Iowa Research Online
The University of Iowa's Institutional Repository

University of Iowa
Iowa Research Online

Theses and Dissertations

Fall 2009

Input model uncertainty and reliability-based design optimization with associated confidence level

Yoojeong Noh
University of Iowa

Copyright 2009 Yoojeong Noh

This dissertation is available at Iowa Research Online: http://ir.uiowa.edu/etd/412

Recommended Citation
Noh, Yoojeong. "Input model uncertainty and reliability-based design optimization with associated confidence level." PhD (Doctor of Philosophy) thesis, University of Iowa, 2009.
http://ir.uiowa.edu/etd/412.

Follow this and additional works at: http://ir.uiowa.edu/etd

Part of the Mechanical Engineering Commons

INPUT MODEL UNCERTAINTY AND RELIABLITY-BASED DESIGN OPTIMIZATION WITH ASSOCIATED CONFIDENCE LEVEL

by

Yoojeong Noh

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Mechanical Engineering
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisor: Professor Kyung K. Choi

ABSTRACT

The objective of this study is to develop new methods for modeling of input uncertainty of correlated variables and to carry out reliability-based design optimization (RBDO) using the identified input uncertainty model with associated confidence level. The proposed research involves: (1) use of copulas to model joint CDFs of input variables; (2) use of the Bayesian method to identify marginal cumulative distribution functions (CDF) and joint CDF of input variables using limited experimental data; (3) reduction of the transformation ordering effect on the inverse reliability analysis using the most probable point (MPP)-based dimension reduction method (DRM); and (4) assessment of the confidence level of the input model uncertainty, and implementation of the confidence level in RBDO to offset the inexact quantification of the input uncertainties due to limited data.

It has been well documented that many random variables such as material properties are correlated, but the correlation has not been considered in RBDO because modeling the joint CDF of correlated variables is known to be difficult. In this study, a copula is introduced to model a joint CDF of input variables. The copula requires marginal CDFs and correlation parameters, which can be obtained in real applications, so it is possible to model a joint CDF. Once the joint input CDF is modeled using a copula, the input variables can be transformed to independent Gaussian variables using Rosenblatt transformation for the inverse reliability analysis.

This study proposes a method to identify correct marginal and joint CDFs (copulas) of input variables. In practical applications, since only limited experimental data are available, it is challenging task to correctly identify the marginal and joint CDF

of input variables using the limited data. In this study, a Bayesian method is proposed to identify the marginal and joint CDFs of input variables that best describe given data among candidates. The performance of the proposed method is investigated and compared with an existing method, the goodness-of-fit (GOF) test.

Using the identified input model, the transformation from original input variables into independent Gaussian variables is carried out, and then the first-order reliability method (FORM), which has been commonly used in reliability analysis, is carried out. However, when the input variables are correlated with non-elliptical copulas, the FORM may yield different reliability analysis results with some errors for different transformation orderings of input variables due to the nonlinearities of the transformed constraint functions. For this, the MPP-based DRM, which more accurately and efficiently calculates the probability of failure than the FORM and the second-order reliability method (SORM), respectively, is used to reduce the effect of transformation ordering in the inverse reliability analysis and, thus, RBDO.

However, when the number of experimental data is limited, the estimated input joint CDF will be inaccurate, which will lead to inaccurate RBDO result. Thus, a method to assess the confidence level of the input model uncertainty in RBDO is developed, and the input model with confidence level is implemented for RBDO.

Abstract Approved: _____

Thesis Supervisor

_____

Title and Department

_____

Date

INPUT MODEL UNCERTAINTY AND RELIABLITY-BASED DESIGN OPTIMIZATION WITH ASSOCIATED CONFIDENCE LEVEL

by

Yoojeong Noh

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Mechanical Engineering
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisor: Professor Kyung K. Choi

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph. D thesis of

Yoojeong Noh

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Mechanical Engineering at the December 2009 graduation.

Thesis Committee:

Kyung K. Choi, Thesis Supervisor

Jia Lu

Yong Chen

Shaoping Xiao

Olesya Zhupanska

TABLE OF CONTENTS

LIST OF TABLES v

LIST OF FIGURES vii

LIST OF DEFINITIONS ix

LIST OF ACRONYMS xi

CHAPTER

I. INTRODUCTION 1

1.1 Background and Motivation 2
1.1.1 Identification of Input Uncertainty Model 2
1.1.2 Transformation Ordering Effect 4
1.1.3 Reliability-based Design Optimization with Associated Confidence Level 7
1.2 Objectives of the Proposed Study 9
1.3 Organization of the Thesis Proposal 12

II. AN INTRODUCTION TO COPULAS 14

2.1 Introduction 14
2.2 Engineering Application with Correlated Variables 14
2.3 Basic Definition of Copulas 17
2.4 Correlation Measures 20
2.4.1 Linear Correlation Coefficient (Pearson's Rho) 23
2.4.2 Spearman's Rho 23
2.4.3 Kendall's Tau 25
2.5 Commonly Used Copulas 27
2.5.1 Elliptical Copulas 28
2.5.2 Archimedean Copulas 31

III. IDENTIFICATION OF MARGINAL AND JOINT CUMULATIVE DISTRIBUTION FUNCTIONS 33

3.1 Introduction 33
3.2 Identification of Marginal CDFs 34
3.2.1 Goodness-of-fit Test 34
3.2.2 Bayesian Method 36
3.2.3 Comparision of Two Methods 43
3.3 Identification of Joint CDFs (Copulas) 46
3.3.1 Goodness-of-fit Test 46
3.3.2 Bayesian Method 48
3.3.3 Comparision of Two Methods 52
3.4 Effectiveness of Bayesian Method 53
3.4.1 Identification of Marginal CDFs 53

3.4.2 Identification of Joint CDFs 57

IV. FUNDAMENTAL CONCEPTS IN RELIABILITY BASED-DESIGN OPTIMIZATION.....61

4.1 Introduction.....61
4.2 Reliability Analysis 61
4.2.1 Transformation 62
4.2.2 First-Order Reliability Method (FORM).....65
4.2.3 Second-Order Reliability Method (SORM) 66
4.3 Inverse Reliability Analysis.....67
4.4 MPP-based Dimension Reduction Method 70
4.5 Reliability-Based Design Optimization.....73

V. TRANSFORMATION ORDERING EFFECT 75

5.1 Introduction.....75
5.2 Transformation Ordering for Various Joint CDF Types 76
5.2.1 Transformation Orderings in Rosenblatt Transformation 76
5.2.2 Independent and Elliptical Copula 78
5.2.3 Non-elliptical Copula 81
5.3 Examples.....84
5.3.1 Two Dimensional Example 85
5.3.2 Four Dimensional Example.....92

VI. APPLICATIONS OF COPULA, BAYESIAN METHOD, AND DIMENSION REDUCTION METHOD TO RELIABILITY-BASED DESIGN OPTIMIZATION PROBLEMS.....95

6.1 Introduction.....95
6.2 Mathematical Example 95
6.3 Coil Spring Problem 101
6.4 Roadarm Example 107

VII. RELIABILITY-BASED DESIGN OPTIMIZATION WITH CONFIDENCE LEVEL UNDER INPUT MODEL UNCERTAINTY 114

7.1 Introduction.....114
7.2 Measure of Input Confidence Level 115
7.3 Confidence Intervals of Mean, Standard Deviation, and Correlation Coefficient 119
7.3.1 Confidence Interval of Mean.....119
7.3.2 Confidence Interval of Standard Deviation.....122
7.3.3 Confidence Interval of Correlation Coefficient.....136
7.4 Adjusted Parameters Using Confidence Intervals of Input Parameters.....137
7.4.1 Adjusted Standard Deviation Using Confidence Intervals of Mean and Standard Deviation.....138
7.4.2 Adjusted Correlation Coefficient Using Confidence Interval of Correlation Coefficient 141
7.5 Confidence Level of Input Model.....146
7.6 Numerial Examples 153

7.6.1 Two-dimensional Mathematical Example.....153
7.6.2 Coil Spring Problem.....157

VIII. CONCLUSIONS AND FUTURE WORK.....160

8.1 Conclusions.....160
8.1.1 Identification of Input Model 160
8.1.2 Reduction of Transformation Ordering Effect on Reliability Based Design Optimization (RBDO).....161
8.1.3 Reliability Based Design Optimization with Confidence Level under Input Model Uncertainty.....162
8.2 Future Recommendation.....163

REFERENCES 165

LIST OF TABLES

Table

2.1. Spearman's Rho and Its Domain 25

2.2. Kendall's Tau and Its Domain 27

2.3. Copula Functions, Generator Functions, and Domains of Correlation Parameters.....32

3.1. Mean, Variance, and Their Domains 37

3.2. Marginal PDFs and Domains of Marginal Parameters 39

3.3. Averaged Normalized Weights over 100 Trials Using Two Approaches 43

3.4. Comparison of GOF with Bayesian Method for Identification of Marginal Gamma CDF 45

3.5. Comparison of GOF with Bayesian Method for Identification of Copula 53

4.1. Gaussian Quadrature Points and Weights.....72

5.1. RBDO Results Using Gaussian Copula ($P_F^{Tar} = 2.275\%$).....86

5.2. RBDO Results Using Clayton Copula ($P_F^{Tar} = 2.275\%$).....88

5.3. RBDO Results Using Frank Copula ($P_F^{Tar} = 2.275\%$) 91

5.4. RBDO Results Using Gumbel and A12 Copula ($P_F^{Tar} = 2.275\%$).....94

6.1. Normalized Weights of Candidate Marginal CDFs for Mathematical Example.....96

6.2. Normalized Weights of Candidate Copulas for Mathematical Example.....98

6.3. RBDO Comparisons for Mathematical Example ($P_F^{Tar} = 2.275\%$) 101

6.4. Properties of Random Variables and Parameters for Coil Spring Example 104

6.5. Normalized Weights of Candidate Marginal CDFs for Coil Spring Example 105

6.6. Normalized Weights of Candidate Copulas for Coil Spring Example 105

6.7. Distribution Type and Parameters of Estimated Input Model for Coil Spring Example 106

6.8. RBDO Comparisons for Coil Spring Example ($P_F^{Tar} = 2.275\%$).....107

6.9. Random Variables and Fatigue Material Properties for Roadarm Example 110

6.10. Normalized Weights of Candidate Copulas Using 29 Paired Data of SAE 950X Steel....111

6.11. RBDO Comparisons for Roadarm Example....113

7.1. Bootstrap Procedures....124

7.2. Estimated Confidence Levels of Standard Deviation (Gaussian, $\mu = 5.0$ and $\sigma = 5.0$)....131

7.3. Mean Values of Upper Bound of Confidence Interval of Standard Deviation (Gaussian, $\mu = 5.0$ and $\sigma = 5.0$)....132

7.4. Estimated Confidence Levels of Standard Deviation (Lognormal, $\mu = 5.0$ and $\sigma = 5.0$)....133

7.5. Mean Values of Upper Bound of Confidence Interval of Standard Deviation (Lognormal, $\mu = 5.0$ and $\sigma = 5.0$)....134

7.6. Standard Deviations of Upper Bound of Confidence Interval of Standard Deviation (Lognormal, $\mu = 5.0$ and $\sigma = 5.0$)....134

7.7. Mean Values of Confidence Intervals of Mean....139

7.8. Mean Values of Confidence Intervals of Standard Deviation and Adjusted Standard Deviation (Gaussian)....140

7.9. Mean Values of Confidence Intervals of Standard Deviation and Adjusted Standard Deviation (Lognormal)....141

7.10. Confidence Intervals of Correlation Coefficient and Adjusted Correlation Coefficient (Frank Copula)....146

7.11. Input Confidence Levels (%) Using Upper Bound of Standard Deviation and Sample Correlation Coefficient, $N(5, 0.3^2)$ 151

7.12. Input Confidence Levels (%) Using Adjusted Standard Deviation and Correlation Coefficient, $N(5, 0.3^2)$ 151

7.13. Input Confidence Levels (%) Using Upper Bound of Standard Deviation and Sample Correlation Coefficient, $N(5, 5^2)$ 152

7.14. Input Confidence Levels (%) Using Adjusted Standard Deviation and Correlation Coefficient, $N(5, 5^2)$ 152

7.15. Probabilities of Failure and Output Confidence Levels for Mathematical Example ($\sigma_1 = \sigma_2 = 0.3$) 155

7.16. Probabilities of Failure and Output Confidence Levels for Mathematical Example ($\sigma_1 = 0.3$ and $\sigma_2 = 1.5$)....157

7.17. Probabilities of Failures and Output Confidence Levels for Coil Spring Example 158

LIST OF FIGURES

Figure

1.1. RBDO under Uncertainties 2

2.1. 29 Paired Data Obtained from 950X Steel 15

2.2. Exhaust Manifold 17

2.3. Scatter Plot of Data and PDF Contours between X and Y 17

2.4. PDF Contours of Gaussian Copula with Different Marginal Distributions 19

2.5. PDF Contours of *t* Copula with Marginal Gaussian Distributions for Different Degrees of Freedom. 30

2.6. PDF Contours of Some Archimedean Copulas Using Marginal Standard Gaussian Distributions for τ =0.5. 32

3.1. Marginal PDFs for μ=10.0 and σ=0.5 44

3.2. Sum of Normalized Weights over 100 Trials for μ=2.0 and σ=0.5 54

3.3. Marginal PDFs for μ=2.0 and σ=0.5 55

3.4. Sum of Normalized Weights over 100 Trials for μ=10.0 and σ=0.5 56

3.5. PDF Contours and Scatter Plots of Samples (*ns*=600) for τ =0.2 58

3.6. Sum of Normalized Weights over 100 Trials for τ =0.2 58

3.7. PDF Contours and Scatter Plots of Samples (*ns*=300) for τ =0.7 59

3.8. Sum of Normalized Weights over 100 Trials for τ =0.7 59

4.1. MPP and Reliability Index in **U**-space 66

5.1. Joint PDFs and β_t-contours in **U**-space and **X**-space 83

5.2. β_t-contours and Constraint Functions with Different Transformation Orderings Using Gaussian Copula 86

5.3. β_t-contours and Constraint Functions with Different Transformation Orderings Using Clayton Copula 87

5.4. β_t-contours and Constraint Functions with Different Transformation Orderings Using Frank Copula 90

6.1. Likelihood Functions of μ for Mathematical Example 96

6.2. Likelihood Functions of τ for Mathematical Example 97

6.3. Scatter Plot of Data and PDF Contour of Frank Copula.....98

6.4. β_t-contours Using Different Input Models for Mathematical Example 99

6.5. Coil Spring.....102

6.6. Army Tracked Vehicle.....108

6.7. Finite Element Model of Roadarm 108

6.8. Fatigue Life Contour and Critical Nodes of Roadarm.....108

6.9. Shape Design Variables for Roadarm.....109

6.10. Joint PDF Contours of Gaussian and Frank Copulas Identified from 29 Paired Data of 950X Steel.....111

7.1. Estimated β_t-contour Fully Covering True β_t-contour.....117

7.2. Estimated β_t-contour Partially Covering True β_t-contour.....117

7.3. β_t-contours Using Lower and Upper Bounds of the Confidence Interval of Mean, Standard Deviation, and Correlation Coefficient 118

7.4. Histograms of Upper Bound of Confidence Interval of Standard Deviation 135

7.5. β_t-contours for Small and Large Correlation Coefficient.....142

7.6. β_t-contours Using Underestimated and Overestimated Correlation Coefficient 143

7.7. Adjusted Correlation Coefficient.....144

7.8. Flow Chart of Assessment of Input Confidence Level.....147

7.9. β_t-contours Using Obtained Input Models with Upper Bound of Standard Deviation and Sample Correlation coefficient.....148

7.10. β_t-contours Using Obtained Input Models with Adjusted Parameters.....150

LIST OF DEFINITIONS

n	Number of random variables
ndv	Number of design variables
$\boldsymbol{d}$	Vector of design variables, $\mathbf{d} = \left[d_1, \cdots, d_{ndv}\right]^T$
$\mathbf{X}$	Vector of random variables, $\mathbf{X} = \left[X_1, \cdots, X_n\right]^T$
$\mathbf{x}$	Realization of vector $\mathbf{X}$, $\mathbf{x} = \left[x_1, \cdots, x_n\right]^T$
$\mathbf{x}^*$	FORM-based MPP
$\mathbf{x}^*_{DRM}$	DRM-based MPP
$\mathbf{U}$	Vector of independent standard Gaussian variables, $\mathbf{U} = \left[U_1, \cdots, U_n\right]^T$
$\mathbf{Z}$	Vector of uniform variables, $\mathbf{Z} = \left[Z_1, \cdots, Z_n\right]^T$
$\mathbf{z}$	Vector of uniform variables, $\mathbf{z} = \left[z_1, \cdots, z_n\right]^T$
$\mathbf{E}$	Vector of correlated standard elliptical variables, $\mathbf{E} = \left[E_1, \cdots, E_n\right]^T$
$\boldsymbol{e}$	Realization of vector $\mathbf{E}$, $\mathbf{e} = \left[e_1, \cdots, e_n\right]^T$
ρ_{ij}	Pearson's correlation coefficient between X_i and X_j
$\mathbf{P}$	Covariance matrix of $\mathbf{X}$, $\left\{\rho_{ij}\right\}$
$\mathbf{P'}$	Covariance matrix of $\mathbf{E}$, $\left\{\rho'_{ij}\right\}$
$\Phi\left(\cdot\right)$	Marginal Gaussian CDF
$\Phi_{\mathbf{P'}}\left(\cdot \middle\vert \mathbf{P'}\right)$	Multivariate Gaussian CDF with $\mathbf{P'}$
$F_{X_i}\left(\cdot\right)$	Marginal CDF of X_i

$F_{X_1 \cdots X_n}(\cdot)$	Joint CDF of $X_1, \cdots, X_n$
$f_{X_1, \cdots, X_n}(\cdot)$	Joint PDF of $X_1, \cdots, X_n$
$F_{X_i}(\cdot \mid \cdot)$	Conditional CDF of X_i
$\boldsymbol{\theta}$	Matrix of correlation parameters of $X_1, \cdots, X_n$
$C(\cdot \mid \boldsymbol{\theta})$	Copula with $\boldsymbol{\theta}$
τ	Kendall's tau

LIST OF ACRONYMS

AMH	Ali-Mikhail-Haq
AMV	Advanced mean value
CDF	Cumulative distribution function
CMV	Conjugate mean value
DRM	Dimension reduction method
FGM	Farlie-Gumbel-Morgenstern
FORM	First-order reliability method
GOF	Goodness-of-fit
HMV	Hybrid mean value
MPP	Most probable point
MV	Mean value
PDF	Probability density function
PMA	Performance measure approach
RBDO	Reliability-based design optimization
RIA	Reliability index approach
SORM	Second-order reliability method
BC	Bias corrected
BCa	Bias corrected accelerated

CHAPTER I

INTRODUCTION

This study presents new methods for identifying and quantifying an input model with correlated input variables and for carrying out reliability-based design optimization (RBDO) with associated confidence level. As shown in Figure 1.1, the input model uncertainty, consisting of statistical and physical uncertainty, is characterized by identifying and quantifying the input model. To identify the joint distribution, first, a copula is introduced to model a joint cumulative distribution function (CDF) of input variables, which is difficult to obtain in practical industrial applications where only limited experimental data are available. Second, a Bayesian method is proposed to correctly identify and quantify the input model, i.e., marginal CDFs and joint CDFs (copulas) from given experimental data. Third, if the input variables are multi-dimensional and correlated, there exists a transformation ordering effect on RBDO. To reduce the transformation ordering effect, an accurate reliability analysis needs to be used as shown in Figure 1.1. In this study, the MPP-based dimension reduction method (DRM) is used. Finally, if the identified and quantified input model is inaccurate, the obtained RBDO results may not satisfy the required target reliability. Thus, the input model with a confidence level is developed in RBDO to obtain RBDO results with associated confidence level as shown in Figure 1.1. The simulation model uncertainty, which is related to verification and validation of surrogate models, is not considered in this study.

Section 1.1 presents background and motivation for the proposed research, Section 1.2 provides the objectives of the proposed research, and Section 1.3 describes the thesis proposal organization.

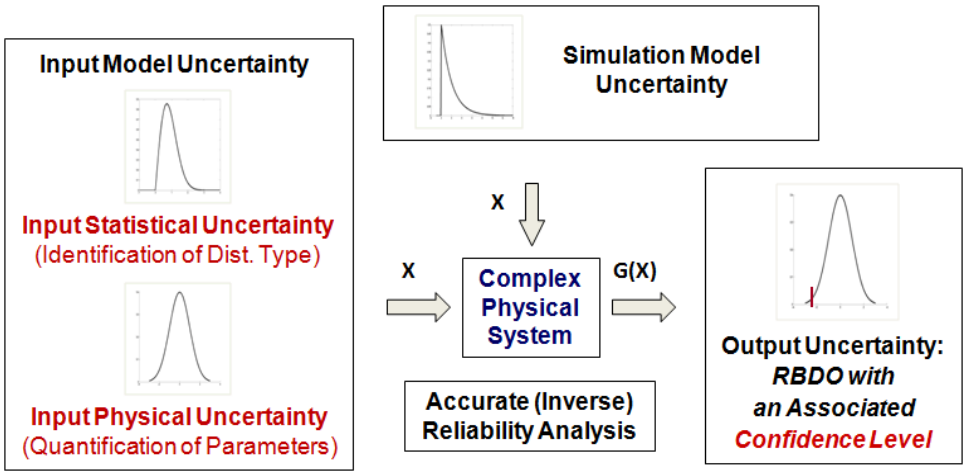


Figure 1.1 RBDO under Uncertainties

1.1 Background and Motivation

1.1.1 Identification of Input Model Uncertainty

In many practical industrial applications, input random variables such as the material properties of the modern composite, nano, and micro materials; fatigue properties; etc. are correlated (Annis, 2004; Socie, 2003; Nikolaidis et al., 2005). However, only limited data are available in practical applications, so it is difficult to obtain a true joint CDF of the input variables, which requires infinite data to obtain. Due to this difficulty, it has been often assumed that the input variables are independent or correlated with a joint Gaussian CDF even though they may be correlated with a joint non-Gaussian CDF (Ditlevsen and Madsen, 1996; Liu and Der Kiureghian, 1986; Madsen et al., 1986; Melchers, 1999). Such erroneous assumptions of the input model lead to wrong RBDO results that could be quite different from true optimum designs (Noh, 2007).

To model the joint CDF of the correlated input random variables, the copula has been used in non-engineering disciplines such as biostatistics, finance, actuarial science, economics, and hydrology (Bouyé et al., 2000; Embrechts et al., 2003; Frees and Valdez, 1998; Kolev et al., 2006; Pham, 2006; Wang, 2007), but it has not been used in the engineering discipline especially for the purpose of design. Since the copula only requires the marginal CDFs and correlation parameters (Nelson, 1999), which can be obtained in practical applications, the joint CDF can be readily obtained. In addition, the copula decouples the joint CDF and marginal CDFs, so that a joint CDF with various types of marginal CDFs, which often occurs in practical applications, can be obtained. Thus, it is proposed to use the copula to model a joint CDF of correlated input variables for RBDO.

To obtain accurate RBDO results, the information of the input variables such as marginal and joint CDFs (copulas) needs to be correctly identified using given experimental data. If the marginal CDF types of the input variables are unknown and only the experimental data are given, then the marginal CDFs should be identified first. For the identification of the marginal CDFs, the goodness-of-fit (GOF) tests such as Chi-square, Cramér-von Mises, and Kolmogorov-Smirnov (K-S) tests have been widely used (Haldar and Mahadevan, 2000; Sterne and Smith, 2000). However, the GOF test relies on the mean and standard deviation calculated from data, and thus if these parameters are incorrectly calculated, a wrong marginal CDF could be identified. Likewise, the GOF for the identification of the copula have been recently developed (Genest and Favre, 2007; Genest and Rémillard, 2005; Wang and Well, 2000). However, since the GOF test depends on the correlation coefficient estimated from data, if the correlation parameter is incorrectly estimated, then a wrong copula could be identified.

To overcome the shortcoming of the GOF test, the Bayesian method (Huard et al., 2006) has been developed to identify the joint CDF (copula). Unlike the GOF test, the Bayesian method calculates the weights of candidate copulas by integrating the likelihood function over the correlation coefficient. Thus, the Bayesian method is independent of the choice of the parameter, and thus, it is preferred over the GOF test. However, since the Bayesian method has been developed only for identifying the copula, it is necessary to develop a Bayesian method for identifying marginal CDFs.

1.1.2 Transformation Ordering Effect

The RBDO aims to find an optimum design achieving the confidence in product reliability at a given probabilistic level. For this, the RBDO involves evaluation of probabilistic constraints such that the probability of failure for each constraint is less than the target probability of failure. However, since the probability of failure is obtained by a multiple integral of the joint probability density function (PDF) of the input variables over the failure region, it is difficult to analytically compute these multiple integrals. Hence, approximation methods such as the first-order reliability method (FORM) (Hasofer and Lind, 1974; Palle and Michael, 1982; Madsen et al., 1986; Haldar and Mahadevan, 2000) or the second-order reliability method (SORM) (Breitung, 1984; Hohenbichler and Rackwitz, 1988) are used. The FORM or SORM computes the probability of failure by approximating the constraint function $G(\mathbf{X})$ using the first- or second-order Taylor series expansion at MPP. Since FORM often provides adequate accuracy and is much easier to use than SORM, it is commonly used in RBDO. Since FORM and SORM are carried out in the standard Gaussian space, a transformation of the

correlated input random variables into the independent standard Gaussian variables is necessary.

There are several methods to transform the correlated variables into independent standard normal variables: the Hermite polynomial transformation, the Winterstein approximation, the Rosenblatt transformation, the Nataf transformation, and orthogonal transformation (Ditlevsen and Madsen, 1996). The Hermite polynomial transformation expresses correlated variables as a linear combination of Hermite polynomials of the standard Gaussian variables using covariance and estimated moments such as mean, variance, skewness, and kurtosis. The Winterstein approximation is a specific type of the Hermite polynomial, which uses a linear combination of three Hermite polynomials. However, the accuracy of these two transformations is directly determined by accuracy of the estimated statistical moments, especially kurtosis and skewness, which are difficult to obtain accurately when the available data are limited. On the other hand, the input marginal CDFs and their parameters, which are required by the Rosenblatt, orthogonal, and Nataf transformations, can be more correctly determined using statistical methods based on samples than higher order moments.

The Rosenblatt transformation (Rosenblatt, 1952) requires complete information about the input variables such as a joint CDF (Melchers, 1999; Ditlevsen and Madsen, 1996). Unlike the Rosenblatt transformation, the Nataf transformation approximates the joint CDF using the Nataf model (Nataf, 1962), which is a Gaussian copula. Thus, the Nataf transformation can be used only for the correlation input variables with a Gaussian copula. Like the Nataf transformation, the orthogonal transformation is linear, so it can be used only if the input variables are correlated with Gaussian copula. If the true joint CDF

is different from the Gaussian copula, the Nataf and orthogonal transformation might yield erroneous reliability analysis results (Noh, 2007). Recently, a generalized Nataf transformation has been developed that can be applied for elliptical copulas, which include Gaussian and *t* copulas (Lebrun and Dutfoy, 2008). However, it is still limited to applications for the elliptical copula families. On the other hand, once the input joint CDF is obtained using copula, the Rosenblatt transformation method, which is analytically exact, can be used.

In the Rosenblatt transformation, there are $n!$ ways of transforming from n correlated input random variables into n independent Gaussian variables. The inverse reliability analysis results will be independent of the transformation ordering, if the probability of failure is calculated exactly in the independent standard Gaussian **U**-space (Noh, 2007). However, if the FORM is used for the inverse reliability analysis, certain orders of transformation could cause more errors than other orders of transformation according to the input joint CDF type. For example, if the input variables have a joint CDF modeled by a non-elliptical copula, which often occurs in industrial applications, then the Rosenblatt transformation becomes highly nonlinear for certain orders of transformation. This makes the transformed constraint functions become highly nonlinear. Thus, if the FORM is used, the inverse reliability analysis results could be very different for the different ordering since the FORM uses a linear approximation of the constraint function to estimate the probability of failure. This effect of transformation ordering in RBDO will be unacceptable and needs to be eliminated or reduced.

Thus, a more accurate method than the FORM is necessary to be able to handle highly nonlinear constraint functions. Reliability analysis using the SORM might be more

accurate than the FORM because it uses the second-order approximation. However, since the second-order derivatives are very difficult and expensive to obtain, the SORM cannot be used practically. To achieve both the efficiency of the FORM and the accuracy of the SORM, the MPP-based DRM is developed (Lee, 2008). The MPP-based DRM is used to accurately calculate the probability of failure after finding the FORM-based MPP, which is then used to develop an enhanced inverse reliability analysis method (i.e., MPP search) that is accurate for highly nonlinear and multi-dimensional problems. Hence, the MPP-based DRM, which yields accurate RBDO results without requiring the second-order sensitivity, can be used to reduce the effect of the transformation ordering on RBDO results.

1.1.3 Reliability-Based Design Optimization with Associated Confidence Level

In the RBDO design process, the input statistical uncertainty is modeled by identification of marginal CDFs and joint CDF using the Bayesian method (Noh et al., 2009a); the input physical uncertainty is modeled by quantification of parameters of the identified marginal CDFs and joint CDF, such as the mean, standard deviation, and correlation coefficient (Noh et al., 2009b). To obtain accurate optimum results, the input model needs to be correctly identified and accurately quantified; if not, the optimum results might be significantly unreliable, no matter how accurate the reliability analysis method is. Research on the identification of marginal and joint distribution types for RBDO has been recently carried out (Noh et al., 2009), but the input model with confidence level has not been considered for RBDO.

To offset the incorrect identification and inexact quantification of the input model, the input model with a confidence level needs to be implemented for RBDO. For this, confidence intervals of the input parameters, i.e., the mean, standard deviation, and correlation coefficient, need to be considered in order to have a desirable confidence level of the input model. The confidence intervals of the mean and standard deviation are usually calculated from the normality assumption that an input random variable has Gaussian distribution (Haldar and Mahadevan, 2000). However, the calculation of confidence intervals using the normality assumption is not accurate for non-Gaussian distributions (Mooney and Duval, 1993).

Thus, use of a bootstrap method, which does not require the normality assumption on the random variable, is necessary for non-Gaussian distribution. The bootstrap method first constructs a distribution of a random variable using given samples in either a nonparametric or parametric way. From the constructed distribution of the random variable, a distribution of a parameter such as mean or standard deviation is obtained using the normal approximation, percentile, bias corrected, bias corrected accelerated, or percentile-*t* methods (Mooney and Duval, 1993; Efron and Tibshirani, 1993). Using the distribution of the parameter constructed from those methods, the confidence interval of the parameter in which we are interested is obtained.

Use of the upper bound of the confidence interval of the standard deviation yields more reliable design than use of the estimated standard deviation, so that it is used in order to obtain a desirable confidence level for the input model. On the other hand, either the lower or upper bounds of the confidence intervals of the mean and correlation coefficient cannot be directly used because they do not necessarily yield more reliable

design. Thus, it is necessary to develop some adjusted parameters involving the confidence intervals of all input parameters such that the adjusted parameters offset the inaccurate estimation of the parameters. Accordingly, it is desirable to use the input model using the adjusted parameters to have a desirable confidence level of the input model and output performance, i.e., target probability of failure at the optimum design.

1.2 Objectives of the Proposed Study

The first objective of this study is to develop a Bayesian method to identify (1) the marginal CDFs and (2) the joint CDF, which are the input information to carry out the RBDO, from experimental data. A copula is used to model the joint CDF of the input variables using limited information such as marginal CDFs and given test data. To accurately model the input joint CDF, the correct copula should be identified. For this, statistical methods such as the GOF test (Genest and Favre, 2007; Genest and Rémillard, 2005; Wang and Well, 2000) and the Bayesian method (Huard et al., 2006) have been recently developed. To find which method more accurately identifies the correct joint CDF (copula), a comparison study will be carried out. Further, the performance of the Bayesian method will be tested for a different number of samples and different parameters such as correlation coefficients.

Likewise, the identification of marginal CDFs is as important as the identification of the copula. The marginal CDFs are sometimes known, but if not, they should be identified from the experimental data. Thus, a Bayesian method is proposed for identification of a marginal CDF that best describes the data among candidates. To see advantages of the proposed method, comparison studies between the proposed Bayesian

method and the GOF test (Haldar and Mahadevan, 2000; Sterne and Smith, 2000) will be carried out. Using simulated data sets, the performance of the Bayesian method is tested for different numbers of samples and different parameters such as mean. Using the Bayesian method, it will be shown how the marginal CDFs and joint CDF of input variables are identified through numerical examples and fatigue problems.

The second objective of the study is to reduce the effect of transformation ordering on RBDO results using the MPP-based DRM. Once the joint CDF is identified by the Bayesian method, the Rosenblatt transformation is utilized to transform the original random variables into the independent standard Gaussian variables for the inverse reliability analysis. In doing so, the order of the transformation can be arbitrary. If a different order of Rosenblatt transformation is used, even the mildly nonlinear constraint function in the original space could become highly nonlinear in the standard Gaussian space depending on the types of the joint input CDF.

First, obviously, if the input variables are independent (i.e., the joint CDF is a simple multiplication of the marginal CDFs), there is no effect of transformation ordering. Second, if the input variables have the joint CDF modeled by an elliptical copula, the effect of transformation ordering still does not exist because the elliptical copula makes the Rosenblatt transformation become linear, which is independent of orderings. The proof that the Rosenblatt is independent of the transformation ordering for independent and elliptical copula will be illustrated. Finally, when the input variables are correlated with joint CDFs modeled by non-elliptical copulas, the constraint function in the original space could lead to a highly nonlinear constraint function in the Gaussian space. Accordingly, it becomes a significant challenge to accurately carry out the inverse

reliability analysis using the FORM, which cannot accurately estimate the probability of failure for nonlinear functions. In this study, the MPP-based DRM, which can accurately handle the nonlinear constraint function, is used for the RBDO of problems with correlated input variables having non-elliptical copulas.

The third objective of this study is to develop an input model with a confidence level for RBDO to compensate the incorrect identification and inexact quantification of the input model uncertainty due to limited input data. For this, instead of using the estimated standard deviations of marginal CDFs, the adjusted standard deviation and correlation coefficient that compensate inaccurate estimation of all parameters are used because they provide more reliable optimum design with desired confidence level than the estimated parameters. In addition, the calculation of the adjusted standard deviation includes the confidence interval of the mean and upper bound of the confidence interval of standard deviation, which are usually obtained from a normality assumption that the input variable follows Gaussian distribution. If the input variable follows a non-Gaussian distribution, the confidence interval of the mean and the upper bound of the confidence interval of the standard deviation could be inaccurate. The confidence interval of the mean can be accurately calculated compared to the confidence interval of the standard deviation for the non-Gaussian distribution, so that a bootstrap method is proposed for accurate calculation of the confidence intervals of the standard deviation for both Gaussian and non-Gaussian distributions.

1.3 Organization of the Thesis Proposal

Chapter 2 presents engineering applications with correlated variables, fundamental concepts of the copula, correlation measures related to the copula, and commonly used copulas, which are helpful for understanding why the copula is necessary to model a joint CDF of input variables.

Chapter 3 presents Bayesian methods to identify marginal CDFs and joint CDF (copula) from experimental data. First, a Bayesian method and GOF test for identifying marginal CDFs are explained, and then the performances of the two methods are compared through simulated data sets. Second, a Bayesian method and GOF test for identifying the copula are explained. Likewise, comparison study of two methods is carried out in terms of which method identifies the copula more efficiently and correctly.

Chapter 4 illustrates the fundamental concepts of reliability analysis, such as FORM and SORM, and RBDO. It describes how the identified joint CDF and marginal CDFs are used in the transformation of the original input variables into independent Gaussian variables for the reliability analysis. Further, the MPP-based DRM that overcomes the disadvantages of the FORM and SORM is presented.

Chapter 5 presents the MPP-based DRM to alleviate the effect of the transformation ordering, especially when the input variables are correlated with non-elliptical copulas. Using numerical examples, the FORM and MPP-based DRM are compared in terms of how much the effect of transformation ordering is reduced.

Chapter 6 presents a mathematical example, a coil spring, and the roadarm of an M1A1 tank to show how the copula, Bayesian method, and MPP-based DRM work together in these examples.

Chapter 7 illustrates RBDO with associated confidence level to offset the inaccurate estimation of the input model caused by insufficient data. The numerical examples presented in Chapter 6 will be used to show how the input model with confidence level yields a desirable input and output confidence level compared with the one without confidence level.

Chapter 8 provides conclusions from the proposed work and suggestions for future work.

CHAPTER II

AN INTRODUCTION TO COPULAS

2.1 Introduction

This chapter presents the fundamental concepts and basic properties of copulas. Section 2.2 illustrates engineering applications with correlated input variables to show why the copula is important for modeling a joint CDF of correlated input variables. Sections 2.3 and 2.4 explain the basic definition of the copula and various correlation measures related to the copula, respectively. Section 2.5 introduces commonly used copula families such as elliptical copulas and Archimedean copulas. Since the identification of the correct copula is important for correctly modeling a joint CDF of input variables, it will be shown how copulas are identified using experimental data and how the identified copulas are used in RBDO in Chapters 3 and 4, respectively.

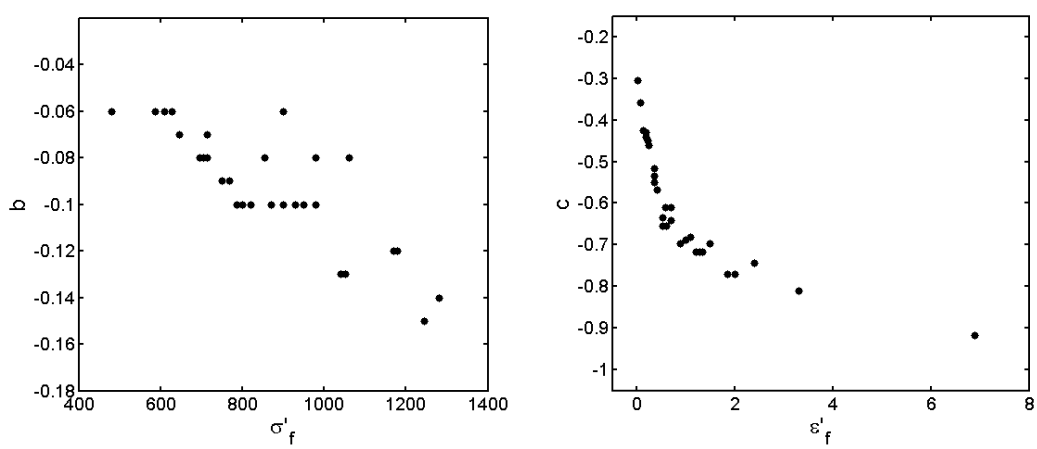
2.2 Engineering Application with Correlated Variables

In many structural RBDO problems, the input random variables such as the material properties and fatigue properties are correlated (Annis, 2004; Socie, 2003; Nikolaidis et al., 2005). For example, in fatigue problems, strain-life fatigue relationship is expressed as

$$\frac{\Delta\varepsilon_f^{'}}{2} = \frac{\sigma_f^{'}}{E}\left(2N_f\right)^b + \varepsilon_f^{'}\left(2N_f\right)^c \tag{2.1}$$

where $\Delta\varepsilon_f^{'}$ is the strain amplitude, E is the Young's modulus, N_f is the crack initiation fatigue life, $\sigma_f^{'}$ and *b* are the fatigue strength coefficient and exponent, and $\varepsilon_f^{'}$ and *c* are fatigue ductility coefficient and exponent, respectively.

Figure 2.1 shows 29 experimental data set of the fatigue strength coefficient $\sigma_f^{'}$ vs. fatigue strength exponent *b*, and fatigue ductility coefficient $\varepsilon_f^{'}$ vs. fatigue ductility exponent *c* of the SAE 950X high strength low alloy (Socie, 2003). As shown in Figure 2.1 (a) and (b), these variables are negatively highly correlated where the correlation coefficients between $\sigma_f^{'}$ and *b* and the one between $\varepsilon_f^{'}$ and *c* are calculated as 0 0.828 and 0 0.906, respectively.



