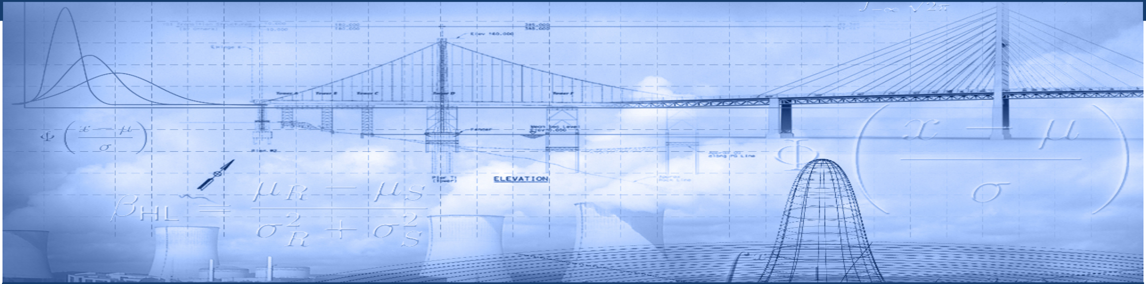
ETH zürich



On constrained distribution-free p-boxes and their propagation

Marco Daub, Stefano Marelli, Bruno Sudret

Chair of Risk, Safety and Uncertainty Quantification | ETH Zürich

Outline

Basics of distribution free p-boxes

Definitions

- X: real-valued random variable
- F_X: cumulative distribution function (CDF)

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

non-decreasing, right-continuous, and converges to 0 for $x \to -\infty$ and to 1 for $x \to \infty$

- $\mathcal{F}$: set of all CDFs on $\mathbb{R}$

Distribution free p-box

A set of possible CDFs for X is provided by a pair of an upper CDF $\overline{F}_X$ and a lower CDF $\underline{F}_X$:

$$[\overline{F}_X, \underline{F}_X] = \{F_X \in \mathcal{F} \mid \underline{F}_X(x) \leq F_X(x) \leq \overline{F}_X(x),\ x \in \mathbb{R}\} \tag{2}$$

Example: Feasible CDFs of a distribution-free p-boxes

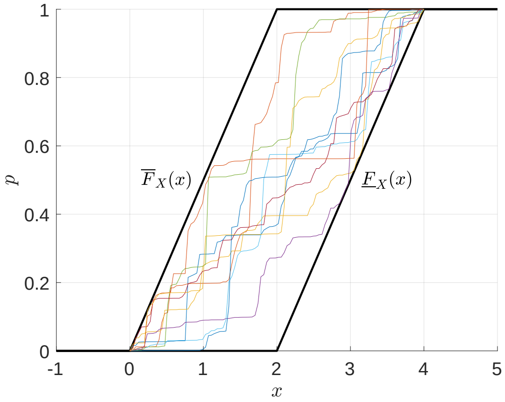


Figure: P-box bounds of a distribution free p-box $[\overline{F}_X, \underline{F}_X]$ with examples of feasible CDFs

Propagation of distribution free p-boxes considering the 2D Rosenbrock function

Input X_1 and X_2 with feasible CDFs within the envelope of normal distributions with $\mu \in [-0.5, 0.5]$ and $\sigma \in [0.7, 1]$:

$$\overline{F}_{X_i}(x_i) = \max_{\mu \in [-0.5, 0.5], \sigma \in [0.7, 1]} F_{\mathcal{N}}(x_i, \mu, \sigma) \qquad (3)$$

$$\underline{F}_{X_i}(x_i) = \min_{\mu \in [-0.5, 0.5], \sigma \in [0.7, 1]} F_{\mathcal{N}}(x_i, \mu, \sigma) \qquad (4)$$

Computational model 2D Rosenbrock function

$$\mathcal{M}(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \qquad (5)$$

Output Y with $F_Y \in [\overline{F}_Y, \underline{F}_Y]$

Example: what are the input CDFs responsible for $\overline{F}_Y(80)$ and $\underline{F}_Y(80)$?

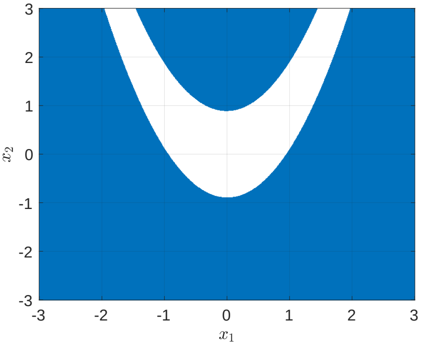
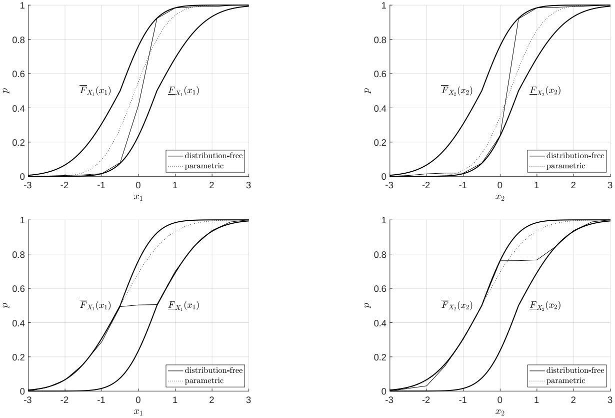


