

# A generalized numerical framework of imprecise probability to propagate epistemic uncertainty

Marco de Angelis, Edoardo Patelli and Michael Beer

*Institute for Risk and Uncertainty, University of Liverpool, UK, marco.de-angelis@liverpool.ac.uk*

**Abstract.** A generalized numerical framework is presented for constructing computational models capable of processing inputs defined as sets of probability distribution functions and sets of intervals. The framework implements a novel solution strategy that couples advanced sampling-based methods and optimization procedures, and provides a credible tool for calculating imprecise measure of failure probability. In this paper, the tool is utilized to perform epistemic uncertainty propagation and to identify the extreme case realizations leading to the bounding values of the failure probability. It has to be noted that the proposed strategy, is insensitive both to the dimension of the problem and to the targeted failure probability, so far as the performance function displays a single failure mode. It is shown by means of examples that the numerical tool is significantly more efficient than a naive approach to the problem of epistemic uncertainty propagation.

**Keywords:** Imprecise probability, Structural reliability, Epistemic uncertainty propagation, Extreme case realizations, Credal sets, Bounded sets

## 1. Introduction

In performance-based engineering decisions often rely on the response of a computational model (Augusti and Ciampoli, 2008). Most often, however, due to insufficient knowledge about the system, also referred to as epistemic uncertainty, it is not possible to create a definite map of values for the inputs of the computational model. In this context assuming a specific probability distribution model can be a strong assumption leading to a possibly wrong decision (Beer and Ferson and Kreinovich, 2013).

In structural reliability assessment, the failure probability, denoted as  $p_F$ , represents the most important quantity. It is of interest computing the effect of epistemic uncertainty on the failure probability and making the least amount of assumptions. This requires the epistemic uncertainty to be propagated throughout the model and consequently quantified in terms of failure probability intervals. Uncertainty propagation can be performed by means of different strategies, but mainly two approaches can be adopted: i) the parametric approach founded on the theory of imprecise probability (Walley, 2013), ii) the non-parametric approach described by the random set theory (Alvarez, 2006). The parametric approach defines sets of probability distribution functions, also known as credal sets (Zaffalon, 2013), and sets of bounded variables or intervals (Ferson et al., 2007; Moens and Vandepitte, 2005), while the non-parametric approach uses only bounding CDFs and copula models, also known as p-boxes (Ferson et al., 2002). Here, the parametric approach to propagate epistemic uncertainty for the failure probability is investigated. The advantages of using

such an approach are manifold, but mainly can be attributed to its efficiency when small target values of failure probability are considered and large scale problems are involved. Limitations of using this approach can also be identified.

## 2. Parametric models for the uncertainty propagation of failure probability

The requirement of treating the epistemic uncertainties in a parametric sense and without making any kind of assumptions, leads to the consideration of bounded sets and credal sets. In a parametric uncertainty model, probability distribution functions are not fixed, but can be chosen from among a set of options. An example can be a set of distribution models, such as Normal, Lognormal, or Beta, with mean 4 and standard deviation 1, or a set of Normal distribution functions with mean in the interval  $[2, 4]$  and standard deviation in the interval  $[0.5, 2]$ .

A bounded set is a set obtained from a sequence of intervals put together by means of dependence functions. When no dependence is defined the bounded set is simply obtained by the Cartesian product of the intervals. A bounded set can also be used for the parameters of a probability distribution model. In this case the probability distribution model is represented by a set of distribution functions (or credal set), where every realization in the bounded set corresponds to only one distribution function. By properly defining these bounded sets, a general framework for uncertainty propagation of failure probability can be identified. Within this framework the uncertainty propagation consists in seeking the minimum and maximum failure probability within these bounded sets.

### 2.1. TRADITIONAL ASSESSMENT OF RELIABILITY BY MEANS OF FAILURE PROBABILITY

In performance-based engineering, the structural system is considered as a collection of performances  $g_i, i = 1, 2, \dots, N_g$ , which are functions of the state variables  $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^n$  (see e.g. (Valdebenito et al., 2014)). Typically, the state variables are the inputs that define the structural system, such as material strength and stiffness, shape and size of structural elements, load magnitudes, etc. The output of the system is identified as specific structural responses, such as frequency and amplitude of vibrations, stresses, deflections and so forth. The performance of the system is obtained comparing single responses against corresponding capacities. If the capacities of the system are included among the state variables, it becomes clear how the performance can be expressed as a function of the state variables only.

The performance function  $g : \mathbb{R}^n \rightarrow g_i \in \mathbb{R}$  maps values from the state space  $\Theta$  to the performance variables of interest. For given criteria on the performance variables,  $g$  defines the failure domain  $\Theta_F = \{\boldsymbol{\theta} \in \Theta \mid g(\boldsymbol{\theta}) \leq 0\}$ , which is identified by the limit state surface  $\tilde{\Theta} = \{\boldsymbol{\theta} \mid g(\boldsymbol{\theta}) = 0\}$ . Points  $\tilde{\boldsymbol{\theta}}$  on the limit state surface are referred to as limit state points. The performance function provides a measure of how far from critical is the state of the system, and in this sense it can also be understood as a safety margin.

An important feature for our development is that the limit state is invariant to the uncertainty set  $\mathcal{M}$ , because the limit state is intrinsic to the structural system, i.e. depends solely on the performance function  $g$ , which in turns is a function of the state variables only. The uncertainty model only determines the probability over the state space, but does not influence the location of limit state points  $\tilde{\boldsymbol{\theta}}$ .

Traditionally, the assessment of structural reliability is based on well-defined (precise) probabilistic models. In this context, the state variables  $\boldsymbol{\theta}$  are all characterized by definite probability functions and the failure probability  $p_F$  can be expressed as

$$p_F = \int_{\Theta_F} h_{\mathcal{D}}(\boldsymbol{\theta}; \mathbf{p}) d\Theta, \quad (1)$$

where,  $h_{\mathcal{D}}$  is the joint probability distribution function of distributional model  $\mathcal{D}$ ,  $\mathbf{p}$  are the distribution parameters that define the probability distribution function, and  $d\Theta$  is the Lebesgue measure of an elementary portion of  $\Theta$ . The computation of Equation (1) can be associated with quite a significant numerical effort. For this reason, mainly advanced sampling-based methods, such as Directional Sampling (Ditlevsen et al., 1988), Advanced Line Sampling (de Angelis and Patelli and Beer, 2014) or Subset Simulation (Au and Beck, 2001) are used to compute  $p_F$ . In general, sampling-based methods modify the integral of Equation (1) as  $p_F = \int_{-\infty}^{\infty} \mathcal{I}_F(\boldsymbol{\theta}) h_{\mathcal{D}}(\boldsymbol{\theta}; \mathbf{p}) d\boldsymbol{\theta}$ ; where,  $\mathcal{I}_F : \mathbb{R}^n \rightarrow \{0, 1\}$  is the indicator function that is equal to 1 if  $\boldsymbol{\theta} \in \Theta_F$  and 0 otherwise.

