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

Simulation Optimization when Facing Input Uncertainty

Enlu Zhou

School of Industrial & Systems Engineering
Georgia Institute of Technology
755 Ferst Drive NW
Atlanta, GA 30332, USA

Wei Xie

Department of Industrial and Systems Engineering
Rensselaer Polytechnic Institute
110 Eighth Street
Troy, NY 12180, USA

ABSTRACT

Simulation optimization usually assumes a known input distribution for the simulation model. However, the input distribution is often estimated from a finite amount of past data and hence is subject to uncertainty, which is usually referred to as "input uncertainty" in the simulation literature. This paper makes an attempt at the question of what is a good formulation for simulation optimization when we face input uncertainty. We propose a risk formulation of simulation optimization that tries to balance the trade-off between optimizing under the estimated input model and hedging against the risk brought by input uncertainty. A simple numerical example that compares the risk formulation with the usual simulation optimization shows that the risk formulation is more preferable under some conditions such as when the data size is small and the objective function value is sensitive to deviation around the optimal solution. However, more rigorous characterizations are still needed in determining which formulation to use for simulation optimization under input uncertainty.

1 INTRODUCTION

We consider the following simulation optimization problem:

$$\min_{x\in\mathscr{X}} H(x) = E_{\xi}[h(x,\xi)], \tag{1}$$

where the solution space $\mathscr{X}$ is a non-empty subset of $\mathbb{R}^d$, and the random variable ξ represents the stochastic effects of the system. In simulation optimization, the system performance (or in other words, the objective function $H(x)$) is evaluated through simulation, and hence only sample performance $h(x,\xi)$ is available. The distribution of ξ, often called the input distribution, is usually estimated from past data and then used to generate samples to drive the simulation. For example, in a queueing network the true distribution of the customer interarrival times is often estimated from the past data of customer arrival times, and in a supply chain system the customer demand distribution is often estimated from past sales data. The finiteness of past data leads to uncertainty in the estimated input distribution. However, this input uncertainty is often ignored in simulation optimization; rather, the estimated input distribution is used as if it were the true distribution of ξ. This approach brings up at least two questions pertaining to simulation optimization when there is input uncertainty.

- First, how to quantify the impact of input uncertainty on the optimization results of (1)? Clearly, each different data set, though from the same unknown distribution, will lead to a different input distribution estimate and hence lead to a different optimal solution of (1). Hence, it is important to know how to interpret such an optimization result and how far (statistically) it is away from the true optimization result. There is a rich body of work studying the impact of input uncertainty on the system performance evaluation without concerning optimization; see, e.g., survey papers by

Henderson 2003, Barton 2012, Song, Nelson, and Pegden 2014. On the other hand, the impact of input uncertainty on the optimization has been studied for the special case when the input distribution is chosen as the empirical distribution based on identically and independently distributed (i.i.d.) data, i.e., (1) is essentially the sample average approximation (SAA) of the original optimization problem under the true input distribution; statistical and convergence properties of SAA haven been studied by, e.g., Shapiro and Nemirovski 2005, Kim, Pasupathy, and Henderson 2015, Lam and Zhou 2015.

- A second question that is quite related with the first is, how to make decisions or optimize the system performance in view of the input uncertainty? The answer to the first question above will probably give us a statistical range (such as a confidence interval) that contains the optimal solution; however, most often we can only apply a single decision (rather than a range of solutions) in practice. So it would be ideal to find one solution that not only optimizes (1) but also hedges against input uncertainty. The simultaneous consideration of these two criteria often results in a trade-off. For example, the distributionally robust optimization (DRO) framework (e.g., Scarf, Arrow, and Karlin 1958, Delage and Ye 2010, Bertsimas, Brown, and Caramanis 2011) is often used to find the optimal solution in the worst case among all possible input distributions.

The focus of this paper is to make an attempt at the second question, i.e., how to carry out simulation optimization when we face input uncertainty. As mentioned above, one way to account for input uncertainty is to use a DRO formulation that looks for the worst-case input distribution among all possibilities supported by the data. Another natural approach would be to optimize an expected objective function that is averaged over all possible input distributions. These two approaches are like the two extremes: the DRO formulation puts all the weight on the worst-case input distribution and is often considered to be overly conservative to risk, and whereas averaging is completely risk neutral to all possibilities. It can be imagined there is a wide spectrum between the two extremes. Indeed we can bridge these two extremes by taking a more flexible attitude towards the risk associated with input uncertainty. Moreover, we have more knowledge than just the set of possible input distributions: we can have the probabilistic structure over the set of input distribution by computing a Bayesian posterior distribution, which represents our belief about the likelihood of input distributions based on data. By utilizing this information, we can impose a risk measure (with respect to the posterior distribution) on the objective function to hedge against the input uncertainty. Therefore, we will propose a risk formulation of simulation optimization when facing input uncertainty. In particular, this new formulation can be shown to include the DRO and averaging formulations as special cases.

The rest of the paper is organized as follows. We will briefly review simulation optimization and input uncertainty quantification in Section 2. In Section 3, we will introduce a new risk formulation of simulation optimization and show its consistency to the usual formulation. In Section 4, we will study a simple numerical example to reveal some insights on the new formulation compared with the usual formulation. Finally we will conclude and outline future directions in Section 5.

