Performance of translation approach for modeling correlated non-normal variables

Dian-Qing Li\textsuperscript{a,}\textsuperscript{*}, Shuai-Bing Wu\textsuperscript{b}, Chuang-Bing Zhou\textsuperscript{a}, K.K. Phoon\textsuperscript{c}

\textsuperscript{a}State Key Laboratory of Water Resources and Hydropower Engineering Science, Key Laboratory of Rock Mechanics in Hydraulic Structural Engineering (Ministry of Education), Wuhan University, 8 Donghu South Road, Wuhan 430072, PR China
\textsuperscript{b}Nanchang Institute of Science and Technology, 998 Chuangye South Road, Nanchang 330108, PR China
\textsuperscript{c}Department of Civil and Environmental Engineering, National University of Singapore, Blk E1A, #07–03, 1 Engineering Drive 2, Singapore 117576, Singapore

A R T I C L E   I N F O

Article history:
Received 27 December 2011
Received in revised form 1 August 2012
Accepted 7 August 2012

Keywords:
Multivariate construction methods
Joint probability distribution
Pearson correlation
Spearman correlation
High order joint moments
Probability of failure

A B S T R A C T

It is common to construct a consistent multivariate distribution from non-normal marginals and Pearson product–moment correlations using the well known translation approach. A practical variant of this approach is to match the Spearman rank correlations of the measured data, rather than the Pearson correlations. In this paper, the performance of these translation methods is evaluated based on their abilities to match the following exact solutions from one benchmark bivariate example where the joint distribution is known: (1) high order joint moments, (2) joint probability density functions (PDFs), and (3) probabilities of failure. It is not surprising to find significant errors in the joint moments and PDFs. However, it is interesting to observe that the Pearson and Spearman methods produce very similar results and neither method is consistently more accurate or more conservative than the other in terms of probabilities of failure. In addition, the maximum error in the probability of failure may not be associated with a large correlation. It can happen at an intermediate correlation.

\textcopyright 2012 Elsevier Ltd. All rights reserved.

1. Introduction

The structural reliability problem is often formulated in terms of a vector of basic random variables characterizing uncertainties in quantities such as loads, material properties, and structural dimensions. To evaluate structural reliability exactly, the complete probability information such as the joint cumulative distribution function (CDF) and the joint probability density function (PDF) of the random vector should be known. In engineering practice, however, the joint PDF is often unknown due to limited data. In most practical applications, only the marginal PDFs and the covariance matrix are available – these incomplete statistical data are quite commonly available [1–3]. Furthermore, the marginal distributions are often non-normal distributions. Based on such limited information, with the exception of the Gaussian case and generalizations such as elliptical and skew elliptical distributions [4], the joint PDFs could not be defined uniquely. Consequently, an exact evaluation of the probability of failure is not possible [5]. To achieve such purpose, the methods for modeling and constructing the approximated multivariate distribution based on incomplete statistical data are needed. It is well accepted that the modeling and simulation of non-Gaussian random vectors are challenging problems [6,7].

In general, there are two common approximate multivariate construction methods [8,9]. One approximate idea is to match the Pearson correlation (referred to as approximate method P hereafter). The other approximation idea is to match the Spearman correlation (referred to as approximate method S hereafter). The approximate method P is based on the Nataf distribution [10,3]. It is well known that the Nataf distribution is the most widely used distribution model to deal with structural reliability problems involving correlated non-normal variables [2,11,12]. For example, Der Kiureghian and Liu [2] derived a set of semi-empirical formulas to calculate the equivalent Pearson correlation coefficients between the transformed standard normal variates using the Nataf model. The approximate method S can also be used for the simulation of non-normal variates [13]. For instance, Phoon [13] used the approximate method S to simulate a non-normal random vector with exponential marginals. Although the two approximate methods are widely used for constructing the multivariate distributions, the accuracy of these approximate methods has not been studied systematically. In particular, it is of practical interest to get a quantitative feel of the errors made by these methods. For example, the approximate methods may estimate the probability of failure with unacceptable errors.

