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.

INPUT PARAMETER UNCERTAINTY QUANTIFICATION WITH ROBUST SIMULATION AND RANKING AND SELECTION

Lewei Shi[1], Zhaolin Hu[1], and Ying Zhong[2]

[1]School of Economics and Management, Tongji University, Shanghai, 200092, CHINA
[2]School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, 611731, CHINA

ABSTRACT

In this paper, we consider an input parameter uncertainty quantification problem, where the true distributions of the input random factors that drive a simulation model are unknown but belong to some parametric families and we aim to find the worst-case expected performance of the simulation model. We propose a framework to calculate the upper and lower bounds of the worst-case expected performance by adopting robust simulation and ranking and selection methods, respectively. The confidence intervals of the upper and lower bounds are further constructed. The framework can address various parametric families in the input modeling and is flexible in the choice of ranking and selection procedure. Numerical experiments show the effectiveness of the framework.

1 INTRODUCTION

Input modeling plays an important role in simulation studies. It refers to selecting reasonable probability distributions for the random factors in a simulation model, such as the interarrival time distribution, the service time distribution in a queuing model, the demand distribution in a supply chain model and so on. However, due to limited information, one often cannot identify the underlying "true" distributions of these random factors, which is called the input uncertainty. For a simulation model, if the distributions of the random factors are incorrectly specified during the simulation process, the observations generated subsequently can be biased. In practice, input uncertainty often poses great challenges to the estimation of the performance measure of a simulation model. How to describe, quantify, and analyze the impact of such uncertainty on the performance measure becomes an important research topic in the simulation studies.

In the simulation literature, input uncertainty has been studied extensively. There are two types of input uncertainties that are frequently studied in the literature. The first one is called distributional uncertainty, which means that one cannot determine the distribution form of the random vector formed by random factors. The second one is called parameter uncertainty, which means that one assumes that the distribution of the random vector belongs to a parametric family but the parameters of the distribution are uncertain (Zouaoui and Wilson 2004). For stochastic simulation under input uncertainty, Corlu et al. (2020) conducted a comprehensive review. We introduce some representative works below.

While studying the input uncertainty, some works considered distributional uncertainty and parameter uncertainty together. Chick (2001) gave a randomly sampling algorithm to determine the input distributions and input parameters according to the Bayesian posterior distribution. Zouaoui and Wilson (2004) developed point and confidence-interval estimators of the posterior mean value under both kinds of uncertainty. Xie et al. (2016) presented a metamodel-assisted bootstrap method that applies to both kinds of uncertainty, and when the parametric family of the input distribution is unknown, the normal-to-anything (NORTA) was implemented. Nelson et al. (2020) put forward an input model averaging method to change the typical input modeling step, in order to reduce the simulation model risk. Lam and Qian (2022) stated that the input variance is commonly used to quantify the two kinds of input uncertainty, and bootstrap is often utilized in estimating the variance. Then, the authors presented a subsampling framework to improve the efficiency of bootstrap. Also, some studies considered the parameter uncertainty separately. Ng and Chick (2006) used asymptotic approximations to get closed-form results for sampling plans, aiming to reduce the parameter uncertainty. Biller and Corlu (2011) used a Bayesian model to capture the parameter uncertainty, and the input distribution is modeled by the NORTA method with Sklar's marginal-copula representation and Cooke's copula-vine specification. Xie et al. (2014) applied Bayesian framework to quantify the input parameter uncertainty, as well as the metamodel uncertainty. Barton et al. (2014) proposed a framework to construct confidence intervals for the mean simulation response, which accounts for input uncertainty in parametric input models. Zhu et al. (2020) studied the risk quantification of extreme scenarios. The authors introduced the nested value-at-risk (VaR) and conditional value-at-risk (CVaR) estimators of the mean response of a stochastic simulation model, and analyzed their properties under input parameter uncertainty.

Besides aforementioned studies, there is another stream of literature focusing on estimating the worst-case expected performance of a simulation model under the input uncertainty. The method is called robust simulation. The rational behind that is similar to the robust optimization by considering the performance on the worst case, so it serves as a way to quantify and evaluate the impact of input uncertainty. In this line of research, Hu et al. (2012) first considered the parameter uncertainty. The authors assumed that the random vector follows a multivariate normal distribution (MVN) where the mean vector and the covariance matrix are unknown and can take values from an ambiguity set. In this situation, finding the worst-case expected performance of a simulation model is equivalent to searching for the combination of the mean vector and the covariance matrix in the ambiguity set under which the simulation model achieves the largest or smallest expected performance. To solve the optimization problem, the authors utilized the properties of the MVN and proposed a sequential quadratic and maxdet programs (SQMP) algorithm. Later, Hu and Hong (2022) developed the robust simulation method for the distributional uncertainty. The authors used the likelihood ratio and ϕ-divergence to construct the ambiguity set, and transformed the original functional optimization problem to a convex stochastic optimization problem. Then, the authors used sample average approximation to solve the problem. Thus, the distributional uncertainty can be well handled and the simulation error can be controlled. Also, the authors discussed various approaches to construct the ambiguity set with specific structure and showed that various performance measures can be addressed in their method.

Similar to the work of Hu et al. (2012), in this paper, we consider the problem of estimating the worst-case expected performance of a simulation model under the input parameter uncertainty. Starting from the multivariate normal distribution, we assume that the random vector can follow more general distributions. Notice that in our problem, while estimating the worst-case expectation and solving the associated optimization problem, the structure of objective function is unclear and the convexity may be missing. The global optimality is difficult to guarantee due to the non-convex nature of the problem. To tackle down this issue, in this paper, we turn to consider how to bound the worst-case expectation. This also coincides with the risk-averse attitude in the robust analysis. To construct the upper bound, we build the relationship between the distributional uncertainty and parameter uncertainty. We extend the robust simulation approach and especially focus on how to calibrate the uncertainty set size when the parametric family belongs to the exponential family and the ϕ-divergence is adopted. On the other side, to obtain a lower bound of the worst-case expectation, we partition the uncertainty set and consider the worst-case expectation when the parameter vector takes values in the grid points. We then formulate a ranking and selection problem and adopt corresponding algorithm to solve the problem. The solution to the ranking and selection problem will provide a lower bound for the worst-case expectation. The upper and lower bounds consist of our input parameter uncertainty quantification framework.