2 Literature Review

Simulation optimization has been a challenging problem due to several reasons such as the expensive evaluation of system models, lack of structure in the performance measure, and the need to balance estimation and optimization. As characterized by Fu, Chen, and Shi 2008, there are four main classes of approaches to simulation optimization over continuous solution space: (i) sample average approximation, e.g., de Mello, Shapiro, and Spearman 1999; (ii) stochastic gradient methods or stochastic approximation, e.g. Kiefer and Wolfowitz 1952, Kushner and Yin 2004; (iii) sequential response surface methodology, e.g., Barton and Meckesheimer 2006, Chang, Hong, and Wan 2013; and (iv) deterministic metaheuristics, a broad category of methods that generalize deterministic metaheuristics to the simulation optimization setting, e.g., Olafsson 2006, Andradóttir 2006. When the solution space is finite and relatively small so

that every solution can be simulated, the problem is often under the name of "ranking and selection" (see Kim and Nelson 2006).

Simulation optimization inevitably requires the estimation of system performance, which itself is an interesting problem that has drawn great attention. Earlier work has focused on efficient estimation, such as variance reduction techniques, while more recent research has studied the impact of input uncertainty on performance evaluation. Numerous methods have been proposed to quantify the uncertainty in performance evaluation and estimation, including analytical methods such as the delta method based on Taylor theorem, e.g. Cheng and Holloand 1997; Bayesian approaches, such as Bayesian model average (BMA) method Chick 2001, Zouaoui and Wilson 2003, and Biller and Corlu 2011; direct and bootstrap sampling methods, e.g., Barton and Schruben 1993, Barton and Schruben 2001; and meta-model assisted approaches, e.g., Barton, Nelson, and Xie 2014, Xie, Nelson, and Barton 2015.

The aforementioned literature consider either simulation optimization under a fixed input distribution or input uncertainty quantification for performance evaluation without concerning optimization. However, they are important building blocks for studying simulation optimization under input uncertainty. Recently Corlu and Biller 2013 investigates a ranking-and-selection problem and develops a subset selection procedure by accounting for parameter uncertainty in the input distribution. On the other hand, distributionally robust optimization (DRO) was first introduced by Scarf, Arrow, and Karlin 1958 in an inventory control problem and provides a nice framework for stochastic optimization under input uncertainty. The surge of data-driven applications in recent years has further advanced the research in DRO. However, different from simulation optimization where problems often lack nice structure and are only evaluated by simulation, research in DRO has put a great emphasis on the construction of uncertainty sets such that the problem is tractable either analytically or computationally by exploiting nice structure properties such as convexity. For example, the uncertainty set can be defined by constraints on the moments of the input distribution, e.g., Scarf, Arrow, and Karlin 1958, Delage and Ye 2010, Wiesemann, Kuhn, and Sim 2014, or by constraints on the support of the input distribution, e.g., Shapiro 2006, or as a set confined by a distance (such as ϕ-divergence) from a nominal distribution, e.g., Ben-Tal, den Hertog, Waegenaere, Melenberg, and Rennen 2013, or a set based on statistical hypothesis tests, e.g., Bertsimas, Gupta, and Kallus 2014.

3 Risk Formulation of Simulation Optimization

Recall that the true distribution of ξ in (1) is unknown, but we are given i.i.d. data $\phi = (\xi_1, \ldots, \xi_n)$ of ξ. Depending on the context throughout the paper, $\xi_1, \ldots, \xi_n$ could denote either realizations of the samples or random variables that are i.i.d. copies of ξ. For ease of exposition, we assume the true distribution of ξ lives in a parameterized family of distributions $\{f(\xi;\theta), \theta \in \Theta\}$, and in particular the true input parameter value θ^c is in the interior of Θ. To estimate the true distribution of ξ from data, we adopt a Bayesian approach by viewing θ as a random variable. The Bayesian posterior distribution will give us a full characterization of the space of all possible input distributions, which is the information we need to hedge against input uncertainty. We pick a prior $p(\theta)$ that represents our initial belief about the parameter value. Then with Bayesian updating, we obtain a posterior distribution

$$p(\theta|\phi) \propto p(\theta)p(\phi|\theta) = p(\theta)\Pi_{i=1}^{n} f(\xi_i;\theta),$$

where $p(\phi|\theta) = \Pi_{i=1}^{n} f(\xi_i;\theta)$ is the likelihood of data ϕ, and the notation $\propto$ denotes equivalence up to a normalization constant. Under some regularity conditions (notably that θ^c is an interior point of Θ), as the data size $n \to \infty$, the posterior distribution approaches normality with mean θ^c and variance $\{nJ(\theta^c)\}^{-1}$, where $J(\theta^c)$ is the Fisher information at θ^c (see Section 4.2 and Appendix B in Gelman et al. 2014). It implies the posterior distribution will become more and more concentrated on the true parameter value as data size increases. It is consistent with our intuition that the input uncertainty should decrease as we have more data, and in the extreme case when we have infinite amount of data we should recover the true input distribution. Please note that the Bayesian approach can also be applied to nonparametric distributions, using Dirichlet processes for example, and the rest of our formulation would still follow.

