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.

ROBUST SIMULATION OF STOCHASTIC SYSTEMS WITH INPUT UNCERTAINTIES MODELED BY STATISTICAL DIVERGENCES

Zhaolin Hu

School of Economics and Management
Tongji University
Shanghai 200092, China

L. Jeff Hong

Department of Management Sciences
City University of Hong Kong
Kowloon, Hong Kong, China

ABSTRACT

Simulation is often used to study stochastic systems. A key step of this approach is to specify a distribution for the random input. This is called input modeling, which is important and even critical for simulation study. However, specifying a distribution precisely is usually difficult and even impossible in practice. This issue is called input uncertainty in simulation study. In this paper we study input uncertainty when using simulation to estimate important performance measures: expectation, probability, and value-at-risk. We propose a robust simulation (RS) approach, which assumes the real distribution is contained in a certain ambiguity set constructed using statistical divergences, and simulates the maximum and the minimum of the performance measures when the distribution varies in the ambiguity set. We show that the RS approach is computationally tractable and the corresponding results can disclose important information about the systems, which may help decision makers better understand the systems.

1 INTRODUCTION

Simulation is an important tool for studying complex stochastic systems. In a typical simulation approach, one builds some model to simulate (approximate) the real system, and then analyzes the model to study the real system. The model is called a simulation model, which together with specified logic maps the inputs to the outputs. Put it in mathematics, we use ξ to denote the input parameters, where ξ is a k-dimensional random vector supported on $\Xi \subset \Re^k$ (We assume k is deterministic throughout the paper). Suppose the output of interest is $H(\xi)$ where $H(\cdot)$ is a single-valued function. Simulation specifies a distribution P_0 for ξ. It then takes samples from P_0, and infers the information about $H(\xi)$ based on the samples. The inferred information is finally used to analyze the real system and to guide the decision making.

In stochastic systems, the output $H(\xi)$ is usually stochastic, which creates difficulty for decisions. To resolve this difficulty, people suggest various performance measures. A performance measure is a functional that maps random quantities (distributions) into deterministic numbers. It is a fundamental notion in decision making and based on it, decision makers can perceive the results in a much easier way. When used to measure the potential risk in risk management, performance measures are often named as risk measures. In this paper, we consider three most important performance measures: expectation, probability, and value-at-risk (VaR). With the setting that ξ follows P_0, we denote the expectation as $\mathrm{E}_{P_0}[H(\xi)]$, where $\mathrm{E}_{P_0}[\cdot]$ indicates that the expectation is taken with respect to (w.r.t.) P_0. Expectation is often the first choice in simulation study, which measures the average value of the output. Probability is another widely used measure. It measures the chance of some random event, desired or undesired. Consider some random event $A(\xi)$ where the randomness is introduced by ξ. We denote the probability of $A(\xi)$ as $\Pr_{P_0}\{A(\xi)\}$ where $\Pr_{P_0}\{\cdot\}$ means the probability is taken w.r.t. P_0. For instance, suppose the loss of a financial activity is $H(\xi)$. Then $A(\xi) := \{H(\xi) \le v\}$ is the event that the loss does not exceed the given threshold v and $\Pr_{P_0}\{A(\xi)\}$ is the probability of this event. Probability is often advocated by decision makers who are risk-averse to randomness. Value-at-risk (VaR), as a risk measure, was proposed in 1990s, and has been considered as a standard risk management tool in the financial industry. Mathematically, VaR is the quantile of a loss distribution. For a given confidence level $\beta \in (0,1)$, the $(1-\beta)$-VaR of a random loss $H(\xi)$ is defined as $\mathrm{VaR}_{1-\beta,P_0}(H(\xi)) := \inf\{v \in \Re : \Pr_{P_0}\{H(\xi) \leq v\} \geq 1-\beta\}$, where, similarly, the subscript P_0 denotes that the VaR is calculated when ξ follows P_0. For a thorough introduction and treatment of VaR, readers are referred to Jorion (2006) and Hong et al. (2014).

Due to the complexity of the performance measures and the system itself, it is often difficult to derive analytical expressions for the performance measures, and in most situations, we have to estimate them via some techniques. Simulation techniques are then often developed. Besides building the system logic, a very important step for simulation is to specify a (joint) distribution for the input parameters which model the randomness of the system. This is called input modeling. Input modeling is fundamental for simulation study. The reason is that simulation outputs depend critically on the input distribution. Actually, the outputs are determined by the input distribution and the system structure (logic). Unfortunately, there is no true distribution just waiting to be found. In real applications, people usually use information available to infer the distribution. When there are data, one can specify a distribution via statistical fitting. When there are no data, a subjective distribution is then often used. In any of the situations, it is a rare case that P_0 can be determined precisely, and there often exist profound uncertainties for P_0. When the selected distribution is not precise, the simulation output becomes unreliable, leading even to incorrect decisions. This issue is named as *input uncertainty* of simulation.

Input uncertainty has for long been a fundamental issue in simulation study. According to Barton (2012), there already existed systematic discussions on this issue in the 1992 Winter Simulation Conference. Since then, the issue has attracted significant investigations, see, e.g., Chick (2001), Henderson (2003), Ng and Chick (2006), Barton et al. (2014), and Xie et al. (2014). As the most recent studies, Barton et al. (2014) investigated how to use bootstrap statistics to handle the issue. Xie et al. (2014) studied applying Bayesian methods to quantify the errors incurred by input uncertainty. Both Barton et al. (2014) and Xie et al. (2014) introduced in great detail the literature on input uncertainty study. We refer the readers to the papers and references therein.

