On constrained distribution-free p-boxes and their propagation

Marco Daub[1)], Stefano Marelli[1)], Bruno Sudret[1)]
[1)] *ETH Zürich, Stefano-Franscini-Platz 5, 8093 Zürich, Switzerland, daub@ibk.baug.ethz.ch, marelli@ibk.baug.ethz.ch, sudret@ethz.ch*

Abstract. Propagating uncertainties through computational models is a key ingredient in uncertainty quantification in engineering. In general, uncertainty can be characterized as aleatoric, caused by variability, and epistemic, caused by lack of knowledge. Imprecise probability theory offers a natural framework to deal with both through, through sets of probability distributions. Among the latter, probability-boxes (p-boxes), which specify upper and lower bounds on admissible cumulative distribution functions (CDFs), are well established in the literature. We hereby introduce a novel class of p-boxes, *constrained distribution-free p-boxes*, that is based on imposing constraints on the admissible distributions (e.g. bound moments, symmetry, derivatives, etc.) on otherwise distribution-free p-boxes. We demonstrate that this class maintains most of the flexibility of classical distribution-free p-boxes, while avoiding most of the non-physical configurations it can be associated with. We also show how constrained distribution-free p-boxes can influence uncertainty bounds in the model predictions, thus improving the quality of the resulting uncertainty estimation.

Keywords: uncertainty quantification, uncertainty propagation, imprecise probabilities, probability-boxes, Monte Carlo simulation

1. Introduction

Over the past decades, *uncertainty quantification (UQ)* for engineering systems has become of great interest, see (Soize, 2017). Popular disciplines of UQ concern, for instance, with the analysis or the selection of a design for reliability or robustness, (Moustapha and Sudret, 2019; Teixeira et al., 2021). Here, computational models that analyze the behavior of mechanical systems together with the increased availability of computational resources have made major contributions: they are used to propagate the uncertainty in input parameters to the output space by performing multiple model evaluations. Usually, the uncertain parameters in the input space are modeled by probabilistic models to account for their variation, i.e., intrinsic randomness. This type of uncertainty is also called *aleatoric uncertainty*, and Monte Carlo simulation (MCS) is usually applied for their propagation. However, there are often situations in engineering in which the actual probabilistic model is unknown. Hence, *epistemic uncertainty*, which is due to lack of knowledge and can be reduced in principle, adds to the aleatoric uncertainty, see (Der Kiureghian and Ditlevsen, 2009).

In general, aleatoric and epistemic uncertainties can be treated in a unified framework called *imprecise probability theory*, see (Augustin et al., 2014; Walley, 1991). For the modeling of uncertain parameters, imprecise probability models consider sets of probability distributions. A popular representative hereof are *(distribution-free) p-boxes*, which provide upper and lower bounds for the unknown cumulative distribution functions (CDF) to define such a set. Because of this simple construction and their intuitive visualization, p-box are especially useful in engineering, (Ferson et al., 2003). The more information about the unknown CDF of the uncertain parameter is available, the tighter the bounding CDFs can be chosen. This yields a unique CDF in the limit case. The bounding CDFs are usually constructed based on given information. In general, it is distinguished between construction methods that are based on incomplete distribution properties, or a given dataset with samples of the uncertain parameters (see (Ferson et al., 2003) for an overview). Unfortunately, the use of these methods may lead to wide bounds and CDFs enclosed by the p-box that do not match the given information, see (Beer et al., 2013). For example, the p-box covering all CDFs of random variables with specific mean and variance values, constructed by the Chebychev inequality, as done in (Oberguggenberger and Fellin, 2008), also covers CDFs of random variables with different mean and variance values. This motivates to constrain the feasible CDFs within the p-box bounds. A possible restriction would be to limit the considerations to a specific CDF family, which yields a so-called *parametric p-box*. However, reasons for this restriction might be lacking and thus, the epistemic uncertainty of the uncertain parameters is not represented well enough. Nevertheless, a remedy can be found by using either generalized distribution families for parametric p-boxes, see (Daub et al., 2021b), or by constraining the feasible CDFs of a distribution-free p-box. In this paper, the latter idea is considered, and *constrained distribution-free p-boxes* are proposed.

This paper is organized as follows: first, the basics of p-boxes are explained, and constrained distribution-free p-boxes are derived from distribution-free p-boxes. Then, both p-box types are discretized for numerics, and a pragmatic propagation method is presented. This is applied to the two-dimensional Rosenbrock function as a computational model and the results for various constrained distribution-free p-boxes are discussed.

2. P-boxes

2.1. From distribution-free to constrained distribution-free p-boxes

In order to describe variables facing aleatoric uncertainty, random variables are used in probability theory. Here, a *real-valued random variable* X is a measurable function which is defined on a probability space and maps from its sample space to $\mathbb{R}$. Conventionally, X is characterized by a *cumulative distribution function (CDF)* F_X, which states the probability that X takes on a value less than or equal to $x \in \mathbb{R}$, i.e.,

$$F_X : \mathbb{R} \to [0, 1],\ F_X(x) = P(X \leq x). \tag{1}$$

A CDF has the property that it is non-decreasing, right-continuous, and converges to 0 for $x \to -\infty$ and 1 for $x \to \infty$. Its derivative $f_X = \frac{\mathrm{d}}{\mathrm{d}x} F_X$ is called *probability density function (PDF)*. Hence, the function value of F_X at $x \in \mathbb{R}$ can be also expressed as

$$F_X(x) = \int_{\mathbb{R}} I_x(x') f_X(x') \,\mathrm{d}x'. \tag{2}$$

where I_x is an indicator function with $I_x(x') = 1$ for $x' \leq x$ and $I_x(x') = 0$ for $x' > x$, $x' \in \mathbb{R}$.

In the context of imprecise probabilities, the actual CDF of X faces epistemic uncertainty, i.e., is unknown due to lack of knowledge, and only sets in which it is contained can be provided. One approach to define such a set of possible CDFs for X is the modeling as a *probability-box*, or short *p-box*. A p-box is defined as a pair of an upper CDF $\overline{F}_X$ and a lower CDF $\underline{F}_X$, see (Ferson et al., 2003). This results in the set

$$[\overline{F}_X, \underline{F}_X] = \{F_X \in \mathcal{F} \mid \underline{F}_X(x) \leq F_X(x) \leq \overline{F}_X(x),\ x \in \mathbb{R}\} \tag{3}$$

for the unknown CDF of X, where $\mathcal{F}$ is the set of all CDFs on $\mathbb{R}$. As $\mathcal{F}$, and thus $[\underline{F}_X, \overline{F}_X]$, allows for arbitrary distribution types, $[\overline{F}_X, \underline{F}_X]$ is also called a *distribution-free p-box*.

