Robust simulation optimization using φ-divergence

Samira Moghaddam[*] and Hashem Mahlooji

Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran

CHRONICLE	ABSTRACT
	We introduce a new robust simulation optimization method in which the probability of occurrence of uncertain parameters is considered. It is assumed that the probability distributions are unknown but historical data are on hand and using φ-divergence functionality the uncertainty region for the uncertain probability vector is defined. We propose two approaches to formulate the robust counterpart problem for the objective function estimated by Kriging. The first method is a minimax problem and the second method is based on the chance constraint definition. To illustrate the methods and assess their performance, numerical experiments are conducted. Results show that the second method obtains better robust solutions with less simulation runs.

1. Introduction

Simulation is a powerful and flexible tool to model problems in which either the closed-form of the objective function or the constraints is not in hand. Examples of such problems are car crash model or aerodynamic wing design. For these problems, by applying optimization tools, the optimal combination of input variables that optimize the simulation output could be found. The area, consisting of the techniques that combine the simulation modeling and optimization tools for solving complicated optimization problems is commonly referred to as simulation optimization. Simulation optimization methods are generally classified into two categories, model-based and metamodel-based approaches (Nezhad & Mahlooji, 2013). In a model-based approach, simulation and optimization components interact with each other and outputs of one are inputs to the other (for more details about model-based approaches one can refer to Bhatnagar et al. (2011), Chang (2012), Kim and Nelson (2007) and Lin et al. (2013). In a metamodel-based approach, there exists a third component called metamodel. Metamodel identifies and estimates the relationship between inputs and outputs of the simulation model in the form of an explicit deterministic function. For expensive simulation models, metamodels such as response-surface methodology, Kriging, radial basis functions, splines or neural networks are viable alternatives to alleviate the magnitude of the computational cost. For more details about metamodel approaches one

* Corresponding author. Tel.: +98 21 6616-5717; Fax:+98 21 6602-2702
E-mail: s_moghaddam@ie.sharif.edu (S. Moghaddam)

can refer to Hurrion (1997), Kleijnen (2015), Liu and Maghsoodloo (2011), Simpson et al. (2001) and Yang (2010). In practice, most of the simulation models are uncertain, i.e., by running the simulation model for the same inputs different outputs are generated. For these models the underlying metamodel would be uncertain too. Usually in the literature, deterministic metamodels are fitted to the combination of the inputs and mean of the simulation output (e.g. Ajdari & Mahlooji, 2014; Kleijnen, 2014; Kleijnen et al., 2010; Quan et al., 2013). If we do not consider uncertainty in metamodel parameters, we may lead to non-optimal solutions.

To deal with this uncertainty, different methods have been proposed in the area of robust simulation optimization. For instance, Myers and Montgomery (2003) assumed that the random (noise) variable ε has $\mu(\varepsilon) = 0$ and $Cov(\varepsilon) = \sigma_\varepsilon^2$, then the mean and variance of the response variable are estimated theoretically. Some methods use metamodels to estimate the mean and the variance of the response variable, for example, in (Dellino et al., 2009, 2010a, 2010b, 2012; Dellino & Meloni, 2015; Kovach & Cho, 2009; Lee et al., 2006) two Kriging metamodels were fitted, one for the mean and one for the variance of the response. To formulate the robust optimization problem, they consider minimizing the mean while satisfying a constraint on the standard deviation (or vice versa).

Few research efforts have been reported so far which employ robust optimization techniques in simulation (for a thorough review on robust optimization techniques one can refer to Bertsimas et al. (2011) and Gabrel et al. (2014). For example, in (Zhou & Zhang, 2010) an interesting approach was presented in which a metamodel and an evolutionary optimization algorithm were combined. In (Stinstra & Den Hertog, 2008) by considering three types of error: simulation error (the difference between reality and computer model), metamodel error (the difference between computer model and metamodel), and implementation error the authors represented the robust counterpart problem based on the Ben-Tal and Nemirovski theory for the regression and Kriging metamodels. In (Marzat et al., 2013) an iterative relaxation procedure was combined with the Kriging-based optimization. In (Bertsimas et al., 2010) a robust optimization method was presented that is suitable for unconstrained problems with a nonconvex cost function as well as for problems based on simulations (see also Cramer et al. 2009; Lung & Dumitrescu, 2011).

The main criticism for the robust optimization methods is that they do not consider the occurrence probabilities of the uncertain parameters and just find a solution which is feasible for all the values in the uncertainty set, so the final solution would be an over conservative solution. A remedy for this is to consider the probability of occurrence of uncertain parameters in a problem. However, in practice, the probability distributions of parameters, *P*, are not known and just historical data may be available. This means that the distributions of the uncertain parameters are uncertain themselves. Considering the uncertainty in the probability distribution, different robust methods have been proposed in the literature. One method is to optimize the objective function based on the worst-case probability distribution (for more details of this method, one can refer to Birge and Wets, (1986), Shapiro and Ahmed (2004) and Shapiro and Kleywegt (2002). Another method aims to construct an uncertainty region for the probability vectors estimated from historical data, in a way that bears a specific confidence level. There are different statistics in the literature for constructing confidence levels for probability vectors. Here we use φ-divergence as it generates a tractable robust counterpart formulation that makes it an interesting method in robust optimization (Ben-Tal et al., 2013).

Works related to φ-divergence in the literature are as follows. Klabjan et al. (2013) employed a special case of φ-divergence to construct uncertainty regions from historical data for the stochastic lot-sizing problem using a $\chi 2$-statistic. Calafiore (2007) formulated the robust portfolio selection problem by considering Kullback-Leibler divergence which aims at finding robust solutions that are feasible for all allowable distributions of the uncertain parameters with bounded supports. Wang et al. (2009) proposed a robust optimization method for newsvendor problem in which the uncertainty set for the unknown distribution is defined as a "likelihood region". Ben-Tal et al. (2013) proposed a method based on φ-

divergence for optimization problems in which the uncertain parameters are probabilities. This is the case when the objective function and/or the constraint functions involve terms with expectations (such as moments of random variables). They applied their method to an investment problem as well as multi-item newsvendor problem. Their results show that uncertainty sets based on φ-divergence were good alternatives for the uncertainty sets such as ellipsoidal, box, and their variations. Yanikoglu and den Hertog (2012) used φ-divergence for safe approximation of chance constraints where the family of distributions, P, is given by a confidence set that is based on historical data. They show that φ-divergence leads to tighter uncertainty sets compared with other existing safe approximation methods.

In this paper, we employ φ-divergence in the simulation optimization context to find robust solutions for problems in which the closed form of the objective function is not in hand. We fit Kriging to the I/O combination of simulation model to estimate the objective function and apply the φ-divergence to the estimated objective function. We assume that the probability distributions of the uncertain parameters are not known but historical data are available, so the probability vector is the uncertain parameter. We propose two methods to employ φ-divergence in robust simulation optimization context .The first method is a minimax problem, based on the work of Ben-Tal et al. (2013). The second one is a chance constraint definition based on the work of Yanikoglu and den Hertog (2012). Our method is different from other robust simulation optimization methods in:

- Using the probabilities of occurrence of uncertain parameters (random variables), that leads to tighter uncertainty sets and less conservative solutions,
- Employing a distribution-free structure in robust optimization, which does not need any assumptions about the probability distribution of the random variables,
- Proposing a method that can be applied to the constrained simulation optimization problems and consider robustness in both optimality and feasibility of a problem,
- Employing a new method for taking samples from the random and decision variables that leads to fewer simulation runs in metamodel construction, which is considered an advantage especially for expensive simulation models).