In this paper, we follow the convention of the economics literature (see, e.g., Ellsberg (1961) and Epstein (1999)) and use the notion of "ambiguity" to describe the phenomenon that a distribution cannot be fully determined. In contrast to the existing literature, we propose a robust simulation (RS) approach to analyze input uncertainty. The basic idea is to quantify the impact of input uncertainty on performance measures, by incorporating simulation and optimization techniques. RS assumes that the real but unknown distribution is contained in a set, which we call an ambiguity set. It could be some confidence region of the real distribution constructed from data or some subjective set reflecting practitioners' cognition on the distribution uncertainty. RS then considers the maximum and the minimum of the performance measure when the distribution varies in the ambiguity set. The two extreme values represent the best and worst values of the performance measure, and thus include important information for decision makers. The notion of "robust simulation" was proposed in Hu et al. (2012), in which the authors considered input uncertainty in environmental policy simulation, and suggested capturing the worst-case performance information. More precisely, what Hu et al. (2012) adopted is a robust ranking-and-selection (R&S) approach, which can be viewed as one side of RS proposed here. A more thorough treatment of robust R&S was provided shortly by Fan et al. (2013). The idea of considering the worst case was also proposed in many other contexts. In financial risk management, Artzner et al. (1999) proposed the notion of coherent risk measure. In economics, Hansen and Sargent (2008) suggested penalizing the ambiguity in economics models. In optimization, the distributionally robust optimization (DRO) approach was suggested to handle the uncertainty in optimization models, see, e.g., Ben-Tal et al. (2013).

As discussed, in simulation study, a distribution P_0 is specified for the random vector ξ, either by statistical fitting or subjective justification. We call P_0 a nominal distribution. Such a nominal distribution is often our best guess and contains valuable information about the stochastic nature of the parameters.

Then, a natural approach to studying the effect of ambiguity is to consider some level of perturbation or deviation of the nominal distribution. This directly motivates our construction of the ambiguity set. In this paper, we model the distributional ambiguity using the so called likelihood ratio (LR). Based on the modeling purpose and information available, different ambiguity sets may be constructed via LR. We mainly consider two classes of constraints imposed on LR using convex functions: uniform constraints and expectation constraints. The uniform constraints are somewhat straightforward. They actually define a uniform band of the LR which we call a band ambiguity set. The expectation constraints naturally lead to the concept of distribution distance. The approach of seeking some distribution distance and then considering a neighborhood of the nominal distribution defined by the distance has been very popular for modeling distributional ambiguity. Many distribution distances have been suggested. Particularly, imposing some minor regularity conditions on the convex functions in the expectation constraints, we obtain the neighborhood defined by a well-known class of distances called ϕ-divergence, which contains many distances including the widely used Kullback-Leibler (KL) divergence, χ^2 distance, Hellinger distance, Variation distance, Burg entropy, and many others. The ϕ-divergence was introduced systematically in Pardo (2006), and was used by Ben-Tal et al. (2013) to model distribution ambiguities in the context of DRO. Our paper has been inspired by the work of Ben-Tal et al. (2013), as well as two other papers Ben-Tal and Teboulle (2007) and Ben-Tal et al. (2010). Ben-Tal and Teboulle (2007) discussed an old-new concept of convex risk measures. Ben-Tal et al. (2010) proposed a soft robust optimization model under ambiguity and related the model to the theory of convex risk measures.

To provide a unified framework, the analysis of this paper is conducted on an ambiguity set which combines the uniform constraints and expectation constraints. We first study the expectation performance measure. Applying the change-of-measure technique, we reformulate the RS problem as a functional convex optimization problem. Implementing the Lagrangian duality, we derive the dual of the functional problem and show that the dual belongs to the conventional stochastic optimization (SO) problems. We discuss how to apply the SO techniques sample average approximation (SAA) and stochastic approximation (SA) to solve the SO problems raised. Following the analysis, we then consider the probability performance measure. We show that the corresponding RS problem can be transformed and put into the framework of expectation and thus a similar approach can be used. Furthermore, we find that the probability function has very nice structures which allow us to convert the RS problems to simple optimization problems. The results show that estimating the maximum (and the minimum) of the probability can be accomplished by estimating the probability under the nominal distribution and by solving a simple optimization problem. We finally discuss RS of VaR. Based on the relation between VaR and probability, we build the results for VaR, which indicate that estimating the maximum (and the minimum) of the VaR can be accomplished by estimating the VaR of a new confidence level under the nominal distribution. The new confidence level can be searched via convex optimization techniques.

The rest of this paper is organized as follows. In Sections 2, 3 and 4 we discuss RS of expectation, probability and VaR respectively. The numerical example studied in Section 5 concludes the paper.

2 EXPECTATION PERFORMANCE MEASURE

We start from the expectation performance measure. Suppose the random output of a system we are interested in is $H(\xi)$. For simplicity of the notation, we suppress the dependence of H on ξ. To ease the analysis, throughout this paper, we assume H is a bounded random variable, although this assumption may be relaxed in many contexts. The RS approach considers the minimal and maximal expectations

$$E_l := \inf_{P\in\mathbb{P}} \mathrm{E}_P[H] \quad \text{and} \quad E_u := \sup_{P\in\mathbb{P}} \mathrm{E}_P[H].$$

The quantities E_l and E_u, which form an interval $[E_l, E_u]$, provide important information for the real system. In particular, they quantify the robustness/sensitivity of the system output to input modeling. In this paper, we discuss in detail how to compute (simulate) the two quantities. Note that $\inf_{P\in\mathbb{P}} \mathrm{E}_P[H] = -\sup_{P\in\mathbb{P}} \mathrm{E}_P[-H]$.

We can transform the problem of computing the infimum to computing a supremum. Therefore, we only focus on the supremum E_u, which corresponds to the following optimization problem

$$\underset{P\in\mathbb{P}}{\text{maximize}}\ \mathrm{E}_P\left[H\right]. \tag{1}$$

The formulation above depends critically on the structure of the ambiguity set $\mathbb{P}$. In what follows, we introduce the structure of $\mathbb{P}$ considered in the paper.