## 2.2. PROPOSED GENERALIZED UNCERTAINTY FRAMEWORK

Within the generalized uncertainty framework, type and extent of uncertainty in the state variables  $\boldsymbol{\theta}$  is defined by the set  $\mathcal{M}$ . The uncertainty set  $\mathcal{M}$  is obtained making the union of a credal set, denoted by  $\mathcal{C}$ , and a bounded set, denoted by  $\mathcal{B}$ , as  $\mathcal{M} = \mathcal{C} \cup \mathcal{B}$ . In order to proceed to a formal definition of bounded and credal sets, first a clear distinction between imprecise random variables  $\boldsymbol{\xi} \in \mathcal{C}$  and intervals  $\mathbf{x} \in \mathcal{B}$  shall be made. The reliability problem is, thus, reformulated to allow for imprecision. The state space of the variables  $\boldsymbol{\theta} \in \mathbb{R}^n$  is split into two independent spaces: namely the space of imprecise random variables  $\boldsymbol{\xi} \in \Omega \subseteq \mathbb{R}^{n_\xi}$  and the space of intervals  $\mathbf{x} \in X \subseteq \mathbb{R}^{n_x}$ , where  $\Theta = \Omega \times X$  and  $n_\xi + n_x = n$ . The space of the intervals is defined by the bounded set  $\mathcal{B}_x = \times_{i=1}^{n_x} [\underline{x}_i, \bar{x}_i]$  obtained from the Cartesian product of the intervals  $[\underline{x}_i, \bar{x}_i] = \bar{x}_i$ , where  $\bar{x}_i > \underline{x}_i$ . Another bounded set  $\mathcal{B}_\xi = \times_{i=1}^{n_\xi} [p_i, \bar{p}_i]$  is defined to model imprecision in the distribution parameters  $\mathbf{p}$ , which in turns builds up the credal set  $\mathcal{C} = \{h_{\mathcal{D}}(\boldsymbol{\xi}; \mathbf{p}) \mid \mathbf{p} \in \mathcal{B}_\xi\}$ . In words,  $\mathcal{C}$  is defined as the set of all probability distribution functions where the distribution parameters range within the set  $\mathcal{B}_\xi$ . As an example, consider the imprecise random variable obtained from a Normal distribution with mean in the interval  $[2, 4]$  and standard deviation in the interval  $[0.5, 2]$ . This imprecise random variable is defined by the credal set  $\{h_{\mathcal{D}}(\boldsymbol{\xi}; \mu, \sigma) \mid \mu \in [2, 4], \sigma \in [0.5, 2]\}$ , and the bounded set  $\mathcal{B}_\xi$  is simply given as  $\mathcal{B} = [2, 4] \times [0.5, 2]$ .

*2.2.0.1. Remarks on dependency among state variables* Dependence amid state variables within the proposed framework can be modelled in a straightforward way. The between random variables can be taken into account by means of a covariance model characterizing the joint probability distribution function or more conveniently by means of a copula function. This does not make the formulation of the problem any harder, because the credal set can still be defined as previously by simply adding to the bounded set also the parameters corresponding to the covariance or copula models. The dependence between the intervals is taken into account by means of dependence functions  $\Phi(\mathbf{x})$  that become part of the bounded set  $\mathcal{B}_x$ . When a dependence function is defined for the intervals, the bounded set  $\mathcal{B}_x$  is no longer obtained by the Cartesian product  $\times_{i=1}^{n_x} \bar{x}_i$ . Often

the dependence functions are defined as a transformation  $\Phi : x \in \mathbb{R}^{n_x} \rightarrow y \in \mathbb{R}^{n_x}$ , where new bounds  $[y_i, \bar{y}_i] = \bar{y}_i$ ,  $i = 1, \dots, n_x$  can be identified. The bounded set thus, can be reformulated as  $\mathcal{B}_x = \left\{ \mathbf{x} \mid \Phi(\mathbf{x}) \in \times_{i=1}^{n_x} \bar{y}_i \right\}$ . Note that this formulation is quite convenient because the search in the bounded sets for the minimum/maximum of the failure probability can be easily operated as a bounded non-constrained optimization.

### 2.3. FAILURE PROBABILITY IN THE GENERALIZED UNCERTAINTY FRAMEWORK

When the uncertainty model comprises only precisely defined probability distributions, i.e.  $n_x = 0$  and  $\mathcal{C}$  degenerates in one distribution function, structural reliability is assessed in terms of a precise failure probability. Imprecise measures of failure probability can be obtained considering the uncertainty set  $\mathcal{M}$ . The failure domain  $\Theta_F$  is now made up of two failure domains as  $\Theta_F = \Omega_F \times X_F$ , where  $\Omega_F(\mathbf{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^{n_\xi} \mid g(\boldsymbol{\xi}, \mathbf{x}) \leq 0\}$  and  $X_F(\boldsymbol{\xi}) = \{\mathbf{x} \in \mathbb{R}^{n_x} \mid g(\boldsymbol{\xi}, \mathbf{x}) \leq 0\}$ . Note that these two domains depend one another through the performance function  $g(\boldsymbol{\xi}, \mathbf{x})$  that is defined over the whole state space  $\Theta$ . Provided the definition of  $\mathcal{C}$ , the imprecise failure probability can be expressed as the interval  $\bar{p}_F(\mathcal{C}, \mathcal{B}_x) = \left[ \underline{p}_F(\mathcal{C}, \mathcal{B}_x), \bar{p}_F(\mathcal{C}, \mathcal{B}_x) \right]$ , where the lower and upper bound of the imprecise failure probability are

$$\underline{p}_F(\mathcal{C}, \mathcal{B}_x) = \inf_{\mathbf{x} \in \mathcal{B}_x} \inf_{\mathbf{p} \in \mathcal{B}_\xi} \int_{\Omega_F(\mathbf{x})} h_{\mathcal{D}}(\boldsymbol{\xi}; \mathbf{p}) d\Omega; \quad \bar{p}_F(\mathcal{C}, \mathcal{B}_x) = \sup_{\mathbf{x} \in \mathcal{B}_x} \sup_{\mathbf{p} \in \mathcal{B}_\xi} \int_{\Omega_F(\mathbf{x})} h_{\mathcal{D}}(\boldsymbol{\xi}; \mathbf{p}) d\Omega, \quad (2)$$

where, the order to which the operations of infimum and supremum are performed can be changed. The inner operand searches the bounds of  $p_F$  within  $\mathcal{C}$ , while the outer one searches the bounds of  $p_F$  within  $\mathcal{B}_x$ . Upper and lower bounds of failure and survival probabilities show a dual relationship. This can be seen clearly in the special case that the uncertainty model is restricted to  $\mathcal{C}$  only. The probability function  $h_{\mathcal{D}}^\circ$  that yields the lower bound  $\underline{p}_F = \underline{p}(\Omega_F)$ , satisfies the equation