If the feasible CDFs can be limited to a specific CDF family $\mathcal{F}_\Theta$ with parameters θ_i, $i = 1, \ldots, n_\theta$, collected in $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$, a *parametric p-box*, also called *distributional p-box* is obtained. Here, bounds are put on the parameters θ_i instead of directly on the CDFs $F_X(\cdot, \boldsymbol{\theta})$. This is usually done in the form of lower bounds $\underline{\theta}_i$ and upper bounds $\overline{\theta}_i$, yielding intervals $[\underline{\theta}_i, \overline{\theta}_i]$ for each parameter θ_i, $i = 1, \ldots, n_\theta$. The envelope of a parametric p-box is a distribution-free p-box defined by the bounding CDFs

$$\overline{F}_X(x) \ = \ \max\{F_X(x, \theta) \mid \underline{\theta}_i \leq \theta_i \leq \overline{\theta}_i,\ i = 1, \ldots, n_\theta\}, \tag{4}$$

$$\underline{F}_X(x) \ = \ \min\{F_X(x, \theta) \mid \underline{\theta}_i \leq \theta_i \leq \overline{\theta}_i,\ i = 1, \ldots, n_\theta\} \tag{5}$$

for $x \in \mathbb{R}$. Note that the parametric p-box is usually a proper subset of its envelope, i.e., they are not equal.

As opposed to focusing on a specific distribution family for using only selected CDFs of a distribution-free p-box, also the feasible CDFs of a distribution-free p-box can be constrained in a non-parametric way. This leads to a *constrained distribution-free p-box*, which is proposed in the following. The idea of this p-box type is to narrow the feasible CDFs by putting further constraints on the p-box than the fundamental p-box constraint in Equation (3). For example, upper and lower bounds can be put on the moments of X, like the mean $\mu = E(X)$ and the variance $\sigma^2 = \mathrm{Var}(X)$, or on the values of the PDF f_X and its further derivatives. In addition, the distributional shape can be constrained for, e.g., symmetry or unimodality. Thus, a set of feasible CDF $\mathcal{F}' \subseteq \mathcal{F}$ that only comprises CDFs fulfilling these constraints can be used to define a constrained distribution-free p-box for the unknown CDF of X via

$$[\overline{F}_X, \underline{F}_X]' = [\overline{F}_X, \underline{F}_X] \cap \mathcal{F}'. \tag{6}$$

Basically, this idea is already included in (Beer et al., 2013), in which a p-box is represented as a quintuple $\langle \overline{F}_X, \underline{F}_X, [\underline{\mu}, \overline{\mu}], [\underline{\sigma}, \overline{\sigma}], \mathcal{F}') \rangle$. Here, the bounds on the mean and the standard deviation, given by $[\underline{\mu}, \overline{\mu}]$ and $[\underline{\sigma}, \overline{\sigma}]$, are separated from further restrictions on the CDFs, given by $\mathcal{F}' \subseteq \mathcal{F}$. However, this remains rather a theoretical construct and appropriate techniques on how to handle such constraints for computations are lacking. To counteract this, a discretization scheme for CDFs that allows the integration of constraints is deployed in the following. First, distribution-free p-boxes with no constraints are considered again.

2.2. Numerics with distribution-free p-boxes

In this paper, the discretization is done with respect to the real space $\mathbb{R}$. Here, a simple approach is to put an equidistant grid on an interval $[\underline{x}, \overline{x}] \subseteq \mathbb{R}$. The lower bound $\underline{x}$ is chosen as the largest values such that $\overline{F}_X(\underline{x}) = 0$ holds and the upper bound $\overline{x}$ as the smallest value such that $\underline{F}_X(\overline{x}) = 1$ holds. In case of an unbounded support, these bounds are chosen application-specific so that the approximations $\overline{F}_X(\underline{x}) \approx 0$ and $\underline{F}_X(\overline{x}) \approx 1$ can be accepted. Let $N \in \mathbb{N}$ be the number of grid points inside the interval $[\underline{x}, \overline{x}]$, then the grid is defined as

$$[\underline{x}, \overline{x}]_h = \{x_h^j \in \mathbb{R} \mid x_h^j = \underline{x} + jh,\ j \in \{0, \dots, N+1\}\} \tag{7}$$

with $h = \frac{\overline{x}-\underline{x}}{N+1}$. Accordingly, the values of the CDF F_X at the grid points x_h^j are denoted by $F_{X,h}^j = F_X(x_h^j) \in [0,1]$ for $j = 0, \dots, N+1$. In order to guarantee that F_X is a proper CDF which is an element of the p-box $[\overline{F}_X, \underline{F}_X]$, the following inequalities must be fulfilled for $F_{X,h}^j$:

$$F_{X,h}^j \le F_{X,h}^{j+1}, \tag{8}$$

$$F_{X,h}^j \le \overline{F}_X(x_h^j), \tag{9}$$

$$F_{X,h}^j \ge \underline{F}_X(x_h^j), \tag{10}$$

$j = 1, \dots, N$. Moreover, $F_{X,h}^0 = F_X(\underline{x}) = 0$ and $F_{X,h}^{N+1} = F_X(\overline{x}) = 1$ hold or are least assumed for $i = 0, N+1$. The unknown $F_{X,h}^j$, $j = 1, \dots, N$ are collected in $\boldsymbol{F}_{X,h} = (F_{X,h}^1, \dots, F_{X,h}^N) \in [0,1]^N$. As the inequalities (8)-(10) are all linear, they can be stated as

$$\boldsymbol{A}\boldsymbol{F}_{X,h} \le \boldsymbol{b}, \tag{11}$$

for $\boldsymbol{F}_{X,h} \in [0,1]^N$, where $\boldsymbol{A} \in \mathbb{R}^{(3N+1)\times N}$ and $\boldsymbol{b} \in \mathbb{R}^{3N+1}$. The set of all $\boldsymbol{F}_{X,h}$ which fulfill the system of linear inequalities (11) is denoted by $\mathcal{S}_{\boldsymbol{F}_{X,h}}$, i.e.,

$$\mathcal{S}_{\boldsymbol{F}_{X,h}} = \{\boldsymbol{F}_{X,h} \in [0,1]^N \mid \boldsymbol{A}\boldsymbol{F}_{X,h} \le \boldsymbol{b}\}. \tag{12}$$

For numerical computations, often the values of F_X at non-grid points are also required. To account for this with a simple approach, linear interpolation is used in this paper, i.e.,

$$F_X(x) = F_{X,h}^j + \frac{F_{X,h}^{j+1} - F_{X,h}^j}{x_h^{j+1} - x_h^j}(x - x_h^j) \tag{13}$$

for $x \in (x_h^j, x_h^{j+1})$. It ensures that F_X is non-decreasing when it fulfills Equation (8). The set of all piecewise linear CDFs defined by Equation (13) for which $\boldsymbol{F}_{X,h} \in \mathcal{S}_{\boldsymbol{F}_{X,h}}$ holds is denoted by $[\overline{F}_X, \underline{F}_X]_h$. For $N \to \infty$, i.e., $h \to 0$, $F_X \in [\overline{F}_X, \underline{F}_X]_h$ is capable to approximate an arbitrary CDF of the p-box $[\overline{F}_X, \underline{F}_X]$ in $(\underline{x}, \overline{x})$. However, by using Equation (8), $F_X \in [\overline{F}_X, \underline{F}_X]_h$ can violate the p-box bounds if $\overline{F}_X$ is not concave or if $\underline{F}_X$ is not convex. This is accepted here, as the maximum error decreases with growing N. Moreover, $F_X \in [\overline{F}_X, \underline{F}_X]_h$ is not differentiable at the grid points. If differentiability is desired, a more profound interpolation approach is monotone (cubic) interpolation, see, e.g., (Fritsch and Carlson, 1980). Below, an example that illustrates the use of piecewise linear CDFs for distribution-free p-boxes is provided.

