Structural Safety 67 (2017) 105–115

Contents lists available at ScienceDirect

Structural Safety

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

Subset simulation for non-Gaussian dependent random variables given incomplete probability information

Fan Wang [a,b,c,*], Heng Li [a]

[a] Department of Building and Real Estate, The Hong Kong Polytechnic University, Kowloon, Hong Kong
[b] School of Resource and Civil Engineering, Wuhan Institute of Technology, Wuhan, PR China
[c] Department of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, PR China

ARTICLE INFO

Article history:
Received 9 December 2016
Received in revised form 27 March 2017
Accepted 30 April 2017

Keywords:
Incomplete probability information
Subset simulation
Parameter sensitivity
Vine copula
Rosenblatt transformation

ABSTRACT

Reliability analysis under incomplete probability information is a challenging task. This paper focuses on the extension of the subset simulation (SS), an efficient reliability approach, to the modeling of any dependent random variables under incomplete probability information. The Nataf transformation is commonly adopted to generate correlated random variables given marginal distributions and correlations; however, it inherently assumes a Gaussian dependence structure. In contrast, the Rosenblatt transformation can be used to generate correlated random variables with any dependence structure; however, the joint probability information must be known. To remove the limitation, the vine copula approach, which is highly flexible in dependence modeling, is used to reconstruct the joint probability information from the prescribed marginal distributions and correlations. The copula parameters in the vine structure are retrieved using an efficient approximation method. Three copula cases including the Gaussian and non-Gaussian dependence structures are investigated by applying the proposed method to a numerical example. The failure probabilities and the effects of the uncertain parameters correspond to different cases are compared, which aims to provide insights into the impact of the dependence structure on the SS results when only the incomplete probability information is given.

1. Introduction

The uncertainty involved in engineering systems necessitates the use of probabilistic methods to enable reliable design. The generic formulation of the failure probability can be expressed as:

$$P_f = \int_{G(\mathbf{X})\leqslant 0} f(\mathbf{X})d\mathbf{X} \tag{1}$$

where $f(\mathbf{X})$ is the multivariate probability density function (PDF) of the random variable vector $\mathbf{X}$, $G(\mathbf{X}) \leqslant 0$ defines the failure domain of interest and, particularly, $G(\mathbf{X}) = 0$ shapes the limit state surface (LSS).

The analytical solution to the above integral is difficult to derive due to the multi-dimensions of the problem and the complicated failure domain. A number of methods have been developed to approximate the failure probability [1–5]. Among the various approximation methods, the Monte Carlo simulation (MCS) is widely used due to its conceptual simplicity and computational robustness [6]. However, when using the crude MCS, the computational cost can be unacceptable if the failure probability to be estimated is small, which motivates many researchers to improve the algorithm efficiency [7–10]. Among these works, the subset simulation (SS) [11–13] has been identified as the most successful one because of its high efficiency in generating rare events and insensitivity to dimensions [14]. Many studies have demonstrated the potential of the SS in estimating small probabilities [15–17] or quantifying the effects of uncertain parameters on failure probability [18–19]. The success of SS lies in the efficient simulation of samples conditioned on a series of nested intermediate events with relatively larger probabilities, which is achieved using the Markov Chain Monte Carlo (MCMC) with a modified Metropolis algorithm (MMA) [11]. Several variants of the SS have also been developed [20–22]; yet, none of the variants present a significant improvement compared to the original algorithm [23].

Compared to the many efforts devoted to developing novel strategies of generating conditional samples, few attentions are given to the generation of correlated random variables. Generally, the correlated random variables can be generated by independent ones through isoprobabilistic transformations [17]. For example, generating a random vector with dependent elements can be

* Corresponding author at: Department of Building and Real Estate, The Hong Kong Polytechnic University, Kowloon, Hong Kong.
E-mail address: wang_fan@hust.edu.cn (F. Wang).

http://dx.doi.org/10.1016/j.strusafe.2017.04.005
0167-4730/

achieved by first generating a sample **U** from the independent standard normal space (*U*-space) and then converted to the original space by using the mapping relationship $\mathbf{U} = T(\mathbf{X})$. If $f(\mathbf{X})$ is available, the Rosenblatt transformation [24] can be applied to achieve this goal. Unfortunately, the data needed for characterization of $f(\mathbf{X})$ are frequently not available to a sufficient extent and quality [25]. In practice, the probabilistic description of **X** is usually given in terms of marginal distributions and correlations, which is referred to as the incomplete probability information [26]. In such case, the Nataf transformation is commonly employed because it can generate a joint PDF (known as the Nataf distribution) that is consistent with the prescribed marginal distributions and correlations and can be simply generalized to multi-dimensions [27]. Essentially, the Nataf transformation assumes a Gaussian dependence structure (normal copula) for correlated random variables. However, recent investigation has shown that this assumption does not always hold [24]. To address the limitation, the copula approach can be used to reconstruct the joint probability distribution with non-Gaussian dependence structure based on the incomplete probability information, and thus generate non-Gaussian dependent random variables [12]. The copula parameter that indicates the strength of dependence is related to the correlation one by one. However, this copula-based reconstruction of joint probability distribution with non-Gaussian dependence structure is mainly restricted to the bivariate cases [28–30]. If the dependence structure has to be characterized by a non-elliptical copula, this reconstructing method cannot be generalized to the multivariate cases [31]. For example, if an *n*-dimensional Archimedean copula is used, there are at most *n*-1 copula parameters, while the number of pair-wise correlation coefficients is $n(n-1)/2$. As a result, the number of the copula parameters is less than the number of correlation coefficients except for the bivariate cases, which implies that it is impossible to establish a one-to-one relationship between them in multi-dimensions. This limitation can be attributed to the inflexibility of the conventional multivariate copulas in representing the complex dependence structure.