2.1 Ambiguity Set

Suppose we have obtained a nominal distribution P_0 for ξ. Suppose the true but unknown distribution is P. We construct the ambiguity set by considering the difference between P_0 and P. Suppose the k-dimensional distributions P and P_0 have densities $p(z)$ and $p_0(z)$ on $\Xi \subset \Re^k$. Note that we do not differentiate P and $p(z)$ throughout this paper: The two notations denote the same distribution if no confusion is caused. Let $L = p/p_0$. Note that L is called a likelihood ratio (LR) in the literature. The definition of LR implicitly assumes that P is absolutely continuous w.r.t. P_0 (denoted as $P \ll P_0$), i.e., for every measurable set A, $P_0(A) = 0$ implies $P(A) = 0$. When P_0 is a discrete distribution, we understand $p_0(z)$ as the probability mass function. When P_0 follows a mixed distribution, $p_0(z)$ is the density at z if P_0 has zero mass at z, and is the probability mass function at z if P_0 has a positive mass at z. Clearly, LR is a good candidate for measuring the perturbation/deviation of the true distribution to the nominal one. As mentioned in Section 1, we use two different classes of constraints on the LR to model the ambiguity. The first is called uniform constraints. Specifically, we consider a convex function $\varphi : \Re \to \Re$, and construct the constraint

$$\varphi(L) \le \rho, \tag{2}$$

where ρ is a positive constant. To guarantee that the nominal distribution satisfies (2), we impose the regularity condition for φ that $\varphi(1) \le \rho$. Because φ is convex and finite valued, the constraint (2) defines a closed interval for L. Furthermore, a finite number of constraints taking the form of (2) still define a closed interval. Therefore, using the uniform constraints we are arriving at a set of p such that the LR falls in an interval, i.e., $a \le L \le b$ for some $0 \le a < 1 < b \le \infty$ (We omit the degenerate case where $a = 1 = b$). Although L is itself a function of ξ, (2) requires the constraint be satisfied for all ξ. This is why we call (2) a uniform constraint.

The second class is called expectation constraints. Specifically, consider a convex function ϕ on $\Re$ and construct the constraint $\mathrm{E}_{P_0}[\phi(L)] \le \eta$. Imposing some minor regularity conditions on ϕ, we are arriving at the famous ϕ-divergence, which has been used frequently in statistics to measure the distance of a distribution to another one. Therefore, imposing constraints on LR using ϕ-divergence admits a clear statistical and practical meaning. Following the notions of Pardo (2006) and Ben-Tal et al. (2013), a ϕ-divergence function is a convex function for $t > 0$, satisfying $\phi(1) = 0$, $0\phi(a/0) := a\lim_{t\to\infty}\phi(t)/t$ for $a > 0$, and $0\phi(0/0) := 0$. For P and P_0 introduced above, the ϕ-divergence from P to P_0 is defined as

$$D_\phi(P\|P_0) = \int_\Xi p_0(z)\phi\left(\frac{p(z)}{p_0(z)}\right)\mathrm{d}z = \mathrm{E}_{P_0}\left[\phi\left(\frac{p(\xi)}{p_0(\xi)}\right)\right] = \mathrm{E}_{P_0}\left[\phi\left(L\right)\right]. \tag{3}$$

Similarly, we understand the integral in (3) as the summation if P_0 is a discrete distribution, and as a mixture of integral and summation if P_0 is a mixed distribution. It can be shown that $D(P\|P_0) \ge 0$ and the equality holds if and only if $p(\xi) = p_0(\xi)$ almost surely (a.s.) under P_0. We now construct a neighborhood $D_\phi(P||P_0) \le \eta$, which from (3) yields an expectation constraint $\mathrm{E}_{P_0}\left[\phi\left(L\right)\right] \le \eta$. As can be seen, instead of requiring L satisfy a constraint for all ξ in uniform constraints, in expectation constraints one only requires L satisfy a constraint on average.

Combining the two classes of constraints, we construct the following ambiguity set of P:

$$\mathbb{P} = \left\{P \in \mathbb{D} : a \le p/p_0 \le b,\ D_{\phi_i}(P||P_0) \le \eta_i, i = 1,\cdots,m\right\},$$

where $\mathbb{D}$ denotes the set of all probability distributions and $D_{\phi_i}(P||P_0)$ denotes the ϕ_i-divergence from P to P_0. In the ambiguity set $\mathbb{P}$, the constants a,b and $\eta_i, i=1,2,\cdots,m$ are indexes of ambiguity, which control the size of $\mathbb{P}$. In terms of L, $\mathbb{P}$ can also be represented as

$$\mathbb{L} = \{L \in \mathbb{L}(a,b) : \mathrm{E}_{P_0}[L] = 1, \mathrm{E}_{P_0}[\phi_i(L)] \le \eta_i, i = 1,\cdots,m\},$$

where $\mathbb{L}(a,b) := \{L : a \le L \le b \text{ a.s.}\}$. In what follows, we discuss how to solve the RS problem (1) with ambiguity set $\mathbb{L}$.

2.2 Robust Simulation of Expectation

Problem (1) is rather abstract for optimization. One of the major difficulties for solving the problem is that the randomness is embedded in the decision variable. A widely used technique that can separate them is the change-of-measure technique. Applying the technique, we obtain that

$$\mathrm{E}_P[H] = \int_\Xi H p(z)\mathrm{d}z = \int_\Xi H \frac{p(z)}{p_0(z)} p_0(z)\mathrm{d}z = \mathrm{E}_{P_0}[HL].$$

Recall the structure of the ambiguity set $\mathbb{P}$ and $\mathbb{L}$. Problem (1) can be written as

$$\begin{aligned} \underset{L\in\mathbb{L}(a,b)}{\text{maximize}} \quad & \mathrm{E}_{P_0}[HL] \\ \text{subject to} \quad & \mathrm{E}_{P_0}[\phi_i(L)] \le \eta_i, i=1,\cdots,m,\ \mathrm{E}_{P_0}[L] = 1. \end{aligned} \tag{4}$$

Problem (4) is a functional optimization problem with L being the decision variable. Note that the objective function in (4) is linear in L and $\phi_i, i=1,\cdots,m$ are all convex. Thus (4) is a convex optimization problem. One standard approach to handling such constrained functional optimization problem is to use the Lagrangian duality, see, e.g., Ben-Tal et al. (2010). We construct the Lagrangian functional associated with (4):

$$\begin{aligned} \ell_0(\lambda,\alpha,L) &:= \mathrm{E}_{P_0}[HL] - \sum_{i=1}^m \alpha_i\left(\mathrm{E}_{P_0}[\phi_i(L)] - \eta_i\right) + \lambda\left(\mathrm{E}_{P_0}[L] - 1\right) \\ &= \mathrm{E}_{P_0}\left[(H+\lambda)L - \sum_{i=1}^m \alpha_i\phi_i(L)\right] + \sum_{i=1}^m \alpha_i\eta_i - \lambda. \end{aligned}$$