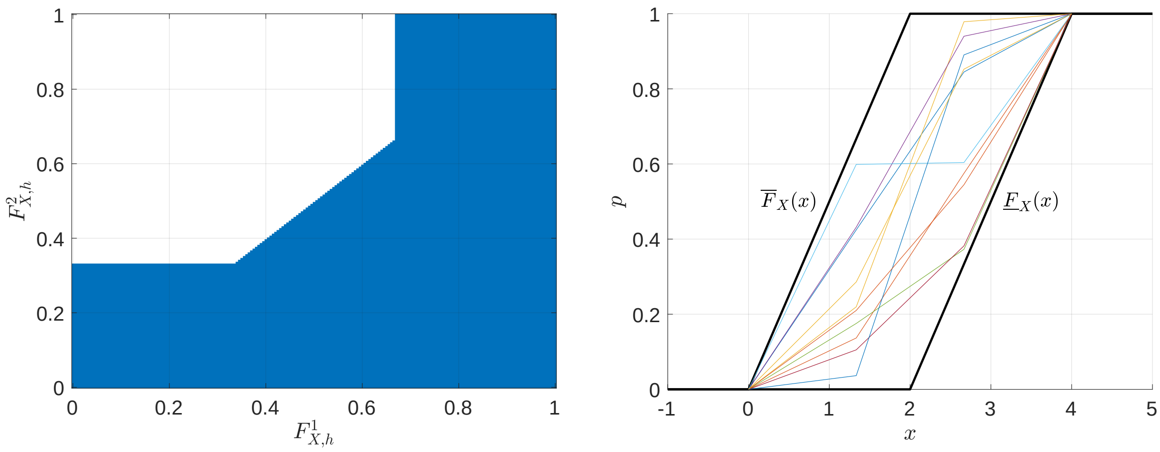


Figure 1. Distribution free p-box $[\overline{F}_X, \underline{F}_X]_h$ for Example 1 and $N = 2$. *Left:* Feasible set $\mathcal{S}_{\boldsymbol{F}_{X,h}}$ (white region) and unfeasible set (blue region) for $\boldsymbol{F}_{X,h}$. *Right:* P-box bounds of $[\overline{F}_X, \underline{F}_X]_h$ with examples of feasible CDFs.

EXAMPLE 1. *Let a distribution-free p-box be defined by the bounding CDFs*

$$\overline{F}_X(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ 0.5x & \text{for } 0 < x \leq 2, \\ 1 & \text{for } x > 2, \end{cases} \qquad \underline{F}_X(x) = \begin{cases} 0 & \text{for } x \leq 2, \\ 0.5x - 1 & \text{for } 2 < x \leq 4, \\ 1 & \text{for } x > 4 \end{cases} \tag{14}$$

for $x \in \mathbb{R}$, *which belong to the uniform distribution family. For the purpose of visualization, a discretization of* $[\underline{x}, \overline{x}] = [0, 4]$ *is done by only* $N = 2$ *inner grid points, first, and the feasible CDFs of* $[\overline{F}_X, \underline{F}_X]$ *are represented by piecewise linear CDFs* $F_X \in [\overline{F}_X, \underline{F}_X]_h$ *defined via* $\boldsymbol{F}_{X,h} = (F^1_{X,h}, F^2_{X,h})$*. CDF examples of* $[\overline{F}_X, \underline{F}_X]_h$ *are shown in Figure 1 together with their feasible set* $\mathcal{S}_{\boldsymbol{F}_{X,h}}$*. Then, the same is done for* $N = 100$ *to illustrate the full capability of the approach, and CDF examples of* $[\overline{F}_X, \underline{F}_X]_h$ *are shown in Figure 2.*

2.3. Numerics with constrained distribution-free p-boxes

The approach for distribution-free p-boxes can be extended for constraint distribution-free p-boxes by introducing, in addition to the inequalities (8)-(10), inequalities accounting for specific constraints. Exemplarily, inequalities to yield upper and lower bounds on the PDF and the mean, and symmetry around $\frac{1}{2}(\underline{x} + \overline{x})$ for the piecewise linear CDFs as defined in Equation (13) are listed in the following:

– *Bounds on the PDF*: As a piecewise linear CDF with $\boldsymbol{F}_{X,h} \in \mathcal{S}_{\boldsymbol{F}_{X,h}}$ is not differentiable at the grid-points in a strong sense, only the derivative values in $x \in (x^j_h, x^{j+1}_h)$, $j = 0, \ldots, N$, are considered. There are a lower bound $\underline{f} \in \mathbb{R}$ and an upper bound $\overline{f} \in \mathbb{R}$ on the PDF if the

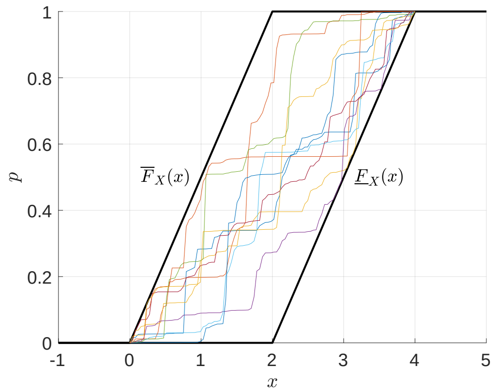


Figure 2. P-box bounds of the distribution free p-box $[\overline{F}_X, \underline{F}_X]_h$ with examples of feasible CDFs for Example 1 and $N = 100$.

inequalities

$$\underline{f} \leq \frac{F^{j+1}_{X,h} - F^{j}_{X,h}}{h} \leq \overline{f} \tag{15}$$

are fulfilled for $j = 0, \ldots, N$.

– *Bounds on the mean*: There are a lower bound $\underline{\mu} \in \mathbb{R}$ and an upper bound $\overline{\mu} \in \mathbb{R}$ on the mean for a piecewise linear CDF with $\boldsymbol{F}_{X,h} \in \mathcal{S}_{\boldsymbol{F}_{X,h}}$ if the inequalities

$$\underline{\mu} \leq \sum_{j=0}^{N} \left(F^{j+1}_{X,h} - F^{j}_{X,h}\right)\left(x^j_h + \frac{h}{2}\right) \leq \overline{\mu} \tag{16}$$

are fulfilled.