Now with the posterior distribution $p(\theta|\phi)$ that represents our current belief about the input parameter, we can use a risk measure $\rho_{p(\theta|\phi)}(\cdot)$ to gauge the risk associated with input uncertainty. Therefore, we propose the following new formulation of simulation optimization to account for input uncertainty:

$$\min_{x\in\mathscr{X}} H^{\rho}(x) = \rho_{\theta}\left\{E_{\xi}[h(x;\xi)]\right\}, \tag{2}$$

where ρ_{θ} is short for $\rho_{p(\theta|\phi)}(\cdot)$, and E_{ξ} short for $E_{f(\xi;\theta)}$. Some salient examples of ρ include expectation, mean-variance, Value-at-Risk (VaR), and Conditional Value-at-Risk (CVaR). We will elaborate on each of them below.

First, when ρ is an expectation, (2) can be written as

$$\min_{x\in\mathscr{X}} \quad E_{\theta}\left\{E_{\xi}[h(x;\xi)]\right\} = E_{\theta,\xi}[h(x;\xi(\theta))], \tag{3}$$

where the expectation $E_{\theta,\xi}$ is with respect to the joint distribution $f(\xi;\theta)p(\theta|\phi)$. It essentially reduces to the usual formulation of simulation optimization but under the joint distribution of θ and ξ. This formulation is neural to the risk due to both the uncertainty associated with the input parameter θ and the uncertainty due to stochastic simulation of ξ.

To incorporate the risk aspect, Markowitz 1952 introduced the mean-variance formulation into portfolio theory, aiming to strike a balance between expected return and the variability. With the mean-variance choice of ρ, our formulation (2) can be written as

$$\min_{x\in\mathscr{X}} \quad E_{\theta}\left\{E_{\xi}[h(x;\xi)]\right\} + a\mathrm{Var}\left\{E_{\xi}[h(x;\xi)]\right\}, \tag{4}$$

where a is a positive constant that can be used to adjust the trade-off.

When ρ is chosen as the α-level VaR, formulation (2) can be written as

$$\min_{x\in\mathscr{X}} \quad \mathrm{VaR}_{\alpha}\left\{E_{\xi}[h(x;\xi)]\right\}, \tag{5}$$

where $\mathrm{VaR}_{\alpha}(l(\theta))$ (here $l(\theta) = E_{\xi}[h(x;\xi)]$) is defined as the α quantile of the loss function $l(\theta)$: $\mathrm{VaR}_{\alpha}(l(\theta)) \triangleq \inf\{t : F(t) \geq \alpha\}$, where $F(\cdot)$ is the cumulative distribution function (c.d.f.) of $l(\theta)$. If $l(\theta)$ is a continuous random variable, then the α-level VaR can be simplified as $\mathrm{VaR}_{\alpha}(l(\theta)) = F^{-1}(\alpha)$.

While VaR has been used and studied extensively, it is not a coherent risk measure because it does not always satisfy the subadditivity axiom (see e.g. Artzner, Delbaen, Eber, and Heath 1999). On the other hand, CVaR is a coherent risk measure and possesses nice properties such as convexity. Letting the risk measure ρ be CVaR, then formulation (2) can be written as

$$\min_{x\in\mathscr{X}} \quad \mathrm{CVaR}_{\alpha}\left\{E_{\xi}[h(x;\xi)]\right\}, \tag{6}$$

where $\mathrm{CVaR}_{\alpha}\left\{E_{\xi}[h(x;\xi)]\right\} = E_{\theta}\left\{E_{\xi}[h(x;\xi)]|E_{\xi}[h(x;\xi)] \geq \mathrm{VaR}_{\alpha}\right\}$, and VaR_{α} is a shorthand notation for $\mathrm{VaR}_{\alpha}\left\{E_{\xi}[h(x;\xi)]\right\}$. It can be shown that $\mathrm{CVaR}_{\alpha}\left\{E_{\xi}[h(x;\xi)]\right\} = \frac{1}{1-\alpha}E_{\theta}\{E_{\xi}[h(x;\xi)]I\{E_{\xi}[h(x;\xi)] \geq \mathrm{VaR}_{\alpha}\}\}$, where $I\{A\}$ is an indicator function whose value is 1 if A is true and 0 otherwise. Intuitively, VaR can be understood as the lower bound of large losses, and CVaR is the conditional expectation of large losses.

To put things into perspective, we point out the connection of our formulation (2) with some existing formulations. The expectation formulation (3) parallels the averaging approach taken by Zouaoui and Wilson 2003 and Chick 2001 for performance evaluation, which takes into account both input uncertainty and stochastic uncertainty (i.e., the uncertainty in stochastic simulation caused by ξ). The VaR formulation (5), when α set to be 100%, becomes DRO with the uncertainty set $\tilde{\Theta} \subseteq \Theta$ being the support of the posterior distribution, i.e.,

$$\min_{x\in\mathscr{X}} \mathrm{VaR}_{100\%}\left\{E_{\xi}[h(x;\xi)]\right\} \quad = \quad \min_{x\in\mathscr{X}} \max_{\theta\in\tilde{\Theta}} E_{\xi}[h(x;\xi)].$$

The Bayesian posterior distribution $p(\theta|\phi)$ can be viewed as a "softer" constraint than the uncertainty set in DRO, as it provides a probability structure over the entire parameter set rather than a zero-or-one partition of the set. Moreover, the choice of α level in formulation (5) allows the freedom to adapt to one's risk preference, as opposed to the DRO formulation that always hedges against the worst case.

