Conference Paper

Multi-level meta-modelling in imprecise structural reliability analysis

Author(s):
Schöbi, Roland; Sudret, Bruno

Publication Date:
2016

Permanent Link:
https://doi.org/10.3929/ethz-a-010671510

Rights / License:
In Copyright - Non-Commercial Use Permitted
Multi-level Meta-modelling in Imprecise Structural Reliability Analysis

R. Schöbi\(^1\) and B. Sudret\(^1\)

\(^1\)Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland. Email: schoebi@ibk.baug.ethz.ch, sudret@ibk.baug.ethz.ch

Abstract: In modern engineering, mechanical processes are modelled using advanced computer simulation tools (e.g. finite element models), which can be characterized as expensive-to-evaluate functions. The cost of quantitative structural reliability analysis is high, when classical sampling-based techniques, such as Monte Carlo simulation, are employed. An important aspect of those analyses is the characterization of the uncertainty in the input. In the case of sparse datasets, probability theory is often not the optimal theory. Imprecise probabilities provide a more general framework accounting for aleatory and epistemic uncertainties in the representation of the dataset. The uncertainty propagation of imprecise probabilities leads to imprecise failure probabilities in structural reliability analysis. In this paper, a multi-level meta-modelling approach is presented, which reduces the computational cost of imprecise structural reliability analysis. Bounds on failure probabilities can then be estimated efficiently.

Keywords: Design enrichment, imprecise probabilities, meta-modelling, probability-box, structural reliability analysis.

1. Introduction
The cost of quantitative structural reliability analysis and reliability-based design optimization is generally high. This is due to the use of advanced simulation tools, such as finite element models, and the repeated simulations required to account for the uncertainties in the process of interest. This limits the applicability of classical sampling-based techniques, such as Monte Carlo simulation, to trivial applications. A solution to this constraint is meta-modelling, which surrogates time-consuming computer simulations with inexpensive function evaluations (Sudret, 2015).

So far, meta-modelling techniques have been developed in the context of probability theory. However, a common situation in practice is having too scarce information to develop a probabilistic input model. Scarcie information introduces epistemic uncertainty (lack of data or imprecise data) along aleatory uncertainty (natural variability). The mix of these two sources of uncertainties is called imprecise probability and can be characterized by e.g. probability-boxes (p-boxes) (Ferson and Ginzburg, 1996). The propagation of p-boxes through the computational model generally results in an imprecise quantity of interest. Hence, an imprecise structural reliability analysis results in an estimate of bounds to the failure probability. In this paper, meta-models are used at different stages of imprecise structural reliability analysis to enhance the overall computational performance.

2. Probability-boxes
Consider the probability space \((\Omega, \mathcal{F}, \mathbb{P})\), where \(\Omega\) denotes an event space equipped with the \(\sigma\)-algebra \(\mathcal{F}\) and a probability measure \(\mathbb{P}\). Random variables, denoted by capital letters, represent the mapping \(X(\omega): \Omega \rightarrow \mathcal{D}_X \subset \mathbb{R}\), where \(\omega \in \Omega\) is an elementary event. Typically, a random variable \(X\) is characterized by its cumulative distribution function (CDF) \(F_X\), which describes the probability \(F_X(x) = \mathbb{P}(X \leq x)\). The CDF provides a single measure for the variability in \(X\). This indicates that the variability is known and quantifiable.

A more general formulation for characterizing \(X\) is a probability-box (p-box), which describes the CDF of \(X\) by a lower \((F_X^-)\) and an upper \((F_X^+)\) boundary curve (Ferson and Ginzburg, 1996; Ferson and Hajagos, 2004). It applies to any \(x \in X\) of which the true but unknown CDF is contained between the
boundary curves, i.e. \( F_X(x) \leq F_X(x) \leq F_X(x) \). The gap between \( F_X \) and \( F_X \) describes the lack of knowledge (epistemic uncertainty) of \( X \), whereas the general shape of the CDF describes the natural variability (aleatory uncertainty) of \( X \). When \( F_X(x) = F_X(x) = F_X(x) \), \( \forall x \in X \), the epistemic uncertainty vanished and the p-box describes solely aleatory uncertainty, reducing to conventional probability theory.

