta_j^*))) \Rightarrow -\frac{\sqrt{\lambda p}}{\lambda p + \lambda_i p_i}\tilde{\mathscr{G}}_i + \frac{\sqrt{\lambda p}}{\lambda p + \lambda_j p_j}\tilde{\mathscr{G}}_j - \frac{\sqrt{\lambda_i p_i}}{\lambda p + \lambda_i p_i}\mathscr{G}_i + \frac{\sqrt{\lambda_j p_j}}{\lambda p + \lambda_j p_j}\mathscr{G}_i.$$

Denote the right-hand-side above as W_{ij}. The random variable (rv) W_{ij} is Gaussian with mean zero and variance

$$\frac{\sigma_i^2}{\lambda p + \lambda_i p_i} + \frac{\sigma_j^2}{\lambda p + \lambda_j p_j} - \frac{2\lambda p\, c_{ij}}{(\lambda p + \lambda_i p_i)(\lambda p + \lambda_j p_j)},$$

where $\sigma_i^2 = \mathrm{Var}\mathscr{G}_i$ and $c_{ij} = \mathrm{Cov}(\tilde{\mathscr{G}}_i, \tilde{\mathscr{G}}_j)$. We further define $\rho_{ij} = \mathrm{Corr}(\tilde{\mathscr{G}}_i, \tilde{\mathscr{G}}_j)$.

3.1 Optimization Problem for Input Data Collection

With the given framework, when comparing systems i and j, the input data collection problem can be summarized into the following optimization problem

$$\min_{p\geq 0,\, p_i\geq 0,\, p_j\geq 0,\, p+p_i+p_j=1} \frac{\sigma_i^2}{\lambda p+\lambda_i p_i}+\frac{\sigma_j^2}{\lambda p+\lambda_j p_j}-\frac{2\lambda p c_{ij}}{(\lambda p+\lambda_i p_i)(\lambda p+\lambda_j p_j)}.$$

Recall that $c_{ij}=\rho_{ij}\sigma_i\sigma_j$. An equivalent formulation is to set $q_i=\lambda_i p_i$, $q=\lambda p$, and minimize

$$\frac{\sigma_i^2}{q+q_i}+\frac{\sigma_j^2}{q+q_j}-\frac{2q\rho_{ij}\sigma_i\sigma_j}{(q+q_i)(q+q_j)} \tag{1}$$

subject to $\frac{q}{\lambda}+\frac{q_i}{\lambda_i}+\frac{q_j}{\lambda_j}=1$.

4 OPTIMAL RESOURCE ALLOCATION FOR INPUT DATA COLLECTION

The optimization problem given by (1) to solve the optimal resource allocation for input data collection turns out to be non-convex. The non-convexity is exactly caused by the correlation feature and creates difficulty in modeling how the correlation ρ_{ij} presented in the input data exactly affects the optimal resource allocation. The following theorem shows that the optimal objective function can be obtained at the boundary of feasible region $\mathscr{S}=\{(q_i,q_j,q)^\top : q_i,q_j,q\geq 0, \frac{q}{\lambda}+\frac{q_i}{\lambda_i}+\frac{q_j}{\lambda_j}=1\}\subset\mathbb{R}^3$.

Theorem 2 For the optimization problem (1), there exists a solution (q_i^*,q_j^*,q^*) that achieves the optimal objective value and has at least one element as zero.

With Theorem 2 in hand, it suffices to explore the resource allocation strategies among $(q_i,0,q)$, $(q_i,q_j,0)$ and $(0,q_j,q)$. As we will show later, this leads to a closed-form representation for the optimal allocation. The proof of Theorem 2 is given as follows.

Proof. For problem (1), let $\mathscr{P}\subset\mathscr{S}$ be the set of globally optimal solutions. Denote $s_i:=\frac{1}{\lambda_i}=\mathbb{E}\,\tau_{i1}$, $s_j:=\frac{1}{\lambda_j}=\mathbb{E}\,\tau_{j1}$, $s:=\frac{1}{\lambda}=\mathbb{E}\,\tilde{\tau}_1$ and $v(q_i,q_j,q):=\frac{\sigma_i^2}{q_i}+\frac{\sigma_j^2}{q_j}-\frac{2q\rho_{ij}\sigma_i\sigma_j}{(q+q_i)(q+q_j)}$. Because all the constraints of problem (1) are linear in q_i,q_j,q, if $(q_i,q_j,q)^\top\in\mathscr{P}$, it satisfies the following Karush–Kuhn–Tucker (KKT) conditions (see, for example, Lemma 5.1.4 from Bazaraa et al. (2013)):

