

DETC2013-13383

ESTIMATION OF PROBABILITY OF FAILURE WITH DEPENDENT VARIABLES USING COPULAS AND SUPPORT VECTOR MACHINES

Peng Jiang and Samy Missoum

Computational Design Optimization of Engineering Systems (CODES) Laboratory
Department of Aerospace and Mechanical Engineering
the University of Arizona
Tucson, Arizona, 85721
Email: {[pjiang](mailto:pjiang@email.arizona.edu), [smissoum](mailto:smissoum@email.arizona.edu)}@email.arizona.edu

ABSTRACT

This paper presents an approach to estimate probabilities of failure in the case of dependent random variables. The approach is based on copulas and support vector machines (SVMs). A copula is used to generate dependent Monte Carlo samples and an SVM is used to construct the explicit boundary of the failure domain. It is shown that this construction of the failure boundary cannot be made in the original space due to the lack of “isotropy” of the probability densities. In this work the SVM is built in the uncorrelated standard normal space and refined using an adaptive sampling scheme. A transformation is used to map SVM training points and Monte-Carlo samples between the original space and the uncorrelated standard normal space. Because SVM is a classification-based approach, it can handle discontinuous responses and, more importantly, several limit states using one single SVM. Several analytical examples are used to demonstrate the methodology.

1 INTRODUCTION

In engineering and biomedical problems, input variables are often statistically dependent. Depending on the strength of the dependence, it can easily be shown that neglecting the dependence can lead to erroneous and non conservative reliability assessments. Although there are efficient techniques to calculate probabilities of failure in specific cases (e.g., normal distributions with inexpensive function evaluations), there is

a need for general approaches. Of paramount importance, techniques are needed to estimate probabilities of failure in the case of highly nonlinear limit-state functions and arbitrary distributions while requiring a reasonable number of function evaluations.

In the case of correlated variables, Nataf [1,2] and Rosenblatt transformations [3] are well-known and commonly used methods. These approaches map the dependent variable space onto the uncorrelated standard normal space. While the Nataf transformation only requires the marginal distributions, the Rosenblatt transformation is based on the knowledge of the joint probability density function, which is not always available. In addition, the Rosenblatt transformation is not invariant with respect to the order by which the variables are transformed. As a consequence, approximation methods such as First Order Reliability Method (FORM) would lead to different results for different orderings. Although the Nataf transformation is attractive since it uses the marginals only, its limitations and accuracy for non-normal distributions have now been recognized [4].

As an alternative to Nataf and Rosenblatt transformations, copulas are a more recent and general approach to model the dependence between variables. The basic idea behind copulas is to create a joint cumulative density function (CDF) of a random vector by “linking” (that is in fact the etymology of the word *copula*) the *marginal* CDFs. The major advantage of copulas

stems from their generality and that they are based on marginals. In addition, copulas can be fitted to data to obtain a dependence model and, if needed, generate dependent samples. There exist several families of copulas (Gaussian, Student's, Archimedean copulas [5]) with the Gaussian and t-copulas being the most widely used. The construction of Archimedean copulas is based on various types of *generator* functions, which implies that theoretically an Archimedean copula can be extended to any dimensions. Construction and random sample generation of an Archimedean copula are already implemented in package 'nacopula' in R [6]. A more in depth description of copulas is provided in Section 3.1.

The ability to generate dependent samples solves only one issue towards the calculation of probabilities of failure. In the case of costly function evaluations, a crude Monte Carlo sampling is not possible. One well-known approach to overcome this difficulty is to approximate the responses using surrogates (response surfaces or metamodels). Alternatively, as we propose in this work, the contour of the failure domain can be estimated explicitly using Support Vector Machines (SVMs) [7, 8, 9]. An SVM classifier has the ability to define highly nonlinear contours representing the boundaries of non-convex and disjoint failure domains. Because it is based on classification, SVM is well suited for problems with multiple failure modes (e.g., for a series system, the failure of one mode allows one to not evaluate the other modes) and discontinuous responses. In addition, an adaptive sampling scheme has been developed to construct an accurate SVM boundary [10, 11], a crucial aspect for the calculation of probabilities of failure.

It is shown in this article that, because of the “radial symmetry” of the distribution of probability densities in the original space, the sampling scheme to obtain an accurate boundary cannot be performed in the original space. For this reason, the boundary of the failure domain is constructed in the uncorrelated standard normal space which offers a radial symmetry of the probability densities. This important and fundamental aspect is developed in Section 4.1. The basic idea of the methodology is therefore to adaptively construct an SVM-based boundary of the failure domain in the uncorrelated standard normal space while samples are generated in the original space using a copula. The Monte-Carlo samples generated using the copula are then mapped onto the standard normal space, allowing one to calculate the probability of failure. A transformation is needed to map the samples to and from the uncorrelated standard normal space. For instance, in the adaptive sampling scheme to construct the SVM in the standard space, the “class” (failure or safe) of each adaptive sample must be determined. For this purpose, the sample is mapped “back” to the original space where a function evaluation (e.g., a finite element analysis) can be performed. In this work, the Nataf transformation is

used as a transformation between the two spaces. Note that the Nataf transformation is *not* used here as a *probabilistic transformation*. Only in a few cases (e.g., normal distributions with “linear” correlation), the Nataf transformation can be used to obtain accurate probability estimates. Note that because the joints distribution can be obtained directly from the copulas, the Rosenblatt transformation could have been used to map the samples to the uncorrelated standard normal space. However, the Rosenblatt transformation is tedious to implement and not easily scalable.

Another incentive for using the standard normal space is that the Monte Carlo samples are located within a hypersphere. As the number of dimensions increases, the volume of a hypersphere becomes much smaller than the corresponding hypercube. Therefore, by constraining the construction of the SVM failure boundary within a hypersphere, the effects of the so-called “curse of dimensionality” are mitigated.

The paper is organized as follows. Section 2 describes the basic methodology. Section 3.1 introduces the concept of copula. Section 3.2 provides background on SVMs. Section 4.2 described the SVM-based adaptive sampling scheme. Finally, Section 5 provides several examples of analytical problems with various probabilistic distributions. The example section also describes an algorithm for generating dependent random samples using a copula. In one of the examples, the probabilities obtained using a copula are compared to the results when Nataf transformation is used as a probabilistic transformation. Finally, the Appendix 6 describes the fitting of copulas from data.

2 SUMMARY OF APPROACH

The basic methodology to calculate probabilities of failure with dependent variables is presented in this section. There are three basic ingredients: the construction of an explicit boundary of the failure domain in the standard normal space, the generation of Monte-Carlo samples in the (original) space of dependent variables, and a transformation to map samples between the original and the uncorrelated standard normal space.

The basic procedure for constructing the failure domain boundary in the uncorrelated standard normal space and estimating the probability of failure is given by the following steps:

1. Generation of dependent samples in the original space.
2. Generation of an initial uniform design of experiments (DOE) in the uncorrelated standard normal space.
3. Transformation of the DOE samples to the original space for response evaluation and class (fail or safe) assignment. In this work, we use the Nataf transformation to forward-

and back map samples between the original space to the uncorrelated standard normal space. Note that the Nataf transformation will *not* be used to generate the samples in the original dependent space. This will be done by a copula.