3.1 Consistency of the risk formulation

The following theorem shows that as the data size increases the risk formulation (2) approaches the original simulation optimization problem under the true input distribution. To simplify notations, we denote the response for any fixed x by $l(\theta) \triangleq E_{f(\xi;\theta)}[h(x;\xi)]$, where we suppress the dependence on x for simplicity. Note that $l(\theta^c)$ is the response under the true input distribution. Let $P_n(\cdot)$ denote the distribution function of $p(\theta|\xi_1,\ldots,\xi_n)$ and $G_n(\cdot)$ the distribution function of $l(\theta)$ conditional on $\xi_1,\ldots,\xi_n$, i.e. $G_n(A) = Pr(l(\theta) \in A|\xi_1,\ldots,\xi_n)$, where A is a measurable set in Θ.

Assumption 1 The parameter set Θ is a compact set, and any small neighborhood of θ^c has a nonzero prior probability.

Assumption 2 The posterior distribution $p(\theta|\phi)$ is a continuous distribution, and the function $l(\cdot)$ is continuous.

Theorem 1 Under Assumptions 1 and 2, for any fixed $x \in \mathscr{X}$, the following convergence results hold in probability with respect to $f(\cdot;\theta^c)$:
(i) For any neighborhood B that contains $l(\theta^c)$, $G_n(B) \triangleq Pr(l(\theta) \in B|\phi) \to 1$ as $n \to \infty$;
(ii) For any of the above choices of ρ (i.e., expectation, mean-variance, VaR, CVaR),

$$\rho_\theta\left\{E_\xi[h(x;\xi)]\right\} \to E_{f(\xi;\theta^c)}[h(x;\xi)] \text{ as } n \to \infty.$$

To show Theorem 1, we will use the following consistency result of the posterior distribution of θ, which is a direct application of a theorem in Appendix B of Gelman et al. 2014.

Lemma 1 (Gelman et al. 2014) Under Assumption 1, if A is a neighborhood of θ^c, then $P_n \triangleq Pr(\theta \in A|\phi) \to 1$ in probability (with respect to $f(\cdot;\theta^c)$) as $n \to \infty$.

All the convergence results stated below are in probability with respect to $f(\cdot;\theta^c)$. To show Theorem 1(i), suppose B is a neighborhood of $l(\theta^c)$. The inverse image of B is defined by $l^{-1}(B) = \{\theta \in \Theta : l(\theta) \in B\}$. Since $l(\cdot)$ is a continuous function, $l^{-1}(B)$ is a neighborhood of θ^c. Hence, $G_n(B) = Pr(l(\theta) \in B|\phi) = Pr(\theta \in l^{-1}(B)|\phi) \to 1$ as $n \to \infty$, where the convergence follows from Lemma 1. Since B can be arbitrarily small, it implies that G_n converges to a point mass on $l(\theta^c)$ as n goes to infinity.

Now we will show the statement in Theorem 1 for the different choices of ρ mentioned above, respectively. First, let ρ be the expectation. Then given a neighborhood A of θ^c, $H^\rho(x) = E_\theta[l(\theta)] = \int_A l(\theta)P_n(d\theta) + \int_{\Theta\setminus A} l(\theta)P_n(d\theta) \to \int_A l(\theta)P_n(d\theta)$, where the converge follows from Lemma 1. Since A can be made arbitrarily small, $H^\rho(x) \to l(\theta^c) = E_{f(\xi;\theta^c)}[h(x;\xi)]$ as $n \to \infty$.

Second, let ρ be the mean-variance. Then $H^\rho(x) = E_\theta[l(\theta)] + c\mathrm{Var}_\theta[l(\theta)]$, where c is a positive constant. It is sufficient to show the variance term goes to 0. With the same approach above, we can show $\mathrm{Var}_\theta[l(\theta)] = E_\theta\left[l(\theta)^2\right] - E_\theta[l(\theta)]^2 \to l(\theta^c)^2 - l(\theta^c)^2 = 0$.

Third, let ρ be the α-level VaR. Then $\mathrm{VaR}_{\theta,\alpha}(l(\theta)) = \inf\{t : G_n((-\infty,t]) \geq \alpha\}$. Using result (i) of Theorem 1, since $l(\theta^c) \in (-\infty, l(\theta^c)]$, $G_n((-\infty, l(\theta^c)]) \to 1$; on the other hand, since $l(\theta^c) \notin (-\infty, l(\theta^c))$, $G_n((-\infty, l(\theta^c)) \to 0$. Hence, $\inf\{t : G_n((-\infty,t]) \geq \alpha\} \to l(\theta^c)$ as $n \to \infty$.