$$\begin{aligned}
&s_iq_i+s_jq_j+sq=1,\\
&\tfrac{\partial v}{\partial q_i}-\mu_i+s_iu=0,\\
&\tfrac{\partial v}{\partial q_j}-\mu_j+s_ju=0,\\
&\tfrac{\partial v}{\partial q}-\mu+su=0,\\
&\mu_iq_i=\mu_jq_j=\mu q=0,\\
&\mu_i,\mu_j,\mu\geq 0,
\end{aligned} \tag{2}$$

where μ_i,μ_j,μ,u are KKT multipliers. The KKT condition describes a necessary condition for the optimality of a feasible point, for which the gradient of the objective function at the feasible point should be orthogonal to the feasible set $\mathscr{S}$. Suppose that $(q_i^*,q_j^*,q^*)^\top\in\mathscr{P}$ satisfies $q_i^*,q_j^*,q^*>0$. Then, according to (2), KKT multipliers μ_i, μ_j and μ are equal to 0. Thus, we have

$$\begin{aligned}
&\Big(\frac{\partial}{\partial q_i}v(q_i^*,q_j^*,q^*),\frac{\partial}{\partial q_j}v(q_i^*,q_j^*,q^*),\frac{\partial}{\partial q}v(q_i^*,q_j^*,q^*)-\frac{\partial}{\partial q_i}v(q_i^*,q_j^*,q^*)-\frac{\partial}{\partial q_j}v(q_i^*,q_j^*,q^*)\Big)\\
=&(-s_iu,-s_ju,(-s+s_i+s_j)u).
\end{aligned} \tag{3}$$

By multiplying both sides of (3) by $(q_1^*+q^*)^2(q_2^*+q^*)^2$, calculating the gradient of $v(q_i,q_j,q)$ and eliminating KKT multiplier u, we have a system of linear equations about q_i^*,q_j^*,q^*:

$$\begin{aligned}
&s_i q_i^* + s_i q_i^* + s q^* = 1,\\
&2\rho_{ij}\sigma_i\sigma_j s_i(q_i^*+q^*) + (s_i+s_j-s)\left(\sigma_i^2\left(q_j^*+q^*\right) - 2\rho_{ij}\sigma_i\sigma_j q^*\right) = 0,\\
&2\rho_{ij}\sigma_i\sigma_j s_j(q_j^*+q^*) + (s_i+s_j-s)\left(\sigma_j^2\left(q_i^*+q^*\right) - 2\rho_{ij}\sigma_i\sigma_j q^*\right) = 0.
\end{aligned} \tag{4}$$

We discuss 3 different cases:

Case I: If $\rho_{ij}\neq 0$, and one of the following equations holds: ① $s_i+s_j=s$, ② $(s_i+s_j-s)\,\sigma_i^2 = 2s_j\rho_{ij}\sigma_i\sigma_j$, ③ $(s_i+s_j-s)\,\sigma_j^2 = 2s_i\rho_{ij}\sigma_i\sigma_j$, then (4) has no solutions. Therefore, there does not exist a KKT point that has three positive elements.

Case II: If $\rho_{ij}\neq 0$ and none of ①, ②, ③ hold, then (4) has a unique solution. We show that this solution is not optimal for (1), which contradicts the optimality of $(q_i^*,q_j^*,q^*)^\top$. Let $d=(d_i,d_j,d_{ij})^\top$ be a feasible direction such that $s_id_i+s_jd_j+sd_{ij}=0$, and there exists $\varepsilon_0>0$ enough small such that $(q_i^*,q_j^*,q^*)^\top+\varepsilon_0 d$ is still feasible. Denote the value, gradient and Hessian matrix of v at $(q_i^*,q_j^*,q^*)^\top$ to be respectively v^*, g^* and Q^*. We have

$$v\left(q_i^*+\varepsilon d_i, q_j^*+\varepsilon d_j, q^*+\varepsilon d_{ij}\right) = v^* + \varepsilon d^\top g^* + \frac{1}{2}\varepsilon^2 d^\top Q^* d + o\left(||\varepsilon d||^2\right).$$

The fact that $(q_i^*,q_j^*,q^*)^\top$ is a KKT point ensures that $d^\top g^*=0$. We have the following subcases:

(i) $s_i+s_j>s$, $\rho_{ij}<0$ or $s_i+s_j<s$, $\rho_{ij}>0$. In this case, we choose $d=(-s,-s,s_i+s_j)^\top$, and find that

$$d^\top Q^* d = \frac{(s_i+s_j-s)\left(\sigma_j^2(s_i+s_j-s)-2\rho_{ij}\sigma_i\sigma_j s_i\right)^2\left(\sigma_i^2(s_i+s_j-s)-2\rho_{ij}\sigma_i\sigma_j s_j\right)^2}{4(\rho_{ij}\sigma_i\sigma_j)^3}.$$

Since $\frac{s_i+s_j-s}{\rho_{ij}^3}<0$, $d^\top Q^* d<0$, there exists $0<\varepsilon_1<\varepsilon_0$ such that $v(q_i^*+\varepsilon_1 d_i, q_j^*+\varepsilon_1 d_j, q^*+\varepsilon_1 d_{ij})<v^*$. Therefore $(q_i^*,q_j^*,q^*)^\top$ is not a minimal point.