The objective of this study is to evaluate the accuracy of the two approximate methods numerically using one benchmark example where the exact joint distribution is known. To achieve this goal, this paper is organized as follows. In Section 2, the concepts of Pearson correlation coefficient and Spearman correlation coefficient are introduced briefly. In Section 3, two approximate methods for constructing multivariate distributions are presented in detail. The performance of the two approximate methods is evaluated based on their abilities to match exact solutions for: (1) high order joint moments, (2) joint
probability density functions (PDFs), and (3) probabilities of failure associated with various performance functions. The performance criteria are detailed in Section 4 while the quantitative errors associated with an illustrative example are presented in Section 5.

2. Two types of correlation coefficients

To facilitate the understanding of the two methods for constructing multivariate distributions, the concepts of correlation coefficients should be introduced first. This study will focus on two kinds of correlation coefficients, namely the usual Pearson product–moment correlation coefficient and the Spearman rank correlation coefficient [14].

These two terms are briefly introduced in the following.

2.1. Pearson correlation coefficient

The Pearson product–moment correlation is also called linear correlation (e.g., [15]). Let \( Y_1 \) and \( Y_2 \) be two real valued random variables with finite variances. The Pearson correlation coefficient between \( Y_1 \) and \( Y_2 \), denoted as \( r_{P,Y} \) (hereafter referred to as \( r_{P,Y} \)), is defined as

\[
r_{P,Y} = \frac{\text{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2}
\]  

(1)

Fig. 1. Relationship between different joint moments for the three multivariate construction methods and those calculated by the Isserlis formula.
in which $\text{Cov}(y_1, y_2)$ is the covariance between $Y_1$ and $Y_2$; $\sigma_1$ and $\sigma_2$ are the standard deviations of $Y_1$ and $Y_2$, respectively.

2.2. Spearman correlation coefficient

The Spearman correlation coefficient is a non-parametric measure of statistical dependence between two variables [14]. The Spearman correlation coefficient between $Y_1$ and $Y_2$, denoted as $r_S(y_1, y_2)$ (hereafter referred to as $r_{S,Y}$), is defined as

$$r_{S,Y} = r_P(F_1(y_1), F_2(y_2))$$

in which $F_1(y_1)$ and $F_2(y_2)$ are the marginal CDFs of $Y_1$ and $Y_2$, respectively. The Spearman correlation is invariant under monotone transform. In this aspect, it is very different from the Pearson correlation.

3. Two approximate methods for constructing multivariate distributions

There are two widely used approximate multivariate construction methods. They are approximate, because they are constructed based on marginal PDFs and incomplete dependency information. They are...
consistent in the Kolmogorov sense, but they may not provide the “correct” model in the presence of incomplete dependency information even if the usual statistical uncertainty is neglected. Hence, these approximate construction methods provide useful multivariate distributions that are plausible in the face of inevitable data limitations.

In the following discussion, four types of distribution, namely, independent standard uniform variables, independent standard normal variables, correlated normal variables, and correlated non-normal variables will be referred. To avoid confusion, the symbols U, X, Y, and Z denote independent standard uniform variables, correlated normal variables, correlated non-normal variables, and independent standard normal variables, respectively. The corresponding spaces spanned by a bivariate vector are referred to as the U, X, Y, and Z spaces. These definitions apply throughout the entire paper unless stated otherwise. We use the term “correlated non-normal variables” loosely to denote a general random vector with non-normal marginal distributions.

Multivariate non-normal distributions, particularly with three or more components, are uncommon compared to the range of marginal distributions. One objective in this paper is to provide some closed-form benchmark examples of multivariate non-normal distributions. Once the joint distributions are known theoretically, it is relatively straightforward to conduct correlation analysis and reliability analysis. We call this analysis based on complete joint information as the “exact method” in this study.

