Reliability Engineering and System Safety 155 (2016) 9–20

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

# Reliability analysis under epistemic uncertainty

Saideep Nannapaneni, Sankaran Mahadevan *

*Department of Civil and Environmental Engineering, Vanderbilt University, Nashville, TN, USA*

ARTICLE INFO

*Article history:*
Received 9 October 2014
Received in revised form
2 May 2016
Accepted 5 June 2016
Available online 16 June 2016

*Keywords:*
Reliability
Epistemic uncertainty
Auxiliary variable
First Order Reliability Method
Correlation
Monte Carlo simulation

ABSTRACT

This paper proposes a probabilistic framework to include both aleatory and epistemic uncertainty within model-based reliability estimation of engineering systems for individual limit states. Epistemic uncertainty is considered due to both data and model sources. Sparse point and/or interval data regarding the input random variables leads to uncertainty regarding their distribution types, distribution parameters, and correlations; this statistical uncertainty is included in the reliability analysis through a combination of likelihood-based representation, Bayesian hypothesis testing, and Bayesian model averaging techniques. Model errors, which include numerical solution errors and model form errors, are quantified through Gaussian process models and included in the reliability analysis. The probability integral transform is used to develop an auxiliary variable approach that facilitates a single-level representation of both aleatory and epistemic uncertainty. This strategy results in an efficient single-loop implementation of Monte Carlo simulation (MCS) and FORM/SORM techniques for reliability estimation under both aleatory and epistemic uncertainty. Two engineering examples are used to demonstrate the proposed methodology.



## 1. Introduction

Reliability analysis is concerned with the assessment of system performance in the presence of uncertainty, which has generally been classified into two types: aleatory (natural variability) and epistemic (lack of knowledge). The reliability estimate is affected by both types of uncertainty; however, extensive previous literature in model-based reliability analysis has predominantly considered the former type and not the latter. While several alternative frameworks have been explored to represent uncertainty, this paper considers the probabilistic framework. In this context, the probability of failure is represented as

$$P_f = \int_{g(\boldsymbol{X})\leq 0} f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} \tag{1}$$

where $P_f$ is the probability of failure, $\boldsymbol{X}$ is the vector of input random variables, $f_{\boldsymbol{X}}(\boldsymbol{x})$ is the joint probability distribution function (PDF) of $\boldsymbol{X}$, $g(\boldsymbol{X})$ is the performance (limit state) function, and $g(\boldsymbol{X})\leq 0$ represents the failure domain. Different types of Monte Carlo simulation methods, as well as analytical integration techniques such as first-order and second-order reliability methods (FORM, SORM), have been developed [1] to evaluate Eq. (1).

The evaluation of the multi-dimensional integral in Eq. (1) can be difficult; therefore, First Order Reliability Methods (FORM) approximate the limit state function, which could be non-linear, with a first-order (linear) approximation while the Second Order Reliability Methods (SORM) estimate the failure probability by employing a second-order approximation to the limit state. Refer to [1] for more details.

Due to insufficient information, uncertainty may arise about the exact values of deterministic variables or the distribution characteristics of random variables in Eq. (1). This is referred to as statistical uncertainty. Several theories, both probabilistic [2–4] and non-probabilistic [5], have been used to represent this type of epistemic uncertainty. Some of the approaches include interval analysis [5], convex models [6], fuzzy sets and possibility theory [7], evidence theory [8], Bayesian probability theory [3] and imprecise probabilities [9].

This paper uses a Bayesian probabilistic approach to model epistemic uncertainty about the input random variables. A random variable may be represented using a parametric (e.g., normal) or a non-parametric distribution. A parametric distribution is associated with a distribution type and distribution parameters. If the distribution type of an input variable $X$ is known but the distribution parameters are uncertain, then $X$ can be represented by a distributional p-box. If the distribution type is also uncertain, then $X$ may be represented by a free p-box [10].

Further, it may be difficult to obtain joint data on all the variables in the system due to limited resources. In such cases, the correlations between variables are also uncertain. In some cases, qualitative information that some variables are positively or

* Corresponding author.
*E-mail addresses:* mahades88@gmail.com, sankaran.mahadevan@vanderbilt.edu (S. Mahadevan).

http://dx.doi.org/10.1016/j.ress.2016.06.005
0951-8320/

negatively correlated might be available. It is desirable to include such information in reliability estimation; however, methods to include correlation uncertainty in reliability estimation are not yet fully explored.

A simple implementation of reliability analysis in the presence of only distribution parameter uncertainty may be through a nested double-loop MCS or a nested MCS-FORM/SORM approach where the distribution parameters are sampled in the outer loop, and for each realization of the parameters, failure probability is calculated in the inner loop using MCS or FORM/SORM. The result of double-loop analysis may be described through an average of all estimates of failure probability or by representing the failure probability itself as a distribution [11]. The hierarchical formulation of uncertainty sources according to the Rosenblatt transformation [12] offers a more convenient approach compared to double-loop sampling. Recently, Sankararaman and Mahadevan [13] proposed a new method of single-loop sampling using the concept of an auxiliary variable, based on the probability integral transform [1], in which samples of a variable are obtained through simultaneous sampling of parameters and CDF values. This approach is used in this paper for faster computation.