(ii) $s_i+s_j<s$, $\rho_{ij}<0$ or $s_i+s_j>s$, $\rho_{ij}>0$. We choose $d=(-s,0,s_i)^\top$, and calculate that

$$d^\top Q^* d = \frac{s_i(s_i-s)\left(\sigma_j^2(s_i+s_j-s)-2\rho_{ij}\sigma_i\sigma_j s_i\right)^2\left(\sigma_i^2(s_i+s_j-s)-2\rho_{ij}\sigma_i\sigma_j s_j\right)^2}{4(s_i+s_j-s)(\rho_{ij}\sigma_i\sigma_j)^3}.$$

Since we have assumed that $s_i<s$, $d^\top Q^* d<0$. Similarly to case (i), $(q_i^*,q_j^*,q^*)^\top$ is not a minimal point.

Case III: When $\rho_{ij}=0$, we discuss about three subcases: $s_i+s_j>s$, $s_i+s_j<s$ and $s_i+s_j=s$. If $s_i+s_j>s$, for any point $(q_i,q_j,q)^\top\in\mathscr{S}$ satisfying $q_i,q_j,q>0$, $d=(-s,-s,s_i+s_j)^\top$ is a descent direction for both $\frac{\sigma_i^2}{q_i+q}$ and $\frac{\sigma_j^2}{q_j+q}$ as a function of (q_i,q_j,q). Therefore, there exists no optimal solution that has three positive elements. If $s_i+s_j<s$, let $d=(s,s,-s_i-s_j)^\top$, we reach the same conclusion as when $s_i+s_j>s$. If $s_i+s_j=s$, define $w_i:=q_i+q$ and $w_j:=q_j+q$. The problem (1) is then converted to

$$\begin{aligned}
\min_{w_i,w_j\geq 0}\quad & \frac{\sigma_i^2}{w_i}+\frac{\sigma_j^2}{w_j}\\
s.t.\quad & s_iw_i+s_jw_j=1.
\end{aligned}$$

The optimal w_i^* and w_j^* satisfies $\frac{\sigma_i}{w_i^*\sqrt{s_i}}=\frac{\sigma_j}{w_j^*\sqrt{s_j}}$. So in this case, $(q_i,q_j,q)^\top\in\mathscr{P}$ if and only if $\frac{\sigma_i}{\sqrt{s_i}(q_i+q)}=\frac{\sigma_j}{\sqrt{s_j}(q_j+q)}$.

Summarizing the conclusions of case I, II and III, whenever $s_i + s_j \neq s$ or $\rho_{ij} \neq 0$, the optimal solution can only be achieved on the boundary of the feasible region. That is, the optimal solution(s) must have at least one element as zero. When $s_i + s_j = s$ and $\rho_{ij} = 0$, there exists an optimal solution that has at least one zero element and also an optimal solution that has all positive elements. □

4.1 Closed-form Optimal Allocation Formula and Interpretation

Denote $s_i := \frac{1}{\lambda_i}$, $s_j := \frac{1}{\lambda_j}$, $s := \frac{1}{\lambda}$, $v(q_i, q_j, q) := \frac{\sigma_i^2}{q_i} + \frac{\sigma_j^2}{q_j} - \frac{2q\rho_{ij}\sigma_i\sigma_j}{(q+q_i)(q+q_j)}$. Theorem 2 implies that

$$\min_{(q_i,q_j,q)^\top \in \mathscr{S}} v(q_i, q_j, q) = \min\left\{b_i^*, b_j^*, \tilde{b}^*\right\},$$

where $b_i^* = \min_{(q_i,q_j,q)^\top \in \mathscr{S}, q_i=0} v(q_i, q_j, q)$, $b_j^* = \min_{(q_i,q_j,q)^\top \in \mathscr{S}, q_j=0} v(q_i, q_j, q)$, $\tilde{b}^* = \min_{(q_i,q_j,q)^\top \in \mathscr{S}, q=0} v(q_i, q_j, q)$. We provide the closed-form value for $b_i^*, b_j^*, \tilde{b}^*$ and the associated optimizers.

$$b_i^* = \begin{cases} \left(\sqrt{(-s_j+s)\,\sigma_i^2} + \sqrt{s_j(\sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j)}\right)^2, \text{ if } s_j\sigma_i^2 < \left(\sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j\right)(s-s_j), \\ s(\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j), \text{ if } s_j\sigma_i^2 \geq \left(\sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j\right)(s-s_j). \end{cases}$$