In the current reliability literature, it is quite common to construct multivariate non-normal distribution using correlation information alone, in part because higher-order statistics cannot be estimated reliably from limited data. The multivariate normal distribution is the most convenient distribution to couple non-normal marginal distributions using a simple covariance/correlation matrix [6,16,8]. The first natural approximation method is to match the Pearson correlation. We call this “approximate method P” for Pearson. The second approximation method is to match the Spearman correlation. We call this “approximate method S” for Spearman. For concreteness, we present the construction of bivariate distributions using these approximate methods below.

3.1. Approximate method P

This approximate method is based on the concept of Nataf transformation [10,3]. Consider the following isoprobabilistic transformation (e.g., [1]),

\[
\begin{align*}
\Phi(x_i) &= F_{Y_i}(y_i) \\
\Phi^{-1}(y_i) &= \Phi^{-1}[F_{Y_i}(y_i)] \\
i &= 1, 2
\end{align*}
\]

(3)

in which \(\Phi(\cdot)\) is the standard normal CDF; \(\Phi^{-1}(\cdot)\) is the inverse normal CDF; \(F_{Y_i}(y_i)\) is the CDF of \(Y_i\). Thus, correlated non-normal variables can be transformed into correlated standard normal variables component by component (called memoryless transformation). According to Nataf transformation theory and the differentiation rules of implicit function, the approximate joint PDF of correlated non-normal random variables \(Y_1\) and \(Y_2\), denoted as \(f_{12}(y_1, y_2; \rho_{12})\), can be expressed as

\[
f_{12}(y_1, y_2; \rho_{12}) = f_1(y_1) f_2(y_2) \frac{\phi_{12}(x_1, x_2; \rho_{XP})}{\phi(\Phi^{-1}(y_1)) \phi(\Phi^{-1}(y_2))}
\]

(4)

In general, this distribution model is referred to as Nataf distribution [3], in which \(f_1(y_1)\) and \(f_2(y_2)\) are the PDFs of \(Y_1\) and \(Y_2\), respectively; \(\phi(x_1)\) and \(\phi(x_2)\) are the standard normal PDFs; \(\phi_{12}(x_1, x_2; \rho_{12})\) is the bivariate normal PDF with correlation \(\rho_{12}\).
For the exact method, the Pearson correlation coefficient in $X$ space can be calculated by first applying the isoprobabilistic transformation
\[ y_1 = F_{y1}^{-1}(\Phi(x_1)) \]
\[ y_2 = F_{y2}^{-1}(\Phi(x_2)) \]

Then, the joint PDF of $Y_1$ and $Y_2$, $f_{y12}(y_1, y_2)$, can be obtained as
\[ f_{y12}(y_1, y_2) = f_{x12}(x_1, x_2) \frac{\phi(x_1) \phi(x_2)}{\Phi(x_1) \Phi(x_2)} \]
\[ f_{y12}(y_1, y_2) = f_{x12}(x_1, x_2) \frac{\phi(x_1) \phi(x_2)}{\Phi(x_1) \Phi(x_2)} \]

According to Eq. (1), the Pearson correlation coefficient between two variables $X_1$ and $X_2$ can be expressed as
\[ r_{p,x}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{x12}(x_1, x_2) \, dx_1 \, dx_2 \]
\[ r_{p,x}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{x12}(x_1, x_2) \, dx_1 \, dx_2 \]
normalized joint central moment considered in this study is

\[ r_{mn}(Y_1, Y_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left( \frac{y_1 - \mu_1}{\sigma_1} \right)^m \left( \frac{y_2 - \mu_2}{\sigma_2} \right)^n f_{12}(y_1, y_2) dy_1 dy_2 \]  

(14)

in which \( r_{mn} \) \((Y_1, Y_2)\) is the \((m, n)\) order normalized joint central moment between \(Y_1\) and \(Y_2\). All joint moments presented in this paper are in the normalized form defined by Eq. (14). For standard normal marginal distributions, Eq. (14) can be further reduced to

\[ r_{mn}(X_1, X_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1^m x_2^n f_{12}(x_1, x_2) dx_1 dx_2 \]  

(15)

where \( r_{mn} \) \((X_1, X_2)\) is the \((m, n)\) order normalized joint central moment between \(X_1\) and \(X_2\); \(f_{12}(x_1, x_2)\) can be obtained using Eq. (12) for the exact method; \(f_{12}(x_1, x_2)\) are the bivariate standard normal PDFs shown in Eqs. (4) and (9) for the methods P and S, respectively.