Then Problem (4) is equivalent to

$$\underset{L\in\mathbb{L}(a,b)}{\text{maximize}}\ \underset{\lambda\in\Re,\alpha\ge 0}{\text{minimize}}\ \ell_0(\lambda,\alpha,L). \tag{5}$$

Interchanging the maximum and minimum in Problem (5), we obtain the Lagrangian dual of Problem (5):

$$\underset{\lambda\in\Re,\alpha\ge 0}{\text{minimize}}\ \underset{L\in\mathbb{L}(a,b)}{\text{maximize}}\ \ell_0(\lambda,\alpha,L). \tag{6}$$

The major concern about the primal and dual problems above are whether they have the same optimal value. Fortunately, the duality gap turns out to be zero. We summarize the result in the following theorem.

Theorem 1 The optimal values of Problems (5) and (6) are equal. The optimal value of Problem (6) is attained at some $\lambda^* \in \Re$ and $\alpha^* \ge 0$.

Theorem 1 guarantees that, to solve (5) it suffices to solve (6). Let $v(\lambda,\alpha)$ denote the optimal value of the inner maximization problem of (6). We discuss first how to derive a simplified form for $v(\lambda,\alpha)$. We take an approach that was adopted by Ben-Tal and Teboulle (2007). It critically utilizes the following lemma, which can be found in Ben-Tal and Teboulle (2007), as well as in Rockafellar and Wets (1998).

Lemma 1 Let Ω be a σ-finite measure space, and let $\mathscr{X} := L^p(\Omega, \mathscr{F}, P), p \in [1, +\infty]$. Let $g : \Re \times \Omega \to (-\infty, +\infty]$ be a normal integrand, and define on $\mathscr{X}$ the integral functional $I_g(x) := \int_\Omega g(x(\omega), \omega) dP(\omega)$. Then,

$$\inf_{x \in \mathscr{X}} \int_\Omega g(x(\omega), \omega) dP(\omega) = \int_\Omega \inf_{s \in \Re} g(s, \omega) dP(\omega)$$

provided the left-hand side is finite. Moreover,

$$\bar{x} \in \arg\min_{x \in \mathscr{X}} I_g(x) \Longleftrightarrow \bar{x}(\omega) \in \arg\min_{s \in \Re} g(s, \omega), a.e.\ \omega \in \Omega.$$

Note that the definitions of $\mathscr{X}$ and normal integrand can be found in Rockafellar and Wets (1998). Lemma 1 guarantees that (Setting $p = 1$ and $g = (H + \lambda) L - \sum_{i=1}^m \alpha_i \phi_i(L)$, it can be verified that $L \in \mathscr{X}$ and g is a normal integrand) we can put the supremum into the expectation in the expression of $v(\lambda, \alpha)$. Therefore,

$$v(\lambda, \alpha) = \mathrm{E}_{P_0} \left[\sup_{L \in \mathbb{L}(a,b)} \left\{ (H + \lambda) L - \sum_{i=1}^m \alpha_i \phi_i(L) \right\} \right] + \sum_{i=1}^m \alpha_i \eta_i - \lambda. \tag{7}$$

To simplify $v(\lambda, \alpha)$, we define an auxiliary function

$$\Psi(s, \alpha) = \sup_{t \in \mathbb{L}(a,b)} \left\{ st - \sum_{i=1}^m \alpha_i \phi_i(t) \right\}. \tag{8}$$

It is not difficult to see that $\Psi(s, \alpha)$ is a well defined deterministic function. Moreover, we have the following proposition.

Proposition 1 $\Psi(s, \alpha)$ is convex in (s, α), is non-decreasing in s, and satisfies $\Psi(s, \alpha) \geq s$.

Proposition 1 summarizes important properties of $\Psi(s, \alpha)$. We will frequently refer to this proposition in the analysis following. With the theory built above, it is easy to prove the following theorem, which summarizes the main result on RS of expectation.

Theorem 2 The optimal value of Problem (1) is equal to that of the following problem

$$\underset{\lambda \in \Re, \alpha \geq 0}{\text{minimize}} \ \mathrm{E}_{P_0} [\Psi(H + \lambda, \alpha)] + \sum_{i=1}^m \alpha_i \eta_i - \lambda. \tag{9}$$

Compared to (1), (9) becomes much more specific, as the expectation is now taken w.r.t. an explicit distribution. Actually, (9) is a standard SO problem, see, e.g., Shapiro et al. (2014). Moreover, Proposition 1 guarantees that the problem is a convex problem. Therefore, (9) is much easier to solve than the original functional problem. Next, we discuss potential solution methods for (9).

2.3 Solution Methods

A number of techniques have been developed for SO problems. Among them the sample average approximation (SAA) method and the stochastic approximation (SA) method are widely used, see, e.g., Shapiro et al. (2014). The idea of SAA is to approximate the SO problem by a deterministic sample problem and then implement deterministic optimization techniques to solve the sample problem. The SA method mimics the steepest decent (ascent) method and iteratively updates the solution based on the sample taken at each iteration. Both methods have their advantages/disadvantages and applicability.

The difficulty of Problem (9) depends on the function $\Psi(\cdot, \cdot)$. If the expression of $\Psi(\cdot, \cdot)$ can be derived analytically, then SAA may be directly applied. When a closed form of $\Psi(\cdot, \cdot)$ is unavailable, it becomes difficult to apply SAA, and in such circumstance SA is often a better choice. In this section, we first show that, for the ambiguity set defined by only one divergence, we can usually obtain the closed form for $\Psi(\cdot, \cdot)$. For this class, we use SAA to solve it. For the general case, we suggest a typical SA algorithm.