The rest of this paper is organized as follows. In Section 2 we present the formulation of our problem and some related results. In Section 3, we propose our framework and discuss the analysis details in the framework. Numerical experiments are included in Section 4. Finally, we conclude the paper in Section 5.

2 PROBLEM STATEMENT

In this section, we firstly formulate the problem studied. Then, we introduce the existed robust simulation method. We will extend this method in Section 3.1.

2.1 Formulation

The problem considered in this paper is formulated as below. Let ξ be the random vector in the simulation model, whose true distribution is p_{θ^*}. We assume that p_{θ^*} comes from a parametric family $\{p_\theta\}_{\theta\in\Theta}$, where Θ is the parameter uncertainty set. Let $H(\xi)$ denote the output of the simulation model, whose closed-form expression is not available. When ξ is given, we can obtain $H(\xi)$ through running the simulation model. Without loss of generality, we assume that lower $H(\xi)$ is better. We aim to find the worst-case expectation of $H(\xi)$. Then the problem is

$$\max_{\theta\in\Theta} E_{p_\theta}[H(\xi)]. \tag{1}$$

Notice that if larger $H(\xi)$ is better, it can be reformulated as $-\max_{\theta\in\Theta} E_{p_\theta}[-H(\xi)]$, so it is sufficient to consider Problem (1). In the following, we will derive the upper and lower bounds of the optimal value of Problem (1).

2.2 Robust Simulation

We introduce the robust simulation approach in Hu and Hong (2022). Hu and Hong (2022) considered the following problem,

$$\max_{P\in\mathbb{P}} E_P[H(\xi)], \tag{2}$$

where $\mathbb{P}$ is an ambiguity set of the distribution P, and other notations are the same as that in Problem (1). They developed a comprehensive method to construct the ambiguity set $\mathbb{P}$, by using the ϕ-divergence. Under this modeling approach, a nominal distribution $p_{\hat{\theta}}$ is firstly chosen. The nominal distribution serves as a reference distribution around which we construct the ambiguity set. After that, choosing the simplified functional space with only one ϕ-divergence constraint, we have that $\mathbb{P} = \{p : p\in\mathbb{D}, D_\phi(p, p_{\hat{\theta}}) \le r\}$, where $\mathbb{D}$ denotes a general probability distribution and $D_\phi(p, p_{\hat{\theta}})$ is the ϕ-divergence between the distribution p and $p_{\hat{\theta}}$. Furthermore, the variable can be changed into the likelihood ratio and Problem (2) with the above ambiguity set $\mathbb{P}$ can be rewritten as

$$\max_{L\in\mathbb{L}} E_{p_{\hat{\theta}}}[H(\xi)L], \tag{3}$$

where $L = \frac{p_\theta}{p_{\hat{\theta}}}$ is the likelihood ratio. And $\mathbb{L} = \left\{L\in\mathbb{L}(0,+\infty) : E_{p_{\hat{\theta}}}[L] = 1, E_{p_{\hat{\theta}}}[\phi(L)] \le r\right\}$ is the ambiguity set for L, where $\mathbb{L}(0,+\infty) := \{0 \le L \le +\infty \; a.s.\}$.

The parameter r represents the size of the ambiguity set, which will be our main focus in the next section. Note that ϕ is a convex function in the ϕ-divergence. Problem (3) is a convex functional optimization problem. Furthermore, it can be transformed to a convex stochastic optimization problem, through strong duality and some optimization techniques. Many ϕ-divergences can be used in this method. We refer readers to Hu and Hong (2022) for details. Particularly, we present the results related to the Kullback-Leibler (KL) divergence, a representative ϕ-divergence, for completeness.

Corollary 1 (Hu and Hong 2022) For the KL divergence, the optimal value of Problem (3) is the optimal value of the following problem

$$\min_{\lambda\in\mathbb{R},\alpha\ge 0} E_{P_{\hat{\theta}}}\left[\alpha\exp\left\{\frac{H(\xi)+\lambda}{\alpha} - 1\right\}\right] + \alpha r - \lambda, \tag{4}$$

and the corresponding sample average approximation (SAA) problem is

$$\begin{aligned} \min \quad & \frac{1}{N}\sum_{j=1}^{N} \exp\{-1\} z_j + \alpha r - \lambda, \\ \text{s.t.} \quad & \alpha \exp\left\{\frac{H(\xi_j)+\lambda}{\alpha}\right\} \le z_j, \\ & z_j \in \mathbb{R}, j=1,\ldots,N, \quad \lambda \in \mathbb{R}, \alpha \ge 0. \end{aligned} \tag{5}$$

Problem (5) is a convex problem called the perspective-of-exponential program and can be solved efficiently using the cvx in MATLAB. Moreover, under mild conditions, the simulation error of the SAA can be controlled by the sample size. One can construct the following approximate confidence interval for the true optimal value of Problem (4),

$$\left[v_N - N^{-\frac{1}{2}} z_{1-\frac{\beta}{2}} \hat{\sigma}_n(x_N^*), v_N + N^{-\frac{1}{2}} z_{1-\frac{\beta}{2}} \hat{\sigma}_n(x_N^*)\right], \tag{6}$$

where N is the sample size, x_N is the optimal solution of sample problem, v_N is the optimal value of sample problem, and $z_{1-\frac{\beta}{2}}$ is the $1-\frac{\beta}{2}$ quantile of standard normal distribution, $\hat{\sigma}_n^2(x_N^*)$ is the variance estimator calculated by $\hat{\sigma}_n^2(x_N^*) = \frac{1}{n-1}\sum_{i=1}^n \left(F(x_N^*,\xi_i) - \frac{1}{n}\sum_{i=1}^n F(x_N^*,\xi_i)\right)^2$, with i.i.d. samples $\{\xi_1,...,\xi_n\}$, which are independent from the sample used in SAA. In this case, $F(x,\xi) = F(\alpha,\lambda,\xi) = \alpha \exp\left\{\frac{H(\xi)+\lambda}{\alpha} - 1\right\} + \alpha r - \lambda$.