The high order joint moments (“normalized” and “central” are dropped for brevity) can also be estimated statistically based on data simulation. The sampling versions of the definition corresponding to Eqs. (14) and (15) are given by

\[ r_{mn}(Y_1, Y_2) = \frac{1}{N} \sum (\frac{y_1 - \overline{y}_1}{\sigma_{y_1}})^m (\frac{y_2 - \overline{y}_2}{\sigma_{y_2}})^n \]  

(16)

and

\[ r_{mn}(X_1, X_2) = \frac{1}{N} \sum x_1^m x_2^n \]  

(17)

where \(N\) is the sample size of Monte Carlo simulation; \(\overline{y}_1\) and \(\overline{y}_2\) are the sample means of \(Y_1\) and \(Y_2\), respectively; \(\sigma_{y_1}\) and \(\sigma_{y_2}\) are the sample standard deviations of \(Y_1\) and \(Y_2\), respectively. If the \(N\) is small, the sample means and standard deviations of \(X_1\) and \(X_2\) are not exactly equal to 0 and 1 due to statistical errors. To eliminate such statistical error, Eq. (17) can be rewritten as

\[ r_{mn}(X_1, X_2) = \frac{1}{N} \sum (\frac{x_1 - \overline{x}_1}{\sigma_{x_1}})^m (\frac{x_2 - \overline{x}_2}{\sigma_{x_2}})^n \]  

(18)

in which \(\overline{x}_1\) and \(\overline{x}_2\) are the sample means of \(X_1\) and \(X_2\), respectively; \(\sigma_{x_1}\) and \(\sigma_{x_2}\) are the sample standard deviations of \(X_1\) and \(X_2\), respectively.

The Isserlis formula [19] can calculate any order joint moment of any number of correlated normal random variables. For the case of two random variables, if \((m + n)\) is odd, the \((m, n)\) order product–moment correlation coefficient between \(X_1\) and \(X_2\) is zero. If \((m + n)\) is even, the \((m, n)\) order product–moment correlation coefficient calculated by the Isserlis formula is given by

\[ r_{mn}(X_1, X_2) = \frac{(m+n-1)!}{m!} \cdot r_{12}^m + \frac{(m+n-3)!}{m!} \cdot r_{12}^2 \cdot (1 - r_{12}^2) + \frac{(m+n-5)!}{m!} \cdot r_{12}^4 \cdot (1 - r_{12}^2) + \cdots \]  

(19)

in which \(r_{12}\) is the Pearson product–moment correlation coefficient between \(X_1\) and \(X_2\); two exclamation marks ! denote the double factorial. It should be noted that the Isserlis formula is exact only for correlated normal random variables. For correlated non-normally distributed random variables, however, it could result in significant errors as demonstrated later.

4.2. Joint probability density functions

The difference in the joint PDF can fundamentally explain the differences in the joint moments, probabilities of failure, and other practical measures. For comparison, only examples with joint PDFs that
are available in closed form are selected for study. Hence, the marginal distributions and the correlation between two variables underlying the joint probability distribution can be determined. The joint PDFs for method P and method S are calculated using Eqs. (4) and (9), respectively.