The remainder of this paper is organized as follows. Section 2 elaborates on Kriging metamodel. In section 3, the φ-divergence concept as well as the proposed robust simulation optimization methods are explained in detail. In section 4, the proposed methods are applied to the well-known EOQ problem, the multi-item newsvendor problem, and seven test functions and results are compared with each other and with those reported by Dellino et al. (2012). Section 5 contains the conclusion and ideas for future research.

2. Kriging metamodel

Kriging gives predictions of unknown values of a random function as the weighted linear combination of the observed values. Kriging has been often applied to deterministic simulation models in engineering. Recently, Kriging has been extended to stochastic simulation models (Ankenman et al., 2010). Kriging assumes the following model:

$$Y(\boldsymbol{x}) = f(\boldsymbol{x})^T \beta + M(\boldsymbol{x}) \quad \text{with} \quad M(\boldsymbol{x}){\sim}NID(0, \sigma^2(\boldsymbol{x})), \tag{1}$$

where $\boldsymbol{x}$ is the vector of design points, $f(\boldsymbol{x})^T\beta$ is the trend model, and $M(\boldsymbol{x})$ is a realization of mean 0 random field which is assumed to exhibit spatial correlation. This means that values $M(\boldsymbol{x})$ and $M(\boldsymbol{x}')$ will tend to be similar if $\boldsymbol{x}$ and $\boldsymbol{x}'$ are close to each other in space.
Ordinary Kriging assumes a stationary covariance process for $Y(\boldsymbol{x})$. This assumption implies that the expected values $E(Y(\boldsymbol{x}))$ are constant (so $f(\boldsymbol{x})^T\beta$ can be replaced by β), and the covariance of $Y(x_i +$

h) and $Y(x_i)$ depends only on the distance (or lag) h. The predictor for the design point, x_0, denoted by $\hat{Y}(x_0)$, is a weighted linear combination of all the (say) n observed output data, w_i:

$$\hat{Y}(x_0) = \sum_{i=1}^{n} \lambda_i w_i \quad \text{with} \quad \sum_{i=1}^{n} \lambda_i = 1 \tag{2}$$

To select the weights, λ_i, in Eq. (2), Kriging minimizes the mean-squared prediction error. Let the $n \times 1$ vector $(Cov[M(x_0), M(x_1)], ..., Cov[M(x_0), M(x_n)])$ be denoted by γ, the $n \times n$ covariance matrix across all design points $x_1, x_2, ..., x_n$ be denoted by Σ_M, and 1 denote a vector of ones. Then the optimal weights turn out to be:

$$\hat{\lambda} = \left(\gamma + 1 \frac{1 - 1^T \Sigma_M \gamma}{1^T \Sigma_M 1} \right)^T \Sigma_M^{-1} . \tag{3}$$

Assuming that M is second-order stationary, leads to the following:

$$\Sigma_M(x, x') = \tau^2 R_M(x - x'; \theta), \tag{4}$$

where τ^2 can be interpreted as the variance of $M(x)$ for all x, and R_M is the correlation that depends only on $x - x'$ and is a function of an unknown parameter θ. We require that $R_M(x - x'; \theta) \to 0$ as the distance between x and x' goes to infinity, and $R_M(0, \theta) = 1$ (for a thorough review on Kriging one can refer to (Kleijnen, 2009)).

3. Robust counterpart with φ-divergence uncertainty

In this section, we first explain the concept of φ-divergence and some useful properties in subsection 3.1. Then, we will discuss the two proposed methods based on φ-divergence in subsections 3.2 and 3.3, respectively.

3.1. Introduction to φ-divergence

Given $\phi(t)$ is a convex function for $t \geq 0$ satisfying $\phi(1) = 0$, $\phi'(1) = 0$, $0\phi(a/0) := a \lim_{t \to \infty} \phi(t)/t$, for $a > 0$, and $0\phi(0/0) := 0$, the φ-divergence (distance) between two vectors $p = (p_1, ..., p_m) \geq 0$ and $q = (q_1, ..., q_m) \geq 0$ is defined as:

$$I_\phi(p, q) = \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right). \tag{5}$$

We consider p to be the unknown true probability vector and q the empirical estimate of p. There are different types of ϕ function in the literature. For a good review one can refer to Pardo (2005). Table 1 taken from (Ben-Tal et al., 2013) presents the most common choices of $\phi(.)$ together with the conjugate function that is defined as follows:

$$\phi^*(s) = sup_{t \geq 0}\{st - \phi(t)\}. \tag{6}$$

If we assume ϕ is twice continuously differentiable in the neighborhood of 1 and $\phi''(1) > 0$, then the test statistic

$$\frac{2N}{\phi''(1)} I_\phi(p, q) \tag{7}$$

follows a $\chi 2$-distribution with $(m-1)$ degrees of freedom (N is the number of historical observations). Using this test statistic, an approximate $(1 - \alpha)$-confidence set for p defines the family of distributions P:

$$P = \{\boldsymbol{p} \in \mathbb{R}^m | \boldsymbol{p} \geq 0, \ \boldsymbol{e}^T \boldsymbol{p} = 1, \qquad I_\phi(\boldsymbol{p}, \boldsymbol{q}) \leq \rho\}, \tag{8}$$

where $\boldsymbol{e}$ denotes a column vector of ones with the same dimension as $\boldsymbol{p}$ and

$$\rho = \rho_\phi(N, m, \alpha) = \frac{\phi''(1)}{2N} \chi^2_{m-1,1-\alpha} \tag{9}$$

Table 1
ϕ-Divergence Examples

Divergence	$\phi(t), t > 0$	$I_\phi(\boldsymbol{p}, \boldsymbol{q})$	$\phi^*(s)$
Kullback-Leibler	$t \log t$	$\sum_i p_i \log(\frac{p_i}{q_i})$	e^{s-1}
Burg entropy	$-\log t$	$\sum_i q_i \log(\frac{p_i}{q_i})$	$-1 - \log(s), s \leq 0$
χ2-distance	$\frac{1}{t}(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{p_i}$	$2 - 2\sqrt{1-s}, s \leq 1$
Modified χ2-distance	$(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{q_i}$	$s + s^2/4, s \geq -2$ $-1, s < -2$
Hellinger distance	$(1 - \sqrt{t})^2$	$\sum_i (\sqrt{p_i} - \sqrt{q_i})^2$	$\frac{s}{1-s}, s \leq 1$

3.2. Constructing the robust counterpart by minimax method

We assume that the probability distributions of random variables are not known and should be estimated based on historical data. We also assume that there is a fixed number, *m*, of given scenarios for random variables and in the case of continuous variables we divide the valid range of random variables to *m* parts and find the center of each range, $\boldsymbol{\varepsilon}_j$, for *j*=1,2,..,*m*. Using historical data we estimate the probability of occurrence of $\boldsymbol{\varepsilon}_j$. We also find *n* sampling points, $\boldsymbol{x}_i$, for *i*=1,2,…,*n,* from solution space using a sampling design like Latin Hypercube Sampling (LHS). Then, we run the simulation model for every combination of sampling points, $\boldsymbol{x}_i$, for *i*=1,2,…,*n* and random variables, $\boldsymbol{\varepsilon}_j$, for *j*=1,2,..,*m* and fit one Kriging metamodel for every scenario of uncertain parameters which amounts to *m* metamodels. The final objective function would be the weighted sum of all the *m* metamodels in which weights are the estimated probability values for the corresponding scenarios. Such an outlook toward the objective function paves the way to use functions like *φ*-divergence in simulation optimization context, which is our main contribution. Using *φ*-divergence leads to robust solutions, which are less conservative than the solutions obtained by common robust optimization methods. We derive the robust reformulation of the optimization problem using Ben-Tal method (Ben-Tal et al., 2013). We describe the steps of the proposed method in more detail below.