Last, let ρ be the α-level CVaR. Then $\mathrm{CVaR}_{\theta,\alpha}(l(\theta)) = \frac{1}{1-\alpha}E_\theta[l(\theta)I\{l(\theta) \geq v_{\alpha,n}\}]$, where $v_{\alpha,n} = \inf\{t : G_n((-\infty,t]) \geq \alpha)$. Denote by $y = l(\theta)$. Thus, $\mathrm{CVaR}_{\theta,\alpha}(l(\theta))$ can be rewritten as

$$\frac{1}{1-\alpha}E_\theta[l(\theta)I\{l(\theta) \geq v_{\alpha,n}\}] = \frac{1}{1-\alpha}\int yI\{y \geq v_{\alpha,n}\}G_n(dy) = \frac{1}{1-\alpha}\int yI\{G_n(y) \geq \alpha\}G_n(dy).$$

Now it is sufficient to show that the truncated distribution function $G'_n(dy) = \frac{I\{G_n(y)\geq\alpha\}G_n(dy)}{1-\alpha}$ converges to a point mass on $l(\theta^c)$. From Theorem 1(i), we know that given a neighborhood B of $l(\theta^c)$, for any

$\varepsilon > 0$, there exists a positive integer N such that for any $n \geq N$, $\int_{\Theta\setminus B} G_n(dy) \leq \varepsilon$. Hence, $\int_{\Theta\setminus B} G'_n(dy) \leq \int_{\Theta\setminus B} \frac{1}{1-\alpha} G_n(dy) \leq \frac{\varepsilon}{1-\alpha}$. It implies $\int_B G'_n(dy) = 1 - \int_{\Theta\setminus B} G'_n(dy) \to 1$ as $n \to \infty$. Therefore, $\int y G'_n(dy) = \int_B y G'_n(dy) + \int_{\Theta\setminus B} y G'_n(dy) \to l(\theta^c)$. The statement is proved.

4 A Numerical Example

We will illustrate the risk formulation of simulation optimization on a simple first-come-first-served M/M/1 queuing system. Customers arrive at a system according to a Poisson process with rate θ^c, and the service time follows an exponential distribution with mean x. There is a cost $c > 0$ per unit increase of service rate; hence, there is a trade-off between decreasing the expected average customer waiting time in system and decreasing the service cost. Moreover, there is often a practical limit on the total cost $M < \infty$, which is much higher than the minimum cost. In particular, when the system is unstable (i.e., server utilization ≥ 1), it will incur the total cost M. The objective is to find a service mean time x that minimizes the total cost:

$$\min_{x>0} H(x) = \begin{cases} \min\left\{E_{\theta^c}[T(x;\xi)] + \frac{c}{x}, M\right\}, & \text{if } \theta^c x < 1; \\ M, & \text{otherwise.} \end{cases} \tag{7}$$

where ξ represents the random interarrival time that follows the exponential distribution $f(\xi;\theta^c) = \exp(1/\theta^c)$, and $T(x;\xi)$ represents the steady-state average customer waiting time. For M/M/1 queue, $E_{\theta^c}[T(x;\xi)]$ has an analytical form $\frac{x}{1-\theta^c x}$. It is easy to see that the objective function is convex, and we can find a unique optimal solution for (7) in closed-form $x^* = \frac{\sqrt{c}}{\sqrt{c}\theta^c+1}$. This analytical solution will be used to provide insight on our numerical solutions.

In the numerical experiment, the value of θ^c is unknown, but the experimenter observes n i.i.d. interarrival time data $\xi_1, \ldots, \xi_n$ from the true underlying distribution $f(\xi;\theta^c) = \exp(1/\theta^c)$. The usual approach is to estimate the parameter by a point estimator $\widehat{\theta} = 1/(\frac{1}{n}\sum_{i=1}^n \xi_i)$, and then solve the simulation optimization problem under the estimated input model:

$$\min_{x>0} H^{\widehat{\theta}}(x) = \begin{cases} \min\left\{\frac{x}{1-\widehat{\theta}x} + \frac{c}{x}, M\right\}, & \text{if } \widehat{\theta}x < 1; \\ M, & \text{otherwise.} \end{cases} \tag{8}$$

Note that here the analytical form $E_{\widehat{\theta}}[T(x;\xi)] = \frac{x}{1-\widehat{\theta}x}$ is plugged into (8), and hence the optimal solution is $\frac{\sqrt{c}}{\sqrt{c}\widehat{\theta}+1}$. We will refer to this approach as "empirical simulation optimization" (ESO).

Instead, we adopt a Bayesian approach and use the posterior distribution to quantify the input uncertainty. Specifically, we use a Gamma distribution $Gamma(a_0, b_0)$ as a prior, which is conjugate with the exponential distribution; hence, the posterior is

$$p(\theta|\phi) \triangleq p(\theta|\xi_1, \ldots, \xi_n) = Gamma(a_0 + n, b_0 + \sum_{i=1}^n \xi_i).$$

We then solve the risk formulation of the simulation optimization problem, which we will refer to as "risk simulation optimization" (RSO):

$$\min_{x>0} H^{\rho}(x) = \begin{cases} \rho_{p(\theta|\phi)}\left[\min\left\{\frac{x}{1-\theta x}, M\right\}\right] + \frac{c}{x}, & \text{if } xE_{p(\theta|\phi)}[\theta] \leq 1; \\ M, & \text{otherwise,} \end{cases} \tag{9}$$

where ρ is one of the four choices: expectation, mean-variance, VaR_α, CVaR_α. Note that (9) is a stochastic optimization problem with expectation constraint. So we use sample average approximation (SAA) to solve the problem. That is, we draw i.i.d. samples $\theta_1, \ldots, \theta_m$ from the posterior $p(\theta|\phi)$, and then solve problem (9) with $p(\theta|\phi)$ replaced by the empirical distribution $\hat{p}(\theta|\phi) = \frac{1}{m}\sum_{i=1}^m I(\theta = \theta_i)$.