Table 1: Some ϕ-Divergence Functions and Their Conjugates

Divergence	$\phi(t), t \geq 0$	$\phi^*(s)$
Kullback-Leibler	$t \log t$	e^{s-1}
Burg entropy	$-\log t$	$-1-\log(-s), s \leq 0$
J-divergence	$(t-1)\log t$	No closed form
χ^2-distance	$\frac{(t-1)^2}{t}$	$2-2\sqrt{1-s}, s \leq 1$
Modified χ^2-distance	$(t-1)^2$	$\begin{cases} -1 & s<-2 \\ s+s^2/4 & s \geq -2 \end{cases}$
Hellinger distance	$(\sqrt{t}-1)^2$	$\frac{s}{1-s}, s<1$
χ-distance of order $\theta>1$	$\lvert t-1\rvert^\theta$	$s+(\theta-1)\left(\frac{\lvert s\rvert}{\theta}\right)^{\theta/(\theta-1)}$
Variation distance	$\lvert t-1\rvert$	$\begin{cases} -1 & s<-1 \\ s & -1 \leq s \leq 1 \end{cases}$
Cressie-Read	$\frac{1-\theta+\theta t-t^\theta}{\theta(1-\theta)}, \theta \neq 0,1$	$\frac{1}{\theta}\left(1-s(1-\theta)\right)^{\theta/(\theta-1)}-\frac{1}{\theta} \quad s<\frac{1}{1-\theta}$

2.3.1 Sample Average Approximation

Suppose $\xi_1, \xi_2, \cdots, \xi_N$ are independent sample generated from P_0. SAA suggests using the following sample problem

$$\underset{\lambda \in \Re, \alpha \geq 0}{\text{minimize}} \ \frac{1}{N}\sum_{j=1}^{N}\Psi(H(\xi_j)+\lambda, \alpha)+\alpha\eta-\lambda \tag{10}$$

to approximate (9). The theory of SAA has been studied for a number of years. With some regularity conditions on $\Psi(\cdot,\cdot)$, the optimal solutions and optimal value of (10) will converge to that of (9). We refer the readers to Shapiro et al. (2014) for details.

Note that the implementation of SAA relies on efficient solution methods for the sample problem (10). We illustrate the tractability of (10) for a very important case. Consider the following special case of $\mathbb{L}$:

$$\mathbb{L}_\phi = \left\{L \in \mathbb{L}(0,+\infty) : \mathrm{E}_{P_0}[L]=1, \mathrm{E}_{P_0}[\phi(L)] \leq \eta\right\}.$$

Then $\mathbb{L}_\phi$ contains the distributions whose distance (measure by ϕ) to the nominal distribution P_0 is within a constant η. It should be a most natural choice in practice. Of particular importance of the ϕ-divergence is its conjugate, which is defined as $\phi^*(s)=\sup_{t\geq 0}\{st-\phi(t)\}$. Table 1 extracted from Ben-Tal et al. (2013) summarizes information about various ϕ-divergence measures. In the table the second column includes various ϕ-divergence functions, whereas the third column shows corresponding conjugates. For the ambiguity set $\mathbb{L}_\phi$, the function $\Psi(s,\alpha)=\sup_{t\geq 0}\{st-\alpha\phi(t)\}=\alpha\phi^*\left(\frac{s}{\alpha}\right)$. Then, Problem (9) takes the following form

$$\underset{\lambda \in \Re, \alpha \geq 0}{\text{minimize}} \ \mathrm{E}_{P_0}\left[\alpha\phi^*\left(\frac{H+\lambda}{\alpha}\right)\right]+\alpha\eta-\lambda.$$

From Table 1, we can see that for most of the divergences, the conjugate function ϕ^* has a closed form. With the expression for ϕ^* given, we can now design efficient procedures to solve the deterministic convex sample problem.

2.3.2 Stochastic Approximation

One of the merits of SA, compared to SAA, is that it does not require a closed form for $\Psi(\cdot,\cdot)$. Therefore, when it is difficult to derive the expression of $\Psi(\cdot,\cdot)$, we often resort to SA. There have been numerous SA procedures in the literature. In this paper, we suggest using the robust stochastic approximation (RSA) procedure proposed by Nemirovski et al. (2009) to solve (9). To describe the procedure, we introduce some notions for (9). Let $x=(\lambda,\alpha)$ denote the decision vector. Suppose Θ is a compact set that includes the optimal solution, and $G(x,\xi)$ is the stochastic subgradient of the objective function. Let $\Pi_\Theta(x)$ denote the projection of x onto Θ. Suppose the number of allowed iterations is N. The RSA procedure works as follows (in the form of Ghadimi and Lan (2015)).

Robust Stochastic Approximation (RSA)

Step 0. Let $x_0 \in \Theta$ be given.

Step k. For $k = 0, 1, \cdots, N-1$, generate ξ_k, and set $x_{k+1} = \Pi_\Theta\left(x_k - \gamma_k G(x_k, \xi_k)\right)$ for some $\gamma_k \in (0, +\infty)$.

Output $\bar{x}_N = \frac{\sum_{k=1}^N \gamma_k x_k}{\sum_{k=1}^N \gamma_k}$.

To implement RSA, we need to provide the step size γ_k and the stochastic subgradient $G(x_k, \xi_k)$. Nemirovski et al. (2009) suggested several choices for γ_k given N. Suppose $st - \sum_{i=1}^m \alpha_i \phi_i(t)$ is strictly convex in t (This is usually guaranteed by the strict convexity of ϕ_i). Then there is a unique optimal solution t^* for (8). It follows from Danskin Theorem (Shapiro et al. 2014) that $\Psi(s, \alpha)$ is differentiable and

$$\nabla \Psi(s, \alpha) = \nabla \left\{ st - \sum_{i=1}^m \alpha_i \phi_i(t) \right\} \Bigg|_{t=t^*}.$$

The stochastic subgradient $G(x, \xi)$ then becomes a stochastic gradient and can be computed accordingly. For further properties (e.g., convergence) and real implementations of RSA, we refer the readers to Nemirovski et al. (2009).

