Spring 2014

Efficient variable screening method and confidence-based method for reliability-based design optimization

Hyunkyoo Cho
University of Iowa

EFFICIENT VARIABLE SCREENING METHOD
AND CONFIDENCE-BASED METHOD
FOR RELIABILITY-BASED DESIGN OPTIMIZATION

by

Hyunkyoo Cho

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

May 2014

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

Hyunkyoo Cho

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Mechanical Engineering at the May 2014 graduation.

Thesis Committee:

Kyung K. Choi, Thesis Supervisor

Jia Lu

Shaoping Xiao

Olesya I. Zhupanska

David A. Lamb

ACKNOWLEDGMENTS

I first of all wish to thank my advisor, Professor Kyung K. Choi, who has not only taught me, but has inspired me to do this study. It has been a great honor to be his student, and he is the best advisor I can imagine. He spent an enormous amount of time and energy to help me make progress in this research. I especially appreciate his enthusiasm for research, which encouraged me to finish this study and my Ph.D. degree.

I also would like to thank Professor Seonho Cho, who led me to pursue a Ph.D. degree at the University of Iowa. All his advice became my source of energy for research. My special thanks go to my former and current colleagues. Professor Ikjin Lee helped me greatly in establishing the foundation of my research and making progress. Drs. Hyeongjin Song and Liang Zhao gave me invaluable advice and help. Mr. Nicholas Gaul provided me great help on software development. I also had the pleasure of working with Dr. Yoojeong Noh, Mr. Weifei Hu, Ms. Minyeong Moon, and Mr. Sangjune Bae.

Hearty thanks for advice on this study as well as thesis writing go to the committee members. Professors Jia Lu, Shaoping Xiao, and Olesya Zhupanska have shared their expertise in all matters of this study. They also gave invaluable advice, comments, and suggestions. Dr. David Lamb is also appreciated for all his support in my research and Ph.D. pursuit.

I wish to acknowledge the technical and financial support of the Automotive Research Center (ARC) in accordance with Cooperative Agreement W56HZV-04-2-0001, the U.S. Army Tank Automotive Research, Development and Engineering Center (TARDEC).

Finally, I would like to thank my parents and other members of my family whose support, patience, and love enabled this study as well as my Ph.D. pursuit.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

α	Significance Level
β	False Negative Rate
β_{HL}	Hasofer-Lint Reliability Index
β_t	Target Reliability Index
c, $c(u,v;\ \theta)$	Copula Density Function
CDF	Cumulative Distribution Function
CL^{Tar}	Target Confidence Level
CLT	Central Limit Theorem
C-RBDO	Confidence-Based RBDO
$\mathbf{d}$, d_i	Design Variable
DDO	Deterministic Design Optimization
DKG	Dynamic Kriging
DoE	Design of Experiment
DRM	Dimension Reduction Method
DSA	Design Sensitivity Analysis
$f(\mathbf{X})$, $f_{\mathbf{X}}(\mathbf{X})$	Probability Density Function of $\mathbf{X}$
FDM	Finite Difference Method
FORM	First-Order Reliability Method
G, $G(\mathbf{X})$	(Deterministic) Performance Measure
g, $g(\mathbf{U})$	Performance Measure in U-Space
γ	User Specified Parameter to Control Threshold in Variable Screening
GSA	Global Sensitivity Analysis
I-RBDO	Iowa Reliability-Based Design Optimization (Code)

LSA	Local Sensitivity Analysis
M	Input Distribution Type
MCS	Monte Carlo Simulation
MLE	Maximum Likelihood Estimator
MPP	Most Probable Point
MSE	Mean Square Error
$\boldsymbol{\mu}$, μ	Mean of Random Variable
N	Number of Random Variables (Dimension of Random Space)
NC	Number of Constraints
ND	Number of Data
NDV	Number of Design Variables
$NMCS_M$	Number of MCS Samples for Input Distribution Types
$NMCS_{\boldsymbol{\psi}}$	Number of MCS Samples for Input Distribution Parameters
nr	Number of Repeated Calculation of Partial Output Variance
ns	Number of Samples to Calculate Partial Output Variance
NVH	Noise, Vibration and Harshness
PDF	Probability Density Function
P_F, p_F	Probability of Failure = Reliability Output
p_F^{Tar}	Target Probability of Failure, Target Reliability Output
PMA	Performance Measure Approach
PRS	Polynomial Response Surface
$\boldsymbol{\psi}$	Input Distribution Parameter
RBDO	Reliability-Based Design Optimization
RBF	Radial Basis Function

RIA	Reliability Index Approach
RV	Random Variable
σ	Standard Deviation of Random Variable
$\mathbf{\Sigma}$	Covariance Matrix of Random Variable
$\sigma_{Y_i}^2$, $s_{Y_i}^2$	Partial Output Variance
SORM	Second-Order Reliability Method
SQP	Sequential Quadratic Programming
STDEV	Standard Deviation
$s_{v_i}^2$	Variance of Partial Output Variance
SVR	Support Vector Regression
τ	Kendall's Tau of Correlated Random Variables
θ	Correlation Coefficient for Copula
$\mathbf{u}$*	Most Probable Point in *U*-Space
U-Space	Independent Standard Normal Space
$\overline{v}_i$	Mean of Partial Output Variance
X, *X*	(Input) Random Variable
$\mathbf{x}$*	Most Probable Point in *X*-Space
X-Space	Original Random Variable Space
*$\mathbf{x}$	Input Data
*$\tilde{\mathbf{x}}$	Dispersion of Input Data (Invariant in C-RBDO Process)

CHAPTER 1
INTRODUCTION

In this study, newly developed methods to overcome curse of dimensionality and insufficient data problems in reliability-based design optimization (RBDO) are presented. To control the curse of dimensionality, a new variable screening method for RBDO is developed. In the variable screening method, output (a performance measure) variance is selected as the measure to identify important variables for RBDO. It is then found that partial output variance, which is calculated using a one-dimensional surrogate model, can approximate the output variance efficiently based on the univariate dimension reduction method (DRM). Large partial output variance indicates that the corresponding variable significantly affects variance of output. To choose important variables effectively, hypothesis testing is used to determine which variable has large partial output variance. Lastly, it is found that even quadratic interpolation is accurate enough for the one-dimensional surrogate model in the developed variable screening method.

To alleviate problems caused by lack of input data in RBDO, a confidence-based method for RBDO is developed. In confidence-based RBDO (C-RBDO), the reliability output (probability of failure) follows a certain distribution. To obtain the confidence level of the reliability output, the probability of the reliability output is broken down into the probability of input distribution parameters and the probability of input distribution types. Then, each probability is obtained using the Bayesian method and the data. The probability of the reliability output at a target reliability output is the confidence level of the target reliability output, and this is the probabilistic constraint of C-RBDO. Confidence-based RBDO is formulated to secure a target confidence level at the target reliability output at an optimum design, so that the optimum provides a conservative design even with limited data. Moreover, the design sensitivity of the confidence level is derived for an efficient and effective optimization process.

In Section 1.1, the background and motivation of this study is presented. Section 1.2 provides the objective of the proposed research, and Section 1.3 describes the organization of this thesis.

1.1 Background and Motivation

1.1.1 Reliability-Based Design Optimization

In the systems that engineers deal with, there are many sources of uncertainty. Material properties such as elastic modulus and Poisson's ratio, ocean waves, wind velocity, ground surface, electric current, climate, sunlight, and even human and machine error show an uncertain nature. Moreover, a group of these input uncertainties propagates through a performance measure to an output uncertainty. Therefore, an engineer who designs such a system should take the output uncertainty into consideration. If not, the designed system may not satisfy the required properties such as speed, fatigue life, strength, safety, noise, etc. Hence, reliability analysis methods, which can identify output uncertainty, and reliability-based design optimization (RBDO), which can find an optimum design considering output uncertainty, have been developed and used recently for many engineering applications.

Reliability analysis methods can be categorized into two groups: (1) sensitivity-based methods and (2) sampling-based methods. The sensitivity-based reliability analysis method requires the sensitivity (gradient) of performance measures to find the most probable point (MPP). The first-order reliability method (FORM) (Haldar and Mahadevan, 2000; Hasofer and Lind, 1974; Tu et al., 1999; Tu et al., 2001), the second-order reliability method (SORM) (Hohenbichler and Rackwitz, 1988; Breitung, 1984), and the dimension reduction method (DRM) (Rahman and Wei, 2006; Lee et al., 2010a) are representative sensitivity-based reliability analysis methods. Sensitivity-based methods first approximate a performance measure in independent standard normal space (*U*-space) and find the closest point on the limit state function (performance measure equals zero) from the origin in *U*-space. Then the closest point is the MPP, and the distance between the MPP and the origin represents the probability of failure. Then, FORM and SORM can approximate a performance measure at the MPP using first- and second-order Taylor series expansion, respectively, and DRM approximates a multi-dimensional performance function with a sum of lower-dimensional functions to calculate the probability of failure.

Reliability-based design optimization can be performed using sensitivity-based reliability analysis. Sensitivity-based reliability analysis can be categorized into (1) the reliability index approach (RIA) and (2) the performance measure approach (PMA). The RIA and PMA use different meanings of the MPP. The MPP in the RIA is one in the above-mentioned sensitivity-based reliability analysis; it represents the probability of failure at the current design. By contrast, the MPP in the PMA represents the target probability of failure (Tu et al., 1999). The PMA shows a more stable and robust nature because it searches for the MPP in a smaller area than the RIA. Both sensitivity-based RBDO methods require the additional information of the design sensitivity of a probabilistic constraint at the MPP. The design sensitivity for the RIA has been developed for FORM-based RBDO (Haldar and Mahadevan, 2000; Tu et al., 1999; Ditlevsen and Madsen, 1996; Hou, 2004; Gumbert et al., 2003; Hohenbichler and Rackwitz, 1988) and for the DRM-based RBDO (Rahman and Wei, 2008). The design sensitivity for the PMA has also been developed for FORM- and DRM-based RBDO (Gumbert et al., 2003; Hou, 2004; Lee et al., 2010b). All the methods for the design sensitivity of probabilistic constraints require the sensitivity (gradient) of the corresponding performance measure. However, the sensitivity of the performance measure is not always available, especially when the performance measure is complicated, highly nonlinear, and an implicit function.

In spite of great advances in sensitivity-based methods, they are not suitable for highly nonlinear problems. The MPP does not represent the correct (target) probability of failure in highly nonlinear problems. On the other hand, sampling-based reliability analysis and RBDO methods do not face such troubles because they calculate the probability of failure using Monte Carlo simulation (MCS). Monte Carlo simulation is known to be independent of nonlinearity and dimensionality. For the sampling-based method, the design sensitivity of probabilistic constraints has been developed for RBDO (Lee et al., 2011a; Lee et al., 2011b). Moreover, the sensitivity method does not require the sensitivity of a performance measure, so the sampling-based methods have more merit for complicated and nonlinear problems. The crucial demerit of the sampling-based method is its inefficiency. Monte Carlo simulation may require thousands of analyses of a performance measure, and it is almost impossible to provide so many analyses for a complicated problem. The surrogate model methods can alleviate the limit as they approximate performance measures with a small number of analyses. However, creating the surrogate model itself becomes challenging for high-dimensional problems. This will be discussed in detail in Section 1.1.2.

All of the aforementioned methods require an accurate input probabilistic model to assure the target reliability output (probability of failure) at the RBDO optimum design. The inaccuracy of the input probabilistic model has a significant impact on the confidence of the RBDO design. Due to cost and time, often insufficient data are provided for creation of an input probabilistic model in practical engineering problems. The limited data induces uncertainty in the input probabilistic model, and this uncertainty propagates to the reliability output. Therefore, the optimum design obtained using the aforementioned RBDO methods may not be reliable due to the lack of data. Hence, an appropriate conservativeness should be incorporated in the optimum design when only limited data are provided. This will be discussed in Section 1.1.3.

1.1.2 Variable Screening Method for RBDO

A design optimization problem is formulated using a set of design variables that describes the design system (Arora, 2004). The number of independent design variables is called the dimensionality of the optimization problem. To obtain an accurate optimum design, the system should be defined with correctly identified design variables. For this reason, a certain number of design variables will be required in the optimization process. On the other hand, in RBDO, the number, dimensionality, becomes crucial because RBDO requires a large number of analyses for reliability analyses and the design sensitivities of probabilistic constraints compared to deterministic design optimization (DDO).

As explained in Section 1.1.1, surrogate models have been used to reduce the number of analyses in RBDO. For decades, several surrogate model methods have been developed, such as the radial basis function (RBF), polynomial response surface (PRS), support vector regression (SVR), Kriging, and dynamic Kriging methods (Cressie, 1991; Barton, 1994; Jin et al., 2001; Simpson et al., 2001; Queipo et al., 2005; Wang and Shan, 2007; Forrester et al., 2008; Forrester and Keane, 2009; Zhao et al., 2011). Still, the number of design variables is a critical factor even for the surrogate models because generating them is a challenging task for high-dimensional problems, due to the curse of dimensionality. Therefore, variable screening is needed to create effective surrogate models of RBDO.

Several variable screening methods have been studied in statistics. An accurate Gaussian process model of computer simulation was generated using a reduced number of variables, which are found according to the maximum likelihood estimator (MLE) of the correlation parameters of the model for a deterministic problem (Welch et al., 1992). Based on the data, vital variables were efficiently identified in the pool of variables using a regression model (Duarte Silva, 2001; Wang, 2009). Moreover, feature selection methods, which select important variables for effective representation of output, have been researched (Guyon and Elisseeff, 2003). Furthermore, manifold learning has been studied to capture input information in a reduced dimension to perform the following statistical analysis efficiently (Izenman, 2008). In physics and engineering, variable screening methods have been researched as well. Significant variables were identified for the quasi-molecular treatment of ion-atom collision (Eichler and Wille, 1975). For a DDO problem of car crashworthiness, important variables were detected using the confidence intervals of the coefficients of a linear surrogate model (Craig et al., 2005). To assess the long-term performance of a geologic repository for high-level radioactive waste with small data and few variables, the sampling-based sensitivity measure ranked the importance of variables (Wu and Mohanty, 2006). Apart from the aforementioned variable screening methods, the design sensitivity analysis (DSA) method can be used for the same purpose because key variables have larger design sensitivity. For deterministic problems, the analytical DSA methods for the performance measures of various engineering problems are already well developed (Choi and Kim, 2005a; Choi and Kim, 2005b), and these methods have been categorized as local sensitivity analysis (LSA; Reedijk, 2000; Chen et al., 2005). For reliability problems, the randomness of the input variables is considered in probabilistic constraints. Therefore, the design sensitivity of the probabilistic constraints mentioned in Section 1.1.1 could be used to find out important variables for reliability problems. In addition, the correlation ratio (McKay et al., 1999), global sensitivity indices (Sobol, 2001), and analytical global sensitivity analysis (GSA) methods (Chen et al., 2005), which are categorized as GSA, can be effective variable screening methods as well (Mack et al., 2007).

However, the above-mentioned methods might not be directly applicable to creating an accurate surrogate model for RBDO. Methods totally based on existing data (Duarte Silva, 2001; Wang, 2009; Guyon and Elisseeff, 2003; Izenman, 2008) could not be applied to RBDO due to the change of the design in the optimization process. Since RBDO utilizes computer simulations such as the finite element method (FEM) and computational fluid dynamics (CFD), the relationship between input and output is known through the computer simulation. Hence, it may not be useful for RBDO to identify input variables from among all the variables that may be irrelevant to output (Guyon and Elisseeff, 2003). It is more of interest to find significant variables among relative variables using the computer simulation in this study. In view of the fact that RBDO requires estimation of how much output uncertainty is induced by the randomness of input variables, capturing input information in reduced dimension (Izenman, 2008) is not a topic for RBDO. By the same token, variable screening and DSA methods for deterministic problems (Welch et al., 1992; Craig et al., 2005; Choi and Kim, 2005a; Choi and Kim, 2005b) could not be used for RBDO. For broad application, a method developed for a specific problem (Eichler and Wille, 1975) might not be adequate. If a method requires a substantial number of analyses (McKay et al., 1999; Sobol, 2001), the analyses may not be affordable in computationally challenging RBDO problems. Also, methods could become unstable due to an insufficient number of analyses (Wu and Mohanty, 2006). As searching MPP in a high-dimensional problem could be expensive, so the design sensitivity of the sensitivity-based reliability method (Haldar and Mahadevan, 2000; Tu et al., 1999; Ditlevsen and Madsen, 1996; Hou, 2004; Gumbert et al., 2003; Hohenbichler and Rackwitz, 1988; Rahman and Wei, 2008; Lee et al., 2010b) does not offer an efficient variable screening method for RBDO. If a method requires accurate full-dimensional surrogate models (Chen et al., 2005; Lee et al., 2011a; Lee et al., 2011b), the surrogate models can be directly used for RBDO without variable screening. In summary, key desirable properties of a variable screening method for RBDO with a surrogate model are found from previous works: it should (1) be efficient, (2) consider input randomness, (3) not require a full-dimensional surrogate model, and (4) be applicable to broader problems.

1.1.3 Confidence-based RBDO

An input probabilistic model in a reliability problem follows a joint probability distribution due to the randomness of input variables. In addition, as briefly discussed in Section 1.1.1, the aforementioned RBDO methods require an accurate input probabilistic model – a so-called "true" input probabilistic model. Theoretically, a true input probabilistic model is obtainable only from population data, which are the test results of all subjects in the system. For example, the true model of the weight of elephants means literally the weights of "all" elephants on earth. Therefore, it is practically impossible to obtain true input probabilistic models in realistic problems. As a result, sampled data instead of population data are usually used to create an input probabilistic model. Of course, how the data is sampled is important, but the amount of data still dictates the accuracy of the input probabilistic model. Unfortunately, due to cost and time constraints, it is highly probable that only limited data are available for an input probabilistic model in a realistic problem. Then, the input probabilistic model generated using the limited data becomes uncertain. That is, the input probabilistic model shows a random feature on top of the randomness of the input variables. As a result, the uncertainty in the input probabilistic model forces the reliability output to be less confident.

Consequently, conservative, but not overly conservative, RBDO methods need to be investigated when only limited data are provided. A safety factor approach could be an intuitive start to considering the uncertainty in the input probabilistic model (Elishakoff, 2004). The P-boxes and the probability bounds, which are essentially a new input probabilistic model at a certain confidence level based on the data, are developed to capture the uncertainty in the input probabilistic model (Tucker and Ferson, 2003; Aughenbaugh and Paredis, 2006; Utkin and Destercke, 2009). The uncertainty in the input probabilistic model and the randomness of the input variables are combined in a modified input probabilistic model by using intentionally enlarged input variances (Noh et al., 2011a; Noh et al., 2011b). Here, it can be seen that all these methods adjust the input probabilistic model to consider the uncertainty in it. However, the uncertainty in the input probabilistic model transfers to the reliability output through a performance measure in a manner similar to how the randomness of the input variables propagates to the output uncertainty. When the performance is nonlinear, it is hardly possible to accurately estimate the uncertainty of the reliability output, which is induced by the uncertainty in the input probabilistic model, by simply altering the input probabilistic model. Moreover, modifying the input probabilistic model may mix the effect of the uncertainty in the input probabilistic model and the randomness of the input variables, which are essentially two different sources of uncertainty.

The Bayesian approach may satisfy the need to directly access the reliability output and to separate the effect of the uncertainty in the input probabilistic model and the randomness of the input variables. The fatigue reliability of a steel bridge was estimated by combining several input probabilistic models and two crack propagation models with the Bayesian method and nondestructive inspection (NDI) data (Zhang and Mahadevan, 2000), and the reliability could be updated as more NDI data became available. The mean of the simulation output was qualified in the presence of the uncertainty in the input probabilistic model using the Bayesian model average (BMA) approach (Chick, 2001), but the two uncertainty sources were not clearly distinguished yet. Later, Gunawan and Papalambros (2006) successfully separated the two sources and assumed that the reliability output follows beta distribution. That is, the reliability output, which quantifies the randomness of the input variables, also follows another distribution (the beta distribution in this case) due to the uncertainty in the input probabilistic model. The cumulative distribution function (CDF) of the beta distribution at certain reliability output is the confidence level of the reliability output. Using these observations, an RBDO problem of minimizing cost and maximizing confidence level was performed. Youn and Wang (2008) obtained an extreme case of the beta distribution using extreme distribution theory, and the median value of the extreme case was used as probabilistic constraints for RBDO. In addition, the design sensitivity of the probabilistic constraints was developed. However, the distribution of the reliability output still has not been fully utilized. Once the distribution is obtained, the confidence level of the reliability output is directly accessible. Therefore, a method that can find an optimum satisfying target reliability output as well as target confidence level can be developed.

Finally, from the previous works, the required properties of a new RBDO method for limited data can be found. The new method should (1) directly access the reliability output using the Bayesian approach, (2) distinguish the uncertainty in the input probabilistic model and the randomness of input variables, and (3) use the two target values of target reliability output and target confidence level. Ultimately, it would be better to (4) develop a design sensitivity for the new RBDO method to secure an effective and efficient optimization process.

1.2 Objectives of the Proposed Study

The first objective of this study is to develop a new variable screening method for RBDO with surrogate models. As discussed in Section 1.1.2, the new method pursues (1) efficiency, (2) consideration of randomness of input variables, (3) a method without a full-dimensional surrogate model, and (4) broad application. To reduce the dimensionality of a reliability problem, screened-out random variables should be fixed at a deterministic value. However, even if only one random variable is fixed, the total randomness of the input variable and the corresponding output uncertainty will be reduced. Hence, to correctly consider input randomness and solve a reliability problem with a reduced number of variables, the reduction of output uncertainty should be minimized. This is why output variance, which represents the output uncertainty, is used in this research as a measure to identify important variables. The variables that induce large output variance will be selected as important variables (Bae, 2012). To improve efficiency, the output variance is approximated with partial output variances, which is the output variance when a variable is random and others are fixed at their mean values, based on the univariate dimension reduction method (DRM; Rahman and Xu, 2004). Furthermore, the partial output variances require only one-dimensional surrogate models, which can be readily obtained with ease. Because the partial output variances can be calculated for any performance measure, they are applicable to broad problems. The variable screening method in this study emphasizes efficiency and practicality because those are very important for large-scale problems and the partial output variance satisfies the requirement.

The second objective of this study is to develop a new RBDO method that is able to find a reliable design even with limited data. As described in Section 1.1.3, the required properties for the new method are (1) direct access to the reliability output using the Bayesian approach, (2) separation of the uncertainty in the input probabilistic model and the randomness of input variables, (3) usage of two target values, and (4) design sensitivity. Uncertainty in the input probabilistic model forces the reliability output to follow a certain probabilistic distribution. Using Bayes' theorem and the given input data, the distribution of reliability output is decomposed into successive conditional probabilities of input distribution parameters and input distribution types (Cho et al., 2012). The probability of input distribution parameters is derived using the Bayesian method with non-informative priors and assuming that the limited data are under a normality condition. Furthermore, the probability of the input distribution type is acquired as well based on Bayes' theorem for the given input distribution parameters and the data. Then, the distribution of reliability output becomes directly accessible using MCS. Finally, knowing that the probability of the reliability output is the confidence level at a reliability output, a new confidence-based RBDO can be formulated to find an optimum that minimizes a cost function and satisfies the target confidence level at the target reliability output. As a result, a reliable, but not overly conservative, optimum design can be obtained using C-RBDO even with limited data. In addition, the analytical design sensitivity for the probabilistic constraint of C-RBDO, which is the confidence level at the target reliability output, has been developed as well. The developed C-RBDO method has great importance because it can solve more realistic RBDO problems.

1.3 Organization of Thesis

Chapter 2 presents basic concepts of reliability analysis and RBDO, which could be helpful to understand the proposed methods.

Chapter 3 proposes a new variable screening method for RBDO with surrogate models. First, output variance is introduced as a benchmark to determine important variables. Second, the developed method is explained in detail. Third, the output variance is approximated with partial output variances. Fourth, hypothesis testing is used to minimize error in the variable screening. Finally, the one-dimensional (1-D) surrogate model is tailored for the developed method.

Chapter 4 verifies the performance of the developed variable screening method for RBDO using analytical examples and a 44-dimensional (44-D) engineering problem. The developed method is compared with global sensitivity analysis (GSA) method. Reliability-based design optimization is performed to confirm the effectiveness of the developed method.

Chapter 5 proposes a new method to estimate the confidence level of reliability output. First, the probability of the reliability output is decomposed into successive conditional probabilities using the Bayesian method. Second, the conditional probabilities are obtained with reasonable assumptions, and a numerical method is proposed to calculate the confidence level.

Chapter 6 presents C-RBDO and its design sensitivity method. First, C-RBDO is formulated using the developed confidence level estimation method. Second, the design sensitivity of the confidence level is developed. Third, efficiency improvement methods are introduced to alleviate the computational cost of C-RBDO. Finally, C-RBDO and the design sensitivity method are tested using numerical examples.

Chapter 7 presents conclusions of this study and future recommendations.

CHAPTER 2

DESIGN UNDER UNCERTAINTY

In this chapter, fundamental ideas of design under uncertainty are reviewed. In Section 2.1, the reliability analysis method, which estimates the probability of failure or the reliability output, is explained. Then, the performance measure approach (PMA), which is an inverse reliability method, is presented in Section 2.2. Sensitivity-based RBDO and its design sensitivity are discussed in Section 2.3. Finally, recently developed sampling-based RBDO methods and stochastic design sensitivity are explained in Section 2.4.

2.1 Reliability Analysis

Reliability analysis is a method that calculates the probability of failure. The probability of failure p_F can be defined using a multi-dimensional integral as (Madsen et al., 1986)

$$p_F \equiv P[G(\mathbf{X}) > 0] = \int_{G(\mathbf{X})>0} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} \tag{2.1}$$

where $\mathbf{X} = [X_1, X_2, \ldots, X_N]^T$ is an N-dimensional vector of input random variables X_i, $G(\mathbf{x})$ is a performance measure function such that $G(\mathbf{x}) > 0$ is defined as failure, and $f_{\mathbf{X}}(\mathbf{x})$ is a joint probability density function (PDF) of the input random variables. In most engineering problems, the type of joint PDF is not Gaussian (multivariate normal distribution), and the performance measure $G(\mathbf{x})$ is nonlinear. In such problems, it is hardly possible to evaluate Eq. (2.1) analytically. There are two approaches to resolve this problem. In the first approach, the random variable vector $\mathbf{X}$, which follows a non-Gaussian PDF, is transformed into independent standard normal space (U-space) and the nonlinear performance measure is approximated with Taylor series expansion in the U-space. This first approach is called “sensitivity-based reliability analysis” since Taylor series expansion uses the sensitivity (gradient) of the performance measure $G(\mathbf{x})$. The transformation and approximation are explained in Sections 2.1.1 and 2.1.2. The second approach is to use the Monte Carlo simulation (MCS) method. Monte Carlo simulation can calculate Eq. (2.1) using samples that are drawn from input joint PDF $f_{\mathbf{X}}(\mathbf{x})$. Hence, the second approach is called "sampling-based reliability analysis" and is explained in Section 2.1.3.

2.1.1 Transformation

Let an N-dimensional random variable vector $\mathbf{X}$ has a joint cumulative distribution function (CDF) $F_{\mathbf{X}}(\mathbf{x})$. Then, the Rosenblatt transformation, which is denoted as $T : \mathbf{X} \rightarrow \mathbf{U}$, can transform $\mathbf{X}$ in X-space to a new random variable vector $\mathbf{U}$ in U-space as (Rosenblatt, 1952)

$$T : \begin{cases} u_1 = \Phi^{-1}\left[F_{X_1}\left(x_1\right)\right] \\ u_2 = \Phi^{-1}\left[F_{X_2}\left(x_2 \middle| x_1\right)\right] \\ \qquad \vdots \\ u_N = \Phi^{-1}\left[F_{X_N}\left(x_N \middle| x_1, x_2, \cdots, x_{N-1}\right)\right] \end{cases} . \tag{2.2}$$

where the conditional CDF $F_{X_i}\left(x_i \middle| x_1, x_2, \cdots, x_{i-1}\right)$ and standard normal CDF $\Phi(\bullet)$ are defined respectively as

$$F_{X_i}\left(x_i \middle| x_1, x_2, \cdots, x_{i-1}\right) = \frac{\int_{-\infty}^{x_i} f_{\mathbf{X}}(x_1, x_2, \cdots, x_{i-1}, \xi)\, d\xi}{f_{X_1 X_2 \cdots X_{i-1}}(x_1, x_2, \cdots, x_{i-1})}, \tag{2.3}$$

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \exp\left(-\frac{1}{2}\xi^2\right) d\xi . \tag{2.4}$$

In the transformed U-space, all random variables U_i follow standard normal distribution (normal distribution with zero mean and unit standard deviation), and they are statistically independent. Therefore, the problem due to non-Gaussian distribution of $\mathbf{X}$ is resolved in the new random variable **U**. Moreover, from Eq. (2.2), the inverse transformation can be obtained as

$$T^{-1}:\begin{cases} x_1 = F_{X_1}^{-1}\left[\Phi\left(u_1\right)\right] \\ x_2 = F_{X_2}^{-1}\left[\Phi\left(u_2 \middle| x_1\right)\right] \\ \qquad\qquad \vdots \\ x_N = F_{X_N}^{-1}\left[\Phi\left(u_N \middle| x_1, x_2, \cdots, x_{N-1}\right)\right] \end{cases} . \tag{2.5}$$

If all X_i are statistically independent from each other, the joint PDF of **X** becomes simply a multiplication of all the marginal PDFs of X_i as

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{N} f_{X_i}(x_i) \tag{2.6}$$

where $f_{X_i}(x_i)$ is the marginal PDF of X_i. Then, the conditional CDF in Eq. (2.3) becomes

$$F_{X_i}\left(x_i \middle| x_1, x_2, \cdots, x_{i-1}\right) = F_{X_i}(x_i) . \tag{2.7}$$

Consequently, the Rosenblatt transformation and its inverse in Eqs. (2.2) and (2.5) are simplified for independent random variables X_i as

$$u_i = \Phi^{-1}\left[F_{X_i}\left(x_i\right)\right] \text{ and } x_i = F_{X_i}^{-1}\left[\Phi\left(u_i\right)\right]. \tag{2.8}$$

For five commonly used distribution types, the Rosenblatt transformation and its inverse for independent random variables are shown in Table 2.1.

Table 2.1 The Rosenblatt Transformation and Its Inverse for Independent Random Variables

Types	Parameters	PDF	Transformations
Normal	μ : mean σ : standard deviation	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$, $x\in(-\infty,\infty)$	$u=(x-\mu)/\sigma$, $x=\mu+\sigma u$
Log-normal	$\bar{\sigma}^2=\ln\left[1+\left(\sigma/\mu\right)^2\right]$, $\bar{\mu}=\ln\mu-0.5\bar{\sigma}^2$	$f(x)=\dfrac{1}{\sqrt{2\pi}\bar{\sigma}x}e^{-\frac{1}{2}\left[\frac{\ln x-\bar{\mu}}{\bar{\sigma}}\right]^2}$, $x\in(0,\infty)$	$u=(\ln x-\bar{\mu})/\bar{\sigma}$, $x=\exp\left(\bar{\mu}+\bar{\sigma}u\right)$
Weibull	$\mu=v\Gamma\left(1+1/k\right)$, $\sigma^2=v^2\,\Gamma\left(1+2/k\right)-\mu^2$	$f(x)=\dfrac{k}{v}\left(\dfrac{x}{v}\right)^{k-1}e^{-(x/v)^k}$, $x\in[0,\infty)$	$u=\Phi^{-1}\left[1-e^{-(x/v)^k}\right]$, $x=v\left[-\ln\left\{1-\Phi\left(U\right)\right\}\right]^{1/k}$
Gumbel	$\mu=v+0.577/\alpha$, $\sigma=\pi/\sqrt{6}\alpha$	$f(x)=\alpha e^{-\alpha(x-v)-e^{-\alpha(x-v)}}$, $x\in(-\infty,\infty)$	$u=\Phi^{-1}\left[\exp\left\{-e^{-\alpha(x-v)}\right\}\right]$, $x=v-\ln\left[-\ln\left\{\Phi(u)\right\}\right]/\alpha$

2.1.2 First- and Second-Order Reliability Methods (FORM/SORM)

To calculate the probability of failure in Eq. (2.1), first the performance measure $G(\mathbf{x})$ is transformed to $g(\mathbf{u})$ in U-space using the Rosenblatt transformation. Then, the performance measure $g(\mathbf{u})$ is approximated at the most probable point (MPP) in U-space using Taylor series expansion. The MPP is the closest point on the limit state function, which means $g(\mathbf{u}) = 0$, to the origin in U-space and is denoted as $\mathbf{u}^*$. Hence, the MPP can be obtained by solving an optimization problem of

$$\begin{aligned} &\text{minimize} && \|\mathbf{u}\| \\ &\text{subject to} && g(\mathbf{u})=0. \end{aligned} \tag{2.9}$$

The concept of MPP is shown graphically in Figure 2.1. It is noted that $y(\mathbf{v})$ indicates $g(\mathbf{u})$ in Figure 2.1. Once MPP is found, the distance between MPP and the origin is the Hasofer-Lind reliability index β_{HL} (Hasofer and Lind, 1974).