4. Construction of an initial SVM approximation of the boundary of the failure space in the uncorrelated standard space.
5. Calculation of initial probability of failure estimate based on the SVM approximation, using all the Monte Carlo samples (generated by the copula) transformed from the original space.
6. Refinement of the limit-state approximation using adaptive sampling (see Section 4.2). Each additional sample is transformed back to the original space to obtain the corresponding responses. The probability of failure estimate is updated after evaluation of each sample.

The authors reiterate that the Nataf transformation is not used to generate the dependent samples. In other words, it is not used as a probabilistic transformation but as one possible (arbitrary) mapping. This point is essential as Nataf can be shown to introduce large errors in the probability estimate if used as a probabilistic transformation.

3 BACKGROUND

3.1 Copulas

Copulas are used to model the dependence of multiple random variables [5]. The basic idea is to “link” the marginal cumulative distributions F_{X_i} distributions instead of the variables themselves. This aspect makes copulas one of the most general and flexible tools to model dependencies as only individual marginal distributions are needed. It has been used successfully in disciplines such as economics [12], biostatistics [13], finance [14], hydrology [15], as well as engineering design [4].

Definition of a copula is given as:

Consider n variables with marginal cumulative distribution functions of $F_{X_1}, F_{X_2}, \dots, F_{X_n}$ assumed continuous. From the Sklar theorem [5], there exists a copula C , such that

$$H(x_1, x_2, \dots, x_n) = C(F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)) \quad (1)$$

where H is the joint cumulative distribution function of the variables. Therefore $C: [0, 1]^n \rightarrow [0, 1]$ is a joint cumulative function with marginal cumulative distribution functions as input variables.

There are several types of copulas available. The commonly used ones are Gaussian and Student's t copulas which are also referred to as elliptical copulas due to the shape of iso-probability contour. Another family of copulas are the Archimedean ones [5], which can then be classified into one-parameter families and

two-parameter families. Some of the copula functions and their parameters are listed in Tab. 1. Note that the elliptical copulas are the ones whose extension to multidimensional problems ($n > 2$) is the simplest. In practice, Archimedean copulas can be constructed in higher dimensions using a “nested” approach [6].

Once a copula has been constructed, Monte-Carlo samples can be generated following various approaches [12, 16]. In addition, for real world applications, the copula can be determined by fitting it to actual data. Comparison between various types of copula to a certain dataset can be found in [17]. A basic description of the fitting process is provided in Section 6.

3.2 Support Vector Machines (SVMs)

SVM [7, 8] is a machine learning technique widely used for classification. In reliability assessment, the use of SVM to approximate highly nonlinear limit-state functions has been shown in several papers by the authors [18, 9, 10, 11]. The two most important features of SVMs are their ability to handle multiple failure modes using a single classifier and to be insensitive to discontinuities.

An SVM defines an explicit boundary that separates samples belonging to two classes labeled as $+1$ and -1 . Given a set of N training samples \mathbf{u}_i in an n -dimensional space and the corresponding class labels, an SVM boundary is given as:

$$s(\mathbf{u}) = b + \sum_{i=1}^N \lambda_i y_i K(\mathbf{u}_i, \mathbf{u}) = 0 \quad (2)$$

where b is a scalar referred to as the bias, λ_i are Lagrange multipliers obtained from the quadratic programming optimization problem used to construct the SVM, and K is a kernel function. The classification of any arbitrary point \mathbf{u} is given by the sign of $s(\mathbf{u})$. The training samples for which the Lagrange multipliers are non-zero are referred to as the *support vectors*. The kernel function K in Eqn. (2) can have several forms, such as polynomial or Gaussian radial basis kernel. The Gaussian kernel (Eqn. (3)) is used in this article.

$$K(\mathbf{u}_i, \mathbf{u}_j) = \exp\left(-\frac{\|\mathbf{u}_i - \mathbf{u}_j\|^2}{2\sigma^2}\right) \quad (3)$$

where σ is the width parameter.

4 BASIC METHODOLOGY

4.1 Why not construct the limit state in the original space?

Using copulas, observations of dependent variables (X_1, X_2, \dots, X_n) can be generated in the original space. Therefore,

TABLE 1: TYPES OF COPULA FUNCTIONS

Type	$C(v_1, v_2, \dots, v_n)$	Parameter domain
<i>Gaussian</i>	$\Phi_R(\Phi^{-1}(v_1), \Phi^{-1}(v_2), \dots, \Phi^{-1}(v_n))^*$	$R_{ij} \in [-1, 1]$
<i>Student</i>	$t_{R,v}(t_v^{-1}(v_1), t_v^{-1}(v_2), \dots, t_v^{-1}(v_n))^{\dagger}$	$R_{ij} \in [-1, 1], v \in (0, \infty)$
<i>Clayton</i>	$[\max(v_1^{-\theta} + v_2^{-\theta} - 1, 0)]^{-1/\theta}$	$\theta \in (0, \infty)$
<i>Gumbel-Hougaard</i>	$\exp\left(-\left[(-\ln v_1)^\theta + (-\ln v_2)^\theta\right]^{1/\theta}\right)$	$\theta \in [1, \infty)$
<i>Frank</i>	$-(1/\theta) \ln(1 + (e^{-\theta v_1} - 1)(e^{-\theta v_2} - 1)/(e^{-\theta} - 1))$	$\theta \in (\infty, \infty) \setminus \{0\}$
<i>FGM</i> [‡]	$v_1 v_2 + \theta v_1 v_2 (1 - v_1)(1 - v_2)$	$\theta \in [-1, 1]$

* Φ_R is the standard multivariate normal distribution with correlation matrix R .

[†] $t_{R,v}$ is the standard multivariate Student's t distribution with correlation matrix R and v degrees of freedom.

[‡] Short for Farlie-Gumbel-Morgenstern.

it might seem natural and intuitive to build the boundary of the failure domain in the original dependent space and use the Monte-Carlo samples to calculate the probability of failure. However, the original space might exhibit a pronounced non-uniformity of Monte-Carlo sample densities. Therefore, any adaptive sampling scheme to construct the SVM boundary (see Section 4.2) would need to locate samples in the vicinity of high density regions. To avoid this “lack of radial symmetry” of the densities, this article proposes to construct the SVM in the standard normal space which presents an isotropy (more precisely, a radial symmetry) of the densities around the means.

The latter point is more effectively demonstrated graphically. Fig. 1 and 2 depicts an example of the transformation of the iso-density contour in the original space and the standard space. In the standard space, consider 8 lines with 15 points uniformly distributed on each. The blue circle represents an iso-density contour. Using a Gumbel bivariate copula with exponential distributions ($\theta = 1$), the contour and the points are transformed back to the original space (Fig. 2). The 8 lines become 8 curves and points on each line are clearly not uniform any longer. In fact, inspection shows that the density of points is much larger close to the origin. Therefore, an adaptive sampling scheme, would require more samples around the origin.