The optimal solution to achieve b_i^* is

$$(q_i, q_j, q) = \left(0, \frac{\sqrt{(\sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j)(s-s_j)} - \sqrt{s_j\sigma_i^2}}{(s-s_j)\sqrt{s_j\sigma_i^2} + s_j\sqrt{(\sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j)(s-s_j)}}, \right.$$
$$\left. \frac{1}{s} - \frac{s_j\sqrt{\left(\sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j\right)(s-s_j)} - \sqrt{s_j^3\sigma_i^2}}{s(s-s_j)\sqrt{s_j\sigma_i^2} + s_j s\sqrt{(\sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j)(s-s_j)}}\right)$$

and $(q_i, q_j, q) = (0, 0, \frac{1}{s})$ respectively under the two conditions.

$$b_j^* = \begin{cases} \left(\sqrt{(-s_i+s)\,\sigma_j^2} + \sqrt{s_i(\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j)}\right)^2, \text{ if } s_i\sigma_j^2 < \left(\sigma_i^2 - 2\rho_{ij}\sigma\sigma_j\right)(s-s_i), \\ s(\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j), \text{ if } s_i\sigma_j^2 \geq \left(\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j\right)(s-s_i). \end{cases}$$

The optimal solution to achieve b_j^* is

$$(q_i, q_j, q) = \left(\frac{\sqrt{\left(\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j\right)(s-s_i)} - \sqrt{s_i\sigma_j^2}}{(s-s_i)\sqrt{s_i\sigma_j^2} + s_i\sqrt{\left(\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j\right)(s-s_i)}}, 0, \right.$$
$$\left. \frac{1}{s} - \frac{s_i\sqrt{\left(\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j\right)(s-s_i)} - \sqrt{s_i^3\sigma_j^2}}{s(s-s_i)\sqrt{s_i\sigma_j^2} + s_i s\sqrt{\left(\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j\right)(s-s_i)}}\right)$$

and $(q_i, q_j, q) = (0, 0, \frac{1}{s})$ respectively under the two conditions.

$$\tilde{b}^* = (\sqrt{s_i}\sigma_i + \sqrt{s_j}\sigma_j)^2.$$

The optimal solution to achieve $\tilde{b}^*$ is $(q_i, q_j, q) = (\frac{\sigma_i}{\sqrt{s_i}(\sqrt{s_i}\sigma_i + \sqrt{s_j}\sigma_j)}, \frac{\sigma_j}{\sqrt{s_j}(\sqrt{s_i}\sigma_i + \sqrt{s_j}\sigma_j)}, 0)$.

As shown in the closed-form solution for $\min_{(q_i,q_j,q)^\top \in \mathscr{S}} v(q_i,q_j,q)$, the optimal resource allocation critically depends on the sign and magnitude of the correlation ρ_{ij}. We discuss how the different value of ρ_{ij} affects the optimal allocation. First, consider the case when $s = s_i + s_j$. This corresponds to the case of additive input data collection costs. If $\rho_{ij} \leq 0$, the optimal allocation is to independently collect data for system i and j, and the optimal fraction q_i^*, q_j^* satisfies $\frac{\sigma_i}{q_i^*\sqrt{s_i}} = \frac{\sigma_j}{q_j^*\sqrt{s_j}}$. Specifically if $\rho_{ij} = 0$, any allocation (q, q_i, q_j) that satisfies $\frac{\sigma_i}{\sqrt{s_i}(q_i+q)} = \frac{\sigma_j}{\sqrt{s_j}(q_j+q)}$ is optimal. If $0 < \rho_{ij} < \max\left\{\frac{1}{2\sigma_i\sigma_j}(\sigma_j^2 - \frac{s_j\sigma_i^2}{s_i}), \frac{1}{2\sigma_i\sigma_j}(\sigma_i^2 - \frac{s_i\sigma_j^2}{s_j})\right\}$, the optimal allocation assigns a fraction of budget to collecting common observations and assigns the rest of budget to solely collecting data from system i if $\sigma_i^2/s_i > \sigma_j^2/s_j$, or solely system j if otherwise $\sigma_j^2/s_j > \sigma_i^2/s_i$. When $\rho_{ij} \geq \max\left\{\frac{1}{2\sigma_i\sigma_j}(\sigma_i^2 - \frac{s_i\sigma_j^2}{s_j}), \frac{1}{2\sigma_i\sigma_j}(\sigma_j^2 - \frac{s_j\sigma_i^2}{s_i})\right\}$, the optimal allocation assigns all the budget to collecting common observations of data. In summary, when $s = s_i + s_j$, there are three different regimes of optimal allocation depending on the sign and value of ρ_{ij}.