– *Symmetry around* $\frac{1}{2}(\underline{x} + \overline{x})$: A piecewise linear CDF with $\boldsymbol{F}_{X,h} \in \mathcal{S}_{\boldsymbol{F}_{X,h}}$ is symmetric around $\frac{1}{2}(\underline{x} + \overline{x})$ if the equations

$$F^{j}_{X,h} = 1 - F^{N+1-i}_{X,h} \tag{17}$$

are fulfilled for $i = 1, \ldots, \frac{N}{2}$ and even N, and for $i = 1, \ldots, \frac{N+1}{2}$ and odd N.

Similar to the inequalities (8)-(10), the inequalities (15) and (16) are linear. Hence, they can be integrated into the system of linear inequalities (11), yielding another system of linear inequalities

$$\boldsymbol{A}'\boldsymbol{F}_{X,h} \leq \boldsymbol{b}', \tag{18}$$

for $\boldsymbol{F}_{X,h} \in [0,1]^N$, where $\boldsymbol{A} \in \mathbb{R}^{N' \times N}$, $\boldsymbol{b} \in \mathbb{R}^{N'}$, and $N' \geq 3N + 1$. Here, the set of all $\boldsymbol{F}_{X,h}$ which fulfill the system of linear inequalities (18) is denoted by $\mathcal{S}'_{\boldsymbol{F}_{X,h}}$, i.e.,

$$\mathcal{S}'_{\boldsymbol{F}_{X,h}} = \{\boldsymbol{F}_{X,h} \in [0,1]^N \mid \boldsymbol{A}'\boldsymbol{F}_{X,h} \leq \boldsymbol{b}'\}, \tag{19}$$

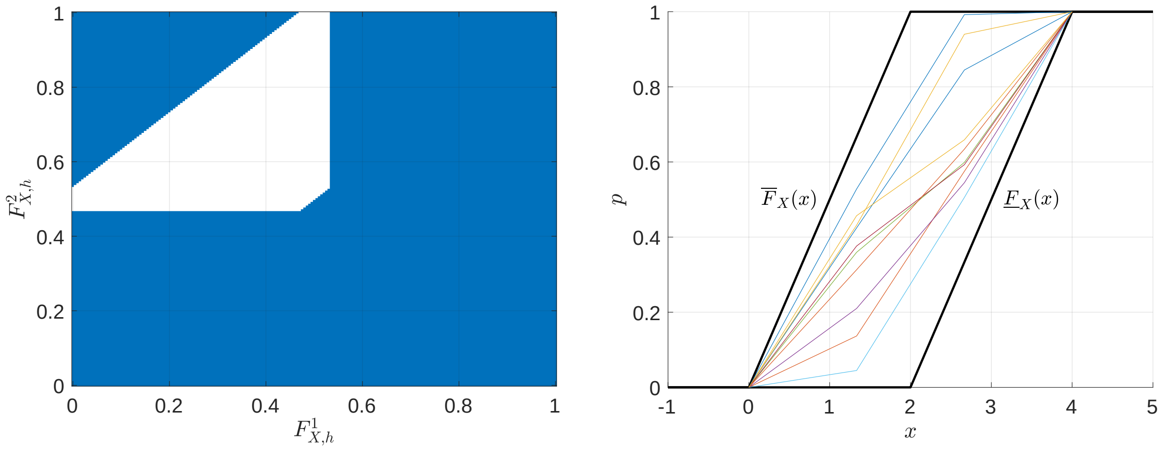


Figure 3. Constrained distribution-free p-box $[\underline{F}_X, \overline{F}_X]'_h$ for Example 2 and $N = 2$. *Left: Feasible set* $\mathcal{S}'_{\boldsymbol{F}_{X,h}}$ *(white region) and unfeasible set (blue region) for* $\boldsymbol{F}_{X,h}$. *Right: P-box bounds of* $[\underline{F}_X, \overline{F}_X]'_h$ *with examples of feasible CDFs.*

and it holds $\mathcal{S}'_{\boldsymbol{F}_{X,h}} \subseteq \mathcal{S}_{\boldsymbol{F}_{X,h}}$. Note that Equation (17) can be considered either in an additional system of linear equations, as each two inequalities in the system of linear inequalities (18), or by reducing the variables $F^j_{X,h}$, $j = 1, \ldots, N$.

Also note that a piecewise linear CDF with $\boldsymbol{F}_{X,h} \in \mathcal{S}_{\boldsymbol{F}_{X,h}}$ has an intrinsic upper bound on the PDF given by $\overline{f} = \frac{1}{h}$ because of the discretization. Accordingly, further inequalities for specific CDF properties can be defined. The set of all $F_X \in [\overline{F}_X, \underline{F}_X]_h$ with $\boldsymbol{F}_{X,h} \in \mathcal{S}'_{\boldsymbol{F}_{X,h}}$ is denoted by $[\overline{F}_X, \underline{F}_X]'_h \subseteq [\overline{F}_X, \underline{F}_X]_h$. Subsequently, the bounding CDFs of Example 1 are used to visualize constrained distribution-free p-boxes with piecewise linear CDFs.

EXAMPLE 2. *Given the bounding CDFs of Example 1, defined by Equation (14). In this example, a constrained distribution-free p-box* $[\underline{F}_X, \overline{F}_X]'$ *that accounts for an upper bound of the PDF with* $\overline{f} = 0.4$ *is considered. Again, the discretization of* $[\underline{x}, \overline{x}] = [0, 4]$ *is done by only* $N = 2$ *inner grid points, first, and the feasible CDFs of* $[\underline{F}_X, \overline{F}_X]'$ *are represented by linear CDFs* $F_X \in [\overline{F}_X, \underline{F}_X]'_h$ *defined via* $\boldsymbol{F}_{X,h} = (F^1_{X,h}, F^2_{X,h})$*. CDF examples of* $[\overline{F}_X, \underline{F}_X]'_h$ *are shown in Figure 3 together with the feasible set* $\mathcal{S}'_{\boldsymbol{F}_{X,h}}$*. Then, the same is done for* $N = 100$ *to illustrate the full capability of the approach, and CDF examples of* $[\overline{F}_X, \underline{F}_X]'_h$ *are shown in Figure 4. In comparison with Example 1, it can be seen that CDFs with large PDF values, i.e. derivative values, are excluded from* $[\underline{F}_X, \overline{F}_X]'_h$.

2.4. Propagating p-boxes

Next, it is considered how constrained distribution-free p-boxes can be propagated through a computational model $\mathcal{M}$ that maps from the n_x-dimensional input space $\mathbb{R}^{n_x}$ to the one-dimensional output space $\mathbb{R}$, i.e.,

$$\mathcal{M} : \mathbb{R}^{n_x} \to \mathbb{R}, \ (x_1, \ldots, x_{n_x}) = \boldsymbol{x} \mapsto y = \mathcal{M}(\boldsymbol{x}). \tag{20}$$

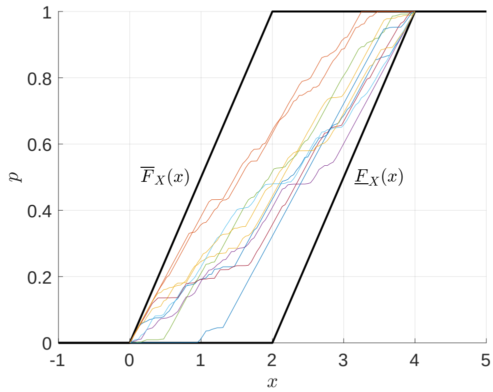


Figure 4. P-box bounds of the distribution free p-box $[\overline{F}_X, \underline{F}_X]'_h$ with examples of feasible CDFs for Example 1 and $N = 100$.