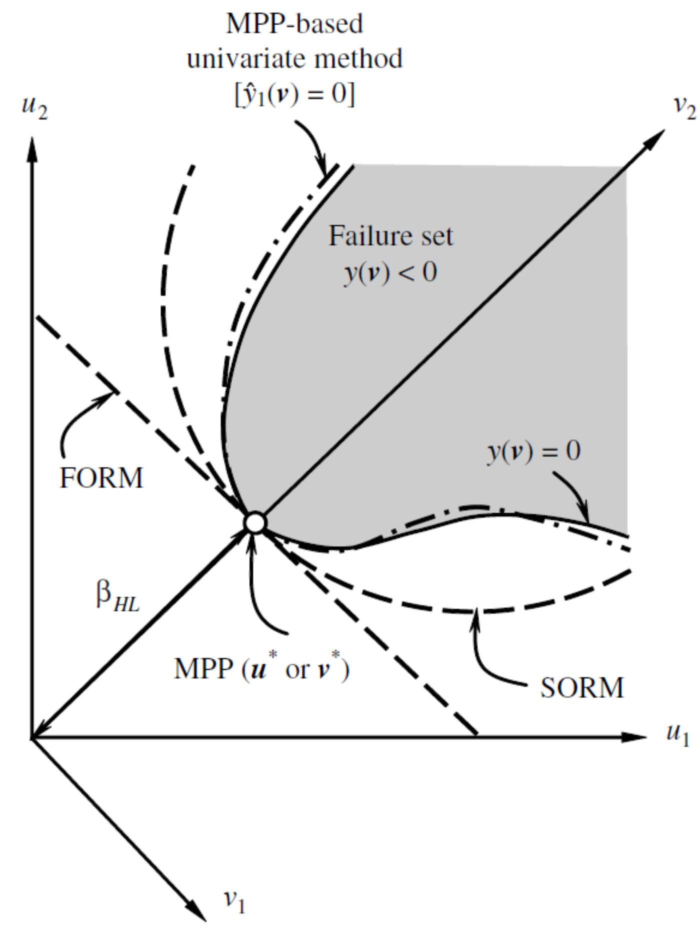


Figure 2.1 Concepts of MPP, Reliability Index β_{HL}, FORM and SORM

Source: Rahman, S., and Wei, D., "A Univariate Approximation at Most Probable Point for Higher-Order Reliability Analysis," *International Journal of Solids and Structures*, Vol. 43, No. 9, pp. 2820-2839, 2006.

The first-order reliability method (FORM) approximates the limit state function $g(\mathbf{u}) = 0$ using first-order Taylor series expansion at the MPP. Hence, the approximation requires the sensitivity (gradient) of the performance measure $g(\mathbf{u})$, and the approximated limit state function is linear in U-space (see Figure 2.1). Then the probability of failure in Eq. (2.1) can be estimated using FORM and the reliability index β_{HL} as

$$p_F^{\text{FORM}} \cong \Phi(-\beta_{HL}) \,. \tag{2.10}$$

For a highly nonlinear performance measure, FORM may not estimate the probability of failure correctly. For example, FORM overestimates the probability of failure in Figure 2.1. The failure region (or failure set) defined by FORM is much larger than the actual one. Hence, for more precise estimation, the second-order reliability method (SORM), which approximates the limit state function with second-order Taylor series expansion, can be used as (Breitung, 1984; Hohenbichler and Rackwitz, 1988)

$$p_F^{\text{SORM}} \cong \Phi(-\beta_{\text{HL}}) \left| \mathbf{I}_{N-1} - 2\frac{\phi(\beta_{\text{HL}})}{\Phi(-\beta_{\text{HL}})} \tilde{\mathbf{A}}_{N-1} \right|^{-\frac{1}{2}} \tag{2.11}$$

where N is number of random variables (dimension of given problem), the matrix $\mathbf{A}$ and its partitions are defined as

$$\mathbf{A} = \frac{1}{2\left\|\nabla_U g\right\|}\mathbf{R}^{\mathrm{T}}\mathbf{H}\mathbf{R} = \begin{bmatrix} \tilde{\mathbf{A}}_{N-1} & \tilde{\mathbf{A}}_{1N} \\ \tilde{\mathbf{A}}_{N1} & \tilde{\mathbf{A}}_{NN} \end{bmatrix}, \tag{2.12}$$

$\mathbf{H}$ is the Hessian matrix at the MPP in U-space, $\mathbf{R}$ is the rotation matrix such that $\mathbf{u} = \mathbf{R}\mathbf{v}$ (see Figure 2.1), and the PDF of the standard normal distribution $\phi(\bullet)$ is defined as

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right). \tag{2.13}$$

Still, it can be seen in Figure 2.1 that SORM might not be able to calculate the probability of failure when the performance measure $g(\mathbf{u})$ is highly nonlinear. To alleviate the problem caused by nonlinearity, the dimension reduction method (DRM) has been developed (Rahman and Xu, 2004; Xu and Rahman, 2004; Rahman and Wei, 2006; Rahman and Wei, 2008). As shown in Figure 2.1, DRM is able to approximate the limit state more accurately than FORM and SORM; hence it provides an accurate reliability analysis result.

2.1.3 Sampling-Based Reliability Analysis

The probability of failure in Eq. (2.1) can be directly calculated by applying the MCS method as (Lee et al., 2011a; Lee et al., 2011b)

$$\begin{aligned} p_F &= \int_{G(\mathbf{X})>0} f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x} = \int_{\mathbb{R}^N} I_{\Omega_F}(\mathbf{x})\,f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x} \\ &\cong \frac{1}{NMCS}\sum_{i=1}^{NMCS} I_{\Omega_F}[\mathbf{x}^{(i)}] \end{aligned} \tag{2.14}$$

where $\mathbf{x}^{(i)}$ is the i-th realization of $\mathbf{X}$ (i-th MCS sample), *NMCS* is the number of MCS samples, Ω_F is the failure domain such that $G(\mathbf{x}) > 0$, and $I_{\Omega_F}(\bullet)$ is an indicator function defined as

$$I_{\Omega_F}(\mathbf{x}) \equiv \begin{cases} 1, & \text{for } \mathbf{x} \in \Omega_F \\ 0, & \text{otherwise} \end{cases}. \tag{2.15}$$

Sampling-based reliability analysis is very straightforward, as shown in Eq. (2.14), and it does not require finding the MPP. Moreover, it does not involve sensitivity or the Hessian matrix of the performance measure, whereas they should be provided for FORM and SORM. However, the accuracy of MCS depends on the number of MCS samples *NMCS*. To calculate the probability of failure accurately, a large number of samples are required. This is further discussed in Section 2.4.

2.2 Performance Measure Approach (PMA)

The sensitivity-based reliability analysis in Section 2.1.2 is called the reliability index approach (RIA; Tu et al., 1999) because it finds the Hasofer-Lind reliability index β_{HL}. It is also shown that the probability of failure can be calculated with β_{HL} using FORM and SORM. On the other hand, the performance measure approach (PMA; Tu et al., 1999) does not estimate the probability of failure; it only judges whether a design satisfies a given target probability of failure p_F^{Tar}. Using Eq. (2.10) in FORM, the target probability of failure is transformed to target reliability index β_t as

$$\beta_t = -\Phi^{-1}(p_F^{Tar}). \tag{2.16}$$

Then, an optimization problem can be formulated using the target reliability index β_t as

$$\begin{aligned} &\text{maximize} && g(\mathbf{u}) \\ &\text{subject to} && \|\mathbf{u}\| = \beta_t. \end{aligned} \tag{2.17}$$

Here, the optimum point of Eq. (2.17) is also called "MPP" and denoted by $\mathbf{u}^*$. Let $g(\mathbf{u}) > 0$ mean failure of the design. If $g(\mathbf{u}^*)$ is less than zero, the design satisfies the target probability of failure. The PMA is well known for its efficiency and robustness (Tu et al., 1999; Tu et al., 2001; Choi et al., 2001; Youn et al., 2003). Several techniques have been developed to find the MPP effectively in the PMA such as the mean value (MV) method, advanced mean value (AMV) method (Wu et al., 1990; Wu, 1994), hybrid mean value (HMV) method (Youn et al., 2003), and enhanced hybrid mean value (HMV+) method (Youn et al., 2005). For the nonlinear performance measures, higher-order approximation is required for the PMA. Hence, the DRM has been applied also to the PMA for the nonlinear performance measures (Lee et al., 2010a).

2.3 Sensitivity-Based RBDO

A reliability-based design optimization (RBDO) problem can be formulated in general form as

$$\begin{aligned} &\text{minimize} && \text{Cost}(\mathbf{d}) \\ &\text{subject to} && P[G_i(\mathbf{X})>0] \le p_{F_i}^{Tar},\ \ i=1,\cdots,NC \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \quad \mathbf{d}\in\mathbb{R}^{NDV} \text{ and } \mathbf{X}\in\mathbb{R}^N \end{aligned} \tag{2.18}$$

where $\mathbf{X}$ is the N-dimensional random variable vector, $\mathbf{d}$ is the NDV-dimensional design variable vector, G_i is the i-th constraint function, $p_{F_i}^{Tar}$ is the target probability of failure for the i-th constraint, and NC is the number of constraints. In RBDO, a design variable d_i is the mean of the corresponding random variable X_j; hence $d_i = \mu(X_j)$. Among random variables, not all of them are related to the design variables. The variables that are not related to the design variables are called "random parameters." It is noted that the probabilistic constraint $P[G_i(\mathbf{X}) > 0]$ is used in Eq. (2.18), not the (deterministic) constraint function G_i.

Sensitivity-based RBDO solves Eq. (2.18) subject to a modified probabilistic constraint using sensitivity-based reliability design optimization. If RIA is used, the probabilistic constraint in Eq. (2.18) is changed to

$$P[G_i(\mathbf{X})>0] \le p_{F_i}^{Tar} \quad \rightarrow \quad \beta_i \le \beta_{t_i}\,, \tag{2.19}$$

where β_i and β_{t_i} are the probability index and target probability index for the i-th constraint, respectively. β_i is the same as β_{HL} for the i-th constraints in the FORM. For the SORM and DRM, β_i can be calculated as

$$\beta_i = -\Phi^{-1}(p_{F_i})\,, \tag{2.20}$$

where p_{F_i} is the probability of failure for the i-th constraint using the SORM or DRM. For an effective and efficient RBDO optimization process, the design sensitivity of the probabilistic constraint is necessary. As mentioned in Chapter 1, the design sensitivity for RIA has been developed for the FORM (Haldar and Mahadevan, 2000; Tu et al., 1999; Ditlevsen and Madsen, 1996; Hou, 2004; Gumbert et al., 2003; Hohenbichler and Rackwitz, 1988) and DRM (Rahman and Wei, 2008).

In RBDO, the actual value of the probability of failure is of no interest since eventually all the probabilistic constraints will satisfy the target probability of failure at the optimum design. Hence, the PMA, which judges only whether a design is safe or not, is more robust and adapts RBDO better than RIA. When the PMA is used, the constraint is changed to

$$P[G_i(\mathbf{X}) > 0] \le p_{F_i}^{Tar} \quad \rightarrow \quad G_i\,(\mathbf{x}^*) \le 0\,, \tag{2.21}$$

where $\mathbf{x}^*$ is the MPP of the PMA in X-space. The design sensitivity for PMA is also developed for FORM and DRM (Gumbert et al., 2003; Hou, 2004; Lee et al., 2010b). For FORM, the expression of the design sensitivity is

$$\frac{\partial G(\mathbf{x}^*)}{\partial \mathbf{d}} = \left[\frac{\partial \mathbf{x}}{\partial \mathbf{d}}\right]^T_{\mathbf{x}=\mathbf{x}^*} \left.\frac{\partial G}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^*} = \left.\frac{\partial G}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^*}. \tag{2.22}$$

2.4 Sampling-Based RBDO

For complicated engineering problems, the sensitivity or Hessian matrix of the constraint function $G(\mathbf{x})$ may not be available. Then, the sampling-based reliability analysis in Section 2.1.3 becomes prominent since it does not require any of them. Since the probability of failure is directly calculated using Eq. (2.14) in the sampling-based reliability analysis, no modification is needed for the probabilistic constraint in Eq. (2.18). The design sensitivity of the probabilistic constraint is developed using the score function and the MCS method (Lee et al., 2011a; Lee et al., 2011b). The design sensitivity method does not require extra MCS; it simultaneously calculates the design sensitivity during estimation of the probability of failure using the same MCS samples and constraint function evaluations.

First, the following four regularity conditions should be satisfied before derivation of the design sensitivity (Rubinstein and Shapiro, 1993; Rahman, 2009).

1. The joint PDF $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})$ is continuous.
2. The mean $\mu_i \in \mathrm{M}_i \subset \mathbb{R},\ i = 1, \cdots, N,$ where M_i is an open interval on $\mathbb{R}$.
3. The partial derivative $\partial f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})/\partial \mu_i$ exists and is finite for all $\mathbf{x}$ and μ_i. In addition, $p_F(\boldsymbol{\mu})$ is a differentiable function of $\boldsymbol{\mu}$.
4. There exists a Lebesgue integrable dominating function $r(\mathbf{x})$ such that

$$\left| h(\mathbf{x}) \frac{\partial f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} \right| \le r(\mathbf{x}) \tag{2.23}$$

for all $\boldsymbol{\mu}$. $h(\mathbf{x})$ is a general function and can be $I_{\Omega_F}(\mathbf{x})$.

If the four conditions are satisfied, taking the derivative of Eq. (2.14) with respect to μ_i yields (Lee et al., 2011a; Lee et al., 2011b)

$$\begin{aligned} \frac{\partial p_F(\boldsymbol{\mu})}{\partial \mu_i} &= \frac{\partial}{\partial \mu_i} \int_{\mathbb{R}^N} I_{\Omega_F}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu}) d\mathbf{x} \\ &= \int_{\mathbb{R}^N} I_{\Omega_F}(\mathbf{x}) \frac{\partial f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} d\mathbf{x} \\ &= \int_{\mathbb{R}^N} I_{\Omega_F}(\mathbf{x}) \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu}) d\mathbf{x} \\ &= E\left[I_{\Omega_F}(\mathbf{x}) \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} \right] \end{aligned} . \tag{2.24}$$

In Eq. (2.24), the derivative of the natural logarithm of the joint PDF $f_{\mathbf{X}}(\mathbf{x};\ \boldsymbol{\mu})$ with respect to μ_i is the first-order score function for μ_i, which is defined as

$$s_{\mu_i}^{(1)}(\mathbf{x};\boldsymbol{\mu}) \equiv \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} . \tag{2.25}$$

As explained before, the design sensitivity using the first-order score function for μ_i in Eq. (2.24) does not involve the sensitivity of the constraint function $G(\mathbf{X})$. It uses only the score function in Eq. (2.25), which can be obtained analytically. The reason is well illustrated in Figure 2.2. The horizontal axis is a random variable vector in a certain direction, and the vertical axis is the constraint function $G_j(\mathbf{X})$. Also, two examples of the input joint PDF $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu})$ are shown in Figure 2.2. If the failure region is set as $G_j(\mathbf{x}) > 0$, the gray area in Figure 2.2 represents the probability of failure. When the design variable $\boldsymbol{\mu}$ changes in the optimization process, the constraint function $G_j(\mathbf{X})$ holds its position, whereas the joint PDF $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu})$ moves along with the design variable. Hence, the gray area expands or shrinks due to the PDF movement, and the change rate of the gray area depends on the shape (slope) of the PDF on the limit state, which is related to the score function. This is why the design sensitivity is related to the score function, not the sensitivity of the constraint function $G_j(\mathbf{X})$ in Eq. (2.24).

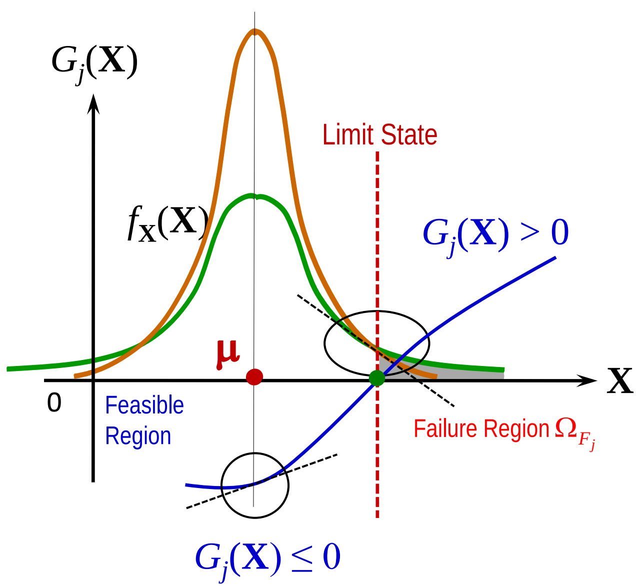


Figure 2.2 Concept of Design Sensitivity for Sampling-Based RBDO

For an independent input random variable X_i, the first-order score function for μ_i in Eq. (2.25) can be expressed with the marginal PDF $f_{X_i}(x_i;\mu_i)$ as

$$s_{\mu_i}^{(1)}(\mathbf{x};\boldsymbol{\mu}) \equiv \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} = \frac{\partial \ln f_{X_i}(x_i;\mu_i)}{\partial \mu_i}. \tag{2.26}$$

If the marginal PDF has an analytical expression as shown in Table 2.1, the derivation of the first-order score function for μ_i of the statistically independent random input is straightforward, as shown in Table 2.2.

Table 2.2 First-Order Score Function for μ_i of Independent Random Variables

Marginal dist. type	First-order score function, $s_{\mu_i}^{(1)}(\mathbf{x};\boldsymbol{\mu})$
Normal	$\dfrac{x_i-\mu_i}{\sigma_i^2}$
Log-normal	$-\dfrac{1}{\bar{\sigma}_i}\dfrac{\partial \bar{\sigma}_i}{\partial \mu_i}+\dfrac{1}{\bar{\sigma}_i^2}\left(\dfrac{\ln x_i-\bar{\mu}_i}{\bar{\sigma}_i}\right)\times\left[\bar{\sigma}_i\dfrac{\partial \bar{\mu}_i}{\partial \mu_i}+(\ln x_i-\bar{\mu}_i)\dfrac{\partial \bar{\sigma}_i}{\partial \mu_i}\right]$
Gumbel	$\alpha_i-\alpha_i e^{-\alpha_i(x_i-\nu_i)}$
Weibull	$\dfrac{1}{k_i}\dfrac{\partial k_i}{\partial \mu_i}-\dfrac{1}{\nu_i}\dfrac{\partial \nu_i}{\partial \mu_i}+\dfrac{\partial k_i}{\partial \mu_i}\ln\dfrac{x_i}{\nu_i}-\dfrac{(k_i-1)}{\nu_i}\dfrac{\partial \nu_i}{\partial \mu_i}-\left(\dfrac{x_i}{\nu_i}\right)^{k_i}\left(\dfrac{\partial k_i}{\partial \mu_i}\ln\dfrac{x_i}{\nu_i}-\dfrac{k_i}{\nu_i}\dfrac{\partial \nu_i}{\partial \mu_i}\right)$

Source: Lee, I., Choi, K. K., Noh, Y., Zhao, L., and Gorsich, D., "Sampling-Based Stochastic Sensitivity Analysis using Score Functions for RBDO Problems with Correlated Random Variables," *Journal of Mechanical Design*, Vol. 133, No. 2, pp. 021003, 2011.

When two random variables X_i and X_j are statistically correlated, the joint PDF of $\mathbf{X}=[X_i, X_j]^T$ can be expressed using marginal PDFs and copula as (Noh et al., 2009; Noh et al., 2010; Lee et al., 2011b)

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu}) &= \frac{\partial^2 C(u,v;\theta)}{\partial u \partial v} f_{X_i}(x_i;\mu_i) f_{X_j}(x_j;\mu_j) \\ &= C_{,uv}(u,v;\theta) f_{X_i}(x_i;\mu_i) f_{X_j}(x_j;\mu_j) \end{aligned} \tag{2.27}$$

where C is a copula function, $u = F_{X_i}(x_i;\mu_i)$ and $v = F_{X_j}(x_j;\mu_j)$ are marginal CDFs for X_i and X_j, respectively, and θ is the correlation coefficient between X_i and X_j for the copula function. Furthermore, the partial derivative of the copula function to u and v is the copula density function c as

$$c(u,v;\theta) \equiv \frac{\partial^2 C(u,v;\theta)}{\partial u \partial v} = C_{,uv}(u,v;\theta)\,. \tag{2.28}$$

Using Eqs. (2.27) and (2.28), the first-order score function for μ_i of the correlated random bivariate variable $\mathbf{X}$ can be derived as

$$s_{\mu_i}^{(1)}(\mathbf{x};\boldsymbol{\mu}) \equiv \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial \mu_i} = \frac{\partial \ln c(u,v;\theta)}{\partial \mu_i} + \frac{\partial \ln f_{X_i}(x_i;\mu_i)}{\partial \mu_i}\,. \tag{2.29}$$

The first term of the right-hand side of Eq. (2.29) for well-known copula types is listed in Table 2.3; the second term is already shown in Table 2.2. In Table 2.3, it can be seen that another derivative term, which is the derivative of the marginal CDF u to design variable μ_i, is required, and the term is derived for the representative distribution types in Table 2.4.

Table 2.3 Log-Derivative of Copula Density Function

Copula type	$\dfrac{\partial \ln c(u,v;\theta)}{\partial \mu_i}$
Clayton	$\left(-\dfrac{1+\theta}{u}+\dfrac{(2\theta+1)u^{-(1+\theta)}}{u^{-\theta}+v^{-\theta}-1}\right)\dfrac{\partial u}{\partial \mu_i}$
AMH	$\left[\dfrac{-\theta^2\left(1-v\right)+\theta(v+1)}{1+\theta^2\left(1-u\right)\left(1-v\right)-\theta\left(2-u-v-uv\right)}-\dfrac{3\theta(1-v)}{1-\theta\left(1-u\right)\left(1-v\right)}\right]\dfrac{\partial u}{\partial \mu_i}$
Frank	$\theta\left[\dfrac{2\left(e^{\theta(1+u)}-e^{\theta(u+v)}\right)}{e^{\theta}-e^{\theta(1+u)}-e^{\theta(1+v)}+e^{\theta(u+v)}}+1\right]\dfrac{\partial u}{\partial \mu_i}$
FGM	$\left[\dfrac{2\theta(2v-1)}{1+\theta\left(1-2u\right)\left(1-2v\right)}\right]\dfrac{\partial u}{\partial \mu_i}$
Gaussian	$\left[\dfrac{\Phi^{-1}(u)}{\phi(\Phi^{-1}(u))}+\dfrac{\theta\Phi^{-1}(v)-\Phi^{-1}(u)}{\phi(\Phi^{-1}(u))(1-\theta^2)}\right]\dfrac{\partial u}{\partial \mu_i}$

Source: Lee, I., Choi, K. K., Noh, Y., Zhao, L., and Gorsich, D., "Sampling-Based Stochastic Sensitivity Analysis using Score Functions for RBDO Problems with Correlated Random Variables," *Journal of Mechanical Design*, Vol. 133, No. 2, pp. 021003, 2011.

Table 2.4 Partial Derivatives of Marginal CDF with Respect to μ_i

Marginal distribution type	Partial derivatives of marginal distribution, $\dfrac{\partial u}{\partial \mu_i}$
Normal	$-\dfrac{1}{\sigma_i}\phi\left(\dfrac{x-\mu_i}{\sigma_i}\right)$
Log-normal	$-\dfrac{1}{\bar{\sigma}_i^2}\left(\dfrac{\partial\bar{\mu}_i}{\partial\mu_i}\bar{\sigma}_i+\dfrac{\partial\bar{\sigma}_i}{\partial\mu_i}(\ln x-\bar{\mu}_i)\right)\phi\left(\dfrac{\ln x-\bar{\mu}_i}{\bar{\sigma}_i}\right)$
Gumbel	$-\alpha_i e^{-\alpha_i(x-v_i)-e^{-\alpha_i(x-v_i)}}$
Weibull	$e^{-\left(\frac{x}{v_i}\right)^{k_i}}\left(\dfrac{x}{v_i}\right)^{k_i}\left(\dfrac{\partial k_i}{\partial\mu_i}\ln\dfrac{x}{v_i}-\dfrac{k_i}{v_i}\dfrac{\partial v_i}{\partial\mu_i}\right)$

Source: Lee, I., Choi, K. K., Noh, Y., Zhao, L., and Gorsich, D., "Sampling-Based Stochastic Sensitivity Analysis using Score Functions for RBDO Problems with Correlated Random Variables," *Journal of Mechanical Design*, Vol. 133, No. 2, pp. 021003, 2011.

Under the assumption that statistical correlation happens in between only two random variables, the joint PDF of a general random input **X**=[X_1, …, X_N]^T, which has *NCORR* correlated pairs, can be expressed as

$$f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})=\prod\nolimits_{j=1}^{NCORR}c_j(u,v;\theta_j)\prod\nolimits_{i=1}^{N}f_{X_i}(x_i;\mu_i) \tag{2.30}$$

Then, the first-order score function for μ_i of the general input **X** is derived as

$$s_{\mu_i}^{(1)}(\mathbf{x};\boldsymbol{\mu})\equiv\frac{\partial\ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu})}{\partial\mu_i}=\begin{cases}\dfrac{\partial\ln f_{X_i}(x_i;\mu_i)}{\partial\mu_i} & \text{for independent } X_i\\[2ex] \dfrac{\partial\ln c_j(u,v;\theta_j)}{\partial\mu_i}+\dfrac{\partial\ln f_{X_i}(x_i;\mu_i)}{\partial\mu_i} & \text{for correlated } X_i\end{cases}. \tag{2.31}$$

It can be seen that the terms on the right side of Eq. (2.31) are the same as those in Eqs. (2.26) and (2.29). Hence, the first-order score function can be obtained, and the design sensitivity in Eq. (2.24) becomes available for the general input random variable **X**.

For an effective and efficient RBDO process, the probability of failure and its design sensitivity need to be calculated accurately with MCS. The accuracy of the MCS depends on the number of MCS samples and the statistic under consideration. For the reliability analysis and RBDO, the statistic is the target probability of failure. Assuming that the error level of MCS is the same as the 95% confidence interval of the estimated probability of failure, the percentage error can be defined as (Haldar and Mahadevan, 2000)

$$\varepsilon\% = \sqrt{\frac{(1 - p_F^{Tar})}{NMCS \times p_F^{Tar}}} \times 200\% \tag{2.32}$$

where *NMCS* is the number of MCS samples and p_F^{Tar} is the target probability of failure. In Eq. (2.32), it is shown that *NMCS* should be increased to maintain a certain level of accuracy for a small target probability of failure. Since many analyses of the MCS samples are not affordable, surrogate models are commonly used for the sampling-based reliability analysis and RBDO as explained in Section 1.1.

Though the sampling-based reliability analysis and RBDO have efficiency issues, they still have attractive features already mentioned in this section. Hence, in the following chapters and sections, the sampling-based reliability analysis and RBDO are used as the main methods.

CHAPTER 3

VARIABLE SCREENING METHOD FOR RBDO

In this chapter, a new variable screening method for reliability-based design optimization (RBDO) is introduced. In Section 3.1, variable screening for deterministic design optimization (DDO) and RBDO are compared, and output variability (output variance) is found to be a benchmark to determine key variables. In Section 3.2, the developed method is explained in detail. First, output variability is quantified with partial output variances, using only one-dimensional (1-D) surrogate models. Then hypothesis testing is used to select key variables with minimum error. Finally, the 1-D surrogate model is tailored especially for the developed method.

3.1 Variable Screening

Screening out variables means finding important variables among all random variables. Here, the word "important" could have different meanings depending on the problem we are dealing with. In the following two sections, the differences between variable screening for DDO and RBDO are explained. Based on the difference, the required properties of variable screening for RBDO are introduced.

3.1.1 Variable Screening for DDO

A DDO problem can be formulated as

$$\begin{aligned} &\text{minimize} && \text{Cost}(\mathbf{d}) \\ &\text{subject to} && G_i(\mathbf{d}) \le 0, \ \ i = 1, \cdots, NC \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^{NDV} \end{aligned} \tag{3.1}$$

where $\mathbf{d}$, G_i, NC, and NDV are the design variable vector, i-th constraint function, number of constraints, and number of design variables, respectively.

In the DDO problem, the input design variables do not have uncertainty, and thus the design sensitivity can be used as a barometer to determine the importance ranking of design variables with respect to the performance measure (constraint function). The question is: "Where should the importance ranking of design variables be determined?" or "Where should the design sensitivity be calculated?"

Local sensitivity analysis (LSA) calculates the design sensitivity at a given design point (Reedijk, 2000; Chen et al., 2005). Usually, LSA is used to provide the direction of design movement in the optimization process. For variable screening, LSA can provide the importance ranking of design variables at the current design point. However, the importance ranking at the given design point could be different from the ranking at other design points if the performance measure is a nonlinear function of design variables. On the other hand, global sensitivity analysis (GSA) is used to calculate overall design sensitivity on the entire design domain. The GSA is like averaged design sensitivity in the design domain. As it is an average, the importance ranking using GSA could mislead at specific points or even regions. Hence, LSA and GSA have advantages and disadvantages for variable screening (Reedijk, 2000).

3.1.2 Variable Screening for RBDO

A general RBDO formulation in Eq. (2.18) is recalled as

$$\begin{aligned} &\text{minimize} && \text{Cost}(\mathbf{d}) \\ &\text{subject to} && P[G_i(\mathbf{X})>0] \le p_{F_i}^{Tar}, \ \ i=1,\cdots,NC \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^{NDV} \ \text{and} \ \mathbf{X} \in \mathbb{R}^N \end{aligned} \tag{3.2}$$

where $\mathbf{X}$ is an N-dimensional random variable vector, $\mathbf{d}$ is an NDV-dimensional design variable vector, G_i is the i-th constraint function, $p_{F_i}^{Tar}$ is the target probability of failure for the i-th constraint, and NC is the number of constraints.

In the RBDO process, design variable vector $\mathbf{d}$ is the mean vector of the corresponding random variable $\mathbf{X}$. Though the design variable $\mathbf{d}$ is deterministic, the design sensitivity for RBDO should consider the randomness of $\mathbf{X}$ because the constraints are based on the probabilistic performance measure $P[G_i(\mathbf{X})>0]$ as shown in Eq. (3.2).

Therefore, the design sensitivity of the performance measure alone cannot be used as a barometer. The design sensitivity of the probabilistic performance measure is introduced in Sections 2.3 and 2.4, and can be used for variable screening for RBDO. The design sensitivities by those methods are LSA because they provide different sensitivities at different designs. The GSA method is also applicable for variable screening in RBDO problems as it is in DDO problems. Again, both LSA and GSA methods have advantages and disadvantages.

The random parameters will not increase the dimensionality of the optimization problem because they are not random design variables. However, the surrogate model that includes random parameters is still required because they affect the output distribution. The main objective of this study is to select important design variables so that accurate surrogate models can be generated and, at the same time, an appropriate optimum design (i.e., not suboptimum) can be obtained in the RBDO process. Hence, once variable screening is done, the screened-out random design variables need to be fixed, not to be random parameters. However, fixing a random variable as a deterministic variable will reduce the total output variability.

Consider a simple example:

$$\begin{gathered} X_i \sim \mathcal{N}(5,3^2), \qquad i=1,2,\ldots,10 \\ Y=\sum_{i=1}^{10} X_i \sim \mathcal{N}\left(50,\left(3\sqrt{10}\right)^2\right). \end{gathered} \tag{3.3}$$

If the probabilistic performance measure is $P[Y > 60]$, then the reliability analysis result is

$$P\left[Y>60\right]=1-\Phi\left(\frac{60-50}{3\sqrt{10}}\right)=0.1459 \tag{3.4}$$

where $\Phi(\bullet)$ is cumulative distribution function (CDF) of standard normal distribution. However, if one dimension is reduced by screening out $X_{10} = \mu_{10} = 5$ while the other variables remain random, then the probabilistic performance measure changes to

$$\tilde{Y} = \sum_{i=1}^{9} X_i + 5 \sim \mathcal{N}(50, 9^2)\,. \tag{3.5}$$

It can be seen that output variability (variance of Y) is reduced from $\left(3\sqrt{10}\right)^2$ to 9^2. As a consequence, the reliability analysis result also changes as

$$P\left[\tilde{Y} > 60\right] = 1 - \Phi\left(\frac{60-50}{9}\right) = 0.1333\,. \tag{3.6}$$

From Eqs. (3.4) and (3.6), 0.0126 (1.26%) of the reliability output is decreased by screening out one variable. A more fundamental problem is that the lost amount 1.26% cannot be estimated without the full-dimensional reliability analysis result of Eq. (3.4). On the other hand, let's assume that X_{10} has a smaller variance of one. Then, the full-dimensional reliability analysis yields

$$P(Y > 60) = 1 - \Phi\left(\frac{60-50}{\sqrt{82}}\right) = 0.1347\,. \tag{3.7}$$

From Eqs. (3.6) and (3.7), the difference is 0.0014 (0.14%), which could be acceptable. Therefore, in this case, X_{10} could be fixed at the mean value. As shown in the example, the output variability decreases if any random variable is fixed at a deterministic value. However, there are some variables that affect the output variability a small amount. The variable screening method for effective surrogate models for RBDO is to find those variables that have small effects on the output variability. It is noted that the random parameters are considered as much as the random design variables in this study. Even though the random parameters are not changing during the RBDO process, they will influence the output variability. Hence they should be considered in the variable screening process, so that reliability of the performance measure can be accurately approximated using reduced dimension.

3.2 Variable Screening with 1-D Surrogate Model

The probability of failure cannot be solely determined by the output variability. To obtain accurate probability of failure, the output distribution is needed, that is, all statistical information of the output is required. However, even though an input distribution is known, it is very difficult to obtain complete output distribution since the performance measure could be implicit, a nonlinear function, or even both. For example, for a given normal input distribution, the output distribution could be bimodal as well as asymmetric. Consequently, it is impractical to select a reduced number of input variables based solely on probability of failure. As discussed in previous sections, a screened-out variable will be fixed at its mean value. Then the change of output mean will be minimized. As a result, the output variability becomes the measure that can determine a probability of failure. Of course, other statistical moments or parameters, such as skewness and kurtosis, could affect probability of failure. However, either of these statistical parameters cannot be a measure by itself. For example, a variable that induces larger (or smaller) output skewness may not be an important variable, but it could be an important variable when it induces larger (or smaller) output skewness and very similar output variability. We could consider a combination of moments as a measure, but there are too many possible combinations to consider. Hence, under the assumption that the output mean is similar, the output variability is chosen as the measure to select vital variables for RBDO in this study.

The output variability can be quantified by the output variance as shown in Section 3.1.2. The exact output variance of a nonlinear implicit performance measure is very difficult to obtain. Hence, an approximated output variance is used in this study. In the following sections, the output variance is decomposed into partial output variances, which are the output variances when each input variable is random and the others are fixed at their mean values. Then, a method to find the design variables that have a large impact on output variance is developed using a hypothesis testing.