When $s < s_i + s_j$, two critical values that affect the optimal allocation are $s/\min\{s_i, s_j\}$ and ρ_{ij}. Note that in practice $s/\min\{s_i, s_j\}$ is always greater than 1. Theorem 2 shows that there exists a threshold value $\gamma > 1$ such that if $s/\min\{s_i, s_j\} > \gamma$, the optimal allocation strategy adopts three different regimes. When ρ_{12} is close to -1, the optimal allocation assigns all the budget to independently collecting data for system i and j. As ρ_{ij} increases, the optimal allocation assigns a fraction of budget to collecting common observations of data simultaneously and assigns the rest budget to independently collecting data from one of the two systems (in a way analogous to the case of $s = s_i + s_j$). When ρ_{ij} further increases and exceeds $\max\left\{\frac{1}{2\sigma_i\sigma_j}(\sigma_j^2 - \frac{s_j\sigma_i^2}{s-s_j}), \frac{1}{2\sigma_i\sigma_j}(\sigma_i^2 - \frac{s_i\sigma_j^2}{s-s_i})\right\}$, the optimal allocation is to collecting common observations of data simultaneously. The above summarizes the three regimes for scenarios where $s/\min\{s_i, s_j\} > \gamma$. On the other hand, if $1 < s/\min\{s_i, s_j\} < \gamma$, there are two different forms. Specifically, there exists a threshold $\rho' < 0$ such that, if $-1 < \rho_{ij} < \rho'$, the optimal allocation assigns the budget to collecting data independently for system i and j; if $\rho_{ij} > \rho'$, the optimal allocation assigns all the budget to collecting common observations of data.

5 JOINT RESOURCE ALLOCATION FOR INPUT DATA COLLECTION AND SIMULATION

In this section, we consider scenarios where the simulation cost is not negligible compared with the input data collection cost. These scenarios arise in performance evaluation for complicated systems, in which high performance computing resources are needed. The simulation may take too long time if not using designated high performance computing resources. The simulation costs therefore may be evaluated by monetary costs for purchasing computing resources or by the opportunity costs for long simulation time. In these scenarios, it is unrealistic to assume that the expected performances $\alpha_i(\cdot)$'s are immediately available at negligible cost. In this section, we extend the general framework introduced in Section 3 to include both input data collection and simulation generation.

Recall that as defined in Section 3, the expected performance of system i is given by $\alpha_i(\theta_i^*)$ for $i \in [m]$, where $\theta_i^* \in \mathbb{R}^{d_i}$ is the true input distribution parameter. Input data is collected and used to derive maximum likelihood estimators for θ_i^*'s, denoted by $\hat{\theta}_i$ for $i \in [m]$. Simulation needs to be run to estimate the expected performance $\alpha_i(\hat{\theta}_i)$'s given the estimated input distribution parameters. The simulation can be done by independently running system i and the sequence of simulation output is $(Y_{ij}(\hat{\theta}_i) : j \geq 1)$ for $i \in [m]$. Alternatively, the technique of common random numbers (CRN) can be used to evaluate the m systems simultaneously. The sequence of simulation output using CRN is $((\tilde{Y}_{ij}(\hat{\theta}) : i \in [m]) : j \geq 1)$. When using CRN, we assume that the cost of obtaining the j-th simulation replication of a set of simultaneous evaluations $(\tilde{Y}_{1j}(\hat{\theta}), \tilde{Y}_{2j}(\hat{\theta}), ...\tilde{Y}_{mj}(\hat{\theta}))$ is given by $\tilde{\eta}_j$, and the cost of generating individual evaluation $Y_{ij}(\hat{\theta}_i)$ for system i is given by η_{ij}. The simulation costs can be random. Then, given a simulation cost budget $\tilde{c}$,

we can either generate

$$\tilde{M}(\tilde{c}) = \max\{n \geq 0 : \tilde{\eta}_1 + \ldots + \tilde{\eta}_n \leq \tilde{c}\}$$

simulation replications using CRN or

$$M_i(\tilde{c}) = \max\{n \geq 0 : \eta_{i1} + \ldots + \eta_{in} \leq \tilde{c}\}$$