For uncertainty in the input variables x_i, represented by the random variables X_i with marginal CDFs F_{X_i}, $i = 1, \ldots, n_x$, there is uncertainty in the output variable y, too. Hence, it is $Y = \mathcal{M}(\boldsymbol{X})$, where $\boldsymbol{X}$ is the n_x-dimensional random vector comprising the real-valued random variables X_i, $i = 1, \ldots, n_x$, and Y is the real-valued random variable representing the uncertainty in the output variable. The CDF F_Y of Y depends on the joint CDF $F_{\boldsymbol{X}}$, or rather the joint PDF $f_{\boldsymbol{X}}$, of the random variables X_i, $j = 1, \ldots, n_x$, i.e.,

$$F_Y(y) = P(\mathcal{M}(X) \leq y) = \int_{\mathbb{R}} I_y(\mathcal{M}(\boldsymbol{x})) f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}x, \tag{21}$$

where Equation (2) was used. As equation (21) is in general hard to solve, numerical techniques like *Monte Carlo simulation (MCS)* can be used, see (Kalos, 2008) for an overview. Here, realizations $\boldsymbol{x}^{(j)}$, $j = 1, \ldots, M$, of a independent random sample from the CDF $F_{\boldsymbol{X}}$ of length M are required. Then, $F_Y(y)$ can be estimated as

$$F_Y(y) \approx \frac{1}{M} \sum_{j=1}^{M} I_y(\mathcal{M}(\boldsymbol{x}^{(j)})) \tag{22}$$

for $y \in \mathbb{R}$. In general, the entries of $\boldsymbol{x}^{(j)}$, denoted by $x_i^{(j)}$, $i = 1, \ldots, n_x$, can be generated by *inverse transform sampling* for $j = 1, \ldots, M$, (Kalos, 2008). For reasons of simplicity, statistical independence between the random variables X_i, $i = 1, \ldots, n_x$ is assumed in this paper. Then, the joint PDF can be computed as the product of the marginal PDFs, and realizations $u_i^{(j)}$ of an independent random sample from the standard uniform distribution of length $n_x M$ can be used to obtain

$$x_i^{(j)} = F_{X_i}^{-1}(u_i^{(j)}), \tag{23}$$

$i = 1, \ldots, n_x$, $j = 1, \ldots, M$. Note that statistical dependence could be also considered using an *isoprobabilistic transform*, see, e.g., (Torre et al., 2019).

In the context of p-boxes with unknown marginal CDFs $F_{X_i} \in [\overline{F}_{X_i}, \underline{F}_{X_i}]$, $i = 1, \dots, n_x$, the CDF F_Y is an element of a set of feasible CDFs for Y. This set can be grasped by its envelope $[\overline{F}_Y, \underline{F}_Y]$, which is a distribution-free p-box. Then, the challenge in propagating the p-boxes from the input to the output space becomes the computation of the bounding CDFs $\overline{F}_Y$ and $\underline{F}_Y$. In general, there are various methods to compute these bounds numerically if parametric or distribution-free p-boxes are considered, see (Faes et al., 2020) for an overview. However, because of the CDF constraints for constrained distribution-free p-boxes $[\overline{F}_{X_i}, \underline{F}_{X_i}]_h$, $i = 1, \dots, n_x$, they are often not transferable. In the following, an intuitive approach, based on MCS and the presented discretization with linear interpolation, is proposed.

For constrained distribution-free p-boxes $[\overline{F}_{X_i}, \underline{F}_{X_i}]'_h$, $i = 1, \dots, n_x$, the dependency of F_Y reduces to the function values of the marginal CDFs F_{X_i}, $i = 1, \dots, n_x$, at the grid points. Without loss of generality, it is assumed that they are all represented by N-dimensional tuples $\boldsymbol{F}_{X_i,h} \in \mathcal{S}'_{\boldsymbol{F}_{X_i,h}}$, $i = 1, \dots, n_x$. Thus, these N-dimensional tuples can be summarized in an $n_x N$-dimensional tuple $\boldsymbol{F}_{\boldsymbol{X},h} = (\boldsymbol{F}_{X_1,h}, \dots, \boldsymbol{F}_{X_{n_x},h}) \in \mathcal{S}'_{\boldsymbol{F}_{\boldsymbol{X},h}}$ with $\mathcal{S}'_{\boldsymbol{F}_{\boldsymbol{X},h}} = \mathcal{S}'_{\boldsymbol{F}_{X_1,h}} \times, \dots \times \mathcal{S}'_{\boldsymbol{F}_{X_{n_x},h}}$. Then, the dependency of F_Y on $\boldsymbol{F}_{\boldsymbol{X},h}$, can be expressed as $F_Y(\cdot, \boldsymbol{F}_{\boldsymbol{X},h})$ and the bounding CDFs $\overline{F}_Y$ and $\underline{F}_Y$ can be determined by

$$\overline{F}_Y(y) = \max_{\boldsymbol{F}_{\boldsymbol{X},h} \in \mathcal{S}'_{\boldsymbol{F}_{\boldsymbol{X},h}}} F_Y(y, \boldsymbol{F}_{\boldsymbol{X},h}), \tag{24}$$

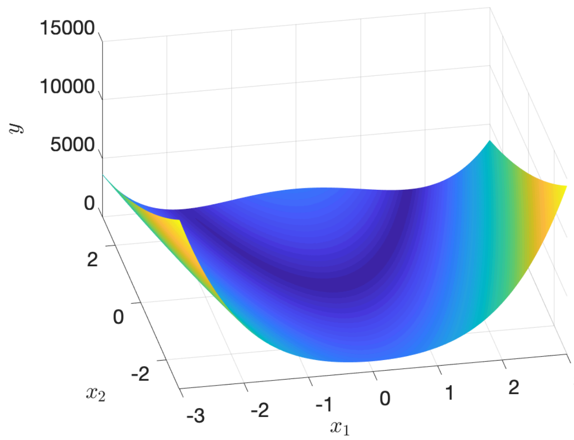
$$\underline{F}_Y(y) = \min_{\boldsymbol{F}_{\boldsymbol{X},h} \in \mathcal{S}'_{\boldsymbol{F}_{\boldsymbol{X},h}}} F_Y(y, \boldsymbol{F}_{\boldsymbol{X},h}) \tag{25}$$

for $y \in \mathbb{R}$. Computationally, these optimization problems can be treated using MCS, i.e., Equation (22), which results the optimization problem

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{F}_{\boldsymbol{X},h}} & \frac{1}{M} \sum_{j=1}^{M} I_y \left(\mathcal{M} \left((F^{-1}_{X_1}(u_1^{(j)}, \boldsymbol{F}_{X_1,h}), \dots, F^{-1}_{X_{n_x}}(u_{n_x}^{(j)}, \boldsymbol{F}_{X_{n_x},h})) \right) \right) \\ \text{subject to} & \boldsymbol{A}' \boldsymbol{F}_{\boldsymbol{X},h} \leq \boldsymbol{b}' \end{array} \tag{26}$$

for estimating $\overline{F}_Y(y)$, and the same optimization problem where "maximize" is replaced by "minimize" for estimating $\underline{F}_Y(y)$, $y \in \mathbb{R}$. Note that the quality of these estimations increases with both the number of grid points N and the number of random samples M.