Step 0. Determining valid ranges for the random and decision variables: We first determine the random variables (e.g., demand and lead-time are random variables in an inventory problem) and their valid ranges. We assume that the probability distributions of the random variables are not known. Given *N* historical observations on the random variables, $\boldsymbol{\varepsilon}$, we partition the range of them into *m* cells in a way

that the number of observations is at least five in every cell. Then we calculate the frequency for cell j using the number of observations, o_j, as an estimate of the occurrence probability in every part as follows:

$$\hat{p}_j = q_j = \frac{o_j}{N} \quad \text{for } j = 1, 2, \dots, m \tag{10}$$

where p_j denotes the probability of $\boldsymbol{\varepsilon}$ falling into cell j and $\hat{p}_j = q_j$ denotes the estimate of p_j.

We should also determine the decision variables (for example economic order quantity is the decision variable in an inventory problem) and the valid ranges for them. We sample n points from the solution space by a uniform sampling design like LHS to form the simulation points set.

Step 1. Constructing the objective function: For every combination of design points, x_i, for $i=1,2,\dots,n$ and random variables, $\boldsymbol{\varepsilon}_j$, for $j=1,2,\dots,m$ we run the simulation model one time and show the simulation outputs by w_{ij}. To estimate the objective function, Kriging metamodels are fitted to the I/O combination of simulation model for every random variable, $\boldsymbol{\varepsilon}_j$, for $j=1,2,\dots,m$. The resulting metamodels would be of the form:

$$Y_j(\boldsymbol{x}_0) = \sum_{i=1}^{n} w_{ij} \lambda_{ij}(\boldsymbol{x}_0) \text{ for } j = 1, 2, \dots, m \tag{11}$$

We define the final objective function as the weighted sum of m fitted metamodels, whose weights are equal to the probability of occurrence of random variables. Therefore, the final model would be as follows:

$$\hat{Y}(\boldsymbol{x}_0) = \sum_{j=1}^{m} \hat{p}_j \sum_{i=1}^{n} \lambda_{ij}(\boldsymbol{x}_0) w_{ij} \quad \boldsymbol{p} \in U \tag{12}$$

where:

$$U = \left\{ \boldsymbol{p} \mid \boldsymbol{p} \geq 0, \ \boldsymbol{e}^T \boldsymbol{p} = 1, \quad I_\varphi(\boldsymbol{p}, \boldsymbol{q}) \leq \rho \right\} \tag{13}$$

The nominal mathematical programming model is as follows,

$$min_{\boldsymbol{x} \in X} \sum_{j=1}^{m} \hat{p}_j \sum_{i=1}^{n} \lambda_{ij}(\boldsymbol{x}) w_{ij} \tag{14}$$

where X is the feasible set given by deterministic constraints.

Step 2. Formulating the robust counterpart problem: We write the robust counterpart for Eq. (14) as:

$$min_{\boldsymbol{x} \in X} max_{\boldsymbol{p} \in U} \sum_{j=1}^{m} \hat{p}_j \sum_{i=1}^{n} \lambda_{ij}(\boldsymbol{x}) w_{ij} \tag{15}$$

We can rewrite Eq. (15) as the following mathematical programming problem:

$$min_{\boldsymbol{x} \in X} z$$
$$\text{s.t. } \sum_{j=1}^{m} \hat{p}_j \sum_{i=1}^{n} \lambda_{ij}(\boldsymbol{x}) w_{ij} \leq z \ \forall \boldsymbol{p} \in U \tag{16}$$

Step 3. Solving the counterpart problem: The mathematical problem in Eq. (16) is a difficult optimization problem that has infinitely many constraints ($\forall \boldsymbol{p} \in U$), and includes quadratic terms in $\boldsymbol{p}$.

Ben-Tal et al. (Ben-Tal et al., 2013) proposed a tractable semidefinite robust counterpart of an optimization problem where the uncertain parameters are probabilities. Let $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{B} \in \mathbb{R}^{n \times m}$, and $b \in \mathbb{R}$ as given parameters, $\boldsymbol{p} \in \mathbb{R}^m$ as the vector of uncertain parameters, and $f(\boldsymbol{x})$ as a function (linear or nonlinear) in $\boldsymbol{x}$, the following robust constraint is considered:

$$(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{p})^T f(\boldsymbol{x}) \le b \quad \forall \boldsymbol{p} \in U \tag{17}$$

With uncertainty region U given by:

$$U = \left\{ \boldsymbol{p} \in \mathbb{R}^m \mid \boldsymbol{p} \ge 0, \boldsymbol{C}\boldsymbol{p} \le d, I_\varphi(\boldsymbol{p}, \boldsymbol{q}) \le \rho \right\}, \tag{18}$$

where $\boldsymbol{C} \in \mathbb{R}^m$ and $d \in \mathbb{R}$. They proved that a vector $\boldsymbol{x} \in \mathbb{R}^n$ satisfies Eq. (17) with uncertainty region U given by Eq. (18) if and only if there exist $\eta \in \mathbb{R}$ and $\delta \in \mathbb{R}$ such that $(\boldsymbol{x}, \delta, \eta)$ satisfies

$$\begin{cases} \boldsymbol{a}^T f(\boldsymbol{x}) + d\eta + \rho\delta + \delta \sum_{i=1}^m q_i \phi^* \left(\dfrac{b_i^T f(\boldsymbol{x}) - c_i^T \eta}{\delta} \right) \le b \\ \eta \ge 0, \delta \ge 0, \end{cases} \tag{19}$$

where b_i and c_i are the i-th columns of $\boldsymbol{B}$ and $\boldsymbol{C}$, respectively and ϕ^* is the conjugate function given by Eq. (6), with $(\delta\phi)^*(s) = \delta\phi^*(s/\delta)$ for $\delta \ge 0$ and $0\phi^*(s/0) = (0\phi^*)(s)$, which equals 0 if $s \le 0$ and $+\infty$ if $s > 0$. In (Ben-Tal et al., 2013) it has been shown that Eq. (19) is a convex optimization problem and is computationally tractable. In addition, the size of the uncertainty region can be controlled by the confidence level. These two features, make the approach appealing in robust simulation optimization. Here we use this method to find the approximate robust reformulation for the model obtained in Eq. (16). Adopting the formulation of Eq. (17) and Eq. (18), the approximate robust reformulation of Eq. (16) is the following semidefinite optimization problem:

$$\begin{aligned} & \min_{x \in X} z \\ & s.t.\ \eta + \rho\delta + \delta \sum_j q_j \phi^* \left(\frac{f_j(\boldsymbol{x}) - \eta}{\delta} \right) \le z \\ & \eta, \delta \ge 0, \end{aligned} \tag{20}$$

where $f_j(\boldsymbol{x}) = \sum_{i=1}^n \lambda_{ij}(\boldsymbol{x}) w_{ij}$ for $j=1,\ldots,m$.

It is noteworthy that although the final robust counterpart problem given by Eq. (19) is convex but as we use Kriging as $f(\boldsymbol{x})$ in eq. (19) and Kriging is not necessarily a convex mapping, the final robust counterpart problem (Eq. (20)) is not convex necessarily. Convexity of the model obtained by Kriging ($f(\boldsymbol{x})$) depends on the convexity of the I/O function and even for convex functions Kriging may give nonconvex estimates. "Kriging may even give metamodels that contradict our prior qualitative (structural) knowledge of the characteristics (e.g., convexity or monotonicity) of the I/O function that is implicitly defined by underlying simulation model"(Kleijnen et al., 2012).