individual simulation replications for system i. We assume that

1. Conditional on the estimated input distribution parameters $\hat{\theta} = \{\hat{\theta}_i : i \in [m]\}$, $(\tilde{\eta}_j, (\tilde{Y}_{ij}(\hat{\theta}) : i \in [m]) : j \geq 1), (\eta_{1j}, Y_{1j}(\hat{\theta}_1) : j \geq 1), \ldots, (\eta_{mj}, Y_{mj}(\hat{\theta}_m) : j \geq 1)$ are independent sequences.
2. Conditional on $\hat{\theta}$, $(\tilde{\eta}_j, (\tilde{Y}_{ij}(\hat{\theta}) : i \in [m]))$ is independent and identically distributed (iid) in j, and $(\eta_{ij}, Y_{ij}(\hat{\theta}_i))$ is iid in j for each $i \in [m]$.
3. $\mathbb{E}[Y_{i1}|\hat{\theta}_i] = \mathbb{E}[\tilde{Y}_{i1}|\hat{\theta}] = \alpha_i(\hat{\theta}_i)$, and $\mathrm{Var}(Y_{i1}|\hat{\theta}_i) = \mathrm{Var}(\tilde{Y}_{i1}|\hat{\theta}) = D_i(\hat{\theta}_i)$ for $i \in [m]$. $\mathrm{Cov}(\tilde{Y}_{i1}, \tilde{Y}_{j1}) = D_{ij}(\hat{\theta})$ for $i, j \in [m]$, where $D_i(\cdot)$ and $D_{ij}(\cdot)$ are continuous functions with respect to θ_i and θ.
4. $\mathbb{E}\,\tilde{\eta}_1 < \infty$, $\mathbb{E}\,\eta_{i1} < \infty$, for $i \in [m]$. In general, $\mathbb{E}\,\eta_{i1} < \mathbb{E}\,\tilde{\eta}_1$ for $i \in [m]$.

Since $\tilde{M}(\cdot)$ and $M_i(\cdot)$'s are renewal counting processes, we have as $\tilde{c} \to \infty$, $\frac{1}{\tilde{c}}\tilde{M}(\tilde{c}) \overset{a.s.}{\to} \mu \triangleq \frac{1}{\mathbb{E}\,\tilde{\eta}_1}$ and $\frac{1}{\tilde{c}}M_i(\tilde{c}) \overset{a.s.}{\to} \mu_i \triangleq \frac{1}{\mathbb{E}\,\eta_{i1}}$. Given a simulation budget $\tilde{c}$, suppose we allocate a fraction r to simultaneous evaluations using CRN, and a fraction r_i to independent simulation evaluation for system i, where

$$r + r_1 + r_2 + \ldots + r_m = 1,$$

with $r \geq 0$, $r_i \geq 0$, $i \in [m]$.

With a given simulation budget and allocation, the simulation estimators for $\alpha_i(\hat{\theta}_i)$'s are given by

$$\hat{\alpha}_i(\hat{\theta}_i) = \frac{\sum_{j=1}^{\tilde{M}(r\tilde{c})} \tilde{Y}_{ij}(\hat{\theta}) + \sum_{j=1}^{M_i(r_i\tilde{c})} Y_{ij}(\hat{\theta}_i)}{\tilde{M}(r\tilde{c}) + M_i(r_i\tilde{c})}.$$

Then,

$$\tilde{c}^{1/2}(\hat{\alpha}_i(\hat{\theta}_i) - \alpha_i(\hat{\theta}_i)) = -\frac{1}{\mu r + \mu_i r_i}\left(\sqrt{\mu r}\tilde{Z}_i(\hat{\theta}) + \sqrt{\mu_i r_i}Z_i(\hat{\theta}_i)\right) + o_p(1),$$

where, conditional on $\hat{\theta}$,

- $(\tilde{Z}_1(\hat{\theta}), \ldots, \tilde{Z}_m(\hat{\theta}))$ is jointly Gaussian with mean 0. $\mathrm{Var}(\tilde{Z}_i(\hat{\theta})) = D_i(\hat{\theta}_i)$ and $\mathrm{Cov}(\tilde{Z}_i(\hat{\theta}), \tilde{Z}_j(\hat{\theta})) = D_{ij}(\hat{\theta})$ for $1 \leq i \leq j \leq m$.
- $\tilde{Z}_i(\hat{\theta}) \overset{\mathscr{D}}{=} Z_i(\hat{\theta}_i)$, and specifically $\mathrm{Var}(\tilde{Z}_i(\hat{\theta})) = \mathrm{Var}(Z_i(\hat{\theta}_i))$.
- The rv's $Z_1(\hat{\theta}_1), Z_2(\hat{\theta}_2), \ldots, Z_m(\hat{\theta}_m)$ are independent and independent of $(\tilde{Z}_1(\hat{\theta}), \ldots, \tilde{Z}_m(\hat{\theta}))$.