4.3. Probabilities of failure

For engineers, the probability of failure may be of the greatest interest. For simplicity and illustrative purpose, only the component probability of failure is studied herein. For comparison, the same performance functions are adopted for the example studied. There exists a symmetric plane for the bivariate joint PDF surface when the joint probability distribution involves the same marginal distributions. In this situation, the component probabilities of failure at two sides of the symmetric plane remain the same. Therefore, to examine the difference in the joint PDF systematically, the following three performance functions are considered:

\[ g(Y) = C_1 - y_1 - y_2 \] (20)
\[ g(Y) = C_2 + y_1 + y_2 \] (21)
\[ g(Y) = C_3 + y_1 - y_2 \] (22)

in which \( C_1, C_2, \) and \( C_3 \) are constants. For convenience, these three performance functions are hereafter referred to as performance functions I, II, and III, respectively. Theoretically, the above performance functions are able to scan the entire area of the joint PDF surface provided that the three constants are varied over a wide range. It is evident from Eqs. (20) to (22) that the considered performance functions are relatively simple. The resulting component probabilities of failure associated with Eqs. (20) to (22) can be solved by direct integration, which can be taken as the exact solutions for benchmarking purposes. Due to space limitation, the derivation of formulae for calculating probabilities of failure using the direct integration method are not given herein.

5. Results and discussion

An example is presented below to study the accuracy of computing a joint moment, the accuracy of constructing a joint PDF, and the accuracy of computing a probability of failure based on approximate multivariate constructions.

A general bivariate extreme value distribution-type B distribution is considered [20]. Its joint CDF is

\[ F_{12}(y_1, y_2) = \exp \left( -\exp \left( -\eta \alpha_1 (y_1 - \lambda_1) \right) + \exp \left( -\eta \alpha_2 (y_2 - \lambda_2) \right) \right) \] (23)

where \( \alpha_1 \) and \( \alpha_2 \) are the shape parameters of \( Y_1 \) and \( Y_2 \), respectively; \( \lambda_1 \) and \( \lambda_2 \) are the location parameters of \( Y_1 \) and \( Y_2 \), respectively; \( \eta \) is the parameter representing the correlation between \( Y_1 \) and \( Y_2 \). Its joint PDF is expressed as

\[ f_{12}(y_1, y_2) = \alpha_1 \alpha_2 \exp \left( -\eta \alpha_1 (y_1 - \lambda_1) - \eta \alpha_2 (y_2 - \lambda_2) \right) \] (24)

The marginal CDFs of \( Y_1 \) and \( Y_2 \) can be derived as

\[ F_1(y_1) = \exp \left( -\exp \left( -\alpha_1 (y_1 - \lambda_1) \right) \right) \]
\[ F_2(y_2) = \exp \left( -\exp \left( -\alpha_2 (y_2 - \lambda_2) \right) \right) \] (25)

Based on Eq. (25), one can readily see that the marginal distributions of each \( Y_1 \) and \( Y_2 \) are general type I extreme value distributions. The means of \( Y_1 \) and \( Y_2 \) are \( \mu_1 = \lambda_1 + 0.5772 / \alpha_1 \) and \( \mu_2 = \lambda_2 + 0.5772 / \alpha_2 \), respectively; Their standard deviations are \( \sigma_1 = \pi / (6 \sqrt{3} \alpha_1) \) and \( \sigma_2 = \pi / (6 \sqrt{3} \alpha_2) \), respectively. For the purpose of illustration, the parameters \( \alpha_1 = \alpha_2 = 1 \) and \( \lambda_1 = \lambda_2 = 0 \) are assumed for the subsequent analyses. Accordingly, the means and standard deviations are obtained as \( \mu_1 = \mu_2 = 0.5772 \) and \( \sigma_1 = \sigma_2 = 1.28 \), respectively.

The Pearson correlation coefficient between \( Y_1 \) and \( Y_2 \) can be calculated as (e.g., [20])

\[ r_{P,Y} = 1 - \frac{1}{\eta^2} \] (26)

For various values of \( \eta \), the Pearson correlation coefficients calculated by Eq. (26) are shown in Table 1. The Pearson correlation coefficient increases from 0 to 0.95 as the parameter \( \eta \) increases from 1 to 0.05 to -0.5. Theoretically, the Pearson correlation coefficient will approach 1.0 when the \( \eta \) is large enough.