Another important motivation to perform the construction in the standard space, is that the adaptive sampling can be performed within a hypersphere since the distributions are normal. Because the volume of a hypersphere is far less than the volume of the corresponding hypercube, this is beneficial to mitigate the “curse of dimensionality” in higher dimensions. The following section provides the detail of the adaptive sampling scheme constrained within a hypersphere.

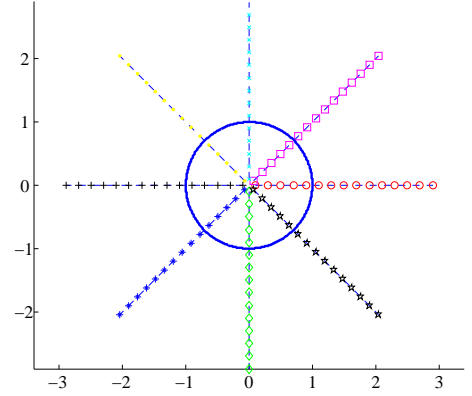


FIGURE 1: 8 LINES IN THE STANDARD SPACE WITH UNIFORMLY SPACED SAMPLES ON EACH.

4.2 Adaptive sampling and Explicit Design Space Decomposition (EDSD)

We wish to construct the explicit boundary of the failure space in the standard space using an SVM. For this purpose, the explicit design space decomposition (EDSD) approach will be used [11, 19]. First, an initial uniform DOE such as Central Voronoi Tessellation (CVT) is used. Each sample is then mapped back to the original space and evaluated in order to obtain its class (fail or safe) to build the first SVM boundary. Then the boundary is refined using an adaptive sampling scheme as described below. As explained in Section 5, a novelty of this work is that all static DOEs and adaptive samples are selected within a hypersphere of radius r centering at the origin.

The first type of sample (referred to as primary sample) of

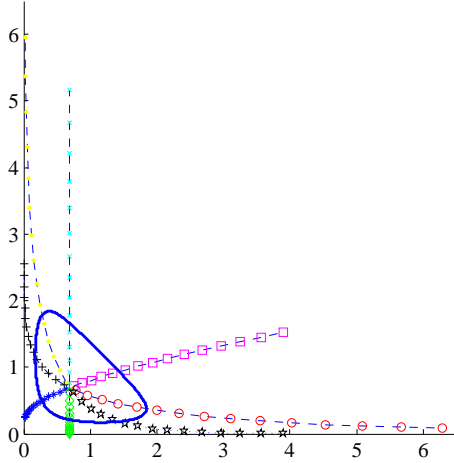


FIGURE 2: TRANSFORMATION OF 8 LINES BACK TO THE ORIGINAL SPACE. THE DENSITIES OF SAMPLES ARE NO LONGER UNIFORM AND ARE LARGER IN THE VICINITY OF THE ORIGIN.

the EDS adaptive sampling scheme is found by searching a point in the standard space that is located as far as possible from existing samples and also has a high probability of misclassification. In addition, the sample should be within a hypersphere. Formally, this sample is found by solving the following optimization problem:

$$\begin{aligned} \max_{\mathbf{u}} \quad & ||\mathbf{u} - \mathbf{u}_{nearest}|| \\ \text{s.t.} \quad & s(\mathbf{u}) = 0 \\ & ||\mathbf{u}|| \leq r \end{aligned} \quad (4)$$

The radius of the hypersphere r is chosen as the maximum distance from the center to all the samples in the uncorrelated standard normal space. These samples are obtained by transforming all the samples generated in the original space to the standard space using the Nataf transformation.

To quantify the advantage of performing EDS in a hypersphere, let us calculate the volume of the hypersphere. The volume of a n -dimensional hypersphere [20] with radius r is expressed as:

$$V_{n_sphere}(r) = \frac{\pi^{n/2}}{\Gamma(n/2 + 1)} r^n \quad (5)$$

where Γ is the Gamma function. Thus, the volume of a n -

dimensional hypersphere with radius r can be rewritten as:

$$V_{n_sphere}(r) = \begin{cases} \frac{2^{(n+1)/2} \pi^{(n-1)/2}}{n!!} r^n & n \text{ is odd} \\ \frac{\pi^{n/2}}{(n/2)!} r^n & n \text{ is even} \end{cases} \quad (6)$$

where $n!!$ is the double factorial of n . Thus, for a n -dimensional hypersphere inscribed in a hypercube, the ratio of the two volumes is:

$$\rho(n) = \begin{cases} \frac{\pi^{(n-1)/2}}{2^{(n-1)/2} n!!} & n \text{ is odd} \\ \frac{\pi^{n/2}}{2^n (n/2)!} & n \text{ is even} \end{cases} \quad (7)$$

The ratio shows that as n increases, the volume of the hypersphere becomes much smaller than the corresponding hypercube.

A secondary sample is selected in a region in the vicinity of SVM boundary that has the highest local unbalance of data from the two classes. Unbalance of data is quantified as $|d_-(\mathbf{u}) - d_+(\mathbf{u})|$, where $d_-(\mathbf{u})$ and $d_+(\mathbf{u})$ are the distances to the closest -1 and $+1$ samples. A high unbalance of data may be an indicator of a phenomenon known as “locking” of SVM, which may result in a slow rate of convergence of SVM to the actual limit-state function. Although change in SVM due to primary samples may be negligible in such cases, secondary samples are useful in removing locking. This is shown conceptually in Fig.

3. Selection of a secondary sample is a two step process:

1. Selection of a point \mathbf{u}_c on the SVM boundary $s(\mathbf{u}) = 0$ with highest measure of unbalance.

$$\begin{aligned} \max_{\mathbf{u}_c} \quad & (d_-(\mathbf{u}_c) - d_+(\mathbf{u}_c))^2 \\ \text{s.t.} \quad & s(\mathbf{u}_c) = 0 \\ & ||\mathbf{u}_c|| \leq r \end{aligned} \quad (8)$$

2. Selection of secondary sample within a hypersphere centered at \mathbf{u}_c .

$$\begin{aligned} \min_{\mathbf{u}} \quad & \text{sign}(d_-(\mathbf{u}_c) - d_+(\mathbf{u}_c))s(\mathbf{u}) \\ \text{s.t.} \quad & ||\mathbf{u} - \mathbf{u}_c|| - \frac{1}{4} |d_-(\mathbf{u}_c) - d_+(\mathbf{u}_c)| \leq 0 \\ & ||\mathbf{u}|| \leq r \end{aligned} \quad (9)$$

In this work, the EDS process adds two primary samples and one secondary sample at each iteration. More details about SVM “locking” can be found in [11].