Because of its non-smooth objective function and type of optimization constraints, problem (26) can be solved by generic population-based metaheuristic optimization, i.e., *evolutionary algorithms (EA)*, (De Jong, 2006). In this paper, the *genetic algorithm (GA)*, a type of EA is used, (Goldberg, 1989). Note, however, that the GA does not necessarily find a global optimum and can become computationally inefficient, in particular, when N is large. This motivates using advanced global optimization techniques for future research, like *efficient global optimization (EGO)*, (Jones et al., 1998). In the next section, the two-dimensional Rosenbrock function is considered as an example of a computational model $\mathcal{M}$. As an analytic function, it can be evaluated fast. For computational models for which this is not the case, e.g., where $\mathcal{M}$ is a black-box function, surrogate models like *Kriging* or *(sparse) polynomial chaos expansion (PCE)* can be used to approximate $\mathcal{M}$, see, e.g., (Lüthen et al., 2020; Schöbi, 2017), before solving problem (26).



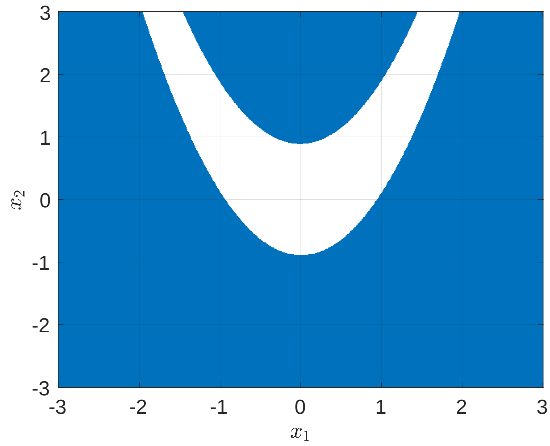


Figure 5. Two-dimensional Rosenbrock function as an example of a computational model $\mathcal{M}$. *Left:* Graph of $\mathcal{M}$ for $-3 \leq x_1, x_2 \leq 3$. *Right:* Set of $\boldsymbol{x}$ with $\mathcal{M}(\boldsymbol{x}) \leq 80$ (white region) and $\mathcal{M}(\boldsymbol{x}) > 80$ (blue region).

3. Application to the Rosenbrock function

3.1. Problem description

The two-dimensional Rosenbrock function, see (Rosenbrock, 1960), is used in (Schöbi, 2017; Schöbi and Sudret, 2017) as an example of a computational model for the propagation of distribution-free p-boxes. It reads

$$\mathcal{M}(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \tag{27}$$

for $\boldsymbol{x} \in \mathbb{R}^2$ and is visualized in Figure 5. The two-dimensional Rosenbrock function has a global minimum at $\boldsymbol{x} = (1, 1)$ with $\mathcal{M}(\boldsymbol{x}) = 0$ and a codomain of $[0, \infty)$.

The variables x_1 and x_2 are modeled as random variables X_1 and X_2, whose feasible CDFs are considered for different cases of constraint distribution-free p-boxes. The CDFs forming the envelope of a parametric p-box from a normal distribution with mean value $\mu \in [-0.5, 0.5]$ and standard deviation value $\sigma \in [0.7, 1]$ are considered as bounding CDFs, i.e.,

$$\overline{F}_{X_i}(x_i) = \max_{\mu \in [-0.5, 0.5], \sigma \in [0.7, 1]} F_{\mathcal{N}}(x_i, \mu, \sigma) = \begin{cases} F_{\mathcal{N}}(x_i, -0.5, 1) & \text{for } x \leq -0.5, \\ F_{\mathcal{N}}(x_i, -0.5, 0.7) & \text{for } x > -0.5, \end{cases} \tag{28}$$

$$\underline{F}_{X_i}(x_i) = \min_{\mu \in [-0.5, 0.5], \sigma \in [0.7, 1]} F_{\mathcal{N}}(x_i, \mu, \sigma) = \begin{cases} F_{\mathcal{N}}(x_i, 0.5, 0.7) & \text{for } x \leq 0.5, \\ F_{\mathcal{N}}(x_i, 0.5, 1) & \text{for } x > 0.5, \end{cases} \tag{29}$$

$i = 1, 2$, where $F_{\mathcal{N}}(\cdot, \mu, \sigma)$ is the CDF of a normal distribution with mean μ and standard deviation σ, see Figures 6 and 7. Note that these bounding CDFs also correspond to the ones considered in (Schöbi, 2017; Schöbi and Sudret, 2017).

Here, the following cases of constraints are considered for both F_{X_1} and F_{X_2}:

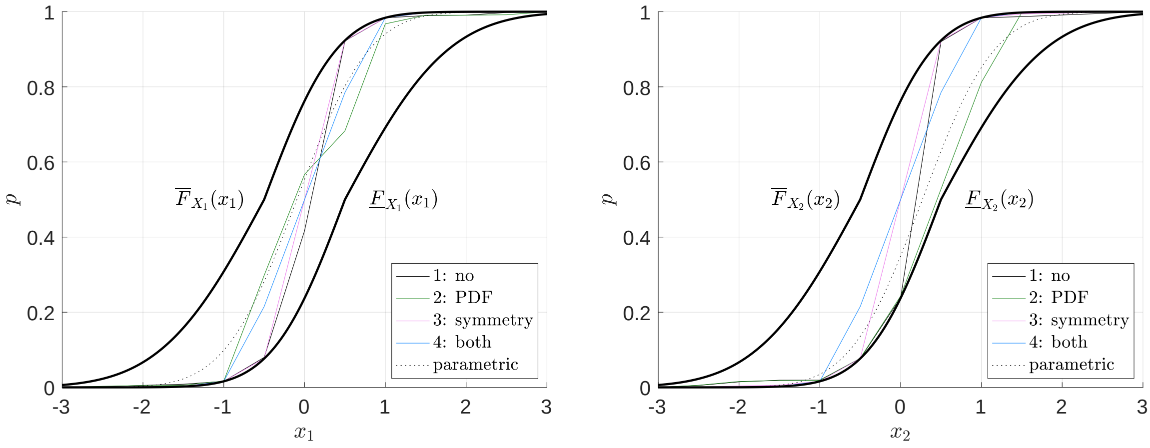


Figure 6. CDFs responsible for $\overline{F}_Y(80)$ of the constrained distribution-free p-boxes (cases 1-4) compared to the one of the parametric p-box for the Rosenbrock example. *Left:* F_{X_1}. *Right:* F_{X_2}.