The high order joint moments in \( X \) space associated with the method P and method S are compared with that of the exact method. The results are summarized in Table 1. Note that the difference among the low order correlation coefficients \( r_{11} \) associated with the three multivariate construction methods is small. The maximum difference in \( r_{11} \) between the method P and the exact method is 0.07. It is 0.09 between the method S and the exact method. The difference in high order joint moments between the two approximate methods and the exact method will be small when \( r_{P,Y} \) approaches 1.0.

As mentioned previously, the Isserlis formula can also produce the high order joint moments exactly for normally distributed random variables. For comparison, Fig. 1 shows the relationship between different correlation coefficients for the three multivariate construction methods and those calculated by the Isserlis formula. Note that the correlation coefficients associated with the methods P and S match well with those obtained from the Isserlis formula for any order of correlation coefficients. The difference in high order joint moments
Fig. 7. Comparison of the component probabilities of failure derived from method P and method S with the exact component probability of failure for various Pearson correlation coefficients (Performance function III).

Table 1
Comparison among higher order joint moments for bivariate extreme value distribution.

<table>
<thead>
<tr>
<th>η</th>
<th>r_{P,Y}</th>
<th>Exact method, ( r_{\text{est}}(X_1, X_2) )</th>
<th>Method P, ( r_{\text{est}}(X_1, X_2) )</th>
<th>Method S, ( r_{\text{est}}(X_1, X_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \begin{array}{c} t_1 \ \ t_2 \ \ t_3 \ \ t_4 \ \ t_5 \ 0 \ \ 0 \ \ 0 \ \ 0 \ \ 0 \end{array} )</td>
<td>( \begin{array}{c} t_1 \ \ t_2 \ \ t_3 \ \ t_4 \ \ t_5 \ 0 \ \ 0 \ \ 0 \ \ 0 \ \ 0 \end{array} )</td>
<td>( \begin{array}{c} t_1 \ \ t_2 \ \ t_3 \ \ t_4 \ \ t_5 \ 0 \ \ 0 \ \ 0 \ \ 0 \ \ 0 \end{array} )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.085 1.094 1.208 1.473 68.0</td>
<td>0.106 1.022 0.959 9.81 24.5</td>
<td>0.080 1.013 0.725 9.46 18.4</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8</td>
<td>0.171 1.197 2.393 20.56 133.5</td>
<td>0.210 1.088 1.948 12.23 52.9</td>
<td>0.164 1.054 1.502 10.95 39.5</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7</td>
<td>0.260 1.311 3.561 26.53 196.8</td>
<td>0.313 1.196 3.004 16.30 89.3</td>
<td>0.252 1.127 2.399 13.65 66.3</td>
</tr>
<tr>
<td>0.7</td>
<td>0.6</td>
<td>0.352 1.439 4.725 32.72 258.9</td>
<td>0.415 1.345 4.165 22.12 137.8</td>
<td>0.343 1.235 3.229 17.80 102.0</td>
</tr>
<tr>
<td>0.6</td>
<td>0.5</td>
<td>0.447 1.584 5.900 39.25 230.7</td>
<td>0.516 1.532 5.464 29.85 202.7</td>
<td>0.439 1.385 4.456 23.75 151.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4</td>
<td>0.545 1.753 7.114 46.33 384.5</td>
<td>0.615 1.757 6.931 39.07 288.5</td>
<td>0.539 1.581 5.792 31.95 220.8</td>
</tr>
<tr>
<td>0.4</td>
<td>0.3</td>
<td>0.648 1.953 8.416 54.33 454.0</td>
<td>0.713 2.017 8.595 51.83 400.3</td>
<td>0.645 1.831 7.407 43.05 319.0</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
<td>0.755 2.198 9.899 64.06 537.7</td>
<td>0.810 2.313 10.481 66.59 543.1</td>
<td>0.756 2.142 9.389 57.93 458.4</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>0.871 2.516 11.782 77.52 658.1</td>
<td>0.906 2.641 12.614 84.25 723.1</td>
<td>0.873 2.526 11.858 77.89 657.3</td>
</tr>
<tr>
<td>0.1</td>
<td>0.05</td>
<td>0.933 2.722 13.057 87.54 754.5</td>
<td>0.953 2.818 13.780 94.28 829.0</td>
<td>0.935 2.750 13.327 90.35 787.2</td>
</tr>
</tbody>
</table>