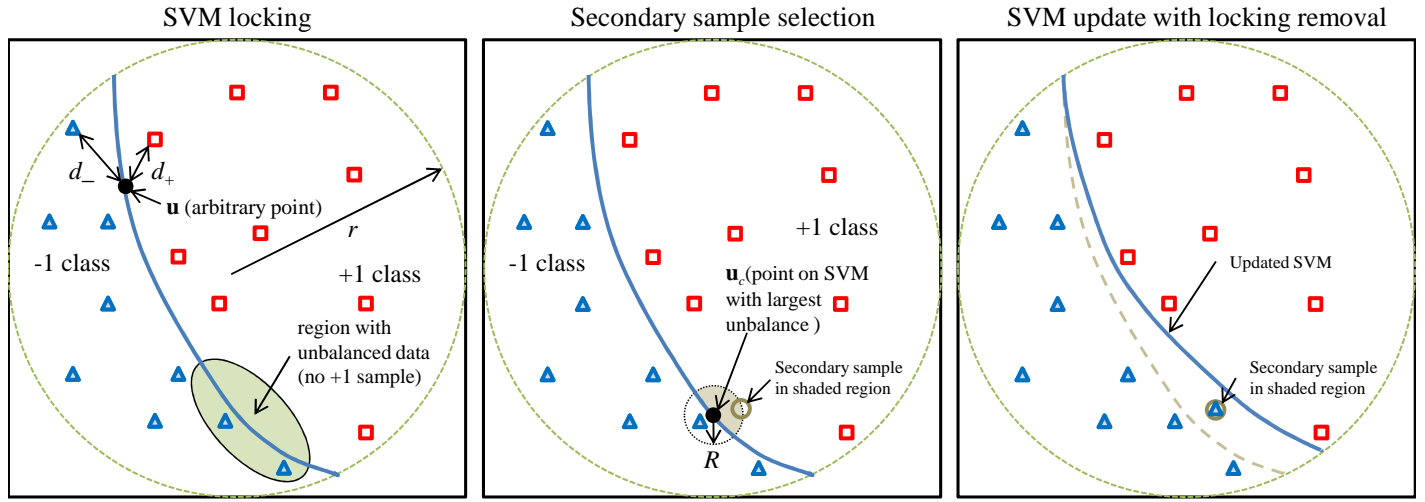


FIGURE 3: LOCKING OF SVM WITH LOCALLY UNBALANCED DATA IN THE VICINITY OF SVM (LEFT), SELECTION OF A SECONDARY SAMPLE (MID), AND UPDATE OF SVM DUE TO THE EVALUATED SECONDARY SAMPLE (RIGHT).(REVISED FROM [11].)

5 EXAMPLES AND RESULTS

Several analytical examples are used to demonstrate the methodology. The examples chosen have known joint distributions from which the marginals can be derived. These marginals are then “linked” using a specific copula. Note that in engineering applications, the joint distribution is typically not known and the examples are only used to demonstrate the methodology. In addition, the proposed analytical examples allow one to obtain the actual probability of failure which would otherwise vary for different copulas used to generate the dependent Monte-Carlo Samples. The first example is tailored to demonstrate the errors that can be made if Nataf is used traditionally as a probabilistic transformation.

For the two-dimensional problems, the dependent samples are generated based on a copula C (exact for our problems). The conditional distribution function of V_2 knowing $V_1 = v_1$ (denoted as $c_{v_1}(v_2)$) can be found as:

$$\begin{aligned} c_{v_1}(v_2) &= P[V_2 \leq v_2 | V_1 = v_1] \\ &= \lim_{\Delta v_1 \rightarrow 0} \frac{C(v_1 + \Delta v_1, v_2) - C(v_1, v_2)}{\Delta v_1} \\ &= \frac{\partial C(v_1, v_2)}{\partial v_1} \end{aligned} \quad (10)$$

The algorithm for generating dependent samples for a given joint distribution function can be summarized as follows:

1. Generate two independent uniform $(0, 1)$ variates v_1 and t ;

2. Set $v_2 = c_{v_1}^{-1}(t)$;
3. Transform using inverse marginal distribution function.
 $x_1 = F_{X_1}^{-1}(v_1)$ and $x_2 = F_{X_2}^{-1}(v_2)$
4. The desired samples are (x_1, x_2) .

In each example, the initial DOE is generated in the standard normal space for the construction of the first approximated failure boundary. The DOE samples are selected to be uniformly spaced (using a max-min distance algorithm). As described in the methodology section, the adaptive sampling in the standard space is constrained within a hypersphere centered at the origin with a radius defined as the distance to the farthest Monte-Carlo sample.

The convergence plots of the estimated probability of failure are given for each problem along with a comparison to the reference probability which is calculated using the actual limit state function in the original space. For all the test examples, the Monte-Carlo simulations are performed using 10^6 samples. For the sake of completeness, the efficiency of the adaptive sampling scheme is compared to the results obtained with a Centroidal Voronoi Tessellations (CVT) [21, 22] DOEs of various size.

5.1 Example 1

Consider the Gumbel's bivariate joint distribution function with parameter set to unity [23, 24]:

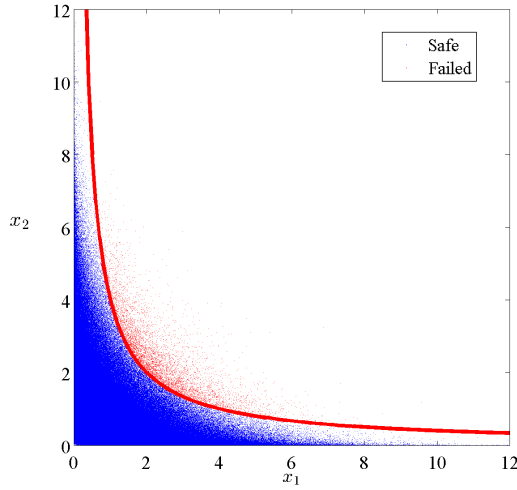


FIGURE 4: MONTE-CARLO SAMPLES IN THE ORIGINAL SPACE USING THE COPULA. FAILURE DOMAIN (RED).

$$F_{\mathbf{X}}(x_1, x_2) = \begin{cases} e^{(-x_1 - x_2 - x_1 x_2)} - e^{-x_1} - e^{-x_2} + 1, & x_1, x_2 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

From which, the marginal distributions for x_1 and x_2 which follow standard exponential distributions are:

$$\begin{aligned} F_{X_1}(x_1) &= 1 - e^{-x_1} \\ F_{X_2}(x_2) &= 1 - e^{-x_2} \end{aligned} \quad (11)$$

For this problem, the copula (i.e., the expression that links the two marginals to the joint distribution $F_{\mathbf{X}}$) can be derived:

$$C(v_1, v_2) = (1 - v_2)(1 - v_1)^{1 - \log(1 - v_2)} + v_1 + v_2 - 1 \quad (12)$$

where $v_1, v_2 \in [0, 1]^2$ are the values of the marginals.

In this problem, the failure domain is defined by one limit-state function is $g(\mathbf{x}) = 4 - x_1 x_2$. A similar example with linear limit-state function was introduced by Hohenbichler and Rackwitz [25], and was discussed by Ditlevsen [26], Madsen [27], and Yoojeong [4] et al.