3 INPUT PARAMETER UNCERTAINTY QUANTIFICATION FRAMEWORK

We now consider Problem (1) and assume that the uncertainty set Θ is provided. In practice, Θ can be determined based on different information and different methods. For example, decision makers may construct Θ based on the objective or subjective information. In the objective uncertainty scenario where data are available, one can build some confidence sets for θ. In the subjective uncertainty scenario, experts may quantify the uncertainty and together determine Θ. When some input modeling methods are used, the uncertainty set for the parameters may also be constructed. For example, when implementing the NORTA method, the correlation parameters of the MVN may have uncertainty and Θ may be considered for the parameters.

3.1 Upper Bound Analysis

We use the robust simulation approach of Hu and Hong (2022) to build the upper bound. Below we provide a slightly more general result. Suppose that $\hat{\theta}$ is a parameter vector appropriately chosen and $p_{\hat{\theta}}$ is the corresponding nominal probability distribution. Note that $\hat{\theta}$ does not have to belong to Θ. Let

$$r = \max_{\theta\in\Theta} D_\phi\left(p_\theta, p_{\hat{\theta}}\right). \tag{7}$$

Suppose that r is finite valued. Then we have the following proposition.

Proposition 1 Suppose that r in Problem (3) is defined by (7) and is finite valued. The optimal value of Problem (3) is an upper bound of the optimal value of Problem (1).

The result in Proposition 1 follows from the fact that the ambiguity set $\mathbb{P} = \{p : p \in \mathbb{D}, D_\phi(p, p_{\hat{\theta}}) \le r\}$ contains $\mathbb{P}_\theta = \left\{p_\theta : \theta \in \Theta, D_\phi(p_\theta, p_{\hat{\theta}}) \le r\right\}$ as a subset. Proposition 1 suggests that by solving Problem (3), we can obtain an upper bound for the worst-case expectation. The bound depends on the selection of $\hat{\theta}$. A natural idea is to choose a center point in the uncertainty set Θ because such selection may achive a small radius r defined by (7). When the computational power is sufficient, we can also select multiple $\hat{\theta}$ and construct multiple approximation problems. By solving these problems, we can obtain multiple upper bounds and select the least conservative one. As introduced in Section 2, Problem (1) can be solved based on the stochastic optimization theory and the statistical error can also be quantified. To implement this bound, we need to obtain r.

Below we discuss how to derive r or derive a conservative bound for r for a number of settings. We analyze the structure of Problem (7) for different combinations of ϕ-divergence and parametric family. Problem (7) can be expressed as the following stochastic programming problem,

$$r = \max_{\theta \in \Theta} E_{p_{\hat{\theta}}} \left[\phi \left(\frac{p_\theta}{p_{\hat{\theta}}} \right) \right]. \tag{8}$$

Problem (8) becomes a deterministic optimization problem when there exists analytic expression for the objective function. There are alternative ϕ-divergences to be chosen, corresponding to different ϕ-functions. Table 1 summarizes some common ϕ-divergences (Hu and Hong 2022). Because that we are going to consider the exponential family, the natural logarithm is taken hereafter.

Table 1: Some ϕ-divergences and respective ϕ-functions.

Divergence	$\phi(t), t \geq 0$
Kullback-Leibler	$t \log t$
Burg entropy	$-\log t + t - 1$
J-divergence	$(t-1)\log t$
Neyman χ^2-distance	$\frac{1}{t}(t-1)^2$
χ^2-distance	$(t-1)^2$
Hellinger distance	$(\sqrt{t}-1)^2$
Cressie–Read	$\frac{1-\theta+\theta t - t^\theta}{\theta(1-\theta)},\ \theta \neq 0, 1$

The exponential family is widely used in the statistical modeling and has good properties for approximation to other distributions (Nielsen and Nock 2013). It includes the distribution whose probability density (mass) function can be written as $p(z|\eta) = e^{t(z)^T \eta - A(\eta)} h(z)$, where η is the natural parameter, $t(z)$ is the sufficient statistic, the superscript T denotes the transpose operator, $A(\eta)$ is the cumulant function, and $h(z)$ is the normalizer. For example, for the normal distribution $N(\mu, \sigma^2)$, $\eta = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^T$, $t(z) = (z, z^2)^T$, $A(\eta) = \log\sigma + \frac{\mu^2}{2\sigma^2}$ and $h(z) = \frac{1}{\sqrt{2\pi}}$.

It is worthwhile noting that a distribution may be parameterized by different parameters. Taking the normal distribution $N(\mu, \sigma^2)$ as an example, it can be parameterized by the mean μ and the standard deviation σ. Alternatively, we can also use the natural parameter mentioned above to parameterize the same distribution, where the natural parameter is $\eta = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^T$. The parameter space in the (μ, σ) parameterization is $(-\infty, \infty) \otimes (0, \infty)$, and the parameter space in the $\eta = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^T$ parameterization is $(-\infty, \infty) \otimes (-\infty, 0)$. The former parameterization is more intuitive, while the latter is more convenient to analyze in a unified manner in our framework. Moreover, the uncertainty set of parameters can be transformed between the different parameterization methods. We combine these two parameterization methods to illustrate our findings.