We can extend the proposed approach to constrained problems in which the closed forms of some constraints are not in hand. The algorithm is the same and the only difference is that we should fit m Kriging metamodels for every constraint and work with multiple constraints rather than a single one. The approximate robust reformulation is as follows:

$$\begin{aligned} & \min_{x \in X} z \\ & s.t.\ \eta_h + \rho\delta_h + \delta_h \sum_j q_j \phi^* \left(\frac{f_{jh}(\boldsymbol{x}) - \eta_h}{\delta_h} \right) \le b_h \quad \forall h = 0, 1, \ldots, H \\ & \eta_h, \delta_h \ge 0, \end{aligned} \tag{21}$$

where h denotes the constraint index ($h = 0$ relates to the objective function) and $f_{jh}(x) = \sum_{i=1}^{n} \lambda_{ijh}(x) w_{ijh}$ for $j=1,2,\dots,m$ and $h=0,1,\dots,H$.

3.3. Constructing the robust counterpart by chance constraint method

Here we use the concept of chance constraint to find the robust counterpart for the objective function found in Eq. (14) in a different way. Let $f(x, \varepsilon)$ be a function of a decision vector, $x,$ and a random vector, $\varepsilon, \beta \in [0,1]$ as the prescribed probability bound and P as the known probability distribution of ε, a chance constraint is given by

$$Pr_{\varepsilon \sim P}\{\varepsilon: f(x, \varepsilon) \leq 0\} \geq \beta \tag{22}$$

Unlike robust optimization, in chance-constrained optimization it is assumed that the uncertain parameters are distributed according to a known distribution and if we have partial or no information on the probability distribution, P, it should be estimated. In such a situation, a safe approximation of ambiguous chance constraints is used. Here we use the safe approximation based on φ-divergence proposed by Yanikoglu and den Hertog (2012). They presented an algorithm to find the solution in a chance constraint problem using φ-divergence. Here we employ the algorithm and propose a new algorithm for problems whose objective function is obtained by Kriging metamodel. Following the proposed algorithm is described in more details.

Step 0. **Determining valid ranges for the random and decision variables:** Like previous section, we first determine the valid ranges for the random variables and divide the space into m cells. Let $V = \{1, \dots ,m\}$ be the set of cell indices. We then calculate q_j (the estimation of occurrence probability) for $j=1,2,\dots,m$ by Eq. (10). We should also determine the decision variables, $x,$ and their valid ranges. We sample n points from the solution space for the random and decision variables using a uniform sampling design like LHS and construct the simulation points set.

Step 1. **Constructing the objective function:** For every point of experimental design set, (x_i, ε_i), for $i=1,2,\dots,n$ we run the simulation model one time and show the simulation output by w_i. Therefore, we run the simulation model n times in total which is fewer than the $n \times m$ runs which is usually needed by simulation optimization methods. To estimate the objective function, a Kriging metamodel is fitted to the I/O combination of simulation model for n simulation points. The resulting metamodel is as follows:

$$Y(x, \varepsilon) = \sum_{i=1}^{n} w_i \lambda_i(x, \varepsilon), \tag{23}$$

where $\lambda_i(x, \varepsilon)$'s are the coefficients which are functions of the decision and random variables.

Step2. **Formulating the robust counterpart problem:** We consider the following chance constrained optimization problem for Kriging metamodel:

$$\min_{x \in X} z$$
$$s.t. \ Pr_{\varepsilon}\{\sum_{i=1}^{n} w_i \lambda_i(x, \varepsilon) \leq z\} \geq \beta, \tag{24}$$

where X is the feasible set given by deterministic constraints and β is the given probability bound. The robust counterpart problem is as follows:

$$\min_{x \in X} z$$
$$s.t. \ \sum_{i=1}^{n} w_i \lambda_i(x, \varepsilon) \leq z \qquad \forall \varepsilon \in Z, \tag{25}$$

where Z is the uncertainty set given by:

$$Z = \left\{ \boldsymbol{\varepsilon} \in \mathbb{R}^m : \left\| \boldsymbol{\varepsilon} \right\|_2 \leq \Omega \right\}, \tag{26}$$

where Ω is the radius of the ball in the uncertainty region which is set to zero in the first iteration (so Z is equal to the center point of the random variables space and the problem is the same as the nominal problem) and in the following iterations, it will be increased according to step five. By solving the above programming problem, the robust optimal solution, $x*$, is found.

For solving Eq. (25), which is a nonlinear robust optimization problem (when $\Omega \neq 0$), some methods are proposed in the literature (see e.g. (Ben-Tal et al., 2015; Bertsimas et al., 2010)). Here we apply the method proposed by Bertsimas et al. (2010) which is applicable to problems whose objective functions are nonconvex and given by a numerical simulation-driven model such as response surface and Kriging metamodels. Their proposed method just requires the value and the gradient of the objective function and is analogous to local search techniques such as gradient descent. (for more details about the method one can refer to (Bertsimas et al., 2010)).

Step 3. Constructing the valid set: We identify the set of cells for which the following equation holds:

$$S = \left\{ j \in V : \sum_{i=1}^{n} w_i \lambda_i \left(\boldsymbol{x}^*, \boldsymbol{\varepsilon}_j \right) \leq z \right\}, \tag{27}$$

where $\boldsymbol{\varepsilon}_j$ is the center point of cell $j \in V$. If the center point of a cell satisfies the constraint in Eq. (27) for a given $x*$, then we assume that all the realizations in the associated cell are feasible for the uncertain constraint.

Step 4. **Calculating $\gamma(S, \alpha)$:** To calculate $\gamma(S, \alpha)$ the following programming problem should be solved:

$$
\begin{aligned}
&\gamma(S, \alpha) = \min \textstyle\sum_{j \in S} p_j \\
&s.t. \ \ I_\phi(\boldsymbol{p}, \boldsymbol{q}) \leq \frac{\phi''(1)}{2N} \chi^2_{m-1,1-\alpha}(= \rho) \\
&\textstyle\sum_{j=1}^{m} p_j = 1 \\
&\boldsymbol{p} \geq 0
\end{aligned}
\tag{28}
$$

Note that Eq. (28) is a convex optimization problem in $\boldsymbol{p}$ since φ-divergence functions are convex. According to the above problem, the probability that the random variable is in the region defined by S, is at least $\gamma(S, \alpha)$ with a $(1-\alpha)$-confidence level. If $\gamma(S, \alpha) \geq \beta$ then the region determined by S is selected as the uncertainty set and the algorithm is terminated. Otherwise, we go to Step 5. In (Yanikoglu & den Hertog, 2012) it has been shown that an alternative way of calculating $\gamma(S, \alpha)$ is using the dual problem of Eq. (28) which results in the following convex problem:

$$max_{\eta \geq 0, \delta} \left\{ -\eta\rho - \delta - \eta \left[\phi^* \left(-\frac{\delta+1}{\eta} \right) \textstyle\sum_{j \in S} q_j + \phi^* \left(-\frac{\delta}{\eta} \right) \textstyle\sum_{j \in V \backslash S} q_j \right] \right\} \tag{29}$$

in which $\phi^*(s) = sup_{t \geq 0} \{ st - \phi(t) \}$.

Step 5. **Increasing Ω**: We increase Ω by the step size ω and go to Step 2.

4. Applications and results

We apply the proposed methods to the well-known EOQ problem, the multi-item newsvendor problem, and seven test functions. Numerical results are obtained on a CORE 2 duo notebook with 2.53GHZ processor and 3GB RAM. To fit Kriging metamodel, we employ the DACE toolbox of MATLAB.

Kullback-Leibler is used as the φ-convergence function. We use *lhsnorm* from the MATLAB statistics toolbox to pick points from the interval of decision variables including their extreme values. As the related robust counterpart is a nonlinear optimization problem (NLP), we apply MATLAB's function *fmincon* to compute the optimum robust solution, say x^*, for the counterpart problem of Kriging metamodel.