Fig. 4 depicts the Monte Carlo samples in the original space which are generated from the joint distribution as explained in Section 3.1 and 5. And the algorithm is given at the beginning of this section. The figure also shows the failure region.

Contours of this set of limit-state functions in the physical space and the standard space are depicted in Fig. 5 and 6, respectively.

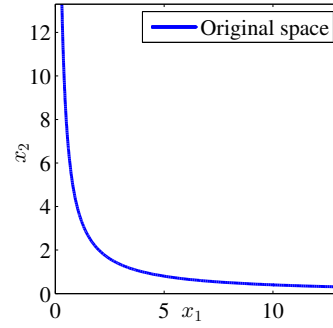


FIGURE 5: LIMIT-STATE FUNCTION IN THE ORIGINAL SPACE FOR EXAMPLE 1

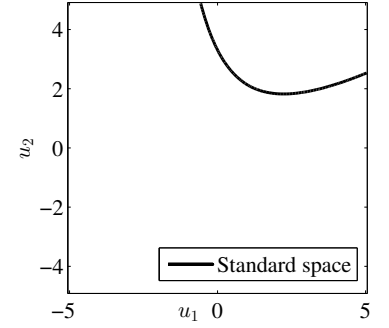


FIGURE 6: LIMIT-STATE FUNCTION IN THE STANDARD NORMAL SPACE FOR EXAMPLE 1

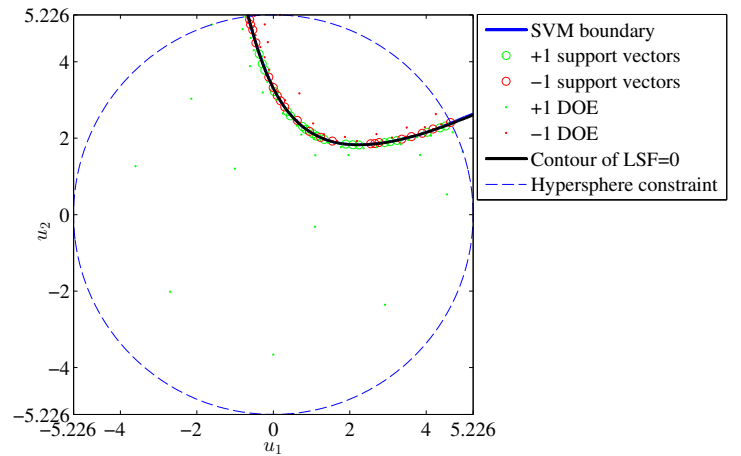


FIGURE 7: SVM APPROXIMATING THE BOUNDARY OF THE FAILURE REGION FOR EXAMPLE 1.

5.1.1 Dependent samples generated using the copula

The final SVM boundary (blue) and the actual contour of limit-state function $g(\mathbf{u}) = 0$ (black) are depicted in Fig. 7 as well as the constraining circle. Failure probabilities are calculated after adding each adaptive sample and relative errors of these probabilities are shown in Fig. 8. The relative error of the probability of failure is under 5% when the number of samples reaches 38. After adding 54 adaptive samples, it is observed that the probability of failure has an error less than 3% compared to the reference failure probability.

The accuracy of the probability of failure is also compared to the construction of the SVM using a CVT DOE only. For this two dimensional problem, the size of the DOE is increased by increments of 5 samples. The results are gathered in Fig. 9. From the results, it appears that the proposed scheme is more

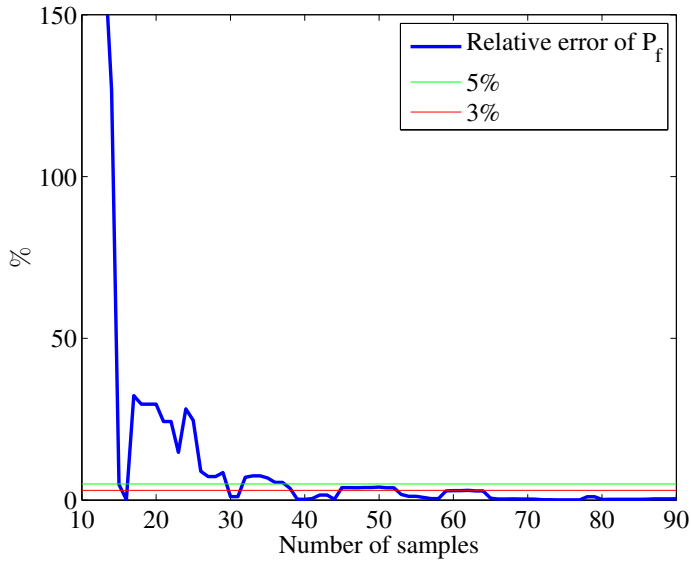


FIGURE 8: RELATIVE ERROR OF PROBABILITY OF FAILURE V.S. NUMBER OF SAMPLES.

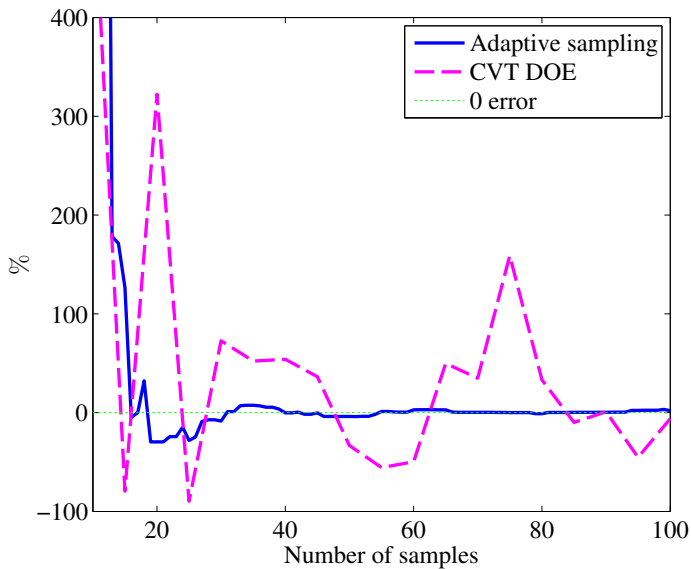


FIGURE 9: COMPARISON OF P_f CALCULATION USING ADAPTIVE SAMPLING SCHEME AND CVT DOE FOR EXAMPLE 1.

accurate and robust than the DOE approach.