For the parametric family in the exponential family, the expression of KL divergence and χ^2-distance were studied (Nielsen and Nock 2013). In this paper, we also derive closed-form formulas for other ϕ-divergences in Table 1. The results are listed in Table 2. The two distributions p_η and $p_{\hat{\eta}}$ come from the same parametric family with different natural parameters η and $\hat{\eta}$. The distribution $p_{\hat{\eta}}$ is the nominal distribution so that $\hat{\eta}$ is prespecified. The mathematical details are omitted due to the space limit.

With the closed-form formulas, we may often simplify Problem (8). Especially, for the one parameter setting, we have the following result.

Table 2: Closed-form formulas for several ϕ-divergences of the parametric family in exponential family.

Divergence	Closed-form formula for $D_\phi\left(p_\eta, p_{\hat\eta}\right)$
Kullback-Leibler	$A(\hat\eta) - A(\eta) - (\hat\eta - \eta)^T \nabla A(\eta)$
Burg entropy	$A(\eta) - A(\hat\eta) + (\hat\eta - \eta)^T \nabla A(\hat\eta)$
J-divergence	$(\nabla A(\eta) - \nabla A(\hat\eta))(\eta - \hat\eta)^T$
Neyman χ^2-distance	$\exp\{A(2\hat\eta - \eta) - 2A(\hat\eta) + A(\eta)\} - 1$
χ^2-distance	$\exp\{A(2\eta - \hat\eta) - 2A(\eta) + A(\hat\eta)\} - 1$
Hellinger distance	$2 - 2\exp\left\{A(\frac{\eta+\hat\eta}{2}) - \frac{1}{2}A(\eta) - \frac{1}{2}A(\hat\eta)\right\}$
Cressie–Read	$\frac{1}{\theta(1-\theta)}\left(1 - \exp\{A(\theta\eta - (\theta-1)\hat\eta) - \theta A(\eta) + (\theta-1)A(\hat\eta)\}\right),\ \theta \neq 0,1$

Theorem 1 Suppose that the parametric family is the one parameter exponential family, i.e., η is a scalar, and $\eta \in \Theta := [\eta_l, \eta_u]$. Problem (8) achieves the optimal value at η_l or η_u.

Proof. For the exponential family, it is known that $A(\eta)$ is a convex function. Let $f(\eta)$ denote the obtained formula for a D_ϕ in Table 2. Then for the one parameter exponential family, all the $f(\eta)$ in Table 2 follow the same structure, i.e., $f'(\eta) \leq 0$ when $\eta < \hat\eta$, $f'(\eta) = 0$ when $\eta = \hat\eta$, and $f'(\eta) \geq 0$ when $\eta > \hat\eta$. When η is a scalar, it implies that the optimal value of Problem (8) can be found by taking the larger objective value at η_l and η_u. □

The one parameter exponential family contains lots of common distributions, such as the Bernoulli distribution, Poisson distribution, exponential distribution and so on. Moreover, when the parametric families have more than one parameter but we treat one of them as uncertain, this parametric family also belongs to the one parameter exponential family. Otherwise, the analysis needs to be conducted specifically. We give some examples for solving Problem (8) using some parametric families from the exponential family and the KL divergence.

3.1.1 Univariate Case

We first consider the univariate case, which is relatively simple and forms the basis of the multivariate case. The expressions of the KL divergence for different parametric families from the exponential family are summarized in Table 3.

In the Gamma and Beta parametric families, $\Gamma(\cdot)$ is the gamma function, $\psi(\cdot)$ is the digamma function defined by $\psi(\alpha) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\log\Gamma(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$, and $B(\alpha,\beta)$ is the Beta function defined by $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. We consider that $\theta \in \Theta := [\theta_l, \theta_u]$. For many parametric families, the optimal solutions of Problem (8) are straightforward by using the calculus techniques. We introduce them briefly. For the Bernoulli, binomial, Poisson and exponential parametric families, the maximum value is attained at the endpoint with larger objective value. (In the binomial parametric family, we treat parameter n as known.) For the normal parametric family, we can treat parameters μ and σ separately, then its maximum value also lies at the endpoint with larger objective value, the same as the log-normal and inverse Gaussian parametric family.

For the Gamma and Beta parametric families, the analyses are more complicated, and we have the following findings. For the Gamma parametric family, the maximum value of Problem (8) can be obtained at the endpoints, when either one of the parameters α and β has uncertainty, or both of them have uncertainty. For the Beta parametric family, the maximum value of Problem (8) can be obtained at the endpoints, when either parameter α or β has uncertainty.

It is worthwhile to mention that the normal, log-normal, inverse Gaussian, and Gamma parametric families come from the two parameter exponential family. When both two parameters in the above parametric family are uncertain, the optimal value of Problem (8) can also be found analytically. It indicates that the application scenarios of our method may include more parametric families.

Table 3: Univariate parametric family from the exponential family and their related KL divergences.