(a) $\sigma_f^{'}$ and *b* (b) $\varepsilon_f^{'}$ and *c*

Figure 2.1. 29 Paired Data Obtained from 950X Steel (Socie, 2003)

It is known that $\sigma_f^{'}$ and $\varepsilon_f^{'}$ follow lognormal CDF and *b* and *c* follow Gaussian CDF (Socie, 2003). However, since the marginal distribution types of two correlated variables are different, i.e., $\sigma_f^{'}$ has a lognormal CDF and *b* has a Gaussian CDF, the joint Gaussian CDF, which is commonly used as a joint CDF of correlated variables, cannot be used because the joint Gaussian CDF means that the marginal CDFs of both random variables are Gaussian. Thus, when the marginal distribution types are mixed and random variables are correlated, a joint CDF that has different types of marginal CDFs needs to be used.

In addition, it could happen in real applications that marginal CDFs are Gaussian, but the joint CDF is not Gaussian. The following example shows an exhaust manifold used on a Chrysler 5.2L engine, as shown in Figure 2.2. The paired data between two random variables X and Y are collected from a machine capability study performed on the machine builder's floor at Chrysler Corporation (Pham, 2001). Figure 2.3 (a) shows the scatter plot of the data. Using the data, the mean values and standard deviations of X and Y are obtained as $\mu_X = 8.945$, $\mu_Y = 0.817$, $\sigma_X = 1.569E-3$, and $\sigma_Y = 1.514E-3$, respectively, and the correlation coefficient between the two variables is calculated as 0.401. As seen in Figure 2.3, the normality assumption for each margin seems to be acceptable from the normal Q-Q plots, which are graphical ways of diagnosing how well the Gaussian distribution describes given data (Pham, 2006). However, the joint PDF does not seem to fit the joint Gaussian PDF because the data are not symmetrically distributed. The data are densely distributed in the lower-left end, but they are widely spread out in the upper-right end. Based on the scatter plot of the data, a non-Gaussian joint PDF, which is modeled by a copula, might better describe the data than the joint

Gaussian PDF. Therefore, the copula needs to be used to model various types of joint CDFs of input random variables.

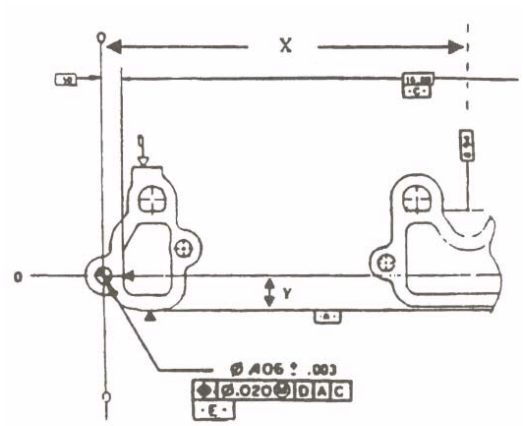
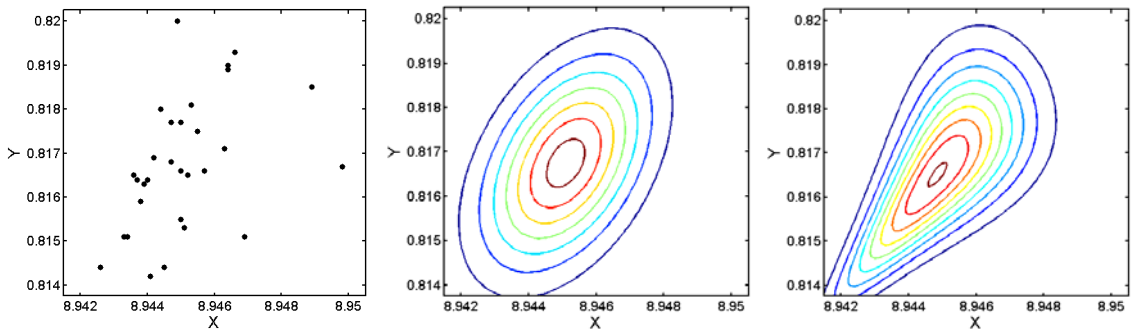


Figure 2.2. Exhaust Manifold (Pham, 2001)



(a) Scatter Plot (b) Gaussian PDF (c) non-Gaussian PDF

Figure 2.3. Scatter Plot of Data and PDF Contours between X and Y (Pham, 2006)

2.3 Basic Definition of Copulas

The word *copula* originates from a Latin word for “link” or “tie” that connects different things. In statistics, the definition of copula is stated by Nelson (1999):

"Copulas are functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions. Alternatively, copulas are multivariate distribution functions whose one-dimensional margins are uniform on the interval [0, 1]."

According to Sklar's theorem (Nelson, 1999), if the random variables $\mathbf{X} = \left[X_1, \cdots, X_n\right]^T$ have marginal distributions $F_{X_1}\left(x_1\right), \cdots$, and $F_{X_n}\left(x_n\right)$, then there exists an *n*-dimensional copula *C* such that

$$F_{X_1,\ldots,X_n}\left(x_1,\ldots,x_n\right) = C\left(F_{X_1}\left(x_1\right),\ldots,F_{X_n}\left(x_n\right)\middle|\boldsymbol{\theta}\right) \tag{2.2}$$

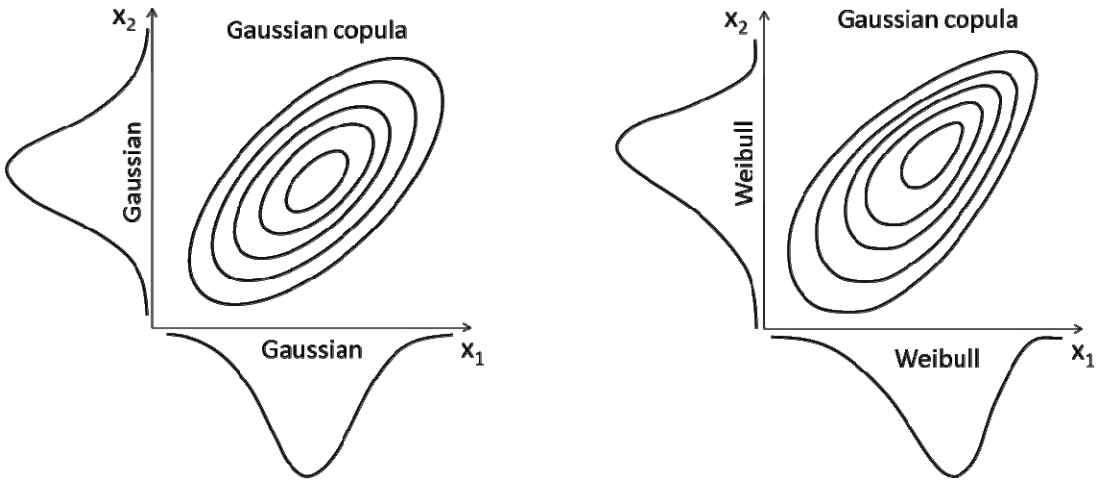
where $F_{X_1,\ldots,X_n}\left(x_1,\ldots,x_n\right)$ is the joint CDF of $\mathbf{X}$ and $\boldsymbol{\theta}$ is the matrix of correlation parameters between $X_1,\ldots,X_n$. If marginal distributions are all continuous, then *C* is unique. Conversely, if *C* is an *n*-dimensional copula and marginal CDFs are given, then the joint distribution is an *n*-dimensional function of marginal CDFs (Nelson, 1999). By taking the derivative of Eq. (2.2), the joint PDF is obtained as

$$f_{X_1,\ldots,X_n}\left(x_1,\cdots,x_n\right) = c\left(F_{X_1}\left(x_1\right),\cdots,F_{X_n}\left(x_n\right)\middle|\boldsymbol{\theta}\right)\prod_{i=1}^{n} f_{X_i}\left(x_i\right) \tag{2.3}$$

where $c\left(z_1,\cdots,z_n\middle|\boldsymbol{\theta}\right) = \dfrac{\partial^n C\left(z_1,\cdots,z_n\middle|\boldsymbol{\theta}\right)}{\partial z_1 \cdots \partial z_n}$ is the copula density function with $z_i = F_{X_i}\left(x_i\right)$, and $f_{X_i}\left(x_i\right)$ is the marginal PDF of X_i for $i = 1,\cdots,n$.

Since the joint CDF is expressed as a function of marginal CDFs, it is readily possible to model a joint CDF using marginal CDFs and correlation parameters $\boldsymbol{\theta}$ that can be obtained from the experimental data in real applications. Moreover, since the copula decouples marginal CDFs and the joint CDF, the joint CDF modeled by the copula can be expressed for any types of marginal CDFs. For example, Figure 2.4 shows PDF contours obtained from the Gaussian copula but with different types of marginal

distributions. Even though the same copula is used, various types of joint distributions can be generated according to the marginal distribution types. Thus, the copula can be used for constructing the joint CDF in real applications that have correlated input variables with a joint CDF but with different types of marginal CDFs of the input variables.



(a) Gaussian Copula + Gaussian Margins (b) Gaussian Copula + Weibull Margins

Figure 2.4. PDF Contours of Gaussian Copula with Different Marginal Distributions

To model the joint CDF using the copula, the matrix of correlation parameters $\boldsymbol{\theta}$ needs to be obtained from the experimental data. Since various types of copulas have their own correlation parameters, it is desirable to have a common correlation measure to obtain the correlation parameters from the experimental data.

2.4 Correlation Measures

Since the correlation parameters among random variables are used in a copula, a natural way to measure the correlation between random variables is through a copula. Many of the properties and correlation measures associated with copulas are scale invariant under monotonically increasing transformations of the marginal distributions because copulas are invariant under monotonically increasing transformation of the marginal distributions. This is a desirable property for correlation measures because if the correlation measure depends on the scale, it cannot be generally used to measure the dependence between random variables. For example, Pearson's rho is most commonly used in practice as a correlation measure. However, since Pearson's rho is not a copula-based correlation measure and only measures a linear correlation between random variables, it needs to be limitedly used as a correlation measure for the joint elliptical distribution. On the other hand, copula-based correlation measures such as Spearman's rho and Kendall's tau, which are also called rank correlation coefficients, can be widely used for various types of copulas.

The rank correlation coefficient was first introduced by Spearman, who was a psychologist, in 1904. In psychology, more than in any other science, it is hard to find a measure that estimates correlation between two variables because there are some cases where correlation cannot be measured quantitatively. For instance, the dependence between hereditary qualities of brothers cannot be quantitatively measured if Pearson's rho, which requires specific values of two variables, is used. On the other hand, if children of a school are divided into conscientious and non-conscientious group, the correlation can be measured by counting how much brothers tend to be in the same

division. Thus, in that case, comparison (ranking) of two groups is a better way to measure the dependence rather than measuring Pearson's rho. The rank does not change under strictly increasing function; hence it can be expressed as copulas.

To measures the degree of correspondence between two variables, the correspondence should be mathematically defined. Concordance is one way of expression for correspondence. If large values of one tend to be associated with large values of the other and small values of one with small values of the other, two variables are said to be concordant. Likewise, if large values of one tend to be associated with small values of the other, two variables are called discordant. Since the copula plays an important role in concordance, and dependence measures are also associated with the concordance, a concordance function Q needs to be introduced. The concordance is the difference between the probability of concordance and the probability of discordance for a pair of random vectors (X_1, Y_1) and (X_2, Y_2), and is defined as (Nelson, 1999)

$$Q = P\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - P\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right] \tag{2.4}$$

where (X_1, Y_1) and (X_2, Y_2) are independent vectors of continuous random variables with joint distributions $H_1(x, y) = C_1\left(F(x), G(y) \middle| \theta_1\right)$ and $H_2 = C_2\left(F(x), G(y) \middle| \theta_2\right)$, respectively, with same marginal distributions $u = F(x)$ and $v = G(y)$.

Since the random variables are continuous, $P\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right] = 1 - P\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right]$. Thus, Eq. (2.4) can be rewritten as

$$Q = 2P\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - 1 \tag{2.5}$$

where $P\left[\left(X_1 - X_2\right)\left(Y_1 - Y_2\right) > 0\right] = P\left[X_1 > X_2, Y_1 > Y_2\right] + P\left[X_1 < X_2, Y_1 < Y_2\right]$, and $P\left[X_1 > X_2, Y_1 > Y_2\right]$ is obtained as (Nelsen, 1999)

$$\begin{aligned} P\left[X_1 > X_2, Y_1 > Y_2\right] &= P\left[X_2 < X_1, Y_2 < Y_1\right] \\ &= \iint_{\mathbf{R}^2} C_2\left(F\left(x\right), G\left(y\right) \middle| \theta_2\right) dC_1\left(F\left(x\right), G\left(y\right) \middle| \theta_1\right) \end{aligned} \tag{2.6}$$

by integrating over the distribution of one of vectors, either $\left(X_1, Y_1\right)$ or $\left(X_2, Y_2\right)$ with $\mathbf{R}^2 = \left(-\infty, \infty\right) \times \left(-\infty, \infty\right)$. Employing $u = F\left(x\right)$ and $v = G\left(y\right)$, Eq. (2.6) is written as

$$P\left[X_1 > X_2, Y_1 > Y_2\right] = \iint_{I^2} C_2\left(u, v \middle| \theta_2\right) dC_1\left(u, v \middle| \theta_1\right) \tag{2.7}$$

where $I^2 = I \times I$ ($I = \left[0,1\right]$). Likewise,

$$P\left[X_1 < X_2, Y_1 < Y_2\right] = \iint_{I^2} \left[1 - u - v + C_2\left(u, v \middle| \theta_2\right)\right] dC_1\left(u, v \middle| \theta_1\right) \tag{2.8}$$

Since $C_1\left(u, v \middle| \theta\right)$ is the joint CDF of uniform variables U and V, $E\left(U\right) = E\left(V\right) = \frac{1}{2}$.

Thus, Eq. (2.8) becomes

$$\begin{aligned} P\left[X_2 < X_1, Y_1 < Y_2\right] &= 1 - \frac{1}{2} - \frac{1}{2} + \iint_{\mathbf{I}^2} C_2\left(u, v \middle| \theta_2\right) dC_1\left(u, v \middle| \theta_1\right) \\ &= \iint_{I^2} C_2\left(u, v \middle| \theta_2\right) dC_1\left(u, v \middle| \theta_1\right) \end{aligned} \tag{2.9}$$

Substituting Eq. (2.7) and (2.9) into Eq. (2.5), Eq. (2.5) can be expressed as

$$Q = Q\left(C_1, C_2\right) = 4\int\int_{I^2} C_2\left(u, v \middle| \theta_2\right) dC_1\left(u, v \middle| \theta_1\right) - 1 \tag{2.10}$$

Using Eq. (2.10), correlation measures associated with copulas, Spearman's rho and Kendall's tau, can be expressed in terms of the concordance function.

2.4.1 Linear Correlation Coefficient (Pearson's Rho)

Pearson's rho, which is also called a product moment correlation coefficient, was first discovered by Bravais (1846) and was developed by Pearson (1896). Pearson's rho indicates the degree of linear relationship between two random variables as follows:

$$\rho = \frac{Cov\left(X,Y\right)}{\sigma_X \sigma_Y} \tag{2.11}$$

where σ_X and σ_Y are standard deviations of X and Y, respectively and $Cov\left(X,Y\right)$ is the covariance between X and Y. Since Pearson's rho only indicates the linear relationship between two random variables, it is not a good measure for nonlinear relationship between two random variables, which often occurs in practical engineering applications. If the given data follows a joint CDF modeled by a non-elliptical distribution, another measure needs to be introduced to estimate the dependency between random variables. Therefore, Pearson's rho is only valid when the joint distribution is modeled by an elliptical.

2.4.2 Spearman's Rho

The Spearman's rho is the special case of the Pearson's rho in which paired samples are converted to rankings before calculating Eq. (2.11). Thus, the sample version of the Spearman's rho is given as (Spearman, 1904; Kendall, 1938)

$$r_s = 1 - \frac{6\sum_{i=1}^{ns} d_i^2}{ns\left(ns^2 - 1\right)} \tag{2.12}$$

where d_i is the difference of two rankings and *ns* is the number of samples.

Employing the concordance function shown in Eq. (2.4), the Spearman's rho is defined to be proportional to the probability of concordance minus the probability of discordance between two random vectors (X_1, Y_1) and (X_2, Y_3) with same marginal distributions $u = F(x)$ and $v = G(y)$, but with different copulas, $H_1(x, y) = C(F(x), G(y)|\theta)$ of (X_1, Y_1) and $H_2(x, y) = \Pi(x, y) = F(x)G(y)$ of (X_2, Y_3) where Π is an independent copula, $\Pi = uv$. Thus, the population version of Spearman's rho is defined as (Nelsen, 1999)

$$\rho_s = 3\left(P\left[(X_1 - X_2)(Y_1 - Y_3) > 0\right] - P\left[(X_1 - X_2)(Y_1 - Y_3) < 0\right]\right) \tag{2.13}$$

Here, the multiplication of 3 is to make Spearman's rho have ranges –1 to 1. Using Eq. (2.10), Eq. (2.13) can be rewritten in terms of a copula as

$$\rho_s = 3Q(C, \Pi) = 12\int\int_{I^2} uvdC(u, v|\theta) - 3 = 12\int\int_{I^2} C(u, v|\theta)dudv - 3 \tag{2.14}$$

Using Eq. (2.14), the Spearman's rho ρ_s can be expressed as a function of the correlation parameter θ if the copula type is known as shown in Table 2.1. Table 2.1 shows Spearman's rho expressed as a function of the correlation parameter, $\rho_s = g(\theta)$, for some copulas such as Frank, FGM (Farlie-Gumbel-Morgenstern), and Gaussian (Bacigál and Komorníková, 2006; Trevedi and Zimmer, 2007). In practical applications, since only samples are available, the Spearman's rho is calculated using Eq. (2.12) and then the correlation parameter is estimated using Eq. (2.14) or explicit functions such as Table 2.1. Still, many other copulas do not have explicit functions or have complicated formulations as shown in Table 2.1. Thus, the correlation parameters of the copulas not

having explicit formulations of Spearman's rho should be calculated by implicitly solving Eq. (2.14).

Table 2.1. Spearman's Rho and Its Domain

Copula	$\rho_s = g(\theta)$	$\rho_s \in \Omega_s$
Clayton	Complicated form	$(0,1]$
AMH*	Complicated form	$[-0.271064, 0.478407]$
Gumbel	No closed form	$[0,1]$
Frank	$1-\frac{12}{\theta}\left\{D_1^{**}(\theta)-D_2^{**}(\theta)\right\}$	$[-1,1]\setminus\{0\}$
FGM	$\frac{\theta}{3}$	$[-1/3,1/3]$
Gaussian	$\frac{6}{\pi}\arcsin\frac{\theta}{2}$	$[-1,1]$

*AMH: Ali-Mikhail-Haq, ${}^{**}D_k(\theta)=\frac{k}{\theta^k}\int_0^x \frac{t^k}{e^t-1}dt$

2.4.3 Kendall's tau

The sample version of Kendall's tau is defined in terms of concordance (Kruskal, 1958; Hollander and Wolfe, 1973; Lehmann, 1975). Let ns paired samples be $\left[(x_1,y_1),(x_2,y_2),\cdots,(x_{ns},y_{ns})\right]^T$ randomly generated from continuous random variables of X and Y. Among $\binom{ns}{2}$ distinct paired samples, two paired samples (x_i,y_i) and (x_j,y_j) are concordant if $x_i<x_j$ and $y_i<y_j$ or if $x_i>x_j$ and $y_i>y_j$. Otherwise, those are discordant. Let c denote the number of concordant pairs and d denote the number of

discordant pairs. Since the Kendall's tau is the probability of concordance minus the probability of discordance, the sample version of Kendall's tau is defined as

$$t = \frac{c-d}{c+d} = (c-d)/\binom{ns}{2} \tag{2.15}$$

The population version of Kendall's tau for two continuous random variables X and Y with a joint distribution is similarly defined as

$$\tau = P\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - P\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right] \tag{2.16}$$

According to the definition of the Kendall's tau, copulas of (X_1, Y_1) and (X_2, Y_2) have a common copula, $C(F(x), G(y)|\theta)$, the population version of the Kendall's tau can be defined as

$$\tau = Q(C, C) = 4\int\int_{I^2} C(u, v|\theta)\, dC(u, v|\theta) - 1 \tag{2.17}$$

using the concordance function in Eq. (2.10).

If only the number of samples is available, the Kendall's tau can be calculated using Eq. (2.15). If Kendall's tau is known, then the correlation parameter of the copula can also be calculated by using the explicit function of Kendall's tau as shown in Table 2.2 or by using Eq. (2.17) implicitly. In most cases, Kendall's tau can be expressed as a function of the correlation parameter as shown in Table 2.2. The domain of Kendall's tau is defined according to the copula type. More detailed information on Spearman's rho and Kendall's tau is presented by Kruskal (1958).

Table 2.2. Kendall's Tau and Its Domain

Copula	$\tau = g(\theta)$	$\tau \in \Omega$
Clayton	$1-\frac{2}{2+\theta}$	$(0,1]$
AMH	$1-\frac{2}{3}\frac{(\theta-1)^2\ln(1-\theta)+\theta}{\theta^2}$	$[-0.181726,1/3]$
Gumbel	$1-\theta^{-1}$	$[0,1]$
Frank	$1-\frac{4}{\theta}\left(1-\frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}dt\right)$	$[-1,1]\setminus\{0\}$
A12	$1-\frac{2}{3\theta}$	$[1/3,1]$
A14	$1-\frac{2}{1+2\theta}$	$[1/3,1]$
FGM	$2\theta/9$	$[-2/9,2/9]$
Gaussian	$\frac{2}{\pi}\arcsin\theta$	$[-1,1]$

2.5 Commonly Used Copulas

In this section, two commonly used copulas—the elliptical copulas and Archimedean copulas—are introduced. These copulas are the most popular because they can be extended to multivariate distributions and are easy to handle because of their tractable characteristics. For example, elliptical copulas provide a linear transformation from original variables to standard Gaussian variables, and thus it is easy to transform from one to another. In the case of Archimedean copulas, each copula has a unique generator function, and the generator is used to calculate the correlation measure, Kendall's tau. Thus, Kendall's tau can be easily obtained without using Eq. (2.17), which requires double integration on I^2, as will be explained in Section 2.5.2.

2.5.1 Elliptical Copulas

Elliptical copulas are copulas of elliptical distributions. A random variable $\mathbf{X}$ has an elliptical distribution, if and only if there exists $\mathbf{S}$, R, and $\mathbf{A}$ that satisfy

$$\mathbf{X} = \boldsymbol{\mu} + R\mathbf{AS} = \boldsymbol{\mu} + \mathbf{AY} \tag{2.18}$$

where R is a non-negative random variable, which is independent of *k*-dimensional uniformly distributed random vector $\mathbf{S}$, and a $d \times k$ matrix $\mathbf{A}$ with $\mathbf{AA}^T = \mathbf{P}$ where $\mathbf{P}$ is covariance matrix, and d and k are the number of random variables in $\mathbf{X}$ and $\mathbf{S}$, respectively (Embrechts et al., 2003; Stander, 2006). It is only considered that rank (**A**) = $d \le k$ and $\mathbf{P}$ is a positive definite matrix with a full rank. When R^2 has a chi-square distribution with d degree-of-freedom, $\mathbf{Y}$ has a standard Gaussian distribution $N(\mathbf{0},\mathbf{1})$, and, accordingly, $\mathbf{X}$ has a Gaussian distribution $N(\boldsymbol{\mu},\mathbf{P})$ for *d*-dimensional variables. If R^2/d has F distribution with d and ν degrees-of-freedom, $\mathbf{Y}$ has a standard *t* distribution $T_\nu(\nu,\mathbf{0},\mathbf{1})$ and $\mathbf{X}$ has a *t* distribution with $T_{\mathbf{P},\nu}(\nu,\boldsymbol{\mu},\mathbf{P})$ for d dimensional variables. The *t* distribution and Gaussian distribution originate the *t* copula and the Gaussian copula, respectively (Pham, 2006).

The Gaussian copula is defined as the joint Gaussian CDF of standard Gaussian variables $\Phi^{-1}(z_1)$, $\cdots$, $\Phi^{-1}(z_n)$ as

$$C_\Phi(z_1,\cdots,z_n|\mathbf{P}) = \Phi_{\mathbf{P}}(\Phi^{-1}(z_1),\ldots,\Phi^{-1}(z_n)|\mathbf{P}), \quad \mathbf{z} \in I^n \tag{2.19}$$

where $z_i = F_{X_i}(x_i)$ is the marginal CDF of X_i for $i = 1,\cdots,n$, and $\mathbf{P}$ is the covariance matrix consisting of correlation coefficients, Pearson's rho, between correlated random variables. $\Phi^{-1}(\cdot)$ represents the inverse of marginal standard Gaussian CDF defined as

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \tag{2.20}$$

and $\Phi_{\mathbf{P}}(\cdot)$ is the joint Gaussian CDF, which is expressed as a multi-dimensional integral as

$$\Phi_{\mathbf{P}}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\mathbf{t}-\boldsymbol{\mu})^T (\mathbf{P})^{-1} (\mathbf{t}-\boldsymbol{\mu})\right] d\mathbf{t} \tag{2.21}$$

for $\mathbf{x} = [x_1, \cdots, x_n]^T$ with a mean vector $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_n]^T$. If the marginal CDFs are Gaussian, then the joint CDF modeled by the Gaussian copula is the joint Gaussian CDF as shown in Figure 2.4 (a). When the marginal CDFs are not Gaussian, various types of joint distributions can be generated using the Gaussian copula as seen in Figure 2.4 (b).

The *t* copula can be constructed using *t* distribution like the Gaussian copula originated from the Gaussian distribution. The *t* copula is defined as a multivariate *t* distribution of standard *t* variables with $T_{\mathbf{P},\nu}(\mathbf{0}, \mathbf{P}, \nu)$ where ν is the degree of freedom.

$$C_T(z_1, \cdots, z_n | \mathbf{P}, \nu) = T_{\mathbf{P},\nu}\left(T_\nu^{-1}(z_1), \cdots, T_\nu^{-1}(z_n) | \mathbf{P}, \nu\right) \tag{2.22}$$

where $\mathbf{P}$ is the covariance matrix and $T_\nu^{-1}(\cdot)$ is the inverse of the standard student *t* distribution with ν degree of freedom defined as

$$T_\nu(x) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt \tag{2.23}$$

where $\Gamma(\cdot)$ is the gamma function, which is defined as

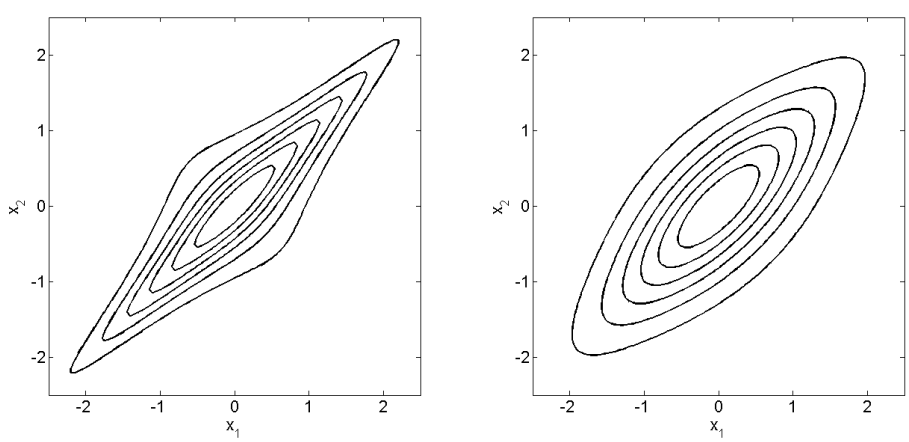
$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \tag{2.24}$$

The joint *t* distribution is given by

$$T_{\mathbf{P},\nu}\left(\mathbf{x}\right)=\int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1}\frac{\Gamma\left(\left(\nu+n\right)/2\right)}{\Gamma\left(\nu/2\right)\sqrt{\nu^n\pi^n\left|\mathbf{P}\right|}}\left\{1+\frac{\left(\mathbf{t}-\boldsymbol{\mu}\right)^T\mathbf{P}^{-1}\left(\mathbf{t}-\boldsymbol{\mu}\right)}{\nu}\right\}^{-\frac{n+\nu}{2}}d\mathbf{t} \qquad (2.25)$$

Note that a Gaussian copula with Gaussian marginal CDFs is the same as a joint Gaussian CDF. However, a *t* copula with *t* marginal CDFs might not be a joint *t* distribution. A joint *t* distribution must have same degrees of freedom at all marginal CDFs. In contrast, the *t* copula with *t* marginal CDFs can have different degrees of freedom at each marginal CDF, which provides more flexibility in modeling heavily tailed data (Pham, 2006).

Figure 2.5 shows the PDF contours obtained from *t* copula with marginal standard Gaussian distributions for different degrees of freedom, ν=1 and ν=5. The same copula and marginal distributions are used, but the shapes of the joint distributions can be quite different because of the different degrees of freedom.



(a) ν=1 (b) ν=5

Figure 2.5. PDF Contours of *t* Copula with Marginal Gaussian Distributions for Different Degrees of Freedom

2.5.2 Archimedean Copulas

The Archimedean copula is constructed in a completely different way than the elliptical copula. An important component of constructing Archimedean copulas is a generator function, which is a complete monotone decreasing function. The function $\varphi_\theta(t)$ is completely monotonic on an interval $I=[0,1]$ if it is continuous and has derivatives of all orders that alternate in sign

$$(-1)^k \frac{d}{d^k}\varphi_\theta(t) \geq 0, \qquad k=1,2,\cdots \tag{2.26}$$

If $\varphi_\theta(\cdot)$ is a continuous and strictly decreasing function from $[0,1]$ to $[0,\infty)$ such that $\varphi_\theta(0)=\infty$ and $\varphi_\theta(1)=0$ and the inverse φ_θ^{-1} is completely monotonic on $[0,\infty)$, then an *n*-copula, which is called an Archimedean copula, for all $n \geq 2$ can be defined as

$$C(z_1,\cdots,z_n|\theta)=\varphi_\theta^{-1}\left[\varphi_\theta(z_1)+\cdots+\varphi_\theta(z_n)\right] \tag{2.27}$$

Each Archimedean copula has a corresponding unique generator function, and the generator function provides copulas, as seen in Eq. (2.27). In case of the Archimedean copulas, once the generator is provided, Kendall's tau can be obtained as

$$\tau = 1+4\int_0^1 \frac{\varphi_\theta(t)}{\varphi_\theta^{'}(t)}dt \tag{2.28}$$

Unlike Kendall's tau, there is no explicit formulation that estimates Spearman's rho using the generator. Table 2.3 presents copula functions, generator functions, and domains of correlation parameters for selected Archimedean copulas. FGM and Gaussian copulas are not shown in this table because those are not Archimedean copulas. Figure 2.6 shows PDF contours obtained from Clayton, Gumbel, Frank, A12, and A14 using marginal standard Gaussian distributions for $\tau = 0.5$.

Table 2.3. Copula Functions, Generator Functions, and Domains of Correlation Parameters

Copula	$C(u,v\|\theta)$	$\varphi_\theta(t)$	$\theta \in \Omega^\theta$
Clayton	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$	$\frac{1}{\theta}\left(t^{-\theta}-1\right)$	$(0,\infty)$
AMH	$uv/\{1-\theta(1-u)(1-v)\}$	$\ln\{(1-\theta+\theta t)/t\}$	$[-1,1)$
Gumbel	$\exp\left\{-\left[(-\ln u)^\theta+(-\ln v)^\theta\right]^{1/\theta}\right\}$	$(-\ln t)^\theta$	$[1,\infty)$
Frank	$-\frac{1}{\theta}\ln\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right)$	$-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$(-\infty,\infty)$
A12	$\left\{1+\left[\left(u^{-1}-1\right)^\theta+\left(v^{-1}-1\right)^\theta\right]^{1/\theta}\right\}^{-1}$	$\left(\frac{1}{t}-1\right)^\theta$	$[1,\infty)$
A14	$\left\{1+\left[\left(u^{-1/\theta}-1\right)^\theta+\left(v^{-1/\theta}-1\right)^\theta\right]^{1/\theta}\right\}^{-\theta}$	$\left(t^{-1/\theta}-1\right)^\theta$	$[1,\infty)$

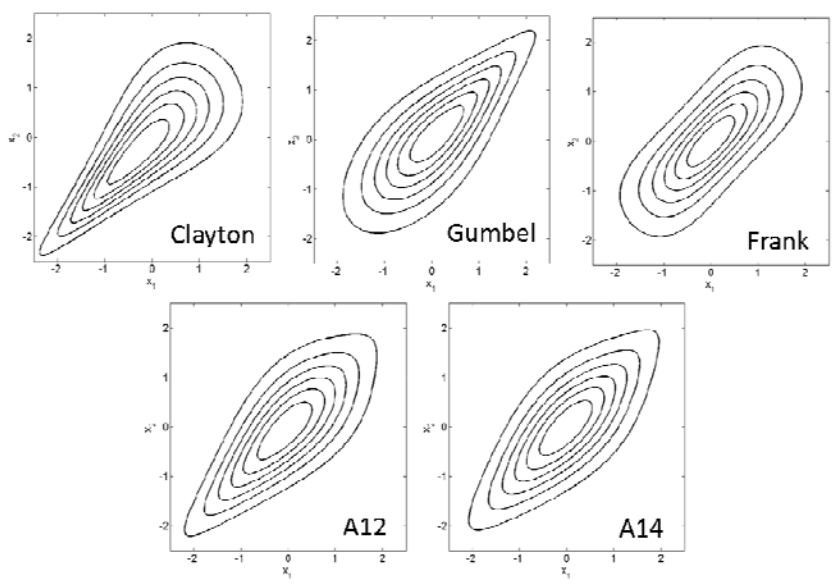


Figure 2.6. PDF Contours of Selected Archimedean Copulas Using Marginal Standard Gaussian Distributions for $\tau = 0.5$

CHAPTER III

IDENTIFICATION OF MARGINAL AND JOINT CUMULATIVE DISTRIBUTION FUNCTIONS

3.1 Introduction

To carry out the RBDO, marginal and joint distributions of input random variables need to be correctly identified and accurately quantified, which is challenging especially when the input variables are correlated and only limited data are available. In this study, a copula is utilized to model the joint CDF of the input variables using limited information such as marginal CDF types (if they are known) and given test data. Since the correct identification of the copula is necessary to model the correct joint CDF, two statistical methods, the goodness-of-fit test (GOF) and the Bayesian method, can be used to identify a copula that best describes given experimental data. In practical applications, certain input variables are known to follow specific marginal CDF types; for example, some fatigue material properties are known to follow lognormal CDFs. However, if the input marginal CDF type is not known, it is necessary to identify the marginal CDF type from given data. As the GOF test and the Bayesian method are used to identify the copula, they can also be used to identify the marginal CDFs.

Section 3.2 illustrates the GOF test and the Bayesian method for the identification of marginal CDFs. Section 3.3 explains the GOF test and the Bayesian method for the identification of the copula. Section 3.4 compares the GOF test and the Bayesian method to show which method more effectively identifies correct marginal CDFs and copula from data sets generated from given true joint CDFs. Section 3.5 illustrates numerical

examples and a fatigue problem to show how the Bayesian method, which is proposed to be used in this study, works.

3.2 Identification of Marginal CDFs

The two most commonly used methods for determining marginal CDFs for the given data are the GOF test and the Bayesian method. The GOF test has been developed and widely used to identify the marginal CDF and calculate its parameters, such as mean and standard deviation. However, since the GOF test relies on those parameters estimated from samples, if the parameters are incorrectly estimated, then the wrong marginal CDF might be identified. On the other hand, since the Bayesian method calculates weights to identify the marginal CDF by integrating the likelihood function over one parameter, either mean or standard deviation, it is less dependent on the estimation of the parameter than the GOF test. Therefore, the Bayesian method is preferred over the GOF test.

3.2.1 Goodness-of-fit Test

The most natural way of checking the adequacy of a hypothesized CDF is to compare the empirical CDF and the hypothesized CDF. There are several types of GOF tests: Chi-square, Cramér-von Mises, Kolmogorov-Smirnov (K-S), etc. The Chi-square test, which compares the difference between the empirical PDF and the hypothesized PDF, requires sufficient data, so that it can be used only for a large number of data. The Cramér-von Mises test is based on the integrated difference between an empirical CDF and a hypothesized CDF, weighted by the hypothesized PDF. The Cramér-von Mises test is known as a more powerful method than the K-S test, but its application is limited to the

symmetric and right-skewed CDFs, unlike the K-S test, which is applicable to all types of CDFs (Cirrone et al., 2004). Thus, in this study, the K-S test is compared with the Bayesian method.

The K-S test compares the empirical CDF and the hypothesized (or theoretical) CDF as

$$D_n = \max_x \left| F_X(x) - G_n(x) \right| \tag{3.1}$$

where $F_X(x)$ and $G_n(x)$ are the hypothesized CDF and the empirical CDF, respectively. Since D_n is a random variable, the CDF of D_n is related to a significance level α as

$$P\left(D_n \le D_n^{\alpha}\right) = 1 - \alpha \tag{3.2}$$

for the confidence level, $1-\alpha$. D_n^{α} is a critical value obtained from a standard mathematical table presented by Haldar and Mahadevan (2000). The maximum difference between $F_X(x_i)$ and $G_n(x_i)$ of the ith sample x_i for $i = 1, \cdots, ns$ is calculated as

$$D_n^{\max} = \max_{1 \le i \le ns} \left| F_X(x_i) - G_n(x_i) \right| \tag{3.3}$$

If the maximum difference $D_n^{\max}$ calculated from samples is smaller than D_n^{α}, the null hypothesis that the given samples come from the hypothesized CDF is accepted; otherwise, it is rejected. Likewise, the p-value also can be used to test the null hypothesis. The p-value is a measure of how much evidence we have against the null hypothesis, and calculated as the CDF value of D_n. If the p-value is larger than α, then the null hypothesis is accepted; otherwise, it is rejected. The larger the p-value, the more strongly the test accepts the null hypothesis. Much like the Bayesian method uses the calculated

weights of candidate marginal CDFs, which will be explain in Section 3.2.2, to identify a marginal CDF, the K-S test uses the calculated p-values of candidate marginal CDFs to identify a marginal CDF. In the K-S test, since D_n is known to follow the Kolmogorov distribution, the p-value is calculated from the Kolmogorov CDF at $D_n^{\max}$. Accordingly, using the calculated p-values, a marginal CDF with the highest p-value is selected as the best fitted marginal CDF in the K-S test.

3.2.2 Bayesian Method

The Bayesian method can be used to identify the best fitted marginal CDF among candidate CDFs by calculating the weights of the candidate CDFs to select the one with the highest weight. Consider a finite set $s_q \subset s$ consisting of candidate marginal CDFs M_k, $k = 1, \cdots, q$, where s is a set of all marginal CDFs and q is the number of the candidate marginal CDFs. The Bayesian method consists of defining q hypotheses:

h_k: The data come from marginal CDF M_k, $k = 1, \cdots, q$.

The probability of each hypothesis h_k given the data D is defined as

$$\Pr\left(h_k \middle| D, I\right) = \frac{\Pr\left(D \middle| h_k, I\right)\Pr\left(h_k \middle| I\right)}{\Pr\left(D \middle| I\right)} \tag{3.4}$$

where $\Pr\left(D \middle| h_k, I\right)$ is the likelihood function, $\Pr\left(h_k \middle| I\right)$ is the prior on the marginal CDF, and $\Pr\left(D \middle| I\right)$ is the normalization constant with any relevant additional knowledge I. These terms will be explained in detail in the following.

First, consider the likelihood function, $\Pr\left(D\middle|h_k, I\right)$ in Eq. (3.4). Under the hypothesis h_k that the data D come from the marginal CDF M_k, the probability of drawing the data D for the hypothesis on M_k is expressed as a likelihood function as

$$\Pr\left(D\middle|h_k, \mu, \sigma, I\right) = \prod_{i=1}^{ns} f_k\left(x_i \middle| a\left(\mu,\sigma\right), b\left(\mu,\sigma\right)\right) \tag{3.5}$$

where x_i is the ith sample value. Since each marginal PDF f_k has its own parameters *a* and *b*, common parameters such as mean or standard deviation need to be used. For most marginal CDFs, their own parameters (*a* and *b*) are expressed as functions of mean and standard deviation as shown in Table 3.1, and thus the likelihood function can be expressed in terms of mean and standard deviation for given samples.