$$\int_{\Omega_F} h_{\mathcal{D}}^\circ(\boldsymbol{\xi}) d\Omega + \int_{\Omega_S} h_{\mathcal{D}}^\circ(\boldsymbol{\xi}) d\Omega = 1, \quad (3)$$

where  $\Omega_S$  denotes the survival domain, and  $\Omega_S \cup \Omega_F = \Omega$ . Therefore,  $h^\circ$  is also the function for which the upper bound  $\bar{p}(\Omega_S)$  is obtained. Thus, the Equation  $\underline{p}(\Omega_F) = 1 - \bar{p}(\Omega_S)$  establishes a *conjugate* (or dual) relationship between lower and upper probability functions. This relationship allows to identify the upper probability function when the lower probability function is known and vice versa. Note, however, that the complete function, which may also have an infinite support, is needed in order for the relationship to be used. From the definition of lower and upper probability follows that  $\underline{p}_F \leq \bar{p}_F$ . When  $\mathcal{C}$  degenerates into a single probability distribution function, precise measures of probability  $p_F = \underline{p}_F = \bar{p}_F$  are obtained.

### 3. Sampling-based estimation of interval failure probability

When imprecision is considered, the failure probability is obtained as interval  $\overline{p_F}$ . In order to calculate the bounds of the failure probability, a global search in the bounded sets  $\mathcal{B}_\xi$  and  $\mathcal{B}_x$  is performed. Each step of the search procedure requires the estimation of a failure probability. A naive approach to the problem for searching in the above sets would be prohibitive in the majority of cases due to the numerical effort incurred. In fact, two nested loops are required, where the inner loop estimates the failure probability and the outer loop searches for its bounds. In this section a sampling-based method called Advanced Line Sampling (ALS) (de Angelis and Patelli and Beer, 2014) is proposed. Compared to Monte Carlo, ALS is far more efficient and is insensitive to the failure probability target. Moreover, ALS can be exploited to make the search in the bounded sets several orders of magnitude faster. This applies when the performance function displays a single failure mode thus an averaged important direction can be established in the original state space.

#### 3.1. THE GLOBAL SEARCH FOR LOWER AND UPPER FAILURE PROBABILITIES

The objective function for the global search in the sets  $\mathcal{B}_\xi$  and  $\mathcal{B}_x$  is the failure probability. In order to identify an approximation of the  $p_F$  bounds, the minimum and maximum values of  $p_F$  within the two bounded sets are sought.

##### 3.1.1. *The search in the bounded set $\mathcal{B}_\xi$ of distribution parameters*

The set  $\mathcal{B}_\xi$  of distribution parameters defines the set of all probability distribution functions to be considered in the analysis. Any element of  $\mathcal{B}_\xi$  is associated with a different value of failure probability. Nonetheless, the limit state in the original space  $\Theta$  does not change as we search in  $\mathcal{B}_\xi$ . This is because the limit state depends upon the structural system and not upon the uncertainty model that defines the probability distribution over the state variables. Since the important direction, denoted as  $\alpha$ , is defined as any direction pointing towards the failure domain, during the search in  $\mathcal{B}_\xi$ , an averaged  $\alpha$  can be set for the entire analysis, independently from the distribution functions of the random variables. Changing the distribution functions modifies the location of the most probable point on the limit state surface. Hence, the direction  $\alpha$ , set at beginning of the analysis, might not be the optimal one for all the distributions analysed. However, it has to be noted that in Line Sampling it is not required the important direction to be pointing precisely towards the design point, even though sometimes this may have an influence on the accuracy of the estimation. In order to ensure high accuracy, Advanced Line Sampling implements an adaptive algorithm capable of updating to better directions.

Once the important direction is defined, an estimation of  $p_F$  would require just few runs of the performance function. Moreover, the signs of the important direction, as it stands in the original space, allow to identify the corners of the hypercube where the state space is nearest and furthest from the limit state. These two corners identify lower and upper conjugate states, where the search for minimum/maximum failure probability and can be intensified. Using the information about the conjugate states, it becomes clear how searching for the bounds is now sensibly easier, having restricted the search domain to just two limited regions of the state space.

### 3.1.2. The search in the bounded set $\mathcal{B}_x$ of structural parameters

Imprecision of structural parameters, characterized by the bounded set  $\mathcal{B}_x$ , requires an extension of the procedure developed so far. In fact, the bounded variables  $\mathbf{x} \in \mathbb{R}^{n_x}$  change the shape of the limit state boundary, which needs to be addressed with a second search as described in Equation (2). In this section, we propose a strategy to include the variables  $\mathbf{x} \in \mathcal{B}_x$  in the numerical framework presented so far. The strategy consists of an extension to an augmented probability space, where the interval variables are treated as dummy normal random variables having imprecise mean values and fixed standard deviations. In simple terms, this permits a combined consideration of the bounded set  $\mathcal{B}_x$  together with the bounded set  $\mathcal{B}_\xi$  in the same manner. Each dummy imprecise random variable has an interval mean value  $\overline{\mu_x} = \overline{x}$ , and a real-valued standard deviation  $\sigma_x$  to be fixed with some convenient value. By defining these dummy imprecise random variables a thorough search can be performed in both sets  $\mathcal{B}_x$  and  $\mathcal{B}_\xi$  simultaneously. The only requirement for the dummy imprecise random variables is that the chosen value of  $\sigma_x$  should neither be too large nor be too small to avoid numerical issues in computing the failure probability. The standard deviation  $\sigma_x$  can be set, for example, as a fraction of the interval radius  $\sigma_x = k(\overline{x} - \underline{x})/2$ , where  $k$  can be any value between 0 and 1. Once the argument optima in the sets  $\mathcal{B}_\xi$  and  $\mathcal{B}_x$  are found, the associated bounds on the failure probability are also known. Two more reliability analyses at the end of the search, run on the argument optima, will be needed to find the failure probability bounds. Note that during this procedure sampling outside the intervals may occur. However, points outside the intervals are solely used to drive the search process. In cases where the physical model restricts the evaluation to the range of the intervals, truncated normal random variables are used as the dummy imprecise variables, which lower and upper limits are equal to the endpoints of the intervals.

## 4. Numerical examples

To show the applicability and efficiency of the proposed method two examples are presented. In the first example an *explicit* performance function is considered to compare the proposed method against a naive approach. In the second example an *implicit* performance function obtained from a large scale finite element model is analyzed. This second example demonstrates the efficiency of the proposed method on large problems involving several state variables and small target probabilities.