In the context of Dempster-Shafer’s theory of evidence (Dempster, 1967; Shafer, 1976), \( F_X \) and \( F_X \) can be interpreted as the belief and plausibility measure, respectively. The belief describes the minimum amount of probability that must be associated to the event \( \{X(\omega) \leq x\} \), whereas the plausibility describes the maximum amount of probability that might be associated to the same event \( \{X(\omega) \leq x\} \).

3. Structural reliability analysis

3.1 Limit-state function

A limit-state function describes the performance of a process as a function of a set of input parameters. Mathematically speaking, it is defined as the following mapping:

\[
G: x \in D_X \subset \mathbb{R}^M \rightarrow y = G(x) \in \mathbb{R},
\]

where \( x \) is an \( M \)-dimensional vector defined in the input domain \( D_X \) (support of basic random variables) and \( y \) is the output scalar. The sign of \( G(x) \) determines if a realization \( x \) of \( X \) corresponds to a safe system \( (G(x) > 0) \) or a failed system \( (G(x) \leq 0) \). It is assumed that the limit-state function \( G \) is a black box, i.e. for each realization \( x \), only the corresponding response \( y = G(x) \) is accessible. In other words, the governing equations are not observable, such as in the case of assessing the system performance through finite element models.

3.2 Failure probability estimation

In the context of a probabilistic input vector \( X \), the failure probability \( P_f \) is defined as the probability that the limit-state function takes negative values:

\[
P_f = \mathbb{P}(G(X) \leq 0),
\]

The failure probability can be recast as the following integral:

\[
P_f = \int_{D_f} f_X(x) \, dx,
\]

where \( D_f = \{x \in D_X: G(x) \leq 0\} \) is the failure domain and \( f_X \) is the joint probability density function of the input vector \( X \).

Due to the generally complex shape of the failure domain \( D_f \) and the black-box limit-state function, the integration in Eq. (3) cannot be performed analytically. A numerical estimate of the failure probability \( P_f \) can be obtained by Monte Carlo simulation. Considering a large sample of \( X \) denoted by \( \mathcal{S} = \{x_1, ..., x_n\} \), the failure probability can be estimated by:

\[
P_f \approx \frac{n_f}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{G(x_i) \leq 0},
\]

where \( n_f \) is the number of failure samples \( x_i \in D_f \), \( n = |\mathcal{S}| \) is the total number of samples and \( \mathbb{I}_{G(x_i) \leq 0} \) is the indicator function for failure such that \( \mathbb{I} = 1 \) for \( G(x) \leq 0 \) and \( \mathbb{I} = 0 \) otherwise. Note that the use of \( \mathbb{I} \) transforms the structural reliability problem into a classification problem where only the sign of \( G(x) \) is required.

3.3 Imprecise structural reliability analysis

In the case of p-box modelled input distributions, the computation of the failure probability in Eq. (3) leads to an interval for \( P_f \) rather than a constant value, i.e. \( P_f \in [P_f, \overline{P}_f] \). It has been shown in previous work (Schöbi and Sudret, 2015; Alvarez et al., 2014) that the imprecise structural reliability problem can be recast into two standard structural reliability problems with probabilistic input, which can then be solved as described in the previous section.