The statistical uncertainty discussed above constitutes one category of epistemic uncertainty; another category is model uncertainty [14]. Models are built to explain the real world phenomena and frequently involve assumptions, simplifications and generalizations. Models may be based on first principles (physics-based) or derived from data (data-driven). Model uncertainty represents the inability of these models to accurately represent the true physical behavior of the system. Uncertainty due to a model may be due to three sources: (1) model parameters, due to limited data; (2) numerical solution errors that arise from the methodology adopted in solving the model equations; and (3) model form errors, which arise due to assumptions and simplifications made in the development of models. Model calibration is used to estimate the model parameters using input-output data. Model verification can be used to quantify numerical solution errors (e.g., finite element discretization error, surrogate model error, round-off error, etc.). Model form errors can be estimated by comparing the model predictions against physical observations (e.g., model validation tests). Whenever physical observations are used for either model calibration or model validation, measurement uncertainties also arise, and these contribute to the uncertainty in the model prediction. Discretization error arises when the solution of the continuum domain is computed using numerical techniques (e.g., finite element methods) which involve discretization of the continuum domain. Surrogate models are often used in uncertainty quantification, reliability analysis and design optimization when high fidelity physics models are computationally expensive. The estimation of surrogate model error involves comparing the output of the original model with the surrogate model.

The next stage after quantification of different types of epistemic uncertainty is their inclusion in a framework for reliability estimation. This paper proposes a probabilistic framework to include both forms of epistemic uncertainty and aleatory uncertainty in reliability analysis. The main issue is that reliability analysis techniques such as MCS, FORM, etc. are wrapped around deterministic physics models, i.e., for a fixed input value, the model output is deterministic. In the presence of model uncertainty, the model output is not deterministic even for a fixed input. When variability and input statistical uncertainty are added, the model output is in the form of multiple probability distributions. Further, the various uncertainty sources and errors do not combine in a simple manner; they occur at different stages of the analysis, and their combination could be nonlinear, nested or iterative. Reliability analysis in the presence of multiple sources and types of uncertainty is thus not straightforward; this paper seeks to overcome this challenge. Current FORM-based reliability analysis methods have included either parameter uncertainty [15] or model errors [16] but distribution type uncertainty, uncertain correlations or combination of several uncertainty sources have not been considered. Similarly, Monte Carlo-based methods have not considered the various epistemic uncertainty sources in reliability analysis.

The overall contribution of the paper is a comprehensive and systematic framework for quantifying and aggregating the contributions of different types of epistemic uncertainty (statistical and model uncertainties) in a manner suitable for *reliability analysis* using FORM and Monte Carlo sampling. The key contributions of this paper can be summarized as follows – (1) quantification of different types of statistical uncertainty (distribution parameter and distribution type uncertainty, and uncertainty about correlations) and model uncertainty (model form and numerical solution errors) within a probabilistic framework; (2) development of a novel FORM-based approach to include different types of epistemic uncertainty (data, model) along with aleatory uncertainty within reliability analysis by utilizing the concepts of auxiliary variable and theorem of total probability; and (3) development of a single-loop Monte Carlo sampling approach for the inclusion of both aleatory and epistemic uncertainty in reliability estimation.

The rest of the paper is organized as follows. In Section 2, procedures to quantify various types of epistemic uncertainty are presented. Section 3 develops the proposed methodologies (using FORM and Monte Carlo sampling) for reliability estimation in the presence of aleatory and epistemic uncertainty. In Section 4, a structural reliability example and a fluid–structure interaction problem (airplane wing) are used to demonstrate the application of the proposed methods. Concluding remarks are provided in Section 5.

## 2. Representation of epistemic uncertainty

In this section, procedures for the representation of epistemic uncertainty due to data and model sources are discussed in order to facilitate reliability analysis through MCS and FORM techniques.

### 2.1. Distribution parameter uncertainty

In the presence of sparse point data on $X$, two approaches may be used to construct the probability distributions of distribution parameters $\Theta$ (using a Bayesian perspective). The first approach is to use resampling methods such as Jack-knife and Bootstrap [17] to generate multiple values of $\Theta$ that are used to construct their distributions; the second approach is to use a likelihood-based representation of the available data to construct distributions of $\Theta$ using Bayes' theorem [18]. The likelihood-based approach can be extended to accommodate interval data and to construct parametric as well as non-parametric distributions [19]; this approach is adopted in this paper.

Let a dataset $D$ for a variable $X$ consist of $n$ point data $p_i\,(i = 1 \text{ to } n)$ and $m$ interval data $[a_j, b_j]\,(j = 1 \text{ to } m)$. The likelihood function for the distribution parameters $\boldsymbol{\Theta}$ can be constructed as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x = p_i \mid \boldsymbol{\theta}) \prod_{j=1}^{m} \left[ F_X(x = b_j \boldsymbol{\theta}) - F_X(x = a_j \mid \boldsymbol{\theta}) \right] \tag{2}$$

where $f_X(x)$ and $F_X(x)$ represent the PDF and CDF of variable $X$ respectively. After constructing the likelihood function, the distributions of the distribution parameters are obtained using Bayes' theorem as