### 4.1. LINEAR PERFORMANCE FUNCTION WITH NOISE

This example is solved both with the proposed method, namely approach A, and with a naive approach, or approach B. Approach A computes the interval probability  $\overline{p}_F$  by identifying the conjugate states using the information of an averaged important direction defined in the original state space. Note that when the distributions are given in terms of moments the conjugate states coincide with the corners of the search domain delimited by  $\mathcal{B}_\xi \cup \mathcal{B}_x$ . Approach B computes  $\overline{p}_F$  by blindly searching in the above bounded sets. The search is driven as an optimization process that looks at both minimum and maximum of the failure probability. A blind search is effective only when the number of search variables is small (less than 5), thus, the example is solved twice: first, case (a), 4 state variables are considered as imprecise random variable and 2 as intervals for a total

of 10 search variables, and second, case (b), just one imprecise random variable and one interval are considered for a total of 3 search variables.

4.1.0.1. *Case (a): performance function with 4 imprecise random variables and 2 intervals* Here, we analyse the performance function (Grooteman, 2011; Harbitz, 1986)

$$g(\boldsymbol{\xi}, \mathbf{x}) = -200 + \xi_1 + 2\xi_2 + 2\xi_3 + \xi_4 - 5x_1 - 5x_2 + 0.001 \sum_{i=1}^4 \sin(100\xi_i) + 0.001 \sum_{j=1}^2 \sin(100x_j); \quad (4)$$

where, the state variables  $\boldsymbol{\theta} = (\boldsymbol{\xi}, \mathbf{x})$  are defined as in table I. An averaged important direction can be identified computing the gradient  $\nabla g' = -(1, 2, 2, 1, -5, -5)$ , where  $g'$  is obtained taking off the noise from Equation 4. With approach A, the sign vector of the important direction  $\text{sign}(\boldsymbol{\alpha}) = (-1, -1, -1, -1, 1, 1)$  identifies the following conjugate states  $\boldsymbol{\theta}_{p_F} = (\overline{\mu_1}, \overline{\mu_2}, \overline{\mu_3}, \overline{\mu_4}, \underline{x_1}, \underline{x_2})$ , and  $\boldsymbol{\theta}_{\overline{p_F}} = (\underline{\mu_1}, \underline{\mu_2}, \underline{\mu_3}, \underline{\mu_4}, \overline{x_1}, \overline{x_2})$ . Since the probability distributions are defined in terms of moments, and no correlation is defined amongst the variables, the minimum and maximum of the failure probability is attained at the corners of the bounded sets, where the standard deviation of the state variables is minimum and maximum respectively. With this approach the argument optima for this problem can be identified as follows:

$$\arg \min_{\mathbf{p} \in \mathcal{B}_\xi, \mathbf{x} \in \mathcal{B}_x} p_F = (\overline{\mu_1}, \overline{\sigma_1}, \overline{\mu_2}, \overline{\sigma_2}, \overline{\mu_3}, \overline{\sigma_3}, \overline{\mu_4}, \overline{\sigma_4}, \underline{x_1}, \underline{x_2}) \quad (5)$$

$$\arg \max_{\mathbf{p} \in \mathcal{B}_\xi, \mathbf{x} \in \mathcal{B}_x} p_F = (\underline{\mu_1}, \underline{\sigma_1}, \underline{\mu_2}, \underline{\sigma_2}, \underline{\mu_3}, \underline{\sigma_3}, \underline{\mu_4}, \underline{\sigma_4}, \overline{x_1}, \overline{x_2}). \quad (6)$$

Approach B provides only an approximation of the solution just found. In fact, a global optimization procedure requires several thousands of function evaluations to be accurate. Moreover, the larger the scale of the problem in terms of search variables, the more evaluations are needed. Here, where the search domain has 10 dimensions, at least 1024 iterations are required to obtain an estimate. The search is driven by a Latin Hypercube Sampling (McKay and Beckman and Conover, 1979) scheme that allows to select the critical points in the search domain. Results from the two approaches are shown in Table II, where the significance of the approximation introduced by approach B can be appreciated.

4.1.0.2. *Case (b): performance function with 1 imprecise random variable and 1 interval* To demonstrate that the approximation introduced by approach B is due to the number of variables in the search domain, the same example is solved with just one imprecise random variable and one interval. Here, the first state variable is  $\theta_1 = x$  and the sixth state variable is  $\theta_6 = \xi$ ; all of the other state variables are precise random variables. The performance function now is:

$$g(\boldsymbol{\theta}, \xi, x) = -200 + x + 2\theta_2 + 2\theta_3 + \theta_4 + 5\theta_5 - 5\xi + 0.001(\sin(100x) + \sin(100\xi) + \sum_{i=2}^5 \sin(100\theta_i)); \quad (7)$$

Table I. Parametric uncertainties for the input state variables (SV) of Case (a)

SV #	Symbol	Uncert. type	Mean/Interval	Stand. dev.
1	$\xi_1$	$\text{LN}(\underline{\mu}_1, \underline{\sigma}_1)$	$\underline{\mu}_1 = [110, 125]$	$\underline{\sigma}_1 = [10, 14]$
2	$\xi_2$	$\text{LN}(\underline{\mu}_2, \underline{\sigma}_2)$	$\underline{\mu}_2 = [115, 130]$	$\underline{\sigma}_2 = [10, 14]$
3	$\xi_3$	$\text{LN}(\underline{\mu}_3, \underline{\sigma}_3)$	$\underline{\mu}_3 = [115, 130]$	$\underline{\sigma}_3 = [10, 14]$
4	$\xi_4$	$\text{LN}(\underline{\mu}_4, \underline{\sigma}_4)$	$\underline{\mu}_4 = [115, 130]$	$\underline{\sigma}_4 = [10, 14]$
5	$x_1$	Interval $\underline{x}_1$	$\underline{x}_1 = [45, 52]$	-
6	$x_2$	Interval $\underline{x}_2$	$\underline{x}_2 = [35, 43]$	-

Table II. Case (b): comparison of results in terms of failure probability and total number of samples required

Approach A		Approach B (LHS)	
$\underline{p}_F$	$N_s$	$\underline{p}_F$	$N_s$
[1.4 10 <sup>-10</sup> , 0.43]	252	[3.2 10 <sup>-6</sup> , 8.4 10 <sup>-2</sup> ]	2.1 10 <sup>6</sup>

where the state variables are defined as in Table III. Again, using the vector sign of the important direction  $\text{sign}(\boldsymbol{\alpha}) = (-1, -1, -1, -1, 1, 1)$  the lower and upper conjugate states can be identified as  $\boldsymbol{\theta}_{\underline{p}_F} = (\underline{x}, \underline{\xi})$  and  $\boldsymbol{\theta}_{\overline{p}_F} = (\underline{x}, \underline{\xi})$  respectively. Thus, approach A leads to the optima

$$\arg \min_{\boldsymbol{p} \in \mathcal{B}_\xi} p_F = (\underline{\mu}_1, \underline{\sigma}_1, \underline{x}) \quad \arg \max_{\boldsymbol{p} \in \mathcal{B}_\xi} p_F = (\underline{\mu}_1, \underline{\sigma}_1, \underline{x}). \quad (8)$$

Comparison of results from approach A and B is shown in Table IV. This time the search process of approach B produces an approximation of the bounds quite close to the exact solution obtained with approach A.