Each formulation yields the respective optimal solution $\widehat{x^*}$. To assess the performance of these solutions, we define a performance measure by the expected square-deviation in the function value of each solution from the true optimal function value $H(x^*)$:

$$D(\widehat{x^*}) = E\left[\left(\frac{H(\widehat{x^*}) - H(x^*)}{H(x^*)}\right)^2\right],$$

where the expectation is with respect to the joint distribution of $\{\xi_1, \ldots, \xi_n\}$. Hence, in implementation we will run K independent replications: for each replication k, we simulate a set of i.i.d. data $\xi_1, \ldots, \xi_n$ from the underlying distribution and proceed as described above to solve each formulation to obtain $\widehat{x^{*,k}}$, and compute the average square-deviation $D = \frac{1}{K}\sum_{k=1}^{K}\left(H(\widehat{x^{*,k}})/H(x^*) - 1\right)^2$. Hence, a larger D value implies a more significant deviation from the true optimal performance in average and thus more risk of the corresponding formulation due to input uncertainty. Of course there are other measures besides D, which might give us slightly different interpretation of the results.

The parameter setting is as follows: true input parameter $\theta^c = 10$ in the first case and $\theta^c = 1$ in the second case, unit service cost $c = 1$, cost limit $M = 500$, number of replications $K = 100$, weight in the mean-variance formulation $a = 20$, level in the VaR and CVaR formulation $\alpha = 0.95$, parameters in the prior Gamma distribution $a_0 = 2$ and $b_0 = 0$, sample size of SAA $m = 1000$. We use the same 1000 samples from the posterior in all risk formulations. In the first case, the true optimal solution is $x^* \approx 0.091$ and the optimal function value is $H(x^*) = 12$. In the second case, the true optimal solution is $x^* = 0.5$ and the optimal function value is $H(x^*) = 3$.

Tables 1 and 2 show the numerical results of the two cases respectively. The first column of the table shows the data size n, which varies from 10 to 1000; under each formulation, the first subcolumn shows the average of solved optimal solutions over all replications with the standard error in parentheses below, and the second subcolumn shows the average square-deviation D. From the numerical results, we have made the following observations.

- In case 1 (see Table 1), although the average of the solutions of ESO is closer to the true optimal solutions, the associated function value deviation is much larger compared with the risk formulations for all data sizes tested except 1000. If we scrutinize the objective function, it is not hard to see that the objective function has a steep (positive) slope to the right of the optimal solution (see left panel of Fig. 1). Intuitively, since in this case the optimal server utilization is $x^*\theta^c \approx 0.91$, a deviation from the true optimal solution to the right (due to an estimator $\widehat{\theta}$ smaller than θ^c) will drive the utilization closer to or even higher than 1, which causes the average waiting time to explode. The risk formulations, however, exhibit a strong resistance to such a deviation from the optimal solution in order to avoid the extremely large cost; hence, their solutions all tend to be smaller than the true optimal solution.
- In contrast to case 2 (see Table 2), the ESO formulation yields smaller function value deviation from the beginning even when the data size is only 10. That is because in this case, the objective function is very flat around the true optimal solution $x^* = 0.5$ (see the right panel of Fig. 1); and hence, a reasonable deviation in the solution will not cause much deviation in the functional value. In other words, here the optimal server utilization is $x^*\theta^c = 0.5$, and therefore the average waiting time is relatively stable around this utilization value. The risk formulations still yield smaller solutions due to their conservativeness in hedging against larger costs, and the associated function deviations are slightly higher than ESO.
- As data size n increase, the differences between all formulations become smaller and smaller, which is guranteed by Theorem 1. Specifically, solutions of all formulations approach the true optimal solution with standard errors going to zero, and the associated functional value deviations also

approach zero. When the data size is large (e.g. $n = 1000$), ESO performs the best compared to risk formulations, which is probably due to a faster convergence of ESO to the original optimization problem. Convergece rates of the different formulations to the orignal optimization problem will be a future study.

- Among all risk formulations, the expecation formulation appears to have the smallest fuction value deviations and least conservative solutions when the data size is not too small (e.g. $n > 20$). This is because VaR and CVaR formulations try to avoid the extreme large cost (in the right tail of the true objective function) by choosing a smaller mean service time x, which is usually smaller than the true optimal x^* and thus leads to a slightly increasing cost compared to the expectation formulation. When the data size is small ($n = 10$ in this example), we observe in experiments that the VaR and CVaR formulations are more robust while the expectation formulation is better in average.

From this simple example, we can make a rough conclusion that it is better to use the risk formulations of simulation optimization when the data size is relatively small and when the variability of the objective function value around the optimal solution is relatively large. However, a more precise set of conditions will be needed to make recommendations on when to use the risk formulations over the traditional simulation optimization formulation.