Table 3.1. Mean, Variance, and Their Domains

Distribution	μ and σ^2	$\mu \in \Omega^{\mu}$	$\sigma \in \Omega^{\sigma}$
Gaussian	μ, σ^2	$\left(-\infty, \infty\right)$	$\left[0, \infty\right)$
Weibull	$\mu = a\Gamma\left(1+\frac{1}{b}\right),\ \sigma^2 = a^2\Gamma\left(1+\frac{2}{b}\right) - \mu^2$	$\left[0, \infty\right)$	$\left[0, \infty\right)$
Gamma	$\mu = ab,\ \sigma^2 = ab^2$	$\left[0, \infty\right)$	$\left[0, \infty\right)$
Lognormal	$\mu = e^{a+b^2/2},\ \sigma^2 = \left(e^{b^2} - 1\right)e^{2a+b^2}$	$\left[0, \infty\right)$	$\left[0, \infty\right)$
Gumbel	$\mu = a + 0.5772b,\ \sigma^2 = b^2\pi^2/6$	$\left(-\infty, \infty\right)$	$\left[0, \infty\right)$
Extreme	$\mu = a - 0.5772b,\ \sigma^2 = b^2\pi^2/6$	$\left[0, \infty\right)$	$\left[0, \infty\right)$
Extreme Type-II	$\mu = b\Gamma\left(1-\frac{1}{a}\right),\ \sigma^2 = b^2\left[\Gamma\left(1-\frac{2}{b}\right) - \Gamma^2\left(1-\frac{1}{b}\right)\right]$	$\left(-\infty, \infty\right)$	$\left[0, \infty\right)$

Introducing the mean or standard deviation as the nuisance variable, Eq. (3.4) can be rewritten as

$$\begin{aligned}\Pr\left(h_k \middle| D, I\right) &= \int_{-\infty}^{\infty} \Pr\left(h_k, \mu, \sigma \middle| D, I\right) d\mu \\ &= \int_{-\infty}^{\infty} \frac{\Pr\left(D \middle| h_k, \mu, \sigma, I\right) \Pr\left(h_k \middle| \mu, I\right) \Pr\left(\mu \middle| I\right)}{\Pr\left(D \middle| I\right)} d\mu\end{aligned} \tag{3.6}$$

or

$$\begin{aligned}\Pr\left(h_k \middle| D, I\right) &= \int_{0}^{\infty} \Pr\left(h_k, \mu, \sigma \middle| D, I\right) d\sigma \\ &= \int_{0}^{\infty} \frac{\Pr\left(D \middle| h_k, \mu, \sigma, I\right) \Pr\left(h_k \middle| \sigma, I\right) \Pr\left(\sigma \middle| I\right)}{\Pr\left(D \middle| I\right)} d\sigma\end{aligned} \tag{3.7}$$

In Eq. (3.6), $\Pr\left(h_k, \mu, \sigma \middle| D, I\right)$ is a function of mean with the standard deviation calculated from samples. Conversely, in Eq. (3.7), $\Pr\left(h_k, \mu, \sigma \middle| D, I\right)$ is a function of standard deviation with the mean calculated from samples. For Eqs. (3.6) and (3.7), Eq. (3.5) is used as the likelihood function. The candidate marginal CDFs in the set s_q are Gaussian, Weibull, Gamma, Lognormal, Gumbel, Extreme, and Extreme Type II in this study. The formulas of candidate marginal PDFs are shown in Table 3.2 with domains Ω^a and Ω^b of parameters *a* and *b*, respectively.

Second, consider the prior on the marginal CDF, $\Pr\left(h_k \middle| I\right)$, in Eq. (3.4). Let the additional information I be as follows:

I_1: Mean or standard deviation belongs to the set Λ^μ or Λ^σ, respectively, and each estimated $\mu \in \Lambda^\mu$ or $\sigma \in \Lambda^\sigma$ is equally likely;

I_2: For given μ or σ, all marginal CDFs satisfying $\mu \in \Omega_k^\mu$ or $\sigma \in \Omega_k^\sigma$ are equally probable, where Ω_k^μ and Ω_k^σ are domains of μ and σ for the marginal CDF M_k.

Table 3.2. Marginal PDFs and Domains of Marginal Parameters

Distribution	$f\left(x \middle\| a,b\right)$	$a \in \Omega^a$	$b \in \Omega^b$
Gaussian	$\frac{1}{b\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right]$	$(-\infty,\infty)$	$[0,\infty)$
Weibull	$\frac{b}{a}\left(\frac{x}{a}\right)^{b-1}\exp\left[-\left(\frac{x}{a}\right)^b\right]$	$[0,\infty)$	$[0,\infty)$
Gamma	$x^{a-1}\frac{\exp[-x/b]}{\Gamma(a)b^a}$	$[0,\infty)$	$[0,\infty)$
Lognormal	$\frac{1}{bx\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln x-a}{b}\right)^2\right]$	$(-\infty,\infty)$	$[0,\infty)$
Gumbel	$\frac{1}{b}\exp\left[-\frac{x-a}{b}-\exp\left[-\frac{x-a}{b}\right]\right]$	$(-\infty,\infty)$	$[0,\infty)$
Extreme	$\frac{1}{b}\exp\left[\frac{x-a}{b}-\exp\left[\frac{x-a}{b}\right]\right]$	$(-\infty,\infty)$	$[0,\infty)$
Extreme Type-II	$\frac{a}{b}\left[\frac{b}{x}\right]^{a+1}\exp\left[-\left(\frac{b}{x}\right)^a\right]$	$(-\infty,\infty)$	$[0,\infty)$

The set Λ^μ or Λ^σ provides information on the interval of mean or standard deviation, respectively, that the user might know. For example, if the user knows the specific domain of Λ^μ or Λ^σ, the domain can be used to integrate the likelihood function for calculation of the weights of each candidate marginal CDF. However, if information on specific domain of Λ^μ or Λ^σ is not known, it can be assumed that $\Lambda^\mu = (-\infty,\infty)$ or $\Lambda^\sigma = [0,\infty)$. In that case, the infinite domain cannot practically be used to integrate the likelihood function, and thus a finite range of Λ^μ or Λ^σ needs to be determined from samples such that Λ^μ or Λ^σ cover the wide range of the parameter. Using the first additional information I_1, the prior on μ and σ can be defined as

$$\Pr\left(\mu \middle| I_1\right) = \begin{cases} \dfrac{1}{\lambda\left(\Lambda^{\mu}\right)}, & \mu \in \Lambda^{\mu} \\ 0, & \mu \notin \Lambda^{\mu} \end{cases} \tag{3.8}$$

or

$$\Pr\left(\sigma \middle| I_1\right) = \begin{cases} \dfrac{1}{\lambda\left(\Lambda^{\sigma}\right)}, & \sigma \in \Lambda^{\sigma} \\ 0, & \sigma \notin \Lambda^{\sigma} \end{cases} \tag{3.9}$$

where $\lambda(\cdot)$ is the Lebesgue measure, which is the interval length of Λ^{μ} or Λ^{σ}. Likewise, since all marginal CDFs are equally probable for $\mu \in \Omega_k^{\mu}$ or $\sigma \in \Omega_k^{\sigma}$, the prior on the marginal CDF is defined as

$$\Pr\left(h_k \middle| \mu, \sigma, I_2\right) = \begin{cases} 1, & \mu \in \Omega_k^{\mu} \\ 0, & \mu \notin \Omega_k^{\mu} \end{cases} \tag{3.10}$$

or

$$\Pr\left(h_k \middle| \mu, \sigma, I_2\right) = \begin{cases} 1, & \sigma \in \Omega_k^{\sigma} \\ 0, & \sigma \notin \Omega_k^{\sigma} \end{cases} \tag{3.11}$$

In this study, it is assumed that the prior follows a uniform distribution, which means there is no information on the distribution of μ or σ. If it is known that the prior of μ or σ follows a specific distribution, $\Pr(\mu)$ or $\Pr(\sigma)$ might be expressed as a PDF and can be used as the prior instead of Eq. (3.8) or (3.9). However, since the prior of μ or σ is usually unknown and the effect of the prior is negligible when the number of samples is enough (larger than 100 samples), Eq. (3.8) or (3.9) are used in most cases.

Finally, consider normalization constant, $\Pr\left(D \middle| I\right)$ in Eq. (3.4). Substituting Eqs. (3.5) and (3.8) – (3.11) into Eqs. (3.6) and (3.7), Eqs. (3.6) and (3.7) can be rewritten as

$$\Pr\left(h_k \middle| D, I\right) = \frac{1}{\Pr\left(D \middle| I\right)} \frac{1}{\lambda\left(\Lambda^{\mu}\right)} \int_{\Omega_k^{\mu} \cap \Lambda^{\mu}} \prod_{i=1}^{ns} f_k\left(x_i \middle| a(\mu,\sigma), b(\mu,\sigma)\right) d\mu \tag{3.12}$$

or

$$\Pr\left(h_k \middle| D, I\right) = \frac{1}{\Pr\left(D \middle| I\right)} \frac{1}{\lambda\left(\Lambda^{\sigma}\right)} \int_{\Omega_k^{\sigma} \cap \Lambda^{\sigma}} \prod_{i=1}^{ns} f_k\left(x_i \middle| a(\mu,\sigma), b(\mu,\sigma)\right) d\sigma \tag{3.13}$$

In Eqs. (3.12) and (3.13), $\Pr\left(D \middle| I\right)$ can be expressed as

$$\Pr\left(D \middle| I\right) = \sum_{k=1}^{q} \Pr\left(D \middle| h_k, I\right) \Pr\left(h_k \middle| I\right) \tag{3.14}$$

Since $\Pr\left(D \middle| I\right)$ is a constant, for convenience, it is not included to calculate weights. Accordingly, Eqs. (3.12) and (3.13) can be expressed as

$$W_k = \frac{1}{\lambda\left(\Lambda^{\mu}\right)} \int_{\Omega_k^{\mu} \cap \Lambda^{\mu}} \prod_{i=1}^{ns} f_k\left(x_i \middle| a(\mu,\sigma), b(\mu,\sigma)\right) d\mu \tag{3.15}$$

or

$$W_k = \frac{1}{\lambda\left(\Lambda^{\sigma}\right)} \int_{\Omega_k^{\sigma} \cap \Lambda^{\sigma}} \prod_{i=1}^{ns} f_k\left(x_i \middle| a(\mu,\sigma), b(\mu,\sigma)\right) d\sigma \tag{3.16}$$

The normalized weight for the marginal CDF M_k is calculated as

$$w_k = \frac{W_k}{\sum_{i=1}^{q} W_i} \tag{3.17}$$

To numerically calculate the integrals in Eq. (3.15) and (3.16), the MATLAB function “quad”, which is a recursive adaptive Simpson quadrature (Gander and Gautschi, 2000), is used. The recursive adaptive Simpson quadrature integrates a function over an interval, say $[a,b]$, using a simple quadrature rule; and then calculates the error between the integral calculated over $[a,b]$ and the sum of two integrals calculated over

subdomains $\left[a,(a+b)/2\right]$ and $\left[(a+b)/2,b\right]$. If the error is less than the user-specified tolerance, then the integral calculated over $\left[a,b\right]$ is accepted. If it is not, the integral is recalculated by subdividing the interval of integration in two and applying the adaptive Simpson's quadrature to each subinterval in a recursive way. The tolerance that is used in this study is 1.0E-6.

In the Bayesian method for identifying the marginal CDF, there are two approaches that use mean or standard deviation as variables for calculating normalized weights, and thus it might be necessary to select one method. To compare the performances of the two approaches, averaged normalized weights over 100 trials are used. Let "Mean" and "Std." be the methods using mean and standard deviation, respectively, as variables for the calculation of weights. Table 3.3 shows the averaged normalized weights for different means and for different samples when Gaussian, Weibull, Gamma, Lognormal, Gumbel, Extreme, and Extreme Type II are the original CDFs. The larger the normalized weights, the better identified each original CDF is. When Gaussian, Weibull, Gamma, Lognormal, and Gumbel are the original CDFs, the normalized weights using "Mean" are slightly better than those using "Std." On the other hand, when Extreme and Extreme Type II are the original CDFs, the normalized weights using "Std." are slightly better than those using "Mean." For all cases, since one method is not always better than the other and the normalized weights calculated from two approaches are similar, both approaches can be used. However, "Mean" is better than "Std." in more cases, so "Mean" is used in this study.

Table 3.3. Averaged Weights over 100 Trials Using Two Approaches

Original Distribution	ns	$\mu = 2$		$\mu = 10$	
		Mean	Std.	Mean	Std.
Gaussian	30	0.300	0.273	0.264	0.232
	100	0.462	0.454	0.347	0.308
	300	0.731	0.702	0.427	0.389
Weibull	30	0.328	0.261	0.300	0.310
	100	0.544	0.484	0.495	0.477
	300	0.706	0.705	0.636	0.573
Gamma	30	0.240	0.213	0.246	0.200
	100	0.410	0.322	0.339	0.301
	300	0.750	0.617	0.347	0.337
Lognormal	30	0.240	0.233	0.253	0.200
	100	0.362	0.362	0.332	0.291
	300	0.597	0.558	0.399	0.383
Gumbel	30	0.230	0.224	0.308	0.327
	100	0.395	0.388	0.465	0.440
	300	0.636	0.590	0.651	0.616
Extreme	30	0.552	0.602	0.374	0.403
	100	0.885	0.927	0.497	0.527
	300	0.993	1.000	0.584	0.565
Extreme Type-II	30	0.481	0.586	0.338	0.392
	100	0.778	0.787	0.465	0.516
	300	0.875	0.916	0.605	0.669

3.2.3 Comparison of Two Methods

It is stated that the Bayesian method performs better in identifying the correct marginal CDF than the GOF test, but it is still valuable to compare the two methods numerically. Consider a set of random data with a different number of samples: ns=30, 100, and 300, given that the original distribution is Gamma with $GM(400, 0.025)$. The Gaussian, Weibull, Gumbel, Lognormal, Extreme, and Extreme Type II are selected as candidate distributions. All parameters of each CDF are calculated from μ=10.0 and

σ=0.5 using Table 3.1. The PDF shapes of the candidate distributions are shown in Figure 3.1.

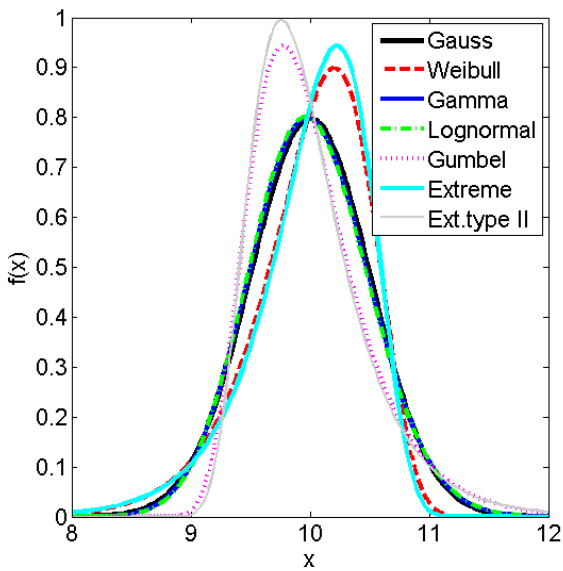


Figure 3.1. Marginal PDFs for μ=10.0 and σ=0.5

In the GOF test, p-values are calculated and used to test the null hypothesis of each candidate CDF. In the Bayesian method, normalized weights are calculated to identify a correct marginal CDF. To compare the two methods, the p-value and normalized weights should be comparable, but they are not. The p-value is the probability of obtaining a value of the test statistic (D_n) at least as extreme as the actually observed value, given that the null hypothesis is true. On the other hand, it is not the probability of the hypothesis (Sterne and Smith, 2000) from which the normalized weights originate. Thus, instead of directly using two values, it might be better to observe how often p-

values and normalized weights are assigned to correct marginal CDF or CDFs with shapes similar to the correct one.

Table 3.4 shows the sum of p-values and normalized weights over 100 trials when both methods are used. For a small number of samples, such as 30, the Bayesian method assigns most normalized weights to Gaussian, Gamma, and Lognormal distributions because the PDF shapes of Gaussian and Lognormal are very close to Gamma for the given mean μ=10, as shown in Figure 3.1. On the other hand, the GOF test assigns p-values to all candidate distributions almost evenly, which means the GOF test does not identify correct CDFs for a small number of samples. Further, since the p-values of all distributions are larger than 5 for a given significance level of 5%, the GOF test accepts all distributions as correct even though the PDF shapes of Weibull, Gumbel, Extreme, and Extreme Type II are quite different from Gamma (original), as shown in Figure 3.1.

Table 3.4. Comparison of GOF with Bayesian Method for Identification of Marginal Gamma CDF

Distribution	*ns* =30		*ns* =100		*ns* =300	
	GOF	Bay.	GOF	Bay.	GOF	Bay.
Gaussian	16.3	24.1	21.9	32.9	29.3	31.0
Weibull	13.1	8.59	10.4	1.12	3.29	0.00
Gamma	16.4	24.6	22.0	33.0	30.8	34.2
Lognormal	16.5	24.7	21.9	32.4	30.7	34.8
Gumbel	13.4	7.46	9.78	0.42	3.01	0.00
Extreme	12.2	6.08	8.01	0.20	1.56	0.00
Extreme type II	12.0	4.45	5.97	0.00	1.39	0.00

For a large number of samples, such as 300, the GOF test identifies Gaussian, Gamma, and Lognormal CDFs. Still, the Bayesian method shows better performance than

the GOF tests in this example. It might be of some concern that the normalized weight of the original distribution is still not high, even for a large number of samples. However, the PDF shapes are very close to each other even at the tail end, which lead to similar optimum design in RBDO. Thus, it does not matter which of the three distributions is selected.

3.3 Identification of Joint CDFs (Copulas)

Just as two methods–the GOF test and the Bayesian method–are used to identify marginal CDFs, two methods can be used to identify the correct copula. Since the GOF test relies on the correlation coefficient calculated from samples, the identification of the copula might not be correct due to the incorrectly calculated correlation coefficient. On the other hand, the Bayesian method calculates weights to identify the copula by integrating the likelihood function over the correlation coefficient, so that it is independent of the estimation of the parameter. Thus, the Bayesian method is preferred over the GOF test.

3.3.1 Goodness-of-fit Test

The GOF test compares the empirical copula C_{ns} calculated from given ns samples and a theoretical copula $C_{\theta_{ns}}$ with some parameter θ_{ns} calculated from the data. The GOF test for the identification of copula can be carried out by using the parametric bootstrap (Genest and Rémillard, 2005). Using the Cramér-von Mises statistic S_{ns}, the null hypothesis H_0 (that is, the data follow a specific copula type C_k) can be tested by

$$S_{ns} = \sum_{i=1}^{ns} \left\{ C_{ns}\left(\hat{U}_i\right) - C_{\theta_{ns}}\left(\hat{U}_i\right) \right\}^2 \tag{3.18}$$

where $\hat{U}_i = R_i / (ns+1)$ for $i = 1, \cdots, ns$ is the empirical marginal CDF, and $R_i = [R_{i1}, \cdots, R_{in}]^T$ is the rank of the samples for $X_1, \cdots, X_n$. The empirical copula is defined as

$$C_{ns}(u) = \frac{1}{ns} \sum_{i=1}^{ns} 1\left(\hat{U}_i \le u\right), \qquad u \in [0,1]^n \tag{3.19}$$

where $1\left(\hat{U}_i \le u\right)$ indicates "1" if $\hat{U}_i \le u$; otherwise, "0". After obtaining the Cramer-von Mises statistic S_{ns} from the samples, the empirical copula $C^*_{ns,k}$ and Cramér-von Mises statistic $S^*_{ns,k}$ need to be repeatedly estimated for every $k = 1, \cdots, N$ where N is a large number (e.g., 1,000 ~10,000).

As a result, an approximate p-value for the test based on the Cramer-von Mises statistic S_{ns} is given by

$$p = \frac{1}{N} \sum_{k=1}^{ns} 1\left(S^*_{ns,k} > S_{ns}\right) \tag{3.20}$$

where $1\left(S^*_{ns,k} > S_{ns}\right)$ indicates "1" if $S^*_{ns,k} > S_{ns}$; otherwise, "0". For some significance level, such as 5%, according to the p-value, the null hypothesis H_0 might be accepted or rejected. When the number of random variables is larger than three, a two-level parametric boot strap similar to the previously described one-level boot strap needs to be used. The detailed algorithm was implemented by Genest and Rémillard (2005).

In recent years, various other GOF tests have been developed. For Archimedean copulas, Genest and Favre (2007) and Genest and Rémillard (2005) proposed a GOF test

statistic that compares the one-dimensional empirical function K_{ns} and the theoretical function $K_{\theta_{ns}}$ as

$$\mathrm{K}(t) = \sqrt{ns}\left\{K_{ns}(t) - K_{\theta_{ns}}(t)\right\} \tag{3.21}$$

where $K_{\theta_{ns}}(t) = \Pr\left(C\left(u,v\middle|\theta_{ns}\right) < t\right)$ and $K_{ns}(t) = \frac{1}{ns}\sum_{k=1}^{ns} 1\left(V_{k,ns} \le t\right)$ with $V_{k,ns} = \frac{1}{ns}\sum_{j=1}^{ns} 1\left(R_{1j} \le R_{1k}, R_{2j} \le R_{2k}\right)$. As in the previous parametric boot strap, using another type of the statistic K, the p-value can be calculated.

The GOF tests suggest a way to identify the correct copula, but they depend on the estimation of the correlation parameter θ_{ns}. Further, for the second method, since the function K is explicitly expressed only for Archimedean copulas, it cannot be used for non-Archimedean copulas because the function K cannot be algebraically formulated for non-Archimedean copulas (Genest and Rémillard, 2005).

3.3.2 Bayesian Method

As the Bayesian method is used to identify the correct marginal CDFs, it can also be used to identify the correct copula by calculating the weights of the candidate copulas to select the one with the highest weight. To calculate the normalized weights, Kendall's tau is used as the variable for integrating the likelihood function.

Let a finite set $S_Q \subset S$ consist of candidate copulas C_k, $k = 1, \cdots, Q$ where S is a set of all copulas and Q is the number of the candidate copulas. The Bayesian method consists of defining Q hypotheses (Huard, 2006):

H_k : The data come from copula C_k, $k = 1, \cdots, Q$.

The probability of each hypothesis H_k given the data D is defined as

$$\Pr\left(H_k \middle| D, I\right) = \frac{\Pr\left(D \middle| H_k, I\right)\Pr\left(H_k \middle| I\right)}{\Pr\left(D \middle| I\right)} \tag{3.22}$$

where $\Pr\left(D \middle| H_k, I\right)$ is the likelihood function, $\Pr\left(H_k \middle| I\right)$ is the prior on the copula family, and $\Pr\left(D \middle| I\right)$ is the normalization constant with any relevant additional knowledge I .

Under the hypothesis H_k (that the data D come from the copula C_k), the probability of drawing the data D for the hypothesis on C_k is expressed as a likelihood function as

$$\Pr\left(D \middle| H_k, \tau, I\right) = \prod_{i=1}^{ns} c_k\left(u_i, v_i \middle| \theta\right) \tag{3.23}$$

where $\left(u_i, v_i\right)$ are ns mutually independent pairs of data and calculated as $u_i = F_X\left(x_i\right)$ and $v_i = F_Y\left(y_i\right)$, where $F_X\left(x_i\right)$ and $F_Y\left(y_i\right)$ are the marginal CDF values obtained from the given paired data $\left(x_i, y_i\right)$. Since it is assumed that the data D come from the copula C_k, the probability of drawing D from the copula C_k (likelihood function) is expressed as a copula density function c_k. The paired data are independent of each other, so the likelihood function is expressed as multiplications of the copula density function values evaluated at all the data. Since each copula C_k has its own correlation parameter θ, a common correlation measure, Kendall's tau, needs to be used. Using the relationship

between the parameter and Kendall's tau $\tau = g_k(\theta)$, as shown in Table 2.2, the correlation parameter can be expressed as $\theta = g_k^{-1}(\tau)$.

Using the Kendall's tau as the nuisance variable, Eq. (3.22) can be rewritten as (Huard, 2006)

$$\Pr(H_k|D,I) = \int_{-1}^{1} \Pr(H_k,\tau|D,I)\,d\tau = \int_{-1}^{1} \frac{\Pr(D|H_k,\tau,I)\Pr(H_k|\tau,I)\Pr(\tau|I)}{\Pr(D|I)}\,d\tau \quad (3.24)$$

where Eq. (3.23) is used as the likelihood function.

Let the additional information I on the copula be as follows:

I_1: Kendall's tau belongs to the set Λ^τ, and each estimated $\tau \in \Lambda^\tau$ is equally likely;

I_2: For a given τ, all copula families satisfying $\tau \in \Omega_k^\tau$ are equally probable, where Ω_k^τ is the domain of τ for C_k.

The set Λ^τ provides information on the interval of Kendall's tau that the user might know. For example, based on the user's experience, it might be known that the range of Kendall's tau estimated between two interesting variables can have only a positive range $\Lambda^\tau = [0,1]$. However, if information on the correlation parameter between variables is not known, it can be assumed as $\Lambda^\tau = [-1,1]$. Using the first additional information I_1, the prior on τ can be defined as (Huard, 2006)

$$\Pr(\tau|I_1) = \begin{cases} \dfrac{1}{\lambda(\Lambda^\tau)}, & \tau \in \Lambda^\tau \\ 0, & \tau \in \Lambda^\tau \end{cases} \quad (3.25)$$

where $\lambda(\cdot)$ is the Lebesgue measure, which is the interval length of Λ^{τ}. Likewise, since all copula families are equally probable for $\tau \in \Omega_k^{\tau}$, the prior on the copula family is defined as

$$\Pr\left(H_k \middle| \tau, I_2\right) = \begin{cases} 1, & \tau \in \Omega_k^{\tau} \\ 0, & \tau \notin \Omega_k^{\tau} \end{cases} \tag{3.26}$$

If it is known that the prior distribution of τ follows a certain distribution, $\Pr(\tau)$ might be expressed as a specific distribution and can be used as the prior instead of Eq. (3.25). However, the prior of τ is usually unknown, so Eq. (3.25) is commonly used in practical applications.

Substituting Eqs. (3.23), (3.25), and (3.26) into Eq. (3.24), Eq. (3.24) can be rewritten as

$$\Pr\left(H_k \middle| D, I\right) = \frac{1}{\Pr\left(D \middle| I\right)} \frac{1}{\lambda\left(\Lambda^{\tau}\right)} \int_{\Omega_k \cap \Lambda^{\tau}} \prod_{i=1}^{ns} c_k\left(u_i, v_i \middle| g_k^{-1}(\tau)\right) d\tau \tag{3.27}$$

In Eq. (3.27), $\Pr\left(D \middle| I\right)$ is expressed as (Huard, 2006)

$$\Pr\left(D \middle| I\right) = \sum_{l=1}^{Q} \Pr\left(D \middle| H_k, I\right) \Pr\left(H_k \middle| I\right) \tag{3.28}$$

However, the problem is that Eq. (3.28) is only valid when the data come from the set of the candidate copulas, and the candidate copulas do not have similar shapes. A procedure to calculate $\Pr\left(D \middle| I\right)$ is explained by Huard (2006), although the procedure is tedious. In this study, since $\Pr\left(D \middle| I\right)$ is a constant, it is not included for convenience. Accordingly, the computation of Eq. (3.27) can be expressed as the computation of the weights as

$$W_k = \frac{1}{\lambda\left(\Lambda^{\tau}\right)} \int_{\Omega_k \cap \Lambda^{\tau}} \prod_{i=1}^{ns} c_k\left(u_i, v_i \middle| g_k^{-1}(\tau)\right) d\tau \tag{3.29}$$

where the "quad" function in MATLAB, which is the recursive adaptive Simpson quadrature, is used to integrate the likelihood function.

The normalized weight of C_k is calculated as

$$w_k = \frac{W_k}{\sum_{i=1}^{Q} W_k} \tag{3.30}$$

Since the Bayesian method selects one marginal CDF or copula that best describes the given data among candidates, the identified marginal CDF, or copula might not be the correct one. On the other hand, since some commonly used marginal CDFs are known and the total number of marginal CDF types is not large, it is easy to determine the candidate marginal CDFs and identify a correct CDF among them. However, since there exist plenty of copula types, it might be possible that the data come from an unknown copula that is not among the given candidate copulas. A way of solving this problem is presented by Bretthorst (1996).

3.3.3 Comparison of Two Methods

Just as the GOF test and Bayesian method were compared for identification of marginal CDFs, two methods for identification of copulas are compared in this section. Consider a set of random data with a different number of samples: ns=30, 100, and 300 for the original copula as Clayton with τ=0.4 where Clayton, Gumbel, Gaussian, Frank, A12, and A14 are selected as candidate copulas.

Table 3.5 shows the sum of p-values and normalized weights over 100 trials. As shown in the table, for a small number of samples, the Bayesian method assigns 39.7 to the Clayton copula (original copula) whereas the GOF test assigns only 17.5 to the

correct copula. Further, the Bayesian method assigns the normalized weights according to the similarity of the copula shape, i.e., A12 is the most similar with Clayton, A14 is the second, and so on. On the other hand, the GOF test accepts all candidate copulas as correct copulas on the average (p-values are larger than 5), even though some copulas such as Gumbel have very distinct shapes with the Clayton copula. As the number of samples is increased up to 300, the performance of the GOF test improves, but the Bayesian method is still better than the GOF test in identifying copula.

Table 3.5. Comparison of GOF with Bayesian Method for Identification of Copula

Copula	*ns* =30		*ns* =100		*ns* =300	
	GOF	Bay.	GOF	Bay.	GOF	Bay.
Clayton	17.5	39.7	34.8	53.4	41.7	68.0
Gumbel	13.4	5.55	0.39	0.00	0.00	0.00
Gaussian	16.3	9.95	0.07	2.13	0.00	0.00
Frank	16.2	9.81	2.58	2.45	0.00	0.00
A12	18.2	20.1	31.0	25.9	34.9	21.6
A14	18.3	14.9	30.5	16.0	23.3	10.4

3.4 Effectiveness of Bayesian Method

In this section, the effectiveness of the Bayesian method in identifying marginal CDFs and copula is studied for different numbers of samples and parameters.

3.4.1 Identification of Marginal CDFs

Given Gaussian, Weibull, Gamma, Lognormal, Gumbel, Extreme, and Extreme type II CDFs, the Bayesian method is tested to identify original CDFs using normalized weights over 100 trials where the number of samples is 30, 100, and 300. Figure 3.2

shows the sum of normalized weights over 100 trials for different numbers of samples when original CDFs, indicated by boxes on the name of the marginal CDFs, are given with μ=2.0 and σ=0.5.

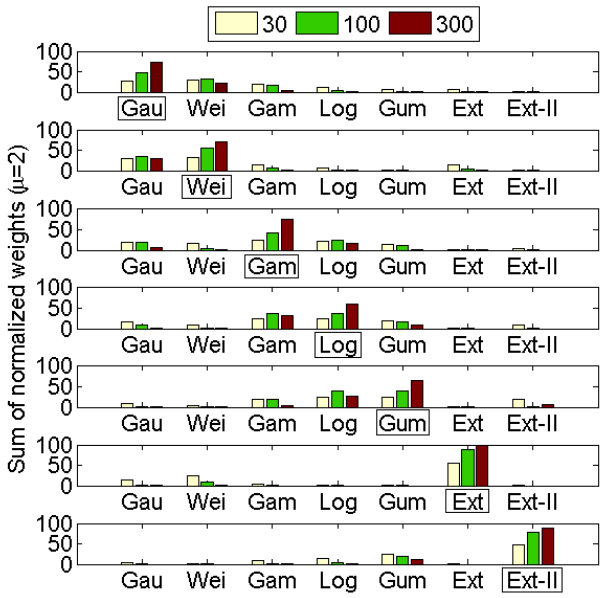


Figure 3.2. Sum of Normalized Weights over 100 Trials for μ=2.0 and σ=0.5

For example, when the Gaussian CDF is original, samples (*ns*=30, 100, and 300) are randomly generated from the Gaussian CDF 100 times. Using the Bayesian method, the normalized weights are calculated from the 100 data sets. Adding up the normalized weights over 100 trials, the sum of normalized weights is approximately 30 for Gaussian, 25 for Weibull, and 45 for the rest of the CDFs when *ns*=30. Since the PDF shapes of Gaussian and Weibull CDFs are similar, as shown in Figure 3.3, the Weibull distribution has the second highest sum of normalized weight among candidate marginal CDFs.

Likewise, when Weibull is the original CDF, the Gaussian distribution has the second highest sum of normalized weight. On the other hand, when the CDFs with distinct PDF shapes such as Extreme and Extreme Type II are original, it is much easier to identify the original CDF.

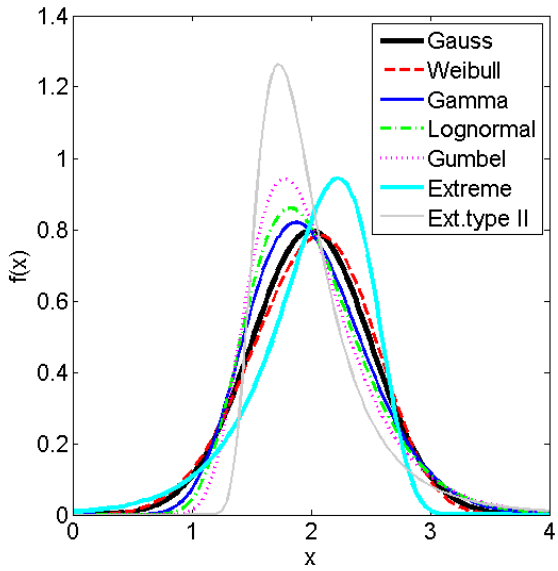


Figure 3.3. Marginal PDFs for μ=2.0 and σ=0.5

When μ=2.0 is changed to 10 with the same σ=0.5, the PDF shapes become different from Figure 3.3. Figure 3.4 shows the sum of normalized weights over 100 trials when μ=10.0 and σ=0.5. In this case, the Gaussian, Gamma, and Lognormal CDFs are very close to each other even at the tail end, as shown in Figure 3.1. Thus, when those three CDFs are original, it is difficult to identify the correct CDF among three CDFs, and the normalized weights of Gaussian, Gamma, and Lognormal are rather similar to each other even if the number of samples is increased to 300. Even though the original CDF

may not be identified, since similar PDF shapes provide similar RBDO results, it does not matter which of the three CDFs is selected. For the CDFs with distinct PDF shapes, such as Weibull, Gumbel, Extreme, and Extreme Type II, it is much easier to identify the correct CDF even for larger mean μ=10.0.

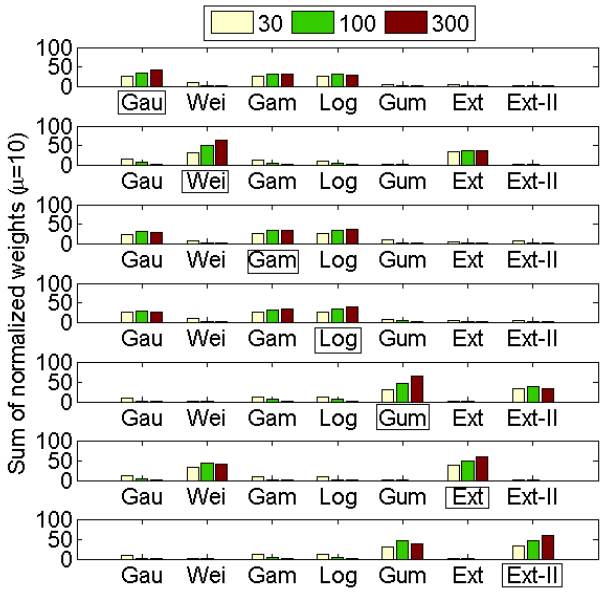


Figure 3.4. Sum of Normalized Weights over 100 Trials for μ=10.0 and σ=0.5

On the other hand, since the identified distribution generally fits to the given data, its tail behavior might be different from the one of the original CDF particularly when the number of samples is small. If the tail behavior of the identified distribution is quite different from the one of the original distribution, different tail behaviors yield different β_t–contours with a target reliability index β_t, which lead to different RBDO results. In practical applications, since experimental data is very limited, it could be hard to select a

distribution that has the same tail as the original distribution. For this, RBDO with a confidence level of the input model uncertainty is currently being investigated so that reliable optimum designs can be obtained.

3.4.2 Identification of Joint CDFs

In this study, different values of Kendall's tau, i.e., τ =0.2, and 0.7, are used to study the effect of Kendall's tau on identification of the correct copulas. The candidate copulas are selected as Clayton, AMH, Gumbel, Frank, A12, A14, FGM, Gaussian, and an independent copula, which is expressed as the multiplication of marginal CDFs, $C(u,v) = uv$. For different Kendall's tau, τ =0.2, and 0.7, the sum of normalized weights over 100 trials are calculated when each original copula is given.

For small correlation coefficients such as τ =0.2, since the PDF contours of most copulas are similar to each other except Clayton and Gumbel, as shown in Figure 3.5, it is not simple to identify the correct one. For instance, even though the original copula is AMH, the normalized weights of incorrect copulas such as the independent copula are high, especially for a small number of samples, ns =30 as shown in Figure 3.6. Therefore, a large number of samples are generally required to identify the correct copula when Kendall's tau is small.

When the correlation between two variables is more significant, for example, $\tau = 0.7$, it is easier to identify the correct copula because the copula shapes are quite distinct from each other, as shown in Figure 3.7. Accordingly, the correct copula can be easily identified with the highest normalized weight seen in Figure 3.8.

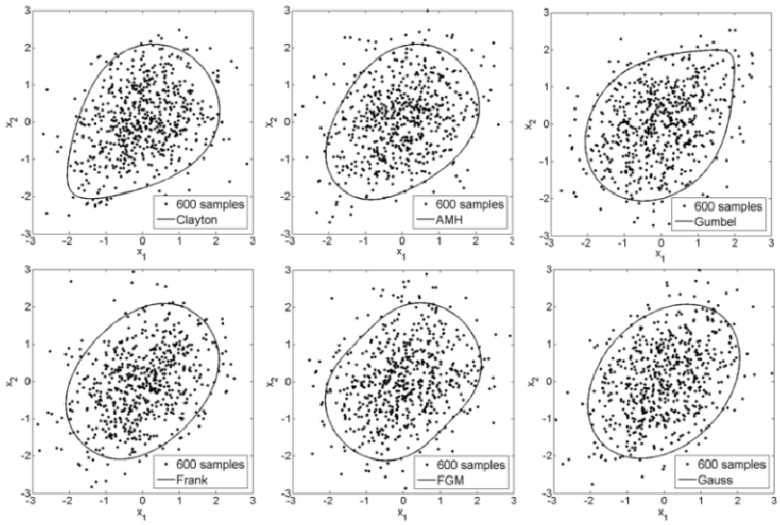


Figure 3.5. PDF Contours and Scatter Plots of Samples (ns =600) for $\tau = 0.2$

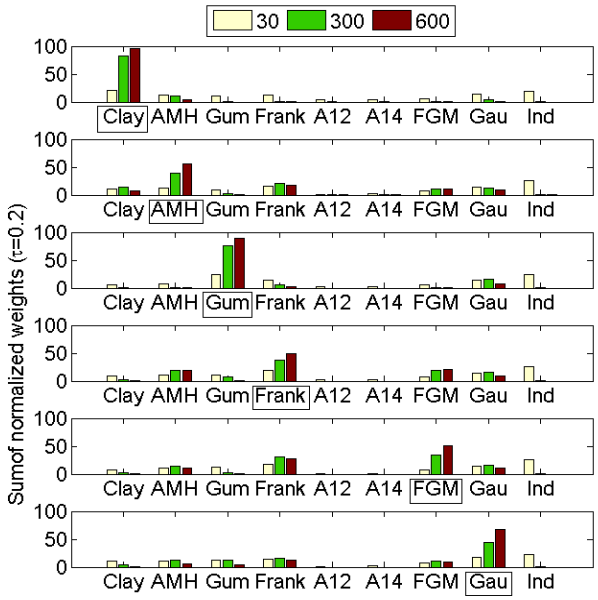


Figure 3.6. Sum of Normalized Weights over 100 Trials for $\tau = 0.2$

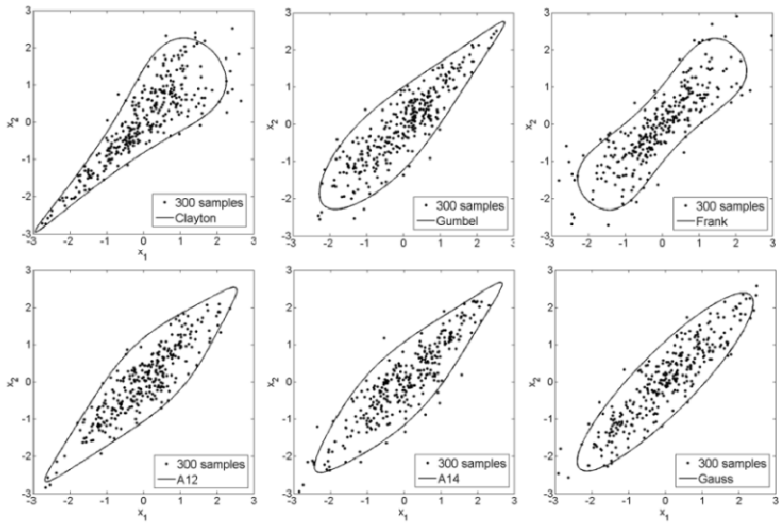


Figure 3.7. PDF Contours and Scatter Plots of Samples (ns=300) for $\tau = 0.7$

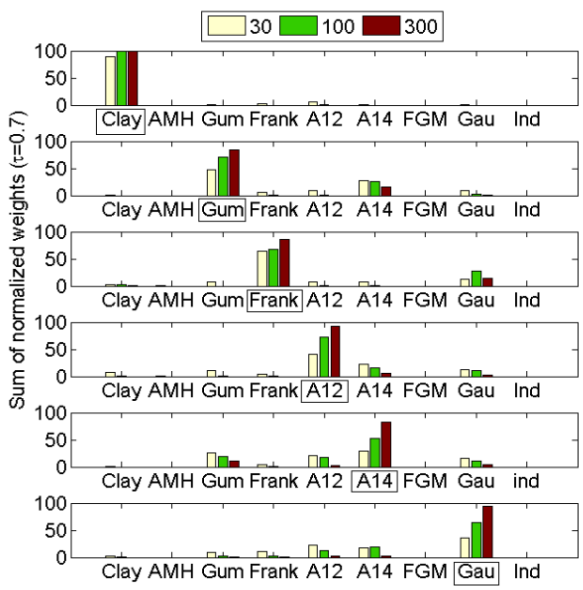


Figure 3.8. Sum of Normalized Weights over 100 Trials for $\tau = 0.7$

In RBDO problems, if the shape of the selected copula is significantly different from the true copula - especially for high correlation - a wrong RBDO result will be

obtained. For instance, if the Clayton copula is the original copula, and other copulas such as Frank or Gaussian are wrongly selected, the RBDO result will be very different from the true RBDO result (Noh et al., 2007). However, if the correlation between input variables is not high, copulas such as AMH, Frank, FGM, and Gaussian, which have similar copula shapes, might provide quite similar RBDO results. Thus, 300 samples could be acceptable to identify the right copula even for small correlation coefficients. For large correlation coefficients, 100, or even 30 samples could be enough to identify correctly for copulas such as Clayton, Gumbel, and Frank.