5.1.2 Testing Nataf as a probabilistic transformation.

In order to show the relevance of using copulas, the

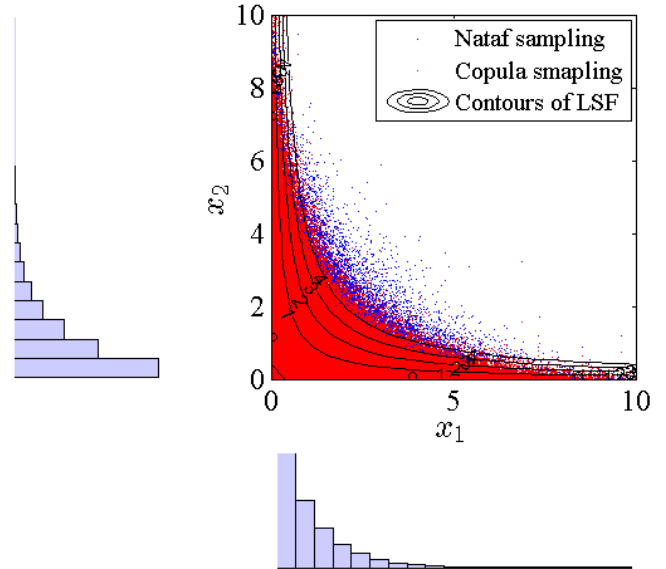


FIGURE 10: SAMPLES GENERATED USING COPULA (RED) METHOD AND INVERSE NATAF TRANSFORMATION (BLUE) AND THE CONTOURS OF LIMIT-STATE FUNCTIONS.

results are compared to the probability estimates if one uses the Nataf transformation as a probability transformation. Practically this means that the actual dependent Monte-Carlo samples are not mapped to the uncorrelated standard normal space but the probability is now calculated using a multivariate uncorrelated standard normal distribution.

In order to compare the copula and the Nataf approaches, several levels of probabilities will be studied. To do so, the limit-state function is parameterized: $g(\mathbf{x}) = a - x_1x_2$, where a varies from 0 to 4. Contours of this set of limit-state functions are depicted in Fig. 10. The contour of the limit-state function moves away from the origin as a increases. The corresponding probability of failure should decrease. The red samples in Fig. 10 are generated from the copula while the blue samples are generated by uncorrelated sampling in the standard normal space and transformed back to the physical space using inverse Nataf transformation. For each a in the limit-state function, probabilities of failure can be calculated using both sets of samples. Difference between the probabilities are studied in Fig. 11. In this case, the copula sampling method uses the information of the joint distribution function while Nataf transformation only uses the (linear) correlation coefficient of variables x_1, x_2 . The copula-based probability being exact, the plot clearly demonstrates that the Nataf probabilistic transformation will lead to large errors as the probability for small probabilities. The plot also depicts the

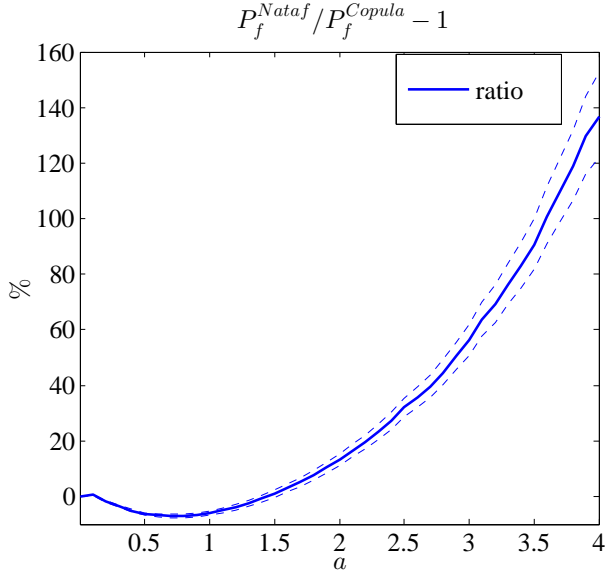


FIGURE 11: RATIO OF PROBABILITIES OF FAILURE CALCULATED USING INVERSE NATAF TRANSFORMATION AND COPULA W.R.T INCREASING a IN LSF. THE DASHED LINES CORRESPOND TO THE 95 % CONFIDENCE INTERVAL.

95% confidence interval.

5.2 Example 2

The parametric bivariate logistic distribution [5] promoted by Ali [28] has the joint distribution function H_θ as

$$F_{\mathbf{x},\theta}(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2} + (1 - \theta)e^{-x_1 - x_2})^{-1} \quad (13)$$

where $x_1, x_2 \in [-\infty, \infty]$ and $\theta \in [-1, 1]$. The marginal distributions are standard logistic distribution as shown in Equation 14.

$$\begin{aligned} F_{X_1}(x_1) &= (1 + \exp(-x_1))^{-1} \\ F_{X_2}(x_2) &= (1 + \exp(-x_2))^{-1} \end{aligned} \quad (14)$$

When $\theta = 1$, the joint distribution function becomes Gumbel's bivariate logistic distribution whose joint distribution function is

$$F_{\mathbf{x}}(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1} \quad (15)$$

and when $\theta = 0$, the input variables X and Y are independent.

The copula of X_1 and X_2 can be found as

$$C_\theta(v_1, v_2) = \frac{v_1 v_2}{1 - \theta(1 - v_1)(1 - v_2)} \quad (16)$$

All the Monte Carlo samples in the physical space which are generated using the copula in Equation 16 are shown in Fig. 12. Histogram plots for x_1 and x_2 are also given.

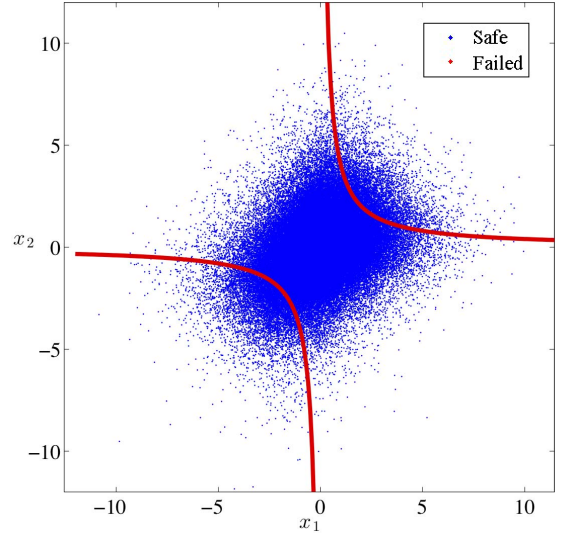


FIGURE 12: SAMPLING IN THE ORIGINAL PHYSICAL SPACE USING SKLAR COPULA.

The contour of limit-state function $g(\mathbf{x}) = 0$ in the original physical space is depicted as the red curves in Fig. 12. And after applying the nonlinear transformation of all the Monte Carlo samples to the standard space, the contour of limit-state function $g(\mathbf{u}) = 0$ is shown in Fig. 13.

For the sake of evaluating the efficiency of adaptive sampling scheme, the same tests with CVT DOE are carried out. Probabilities of failure are calculated for each DOE. The results of the probabilities are shown in Fig. 16.

6 CONCLUSION

Using copulas and SVMs, this work introduces an approach to evaluate probabilities of failure when the random variables are dependent. The technique is based on the explicit construction of the limit-state function in the uncorrelated standard normal space. The Monte Carlo samples are generated in the original dependent space directly with the help of copula and then

TABLE 2: RESULTS OF EXAMPLES.

Results	Example 1	Example 2
$P_f^{Ref*} (\times 10^{-2})$	0.608	1.209
Size of initial DOE	10	30
No. of samples when $\epsilon^{SVM} < 5\%$	38	59
No. of samples when $\epsilon^{SVM} < 3\%$	64	93
95% CI of $P_f^{Ref} (\times 10^{-2})$	[0.593, 0.624]	[1.140, 1.278]

* P_f^{Ref} is calculated using the analytical limit-state function.