1. *No additional constraints*: This case corresponds to the case of distribution-free p-box, where only the fundamental p-box constraints, i.e., inequalities (8)-(10), are considered.

2. *Bounds on the PDF*: An upper bound on the PDF is considered with

$$\overline{f} = \max_{x \in \mathbb{R}, \mu \in [-0.5, 0.5], \sigma \in [0.7, 1]} f_{\mathcal{N}}(x, \mu, \sigma) \approx 0.57,$$

i.e., inequality (15). Here, $f_{\mathcal{N}}(\cdot, \mu, \sigma)$ is the PDF of a normal distribution with mean μ and standard deviation σ. This type of constraint distribution-free p-box is a superset of the parametric p-box from the normal distribution with $\mu \in [-0.5, 0.5]$, $\sigma \in [0.7, 1]$.

3. *Symmetry around* 0: A symmetry constraint around $x_1 = x_2 = 0$ is considered, i.e., Equation (17). Hence, it follows $\mu_1 = \mu_2 = 0$, and this type of a constrained distribution-free p-box is no longer a superset of the parametric p-box with $\mu \in [-0.5, 0.5]$.

4. *Both bounds on the PDF and symmetry around* 0: The bounds on the PDF and the symmetry constraint around $x_1 = x_2 = 0$, as defined above, are considered.

Subsequently, the bounding CDFs $\overline{F}_Y$ and $\underline{F}_Y$ are computed for these cases of constrained distribution-free p-boxes using the proposed computational approach and compared to the ones of the parametric p-box from the normal distribution with $\mu \in [-0.5, 0.5]$, $\sigma \in [0.7, 1]$.

3.2. Results for $\overline{F}_Y$ and $\underline{F}_Y$

For discretizing the constraint distribution-free p-boxes, the grid $[-3, 3]_{h=0.5}$ with $N = 11$ inner grid points and $M = 10^4$ samples for the MCS are used in this paper. The results for $\overline{F}_Y(y)$ and $\underline{F}_Y(y)$, $y \in [0, 1{,}000]$, are visualized in Figure 8. Here, the case $y = 80$ is investigated in more detail.

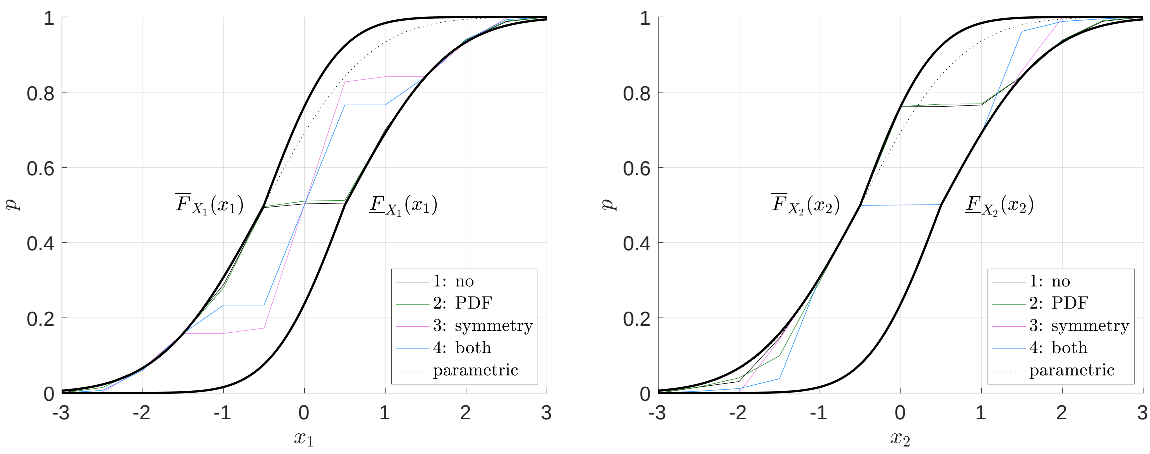


Figure 7. CCDFs responsible for $\underline{F}_Y(80)$ of the constrained distribution-free p-boxes (cases 1-4) compared to the one of the parametric p-box for the Rosenbrock example. *Left:* F_{X_1}. *Right:* F_{X_2}.

Table I. Values of $\overline{F}_Y(80)$ and $\underline{F}_Y(80)$ for the constrained distribution-free p-boxes (cases 1-4) compared to the ones of the parametric p-box for the Rosenbrock example.

Case	1 (no)	2 (PDF)	3 (symmetry)	4 (both)	parametric
$\overline{F}_Y(80)$	0.90	0.76	0.88	0.76	0.69
$\underline{F}_Y(80)$	0.16	0.16	0.21	0.22	0.33

The CDFs responsible for $\overline{F}_Y(80)$ are depicted in Figure 6 and the CDFs responsible for $\underline{F}_Y(80)$ in Figure 7. Furthermore, the numerical values of $\overline{F}_Y(80)$ and $\underline{F}_Y(80)$ are shown in Table I.

In Figure 8, it can be seen that using constrained distribution-free p-boxes (cases 2-4) compared to the distribution-free p-box with no constraints (case 1) in the input space can lead to tighter p-box bounds in the output space. For case 1, $\overline{F}_Y(y)$, $y \in [0, 200]$, raises fast, from 0 to almost 1. In contrast, $\underline{F}_Y(y)$ raises much slower and is below 0.7 for $y \in [0, 1{,}000]$. The results for cases 2-4 follow this overall trend, although in a different manner: While the upper bounding CDF $\overline{F}_Y(y)$, $y \in [0, 1{,}000]$, for case 3 follows the one for case 1 closely, the ones for cases 2 and 4 raise, with a similar rate, slower until $y \in [0, 140]$. For $y > 140$, the upper bounding CDF for case 2 tends faster to 1 than the one for case 4. The lower bounding CDFs for cases 1-4 behave similarly to their upper bounding CDFs for $y \in [0, 60]$. Then for $y > 60$, this effect reverses and $\underline{F}_Y(y)$ for case 2 follows the one for case 1 closer, than the ones for case 3 and 4. Moreover, the bounding CDFs $\overline{F}_Y(y)$ and $\underline{F}_Y(y)$ for the parametric p-box are enclosed between the bounding CDFs for all cases, $y \in [0, 1{,}000]$.