between the exact method and the Isserlis formula does not monotonically increase as the high order joint moments increase. The maximum difference in the high order joint moment can happen at an intermediate high order joint moment. The methods P and S are expected to satisfy the Isserlis formula given the bivariate normal distribution underlying these methods. The Isserlis formula is not expected to work for this example considered, because it is not derived from translation of a bivariate normal. The numerical differences shown in Fig. 1 are however quite interesting.

The joint PDFs for the methods P and S are determined using Eqs. (4) and (9), respectively. The theoretical joint PDF has been given in Eq. (24). Fig. 2 compares the joint PDFs qualitatively for various Pearson correlation coefficients in \( Y \) space. It can be seen that there exists a significant difference in the joint PDFs especially when the correlation is strongly positive.

Fig. 3 compares the component probabilities of failure on log scale associated with the three methods for various values of \( C_1 \) in the performance function I (Eq. (20)). The difference in probabilities of failure between the two approximate methods and the exact method increases with decreasing probability of failure. Both approximate methods underestimate the probability of failure, which is unconservative for structural safety assessment. Note that the probabilities of failure for \( r_{P,Y} = 0.95 \) obtained from the three methods are almost the same. For instance, when \( C_1 = 18 \), the probabilities of failure obtained from the exact method, methods P and S are 1.20E−04, 1.05E−04, and 9.87E−05, respectively.
Fig. 4 shows the component probabilities of failure on log scale for various Pearson correlation coefficients between $Y_1$ and $Y_2$. The constant $C_1$ is assumed to be 18. Note that the method $P$ is slightly closer to the exact method than that of the method $S$. When the two variables $Y_1$ and $Y_2$ are completely independent or fully correlated, there is no difference in the probabilities of failure obtained from the three methods. Furthermore, the difference in the probabilities of failure between the two approximate methods and the exact method will approach its maximum value for an intermediate Pearson correlation coefficient. For $r_{P,Y} = 0.10$, the probabilities of failure obtained from the exact method, method $P$, and method $S$ are $1.87E−05$, $7.66E−07$, and $6.02E−07$, respectively. The probability of failure obtained from the exact method is about 24 and 31 times those obtained from the methods $P$ and $S$, respectively. The resulting errors underlying the exact method is about 24 and 31 times those obtained from the methods $P$ and $S$, respectively. The reason is that when the correlation coefficient is large. For $r_{P,Y}$, the probabilities of failure between the two approximate methods and the exact method increases with decreasing probability of failure. Additionally, the two approximate methods overestimate the probability of failure, which means they are conservative for structural safety assessment in this case.

Fig. 5 compares the component probabilities of failure on log scale for various Pearson correlation coefficients between $Y_1$ and $Y_2$. The constant $C_2$ is assumed to be 4. Compared with the results shown in Fig. 4, the method $S$ is slightly closer to the exact method than that of the method $P$. In addition, there exists a small difference in the probabilities of failure between the two approximate methods and the exact method. Furthermore, such difference does not always increase with increasing Pearson correlation coefficient.

With regard to the performance function $II$ (Eq. (21)), Fig. 7 compares the component probabilities of failure associated with the three methods for various values of $C_2$ in the performance function $II$. Similar to the results shown in Fig. 3, the difference in probabilities of failure between the two approximate methods and the exact method increases with decreasing probability of failure. Additionally, the two approximate methods overestimate the probability of failure, which means they are conservative for structural safety assessment in this case.

Fig. 6 shows the component probabilities of failure on log scale for various Pearson correlation coefficients between $Y_1$ and $Y_2$. The constant $C_2$ is assumed to be 4. Compared with the results shown in Fig. 4, the method $S$ is slightly closer to the exact method than that of the method $P$. In addition, there exists a small difference in the probabilities of failure between the two approximate methods and the exact method. Furthermore, such difference does not always increase with increasing Pearson correlation coefficient.