Once the input models are estimated, the RBDO can be carried, which will be explained in Chapter 4. However, when the input variables are correlated with non-elliptical copulas, different transformation orderings might yield different RBDO results. For this, the more accurate reliability analysis method, MPP-based DRM, is used in Chapter 5. In the Chapter 6, combining all proposed methods, several numerical examples are tested to show how the RBDO using the MPP-based DRM is carried out based on the input model estimated from the Bayesian method.

CHAPTER IV

FUNDAMENTAL CONCEPTS IN RELIBABILITY-BASED DESIGN OPTIMIZATION

4.1 Introduction

This chapter explains fundamental concepts in reliability-based design optimization (RBDO). Sections 4.2 and 4.3 explain the basic ideas of reliability analysis and inverse reliability analysis. For reliability analysis and inverse reliability analysis, the first-order reliability method (FORM) and the second-order reliability method (SORM) have been used most frequently. However, the FORM is not accurate for approximating the highly nonlinear function, and while the SORM is more accurate than the FORM, it is expensive to use. Thus, Section 4.4 illustrates the most probable point (MPP)-based dimension reduction method (DRM), which yields efficient and accurate reliability analysis and design optimization. Section 4.5 explains the RBDO process, which uses inverse reliability analysis.

4.2 Reliability Analysis

A reliability analysis involves calculation of probability of failure. The probability of failure, P_F, is estimated by a multi-dimensional integral (Madsen et al., 1986) as

$$P_F \equiv P\left(G\left(\mathbf{X}\right) > 0\right) = \int_{G(\mathbf{X})>0} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \tag{4.1}$$

where $\mathbf{X}$ is an *n*-dimensional random vector and $G(\mathbf{X})$ is a constraint function where $G(\mathbf{X}) > 0$ is defined as failure. $\mathbf{x}$ is the realization of the random vector $\mathbf{X}$ and $f_{\mathbf{X}}(\mathbf{x})$ is a joint PDF of $\mathbf{X}$. However, it is difficult to compute the multi-dimensional integral analytically. Thus, it is necessary to use analytical approximations of this integral that are simpler to calculate. The FORM and SORM estimate the probability of failure by approximating the constraint function $G(\mathbf{X})$ by first-order Taylor series expansion and second-order Taylor series expansion, respectively. Since the FORM and SORM are carried out in standard Gaussian variables, a transformation of the random variables into standard Gaussian variables will be introduced in the next section.

4.2.1 Transformation

Given *n*-dimensional random vector $\mathbf{X}$ with a joint CDF, $F_{\mathbf{X}}(\mathbf{x})$, transformation from random variables $\mathbf{X}$ into independent standard Gaussian variables $\mathbf{U}$ is defined by the Rosenblatt transformation (Rosenblatt, 1952) as the following successive conditioning:

$$\begin{aligned} u_1 &= \Phi^{-1}\left[F_{X_1}(x_1)\right] \\ u_2 &= \Phi^{-1}\left[F_{X_2}(x_2|x_1)\right] \\ &\vdots \\ u_n &= \Phi^{-1}\left[F_{X_n}(x_n|x_1, x_2, \cdots, x_{n-1})\right] \end{aligned} \tag{4.2}$$

where *n* is the number of random variables and $F_{X_i}(x_i|x_1, x_2, ..., x_{i-1})$ is the conditional CDF of X_i conditional on $X_1 = x_1,\ X_2 = x_2, \cdots, X_{i-1} = x_{i-1}$, which is defined as

$$F_{X_i}\left(x_i \middle| x_1, x_2, \cdots, x_{i-1}\right) = \frac{\int_{-\infty}^{x_i} f_{X_1 X_2 \cdots X_i}(x_1, x_2, \cdots, x_{i-1}, \xi) d\xi}{f_{X_1 X_2 \cdots X_{i-1}}(x_1, x_2, \cdots, x_{i-1})} \tag{4.3}$$

The inverse transformation from the independent Gaussian variables $\mathbf{U}$ into the random variables $\mathbf{X}$ can be obtained from Eq. (4.2) as

$$\begin{aligned} x_1 &= F_{X_1}^{-1}\left[\Phi\left(u_1\right)\right] \\ x_2 &= F_{X_2}^{-1}\left[\Phi\left(u_2 \middle| x_1\right)\right] \\ &\vdots \\ x_n &= F_{X_n}^{-1}\left[\Phi\left(u_n \middle| x_1, x_2, \cdots, x_{n-1}\right)\right] \end{aligned} \tag{4.4}$$

Thus, once the joint CDF is identified using the Bayesian method explained in Section 3, the transformation can be carried out because the conditional CDF can be analytically obtained from the joint CDF.

For the joint CDF modeled by an elliptical copula, since the independent standard Gaussian variable $\mathbf{U}$ corresponds to the uniformly distributed variable $\mathbf{S}$ in Eq. (2.18), the transformation from the correlated standard elliptical variables into independent standard Gaussian variables is linear as

$$\mathbf{u} = \mathbf{L}^{-1}\mathbf{e} \tag{4.5}$$

where $\mathbf{e}$ represents correlated standard elliptical variables, defined as $e_i = \Psi^{-1}\left[F_{X_i}\left(x_i\right)\right]$ for $i = 1, \cdots, n$. $\Psi^{-1}\left(\cdot\right)$ is the inverse of the elliptical CDF. If the elliptical copula is Gaussian, $\mathbf{e}$ represents correlated Gaussian variables, which is the same as $\mathbf{X}$ in Eq. (2.18). If it is a *t* copula, $\mathbf{e}$ is a standard *t* variables $\mathbf{X}$ over a random variable R, defined as $\mathbf{X}/R$ in Eq. (2.18) where R^2 has an inverse gamma distribution $Ig\left(\frac{1}{2}\nu, \frac{1}{2}\nu\right)$. $\mathbf{L}^{-1}$ is the inverse of a lower triangular matrix $\mathbf{L}$ obtained from the Cholesky decomposition of

reduced covariance matrix $\mathbf{P}' = \left\{\rho_{ij}'\right\}$. That is, $\mathbf{P}' = \mathbf{L}\mathbf{L}^{\mathbf{T}}$ and each entry of the matrix $\mathbf{L}$ is obtained as

$$l_{ij} = \begin{cases} \sqrt{\left(1 - \sum_{k=1}^{i-1} l_{ik}^2\right)}, & i = j \\ \left(\rho_{ij}' - \sum_{k=1}^{j-1} l_{ik} l_{jk}\right) / l_{jj}, & i > j \end{cases} \tag{4.6}$$

Since the reduced correlation coefficient ρ_{ij}' between E_i and E_j is different from the correlation coefficient ρ_{ij} between X_i and X_j, it needs to be calculated from ρ_{ij}. The reduced correlation coefficient ρ_{ij}' is obtained from the correlation coefficient ρ_{ij} using the following equation:

$$\rho_{ij} = E\left[\Xi_i \Xi_j\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi_i \xi_j \psi\left(e_i, e_j; \rho_{ij}'\right) de_i de_j \tag{4.7}$$

where $\Xi_i = \left(X_i - \mu_{X_i}\right) / \sigma_{X_i}$ is the normalized random variable of X_i, ξ_i is the realization of Ξ_i, and $\psi(\cdot)$ represents the elliptical PDF. The reduced correlation coefficient ρ_{ij}' is obtained by implicitly solving Eq. (4.7) using a commercial program such as MATLAB.

If the elliptical variable $\mathbf{e}$ is the Gaussian variable, ρ_{ij}' can be approximated by

$$\rho_{ij}' = R_{ij} \rho_{ij} \tag{4.8}$$

where $R_{ij} = a + bV_i + cV_i^2 + d\rho_{ij} + e\rho_{ij}^2 + f\rho_{ij}V_i + gV_j + hV_j^2 + k\rho_{ij}V_j + lV_iV_j$, V_i and V_j are the coefficients of variation $\left(V = \sigma / \mu\right)$ for each variable, and the coefficients depend on the marginal distribution types of input variables. When the marginal CDFs are Gaussian and the joint CDF is Gaussian, the reduced correlation coefficient is the same as the original correlation coefficient, which means $R_{ij} = 1$. For various types of marginal

distribution types, the corresponding coefficients are given by Madsen (1986) and Ditlevsen (1996). The maximum error of the estimated correlation coefficient obtained from Eq. (4.8) is normally much less than 1%, and even if the marginal exponential CDF or negative correlation is involved, the maximum error in the correlation coefficient is at most up to 2% (Ditlevsen, 1996). Therefore, the approximation provides adequate accuracy with less computational effort. When **e** is the Gaussian variable, Eq. (4.5) is called as Nataf transformation (Madsen et al., 1986; Ditlevsen, 1996).

The linear transformation in Eq. (4.5) can be expressed in terms of the eigenvectors of the reduced covariance matrix $\mathbf{P'} = \left\{ \rho_{ij}^{'} \right\}$ and is called an orthogonal transformation (Madsen, 1986). The linear transformation can be limitedly used for elliptical copulas. If the identified copulas are not elliptical, it yields significant errors.

4.2.2 First-Order Reliability Method (FORM)

To calculate the probability of failure of the constraint function using the FORM and SORM, it is necessary to find a contact point on the limit state function, which has the shortest distance between the limit state function and the origin in the independent Gaussian space (**U**-space), as shown in Figure 4.1. The limit state function is transformed from **X**-space to **U**-space using the Rosenblatt transformation, defined as $g(\mathbf{u}) \equiv G(\mathbf{x}(\mathbf{u})) = G(\mathbf{x})$, and the contact point is called a MPP, denoted as $\mathbf{u}^*$. The MPP is obtained by solving the following optimization problem to

$$\begin{aligned} &\text{minimize} && \|\mathbf{u}\| \\ &\text{subject to} && g(\mathbf{u}) = 0 \end{aligned} \tag{4.9}$$

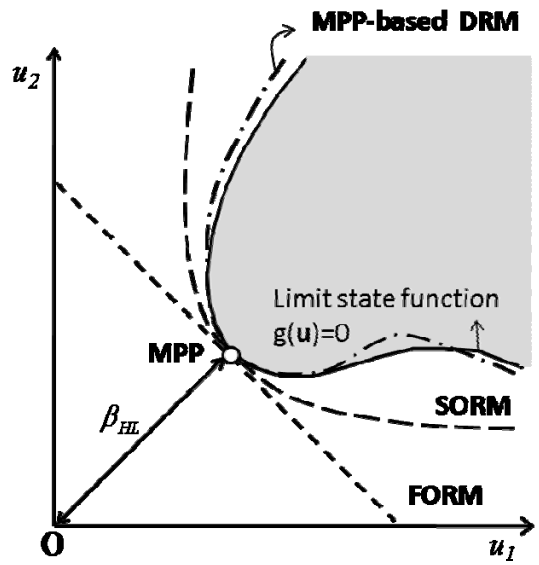


Figure 4.1. MPP and Reliability Index in **U**-space (Wei, 2006; Lee, 2008)

The shortest distance from the MPP to the origin of **U**-space is called the Hasofer-Lind reliability index (Hasofer and Lind, 1974), denoted as $\beta_{HL} = \left\| \mathbf{u}^* \right\|$. Thus, using the reliability index β_{HL}, the probability of failure is approximated as a linear approximation of the constraint function, which is indicated as a dotted line in Figure 4.1,

$$P_F^{FORM} \cong \Phi\left(-\beta_{HL}\right) \tag{4.10}$$

4.2.3 Second-Order Reliability Method (SORM)

The MPP obtained by solving Eq. (4.9) can be used to calculate the probability of failure using the SORM. Using the quadratic approximation of the constraint function and a transformation from **U**-space to rotated standard Gaussian space (**V**-space), the

probability of failure can be obtained using the SORM as (Breitung, 1984; Hohenbichler and Rackwitz, 1988; Rahman and Wei, 2006)

$$P_F^{\mathrm{SORM}} \cong \Phi(-\beta_{\mathrm{HL}}) \left| \mathbf{I}_{N-1} - 2\frac{\phi(\beta_{\mathrm{HL}})}{\Phi(-\beta_{\mathrm{HL}})} \tilde{\mathbf{A}}_{N-1} \right|^{-\frac{1}{2}} \tag{4.11}$$

where $\mathbf{A} = \begin{bmatrix} \tilde{\mathbf{A}}_{N-1} & \tilde{\mathbf{A}}_{1N} \\ \tilde{\mathbf{A}}_{N1} & \tilde{\mathbf{A}}_{NN} \end{bmatrix} = \frac{1}{2\left\|\nabla g(\mathbf{u})\right\|} \mathbf{R}^{\mathrm{T}}\mathbf{H}\mathbf{R}$, $\nabla g(\mathbf{u})$ is the gradient vector of $g(\mathbf{u})$, $\mathbf{H}$ is the Hessian matrix evaluated at the MPP in **X**-space, $\mathbf{R}$ is the rotation matrix such that $\mathbf{u} = \mathbf{R}\mathbf{v}$, and $\phi(\cdot)$ is the PDF of a standard Gaussian random variable.

4.3 Inverse Reliability Analysis

The reliability analysis explained in Section 4.2 is called the reliability index approach (RIA) (Tu and Choi, 1999), in that the probability of failure is calculated by finding the reliability index β_{HL} using Eq. (4.9). The RIA can calculate the probability of failure at a given design, but the RIA yields inefficiency and instability in the reliability analysis.

To solve this shortcoming of the RIA, a performance measure approach (PMA) has been developed; it has been reported that the PMA is much more robust and efficient than the reliability analysis in the RIA (Tu and Choi, 1999; Tu et al., 2001; Choi et al., 2001; Youn et al., 2003). Unlike the way the RIA directly calculates the probability of failure, the PMA judges whether a given design satisfies a probabilistic constraint for the given target probability of failure, P_F^{Tar}. Using the FORM, the target reliability index β_t

can be calculated as $\beta_t = -\Phi^{-1}\left(P_F^{Tar}\right)$. Then, the feasibility of the given design can be checked by solving the following optimization problem to

$$\begin{aligned} &\text{maximize} && \mathrm{g}(\mathbf{u}) \\ &\text{subject to} && \|\mathbf{u}\| = \beta_{\mathrm{t}} \end{aligned} \tag{4.12}$$

Finding the optimum point (MPP) is called inverse reliability analysis in that Eq. (4.12) is an inverse formulation of (4.9). If the constraint function value at the MPP, $\mathrm{g}(\mathbf{u}^*)$, is less than zero, then the feasibility of the design is satisfied for the given target probability of failure, P_F^{Tar}. If not, it is violated. Just as the FORM is used in the inverse reliability analysis, it might be possible to use the SORM. However, since the inverse reliability analysis using the SORM is much more difficult than using the FORM and the SORM requires the second order sensitivity, it has not been developed.

To find the MPP using the inverse reliability analysis, several methods have been developed: the mean value (MV) method, the advanced mean value (AMV) method (Wu et al., 1990; Wu, 1994), the hybrid mean value (HMV) method (Youn et al., 2003), and the enhanced hybrid mean value (HMV+) method (Youn et al., 2005). The MV method linearly approximates the constraint function using the function value and gradient at the mean value in the standard Gaussian space. The AMV method uses a gradient at the MPP obtained by the MV method to find the next MPP, and iteration continues until the approximate MPP converges to a correct MPP. The HMV method uses the AMV method for convex constraint function and the conjugate mean value (CMV) method for concave constraint function. The HMV+ method improved the HMV method by interpolating two previous MPP candidate points when the constraint function is concave instead of using the CMV method.

Since the gradient of the constraint function in standard Gaussian **U**-space is required to find the MPP in all methods, it needs to be calculated as

$$\frac{\partial G_i}{\partial u_j} = \sum_{k=1}^{n} \frac{\partial G_i}{\partial x_k} \cdot \frac{\partial x_k}{\partial u_j} = \sum_{k=1}^{n} \frac{\partial G_i}{\partial x_k} \cdot J_{kj} \tag{4.13}$$

where $J_{kj} = \begin{cases} \dfrac{\partial x_k}{\partial u_k} = \dfrac{f_{X_1 \cdots X_{k-1}}(x_1, \cdots, x_{k-1})\phi(u_k)}{f_{X_1 \cdots X_k}(x_1, \cdots, x_k)}, & for\ k = j \\ \dfrac{\partial x_k}{\partial u_j} = -\dfrac{f_{X_1 \cdots X_{k-1}}(x_1, \cdots, x_{k-1})}{f_{X_1 \cdots X_k}(x_1, \cdots, x_k)} \displaystyle\sum_{i=j}^{k-1} \frac{\partial F_{X_k}(x_k | x_1, \cdots, x_{k-1})}{\partial x_i} \frac{\partial x_i}{\partial u_j}, & for\ k > j \end{cases}$

and each component of Jacobian matrix J_{kj} can be calculated using derivatives of Eq. (4.2).

If the input variables are independent, then the Jacobian matrix is simply expressed as only diagonal terms:

$$J_{kk} = \frac{\phi(u_k)}{f_{X_k}(x_k)} \tag{4.14}$$

If the input variables are correlated with an elliptical copula, the Jacobian matrix is given as

$$J_{kj} = \begin{cases} \dfrac{\partial x_k}{\partial u_k} = \dfrac{\phi(u_k) l_{kk}}{f_{X_k}(x_k)}, & for\ k = j \\ \dfrac{\partial x_k}{\partial u_j} = \dfrac{\phi(u_k) l_{kj}}{f_{X_k}(x_k)}, & for\ k > j \end{cases} \tag{4.15}$$

where l_{kj} is obtained from Eq. (4.6).

When the input variables are correlated with a non-elliptical copula, it is hard to expand *n*-dimensional copula as mentioned in Chapter 2, so that bivariate copulas are considered in most copula applications. In this case, the Jacobian matrix can be simplified as

$$\mathbf{J} = \begin{bmatrix} \dfrac{\phi(u_1)}{f_{X_1}(x_1)} & 0 \\ -\dfrac{f_{X_1}(x_1)\phi(u_1)}{f_{X_1X_2}(x_1,x_2)}\dfrac{\partial^2 C(u,v|\theta)}{\partial u^2} & \dfrac{f_{X_1}(x_1)\phi(u_2)}{f_{X_1X_2}(x_1,x_2)} \end{bmatrix} \tag{4.16}$$

where u and v are marginal CDFs, $u = F_{X_1}(x_1)$ and $v = F_{X_2}(x_2)$. Once the explicit formulation of the joint CDF is given, Eq. (4.16) can be used to carry out the inverse reliability analysis.

4.4 MPP-based Dimension Reduction Method

MPP-based reliability analysis such as the FORM and the SORM has been commonly used for reliability assessment. However, when the constraint function is nonlinear or multi-dimensional, reliability analysis using the FORM could be erroneous because the FORM cannot handle the complexity of nonlinear or multi-dimensional functions. Reliability analysis using the SORM may be accurate, but the second-order derivatives required for the SORM are very difficult and expensive to obtain in practical industrial applications. On the other hand, the MPP-based DRM achieves both the efficiency of the FORM and the accuracy of the SORM (Lee, 2008).

The DRM is developed to accurately and efficiently approximate a multi-dimensional integral. There are several DRMs, depending on the level of dimension reduction: the univariate dimension reduction, bivariate dimension reduction, and multivariate dimension reduction. The univariate, bivariate, and multivariate dimension reduction indicate an additive decomposition of *n*-dimensional performance function into

one-, two-, and s-dimensional functions ($s \le n$), respectively. In this study, the univariate DRM is used for calculating probability of failure due to its simplicity and efficiency.

The univariate DRM is carried out by decomposing an n-dimensional constraint function $G(\mathbf{X})$ into the sum of one-dimensional functions at the MPP as (Rahman and Wei, 2006; Wei, 2006; Lee, 2008)

$$G(\mathbf{X}) \cong \hat{G}(\mathbf{X}) \equiv \sum_{i=1}^{n} G(x_1^*, \cdots, x_{i-1}^*, X_i, x_{i+1}^*, \cdots, x_n^*) - (n-1)G(\mathbf{x}^*) \tag{4.17}$$

where $\mathbf{x}^* = \{x_1^*, x_2^*, \cdots, x_n^*\}^{\mathrm{T}}$ is the FORM-based MPP obtained from Eq. (4.12) and n is the number of random variables.

In the inverse reliability analysis, since the probability of failure cannot be directly calculated in **U**-space, a constraint shift in a rotated standard Gaussian space (**V**-space) needs to be defined as

$$\tilde{G}^s(\mathbf{v}) \equiv \tilde{G}(\mathbf{v}) - \tilde{G}(\mathbf{v}^*) \tag{4.18}$$

where $\mathbf{v}^* = \{0, \cdots, 0, \beta\}^{\mathrm{T}}$ is a MPP in **V**-space and $\tilde{G}(\mathbf{v}) \equiv G(\mathbf{x}(\mathbf{v}))$. Then, using the shifted constraint function and the moment-based integration rule (Xu and Rahman, 2003), that is, similar to Gaussian quadrature (Atkinson, 1989), the probability of failure using the MPP-based DRM is approximated as (Lee, 2008)

$$P_F^{\mathrm{DRM}} = \frac{\prod_{i=1}^{n-1} \sum_{j=1}^{N} w_j \Phi\left(-\beta + \frac{\tilde{G}_i^s(v_i^j)}{b_1}\right)}{\Phi\left(-\beta\right)^{n-2}} \tag{4.19}$$

where $b_1 = \|\nabla_U g\|$, v_i^j represents the jth quadrature point for v_i, w_j denotes weights, and N is the number of quadrature points. The quadrature points and weights for the standard Gaussian random variables v_i are shown in Table 4.1.

Table 4. 1. Gaussian Quadrature Points and Weights (Lee, 2008)

n	Quadrature Points	Weights
1	0.0	1.0
3	$\pm\sqrt{3}$	0.166667
	0.0	0.666667
5	±2.856970	0.011257
	±1.355626	0.222076
	0.0	0.533333

Using the estimated P_F^{DRM} obtained from the MPP-based DRM for the shifted constraint function $G^{\mathrm{s}}(\mathbf{x})$, the corresponding reliability index β_{DRM} can be defined as

$$\beta_{\mathrm{DRM}} = -\Phi^{-1}(P_F^{\mathrm{DRM}}) \tag{4.20}$$

which is not the same as the target reliability index $\beta_t = -\Phi^{-1}(P_F^{Tar})$ because the nonlinearity of the constraint function is reflected in the calculation of P_F^{DRM}. Hence, using β_{DRM}, a new updated reliability index β_{up} can be defined as (Lee, 2008)

$$\beta_{\mathrm{up}} \equiv \beta_{\mathrm{cur}} + \Delta\beta = \beta_{\mathrm{cur}} + (\beta_t - \beta_{\mathrm{DRM}}) \tag{4.21}$$

where β_{cur} is the current reliability index. If the constraint function is concave, β_{up} will be smaller than β_{cur} because $\beta_{\mathrm{DRM}} > \beta_t$, i.e., a smaller reliability index is used to correctly update MPP using DRM for the concave constraint function, and vice versa for the convex constraint function. If the FORM-based reliability analysis overestimates the probability of failure, then the constraint function is defined as concave near the MPP. Otherwise, the constraint function is defined as convex (Lee, 2008).

Using this updated reliability index, the updated MPP satisfying the target probability of failure can be found through a new inverse reliability analysis. After

finding a new MPP, constraint shift is again used to calculate the probability of failure using DRM. After iteratively carrying out this procedure until converged, the DRM-based MPP can be obtained in which P_F^{DRM} is the same as the target probability of failure, P_F^{Tar}. A detailed explanation of the MPP-based DRM is presented by Lee et al. (2008).

4.5 Reliability-Based Design Optimization

In general, the RBDO model can be formulated to

$$\begin{aligned} &\text{minimize} && \text{cost}(\mathbf{d}) \\ &\text{subject to} && P\left(G_i(\mathbf{X})>0\right) \le P_{F_i}^{Tar},\ i=1,\cdots,nc \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U,\quad \mathbf{d}\in \mathrm{R}^{ndv} \text{ and } \mathbf{X}\in \mathrm{R}^n \end{aligned} \tag{4.22}$$

where $\mathbf{X}$ is the vector of random variables; $\mathbf{d}=\boldsymbol{\mu}\left(\mathbf{X}\right)$ is the vector of design variables; $G_i(\mathbf{X})$ represents the ith constraint functions; $P_{F_i}^{Tar}$ is the given target probability of failure for the ith constraint; and *nc*, *ndv*, and *n* are the number of constraints, number of design variables, and number of random variables, respectively.

Using the inverse reliability analysis, the ith probabilistic constraint can be written as

$$P\left(G_i(\mathbf{X})>0\right)-P_{F_i}^{Tar} \le 0 \Rightarrow G_i\,(\mathbf{x}^*) \le 0 \tag{4.23}$$

where $G_i\,(\mathbf{x}^*)$ is the ith constraint function value evaluated at the MPP $\mathbf{x}^*$ in **X**-space. Using the FORM, Eq. (4.23) can be rewritten as

$$P\left(G_i(\mathbf{X})>0\right)-\Phi\left(-\beta_{t_i}\right) \le 0 \Rightarrow G_i\,(\mathbf{x}^*) \le 0 \tag{4.24}$$

where $P_{F_i}^{Tar} = \Phi\left(-\beta_{t_i}\right)$ and β_{t_i} is the target reliability index. Thus, Eq. (4.22) can be written as

$$\begin{aligned} &\text{minimize} && \text{cost}(\mathbf{d}) \\ &\text{subject to} && P\left(G_i(\mathbf{X}) > 0\right) \le P_{F_i}^{Tar} = \Phi(-\beta_{t_i}),\ i = 1, \cdots, nc \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \quad \mathbf{d} \in \mathrm{R}^{ndv} \text{ and } \mathbf{X} \in \mathrm{R}^n \end{aligned} \tag{4.25}$$

where the probabilistic constraint, Eq. (4.24), is changed into

$$P\left(G_i(\mathbf{X}) > 0\right) - \Phi(-\beta_{t_i}) \le 0 \Rightarrow G_i\ (\mathbf{x}_{\mathrm{FORM}}^*) \le 0 \tag{4.26}$$

where $\mathbf{x}_{\mathrm{FORM}}^*$ is FORM-based MPP, which is obtained by solving Eq. (4.12) and transforming $\mathbf{u}_{\mathrm{FORM}}^*$ in **U**-space into $\mathbf{x}_{\mathrm{FORM}}^*$ in **X**-space using Eq. (4.4).

Similar to the FORM, using the DRM-based inverse reliability analysis, the RBDO formulation in Eq. (4.22) can be rewritten as

$$\begin{aligned} &\text{minimize} && \text{cost}(\mathbf{d}) \\ &\text{subject to} && G_i(\mathbf{x}_{\mathrm{DRM}}^*) \le 0,\ i = 1, \cdots, nc \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \quad \mathbf{d} \in \mathrm{R}^{ndv} \text{ and } \mathbf{X} \in \mathrm{R}^n \end{aligned} \tag{4.27}$$

where $\mathbf{x}_{\mathrm{DRM}}^*$ is the MPP obtained from the DRM.

CHAPTER V

TRANSFORMATION ORDERING EFFECT

5.1 Introduction

Once the copula is identified from the Bayesian method, the Rosenblatt transformation is utilized to transform the original random variables into the independent standard Gaussian variables for the inverse reliability analysis as explained in Chapter 4. However, depending on joint input CDF types, different transformation orderings of input variables might yield different RBDO results. For example, if the input variables are independent or have a joint CDF modeled by an elliptical copula, there is no effect of transformation ordering. However, if the input variables have a joint CDF modeled by a non-elliptical copula, which often occurs in industrial applications, since the Rosenblatt transformation becomes highly nonlinear, certain transformation orderings can significantly affect the nonlinearity of the transformed constraints. In this case, if the FORM is used, the inverse reliability analysis results could be very different for the different ordering since the FORM uses a linear approximation of the constraint to estimate the probability of failure. This effect of transformation ordering in RBDO will be unacceptable and make the user significantly concerned.

To reduce the dependency of the inverse reliability analysis result and thus the RBDO result on the ordering of the Rosenblatt transformation, it is proposed to use the MPP-based DRM for the inverse reliability analysis in this study. Section 5.2 provides theoretical proofs to show that the independent and elliptical copulas are independent of transformation orderings, whereas the non-elliptical copula is not. Section 5.3 illustrates

numerical examples to show how much the transformation orderings affect RBDO results and how much the MPP-based DRM reduces the effect of the transformation orderings.

5.2 Transformation Ordering for Various Joint CDF Types

The joint CDFs of input random variables can be categorized as follows: independent joint CDF, a joint CDF modeled by an elliptical copula, and a joint CDF modeled by a non-elliptical copula. To study the effect of transformation ordering in the inverse reliability analysis for various input joint CDF types, it is necessary to investigate whether the same MPPs can be obtained for different transformation orderings. However, since the MPPs depend on constraint functions, it is not convenient to compare the MPPs for all constraint functions. Instead of comparing the MPPs, comparing β_t-contours in **X**-space that are obtained by transforming the β_t-contour, $\|\mathbf{u}\| = \beta_{t_i}$, in **U**-space for different transformation orderings is more appropriate. Obtaining the same β_t-contours means obtaining the same MPPs in **X**-space, which will lead to the same RBDO result.

5.2.1 Transformation Orderings in Rosenblatt Transformation

As explained in Section 4.2.1, the input random variables need to be transformed into the independent standard Gaussian variables using the Rosenblatt transformation, which is defined as the successive conditioning:

$$
\begin{aligned}
u_1 &= \Phi^{-1}\left[F_{X_1}\left(x_1\right)\right] \\
u_2 &= \Phi^{-1}\left[F_{X_2}\left(x_2 \middle| x_1\right)\right] \\
&\vdots \\
u_i &= \Phi^{-1}\left[F_{X_i}\left(x_i \middle| x_1, x_2, \cdots, x_{i-1}\right)\right] \\
&\vdots \\
u_j &= \Phi^{-1}\left[F_{X_j}\left(x_j \middle| x_1, x_2, \cdots x_i, \cdots, x_{j-1}\right)\right] \\
&\vdots \\
u_n &= \Phi^{-1}\left[F_{X_n}\left(x_n \middle| x_1, x_2, \cdots, x_{n-1}\right)\right]
\end{aligned}
\tag{5.1}
$$

If the ordering of the variable x_i is changed into the variable x_j, Eq. (5.1) can be rewritten as

$$
\begin{aligned}
u_1 &= \Phi^{-1}\left[F_{X_1}\left(x_1\right)\right] \\
u_2 &= \Phi^{-1}\left[\left(F_{X_2}\left(x_2 \middle| x_1\right)\right)\right] \\
&\vdots \\
u_i &= \Phi^{-1}\left[F_{X_j}\left(x_j \middle| x_1, x_2, \cdots, x_{i-1}\right)\right] \\
&\vdots \\
u_j &= \Phi^{-1}\left[F_{X_i}\left(x_i \middle| x_1, x_2, \cdots x_{i-1}, x_j, x_{i+1}, \cdots, x_{j-1}\right)\right] \\
&\vdots \\
u_n &= \Phi^{-1}\left[F_{X_n}\left(x_n \middle| x_1, x_2, \cdots, x_{n-1}\right)\right]
\end{aligned}
\tag{5.2}
$$

Thus, if the number of variables is *n*, there are $n!$ ways of transforming the original variables into independent standard Gaussian variables. Even though there are many different ways to transform the original variables into independent standard Gaussian variables, since the Rosenblatt transformation is exact, if the inverse reliability analysis in the independent standard Gaussian space is exact, then we should obtain the same results. However, if the FORM is used for the inverse reliability analysis, certain

orders of transformation might yield more errors than other orders of transformation depending on the input joint CDF type.

5.2.2 Independent and Elliptical Copula

First, when the input variables are independent, the β_t-contour in **U**-space is transformed into **X**-space using the Rosenblatt transformation with a given ordering in Eq. (5.1) as

$$\|\mathbf{u}\|^2 = \left[\Phi^{-1}\left(z_1\right)\right]^2 + \cdots + \left[\Phi^{-1}\left(z_i\right)\right]^2 + \cdots + \left[\Phi^{-1}\left(z_j\right)\right]^2 + \cdots + \left[\Phi^{-1}\left(z_n\right)\right]^2 = \beta_t^2 \quad (5.3)$$

where $z_k = F_{X_k}\left(x_k\right)$ for $k = 1, \cdots, n$ and β_t is the target reliability index.

When the order of the variable x_i is interchanged into the variable x_j for the second ordering, the β_t-contour in **U**-space is transformed into **X**-space using Eq. (5.2) as

$$\|\mathbf{u}\|^2 = \left[\Phi^{-1}\left(z_1\right)\right]^2 + \cdots + \left[\Phi^{-1}\left(z_j\right)\right]^2 + \cdots + \left[\Phi^{-1}\left(z_i\right)\right]^2 + \cdots + \left[\Phi^{-1}\left(z_n\right)\right]^2 = \beta_t^2 \quad (5.4)$$

which results in the same transformed β_t-contour, Eq. (5.3). Thus, there is no effect of transformation ordering for the independent variables.

Second, consider correlated input variables with a joint CDF modeled by the elliptical copula. In the joint CDF modeled by the elliptical copula, each variable $\Psi^{-1}\left[F_{X_i}\left(x_i\right)\right]$ for $i = 1, \cdots, n$ is the correlated standard elliptical variable with the covariance matrix $\mathbf{P}' = \left\{\rho_{ij}'\right\}$, where $\Psi^{-1}\left(\cdot\right)$ is the inverse of the elliptical CDF. Let $e_i = \Psi^{-1}\left[F_{X_i}\left(x_i\right)\right]$. Using the Rosenblatt transformation in Eq. (4.5), the β_t-contour is transformed from **U**- to **X**-space as

$$\mathbf{u}^T\mathbf{u} = \mathbf{e}^T\left(\mathbf{L}^{-1}\right)^T\mathbf{L}^{-1}\mathbf{e} = \mathbf{e}^T\left(\mathbf{P}^{'}\right)^{-1}\mathbf{e} \tag{5.5}$$

where $\mathbf{e} = \left[e_1, \cdots, e_i, \cdots, e_j, \cdots, e_n\right]^T$, and $\mathbf{P}^{'}$ is given as

$$\mathbf{P}^{'} = \begin{bmatrix} 1 & \rho_{12}^{'} & \cdots & \rho_{1i}^{'} & \cdots & \rho_{1j}^{'} & \cdots & \rho_{1n}^{'} \\ & 1 & & \rho_{2i}^{'} & & \rho_{2j}^{'} & & \rho_{2n}^{'} \\ & & & \vdots & & \vdots & & \vdots \\ & & & 1 & & \rho_{ij}^{'} & & \rho_{in}^{'} \\ & & & & & \vdots & & \vdots \\ & & sym. & & & 1 & & \rho_{jn}^{'} \\ & & & & & & & \vdots \\ & & & & & & & 1 \end{bmatrix} \tag{5.6}$$

If the order is changed, i.e., the order of the ith and jth variables are interchanged ($i<j$) in the Rosenblatt transformation, Eq. (4.5) is changed to

$$\bar{\mathbf{u}} = \mathbf{L}_1^{-1}\mathbf{e}_1 \tag{5.7}$$

where the vector of the elliptical variable with the interchanged order represents $\mathbf{e}_1 = \left[e_1, \cdots, e_j, \cdots, e_i, \cdots, e_n\right]^T$. $\mathbf{L}_1$ is obtained from the Cholesky decomposition of $\mathbf{P}_1^{'}$ (i.e., $\mathbf{P}_1^{'} = \mathbf{L}_1\mathbf{L}_1^T$), where $\mathbf{P}_1^{'}$ is

$$\mathbf{P}_1^{'} = \begin{bmatrix} 1 & \rho_{12}^{'} & \cdots & \rho_{1j}^{'} & \cdots & \rho_{1i}^{'} & \cdots & \rho_{1n}^{'} \\ & 1 & & \rho_{2j}^{'} & & \rho_{2i}^{'} & & \rho_{2n}^{'} \\ & & & \vdots & & \vdots & & \vdots \\ & & & 1 & & \rho_{ji}^{'} & & \rho_{jn}^{'} \\ & & & & & \vdots & & \vdots \\ & & sym. & & & 1 & & \rho_{in}^{'} \\ & & & & & & & \vdots \\ & & & & & & & 1 \end{bmatrix} \tag{5.8}$$

and

$$\mathbf{L}_1 = \begin{bmatrix} l_{11} & 0 & & & \cdots & & & 0 \\ l_{21} & l_{22} & 0 & & \cdots & & & 0 \\ \vdots & & \ddots & & & & & \vdots \\ l_{i1} & l_{i2} & \cdots & l_{ii} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & & & \vdots \\ l_{j1} & l_{j2} & \cdots & l_{ji} & \cdots & l_{jj} & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{ni} & \cdots & l_{nj} & \cdots & l_{nn} \end{bmatrix} \tag{5.9}$$

Then, the transformed β_t-contour with the interchanged order is given by

$$\bar{\mathbf{u}}^T \bar{\mathbf{u}} = \mathbf{e}_1^T \left(\mathbf{L}_1^{-1}\right)^T \mathbf{L}_1^{-1} \mathbf{e}_1 = \mathbf{e}_1^T \left(\mathbf{P}_1^{'}\right)^{-1} \mathbf{e}_1 \tag{5.10}$$

To show that the interchanged ordering provides the same transformed β_t-contours, i.e., $\mathbf{u}^T \mathbf{u} = \bar{\mathbf{u}}^T \bar{\mathbf{u}}$ in Eqs. (5.5) and (5.10), Eq. (5.10) needs to be expressed in terms of $\mathbf{e}$ instead of $\mathbf{e}_1$. For this, another matrix $\mathbf{L}_2$ is introduced. Because the vector of the elliptical variable $\mathbf{e}$ with the original order can be obtained by interchanging the ith row with the jth row of $\mathbf{L}_1$ such that $\mathbf{e} = \mathbf{L}_2 \bar{\mathbf{u}}$, the matrix $\mathbf{L}_2$ cab be obtained as

$$\mathbf{L}_2 = \begin{bmatrix} l_{11} & 0 & & & \cdots & & & 0 \\ l_{21} & l_{22} & 0 & & \cdots & & & 0 \\ \vdots & & \ddots & & & & & \vdots \\ l_{j1} & l_{j2} & \cdots & l_{ji} & \cdots & l_{jj} & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots \\ l_{i1} & l_{i2} & \cdots & l_{ii} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{ni} & \cdots & l_{nj} & \cdots & l_{nn} \end{bmatrix} \tag{5.11}$$

The β_t-contours are given as

$$\bar{\mathbf{u}}^T \bar{\mathbf{u}} = \mathbf{e}_1^T \left(\mathbf{P}_1^{'}\right)^{-1} \mathbf{e}_1 = \mathbf{e}^T \left(\mathbf{L}_2^{-1}\right)^T \mathbf{L}_2^{-1} \mathbf{e} = \mathbf{e}^T \left(\mathbf{P}_2^{'}\right)^{-1} \mathbf{e} \tag{5.12}$$

To show that $\mathbf{u}^T\mathbf{u} = \bar{\mathbf{u}}^T\bar{\mathbf{u}}$ in Eqs. (5.5) and (5.12), it needs to be shown that $\mathbf{P}' = \mathbf{P}_2'$ or $\mathbf{L}\mathbf{L}^T = \mathbf{L}_2\mathbf{L}_2^T$. Consider two arbitrary correlation coefficients of $\mathbf{P}_1'$. For any arbitrary kth column, the entry at the ith row of $\mathbf{P}_1'$, $\left(\mathbf{P}_1'\right)_{ik}$ and the one in the jth row of $\mathbf{P}_1'$, $\left(\mathbf{P}_1'\right)_{jk}$ are ρ_{jk}' and ρ_{ik}', respectively, as shown in Eq. (5.8). Since ρ_{jk}' and ρ_{ik}' are the entries at the jth row and kth column of $\mathbf{P}'$ and at the ith row and kth column of $\mathbf{P}'$, respectively, all entries of $\mathbf{P}'$ are the same as those of $\mathbf{P}_2'$, as follows:

$$\left(\mathbf{P}_1'\right)_{ik} = \left(\mathbf{P}'\right)_{jk} = i^{\text{th}} \text{ row of } \mathbf{L}_1 \times k^{\text{th}} \text{ col. of } \mathbf{L}_1^T = j^{\text{th}} \text{ row of } \mathbf{L}_2 \times k^{\text{th}} \text{ col. of } \mathbf{L}_2^T = \left(\mathbf{P}_2'\right)_{jk} \quad (5.13)$$

$$\left(\mathbf{P}_1'\right)_{jk} = \left(\mathbf{P}'\right)_{ik} = j^{\text{th}} \text{ row of } \mathbf{L}_1 \times k^{\text{th}} \text{ col. of } \mathbf{L}_1^T = i^{\text{th}} \text{ row of } \mathbf{L}_2 \times k^{\text{th}} \text{ col. of } \mathbf{L}_2^T = \left(\mathbf{P}_2'\right)_{ik} \quad (5.14)$$

This means the transformed β_t-contours are the same, even for different transformation orderings of input variables. Therefore, the Rosenblatt transformation is independent of ordering for the joint CDF modeled by an elliptical copula.