3.2.1 Approximated Output Variance

A univariate dimension reduction method (DRM) is a well-known approximation method for statistical moments using multiple 1-D integrations (Rahman and Xu, 2004). Consider a performance measure Y and its realization y subject to N-dimensional input random vector $\mathbf{X}=[X_1, \ldots, X_N]^T$:

$$Y(\mathbf{X}) = Y(X_1, \ldots, X_N), \;\; y(\mathbf{x}) = y(x_1, \ldots, x_N)\,. \tag{3.8}$$

Define a function Y_i, which is the performance measure when X_i is random and other variables are fixed at their mean values, as

$$Y_i = Y(\mu_1, \ldots, \mu_{i-1}, X_i, \mu_{i+1}, \ldots, \mu_N)\,. \tag{3.9}$$

The realization of the performance measure at the input mean point $\boldsymbol{\mu}_\mathbf{X}$ is defined as

$$y_0 = y(\boldsymbol{\mu}_\mathbf{X})\,. \tag{3.10}$$

The l-th statistical moment of Y, which is approximated using the univariate DRM, is defined as (Rahman and Xu, 2004)

$$m_l \cong E\left[\left\{\sum_i Y_i - (N-1)y_0\right\}^l\right]. \tag{3.11}$$

Then, the output variance σ_Y^2 can be approximated as

$$\sigma_Y^2 \cong m_2 - m_1^2 = \sum_i \sigma_{Y_i}^2 + 2\sum_{i>j} \rho_{Y_iY_j}\sigma_{Y_i}\sigma_{Y_j} \tag{3.12}$$

where $\sigma_{Y_i}^2$ is the variance of Eq. (3.9), which is the partial output variance when only X_i is random, and $\rho_{Y_iY_j}$ is the correlation coefficient between Y_i and Y_j. As shown in Eq.

(3.12), the partial output variances $\sigma_{Y_i}^2$ are the main variables for approximating the output variance σ_Y^2. When $\sigma_{Y_i}^2$ is larger than other partial output variances, it takes the largest portion in the output variance σ_Y^2. Therefore, if some X_i produces larger partial output variance than others, then X_i should be selected as an important variable. It is noted that calculation of $\sigma_{Y_i}^2$ requires only 1-D integration, and thus only 1-D surrogate models are required to calculate the partial output variances.

Statistical correlation between X_i and X_j yields the term of $\rho_{Y_iY_j}\sigma_{Y_i}\sigma_{Y_j}$ in Eq. (3.12) and affects the output variance. When X_i and X_j are strongly correlated, one could be replaced by the other. To calculate the term $\rho_{Y_iY_j}\sigma_{Y_i}\sigma_{Y_j}$, a two-dimensional surrogate model is required. If there are only a few correlation pairs, calculating the correlation term could be affordable. However, with a practical point of view, the partial output variance $\sigma_{Y_i}^2$ is focused in this study. As we are looking for important variables, not the value of the output variance σ_Y^2, the partial output variance would be enough for variable screening. In Figure 3.1, contours of independent, positively correlated ($\rho = 0.8$), and negatively correlated ($\rho = -0.8$) probability density functions are shown. Correlation determines how the random variables are distributed inside the box (dotted line), whereas the size of the box is determined by variances of X_1 and X_2. It can be seen that the primary effect on output variance is the box size, and then the distribution inside the box follows. Consequently, to perform variable screening efficiently, the first thing we need to consider is the box size, not the distribution of random variables inside the box. Hence, the correlation term is not considered in this study for efficiency and practicality. It is noted that the statistical correlation between X_i and X_j will be considered in reduced-dimensional RBDO if both variables are selected.

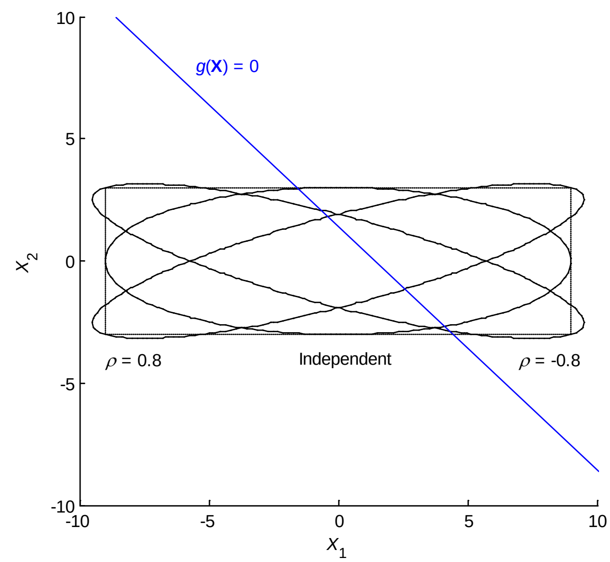


Figure 3.1 Effect of Variance and Correlation of Input Random Variables

The partial output variance of $\sigma_{Y_i}^2$ is like LSA because it can have different values at different input mean points $\boldsymbol{\mu}_{\mathbf{X}}$, which represents the current design point in the RBDO process. Hence, the variable screening result could be changed as the design point changes. There are several recommended points at which to perform variable screening using LSA. The first one is the DDO optimum. As the DDO optimum is usually close to the RBDO optimum, the variable screening result at the DDO optimum is likely to be similar to the result at the RBDO optimum. Also, the design point where most of the deterministic constraints are active can be a good candidate point. It is noted that DDO or the design point where the constraints are active could be obtained using the finite difference method (FDM) in a practical engineering problem. Also, DDO could be achieved using the sensitivity obtained from a 1-D surrogate model because DDO requires only the deterministic LSA, which is 1-D.

3.2.2 Variable Screening Using Hypothesis Testing

Using the 1-D surrogate model, the partial output variance $\sigma_{Y_i}^2$ can be calculated approximately as

$$s_{Y_i}^2 = \frac{1}{ns-1}\sum_{j=1}^{ns}\left\{y_i(\mu_1,\ldots,x_i^{(j)},\ldots,\mu_N)-\bar{y}_i\right\}^2 \tag{3.13}$$

where $x_i^{(j)}$ is the j-th realization of the input random variable X_i, ns is the number of samples, and $\bar{y}_i$ is the mean of y_i as

$$\bar{y}_i = \frac{1}{ns}\sum_{j=1}^{ns} y_i(\mu_1,\ldots,x_i^{(j)},\ldots,\mu_N)\,. \tag{3.14}$$

As explained in previous sections, the partial output variance $s_{Y_i}^2$ can be used to determine important design variables. To make the variable screening procedure systematic, hypothesis testing is applied in this study. Hypothesis testing can prevent undesirable choices that could occur during the decision-making procedure. Calculated partial output variance $s_{Y_i}^2$ depends on the number of samples *ns*. When *ns* is large enough, the variable screening result will be accurate. However, it would require a large computational time. Also, it is hard to determine when *ns* is large enough. When *ns* is small, it will include statistical error. If calculated $s_{Y_i}^2$ are distinctive from each other or with respect to the screening threshold value, then the effect of *ns* may not be significant. However, *ns* could cause an error when some $s_{Y_i}^2$ are similar to each other or are near the screening threshold value. Hypothesis testing can prevent this problem in a statistical manner by letting users control the error level.

Various hypothesis testing methods have been developed for the decision-making problem (Rosner, 2006). Among those methods, we need the one that is not sensitive to distribution type because the distribution type of Y_i or $s_{Y_i}^2$ is not known in general. The one-sample t-test is developed based on the central limit theorem (CLT), which states that the sample mean of non-normal distribution follows normal distribution approximately for a large number of samples. The one-sample *t*-test is not sensitive to underlying distribution types, so it is used in this study. As the *t*-test is a method for sample mean, $s_{Y_i}^2$ is calculated *nr* times for its statistical moments as

$$\overline{v}_i = \frac{1}{nr}\sum_{k=1}^{nr} s_{Y_i}^{2(k)}, \tag{3.15}$$

$$s_{v_i}^2 = \frac{1}{nr-1}\sum_{k=1}^{nr} (s_{Y_i}^{2(k)} - \overline{v}_i)^2, \tag{3.16}$$

where $s_{Y_i}^{2(k)}$ is the *k*-th repetition of $s_{Y_i}^2$ and *nr* is the number of repetitions. Now, the hypothesis is constructed:

$$H_0 : \overline{v}_i \le \mu_0 \text{ versus } H_1 : \overline{v}_i > \mu_0 \tag{3.17}$$

where μ_0 is the criterion of hypothesis testing. According to Eq. (3.17), the design variable that corresponds to $\overline{v}_i$, which is greater than μ_0 (H_1 is true), will be selected as an important variable. Using the one-sample *t*-test, the hypothesis can be tested by checking the following statement:

$$\text{Reject } H_0 \text{ in favor of } H_1 \text{ if } q \ge t_{nr-1,1-\alpha} \tag{3.18}$$

where α is the significance level, $t_{nr-1,\bullet}$ is $t_{nr-1}^{-1}(\bullet)$, and the test statistics q is defined as

$$q \equiv \left(\overline{v}_i - \mu_0\right) \Big/ \frac{s_{v_i}}{\sqrt{nr}}. \tag{3.19}$$

In Eqs. (3.15) and (3.16), the uncertainty induced by *ns* is transferred to *nr*. Hence, *ns* can be a fixed number, whereas *nr* should be decided appropriately. Also, μ_0 needs to be identified in Eqs. (3.17) and (3.19). μ_0 is the key criterion that decides important variables, and it should be a value relative to $\overline{v}_i$ because the relative difference of partial output variances should be checked for variable screening. At the same time,

μ_0 needs to be statistically independent from $\overline{v}_i$ for reasonable hypothesis testing. In this study, preliminary testing is proposed to obtain reasonable *nr* and μ_0 as follows. First, choose nr_0, which is large enough so that the CLT holds. Then, calculate the initial statistical moments of $s_{Y_i}^2$ as

$$\overline{v}_i^{(0)} = \frac{1}{nr_0}\sum_{k=1}^{nr_0} s_{Y_i}^{2(k)} , \tag{3.20}$$

$$s_{v_i}^{2(0)} = \frac{1}{nr_0 - 1}\sum_{k=1}^{nr_0} (s_{Y_i}^{2(k)} - \overline{v}_i^{(0)})^2 . \tag{3.21}$$

Using the value from Eq. (3.20), the testing criterion μ_0 relative to $\overline{v}_i$ can be calculated as

$$\mu_0 = \frac{\gamma}{N}\sum_{i=1}^{N} \overline{v}_i^{(0)} \tag{3.22}$$

where γ is a constant that the user selects. *nr* is calculated by limiting type II error (H_0 is accepted when H_1 is true) at the level of false negative rate β as (Rosner, 2006)

$$nr = \max\left(\frac{s_{v_i}^2 (t_{nr_0-1,1-\alpha} + t_{nr_0-1,1-\beta})}{\overline{v}_i^{(0)} - \mu_0}, nr_0 \right) \tag{3.23}$$

In Eq. (3.23), $t_{nr-1,\bullet}$ should be used instead of $t_{nr_0-1,\bullet}$ for accurate calculation of *nr*. However, Eq. (3.23) requires the value of *nr* on the right side to calculate *nr*. To avoid this problem, $t_{nr_0-1,\bullet}$ is used instead, and $t_{nr_0-1,\bullet}$ produces a conservative result as it is larger than $t_{nr-1,\bullet}$ because *nr* is larger than nr_0 in Eq. (3.23) and α is usually small. Finally, *nr* and μ_0 are determined so that the proposed hypothesis testing can be utilized.

3.2.3 1-D Surrogate Model

In previous sections, the 1-D surrogate model is treated as the given one because it is not difficult to generate. However, efficiently creating a 1-D surrogate model could be an issue. For efficiency, quadratic interpolation is proposed as a basic 1-D surrogate model in this study. Quadratic interpolation may not be an adequate method for creating a surrogate model for a highly nonlinear performance measure. However, a nonlinear performance measure can be effectively approximated by a quadratic function on a small region. If X follows a normal distribution, 99.73% of X is in $(\mu_X - 3\sigma_X, \mu_X + 3\sigma_X)$, which is much smaller than the infinite domain. Even if X does not follow normal distribution, the interval $(\mu_X - 3\sigma_X, \mu_X + 3\sigma_X)$ can cover almost all (approximately 98%) of X. In view of the fact that this study is focused on calculation of partial output variance, the region $(\mu_X - 3\sigma_X, \mu_X + 3\sigma_X)$ is large enough. Hence, the 1-D surrogate model needs to be accurate in the region $(\mu_X - 3\sigma_X, \mu_X + 3\sigma_X)$ so that quadratic interpolation could be the appropriate method to approximate the performance measure in the region.

Quadratic interpolation requires three design of experiments (DoE) samples, and the location of DoE samples affects the accuracy of interpolation. The location of DoE samples determined using the Chebyshev polynomial is known to give uniform error in the domain (Rao, 2002). Because only the region $(\mu_X - 3\sigma_X, \mu_X + 3\sigma_X)$ is of interest, the location of DoE samples is determined as $x_1 = \mu_X - 2.5981\sigma_X$, $x_2 = \mu_X$ and $x_3 = \mu_X + 2.5981\sigma_X$ using the Chebyshev polynomial. Since a random variable X may not be evenly distributed in its domain, providing uniform error does not necessarily mean that the calculated partial output variance is accurate. However, since no unique location of DoE samples is best for accurate partial output variance, the sample location by the Chebyshev polynomial is used in this study because it yields reasonable results for various distribution types of random variable X. If the random variable X has a closed and bounded domain like $[a, b]$, the domain can be directly used for calculation of partial output variance, and the location of DoE samples are $x_1 = 0.93301a + 0.06699b$, $x_2 = (a+b)/2$ and $x_3 = 0.06699a + 0.93301b$, using Chebyshev polynomials.

To check the performance of a selected location of DoE samples, a nonlinear performance measure Y is used as

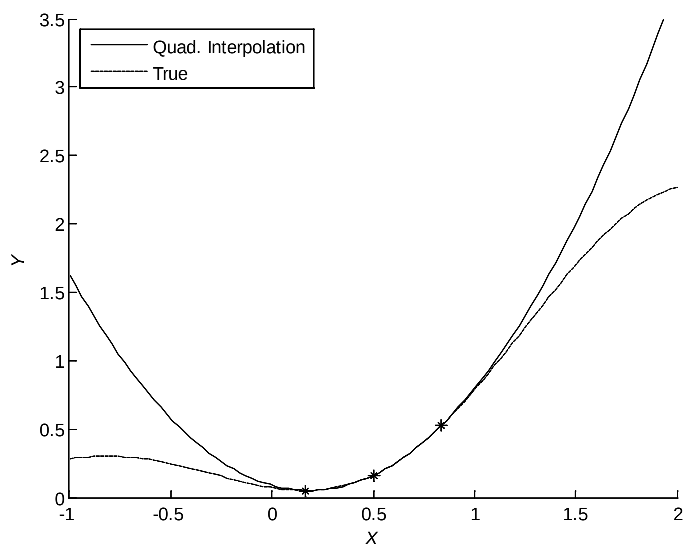
$$Y(X) = 0.3 + \sin(16X/15 - 0.7) + \sin^2(16X/15 - 0.7)\,. \qquad (3.24)$$

Assuming that random variable X follows $\mathcal{N}(0.5, 0.333^2)$, three locations of DoE samples are chosen to compare the accuracy of the partial output variance. The first location is {0.167, 0.5, 0.833}, which is μ_X and $\mu_X \pm \sigma_X$, and the second location is from the Chebyshev polynomial as {−0.365, 0.5, 1.365}. The third location is wider, as {−0.667, 0.5, 1.667}, which is μ_X and $\mu_X \pm 3.5\sigma_X$. Partial output variances are calculated using 100,000 realizations of X, and true partial output variance is calculated by Eq. (3.24) with the same realizations. To check the accuracy of the quadratic interpolation itself, mean square error (MSE) is calculated in the region of (−0.5, 1.5), which is $(\mu_X - 3\sigma_X, \mu_X + 3\sigma_X)$ with 100 uniformly distributed points. The calculated result is shown in Table 3.1, and the shape of quadratic interpolations is shown in Figure 3.2, where asterisk marks (*) represent the DoE sample points. As shown in Table 3.1, the location of the DoE sample using Chebyshev polynomials produces more accurate partial output variance compared to the true one and less MSE than the other cases.

This example cannot represent all performance measures. When a highly nonlinear performance measure is expected, more sophisticated surrogate methods, such as the radial basis function (RBF), polynomial response surface (PRS), support vector regression (SVR), Kriging, and dynamic Kriging (Cressie, 1991; Barton, 1994; Jin et al., 2001; Simpson et al., 2001; Queipo et al., 2005; Wang and Shan, 2007; Forrester et al., 2008; Forrester and Keane, 2009; Zhao et al., 2011) methods, are better. In any case, it is recommended to sample inside the interval of $(\mu_X - 3\sigma_X, \mu_X + 3\sigma_X)$ for the random variable X if the distribution has an infinite domain.

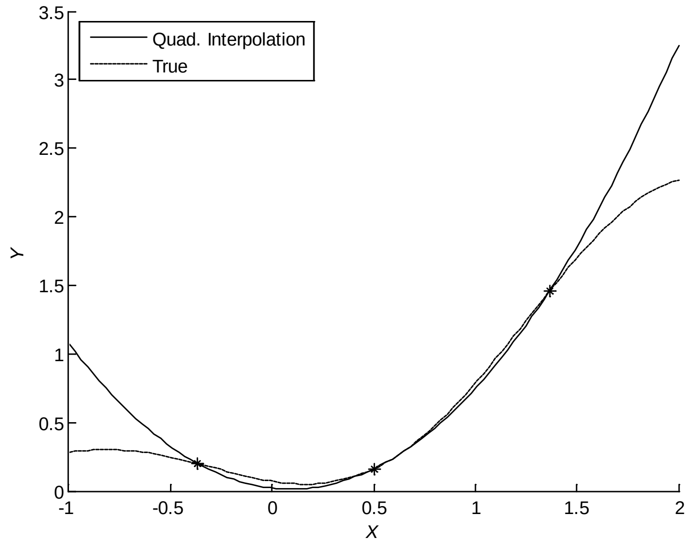
Table 3.1 Quadratic Interpolation with Different Sample Locations

<table>
<tr><td colspan="2" rowspan="2">Location of samples</td><td colspan="3">Quadratic interpolation</td><td rowspan="2">True</td></tr>
<tr><td>{0.167, 0.5, 0.833}</td><td>{−0.365, 0.5, 1.365}</td><td>{−0.667, 0.5, 1.667}</td></tr>
<tr><td rowspan="2">Partial output variance</td><td>Value</td><td>8.74E-02</td><td>7.84E−02</td><td>6.92E−02</td><td>7.84E−02</td></tr>
<tr><td>Accuracy</td><td>111.53%</td><td>100.01%</td><td>88.33%</td><td>100.00%</td></tr>
<tr><td colspan="2">MSE</td><td>1.17E−02</td><td>1.19E−03</td><td>9.00E−03</td><td>0</td></tr>
</table>

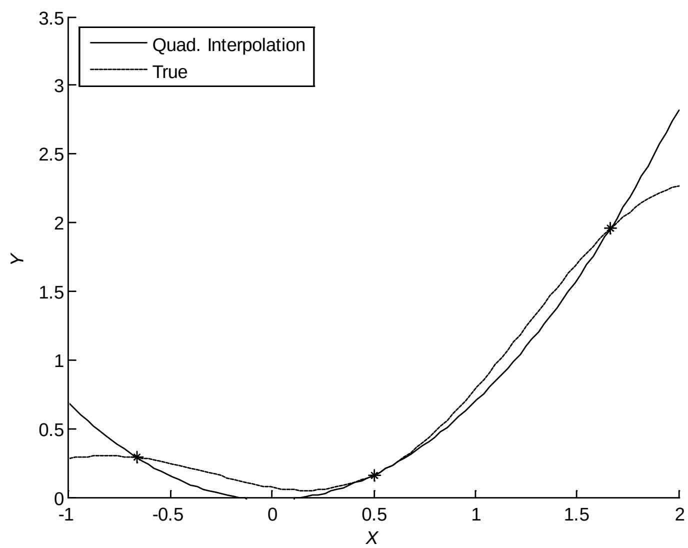


(a) Location of sample: {0.167, 0.5, 0.833}

Figure 3.2 Quadratic Interpolation of *Y* with Different Locations of Samples



(b) Location of sample: {−0.365, 0.5, 1.365}



(c) Location of sample: {−0.667, 0.5, 1.667}

Figure 3.2 Continued

CHAPTER 4
NUMERICAL EXAMPLES OF THE VARIABLE SCREENING METHOD FOR RBDO

In this chapter, the effectiveness of the developed variable screening method is verified using numerical examples. In Section 4.1, user-specified parameters for the developed variable screening method are explained. In Section 4.2, the performance of the present method is tested using two analytical examples. In Section 4.3, the developed variable screening method is applied to a 44-D engineering problem, and the result is compared with the one from a global sensitivity analysis (GSA) method. In addition, reliability-based design optimization (RBDO) is performed for various variable selections, the results are compared, and the effectiveness of the method is confirmed.

4.1 User-Specified Parameters for Variable Screening

For the numerical examples, partial output variances are calculated to select important variables using the 1-D quadratic interpolation presented in Section 3.2.3. To use the variable screening method, five parameters: significance level α, false negative rate β, number of sample *ns*, initial number of repetition $nr_{0,}$ and control parameter γ for threshold value, need to be decided by users. Smaller α and β are better choices because they result in fewer statistical errors in the variable screening method. However, when they are too small, very large *nr* could be required to maintain the error level specified by α and β in Eq. (3.23). Hence, 0.025 to 0.05 would be a reasonable choice for them. For nr_0 and *ns*, a small number could be chosen to reduce computational cost. However, a small value of nr_0 and *ns* will rapidly increase *nr* to maintain the error levels. Hence, an appropriately large number should be used; they are set to 50 or 100 in the numerical examples. The parameter γ in Eq. (3.22) is for user control of the threshold value that determines important variables. In the numerical examples, γ is initially set to 1.0, and the variable screening procedure is performed. Then, the ratio of the sum of partial output variances of selected variables to that of all random variables, which is an estimation of the captured output variance, is checked. If the ratio is less than 85%, γ is lowered to achieve 85%. As explained before, the partial output variance is an approximation method, which is why the ratio of 85% may not mean that 85% of total output variance is actually captured in the selected variables. However, it would be a good estimation with affordable cost because it does not require many design of experiments (DoE) samples or full-dimensional surrogate models.

4.2 Analytical Examples

Hartmann 6-D and Dixon-Price 12-D are well-known analytical functions. They are high-dimensional as well as nonlinear, so they are tested for the variable screening method. Constant terms are added to the original functions to make both functions active (i.e., $G(\mathbf{x}) = 0$) at the mean point of the input random variables. Note that adding a constant term does not change the character of the functions. Input random variables have a variety of marginal distribution types and copula types, so the analytical examples can reveal the effects of different distribution types and correlations.

The functions are tested with three different methods. The first is the developed variable screening method. As mentioned before, parameters of $\alpha = \beta = 0.025$, $nr_0 = ns = 100$ and $\gamma = 1.0$, and 1-D surrogate model with quadratic interpolation are used. γ is initially set to 1.0 and lowered when necessary. The second is screening with accurate partial output variances using the analytical functions directly and 1,000,000 realizations of random variables. The calculated partial output variances are used for reference. When a performance measure is a linear function ($G = \Sigma\alpha_i X_i$) of the input random variables X_i's, the output variance is $\Sigma\alpha_i^2\sigma_{X_i}^2$, where $\sigma_{X_i}^2$ is a variance of X_i. Hence, the partial output variance can be linearly approximated as $\alpha_i^2\sigma_{X_i}^2$ with a design sensitivity (gradient) α_i and input variance $\sigma_{X_i}^2$. Furthermore, an important variable might be selected based on the partial output variances calculated with the sensitivity-variance method, and it is applied to the analytic functions for comparison.

4.2.1 Hartmann 6-D

The first analytical example is the Hartmann 6-D, and a constant term is added as explained earlier. The analytical expression is shown as (Dixon and Szegö, 1978)

$$G(\mathbf{X}) = -\sum_{i=1}^{q} a_i \exp\left(-\sum_{j=1}^{m} b_{ij}(X_j - d_{ij})^2\right) + 3.3082 \tag{4.1}$$

where $0 \le X_i \le 1$, $m = 6$, $q = 4$ and

$$\mathbf{a} = \begin{bmatrix} 1.0 & 1.2 & 3.0 & 3.2 \end{bmatrix}, \tag{4.2}$$

$$\mathbf{b} = \begin{bmatrix} 10.0 & 3.0 & 17.0 & 3.5 & 1.7 & 8.0 \\ 0.05 & 10.0 & 17.0 & 0.1 & 8.0 & 14.0 \\ 3.0 & 3.5 & 1.7 & 10.0 & 17.0 & 8.0 \\ 17.0 & 8.0 & 0.05 & 10.0 & 0.1 & 14.0 \end{bmatrix}, \tag{4.3}$$

$$\mathbf{d} = \begin{bmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{bmatrix}. \tag{4.4}$$

Information about input random variables is listed in Table 4.1. Input random variables have four different marginal distribution types: normal, lognormal, gamma, and Weibull. X_5 and X_6 are correlated with the Clayton copula and Kendall's tau of 0.5.

Table 4.1 Input Random Variables for Hartmann 6-D Example

Random variable	Distribution type	Mean	STDEV	Correlation (Copula)
X_1	Normal	0.20	0.009	-
X_2	Lognormal	0.15	0.007	-
X_3	Gamma	0.48	0.015	-
X_4	Weibull	0.28	0.014	-
X_5	Normal	0.31	0.014	Clayton
X_6	Weibull	0.68	0.013	$\tau = 0.5$

The result of variable screening is shown in Table 4.2. The design sensitivity (gradient) of the Hartmann 6-D in Eq. (4.1) at the input mean point is shown in the second column, and the third through fifth columns show partial output variances using the sensitivity-variance method, the variable screening method, and the accurate method, respectively. As the variable screening method calculates partial output variances *nr* times, the result is the mean value of the calculated partial output variances. In each method, important variables are marked with bold font. It can be seen that the variable screening method finds the same variables as the accurate method, whereas the sensitivity-variance method misses X_4 and X_5. The sixth and seventh columns are the ratios of partial output variances using the sensitivity-variance method and the variable screening method to the accurate partial output variances, respectively. It is evident that the sensitivity-variance method cannot estimate the partial output variances accurately, while the variable screening method does. Overall, the variable-screening method outperforms the sensitivity-variance method. Hence, it is better to use at least quadratic approximation for the 1-D surrogate model to calculate partial output variances.

Table 4.2 Partial Output Variances of Hartmann 6-D Example

Random variable	Design sensitivity	Sensitivity -variance	Variable screening	Accurate	Sens.-Var. /Accurate	Var. Scr. /Accurate
X_1	−5.88E−02	2.80E−07	2.24E−06	2.25E−06	12.4%	99.6%
X_2	−1.36E−02	9.04E−09	6.12E−07	6.17E−07	1.5%	99.2%
X_3	7.15E−02	1.15E−06	1.11E−05	1.07E−05	10.7%	103.7%
X_4	2.63E−01	1.35E−05	**8.88E−05**	**8.67E−05**	15.6%	102.4%
X_5	−1.36E−01	3.61E−06	**1.86E−04**	**1.86E−04**	1.9%	100.0%
X_6	1.17E+00	**2.29E−04**	**1.65E−04**	**1.66E−04**	138.0%	99.4%
Selected		2.29E−04	4.40E−04	4.39E−04	-	-
All		2.48E−04	4.54E−04	4.52E−04	-	-
Selected / All		92.3%	96.9%	97.1%	-	-
Criterion (μ_0)		4.13E−05	7.42E−05	7.54E−05	-	-

In Table 4.2, the bottom four rows show more information about each method. The first and second rows are the sums of partial output variances of selected variables and all variables, respectively. The third row shows the ratio of the first row to the second row. The last row is μ_0, which is the criterion used to select important variables. In the variable screening method, the important variables (X_4, X_5 and X_6) are determined to be larger than μ_0 by hypothesis testing. The sensitivity-variance method and the accurate method select important variables if the variable has partial output variance larger than μ_0. In the third row, the ratio for the variable screening method is larger than 85%, so γ is the initial value of 1.0, and equivalent μ_0 is applied for other methods. The sensitivity-variance method estimates that 92.3% of output variance is contained in X_6 only. This is a very poor estimation, as only 36.7% (=1.66E−04/4.52E−04) of output variance is captured in X_6 according to the result of the accurate method. On the other hand, the variable screening method estimates that 96.9% of output variance is contained in X_4, X_5, and X_6, and this is very accurate compared to the 97.1% determined by the accurate method.

However, the total and captured output variances in Table 4.2 are approximations using the partial output variances, and the correlation term is not considered, as explained in Section 3.2.1. Having the analytical expression of the Hartmann 6-D example in Eq. (4.1), true total output variance induced by multiple input random variables can be calculated as well. The true total output variance is calculated using 1,000,000 realizations of all input random variables, and the calculated value is 4.23E−04, as shown in Table 4.3. Recalling the approximated result in Table 4.2, the variable screening method (4.54E−04) and the accurate method (4.52E−04) well approximate the true total output variance, whereas the sensitivity-variance method (2.48E−04) is not able to do so. In Table 4.3, true captured output variance by the selected variables is also calculated. To calculate the true captured output variance, the realizations, which are generated to calculate the true total output variances, are used. Among them, the realizations of the screened-out variables are fixed at their mean values. Then, the variance of the Hartmann 6-D is calculated using the modified realizations. The true captured output variance in X_6 is 1.66E−04, which is the same as the partial output variance of X_6 found by the accurate method (see Table 4.2). Hence, the sensitivity-variance method captures only 39.2% (1.66E−04/4.23E−04) of the true total output variance in its selection X_6. This will lead a reliability problem to estimate the probability of failure incorrectly. By contrast, the captured output variance in X_4, X_5, and X_6 is 4.10E−04, which is 96.9% (4.10E−04/4.23E−04) of the total output variance. This indicates that a reliability problem could be solved accurately utilizing X_4, X_5, and X_6. From this example, it can be seen that the variable screening method works as it is intended.

Table 4.3 True Total and Captured Output Variances of Hartmann 6-D Example

	Sensitivity -variance	Variable screening & accurate	True total output variance
Selected variables	X_6 (1-D)	X_4, X_5, X_6 (3-D)	All (6-D)
Captured output variance	1.66E−04	4.10E−04	4.23E−04
Ratio to total out. var.	39.2%	96.9%	100.0%

4.2.2 Dixon-Price 12-D

The second analytical example is the Dixon-Price 12-D, and again a constant term is added to its original function. The analytical expression is shown as (Lee, 2007)

$$G(\mathbf{X}) = (X_1 - 1)^2 + \sum_{i=2}^{m} i\,(2X_i^2 - X_{i-1})^2 - 3.5575 \times 10^{-3} \tag{4.5}$$

where $-10 \le x_i \le 10$, $i = 1, 2, \ldots, m$ and $m = 12$. Input random variables shown in Table 4.4 are used for the test. They have five different marginal distribution types of normal, lognormal, Weibull, Gumbel, and gamma. X_1 and X_2 are correlated with the Frank copula and Kendall's tau of 0.7. Also, X_5 and X_6 are correlated with the FGM copula and Kendall's tau of 0.2.

The test result of the Dixon-Price 12-D example is shown in Table 4.5, and selected variables in each method are marked with bold font. In this example, the value of γ is lowered to 0.7 to contain at least 85% of output variance in the selected variables. And it is shown that 86.6% of output variance is estimated using the variable screening method in Table 4.5. Design sensitivities with respect to $X_9 \sim X_{12}$ are zero, and accordingly the partial output variances of $X_9 \sim X_{12}$ using the sensitivity-variance method are zero. Hence, the sensitivity-variance method misses X_9 and X_{12} even though they have large partial output variances. Moreover, the other partial output variances using the sensitivity-variance method have poor accuracy compared to the accurate method (see the sixth column in Table 4.5). Hence, X_2 and X_3 are selected instead of X_6 even though X_6 actually has larger partial output variance than X_2 and X_3. On the contrary, the variable screening method reasonably estimates partial output variances and correctly identifies important variables compared to the accurate method. Therefore, it is confirmed that at least quadratic approximation is needed for the 1-D surrogate model to calculate partial output variances.

Table 4.4 Input Random Variables for Dixon-Price 12-D Example

Random variable	Distribution type	Mean	STDEV	Correlation (Copula)
X_1	Normal	1.00	0.025	Frank
X_2	Normal	0.71	0.02	$\tau = 0.7$
X_3	Lognormal	0.59	0.02	-
X_4	Lognormal	0.55	0.02	-
X_5	Weibull	0.52	0.02	FGM
X_6	Weibull	0.51	0.02	$\tau = 0.2$
X_7	Gumbel	0.51	0.015	-
X_8	Gumbel	0.50	0.015	-
X_9	Normal	0.50	0.015	-
X_{10}	Normal	0.50	0.01	-
X_{11}	Gamma	0.50	0.01	-
X_{12}	Gamma	0.50	0.015	-

Table 4.5 Partial Output Variances of Dixon-Price 12-D Example

Random variable	Design sensitivity	Sensitivity -variance	Variable screening	Accurate	Sens.-Var. /Accurate	Var. Scr. /Accurate
X_1	−3.28E−02	6.72E−07	7.95E−06	7.71E−06	8.7%	103.1%
X_2	1.76E−01	**1.24E−05**	1.41E−04	1.35E−04	9.2%	104.4%
X_3	−3.15E−01	**3.98E−05**	1.56E−04	1.62E−04	24.6%	96.3%
X_4	3.56E−01	**5.07E−05**	**2.95E−04**	**2.81E−04**	18.0%	105.0%
X_5	−1.94E−01	**1.50E−05**	**4.98E−04**	**4.45E−04**	3.4%	111.9%
X_6	−1.38E−01	7.61E−06	**6.22E−04**	**5.51E−04**	1.4%	112.9%
X_7	4.51E−01	**4.58E−05**	**5.27E−04**	**5.57E−04**	8.2%	94.6%
X_8	−3.20E−01	**2.30E−05**	**3.10E−04**	**3.55E−04**	6.5%	87.3%
X_9	0.00E+00	0.00E+00	**2.20E−04**	**2.15E−04**	0.0%	102.3%
X_{10}	0.00E+00	0.00E+00	5.31E−05	5.25E−05	0.0%	101.1%
X_{11}	0.00E+00	0.00E+00	6.42E−05	6.32E−05	0.0%	101.6%
X_{12}	0.00E+00	0.00E+00	**2.45E−04**	**2.39E−04**	0.0%	102.5%
Selected		1.87E−04	2.72E−03	2.64E−03	-	-
All		1.95E−04	3.14E−03	3.06E−03	-	-
Selected / All		95.9%	86.6%	86.3%	-	-
Criterion (μ_0)		1.14E−05	1.68E−04	1.79E−04	-	-

Using Eq. (4.5), the true total and captured output variances of Dixon-Price 12-D example are calculated as shown in Table 4.6. In Table 4.5, the variable screening (3.14E−03) and the accurate methods (3.06E−03) reasonably approximate the true total output variance (3.30E−03 in Table 4.6), while the sensitivity-variance method (1.95E−04) cannot. In Table 4.6, the true captured output variance by selected variables using the sensitivity-variance method is only 2.15E−03, which is 65.2% of the true total output variance. By contrast, the output variance of 2.80E−03 is contained in $X_4 \sim X_9$

and X_{12}, which indicates that 84.8% of the true total output variance is captured. Hence, it is verified that the variable screening method correctly finds the important variables of the Dixon-Price 12-D example. Through analytical examples, it is shown that the partial output variance is a well-performing measure for variable screening purposes, and the proposed variable screening method successfully finds important variables as it is intended.