Consider the random vector \( C \) of length \( M \) which follows a uniform distribution in the unit-hypercube domain \([0,1]^M\). Let the variable \( C_i \) describe the CDF value of \( X_i \). For the sake of simplicity, assume further that the input variables \( X_i \) are independent. Then, each \( c_i \in C_i \) can be transformed into an interval in the domain \( D_{X_i} \) through the inverse CDF of the p-box boundaries:

\[
X_i(c_i) = F_{X_i}^{-1}(c_i), \quad x_i(c_i) = F_{X_i}^{-1}(c_i).
\]
Two new limit-state functions can then be defined describing the lower and upper boundaries of the limit-state p-box:
\[
\begin{align*}
Y &= \mathcal{G}(c) \equiv \min_{x \in [g(x),\bar{c}(e)]} \mathcal{G}(x), \quad (6) \\
\bar{Y} &= \mathcal{G}(c) \equiv \max_{x \in [\underline{g}(x),\bar{c}(e)]} \mathcal{G}(x). \quad (7)
\end{align*}
\]
Finally, the boundaries of the failure probability can be obtained by:
\[
\begin{align*}
P_f &= \mathbb{P}(\bar{Y} \leq 0) = \mathbb{P}(\mathcal{G}(X) \leq 0), \\
\bar{P}_f &= \mathbb{P}(Y \leq 0) = \mathbb{P}(\bar{\mathcal{G}}(X) \leq 0).
\end{align*}
\]

The two boundaries of the failure probability \(P_f\) and \(\bar{P}_f\) can then be estimated with Monte Carlo simulation by sampling the probabilistic input vector \(C\). Evaluation of each boundary is performed independently.

4. Multi-level meta-modelling

4.1 Basic idea

Consider cases with expensive-to-evaluate limit-state functions and low failure probabilities \((10^{-3} - 10^{-6})\), which occur often in engineering practice. Additionally, consider p-boxes in the input introducing optimization operations on the limit-state function, as seen in Eq. (6) and (7). Because these three factors induce high computational costs, imprecise structural reliability analysis may become intractable.

In the following sections, two levels of meta-modeling are introduced for reducing the total computational costs. The first level acts directly on the limit-state function \(\mathcal{G}\), whereas the second level tackles the optimization operations in \(\mathcal{G}\) and \(\bar{\mathcal{G}}\).

4.2 Adaptive Kriging meta-model

4.2.1 Kriging

Kriging (a.k.a. Gaussian process modelling) is a meta-modelling technique that considers the limit-state function to be a realization of a Gaussian process (Santner et al., 2003):
\[
\mathcal{G}^{(k)}(x) = \beta^T f(x) + \sigma^2 Z(x, \omega), \quad (8)
\]
where \(f(x) = \{f_j(x), j = 1, ..., p\}\) are regression functions, \(\beta\) is a vector of coefficients, \(\beta^T f(x)\) and \(\sigma^2\) are the mean and the variance of the Gaussian process respectively, and \(Z(x, \omega)\) is a zero-mean, unit-variance stationary Gaussian process. The Gaussian process is characterized by an autocorrelation function \(R(|x - x'|; \theta)\) and its hyperparameter(s) \(\theta\).

A Kriging meta-model is trained with a set of input realizations \(X = \{\chi^{(i)}, i = 1, ..., N\}\) and the corresponding responses of the limit-state function \(Y = \{Y^{(i)} = \mathcal{G}(\chi^{(i)}), i = 1, ..., N\}\). The Kriging parameters \(\beta, \sigma^2\) are then obtained by the generalized least-squares solution (Santner et al., 2003):
\[
\begin{align*}
\beta(\theta) &= (F^T R^{-1} F)^{-1} F^T R^{-1} y, \quad (9) \\
\sigma^2(\theta) &= \frac{1}{N} (y - F \beta)^T R^{-1} (y - F \beta), \quad (10)
\end{align*}
\]
where \(R_{ij} = R(|\chi^{(i)} - \chi^{(j)}|; \theta)\) is the correlation matrix and \(F_{ij} = f_j(\chi^{(i)})\). In common cases where the correlation parameters \(\theta\) are unknown, their values can be estimated by e.g. maximum likelihood estimation.