#### 4.2. LARGE SCALE FINITE ELEMENT MODEL OF A SIX-STOREY BUILDING

In this example the reliability analysis of a six-story building subject to wind load is carried out (Schuëller and Pradlwarter, 2007). Three different models of uncertainty are considered with increasing level of generality. Firstly, a standard reliability analysis, where the inputs are modelled by precise probability distribution functions, is performed. Secondly, the structural parameters are modelled as imprecise random variables with the credal set  $\mathcal{C}$ . In the third analysis both imprecise random variables and intervals are considered for the structural parameters.

An ABAQUS finite element model (FEM) is built for the six-story building, which includes beam, shell and solid elements. The load is considered as combination of a (simplified) lateral wind load and the self-weight, which are both modelled by deterministic static forces acting on nodes of each floor. The magnitude of the wind load increases with the height of the building. The FEM of the

Table III. Imprecise random variables ( $\xi$ ), precise random variables ( $\theta$ ) and intervals ( $x$ ), for the inputs of Case (b)

SV #	Symbol	Uncert. type	Mean/Interval	Stand. dev.
1	$\xi$	$\text{LN}(\underline{\mu}_1, \underline{\sigma}_1)$	$\underline{\mu}_1 = [115, 145]$	$\underline{\sigma}_6 = [5, 14]$
2	$\theta_2$	$\text{LN}(\mu_2, \sigma_2)$	$\mu_2 = 120$	$\sigma_2 = 12$
3	$\theta_3$	$\text{LN}(\mu_3, \sigma_3)$	$\mu_3 = 120$	$\sigma_3 = 12$
4	$\theta_4$	$\text{LN}(\mu_4, \sigma_4)$	$\mu_4 = 120$	$\sigma_4 = 12$
5	$\theta_5$	$\text{LN}(\mu_5, \sigma_5)$	$\mu_5 = 50$	$\sigma_5 = 15$
6	$x$	Interval $\underline{x}$	$\underline{x} = [5, 45]$	-

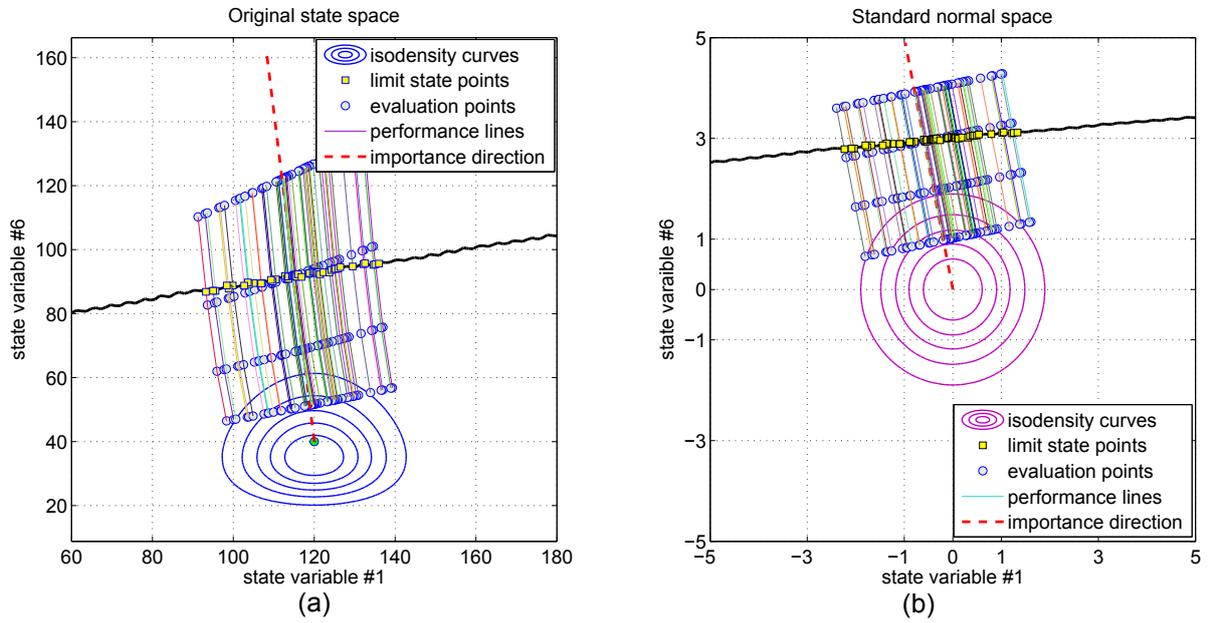


Figure 1. Limit state surface and adaptive lines in the original state space (a) and in the standard normal space (b) for imprecise variables  $\theta_1 = x$  and  $\theta_6 = \xi$

Table IV. Case (b): comparison of results in terms of failure probability and total number of samples required

Approach A		Approach B (LHS)	
$\underline{p}_F$	$N_s$	$\underline{p}_F$	$N_s$
$[5.2 \cdot 10^{-3}, 0.28]$	309	$[8.3 \cdot 10^{-3}, 0.23]$	$1.3 \cdot 10^5$

Table V. Precise distribution models for the input structural parameters.

SV #	Probability dist.	Distribution	Description	Units
1	$N(0.1, 10^{-4})$	Normal	Column's strength	GPa
2 – 193	$\text{Unif}(0.36, 0.44)$	Uniform	Sections size	m
194 – 212	$\text{LN}(35.0, 12.25)$	Lognormal	Young's modulus	GPa
213 – 231	$\text{LN}(2.5, 6.25 \cdot 10^{-2})$	Lognormal	Material's density	kg/dm <sup>3</sup>
232 – 244	$\text{LN}(0.25, 6.25 \cdot 10^{-4})$	Lognormal	Poisson's ratio	-

structure involves approximately 8200 elements and 66,300 DOFs. A total of 244 independent state variables are considered to account for the uncertainty of the structural parameters. The material strength (capacity) is represented by a Normal distribution, while Lognormal distributions are assigned to the Young's modulus, the density and the Poisson ratio. In addition, the cross-sectional width and height of the columns are modelled by independent uniform distributions. A summary of the distribution models is reported in Table V.