This paper aims to provide a method of generating non-Gaussian dependent random variables under incomplete probability information and investigate the impact of different dependence structures on SS results. To this end, the vine copula approach [32] which uses nested bivariate copulas to construct the joint probability distribution is adopted for dependence modeling. The relevant vine copula parameters are retrieved from the incomplete probability information via an efficient approximation method. Using different nested bivariate copulas, the joint probability distributions with different dependence structures are modeled, which is further utilized for generating correlated random variables through Rosenblatt transformation [24]. The proposed approach is illustrated with a retaining wall example involving mutually correlated non-normal trivariates. Three dependence structures including the Gaussian case and non-Gaussian case are investigated and the corresponding SS results are compared.

2. Subset simulation and efficient sensitivity analysis

2.1. Subset simulation

The basic idea of SS [11–13] is to progressively approach the target failure region *F* through a sequence of nested subsets of the random space:

$$F \subset F_m \subset F_{m-1} \subset \ldots \subset F_1 \subset F_0 = \mathbf{R}^n \tag{2}$$

By conditioning on the intermediate failure regions $F_1, F_2, \ldots, F_m$, the failure probability can be expressed as:

$$P(F) = P(F_1)\prod_{j=2}^{m} P(F_j|F_{j-1}) \cdot P(F|F_m) \tag{3}$$

Without loss of generality, suppose that the intermediate and target failure regions are defined as $F_j = \{Y = G(\mathbf{X}) < y_j, j = 1, \ldots, m\}$ and $F = \{Y = G(\mathbf{X}) < y^*\}$, respectively, wherein the thresholds satisfy that $y_1 > y_2 > \ldots > y_m > y^*$ (y^* is the threshold that defines the targeted failure region). Then, Eq. (3) becomes:

$$P(F) = P(Y < y_1)\prod_{j=2}^{m} P(Y < y_j|Y < y_{j-1})P(Y < y^*|Y < y_m) \tag{4}$$

In practice, the response values pertinent to the simulated samples are sorted in an ascending order, and y_j are determined as the p_0-percentile of the sorted values so that $P(F_1)$ and $P(F_j|F_{j-1}) j = 1, \ldots, m$ equal to a constant value of p_0.

The conditional samples cannot be efficiently generated by the crude MCS. Au and Beck [11] proposed a method of sampling from the conditional distribution based on the MMA. The main advantage of the MMA is that it is efficient in simulating samples with independent elements in high-dimensional space [11]. For details of the implementation procedure of SS, readers are referred to [11–13,22,23].

2.2. Sensitivity analysis based on conditional samples

Insights into the effects of parameter uncertainties on the system failure probability are of great value for revealing the failure mechanism. Using the Bayesian Theorem, the conditional failure probability can be computed by:

$$P(F|x) = \frac{P(F)P(x|F)}{P(x)} \tag{5}$$

where $P(F)$ can be calculated by Eq. (3) and $P(x)$ is the prescribed marginal distribution of the uncertain parameter *x*. Thus the critical step to obtain $P(F|x)$ becomes to the estimate of $P(x|F)$, which depicts the distribution of *x* given that the system has failed. Since the SS is efficient in generating failure samples, $P(x|F)$ can be readily obtained from the corresponding histogram. Based on the Total Probability Theorem, $P(x|F)$ is given as:

$$P(x|F) = \sum_{j=1}^{M} P(x|\Omega_j \cap F)P(\Omega_j|F) \tag{6}$$

where $\{\Omega_j, j = 0, 1, \ldots, M\}$ is the mutually exclusive sets partitioning the sample space; *M* is the desired number of subsets and $M \geqslant m$ ($M = m$ indicates that the algorithm stops immediately after that the target region is reached while $M > m$ means that the simulation continues until a desired number of subsets are generated):

$$\begin{aligned} &\Omega_0 = \bar{F}_1 = \{Y \geqslant y_1\} \\ &\Omega_j = F_j - F_{j+1} = \{y_j > Y \geqslant y_{j+1}\}, j = 1, 2, \ldots, M-1 \\ &\Omega_M = F_M = \{Y < y_M\} \end{aligned} \tag{7}$$

$P(\Omega_j|F)$ can be calculated by:

$$P(\Omega_j|F) = \frac{P(F|\Omega_j)P(\Omega_j)}{P(F)} \tag{8}$$

where $P(F|\Omega_j)$ is simply estimated as the fraction of the failure samples in Ω_j, and

$$\begin{aligned} &P(\Omega_0) = 1 - p_0 \\ &P(\Omega_j) = p_0^j - p_0^{j+1}, j = 1, 2, \ldots, M-1 \\ &P(\Omega_M) = p_0^M \end{aligned} \tag{9}$$