The prediction value of the limit-state function at an arbitrary point \(x \in \mathcal{D}_X\) is a Gaussian random variable with the following mean and variance:
\[
\begin{align*}
\mu_y(x) &= f(x)^T \beta + r(x)^T R^{-1} (y - F \beta), \quad (11) \\
\sigma^2_y(x) &= \sigma^2 + (1 - r(x)^T R r(x)) + u(x)^T (F^T R^{-1} F)^{-1} u(x), \quad (12)
\end{align*}
\]
where \(r(x) = R(|x - \chi^{(i)}|; \theta)\) and \(u(x) = F^T R^{-1} r(x) - f(x)\).

4.2.2 Adaptive experimental design

Kriging approximates the limit-state function most accurately close to the points of the experimental design \(X\). However, these points are often not suitable to approximate the limit-state surface \(\mathcal{G}(x) = 0\), and hence to estimate the failure probability. In this context, adaptively enriching the experimental design in a guided way can improve the estimation of \(P_f\).

In this paper, the adaptive-Kriging-Monte Carlo simulation (AK-MCS) proposed by Echard et al. (2011) and further discussed in Schöbi et al. (2015) is used. The main steps of AK-MCS are the following:
(i) Generate a small initial experimental design \( \mathbf{X} \) and evaluate the limit-state function so that \( Y^{(i)} = G(X^{(i)}) \).

(ii) Train a Kriging meta-model \( G^{(K)} \) based on \( \{ \mathbf{X}, Y \} \).

(iii) Generate a large set of candidate samples \( \mathcal{S} = \{x_1, ..., x_n \} \) and evaluate the response of the meta-model \( G^{(K)} \), i.e. \( \mu_f(x_i), \sigma_f(x_i), \ i = 1, ..., n \).

(iv) Compute the following three failure probabilities:

\[
\tilde{P}_f = \mathbb{P}(\mu_f(x) \leq 0), \quad \tilde{P}_f^+ = \mathbb{P}(\mu_f(x) + 2\sigma_f(x) \leq 0).
\]

(v) Check the convergence criterion:

\[
\frac{\tilde{P}_f^+ - \tilde{P}_f^-}{\tilde{P}_f^0} \leq \epsilon_{pf},
\]

If it is not fulfilled, continue with step (vi), otherwise stop the adaptive design algorithm and return the last meta-model \( G^{(K)} \).

(vi) Compute the probability of misclassification for every \( x_i \in \mathcal{S} \):

\[
P_m(x_i) = \Phi\left( \frac{\mu_f(x_i)}{\sigma_f(x_i)} \right),
\]

where \( \Phi(\cdot) \) is the CDF of a standard normal variable (\( \mu = 0, \sigma = 1 \)). Select the next best sample to be added to the experimental design as:

\[
\mathbf{X}^* = \text{argmax}_i P_m(x_i).
\]

(vii) Add \( \mathbf{X}^* \) to the experimental design, i.e. \( \mathbf{X} \leftarrow \{ \mathbf{X}, \mathbf{X}^* \} \), and compute \( Y^* = G(X^*) \), i.e. \( Y \leftarrow \{ Y, Y^* \} \). Then go back to step (ii).

After the termination of the adaptive experimental design algorithm, the failure probability can be estimated by Monte Carlo simulation using the last Kriging meta-model \( G^{(K)} \). It has been shown that \( \epsilon_{pf} = 5% \) gives accurate results at a reasonable cost.

### 4.3 Meta-modelling the limit-state surface of \( G \)

The first meta-model is applied on the limit-state function \( G \). The Kriging meta-model \( G^{(K)} \) shall model accurately the limit-state surface of \( G \). In order to conduct the AK-MCS analysis, a probabilistic input vector \( \mathbf{X} \) is required. For imprecise structural reliability analysis, auxiliary input variables \( \tilde{\mathbf{X}}_i \) are defined, because the input is modelled by p-boxes. \( \tilde{\mathbf{X}}_i \) is characterized by the average value of its p-box boundaries:

\[
\tilde{F}_X(x_i) = (\tilde{F}_x(x_i) + \tilde{F}_x(x_i))/2.
\]

The resulting meta-model \( G^{(K)} \) is then taken as an input for the second layer of meta-modelling in the next section.