Hence, as $\tilde{c} \to \infty$,

$$\begin{aligned}
&\tilde{c}^{\frac{1}{2}}(\hat{\alpha}_i(\hat{\theta}_i) - \hat{\alpha}_j(\hat{\theta}_j) - (\alpha_i(\hat{\theta}_i) - \alpha_j(\hat{\theta}_j)) \\
\Rightarrow\ &-\frac{\sqrt{\mu r}}{\mu r + \mu_i r_i}\tilde{Z}_i(\hat{\theta}) + \frac{\sqrt{\mu r}}{\mu r + \mu_j r_j}\tilde{Z}_j(\hat{\theta}) - \frac{\sqrt{\mu_i r_i}}{\mu r + \mu_i r_i}Z_i(\hat{\theta}_i) + \frac{\sqrt{\mu_j r_j}}{\mu r + \mu_j r_j}Z_j(\hat{\theta}_j).
\end{aligned}$$

Denote the limiting rv as $V_{ij}(\hat{\theta})$. Conditional on $\hat{\theta}$, the rv $V_{ij}(\hat{\theta})$ is Gaussian with mean zero and variance

$$\frac{D_i(\hat{\theta}_i)}{\mu r + \mu_i r_i} + \frac{D_j(\hat{\theta}_j)}{\mu r + \mu_j r_j} - \frac{2\mu r D_{ij}(\hat{\theta})}{(\mu r + \mu_i r_i)(\mu r + \mu_j r_j)}.$$

Therefore, conditional on the input distribution specified by $\hat{\theta}$ and given a simulation budget c, the optimal simulation budget allocation problem in order to differentiate system i and j is given by

$$\min_{r_i\geq 0, r_j\geq 0, r\geq 0, r_i+r_j+r=1} \quad \frac{D_i(\hat{\theta}_i)}{\mu r+\mu_i r_i}+\frac{D_j(\hat{\theta}_j)}{\mu r+\mu_j r_j}-\frac{2\mu r D_{ij}(\hat{\theta})}{(\mu r+\mu_i r_i)(\mu r+\mu_j r_j)}.$$

Suppose that we want to compare the system performances for system i and system j and a total budget C is allocated to both input data collection and simulation experiments. Suppose we allocate a fraction p to collecting common observations, a fraction p_i (or p_j) to collecting independent observations from system i (or j), a fraction r to running simulation replications to evaluate m systems simultaneously using CRN, and a fraction r_i (or r_j) to running individual simulation replications for system i (or j) , where

$$p+p_i+p_j+r+r_i+r_j=1$$

with $p\geq 0,\ p_i\geq 0,\ r\geq 0,\ r_i\geq 0,\ i\in[m]$. The resulted system performance estimations are $\hat{\alpha}_i(\hat{\theta}_i)$ and $\hat{\alpha}_j(\hat{\theta}_j)$. The following central limit theorem is a direct result by noticing that the uncertainty presented in the input data collection and the uncertainty emerged from simulation replications are independent.

Theorem 3 When $C\to\infty$,

$$\begin{aligned}&C^{\frac{1}{2}}[(\hat{\alpha}_i(\hat{\theta}_i)-\hat{\alpha}_j(\hat{\theta}_j))-(\alpha_i(\theta_i^*)-\alpha_j(\theta_j^*))]\\ &=C^{\frac{1}{2}}[(\hat{\alpha}_i(\hat{\theta}_i)-\alpha_i(\hat{\theta}_i))-(\hat{\alpha}_j(\hat{\theta}_j)-\alpha_j(\hat{\theta}_j))]+C^{\frac{1}{2}}[(\alpha_i(\hat{\theta}_i)-\alpha_i(\theta_i^*))-(\alpha_j(\hat{\theta}_j)-\alpha_j(\theta_j^*))]\Rightarrow \quad U_{ij}(\theta^*),\end{aligned}$$

where $U_{ij}(\theta^*)$ is a Gaussian rv with mean zero and variance $\mathrm{Var}_{ij}(\theta^*)$ given by

$$\frac{\sigma_i^2}{\lambda p+\lambda_i p_i}+\frac{\sigma_j^2}{\lambda p+\lambda_j p_j}-\frac{2\lambda p c_{ij}}{(\lambda p+\lambda_i p_i)(\lambda p+\lambda_j p_j)}+\frac{D_i(\theta_i^*)}{\mu r+\mu_i r_i}+\frac{D_j(\theta_j^*)}{\mu r+\mu_j r_j}-\frac{2\mu r D_{ij}(\theta^*)}{(\mu r+\mu_i r_i)(\mu r+\mu_j r_j)}.$$

When comparing two systems and selecting the better, maximizing the asymptotic probability of correct selection is equivalent to minimizing the limiting variance $\mathrm{Var}_{ij}(\theta^*)$ as given above. The associated joint optimal budget allocation problem is given by

$$\begin{aligned}\min_{p_i,p_j,p,r_i,r_j,r} \quad & \mathrm{Var}_{ij}(\theta^*)\\ s.t. \quad & p_i+p_j+p+r_i+r_j+r=1\\ & p_i,p_j,p,r_i,r_j,r\geq 0.\end{aligned} \tag{5}$$

Due to the page limit, we conclude by noting that the joint resource allocation problem 5 in presence of correlation in both input data and simulation can be decoupled into three sub-problems, each admitting a closed-form solution.