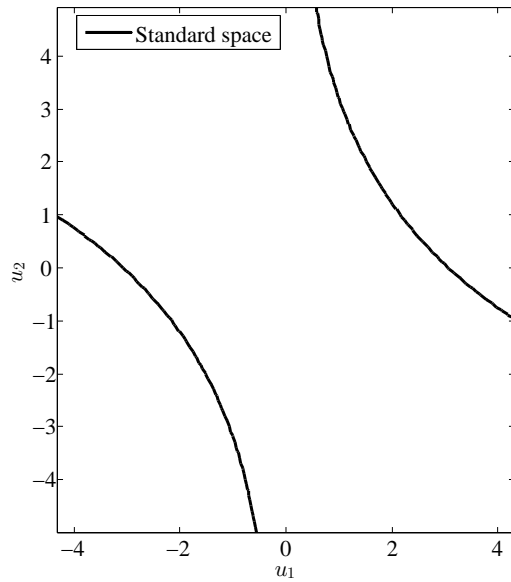


FIGURE 13: LIMIT-STATE FUNCTION IN THE STANDARD SPACE FOR EXAMPLE 2.

transformed to the uncorrelated standard normal space. The construction of an accurate limit-state function is based on SVM explicit design space decomposition with adaptive sampling scheme. One of the key conclusions of this work is that the standard normal space offers radial symmetry but also enables one to mitigate the effect of the curse of dimensionality by sampling within a hypersphere.

The next steps of this research will test this methodology for higher dimensional problems. In addition, we will test this methodology on cases where the copula is constructed from data.

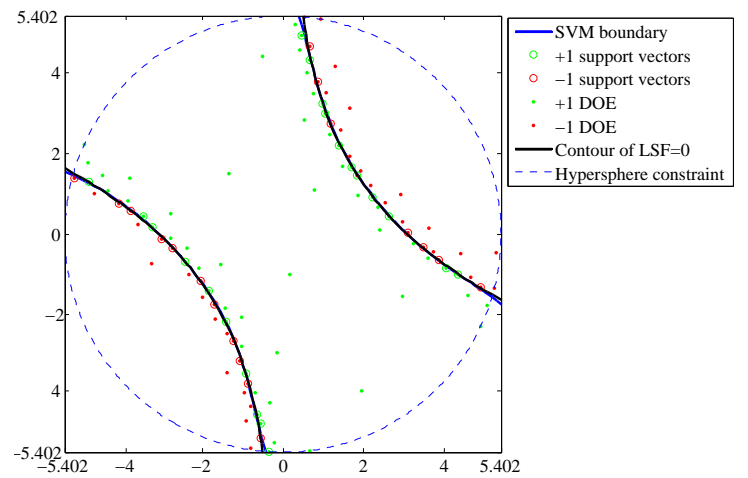


FIGURE 14: SVM BOUNDARY BETWEEN SAFE AND FAILED SAMPLES IN THE STANDARD SPACE.

REFERENCES

- [1] Nataf, A., 1962. “Détermination des distributions de probabilités dont les marges sont données”. *Comptes rendus de l’académie des sciences*, **225**, pp. 42–43.
- [2] Liu, P., and Der Kiureghian, A., 1986. “Multivariate distribution models with prescribed marginals and covariances”. *Probabilistic Engineering Mechanics*, **1**(2), pp. 105–112.
- [3] Rosenblatt, M., 1952. “Remarks on a multivariate transformation”. *The Annals of Mathematical Statistics*, **23**(3), pp. 470–472.
- [4] Noh, Y., Choi, K., and Du, L., 2009. “Reliability-based design optimization of problems with correlated input variables using a Gaussian Copula”. *Structural and Multidisciplinary Optimization*, **38**(1), pp. 1–16.
- [5] Nelsen, R., 2006. *An introduction to copulas*. Springer Verlag.
- [6] Hofert, M., and Mächler, M., 2011. “Nested Archimedean

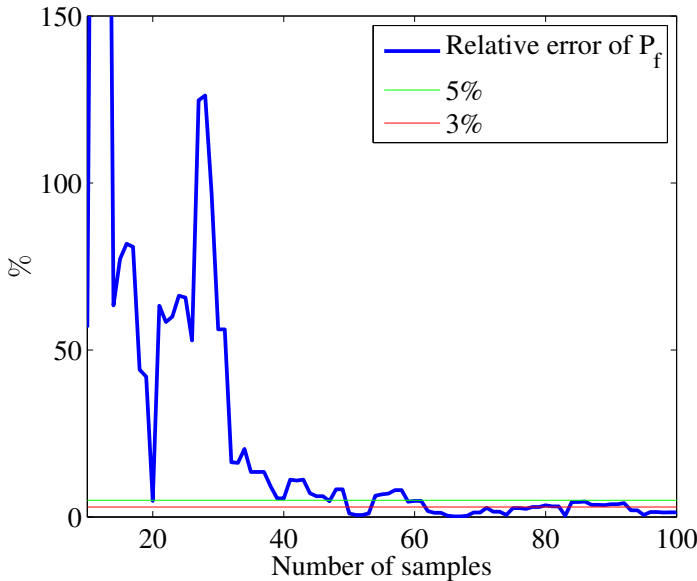


FIGURE 15: RELATIVE ERROR OF PROBABILITY OF FAILURE V.S. NUMBER OF SAMPLES.

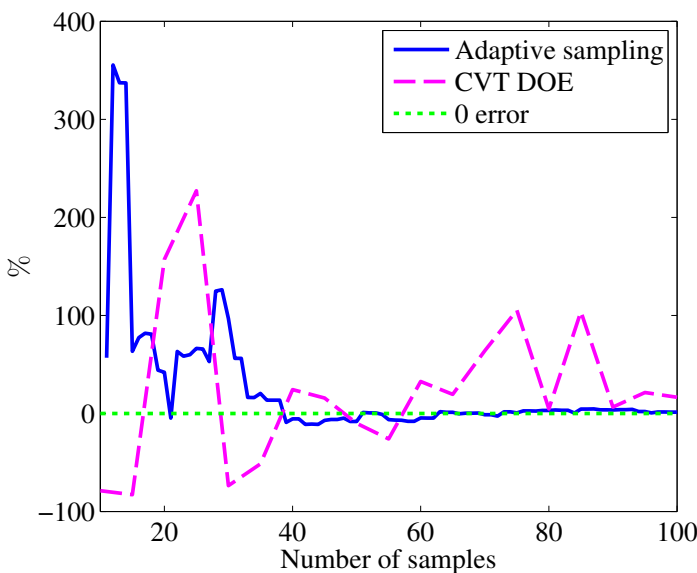


FIGURE 16: COMPARISON OF P_f CALCULATION USING ADAPTIVE SAMPLING SCHEME AND OTHER DOE FOR EXAMPLE 2.

copulas meet R: the nacopula package”. *Journal of Statistical Software*, **39**(9), pp. 1–20.