### 4.4 Meta-modelling the limit-state surface of \( \underline{G} \) and \( \overline{G} \)

The second level of meta-modelling surrogates the two limit-state functions \( \underline{G} \) and \( \overline{G} \), which can be written as functions of the first level meta-model when substituting \( G^{(K)}(x) \) for \( G \) in Eq. (6) and (7). Then, Eq. (6) and (7) are approximated as:

\[
\underline{G}(c) = \min_{x \in \{c \}} G^{(K)}(x), \quad \overline{G}(c) = \max_{x \in \{c \}} G^{(K)}(x).
\]

In order to estimate the boundaries of the failure probability \( P_f \) and \( \overline{P}_f \), AK-MCS analyses are conducted separately for \( \overline{G} \) and \( \underline{G} \) respectively. Note that in this case the input vector consists of probabilistic variables \( C_i \), thus an auxiliary distribution is not required.

The above procedure is summarized in Figure 1. It shows the two levels of AK-MCS meta-modelling and their relation.

![Figure 1. Multi-level meta-modelling scheme.](image-url)
5. Applications

5.1 Four-branch function

The four-branch function describes the failure of a system with four distinct component limit states (Echard et al., 2011):

\[
g_1(x) = \min \left\{ \begin{align*}
3 + 0.1(x_1 - x_2)^2 - (x_1 - x_2)/\sqrt{2} \\
3 + 0.1(x_1 - x_2)^2 + (x_1 - x_2)/\sqrt{2} \\
(x_2 - x_1) + 6/\sqrt{2} \\
(x_1 - x_2) + 6/\sqrt{2}
\end{align*} \right\},
\]

(16)

where the input variables \( X_i \) are modelled by independent p-boxes \( F_{X_i} \in [E_{X_i}, F_{X_i}] \). The boundaries of \( F_{X_i} \) are modelled by Gaussian distributions defined by \( E_{X_i} = F_N(x_i|\mu = 0.25, \sigma = 1) \) and \( F_{X_i} = F_N(x_i|\mu = -0.25, \sigma = 1) \).

The imprecise structural reliability analysis involves an initial experimental design of size \( N_0 = 12 \) Latin-hypercube samples (Step (i) in Section 4.2.2), a Monte Carlo sample \( \mathcal{S} \) of size \( nMC = 10^6 \) (Step (iii)), and Kriging models with Gaussian autocorrelation function. Note that in order to speed up convergence, the Kriging models at the second level are created in the auxiliary input domain defined in Eq. (13), as previously discussed in Schöbi and Sudret (2015). The optimizations (Eq. (14) and (15)) are performed using a genetic algorithm.

The results are summarized in Table 1. The reference values for the boundaries of the failure probability \( P_{f,ref} \) are obtained by Monte Carlo simulation on the exact model \( G \) and \( \bar{G} \) using \( n_{ref} = 10^6 \) samples. The estimates of the failure probabilities \( \bar{P}_f \) for \( P_f \) and \( \bar{P}_f \) are accurate with respect to the reference values. The number of evaluations of \( G \) required to build the first-level meta-model is as low as \( N_1 = 12 + 83 = 95 \), which shows the efficiency of the proposed approach. \( N_2 \) denotes the number of calls to \( G \) and \( \bar{G} \) (Eq. (14) and (15)). The small values of \( N_2 \) indicate a small number of optimization operations caused by the second level of meta-models, i.e. the AK-MCS. Note that although meta-models are used in Eq. (14) and (15), the optimization operations are expensive due to the large number of calls to \( G^{(K)} \) to find the optimal limit-state function values.