REFERENCES

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.

Bazaraa, M. S., H. D. Sherali, and C. M. Shetty. 2013. *Nonlinear Programming: Theory and Algorithms*. Hoboken, New Jersey: John Wiley & Sons.

Chen, C.-H., S. E. Chick, L. H. Lee, and N. A. Pujowidianto. 2015. "Ranking and Selection: Efficient Simulation Budget Allocation". In *Handbook of Simulation Optimization*, edited by M. Fu, 45–80. New York: Springer-Verlag.

Cheng, R. C., and W. Holloand. 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., and B. Biller. 2013. "A Subset Selection Procedure under Input Parameter Uncertainty". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 463–473. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Dai, L., and C. Chen. 1997. "Rates of Convergence of Ordinal Comparison for Dependent Discrete Event Dynamic Systems". *Journal of Optimization Theory and Applications* 94(1):29–54.

Fu, M. C., J.-Q. Hu, C.-H. Chen, and X. Xiong. 2007. "Simulation Allocation for Determining the Best Design in the Presence of Correlated Sampling". *INFORMS Journal on Computing* 19(1):101–111.

Glasserman, P., and P. Vakili. 1994. "Comparing Markov Chains Simulated in Parallel". *Probability in the Engineering and Informational Sciences* 8(3):309–326.

Glynn, P., and S. Juneja. 2015. "Selecting the Best System and Multi-armed Bandits". *arXiv preprint arXiv:1507.04564*. https://arxiv.org/abs/1507.04564, accessed 16th February 2020.

Karnin, Z., T. Koren, and O. Somekh. 2013. "Almost Optimal Exploration in Multi-armed Bandits". In *Proceedings of the 30th International Conference on Machine Learning*, edited by S. Dasgupta and D. McAllester, 1238–1246. Stroudsburg, Pennsylvania: IMLS.

Kaufmann, E., and S. Kalyanakrishnan. 2013. "Information Complexity in Bandit Subset Selection". In *Proceedings of Machine Learning Research*, 228–251.

Kim, S.-H., and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-Zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 11(3):251–273.

Kim, S.-H., and B. L. Nelson. 2007. "Recent Advances in Ranking and Selection". In *Proceedings of the 2007 Winter Simulation Conference*, edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 162–172. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nelson, B. L., and F. J. Matejcik. 1995. "Using Common Random Numbers for Indifference-Zone Selection and Multiple Comparisons in Simulation". *Management Science* 41(12):1935–1945.

Russo, D. 2020. "Simple Bayesian Algorithms for Best Arm Identification". *Operations Research*. To Appear.

Ryzhov, I. O. 2016. "On the Convergence Rates of Expected Improvement Methods". *Operations Research* 64(6):1515–1528.

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 L. J. Hong. 2015. "Input Uncertainty and Indifference-Zone Ranking & Selection". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 414–424. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wu, D., and E. Zhou. 2017. "Ranking and Selection under Input Uncertainty: Fixed Confidence and Fixed Budget". *arXiv preprint arXiv:1708.08526*. https://arxiv.org/abs/1708.08526, accessed 14th February 2020.

Wu, D., and E. Zhou. 2018. "Analyzing and Provably Improving Fixed Budget Ranking and Selection Algorithms". *arXiv preprint arXiv:1811.12183*. https://arxiv.org/abs/1811.12183, accessed 24th February 2020.

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.

Yang, W.-N., and B. L. Nelson. 1991. "Using Common Random Numbers and Control Variates in Multiple-comparison Procedures". *Operations Research* 39(4):583–591.

AUTHOR BIOGRAPHIES

JINGXU XU is a Ph.D. student in the Department of Industrial Engineering & Operations Research at the University of California Berkeley. He has done research in simulation, stochastic modeling, and data analytics.

PETER W. GLYNN is the Thomas Ford Professor in the Department of Management Science and Engineering (MSE) at Stanford University. He is a Fellow of INFORMS and of the Institute of Mathematical Statistics, has been co-winner of Best Publication Awards from the INFORMS Simulation Society in 1993, 2008, and 2016, and was the co-winner of the John von Neumann Theory Prize from INFORMS in 2010. In 2012, he was elected to the National Academy of Engineering. His research interests lie in stochastic simulation, queueing theory, and statistical inference for stochastic processes. His email address is glynn@stanford.edu.

ZEYU ZHENG is an assistant professor in the Department of Industrial Engineering & Operations Research at the University of California Berkeley. He has done research in simulation, stochastic modeling, data analytics, statistical learning, and over-the-counter financial markets. His email address is zyzheng@berkeley.edu.