[7] Vapnik, V.N., 1998. *Statistical Learning Theory*. John Wiley & Sons.

- [8] Gunn, S.R., 1998. Support vector machines for classification and regression. Tech. Rep. ISIS-1-98, Department of Electronics and Computer Science, University of Southampton.
- [9] Basudhar, A., Missoum, S., and Harrison Sanchez, A., 2008. “Limit state function identification using Support Vector Machines for discontinuous responses and disjoint failure domains”. *Probabilistic Engineering Mechanics*, **23**(1), pp. 1–11.
- [10] Basudhar, A., and Missoum, S., 2008. “Adaptive explicit decision functions for probabilistic design and optimization using support vector machines”. *Computers & Structures*, **86**(19–20), pp. 1904–1917.
- [11] Basudhar, A., and Missoum, S., 2010. “An improved adaptive sampling scheme for the construction of explicit boundaries”. *Structural and Multidisciplinary Optimization*, **42**(4), pp. 517–529.
- [12] Frees, E., and Valdez, E., 1998. “Understanding relationships using copulas”. *North American actuarial journal*, **2**(1).
- [13] Li, D., 2000. “On default correlation: a copula function approach”. *Journal of Fixed income*, **9**(4), pp. 43–54.
- [14] Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T., 2000. “Copulas for finance: A reading guide and some applications”. *Manuscript, Financial Econometrics Research Center*.
- [15] Dupuis, D., 2007. “Using copulas in hydrology: Benefits, cautions, and issues”. *Journal of Hydrologic Engineering*, **12**, p. 381.
- [16] Yan, J., 2007. “Enjoy the joy of copulas: With a package copula”. *Journal of Statistical Software*, **21**(4), pp. 1–21.
- [17] Genest, C., and Favre, A., 2007. “Everything you always wanted to know about copula modeling but were afraid to ask”. *Journal of Hydrologic Engineering*, **12**, p. 347.
- [18] Hurtado, Jorge E., 2004. “An examination of methods for approximating implicit limit state functions from the viewpoint of statistical learning theory”. *Structural Safety*, **26**, July, pp. 271–293.
- [19] Basudhar, A., and Missoum, S., 2009. “A sampling-based approach for probabilistic design with random fields”. *Computer Methods in Applied Mechanics and Engineering*, **198**(47–48), pp. 3647 – 3655.
- [20] Hamming, R., 1986. *Coding and information theory*. Prentice-Hall, Inc.
- [21] Romero, Vincente J., Burkardt, John V., Gunzburger, Max D., and Peterson, Janet S., 2006. “Comparison of pure and “latinized” centroidal voronoi tessellation against various other statistical sampling methods”. *Journal of Reliability Engineering and System Safety*, **91**(1266–1280).
- [22] Du, Q., Faber, V., and Gunzburger, M., 1999. “Centroidal Voronoi tessellations: applications and algorithms”. *SIAM review*, **41**(4), pp. 637–676.

- [23] Melchers, R., 1999. *Structural reliability analysis and prediction*, Vol. 437. John Wiley & Sons.
- [24] Dolinski, K., 1983. “First-order second-moment approximation in reliability of structural systems: critical review and alternative approach”. *Structural Safety*, **1**(3), pp. 211–231.
- [25] Hohenbichler, M., and Rackwitz, R., 1981. “Nonnormal dependent vectors in structural reliability”. *Journal of the Engineering Mechanics, ASCE*, **107**(6), pp. 1227–1238.
- [26] Ditlevsen, O., and Madsen, H., 1996. *Structural reliability methods*, Vol. 178. Citeseer.
- [27] Madsen, H., Krenk, S., and Lind, N., 1986. “Methods of structural safety”. *Englewood Cliffs, NJ*.
- [28] Ali, M., Mikhail, N., and Haq, M., 1978. “A class of bivariate distributions including the bivariate logistic”. *Journal of multivariate analysis*, **8**(3), pp. 405–412.
- [29] Genest, C., and Rivest, L., 1993. “Statistical inference procedures for bivariate archimedean copulas”. *Journal of the American Statistical Association*, pp. 1034–1043.
- [30] Deheuvels, P., 1979. “La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d’indépendance”. *Acad. Roy. Belg. Bull. Cl. Sci.*(5), **65**(6), pp. 274–292.
- [31] Joe, H., and Xu, J., 1996. “The estimation method of inference functions for margins for multivariate models”. *Department of Statistics, University of British Columbia, Technical Report*, **166**.
- [32] Noh, Y., Choi, K., and Lee, I., 2010. “Identification of marginal and joint CDFs using bayesian method for RBDO”. *Structural and Multidisciplinary Optimization*, **40**(1), pp. 35–51.

where $\mathcal{L}_i(\theta | \mathbf{x}^{(i)})$ is the likelihood for the i^{th} observation,

$$\mathcal{L}_i(\theta | \mathbf{x}^{(i)}) = c(F_{X_1}^{(i)}(x_1), F_{X_2}^{(i)}(x_2), \dots, F_{X_n}^{(i)}(x_n) | \theta) \prod_{j=1}^n f_{X_j}^{(i)}(x_j) \quad (18)$$

In practice it is more convenient to use the log-likelihood function:

$$\begin{aligned} \ln \mathcal{L}(\theta | \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}) &= \sum_{i=1}^m \ln \mathcal{L}_i(\theta | \mathbf{x}^{(i)}) \\ &= \sum_{i=1}^m \ln c(F_{X_1}^{(i)}(x_1), F_{X_2}^{(i)}(x_2), \dots, F_{X_n}^{(i)}(x_n) | \theta) + \\ &\quad \sum_{i=1}^m \sum_{j=1}^n \ln f_{X_j}^{(i)}(x_j) \end{aligned} \quad (19)$$

Normally, the estimation of the parameters of copulas requires numerical optimization process to maximize the log-likelihood function. Based on the idea of using of maximum likelihood in identifying the appropriate copula for given data, Joe and et al [31] introduced a method called Inference Functions for Margins (IFM) to reduce computational cost and Noh and et al [32] proposed to use Bayesian method to identify a copula that fits the data best.

Once the model of the copula is constructed, the proposed procedures in this article can be applied to assess the reliability using Monte Carlo simulation.

Appendix A: Fitting copula to existing data

In engineering applications, there is often a limited amount of data available thus making it difficult to obtain joint distribution information with traditional approaches. Copulas can be used for this purpose to extract dependence information and a joint distribution. Therefore, it is important to determine an appropriate copula to use. Genest and Rivest [29] have developed a non-parametric way to identify the copula in the Archimedean family based on the empirical copula [30]. The alternative parametric method is based on the maximum likelihood estimator.

Consider θ is the vector of parameters needed to be found for a copula with m observations. The likelihood of the joint distribution function is:

$$\mathcal{L}(\theta | \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}) = \prod_{i=1}^m \mathcal{L}_i(\theta | \mathbf{x}^{(i)}) \quad (17)$$