<table>
<thead>
<tr>
<th>( P_f )</th>
<th>( \bar{P}_f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{f,ref} )</td>
<td>( 1.27 \cdot 10^{-3} )</td>
</tr>
<tr>
<td>( \bar{P}_f )</td>
<td>( 1.34 \cdot 10^{-3} )</td>
</tr>
<tr>
<td>( N_1 )</td>
<td>( 12 + 83 = 95 )</td>
</tr>
<tr>
<td>( N_2 )</td>
<td>( 167 )</td>
</tr>
</tbody>
</table>

5.2 Structural frame

Consider the five-storey frame application previously discussed in Blatman and Sudret (2010). The quantity of interest is the deflection of the top floor at the right corner, denoted by \( u \). The limit-state function is then defined as:

\[
g_{frame}(x) = u_{adm} - u(x),
\]

(17)

where \( u_{adm} = 0.06 \) m is the admissible deflection and \( x \) consists of 21 dependent input variables describing loads, material parameters, and geometrical quantities. The input marginal and correlation structure are defined in Blatman and Sudret (2010) in the probabilistic context. In order to define p-boxes in the input variables, the loads are modelled by p-boxes. In particular, the boundaries of the CDF of each load are defined as:

\[
E_{X_i} = F_{X_i}(x_i; \mu_{X_i}(1 + \alpha), \sigma_{X_i}(1 + \alpha)),
\]

\[
F_{X_i} = F_{X_i}(x_i; \mu_{X_i}(1 - \alpha), \sigma_{X_i}(1 - \alpha)),
\]

where \( i = 1,2,3 \) and \( \alpha = 0.05 \). The remaining variables are kept as dependent probabilistic random variables.

The same settings are used for AK-MCS as in the four-branch function application. In particular, AK-MCS initiates with only 12 samples in the initial experimental design.

The imprecise structural reliability analysis results in the failure probability bounds \([P_f, \bar{P}_f] = [1.25 \cdot 10^{-4}, 3.81 \cdot 10^{-4}] \) obtained with a total number of \( N_1 = 12 + 212 = 224 \) evaluations of \( G \). The number of evaluations on the second-level meta-model is \( N_2 = 12 + 142 = 154 \) and \( N_2 = 12 + 220 = 232 \) for \( P_f \).
and $\overline{P}_f$ respectively. The values of $N_2$ are comparable to $N_1$, indicating an efficient estimation of the bounds of the failure probability.

Note that due to the nature of the problem, the limit-state function is a monotone function with respect to the imprecise loads. Thus, the exact boundaries of failure probabilities can be obtained by simply considering the boundaries of the input p-boxes, i.e. $\underline{F}_X$ and $\overline{F}_X$. Then, an Importance Sampling analysis (based on preliminary FORM) with $n = 10^6$ samples results in $P_f = [1.27 \cdot 10^{-4}, 3.86 \cdot 10^{-4}]$ which are approximated accurately with the proposed approach.

6. Conclusions

In engineering practice, datasets are often sparse due to the high costs of data generation. In this context, p-boxes can be used to characterize the uncertainty in variables, accounting for epistemic and aleatory uncertainties. P-boxes provide a lower and an upper boundary function for the cumulative distribution function of a variable.

This paper discusses the estimation of failure probabilities in the presence of p-boxes in the input. P-boxes increase the complexity of the imprecise structural reliability analysis, which may become intractable for expensive-to-evaluate limit-state functions. In order to reduce the computational effort and to keep the analysis tractable, Kriging meta-models with adaptive experimental designs are employed. In particular, meta-models are used at two different levels of the imprecise structural reliability analysis for an efficient estimation of the imprecise failure probability, i.e. the boundaries of the imprecise failure probability.

Two examples (a benchmark analytical function and a structural frame) illustrate the efficiency of the proposed multi-level meta-modelling approach. In there, a small number of evaluations of the limit-state function is sufficient to estimate the imprecise failure probability accurately.

References