3 PROBABILITY PERFORMANCE MEASURE

We next consider the probability performance measure. Suppose $A(\xi)$ is the random event of concern. RS considers the following quantities,

$$P_l := \inf_{P \in \mathbb{P}} \Pr_P\{A(\xi)\} \quad \text{and} \quad P_u := \sup_{P \in \mathbb{P}} \Pr_P\{A(\xi)\}.$$

Similarly, $[P_l, P_u]$ provides important information for simulation practitioners. Let $A^c(\xi)$ denote the compliment of the event $A(\xi)$, i.e., $A^c(\xi)$ occurs if and only if $A(\xi)$ does not occur. Then

$$\inf_{P \in \mathbb{P}} \Pr_P\{A(\xi)\} = \inf_{P \in \mathbb{P}} 1 - \Pr_P\{A^c(\xi)\} = 1 - \sup_{P \in \mathbb{P}} \Pr_P\{A^c(\xi)\}. \tag{11}$$

The relation suggests it suffices to consider either the supremum or the infimum. In what follows we focus on P_u. Let $\mathbb{1}_{\{A(\xi)\}}$ denote the indicator function, which equals 1 if $A(\xi)$ happens and 0 otherwise. Then $\Pr_P\{A(\xi)\}$ can be rewritten as $\mathrm{E}_P\left[\mathbb{1}_{\{A(\xi)\}}\right]$. For simplicity of notation, we abbreviate $\mathbb{1}_{\{A(\xi)\}}$ by $\mathbb{1}$. Then P_u corresponds to the following optimization problem

$$\underset{P \in \mathbb{P}}{\text{maximize}} \ \mathrm{E}_P[\mathbb{1}]. \tag{12}$$

The reformulation shows that RS of probability can be placed within the framework of RS of expectation, which allows us to implement the results developed in preceding section to handle the RS problem.

3.1 Robust Simulation of Probability

In this section we tailor the functional optimization approach in Section 2 to solve (12). To simplify the notation, we let $\kappa = \Pr_{P_0}\{A(\xi)\}$. Note that κ is the probability of $A(\xi)$ under the nominal distribution P_0. Although κ is typically unknown, we can estimate (simulate) it in a conventional way. Thus we take it as given. Set $H(\xi) = \mathbb{1}$ in (1). Then $H(\xi)$ is naturally bounded by 1. It follows from (7) and (8) that

$$v(\lambda, \alpha) = \mathrm{E}_{P_0}[\Psi(\mathbb{1} + \lambda, \alpha)] + \sum_{i=1}^m \alpha_i \eta_i - \lambda = \Psi(1 + \lambda, \alpha)\kappa + \Psi(\lambda, \alpha)(1 - \kappa) + \sum_{i=1}^m \alpha_i \eta_i - \lambda, \tag{13}$$

where the second equality follows from the definition of the random variable $\mathbb{1}$. From Theorem 2, we obtain the following result on the maximal probability.

Theorem 3 Suppose that the ambiguity set is $\mathbb{L}$. Then P_u equals $\inf_{\lambda\in\Re,\alpha\geq 0}\ v(\lambda,\alpha)$ where $v(\lambda,\alpha)$ is given by (13).

Theorem 3 builds that the maximal probability is equal to the optimal value of an optimization problem with real decision variables. It clearly shows estimating the maximum of the probability can be accomplished by estimating the probability under the nominal distribution and by solving a simple optimization problem. Now we discuss in detail how to solve the problem. To unify the analysis, we define for each $y\in[0,1]$,

$$Z(\lambda,\alpha,y):=y\Psi(1+\lambda,\alpha)+(1-y)\Psi(\lambda,\alpha)+\sum_{i=1}^{m}\alpha_i\eta_i-\lambda. \tag{14}$$

Construct the following problem

$$\underset{\lambda\in\Re,\alpha\geq 0}{\text{minimize}}\ Z(\lambda,\alpha,y), \tag{15}$$

and denote its optimal value by $v^*(y)$. Then clearly, $P_u=v^*(\kappa)$. Suppose we have computed the value of κ. We only need to solve (15) for $y=\kappa$. It follows from Proposition 1 that for any given $y\in[0,1]$, $Z(\lambda,\alpha,y)$ is convex in (λ,α), and thus (15) is a convex optimization problem. Because the function $\Psi(s,\alpha)$ is itself defined by a supremum, we obtain the dual of (8) and build the corresponding strong duality. This results in (using similar techniques in Corollary 4 of Ben-Tal et al. (2013)) the equivalent reformulation of (15):

$$\begin{aligned}\text{minimize}\quad & y\left[\sum_{i=1}^{m}\alpha_i\phi_i^*\left(\frac{s_i}{\alpha_i}\right)-a\mu_1-b\mu_2\right]+(1-y)\left[\sum_{i=1}^{m}\alpha_i\phi_i^*\left(\frac{t_i}{\alpha_i}\right)-a\nu_1-b\nu_2\right]+\sum_{i=1}^{m}\alpha_i\eta_i-\lambda\\ \text{subject to}\quad & \sum_{i=1}^{m}s_i-\mu_1-\mu_2=1+\lambda,\sum_{i=1}^{m}t_i-\nu_1-\nu_2=\lambda,\\ & \lambda\in\Re,\alpha\geq 0,\mu_1\geq 0,\mu_2\leq 0,\nu_1\geq 0,\nu_2\leq 0,s_i\in\Re,t_i\in\Re,i=1,2,\cdots,m.\end{aligned}$$

For each y given, the problem above is a convex optimization problem. With the conjugates ϕ_i^* given explicitly, it can be solved readily.

As mentioned, (11) shows we can use a similar procedure to compute P_l. The procedure requires the input value $\Pr_{P_0}\{A^c(\xi)\}$. Note that $\Pr_{P_0}\{A^c(\xi)\}=1-\kappa$. We have $\sup_{P\in\mathbb{P}}\ \Pr_P\{A^c(\xi)\}=v^*(1-\kappa)$. Therefore, we can also compute P_l by only simulating κ (This is done for P_u). We summarize the result in the following theorem.

Theorem 4 Suppose that the ambiguity set is $\mathbb{L}$. Then $P_l=1-v^*(1-\kappa)$.

Theorems 3 and 4 show that P_u and P_l are determined by κ, the probability of $A(\xi)$ under the nominal distribution. However, as mentioned, in real applications κ also needs estimation. To quantify the estimation error, we usually construct a confidence interval for κ. One may concern that if the estimator of κ contains statistical error, the optimization procedure (15) may significantly amplify the error, leading to unreliable results for RS estimators. Fortunately, we have the following proposition.