Parametric family	$D_{\mathrm{KL}}(p_\theta, p_{\hat{\theta}})$
Bernoulli	$p\log\frac{p}{\hat{p}}+(1-p)\log\frac{1-p}{1-\hat{p}}$
Binomial	$np\log\frac{p}{\hat{p}}+(n-np)\log\frac{1-p}{1-\hat{p}}$
Poisson	$\lambda\log\frac{\lambda}{\hat{\lambda}}+\hat{\lambda}-\lambda$
Normal	$\frac{1}{2}\left[\frac{(\hat{\mu}-\mu)^2}{\hat{\sigma}^2}+\frac{\sigma^2}{\hat{\sigma}^2}-\log\frac{\sigma^2}{\hat{\sigma}^2}-1\right]$
Gamma	$\hat{\alpha}\log\frac{\beta}{\hat{\beta}}-\log\frac{\Gamma(\alpha)}{\Gamma(\hat{\alpha})}+(\alpha-\hat{\alpha})\psi(\alpha)-(\beta-\hat{\beta})\frac{\alpha}{\beta}$
Exponential	$\log\frac{\lambda}{\hat{\lambda}}+\frac{\hat{\lambda}}{\lambda}-1$
Log-normal	$\frac{1}{2}\left[\frac{(\hat{\mu}-\mu)^2}{\hat{\sigma}^2}+\frac{\sigma^2}{\hat{\sigma}^2}-\log\frac{\sigma^2}{\hat{\sigma}^2}-1\right]$
Beta	$\log\frac{B(\hat{\alpha},\hat{\beta})}{B(\alpha,\beta)}-(\hat{\alpha}-\alpha)\psi(\alpha)-(\hat{\beta}-\beta)\psi(\beta)+(\hat{\alpha}-\alpha+\hat{\beta}-\beta)\psi(\alpha+\beta)$
Inverse Gaussian	$\frac{1}{2}\log\frac{\lambda}{\hat{\lambda}}+\frac{\hat{\lambda}}{2\hat{\mu}^2}\mu-\frac{\hat{\lambda}}{\hat{\mu}}+\frac{\hat{\lambda}}{2}\left(\frac{1}{\mu}+\frac{1}{\lambda}\right)-\frac{1}{2}$

3.1.2 Multivariate Case

Now we consider the multivariate case. For the multivariate case, we first mention that the KL divergence has some nice structure when the random variables are independent, i.e., $D_{\mathrm{KL}}(q,p)=\sum_{i=1}^{d}D_{\mathrm{KL}}(q_i,p_i)$ when $q(x)=q_1(x_1)\cdots q_d(x_d)$, $p(x)=p_1(x_1)\cdots p_d(x_d)$ and d is the dimension of the random vector. Under this setting, we can use the bounds for the marginal distributions to construct an overall bound for the joint distribution (Hu and Hong 2022). When the random variables are not independent, the problem becomes challenging. Here, we explore the multivariate normal parametric family. We highlight that the multivariate normal distribution (MVN) plays an important role in the simulation study, because it is widely used in practice and is a major component in some input modeling approach, e.g., the NORTA method (Biller and Corlu 2011; Xie et al. 2016). For the MVN, Problem (8) becomes

$$\max_{\mu\in\Theta_\mu,\Sigma\in\Theta_\Sigma}\frac{1}{2}\left[(\hat{\mu}-\mu)^T\hat{\Sigma}^{-1}(\hat{\mu}-\mu)+\mathrm{tr}(\hat{\Sigma}^{-1}\Sigma)-\log\frac{|\Sigma|}{|\hat{\Sigma}|}-d\right], \tag{9}$$

where Θ_μ and Θ_Σ are the uncertainty sets of the mean vector and covariance matrix, and $\hat{\mu}$ and $\hat{\Sigma}$ are the mean vector and covariance matrix of the nominal distribution. The matrices $\hat{\Sigma}$ and Σ are positive definite. This optimization problem is generally non-convex, but it is possible to get an upper bound of the optimal value of Problem (9) by optimizing the two variables μ and Σ separately.

For $\mu\in\Theta_\mu$, we can solve the following problem to find the optimal solution of μ in Problem (9),

$$\max_{\mu\in\Theta_\mu}(\hat{\mu}-\mu)^T\hat{\Sigma}^{-1}(\hat{\mu}-\mu). \tag{10}$$

Because $\hat{\Sigma}^{-1}$ is also a positive definite matrix, Problem (10) involves maximizing a convex function. For a box set, we can obtain the following result.

Proposition 2 Suppose that $\Theta_\mu=\{\mu|a_i\le\mu_i\le b_i, i=1,\ldots,d,\mu\in\mathbb{R}^d\}$. The optimal value of Problem (10) can be attained at some vertex of Θ_μ that has the maximal function values among all vertices.

Proof. Note that any $\mu\in\Theta_\mu$ can be expressed as the convex combination of the set of vertices $\{v_1,\ldots,v_m\}$ $\left(m\le 2^d\right)$ of Θ_μ, i.e., $\mu=\sum_{j=1}^{m}\theta_j v_j, \sum_{j=1}^{m}\theta_j=1, \theta_j\ge 0, j=1,\cdots,m$. Let $f(\cdot)$ denote the objective function of Problem (10). Due to the convexity, we have $f(\mu) \leq \theta_1 f(v_1) + \cdots + \theta_m f(v_m) \leq \max_j \{f(v_j)\}, j = 1, \ldots, m$. It indicates that the optimal value of Problem (10) does not exceed the maximum of the function values at the vertices, and thus the optimum can be attained at some vertex. $\square$

For $\Sigma \in \Theta_\Sigma$, we can similarly solve the following problem to find the optimal solution of Σ in Problem (9),

$$\max_{\Sigma \in \Theta_\Sigma} \operatorname{tr}(\hat{\Sigma}^{-1}\Sigma) - \log|\Sigma|. \tag{11}$$

We consider a setting where Σ is constrained by $\Sigma_l \preceq \Sigma \preceq \Sigma_u$, where Σ_l and Σ_u are both given positive definite matrices. The objective function of Problem (11) is a convex function of Σ. For this setting, we are able to derive an upper bound of the optimal value of Problem (11). We have the following proposition.

Proposition 3 Suppose that $\Theta_\Sigma = \{\Sigma | \Sigma_l \preceq \Sigma \preceq \Sigma_u, \Sigma \succ 0, \Sigma \in \mathbb{R}^{d\times d}\}$. Then $\operatorname{tr}\left(\hat{\Sigma}^{-1}\Sigma_u\right) - \log|\Sigma_l|$ is an upper bound of the optimal value of Problem (11).