Figure: Set of $\boldsymbol{x}$ with $\mathcal{M}(\boldsymbol{x}) \leq 80$ (white region) and $\mathcal{M}(\boldsymbol{x}) > 80$ (blue region)

$\overline{F}_Y$ and $\underline{F}_Y$ can be a result of input CDFs $F_{\boldsymbol{X}} \in [\overline{F}_{\boldsymbol{X}}, \underline{F}_{\boldsymbol{X}}]$ that could be considered unlikely for specific engineering applications

$$\overline{F}_Y(80) = 0.90$$

$$\underline{F}_Y(80) = 0.16$$



Outline

Limitations of distribution-free p-boxes

Introducing constrained distribution-free p-boxes

Basics of constrained distribution free p-boxes

Idea

- Constrain the feasible CDFs of a distribution-free p-box in a non-parametric way
- For example, by upper and lower bounds for the moments of X, the PDF f_X, or further derivatives
- For example, by constraints on the distributional shape, e.g., for symmetry or unimodality

Constrained distribution free p-box

The distribution free p-box is intersected with a set of feasible CDF $\mathcal{F}' \subseteq \mathcal{F}$ that only comprises CDFs fulfilling additional constraints:

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

Note: There are some approaches in literature that could be considered as constrained distribution free p-boxes. However, appropriate techniques on how to handle additional p-box constraints computationally are lacking.

Distribution free p-boxes vs. constrained distribution free p-boxes

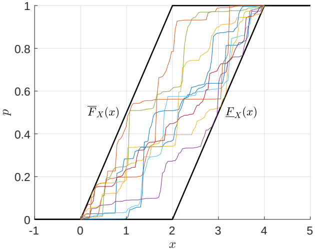


Figure: P-box bounds of a distribution free p-box $[\overline{F}_X, \underline{F}_X]$ with examples of feasible CDFs

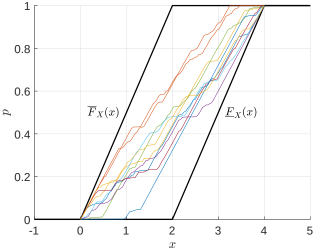


Figure: P-box bounds of a constrained distribution free p-box $[\overline{F}_X, \underline{F}_X]'$ (upper bound on the PDF) with examples of feasible CDFs

Numerics with constrained distribution-free p-boxes

Idea

- An (equidistant) grid on the interval $[\underline{x}, \overline{x}]$ with N inner grid points x_h^j is considered
- $F_{X,h}^j$ denote the values of F_X at the grid points (assume $F_{X,h}^0 = 0$ and $F_{X,h}^{N+1} = 1$)
- To guarantee that F_X is a proper CDF which is an element of the p-box $[\overline{F}_X, \underline{F}_X]$, $F_{X,h}^j$ must fulfill

$$
\begin{aligned}
F_{X,h}^j &\leq F_{X,h}^{j+1}, && (7)\\
F_{X,h}^j &\leq \overline{F}_X(x_h^j), && (8)\\
F_{X,h}^j &\geq \underline{F}_X(x_h^j), && (9)
\end{aligned}
$$

- As a simple approach, linear interpolation is used for the values of F_X at non-grid points
- Additional constraints, like a lower bound $\underline{f} \in \mathbb{R}$ and an upper bound $\overline{f} \in \mathbb{R}$ on the PDF can be introduced as an additional inequality

$$\underline{f} \leq \frac{F_{X,h}^{j+1} - F_{X,h}^j}{h} \leq \overline{f} \tag{10}$$

Propagating constrained distribution-free p-boxes

Approach

- Multiple random variables X_i are considered (assuming statistical independence)
- When propagating X_i through a computational model $\mathcal{M}$, F_Y is determined by the function values of the marginal CDFs F_{X_i} at the grid points and hence:

$$\overline{F}_Y(y) = \max_{\boldsymbol{F}_{\boldsymbol{X},h} \in \mathcal{S}'_{\boldsymbol{F}_{\boldsymbol{X},h}}} F_Y(y, \boldsymbol{F}_{\boldsymbol{X},h}), \tag{11}$$

$$\underline{F}_Y(y) = \min_{\boldsymbol{F}_{\boldsymbol{X},h} \in \mathcal{S}'_{\boldsymbol{F}_{\boldsymbol{X},h}}} F_Y(y, \boldsymbol{F}_{\boldsymbol{X},h}), \tag{12}$$

  where $\mathcal{S}'_{\boldsymbol{F}_{\boldsymbol{X},h}}$ is the set of $F^j_{X,h}$ fulfilling inequalities (7)-(10)
- Monte Carlo simulation (MCS) with an inverse transform sampling is used for propagation *(for more advanced methods see: MGR Faes et al., Engineering analysis with probability boxes: a review on computational methods)*
- The resulting optimization problems are solved by a genetic algorithm