5.2.3. Non-elliptical Copula

Consider correlated input variables with a joint CDF modeled by a non-elliptical copula. For example, let two random variables have a joint CDF modeled by the Clayton copula, which is one of the Archimedean copulas, with the marginal Gaussian CDFs $X_1, X_2 \sim N\left(0,1^2\right)$. The Clayton copula is defined as

$$C\left(z_1, z_2 \middle| \theta\right) = \left[z_1^{-\theta} + z_2^{-\theta} - 1\right]^{-1/\theta}, \quad for\ \theta > 0 \quad (5.15)$$

where the generator is $\varphi_\theta(t) = \frac{1}{\theta}\left(t^{-\theta} - 1\right)$, $z_1 = \Phi(x_1)$, $z_2 = \Phi(x_2)$, and θ is the correlation parameter of the Clayton copula. In the Clayton copula, using the Kendall's tau τ obtained from samples, θ can be expressed as

$$\theta = \frac{2\tau}{1-\tau} \tag{5.16}$$

Using the Clayton copula, the Rosenblatt transformation can be carried out in two different ways as

$$\begin{aligned} \Phi(u_1) &= F_{X_1}(x_1) = \Phi(x_1) \\ \Phi(u_2) &= F_{X_2}(x_2|x_1) = \Phi(x_1)^{-\theta-1}\left[\Phi(x_1)^{-\theta} + \Phi(x_2)^{-\theta} - 1\right]^{-1/\theta-1} \end{aligned} \tag{5.17}$$

and

$$\begin{aligned} \Phi(u_1) &= F_{X_2}(x_2) = \Phi(x_2) \\ \Phi(u_2) &= F_{X_1}(x_1|x_2) = \Phi(x_2)^{-\theta-1}\left[\Phi(x_1)^{-\theta} + \Phi(x_2)^{-\theta} - 1\right]^{-1/\theta-1} \end{aligned} \tag{5.18}$$

Using Eqs. (5.17) and (5.18), the β_t-contour can be expressed in terms of x_1 and x_2

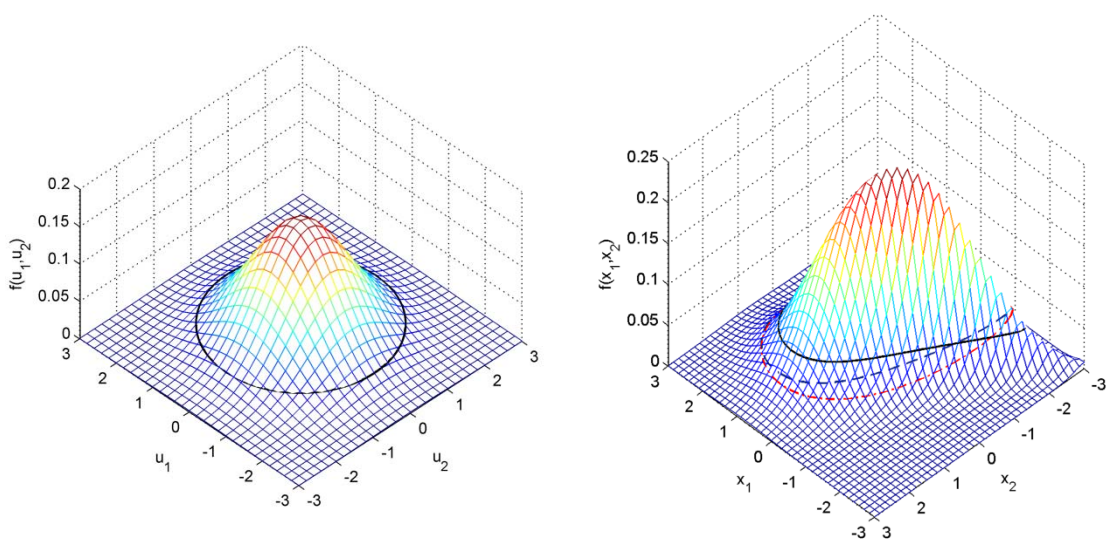
$$\mathbf{u}^{\mathbf{T}}\mathbf{u} = u_1^2 + u_2^2 = x_1^2 + \left(\Phi^{-1}\left(\Phi(x_1)^{-\theta-1}\left[\Phi(x_1)^{-\theta} + \Phi(x_2)^{-\theta} - 1\right]^{-1/\theta-1}\right)\right) = \beta_t^2 \tag{5.19}$$

and

$$\mathbf{u}^{\mathbf{T}}\mathbf{u} = u_1^2 + u_2^2 = x_2^2 + \left(\Phi^{-1}\left(\Phi(x_2)^{-\theta-1}\left[\Phi(x_1)^{-\theta} + \Phi(x_2)^{-\theta} - 1\right]^{-1/\theta-1}\right)\right) = \beta_t^2 \tag{5.20}$$

Figure 5.1 (a) shows the β_t-contour in **U**-space with an independent PDF where $\beta_t = 2.0$. Figure 5.1 (b) shows two β_t-contours transformed from **U**-space to **X**-space indicated as dashed-dot ring (ordering 1) and dashed ring (ordering 2) with a PDF modeled by the Clayton copula. If input variables are independent or correlated with elliptical copula, the β_t-contour in **U**-space, which is a solid ring in Figure 5.1 (a), is

transformed into β_t-contours in **X**-space that are parallel to the X_1-X_2 plane. Therefore, the β_t-contours will be the same in **X**-space for different transformation orderings. However, when input variables are correlated with a non-elliptical copula, because of the nonlinear Rosenblatt transformation, the β_t-contour in **U**-space is differently transformed to **X**-space according to the transformation ordering, which are not parallel to the X_1-X_2 plane, as shown by dashed and dashed-dot rings in Figure 5.1 (b). That is, the β_t-contours projected on the X_1-X_2 plane are different from a β_t-contour with constant PDF values indicated as a solid ring in Figure 5.1 (b), whereas the independent PDF contour for a certain PDF value and the β_t-contours will coincide. However, note that the inverse reliability analysis is carried out in **U**-space, not **X**-space.



(a) **U**-space (b) **X**-space

Figure 5.1. Joint PDFs and β_t-contours in **U**-space and **X**-space

It might be possible that MPPs obtained from different transformation orderings could be the same even though the β_t-contours in **X**-space are different. However, because the MPPs depend on both the β_t-contour shapes and constraint functions, it is not easy to predict how the MPPs depend on the β_t-contour shape. As seen in Fig. 5.1 (b), the two β_t-contours in **X**-space have intersection points, where x_1 and x_2 values are the same. If both MPPs happen to be at the intersection point, then even if the β_t-contours in **X**-space are different, these MPPs will be the same. However, it will be extremely rare that both MPPs are at the same intersection point of two differently transformed β_t-contours.

Thus, for the joint CDF modeled by a non-elliptical copula, the Rosenblatt transformation becomes highly nonlinear, which cannot be handled accurately by the FORM. A more accurate method than the FORM for the estimation of the probability of failure in the reliability analysis is necessary to reduce the effect of the transformation ordering on the RBDO results.

5.3 Examples

To observe how the ordering of transformation affects the RBDO results for the joint CDFs modeled by elliptical copulas and non-elliptical copulas, two- and four-dimensional mathematical problems are tested.

5.3.1 Two-Dimensional Example

Suppose that two input variables are correlated with the Gaussian copula where Pearson's rho is given as $\rho = 0.5$. The RBDO formulation is defined as

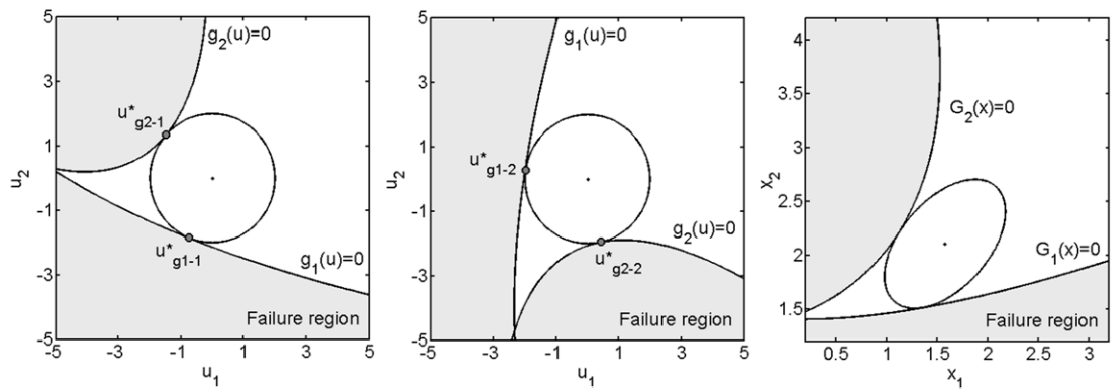
$$
\begin{aligned}
&\text{min.} \quad \text{cost}(\mathbf{d}) = d_1 + d_2 \\
&\text{s.t.} \quad P\left(G_i\left(\mathbf{X}\right) \geq 0\right) \leq P_F^{Tar},\ i = 1,2,3 \\
&\qquad 0 \leq d_1, d_2 \leq 10,\ P_F^{Tar} = 2.275\% \\
&G_1(\mathbf{X}) = 1 - \left(0.4339X_1 - 0.9010X_2 - 1.5\right)^2 \left(0.9010X_1 + 0.4339X_2 + 2\right)/20 \\
&G_2(\mathbf{X}) = 1 - \left(X_1 + X_2 - 2.8\right)^2/30 - \left(X_1 - X_2 + 12\right)^2/120 \\
&G_3(\mathbf{X}) = 1 - 80/\left(8X_1 + X_2^2 + 5\right)
\end{aligned}
\tag{5.21}
$$

where the marginal CDFs are Gaussian given by $X_1, X_2 \sim N(5.0, 0.3^2)$. Denote the initial ordering as ordering 1 and the interchanged ordering ($x_1 \leftrightarrow x_2$) as ordering 2.

When input variables are correlated with the Gaussian copula (even if the MPPs are different in **U**-space for different orderings as shown in Figure 5.2 (a) and (b)), the β_t -contours are the same, which means the MPPs are the same in **X**-space, as shown in Figure 5.2 (c). Thus, the same optimum design points are obtained even for different orderings, which confirms the theoretical proof that there is no transformation ordering effect for the joint CDF modeled by the Gaussian copula, one of the elliptical copula families.

Even though there is no effect of transformation ordering, since the second constraint function is nonlinear, the FORM has some error in estimating the probability of failure for the second constraint, P_{F_2}. Using FORM and DRM with three and five quadrature points (FORM, DRM3, and DRM5, respectively, in Table 5.1), RBDO results including the cost and optimum design results are obtained. The probabilities of failure

P_{F_1} and P_{F_2} are calculated at the obtained optimum design points for active constraints G_1 and G_2 using Monte Carlo simulation (MCS).



(a) Ordering 1 in **U**-space (b) Ordering 2 in **U**-space (c) Two Orderings in **X**-space

Figure 5.2. β_t-contours and Constraint Functions with Different Transformation Orderings Using Gaussian Copula

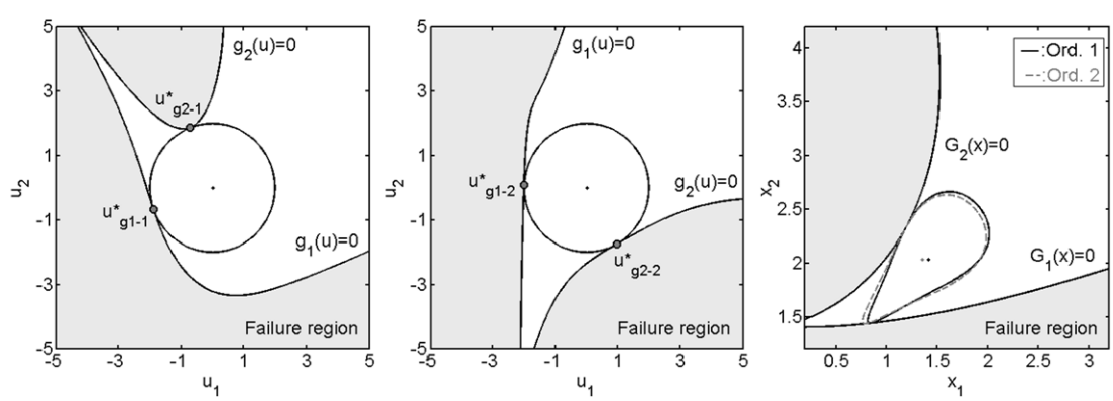
Table 5.1. RBDO Results Using Gaussian Copula ($P_F^{Tar} = 2.275\%$)

Case	Cost	Optimum design points	P_{F_1} (%)	P_{F_2} (%)
FORM	3.678	1.574, 2.104	2.342	1.722
DRM3	3.653	1.548, 2.105	2.208	2.142
DRM5	3.651	1.546, 2.105	2.256	2.276

As shown in the second row of Table 5.1, P_{F_2} is poorly estimated, i.e., less than the target probability of failure, 2.275% compared to P_{F_1}. When the MPP-based DRM with three quadrature points is used, the probabilities of failure for both constraint

functions (P_{F_1} and P_{F_2}) are closer to the target probability of failure. When the MPP-based DRM with five quadrature points is used, P_{F_1} and P_{F_2} are almost the same as the target probability of failure. Thus, the FORM error can be reduced using the MPP-based DRM.

Suppose that two input variables are correlated by a Clayton copula with Kendall's tau $\tau = 0.5$. The same RBDO problem in Eq. (5.21) is tested. As shown in Figures. 5.3 (a) and (b), the β_t-contours in **U**-space are the same, but the constraint functions are differently transformed from **X**-space into **U**-space according to the different transformation orderings. Since the transformation of the non-elliptical copula is highly nonlinear, some transformed constraint functions become highly nonlinear in **U**-space. For ordering 1 (Figure 5.3 (a)), the first constraint function is mildly nonlinear near the MPP, u^*_{g1-1}, but the second constraint function is highly nonlinear near the corresponding MPP, u^*_{g2-1}, which yields a large FORM error.



(a) Ordering 1 in **U**-space (b) Ordering 2 in **U**-space (c) Two Orderings in **X**-space

Figure 5.3. β_t-contours and Constraint Functions with Different Transformation Orderings Using Clayton Copula

On the other hand, for ordering 2 in Figure 5.3 (b), two constraint functions are mildly nonlinear near the MPPs , u^*_{g1-2} and u^*_{g2-2} , so that the FORM estimates the probability of failure more accurately than that of the second constraint function with the first ordering. The β_t-contours in Figure 5.3 (a) and (b) can be transformed as two different β_t-contours in **X**-space according to different orderings, as shown in Figure 5.3 (c). These different β_t-contours provide different optimum design points, as shown in the FORM results of Table 5.2.

Table 5.2. RBDO Results Using Clayton Copula ($P_F^{Tar} = 2.275\%$)

Case	Cost	Optimum design points	$\left\|d_1^{opt} - d_2^{opt}\right\|$	P_{F_1} (%)	P_{F_2} (%)
FORM-1	3.446	1.413, 2.032	0.062	2.531	1.031
FORM-2	3.386	1.352, 2.034		2.354	2.108
DRM3-1	3.417	1.380, 2.037	0.034	2.352	1.582
DRM3-2	3.383	1.347, 2.036		2.280	2.257
DRM5-1	3.400	1.364, 2.036	0.016	2.320	1.881
DRM5-2	3.385	1.348, 2.037		2.276	2.266

In Table 5.2, FORM-1 and FORM-2 indicate the FORM with orderings 1 and 2, respectively. The probability of failures P_{F_1} and P_{F_2} are calculated at the obtained optimum design points for active constraints G_1 and G_2 using the MCS. As expected, when the FORM is used for the ordering 1, FORM-1, the probability of failure for the second constraint P_{F_2} is poorly estimated, i.e., much less than target probability of 2.275%. As a result, the optimum design points obtained using the FORM with different

orderings are indeed different as shown in the third and fourth columns of Table 5.2. If the MPP-based DRM with three quadrature points, denoted as DRM3-1 and DRM3-2 for orderings 1 and 2, respectively, is used, the difference between optimum design results is reduced from 0.062 to 0.034 and the DRM provides a more accurate estimation of the probabilities of failure, i.e., closer to 2.275% for both orderings. If the number of quadrature points is five (DRM5-1 and DRM5-2), then the optimum design points are much closer to each other for both orderings and the probability of failure calculation also becomes more accurate. To estimate the probability of failures more accurately for highly nonlinear constraint functions such as the second constraint shown in Fig. 5.3(a), more than five quadrature points might be necessary. However, increasing the number of quadrature points means increasing computational effort. Therefore, three or five quadrature points for the MPP-based DRM are usually used.

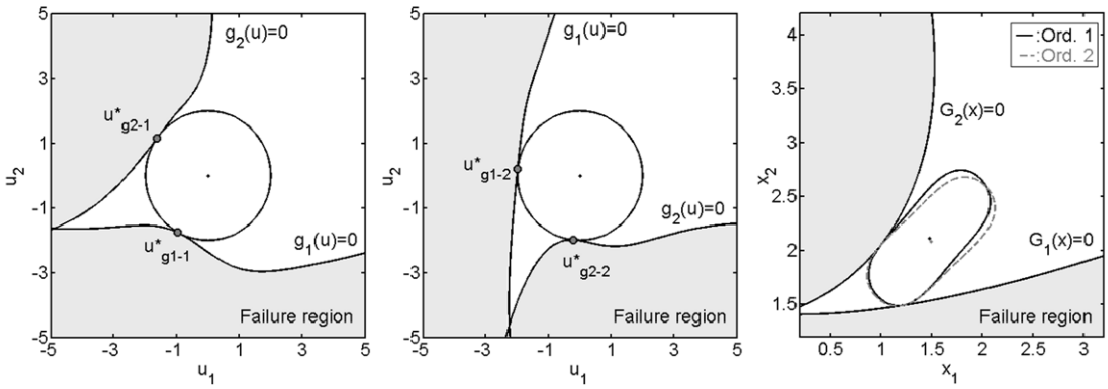
For the same problem in Eq. (5.21), assume that two input variables are now correlated with a Frank copula, which belongs to the Archimedean copula, with Kendall's tau $\tau = 0.5$. The Frank copula is given as

$$C\left(z_1, z_2 \middle| \theta\right) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta z_1} - 1\right)\left(e^{-\theta z_2} - 1\right)}{e^{-\theta} - 1}\right) \tag{5.22}$$

The correlation parameter θ can be calculated from Kendall's tau by solving the following equation:

$$\tau = 1 - \frac{4}{\theta}\left(1 - \frac{1}{\theta}\int_0^t \frac{\theta}{e^t - 1}dt\right) \tag{5.23}$$

As observed in the previous example, when the joint CDF modeled by a non-elliptical copula is used, the MPPs obtained from differently transformed constraint functions provide different RBDO results according to the different ordering of input variables.



(a) Ordering 1 in U-space (b) Ordering 2 in U-space (c) Two Orderings in X-space

Figure 5.4. β_t-contours and Constraint Functions with Different Transformation Orderings Using the Frank Copula

Due to the nonlinear transformation of the joint CDF modeled by a non-elliptical copula, the constraint functions in **X**-space are differently transformed into those in **U**-space for different orderings, and some transformed constraint functions are highly nonlinear in Figures. 5.4 (a) and (b). If the β_t-contour is transformed from **U**-space to **X**-space, it becomes two different β_t-contours in **X**-space for different transformation orderings, which result in different optimum design points as shown in Figure 5.4 (c).

For ordering 1, the first constraint function is highly nonlinear near the MPP, u^*_{g1-1}, in Figure 5.4 (a), while for ordering 2, the second constraint function is highly nonlinear near the MPP, u^*_{g2-2} in Figure 5.4 (b). Therefore, for ordering 1, the FORM error is large for the first constraint (FORM-1), while for the second ordering, it is large for the second constraint (FORM-2). As shown in Table 5.3, probabilities of failure P_{F_2} for ordering 1

and P_{F_1} for ordering 2 are close to the target probability, 2.275%, whereas P_{F_1} for ordering 1 and P_{F_2} for ordering 2 are not.

When the MPP-based DRM with three quadrature points is used, the probability of failure becomes closer to the target probability for both orderings (DRM3-1 and DRM3-2). The DRM with five quadrature points provides the most accurate calculation of the probability of failure (DRM5-1 and DRM5-2). The optimum design points obtained from the DRM are indeed similar to each other compared with those obtained from the FORM for different orderings. Thus, the DRM is necessary to reduce the effect of transformation ordering and to provide accurate RBDO results.

If the number of correlated variables is larger than two, the effect of transformation ordering and inaccurate estimation of probability of failure might be more significant, and thus using the FORM in the reliability analysis might be more inaccurate. In the next section, this issue will be further addressed through a four-dimensional problem.

Table 5.3. RBDO Results Using Frank Copula ($P_F^{Tar} = 2.275\%$)

Case	Cost	Optimum design points	$\left\|d_1^{opt} - d_2^{opt}\right\|$	P_{F_1} (%)	P_{F_2} (%)
FORM-1	3.590	1.477, 2.114	0.036	1.765	2.287
FORM-2	3.572	1.491, 2.081		2.351	1.638
DRM3-1	3.541	1.455, 2.086	0.015	2.171	2.255
DRM3-2	3.551	1.469, 2.081		2.323	1.953
DRM5-1	3.535	1.453, 2.082	0.006	2.242	2.270
DRM5-2	3.539	1.459 2.080		2.285	2.093

5.3.2 Four-Dimensional Example

This example is the four-dimensional modified Rosen-Suzuki problem (Schittkowski, 1987), and the RBDO is formulated to

$$
\begin{aligned}
\text{min.} \quad & \text{cost}(\mathbf{d}) = \frac{d_1(d_1-15)+d_2(d_2-15)+d_3(2d_3-41)+d_4(d_4-3)+245}{245} \\
\text{s.t.} \quad & P\left(G_i(\mathbf{X}) \geq 0\right) \leq P_F^{Tar},\ i=1,2,3 \\
& 0 \leq d_1, d_2, d_3, d_4 \leq 10,\ P_F^{Tar} = 2.275\% \\
& G_1(\mathbf{X}) = 1 - \frac{X_1(9-X_1)+X_2(11-X_2)+X_3(11-X_3)+X_4(11-X_4)}{68} \\
& G_2(\mathbf{X}) = 1 - \frac{X_1(11-X_1)+2X_2(10-X_2)+X_3(10-X_3)+X_4(21-2X_4)}{151} \\
& G_3(\mathbf{X}) = 1 - \frac{2X_1(9-X_1)+X_2(11-X_2)+X_3(10-X_3)+X_4}{95}
\end{aligned}
\tag{5.24}
$$

Assume that the first and second variables are correlated with the Gumbel copula and the third and fourth variables are correlated with the A12 copula, in which the Gumbel and A12 copula belong to the Archimedean copula. The Gumbel copula is defined as

$$
C\left(z_1, z_2 \middle| \theta\right) = \exp\left\{-\left[\left(-\ln z_1\right)^{\theta} + \left(-\ln z_2\right)^{\theta}\right]^{1/\theta}\right\} \tag{5.25}
$$

where $z_1 = \Phi(x_1)$ and $z_2 = \Phi(x_2)$ with $X_1, X_2 \sim N(5.0, 0.3^2)$. Kendall's tau $\tau = 0.5$ is assumed for both copulas, and the correlation parameter is obtained as $\theta = \frac{1}{(1-\tau)}$. The A12 copula is defined as

$$
C\left(z_3, z_4 \middle| \theta\right) = \left\{1 + \left[\left(z_3^{-1}-1\right)^{\theta} + \left(z_4^{-1}-1\right)^{\theta}\right]^{1/\theta}\right\}^{-1} \tag{5.26}
$$

Likewise, $z_3 = \Phi(x_3)$, $z_4 = \Phi(x_4)$ with $X_3, X_4 \sim N(5.0, 0.3^2)$ and $\theta = \frac{2}{3(1-\tau)}$.

Since the number of input variables is four and two pairs of variables are correlated, four different orderings are possible in the transformation. In Table 5.4, FORM-1 indicates FORM with the initial ordering, which means the ordering is not changed. FORM-2 is the case with the interchanged ordering of x_1 and x_2, and FORM-3 is the one with interchanged ordering of x_3 and x_4. FORM-4 is the case where the orderings of all variables are interchanged, which means x_1 and x_2 are interchanged and x_3 and x_4 are interchanged. As shown in Table 5.4, for all orderings, the probabilities of failure P_{F_2} and P_{F_3} are poorly estimated when the FORM is used. Even though the calculation of probability of failure for the fourth ordering is most accurate, P_{F_2} and P_{F_3} are still much larger than the target probability, $P_F^{Tar} = 2.275\%$. Compared with the two-dimensional example, the FORM error for the four-dimensional case is more significant. When the MPP-based DRM with three quadrature points is used (DRM3-1, 2, 3, and 4), the difference between the probabilities of failure becomes smaller than when the FORM is used. When five quadrature points are used (DRM5-1, 2, 3, and 4), the MPP-based DRM estimates the probabilities of the failure more accurately than the case with three quadrature points. Thus, the MPP-based DRM is necessary to reduce the ordering effect on RBDO results.

With the accuracy of the inverse reliability analysis using the DRM even for highly nonlinear constraint functions, it is shown that the RBDO results are becoming less dependent on the Rosenblatt transformation ordering of the input variables.

Table 5.4. RBDO Results Using Gumbel and A12 Copula

Case	Cost	Optimum design points	P_{F_2} (%)	P_{F_3} (%)
FORM-1	– 0.144	4.620, 5.623, 6.618, 4.024	9.635	4.242
FORM-2	– 0.144	4.644, 5.591, 6.618, 4.020	8.192	4.181
FORM-3	– 0.145	4.634, 5.561, 6.639, 4.065	7.110	4.022
FORM-4	– 0.145	4.644, 5.584, 6.646, 4.072	6.298	4.072
DRM3-1	– 0.139	4.636, 5.586, 6.574, 4.119	3.000	2.605
DRM3-2	– 0.139	4.636, 5.561, 6.582, 4.123	2.701	2.612
DRM3-3	– 0.139	4.611, 5.564, 6.592, 4.131	3.109	2.472
DRM3-4	– 0.139	4.619, 5.545, 6.590, 4.128	2.790	2.252
DRM5-1	– 0.139	4.637, 5.584, 6.573, 4.124	2.903	2.605
DRM5-2	– 0.139	4.640, 5.558, 6.582, 4.125	2.617	2.637
DRM5-3	– 0.139	4.610, 5.554, 6.593, 4.149	2.468	2.306
DRM5-4	– 0.138	4.612, 5.544, 6.591, 4.149	2.359	2.133

Note: Method-1: Ordering 1 (original ordering).

Method-2: Ordering 2 ($X_1 \leftrightarrow X_2$).

Method-3: Ordering 3 ($X_3 \leftrightarrow X_4$).

Method-4: Ordering 4 ($X_1 \leftrightarrow X_2$ and $X_3 \leftrightarrow X_4$).

CHAPTER VI

APPLICATIONS OF COPULA, BAYESIAN METHOD, AND DIMENSION REDUCTION METHOD TO RELIABILITY-BASED DESIGN OTIMIZATION PROBLEMS

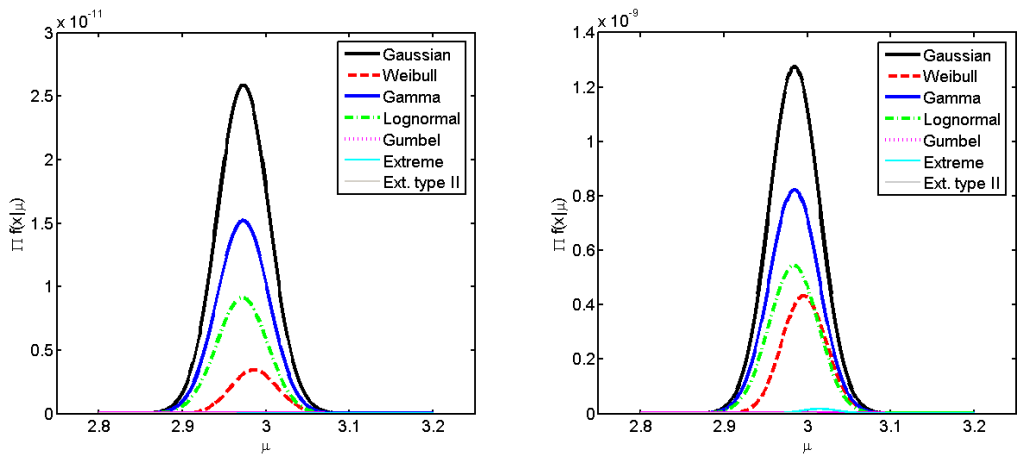
6.1 Introduction

A copula, which is used to model a joint distribution, was illustrated in Chapter 2. Using the copula concept, Chapter 3 explained how an input model, i.e., the marginal and joint distribution modeled by the copula, is identified using the Bayesian method. In Chapter 4, the basic concept of RBDO was introduced, and, for accurate reliability analysis, the MPP-based DRM was explained in Chapter 5. In this chapter, to test the proposed methods together, random samples are generated from a true input model. Using the generated samples, the input models are identified and quantified using the Bayesian method, and then MPP-based DRM for RBDO is carried out. Two examples will illustrate how the combined proposed methods work.

6.2 Mathematical Example

Suppose that true marginal distributions of two input random variables are Gaussian distributions, X_1 and $X_2 \sim N\left(3, 0.3^2\right)$ with a joint CDF modeled by the Frank copula. The Kendall's tau is given as $\tau = 0.8$. From the true input marginal and joint distribution, 100 random samples are generated. It is shown how the Bayesian method identifies the correct marginal and joint CDFs from the samples.

Using the samples, the likelihood functions for candidate marginal distributions for X_1 and X_2 are obtained. As shown in Figure 6.1 (a), the Gaussian distribution has the highest peak, and it is widely spread, which means the normalized weight of the Gaussian distribution is the highest among candidate marginal distributions, as shown in Table 6.1.



(a) X_1 (b) X_2

Figure 6.1. Likelihood Functions of μ for Mathematical Example

Table 6.1. Normalized Weights of Candidate Marginal CDFs for Mathematical Example

Random Variables	Gaussian	Weibull	Gamma	Lognormal	Gumbel	Extreme	Extreme type -II
X_1	0.480	0.081	0.277	0.160	0.000	0.001	0.000
X_2	0.405	0.163	0.258	0.167	0.000	0.001	0.000

Thus, the Bayesian method identifies the Gaussian distribution, which is the true marginal distribution type of X_1. Likewise, for the second variable X_2 in Figure 6.1 (b), the Gaussian distribution has the highest normalized weights in Table 6.1. Thus, the Bayesian method correctly identifies the true marginal distribution of X_2.

Using the identified marginal CDFs, a copula can be identified using the Bayesian method. Figure 6.2 shows the likelihood function of τ for candidate copulas. Since the Frank copula has a distinct shape among candidate copulas, the Frank copula has the highest weight among candidate copulas and the normalized weight is very high, 0.999 as shown in Table 6.2. Thus, the copula is correctly identified. According to the joint PDF contour modeled by the Frank copula shown in Figure 6.3, the Frank copula well describes the given data.

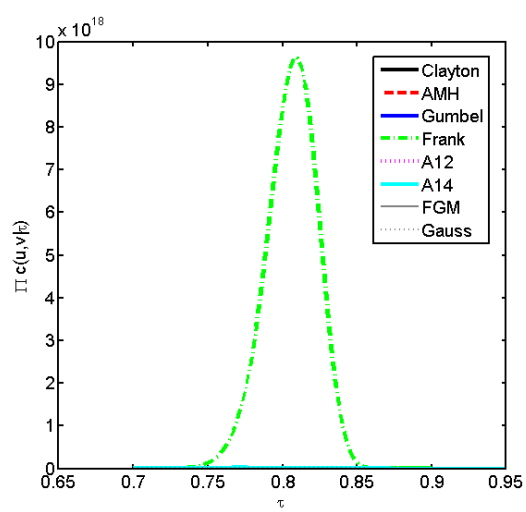


Figure 6.2. Likelihood Functions of τ for Mathematical Example

Table 6.2. Normalized Weights of Candidate Copulas for Mathematical Example

Clayton	AMH	Gumbel	Frank	A12	A14	FGM	Gauss
0.000	0.000	0.000	0.999	0.000	0.001	0.000	0.000

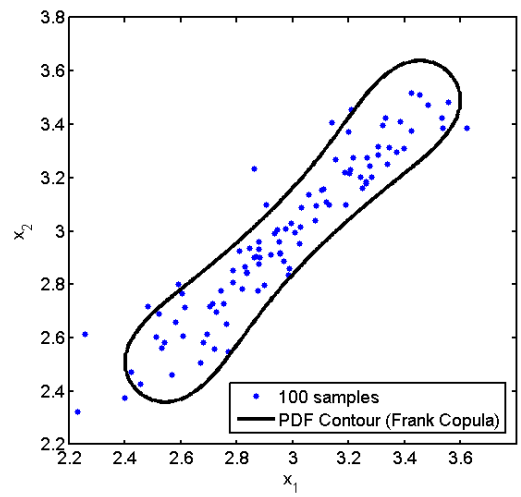


Figure 6.3. Scatter Plot of Data and PDF Contour of Frank Copula

Once the marginal distributions and copula are identified, their parameters need to be calculated from the samples. The means and standard deviations values are calculated as $\mu_1 = 2.973$ and $\sigma_1 = 0.310$ for X_1, and $\mu_2 = 2.985$ and $\sigma_2 = 0.298$ for X_2, where the Kendall's tau is obtained as 0.808 from the samples.

To observe the effect of correlation on RBDO results, the estimated input model without considering the correlation, i.e. $\tau = 0.0$, is first tested, and compared with the estimated model with considering the correlation, $\tau = 0.808$, using the FORM. Second, using the estimated model with $\tau = 0.808$, the FORM and MPP-based DRM are tested to see how the MPP-based DRM yields more accurate optimum design than the FORM does.

An RBDO formulation is defined to

$$
\begin{aligned}
&\text{minimize} \quad \text{cost}(\mathbf{d}) = d_1 + d_2 \\
&\text{subject to} \quad P\left(G_i\left(\mathbf{X}\right) \geq 0\right) \leq P_F^{Tar} (= 2.275\%),\ i = 1,2,3 \\
&\qquad \mathbf{d} = \boldsymbol{\mu}\left(\mathbf{X}\right),\ 0 \leq d_1, d_2 \leq 10 \\
&G_1(\mathbf{X}) = 1 - \left(0.9010X_1 - 0.4339X_2 + 1.5\right)^2 \times \left(0.4339X_1 + 0.9010X_2 + 2\right)/20 \\
&G_2(\mathbf{X}) = 1 - \left(X_1 + X_2 - 2.8\right)^2/30 - \left(X_1 - X_2 - 12\right)^2/120 \\
&G_3(\mathbf{X}) = 1 - 200/\left\{2.5(X_1 - 3)^2 + 8X_2 + 5\right\}
\end{aligned}
\tag{6.1}
$$

Figure 6.4 shows the β_t-contours and optimum designs obtained from three input models – the estimated input models with and without correlation, and the true input model.

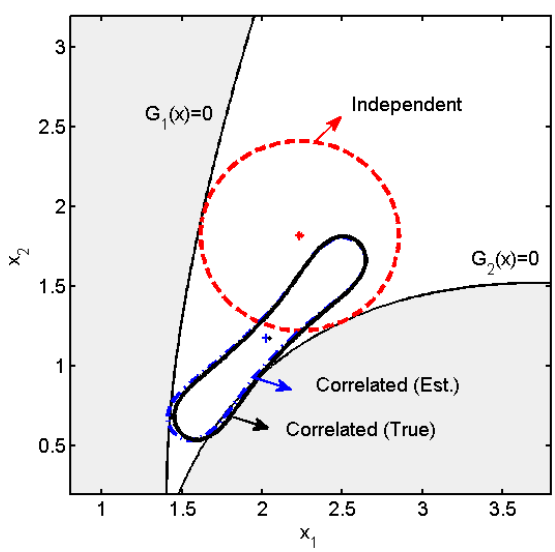


Figure 6.4. β_t-contours Using Different Input Models for Mathematical Example

First, the RBDO results for the estimated input models with and without considering correlation are compared using the FORM. As shown in Figure 6.4, the

estimated input model without correlation, $\tau = 0.0$, yields a β_t-contour (dashed contour), which is rather different from the one using the true input model (solid contour). Accordingly, the obtained optimum design, '*', is also very different from the true optimum, '·'. On the other hand, the estimated input model with correlation, $\tau = 0.808$, yields a β_t-contour, (dashed-dotted contour), which is similar to the one using the true input model (solid contour). Thus, the obtained optimum design, '+', is very close to the true optimum, '·'.

Table 6.3 shows the obtained optimum designs and the probabilities of failure evaluated at those designs using MCS with the true input model. As explained in Figure 6.4, the estimated input model without correlation yields the inaccurate optimum design with high cost while the one with correlation yields more accurate optimum design with the reduced cost. Further, the probabilities of failure for two active constraints (P_{F_1} and P_{F_2}) using the estimated input model without correlation are much smaller than the target probability of failure (2.275%), i.e., the optimum designs are much conservative. On the other hand, the probabilities of failure using the estimated input model with correlation are close to the target probability of failure, but still smaller than 2.275%. Thus, in this example, the estimated input model without correlation yields an unnecessarily conservative design, whereas the input model with correlation yields a reliable design, but with significant cost reduction.