Component failure for the columns of the 6th storey is considered as failure criterion. The performance function is defined as

$$g(\boldsymbol{\theta}) = |\sigma_I(\boldsymbol{\theta}) - \sigma_{III}(\boldsymbol{\theta})|/2 - \sigma_y, \quad (9)$$

i.e. as the difference between the maximum Tresca stress, where  $\sigma_{III} \leq \sigma_{II} \leq \sigma_I$  are the principal stresses, and the yield stress  $\sigma_y$ .

**4.2.0.3. Standard reliability analysis** A reliability analysis is carried out with the precise distribution models reported in Table V, and using both Line Sampling (LS) and Advanced Line Sampling (ALS) for comparison of efficiency. The initial important direction is selected based on the gradient in the origin of the Standard Norm Space. The sign vector of the identified important direction is displayed in figure 2. In this example, performing LS with 30 lines (180 samples) leads to the failure probability of  $\hat{p}_F = 1.30 \cdot 10^{-4}$  and a coefficient of variation of  $CoV = 0.076$ . ALS leads to the probability of failure  $\hat{p}_F = 1.42 \cdot 10^{-4}$  with a coefficient of variation of  $CoV = 0.092$ , but with only 62 samples. Both methods estimate approximately the same value of failure probability, but quite a smaller number of model evaluations were required by ALS.

**4.2.0.4. Imprecision in distribution parameters  $\mathbf{p}$ ; uncertainty set  $\mathcal{M}_I$**  The model of uncertainty is extended to include the credal set

$$\mathcal{C} \{ h_{\mathcal{D}}(\boldsymbol{\theta}; \mathbf{p}) \mid \mathbf{p} \in \mathbb{R}^{488}, \mathbf{p} \in \mathcal{B}_{\xi} \},$$

where,  $\mathcal{D}$  are the probability distribution models from Table V, and  $\mathbf{p} = (\mu_1, \sigma_1, \dots, m_{244}, v_{244})$  are the distribution parameters of these models specified by the bounded set  $\mathcal{B}_{\xi} = \times_i^{488} \underline{p}_i$ . The interval parameters are represented as  $\underline{p} = p_c(1 - \epsilon)$ ,  $\bar{p} = p_c(1 + \epsilon)$ , using the interval center  $p_c = (\bar{p} + \underline{p})/2$

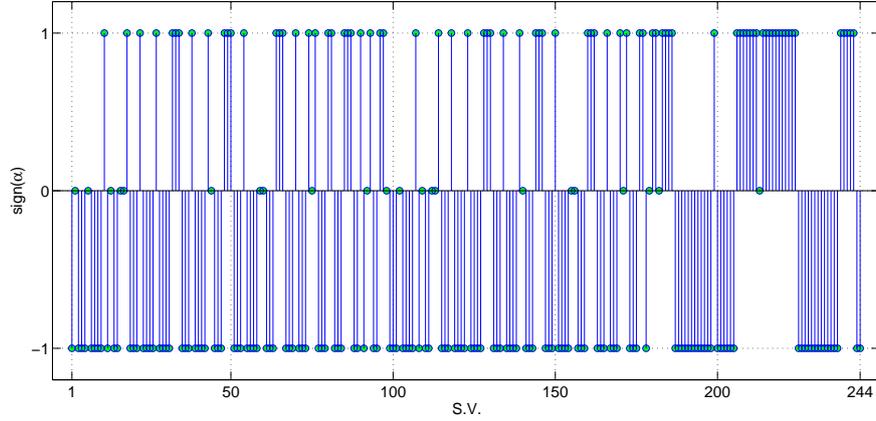


Figure 2. Sign vector of the important direction

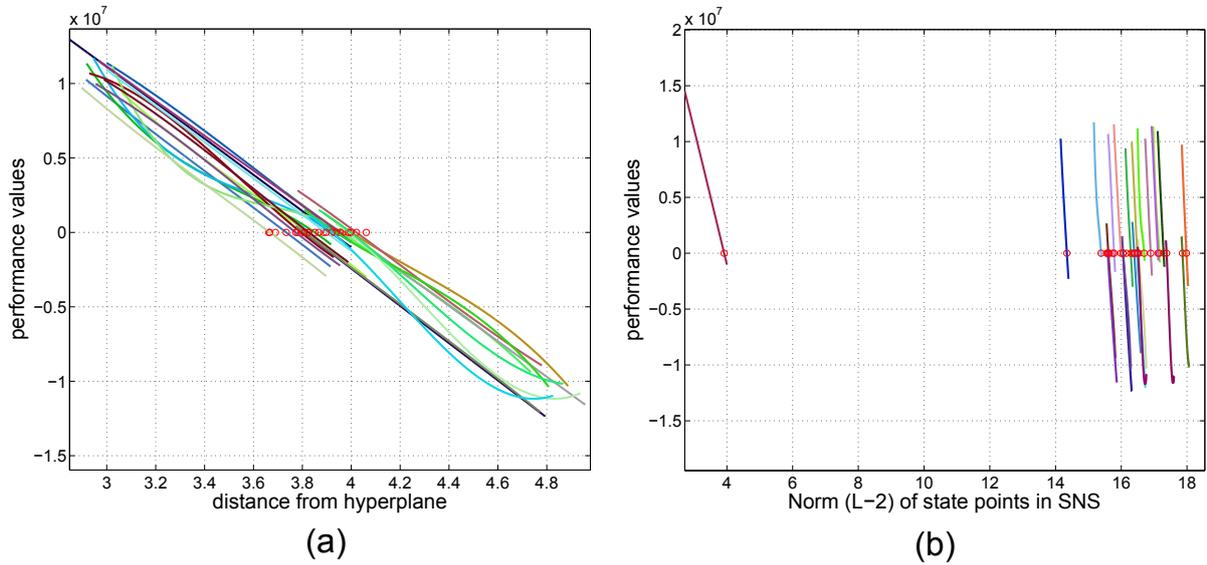


Figure 3. Values of the performance function along the lines in Standard Normal Space for one reliability analysis of the multi-storey building. In Figure (a), lines and distances from the hyperplane are plotted, while in Figure (b) lines and L-2 norm of the limit state points are plotted

and the relative radius of imprecision  $\epsilon$ . The intervals  $[p, \bar{p}]$  are defined by the bounded set  $\mathcal{B}_\epsilon$ . In the example, all interval parameters, are modeled with the same relative imprecision  $\epsilon$ . In order to explore the effects of  $\epsilon$  on the results, a fuzzy set is used to consider a nested set of intervals  $\tilde{p} = \{[p, \bar{p}]\}$  for the parameters in one analysis. The width (amplitude) of the intervals is controlled by  $\epsilon$  to obtain fuzzy sets  $\tilde{p}$ . An upper limit for the relative uncertainty is set as  $\bar{\epsilon} = 0.075$ . Specifically, the intervals for  $\epsilon = \{0, 0.005, 0.01, 0.025, 0.05, 0.075\}$  are considered. The reliability analysis with