With regard to the performance function $III$ (Eq. (22)), Fig. 7 compares the component probabilities of failure associated with the three methods for various values of $C_3$. In comparison with the results for the performance functions $I$ and $II$, there is a significant difference in the probabilities of failure obtained from the three methods especially when the correlation coefficient is large. For $r_{P,Y} = 0.95$ and $C_3 = 2.5$, the probabilities of failure obtained from the exact method, method $P$, and method $S$ are $1.39E−05$, $4.64E−05$, and $1.76E−04$, respectively. The results for the methods $P$ and $S$ are 3.3 and 12.7 times the exact solution.

The component probabilities of failure on log scale versus Pearson correlation coefficients are plotted in Fig. 8. The constant $C_3$ is assumed to be 3. Note that the probabilities of failure obtained from the two approximate methods differ greatly from the exact solutions especially when the $r_{P,Y}$ is large. Furthermore, the probabilities of failure between the method $P$ and $S$ can also differ considerably for large value of $r_{P,Y}$. Taking $r_{P,Y} = 0.95$ as an example, the probabilities of failure obtained from the exact method, method $P$, and method $S$ are $1.49E−06$, $8.01E−06$, and $3.84E−05$, respectively. The probabilities of failure for the methods $P$ and $S$ are 5.4 and 25.8 times the exact solution. The probability of failure for the method $S$ is about five times that for the method $P$. Such findings are very different from those drawn from the previous results. The reason is that when the $r_{P,Y}$ approaches 1.0, the joint PDF at the plane of performance function $III$ is very sensitive to the change of $r_{P,Y}$. A very small change of the $r_{P,Y}$ will cause a significant change in the probability of failure. For example, when the $r_{P,Y}$ increases only from 0.925 to 0.95, the resulting probability of failure obtained from the exact method decreases from $1.75E−05$ to $1.49E−06$. The former is about 10 times larger than the latter.

6. Summary and conclusions

Two multivariate construction methods, namely the approximate method $P$ and the approximate method $S$, have been studied. The performance of these approximations is evaluated based on their abilities to match exact solutions for: (1) high order joint moments, (2) joint PDFs, and (3) probabilities of failure. One benchmark example with known joint distribution (non-translation) is presented to quantify the errors numerically. Several conclusions can be drawn from this study:

1. The difference in the high order joint moments associated with the three multivariate construction methods could be very significant. As to be expected, the Isserlis formula cannot work for correlated non-normal variables not derived from the translation approach. The simulation studies however demonstrate that the errors are generally unacceptable.

2. The only difference in the joint PDFs between the methods $P$ and $S$ is the product–moment correlation coefficient between two correlated standard normal variables. The joint PDF obtained from the method $P$ is slightly different from that obtained from the method $S$. The difference in the joint PDFs between the two approximate methods and the exact method increases as the correlation becomes stronger.

3. The difference in probabilities of failure between the two approximate methods and the exact method highly depends on the level of probability of failure, the form of performance function, and the degree of correlation between two random variables underlying the bivariate distribution. Such difference increases with decreasing probability of failure. Significant difference in probabilities of failure between the two approximate methods and the exact method could be observed when the failure domain coincides with the domain where the joint PDF is poorly approximated and the probability of failure is below $10^{−3}$.

4. The maximum error in the probability of failure may not be associated with a large correlation. It can happen at an intermediate correlation.

5. There exists a small difference in the probabilities of failure between the two approximate methods for the examples studied. Hence, the $P$ and $S$ methods are effectively the same from a numerical viewpoint. The $P$ method is not consistently more accurate or more conservative than the $S$ method or vice-versa.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Project Nos. 51028901 and 51079112) and the Foundation for the Author of National Excellent Doctoral Dissertation of PR
China (Project No. 2007B50).

References