Second, since two input variables are correlated with Frank copula, which is non-Gaussian copula, the Rosenblatt transformation might become highly nonlinear as explained in Chapter 5. In this case, if the FORM is used, the inverse reliability analysis

results could be erroneous. Thus, the RBDO results using FORM and DRM are compared for the estimated input model with $\tau = 0.808$. As shown in Table 6.3, the optimum design using the DRM is closer to the true optimum than the one using the FORM. Accordingly, the probabilities of failure using the DRM , $P_{F_1} = 2.235\%$ and $P_{F_2} = 1.794\%$ for two active constraints, are closer to the target probability of failure, 2.275%, than those using FORM, $P_{F_1} = 1.906\%$ and $P_{F_2} = 1.395\%$ as displayed in Table 6.3. Therefore, it is necessary to consider the correlation in RBDO and use a very accurate reliability analysis such as the MPP-based DRM.

Table 6.3. RBDO Comparisons for Mathematical Example ($P_F^{Tar} = 2.275\%$)

	FORM		DRM	
	Independent	Correlated (Est.)	Correlated (Est.)	True
d_1	2.233	2.051	2.023	2.026
d_2	1.835	1.222	1.175	1.163
P_{F_1}	0.746%	1.906%	2.235%	2.245%
P_{F_2}	0.019%	1.395%	1.794%	2.070%
Cost	4.068	3.273	3.203	3.189

6.3 Coil Spring Problem

The design objective of the coil spring is to minimize the volume to carry a given axial load such that the design satisfies the minimum deflection and allowable shear stress requirement, and the surge wave frequency is above the lower limit (Arora, 2004). The volume of the spring is defined as

$$\text{volume} = \frac{1}{4}\left(N+Q\right)\pi^2(D+d)d^2 \tag{6.2}$$

where N is the number of active coils; Q is the number of inactive coils, given as $Q=2$; D is the mean inner diameter; d is the wire diameter as shown in Figure 6.5.

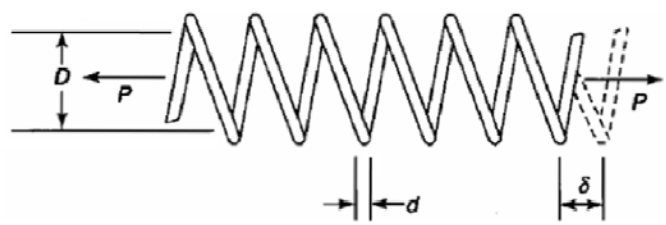


Figure 6.5. Coil Spring

The first constraint is that the deflection δ under the applied load P should be at least the minimum spring deflection, Δ as

$$\delta = \frac{8P\left(D+d\right)^3 N}{d^4 G} \geq \Delta \tag{6.3}$$

where $P = 10lb$, $\Delta = 0.5in.$, and G is the shear modulus. The second constraint is that the shear stress in the wire should not be larger than the allowable shear stress, τ_a, which is formulated as

$$\tau = \frac{8P(D+d)}{\pi d^3}\left(\frac{4(D+d)-d}{4D} + \frac{0.615d}{D+d}\right) \leq \tau_a \tag{6.4}$$

where $\tau_a = 80{,}000lb/in.^2$. The third constraint requires that the surge wave frequency of the spring should be higher than the lower limit of surge wave frequency ω_0 as

$$\omega = \frac{d}{2\pi N(D+d)^2}\sqrt{\frac{G}{2\rho}} \geq \omega_0 \tag{6.5}$$

where w_0 =100Hz , and ρ is the mass density.

In this example, the number of active coils(N), mean inner diameter (D), and wire diameter(d) are selected as random variables, and mass density of material(ρ) and shear modulus(G) are selected as random parameters. Using the constraints for the coil spring problem, the RBDO formulation is defined as

$$\begin{aligned}
&\text{minimize} \quad \text{cost}(\boldsymbol{d}) = \frac{1}{4}(d_1+Q)\pi^2(d_2+d_3)d_3^2 \\
&\text{subject to} \quad P(G_i(\mathbf{X}) \geq 0) \leq P_F^{Tar} = 2.275\%, \quad i = 1,2,3 \\
&G_1(\mathbf{X}) = 1.0 - \frac{8P(X_2+X_3)^3 X_1}{X_3^4 X_5 \Delta} \\
&G_2(\mathbf{X}) = -1.0 + \frac{8P(X_2+X_3)}{\pi X_3^3 \tau_a}\left[\frac{(4X_2+3X_3)}{4X_2} + \frac{0.615X_3}{(X_2+X_3)}\right] \\
&G_3(\mathbf{X}) = 1.0 - \frac{X_3}{2\pi X_1 (X_2+X_3)^2 w_0}\sqrt{\frac{X_5}{2X_4}}
\end{aligned} \tag{6.6}$$

Table 6.4 shows statistical information of the random variables and parameters and their design bounds. The coefficient of variations (COVs), ratios of standard deviation to mean, for two material properties are referred to Schuëller (2007). Since X_4 and X_5 are random parameters, the mean values of two parameters do not change during RBDO. In the manufacturing process, the coil inner diameter (X_2) and the wire diameter (X_3) are correlated, and thus the correlation coefficient between those two variables is assumed to be $\tau = 0.7$ where Clayton copula is used to model the joint distribution of X_2 and X_3 .

Table 6.4. Properties of Random Variables and Parameters for Coil Spring Example

Random Variables and Parameters			Dist. Type	d_L	d	d_U	Std.	COV
Variables	X_1	No. of active coils (no.)	Gaussian	1.0	10.0	100.0	0.5	5%
	X_2	Inner diameter (in.)	Gaussian	0.1	1.0	5.0	0.05	5%
	X_3	Wire diameter (in.)	Gaussian	0.01	0.1	0.5	0.005	5%
Parameters	X_4	Mass density (lb·sec^2/in.^{4})	Gaussian		7.38E-4		2.95E-5	4%
	X_5	Shear modulus (lb/in.^{2})	Gaussian		1.15E+7		1.38E+6	12%

Suppose that the true model has the statistical data as shown in Table 6.4. Using the randomly generated 100 samples from the true model, marginal distributions are first identified using the Bayesian method. Table 6.5 shows the normalized weights of the candidate marginal distributions of all input variables. The marginal distributions of X_1, X_2, X_3, and X_5 are correctly identified as Gaussian distribution, but the marginal distribution of X_4 is identified as a lognormal distribution because the lognormal distribution shape is similar to the Gaussian distribution shape for COV=4%. Therefore, the normalized weights of the Gaussian, Gamma, and lognormal distributions, which have similar PDF shapes, are relatively higher than those of other distributions. Likewise, a copula is identified as shown in Table 6.6. Since the Clayton copula has a very distinct shape among candidates, it is easy to identify, and as a result, the normalized weight for the Clayton copula is very high.

Table 6.5. Normalized Weights of Candidate Marginal CDFs for Coil Spring Example

Random Variables	Gaussian	Weibull	Gamma	Lognormal	Gumbel	Extreme	Extreme type -II
X_1	0.395	0.000	0.330	0.275	0.000	0.000	0.000
X_2	0.425	0.001	0.317	0.256	0.000	0.000	0.000
X_3	0.440	0.002	0.313	0.246	0.000	0.000	0.000
X_4	0.220	0.000	0.361	0.420	0.000	0.000	0.000
X_5	0.495	0.014	0.309	0.182	0.000	0.000	0.000

Table 6.6. Normalized Weights of Candidate Copulas for Coil Spring Example

Clayton	AMH	Gumbel	Frank	A12	A14	FGM	Gauss
1.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000

Once the marginal CDFs and copula are identified, the parameters of the estimated input model need to be estimated. The mean and standard deviations for all input variables are obtained from the 100 samples as shown in Table 6.7. In this case, the mean and standard deviations of the obtained input model are accurately estimated as compared them with those of the true input model in Table 6.4. Further, the estimated correlation coefficient between X_2 and X_3 is calculated as 0.808, which is also close to the true correlation coefficient, 0.8.

Table 6.7. Distribution Type and Parameters of Estimated Input Model for Coil Spring Example

Random Variables	Distribution Type	Mean	Std.
X_1	Gaussian	0.992E+1	0.541E+0
X_2	Gaussian	0.999E+0	0.539E-1
X_3	Gaussian	0.998E-1	0.530E-2
X_4	Lognormal	0.738E-3	0.262E-5
X_5	Gaussian	0.116E+8	0.134E+7

To compare the input models with and without correlation, first, an estimated model without correlation $\tau = 0.0$ and the one with estimated correlation $\tau = 0.808$ are tested using the FORM as shown in Table 6.8. Similar to the mathematical example, the estimated input model with $\tau = 0.0$ yields a wrong optimum design. On the other hand, the estimated input model with $\tau = 0.808$ more approaches the true optimum than the one with $\tau = 0.0$. Further, the estimated input model with $\tau = 0.0$ yields much conservative design with a high cost whereas the one with $\tau = 0.808$ yields a reliable optimum design with a reduced cost.

Second, for the estimated input model with $\tau = 0.808$, the RBDO results using the FORM and the DRM are compared. Since the Rosenblatt transformation becomes nonlinear for the non-Gaussian copula (Clayton), the obtained optimum design using the DRM is closer to the true optimum than the one using the FORM. In addition, the probabilities of failure using the DRM are closer to the target probability of failure (2.275%) than those using the FORM. Thus, for RBDO problems with correlated

variables, it is very important to consider the correlation and use the MPP-based DRM for reliable optimum design.

Table 6.8. RBDO Comparisons for Coil Spring Example ($P_F^{Tar} = 2.275\%$)

	FORM		DRM	
	Independent	Correlated (Est.)	Correlated (Est.)	True
d_1	0.556E+1	0.137E+2	0.147E+2	0.102E+2
d_2	0.956E+0	0.460E+0	0.426E+0	0.522E+0
d_3	0.820E-1	0.643E-1	0.631E-1	0.665E-1
P_{F_1}	0.161%	1.025%	1.773%	2.151%
P_{F_2}	0.218%	1.829%	1.703%	2.106%
Cost	0.130E+0	0.845E-1	0.802E-1	0.785E-1

6.4 Roadarm Example

A roadarm of the army tracked vehicle, the M1A1 tank shown in Figure 6.6, is used to demonstrate the applicability of the Bayesian method and MPP-based DRM for RBDO. The roadarm is modeled using 1572 eight-node isoparametric finite elements (SOLID45) and four beam elements (BEAM44) of a commercial program, ANSYS (Swanson, 1989), as shown in Figure 6.7. The material of the roadarm is S4340 steel with Young's modulus E=3.0×10^7 psi and Poisson's ratio ν=0.3. The durability analysis of the roadarm is carried out to obtain the fatigue life contour using Durability and Reliability Analysis Workspace (DRAW) (Center for Computer-Aided Design, 1999a; Center for Computer-Aided Design, 1999b). The fatigue lives at the critical nodes are selected as design constraints of the RBDO in Figure 6.8.

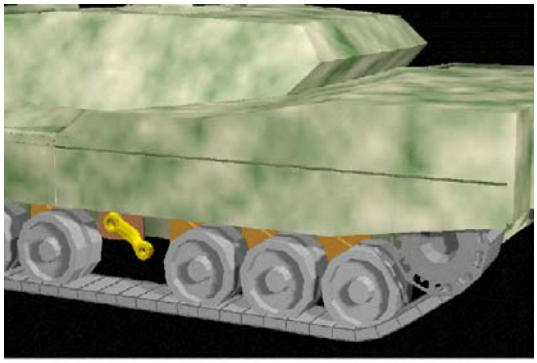


Figure 6.6. Army Tracked Vehicle

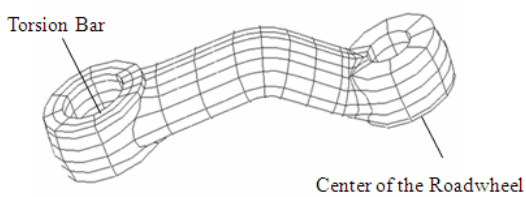


Figure 6.7. Finite Element Model of Roadarm

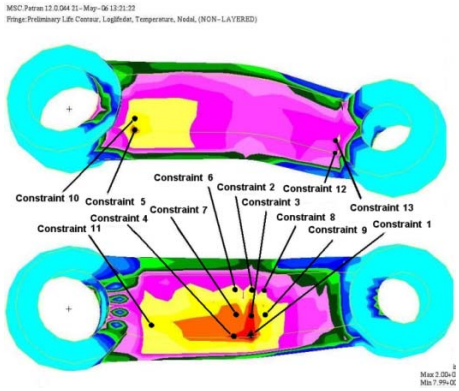


Figure 6.8. Fatigue Life Contour and Critical Nodes of Roadarm

In Figure 6.9, the shape design variables consist of four cross-sectional shapes of the roadarm where the widths (x_1-direction) of the cross-sectional shapes are defined as

design variables d_1, d_3, d_5, and d_7, at intersections 1, 2, 3, and 4, respectively, and the heights (x_3-direction) of the cross-sectional shapes are defined as design variables d_2, d_4, d_6, and d_8.

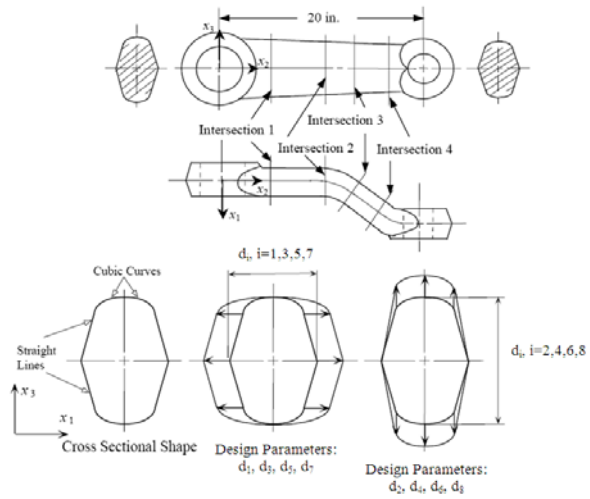


Figure 6.9. Shape Design Variables for Roadarm

Table 6.9 shows the initial design point, and the lower and upper bounds of eight design variables. The eight random variables are assumed to be independent, and 5% of COV are used. For the input fatigue material properties, since the statistical information of S4340 steel is not available other than its nominal value, the statistical information of 950X steel (Socie, 2003) is used to describe the properties of S4340 steel. The COV of SAE 950X is 115% for $\varepsilon_f^{'}$ and 25% for other material properties (Socie, 2003). Since S4340 is a stronger material than SAE 950X, in this study, it is assumed that the COV of S4340 is 50% for $\varepsilon_f^{'}$ and 25% for the other material properties as shown in Table 6.9.

Table 6.9. Random Variables and Fatigue Material Properties for Roadarm Example

Random Variables	Lower Bound $\mathbf{d}^L$	Initial Design $\mathbf{d}^0$	Upper Bound $\mathbf{d}^U$	Std.	Dist. Type
d_1	1.3500	1.7500	2.1500	0.0875	Gaussian
d_2	2.6496	3.2496	3.7496	0.1625	Gaussian
d_3	1.3500	1.7500	2.1500	0.0875	Gaussian
d_4	2.5703	3.1703	3.6703	0.1585	Gaussian
d_5	1.3563	1.7563	2.1563	0.0878	Gaussian
d_6	2.4377	3.0377	3.5377	0.1519	Gaussian
d_7	1.3517	1.7517	2.1517	0.0876	Gaussian
d_8	2.5085	2.9085	3.4085	0.1454	Gaussian

Fatigue Material Properties

Random Parameters	Mean	Std.	COV	Dist. Type
Fatigue Strength Coefficient, σ'_f	177000	44250	25%	Lognormal
Fatigue Strength Exponent, b	−0.073	0.018	25%	Gaussian
Fatigue Ductility Coefficient, ε'_f	0.410	0.205	50%	Lognormal
Fatigue Ductility Exponent, c	−0.600	0.150	25%	Gaussian

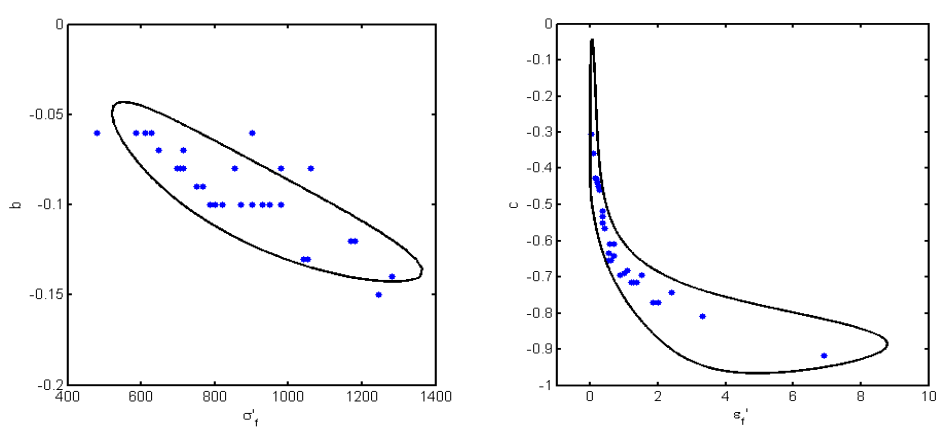
From Socie's study on 950X steel (Socie, 2003), it is shown that σ'_f and b have highly negative correlation, −0.828, and that ε'_f and c also have highly negative correlation, −0.976. Since the marginal distributions of those random parameters are known, copulas only need to be identified for σ'_f and b, and for ε'_f and c, respectively. However, the experimental data of S4340, which is used in the roadarm, is not available, so 29 experimental paired data obtained from SAE 950X steel are used to identify copulas for two pairs of variables, σ'_f and b, and ε'_f and c.

Before identifying the copulas for two pairs of variables, candidate copulas need to be determined. In this example, since two pairs of variables are highly negatively

correlated, only the Frank and Gaussian are selected as candidate copulas because other copulas such as Clayton and Gumbel cannot have negative correlation and some copulas such as AMH and FGM can only have large negative correlation. Using the Bayesian method, the Gaussian copula for $\sigma_f^{'}$ and b and the Frank copula for $\varepsilon_f^{'}$ and c are identified based on given marginal CDFs and data as shown in Table 6.10. As shown in Figure 6.10, the PDF contours of two identified copulas well describe the given experimental data.

Table 6.10 Normalized Weights of Candidate Copulas Using 29 Paired Data of SAE 950X Steel

Material Properties	Frank	Gauss
$\sigma_f^{'}$, b	0.355	0.645
$\varepsilon_f^{'}$, c	0.607	0.393



(a) $\sigma_f^{'}$ and b (b) $\varepsilon_f^{'}$ and c

Figure 6.10. Joint PDF Contours of Gaussian and Frank Copulas Identified from 29 Paired Data of 950X Steel

Using the identified copulas and known marginal CDFs, the RBDO is carried out. The formulation of the roadarm is defined as

$$
\begin{aligned}
&\text{minimize} \quad \text{cost}(\mathbf{d}) \\
&\text{subject to} \quad P(G_i(\mathbf{X}) \geq 0) \leq P_F^{Tar},\ i = 1 \sim 13 \\
&\mathbf{d} = \boldsymbol{\mu}(\mathbf{X}),\ \mathbf{d}_L \leq \mathbf{d} \leq \mathbf{d}_U,\ \mathbf{d} \in R^n, P_F^{Tar} = 2.275\% \\
&G_i(\mathbf{X}) = 1 - \frac{L(\mathbf{X})}{L_t},\ i = 1 \sim 13 \\
&\text{cost}(\mathbf{d}) : \text{Weight of Roadarm} \\
&L(\mathbf{X}) : \text{Crack Initiation Fatigue Life,} \\
&L_t : \text{Crack Initiation Target Fatigue Life (=5 years)}
\end{aligned}
\tag{6.7}
$$

First, the input random variables are assumed to be independent to carry out RBDO. Second, the correlation of input fatigue material properties is used to test the applicability of the copula using the FORM. Finally, the MPP-based DRM with correlated input variables is carried out to obtain more accurate optimum design.

As shown in Table 6.11, the FORM assuming that input variables are independent yields a wrong optimum design with increased weight of the roadarm, 592.22. When the correlation is considered, the weight of the roadarm is significantly reduced as, 519.70, compared to the independent case. Using the MPP-based DRM, the weight of the roadarm is further reduced to 514.02 while satisfying the target probability of failure. Therefore, RBDO results can be significantly improved by using copulas to obtain a better estimation of the joint CDFs; and MPP-based DRM to obtain a better optimum design.

In the mathematical and coil spring example, the generated input data well describe the true input model; thus, the obtained optimum design is close to the true optimum design. However, when the number of samples is very limited, for example,

when the 29 paired data of 950X steel is only available in the roadarm, the identification and quantification error could be significant. In the next section, to offset inaccurate estimation of the input model, the input model with confidence level is implemented for RBDO using the adjusted parameters with confidence intervals of mean, standard deviation, and correlation coefficient.

Table 6.11. RBDO Comparisons for Roadarm Example

	Initial	D.O*	FORM		DRM
			Independent	Correlated	
d_1	1.750	1.588	2.194	1.958	1.928
d_2	3.250	2.650	2.650	2.650	2.650
d_3	1.750	1.922	2.602	2.031	2.067
d_4	3.170	2.570	3.010	2.670	2.577
d_5	1.756	1.477	2.656	1.775	1.776
d_6	3.038	3.292	2.538	3.538	3.535
d_7	1.752	1.630	2.422	2.152	2.075
d_8	2.908	2.508	2.895	2.536	2.512
Cost	515.09	464.56	592.22	519.70	514.02

*D.O means deterministic optimum

CHAPTER VII

RELIABILITY-BASED DESIGN OPTIMIZATION WITH CONFIDENCE LEVEL UNDER INPUT MODEL UNCERTAINTY

7.1 Introduction

The input model is a collection of marginal and joint distributions of input random variables with their associated parameters. When it is uncertain what types of marginal and joint distributions and parameters should be used for the input model, the input model uncertainties arise. Input model uncertainties consist of input statistical uncertainty and input physical uncertainty. The input statistical uncertainty is associated with identification of marginal and joint distribution; the input physical uncertainty is associated with quantification of input parameters of the identified marginal and joint distributions such as mean, standard deviation, and correlation coefficient. In industrial applications, since often limited data on input variables is available due to expensive experimental testing costs, it is difficult to correctly model the input model uncertainties. If the input model uncertainties are not correctly modeled from given data, the estimated input model will yield incorrect reliability-based optimum design.

Therefore, to offset the inexact identification and quantification of the input model uncertainties, an input model with a confidence level is developed for use in RBDO. For this, instead of using the estimated parameters obtained from given limited data, adjusted parameters calculated using confidence intervals of the estimated parameters are proposed. For the calculation of the confidence intervals, a method using normality assumption on an input variable and bootstrap methods are described. To check

whether use of the adjusted parameters provides the desirable confidence level of the output performance, i.e., the target probability of failure, the β_t-contour with target reliability index is used to measure the confidence level of the input model. If the β_t-contour covers the true β_t-contour, then it will yield a reliable optimum design regardless of the location of MPP on the estimated β_t-contour.

Section 7.2 explains why the β_t-contour is used to measure input confidence levels in detail, and how confidence intervals of the input parameters affect the shape of the β_t-contour. Section 7.3 illustrates how to calculate confidence intervals of the input parameters. In Section 7.4, the adjusted parameters using the confidence intervals of the input parameters are introduced to have desirable input confidence level. Section 7.5 shows simulation results for the input confidence levels for different number of samples and parameters. Section 7.6 provides numerical examples to show how the input model with adjusted parameters yields reliable optimum designs.

7.2 Measure of Input Confidence Level

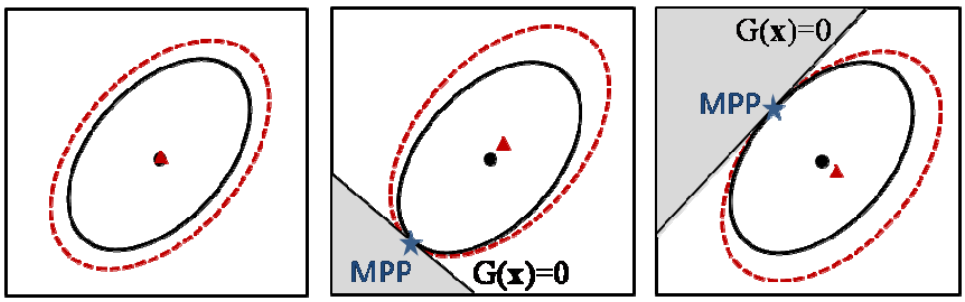
When an input model is estimated from given limited data, we are concerned how much confidence level of output performance the estimated input model provides. However, it is difficult to predict the accurate confidence level of output performance of the RBDO result because the confidence level of output performance can be different for different RBDO problems according to the location of MPP on the β_t-contour at the optimum design even though the same input model is used. Thus, the confidence level of the input model needs to be first estimated before stepping into the estimation of the

confidence of output performance. Even though the confidence level of the input model is not necessarily equivalent to the output confidence level, if a conservative measure for calculating the input confidence level, i.e., a β_t-contour is used, then it can be assured that the confidence level of output performance is at least larger than the confidence level of input model.

Figure 7.1 (a) shows an estimated β_t-contour (dashed contour) that fully covers a true β_t-contour (solid contour). For different shapes of constraint functions as shown in Figures 7.1 (b) and (c), the optimum designs using the estimated β_t-contour and the true β_t-contour are obtained as the triangular mark and the circular dot, respectively, where the gray region indicates the failure region, i.e., $G(\mathbf{x}) > 0$. Since the obtained optimum (triangular mark) is farther away from the active constraint function than the true optimum as shown in Figures 7.1 (b) and 7.1 (c), the probabilities of failure evaluated at the obtained optimum designs are smaller than the one at the true optimum, which means that the estimated β_t-contour yields reliable optimum designs.

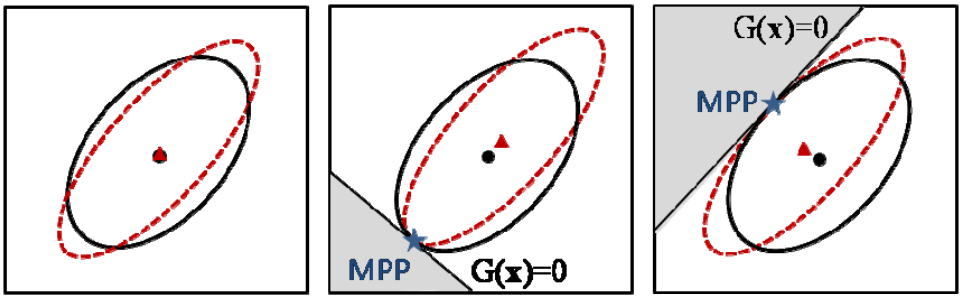
However, when the estimated β_t-contour does not fully cover the true β_t-contour as shown in Figure 7.2 (a), it could yield a reliable optimum design, or not. In Figure 7.2 (b), the optimum design for the estimated β_t-contour is reliable because it is farther away from the active constraint function than the true optimum. However, for the different shape of constraint function in Figure 7.2 (c), the optimum design for the estimated β_t-contour is not reliable because it is closer to the active constraint than the true optimum.

Thus, the estimated β_t-contour that fully covers the true β_t-contour ensures that the obtained optimum design is always reliable.



(a) β_t-contours (b) Case 1 (c) Case 2

Figure 7.1. Estimated β_t-contour Fully Covering True β_t-contour

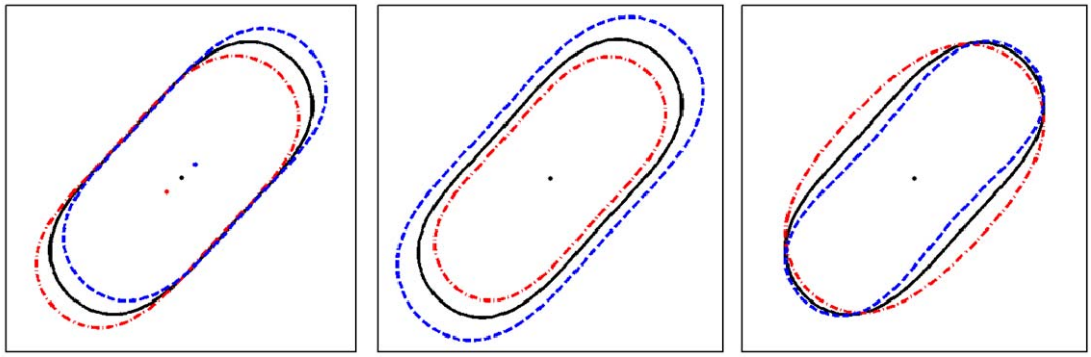


(a) β_t-contours (b) Case 1 (c) Case 2

Figure 7.2. Estimated β_t-contours Partially Covering True β_t–contour

To obtain the estimated β_t-contour covering the true β_t-contour, it is necessary to know how the input parameters affect the β_t-contour shapes. Figure 7.3 shows the β_t-contours using the true parameters and upper and lower bounds of confidence intervals of

mean, standard deviation, and correlation coefficient, indicated as solid, dashed and dashed-dotted contours, respectively. In these figures, when the upper and lower bounds of one parameter are considered, the other two parameters are assumed to be exact.



(a) Mean (b) Standard Deviation (c) Correlation Coefficient

Figure 7.3. β_t-contours Using Lower and Upper Bounds of the Confidence Interval of Mean, Standard Deviation, and Correlation Coefficient

In Figure 7.3 (a), the upper and lower bounds of the confidence interval of the mean are related to the position of the β_t-contour. As shown in the figure, neither the upper nor the lower bounds of the confidence interval of the mean makes the estimated β_t-contour fully cover the true β_t-contour. On the other hand, the upper bound of the confidence interval of the standard deviation makes the estimated β_t-contour fully covers the true β_t-contour as shown in Figure 7.3 (b). Figure 7.3 (c) shows the β_t-contours using the lower and upper bounds of the confidence interval of the correlation coefficient. Like the mean, neither the lower nor the upper bound of the confidence interval of the correlation coefficient makes the estimated β_t-contour fully covers the true β_t-contour.

Thus, the upper bound of the confidence interval of the standard deviation can be only used to obtain the β_t-contour covering the true β_t-contour, which will lead to the reliable optimum design.

However, the prediction error in the mean and correlation coefficient still exists in the input model in which only insufficient data is available. Therefore, instead of using the estimated mean and correlation coefficient, adjusted parameters using the confidence intervals of the mean and correlation coefficient are proposed to enlarge the β_t-contour such that the β_t-contour with the adjusted parameters covers the true β_t-contour, which will lead to a desirable confidence level of input model. In the next section, it will be illustrated how the confidence intervals of the input parameters are calculated.

7.3 Confidence Intervals of Mean, Standard Deviation, and Correlation Coefficient

As stated earlier, the adjusted parameters are used to provide a desirable confidence level for the input model. Since the adjusted parameters are calculated using the confidence intervals of the mean, standard deviation, and correlation coefficient, the confidence intervals of those parameters are described in this section.

7.3.1 Confidence Interval of Mean

Once marginal and joint distribution types are identified using the Bayesian method, it is necessary to evaluate their parameters based on given limited data. To calculate the parameters of the marginal CDFs, the maximum likelihood estimate (MLE)

can be used. However, the MLE might be biased (Baxter, 1980), which means that the expected value of the parameter might not be the same as the parameter being estimated. For instance, the MLE is biased for estimating the variance σ^2 of a Gaussian distribution. On the other hand, a minimum variance unbiased estimator (MVUE), which is a commonly used estimator, is unbiased for estimating the variance σ^2, and the variance of σ^2, $Var\left[\sigma^2\right]$, is smaller than any other unbiased estimators. This estimator was derived by Likes (1969).

The mean and variance that are estimated from given samples are called the sample mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$, respectively. The sample mean is calculated as

$$\tilde{\mu} = \sum_{i=1}^{ns} \frac{x_i}{ns} \tag{7.1}$$

which is the same for the MLE and MVUE. The sample variance for the MLE is calculated as

$$\tilde{\sigma}^2 = \frac{1}{ns} \sum_{i=1}^{ns} \left(x_i - \tilde{\mu}\right)^2 \tag{7.2}$$

and the sample variance for the MVUE is calculated as

$$\tilde{\sigma}^2 = \frac{1}{ns-1} \sum_{i=1}^{ns} \left(x_i - \tilde{\mu}\right)^2 \tag{7.3}$$

To calculate the confidence interval of the mean, suppose that X is a normal random variable with a population mean μ and standard deviation σ. A random sample of size *ns* is collected, i.e., $\mathbf{x} = \left[x_1, \cdots, x_{ns}\right]$, and then the sample mean can be calculated using Eq. (7.1). It is assumed that each sample $x_1, x_2, \cdots, x_{ns}$ comes from a set of random

variables X_1, X_2, $\cdots$, X_{ns} where X_i follows the same distribution as X. Since the sample mean is a random variable, it can be stated as

$$\tilde{\mathrm{M}} = \frac{1}{ns}\sum_{i=1}^{ns} X_i \tag{7.4}$$

The expectation of the sample mean becomes the population mean as

$$E\left(\tilde{\mathrm{M}}\right) = \mu \tag{7.5}$$

The variance of the sample mean is calculated as

$$Var\left(\tilde{\mathrm{M}}\right) = Var\left(\frac{1}{ns}\sum_{i=1}^{ns} X_i\right) = \frac{\sigma^2}{ns} \tag{7.6}$$

When the population standard deviation is unknown, which is the usual case, $\left(\tilde{\mathrm{M}} - \mu\right) / \left(\sigma / \sqrt{ns}\right)$ follows a student's *t*-distribution (Haldar and Mahadevan, 2000). Thus, for the sample mean and standard deviation, the confidence interval for $\left(1-\alpha\right)$ level is obtained as

$$\Pr\left[\tilde{\mu} - t_{\alpha/2,ns-1}\frac{\tilde{\sigma}}{\sqrt{ns}} \le \mu \le \tilde{\mu} + t_{\alpha/2,ns-1}\frac{\tilde{\sigma}}{\sqrt{ns}}\right] = 1-\alpha \tag{7.7}$$

where $t_{\alpha/2,ns-1}$ is the value of a student's *t*-distribution with (*ns*-1) degree of freedom evaluated at the probabilities of α/2. For example, for the 95% confidence level, α = 0.05. Thus, the lower and upper bounds of the confidence interval of the mean, $\tilde{\mu}^L$ and $\tilde{\mu}^U$, are obtained as

$$\tilde{\mu}^L = \tilde{\mu} - t_{\alpha/2,ns-1}\frac{\tilde{\sigma}}{\sqrt{ns}} \text{ and } \tilde{\mu}^U = \tilde{\mu} + t_{\alpha/2,ns-1}\frac{\tilde{\sigma}}{\sqrt{ns}} \tag{7.8}$$

However, since Eq. (7.8) is calculated based on the assumption that the random variable follows a Gaussian distribution, it might be incorrect for a non-Gaussian

distribution. To circumvent this, the bootstrap method, which does not require the normality assumption on the input variable X, can be used. In this study, since the confidence interval of the mean is rather accurately calculated even for the non-Gaussian distribution (Shao and Tu, 1995; Bonate, 2005), only the confidence interval of the standard deviation is tested in this study. The bootstrap method will be further discussed in the next section.

7.3.2 Confidence Interval of Standard Deviation

Suppose that X is a Gaussian random variable and *ns* samples, $x_1, x_2, \cdots, x_{ns}$, are collected to estimate the population variance σ^2, which is an unknown constant. To estimate the population variance σ^2, it is assumed that the samples come from *ns* independent Gaussian random variables, X_1, X_2, $\cdots$, X_{ns}. Using Eq. (7.3), the sample variance can be calculated as (Haldar and Mahadevan, 2000)

$$\Sigma^2 = \frac{1}{ns-1}\sum_{i=1}^{ns}\left(X_i - \tilde{\mathrm{M}}\right)^2 \tag{7.9}$$

where $\tilde{\mathrm{M}}$ is the Gaussian random variable of the sample mean. Multiplying both sides of Eq. (7.9) by (*ns*-1), Eq. (7.9) can be rewritten as

$$\left(ns-1\right)\Sigma^2 = \sum_{i=1}^{ns}\left[\left(X_i-\mu\right)-\left(\tilde{\mathrm{M}}-\mu\right)\right]^2 = \sum_{i=1}^{ns}\left(X_i-\mu\right)^2 - ns\left(\tilde{\mathrm{M}}-\mu\right)^2 \tag{7.10}$$

Dividing both sides of Eq. (7.10) by σ^2, Eq. (7.10) is written as

$$\frac{\left(ns-1\right)\Sigma^2}{\sigma^2} = \sum_{i=1}^{ns}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\tilde{\mathrm{M}}-\mu}{\sigma/\sqrt{ns}}\right)^2 \tag{7.11}$$

Since μ and σ^2 are constant, and the first term on the right side of Eq. (7.11) is a sum of the *ns* squared independent Gaussian variables, it thus has a chi-square distribution with

ns degrees of freedom (Ang and Tang, 1984), denoted as χ^2_{ns}. The second term on the right side has only one squared Gaussian variable and thus has a chi-square distribution with one degree of freedom. Since the sum of two chi-square distributions with *i* and *j* degrees of freedom is also the chi-square distribution with (*i*+*j*) degrees of freedom (Hoel, 1962), the left side of Eq. (7.11) has a chi-square distribution with (*ns*–1) degree of freedom, denoted as χ^2_{ns-1}.

When the PDF of $(ns-1)\tilde{\Sigma}^2/\sigma^2$ has a chi-square distribution with (*ns*–1) degrees of freedom, the two-sided (1–α) confidence interval of the population variance σ^2 is given as

$$\Pr\left[c_{\alpha/2,ns-1} \le \frac{(ns-1)\Sigma^2}{\sigma^2} \le c_{1-\alpha/2,ns-1}\right] = 1-\alpha \tag{7.12}$$

where $c_{\alpha/2,ns-1}$ and $c_{1-\alpha/2,ns-1}$ are the critical values of the chi-square distribution evaluated at the probability levels of α/2 and (1–α/2) with (*ns*–1) degrees of freedom, respectively. Using the realization of Σ^2, denoted as $\tilde{\sigma}^2$, Eq. (7.12) is rewritten as

$$\Pr\left[\frac{(ns-1)\tilde{\sigma}^2}{c_{1-\alpha/2,ns-1}} \le \sigma^2 \le \frac{(ns-1)\tilde{\sigma}^2}{c_{\alpha/2,ns-1}}\right] = 1-\alpha \tag{7.13}$$

Equation (7.13) indicates that the true standard deviation σ is within this interval, which is calculated from given *ns* samples for a significance level α, with $100\times(1-\alpha)\%$ of the probability. For two-sided (1–α), the lower and upper bounds of the confidence interval of the standard deviation, $\tilde{\sigma}^L$ and $\tilde{\sigma}^U$, respectively, are calculated as

$$\tilde{\sigma}^L = \sqrt{\frac{(ns-1)\tilde{\sigma}^2}{c_{1-\alpha/2,ns-1}}} \text{ and } \tilde{\sigma}^U = \sqrt{\frac{(ns-1)\tilde{\sigma}^2}{c_{\alpha/2,ns-1}}} \tag{7.14}$$

However, the confidence interval of the standard deviation in Eq. (7.13) is calculated based on an assumption that the input random variable follows a Gaussian distribution. If it does not, the estimated confidence interval could be erroneous. Thus, the bootstrap method, which does not require the normality assumption on the input variable, needs to be used.

The bootstrap method calculates a confidence interval of an estimated parameter $\tilde{\theta}$ by constructing a distribution of the parameter using a frequency distribution of $\tilde{\theta}^*$'s obtained from randomly generated bootstrap samples based on given data. Table 7.1 shows how to calculate the confidence interval of a parameter using the bootstrap method.

Table 7.1. Bootstrap Procedures

	Bootstrap Procedures
Step 1	Construct an empirical distribution $F_{ns}(\mathbf{x})$ from original data $\mathbf{x}=[x_1, x_2, \cdots, x_{ns}]$.
Step 2	Generate bootstrap samples $\mathbf{x}^* = [x_1^*, x_2^*, \cdots, x_{ns}^*]$ from empirical distribution $F_{ns}(\mathbf{x})$ for non-parametric approach or parametric distribution $\tilde{F}(\mathbf{x}\|a,b)$ for parametric approach.
Step 3	Calculate a statistic of interest $\tilde{\theta}$ from bootstrap samples, yielding $\tilde{\theta}_b^*$.
Step 4	Repeat Steps 2 and 3 B times (i.e., B=1000).
Step 5	Construct a probability distribution $\tilde{G}^*(\tilde{\theta}^*)$ from $\tilde{\theta}_1^*, \tilde{\theta}_2^*, \cdots, \tilde{\theta}_B^*$, and then calculate confidence interval of estimated parameter, $\tilde{\theta}$ using $\tilde{G}^*(\tilde{\theta}^*)$.