Table VI. Inputs definition from model  $\mathcal{M}_I$ ;  $\epsilon = \{0, 0.005, 0.0.01, 0.025, 0.05, 0.075\}$ 

SV #	Prob. dist.	$\bar{p} = p_c [1 - \epsilon, 1 + \epsilon]$		Description	Units
1	$N(\bar{\mu}, \bar{\sigma})$	$\mu_c = 0.1$	$\sigma_c = 0.01$	Columns' strength	GPa
2 – 193	$\text{Unif}(\bar{a}, \bar{b})$	$a_c = 0.36$	$b_c = 0.44$	Sections' size	m
194 – 212	$\text{LN}(\bar{m}, \bar{v})$	$m_c = 35$	$v_c = 12.25$	Young's modulus	GPa
213 – 231	$\text{LN}(\bar{m}, \bar{v})$	$m_c = 2.5$	$v_c = 6.25 \cdot 10^{-2}$	Material's density	kg/dm <sup>3</sup>
232 – 244	$\text{LN}(\bar{m}, \bar{v})$	$m_c = 0.25$	$v_c = 6.25 \cdot 10^{-4}$	Poisson's ratio	-

the generalized model of uncertainty is performed using the important direction determined in the original space.

From a rough search in the set  $\mathcal{B}_\xi$ , it was found that the important direction did not significantly change in the original space. This allowed us to identify the argument optima in the bounded set  $\mathcal{B}_\xi$  as combination of extreme moments as described in Section 3.1.2. Upper and lower conjugate states are associated with the maximum and minimum of the failure probability, respectively. The result of the uncertainty propagation is shown in Table VII. From Table VII can be appreciated that the number of samples required by one robust reliability analysis, on average, is approximately 254, which is even less than number of samples required by two standard reliability analyses using Line Sampling ( $\sim 360$  samples). This is a considerable results considering that a standard approach, driven by two nested loops, would have required several hundreds of thousands of samples. The failure probability is obtained as a fuzzy set, which includes the standard reliability analysis as special case with  $\epsilon = 0$ . Each interval for  $p_F$  corresponds to the respective interval  $\bar{p} = [p, \bar{p}]$  in the input for the same membership level, and each membership level is associated with a different value  $\epsilon$ . In a design context, this result can be used to identify a tolerated level of imprecision for the inputs given a constrain on the failure probability. For example, fixing an allowable failure probability of  $10^{-3}$ , the maximum level of imprecision for the distribution parameters is limited to 1%.

4.2.0.5. *Imprecision in both distribution parameters  $\mathbf{p}$  and structural parameters  $\mathbf{x}$ ; uncertainty set  $\mathcal{M}_{II}$*  In this example the section sizes  $\mathbf{x} \in \mathbb{R}^{192}$  are considered as interval variables, while the remaining structural parameters  $\boldsymbol{\zeta} \in \mathbb{R}^{52}$  are considered as imprecise random variables, see Table VIII. The model of uncertainty comprises the credal set

$$\mathcal{C} = \{h_{\mathcal{D}}(\boldsymbol{\zeta}; \mathbf{p}) \mid \mathbf{p} \in \mathbb{R}^{104}, \mathbf{p} \in \mathcal{B}_\xi\}, \quad (10)$$

and the bounded set  $\mathcal{B}_x = \times_i^{192} \bar{x}_i$ . The imprecise distribution parameters are modeled using the radius of imprecision  $\epsilon$ , as in model case  $\mathcal{M}_I$ . An upper limit for the relative radius of imprecision is set to  $\bar{\epsilon} = 0.03$ . In the analysis, a rough search in the sets  $\mathcal{B}_\xi$  and  $\mathcal{B}_x$  allowed us again to identify a main important direction for determining the argument optima associated with the minimum and maximum value of failure probability. The result is shown in Table IX. From Table IX can be appreciated that the number of samples required by the uncertainty propagation, on average,

Table VII. Results from model  $\mathcal{M}_I$  in terms of bounds on the failure probability and total number of samples

$\epsilon$	Lower Bound		Upper Bound		Ns
	$p_F$	CoV	$\bar{p}_F$	CoV	
0.000	$1.42 \cdot 10^{-4}$	$9.2 \cdot 10^{-2}$	$1.42 \cdot 10^{-4}$	$9.2 \cdot 10^{-2}$	126
0.005	$5.75 \cdot 10^{-5}$	$8.7 \cdot 10^{-2}$	$2.63 \cdot 10^{-4}$	$7.1 \cdot 10^{-2}$	257
0.010	$4.57 \cdot 10^{-5}$	$33.6 \cdot 10^{-2}$	$5.30 \cdot 10^{-4}$	$11.5 \cdot 10^{-2}$	250
0.025	$1.75 \cdot 10^{-6}$	$8.8 \cdot 10^{-2}$	$3.22 \cdot 10^{-3}$	$5.3 \cdot 10^{-2}$	253
0.050	$2.27 \cdot 10^{-8}$	$57.0 \cdot 10^{-2}$	$3.88 \cdot 10^{-2}$	$5.4 \cdot 10^{-2}$	255
0.075	$1.88 \cdot 10^{-11}$	$12.2 \cdot 10^{-2}$	$2.02 \cdot 10^{-1}$	$3.5 \cdot 10^{-2}$	254

Table VIII. Inputs definition from model  $\mathcal{M}_{II}$ ;  $\epsilon = \{0, 0.01, 0.015, 0.020, 0.025, 0.03\}$ 

SV #	Uncertainties type		$\bar{p} = p_c [1 - \epsilon, 1 + \epsilon]$ , $\bar{x} = [\underline{x}, \bar{x}]$	
1	distribution	$N(\underline{\mu}, \underline{\sigma}^2)$	$\mu_c = 0.1$	$\sigma_c = 0.01$
2 – 193	interval	$\bar{x}$	$\underline{x} = 0.36$	$\bar{x} = 0.44$
194 – 212	distribution	$LN(\bar{m}, \bar{v})$	$m_c = 35$	$v_c = 12.25$
213 – 231	distribution	$LN(\bar{m}, \bar{v})$	$m_c = 2.5$	$v_c = 6.25 \cdot 10^{-2}$
232 – 244	distribution	$LN(\bar{m}, \bar{v})$	$m_c = 0.25$	$v_c = 6.25 \cdot 10^{-4}$

is approximately 254. Again, it is necessary to point out that a standard approach, driven by two nested loops, would have required several hundreds of thousands of samples to compute the interval failure probability.

To explore the sensitivity against imprecision of the uncertain parameters, the failure probability is obtained as a fuzzy set. The relative radii of imprecision  $\epsilon = \{0, 0.01, 0.015, 0.020, 0.025, 0.03\}$  are considered to construct a fuzzy model for all parameters. The intervals for the structural parameters  $\bar{x}$  in  $\mathcal{B}_x$ , describing the size of the cross-sections, are independent of  $\epsilon$ , see Table VIII. Once more, the analysis may serve as a design tool to find the tolerable level of imprecision provided a threshold of allowable probability.