All observations for constrained distribution-free p-boxes can be explained by examining the constraints of cases 2-4 in comparison to case 1 for a specific $y \in [0, 1{,}000]$. Exemplary, this is done

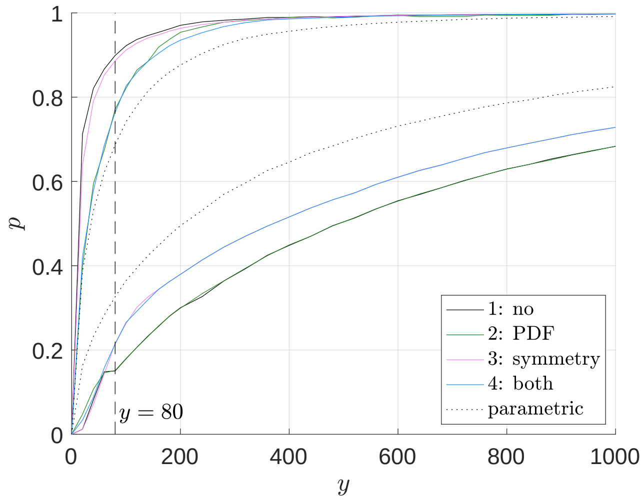


Figure 8. Graph of $\overline{F}_Y$ and $\underline{F}_Y$ of the constrained distribution-free p-boxes (cases 1-4) compared to the one of the parametric p-box for the Rosenbrock example.

for $y = 80$. Here, the CDFs of case 1 are very steep at $x_1 = 0$ and $x_2 = 0$ for $\overline{F}_Y(80)$. This cannot be achieved by cases 2 and 4 due to the bounds on the PDF. For $\underline{F}_Y(80)$, the CDFs of case 1 are moderately steep at both $x_1 \in \{-1, 1\}$ and $x_2 \in \{-1, 1\}$ with a larger focus on -1 for x_2. This cannot be achieved by cases 3 and 4 due to the symmetry around $x_2 = 0$.

Besides the errors in the estimation of $\overline{F}_Y(y)$ and $\underline{F}_Y(y)$, $y \in [0, 1{,}000]$, due to N and M, a further error can occur when the GA does not find the global optimum. In order to increase the computational efficiency for determining $\overline{F}_Y$ and $\underline{F}_Y$, this can be combined with regression methods, compare (Schöbi, 2017; Schöbi and Sudret, 2017). Furthermore, the optimization results of $\overline{F}_Y(y)$ and $\underline{F}_Y(y)$ can be improved by using advanced global optimization strategies, as discussed above. An approach for selecting the grid point non-equidistantly could also be beneficial to further improve the optimization results.

4. Conclusion

In this paper, the propagation of uncertainty modeled as probability-boxes (p-boxes) through a computational model is considered. Traditionally, p-boxes, which are defined as upper and lower bounds on cumulative distribution functions (CDFs), consider either only a specific distribution family or all CDFs between the bounds as feasible. The former are referred to as parametric p-boxes and the latter as distribution-free p-boxes. Because parametric p-boxes might lead to bounds in the output space that are too tight, and distribution-free p-boxes might lead to bounds that are too wide, constrained distribution-free p-boxes are proposed here. The idea of this new p-box type is based on distribution-free p-boxes plus a narrowing of the feasible CDFs. This is achieved by putting constraints on the CDFs, e.g., bounds on their derivatives, bounds their moments, or restrictions on their distributional shape. In order to treat constrained distribution-free p-boxes numerically, an approach, which is based on a discretization of the input space and on linear interpolation for CDFs defined at the resulting grid points, is presented. Then, the discretized p-boxes are propagated to p-boxes in the output space using Monte Carlo simulation and optimization. The effects of considering constraint distribution-free p-boxes are demonstrated using a two-dimensional analytical example. They influence the bounds of the p-box in the output space depending on the type of constraints, resulting in tighter bounds than the ones from distribution-free p-boxes but also wider than the ones from parametric p-boxes if the constraints allow. Hence, constrained distribution-free p-boxes can enhance uncertainty quantification for practical applications whenever more information than the information of bounds on the CDFs is available, but a limitation to a specific distribution is too restrictive.

For the application of constrained distribution-free p-boxes in more complex problems, it could be helpful to improve the numerical propagation by adjustment concerning, e.g., the selection of grid points, the interpolation scheme, or the optimization algorithm. Pointers on how this could be achieved are given throughout this paper.

Acknowledgements

Marco Daub acknowledges the support of the German Academic Exchange Service (DAAD) with a postdoc fellowship.

References

Augustin, T., F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes. *Introduction to imprecise probabilities.* Wiley Series in Probability and Statistics. Wiley, 2014.

Beer, M., S. Ferson, and V. Kreinovich. Imprecise probabilities in engineering analyses. *Mechanical Systems and Signal Processing*, 37(1-2):4–29, 2013.

Daub, M., S. Marelli, and B. Sudret. Uncertainty propagation with constrained distribution-free vs. generalized parametric p-boxes for engineering applications. Technical report, ETH Zürich, Zurich, Switzerland, 2021.

De Jong, K. A. *Evolutionary computation: A unified approach.* MIT Press, Cambridge, MA, United States, 2006.

Der Kiureghian, A. and O. Ditlevsen. Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112, 2009.

Faes M., M. Daub, and M. Beer. Engineering analysis with imprecise probabilities: a state-of-the-art review on p-boxes. *Proceedings of the 7th Asian-Pacific Symposium on Structural Reliability and its Applications (APSSRA2020)*, Tokyo, Japan, 2020.

Ferson, S., V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical report, Sandia National Laboratories, Albuquerque and Livermore, United States, 2003.

Fritsch, F. N. and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.

Goldberg, D. E. . *Genetic algorithms in search, optimization and machine learning.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, United States, 1989.

Jones, D. R., M. Schonlau and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455—492, 1998.

Kalos, M. H. and P. A. Whitlock. *Monte Carlo methods.* Wiley-VCH, Berlin, 2008.

Lüthen, N., S. Marelli, and B. Sudret. Sparse polynomial chaos expansions: Literature survey and benchmark. *SIAM/ASA Journal on Uncertainty Quantification* (accepted), arXiv:2002.01290.

Moustapha, M., and B. Sudret. Reliability bounds through random sets: Non-parametric methods and geotechnical applications. *Structural and Multidisciplinary Optimization*, 60:2157–2176, 2019.

Oberguggenberger, M. and W. Fellin. Reliability bounds through random sets: Non-parametric methods and geotechnical applications. *Computers & Structures*, 86(10):1093–1101, 2008.

Rosenbrock, H. H. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3):175–184, 1960.

Schöbi, R. *Surrogate models for uncertainty quantification in the context of imprecise probability modeling.* PhD thesis, ETH Zürich, Zurich, Switzerland, 2017.

Schöbi, R. and B. Sudret. Uncertainty propagation of p-boxes using sparse polynomial chaos expansions. *Journal of Computational Physics*, 339:307–327, 2017.

Soize, C. *Uncertainty Quantification.* Springer International Publishing, Cham, Switzerland, 2017.

Teixeira, R., M. Nogal, and A. O'Connor. Adaptive approaches in metamodel-based reliability analysis: A review *Structural Safety*, 102019, 2021.

Torre, E., S. Marelli, P. Embrechts, and B. Sudret. A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas *Probabilistic Engineering Mechanics*, 55:1–16, 2019.

Walley, P. *Statistical Reasoning with Imprecise Probabilities.* Chapman & Hall, London, UK, 1991.