Table 4.6 True Total and Captured Output Variances of Dixon-Price 12-D Example

	Sensitivity -variance	Variable screening & accurate	True total output variance
Selected variables	$X_2 \sim X_5$, X_7, X_8 (6-D)	$X_4 \sim X_9$, X_{12} (7-D)	All (12-D)
Captured output variance	2.15E−03	2.80E−03	3.30E−03
Ratio to total out. var.	65.2%	84.8%	100.0%

4.3 Engineering Example

A car noise, vibration, and harshness (NVH) and crash safety problem is considered to demonstrate the performance and efficiency of the proposed method. The problem includes full frontal impact, 40% offset frontal impact, and NVH as constraints. There are a total of 11 performance measures as shown in Table 4.7: nine safety measures and two NVH measures.

Table 4.7 Performance Measure Description

<table>
<tr><th colspan="2">Mode</th><th>Function</th><th>Value</th><th>Feasibility decision</th></tr>
<tr><td rowspan="9">Safety</td><td rowspan="2">Full frontal impact</td><td>G_1</td><td>Chest G</td><td rowspan="11">$\leq Baseline_i$</td></tr>
<tr><td>G_2</td><td>Crush displacement</td></tr>
<tr><td rowspan="7">40% offset impact</td><td>G_3</td><td>Brake pedal</td></tr>
<tr><td>G_4</td><td>Footrest</td></tr>
<tr><td>G_5</td><td>Left toepan</td></tr>
<tr><td>G_6</td><td>Center toepan</td></tr>
<tr><td>G_7</td><td>Right toepan</td></tr>
<tr><td>G_8</td><td>Left IP</td></tr>
<tr><td>G_9</td><td>Right IP</td></tr>
<tr><td colspan="2" rowspan="2">NVH</td><td>G_{10}</td><td>Torsion mode</td></tr>
<tr><td>G_{11}</td><td>Vertical bending mode</td></tr>
</table>

In this example, it is assumed that the only source of uncertainty is the thickness of the body plates. The 44 random variables shown in Table 4.8 are used to represent the thicknesses. All random variables follow normal distribution and are statistically independent. The baseline design $\mathbf{d}^B$ is the mean vector of the 44 random variables, and there is no random parameter in this example. Among those random variables, six random variables (X_1 to X_5 and X_8) are common variables for both safety and NVH measures, two (X_6 and X_7) are variables only for safety, and the other 36 random variables are only for NVH measures.

This problem requires three and a half hours for the impact dynamic analysis for crash safety and the modal analysis for NVH. Thus, the actual analysis takes too much time to test the proposed method thoroughly. Ford Motor Company provided full-dimensional global (considering the entire design domain) surrogate models so that we could use them to demonstrate the proposed method of variable screening. The full-dimensional surrogate models may not be accurate, since 44-D is too high to create accurate surrogate models, especially for RBDO. However, to test the proposed method of variable screening, the responses from the 44-D global surrogate models are treated as true responses in this example. The maximum dimension at which accurate surrogate models can be generated depends on the computational power and nonlinearity of a given problem. In this study, the dynamic Kriging (DKG) method (Zhao et al., 2011) is used for an accurate surrogate model, and 18-D is targeted as the maximum degrees of freedom of DKG models. The Iowa-Reliability-Based Design Optimization (I-RBDO) code (Choi et al., 2012) has been used to generate DKG models and carry out RBDO.

Table 4.8 Input Random Variables

RVs	Dist. type	$\mathbf{d}^B$	ST DEV	$\mathbf{d}^L$	$\mathbf{d}^U$	RVs	Dist. type	$\mathbf{d}^B$	ST DEV	$\mathbf{d}^L$	$\mathbf{d}^U$
X_1	Normal	1.9	0.05	1.5	2.3	XN_1	Normal	0.9	0.03	0.7	1.1
X_2	Normal	1.91	0.05	1.5	2.3	XN_2	Normal	1.1	0.03	0.8	1.4
X_3	Normal	2.51	0.06	2.0	3.0	XN_3	Normal	1.55	0.05	1.2	1.9
X_4	Normal	2.4	0.06	1.9	2.9	XN_4	Normal	0.9	0.03	0.7	1.1
X_5	Normal	2.55	0.06	2.0	3.1	XN_5	Normal	1.5	0.03	1.2	1.8
X_6	Normal	2.25	0.06	1.8	2.7	XN_6	Normal	1.2	0.03	0.9	1.5
X_7	Normal	2.25	0.06	1.8	2.7	XN_7	Normal	1.1	0.03	0.8	1.4
X_8	Normal	1.5	0.03	1.2	1.8	XN_8	Normal	1.52	0.05	1.2	1.9
X_{10}	Normal	1.28	0.03	0.9	1.6	XN_9	Normal	0.8	0.03	0.6	1.0
X_{11}	Normal	1.4	0.03	1.0	1.8	XN_{10}	Normal	0.8	0.03	0.6	1.0
X_{12}	Normal	1.1	0.03	0.8	1.4	XN_{11}	Normal	1.2	0.03	0.9	1.5
X_{13}	Normal	2.2	0.06	1.7	2.7	XN_{12}	Normal	0.75	0.03	0.6	0.9
X_{14}	Normal	1.5	0.03	1.2	1.8	XN_{13}	Normal	0.75	0.03	0.6	0.9
X_{15}	Normal	1.25	0.03	0.9	1.6	XN_{14}	Normal	0.75	0.03	0.6	0.9
X_{16}	Normal	2.5	0.06	2.0	3.0	XN_{15}	Normal	1.0	0.03	0.8	1.2
X_{17}	Normal	2.0	0.05	1.5	2.5	XN_{16}	Normal	1.14	0.03	0.9	1.4
X_{18}	Normal	1.4	0.03	1.1	1.7	XN_{17}	Normal	1.2	0.03	0.9	1.5
X_{20}	Normal	1.22	0.03	0.9	1.5	XN_{18}	Normal	1.4	0.03	1.1	1.7
X_{23}	Normal	0.75	0.03	0.6	1.0	XN_{19}	Normal	1.2	0.03	0.9	1.5
X_{24}	Normal	1.9	0.05	1.5	2.3	XN_{20}	Normal	1.4	0.03	1.1	1.7
X_{25}	Normal	0.65	0.03	0.5	0.8	XN_{21}	Normal	2.13	0.06	1.7	2.6
X_{26}	Normal	0.85	0.03	0.6	1.1						
X_{27}	Normal	0.85	0.03	0.6	1.1						

4.3.1 Variable Screening

At the baseline design $\mathbf{d}^B$, which is the initial design as shown in Table 4.8, all 11 performance measures in Table 4.7 are active. That is, the value of every performance measure at the baseline design is the same as the baseline values, with $G_i^B = Baseline_i$, i = 1~11. Therefore, the proposed variable screening method is performed for the problem at the baseline design. The number of samples *ns*, initial number of repeated calculations nr_0, significance level α, false negative rate β, and threshold value γ are set as 50, 50, 0.05, 0.05, and 1.0, respectively. Four hundred eighty-four (44 design variables × 11 performance measures) 1-D surrogate models with quadratic interpolation are generated using 89 DoE analyses (i.e., simulation samples). It is noted that 11 values of performance measures are obtained from one DoE analysis. The results of partial output variances $\overline{v}_i$ are listed in Table 4.9 and Table 4.10 for every performance measure. The partial output variances of the important variables for each performance measure are marked with bold font. It is noted that only partial output variances of X_1~X_8 are listed in Table 4.9 since G_1~G_9 are only a function of X_1~X_8 as the variable screening method identified the partial output variances to be zero for other random variables. In Table 4.10, X_6 and X_7 have zero partial output variances as G_{10} and G_{11} are not functions of X_6 and X_7. In the last three rows of Table 4.9 and Table 4.10, the sums of partial output variances of selected variables, the sums of all partial output variances, and their ratios are listed. As explained in previous sections, this is the estimated ratio between the captured output variance in selected variables to the total output variance. It is estimated that a minimum of 90.3% of the total output variance is captured in the selected variables. In total, 14 random variables: X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{10}, X_{20}, X_{23}, X_{25}, X_{26}, and XN_1, are selected as important variables. Accordingly, 14 design variables, which are the means of the selected random variables, are considered as important design variables.

Table 4.9 Partial Output Variances $\overline{v}_i$ (G_1~G_9)

RVs	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9
$\boldsymbol{X_1}$	**2.37E−02**	**3.75E+00**	**4.06E+01**	**1.11E+01**	**1.94E+00**	**1.73E+00**	**6.21E−01**	**3.45E−01**	**5.75E−01**
$\boldsymbol{X_2}$	**1.88E−02**	**1.01E+00**	3.55E−01	**1.86E+00**	**2.07E+00**	**2.06E+00**	**1.09E+00**	**4.03E−01**	**2.69E−01**
$\boldsymbol{X_3}$	5.08E−05	1.76E−02	**1.05E+01**	**9.74E+00**	**5.95E+00**	**4.38E+00**	**2.00E+00**	**2.37E−01**	2.24E−02
$\boldsymbol{X_4}$	1.66E−04	9.27E−02	**3.91E+00**	**1.79E+00**	2.65E−01	6.65E−02	1.73E−03	4.78E−03	9.65E−03
$\boldsymbol{X_5}$	1.12E−04	1.37E−04	**2.74E+00**	**2.23E+00**	**4.90E−01**	**9.11E−01**	**9.30E−01**	**1.79E−01**	**7.94E−02**
$\boldsymbol{X_6}$	4.58E−05	**2.77E+00**	1.64E−01	8.14E−02	**8.18E−01**	1.60E−01	**1.70E−01**	**3.95E−02**	**3.27E−02**
$\boldsymbol{X_7}$	**2.32E−03**	1.34E−01	1.91E−01	1.50E−01	**2.75E+00**	**1.71E+00**	**5.66E−01**	**1.10E−01**	**2.77E−01**
$\boldsymbol{X_8}$	**1.23E−03**	6.28E−02	4.39E−02	3.02E−01	**5.35E−01**	**5.45E−01**	**3.27E−01**	**6.25E−02**	**4.87E−02**
Selected	4.61E−02	7.53E+00	5.78E+01	2.67E+01	1.46E+01	1.13E+01	5.70E+00	1.38E+00	1.28E+00
All	4.64E−02	7.84E+00	5.85E+01	2.73E+01	1.48E+01	1.16E+01	5.71E+00	1.38E+00	1.31E+00
Selec./All	99.4%	96.0%	98.8%	97.8%	98.6%	97.4%	99.8%	100.0%	97.7%

Table 4.10 Partial Output Variances $\overline{v}_i$ (G_{10} and G_{11})

RVs	G_{10}	G_{11}	RVs	G_{10}	G_{11}	RVs	G_{10}	G_{11}
$\boldsymbol{X_1}$	**1.30E−03**	**1.48E−03**	X_{17}	4.24E−07	2.01E−04	XN_8	2.06E−05	5.45E−05
X_2	4.39E−05	3.44E−05	X_{18}	9.71E−06	3.33E−05	XN_9	1.61E−06	1.89E−05
$\boldsymbol{X_3}$	**8.95E−04**	3.05E−04	$\boldsymbol{X_{20}}$	**2.21E−03**	**1.84E−03**	XN_{10}	5.11E−06	1.83E−06
X_4	8.74E−05	2.63E−05	$\boldsymbol{X_{23}}$	**2.67E−03**	**1.13E−03**	XN_{11}	1.56E−07	1.12E−07
X_5	3.06E−04	4.23E−05	X_{24}	2.35E−05	6.89E−05	XN_{12}	1.83E−06	3.10E−05
X_6	0	0	$\boldsymbol{X_{25}}$	**4.85E−04**	**7.31E−03**	XN_{13}	1.69E−07	8.19E−07
X_7	0	0	$\boldsymbol{X_{26}}$	**4.59E−03**	**1.06E−02**	XN_{14}	9.08E−08	3.75E−05
$\boldsymbol{X_8}$	**5.56E−04**	1.89E−04	X_{27}	3.43E−05	3.76E−04	XN_{15}	1.84E−06	1.66E−05
$\boldsymbol{X_{10}}$	**5.49E−04**	3.87E−04	$\boldsymbol{XN_1}$	9.33E−05	**6.38E−03**	XN_{16}	4.35E−08	2.88E−08
X_{11}	3.32E−05	2.05E−04	XN_2	4.49E−07	7.48E−05	XN_{17}	3.24E−07	1.84E−06
X_{12}	9.35E−05	3.98E−04	XN_3	9.59E−08	2.00E−05	XN_{18}	1.04E−07	1.18E−07
X_{13}	7.23E−06	3.82E−04	XN_4	6.04E−08	8.36E−05	XN_{19}	3.03E−08	3.37E−07
X_{14}	5.76E−07	1.17E−05	XN_5	2.60E−07	1.42E−07	XN_{20}	2.06E−07	2.98E−06
X_{15}	1.16E−07	1.13E−05	XN_6	1.29E−07	2.22E−07	XN_{21}	2.44E−07	9.41E−06
X_{16}	9.14E−08	1.63E−06	XN_7	9.46E−08	7.94E−06			
Selected							1.33E−02	2.87E−02
All							1.40E−02	3.18E−02
Selec./All							95.0%	90.3%

Sensitivity-variance method introduced in Section 4.2 is applied to performance measure G_5, and the result is shown in Table 4.11. Selected random variables for G_5 are X_1~X_3 and X_5~X_8 out of X_1~X_8. Among X_1~X_8, variables X_3~X_7 have the largest standard deviation of 0.06. However, X_4 is not selected among them because it has small design sensitivity compared to others. Here, the design sensitivities are calculated at the design point using the forward finite difference method (FDM) with 0.1% perturbation. By contrast, X_8 is selected as an important variable even though it has smallest standard deviation of 0.03. Again, this is because it has relatively large sensitivity and induces large output variances.

Table 4.11 Result of Sensitivity-Variance Method in G_5

RVs	STDEV (σ_{X_i})	Design Sensitivity (α_i)	Partial Output Variance ($\alpha_i^2\sigma_{X_i}^2$)
X_1	**0.05**	**27.9105**	**1.95E+00**
X_2	**0.05**	**29.2723**	**2.14E+00**
X_3	**0.06**	**−40.7490**	**5.98E+00**
X_4	0.06	8.7917	2.78E−01
X_5	**0.06**	**−11.6078**	**4.85E−01**
X_6	**0.06**	**14.1289**	**7.19E−01**
X_7	**0.06**	**27.8533**	**2.79E+00**
X_8	**0.03**	**24.3000**	**5.31E−01**

Interestingly, the partial output variances using the sensitivity-variance method of G_5 shown in Table 4.11 are close to the result shown in the sixth column of Table 4.9. In fact, the same variables as those in the variable screening method will be selected by using the sensitivity-variance method throughout all 11 constraints. However, the sensitivity-variance method has the possibility of choosing undesirable variables as shown in the analytical examples in Section 4.2. In Figure 4.1, the shape of G_5 when each X_i is random is shown. It is easily anticipated that the design sensitivity of G_5 with respect to X_i could be very small or even zero so that the sensitivity-variance method may provide inaccurate partial output variance; this can be prevented if the variable screening method is used.

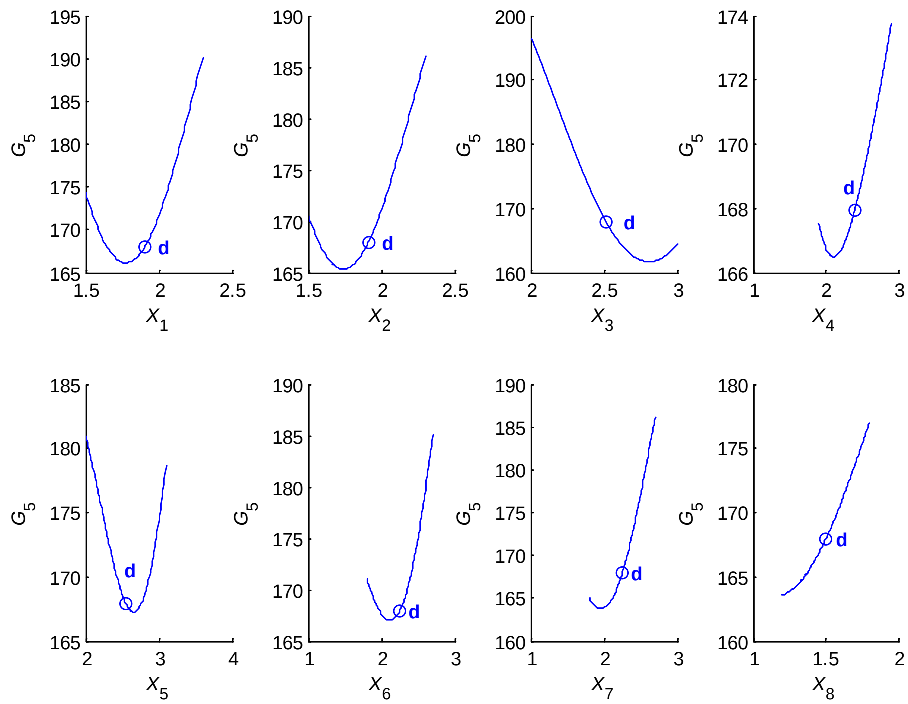


Figure 4.1 Shape of G_5 when Each X_i Is Random

Table 4.12 Partial Output Variance of X_2 in G_6 at DDO Optimum Design

Method	Perturbation	Selection	Design sensitivity	Partial output variance	Accuracy
Sensitivity-variance	Forward, 1%	No	−5.83E+00	8.49E−02	56.6%
	Forward, 0.5%	Yes	−6.30E+00	9.94E−02	66.3%
	Forward, 0.1%	Yes	−6.68E+00	1.12E−01	74.7%
	Central, 1%	Yes	−6.79E+00	1.15E−01	76.7%
	Central, 0.5%	Yes	−6.78E+00	1.15E−01	76.7%
	Central, 0.1%	Yes	−6.78E+00	1.15E−01	76.7%
Variable screening	-	Yes	-	1.51E−01	100.7%
Accurate	-	Yes	-	1.50E−01	100.0%

The sensitivity-variance method requires accurate design sensitivity. In practical engineering problems, design sensitivity might be calculated using FDM. To use FDM, a user determines perturbation method (forward, backward, or central) and perturbation size, and the result of the sensitivity-variance method could depend on the user's choice. In Table 4.12, partial output variance of X_2 in G_6 at deterministic design optimization (DDO) optimum design with various methods is shown. It can be seen that X_2 is not selected as an important variable when design sensitivity is calculated using forward FDM with 1% perturbation. The partial output variance is only 56.6% of that found using the accurate method with 100,000 realizations of X_2. To obtain more accurate design sensitivity with forward FDM, small perturbation is required as shown in Table 4.12. However, small perturbation does not always provide accurate design sensitivity, and determining appropriate perturbation size would require extra DoE samples. Central FDM provides more accurate design sensitivity, and it is insensitive to perturbation size. However, the partial output variance with central FDM sensitivity shows at most 76.7% accuracy compared to the accurate method. It is noted that the variable screening method would not require perturbation size determination. Moreover, a user can perform the proposed variable screening method using only one more DoE sample than the sensitivity-variance method with central FDM.

Since we have 44-D global surrogate models for this example, GSA can be carried out to verify the effectiveness of the proposed method. Among various GSA methods, the global sensitivity index method, which can identify the global effect of the variables of interest on the output, is used here. The main strength of the global sensitivity index method is that it can find interactions between all variables (not statistical correlation between random variables). All random variables are assumed to follow uniform distribution in their corresponding design domain of $\mathbf{d}^L$ and $\mathbf{d}^U$, and global sensitivity indices are calculated using the Monte Carlo simulation (MCS) method with 1 million MCS samples (Sobol, 2001). There are many global sensitivity indices in this 44-D problem; the total sensitivity index S_i^{tot} is used for variable ranking and screening. The total sensitivity index S_i^{tot} is a "total influence of the *i*-th random variable" to the output. That is, it indicates the main effect plus interactions of the *i*-th random variable with other random variables (Chen et al., 2005). The results are listed in Table 4.13 and Table 4.14. To identify important random variables, the mean value of S_i^{tot} is calculated for each constraint, and the random variable, which yields larger S_i^{tot} than the mean value, is selected as an important variable and marked in bold font in these tables. In Table 4.13, S_i^{tot} only for $X_1 \sim X_8$ are listed because S_i^{tot} for other variables are zero due to the fact that $G_1 \sim G_9$ are functions of $X_1 \sim X_8$. Also, the sum of S_i^{tot} for the selected random variables, the sum of S_i^{tot} for all random variables, and their ratios are listed in the last three rows, respectively.

Table 4.13 Global Sensitivity Indices S_i^{tot} (G_1~G_9)

RVs	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9
$\boldsymbol{X_1}$	**4.16E−01**	**3.84E−01**	**5.23E−01**	**2.68E−01**	**1.74E−01**	**1.89E−01**	**2.07E−01**	**2.56E−01**	**4.35E−01**
$\boldsymbol{X_2}$	**3.46E−01**	**1.18E−01**	**1.04E−01**	**1.88E−01**	**1.98E−01**	**2.09E−01**	**1.66E−01**	**2.62E−01**	**2.07E−01**
$\boldsymbol{X_3}$	7.83E−03	5.75E−03	**3.28E−01**	**5.07E−01**	**5.18E−01**	**4.97E−01**	**4.92E−01**	**2.69E−01**	**4.96E−02**
$\boldsymbol{X_4}$	1.89E−02	**2.23E−02**	**7.92E−02**	**1.34E−01**	**1.25E−01**	**1.38E−01**	**1.42E−01**	**7.38E−02**	**2.87E−02**
$\boldsymbol{X_5}$	7.50E−03	−7.22E−04	**1.40E−01**	**2.10E−01**	**1.53E−01**	**1.76E−01**	**1.78E−01**	**1.90E−01**	**9.10E−02**
$\boldsymbol{X_6}$	**5.65E−02**	**2.67E−01**	2.08E−02	**5.01E−02**	**1.11E−01**	**1.15E−01**	**1.67E−01**	**5.13E−02**	**2.74E−02**
$\boldsymbol{X_7}$	**8.13E−02**	5.65E−03	1.77E−02	**4.52E−02**	**1.59E−01**	**1.60E−01**	**1.76E−01**	**6.73E−02**	**1.73E−01**
$\boldsymbol{X_8}$	**3.23E−02**	**2.06E−02**	4.02E−03	2.33E−02	**3.82E−02**	**4.09E−02**	**5.21E−02**	**7.13E−02**	**6.14E−02**
Selected	9.32E−01	8.12E−01	1.17E+00	1.40E+00	1.48E+00	1.52E+00	1.58E+00	1.24E+00	1.07E+00
All	9.66E−01	8.23E−01	1.22E+00	1.43E+00	1.48E+00	1.52E+00	1.58E+00	1.24E+00	1.07E+00
Selec./All	96.5%	98.7%	95.9%	97.9%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 4.14. Global Sensitivity Indices S_i^{tot} (G_{10} and G_{11})

RVs	G_{10}	G_{11}	RVs	G_{10}	G_{11}	RVs	G_{10}	G_{11}
$\boldsymbol{X_1}$	**1.05E−01**	**7.09E−02**	X_{17}	3.30E−04	1.46E−02	XN_8	−8.85E−04	−1.34E−03
$\boldsymbol{X_2}$	1.01E−03	2.38E−03	X_{18}	−4.77E−04	−2.50E−04	XN_9	8.54E−04	−1.71E−03
$\boldsymbol{X_3}$	**5.72E−02**	1.01E−02	$\boldsymbol{X_{20}}$	**2.07E−01**	**9.17E−02**	XN_{10}	1.28E−04	−1.07E−04
$\boldsymbol{X_4}$	8.66E−03	2.65E−03	$\boldsymbol{X_{23}}$	**1.34E−01**	**3.57E−02**	XN_{11}	−2.72E−04	−2.85E−05
$\boldsymbol{X_5}$	**3.29E−02**	5.69E−03	X_{24}	−1.60E−03	−8.39E−04	XN_{12}	−1.51E−04	−1.94E−04
X_6	0	0	$\boldsymbol{X_{25}}$	1.12E−02	**9.79E−02**	XN_{13}	−5.24E−05	1.59E−04
X_7	0	0	$\boldsymbol{X_{26}}$	**2.96E−01**	**3.80E−01**	XN_{14}	−3.88E−05	−6.91E−05
$\boldsymbol{X_8}$	**4.73E−02**	5.89E−03	$\boldsymbol{X_{27}}$	6.04E−03	**2.38E−02**	XN_{15}	1.20E−04	4.59E−04
$\boldsymbol{X_{10}}$	**5.98E−02**	1.99E−02	$\boldsymbol{XN_1}$	5.62E−03	**1.65E−01**	XN_{16}	8.87E−05	2.85E−04
X_{11}	3.86E−03	1.67E−02	XN_2	−1.11E−04	2.31E−03	XN_{17}	−6.09E−05	−3.61E−04
$\boldsymbol{X_{12}}$	1.34E−02	**2.88E−02**	XN_3	1.13E−04	1.33E−03	XN_{18}	−1.12E−05	−9.18E−05
X_{13}	3.20E−04	1.50E−02	XN_4	1.54E−04	−2.50E−03	XN_{19}	1.10E−04	1.23E−03
X_{14}	2.24E−04	−4.65E−04	XN_5	−2.96E−04	9.10E−04	XN_{20}	−1.67E−04	−8.38E−04
X_{15}	−1.33E−04	1.36E−03	XN_6	2.34E−04	−2.32E−05	XN_{21}	6.77E−05	−6.44E−04
X_{16}	−1.94E−04	2.42E−03	XN_7	−1.61E−05	−3.27E−04			
Selected							9.39E−01	8.94E−01
All							9.87E−01	9.87E−01
Selec./All							95.1%	90.6%

Using the global sensitivity index method, 16 random variables are selected as shown in Table 4.15. Those 16 random variables include all 14 random variables selected using the proposed method as shown in Table 4.15. Moreover, if we limit it to 14 random variables to be selected, X_{12} and X_{27} will not be selected as they have the least S_i^{tot} among the selected variables for G_{11} as shown in Table 4.14. Thus, the selected 14 random variables by both methods are identical. The ratio between the sensitivity indices of selected variables and all variables has no physical meaning. However, it is an indicator that shows how much variance is captured by the selected random variables. The results are quite similar to those of the proposed method as shown in Tables 4.9 and 4.10 and Tables 4.13 and 4.14, respectively. Hence, it is demonstrated that the proposed variable screening method is quite effective even though it does not require global surrogate models like the global sensitivity index method.

Table 4.15 Selected Random Variables

Method	Selected variables
Proposed method	X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{10}, X_{20}, X_{23}, X_{25}, X_{26}, XN_1 (14 RVs)
Global sensitivity index	X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{10}, X_{12}, X_{20}, X_{23}, X_{25}, X_{26}, X_{27}, XN_1 (16 RVs)

4.3.2 Reliability-Based Design Optimization

For this example, RBDO is formulated as

$$\begin{aligned} &\text{minimize} && \text{Weight}(\mathbf{d}) \\ &\text{subject to} && P[G_i^B(\mathbf{X}) > \text{Baseline}_i] \le 10\%,\ \ i = 1, \cdots, 11. \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^{44} \text{ and } \mathbf{X} \in \mathbb{R}^{44} \end{aligned} \tag{4.6}$$

For a comparison study, we considered three cases: (1) a set of 14 random variables is selected based on experience without using the proposed variable screening method, (2) another set of 14 random variables is selected using the proposed variable screening method as shown in Section 4.3.1, and (3) in addition to the 14 random variables selected in case 2, four more random variables are selected using the cost function sensitivity for a total of 18 design variables to test the effectiveness of the proposed variable screening method and the accuracy of the DKG generated by I-RBDO. The selected design variables are listed in Table 4.16.

Table 4.16 Selected Random Variables for RBDO

Cases	Common selection	Different selection
(1) Based on experience	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{20}, X_{25}, XN_1$ (11 RVs)	XN_9, XN_{10}, XN_{11}
(2) Variable screening		X_{10}, X_{23}, X_{26}
(3) Variable screening + Cost Function		X_{10}, X_{23}, X_{26} + $XN_4, XN_9, XN_{10}, XN_{11}$

Because the cost function, which is the weight in this problem, is a function of design variables **d**, not random variables **X**, the function is deterministic. Therefore, the design sensitivity of the cost function with respect to the design variable is calculated by the FDM, and the four design variables (and related random variables XN_4, XN_9, XN_{10}, and XN_{11}) that show the largest sensitivity among the unselected design variables are chosen. Reliability-based design optimization (RBDO) is carried out with three sets of selected random variables. The optimum design results are summarized in Table 4.17. The bold font indicates chosen design variables, and others are fixed at the baseline design value. Also, probabilities of failure, cost function values, and design iteration details are listed in Table 4.18. All RBDOs are carried out using the sampling-based RBDO method in I-RBDO with 500,000 MCS samples. For the three cases, DKG surrogate models are created, using the DoE sample responses obtained from the 44-D global surrogate models, which are treated as true responses, for RBDO. In addition, the full-dimensional RBDO is performed as well using the 44-D surrogate models. The 44-D RBDO result is treated as the true RBDO optimum and used for the purpose of validation of RBDO results obtained for the three reduced-dimensional cases.

Indeed, the optimum design values for d_1~d_8 are very close to the full-dimensional 44-D case, as shown in Table 4.17, which shows that the generated DKG surrogate models are accurate. In the three cases, the random variables X_1~X_8 are selected because they have large partial output variance for performance measures G_1~G_9. This means that they contribute a large portion of the output variance. Hence, finding optimum values for them is the most effective way to reduce the probabilities of failure of G_1~G_9. Similarly, d_{22} (corresponding to X_{26}) moves to the upper bound of 1.1 when it is selected because it has the largest partial output variances of G_{10} and G_{11}. The design variables d_{27}, d_{32}, d_{33}, and d_{34} (corresponding to XN_4, XN_9, XN_{10}, and XN_{11}), which are selected by the sensitivity of the cost function, move to their lower bounds of 0.7, 0.6, 0.6, and 0.9, respectively, to reduce the cost function without significantly affecting the reliability of the optimum design. On the other hand, even though some design variables are selected due to the partial output variances of some constraints, they move to their lower bounds. For example, d_{24} (corresponding to XN_1) moves to the lower bound of 0.7 because it has the largest sensitivity for the cost function, even though it has the third-largest partial output variance of G_{11}. That is, via the trade-offs in the optimization process, it is moved to the lower bound to minimize the cost function rather than to reduce the probability of failure.