The first step is to construct an empirical distribution $F_{ns}(\mathbf{x})$ generated from given samples $\mathbf{x}=[x_1, x_2, \cdots, x_{ns}]$. In the second step, if a random sample of size *ns* with replacement is drawn from the empirical distribution $F_{ns}(\mathbf{x})$, then this is called a non-

parametric approach. If the re-sample is drawn from a specified model $\tilde{F}\left(\mathbf{x}\middle|a,b\right)$ determined from the given samples, this is called a parametric approach. In this study, the distribution type of the parametric model is identified using the Bayesian method, and its parameters, a and b, are quantified using Table 3.1. The third step is to calculate the statistic of interest $\tilde{\theta}$ from the re-sample, drawn from either empirical or parametric distribution, yielding $\tilde{\theta}_b^*$. In the fourth step, the second and third steps are repeated B times (i.e., B=1000). Then, the fifth step is to construct a probability distribution from $\tilde{\theta}_1^*, \tilde{\theta}_2^*, \cdots, \tilde{\theta}_B^*$. This distribution is the bootstrap sampling distribution of $\tilde{\theta}$, $\tilde{G}^*\left(\tilde{\theta}^*\right)$, which is used to calculate the confidence interval of $\tilde{\theta}$. To obtain the bootstrap sampling distribution of $\tilde{\theta}$, the normal approximation, percentile, bias corrected (BC), percentile-*t*, and bias corrected accelerated (BCa) methods can be used.

The normal approximation method assumes that the distribution of $\tilde{\theta}$ is a Gaussian distribution. Using the assumption, the confidence interval of $\tilde{\theta}$ is obtained as (Mooney and Duval, 1993)

$$\tilde{\theta} - z_{\alpha/2}\tilde{\sigma}_{\tilde{\theta}}^* < \theta < \tilde{\theta} + z_{\alpha/2}\tilde{\sigma}_{\tilde{\theta}}^* \tag{7.15}$$

where $\tilde{\sigma}_{\tilde{\theta}}^* = \sqrt{\sum_{b=1}^{B}\left[\tilde{\theta}_b^* - \tilde{\theta}^*\right]^2 / \left(B-1\right)}$, $\tilde{\theta}^* = \sum_{b=1}^{B}\tilde{\theta}_b^* / B$, and $z_{\alpha/2}$ is the value of standard Gaussian distribution CDF with $\alpha/2$. This normal approximation assumes that the distribution of $\tilde{\theta}$ is Gaussian, so that the confidence interval is not accurate for non-Gaussian distribution that has a very different PDF shape from the Gaussian distribution.

The percentile method calculates the confidence interval of the parameter based on that the bootstrap sampling distribution $\tilde{G}\left(\tilde{\theta}^*\right)$ approximates the population distribution $G\left(\tilde{\theta}\right)$. The basic idea of this method is that the confidence interval for $\left(1-\alpha\right)$ level includes all the values of $\tilde{\theta}^*$ between $\left(\alpha/2\right)$ and $\left(1-\alpha/2\right)$ percentiles of $\tilde{G}^*\left(\tilde{\theta}^*\right)$. The sorting vector of $\tilde{\theta}_b^*$'s obtained from each bootstrap sample for $b=1,\cdots,B$, the values of $\tilde{\theta}_b^*$ evaluated at $\left(\alpha/2\times100\right)^{\text{th}}$ and $\left(1-\alpha/2\right)\times100^{\text{th}}$ percentiles of $\tilde{G}^*\left(\tilde{\theta}^*\right)$ are used as the lower and upper bounds of $\tilde{\theta}$,

$$\tilde{\theta}_{\alpha/2}^* < \theta < \tilde{\theta}_{1-\alpha/2}^* \tag{7.16}$$

Since the percentile method does not assume that the bootstrap sampling distribution follows a Gaussian distribution like the normal approximation method, it allows $\tilde{G}^*\left(\tilde{\theta}^*\right)$ conforming to any shape that the data follow. For this reason, it is the most widely used bootstrap technique among applied statisticians (Liu and Singh, 1988). However, when the number of samples is small, $\tilde{G}^*\left(\tilde{\theta}^*\right)$ might be a biased estimator of $G\left(\tilde{\theta}\right)$, i.e., $\tilde{\theta}^*$ is a biased estimator of $\tilde{\theta}$. In that case, the percentile method can be inaccurate.

The bias corrected (BC) method corrects the bias term by introducing an adjusted parameter z_0. Suppose that there exists some monotonic transformations of $\tilde{\theta}^*$ and $\tilde{\theta}$, say $\tilde{\varphi}$ and φ, respectively. Instead of assuming that $\tilde{\theta}^*-\tilde{\theta}$ be centered at zero, the BC method assumes that $\varphi\left(\tilde{\theta}^*\right)-\varphi\left(\tilde{\theta}\right)+z_0=z$ follows a standard Gaussian distribution.

Since $\tilde{\varphi}$ and φ are monotonic functions, it holds that $\Pr\left(\tilde{\theta}^* \le \tilde{\theta}\right) = \Pr\left(z \le z_0\right) = \Phi\left(z_0\right)$ (Efron, 1982). Accordingly, z_0 is calculated using (Mooney and Duval, 1993)

$$z_0 = \Phi^{-1}\left(\Pr\left(\tilde{\theta}^* \le \tilde{\theta}\right)\right) \tag{7.17}$$

where z_0 is a biasing constant that compensates for the bias between $\tilde{\theta}^*$ and $\tilde{\theta}$. Since $\tilde{G}^*\left(\tilde{\theta}^*\right)$ is invariant to the transformation, the transformation does not need to be known. Using z_0, the confidence interval of $\tilde{\theta}$ is obtained as

$$\tilde{\theta}^*_{\Phi\left(2z_0+z_{\alpha/2}\right)} < \theta < \tilde{\theta}^*_{\Phi\left(2z_0+z_{1-\alpha/2}\right)} \tag{7.18}$$

where $\tilde{\theta}^*_{\Phi\left(2z_0+z_{\alpha/2}\right)}$ is the value of $\tilde{\theta}^*$ evaluated at the $\Phi\left(2z_0+z_{\alpha/2}\right)\times 100^{\text{th}}$ percentile, and $\tilde{\theta}^*_{\Phi\left(2z_0+z_{1-\alpha/2}\right)}$ is the value of $\tilde{\theta}^*$ evaluated at the $\Phi\left(2z_0+z_{1-\alpha/2}\right)\times 100^{\text{th}}$ percentile. The BC method corrects the bias term, but it still requires the parametric assumption that there exist monotonic transformations of $\tilde{\theta}^*$ and $\tilde{\theta}$.

The bias corrected and accelerated (BCa) method generalizes the BC method. The BC method only corrects the bias, whereas the BCa method corrects both the bias and the skewness. The BCa method assumes that for some monotone transformations, $\tilde{\varphi}$ and φ, some bias constant z_0 and acceleration constant a result in (Efron and Tibshirani, 1993)

$$\frac{\left(\tilde{\varphi}-\varphi\right)}{\sigma_{\tilde{\varphi}}} \sim N\left(-z_0\sigma_h, \sigma_h^2\right), \qquad \sigma_h = 1 + ah \tag{7.19}$$

where $\sigma_{\tilde{\varphi}}$ is the constant standard error of $\tilde{\varphi}$, and the acceleration a is defined as

$$a = \frac{\sum_{i=1}^{ns}\left(\tilde{\theta} - \tilde{\theta}_{(i)}\right)^3}{6\left\{\sum_{i=1}^{ns}\left(\tilde{\theta} - \tilde{\theta}_{(i)}\right)^2\right\}^{3/2}} \tag{7.20}$$

where $\tilde{\theta}_{(i)}$ is the estimated parameter of $\mathbf{x}_{(i)} = \left(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{ns}\right)$ without the ith point x_i and $\tilde{\theta} = \sum_{i=1}^{ns}\tilde{\theta}_{(i)} / ns$. Using Eq. (7.19), the BCa confidence interval is defined as

$$\tilde{\theta}^*_{\alpha_1} < \theta < \tilde{\theta}^*_{\alpha_2} \tag{7.21}$$

where $\alpha_1 = \Phi\left(z_0 + \frac{z_0 + z_{\alpha/2}}{1 - a \times \left(z_0 + z_{\alpha/2}\right)}\right)$, $\alpha_2 = \Phi\left(z_0 + \frac{z_0 + z_{1-\alpha/2}}{1 - a \times \left(z_0 + z_{1-\alpha/2}\right)}\right)$, and $z_0 = \Phi^{-1}\left(\Pr\left(\hat{\theta}^* \le \hat{\theta}\right)\right)$.

However, since the BCa method highly depends on the acceleration a, if it is not accurate, the BCa method also might be inaccurate.

The percentile-*t* method uses the distribution of a standardized estimator to calculate the confidence interval. The percentile-*t* interval is expected to be accurate to the extent that standardizing depends less on the boot sampling estimator $\tilde{\theta}^*$ than the percentile method. Transforming the estimated parameter $\tilde{\theta}^*_b$ to the standardized parameter t^*_b (Mooney and Duval, 1993),

$$t^*_b = \left(\tilde{\theta}^*_b - \tilde{\theta}\right) / \tilde{\sigma}_{\tilde{\theta}^*_b} \tag{7.22}$$

where $\tilde{\theta}^*_b$ is the estimated parameter from each re-sampled data $\mathbf{x}^*_b = \left[x^*_{1b}, x^*_{2b}, \cdots, x^*_{nsb}\right]$ for $b = 1, \cdots, B$. $\tilde{\theta}$ is the estimated parameter from the original data, $\mathbf{x} = \left[x_1, x_2, \cdots, x_{ns}\right]$; and $\tilde{\sigma}_{\tilde{\theta}^*_b}$ is the standard deviation of $\tilde{\theta}$ obtained from a double bootstrap, which is

another level of re-sampling. That is, the double bootstrap sample $\mathbf{x}_d^* = \left[x_{1d}^{**}, x_{2d}^{**}, \cdots, x_{nsd}^{**} \right]$, $d = 1, \cdots, D$ is re-sampled from the bootstrap samples $\mathbf{x}_b^* = \left[x_{1b}^*, x_{2b}^*, \cdots, x_{nsb}^* \right]$, $b = 1, \cdots, B$. Thus, the percentile-*t* method requires many bootstrap samples ($D \times B$). Using double bootstrap samples, $\tilde{\sigma}_{\tilde{\theta}_b^*}$ is obtained as

$$\tilde{\sigma}_{\tilde{\theta}_b^*} = \sqrt{\frac{\sum_{d=1}^{D} \left[\tilde{\theta}_d^{**} - \tilde{\theta}^{**} \right]^2}{D-1}} \tag{7.23}$$

where $\tilde{\theta}^{**} = \sum_{d=1}^{D} \tilde{\theta}_d^{**} / D$.

The confidence interval of the parameter $\tilde{\theta}$ is obtained as

$$\tilde{\theta} + t_{\alpha/2}^* \tilde{\sigma}_{\tilde{\theta}} < \theta < \tilde{\theta} + t_{1-\alpha/2}^* \tilde{\sigma}_{\tilde{\theta}} \tag{7.24}$$

where $t_{\alpha/2}^*$ and $t_{1-\alpha/2}^*$ are the values of t^* evaluated at $\alpha/2 \times 100$ th and $(1-\alpha/2) \times 100$ th percentiles, respectively. In Eq. (7.24), $\tilde{\sigma}_{\tilde{\theta}}$ is the population standard deviation, which is calculated as $\tilde{\sigma}_{\tilde{\theta}} = \sqrt{\sum_{d=1}^{D} \left[\tilde{\theta}_d^* - \tilde{\theta}^* \right]^2 / (D-1)}$, where $\tilde{\theta}_d^*$ is the estimator obtained from re-sampled data $\mathbf{x}^* = \left[x_1^*, x_2^*, \cdots, x_D^* \right]$, which is randomly generated from the original data.

The five bootstrap methods are combined with the nonparametric and the parametric approach as explained in Table 7.1. In this study, for the parametric approach, both the true CDF and the identified CDF using the Bayesian method are used. To test the bootstrap methods, the Gaussian marginal distribution is considered as a true model with $\mu = 5.0$ and $\sigma = 5.0$ (COV=1.0) and 1000 data sets with *ns*=30, 100, 300 are generated from the true model. Note that COV of 1.0 seems too large, but it is still

observed in real engineering applications, such that the COV of the fatigue ductility coefficient of 950X steel is obtained as 1.15 from 29 data (Socie, 2003). Thus, it is valuable to test the problem with a large COV.

The performance of the bootstrap method is assessed by the estimated confidence level of the standard deviation. The confidence level of the standard deviation is estimated by the probability that the upper bounds of the confidence interval of the standard deviation, calculated from 1000 data sets, are larger than the true standard deviation. Thus, it is the most desirable that the estimated confidence level is larger than the target confidence level. In this study, since 95% of the two-sided target confidence level corresponds to 97.5% of the upper side of the confidence interval, the corresponding target confidence level is 97.5%. On the other hand, unnecessarily large upper bound of the standard deviation is not good because it yields too conservative optimum design. Therefore, the most desirable confidence interval should include the true standard deviation with the target confidence level, but with an adequately size.

Table 7.2 shows the estimated confidence level of the standard deviation using different bootstrap methods. The parametric approach using the true marginal CDF has the best performance, and the parametric approach with identified CDF is not as good as the one with true CDF due to the identification and quantification error. However, it is better than the nonparametric approach because it yields more desirable confidence level than the nonparametric approach when the identified CDFs are correct.

Among the five bootstrap methods, the percentile-*t* has almost 97.5% confidence level even for *ns*=30 and the BCa method has the second best performance. As the number of samples increases, all methods converge toward 97.5%. In this case, since the

true distribution is Gaussian, the method using a normality assumption on the input variable X, i.e., Eq. (7.14), has 97.5% regardless of the number of samples. Notice that Eq. (7.14) uses the normality assumption on X, whereas the normal approximation (bootstrap) uses the normality assumption on the parameter θ.

Table 7.3 shows the mean values of the upper bound of the confidence intervals of standard deviation obtained from 1000 data sets using all the methods. Since the parametric approach has larger confidence levels than the nonparametric approach as shown in Table 7.2, the confidence intervals using the parametric approach are accordingly larger than those using the nonparametric approach as shown in Table 7.3.

Since the obtained upper bounds using all bootstrap methods are adequately large corresponding to estimated confidence levels, all methods seem to have adequately sized confidence intervals. Thus, the percentile-*t* method, which has the most desirable confidence level, is the most preferred among five bootstrap methods, followed by the BCa method.

Table 7.2. Estimated Confidence Levels of Standard Deviation (Gaussian, $\mu = 5.0$ and $\sigma = 5.0$)

ns	Approach	Nor. Approx.	Percentile	BC	BCa	Percentile-*t*
30	Nonpar.	92.4	89.9	94.3	96.2	97.6
	Par. (Iden.)	93.0	93.1	94.4	96.0	96.7
	Par. (True)	94.0	94.1	95.9	97.1	97.7
100	Nonpar.	94.1	93.3	94.9	95.6	96.7
	Par. (Iden.)	96.4	96.5	97.0	98.0	98.1
	Par. (True)	95.5	95.4	96.3	96.8	97.4
300	Nonpar.	96.4	96.1	96.6	97.9	97.5
	Par. (Iden.)	97.0	96.7	97.0	97.5	97.7
	Par. (True)	96.8	96.9	97.1	97.8	97.6

Note: Method Using Normality Assumption on X : 97.5% for *ns*=30, 100, 300

Table 7.3. Mean Values of Upper Bound of Confidence Interval of Standard Deviation (Gaussian, $\mu = 5.0$ and $\sigma = 5.0$)

ns	Approach	Nor. Approx.	Percentile	BC	BCa	Percentile-*t*
30	Nonpar.	6.133	6.001	6.213	6.407	6.901
	Par. (Iden.)	6.346	6.448	6.491	6.735	6.826
	Par. (True)	6.201	6.209	6.325	6.536	6.753
100	Nonpar.	5.662	5.633	5.708	5.779	5.843
	Par. (Iden.)	5.705	5.720	5.727	5.828	5.828
	Par. (True)	5.696	5.697	5.728	5.796	5.841
300	Nonpar.	5.389	5.381	5.406	5.430	5.442
	Par. (Iden.)	5.396	5.397	5.405	5.428	5.440
	Par. (True)	5.398	5.400	5.408	5.431	5.441

Note: Method Using Normality Assumption on X: 6.680 for *ns*=30, 5.793 for *ns*=100, and 5.414 for *ns*=300

Consider a non-Gaussian distribution, lognormal distribution with $\mu = 5.0$ and $\sigma = 5.0$ (COV=1.0). Since the distribution of the standard deviation is highly skewed due to the lognormal distribution with large COV, all methods have poor confidence levels of the standard deviation especially for *ns*=30 as shown in Table 7.4. However, as the number of samples increases, the obtained confidence levels approach the target confidence level whereas the method using the normality assumption on X is nearly only 65% regardless of the number of samples.

As shown in Table 7.4, when the nonparametric approach is used, the percentile-*t* has the highest confidence level, followed by the BCa method among five bootstrap methods. For the parametric approach using identified CDF, the percentile-*t* method is the best for *ns*=30 and 100, and the percentile method is the best for *ns*=300. When the parametric approach using the true CDF is used, the percentile method has the best performance for *ns*=30, 100, and 300.

Table 7.4. Estimated Confidence Levels of Standard Deviation (Lognormal, $\mu = 5.0$ and $\sigma = 5.0$)

ns	Approach	Nor. Approx.	Percentile	BC	BCa	Percentile-*t*
30	Nonpar.	64.9	61.3	64.8	70.4	83.7
	Par. (Iden.)	74.9	79.0	85.0	88.9	89.6
	Par. (True)	89.3	95.2	88.3	92.4	88.3
100	Nonpar.	76.0	75.1	78.1	83.1	88.8
	Par. (Iden.)	87.2	87.5	87.4	89.9	92.0
	Par. (True)	95.2	97.7	86.9	89.9	87.1
300	Nonpar.	82.3	82.7	84.5	87.9	91.3
	Par. (Iden.)	95.9	97.5	86.8	88.4	85.7
	Par. (True)	96.7	99.0	86.3	88.2	85.3

Note: Method Using Normality Assumption on X : 65.8% for *ns*=30, 66.0% for *ns*=100, and 65.0% for *ns*=300

Even though the percentile-*t* method provides more desirable confidence levels than other methods for the nonparametric approach, it provides unnecessarily large upper bounds of standard deviation. For example, the mean value of the upper bound of standard deviation using the nonparametric percentile-*t* method is 20.01, which is significantly larger than true standard deviation 5.0 as shown in Table 7.5. In addition, it has a large standard deviation, 32.78, of the upper bounds of confidence intervals for standard deviation as shown in Table 7.6. That is, the upper bounds of confidence intervals of the standard deviation are overestimated and widely spread over the wide range of the standard deviation. As the number of samples increases, the upper bound of the standard deviation approaches the true standard deviation and its variation is reduced, but it converges very slowly to the true standard deviation compared to other methods. Thus, the BCa method, which has second highest confidence level and adequate value of upper bound, is preferred for the nonparametric approach.

Table 7.5. Mean Values of Upper Bound of Confidence Interval of Standard Deviation (Lognormal, $\mu = 5.0$ and $\sigma = 5.0$)

ns	Approach	Nor. Approx.	Percentile	BC	BCa	Percentile-*t*
30	Nonpar.	6.900	6.458	6.803	7.379	20.01
	Par. (Iden.)	7.830	8.396	12.72	15.22	16.62
	Par. (True)	8.742	9.998	12.57	15.83	16.04
100	Nonpar.	6.482	6.367	6.604	7.144	10.78
	Par. (Iden.)	7.167	7.596	9.204	10.43	10.83
	Par. (True)	7.367	7.978	9.342	10.60	10.77
300	Nonpar.	6.088	6.078	6.219	6.571	7.810
	Par. (Iden.)	6.416	6.763	7.324	7.877	7.642
	Par. (True)	6.491	6.824	8.034	7.422	7.984

Note: Method Using Normality Assumption on X : 6.451 for *ns*=30, 5.708 for *ns*=100, and 5.401 for *ns*=300

Table 7.6. Standard Deviations of Upper Bound of Confidence Interval of Standard Deviation (Lognormal, $\mu = 5.0$ and $\sigma = 5.0$)

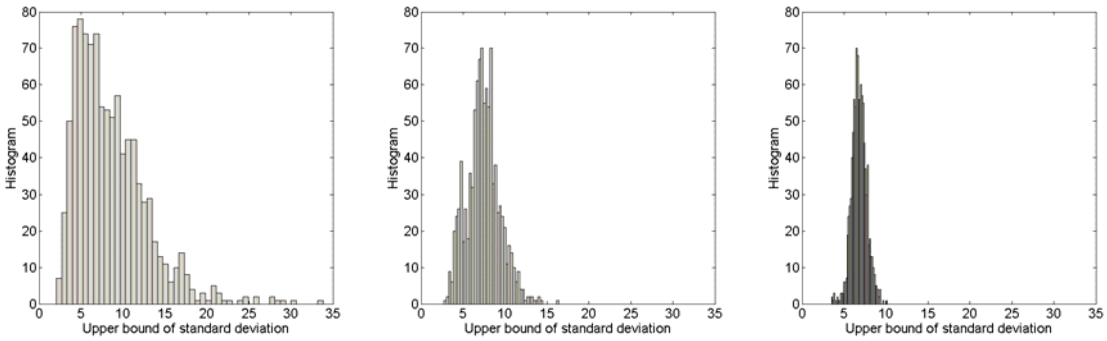
ns	Approach	Nor. Approx.	Percentile	BC	BCa	Percentile-*t*
30	Nonpar.	3.629	3.130	3.491	3.990	32.78
	Par. (Iden.)	3.873	4.316	11.62	12.96	20.66
	Par. (True)	3.587	3.958	10.18	11.81	21.21
100	Nonpar.	2.272	2.132	2.365	2.813	11.71
	Par. (Iden.)	1.977	2.019	4.725	5.373	12.34
	Par. (True)	1.917	1.663	4.790	5.520	12.77
300	Nonpar.	1.219	1.200	1.308	1.594	3.930
	Par. (Iden.)	0.971	0.873	2.728	2.978	4.692
	Par. (True)	0.929	0.815	2.598	2.946	4.695

Note: Method Using Normality Assumption on X : 2.484 for *ns*=30, 1.488 for *ns*=100, and 0.834 for *ns*=300

Likewise, for the parametric approach, the BC, BCa, and percentile-*t* methods have relatively higher confidence levels than other methods, but they have very large upper bounds and variations, as shown in Tables 7.5 and 7.6. Further, the estimated confidence levels using BC, BCa, and percentile-*t* methods do not converge to the target

confidence level even for a large number of samples. On the other hand, the percentile method yields somewhat less desirable confidence levels than the three methods, but it has adequate values of upper bounds for standard deviation and small variations as shown in Table 7.5 and 7.6. Figure 7.4 shows the histograms of the upper bounds of confidence interval of the standard deviation using the parametric percentile bootstrap method with identified CDF. The estimated upper bounds of standard deviations mostly centered at the true standard deviation, 5.0 even for *ns*=30. As the number of samples increases, a large amount of the upper bounds of standard deviation tends to be very close to the true standard deviation with a small variation.

Since the parametric percentile method provides the better confidence levels than the best nonparametric approach, BCa, the parametric percentile method is preferred one for non-Gaussian distribution with high skewness.



(a)*ns*=30 (b) *ns*=100 (b) *ns*=300

Figure 7.4. Histograms of Upper Bound of Confidence Interval of Standard Deviation

In summary, when the random variable has the Gaussian distribution, both the method using a normality assumption on X and the percentile-*t* bootstrap method can be

used. When the input marginal distribution is not Gaussian and highly skewed, the parametric percentile method has the most desirable performance. Even though the bootstrap methods do not achieve the target confidence level for a small number of samples, as the number of samples increases, the obtained confidence levels converge to the target confidence level while the method using normality assumption on X does not. The bootstrap method has reasonably good performance for non-Gaussian distribution with small COV, i.e., less than 0.2. The bootstrap method can be applied to any types of distribution, and the test results for various types of distributions are presented by Shao and Du (1995).

The upper bound of the confidence interval calculated from the method using a normality assumption on X and the bootstrap method will be used to calculate the adjusted standard deviation, which is introduced in Section 7.4.

7.3.3 Confidence Interval of Correlation Coefficient

To calculate the confidence interval of the correlation coefficient, proposition 3.1 from Genest and Rivest (1993) is used as

$$\sqrt{ns}\,\frac{\tilde{\tau}-\tau}{4w} \sim N\left(0,1\right) \tag{7.25}$$

where τ is the population correlation coefficient, $w=\sqrt{\frac{1}{ns}\sum_{i=1}^{ns}\left(w_i+\tilde{w}_i-2\bar{w}\right)^2}$, $w_i=\frac{1}{ns}\sum_{j=1}^{ns}I_{ij}$, $\tilde{w}_i=\frac{1}{ns}\sum_{j=1}^{ns}I_{ji}$, and $\bar{w}=\frac{1}{ns}\sum_{i=1}^{ns}w_i$. If $x_{1j}<x_{1i}$ and $x_{2j}<x_{2i}$ (note x_{1i} and x_{2i} are ith sample point for X_1 and X_2), then $I_{ij}=1$. Otherwise, $I_{ij}=0$.

Let the correlation parameter be a function of the correlation coefficient, Kendall's tau, $\theta = g^{-1}(\tau)$. It is known that as *ns* approaches infinity, the sample correlation parameter follows a Gaussian distribution (Genest and Favre, 2007).

$$\tilde{\theta} \sim N\left(\theta, \frac{1}{ns}\left\{4w\frac{dg^{-1}(\tilde{\tau})}{d\tilde{\tau}}\right\}^2\right) \tag{7.26}$$

Thus, the confidence interval of the correlation parameter for $100\times(1-\alpha)$ of the confidence level is obtained as

$$\Pr\left[\tilde{\theta} - z_{\alpha/2}\frac{1}{\sqrt{ns}}4w\left|\frac{dg^{-1}(\tilde{\tau})}{d\tilde{\tau}}\right| \le \theta \le \tilde{\theta} + z_{\alpha/2}\frac{1}{\sqrt{ns}}4w\left|\frac{dg^{-1}(\tilde{\tau})}{d\tilde{\tau}}\right|\right] = 1-\alpha \tag{7.27}$$

where $z_{\alpha/2}$ is the CDF value of Gaussian distribution evaluated at $\alpha/2$.

Using the upper and lower bounds of the confidence interval of the correlation parameter θ, the upper and lower bounds of confidence interval of the correlation coefficient τ are calculated from $\tau = g(\theta)$ using Eq. (2.17) or explicit functions such as Table 2.2.

7.4 Adjusted Parameters Using Confidence Intervals of Input Parameters

As explained in the previous section, the upper bound of the confidence interval of the standard deviation yields a large β_t-contour, which leads to a reliable design, whereas upper and lower bounds of mean and correlation coefficient do not. Thus, the confidence intervals of the mean and correlation coefficient cannot be directly used for the input model, unlike the one for the standard deviation.

To resolve this problem, an adjusted standard deviation that includes the confidence interval of the mean and the upper bound of the confidence interval of the standard deviation is proposed to enlarge the β_t-contour, so that it covers the prediction errors of both the mean and standard deviation in estimating the input confidence level. Likewise, using the confidence interval of the correlation coefficient, the adjusted correlation coefficient is proposed. The adjusted correlation coefficient combined with the adjusted standard deviation yields a β_t-contour, which is large enough to cover the true β_t-contour.

7.4.1 Adjusted Standard Deviation Using Confidence Intervals of Mean and Standard Deviation

A small standard deviation indicates that the samples tend to be close to the mean while large standard deviation indicates that the samples are spread out with a large range of the sample values. Table 7.7 shows the expected values of the sample mean, the lower and upper bounds of the confidence intervals of the mean obtained from 1000 data sets where the true distribution is Gaussian with the true mean, $\mu = 5.0$. To see the effect of standard deviation on the estimation of the mean, small and large standard deviations, 0.3 and 5.0, are tested.

As displayed in Table 7.7, the small standard deviation yields a small confidence interval of the mean because most of the samples center on the true mean. On the other hand, the large standard deviation yields a large confidence interval of the mean because the samples are widely spread on a large domain of X. Thus, for large standard deviation,

it is important to consider the prediction error of mean values in order to have a desirable confidence level for the input model.

Table 7.7. Mean Values of Confidence Intervals of Mean

True Mean and Std.	*ns*	Lower Bound	Sample Mean	Upper Bound
μ=5, σ=0.3	30	4.887	4.999	5.112
	100	4.937	4.996	5.056
	300	4.965	4.999	5.033
μ=5, σ=5.0	30	3.121	5.000	6.879
	100	4.018	5.010	6.002
	300	4.430	5.000	5.570

The adjusted standard deviation can be defined by adding the confidence intervals of the mean to the upper bound of the confidence intervals of the standard deviation as

$$\tilde{\sigma}^A = \tilde{\sigma}^U + \Delta\tilde{\mu} \tag{7.28}$$

where $\tilde{\sigma}^U$ is the upper bound of the confidence interval of the standard deviation and $\Delta\tilde{\mu} = \left|\tilde{\mu}^U - \tilde{\mu}\right| = \left|\tilde{\mu} - \tilde{\mu}^L\right|$ since the sample mean is the middle point of the confidence interval for the Gaussian distribution.

However, since the confidence interval of the mean could yield an unnecessarily large effect on the adjusted standard deviation, the COV is used as a scale factor such that the effect of the mean is small for small COV and large for large COV. Thus, the adjusted standard deviation is defined as

$$\tilde{\sigma}^A = \tilde{\sigma}^U + \frac{\tilde{\sigma}}{\tilde{\mu}} \times \Delta\tilde{\mu} \tag{7.29}$$

Table 7.8 shows the mean values of the sample standard deviation and the lower and upper bounds of the confidence intervals of the standard deviation obtained from

1000 data sets where samples are generated from Gaussian distribution, $X \sim N\left(5.0, 0.3^2\right)$ and $X \sim N\left(5.0, 5.0^2\right)$. Since the true distribution is Gaussian, the method using a normality assumption on X is used.

Table 7.8. Mean Values of Confidence Intervals of Standard Deviation and Adjusted Standard Deviation (Gaussian)

True Std.	*ns*	Lower Bound	Sample Std.	Upper Bound	Adjusted Std.
	30	0.239	0.300	0.404	0.410
0.3	100	0.263	0.299	0.347	0.351
	300	0.277	0.299	0.325	0.327
	30	3.998	5.020	6.748	8.698
5.0	100	4.390	4.999	5.808	6.815
	300	4.645	5.017	5.454	6.028

For small standard deviation, σ=0.3, due to small prediction error of the mean, the adjusted standard deviation is slightly larger than the upper bound of the confidence intervals of the standard deviation. On the other hand, for large standard deviation, $\sigma = 5.0$, due to large prediction error of the mean, the adjusted standard deviation is fairly larger than the upper bound of the confidence interval of the standard deviation. As the number of samples increases, the adjusted standard deviation, lower and upper bounds of the confidence interval of standard deviation tend to converge to the true standard deviation.

Suppose the samples are generated from lognormal with same mean, $\mu = 5.0$, and two different standard deviations, $\sigma = 0.3$ and $\sigma = 5.0$. As shown in Table 7.9, when the standard deviation is small, σ=0.3, the lognormal distribution is almost identical to the Gaussian distribution, so that the estimated confidence interval (Table 7.9) is almost the

same as the one for the Gaussian distribution (Table 7.8). When the standard deviation is large, σ=5.0, the lognormal distribution becomes right skewed, which leads to a large confidence interval, as shown in Table 7.9, compared to the one for the Gaussian distribution in Table 7.8. Therefore, the use of the adjusted standard deviation covers the prediction error of both the mean and standard deviation whereas the use of the upper bound of confidence intervals of standard deviation only covers the prediction error of the standard deviation.

Table 7.9. Mean Values of Confidence Intervals of Standard Deviation and Adjusted Standard Deviation (Lognormal)

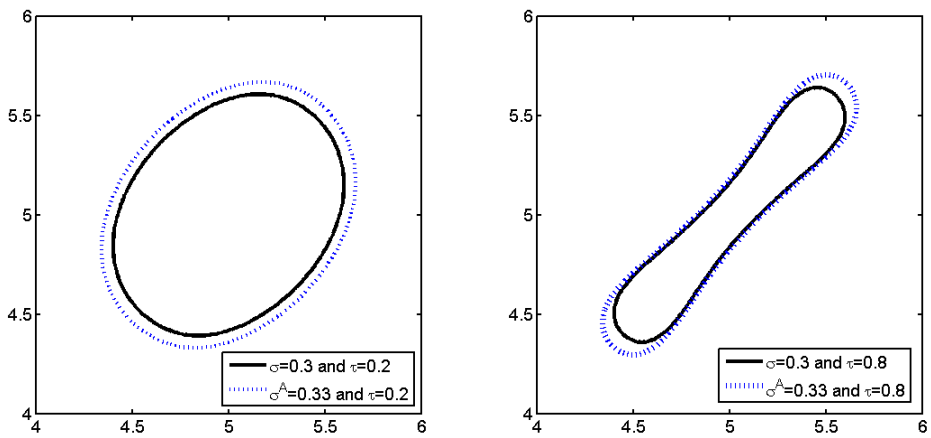
True Std.	*ns*	Lower Bound	Sample Std.	Upper Bound	Adjusted Std.
	30	0.218	0.299	0.386	0.393
0.3	100	0.257	0.299	0.343	0.355
	300	0.274	0.298	0.322	0.325
	30	2.440	4.900	8.396	10.51
5.0	100	3.236	4.950	7.596	8.784
	300	3.778	4.977	6.763	7.361

7.4.2 Adjusted Correlation Coefficient Using Confidence Intervals of Correlation Coefficient

As shown in Figure 7.3 (c), the β_t-contours for the lower and upper bounds of the correlation coefficient do not yield reliable design when the sample standard deviation is not large enough to cover the true β_t-contour. To resolve this problem, the adjusted correlation coefficient is proposed in this study.

Suppose that a true standard deviation σ is 0.3 and the adjusted standard deviation $\tilde{\sigma}^A$ is 0.33. If the small correlation coefficient ($\tau = 0.2$) and the adjusted

standard deviation ($\tilde{\sigma}^A = 0.33$) are used, they enlarge both the width and length of the β_t-contour, as shown in Figure 7.5 (a). On the other hand, if the large correlation coefficient ($\tau = 0.8$) and the adjusted standard deviation ($\tilde{\sigma}^A = 0.33$) are used, they enlarge the length of the β_t-contour much more than the width of the β_t-contour, as shown in Figure 7.5 (b). Thus, when the true correlation is large, $\tau = 0.8$, if the sample correlation coefficient is estimated to be larger than 0.8, the adjusted standard deviation ($\tilde{\sigma}^A = 0.33$) cannot cover the true β_t-contour.

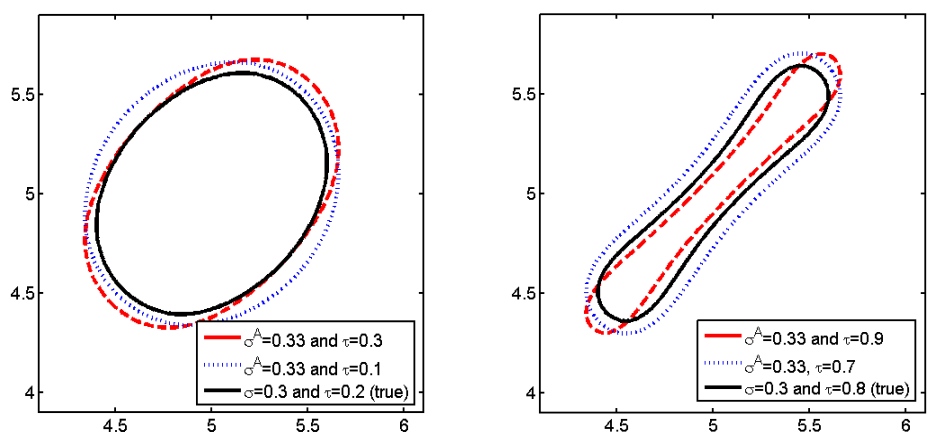


(a) τ=0.2 (b) τ=0.8

Figure 7.5. β_t-contours for Small and Large Correlation Coefficient

Suppose that two random variables follow Gaussian distributions, X_1 and $X_2 \sim N(5, 0.3^2)$. If samples are generated from the true model, the sample correlation coefficient might be underestimated or overestimated. Figure 7.6 (a) and (b) shows the β_t-contours with underestimated and overestimated correlation coefficients,

indicated as the dotted and dashed contours, respectively, for τ=0.2 and 0.8. For τ=0.2 in Fig. 7.6 (a), if the adjusted standard deviation is 0.33, the obtained β_t-contour covers the true β_t-contour for both underestimated ($\tilde{\tau}$=0.1) and overestimated ($\tilde{\tau}$=0.3). However, for $\tau = 0.8$ in Fig. 7.6 (b), even though the same adjusted standard deviation is used ($\tilde{\sigma}^A = 0.33$), the β_t-contour with underestimated correlation coefficient ($\tilde{\tau}$=0.7) covers the true β_t-contour, whereas the overestimated correlation coefficient ($\tilde{\tau}$=0.9) does not. Thus, the correlation coefficient needs to be adjusted such that the obtained β_t-contour covers the true β_t-contour for both small and large correlation coefficients.



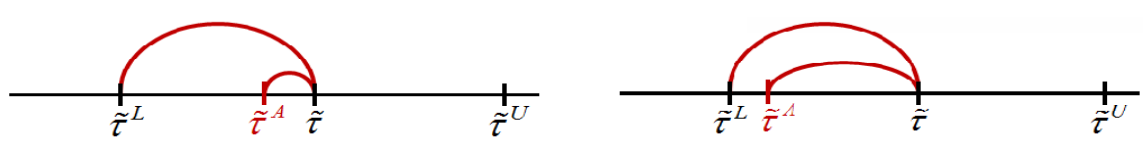
(a) τ=0.2 (b) τ=0.8

Figure 7.6. β_t-contours Using Underestimated and Overestimated Correlation Coefficient

Since the underestimated correlation coefficient and large standard deviation yield a reliable design, the lower bound of the confidence intervals of the correlation coefficient, $\tilde{\tau}^L$, seems to be adequate. However, when the correlation coefficient is small

positive, for example, for τ =0.2, its lower bound could be a negative correlation coefficient. In that case, the obtained β_t-contour shape will be very different from a true β_t-contour, which leads to an inaccurate optimum design.

Therefore, when the correlation coefficient is small, the sample correlation coefficient can be used to obtain the desirable input confidence level because the underestimated or overestimated correlation coefficient combined with the adjusted standard deviation $\tilde{\sigma}^A$ covers well the true β_t-contour as shown in Figure 7.6 (a). Still, the smaller correlation coefficient covers the true β_t-contour better, so that the adjusted correlation coefficient $\tilde{\tau}^A$ needs to be closer to the sample correlation coefficient $\tilde{\tau}$, but smaller than $\tilde{\tau}$ as shown in Figure 7.7 (a). On the other hand, when the correlation coefficient is large, the sample correlation coefficient cannot be used to obtain the desirable input confidence level especially when it is overestimated. The overestimated correlation coefficient cannot cover the true β_t-contour even though the adjusted standard deviation is large as shown in Figure 7.6 (b). Thus, for the large correlation coefficient, the adjusted correlation coefficient $\tilde{\tau}^A$ needs to be closer to the lower bound of correlation coefficient, $\tilde{\tau}^L$, but larger than $\tilde{\tau}^L$ as shown in Figure 7.7 (b).