4.1. EOQ Inventory Model

Like Dellino et al. (2010b) we apply our methodology to the classic EOQ inventory model. Total costs per time unit C is considered as the objective function, which consists of setup cost per order K, cost per unit purchased c, and holding cost per inventory unit per time unit h. The delivery lead-time is zero, i.e., goods arrive into inventory as soon as an order is placed. Continuous review is assumed, i.e., an order is placed as soon as the inventory level falls below the prescribed reorder point, which is set to zero. Simulation uses 10,000 periods per replicate. In our example we use the parameter values of $K = 12000$, $c = 10$, and $h = 0.3$. To study robust simulation optimization, we assume that the customer demand in each period takes integer values between 6000 and 10000. We set α equal to 0.05. Here, the objective is to determine an EOQ value between $15000 \leq Q \leq 45000$ to minimize the total cost per period.

The inventory models were set up in ARENA 12. We pick nine equally spaced points; including the extreme points, 15000 and 45000 (see the first column of Table 2). To obtain the frequencies for random variable (demand), we divide the range of the parameter into nine equal intervals of size 500 (the second row of Table 2). Therefore; any distribution on the demand space can be represented by a vector $\boldsymbol{p} = (p_1, \dots, p_m)$ for $m=9$, with the constraint $\sum_{j=1}^{m} p_j = 1; p_j \geq 0$. Frequencies of the intervals are presented in the first row of Table 2. Running the simulation model, the total costs C (objective function value) are obtained for every combination of order quantities (decision variable) and demand range (random variable) that are listed in Table 2.

Table 2

Results of simulation model for EOQ problem

q	0.01	0.04	0.10	0.20	0.30	0.20	0.10	0.04	0.01
x	d=6000	d=6500	d=7000	d=7500	d=8000	d=8500	d=9000	d=9500	d=10000
15000	67048	72447	77853	83242	88654	94052	99447	104880	110242
20000	66601	71902	77200	82520	87809	93092	98412	103679	109023
25000	66633	71872	77118	82351	87581	92824	98057	103300	108560
30000	66902	72096	77290	82500	87688	92901	98101	103271	108509
35000	67300	72475	77654	82820	87990	93163	98335	103577	108687
40000	67810	72961	78107	83261	88404	93539	98703	103800	109012
45000	68355	73480	78616	83743	88880	94013	99146	104281	109379
50000	68941	74056	79173	84311	89420	94535	99662	104780	109864
55000	69559	74665	79780	84877	89997	95100	100230	1067798	110486

In the first method, nine Kriging metamodel $Y_j(\boldsymbol{x})$, for $j = 1,2, \dots ,9$ are fitted to the I/O combinations for every value of demand. The probability of occurrence of every metamodel is equal to the probability of occurrence of the random variable (demand). Employing the proposed method, the counterpart problem is written using (16) and is solved according to (20). The optimal solution is $x^* = 25000$ and the robust optimal objective function $C^* = 87590$.

For the second method, we consider probability bound $\beta = 0.95$, step size $\omega=500$, and $m=9$. We set the number of sampling points at $n=20$ and using LHS we select the sampling points from the valid range of the decision and random variables and run the simulation model for them. We then fit a Kriging metamodel to the combination of I/O of simulation. In Table 3 the number of iterations, It., the value of demand, d, the radius of the tight uncertainty region calculated by the algorithm, Ω, the optimal solution,

x^*, the objective function, C^*, the set of cells, S, and the probability bounds satisfied by the algorithm, $\gamma(S, \alpha)$, are shown for 3 iterations at 0.9 confidence level.

As can be seen, the final optimal robust solution is $x^* = 27389$ and the robust objective function is equal to $C^* = 101070$. The first row in Table 3 shows the answer for the nominal problem and the last row corresponds to the optimal robust solution with respect to the defined probability bound. Between the nominal solution and the solution satisfying a bound of 0.95, the optimal objective value increases by 15% while there is about 30% increase in the immunity to uncertainty.

Table 3
The optimal solution and the objective function obtained by the second method for EOQ problem

It.	d	Ω	x^*	C^*	S	$\gamma(S, \alpha)$
1	8000	0	25300	87952	{6000,6500,7000,7500,8000}	0.65
2	[7500,8500]	500	26760	96830	{6000,6500,7000,7500,8000,8500}	0.85
3	[7000,9000]	1000	27389	101070	{6000,6500,7000,7500,8000,8500,9000}	0.95

We compare the solutions to those obtained by Dellino method from (Dellino et al., 2012). In their method, two Kriging metamodels are fitted one for the mean and one for the variance of the response. To formulate the robust optimization problem, they consider minimizing the mean of the response as the objective function while satisfying a constraint on the standard deviation. In (Dellino et al., 2012) it is assumed that the demand has a Normal distribution with mean $\mu = 8000$ and standard deviation $\sigma = 800$. Here we obtain the robust solution by Dellino's method for $S_c=8200$. The optimal robust solution is $x^* = 32000$ and the robust objective function equals to $C^* = 92883$.

For comparison purposes, we also calculate the reference minimax value for EOQ problem. It is easy to verify that this problem has the following true I/O function, which we shall use to check our results:

$$C = \frac{aK}{Q} + ac + \frac{hQ}{2} \tag{30}$$

where a is the random variable (demand) and Q is the decision variable (order quantity). We write the counterpart problem as:

$$\min_Q \max_a \frac{12000a}{Q} + 10a + \frac{0.3Q}{2}$$
$$\text{s.t. } 15000 \leq Q \leq 45000 \tag{31}$$
$$5600 \leq a \leq 10400$$

We solve the above problem by a numerical simple algorithm to find the exact solution. The optimal solution yields $Q^* = 28844$ and $C^* = 112650$ which we use as a reference to compare the results of our methods to those proposed by Dellino et al. (2012). Results for the proposed methods and Dellino et al. (2012) method are compared by computing the absolute value of their relative deviation from reference values for the solution and the function value. The results are reported in Table 4.

Table 4
Results and relative deviation from reference for three methods on EOQ problem

	first method		second method		Dellino et al. method	
	result	relative deviation	result	relative deviation	result	relative deviation
x^*	25000	13.32%	27389	5.04%	32000	10.94%
C^*	87590	22.24%	101070	10.27%	92883	17.54%
N	81	-	30	-	75	-

As it is shown the first method leads to less conservative solutions compared with other two methods but relative deviation of the second method is less than other two methods. The relative deviation of the Dellino et al. (2012) method is less than the first method. Moreover, the second method requires the least number of simulation runs which is considered an advantage especially for expensive simulation problems.

4.2. The multi-item newsvendor problem

The newsvendor problem is a single period stochastic inventory problem. In its simplest form, there is a single item with a probabilistic demand that can only be sold in one period. Due to the uncertain demand, this newsvendor can face both unsold items and unmet demand. The unsold items will return a loss, and unmet demand generates a cost of lost sales. The objective is to select the replenishment quantity to maximize the expected profit. One extension to the basic newsvendor problem is the multi-item problem that deals with optimizing the replenishment quantity of several items. For each item i, we define the purchase cost c_i, salvage value s_i, selling price v_i, and the cost of lost sales l_i. Moreover, we denote Q_i for the order quantity for item i. We assume that the demand for each item i during the period is a random variable that can take on m values, d_{ij} for $i = 1,2, ..., n$ and $j = 1,2, ..., m$. We denote p_{ij} for the unknown probability that demand for item i equals d_j. We consider n=3 items, and m=3 scenarios for each item including: d_{i1}=4 (low demand), $d_{i2} = 8$(medium demand), and $d_{i3} = 10$(high demand). The parameter values of the problem, as well as the values of q_{ij} are given in Table 5.