Proof. From Section 3.6 of Boyd and Vandenberghe (2004), it holds that, a differentiable function f with convex domain is K-increasing if $\nabla f(x) \succ_{K^*} 0$ for all $x \in \operatorname{dom} f$, where K^* is the dual cone of K. Note that the semidefinite cone is self-dual. By introducing a new variable $\bar{\Sigma} = \Sigma - \Sigma_l$, we have that $0 \preceq \bar{\Sigma} \preceq \Sigma_u - \Sigma_l$. The gradients of $\operatorname{tr}(\hat{\Sigma}^{-1}\bar{\Sigma} + \hat{\Sigma}^{-1}\Sigma_l) = \operatorname{tr}(\hat{\Sigma}^{-1}\bar{\Sigma}) + \operatorname{tr}(\hat{\Sigma}^{-1}\Sigma_l)$ and $-\log|\bar{\Sigma} + \Sigma_l|$ with respect to $\bar{\Sigma}$ are $\hat{\Sigma}^{-1}$ and $-\Sigma^{-1}$ respectively. Note that $\hat{\Sigma}^{-1} \succ 0$ and $\Sigma^{-1} \succ 0$. Then we have $\operatorname{tr}(\hat{\Sigma}^{-1}(\bar{\Sigma} + \Sigma_l))$ is matrix increasing and $-\log|\bar{\Sigma} + \Sigma_l|$ is matrix decreasing. Therefore, an upper bound can be obtained by taking the largest value of each part. $\square$

Combining Propositions 2 and 3, we are able to derive an upper bound of the optimal value of Problem (9) with the given uncertainty set. Then, the bound can be used to construct the ambiguity set in Problem (3).

3.2 Lower Bound Analysis

We aim to find a lower bound of the optimal value of Problem (1) in this section. Recall that Problem (1) is a maximization problem. So the objective function value of any feasible solution is a lower bound of the optimal value of Problem (1), i.e., the worst-case expectation. Our method is to partition the parameter space Θ, and obtain a set of parameter vectors $\{\theta_1, \ldots, \theta_n\}$. Then we aim to find the parameter vector θ_* with the largest objective value $E_{p_{\theta_*}}[H(\xi)]$. This problem can be viewed as a discrete optimization via simulation (DOvS) problem or a ranking and selection (R&S) problem. The main difference is whether to consider the relationship among solutions. DOvS algorithms usually have the convergence guarantee, but they cannot provide the predetermined probability guarantee of the final solution. The latter guarantee conforms to our framework, and it is possible for R&S procedure to achieve such guarantee.

Therefore, we treat the problem as a ranking and selection problem. There exist various procedures in R&S problem, and each procedure has its advantages and disadvantages. Considering the frontier of R&S procedure, the scalability and the statistical guarantee of the chosen alternative, we find the Knockout-Tournament (KT) procedure. This procedure takes use of the rule of Knockout-Tournament in tennis, which can avoid comparison between every two alternatives and then enhance efficiency in solving large-scale R&S problem. The parallel version of KT procedure is KT^+. Based on the rule of KT, KT^+ divides the task into different processors to sufficiently utilize the computational resource. In our circumstance, the alternative in R&S is the combination of parameters, and it depends on the partition rule. When the dimension of parameters is high, the number of alternatives easily becomes large if the partition becomes dense. Considering this, we choose the procedure KT^+ with probability of correct selection (PCS) guarantee in this work. For details on this procedure, we refer readers to Zhong and Hong (2022). After the procedure selects the best alternative θ_*, we resample θ_* with another group of samples to form a confidence interval of the lower bound.

Under the normality assumption of the objective value, a confidence interval $[\mu_l, \mu_u]$ for that at the selected parameter θ_* can be constructed using additional samples $\{H_1, \dots, H_m\}$, where m is the resampling size. The following relationship holds,

$$\frac{\bar{H} - \mu}{S/\sqrt{m}} \sim t(m-1),$$

where $\bar{H} = \frac{1}{m}\sum_{i=1}^{m} H_i$ is the sample mean, $S = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(H_i - \bar{H})^2}$ is the sample standard deviation, and $t(m-1)$ is the t distribution with $m-1$ degrees of freedom. Then we can construct a $1-\gamma$ confidence interval for μ as follows, where $t_{1-\frac{\gamma}{2}}(m-1)$ is the $1-\frac{\gamma}{2}$ quantile of the distribution $t(m-1)$,

$$\left[\bar{H} - \frac{S}{\sqrt{m}} t_{1-\frac{\gamma}{2}}(m-1), \bar{H} + \frac{S}{\sqrt{m}} t_{1-\frac{\gamma}{2}}(m-1)\right], \tag{12}$$

which can be used to quantify the simulation error.

3.3 General Framework

We summarize the general framework in the following algorithm. It is worthwhile to note that the choice of the ranking and selection procedure is flexible and can vary according to the properties of the problem.

Algorithm 1 (Input Parameter Uncertainty Quantification Framework)

Input: The black-box simulation model; the ranking and selection procedure; the selected performance measure; the uncertainty set (with confidence level α); the confidence level of the construction of upper bound β; the confidence level of the construction of lower bound γ; the resampling size of the best alternative after the ranking and selection n_1; the sample size of the nominal distribution to conduct robust simulation n_2, and additional sample size to construct the confidence interval n_3.

Step 1: Generate alternatives in the parameter uncertainty set according to some partition rules.

Step 2: Conduct ranking and selection for the alternatives and identify the best alternative.

Step 3: Generate additional samples of the best alternative to construct the confidence interval of lower bound using equation (12).

Step 4: Generate samples of the nominal distribution to conduct robust simulation, i.e., constructing and solving the related optimization problem (such as Problem (5) for the KL divergence), and use additional samples to construct the confidence interval using equation (6).