Proposition 2 Suppose that $[\hat{\kappa}_l,\hat{\kappa}_u]$ is a $1-\gamma$ confidence interval for κ, i.e., $\Pr\{\hat{\kappa}_l\leq\kappa\leq\hat{\kappa}_u\}\geq 1-\gamma$. Then $[v^*(\hat{\kappa}_l),v^*(\hat{\kappa}_u)]$ is a $1-\gamma$ confidence interval for P_u, and $[1-v^*(1-\hat{\kappa}_l),1-v^*(1-\hat{\kappa}_u)]$ is a $1-\gamma$ confidence interval for P_l.

Proposition 2 shows that, once the confidence interval of the nominal probability is ready, we can easily obtain confidence intervals for the maximal probability and minimal probability accordingly, with the same confidence level.

4 VALUE-AT-RISK PERFORMANCE MEASURE

We now turn our discussion to VaR. The distribution mis-specification issue for VaR has received much attention in real applications. In insurance and actuarial literature, people have suggested considering the worst-case VaR and using it to guide the reservation of capital, to protect against both risks and uncertainties, see, for instance, Wang et al. (2013) and references therein. Because VaR and probability are inherently related, in our framework the results derived for RS of probability may be extended to VaR.

4.1 Robust Simulation of Value-at-Risk

Suppose $H(\xi)$ is the random loss. Because VaR is a risk measure which is used to quantify the undesired risk, decision makers usually only concern the worst case, i.e., the maximal VaR. But for completeness, we still consider the two quantities, the minimal VaR and the maximal VaR

$$V_l := \inf_{P\in\mathbb{P}} \mathrm{VaR}_{1-\beta,P}(H(\xi)) \quad \text{and} \quad V_u := \sup_{P\in\mathbb{P}} \mathrm{VaR}_{1-\beta,P}(H(\xi)).$$

By relating VaR to the probability performance measure in Section 3, we have the following theorem.

Theorem 5 Suppose that the ambiguity set is $\mathbb{L}$. Then $V_u = \mathrm{VaR}_{1-\beta_l,P_0}(H(\xi))$ and $V_l = \mathrm{VaR}_{\beta_u,P_0}(H(\xi))$, where

$$\beta_l = \sup_{\lambda\in\Re,\alpha\geq 0} \frac{\beta-(\Psi(\lambda,\alpha)+\sum_{i=1}^m \alpha_i\eta_i-\lambda)}{\Psi(1+\lambda,\alpha)-\Psi(\lambda,\alpha)} \tag{16}$$

and

$$\beta_u = \sup_{\lambda\in\Re,\alpha\geq 0} \frac{1-\beta-(\Psi(\lambda,\alpha)+\sum_{i=1}^m \alpha_i\eta_i-\lambda)}{\Psi(1+\lambda,\alpha)-\Psi(\lambda,\alpha)}. \tag{17}$$

Theorem 5 shows that the maximal VaR and minimal VaR are equal to some pure VaR under the nominal distribution with only the confidence level being adjusted from the original one. It is not difficult to verify that $\beta_l \leq \beta$ and $\beta_u \leq 1-\beta$. Theorem 5 shows in VaR, risk and ambiguity, the two sides of a coin, are indeed interrelated.

4.2 Computation of New Confidence Level

To simulate the $1-\beta_l$ VaR and β_u VaR, we still need to derive the new confidence levels β_l and β_u, which are defined by (16) and (17) respectively. Because the corresponding optimization problems are typically non-convex, it may be difficult to obtain β_l and β_u by directly solving (16) and (17). To this end, we go back to the definition of $v(\lambda,\alpha)$, and show that β_l and β_u can be obtained via solving a sequence of convex optimization problems. Because the two quantities are defined by the same structure, we only consider β_l.

Note that $\beta_l \in [0,1]$. Therefore, we only need to seek β_l from $[0,1]$ and β_l is equal to the optimal value of the following optimization problem:

$$\begin{aligned} &\underset{0\leq y\leq 1,\lambda\in\Re,\alpha\geq 0}{\text{maximize}} && y \\ &\text{subject to} && y \leq \frac{\beta-(\Psi(\lambda,\alpha)+\sum_{i=1}^m \alpha_i\eta_i-\lambda)}{\Psi(1+\lambda,\alpha)-\Psi(\lambda,\alpha)}, \end{aligned}$$

which can be reformulated as

$$\underset{0\leq y\leq 1,\lambda\in\Re,\alpha\geq 0}{\text{maximize}} \quad y \quad \text{subject to} \quad Z(\lambda,\alpha,y) \leq \beta, \tag{18}$$

where Z is defined by (14). From preceding section, we have for any given $y\in[0,1]$, $Z(\lambda,\alpha,y)$ is convex in (λ,α). Furthermore, from Proposition 1, it is not difficult to show that $Z(\lambda,\alpha,y)$ is nondecreasing in y. The nice structures allow for the following Bisection Search procedure to solve Problem (18).

Bisection Search
Step 0. Set $i=0$. Set $y_l := 0$ and $y_u := 1$

Step i. Set $y_i = \frac{y_l + y_u}{2}$ and solve the following problem to obtain its optimal value v:

$$\underset{\lambda \in \Re, \alpha \geq 0}{\text{minimize}} \quad Z(\lambda, \alpha, y_i).$$

If $v \leq \beta$, update $y_l =: y_i$, otherwise, update $y_u =: y_i$. Set $i = i + 1$.

It is not difficult to verify that the sequence $\{y_i\}$ generated by the Bisection Search procedure converges to the optimal value of Problem (18), i.e., β_l, and the convergence rate is in an exponential order. To implement the Bisection Search procedure, we need to solve a sequence of convex optimization problems in Step i. Because the problem in Step i is an instance of Problem (15), it may be solved readily.