Table 5
The parameter values of the multi-item newsvendor problem

Item (i)	c	v	s	l	q_{i1}	q_{i2}	q_{i3}
1	4	6	2	4	0.375	0.375	0.250
2	5	8	2.5	3	0.250	0.250	0. 500
3	6	9	1.5	5	0.375	0.250	0.375

points; including the extreme points (4 and 10), for Q_i, i=1,2,3. Running the simulation model, the total costs C (objective function value) are obtained for every combination of order quantities (decision variable) and demand range (random variable). In the first method, 27 Kriging metamodel $Y_j(\boldsymbol{x})$, for $j = 1,2, ...,27$ are fitted to the I/O combinations for every value of demand. The probability of occurrence of every metamodel is equal to the probability of occurrence of the random variable (demand). Employing the proposed method, the counterpart problem is written using Eq. (16) and is solved according to Eq. (20). The optimal solution is $Q^* = (8.9, 9.2, 8.9)$ and the robust optimal objectives function is $C^* = 34.37$. For the second method, we consider probability bound $\beta = 0.90$, step size ω=2. We set the number of sampling points at n=36 and using LHS we select sampling points from the valid range of the decision and random variables and run the simulation model for them. We then fit a Kriging metamodel to the combination of I/O of simulation. The final optimal robust solution, for 3 iterations and 0.95 confidence level, is $x^* = (8.1, 7.3, 7.2)$ and the robust objective function equals to $C^* = 9.5$. We compare the solutions to those obtained by Dellino method. Here we obtain the robust solution by Dellino's method for S_c=8. The optimal robust solution is $x^* = (6,6,6)$ and the robust objective function equals to $C^* = 12.346$. To calculate the reference minimax value for the problem, we use the following true I/O function:

$$f_{i,j}(Q_i, d_j) = v_i \min\{d_j, Q_i\} + s_i \max\{0, Q_i - d_j\} - l_i \max\{0, d_j - Q_i\} - c_i Q_i \, , \tag{32}$$

We write the counterpart problem as:

$$\max_Q \min_d f(Q, d) \tag{33}$$

s.t. $4 \le Q \le 10$

$d \in \{4,8,10\}$

We solve the above problem by a numerical simple algorithm to find the exact solution. The optimal solution is $Q^* = (7,7,7)$ and $C^* = 5$ which we use as a reference to compare the results of our methods to those of Dellino et al.'s. Results for the proposed methods and Dellino et al. (2012) method are compared by computing the absolute value of their relative deviation from reference values for the solution and the function value. The results are reported in Table 6. As it is shown the first method leads to less conservative solutions compared to other two methods but relative deviation of the second method is less than the other two methods. The relative deviation of the Dellino et al. method is less than the first method. Moreover, the number of simulation runs of the second method is the least which is considered an advantage especially for expensive simulation problems.

Table 6
Results and relative deviation from reference for three methods on multi-item newsvendor problem

	first method		second method		Dellino et al. method	
	result	relative deviation	result	relative deviation	result	relative deviation
x^*	(8.9,9.2,8.9)	0.85	(8.1,7.3,7.2)	0.22	(6.2,6.3,6.3)	0.32
C^*	34.37	5.874	9.5	0.9	12.345	1.46
N	540	-	150	-	270	-

4.3. Test Functions

Here we employ seven test functions from Marzat et al. (2013) and Rustem and Howe (2009) to show the performance of the proposed methods. These test functions are taken as black-box objective functions and the proposed method is applied to evaluate the minimax solution. The i-th decision variable is denoted by x_i, and the j-th random variable by e_j. The analytical expressions of functions are:

$f_1(x,\varepsilon) = 5(x_1^2 + x_2^2)^2 - (\varepsilon_1^2 + \varepsilon_2^2)^2 + x_1(-\varepsilon_1 + \varepsilon_2 + 5) + x_2(\varepsilon_1 - \varepsilon_2 + 3)$

$f_2(x,\varepsilon) = 4(x_1 - 2)^2 - 2\varepsilon_1^2 + x_1^2\varepsilon_1 - \varepsilon_2^2 + 2x_2^2\varepsilon_2$

$f_3(x,\varepsilon) = x_1^4\varepsilon_2 + 2x_1^3\varepsilon_1 - x_2^2\varepsilon_2(\varepsilon_2 - 3) - 2x_2(\varepsilon_1 - 3)^2$

$f_4(x,\varepsilon) = -\sum_{i=1}^{3}(\varepsilon_i - 1)^2 + \sum_{i=1}^{2}(x_i - 1)^2 + \varepsilon_3(x_2 - 1) + \varepsilon_1(x_1 - 1) + \varepsilon_2 x_1 x_2$

$f_5(x,\varepsilon) = -(x_1 - 1)\varepsilon_1 - (x_2 - 2)\varepsilon_2 - (x_3 - 1)\varepsilon_3 + 2x_1^2 + 3x_2^2 + x_3^2 - \varepsilon_1^2 - \varepsilon_2^2 - \varepsilon_3^2$

$f_6(x,\varepsilon) = \varepsilon_1(x_1^2 - x_2 + x_3 - x_4 + 2) + \varepsilon_2(-x_1 + 2x_2^2 - x_3^2 + 2x_4 + 1) + \varepsilon_3(2x_1 - x_2 + 2x_3 - x_4^2 + 5) + 5x_1^2 + 4x_2^2$
$\quad + 3x_3^2 + 2x_4^2 - \sum_{i=1}^{3}\varepsilon_i^2$

$f_7(x,\varepsilon) = 2x_1x_5 + 3x_4x_2 + x_5x_3 + 5x_4^2 + 5x_5^2 - x_4(\varepsilon_4 - \varepsilon_5 - 5) + x_5(\varepsilon_4 - \varepsilon_5 + 3) + \sum_{i=1}^{3}(\varepsilon_i(x_i^2 - 1)) - \sum_{i=1}^{5}\varepsilon_i^2$

We employ the first method to the test functions and set α equal to 0.05. Table 7 shows the test functions, solution space of the decision variables, *X,* and random variables, *E,* the number of simulation points, *n,* the number of ranges for the random variables, *m,* the optimal robust solution, x^*, the corresponding value of the objective function, C^*, and the number of simulation runs, *N*.

Table 7
Results of the first method for the test functions

Test function	X	E	n	m	x^*	C^*	N
$f_1(x,\varepsilon)$	$[-5,5]^2$	$[-5,5]^2$	25	25	(-0.075,0.201)	-149.36	625
$f_2(x,\varepsilon)$	$[-5,5]^2$	$[-5,5]^2$	25	25	(2.066,-0.656)	-9.74	625
$f_3(x,\varepsilon)$	$[-5,5]^2$	$[-3,3]^2$	25	9	(-4.277, 4.732)	-175.4	225
$f_4(x,\varepsilon)$	$[-5,5]^2$	$[-3,3]^3$	25	27	(1.002,1.283)	-3.2	675
$f_5(x,\varepsilon)$	$[-5,5]^3$	$[-1,1]^3$	50	8	(0.074,0.367,0.044)	-7.62	400
$f_6(x,\varepsilon)$	$[-5,5]^4$	$[-2,2]^3$	80	8	(-0.134,-0.032,-0.188,0.102)	-6.9	640
$f_7(x,\varepsilon)$	$[-5,5]^5$	$[-3,3]^5$	3125	243	(0.91,0.88,0.54,-1.76,-1.98)	-9.54	759375

In second method, probability bound β is set to 0.9 and ω is set to 1. The first four columns of Table 8 show the test functions, number of iterations, the optimal robust solution, and the corresponding value of the objective function respectively. The fifth column presents the radius of the minimal ball in the uncertainty region, Ω, and the sixth column gives the probability bounds satisfied by the algorithm, $\gamma(S, \alpha)$. The seventh column corresponds to the number of simulation runs required to obtain the optimal solution.