Table 4.17 RBDO Optimum Design

Design variables	Corres-ponding RVs	Baseline design	Based on experi-ence	Variable screen-ing	Variable screening + Cost function	Full dimen-sion
d_1	$\boldsymbol{X_1}$	1.9	**1.8343**	**1.8366**	**1.8336**	**1.8491**
d_2	$\boldsymbol{X_2}$	1.91	**2.1810**	**2.1804**	**2.1806**	**2.1692**
d_3	$\boldsymbol{X_3}$	2.51	**2.8528**	**2.8561**	**2.8540**	**2.8803**
d_4	$\boldsymbol{X_4}$	2.4	**1.9817**	**1.9810**	**1.9856**	**1.9507**
d_5	$\boldsymbol{X_5}$	2.55	**2.7195**	**2.7228**	**2.7261**	**2.7296**
d_6	$\boldsymbol{X_6}$	2.25	**2.2543**	**2.2497**	**2.2558**	**2.2408**
d_7	$\boldsymbol{X_7}$	2.25	**2.3199**	**2.3185**	**2.3207**	**2.3284**
d_8	$\boldsymbol{X_8}$	1.5	**1.7904**	**1.7966**	**1.7860**	**1.8**
d_9	$\boldsymbol{X_{10}}$	1.28	1.28	**0.9**	**0.9**	**1.5336**
d_{10}	X_{11}	1.4	1.4	1.4	1.4	**1.0038**
d_{11}	X_{12}	1.1	1.1	1.1	1.1	**0.9868**
d_{12}	X_{13}	2.2	2.2	2.2	2.2	**1.7006**
d_{13}	X_{14}	1.5	1.5	1.5	1.5	**1.2**
d_{14}	X_{15}	1.25	1.25	1.25	1.25	**0.9**
d_{15}	X_{16}	2.5	2.5	2.5	2.5	**2.0**
d_{16}	X_{17}	2.0	2.0	2.0	2.0	**1.5033**
d_{17}	X_{18}	1.4	1.4	1.4	1.4	**1.1**
d_{18}	$\boldsymbol{X_{20}}$	1.22	**0.9**	**0.9**	**0.9**	**0.9**
d_{19}	$\boldsymbol{X_{23}}$	0.75	0.75	**0.6**	**0.6**	**0.6**
d_{20}	X_{24}	1.9	1.9	1.9	1.9	**1.5**
d_{21}	$\boldsymbol{X_{25}}$	0.65	**0.6897**	**0.5429**	**0.5871**	**0.8**
d_{22}	$\boldsymbol{X_{26}}$	0.85	0.85	**1.1**	**1.1**	**1.1**
d_{23}	X_{27}	0.85	0.85	0.85	0.85	**1.1**
d_{24}	$\boldsymbol{XN_1}$	0.9	**0.7**	**0.7**	**0.7**	**0.7**
d_{25}	XN_2	1.1	1.1	1.1	1.1	**0.8**
d_{26}	XN_3	1.55	1.55	1.55	1.55	**1.2039**
d_{27}	$\boldsymbol{XN_4}$	0.9	0.9	0.9	**0.7**	**0.7**
d_{28}	XN_5	1.5	1.5	1.5	1.5	**1.2**
d_{29}	XN_6	1.2	1.2	1.2	1.2	**0.9**
d_{30}	XN_7	1.1	1.1	1.1	1.1	**0.8**

Table 4.17 Continued

Design variables	Corres-ponding RVs	Baseline design	Based on experi-ence	Variable screen-ing	Variable screening + Cost function	Full dimen-sion
d_{31}	XN_8	1.52	1.52	1.52	1.52	**1.6925**
d_{32}	$\boldsymbol{XN_9}$	0.8	**0.6**	0.8	**0.6**	**0.6**
d_{33}	$\boldsymbol{XN_{10}}$	0.8	**0.6**	0.8	**0.6**	**0.6**
d_{34}	$\boldsymbol{XN_{11}}$	1.2	**0.9**	1.2	**0.9**	**0.9**
d_{35}	XN_{12}	0.75	0.75	0.75	0.75	**0.6**
d_{36}	XN_{13}	0.75	0.75	0.75	0.75	**0.6**
d_{37}	XN_{14}	0.75	0.75	0.75	0.75	**0.6**
d_{38}	XN_{15}	1.0	1.0	1.0	1.0	**0.8**
d_{39}	XN_{16}	1.14	1.14	1.14	1.14	**0.9083**
d_{40}	XN_{17}	1.2	1.2	1.2	1.2	**0.9**
d_{41}	XN_{18}	1.4	1.4	1.4	1.4	**1.1**
d_{42}	XN_{19}	1.2	1.2	1.2	1.2	**0.9**
d_{43}	XN_{20}	1.4	1.4	1.4	1.4	**1.1**
d_{44}	XN_{21}	2.13	2.13	2.13	2.13	**1.7**

Table 4.18 Cost, Probabilities of Failure at RBDO Optimum Design, and Optimization Details

Performance measure	Based on experience	Variable screening	Variable screening + cost function	Full dimension
Cost	249.10	259.83	244.17	225.68
G_1	10.06%	9.94%	9.96%	9.96%
G_2	10.03%	10.11%	10.04%	9.99%
G_3	0.00%	0.00%	0.00%	0.00%
G_4	0.11%	0.11%	0.09%	0.10%
G_5	1.96%	1.95%	2.04%	1.91%
G_6	9.95%	10.01%	10.02%	9.99%
G_7	9.89%	9.93%	9.95%	10.01%
G_8	10.03%	9.98%	9.91%	9.07%
G_9	10.04%	9.99%	9.87%	9.92%
G_{10}	0.00%	0.00%	0.00%	0.00%
G_{11}	10.02%	9.97%	9.93%	10.02%
No. of design iterations	20	30	20	21
No. of DoE samples	2,666	4,358	3,306	-

To verify once again that the surrogate model generated by DKG is accurate, the same three cases are performed using responses from the 44-D global surrogate model directly while fixing screened-out variables at their baseline design points. As shown in Table 4.19, the optimums found using DKG and 44-D global surrogate models are very close to each other. Hence, it is confirmed that DKG generated accurate surrogate models. Moreover, it is also verified that RBDO can be conducted based on an accurate surrogate model even for a moderately large-dimensional problem (14 and 18 dimensions).

For the three reduced-dimensional cases, the probabilities of failure are calculated using only selected variables as random variables, since the other design variables are treated as deterministic as explained in Section 3.1.2 with fixed values at the baseline design. In Table 4.18, it can be seen that, in all cases, the target design constraints of 10% probability failure are closely satisfied as expected at these optimum designs since the RBDO considers only the selected variables as random variables. On the other hand, to check correct reliabilities, reliability analyses are carried out at these optimum designs, treating all variables as random using the 44-D surrogate models and MCS with 1 million samples, as shown in Table 4.20. It is noted that the probabilities of failure of the full-dimensional optimum in Table 4.18 and Table 4.20 are different even though the discrepancy is negligible. Theoretically, they should be the same; however, they are not equal because different numbers of MCS samples (500,000 and 1 million) are used and MCS error is induced. At the baseline design, all constraints have an approximately 50% probability of failure, and this is reasonable because all constraints are active at the baseline design. However, they are not exactly 50% because the constraint functions are nonlinear. Probabilistic constraint results corresponding to G_1~G_9 are active or feasible in both Table 4.18 and Table 4.20. Due to the fact that G_1~G_9 are functions of X_1~X_8, and all of them are selected as important variables, the RBDO result with reduced dimension and the reliability analysis result with full dimension are very close to each other, considering MCS errors. The constraint G_{10} shows inactive results regardless of which selection design variable set is used.

All probabilities of failure for the constraint G_{11} in Table 4.18 satisfy the target probability of failure 10%, which makes it obvious that these are the reliability analyses results of reduced-dimensional problems. However, full-dimensional reliability analyses at optimum designs show quite different values, as shown in Table 4.20. Selection based on experience shows a 17.70% probability of failure, which violates the target probability of failure significantly. The variables selected based on experience contain only 55.4% (=1.76E−02/3.18E−02 × 100%) of the total output variance of G_{11}. Hence, it cannot find any safe design once dimension is reduced. On the other hand, the probabilities of failure for the proposed variable screening method (Case 2) and also the one considering cost function (Case 3) are close to the target probability of failure. The selected variables contain 93.4% (=2.97E−02/ 3.18E−02 × 100%) and 93.7% (= 2.98E−02/ 3.18E−02 × 100%) of the total output variance of G_{11} for Case 2 and Case 3, respectively. Hence it can find a correct optimum even with reduced dimension.

Table 4.19 RBDO Optimum Design with I-RBDO and True Model

Design variables	Corresponding RVs	Based on experience		Variable screening		Var. screen. + cost fn.	
		DKG	44-D surrogate	DKG	44-D surrogate	DKG	44-D surrogate
d_1	X_1	1.8343	1.8371	1.8366	1.8359	1.8336	1.8425
d_2	X_2	2.1810	2.1799	2.1804	2.1807	2.1806	2.1771
d_3	X_3	2.8528	2.8578	2.8561	2.8576	2.8540	2.8654
d_4	X_4	1.9817	1.9807	1.9810	1.9851	1.9856	1.9525
d_5	X_5	2.7195	2.7222	2.7228	2.7233	2.7261	2.7209
d_6	X_6	2.2543	2.2488	2.2497	2.2501	2.2558	2.2464
d_7	X_7	2.3199	2.3173	2.3185	2.3169	2.3207	2.3265
d_8	X_8	1.7904	1.8	1.7966	1.7985	1.786	1.8
d_9	X_{10}	-	-	0.9	0.9	0.9	0.9
d_{18}	X_{20}	0.9	0.9	0.9	0.9	0.9	0.9
d_{19}	X_{23}	-	-	0.6	0.6	0.6	0.6
d_{21}	X_{25}	0.6897	0.6875	0.5429	0.5424	0.5871	0.5826
d_{22}	X_{26}	-	-	1.1	1.1	1.1	1.1
d_{24}	XN_1	0.7	0.7	0.7	0.7	0.7	0.7
d_{27}	XN_4	-	-	-	-	0.7	0.7
d_{32}	XN_9	0.6	0.6	-	-	0.6	0.6
d_{33}	XN_{10}	0.6	0.6	-	-	0.6	0.6
d_{34}	XN_{11}	0.9	0.9	-	-	0.9	0.9

Table 4.20 Reliability Analysis Result Using Full-Dimensional Surrogate Model

Performance measure	Baseline design	Based on experience	Variable screening	Variable screening + cost function	Full dimension
Cost	269.47	249.10	259.83	244.17	225.68
G_1	48.25%	10.06%	9.96%	10.00%	10.05%
G_2	51.34%	10.02%	10.11%	10.04%	10.09%
G_3	54.14%	0.00%	0.00%	0.00%	0.00%
G_4	55.57%	0.10%	0.12%	0.09%	0.12%
G_5	58.94%	1.96%	1.93%	1.98%	1.91%
G_6	59.70%	10.08%	10.05%	10.05%	10.00%
G_7	59.86%	10.20%	10.04%	9.91%	10.06%
G_8	53.23%	10.02%	10.03%	9.97%	9.14%
G_9	51.15%	10.02%	9.96%	9.96%	9.96%
G_{10}	49.10%	0.00%	0.00%	0.00%	0.00%
G_{11}	52.46%	17.70%	11.23%	11.17%	10.05%

CHAPTER 5
CONFIDENCE LEVEL OF RELIABILITY OUTPUT FOR CONFIDENCE-BASED RBDO

In this chapter, a newly developed method to estimate the confidence level of reliability output (probability of failure) is presented. In Section 5.1, the reliability output is represented in the general form for further derivations. In Section 5.2, the probability of the reliability output is decomposed into successive conditional probabilities using the Bayesian method, and those conditional probabilities are described in Sections 5.3, 5.4, and 5.5. In Section 5.6, a numerical method is proposed to calculate the confidence level. Finally, the proposed method is applied to a two-dimensional mathematical example in Section 5.7.

5.1 Reliability Output and Limited Data

In Section 2.1.3, the reliability output, which is the probability of failure p_F, was defined using a multi-dimensional integral and an indicator function as

$$p_F = \int_{G(\mathbf{X})>0} f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x} = \int_{\mathbb{R}^N} I_{\Omega_F}(\mathbf{x})\,f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x}\,. \tag{5.1}$$

where Ω_F is the failure domain such that a performance measure $G(\mathbf{x})$ is larger than zero (i.e., $G(\mathbf{x}) > 0$) and $I_{\Omega_F}(\bullet)$ is an indicator function defined as

$$I_{\Omega_F}(\mathbf{x}) \equiv \begin{cases} 1, & \text{for } \mathbf{x} \in \Omega_F \\ 0, & \text{otherwise} \end{cases}. \tag{5.2}$$

In Eq. (5.1), the input joint probability density function (PDF) $f_{\mathbf{X}}(\mathbf{x})$ has a distribution type and distribution parameters, which determine the probability of failure. Hence, Eq. (5.1) can be represented in the general form as

$$p_F(M,\boldsymbol{\psi}) \equiv P[\mathbf{X} \in \Omega_F] = \int_{\Omega_F} I_{\Omega_F}(\mathbf{x})\, f(\mathbf{x}\,|\,M,\boldsymbol{\psi})\,d\mathbf{x} \tag{5.3}$$

where M and $\boldsymbol{\psi}$ are the input distribution type and input distribution parameters, respectively. If the population data are available, the true input type M and the true values of $\boldsymbol{\psi}$ can be obtained. If this is the case, Eq. (5.3) produces a reliability output value. However, in practical engineering problems, only limited data are available, so M and $\boldsymbol{\psi}$ follow probability distributions instead of being fixed types or values. Consequently, the reliability output follows a probabilistic distribution, which is affected by M, $\boldsymbol{\psi}$, and the size of the limited data.

5.2 Probability of Reliability Output

Consider a limited input data set $*\mathbf{x}$. The input distribution type M and input distribution parameters $\boldsymbol{\psi}$ might be inferred from the given data set $*\mathbf{x}$. In this study, it is assumed that input distribution type M and parameters $\boldsymbol{\psi}$ follow certain probability distributions that can be analogized from the given data set $*\mathbf{x}$. Under the assumption, and using the Bayesian approach with the given data $*\mathbf{x}$, a joint PDF of the reliability output P_F, input distribution type M, and input distribution parameters $\boldsymbol{\psi}$ is obtained as

$$f(p_F, M, \boldsymbol{\psi} \mid *\mathbf{x}) = f(p_F \mid M, \boldsymbol{\psi}, *\mathbf{x})\, P(M \mid \boldsymbol{\psi}, *\mathbf{x})\, f(\boldsymbol{\psi} \mid *\mathbf{x})\,. \tag{5.4}$$

In Eq. (5.4), it can be seen that the joint PDF is a product of three successive conditional probabilities. If all terms in Eq. (5.4) are available, the marginal PDF of the reliability output P_F can be obtained by integrating M and $\boldsymbol{\psi}$ in Eq. (5.4) as

$$f_{P_F}(p_F \mid *\mathbf{x}) = \sum_M \int_{\Omega_{\boldsymbol{\psi}}} f(p_F, M, \boldsymbol{\psi} \mid *\mathbf{x})\, d\boldsymbol{\psi}, \quad p_F \in [0,1]\,. \tag{5.5}$$

Furthermore, the cumulative distribution function (CDF) of P_F is obtained by integrating Eq. (5.5) with respect to the reliability output as

$$F_{P_F}(p_F \mid *\mathbf{x}) = \int_0^{p_F} \sum_M \int_{\Omega_{\boldsymbol{\psi}}} f(\rho, M, \boldsymbol{\psi} \mid *\mathbf{x})\, d\boldsymbol{\psi}\, d\rho, \quad p_F \in [0,1] \tag{5.6}$$

where ρ is the variable that corresponds to the reliability output P_F. The value of CDF of P_F in Eq. (5.6) represents the probability that P_F of a design with limited data is less than the specified value p_F. That is, the CDF value is the probability that a design is safer (more conservative) than p_F. Hence, it is called the "confidence level" of the reliability output p_F in this study.

To obtain the probability of P_F in Eqs. (5.5) and (5.6), all terms on the right side of Eq. (5.4) need to be identified. The first term is the probability of P_F with the given M, $\boldsymbol{\psi}$ and the data $*\mathbf{x}$. As shown in Eq. (5.3), the reliability output is determined by the input distribution type M and parameters $\boldsymbol{\psi}$. Consequently, when M and $\boldsymbol{\psi}$ are given, the reliability output is a deterministic value, and the probability of it becomes a Dirac-delta measure as

$$f_{P_F}(p_F \mid M, \boldsymbol{\psi}, *\mathbf{x}) = \delta[p_F - p_F(M, \boldsymbol{\psi})]. \tag{5.7}$$

The second and third terms on the right side of Eq. (5.4) are obtained in the following sections.

5.3 Input Data

Before explaining the probabilities of input distribution type M and parameters $\boldsymbol{\psi}$, the given data set $*\mathbf{x}$ is described in this section. For simplicity of explanation, the number of data for each input random variable is set to ND. This can be easily extended to a case in which the numbers of data are not the same. The input data set $*\mathbf{x}$ could contain the following data subsets:

$$*\mathbf{x} = \{*\mathbf{x}_1, \ldots, *\mathbf{x}_N\}, \tag{5.8}$$

where N is the number of input random variables, and the data subset $*\mathbf{x}_i$ for the i-th random variable X_i is a column vector with size ND as

$$*\mathbf{x}_i = \begin{bmatrix} *x_i^{(1)} & *x_i^{(2)} & \cdots & *x_i^{(ND)} \end{bmatrix}^T. \tag{5.9}$$

The data subset $*\mathbf{x}_i$ can be decomposed into two parts as

$$*\mathbf{x}_i = *\overline{\mathbf{x}}_i + *\tilde{\mathbf{x}}_i \tag{5.10}$$

where $*\overline{\mathbf{x}}_i$ is a column vector with size ND whose total entities are the mean of the data as

$$*\overline{\mathbf{x}}_i = \begin{bmatrix} *\overline{x}_i & *\overline{x}_i & \cdots & *\overline{x}_i \end{bmatrix}^T \text{ such that } \quad *\overline{x}_i = \frac{1}{ND}\sum_{m=1}^{ND} *x_i^{(m)}. \tag{5.11}$$

Some of the input random variables are related to design variables. If X_i is related to a design variable d_i, the i-th data subset $*\mathbf{x}_i$ is changed in the reliability-based design optimization (RBDO) process as

$$*\mathbf{x}_i^{RBDO} = \mathbf{d}_i + *\tilde{\mathbf{x}}_i \tag{5.12}$$

where $\mathbf{d}_i$ is the i-th design point vector defined as

$$\mathbf{d}_i = \begin{bmatrix} d_i & d_i & \cdots & d_i \end{bmatrix}^T. \tag{5.13}$$

In Eq. (5.12), the input data in the RBDO process is changed to be centered at the current design point. However, $*\tilde{\mathbf{x}}$, which is the dispersion of the data with respect to the design point, is maintained in the RBDO process. The data decomposition in Eqs. (5.10) and (5.12) is a usual practice in the RBDO process, but this is discussed in this section to note that $*\tilde{\mathbf{x}}$ is the part where the input uncertainty exists and that it remains the same while the design is changed during the RBDO process. This is an important fact in the following sections.

5.4 Probability of Input Distribution Parameters

As discussed earlier, the input distribution parameters follow a certain distribution instead of being fixed values when only limited data is available. However, the exact distribution of input distribution parameters is not known. The exact distribution of the parameters is available only when population data of the parameters are given. However, the population data of the parameters can be obtained only when the population data of the input random variable **X** are available. If the population data of **X** are available, there is no issue of limited data. Otherwise, only an approximated distribution of the input distribution parameters can be obtained. There could be a number of parameters for an input model, such as mean, variance (or standard deviation), skewness, kurtosis, Spearman's rho, Kendall's tau, etc. In this study, two parameters of mean and variance are considered for marginal PDF and CDF, and Kendall's tau is used for statistical correlation between two random variables.

5.4.1 Probability Distribution of Input Mean and Variance

The central limit theorem (CLT) is a widely used method for obtaining the distribution of the input mean (mean of input random variables **X**) with the given input data. Though CLT produces the distribution of the input mean under the assumption that the data follow normal distribution, it produces a well-approximated distribution of the input mean when the input data follow other distributions. In the same sense, the distributions of the input mean and variance are obtained using Bayes' theorem under the assumption that the given input data *$\mathbf{x}$ follow normal distribution in this study. This does not mean that the input distribution type M is normal distribution; this is only an intermediate assumption to find the approximate distribution of the input mean and variance. It will be shown that the result of Bayes' theorem is the same as the one from CLT. Also, the non-informative prior, which means that there is no information except the given input data, is used for Bayes' theorem.

Under the normality assumption described above and with the non-informative prior, the input variance v_i, for the i-th independent random variable X_i and the given data subset $*\mathbf{x}_i$, follows inverse-gamma distribution as (Gelman et al., 2004)

$$v_i \mid *\mathbf{x}_i \sim \text{Inv-}\chi^2(ND-1,\ s_i^2) = \text{IG}\left(\frac{ND-1}{2}, \frac{(ND-1)s_i^2}{2}\right) \tag{5.14}$$

where the sample variance s_i^2 can be calculated as

$$s_i^2 = \frac{1}{ND-1}\sum_{m=1}^{ND}(*\tilde{x}_i^{(m)})^2 = \frac{1}{ND-1} *\tilde{\mathbf{x}}_i^T *\tilde{\mathbf{x}}_i\ . \tag{5.15}$$

In Eq. (5.14), s_i^2 and the amount of data ND consist of the parameters for the inverse-gamma distribution. It is noted that s_i^2 is determined by $*\tilde{\mathbf{x}}$ and ND, which do not change in the RBDO process. Hence, the distribution of v_i in Eq. (5.14) does not change during the RBDO process. Smaller ND produces larger variability in the input variance in Eq. (5.14). Therefore, the confidence level of reliability output decreases when less input data is available.

In this study, statistical correlation is assumed to happen only between two random variables. When the j-th and k-th input random variables X_j and X_k are correlated, the input covariance matrix $\mathbf{\Sigma}_{jk}$, for the given data subsets $*\mathbf{x}_j$ and $*\mathbf{x}_k$, follows inverse-Wishart distribution as (Gelman et al., 2004)

$$\mathbf{\Sigma}_{jk} \mid *\mathbf{x}_j, *\mathbf{x}_k \sim \text{Inv-Wishart}_{ND-1}\left(\mathbf{S}_{jk}^{-1}\right) \tag{5.16}$$

where $\mathbf{S}_{jk}$ is $(ND-1)$ times the sample covariance matrix for $*\mathbf{x}_j$ and $*\mathbf{x}_k$ as

$$\mathbf{S}_{jk} = \sum_{m=1}^{ND}\left[\begin{pmatrix} *\tilde{x}_j^{(m)} \\ *\tilde{x}_k^{(m)} \end{pmatrix}\begin{pmatrix} *\tilde{x}_j^{(m)} & *\tilde{x}_k^{(m)} \end{pmatrix}\right] = \begin{bmatrix} *\tilde{\mathbf{x}}_j & *\tilde{\mathbf{x}}_k \end{bmatrix}^T \begin{bmatrix} *\tilde{\mathbf{x}}_j & *\tilde{\mathbf{x}}_k \end{bmatrix}. \tag{5.17}$$

In Eq. (5.16), it can be seen that the distribution of the input covariance matrix $\mathbf{\Sigma}_{jk}$ also does not change in the RBDO process because $\mathbf{S}_{jk}$ and ND remain the same. Smaller ND will decrease the confidence level of the reliability output as well in Eq. (5.16).

The input mean μ_i of the i-th independent variable X_i, for the given input variance v_i and data $^*\mathbf{x}_i$, follows normal distribution as (Gelman et al., 2004)

$$\mu_i \mid v_i, {}^*\mathbf{x}_i \sim \mathcal{N}({}^*\overline{x}_i, v_i / ND). \tag{5.18}$$

As Eq. (5.18) is the same as the distribution from CLT, the distributions of the input variance and the mean in Eqs. (5.14) and (5.18) are reasonable and trustworthy. When X_i is related to a design variable d_i, Eq. (5.18) can be expressed as

$$\mu_i \mid v_i, {}^*\mathbf{x}_i^{RBDO} \sim \mathcal{N}(d_i, v_i / ND). \tag{5.19}$$

In Eqs. (5.18) and (5.19), smaller ND makes the input mean μ_i have larger variability, so the confidence level of the reliability output decreases. For the given covariance matrix $\mathbf{\Sigma}_{jk}$ as well as input data $^*\mathbf{x}_j$ and $^*\mathbf{x}_k$, the input mean vector $\boldsymbol{\mu}_{jk}$ of the correlated variables X_j and X_k follows bivariate normal distribution as

$$\boldsymbol{\mu}_{jk} \mid \mathbf{\Sigma}_{jk}, {}^*\mathbf{x}_j, {}^*\mathbf{x}_k \sim \mathcal{N}\left(\begin{pmatrix} {}^*\overline{x}_j \\ {}^*\overline{x}_k \end{pmatrix}, \frac{\mathbf{\Sigma}_{jk}}{ND}\right). \tag{5.20}$$

Again, Eq. (5.20) coincides with two-dimensional CLT. Hence, the distributions of the input covariance matrix and mean vector in Eqs. (5.16) and (5.20) are valid. If the correlated pair X_j and X_k is related to design variables d_j and d_k, Eq. (5.20) can be represented as

$$\boldsymbol{\mu}_{jk} \mid \mathbf{\Sigma}_{jk}, {}^*\mathbf{x}_j^{RBDO}, {}^*\mathbf{x}_k^{RBDO} \sim \mathcal{N}\left(\begin{pmatrix} d_j \\ d_k \end{pmatrix}, \frac{\mathbf{\Sigma}_{jk}}{ND}\right). \tag{5.21}$$

Also, smaller ND induces larger variability in the input mean vector; as a result, the confidence level of the reliability output decreases.

5.4.2 Probability Distribution of Kendall's Tau

In this study, the copula is used to describe statistical correlation between input random variables. The copula requires a measurement of dependence in the correlated input variables. The Kendall's tau can be used for the dependence measure; it estimates the dependence based on ranks in data pairs. Consider a correlated input random variable pair X_j and X_k. Then the dependence in the pair can be calculated using the Kendall's tau as (Genest and Favre, 2007)

$$t_{jk} = \frac{P_{ND} - Q_{ND}}{\binom{ND}{2}} = \frac{4}{ND(ND-1)} P_{ND} - 1 \tag{5.22}$$

where P_{ND} and Q_{ND} are number of concordant and discordant pairs, respectively. Here, two data pairs $({}^*\tilde{x}_j^{(m)}, {}^*\tilde{x}_k^{(m)})$ and $({}^*\tilde{x}_j^{(n)}, {}^*\tilde{x}_k^{(n)})$ for X_j and X_k are said to be concordant when $({}^*\tilde{x}_j^{(m)} - {}^*\tilde{x}_j^{(n)})({}^*\tilde{x}_k^{(m)} - {}^*\tilde{x}_k^{(n)}) > 0$ and discordant if $({}^*\tilde{x}_j^{(m)} - {}^*\tilde{x}_j^{(n)})({}^*\tilde{x}_k^{(m)} - {}^*\tilde{x}_k^{(n)}) < 0$.

As it is calculated from the data, the calculated value from Eq. (5.22) is a sample Kendall's tau (t_{jk}), not the input Kendall's tau (τ_{jk}). The insufficient data forces the input Kendall's tau to follow a certain distribution as well. An approximated distribution of τ_{jk} is given as (Genest and Favre, 2007)

$$\sqrt{ND}\,\frac{t_{jk} - \tau_{jk}}{4S} \sim \mathcal{N}(0, 1^2), \tag{5.23}$$

and it can be easily transformed to

$$\tau_{jk} \sim \mathcal{N}\left(t_{jk}, \left(\frac{4S_{jk}}{\sqrt{ND}} \right)^2 \right), \tag{5.24}$$

where

$$S_{jk}^2 = \frac{1}{ND} \sum_{m=1}^{ND} (W_m + \tilde{W}_m - 2\overline{W})^2, \tag{5.25}$$

$$W_m = \frac{1}{ND} \#\{n : {}^*\tilde{x}_j^{(n)} \le {}^*\tilde{x}_j^{(m)},\ {}^*\tilde{x}_k^{(n)} \le {}^*\tilde{x}_k^{(m)}\}, \tag{5.26}$$

$$\tilde{W}_m = \frac{1}{ND} \#\{n : {}^*\tilde{x}_j^{(m)} \le {}^*\tilde{x}_j^{(n)},\ {}^*\tilde{x}_k^{(m)} \le {}^*\tilde{x}_k^{(n)}\}, \text{ and} \tag{5.27}$$

$$\bar{W} = (W_1 + \cdots + W_{ND}) / ND. \tag{5.28}$$

The distribution of the Kendall's tau τ_{jk} in Eq. (5.24) is still an approximation, and it is known to be closer to the true distribution of τ_{jk} as the number of the data *ND* approaches infinity. When *ND* is small, the distribution of τ_{jk} becomes wide. Then, τ_{jk} can be out of its theoretical range [−1, 1]. To alleviate this problem, a 95% probability interval of the distribution of τ_{jk} in Eq. (5.24) is used in this study. This is discussed further in Section 5.6.

The Kendall's tau, the input mean, and the variance are statistically independent. Hence, the joint PDF of the input distribution parameters is a product of all PDFs identified in Sections 5.4.1 and 5.4.2 as

$$\begin{aligned} f(\boldsymbol{\psi} \mid {}^*\mathbf{x}) = &\prod_i \left[f(\mu_i \mid v_i, {}^*\mathbf{x}_i) f(v_i \mid {}^*\mathbf{x}_i) \right] \\ &\times \prod_{j,k} \left[f(\boldsymbol{\mu}_{jk} \mid \boldsymbol{\Sigma}_{jk}, {}^*\mathbf{x}_j, {}^*\mathbf{x}_k) f(\boldsymbol{\Sigma}_{jk} \mid {}^*\mathbf{x}_j, {}^*\mathbf{x}_k) \right]. \\ &\times \prod_{j,k} f(\tau_{jk} \mid {}^*\mathbf{x}_j, {}^*\mathbf{x}_k) \end{aligned} \tag{5.29}$$

5.5 Probability of Input Distribution Type

The probability of an input distribution type M with the given input data $*\mathbf{x}$ and parameters $\boldsymbol{\psi}$ is obtained using Bayes' theorem as

$$P(M \mid \boldsymbol{\psi}, *\mathbf{x}) = \frac{P(*\mathbf{x} \mid M, \boldsymbol{\psi})\, P(M \mid \boldsymbol{\psi})}{\sum_M P(*\mathbf{x} \mid M, \boldsymbol{\psi})\, P(M \mid \boldsymbol{\psi})} = \frac{L(*\mathbf{x}; M, \boldsymbol{\psi})\, P(M \mid \boldsymbol{\psi})}{\sum_M L(*\mathbf{x}; M, \boldsymbol{\psi})\, P(M \mid \boldsymbol{\psi})} \quad (5.30)$$

where the likelihood function $L(*\mathbf{x};\, M,\, \boldsymbol{\psi})$ is a product of the PDF value at each input data point as

$$\begin{aligned} L(*\mathbf{x}; M, \boldsymbol{\psi}) &= \prod_i \prod_m f(*x_i^{(m)} \mid M_i, \boldsymbol{\psi}_i) \\ &\times \prod_{j,k} \prod_n f(*x_j^{(n)}, *x_k^{(n)} \mid M_{jk}, \boldsymbol{\psi}_{jk}) \end{aligned} . \quad (5.31)$$

Here, M_i is the input distribution type for the i-th independent random variable X_i, $\boldsymbol{\psi}_i$ is $\{\mu_i,\, v_i\}$, M_{jk} is the input distribution type for the j-th and k-th correlated random variables X_j and X_k, and $\boldsymbol{\psi}_{jk}$ is $\{\boldsymbol{\mu}_{jk},\, \boldsymbol{\Sigma}_{jk},\, \tau_{jk}\}$.

The first term on the right side in Eq. (5.31) is a marginal PDF for X_i, which is expressed as

$$f(x_i \mid M_i, \boldsymbol{\psi}_i) = f_{M_i}(x_i; \mu_i, v_i) . \quad (5.32)$$

The second term in Eq. (5.31) is a joint PDF for the correlated pair X_j and X_k as (Nelson, 2006)

$$\begin{aligned} & f(x_j, x_k \mid M_{jk}, \boldsymbol{\psi}_{jk}) \\ & = c_{M_{jk}}(u_j, u_k; \theta_{jk})\, f_{M_j}(x_j; \mu_{jk,1}, \Sigma_{jk,11})\, f_{M_k}(x_k; \mu_{jk,1}, \Sigma_{jk,22}) \end{aligned} \quad (5.33)$$

where $c_{M_{jk}}$ is the copula density function; θ_{jk} is the correlation coefficient for the copula, which is a function of the Kendall's tau τ_{jk}; and u_j and u_k are the marginal CDF values of the marginal distribution at x_j and x_k, respectively. $\mu_{jk,m}$ is the m-th entry of the input mean vector $\boldsymbol{\mu}_{jk}$, and $\boldsymbol{\Sigma}_{jk,mn}$ is the mn-th entry of the input covariance matrix $\boldsymbol{\Sigma}_{jk}$. It is noted that the correlation between the input means $\mu_{jk,1}$ and $\mu_{jk,2}$ is already considered in Eqs. (5.20) and (5.21) by using the covariance matrix $\mathbf{\Sigma}_{jk}$, which has a correlation part.

Assuming that there is no prior information about which input distribution type is preferable, that is, all candidate distribution types are equally probable, the term $P(M\,|\boldsymbol{\psi})$ in Eq. (5.30) is a constant. Then, Eq. (5.30) can be simplified as

$$P(M\,|\,\boldsymbol{\psi},{}^*\mathbf{x}) = \frac{L({}^*\mathbf{x};M,\boldsymbol{\psi})}{\sum_{M} L({}^*\mathbf{x};M,\boldsymbol{\psi})}. \tag{5.34}$$

As mentioned before, there could be a case in which each input data subset has a different number of data. In this case, equations in this chapter can be generalized by replacing ND with ND_i for the i-th data subset ${}^*\mathbf{x}_i$.

5.6 Calculation of the Confidence Level of the Reliability Output

Since all terms on the right side of Eq. (5.4) are now available in Eqs. (5.7), (5.29), and (5.34), the confidence level of the reliability output in Eq. (5.6) at a reliability output value p_F can be calculated. When p_F is given, Eq. (5.7) needs to be evaluated to calculate the confidence level in Eq. (5.6). However, it is too complicated to solve Eq. (5.7) analytically as Eq. (5.7) involves the Dirac-delta measure. Therefore, the confidence level is calculated numerically using Monte Carlo simulation (MCS) as

$$\begin{aligned}
&F_{P_F}(p_F\,|\,{}^*\mathbf{x}) \\
&\simeq \tfrac{1}{NMCS_M\;NMCS_\psi}\int_0^{p_F}\sum_{n=1}^{NMCS_\psi}\sum_{m=1}^{NMCS_M} f_{P_F}(\rho\,|\,M^{(m)},\boldsymbol{\psi}^{(m)},{}^*\mathbf{x})\,d\rho \\
&= \tfrac{1}{NMCS_M\;NMCS_\psi}\int_0^{1}\sum_{n=1}^{NMCS_\psi}\sum_{m=1}^{NMCS_M} I_{[0,p_F]}(\rho)\,\delta[\rho - p_F(M^{(m)},\boldsymbol{\psi}^{(n)})]\,d\rho \\
&= \tfrac{1}{NMCS_M\;NMCS_\psi}\sum_{n=1}^{NMCS_\psi}\sum_{m=1}^{NMCS_M} I_{[0,p_F]}\left[\,p_F(M^{(m)},\boldsymbol{\psi}^{(n)})\right]
\end{aligned} \tag{5.35}$$

where $NMCS_\psi$, $NMCS_M$, $M^{(m)}$, and $\boldsymbol{\psi}^{(n)}$ are the MCS sample size for M and $\boldsymbol{\psi}$, the m-th realization of (M | $\boldsymbol{\psi}$,*$\mathbf{x}$), and the n-th realization of ($\boldsymbol{\psi}$ |*$\mathbf{x}$), respectively; and $I_{[0,p_F]}(\rho)$ is an indicator function whose value is 1 when ρ is in between 0 and p_F, and 0 otherwise. Here, the realizations of (M | $\boldsymbol{\psi}$,*$\mathbf{x}$) and ($\boldsymbol{\psi}$ |*$\mathbf{x}$) are drawn in accordance with the probabilities in Eqs. (5.29) and (5.34). As explained in Section 5.4.2, the distribution of the Kendall's tau may be out of the bound [−1, 1] when the amount of data is insufficient, and it can cause numerical error in the calculation of Eq. (5.35). To avoid the error, the realizations of Kendall's tau are drawn from Eq. (5.24), and those out of the 95% probability interval of Eq. (5.24) are moved to the closest point in the interval. The overall procedure to evaluate Eq. (5.35) is shown in Figure 5.1.

