Mechanical Systems and Signal Processing 37 (2013) 137–151

Contents lists available at SciVerse ScienceDirect

Mechanical Systems and Signal Processing

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

Structural reliability analysis on the basis of small samples: An interval quasi-Monte Carlo method

Hao Zhang [a,*], Hongzhe Dai [b], Michael Beer [c], Wei Wang [b]

[a] School of Civil Engineering, University of Sydney, NSW 2006, Australia
[b] School of Civil Engineering, Harbin Institute of Technology, Harbin 150090, China
[c] Institute for Risk & Uncertainty, University of Liverpool, Liverpool, UK

ARTICLE INFO

Article history:
Received 15 August 2011
Received in revised form
28 February 2012
Accepted 4 March 2012
Available online 22 March 2012

Keywords:
Epistemic uncertainty
Imprecise probability
Low-discrepancy sequence
Probability box
Quasi-Monte Carlo
Structural reliability

ABSTRACT

In practice, reliability analysis of structures is often performed on the basis of limited data. Under this circumstance, there are practical difficulties in identifying unique distributions as input for a probabilistic analysis. But the selection of realistic probabilistic input is critical for the quality of the results of the reliability analysis. This problem can be addressed using an entire set of plausible distribution functions rather than one single distribution for random variables based on limited data. The uncertain nature of the available information is then reflected in the probabilistic input. An imprecise probability distribution can be modeled by a probability box, i.e., the bounds of the cumulative distribution function for the random variable. Sampling-based methods have been proposed to perform reliability analysis with probability boxes. However, direct sampling of probability boxes requires a large number of samples. The computational cost can be very high as each simulation involves an interval analysis (a range-finding problem). This study proposes an interval quasi-Monte Carlo simulation methodology to efficiently compute the bounds of structure failure probabilities. The methodology is based on deterministic low-discrepancy sequences, which are distributed more regularly than the (pseudo) random points in direct Monte Carlo simulation. The efficiency and accuracy of the present method is illustrated using two examples. The reliability implications of different approaches for construction of probability boxes are also investigated through the example.

1. Introduction

The building process of civil engineering structures and infrastructure is complicated by the various sources of uncertainties in structural resistance and loads, as well as in computational models. These uncertainties are treated as random variables when using established probabilistic methods for reliability analysis. Structural reliability is then measured by a probability of failure, P_f [28].

Although the mathematical formulation as well as the basic numerical techniques for the calculation of P_f appears to be straightforward, difficulties appear in practical applications. For one thing, the evaluation of the involved multi-dimensional integral can be challenging. Development of numerical techniques target at a high numerical efficiency and concern advanced concepts of Monte Carlo (MC) simulation such as subset simulation [2] and line sampling [26].

* Corresponding author. Tel.: +61 2 93513923; fax: +61 2 93513343.
E-mail addresses: hao.zhang@sydney.edu.au (H. Zhang), hzdai@hit.edu.cn (H. Dai), mbeer@liverpool.ac.uk (M. Beer), wwang@hit.edu.cn (W. Wang).

0888-3270/$ - see front matter
doi:10.1016/j.ymssp.2012.03.001

Particular attention is currently paid to simulation schemes for high-dimensional problems. Moreover, there are practical difficulties in identifying the proper distribution for the random variables particularly to model their extremes ("tails") which are of the greatest concern in reliability assessment. Structural failure probabilities, generally very small, are sensitive to the choice of probability distributions [11]. However, available data on structural strength and loads are typically limited, and competing distributions often cannot be distinguished by standard statistical tests. When observational data are limited, the analyst may not be able to identify the type of the distribution of a random variable, or precise values for the distribution parameters, or there may be competing probabilistic models. The selection of a distribution for the probabilistic input is so generally realized based on ambiguous information and indications. This may lead to a wrong model choice and a strong overestimation of structural reliability resulting in critical decisions with severe consequences.