Table 8
Results of the second method for the test functions

Test function	It.	x^*	C^*	Ω	$\gamma(S, \alpha)$	N
$f_1(x, \varepsilon)$	4	(-0.311,-0.106)	-3.56	3	1	50
$f_2(x, \varepsilon)$	4	(1.943,-0.032)	-1.42	3	0.96	50
$f_3(x, \varepsilon)$	3	(-1.273, 1.043)	-3.745	2	0.92	30
$f_4(x, \varepsilon)$	3	(0.464,0.024)	-1.4	2	0.94	60
$f_5(x, \varepsilon)$	2	(0.320,0.198,-0.276)	1.02	1	1	50
$f_6(x, \varepsilon)$	2	(-0.097,-0.278,0.137,-0.072)	3.234	1	1	100
$f_7(x, \varepsilon)$	4	(1.22,1.54,1.13,-1.04,-1.11)	-5.45	3	0.96	1000

We compare the resulting solutions to those obtained by Dellino et al. from (Dellino et al., 2012). Here we assume that the random variables have a Normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.15R$, where R is the range of the random variable. We set the number of replications at each design point at 10. The number of sampling points is set to $10d$ (d is the number of decision variables). We obtain the robust solution by Dellino et al. method for standard deviation Sc equal to the minimum standard deviation obtained from the 20 first simulated design points for every test function.

For comparison purposes we employ the results of Rustem and Howe (2009) for these test functions as reference solutions. In (Rustem & Howe, 2009) three descent methods have been compared on these test functions and lead to similar results. Table 9 shows the results for reference values and Dellino et al. method.

Table 9
Reference solutions and results of the Dellino et al. method for the seven test functions

Test functions	Reference solutions		Dellino et al.'s method		
	x^*	C^*	x^*	C^*	N
$f_1(x, \varepsilon)$	(-0.483,-0.316)	-1.683	(0.5,0.3)	-741.34	200
$f_2(x, \varepsilon)$	(1.695,-0.003)	1.403	(0.05,0.21)	-14.88	200
$f_3(x, \varepsilon)$	(-1.180,0.912)	-2.468	(0.23,0.02)	-0.53	200
$f_4(x, \varepsilon)$	(0.418,0.418)	-0.134	(1.05,1.90)	-5.002	200
$f_5(x, \varepsilon)$	(0.111,0.153,0.2)	1.345	(1.22,2.14,0.92)	4.67	300
$f_6(x, \varepsilon)$	(-0.231,0.222,-0.675,-0.083)	4.543	(0.07,-1.83,-3.16,0.1)	30.53	400
$f_7(x, \varepsilon)$	(1.42,1.66,1.25,-0.97,-0.73)	-6.35	(1.07,1.32,0.91,0.04,-0.10)	-7.43	31250

The results for the proposed methods and Dellino et al. (2012) method are compared by computing the absolute value of their relative deviations from reference values for the solution and the function values. The results are reported in Table 10. As it is shown, the results of the second method are closer to the reference values than those of the first method and Dellino et al. method. The mean of deviation for the values of the objective function is 1.96 for the second method while this amount is 28.40 for the first method and 71.06 for Dellino et al. method. Moreover, the mean deviation for the values of the robust solution is 2.40 for the second method, 32.97 for the first method and 18.00 for Dellino et al. method. Therefore, the second method obtains better minimax solutions than the first method and Dellino et al.

method. Moreover, number of simulation runs in second method is significantly fewer than the two other methods.

We also study the effect of β in relative deviation of the objective function and the robust solution for $f_1(x, \varepsilon)$. As it is shown in Fig. 1, by increasing β toward 1, the relative deviation of the objective function and robust solution decreases. This is so by increasing β, the second method will become similar to the mini-max method which is employed as reference for evaluation of solutions. It is worth noting that the comparison of the solutions of robust methods is not straightforward as they use different logics to obtain the robust solutions. However, we can mention the following as the main advantages of the proposed methods in comparison to other robust simulation optimization methods:

- We have considered the probability of occurrence of random variables and our methods give solutions that are less conservative than those of the worst-case methods.
- Our proposed methods are based on the distribution-free structure of robust optimization, i.e., unlike the common robust simulation optimization approaches, we do not make any assumptions about the probability distribution of the random variables.
- The first method leads to tighter uncertainty sets and less conservative solutions. Furthermore, it can be applied to the constrained problems and can consider robustness in feasibility and optimality of problems. While most of the existing methods are suitable for unconstrained problems and can just consider robustness dealing with the objective function.
- The second method's solutions are close to the minimax solutions (but less conservative) and its level of conservatism can be controlled by the probability bound, β .Furthermore, the second method requires fewer points for simulation than common robust simulation optimization methods.

Table 10
Relative deviations of the two proposed methods and Dellino et al. method from reference value

Test function	First method		Second method		Dellino et al. method	
	$x(\%)$	$f_i(x,\varepsilon)(\%)$	$x(\%)$	$f_i(x,\varepsilon)(\%)$	$x(\%)$	$f_i(x,\varepsilon)$
$f_1(x,\varepsilon)$	2.48	87.70	1.02	1.11	3.98	439.48
$f_2(x,\varepsilon)$	217.44	7.94	9.52	2.01	71.97	11.60
$f_3(x,\varepsilon)$	1.56	70.06	0.06	0.51	2.17	0.78
$f_4(x,\varepsilon)$	3.46	22.88	1.05	9.44	5.05	36.32
$f_5(x,\varepsilon)$	2.51	6.65	4.55	0.24	23.51	2.47
$f_6(x,\varepsilon)$	2.22	2.51	0.33	0.28	16.43	5.72
$f_7(x,\varepsilon)$	1.12	0.50	0.28	0.14	2.93	0.17
mean	32.97	28.33	2.40	1.96	18.00	70.9
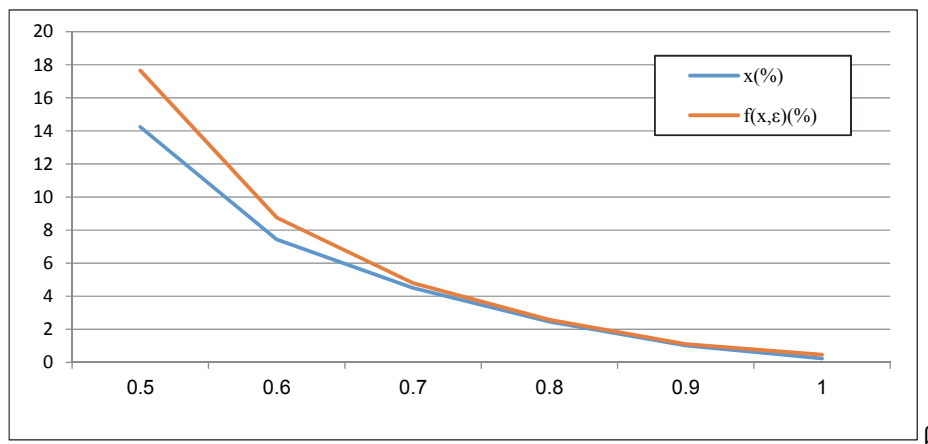
Fig. 1. The relative deviations of the objective function and robust solution for different values of β for $f_1(x,\varepsilon)$

5. Conclusions and ideas for future research

We have proposed two methods based on φ-divergence uncertainty to obtain the robust counterpart for metamodel-based simulation optimization problems (problems that the closed form of the objective functions or constraints are not in hand and are estimated by metamodels). The methods are applicable to problems whose probability distributions of the random variables are unknown but historical data are available. Considering the probability of occurrence of random variables leads to a tighter uncertainty set and less conservative robust solutions.

In the first method, we have found the minimax solution based on Ben-Tal method. In the second method, we use the chance constraint methodology to find the robust solution. We employed Kriging metamodel to estimate the objective function (and constraints) but other metamodels like regression and artificial neural networks can be employed too.