Table 1: Comparison of different formulations (Case 1: true input parameter $\theta^c = 10$, true optimal server utilization $x^*\theta^c \approx 0.91$)

	ESO		Mean RSO		Mean-Var RSO		VaR RSO		CVaR RSO	
n	$\widehat{x^*}$	$D(\widehat{x^*})$	$\widehat{x^*}$	$D(\widehat{x^*})$	$\widehat{x^*}$	$D(\widehat{x^*})$	$\widehat{x^*}$	$D(\widehat{x^*})$	$\widehat{x^*}$	$D(\widehat{x^*})$
10	0.092 (0.003)	662	0.052 (0.002)	0.910	0.043 (0.002)	18.4	0.061 (0.002)	33.6	0.048 (0.002)	1.17
20	0.091 (0.002)	463	0.059 (0.001)	0.320	0.052 (0.001)	0.636	0.067 (0.001)	33.2	0.054 (0.001)	0.527
50	0.090 (0.001)	281	0.068 (0.001)	0.094	0.064 (0.001)	0.150	0.074 (0.001)	0.047	0.064 (0.001)	0.138
100	0.090 (0.0008)	167	0.075 (0.0007)	0.032	0.071 (0.0007)	0.058	0.078 (0.0007)	0.018	0.072 (0.0007)	0.051
1000	0.091 (0.0003)	0.0008	0.089 (0.0003)	0.0002	0.085 (0.0002)	0.001	0.087 (0.0002)	0.0005	0.086 (0.0002)	0.0009

5 Conclusion and future directions

In this paper, we proposed a new risk formulation for simulation optimization in order to account for input uncertainty. We compared different risk formulations and the usual simulation optimization formulation on a simple numerical example, and confirmed that the risk formulations can yield more robust solutions when the objective function value is more sensitive to small deviations from the true optimal solution. However, the risk formulation may tend to be overly conservative otherwise.

There are several research directions to go. First, how to solve the risk formulations numerically is a challenging problem, given that the plain simulation optimization without input uncertainty is already quite difficult. Two common approaches are the sample average approximation and stochastic approximation, which can be extended to the risk formulations. Second, when to use which formulation is an interesting question. As demonstrated in this very simple example in the paper, there seems not a single choice of

Table 2: Comparison of different formulations (Case 2: true input parameter $\theta^c = 1$, true optimal server utilization $x^*\theta^c = 0.5$)

	ESO		Mean RSO		Mean-Var RSO		VaR RSO		CVaR RSO	
n	$\widehat{x^*}$	$D(\widehat{x^*})$	$\widehat{x^*}$	$D(\widehat{x^*})$	$\widehat{x^*}$	$D(\widehat{x^*})$	$\widehat{x^*}$	$D(\widehat{x^*})$	$\widehat{x^*}$	$D(\widehat{x^*})$
10	0.495 (0.008)	0.004	0.423 (0.011)	0.043	0.338 (0.008)	0.097	0.387 (0.007)	0.032	0.351 (0.008)	0.079
20	0.494 (0.006)	0.001	0.464 (0.007)	0.004	0.377 (0.005)	0.022	0.412 (0.005)	0.008	0.388 (0.005)	0.017
50	0.4984 (0.003)	0.0001	0.490 (0.003)	0.0001	0.423 (0.003)	0.002	0.444 (0.003)	0.001	0.430 (0.003)	0.002
100	0.498 (0.003)	5e-05	0.4941 (0.003)	6e-05	0.447 (0.003)	0.0008	0.459 (0.003)	0.0004	0.449 (0.003)	0.0007
1000	0.499 (8e-04)	4e-07	0.500 (8e-04)	4e-07	0.490 (8e-04)	2e-06	0.486 (8e-04)	4e-06	0.483 (8e-04)	6e-06

the best formulation(s). More precise conditions and more insights about the problem structure may be needed to determine a good formulation to use.

ACKNOWLEDGMENTS

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

REFERENCES

Andradóttir, S. 2006. "Chapter 20 An overview of simulation optimization with random search". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. Henderson and B. Nelson, 617–632: Elsevier.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. "Coherent measures of risk". *Mathematical Finance* 9 (3): 203–228.

Barton, R., and M. Meckesheimer. 2006. *Handbooks in Operations Research and Management Science: Simulation*, Chapter 18: Metamodel-Based Simulation Optimization, 535 – 574. Elsevier.

Barton, R., B. Nelson, and W. Xie. 2014. "Quantifying Input Uncertainty via Simulation Confidence Intervals". *INFORMS Journal on Computing* 26 (1): 74–87.

Barton, R. R. 2012. "Tutorial: Input Uncertainty in Output Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, 67–78.

Barton, R. R., and L. W. Schruben. 1993. "Uniform and Bootstrap Resampling of Input Distributions.". In *Proceedings of the 1993 Winter Simulation Conference*, 503–508.

Barton, R. R., and L. W. Schruben. 2001. "Resampling Methods for Input Modeling". In *Proceedings of the 2001 Winter Simulation Conference*, edited by D. J. M. B. A. Peters, J. S. Smith and e. M. W. Rohrer, 372–378.

Ben-Tal, A., D. den Hertog, A. D. Waegenaere, B. Melenberg, and G. Rennen. 2013. "Robust solutions of optimization problems affected by uncertain probabilities". *Management Science* 59 (2): 341–357.