It is advisable to consider the distribution itself as uncertain. These statistical uncertainties are epistemic (knowledge-based) in nature [10]. Because of the distribution sensitivity problem, the failure probability calculated on the basis of small data samples is only a "notional" value and should not be interpreted as a relative frequency in an absolute sense [11,10]. To overcome the distribution sensitivity problem, the development of the first-generation probability-based structural design codes utilized a process of calibration to existing practice. The notional reliabilities associated with existing and new practices were computed in the same model universe of probability distributions and used as a means of comparison [13]. If the calculated reliabilities are notional and only used for the purpose of reliability comparison, then the uncertainty in the distributions is generally of minor importance and can be ignored. A reliability comparison in this sense represents a ranking of alternatives, whereby the uncertainties in the distributions associated with the alternatives are normally not significantly different in magnitude since they origin from the same source (same model universe of probability distributions). The effects of the uncertainties in the distributions, hence, cancel out one another on an ordinal (ranking) scale almost completely.

However, there are circumstances where epistemic uncertainties due to limited availability of data need to be included explicitly in reliability analysis and, further on in risk assessment. One such case is the risk-informed decision-making, in which the regulatory authorities often see a need to quantify the confidence in the results of the risk assessment, particularly if the event is rare but the consequence is severe. Another case is the performance-based approach to structural design. In this new paradigm of structural engineering, it is necessary to establish explicit reliability/risk terms to rationalize the selection of performance levels in structural design. Designers using innovative building materials and technologies are also concerned with computing realistic structural reliabilities because they cannot rely on past experience to calibrate the reliabilities. In all these circumstances a notional reliability measure without taking into account the epistemic uncertainty in the distributions is not very helpful.

Within a pure probabilistic framework, epistemic uncertainty can be handled with Bayesian approaches. Uncertain parameters of a probabilistic model can be described with prior distributions and updated by means of limited data. They can then be introduced formally, with the remaining (aleatory) uncertainties, in the reliability calculation [10,12]. In the case where competing probabilistic models exist, each model is considered separately with an assigned probability mass. A failure probability can be computed for each probabilistic model. The expectation of the failure probability can then be calculated as a characteristic result, and the frequency distribution (or variance) of the failure probability can be evaluated to separate the effects of aleatory and epistemic uncertainty. In practical applications, this requires a high numerical effort or statistical approximations.

Alternatively, an imprecisely known probability distribution can be modeled by a family of all candidate probability distributions which are compatible with available data. This is the idea of the theory of imprecise probabilities [41]. Dealing with a set of probability distributions is essentially different from a Bayesian approach. A practical way to represent the distribution family is to use a probability bounding approach by specifying the lower and upper bounds of the imprecise probability distribution. This corresponds to the use of an interval to represent an unknown but bounded number. Consequently, a unique failure probability cannot be determined. Instead, the failure probability is obtained as an interval whose width reflects the imprecision of the distribution model in the calculated reliability.

A popular uncertainty model using the probability bounding approach is the probability box (p-box for short) structure [16]. A p-box is closely related to other set-based uncertainty models such as random sets, Dempster–Shafer evidence theory and random intervals. In many cases, these uncertainty models can be converted into each other, and thus considered to be equivalent [42,16,21,31,5]. Therefore, the methodology presented in this paper is also applicable to other set-based uncertainty models.

Within the reliability analysis context, simulation-based methods have been suggested to propagate p-boxes (e.g. [1,3,4,44]). Direct sampling of p-boxes requires a large number of samples to control the sampling uncertainty. The total computational cost can be very high as each simulation may involve an expensive range-finding problem. The issue of computational cost becomes more serious when the limit state function is only implicitly known through a computational model. Another concept follows the idea of global optimization to directly identify the bounds of probabilistic results [30]. Although this concept is general and can be applied in association with variance-reduction methods (e.g. [45]), the computations are still numerically demanding. There is some urgency for further developments towards efficient methodologies for analysis with imprecise probabilities.

This study focuses on the reduction of sampling uncertainty with quasi-Monte Carlo technique. Quasi-MC method is typically used for multidimensional numerical integration problems. It performs in a similar manner as the Monte Carlo