We have applied the proposed methods to the well-known EOQ problem, the multi-item newsvendor problem, and seven test functions and compared the results with each other and with those of Dellino et al. method. Numerical results show that the first method leads to less conservative solutions compared to the two other methods. However, the solutions of the second method are closer to the minimax solutions and its number of simulation runs are fewer than the two other methods.

An interesting subject for further research is employing other goodness-of-fit statistics in the simulation optimization context and investigating the tractability of the corresponding robust counterpart problem. Studying the applicability of the proposed method to the complicated real life problems where the probability distribution is not in hand and only some historical demand data are given is among other extensions.

References

Ajdari, A., & Mahlooji, H. (2014). An adaptive exploration-exploitation algorithm for constructing metamodels in random simulation using a novel sequential experimental design. *Communications in Statistics-Simulation and Computation, 43*(5), 947-968.

Ankenman, B., Nelson, B. L., & Staum, J. (2010). Stochastic kriging for simulation metamodeling. *Operations research, 58*(2), 371-382.

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

Ben-Tal, A., Den Hertog, D., & Vial, J.-P. (2015). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical programming, 149*(1-2), 265-299.

Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM review, 53*(3), 464-501.

Bertsimas, D., Nohadani, O., & Teo, K. M. (2010). Robust optimization for unconstrained simulation-based problems. *Operations research, 58*(1), 161-178.

Bhatnagar, S., Mishra, V. K., & Hemachandra, N. (2011). Stochastic algorithms for discrete parameter simulation optimization. *Automation Science and Engineering, IEEE Transactions on, 8*(4), 780-793.

Birge, J. R., & Wets, R. J.-B. (1986). *Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse*: Springer.

Calafiore, G. C. (2007). Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization, 18*(3), 853-877.

Chang, K.-H. (2012). Stochastic Nelder–Mead simplex method–A new globally convergent direct search method for simulation optimization. *European journal of operational research, 220*(3), 684-694.

Cramer, A. M., Sudhoff, S. D., & Zivi, E. L. (2009). Evolutionary algorithms for minimax problems in robust design. *Evolutionary Computation, IEEE Transactions on, 13*(2), 444-453.

Dellino, G., Kleijnen, J. P., & Meloni, C. (2009). *Robust simulation-optimization using metamodels.* Paper presented at the Winter Simulation Conference.

Dellino, G., Kleijnen, J. P., & Meloni, C. (2010a). *Parametric and distribution-free bootstrapping in robust simulation-optimization.* Paper presented at the Proceedings of the Winter Simulation Conference.

Dellino, G., Kleijnen, J. P., & Meloni, C. (2010b). Robust optimization in simulation: Taguchi and response surface methodology. *International Journal of Production Economics, 125*(1), 52-59.

Dellino, G., Kleijnen, J. P., & Meloni, C. (2012). Robust optimization in simulation: Taguchi and Krige combined. *INFORMS Journal on Computing, 24*(3), 471-484.

Dellino, G., & Meloni, C. (2015). Metamodel Variability in Simulation-Optimization: a Bootstrap Analysis.

Gabrel, V., Murat, C., & Thiele, A. (2014). Recent advances in robust optimization: An overview. *European journal of operational research, 235*(3), 471-483.

Hurrion, R. (1997). An example of simulation optimisation using a neural network metamodel: finding the optimum number of kanbans in a manufacturing system. *Journal of the Operational Research Society, 48*(11), 1105-1112.

Kim, S.-H., & Nelson, B. L. (2007). *Recent advances in ranking and selection.* Paper presented at the Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come.

Klabjan, D., Simchi-Levi, D., & Song, M. (2013). Robust Stochastic Lot-Sizing by Means of Histograms. *Production and Operations Management, 22*(3), 691-710.

Kleijnen, J. P. (2009). Kriging metamodeling in simulation: A review. *European journal of operational research, 192*(3), 707-716.

Kleijnen, J. P. (2014). Simulation-optimization via Kriging and bootstrapping: a survey. *Journal of Simulation, 8*(4), 241-250.

Kleijnen, J. P. (2015). Regression and kriging metamodels with their experimental designs in simulation: review (Vol. 35): Tilburg: CentER, Center for Economic Research.

Kleijnen, J. P., Beers, W. v., & Nieuwenhuyse, I. v. (2010). Constrained optimization in expensive simulation: Novel approach. *European journal of operational research, 202*(1), 164-174.

Kleijnen, J. P., Mehdad, E., & van Beers, W. (2012). *Convex and monotonic bootstrapped kriging.* Paper presented at the Proceedings of the Winter Simulation Conference.

Kovach, J., & Cho, B. R. (2009). A D-optimal design approach to constrained multiresponse robust design with prioritized mean and variance considerations. *Computers & Industrial Engineering, 57*(1), 237-245.

Lee, K.-H., Park, G.-J., & Joo, W.-S. (2006). A global robust optimization using kriging based approximation model. *JSME International Journal, Series C: Mechanical Systems Machine Elements & Manufacturing, 49*(3), 779-803.

Lin, R.-C., Sir, M. Y., & Pasupathy, K. S. (2013). Multi-objective simulation optimization using data envelopment analysis and genetic algorithm: Specific application to determining optimal resource levels in surgical services. *Omega, 41*(5), 881-892.

Liu, H., & Maghsoodloo, S. (2011). Simulation optimization based on taylor kriging and evolutionary algorithm. *Applied Soft Computing, 11*(4), 3451-3462.

Lung, R. I., & Dumitrescu, D. (2011). *A new evolutionary approach to minimax problems.* Paper presented at the Evolutionary Computation (CEC), 2011 IEEE Congress on.

Marzat, J., Walter, E., & Piet-Lahanier, H. (2013). Worst-case global optimization of black-box functions through Kriging and relaxation. *Journal of Global Optimization, 55*(4), 707-727.

Myers, W. R., & Montgomery, D. C. (2003). Response surface methodology. *Encycl Biopharm Stat, 1*, 858-869.

Nezhad, A. M., & Mahlooji, H. (2013). An artificial neural network meta-model for constrained simulation optimization. *Journal of the Operational Research Society, 65*(8), 1232-1244.

Pardo, L. (2005). *Statistical inference based on divergence measures*: CRC Press.

Quan, N., Yin, J., Ng, S. H., & Lee, L. H. (2013). Simulation optimization via kriging: a sequential search using expected improvement with computing budget constraints. *Ie Transactions, 45*(7), 763-780.

Rustem, B., & Howe, M. (2009). *Algorithms for worst-case design and applications to risk management*: Princeton University Press.

Shapiro, A., & Ahmed, S. (2004). On a class of minimax stochastic programs. *SIAM Journal on Optimization, 14*(4), 1237-1249.

Shapiro, A., & Kleywegt, A. (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software, 17*(3), 523-542.

Simpson, T. W., Poplinski, J., Koch, P. N., & Allen, J. K. (2001). Metamodels for computer-based engineering design: survey and recommendations. *Engineering with computers, 17*(2), 129-150.

Stinstra, E., & Den Hertog, D. (2008). Robust optimization using computer experiments. *European journal of operational research, 191*(3), 816-837.

Wang, Z., Glynn, P., & Ye, Y. (2009). Likelihood robust optimization for data-driven newsvendor problems: Working paper.

Yang, F. (2010). Neural network metamodeling for cycle time-throughput profiles in manufacturing. *European journal of operational research, 205*(1), 172-185.

Yanikoglu, I., & den Hertog, D. (2012). Safe approximations of ambiguous chance constraints using historical data. *INFORMS Journal on Computing, 25*(4), 666-681.

Zhou, A., & Zhang, Q. (2010). *A surrogate-assisted evolutionary algorithm for minimax optimization*. Paper presented at the Evolutionary Computation (CEC), 2010 IEEE Congress on.