Back to the 2D Rosenbrock problem with constrained distribution-free p-boxes

Additional constraints for both F_{X_1} and F_{X_2} are introduced:

1. *No additional constraints*: This case corresponds to the case of distribution-free p-box (and was considered before).
2. *Bounds on the PDF*: An upper bound on the PDF is considered with

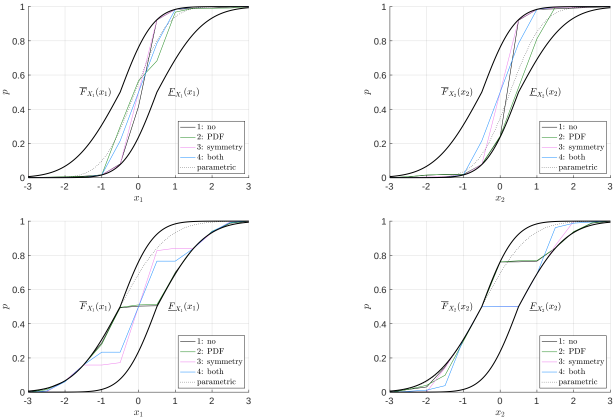
$$\overline{f} = \max_{x \in \mathbb{R}, \mu \in [-0.5, 0.5], \sigma \in [0.7, 1]} f_{\mathcal{N}}(x, \mu, \sigma) \approx 0.57,$$

   Here, $f_{\mathcal{N}}(\cdot, \mu, \sigma)$ is the PDF of a normal distribution with mean μ and standard deviation σ.
3. *Symmetry around* 0: A symmetry constraint around $x_1 = x_2 = 0$ is considered.
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.

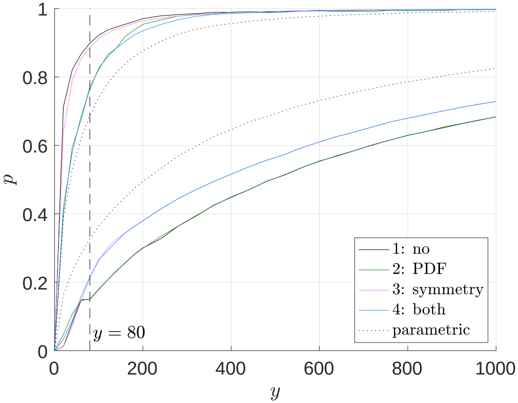
Input CDFs responsible for $\overline{F}_Y(80)$ and $\underline{F}_Y(80)$ (cases 1-4)

$\overline{F}_Y(80)$			
1 (no)	2 (PDF)	3 (sym)	4 (both)
0.90	0.76	0.88	0.76

$\underline{F}_Y(80)$			
1 (no)	2 (PDF)	3 (sym)	4 (both)
0.16	0.16	0.21	0.22



Graph of $\overline{F}_Y$ and $\underline{F}_Y$ of the constrained distribution-free p-boxes (cases 1-4)



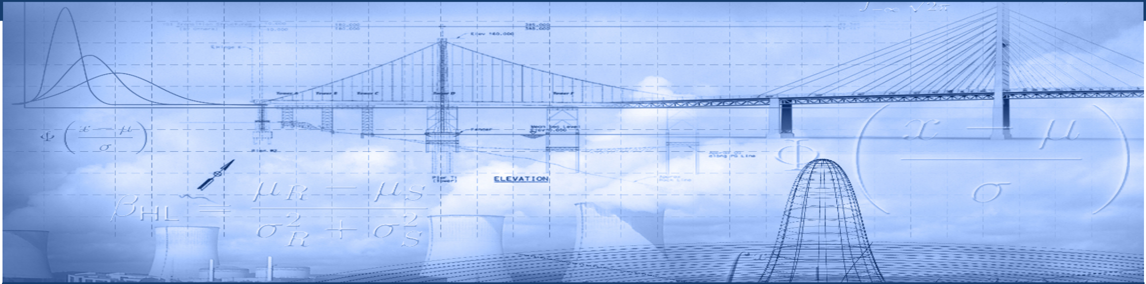
Conclusion

Recap:

- A new p-box type, the constrained distribution-free p-box, which is based on distribution-free p-boxes plus a narrowing of the feasible CDFs, was proposed
- An approach, which is based on a discretization of the input space and on linear interpolation, was presented to treat them numerically
- The discretized p-boxes were propagated to p-boxes in the output space using Monte Carlo simulation and optimization

Key takeaways:

- Constraint distribution-free p-boxes lead to bounds of the p-box in the output space that might be tighter than the ones from distribution-free p-boxes but also wider than the ones from parametric p-boxes
- 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



Thank you for your attention!