5 AN EMERGENCY MEDICAL SERVICE MODEL

We consider a toy model on emergency medical service (EMS), to illustrate the computation of RS, and also to conclude the paper. Suppose the emergency call may occur at any point $\xi = (\xi_1, \xi_2)$ of a geographic region (the whole plane, measured by km) with a joint normal distribution $P := N(\mu, \Sigma)$. Suppose there are five EMS bases located at $O(0,0), A(12,0), B(0,12), C(-12,0)$, and $D(0,-12)$. Once a call arrives, an ambulance in the nearest base will be set to serve it. Let $(\zeta_1(\xi), \zeta_2(\xi))$ denote the location of the response base for ξ. For simplicity, assume the speed of any ambulance is a constant $v = 40$km/h. Then, the response time for ξ is $H(\xi) = v^{-1}\sqrt{(\xi_1 - \zeta_1(\xi))^2 + (\xi_2 - \zeta_2(\xi))^2}$. The percentage of late calls for EMS is often studied in healthcare management practice. A call is taken to be late if the response time exceeds a threshold, see, e.g., Maxwell et al. (2014). In this example, we consider the nine minutes threshold. We are interested in the late percentage, i.e., $\Pr_P\{H(\xi) > 9/60\}$. While it is not impossible to derive an analytical expression for the probability, we estimate it via simulation.

Assume a nominal estimate for P is $P_0 = N(\mathbf{0}, 10 \times \mathbf{I})$ where $\mathbf{0}$ is the zero vector and $\mathbf{I}$ is the identity matrix. The nominal late percentage under P_0 is estimated to be 0.0912. Suppose the true joint normal distribution P falls within a neighborhood $\mathbb{L}_\phi$ of P_0 where we assume ϕ is the χ^2-distance. We conduct RS of the probability measure with ambiguity set $\mathbb{L}_\phi$, using the approach developed in Section 3, and report the computational results for different values of η in Table 2.

Table 2: RS for Probability

η	P_l	P_u	η	P_l	P_u
1.0000	0.0071	0.5841	0.0100	0.0663	0.1242
0.1000	0.0339	0.2228	0.0010	0.0825	0.1007

From the table, we can see that the late percentage is quite sensitive to the input distribution, reflecting the importance of the input uncertainty issue. On the other hand, when the index of ambiguity η reduces, the maximal probability and minimal probability become closer and closer to the nominal value, supporting that the uncertainty in output may be suppressed by reducing the input uncertainty.

REFERENCES

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. “Coherent Measures of Risk”. *Mathematical Finance*, 9(3): 203-228.

Barton, R. R. 2012. “Tutorial: Input Uncertainty in Output Analysis”. In *Proceedings of the 2003 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 67-78. IEEE Computer Society, Washington, DC.

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.

Ben-Tal, A., D. Bertsimas, and D. Brown. 2010. “A Soft Robust Model for Optimization under Ambiguity”. *Operations Research*, 58(4): 1220-1234.

Ben-Tal, A., D. den Hertog, A. M. B. de Waegenaere, B. Melenberg, and G. Rennen. 2013. "Robust Solutions of Optimization Problems Affected by Uncertain Probabilities". *Management Science*, 59(2): 341-357.

Ben-Tal, A., and M. Teboulle. 2007. "An Old-new Concept of Convex Risk Measures: The Optimized Certainty Equivalent". *Mathematical Finance*, 17(3): 449-476.

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

Ellsberg, D. 1961. "Risk, Ambiguity, and the Savage Axioms". *The Quarterly Journal of Economics*, 75: 643-669.

Epstein, L. G. 1999. "A Definition of Uncertainty Aversion". *Review of Economic Studies*, 66: 579-608.

Fan, W. W., L. J. Hong, and X. W. Zhang. 2013. "Robust Selection of the Best". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 868-876. IEEE Press Piscataway, NJ.

Ghadimi, S., and G. Lan. 2015. "Stochastic Approximation Methods and Their Finite-time Convergence Properties". Chapter 7, *Handbook of Simulation Optimization*, Fu, M. C. Editor, Springer.

Hansen, L. P., and T. J. Sargent. 2008. *Robustness*. Princeton University Press.

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*, edited by S. Chick, P. J. Sanchez, D. Ferrin, and D. J. Morrice, 90-100. IEEE, Piscataway NJ.

Hong, L. J., Z. Hu., and G. Liu. 2014. "Monte Carlo Methods for Value-at-risk and Conditional Value-at-risk: A Review". *ACM Transactions on Modeling and Computer Simulation*, 24(4): Article 22.

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.

Jorion, P. 2006. *Value at Risk*, Second Edition. New York: McGraw-Hill, Inc.

Maxwell, M. S., E. C. Ni, C. Tong, S. R. Hunter, S. G. Henderson, and H. Topaloglu. 2014. "A Bound on the Performance of An Optimal Ambulance Redeployment Policy". *Operations Research*, 62(5): 1014-1027.

Nemirovski, A., A. Juditsky, G. Lan, and A. Shapiro. 2009. "Robust Stochastic Approximation Approach to Stochastic Programming". *SIAM Journal on Optimization*, 19(4): 1574-1609.

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

Pardo, L. 2006. *Statistical Inference Based on Divergence Measures*, Chapman & Hall/CRC, Florida.

Rockafellar, R. T., and R. J.-B. Wets. 1998. *Variational Analysis*, Springer-Verlag, New York.

Shapiro, A., D. Dentcheva, and A. Ruszczyński. 2014. *Lectures on Stochastic Programming: Modeling and Theory*, Second Edition. SIAM, Philadelphia.

Wang, R., L. Peng, and J. Yang. 2013. "Bounds for the Sum of Dependent Risks and Worst Value-at-Risk with Monotone Marginal Densities". *Finance and Stochastics*, 17(2): 395-417.

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.

AUTHOR BIOGRAPHIES

ZHAOLIN HU is an Associate Professor at the School of Economics and Management of Tongji University. His current research interests include simulation theory and practice, stochastic optimization, and risk management.

L. JEFF HONG is a Chair Professor of Management Sciences in the College of Business at the City University of Hong Kong. His research interests include stochastic simulation, stochastic optimization, and financial risk management. He is currently an associate editor of *Operations Research*, *Naval Research Logistics* and *ACM Transactions on Modeling and Computer Simulation*.