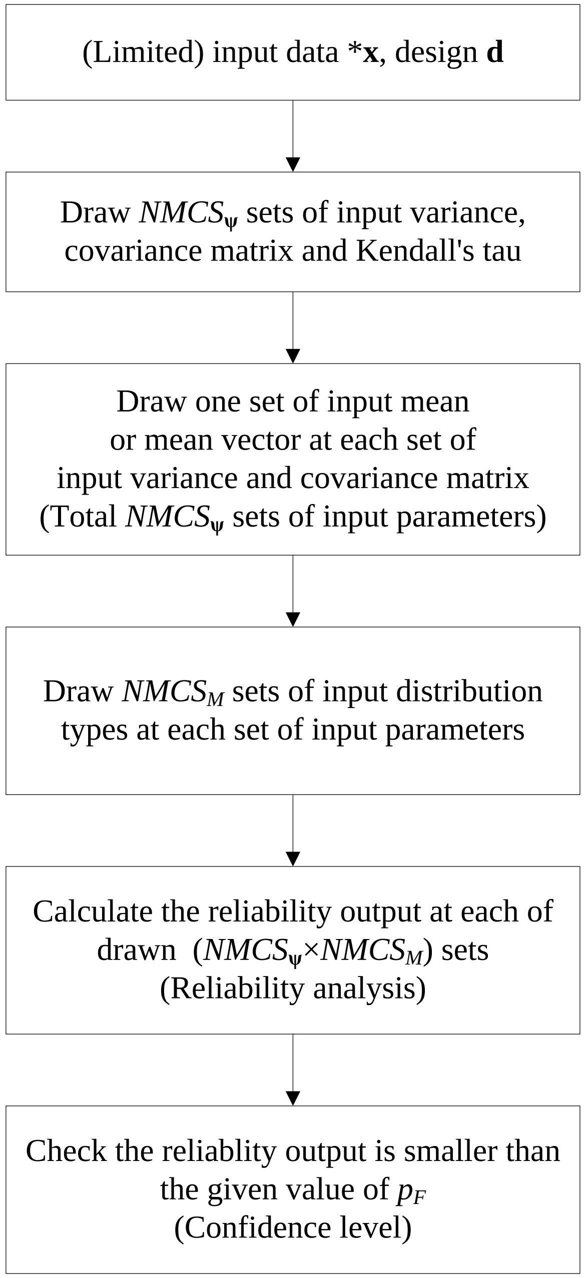


Figure 5.1 Flowchart of Confidence Level Calculation

5.7 Numerical Example

To check the confidence levels of the reliability output, three performance measures of the Iowa 2-D mathematical problem are considered as

$$
\begin{aligned}
G_1(\mathbf{X}) &= 1 - \frac{X_1^2 X_2}{20} \\
G_2(\mathbf{X}) &= -1 + (0.9063X_1 + 0.4226X_2 - 6)^2 + (0.9063X_1 + 0.4226X_2 - 6)^3 \\
&\quad - 0.6(0.9063X_1 + 0.4226X_2 - 6)^4 - (-0.4226X_1 + 0.9063X_2) \\
G_3(\mathbf{X}) &= 1 - \frac{80}{X_1^2 + 8X_2 + 5}
\end{aligned}
\tag{5.36}
$$

where X_1 and X_2 are input random variables. The limit states ($G_i = 0$) of Eq. (5.36) are shown in Figure 5.2, and $G_i < 0$ refers to the feasible area in this example. If X_1 and X_2 follow a benchmark distribution shown in Table 5.1 and the distribution is assumed to be known, the point $\mathbf{d} = [d_1 \;\; d_2]^T = [5.0541 \;\; 1.5918]^T$ in Figure 5.2 is the RBDO optimum with a target reliability output of 2.275%. Hence, the probabilities of failure for G_1 and G_2 are 2.275%. It can be seen that G_3 is an inactive constraint in Figure 5.2.

Table 5.1 Benchmark Input Distribution

Random Variable	Marginal Distribution	Mean	STDEV	Copula	Kendall's Tau
X_1	Normal	d_1 = 5.0541	0.3	Clayton	0.5
X_2	Normal	d_2 = 1.5918	0.3		

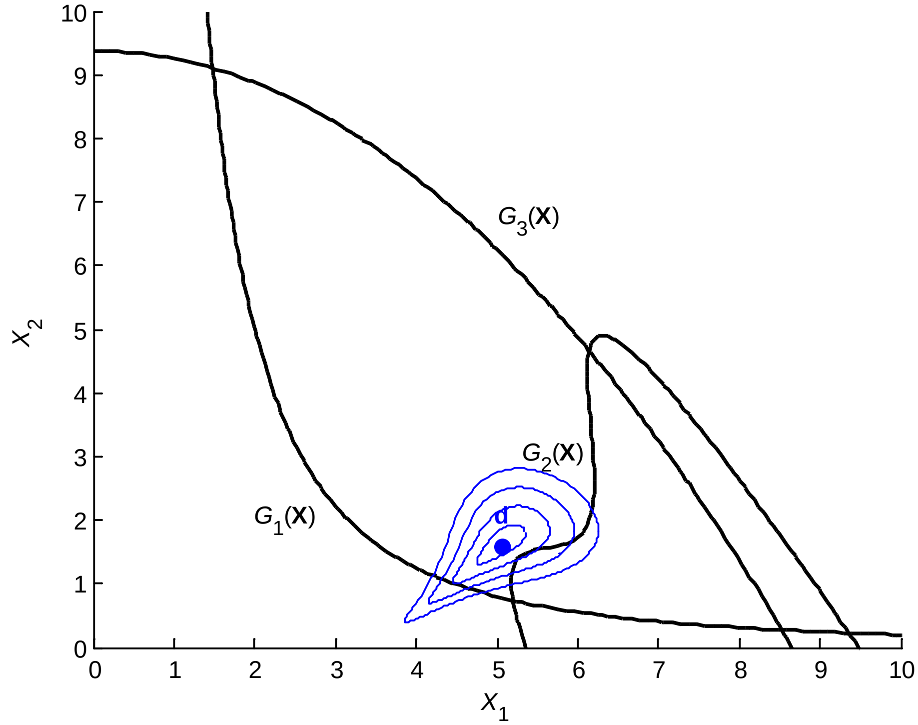


Figure 5.2 Limit States of Iowa 2-D Example
and Contour of Benchmark Distribution

5.7.1 Confidence Level Calculation Using 10 Data Pairs

As the proposed method is for a limited data problem, 10 pairs of input data are randomly drawn from the benchmark distribution in Table 5.1. To briefly explain random data generation, first, two sets of 10 random values are generated from a uniform distribution between 0 and 1, i.e., $U(0,1)$, using MATLAB's "rand" function, and they are treated as independent. Second, they are correlated according to the copula and Kendall's tau in Table 5.1 (Nelsen, 2006). Finally, they are transformed to random variable space (X-space) using the inverse CDF function of the marginal distributions in Table 5.1. The $^*\tilde{\mathbf{x}}_1$ and $^*\tilde{\mathbf{x}}_2$ of the drawn data are shown as

$$
{}^{*}\tilde{\mathbf{x}}_1 = \begin{bmatrix} 1.325\times10^{-1} \\ 2.585\times10^{-1} \\ -4.784\times10^{-1} \\ 2.724\times10^{-1} \\ -3.470\times10^{-2} \\ -5.248\times10^{-1} \\ -3.123\times10^{-1} \\ -1.008\times10^{-1} \\ 3.806\times10^{-1} \\ 4.070\times10^{-1} \end{bmatrix}, \quad {}^{*}\tilde{\mathbf{x}}_2 = \begin{bmatrix} -2.366\times10^{-1} \\ 4.678\times10^{-1} \\ -1.379\times10^{-1} \\ 5.664\times10^{-4} \\ 9.555\times10^{-2} \\ -6.767\times10^{-1} \\ -3.654\times10^{-1} \\ 1.970\times10^{-1} \\ 2.111\times10^{-1} \\ 4.445\times10^{-1} \end{bmatrix}. \tag{5.37}
$$

For the input distribution type M, the 20 candidate types listed in Table 5.2 are used. In this example, seven marginal distribution types (normal, lognormal, Weibull, Gumbel, gamma, extreme, and extreme type-II) and eight copula types (Clayton, Frank, FGM, Gaussian, AMH, Gumbel, A12, and A14) are considered. Hence, in a bivariate and correlated problem, there are 392 (= 7×7×8) combinations. However, considering all 392 combinations is ineffective and inefficient since a lot of them have very small probability. Hence, it is reasonable to narrow the number of candidates down. In this example, the 20 most probable types (according to their likelihoods) are selected from among those 392 at the design point **d** using the drawn data, means, and variances of the data. The most probable type has 1.58% of probability and the 20th most probable type has 1.10%. Since they have meaningful probabilities, the 20 of them are selected. Both numbers of MCS samples, $NMCS_{\Psi}$ and $NMCS_M$, are set to 10,000. Finally, following the procedure shown in Figure 5.1, the confidence level of the reliability output is calculated with the drawn 10 data pairs and 20 candidate distribution types, and the obtained result is shown in Figure 5.3.

Table 5.2 Candidate Input Distribution Types at **d** with 10 Data Pairs

No.	Marginal type for X_1	Marginal type for X_2	Copula
1	Weibull	Weibull	Gaussian
2	Extreme	Weibull	Gaussian
3	Extreme	Extreme	Gaussian
4	Extreme	Extreme	A12
5	Weibull	Extreme	Gaussian
6	Weibull	Extreme	A12
7	Weibull	Weibull	A14
8	Weibull	Weibull	A12
9	Weibull	Normal	Gaussian
10	Extreme	Weibull	A14
11	Extreme	Weibull	A12
12	Extreme	Normal	Gaussian
13	Normal	Weibull	Gaussian
14	Extreme	Extreme	A14
15	Normal	Weibull	A14
16	Gamma	Weibull	Gaussian
17	Weibull	Extreme	A14
18	Normal	Weibull	A12
19	Gamma	Weibull	A14
20	Lognormal	Weibull	Gaussian

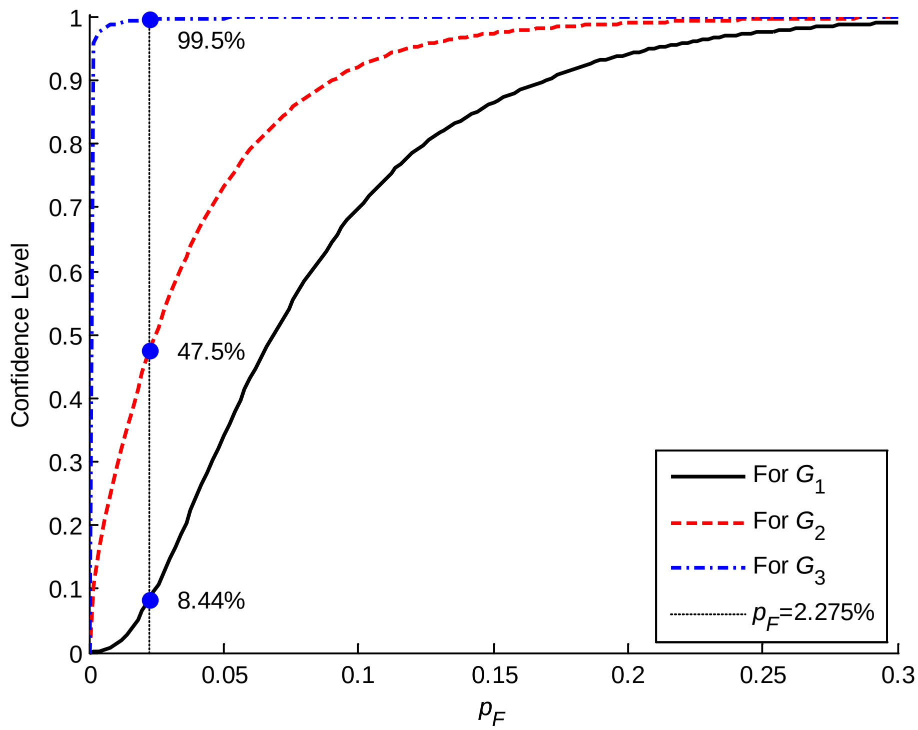


Figure 5.3 Confidence Level of Reliability Output with 10 Data Pairs

As mentioned before, the reliability output (probability of failure) at the point **d** is approximately 2.275% for the active constraints G_1 and G_2 assuming the benchmark distribution is known. However, the confidence levels, at the p_F of 2.275%, are 8.44% and 47.5% for G_1 and G_2, respectively, as shown in Figure 5.3. That means, at the design **d**, we have less than 50% confidence that the design will meet the target reliability output of 2.275% due to the limited input data. Consequently, conservativeness has to be applied to the design if only the 10 data pairs are available to assure that the target reliability output is satisfied. For the third constraint G_3, which is quite inactive, the confidence level increases rapidly from p_F = 0 and shows 99.5% confidence level at p_F = 2.275%. This is a reasonable result because the limit state of G_3 is far away enough from

d and thus a very conservative design with respect to G_3. As the input data in Eq. (5.37) are positively correlated, the contour of the joint PDF of **X** is wider in the $X_1 = X_2$ direction, whereas it is shorter in the $X_1 = -X_2$ direction (see Figure 5.2). That is, more uncertainties are induced in $X_1 = X_2$ direction than in the $X_1 = -X_2$ direction. Hence, the confidence level for G_1 (8.44%), which is affected by uncertainties in the $X_1 = X_2$ direction, is much smaller than the confidence level for G_2 (47.5%), which is affected by uncertainties in the $X_1 = -X_2$ direction as shown in Figures 5.2 and 5.3.

5.7.2 Confidence Level Calculation Using 20 Data Pairs

To understand how the amount of data affects the confidence level of the reliability output, 20 pairs of data are drawn again from the benchmark distribution in Table 5.1. Using the drawn data, 20 new candidate distribution types are chosen according to the same procedure used for the 10 pairs of data. ${}^{*}\tilde{\mathbf{x}}_1$ and ${}^{*}\tilde{\mathbf{x}}_2$ of the 20 pairs of data and the candidate input distribution types are shown in Eq. (5.38) and Table 5.3, respectively.

The obtained confidence level result is shown in Figure 5.4. At the same design point **d**, the confidence levels, at $p_F = 2.275\%$, are 36.7% and 65.5% for G_1 and G_2, respectively. Therefore, the result agrees with the expectation that the confidence level is more assured as more data are available. For the inactive constraint G_3, the confidence level is 100%, which has not changed much from the case of the 10 data pairs. This is reasonable because the confidence level is already maximized even with the 10 data pairs. The confidence level for G_1 (36.7%) is still smaller than the confidence level for G_2 (65.5%) for the same reason as for the case of the 10 data pairs.

Throughout two examples with 10 and 20 data pairs, it is shown that conventional RBDO, which requires true input distribution, cannot assure target reliability output when insufficient data are provided. Therefore, the confidence level of the reliability output should be incorporated in the RBDO for limited input data. It also can be seen that the developed method considers number of provided data appropriately when it is estimating the confidence level.

$$\tilde{\mathbf{x}}_1 = \begin{bmatrix} 2.587\times10^{-1} \\ -4.704\times10^{-1} \\ 2.800\times10^{-1} \\ -5.873\times10^{-1} \\ 2.603\times10^{-1} \\ 3.195\times10^{-1} \\ -2.031\times10^{-1} \\ -1.589\times10^{-1} \\ 1.648\times10^{-1} \\ -1.918\times10^{-1} \\ 7.165\times10^{-2} \\ 1.227\times10^{-1} \\ -2.250\times10^{-1} \\ -2.813\times10^{-1} \\ 2.538\times10^{-1} \\ 1.140\times10^{-1} \\ 1.608\times10^{-1} \\ 4.148\times10^{-1} \\ -2.760\times10^{-1} \\ -2.732\times10^{-2} \end{bmatrix},\ {}^{*}\tilde{\mathbf{x}}_2 = \begin{bmatrix} 2.198\times10^{-1} \\ -5.496\times10^{-1} \\ 4.337\times10^{-2} \\ -6.646\times10^{-1} \\ 6.284\times10^{-1} \\ 2.777\times10^{-1} \\ -1.637\times10^{-1} \\ -2.685\times10^{-1} \\ 2.329\times10^{-1} \\ 4.845\times10^{-1} \\ -1.100\times10^{-2} \\ -1.032\times10^{-1} \\ -1.651\times10^{-1} \\ -1.248\times10^{-1} \\ -2.317\times10^{-2} \\ 2.930\times10^{-1} \\ -1.171\times10^{-1} \\ 3.857\times10^{-1} \\ -3.762\times10^{-1} \\ 1.781\times10^{-3} \end{bmatrix} \tag{5.38}$$

Table 5.3 Candidate Input Distribution Types at **d** with 20 Data Pairs

No.	Marginal type for X_1	Marginal type for X_2	Copula
1	Lognormal	Gamma	Clayton
2	Normal	Normal	Clayton
3	Gamma	Gamma	Clayton
4	Weibull	Weibull	Clayton
5	Gamma	Normal	Clayton
6	Weibull	Normal	Clayton
7	Extreme	Weibull	Clayton
8	Normal	Gamma	Clayton
9	Lognormal	Normal	Clayton
10	Extreme	Normal	Clayton
11	Lognormal	Gamma	A12
12	Gamma	Gamma	A12
13	Normal	Weibull	Clayton
14	Normal	Gamma	A12
15	Lognormal	Lognormal	A12
16	Normal	Normal	A12
17	Lognormal	Lognormal	Clayton
18	Gamma	Lognormal	A12
19	Gamma	Weibull	Clayton
20	Gamma	Normal	A12

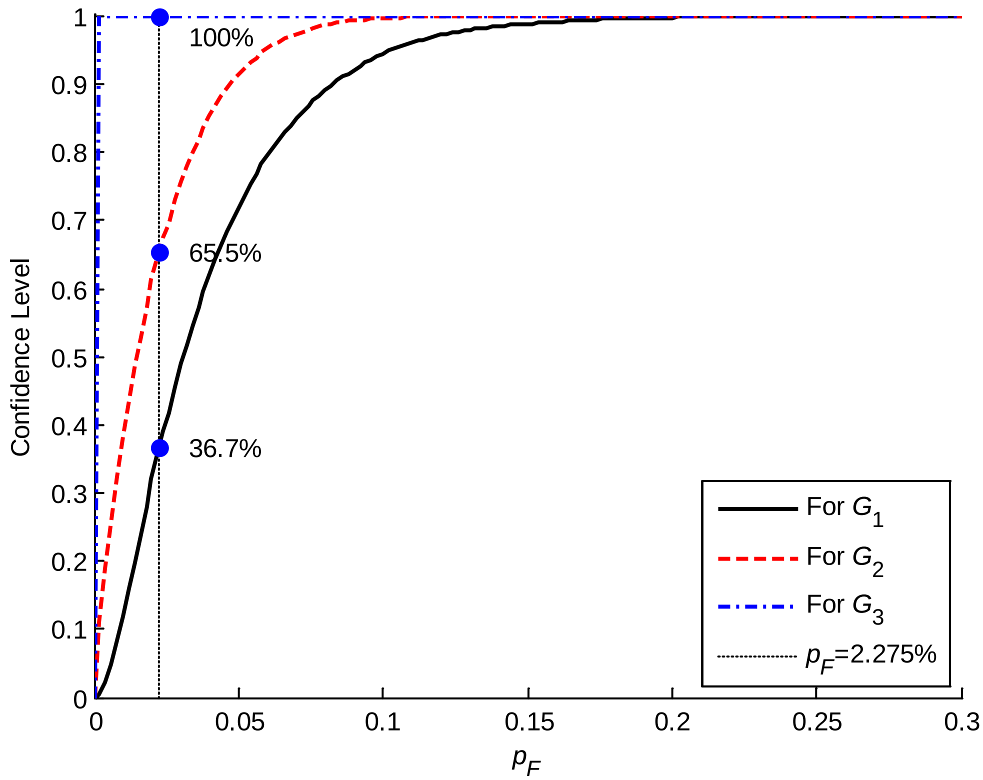


Figure 5.4 Confidence Level of Reliability Output with 20 Data Pairs

CHAPTER 6
FORMULATION AND DESIGN SENSITIVITY FOR CONFIDENCE-BASED RBDO

In this chapter, confidence-based RBDO (C-RBDO) and its design sensitivity are presented. In Section 6.1, C-RBDO is formulated using the confidence level estimation of the reliability output developed in Chapter 5 as its probabilistic constraints. In Sections 6.2 and 6.3, the design sensitivity of the confidence level is derived for the efficient and effective C-RBDO process. In Section 6.4, efficiency improvement methods are introduced for the C-RBDO. Finally, performances of the C-RBDO and the design sensitivity method are verified in Section 6.5 using numerical examples.

6.1 Confidence-based RBDO Formulation

The formulation of an RBDO problem has been shown in Section 2.3 as

$$\begin{aligned} &\text{minimize} && \text{Cost}(\mathbf{d}) \\ &\text{subject to} && P[G_i(\mathbf{X}) > 0] \le p_{F_i}^{Tar}, \; i = 1, \cdots, NC \\ & && \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^{NDV} \text{ and } \mathbf{X} \in \mathbb{R}^N \end{aligned} \tag{6.1}$$

where $\mathbf{X}$ is the N-dimensional random variable vector, $\mathbf{d}$ is the NDV-dimensional design variable vector, G_i is the i-th constraint function, $p_{F_i}^{Tar}$ is the target reliability output for the i-th constraint, and NC is the number of constraints. The formulation in Eq. (2.18) is based on an assumption that there is no uncertainty in the input probabilistic model. That is, an input distribution type and a set of parameters are selected and assumed to be “true” input type and parameters. However, it is no longer valid when the uncertainty in the input probabilistic model exists due to the limited data.

In Chapter 5, the confidence level of the reliability output has been obtained for the limited data. Therefore, using the confidence level as a probabilistic constraint, the RBDO formulation can be changed to a C-RBDO formulation as

$$\begin{array}{ll} \text{minimize} & \text{Cost}(\mathbf{d}) \\ \text{subject to} & F_{P_{F_i}}(p_{F_i}^{Tar} \mid {}^*\mathbf{x}) \geq CL_i^{Tar} \qquad i = 1, \ldots, NC \\ & \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^{NDV} \text{ and } \mathbf{X} \in \mathbb{R}^N \end{array} \tag{6.2}$$

where $p_{F_i}^{Tar}$ and CL_i^{Tar} are the target reliability output and target confidence level for the *i*-th constraint, respectively. By using the two target values, C-RBDO is able to obtain an optimum design that satisfies the target reliability output $p_{F_i}^{Tar}$ with the target confidence level CL_i^{Tar}. Therefore, the optimum point can secure appropriate conservativeness even with a finite number of data.

In the C-RBDO formulation, both the number of input data and the design point affect the confidence level of the reliability output. The target value CL_i^{Tar} at $p_{F_i}^{Tar}$ can be achieved by providing more input data to reduce uncertainty in the input probabilistic model or by finding a more reliable optimal design. If the feasible region of a given problem is too small to find a feasible optimum in accordance with Eq. (6.2), increasing the amount of input data would be inevitable to satisfy the target criteria.

6.2 Design Sensitivity of Confidence Level

The confidence-based RBDO in Eq. (6.2) uses the confidence level of the reliability output as its probabilistic constraints. Consequently, a new design sensitivity method for the confidence level is necessary to obtain a C-RBDO optimum design accurately as well as efficiently. For further derivation, the analytical form for the cumulative distribution function (CDF) of the reliability output is recalled as

$$\begin{aligned} F_{P_F}(p_F \mid {}^*\mathbf{x}) &= \int_0^{p_F} \sum_M \int_{\Omega_\psi} f(\rho, M, \boldsymbol{\psi} \mid {}^*\mathbf{x})\, d\boldsymbol{\psi}\, d\rho \\ &= \int_0^{p_F} \sum_M \int_{\Omega_\psi} f(\rho \mid M, \boldsymbol{\psi}, {}^*\mathbf{x})\, P(M \mid \boldsymbol{\psi}, {}^*\mathbf{x})\, f(\boldsymbol{\psi} \mid {}^*\mathbf{x})\, d\boldsymbol{\psi}\, d\rho \end{aligned}, \tag{6.3}$$

and the derivative of Eq. (6.3) with respect to a design variable d_i yields

$$\begin{aligned}
&\frac{\partial}{\partial d_i} F_{P_F}(p_F \mid {}^*\mathbf{x}) \\
&\quad = \frac{\partial}{\partial d_i} \int_0^{p_F} \sum_M \int_{\Omega_\psi} f(\rho \mid M, \boldsymbol{\psi}, {}^*\mathbf{x}) P(M \mid \boldsymbol{\psi}, {}^*\mathbf{x}) f(\boldsymbol{\psi} \mid {}^*\mathbf{x})\, d\boldsymbol{\psi}\, d\rho \\
&\quad = \int_0^{p_F} \sum_M \left\{ \int_{\Omega_\psi} f(\rho \mid M, \boldsymbol{\psi}, {}^*\mathbf{x}) P(M \mid \boldsymbol{\psi}, {}^*\mathbf{x}) f(\boldsymbol{\psi} \mid {}^*\mathbf{x}) \right. \\
&\qquad\qquad \left. \times \frac{\partial}{\partial d_i} \left[\ln P(M \mid \boldsymbol{\psi}, {}^*\mathbf{x}) + \ln f(\boldsymbol{\psi} \mid {}^*\mathbf{x}) \right] d\boldsymbol{\psi} \right\} d\rho
\end{aligned} \quad . \tag{6.4}$$

Compared with Eq. (6.3), there are two additional terms in Eq. (6.4). The first additional term is the change rate of the natural logarithm of the probability of input distribution type to the design change. This term is explained in detail in Section 6.3 and is defined for now as

$$\frac{\partial}{\partial d_i} \ln P(M \mid \boldsymbol{\psi}, {}^*\mathbf{x}) \equiv S_M(M, \boldsymbol{\psi}, {}^*\mathbf{x}, d_i)\,. \tag{6.5}$$

As discussed in Section 5.3, a design change in the optimization process does not affect the dispersion of input data. That is why the probabilities of the input variance and covariance matrix are independent of the design variable d_i. Likewise, the probability of the input Kendall's tau is not related to the design variable. Therefore, the second additional term in Eq. (6.4), which is the change rate of the natural logarithm of the probability of input parameters to the design change, is derived when d_i is the design variable that corresponds to an independent random variable X_i as

$$\begin{aligned}
\frac{\partial}{\partial d_i} \ln f(\boldsymbol{\psi} \mid {}^*\mathbf{x}) = \frac{\partial}{\partial d_i} \ln f(\mu_i \mid v_i, {}^*\mathbf{x}_i) &= \frac{ND\left(\mu_i - d_i\right)}{v_i} \\
\equiv S_{\boldsymbol{\psi}}^I(\mu_i, v_i, d_i, ND)
\end{aligned} \quad . \tag{6.6}$$

where μ_i, v_i and ND are given input mean, variance and number of data for the input random variable X_i, respectively. Consider a correlated random variable pair X_j and X_k

and a design variable d_j related to X_j. Then, the second additional term in Eq. (6.4) for d_j is derived as

$$\begin{aligned}\frac{\partial}{\partial d_j}\ln f(\boldsymbol{\psi}\,|\,{}^*\mathbf{x}) &= \frac{\partial}{\partial d_j}\ln f(\boldsymbol{\mu}_{jk}\,|\,\boldsymbol{\Sigma}_{jk},{}^*\mathbf{x}_j,{}^*\mathbf{x}_k) \\ &= \frac{ND}{1-\left(\Sigma_{jk,12}/\sqrt{\Sigma_{jk,11}\Sigma_{jk,22}}\right)^2}\left[\frac{\mu_{jk,m}-d_j}{\Sigma_{jk,mm}}-\frac{\Sigma_{jk,12}(\mu_{jk,n}-d_k)}{\Sigma_{jk,11}\Sigma_{jk,22}}\right] \\ &\equiv S_{\psi}^{C}(\boldsymbol{\mu}_{jk},\boldsymbol{\Sigma}_{jk},d_j,d_k,ND)\end{aligned} \tag{6.7}$$

where $\boldsymbol{\mu}_{jk}$ and $\mu_{jk,m}$ are given input mean vector and its m-th entry, respectively, $\boldsymbol{\Sigma}_{jk}$ and $\Sigma_{jk,mn}$ are given input covariance matrix and its mn-th entry, respectively, m = 1 and n = 2. For the other design variable d_k, Eq. (6.7) can be used after changing the position of d_j and d_k and setting m = 2 and n = 1.

Although both of the additional terms in Eq. (6.4) are obtained in Eqs. (6.5), (6.6), and (6.7), the design sensitivity cannot be calculated analytically. The reason is the same as why Eq. (6.3) is evaluated using the Monte Carlo simulation (MCS) method in Section 5.6. Hence, the design sensitivity in Eq. (6.4) is calculated using the MCS method. The design sensitivity for the design variable d_i, which corresponds to the independent input random variable X_i, is

$$\begin{aligned}\frac{\partial}{\partial d_i}F_{P_F}(p_F\,|\,{}^*\mathbf{x}) \simeq \frac{1}{NMCS_M\,NMCS_{\psi}}\sum_{n}^{NMCS_{\psi}}\sum_{m}^{NMCS_M}&\Big\{I_{[0,p_F]}\left[p_F(M^{(m)},\boldsymbol{\psi}^{(n)})\right] \\ &\times\left[S_M(M^{(m)},\boldsymbol{\psi}^{(n)},{}^*\mathbf{x},d_i)+S_{\psi}^{I}(\mu_i^{(n)},v_i^{(n)},d_i,ND)\right]\Big\},\end{aligned} \tag{6.8}$$

and the design sensitivity for the design variable d_j, which corresponds to the correlated input random variable X_j, is

$$\begin{aligned}\frac{\partial}{\partial d_j}F_{P_F}(p_F\,|\,{}^*\mathbf{x}) \simeq \frac{1}{NMCS_M\,NMCS_{\psi}}\sum_{n}^{NMCS_{\psi}}\sum_{m}^{NMCS_M}&\Big\{I_{[0,p_F]}\left[p_F(M^{(m)},\boldsymbol{\psi}^{(n)})\right] \\ &\times\left[S_M(M^{(m)},\boldsymbol{\psi}^{(n)},{}^*\mathbf{x},d_j)+S_{\psi}^{C}(\boldsymbol{\mu}_{jk}^{(n)},\boldsymbol{\Sigma}_{jk}^{(n)},d_j,d_k,ND)\right]\Big\}.\end{aligned} \tag{6.9}$$

Equations (6.8) and (6.9) are quite similar to the confidence level of the reliability output. Only the additional terms in Eqs. (6.5), (6.6), and (6.7) have to be calculated at each MCS sample, and they are computationally inexpensive. Hence, the design sensitivity can be calculated with little additional effort during the calculation of the confidence level of the reliability output. It is noted that the equations in this section can be easily generalized by replacing *ND* with ND_i for the *i*-th data subset $^*\mathbf{x}_i$ in a case where each input data subset has a different number of data.

6.3 Derivative of Natural Logarithm of Probability of Input Distribution Type

In Section 6.2, Eq. (6.5), which is the change rate of the natural logarithm of the probability of input distribution type to the design change, is required for the design sensitivity of the confidence level in Eqs. (6.8) and (6.9). The easiest way to calculate Eq. (6.5) is with the finite difference method (FDM) because it only requires evaluations of the probability of input distribution type in Eq. (5.34) at perturbed design and current design. However the FDM could be inaccurate when appropriate perturbation size is not provided. Specifically, when Eq. (5.34) involves several different input distribution types, there may be no unique perturbation size that is appropriate for all types. Hence, determining perturbation size could cause unnecessary difficulty and inaccuracy when calculating Eq. (6.5) using the FDM.