(a) τ =0.2 (b) τ =0.8

Figure 7.7. Adjusted Correlation Coefficient

As shown in Figure 7.7, the adjusted correlation coefficient $\tilde{\tau}^A$ is selected such that the ratio of $\tilde{\tau}-\tilde{\tau}^A$ to $\tilde{\tau}-\tilde{\tau}^L$ is equivalent to the sample correlation coefficient $\tilde{\tau}$ as

$$\tilde{\tau}=\frac{\tilde{\tau}-\tilde{\tau}^A}{\tilde{\tau}-\tilde{\tau}^L} \tag{7.30}$$

Thus, using Eq. (7.30), the adjusted correlation coefficient $\tilde{\tau}^A$ can be obtained as

$$\tilde{\tau}^A=\tilde{\tau}-\tilde{\tau}\times\left(\tilde{\tau}-\tilde{\tau}^L\right) \tag{7.31}$$

However, if a copula has an asymmetric shape such as Clayton or Gumbel, $\tilde{\tau}-\tilde{\tau}^L$ might not be large enough to obtain the adequately small adjusted correlation coefficient that yields reliable designs. In that case, since the difference between the upper bound and the sample correlation coefficient $\tilde{\tau}-\tilde{\tau}^A$ is larger than $\tilde{\tau}-\tilde{\tau}^L$, if the larger value of $\tilde{\tau}-\tilde{\tau}^A$ is used instead of $\tilde{\tau}-\tilde{\tau}^L$ in Eq. (7.31), it yields more appropriately small adjusted correlation coefficient than $\tilde{\tau}-\tilde{\tau}^L$. Thus, for general cases, the maximum of $\tilde{\tau}-\tilde{\tau}^A$ and $\tilde{\tau}-\tilde{\tau}^L$ needs to be used instead of $\tilde{\tau}-\tilde{\tau}^L$ in Eq. (7.31) as

$$\tilde{\tau}^A=\tilde{\tau}-\tilde{\tau}\times\max\left(\tilde{\tau}^U-\tilde{\tau},\tilde{\tau}-\tilde{\tau}^L\right) \tag{7.32}$$

Table 7.10 shows the mean values of the confidence intervals of the correlation coefficients using 1000 data sets obtained from the Frank copula. To calculate the confidence interval of the correlation coefficient, copulas are identified for each data set. When the true correlation coefficient is 0.2 and *ns*=30, since independent copulas are frequently identified, the mean of sample correlation coefficients is only 0.137, which is fairly smaller than 0.2, and the adjusted correlation coefficient is close to the sample correlation coefficient. For a large number of samples, *ns*=300, since the Frank copula is correctly identified and sample correlation coefficient is correctly estimated, the mean value of the sample correlation coefficient is close to 0.2, and the adjusted correlation

coefficient also approaches the true correlation coefficient. When the correlation coefficient is large ($\tau = 0.8$), the adjusted correlation coefficient is close to the lower bound of the confidence intervals of the correlation coefficient. As the number of samples increases, it approaches the true correlation coefficient, 0.8.

This adjusted correlation coefficient will be used combined with the adjusted standard deviation, and those adjusted parameters will be used to calculate confidence levels of input models in Section 7.5.

Table 7.10. Confidence Intervals of Correlation Coefficient and Adjusted Correlation Coefficient (Frank Copula)

True	*ns*	Lower Bound	Sample	Upper Bound	Adjusted
	30	0.024	0.137	0.226	0.102
0.2	100	0.071	0.179	0.275	0.155
	300	0.127	0.201	0.268	0.186
	30	0.269	0.494	0.642	0.379
0.5	100	0.393	0.499	0.580	0.446
	300	0.444	0.501	0.550	0.473
	30	0.639	0.800	0.866	0.669
0.8	100	0.747	0.799	0.833	0.758
	300	0.775	0.799	0.819	0.780

7.5 Confidence Level of Input Model

Once the adjusted standard deviation and correlation coefficient are obtained, the β_t-contour with the adjusted parameters can be generated to estimate the confidence level of the input model.

Figure 7.8 shows a flowchart for assessment of the confidence level of the input model. When a true input model is given – for example, the Frank copula with Gaussian

marginal CDFs is the true joint CDF in this testing – different samples, *ns*=30, 100, and 300, can be generated from the true input model for a sufficient number of trials, such as 300. Based on the generated samples, mean and standard deviation are quantified, and the marginal distribution types of X_1 and X_2 are identified using the Bayesian method. Using the estimated marginal distributions, a copula that best describes the given data is identified, and the correlation coefficient is calculated. Using the estimated input model, the β_t-contour can be drawn. By comparing the estimated β_t-contour and the true β_t-contour, the confidence level of the input model is assessed by calculating the probability that the obtained β_t-contour is larger than the true β_t-contour by testing 300 trials.

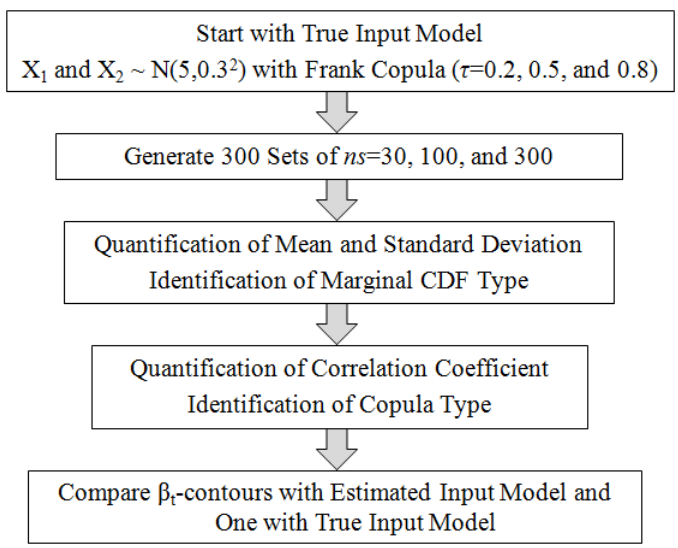


Figure 7.8. Flow Chart of Assessment of Input Confidence Level

Using the above procedure, two input models – one model with the upper bound of standard deviation and sample correlation coefficient (only with confidence level of

standard deviation) and another model with the adjusted parameters (with confidence level of all input parameters) – are tested. Figures 7.9 and 7.10 show the true β_t-contour and estimated β_t-contours using the two input models where the target confidence level is 97.5%.

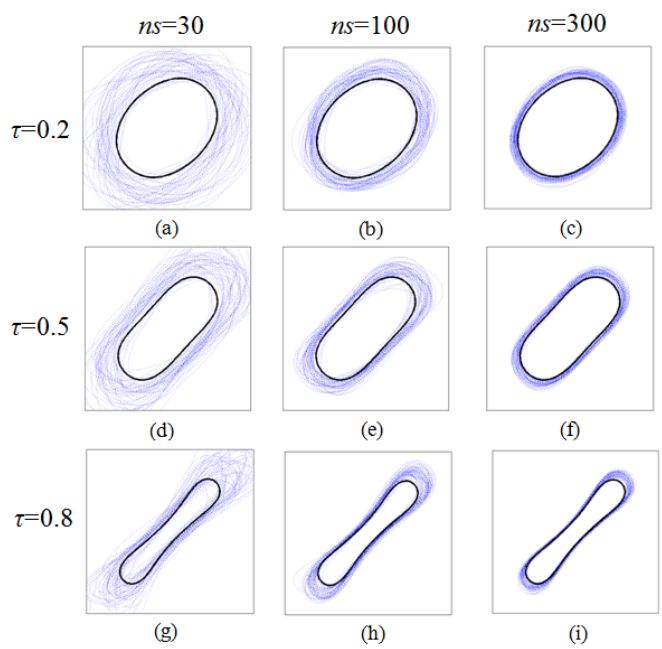


Figure 7.9. β_t-contours Using Obtained Input Models with Upper Bound of Standard Deviation and Sample Correlation coefficient

First, consider the input model with the upper bound of standard deviation and the sample correlation coefficient. As expected, when the number of samples is small, *ns*=30, the obtained β_t-contour shapes vary significantly and sometimes do not cover the true β_t-contour due to wrong estimation of the input model as shown in Figure 7.9 (a), (d), and (g). When the true correlation coefficient is small, τ=0.2, as shown in Figure 7.9 (a),

the obtained β_t-contour mostly covers the true β_t-contour even for a small number of samples. On the other hand, when the true correlation coefficient is large, τ=0.8, as shown in Figure 7.9 (g), the obtained β_t-contours do not cover the true β_t-contour well, especially when the correlation coefficient is estimated as higher than 0.8. As the number of samples increases, since the correlation coefficient is accurately estimated, and copulas and marginal distributions are correctly identified in most cases, the obtained β_t-contours mostly cover the true β_t-contour as shown in Figure 7.9 (c), (f), and (i).

Figure 7.10 shows the true β_t-contour and obtained β_t-contours using the adjusted standard deviation and correlation coefficient. Consider a small number of samples, *ns*=30. For the small number of samples with $\tau = 0.2$, the obtained β_t-contours using the adjusted parameters in Figure 7.10 (a) mostly cover the true β_t-contour like those using the upper bound of confidence intervals of standard deviation with the sample correlation coefficient as shown in Figure 7.9 (a). However, for the large correlation coefficient, $\tau = 0.8$, the obtained β_t-contours with adjusted parameters in Figure 7.10 (g) cover the true β_t-contour better than those with the upper bound of standard deviation and the sample correlation coefficient in Figure 7.9 (g). As the number of samples increases, the obtained β_t-contours are close to the true β_t-contour as shown in Figure 7.10 (c), (f), and (i).

Thus, when the adjusted parameters are used, the obtained β_t-contours cover well the true β_t-contour regardless of correlation coefficient values, which lead to have a

more desirable input confidence level than using the upper bound for standard deviation and the sample correlation coefficient.

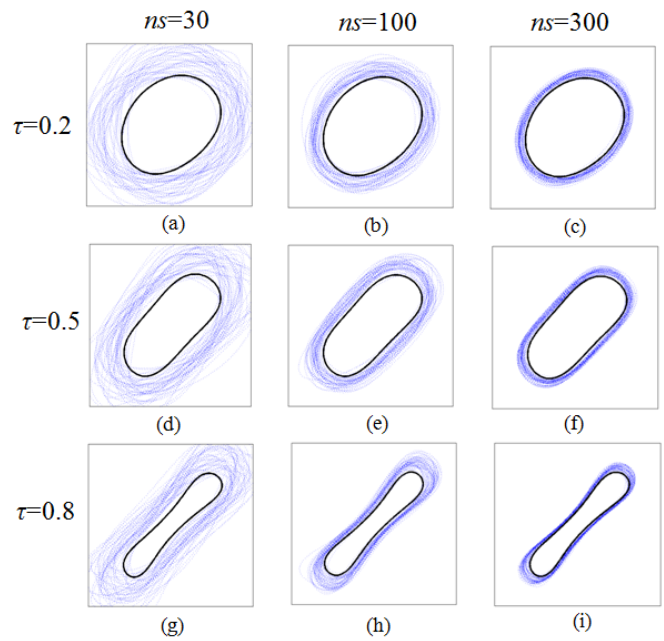


Figure 7.10. β_t-contours Using Obtained Input Models with Adjusted Parameters

Calculating the probability that the obtained β_t-contours in Figures 7.9 and 7.10 are larger than the true β_t-contour for 300 data sets, the confidence levels of input models are estimated as shown in Tables 7.11 and 7.12, respectively.

When the upper bound of standard deviation and the sample correlation coefficient are used as listed in Table 7.11, the input confidence level for $\tau = 0.8$ is much lower than for $\tau = 0.2$ because, as shown in Figure 7.9, the overestimated correlation coefficient for the high correlation coefficient makes the obtained β_t-contour often do not cover the true β_t-contour, whereas the ones for the small correlation coefficient do.

Using the adjusted standard deviation and correlation coefficient, the input confidence level is significantly increased especially for large correlation coefficient, as shown in Table 7.12. As the number of samples increases, the estimated confidence levels using the upper bound of standard deviation and the sample correlation coefficient are getting close to the target confidence level, 97.5%, but those are not as much close as confidence levels using the adjusted parameters.

Table 7.11. Input Confidence Levels (%) Using Upper Bound of Standard Deviation and Sample Correlation Coefficient, $N(5, 0.3^2)$

ns	τ=0.2	τ=0.5	τ=0.8
30	88	77	52
100	92	83	75
300	95	94	93

Table 7.12. Input Confidence Levels (%) Using Adjusted Standard Deviation and Correlation Coefficient, $N(5, 0.3^2)$

ns	τ=0.2	τ=0.5	τ=0.8
30	89	90	88
100	92	92	92
300	96	95	95

To observe the effect of large standard deviation on the input confidence level, suppose that two random variables follow Gaussian distribution with large standard deviations, X_1 and $X_2 \sim N\left(5, 5^2\right)$ correlated with the Frank copula.

Table 7.13 shows the input confidence levels using the upper bound of the confidence intervals of standard deviation and the sample correlation coefficient. As

expected, the input confidence level is very small especially for large correlation with a small number of samples. When the adjusted parameters are used, the input confidence level is significantly improved, as shown in Table 7.14. For the large standard deviation, the adjusted standard deviation becomes very large due to the large prediction error of the mean, which leads to offset the wrong identification and quantification of the parameter. Therefore, the obtained input confidence levels for the large standard deviation are very close to the target confidence level, 97.5% even for *ns*=30. However, as the number of samples increase, since confidence intervals of the mean and standard deviation for the large standard deviation are more slowly converged to the true mean and standard deviation than for small standard deviation, the obtained confidence levels are somewhat higher than the target confidence level.

Table 7.13. Input Confidence Levels (%) Using Upper Bound of Standard Deviation and Sample Correlation Coefficient, $N(5,5^2)$

ns	τ=0.2	τ=0.5	τ=0.8
30	85	80	50
100	90	86	78
300	95	94	95

Table 7.14. Input Confidence Levels (%) Using Adjusted Standard Deviation and Correlation Coefficient, $N(5,5^2)$

ns	τ=0.2	τ=0.5	τ=0.8
30	92	97	97
100	98	97	98
300	99	99	99

As stated earlier, even if we use an input model with confidence level that is less than the target confidence level, the confidence level of the reliability-based optimum design meeting the target reliability β_t should be higher than the input confidence level. In Section 7.6, it is shown how the confidence level of the reliability-based optimum design meeting the target reliability β_t, i.e., confidence level of output performance, is obtained through numerical examples.

7.6 Numerical Examples

A two-dimensional mathematical example and a coil spring example, which are used in Sections 6.2 and 6.3, will show how the input model with confidence level yields confidence levels of output performances.

7.6.1 Two-dimensional Mathematical Example

Consider a two-dimensional mathematical example where two input variables follow Gaussian distributions, X_1 and $X_2 \sim N(3, 0.3^2)$, as presented in Sections 6.2. It is assumed that X_1 and X_2 are correlated with the Frank copula with $\tau = 0.8$. An RBDO formulation for this example is defined in Eq. (6.1). For the comparison study, three types of input models – one model with the estimated parameters (without confidence level of input parameters), another model with the upper bound for standard deviation and the sample correlation coefficient (only with confidence level of standard deviation), and the other model with the adjusted parameters (with confidence level of all input parameters) – are tested.

In Chapter 6, it is shown how the marginal distribution and copula are identified using a data set obtained from a true model, and then the MPP-based DRM is carried out. In this section, 100 data sets with *ns*=30, 100, and 300 are generated from the true input model. Then, the marginal distribution, and copula type, and their parameters are determined from each data set. Using the estimated input models from 100 data sets, the optimum designs are obtained using MPP-based DRM. Finally, using the true input model, the output confidence levels are estimated by calculating the probability that the probability of failure evaluated at the obtained optimum design using MCS is smaller than the target probability of failure 2.275%.

Table 7.15 shows the minimum, mean, maximum values of the probabilities of failure for two active constraints P_{F_1} and P_{F_2}, and the output confidence level of optimum designs, which are obtained from 100 data sets. As shown in Table 7.15, when the input model with the estimated parameters is used for *ns*=30, the mean values of P_{F_1} and P_{F_2} (3.426% and 2.594%, respectively) are larger than the target probability of failure 2.275%. In particular, notice that the maximum values of P_{F_1} and P_{F_2} (14.29% and 7.752%, respectively) are much larger than 2.275%. Thus, the confidence levels of the output performance are significantly lower than the target confidence level 97.5% as shown in Table 7.15. That is, the input model with the estimated parameters yields unreliable designs in most cases.

On the other hand, when the input model with the upper bound of the standard deviation and sample correlation is used for *ns*=30, the mean values of P_{F_1} and P_{F_2} (1.074% and 1.562%, respectively) are smaller than 2.275%, which yield more reliable design than those using the estimated parameters. The maximum values of P_{F_1} and P_{F_2} (5.918% and

4.719%, respectively) are also much smaller than those using the estimated parameters. Accordingly, the obtained output confidence levels are much improved over those using estimated parameters.

When the input model with the adjusted parameters is used, the output confidence levels using the adjusted parameters become much closer to the target confidence level. Still, the estimated output confidence levels are still smaller than the target confidence level due to the incorrect identification of the marginal distribution and copula for a small number of samples, *ns*=30. However, as the number of samples increases, the minimum, mean, and maximum values of P_{F_1} and P_{F_2} using the input model with the adjusted parameters approaches the target probability of failure from a more reliable side than those using other input models. Moreover, the input model with adjusted parameters provides better output confidence levels than other input models.

Table 7.15. Probabilities of Failure and Output Confidence Levels for Mathematical Example ($\sigma_1 = \sigma_2 = 0.3$)

ns	Parameters	Estimated Par.		Upper Bound for Std.		Adjusted Par.	
		P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)
30	Min	0.142	0.315	0.007	0.119	0.004	0.026
	Mean	3.426	2.594	1.074	1.562	0.908	0.444
	Max	14.29	7.752	5.918	4.719	5.082	2.595
	Conf. level	35	54	83	81	90	98
100	Min	0.640	0.516	0.231	0.393	0.192	0.255
	Mean	2.279	2.087	1.425	1.573	1.282	0.947
	Max	5.829	5.941	3.694	4.091	3.445	2.300
	Conf. level	37	65	89	88	93	99
300	Min	0.854	1.064	0.510	0.894	0.473	0.719
	Mean	2.394	1.924	1.615	1.624	1.518	1.262
	Max	3.835	3.599	2.846	3.150	2.732	2.285
	Conf. level	43	77	94	90	95	98

Note: Optimum Design: (2.026, 1.163) ; P_{F_1} = 2.245%, P_{F_2} =2.070%

To test the effect of the large standard deviation on the output confidence level, consider a true input model with $X_1 \sim N(3, 0.3^2)$ and $X_2 \sim N(3, 1.5^2)$. As shown in Table 7.16, the input model with the estimated parameters still yields large probabilities of failure and small output confidence levels. The input model with the upper bound for standard deviation and the sample correlation coefficient provides better results than the one with the estimated parameters, but the output confidence levels are not as good as the one with the adjusted parameters especially for a small number of samples. For *ns*=300, the output confidence level using upper bound for standard deviation is larger than the target confidence level, 97.5%, which means it yields somewhat conservative optimum designs. This is because, for large standard deviation (i.e., $\sigma_2 = 1.5$), the upper bound for standard deviation is slowly converging to the true standard deviation rather than the one for small standard deviation. For the same reason, the input model with the adjusted parameters yields 100% for *ns*=300. Even though the output confidence levels are high, the minimum, mean, and maximum values of probabilities of failure tend to approach the target probability of failure, 2.275%, as the number of samples increases as shown in Table 7.16.

In summary, the input model with the adjusted parameters has the most desirable confidence level of output performance comparing with input models with the estimated parameters and those with the upper bound for standard deviation and the sample correlation coefficient. For a large standard deviation, it yields somewhat conservative design due to slow convergence of confidence intervals of the mean and standard deviation, but the obtained probabilities of failure approach the target probability of failure as the number of samples increases.

Table 7.16 Probabilities of Failure and Output Confidence Levels for Mathematical Example ($\sigma_1 = 0.3$ and $\sigma_2 = 1.5$)

ns	Parameters	Estimated Par.		Upper Bound for Std.		Adjusted Par.	
		P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)
30	Min	0.293	0.045	0.023	0.001	0.001	0.001
	Mean	2.583	2.895	0.992	1.256	0.313	0.380
	Max	10.92	9.732	6.961	4.723	4.324	4.457
	Conf. level	58	48	90	94	96	96
100	Min	1.035	0.871	0.507	0.296	0.196	0.121
	Mean	2.261	2.338	1.309	1.117	0.715	0.667
	Max	8.804	5.058	6.564	2.880	3.385	1.985
	Conf. level	66	56	95	96	99	100
300	Min	1.153	1.225	0.800	0.767	0.531	0.483
	Mean	2.006	2.154	1.496	1.444	1.058	1.074
	Max	2.989	3.383	2.344	2.449	1.744	1.904
	Conf. level	79	68	99	99	100	100

Note: Optimum Design: (2.977, 4.280) ; P_{F_1} =2.075%, P_{F_2} = 2.213%

7.6.2 Coil Spring Problem

Consider the coil spring problem, which was illustrated in Section 6.3. As stated earlier, in this example, the number of active coils, mean inner diameter of coil spring, and wire diameter are selected as random variables; mass density of material and shear modulus are selected as random parameters for RBDO. All variables are assumed to follow the Gaussian distribution. The coil inner diameter and the wire diameter are assumed to be correlated with the Clayton copula and $\tau = 0.7$. The statistical information of five random variables and parameters is presented in Table 6.4. Like the mathematical example, three input models – one model with the estimated parameters, another model with the upper bound of standard deviation and the sample correlation coefficient, and the other model with the adjusted parameters – are tested. The input models are obtained using 100 data sets, which are generated from the true input model. Using the three input

models obtained from 100 data sets, MPP-based DRM is carried out for RBDO where the RBDO formulation is given in Eq. (6.6).

As displayed in Table 7.17, when the estimated parameters are used, the obtained output confidence levels are much lower than the target confidence level 97.5%. The output confidence levels using the upper bound for standard deviation and sample correlation coefficient are closer to 97.5%, but its performance is still not as satisfactory as those using adjusted parameters. As the number of samples increases, the input model with adjusted parameters still has the best performance. The minimum, mean, and maximum values of probabilities of failure using all three input models get closer to the target confidence level, but the input model with the adjusted parameters approach it from more reliable side than other input models.

Table 7.17 Probabilities of Failures and Output Confidence Levels for Coil Spring Example

ns	Parameters	Estimated Par.		Upper Bound for Std.		Adjusted Par.	
		P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)
30	Min	0.544	0.027	0.103	0.000	0.019	0.000
	Mean	3.011	3.566	0.927	0.933	0.251	0.686
	Max	15.46	14.17	9.115	7.172	1.328	4.049
	Conf. level	47	39	95	85	100	95
100	Min	1.179	0.116	0.588	0.016	0.365	0.020
	Mean	2.169	2.487	1.163	1.186	0.697	1.043
	Max	5.302	5.586	2.535	3.117	1.608	2.921
	Conf. level	62	46	99	91	100	95
300	Min	1.496	1.062	0.996	0.698	0.831	0.534
	Mean	2.116	2.376	1.523	1.542	1.171	1.480
	Max	3.119	4.634	2.481	3.270	2.108	3.148
	Conf. level	71	49	99	93	100	97

Note: Optimum Design: (0.102E+2, 0.522E+0, 0.665E-1, 0.738E-3, 0.115E+8) ;
P_{F_1} =2.151%, P_{F_2} = 2.106%

The input model with the adjusted parameters has the desirable input and output confidence levels, as shown in the two numerical examples. For future research, the input confidence level for non-Gaussian distributions needs to be tested using the bootstrap method. More RBDO problems with non-Gaussian distributions need to be tested to check whether the obtained output confidence level is close to the target confidence level.

CHAPTER VIII

CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

8.1.1 Identification of Input Model

To carry out the RBDO, the information of the input variables such as marginal and joint CDFs need to be correctly identified and accurately quantified. In this study, a copula is utilized to model the joint CDF of the correlated input variables using limited information such as marginal CDF types and given test data. Since the correct identification of the copula is necessary to correctly model the joint CDF, the Bayesian method is proposed to identify a copula that best describes given experimental data. Likewise, the identification of marginal CDFs is as important as the identification of the copula, and thus a Bayesian method is proposed for the identification of correct marginal CDFs.

Using simulated data sets, the performance of the Bayesian method is tested for different numbers of samples and is compared with the GOF test. It is shown that the performance of the Bayesian method is much better than the one of the GOF test in identifying the correct marginal distributions and copulas, even if the number of samples is small. As the number of samples is increased up to 300 samples, the performance of the GOF test improves, but the Bayesian method is still better than the GOF test. Several examples are used to demonstrate how the Bayesian method is carried out to identify best fitted marginal CDFs and the copula.

8.1.2 Reduction of Transformation Ordering Effect on Reliability Based Design Optimization (RBDO)

In RBDO problems, once the joint CDFs are identified using the Bayesian method, they are used in the Rosenblatt transformation for the inverse reliability analysis. Incorporating the copula concept, the joint CDF can be categorized as an independent joint CDF, a joint CDF modeled by an elliptical copula, and a joint CDF modeled by a non-elliptical copula. It is found that the transformation ordering of input variables might have a significant effect on the reliability analysis according to joint CDF types. To test the effect of the transformation ordering on the inverse reliability analysis, β_t-contour with target reliability index, which are transformed from a standard Gaussian space (**U**-space) to an original space (**X**-space) using the Rosenblatt transformation, are compared for different transformation orderings.

When the input variables are independent or correlated with elliptical copulas, the β_t-contours are the same for different transformation orderings because the Rosenblatt transformation does not change nonlinearity of the constraint function. Accordingly, the inverse analysis results are the same for different transformation orderings of the input variables. However, when the correlated input variables have joint CDFs modeled by non-elliptical copulas, the transformed β_t -contours are different for different transformation orderings because of the nonlinear Rosenblatt transformation. In the inverse reliability analysis, the nonlinearities of the transformed constraint function from **X**-space to **U**-space become different for different transformation orderings, and, as a result, the different transformation orderings yield different RBDO results when an inaccurate method such as the FORM is used. Especially, the joint CDFs modeled by

non-elliptical copulas could lead to highly nonlinear constraint functions that the FORM cannot handle. Thus, in this study, it is proposed to use the MPP-based DRM. Numerical examples show that when the MPP-based DRM is used, the difference between the RBDO results using different transformation orderings is reduced and the accurate estimation of probability of failure is achieved.

8.1.3 Reliability Based Design Optimization with Confidence Level under Input Model Uncertainty

In broad industrial applications, only limited experimental data are usually available, so the input model obtained from the limited data might be inaccurate, which lead to inaccurate RBDO results. To offset the inaccurate estimation of the input model, an adjusted standard deviation and correlation coefficient involving confidence intervals of all input parameters (mean, standard deviation, and correlation coefficient) are proposed such that they offset the inaccurate estimation of the input parameters.

The calculation of the adjusted standard deviation involves the upper bound of the confidence interval for standard deviation. In many cases, the method using the normality assumption on the input variable has been used, but it is not accurate for non-Gaussian distributions. Thus, the bootstrap method, which does not require the normality assumption, is used for calculation of the confidence interval of the standard deviation. Simulations results show that the method using the normality assumption on the input variable has a desirable confidence level of the standard deviation for Gaussian distribution, but it does not for non-Gaussian distribution. On the other hand, the bootstrap method yields the desirable confidence level of the standard deviation for both

Gaussian and non-Gaussian distributions. For the highly skewed non-Gaussian distribution, the obtained confidence level of the standard deviation is somewhat smaller than the target confidence level, but it approach the target confidence level for a large number of samples whereas the method on the normality assumption does not.

To check whether use of the adjusted parameters provides the desirable confidence level of the output performance, i.e., the desired target probability of failure, the confidence level of the input model is estimated using the β_t-contour. Numerical results show that the input models without confidence level of input parameters and those only with confidence level of standard deviation do not yield desirable confidence levels for the input model and output performance on RBDO results. On the other hand, the input models with adjusted parameters, which include confidence level of all input parameters, yield desirable input confidence levels and the obtained RBDO results are considerably reliable, which leads to desirable confidence levels of the output performances.

8.2 Future Recommendation

The bootstrap method has a good performance of calculating the confidence interval of standard deviation for non-Gaussian distributions. However, it has not been tested to obtain the input and output confidence level for problems with non-Gaussian distributions.

For future research, the input and output confidence levels need to be tested through numerical examples with non-Gaussian distributions using the bootstrap method.

Further, a real engineering problem such as the M1A1 tank roadarm needs to be tested to show how the optimum design is changed when the confidence level is implemented in RBDO.

REFERENCES

Ang, A.H-S., and Tang, W.H., *Probability Concepts in Engineering Design*, Vol. I: Decision, Risk and Reliability, Wiley, New York, 1984.

Annis, C., "Probabilistic Life Prediction Isn't As Easy As It Looks," *Journal of ASTM International*, Vol. 1, No.2, pp. 3-14, 2004.

Arora, J. S., *Introduction to optimum design*, Elsevier, Amsterdam, 2004.

Atkinson K. E., *An Introduction to Numerical Analysis*, John Wiley & Sons, Chapter 5, New York, NY, 1989.

Baxter, M., "Minimum Variance Unbiased Estimation of the Parameters of the Pareto Distribution," *Metrika*, Vol. 29, No. 1, pp. 133-138, 1980.

Bonate, B.L., *Phamacokinetic-pharmacodynamic Modeling and Simulation*, Springer, 2005.

Bouyé, E., Durrleman, A., Nikeghbali, A., Riboulet, G., and Roncalli, T., "Copulas for Finance-a Reading Guide and Some Applications," *Technical report. Groupe de Recherche Opérationnelle*, Crédit Lyonnais, 2000.

Bravais, A., "Analyse Mathematique sur les Probabilites des Erreurs de Situation d'un Point," *Memoires par divers Savan*, Vol. 9, pp. 255-332, 1846.

Breitung, K., "Asymptotic Approximations for Multinormal Integrals," *ASCE Journal of Engineering Mechanics*, Vol. 110, No. 3, pp. 357-366, 1984.

Bretthorst, G.L., "An Introduction to Model Selection Using Probability Theory as Logic," *Maximum Entropy and Bayesian Methods*, Heidbreder, G. ed., Kluwer Academic Publishers, pp. 1-42, 1996.

Center for Computer-Aided Design, College of Engineering, *DRAW Concept Manual*, The University of Iowa, Iowa City, IA, 1999a.

Center for Computer-Aided Design, College of Engineering, *DRAW User Reference*, The University of Iowa, Iowa City, IA, 1999b.

Choi, K. K., Tu, J., and Park, Y. H., "Extensions of Design Potential Concept for Reliability-Based Design Optimization," *Journal of Structural and Multidisciplinary Optimization*, Vol. 22, No. 3, pp. 335-350, 2001.

Cirrone, G. A. P., Donadio, S., Guatelli, S., et al., "A Goodness-of-Fit Statistical Toolkit," *IEEE Transactions on Nuclear Science*, Vol. 51, No. 5, pp. 2056 – 2063, 2004.

Ditlevsen, O., and Madsen, H.O., *Structural Reliability Methods*, Wiley, New York, 1996.

Efron, B., *The Jackknife, the Bootstrap, and Other Resampling Plans*, SIAM, Philadelphia, 1982.

Efron, B., and Tibshirani, R., An Introduction to the Bootstrap, Chapman&Hall/CRC, 1993.

Embrechts, P., Lindskog, F., and McNeil, A., "Modeling Dependence with Copulas and Applications to Risk Management," *Handbook of Heavy Tailed Distributions in Finance*, Elsvier, Armsterdam, The Netherlands, Chap. 8., 329-384, 2003.

Frees, E. W., and Valdez, E. A., "Understanding Relationships Using Copulas," *North American Actuarial Journal*, Vol. 2, No. 1, pp. 1-25, 1998.

Gander, W., and Gautschi, W., "Adaptive Quadrature – Revisited," *BIT*, Vol. 40, pp. 84-101, 2000.

Genest, C., and Favre, A-C., "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask." *Journal of Hydrologic Engineering*, Vol. 12, No. 4, pp. 347-368, 2007.

Genest, C., and Rémillard, B., "Validity of the Parametric Bootstrap for Goodness-of-fit Testing in Semiparametric Models," *Technical Rep. G-2005-51*, Group d'Études et de Recherche en Analyse des Décision, 2005.

Genes, C., and Rivest, L.-P., "Statistical Inference Procedures for Bivariate Archimedean Copulas," *Journal of the American Statistical Association*, Vol. 88, pp. 1034-1043, 1993.

Haldar, A., and Mahadevan, S., *Probability, Reliability, and Statistical Methods in Engineering Design*, John Wiley & Sons, New York, 2000.

Hahn, G. J., and Shapiro, S. S., *Statistical Models in Engineering*, John Wiley & Sons, 1994.

Hasofer, A. M., and Lind, N. C., "An Exact and Invariant First Order Reliability Format," *ASCE Journal of the Engineering Mechanics Division*, Vol. 100, No. 1, pp. 111-121, 1974.

Hoel, P. G., *Introduction to Mathematical Statistics (3rd Edition)*, Wiley, New York, 1962.

Hohenbichler, M., and Rackwitz, R., "Improvement of Second-Order Reliability Estimates by Importance Sampling," *ASCE Journal of Engineering Mechanics*, Vol. 114, No. 12, pp. 2195-2199, 1988.

Hollander, M., and Wolfe, D. A., *Nonparametric Statistical Methods*, John Wiley & Sons, New York, 1973.

Huard, D., Évin, G., and Favre A-C., "Bayesian Copula Selection," *Computational Statistics & Data Analysis*, Vol. 51, pp. 809-822, 2006.

Kendall, M., "A New Measure of Rank Correlation," *Biometrika*, Vol. 30, pp. 81-89, 1938.

Kolev, N., Anjos, U., and Mendes, B., "Copulas: a Review and Recent Developments," *Stochastic Models*, Vol. 22, pp. 617-660, 2006.

Kruskal, W.H., "Ordinal Measures of Associations," *American Statistical Association Journal*, Vol. 53, No. 284, pp. 814-861, 1958.

Lebrun, R., and Dutfoy, A., "A Generalization of the Nataf Transformation to Distributions with Elliptical Copula," *Probabilistic Engineering Mechanics*, to appear, 2008.

Lee, I., "Reliability-based Design Optimization and Robust Design Optimization Using Univariate Dimension Reduction Method," Ph. D. Thesis, The University of Iowa, 2008.

Lehmann, E. L., *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, Inc., San Francisco, 1975.

Likes, J., "Minimum Variqance Unbiased Estimates of the Parameters of Power-Function and Pareto's Distribution," *Statistische Hefte*, Vol. 10, pp. 104-110, 1969.

Liu, P-L., and Der Kiureghian A., "Multivariate Distribution Models with Prescribed Marginals and Covariances," *Probabilistic Engineering Mechanics*, Vol. 1, No. 2, pp. 105-112, 1986.

Liu, R.Y., and Singh, K., "Invited Discussion on Theoretical Comparison of Bootstrap Confidence Intervals", *Annals of Statistics*, Vol. 16, pp.978-979, 1988.

Madsen, H. O., Krenk, S., and Lind, N. C., *Methods of Structural Safety*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1986.

Meggiolaro, M. A., and Castro, J. T. P., "Statistical Evaluation of Strain-Life Fatigue Crack Initiation Predictions," *International Journal of Fatigue*, Vol. 26, No. 5, pp.463-476, 2004.

Melchers, R.E., *Structural Reliability Analysis and Prediction*, 2nd edition, J. Wiley & Sons, New York, 1999.

Mooney, C.Z., and Duval, R.D., *Bootstrapping: A Nonparametric Approach to Statistical Inference*, Sage Publication, 1993.

Nataf, A., "Détermination des CDFs de probabilités dont les marges sont données," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, Vol. 255, pp. 42-43, 1962.

Nelson, R. B., *An Introduction to Copulas*, Springer, New York, 1999.

Nikolaidis, E., Ghiocel, D. M., and Singhal, S., *Engineering Design Reliability Handbook*, CRC press, New York, 2005.

Noh, Y., Choi, K.K., and Du, L., "New Transformation of Dependent Input Variables Using Copula for RBDO," *7th World Congress of Structural and Multidisciplinary Optimiz*ation, Seoul, Korea, May 21-25, 2007.

Noh, Y., Choi, K.K., and Du, L., "Reliability Based Design Optimization of Problems with Correlated Input Variables Using Copulas," *Structural and Multidisciplinary Optimization*, Vol. 38, No. 1, pp. 1-16, 2008.

Noh, Y., Choi, K.K., and Lee, I., "Identification of Marginal and Joint CDFs Using the Bayesian Method for RBDO," *Structural and Multidisciplinary Optimization*, Vol. 40, No. 1, pp. 35-51, 2009a.

Noh, Y., Choi, K.K., and Lee, I., "Reliability-Based Design Optimization with Confidence Level under Input Uncertainty," *35th ASME Design Automation Conference*, San Diego, California, August 30- September 2, 2009b.

Palle, T. C., and Michael, J. B., *Structural Reliability Theory and Its Applications*, Springer-Verlag, Berlin, Heidelberg, 1982.

Pearson, K., "Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity and Panmixia," *Philos. Trans. Royal Soc. London Ser. A* , Vol. 187, pp. 253-318, 1896.

Pham, H., *Recent Advances in Reliability and Quality Engineering*, World Scientific, Singapore, 2001.

Pham, H., *Springer Handbook of Engineering Statistics*, Springer, London. 2006.

Rahman, S., and Wei, D., "A Univariate Approximation at Most Probable Point for Higer-Order Reliability Analysis," *International Journal of Solids and Structures*, Vol. 43, pp. 2820-2839, 2006.

Rosenblatt, M., "Remarks on A Multivariate Transformation," *Annal of Mathematical Statistics*, Vol. 23, pp. 470-472, 1952.

Schittkowski, K., *More Test Examples for Nonlinear Programming Codes*, Springer-Verlag, New York, 1987.

Shao, J., and Tu, D., *The Jackknife and Bootstrap*, Springer, 1995.

Schuëller, G.I., "On the treatment of uncertainties in structural mechanics and analysis",

Computers & structures, Vol. 85, pp. 235-243, 2007.

Socie, D. F., Seminar notes: Probabilistic Aspects of Fatigue, 2003. http://www.fatiguecaculator.com.

Spearman, C., "The Proof and Measurement of Association between Two Things," *The American Journal of Psychology*, Vol. 15, pp. 72-101, 1904.

Stander, Y., "The Use of Copulas in Risk Management," M.S. Thesis, University of Johannesburg, 2006.

Sterne, J. A., and Smith, G. D., "Sifting the Evidence — What's Wrong with Significance Tests?" *British Medical Journal*, Vol. 322, pp. 226–231, 2000.

Swanson Analysis System Inc., *ANSYS Engineering Analysis System User's Manual*, Vol. I & II, Houston, PA, 1989.

Tobias, P.A., and Trindade, D.C., *Applied Reliability*, 2nd edition, CRC Press, 1995.

Tu, J., and Choi, K. K., "A New Study on Reliability-Based Design Optimization," *Journal of Mechanical Design*, Vol. 121, No. 4, pp. 557–564, 1999.

Tu, J., Choi, K. K., and Park, Y. H., "Design Potential Method for Reliability-Based System Parameter Design Using Adaptive Probabilistic Constraint Evaluation," *AIAA Journal*, Vol. 39, No. 4, pp. 667-677, 2001.

Wang, A., "The Analysis of Bivariate Truncated Data Using the Clayton Copula Model," *The International Journal of Biostatistics*, Vol. 3, No. 1, 2007.

Wang, W., and Wells, M. T., "Model Selection and Semiparametric Inference for Bivariate fFilure-time Data with Discussion," *J. Am. Stat. Assoc.*, Vol. 95, No. 1, pp. 62-76, 2000.

Wei, D., "A Univariate Decomposition Method For Higher-Order Reliability Analysis And Design Optimization," Ph. D. Thesis, University of Iowa, 2006.

Wu, Y. T., Millwater, H.R., and Cruse, T.A., ‘‘Advanced Probabilistic Structural Analysis Method for Implicit Performance Functions,’’ *AIAA Journal*, Vol. 28, No. 9, pp. 1663–1669, 1990.

Wu, Y. T., ‘‘Computational Methods for Efficient Structural Reliability and Reliability Sensitivity Analysis,’’ *AIAA Journal*, Vol. 32 No. 8, pp. 1717–1723, 1994.

Xu, H., and Rahman, S., “A Moment-Based Stochastic Method for Response Moment and Reliability Analysis,” *Proceedings of 2nd MIT Conference on Computational Fluid and Solid Mechanics*, Cambridge, MA, July 17-20, 2003.

Youn, B. D., Choi, K. K., and Park, Y. H., “Hybrid Analysis Method for Reliability-Based Design Optimization,” *Journal of Mechanical Design*, Vol. 125, No. 2, pp. 221-232, 2003.

Youn, B.D., Choi, K. K., and Du, L., “Enriched Performance Measure Approach for Reliability-Based Design Optimization,” *AIAA Journal*, Vol. 43, No. 4, pp. 874-884, 2005.