Step 5: Obtain the upper and lower bounds, each with a confidence interval to quantify the simulation error.

Output: The upper and lower bounds of the worst-case value of the selected performance measure, each with a confidence interval.

4 NUMERICAL EXPERIMENT

4.1 Test Problem

We consider an instance of Problem (1). To validate our approach, we consider a case where H is known and we can solve Problem (1) analytically. Consider the following setting, where x is the decision variable, ξ is the random vector, $\xi \sim MVN(\mu, \Sigma)$, $\xi^T x \sim N(\mu^T x, x^T \Sigma x)$ and $H(\xi) = e^{-\xi^T x}$. Then we know that

$$E[e^{-\xi^T x}] = e^{-\mu^T x + \frac{1}{2} x^T \Sigma x}.$$

We consider the uncertainty set $\Theta_\mu = \left\{\mu | (1-\rho)\hat{\mu} \le \mu \le (1+\rho)\hat{\mu}, \mu \in \mathbb{R}^d\right\}$ for the mean vector and $\Theta_\Sigma = \left\{\Sigma | (1-\varepsilon)\hat{\Sigma} \preceq \Sigma \preceq (1+\varepsilon)\hat{\Sigma}, \Sigma \succ 0, \Sigma \in \mathbb{R}^{d\times d}\right\}$ for the covariance matrix. The parameters $0 < \rho < 1$

and $0 < \varepsilon < 1$ reflect the size of two uncertainty sets. For convenience, we partition the parameter space through partitioning ρ and ε equidistantly in the analysis of lower bound. Specifically, we set that $\hat{\mu}$ is a vector with all elements to be 0.2, $\hat{\Sigma} = \mathrm{diag}(0.05^2)$, and x is a vector with all elements being -2. It is straightforward that the true worst-case value can be calculated exactly, i.e., $e^{-(1+\rho)\hat{\mu}^T x + \frac{1}{2}x^T(1+\varepsilon)\hat{\Sigma}x}$. On the other hand, we use our framework to conduct experiments, where r comes from Problem (9) and Propositions 2 and 3.

We mainly focus on the gap between the upper and lower bounds, as well as the position of the true worst-case value and the nominal average value, with some different settings. The main objective is to illustrate that the framework is useful to quantify the input parameter uncertainty, and the simulation error can be controlled with the increase of sample size. We note that the worst-case alternative is always involved in the alternatives in ranking and selection because of simplification of this case, however, it is not the usual situation, so we eliminate that alternative manually.

4.2 Results

We present the experiment result below. In the experiments, the confidence levels for both the upper and lower bounds are set to be 0.95. In the KT^+ procedure, the processor used is set to be 3, the alternative number in one group is set to be 3, the number of first-stage samples is set to be 5,000, and the probability of correct selection (PCS) is set to be 0.995. For simplicity, the parameters n_1, n_2 and n_3 are set to be equal, i.e., $n_1 = n_2 = n_3 = n$, and we call n the sample size uniformly. The nominal average value is $\frac{1}{n}\sum_{i=1}^{n} H(\xi_i)$, where $\xi_i \overset{i.i.d.}{\sim} MVN(\hat{\mu}, \hat{\Sigma})$. We vary the sample size, the dimension of random vector ξ, the partition number of parameter space, and the size of uncertainty set, to test the validity of our framework.

Firstly, we consider different settings of uncertainty set, i.e., different combinations of parameters ρ and ε. Specifically, Table 4 presents the results. The term CI is the abbreviation of confidence interval hereafter.

Table 4: Results of different combinations of parameters ρ and ε. (dimension = 5; indifference zone parameter = 0.01; sample size = 5000)

Combination of ρ & ε	True worst-case value	Nominal average value	95% CI of lower bound	95% CI of upper bound
0 & 5%	7.58559	7.54581	[7.54671, 7.64506]	[8.81399, 8.93965]
5% & 0	8.37290	7.59175	[8.28819, 8.39295]	[8.35634, 8.47960]
5% & 5%	8.38337	7.55554	[8.31408, 8.42054]	[9.03340, 9.22893]
10% & 5%	9.26506	7.57807	[9.18970, 9.30911]	[9.55579, 10.02675]
5% & 10%	8.39386	7.56571	[8.34758, 8.45718]	[9.61177, 9.81705]
10% & 10%	9.27665	7.56460	[9.22583, 9.34639]	[10.10984, 10.43734]

From the above table, we can conclude that different structures of uncertainty set can be well handled, which shows the scalability of our framework.

Secondly, we consider different dimensions of the random vector ξ. The results are summarized in Table 5.

It can be found that the dimension of the random vector has high impact on this framework, especially on the upper bound. When the dimension is higher, the results become more conservative. It implies that higher dimension brings larger uncertainty.

Thirdly, we consider different partition numbers of the parameter space. When the partition number is five, it means that both the mean vector and the covariance matrix are selected from five different options. Then the number of alternative is $5 * 5 - 1 = 24$. The results are summarized in Table 6.

It is shown that the denser partition can achieve the improvement of lower bound. There exists a trade-off between the partition and the simulation efforts. The design of partition should be determined

Table 5: Results of different dimensions of the random vector. ($\rho = 0.1$, $\varepsilon = 0.1$; partition number = 5; sample size = 5000; indifference zone parameter is adjusted as needed, appearing in the bracket behind the dimension.)