If analytical expressions of marginal probability density functions (PDFs) and copulas are available, Eq. (6.5) could be derived analytically by taking the derivative of the natural logarithm of Eq. (5.34) with respect to the design variable. First, the expression of data in Eq. (5.12) is recalled and the superscript “*RBDO*” is dropped for simplicity:

$$^*\mathbf{x}_i = \mathbf{d}_i + {}^*\tilde{\mathbf{x}}_i \tag{6.10}$$

where

$$*\mathbf{x}_i = \begin{bmatrix} *x_i^{(1)} & *x_i^{(2)} & \cdots & *x_i^{(ND)} \end{bmatrix}^T \text{ and} \tag{6.11}$$

$$\mathbf{d}_i = \begin{bmatrix} d_i & d_i & \cdots & d_i \end{bmatrix}^T . \tag{6.12}$$

Knowing that $*\tilde{\mathbf{x}}_i$ is invariant in the C-RBDO process, the derivative of a function $h(*\mathbf{x})$ with respect to design d_i is the summation of the derivative of the function with respect to data $*x_i^{(j)}$:

$$\frac{\partial}{\partial d_i} h(*\mathbf{x}) = \sum_{j=1}^{ND} \frac{\partial}{\partial * x_i^{(j)}} h(*\mathbf{x}) . \tag{6.13}$$

In Eq. (5.33), the probability of input distribution type is obtained as

$$P(M \mid \boldsymbol{\psi}, *\mathbf{x}) = \frac{L(*\mathbf{x}; M, \boldsymbol{\psi})}{\sum_M L(*\mathbf{x}; M, \boldsymbol{\psi})} \tag{6.14}$$

where the likelihood function $L(*\mathbf{x}; M, \boldsymbol{\psi})$ is defined as

$$L(*\mathbf{x}; M, \boldsymbol{\psi}) = \prod_i \prod_m f(*x_i^{(m)} \mid M_i, \boldsymbol{\psi}_i) \prod_{j,k} \prod_n f(*x_j^{(n)}, *x_k^{(n)} \mid M_{jk}, \boldsymbol{\psi}_{jk}) . \tag{6.15}$$

If the p-th random variable X_p is statistically independent, PDFs of X_p can be set apart from Eq. (6.15) for further derivation as

$$\begin{aligned} L(*\mathbf{x}; M, \boldsymbol{\psi}) = & \prod_{i \neq p} \prod_m f(*x_i^{(m)} \mid M_i, \boldsymbol{\psi}_i) \\ & \times \prod_{j,k} \prod_n f(*x_j^{(n)}, *x_k^{(n)} \mid M_{jk}, \boldsymbol{\psi}_{jk}) \prod_m f_{M_p}(*x_p^{(m)}; \mu_p, \nu_p) \end{aligned} . \tag{6.16}$$

Furthermore, if the q-th random variable is statistically correlated with the r-th random variable, Eq. (6.15) also can be expressed by separating out joint PDFs of X_q and X_r as

$$\begin{aligned} & L(^*\mathbf{x}; M, \boldsymbol{\psi}) \\ & = \prod_i \prod_m f(^*x_i^{(m)} \mid M_i, \boldsymbol{\psi}_i) \prod_{j \neq q, k \neq r} \prod_n f(^*x_j^{(n)}, ^*x_k^{(n)} \mid M_{jk}, \boldsymbol{\psi}_{jk}) \cdot \\ & \quad \times \prod_n \left\{ c_{M_{qr}}(^*u_q^{(n)}, ^*u_r^{(n)}; \theta) f_{M_q}(^*x_q^{(n)}; \mu_q, v_q) f_{M_r}(^*x_r^{(n)}; \mu_r, v_r) \right\} \end{aligned} \tag{6.17}$$

where $^*u_q^{(n)} = F_{M_q}(^*x_q^{(n)}; \mu_q, v_q)$ which is the CDF of X_q at $^*x_q^{(n)}$. Then, the derivatives of the natural logarithms of Eqs. (6.16) and (6.17) are derived using Eq. (6.13) as

$$\begin{aligned} \frac{\partial}{\partial d_p} \ln L(^*\mathbf{x}; M, \boldsymbol{\psi}) &= \sum_{m=1}^{ND} \frac{\partial}{\partial ^* x_p^{(m)}} \ln f_{M_p}(^*x_p^{(m)}; \mu_p, v_p) \\ &= \sum_{m=1}^{ND} \frac{1}{f_{M_p}(^*x_p^{(m)}; \mu_p, v_p)} \frac{\partial f_{M_p}(^*x_p^{(m)}; \mu_p, v_p)}{\partial ^* x_p^{(m)}} \quad \text{and} \end{aligned} \tag{6.18}$$

$$\begin{aligned} & \frac{\partial}{\partial d_q} \ln L(^*\mathbf{x}; M, \boldsymbol{\psi}) \\ & = \sum_{n=1}^{ND} \frac{\partial}{\partial ^* x_q^{(n)}} \left[\ln \left\{ c_{M_{qr}}(^*u_q^{(n)}, ^*u_r^{(n)}; \theta) \right\} + \ln \left\{ f_{M_q}(^*x_q^{(n)}; \mu_q, v_q) \right\} \right] \\ & = \sum_{n=1}^{ND} \left[\frac{\partial \ln \left\{ c_{M_{qr}}(^*u_q^{(n)}, ^*u_r^{(n)}; \theta) \right\}}{\partial ^* u_q^{(n)}} f_{M_q}(^*x_q^{(n)}; \mu_q, v_q) \right. \\ & \qquad \left. + \frac{1}{f_{M_q}(^*x_q^{(n)}; \mu_q, v_q)} \frac{\partial f_{M_q}(^*x_q^{(n)}; \mu_q, v_q)}{\partial ^* x_q^{(n)}} \right] , \end{aligned} \tag{6.19}$$

respectively. In Eqs. (6.18) and (6.19), the derivatives of PDF *f* and copula density function *c* are required, and the derivatives of commonly used PDFs and copula density functions are shown in Table 6.1 and Table 6.2, respectively. Using the derivatives, Eqs. (6.18) and (6.19), the derivative of the natural logarithm of Eq. (6.14) can be obtained as

$$\begin{aligned} \frac{\partial}{\partial d_i} \ln P(M \mid \boldsymbol{\psi}, ^*\mathbf{x}) &= \frac{\partial}{\partial d_i} \left[\ln L(^*\mathbf{x}; M, \boldsymbol{\psi}) - \ln \sum_M L(^*\mathbf{x}; M, \boldsymbol{\psi}) \right] \\ &= \frac{\partial}{\partial d_i} \ln L(^*\mathbf{x}; M, \boldsymbol{\psi}) - \frac{1}{\sum_M L(^*\mathbf{x}; M, \boldsymbol{\psi})} \sum_M \left[L(^*\mathbf{x}; M, \boldsymbol{\psi}) \frac{\partial}{\partial d_i} \ln L(^*\mathbf{x}; M, \boldsymbol{\psi}) \right] . \end{aligned} \tag{6.20}$$

Also, Eq. (6.20) is exactly the same as Eq. (6.5). Hence, the sensitivity for C-RBDO in Eqs. (6.8) and (6.9) can be obtained using Eq. (6.20).

Table 6.1 PDFs and Derivatives of PDFs

Distribution type	PDF, $f_X(x)$	Derivative of PDF, df/dx
Normal	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$	$-\frac{(x-\mu)}{\sigma^2} f_X(x)$
Lognormal	$\frac{1}{\sqrt{2\pi}\bar{\sigma}x} e^{-\frac{1}{2}\left[\frac{\ln x-\bar{\mu}}{\bar{\sigma}}\right]^2}$	$-\frac{1}{x}\left(1+\frac{\ln x-\bar{\mu}}{\bar{\sigma}^2}\right) f_X(x)$
Weibull	$\frac{k}{v}\left(\frac{x}{v}\right)^{k-1} \exp\left[-\left(\frac{x}{v}\right)^{k}\right]$	$\left\{\frac{k-1}{x}-\frac{k}{v}\left(\frac{x}{v}\right)^{k-1}\right\} f_X(x)$
Gumbel	$\alpha \exp\left[-\alpha(x-v)-e^{-\alpha(x-v)}\right]$	$\alpha\left(e^{-\alpha(x-v)}-1\right) f_X(x)$
Gamma	$x^{a-1}\frac{e^{-x/b}}{\Gamma(a)b^a}$	$\left(\frac{a-1}{x}-\frac{1}{b}\right) f_X(x)$
Extreme	$\frac{1}{b}\exp\left[\frac{x-a}{b}-e^{\frac{x-a}{b}}\right]$	$\frac{1}{b}\left\{1-\exp\left(\frac{x-a}{b}\right)\right\} f_X(x)$
Extreme type-II	$\frac{a}{b}\left(\frac{b}{x}\right)^{a+1} \exp\left[-\left(\frac{b}{x}\right)^{a}\right]$	$\frac{1}{x}\left\{a\left(\frac{b}{a}\right)^{a}-a-1\right\} f_X(x)$

Table 6.2 Copula Density Functions and Derivatives

Copula	Copula density function, $c(u,v;\theta)$	Derivative, $d\ln c/du$
Clayton	$(1+\theta)(uv)^{-(1+\theta)}(-1+u^{-\theta}+v^{-\theta})^{-(2+1/\theta)}$	$\frac{(2\theta+1)u^{-(1+\theta)}}{-1+u^{-\theta}+v^{-\theta}}-\frac{1+\theta}{u}$
Frank	$\frac{\theta e^{\theta(1+u+v)}(e^{\theta}-1)}{\{e^{\theta}-e^{\theta(1+u)}-e^{\theta(1+v)}+e^{\theta(u+v)}\}^2}$	$\theta\left[1-\frac{2\left\{e^{\theta(u+v)}-e^{\theta(1+u)}\right\}}{e^{\theta}-e^{\theta(1+u)}-e^{\theta(1+v)}+e^{\theta(u+v)}}\right]$
FGM	$1+\theta(1-2u)(1-2v)$	$-\frac{2\theta(1-2v)}{1+\theta(1-2u)(1-2v)}$
Gaussian	$\frac{1}{\sqrt{1-\theta^2}}\exp\left[\frac{-\theta^2\left\{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2\right\}+2\theta\Phi^{-1}(u)\Phi^{-1}(v)}{2(1-\theta^2)}\right]$	$\frac{\theta}{1-\theta^2}\left\{\Phi^{-1}(v)-\theta\Phi^{-1}(u)\right\}\frac{1}{\phi(\Phi^{-1}(u))}$
AMH	$\frac{1+\theta^2(1-u)(1-v)-\theta(2-u-v-uv)}{[1-\theta(1-u)(1-v)]^3}$	$\frac{\theta(\theta v+v-\theta+1)}{1+\theta^2(1-u)(1-v)-\theta(2-u-v-uv)}-\frac{3\theta(1-v)}{1-\theta(1-u)(1-v)}$

*Φ: CDF of standard normal distribution, ϕ: PDF of standard normal distribution

Table 6.2 Continued

Copula	Copula density function, $c(u,v;\theta)$	Derivative, $d\ln c\,/\,du$
Gumbel	$\exp(A^{-1/\theta})\dfrac{(-\ln u)^{\theta-1}(-\ln v)^{\theta-1}}{uv}A^{1/\theta-2}(A^{1/\theta}+\theta-1)$ where $A=(-\ln u)^{\theta}+(-\ln v)^{\theta}$	$\dfrac{1}{u}\left[-1+\dfrac{\theta-1}{\ln u}+(-\ln u)^{\theta-1}\left\{A^{1/\theta-1}-\dfrac{A^{1/\theta-1}}{A^{1/\theta}+\theta-1}-\dfrac{1-2\theta}{A}\right\}\right]$
A12	$\dfrac{(u^{-1}-1)^{\theta-1}(v^{-1}-1)^{\theta-1}A^{1/\theta-2}\{(\theta+1)A^{1/\theta}+\theta-1\}}{u^{2}v^{2}(1+A^{1/\theta})^{3}}$ where $A=(u^{-1}-1)^{\theta}+(v^{-1}-1)^{\theta}$	$\dfrac{\theta-1}{u(u-1)}-\dfrac{(1-2\theta)(u^{-1}-1)^{\theta-1}}{u^{2}A}$ $-\dfrac{(1+\theta)A^{1/\theta-1}(u^{-1}-1)^{\theta-1}}{u^{2}\{(1+\theta)A^{1/\theta}+\theta-1\}}-\dfrac{2}{u}+\dfrac{3A^{1/\theta-1}(u^{-1}-1)^{\theta-1}}{u^{2}(A^{1/\theta}+1)}$
A14	$\dfrac{1}{\theta}(u^{-1/\theta}-1)^{\theta-1}(\mathrm{v}^{-1/\theta}-1)^{\theta-1}u^{-1(1/\theta+1)}v^{-1(1/\theta+1)}$ $\times\left(1+A^{1/\theta}\right)^{-\theta-2}A^{1/\theta-2}\left(2\theta A^{1/\theta}+\theta-1\right)$ where $A=(u^{-1/\theta}-1)^{\theta}+(v^{-1/\theta}-1)^{\theta}$	$-\left(\dfrac{1}{\theta}+1\right)\dfrac{1}{u}-\left(1-\dfrac{1}{\theta}\right)\dfrac{u^{-(1/\theta+1)}}{u^{-1/\theta}-1}+A^{1/\theta-1}(u^{-1/\theta}-1)^{\theta-1}u^{-(1/\theta+1)}$ $\times\left\{\dfrac{2-1/\theta}{A^{1/\theta}}+\dfrac{1+2/\theta}{A^{1/\theta}+1}-\dfrac{2}{2\theta A^{1/\theta}+\theta-1}\right\}$

6.4 Efficiency Improvements

Confidence-based RBDO is rather computationally expensive because it requires $NMCS_\Psi \times NMCS_M$ calculations of the reliability output in Eq. (5.35). Of course, the reliability analysis is required $NMCS_\Psi \times$(# of candidate distribution types) times practically; however it is still more computationally expensive than conventional RBDO. To alleviate the burden, two methods to improve efficiency of C-RBDO are used in this study.

6.4.1 Two-Step Reliability Analysis

For the C-RBDO process, accurate values of the reliability outputs are not necessary because C-RBDO uses the information whether or not a reliability output is larger than target reliability output. This fact could be used to reduce computational cost. Assuming the reliability output would be calculated using the MCS method, the error of the reliability output is estimated with 95% confidence level as (Haldar and Mahadevan, 2000)

$$\varepsilon\% = \sqrt{\frac{(1-p_F^{Tar})}{NMCS \times p_F^{Tar}}} \times 200\% \cdot \tag{6.21}$$

It is noted that the error means ε% of the target reliability output p_F^{Tar} and the error is very small as p_F^{Tar} is usually a small number. By rearranging Eq. (6.21), an appropriate number of MCS samples is obtained as

$$NMCS(\varepsilon\%) = \frac{40000(1-p_F^{Tar})}{\varepsilon^2\, p_F^{Tar}}. \tag{6.22}$$

In Eq. (6.22), it can be seen that $NMCS$ is inversely proportional to ε^2, while computational cost increases proportional to $NMCS$. Hence, the cost could be reduced by controlling ε and calculating the reliability output in two steps.

First, set a relatively large error level ε_p for the preliminary test. For example, 15% could be a reasonable number for ε_p. Then the number of MCS samples for preliminary test $NMCS_p$ becomes $NMCS$(15%) in Eq. (6.22). If a calculated reliability output with $NMCS_p$ samples is out of the range $[p_F^{Tar}(1 \pm \varepsilon_P/100)]$, the calculated value can be used for confidence level estimation. If the value is in the range, more MCS samples satisfying small error level of ε_a in Eq. (6.22) are used to accurately determine whether or not the calculated reliability is larger than target reliability output p_F^{Tar}. For example, ε_a could be set to 3%. Then the number of MCS samples for accurate test $NMCS_a$ becomes $NMCS$(3%) in Eq. (6.22). It is noted that $NMCS_p$=$NMCS$(15%) is only 4% of $NMCS_a$ =$NMCS$(3%) in Eq. (6.22). If approximately 20% of $NMCS_\Psi$×(# of candidate distribution types) falls into the range, the two-step method requires only 24% (=20% + 4%) of the computational time of the previous method; this is significant improvement.

6.4.2 Reusable Monte Carlo Simulation

Monte Carlo simulation requires different sample sets to estimate reliability outputs of different input distribution types. In conventional RBDO, reliability analysis using the MCS method would be performed for one input distribution type; however multiple distribution types are used in C-RBDO as it involves several candidate distribution types. Hence, it would be quite efficient if the MCS is performed for some input distribution types and the results propagate to other distribution types without any more simulations.

Important sampling could be applied to reuse an MCS result for other distribution types. The main idea of the important sampling is shown as (Rubinstein and Kroese, 2008)

$$\begin{aligned} p_F &= \int_\Omega I_{\Omega_F}(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} = \int_\Omega I_{\Omega_F}(\mathbf{x}) \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x})\, d\mathbf{x} \\ &= E_g\left[I_{\Omega_F}(\mathbf{X}) \frac{f(\mathbf{X})}{g(\mathbf{X})} \right] = E_g\left[I_{\Omega_F}(\mathbf{X}) w(\mathbf{X}) \right] \end{aligned} \tag{6.23}$$

where $f(\mathbf{X})$ is the PDF of the input random variable $\mathbf{X}$, $g(\mathbf{X})$ is the PDF of the MCS samples, Ω_F is the failure region of a constraint such that $G(\mathbf{x}) > 0$, $I_{\Omega_F}(\bullet)$ is an indicator function defined as

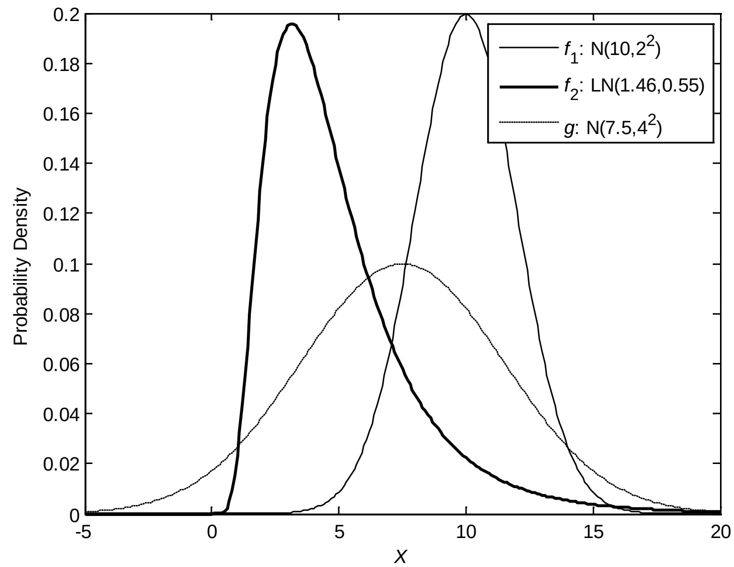
$$I_{\Omega_F}(\mathbf{x}) \equiv \begin{cases} 1, & \text{for } \mathbf{x} \in \Omega_F \\ 0, & \text{otherwise} \end{cases}, \tag{6.24}$$

and $w(\mathbf{X}) = f(\mathbf{X}) / g(\mathbf{X})$ is the weight function. Here, it is noted that MCS samples do not follow the input distribution $f(\mathbf{X})$. The samples follow $g(\mathbf{X})$, which is called “sampling distribution” in this study, and the constraint G is evaluated at the samples. Then, the evaluated result can transfer to the reliability output when the input random variable follows $f(\mathbf{X})$ using Eq. (6.23). Hence, several reliability output values can be obtained for different $f(\mathbf{X})$ just calculating the weight function $w(\mathbf{X})$ at the MCS samples. It is noted that the weight function is an analytical function, so it requires small computational time. The method in Eq. (6.23) is called “reusable MCS” in this study from now on.

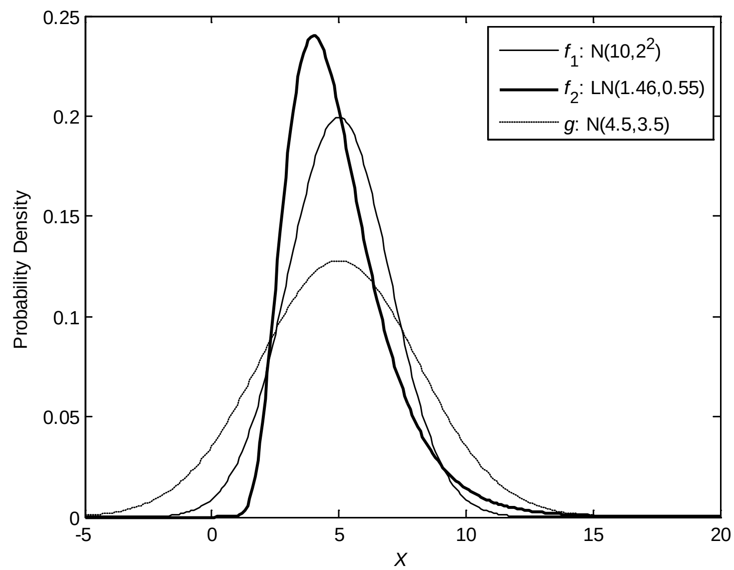
Reusable MCS has advantages and disadvantages. The main advantage is the efficiency improvement explained earlier. The disadvantage is loss of accuracy. The original MCS, which is the case of $f(\mathbf{X}) = g(\mathbf{X})$ in Eq. (6.23), is insensitive to dimensionality of $\mathbf{X}$. Moreover, the reliability result using original MCS is consistent when appropriate number of samples is used, as shown in Eq. (6.21). The key factor that determines the accuracy of reusable MCS is the sampling distribution $g(\mathbf{X})$. If $g(\mathbf{X})$ is appropriately selected, the loss of accuracy will be minimized. Theoretically, reusable MCS can estimate the reliability output of two different distributions with one sampling distribution. However, covering two different distributions of $f_1(\mathbf{X})$ and $f_2(\mathbf{X})$ could make sampling distribution $g(\mathbf{X})$ too wide, as shown in Figure 6.1(a). Then, the result might be inaccurate or inconsistent because the samples generated from $g(\mathbf{X})$, which cover the failure domain of $f_1(\mathbf{X})$ and $f_2(\mathbf{X})$, could be very small. Consequently, reusable MCS is used for cases with different distribution types but the same statistical parameters (mean and variance) in this study. As shown in Figure 6.1(b), sampling distribution is not very wide when it tries to cover distributions with different types but the same parameters.

For the sampling distribution, its distribution type and parameters should be determined. Normal distribution is a symmetric distribution with respect to the mean value of the input random variable. Hence it can be an ideal distribution type for the sampling distribution because it can distribute MCS samples evenly on both sides of the mean point. For the same reason, the Gaussian copula is selected to cover other copula types because it is symmetric to the correlation axis.

Statistical parameters of the sampling distribution should be selected to cover candidate distribution types. First, the mean value of the sampling distribution is selected as the same given mean value for the input random variable. When they are different, sampling distribution may not appropriately cover the domain of the input random variable, especially for correlated variables as shown in Figure 6.2(a). Given variances for input random variables are increased to cover all the candidate types; however, the ratio of increment in each dimension should be the same. If they are different, there is chance that the sampling distribution may not cover some candidate distribution types especially when there are correlated variables as shown in Figure 6.2(b). The increment of variance is selected as follows. First, the 95% probability interval of each candidate distribution type in each dimension is calculated. Then the variance of sampling distribution in each dimension is determined so the 95% probability interval of the sampling distribution can cover all the intervals of the candidate types.



(a) Covering distributions with different types and parameters

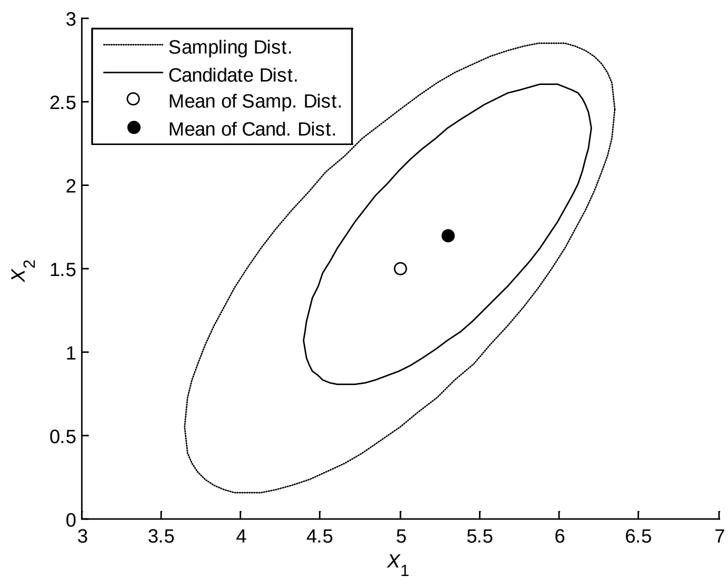


(b) Covering distributions with different types but same parameters

Figure 6.1 Coverage of Sampling Distribution

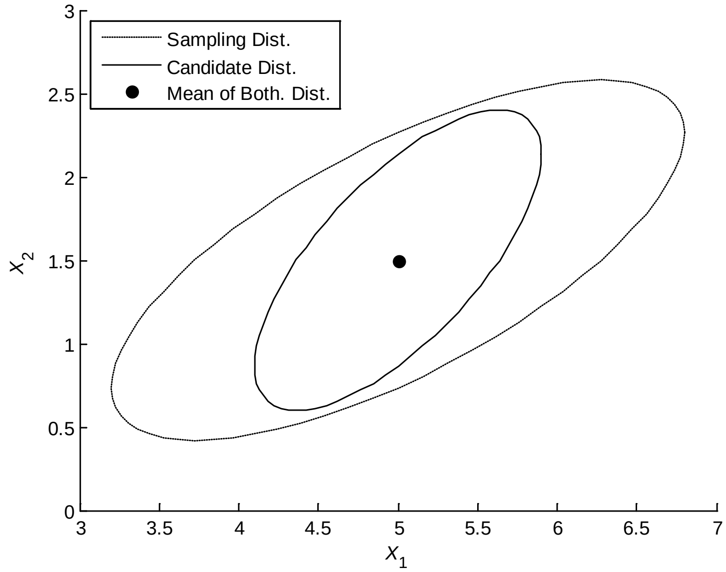
Among the increment in each dimension, the maximum ratio of increased variance to original is selected and applied to all dimensions to avoid the problem explained earlier (see Figure 6.2(b)). In this way, the sampling distribution shown in Figure 6.2(c) could be determined. Here, matching the “95%” probability interval may not be valid for the cases in which the target reliability output p_F^{Tar} is larger than 2.275% (2 or less-sigma design). When it is less, a larger probability interval should be used. Same Kendall’s tau given for input random variables is also used for the sampling distribution because the coverage of the sampling distribution is already achieved by increasing variance.

If reusable MCS is applied, the computational time due to number of candidate distribution types is minimized because it is replaced by analytic calculation. Therefore, reusable MCS could provide substantial improvements in the efficiency of C-RBDO.

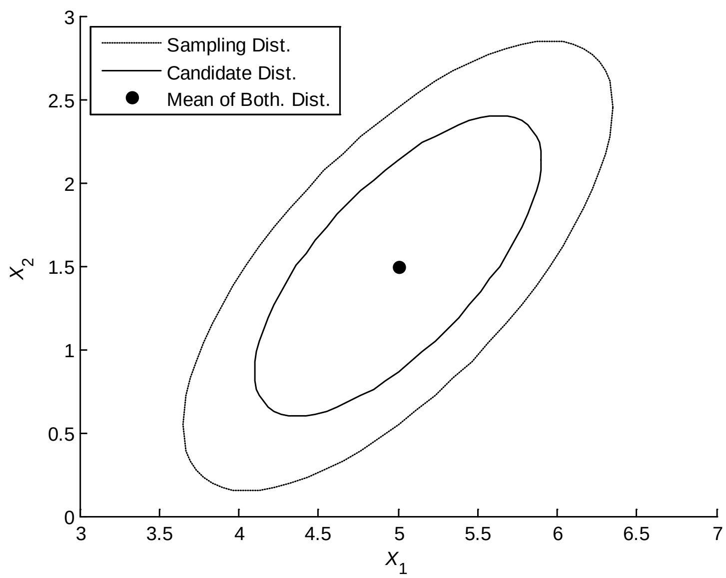


(a) Different mean value

Figure 6.2 Properties of Parameters for Sampling Distribution



(b) Different variance increment in each dimension



(c) Same mean and same variance increment

Figure 6.2 Continued

6.5 Numerical Example

A 2-D mathematical and a 7-D engineering example have been tested to verify the performance of the developed C-RBDO. The Iowa 2-D example is used to check the accuracy of the derived design sensitivity and explain the overall process of C-RBDO. In addition, the effectiveness and convergence of C-RBDO have been tested using the Iowa 2-D example. Using the speed-reducer 7-D problem, the performance of C-RBDO in a high-dimensional problem has been tested.

6.5.1 Iowa 2-D Example

The C-RBDO formulation for the Iowa 2-D example is formulated as

$$\begin{aligned} &\text{minimize} && \text{Cost}(\mathbf{d}) = -\frac{(d_1+d_2-10)^2}{30} - \frac{(d_1-d_2+10)^2}{120} \\ &\text{subject to} && F_{P_{F_i}}(p_{F_i}^{Tar} = 2.275\% \,|\, {}^*\mathbf{x}) \geq CL_i^{Tar} = 90\% \quad i=1,2,3\,. \\ & && \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^2 \text{ and } \mathbf{X} \in \mathbb{R}^2 \end{aligned} \tag{6.25}$$

where $\mathbf{d}^L$=[0 0]^T, $\mathbf{d}^U$=[10 10]^T, the feasible region is defined as $G_i \leq 0$, and the three performance measures (constraints) are defined as

$$\begin{aligned} G_1(\mathbf{X}) &= 1 - \frac{X_1^2 X_2}{20} \\ G_2(\mathbf{X}) &= -1 + (0.9063X_1 + 0.4226X_2 - 6)^2 + (0.9063X_1 + 0.4226X_2 - 6)^3 \\ &\quad - 0.6(0.9063X_1 + 0.4226X_2 - 6)^4 - (-0.4226X_1 + 0.9063X_2) \\ G_3(\mathbf{X}) &= 1 - \frac{80}{X_1^2 + 8X_2 + 5} \end{aligned} \,. \tag{6.26}$$

The contour of the cost function in Eq. (6.25) and the limit states (G_i = 0) of Eq. (6.26) are shown in Figure 6.3.

6.5.1.1 Design Sensitivity of Confidence Level

In this section, the accuracy of the derived design sensitivity of the confidence level is verified. As the design variable **d** in the Iowa 2-D example corresponds to the correlated random variable **X**, Eq. (6.9) is used to calculate the sensitivity. In Eq. (6.9), the two additional terms of S_M and S_ψ^C are the key parts of the sensitivity. The same 10 input data pairs used in Section 5.7.1 are used for the test of the design sensitivity, and their $^*\tilde{\mathbf{x}}_1$ and $^*\tilde{\mathbf{x}}_2$ are recalled here as

$$^*\tilde{\mathbf{x}}_1 = \begin{bmatrix} 1.325\times10^{-1} \\ 2.585\times10^{-1} \\ -4.784\times10^{-1} \\ 2.724\times10^{-1} \\ -3.470\times10^{-2} \\ -5.248\times10^{-1} \\ -3.123\times10^{-1} \\ -1.008\times10^{-1} \\ 3.806\times10^{-1} \\ 4.070\times10^{-1} \end{bmatrix}, \quad ^*\tilde{\mathbf{x}}_2 = \begin{bmatrix} -2.366\times10^{-1} \\ 4.678\times10^{-1} \\ -1.379\times10^{-1} \\ 5.664\times10^{-4} \\ 9.555\times10^{-2} \\ -6.767\times10^{-1} \\ -3.654\times10^{-1} \\ 1.970\times10^{-1} \\ 2.111\times10^{-1} \\ 4.445\times10^{-1} \end{bmatrix}, \tag{6.27}$$

Even though the S_M and S_ψ^C are related to the input covariance matrix, its effect is limited since the probability of the input covariance matrix remains the same when the design changes. Therefore, the effect of the input covariance matrix on the design and on the perturbed design are the same, whereas the change in the input mean vector is significant. Moreover, for the same reason, the effect of the Kendall's tau is limited as well. As the FDM method is too expensive to calculate very accurate sensitivity to be used as a reference, simplification of the problem is required to check the accuracy of the sensitivity with affordable computation power and time. In this sense, the input covariance matrix and the Kendall's tau are fixed at the sample covariance matrix and sample Kendall's tau. As explained, this is a valid and necessary simplification for an accuracy check of the sensitivity with FDM. It is noted that this simplification is for fast convergence of the FDM sensitivity, not for the developed sensitivity method. Therefore, the simplification is not applied to the actual C-RBDO process in the following section. In addition, the five candidate distribution types listed in Table 6.3 are used instead of the 20 types that are used in Section 5.7.1.

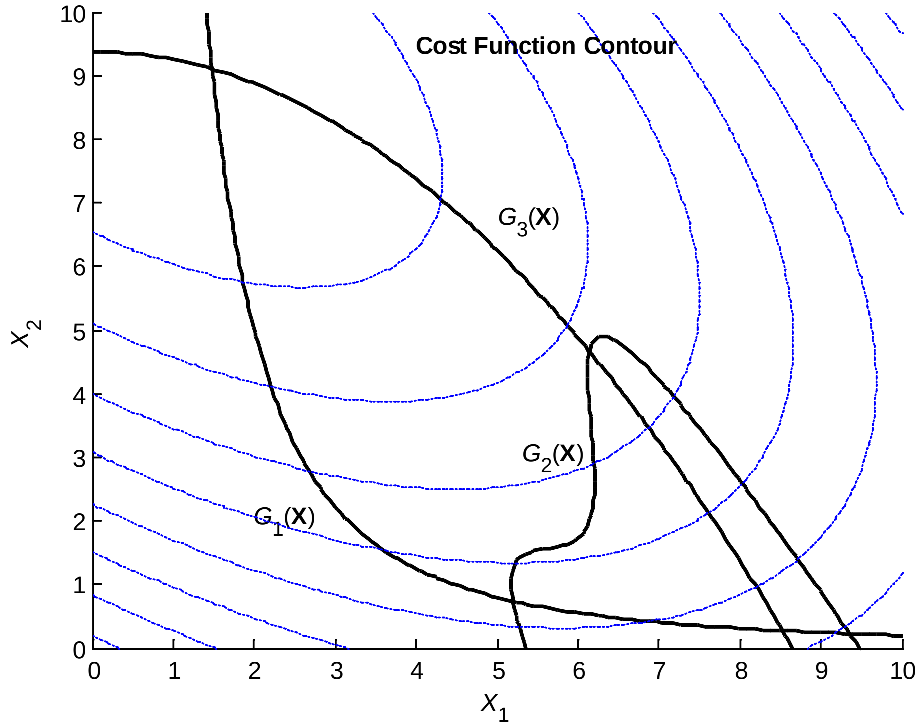


Figure 6.3 Cost Function Contour and Limit States in Iowa 2-D Example

Table 6.3 Candidate Input Distribution Types for Sensitivity Calculation

No.	Marginal type for X_1	Marginal type for X_2	Copula
1	Extreme	Extreme	Gaussian
2	Weibull	Extreme	Gaussian
3	Extreme	Weibull	Gaussian
4	Extreme	Extreme	A14
5	Weibull	Weibull	Gaussian

At the point $\mathbf{d} = [d_1 \ \ d_2]^T = [5.0541 \ \ 1.5918]^T$, the design sensitivity of the confidence level is calculated and compared with the FDM result. The FDM result is carried out by perturbing d_1 by 0.1% and using 2 million MCS samples for the input distribution parameter ($NMCS_\Psi$ = 2,000,000). Because the FDM requires evaluations at both the original design and the perturbed design, a total of 4 million MCS samples are actually used to calculate the sensitivity using the FDM. On the other hand, $NMCS_\Psi$ is set to 10,000 for the analytical sensitivity and needs only one evaluation. Thus, only 0.25% of the MCS samples for the FDM are used for the analytical sensitivity. For both the analytical and FDM sensitivities, $NMCS_M$ is set to 1,000. The accuracy check result is summarized in Table 6.4 and shows that the derived sensitivity agrees well with the FDM result. It is also noted that the FDM sensitivity is still converging as $NMCS_\Psi$ increases. Hence, the more accurate FDM sensitivity could be calculated with more MCS samples, and the discrepancy of 4.3% and 10.9% in Table 6.4 may be due to the inaccuracy of the FDM sensitivity. The sensitivity for the third constraint is not considered because the design point $\mathbf{d}$ is so far from the third constraint that the sensitivity is meaningless for design optimization.

Table 6.4 Analytical Sensitivity and FDM Sensitivity

Method	Sensitivity for $G_1(\mathbf{X})$	Accuracy	Sensitivity for $G_2(\mathbf{X})$	Accuracy
Analytic	0.709119	104.3%	-4.94209	110.9%
FDM	0.679895		-4.45638	

6.5.1.2 Confidence-Based RBDO for Iowa 2-D Example

Confidence-based RBDO is performed for the Iowa 2-D example in Eq. (6.25). Sequential quadratic programing (SQP) is used as the optimization method, and the design sensitivity in Eq. (6.9) is used. It is noted that the assumption in Section 6.5.1.1 is not imposed for the design sensitivity. For numerical efficiency, deterministic design optimization (DDO) is carried out first using the Iowa Reliability-Based Design Optimization (I-RBDO) code (Choi et al., 2012). Then, the conventional RBDO is performed using the I-RBDO from the deterministic optimum with the most likely distribution type and parameters, which are obtained using the given input data. Finally, the C-RBDO is performed at the conventional RBDO optimum. In this way, the computational effort for design iterations of the C-RBDO process is minimized.