Here, the uncertainty due to imprecision is larger, because the whole range of the intervals is taken into account for the cross-sections. As in the previous case, a rough search in the sets  $\mathcal{B}_\xi$  and  $\mathcal{B}_x$  allowed us to identify a main important direction for selecting the argument optima producing minimum and maximum value of failure probability. Values of failure probability, obtained with  $\epsilon = \{0, 0.01, 0.015, 0.020, 0.025, 0.03\}$ , are shown in Table IX.

Table IX. Results from model  $\mathcal{M}_{II}$  in terms of bounds on the failure probability and total number of samples

$\epsilon$	Lower Bound		Upper Bound		Ns
	$\underline{p}_F$	CoV	$\bar{p}_F$	CoV	
0.000	$4.70 \cdot 10^{-7}$	$10.2 \cdot 10^{-2}$	$6.73 \cdot 10^{-3}$	$11.5 \cdot 10^{-2}$	259
0.010	$2.28 \cdot 10^{-7}$	$13.4 \cdot 10^{-2}$	$9.71 \cdot 10^{-3}$	$12.2 \cdot 10^{-2}$	247
0.015	$1.10 \cdot 10^{-7}$	$10.3 \cdot 10^{-2}$	$1.11 \cdot 10^{-2}$	$7.6 \cdot 10^{-2}$	255
0.020	$5.19 \cdot 10^{-8}$	$13.1 \cdot 10^{-2}$	$2.08 \cdot 10^{-2}$	$14.6 \cdot 10^{-2}$	255
0.025	$2.51 \cdot 10^{-8}$	$9.97 \cdot 10^{-2}$	$2.72 \cdot 10^{-2}$	$15.3 \cdot 10^{-2}$	249
0.030	$1.40 \cdot 10^{-8}$	$9.94 \cdot 10^{-2}$	$3.21 \cdot 10^{-2}$	$6.5 \cdot 10^{-2}$	254

## 5. Conclusions

In this paper a generalized uncertainty framework is formulated and a numerical strategy is proposed to propagate the epistemic uncertainty in terms of failure probability. Parametric models of uncertainty that comprise bounded sets and credal sets are formulated as a sound way to account for epistemic uncertainty. This formulation finds a natural collocation in the general theory of imprecise probability. The strategy couples advanced sampling-based methods with optimisation procedures. The use of Advanced Line Sampling as a method for estimating precise failure probabilities proves to be essential not only to provide accurate estimates, but also for easing the search process. An adaptive algorithm was developed to increase the accuracy of the sampling method. By means of this strategy, based on Advanced Line Sampling, the lower and upper bounds of the failure probability  $p_F$  can be identified by searching for the minimum and maximum value of  $p_F$  within the feasible domain. The feasible domain is naturally defined by the bounded sets limiting the values of the distribution and structural parameters.

Within this framework the uncertainty propagation can be efficiently performed as far as a single failure mode is concerned. The efficiency of the proposed strategy was demonstrated by means of numerical examples, where the uncertainty propagation resulted several orders of magnitude faster compared to a naive approach based on global optimization. Moreover, the proposed approach shows that, using parametric models, the uncertainty propagation of failure probability can be performed with a quite limited numerical effort. In practice, with this approach the time required by the uncertainty propagation is comparable to the time of a single Monte Carlo analysis.

Limitations of the proposed approach can also be identified, as the efficiency plummets when problems with multiple failure modes are considered. Multiple failure modes can be found in series and parallel systems as well as in systems where the performance function is highly nonlinear. Moreover, parametric models limit the analyst to consider families of parental distributions, whereas often very few information are available and only bounds on empirical CDFs can be identified.

## References

- Alvarez, D. On the calculation of the bounds of probability of events using infinite random sets. *International journal of approximate reasoning*, 43(3):241–267, 2006. Elsevier
- Au, S. and Beck, J. L. *Estimation of small failure probabilities in high dimensions by subset simulation*. *Probabilistic Engineering Mechanics*, 16(4):263–277, 2001. Elsevier
- Augusti, G. and Ciampoli, M. *Performance-based design in risk assessment and reduction*. *Probabilistic Engineering Mechanics*, 23(4):496–508, 2008. Elsevier
- Beer, M. and Ferson, S. and Kreinovich, V. *Imprecise probabilities in engineering analyses*. *Mechanical Systems and Signal Processing*, 37(1):4–29, 2013. Elsevier
- de Angelis, M. and Patelli, E. and Beer, M. *Advanced line sampling for efficient robust reliability analysis*. *Structural safety*, submitted, 2014. Elsevier
- Ditlevsen, O. and Bjerager, P. and Olesen, R. and Hasofer, A.M. *Directional simulation in Gaussian processes*. *Probabilistic Engineering Mechanics*, 3(4):207–217, 1988.
- Ferson, S. and Kreinovich, V. and Ginzburg, L. and Myers, D. S. and Sentz, K. *Constructing probability boxes and Dempster-Shafer structures*. *Sandia National Laboratories*, Volume 835, 2002.
- Ferson, S. and Kreinovich, V. and Hajagos, J. and Oberkampf, W. and Ginzburg, L. *Experimental uncertainty estimation and statistics for data having interval uncertainty*. *Sandia National Laboratories*, 2007.
- Grooteman, F. *An adaptive directional importance sampling method for structural reliability*. *Probabilistic Engineering Mechanics*, 26(2):134–141, 2011. Elsevier
- Harbitz, A. *An efficient sampling method for probability of failure calculation*. *Structural safety*, 3(2):109–115, 1986. Elsevier
- McKay, M. D. and Beckman, R. J. and Conover, W. J. *Comparison of three methods for selecting values of input variables in the analysis of output from a computer code*. *Technometrics*, 21(2):239–245, 1979. Taylor & Francis
- Moens, D. and Vandepitte, D. *A survey of non-probabilistic uncertainty treatment in finite element analysis*. *Computer methods in applied mechanics and engineering*, 194(12):1527–1555, 2005. Elsevier
- Schuëller, G.I. and Pradlwarter, H.J. *Benchmark study on reliability estimation in higher dimensions of structural systems—an overview*. *Structural Safety*, 29(3):167–182, 2007. Elsevier
- Valdebenito, M.A. and Pradlwarter, H.J. and Schuëller, G.I. *The role of the design point for calculating failure probabilities in view of dimensionality and structural nonlinearities*. *Structural Safety*, 32(2):101–111, 2010. Elsevier
- Walley, P. *Statistical reasoning with imprecise probabilities*. Chapman and Hall London
- Zaffalon, M. *The naive credal classifier*. *Journal of statistical planning and inference*, 105(1):5–21, 2002. Elsevier