Dimension	True worst-case value	Nominal average value	95% CI of lower bound	95% CI of upper bound
2 (0.001)	2.43757	2.24211	[2.42745, 2.44735]	[2.51785, 2.54463]
5 (0.01)	9.27665	7.54788	[9.20842, 9.32883]	[9.68858, 10.64518]
10 (0.1)	86.05615	57.44255	[84.90523, 86.45918]	[102.18208, 108.87875]

Table 6: Results of different partition numbers of the parameter space. ($\rho = 0.1$, $\varepsilon = 0.1$; dimension = 5; indifference zone parameter = 0.01; sample size = 5000)

Partition number	True worst-case value	Nominal average value	95% CI of lower bound	95% CI of upper bound
5	9.27665	7.57785	[9.15969, 9.27839]	[9.82530, 10.60717]
8	9.27665	7.58580	[9.18132, 9.30378]	[10.16762, 10.35309]
10	9.27665	7.62052	[9.22340, 9.34455]	[10.20638, 10.48620]

by balancing the accuracy and the simulation efforts. Moreover, the last column indicates that the result of solving robust simulation problem (Problem (5)) has some deviation under the same sample size.

Fourthly, we consider different sample sizes. Due to the deviation mentioned above, the robust simulation problem is solved ten times independently, under each sample size. And the average results corresponding with nominal average value and confidence interval of upper bound are reported. The results are summarized in Table 7.

Table 7: Results of different sample sizes. ($\rho = 0.1$, $\varepsilon = 0.1$; partition number = 5; dimension = 5; indifference zone parameter = 0.01)

Sample size	True worst-case value	Nominal average value	95% CI of lower bound	95% CI of upper bound
50	9.27665	7.59612	[8.71712, 10.14399]	[9.38429, 10.85430]
500	9.27665	7.58129	[9.14001, 9.52986]	[10.00363, 10.58508]
5000	9.27665	7.56601	[9.19203, 9.31115]	[10.11432, 10.43785]

From Table 7, we can conclude that both the confidence intervals become narrower with the increase of the sample size. It indicates that the simulation error can be controlled and the required accuracy can be achieved as long as the sampling cost is affordable.

Finally, from all the above tables, we can find that the nominal average value is lower than the true worst-case value, which means that the expected performance under nominal distribution may be far from the worst-case expected performance, and our framework contributes for evaluating the latter one.

5 CONCLUSION

In this paper, we propose a framework to quantify the impact of the input parameter uncertainty. By establishing the relationship between the distributional uncertainty and the parameter uncertainty, we use the robust simulation method to get an upper bound of the worst-case expected performance. Moreover, we use the ranking and selection method to get a lower bound of the worst-case expected performance. We show that the framework can address various scenarios. The numerical experiments show that our framework is valid for quantifying the input parameter uncertainty.

ACKNOWLEDGMENTS

This work was partially supported by the National Natural Science Foundation of China [No. 72071148].

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.

Biller, B. and C. G. Corlu. 2011. "Accounting for Parameter Uncertainty in Large-Scale Stochastic Simulations with Correlated Inputs". *Operations Research* 59(3):661–673.

Boyd, S. and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge: Cambridge University Press.

Chick, S. E. 2001. "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty". *Operations Research* 49(5):744–758.

Corlu, C. G., A. Akcay, and W. Xie. 2020. "Stochastic Simulation under Input Uncertainty: A Review". *Operations Research Perspectives* 7.

Hu, Z., J. Cao, and L. J. Hong. 2012. "Robust Simulation of Global Warming Policies Using the DICE Model". *Management Science* 58(12):2190–2206.

Hu, Z. and L. J. Hong. 2022. "Robust Simulation with Likelihood-Ratio Constrained Input Uncertainty". *INFORMS Journal on Computing* 34(4):2350–2367.

Lam, H. and H. Qian. 2022. "Subsampling to Enhance Efficiency in Input Uncertainty Quantification". *Operations Research* 70(3):1891–1913.

Nelson, B. L., A. T. K. Wan, G. Zou, X. Zhang and X. Jiang. 2020. "Reducing Simulation Input-Model Risk via Input Model Averaging". *INFORMS Journal on Computing* 33(2):672–684.

Ng, S. H. and S. E. Chick. 2006. "Reducing Parameter Uncertainty for Stochastic Systems". *ACM Transactions on Modeling and Computer Simulation* 16(1):26–51.

Nielsen, F. and R. Nock. 2013. "On the Chi Square and Higher-Order Chi Distances for Approximating f-Divergences". *IEEE Signal Processing Letters* 21(1):10–13.

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.

Xie, W., B. L. Nelson, and R. R. Barton. 2016. "Multivariate Input Uncertainty in Output Analysis for Stochastic Simulation". *ACM Transactions on Modeling and Computer Simulation* 27(1):1–22.

Zhong, Y. and L. J. Hong. 2022. "Knockout-Tournament Procedures for Large-Scale Ranking and Selection in Parallel Computing Environments". *Operations Research* 70(1):432–453.

Zhu, H., T. Liu, and E. Zhou. 2020. "Risk Quantification in Stochastic Simulation under Input Uncertainty". *ACM Transactions on Modeling and Computer Simulation* 30(1):1–24.

Zouaoui, F. and J. R. Wilson. 2004. "Accounting for Input-Model and Input-Parameter Uncertainties in Simulation". *IIE Transactions* 36(11):1135–1151.

AUTHOR BIOGRAPHIES

LEWEI SHI is a Ph.D. student in the School of Economics and Management at Tongji University. He received his bachelor's degree from Tongji University. His research interests include uncertainty quantification, robust simulation, ranking and selection and stochastic programming. His email address is transformer-shi@tongji.edu.cn.

ZHAOLIN HU is a professor in the School of Economics and Management at Tongji University. His current research interests include stochastic optimization, simulation theory and practice, machine learning, and risk management. His email address is russell@tongji.edu.cn.

YING ZHONG is an associate professor in the Department of Management Science and E-commerce at the University of Electronic Science and Technology of China. His research interests include simulation optimization, the multiarmed bandit problem with covariates, and parallel computing. His email address is yzhong4@uestc.edu.cn.