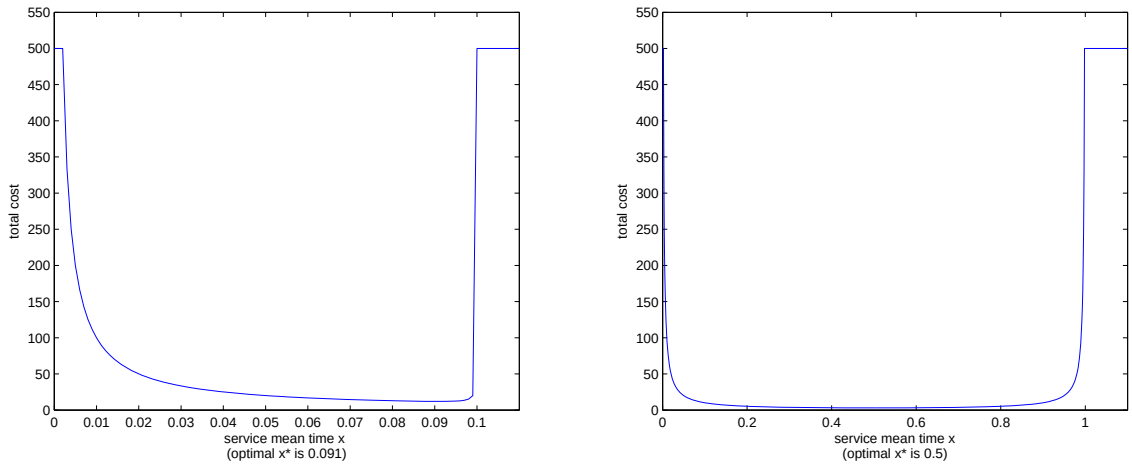

Figure 1: Objective functions for case 1 (left) and case 2 (right)

Bertsimas, D., D. Brown, and C. Caramanis. 2011. "Theory and applications of robust optimization". *SIAM Review* 53 (3): 4614–501.

Bertsimas, D., V. Gupta, and N. Kallus. 2014. "Data-driven robust optimization". Submitted.

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.

Chang, K.-H., L. J. Hong, and H. Wan. 2013. "Stochastic trust-region response-surface method (STRONG) A new response surface framework for simulation optimization". *INFORMS Journal on Computing* 25:230–243.

Cheng, R. C., and W. Holloand. 1997. "Sensitivity of computer simulaiton experiments to errors in input data". *Journal of Statistical Computation and Simulation* 57 (1-4): 219–41.

Chick, S. 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*, 463–473.

de Mello, T. H., A. Shapiro, and M. L. Spearman. 1999. "Finding optimal material release times using simulation-based optimization". *Management Science* 45:86–102.

Delage, E., and Y. Ye. 2010. "Distributionally robust optimization under moment uncertaity with applications to data-driven problems". *Operations Research* 58 (3): 595–612.

Fu, M. C., C.-H. Chen, and L. Shi. 2008. "Some Topics for Simulation Optimization". In *Proceedings of the 2008 Winter Simulation Conference*, 27–38.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis*. Chapman & Hall.

Henderson, S. G. 2003. "Input model uncertainty: Why do we care and what should we do about it". In *Proceedings of the 2003 Winter Simulation Conference*, 90–100.

Kiefer, J., and J. Wolfowitz. 1952. "Stochastic Estimation of the Maximum of a Regression Function". *Annals of Mathematical Statistics* 23:462–466.

Kim, S., R. Pasupathy, and S. G. Henderson. 2015. "A guide to sample average approximation". In *Handbook of Simulation Optimization*, 207–243. Springer.

Kim, S.-H., and B. Nelson. 2006. "Selecting the best system". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. Henderson and B. Nelson, Chapter 17, 501–534: Elsevier.

Kushner, H. J., and G. G. Yin. 2004. *Stochastic Approximation Algorithms and Applications*. 2nd ed. New York, NY: Springer-Verlag.

Lam, H., and E. Zhou. 2015. "Quantify Uncertainty in Stochastic Optimization". In *Proceedings of the 2015 Winter Simulation Conference*. Submitted.

Markowitz, H. M. 1952. "Portfolio Selection". *The Journal of Finance* 7 (1): 77–91.

Olafsson, S. 2006. *Handbooks in operations research and management science*, Chapter 13: Metaheuristics, 633 – 654. Elsevier.

Scarf, H., K. Arrow, and S. Karlin. 1958. "A min-max solution of an inventory problem". *Studies in the mathematical theory of inventory and production* 10:201–209.

Shapiro, A. 2006. "Worst-case distribution analysis of stochastic programs". *Mathematical Programming* 107 (1): 91–96.

Shapiro, A., and A. Nemirovski. 2005. "On complexity of stochastic programming problems". In *Continuous optimization*, 111–146. Springer.

Song, E., B. L. Nelson, and D. C. Pegden. 2014. "Advanced Tutorial: Input Uncertainty Quantification". In *Proceedings of the 2014 Winter Simulation Conference*, 162–176.

Wiesemann, W., D. Kuhn, and M. Sim. 2014. "Distributionally robust convex optimization". *Operations Research* 62 (6): 1358 – 1376.

Xie, W., B. L. Nelson, and R. R. Barton. 2015. "A Bayesian Framework for Quantifying Uncertainty in Stochastic Simulation". *Operations Research*. Accepted.

Zouaoui, F., and J. R. Wilson. 2003. "Accounting for parameter uncertainty in simulation input modeling". *IIE Transactions* 35 (9): 781–792.

AUTHOR BIOGRAPHIES

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

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