Using the 10 data pairs in Eq. (6.27), C-RBDO has been performed. In addition, the optimization is also conducted for the 20 data pairs used in Section 5.7.2 and recalled in Eq. (6.28). Both $NMCS_{\Psi}$ and $NMCS_M$ are set to 10,000. Assuming that the conventional RBDO optimum design is close to the C-RBDO optimum design, candidate input distribution types are determined at the conventional RBDO optimum using the given input data. Following the same procedure in Section 5.7.1, the 20 candidate types listed in Table 6.5 and Table 6.6 are selected from among 392 (= 7×7×8) combinations at each of the conventional RBDO optimum designs using the data, means, and variances of the data. As a benchmark, the conventional RBDO optimum based on the benchmark input distribution, which is also used in Section 5.7 and recalled in Table 6.7, is used. As explained in the Section 5.7, the 10 and 20 data pairs are drawn from the benchmark input distribution. The results of C-RBDO and the benchmark are summarized in Table 6.8.

$$
{}^*\tilde{\mathbf{x}}_1 = \begin{bmatrix} 2.587\times10^{-1} \\ -4.704\times10^{-1} \\ 2.800\times10^{-1} \\ -5.873\times10^{-1} \\ 2.603\times10^{-1} \\ 3.195\times10^{-1} \\ -2.031\times10^{-1} \\ -1.589\times10^{-1} \\ 1.648\times10^{-1} \\ -1.918\times10^{-1} \\ 7.165\times10^{-2} \\ 1.227\times10^{-1} \\ -2.250\times10^{-1} \\ -2.813\times10^{-1} \\ 2.538\times10^{-1} \\ 1.140\times10^{-1} \\ 1.608\times10^{-1} \\ 4.148\times10^{-1} \\ -2.760\times10^{-1} \\ -2.732\times10^{-2} \end{bmatrix}, \quad {}^*\tilde{\mathbf{x}}_2 = \begin{bmatrix} 2.198\times10^{-1} \\ -5.496\times10^{-1} \\ 4.337\times10^{-2} \\ -6.646\times10^{-1} \\ 6.284\times10^{-1} \\ 2.777\times10^{-1} \\ -1.637\times10^{-1} \\ -2.685\times10^{-1} \\ 2.329\times10^{-1} \\ 4.845\times10^{-1} \\ -1.100\times10^{-2} \\ -1.032\times10^{-1} \\ -1.651\times10^{-1} \\ -1.248\times10^{-1} \\ -2.317\times10^{-2} \\ 2.930\times10^{-1} \\ -1.171\times10^{-1} \\ 3.857\times10^{-1} \\ -3.762\times10^{-1} \\ 1.781\times10^{-3} \end{bmatrix} \tag{6.28}
$$

Table 6.5 Candidate Input Distribution Types at the Conventional RBDO with 10 Data Pairs

No.	Marginal type for X_1	Marginal type for X_2	Copula
1	Weibull	Weibull	Gaussian
2	Extreme	Weibull	Gaussian
3	Extreme	Extreme	Gaussian
4	Extreme	Extreme	A12
5	Weibull	Weibull	A12
6	Weibull	Extreme	Gaussian
7	Weibull	Weibull	A14
8	Weibull	Extreme	A12
9	Extreme	Weibull	A12
10	Extreme	Weibull	A14
11	Weibull	Normal	Gaussian
12	Extreme	Normal	Gaussian
13	Normal	Weibull	Gaussian
14	Normal	Weibull	A12
15	Extreme	Extreme	A14
16	Normal	Weibull	A14
17	Gamma	Weibull	Gaussian
18	Weibull	Extreme	A14
19	Gamma	Weibull	A12
20	Gamma	Weibull	A14

Table 6.6 Candidate Input Distribution Types at the Conventional RBDO with 20 Data Pairs

No.	Marginal type for X_1	Marginal type for X_2	Copula
1	Lognormal	Gamma	Clayton
2	Normal	Normal	Clayton
3	Gamma	Gamma	Clayton
4	Weibull	Weibull	Clayton
5	Gamma	Normal	Clayton
6	Weibull	Normal	Clayton
7	Extreme	Weibull	Clayton
8	Normal	Gamma	Clayton
9	Lognormal	Normal	Clayton
10	Extreme	Normal	Clayton
11	Lognormal	Gamma	A12
12	Gamma	Gamma	A12
13	Normal	Weibull	Clayton
14	Normal	Gamma	A12
15	Lognormal	Lognormal	A12
16	Normal	Normal	A12
17	Lognormal	Lognormal	Clayton
18	Gamma	Lognormal	A12
19	Gamma	Weibull	Clayton
20	Gamma	Normal	A12

Table 6.7 Benchmark Input Distribution

Random variable	Marginal distribution	Mean	STDEV	Copula	Kendall's tau
X_1	Normal	d_1	0.3	Clayton	0.5
X_2	Normal	d_2	0.3		

Table 6.8 Optimum Design of Confidence-based RBDO and Benchmark

Case	Design iteration	Conf. level eval.	Optimum design		Cost	Confidence level (%)		
			d_1	d_2		G_1	G_2	G_3
10 data pairs	8	34	5.1527	2.5260	−1.5082	89.1	89.8	98.9
20 data pairs	7	8	5.2429	1.8835	−1.7626	91.0	91.3	100.0
Benchmark	-	-	5.0541	1.5918	−1.8853	-	-	-

As shown in Table 6.8, the C-RBDO optimum designs satisfy the given target confidence level of 90% for both the 10 and 20 input data pairs cases. Moreover, the C-RBDO processes are converged under eight design iterations, which is efficient enough. This could be an indication that the provided design sensitivity is quite accurate. Because G_3 is far enough from both optimums, the confidence levels for G_3 are almost 100% as shown in Table 6.8. Comparing both C-RBDO optimum designs with the benchmark design, the optimum cost is increased. For the 10 input data pairs case, 20% more cost (−1.8853 vs. −1.5082) is required to meet the target confidence level since significant uncertainty arises in the input probabilistic model due to the limited data. Hence, a more conservative design is obtained in C-RBDO using higher optimum cost. However, in the 20 data pairs case, the optimum cost value increases only 6.5% compared to the benchmark (−1.8853 vs. −1.7626) as it has more data. That is, less uncertainty is induced in the input probabilistic model, so a less conservative design is obtained than for the 10 data pairs case to satisfy the target confidence level.

To understand how conservative the C-RBDO optimum designs are, a conventional reliability analysis is performed at each of the C-RBDO optimums using the benchmark distribution (correct distribution) in Table 6.7. The calculated results are summarized in Table 6.9, and the optimum designs are graphically shown in Figure 6.4 as well. As shown in Table 6.9, the optimum design of the 10 data pairs case has probabilities of failure of 0.0022% and 0.0241% for G_1 and G_2, respectively, which are under 1/100 (0.0241/2.2791) of the result of the benchmark case. However, in the 20 data pairs case, the probabilities of failure are only 0.2190% and 0.3307% for G_1 and G_2, respectively, which are approximately 1/10 (0.3307/2.2791) of the result of the benchmark. Therefore, it can be concluded that the number of data is a crucial factor in the uncertainty in the input probabilistic model, especially when the data is limited.

The trend is also shown clearly in Figure 6.4. The deterministic design optimum usually has a probability of failure around 50%. To reduce the probability of failure, the conventional RBDO optimum with the given benchmark distribution is pushed inside the feasible region. It can be seen that the C-RBDO optimums are further pushed inside. However, the optimum of the 20 data pairs case is much closer to the benchmark optimum than that of the 10 data pairs case. Hence, it is confirmed that the 20 data pairs induce much less uncertainty than the 10 data pairs.

Table 6.9 Conventional Reliability Results

Case	Optimum design		Cost	Probability of failure (%)		
	d_1	d_2		G_1	G_2	G_3
10 data pairs	5.1527	2.5260	−1.5082	0.0022	0.0241	0.0
20 data pairs	5.2429	1.8835	−1.7626	0.2190	0.3307	0.0
Benchmark	5.0541	1.5918	−1.8853	2.2912	2.2791	0.0

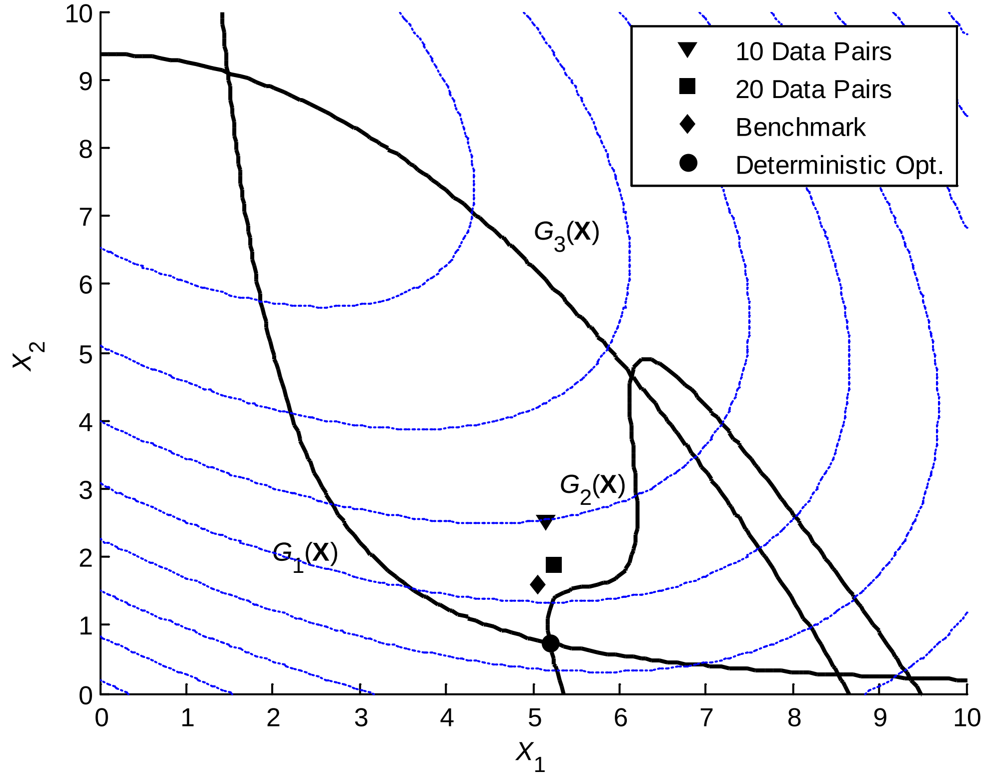


Figure 6.4 Optimum Designs in Iowa 2-D Example

6.5.1.3 Convergence Test

In the C-RBDO process, the only parameter that determines whether or not the C-RBDO optimum converges to the benchmark optimum is the number of data *ND*. Changing the target confidence level may expedite the convergence, but the opposite is also possible. In this section, the number of data pairs in the Iowa 2-D example is increased to 1000 to see if the C-RBDO optimum moves closer to the benchmark optimum.

For the convergence test, 50, 100, 200, 500, and 1000 pairs of data are drawn from the benchmark distribution. Statistical parameters calculated from the drawn data pairs are summarized in Table 6.10. The same procedure used in Section 6.5.1.2 has been followed for the large data pairs cases. First, the initial design is set to DDO as DDO is the same regardless of the number of data pairs. Conventional RBDO is performed for each case with the most likely distribution types and the parameters listed in Table 6.10. Then, the 20 candidate distribution types are determined at the conventional RBDO optimum, and C-RBDO is performed from the conventional RBDO optimum.

Table 6.10 Parameters Calculated from Data in Large Data Cases

Case	X_1		X_2		Kendall's tau
	Mean	STDEV	Mean	STDEV	
50 Data Pairs	4.957	0.303	4.950	0.286	0.564
100 Data Pairs	5.029	0.302	5.031	0.290	0.458
200 Data Pairs	4.974	0.287	4.984	0.287	0.485
500 Data Pairs	4.996	0.285	4.990	0.284	0.496
1000 Data Pairs	4.990	0.287	5.003	0.294	0.507
Benchmark	5.0	0.3	5.0	0.3	0.5

The obtained C-RBDO optimums as well as 10 and 20 data pairs and the benchmark optimums in Section 6.5.1.2 are listed in Table 6.11. Target confidence level CL^{Tar} is set to 90% for all cases, and 95% and 97.5% of the target confidence level are used for the 500 and 1000 data pairs cases. The seventh column of Table 6.11 shows the distance from each C-RBDO optimum to the benchmark optimum. Overall, the distance decreases as more data are provided in the C-RBDO procedure. Hence, it can be seen that the C-RBDO optimum converges to the benchmark optimum as the number of data increases. However, it is not always true, as shown in the cases of the 50, 100, and 200 data pairs for the 90% target confidence level. The C-RBDO optimum of the 50 data pairs case is closer to the benchmark optimum than the 100 data pairs case and equally close with the 200 data pairs case. The reason is that the 50 data pairs case has a relatively larger sample Kendall's tau (0.564) than the two other cases (0.458 and 0.485), as shown in Table 6.10. Large Kendall's tau allows an optimum to be close to the limit state of the second constraint G_2, which is less conservative. Hence, the optimum of the 50 data pairs case has a distance shorter than or equal to the benchmark optimum even though it has smaller data. The statistical information contained in data affects the conservativeness of the C-RBDO optimum, and consequently it may expedite or prolong the convergence of C-RBDO.

The C-RBDO optimums for the 90% target confidence level are graphically shown in Figure 6.5. It is evident that the optimums are gathered near the benchmark optimum when more than 50 data pairs are given. To see the convergence feature in detail, a small region near the benchmark optimum in Figure 6.5 is magnified in Figure 6.6. In the magnified plot, it is shown that optimum closer to the benchmark optimum is obtained as more data are provided except in the case of 50 data pairs, as addressed earlier. The results for the 95% and 97.5% target confidence levels are shown in Figure 6.6 as well. For the 500 data pairs case, increasing the target confidence level makes the optimums move upward, and the design becomes more conservative and requires more cost. However, they are positioned a similar distance from the benchmark optimum, which indicates that increasing the target confidence level does not affect the convergence of the optimums in this case. On the contrary, the C-RBDO optimums for 1000 data pairs are in-lined to the direction of the benchmark optimum. In this case, controlling the target confidence level expedites the convergence. Hence, it is evident that target confidence level is not a consistent parameter that affects the convergence of

C-RBDO optimums. Again, it is verified that the C-RBDO optimums converge to the benchmark optimum as number of data increases in Figures 6.5 and 6.6.

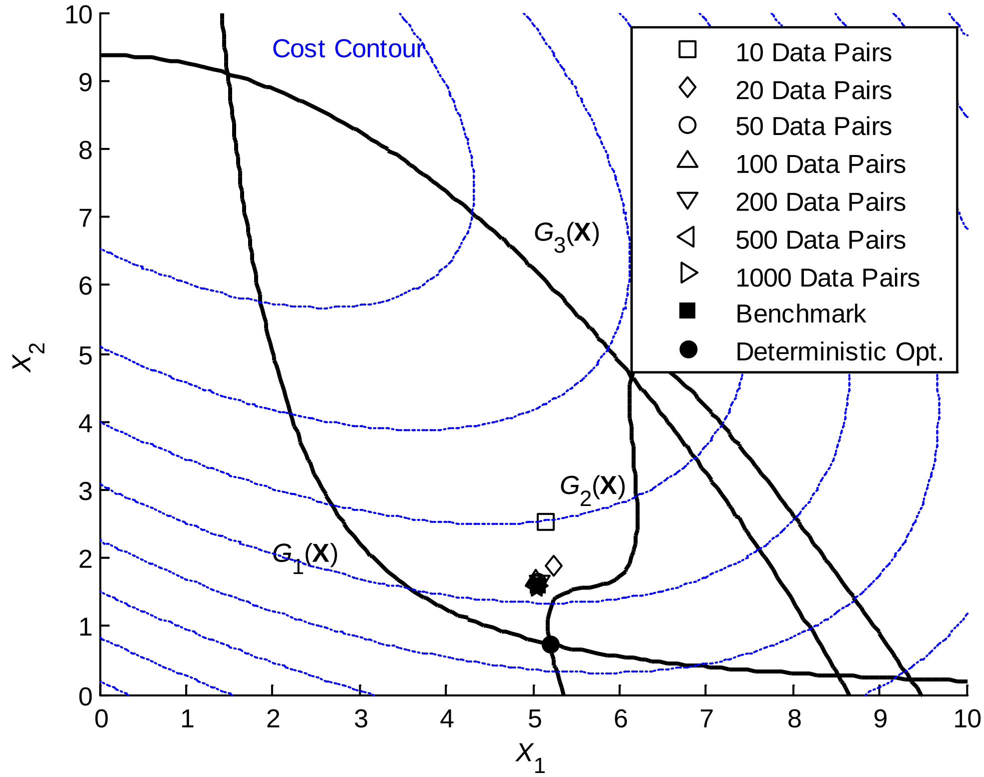


Figure 6.5 C-RBDO Optimum Designs for 90% CL^{Tar}

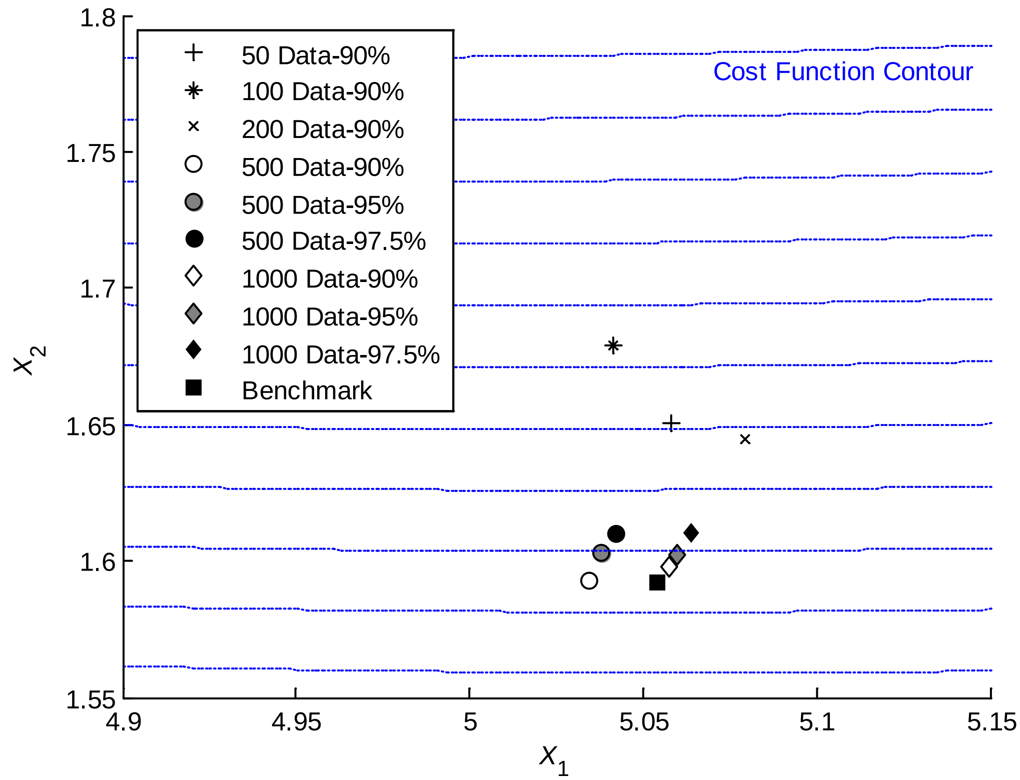


Figure 6.6 C-RBDO Optimum Designs near Benchmark Optimum

Table 6.11 Optimum Design of C-RBDO for Different Numbers of Data Pairs

Case	CL^{Tar}	Design Iter.	CL Eval.	Optimum design		Distance	Cost	Confidence level (%)		
				d_1	d_2			G_1	G_2	G_3
10 data pairs	90%	8	34	5.1527	2.5260	0.939	−1.5082	89.1	89.8	98.9
20 data pairs	90%	6	8	5.2429	1.8835	0.347	−1.7625	91.0	91.3	100.0
50 data pairs	90%	2	4	5.0581	1.6506	0.059	−1.8591	89.8	89.5	100.0
100 data pairs	90%	4	8	5.0413	1.6790	0.088	−1.8465	89.4	90.0	100.0
200 data pairs	90%	4	6	5.0793	1.6449	0.059	−1.8617	90.3	90.0	100.0
500 data pairs	90%	3	5	5.0344	1.5924	0.020	−1.8850	89.6	90.0	100.0
	95%	6	10	5.0385	1.6025	0.019	−1.8805	95.5	95.2	100.0
	97.5%	4	6	5.0421	1.6098	0.022	−1.8772	97.4	97.0	100.0
1000 data pairs	90%	3	5	5.0577	1.5979	0.007	−1.8826	89.5	91.3	100.0
	95%	3	5	5.0599	1.6018	0.012	−1.8808	94.1	94.3	100.0
	97.5%	4	5	5.0643	1.6103	0.021	−1.8770	97.7	97.6	100.0
Benchmark	-	-	-	5.0541	1.5918	0.0	−1.8853	-	-	-

6.5.2 Speed-Reducer 7-D Example

The speed-reducer example is a 7-D engineering problem that is frequently used for development of a design optimization method. The original problem is for deterministic design, and it has been changed for C-RBDO. A C-RBDO problem is formulated for the speed-reducer 7-D example using cost function (Golinski, 1970) and probabilistic constraints as

$$\begin{aligned} \text{minimize} \quad & \text{Cost}(\mathbf{d}) = 0.7854\,d_1 d_2^2 (3.3333\,d_3^2 + 14.9334\,d_3 - 43.0934) \\ & \qquad -1.508\,d_1(d_6^2 + d_7^2) + 7.477(d_6^3 + d_7^3) \\ & \qquad +0.7854(d_4 d_6^2 + d_5 d_7^2) \qquad . \\ \text{subject to} \quad & F_{P_{F_i}}(p_{F_i}^{Tar} = 2.275\% \,|\, {}^*\mathbf{x}) \geq CL_i^{Tar} \quad i = 1, \ldots, 11 \\ & \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^7 \text{ and } \mathbf{X} \in \mathbb{R}^7 \end{aligned} \tag{6.29}$$

where $\mathbf{d}^L$=[2.0 0.5 12 6.8 6.8 2.4 4.8]^T, $\mathbf{d}^U$=[4.0 0.9 32 8.8 8.8 4.4 5.8]^T, the feasible region is defined as $G_i \leq 0$, and the eleven constraints are defined as (Golinski, 1970)

$$\begin{aligned} & G_1(\mathbf{X}) = 27\,X_1^{-1} X_2^{-2} X_3^{-1} - 1 \\ & G_2(\mathbf{X}) = 397.5\,X_1^{-1} X_2^{-2} X_3^{-2} - 1 \\ & G_3(\mathbf{X}) = 1.93\,X_2^{-1} X_3^{-1} X_4^3 X_6^{-4} - 1 \\ & G_4(\mathbf{X}) = 1.93\,X_2^{-1} X_3^{-1} X_5^3 X_7^{-4} - 1 \\ & G_5(\mathbf{X}) = \sqrt{745^2 X_2^{-2} X_3^{-2} X_4^2 + 16.9 \times 10^6} \Big/ 110 X_6^3 - 1 \\ & G_6(\mathbf{X}) = \sqrt{745^2 X_2^{-2} X_3^{-2} X_5^2 + 157.5 \times 10^6} \Big/ 85 X_7^3 - 1 \; . \\ & G_7(\mathbf{X}) = X_2 X_3 / 40 - 1 \\ & G_8(\mathbf{X}) = 5X_2 - X_1 \\ & G_9(\mathbf{X}) = X_1 - 12X_2 \\ & G_{10}(\mathbf{X}) = 1.5X_6 - X_4 + 1.9 \\ & G_{11}(\mathbf{X}) = 1.1X_7 - X_5 + 1.9 \end{aligned} \tag{6.30}$$

The constraints include limitations on the bending stress of gear tooth, surface stress, transverse deflection of shafts 1 and 2 due to transmitted force, and stresses in shafts 1 and 2. Input random variables of the speed-reducer problem represent the face width (X_1), module of teeth (X_2), number of teeth on pinion (X_3), length of shaft 1 and 2 between bearings (X_4, X_5), and diameter of shaft 1 and 2 (X_6, X_7). All random variables are assumed to be independent in this example. A schematic of the speed reducer is shown in Figure 6.7.

A benchmark distribution in Table 6.12 is used as an exact input probabilistic model, and two sets of input data are drawn from the benchmark distribution as listed in Table 6.13 and Table 6.14. In this example, different random variables have different numbers of data. X_1 and X_2 have 20 data, X_3 to X_5 have 30 data, and X_6 and X_7 have 15 data. However, the two sets have the same numbers of data. In the bottom two rows of Table 6.13 and Table 6.14, sample means and standard deviations (means and standard deviations calculated using data) are shown. The ratio of the sample mean (or standard deviation) to the benchmark mean (or standard deviation) is shown as well. It can be seen that sample means are estimated very accurately as their ratios are in the range of 99.1% to 100.3% of the benchmark mean. On the contrary, the sample standard deviations are in the range of 67.3% to 107.3% of the benchmark standard deviations. That is, the data could not estimate standard deviations accurately. In particular, small standard deviations could yield not-enough-conservativeness as they are underestimating input randomness in the RBDO process. The small standard deviations are a typical example of uncertainty in the input probabilistic model due to limited data. C-RBDO can secure appropriate conservativeness by considering the uncertainty in the input probabilistic model.

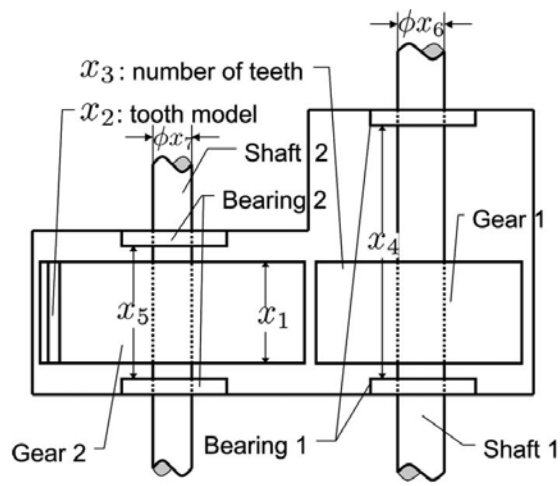


Figure 6.7 Schematic of Speed-Reducer

Source: Lu, S., and Kim, H. M., "A Regularized Inexact Penalty Decomposition Algorithm for Multidisciplinary Design Optimization Problems with Complementarity Constraints," *Journal of Mechanical Design*, Vol. 132, No. 4, pp. 041005-041005, 2010.

Table 6.12 Benchmark Input Distribution

RV	Type	Mean	STDEV
X_1		3.0	0.06
X_2		0.7	0.015
X_3		22	0.4
X_4	Normal	7.8	0.15
X_5		7.8	0.15
X_6		3.4	0.06
X_7		5.3	0.1

Table 6.13 Input Data Set 1

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	2.9162	0.7063	21.3495	7.5622	7.8251	3.3778	5.2981
2	2.9759	0.6929	21.9324	7.5080	7.6454	3.4206	5.3161
3	2.9777	0.7103	22.1643	7.9300	7.7038	3.3332	5.2435
4	3.0407	0.7113	22.0967	7.8282	7.9806	3.3511	5.3331
5	3.0067	0.6830	21.6775	7.5644	7.8135	3.4148	5.1677
6	2.9946	0.7026	22.0426	7.8667	7.8068	3.4065	5.2387
7	3.0908	0.6932	21.4084	7.4867	7.7151	3.4452	5.3234
8	3.0106	0.7196	22.6049	7.7064	7.9453	3.3940	5.2609
9	2.9834	0.7029	21.7892	7.9942	8.1287	3.4585	5.4177
10	2.9421	0.7023	21.9633	7.6974	7.8279	3.4162	5.3472
11	2.9138	0.7151	21.8453	7.7979	7.8020	3.3530	5.4161
12	2.9821	0.7190	22.2193	7.8859	7.7373	3.4331	5.1409
13	2.9055	0.6842	21.8602	7.9514	7.7184	3.3354	5.2672
14	2.9826	0.7104	22.4780	7.8078	7.5923	3.4398	5.3350
15	3.0099	0.6604	21.6022	7.9582	7.6216	3.3964	5.2535
16	2.9758	0.6898	21.5932	7.8407	7.8276		
17	3.0347	0.6917	22.3571	7.8958	7.8820		
18	2.9424	0.6839	21.5220	7.9844	8.0628		
19	3.0631	0.7126	22.0974	7.8843	7.9937		
20	2.9442	0.7057	22.2700	7.8570	8.0595		
21			22.4971	7.6956	7.6727		
22			21.3538	8.0374	7.8898		
23			21.6968	8.0829	7.6195		
24			22.1801	7.7731	7.9242		
25			22.2690	7.8197	7.6111		
26			22.4848	7.9012	7.5856		
27			21.6872	7.8424	8.0436		
28			20.8836	7.7129	7.7266		
29			22.4629	7.6647	7.6988		
30			22.1971	7.8187	7.7650		
Mean	2.9846 (99.5%)	0.6999 (100.0%)	21.9529 (99.8%)	7.8119 (100.2%)	7.8075 (100.1%)	3.3984 (100.0%)	5.2906 (99.8%)
STDEV	0.0494 (82.3%)	0.0149 (99.3%)	0.4169 (104.2%)	0.1517 (101.1%)	0.1553 (103.5%)	0.0404 (67.3%)	0.0783 (78.3%)

Table 6.14 Input Data Set 2

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	2.9203	0.7025	22.0096	7.8025	7.9095	3.3053	5.2558
2	3.0313	0.6938	22.3822	7.6630	7.6668	3.4074	5.2584
3	2.9503	0.6837	21.8453	7.9339	7.6510	3.3389	5.3148
4	3.0280	0.7033	21.9388	7.8259	7.8232	3.2893	5.2958
5	3.0461	0.6740	22.2043	7.8492	7.9829	3.3938	5.3992
6	3.0064	0.6925	22.4586	7.5509	7.9383	3.4217	5.4540
7	3.0507	0.6974	21.6787	7.6625	7.7680	3.4144	5.2811
8	3.0657	0.7174	21.7482	7.7282	7.5703	3.4106	5.3023
9	3.0117	0.7123	21.8987	7.8715	7.7142	3.2963	5.2589
10	3.0412	0.7021	21.8233	7.7388	7.8396	3.3547	5.2006
11	2.9139	0.7327	22.1498	7.8177	7.6579	3.4768	5.3028
12	2.9632	0.7284	21.6523	7.8034	7.7336	3.4475	5.2508
13	2.9368	0.6807	21.0246	7.8330	7.7705	3.4571	5.2411
14	3.0484	0.6971	22.4264	7.5537	7.4158	3.4030	5.1304
15	3.0993	0.6891	22.1151	7.9493	7.9377	3.4129	5.2804
16	3.0113	0.7011	22.0429	7.8429	7.7072		
17	3.0981	0.6946	21.9851	7.8243	7.9347		
18	3.0369	0.6926	21.5144	7.6097	7.7409		
19	2.9672	0.6931	22.1310	7.5276	7.7911		
20	2.9641	0.6685	22.0537	7.5023	7.9542		
21			21.9131	7.8946	8.1346		
22			22.5774	7.7433	7.9788		
23			22.6514	7.6291	7.5750		
24			21.7253	7.5455	7.8447		
25			21.7726	7.6574	7.9804		
26			22.4377	7.5837	7.6524		
27			21.2991	7.7287	7.7889		
28			21.8885	7.8490	7.6391		
29			21.8662	7.7392	7.5552		
30			21.8235	7.7104	7.8646		
Mean	3.0095 (100.3%)	0.6978 (99.7%)	21.9679 (99.9%)	7.7324 (99.1%)	7.7840 (99.8%)	3.3887 (99.7%)	5.2817 (99.7%)
STDEV	0.0553 (92.2%)	0.0161 (107.3%)	0.3614 (90.4%)	0.1281 (85.4%)	0.1605 (107.0%)	0.0587 (97.8%)	0.0753 (75.3%)

Similar to the Iowa 2-D example in Section 6.5.1, the DDO optimum is obtained first using the I-RBDO code. Then conventional RBDO is performed using the most likely distribution types chosen at the DDO optimum, the parameters shown at the bottom rows of Table 6.13 and Table 6.14, and the I-RBDO code. The most likely distributions for input data set 1 are lognormal, extreme, Weibull, Weibull, Gumbel, Weibull, and normal for X_1 to X_7, respectively. The distributions for input data set 2 are normal, lognormal, normal, normal, normal, Weibull, and lornormal for X_1 to X_7, respectively. After a conventional RBDO optimum is obtained, candidate distribution types for C-RBDO are chosen. Here, a different method than the one for the Iowa 2-D example (20 most likely types) is required because there are 823,543 (=7^7) possible combinations if seven marginal distribution types are considered for each random variable. The 20 most likely types among the 823,543 combinations have the same probability. That means they are essentially the same distribution type. Therefore, it may not be appropriate to consider them as they cannot provide enough flexibility of input distribution types to C-RBDO.

To determine a reasonable number of candidate distribution types, the probabilities of 823,543 combinations are calculated first using the parameters in Table 6.13 and Table 6.14 at the conventional RBDO optimums. Then the probabilities when X_i follows a certain marginal distribution type, regardless of other random variables, are summed. For example, among the 823,543 probabilities, the probabilities when X_1 follows normal distribution are summed. Hence, there are total 49 cumulative probabilities (seven random variables by seven marginal distribution types), as shown in Table 6.15 and Table 6.16. The distribution types for each random variable could be categorized into three groups according to their cumulative probabilities. For example, marginal distribution types for X_7 in input data set 1 are categorized into three groups. The first group contains normal, lognormal, and gamma, which have a cumulative probability around 24%. The second group of Weibull and extreme has approximately

10% cumulative probability. The last group has Gumbel and extreme type-II, and they have small cumulative probability.

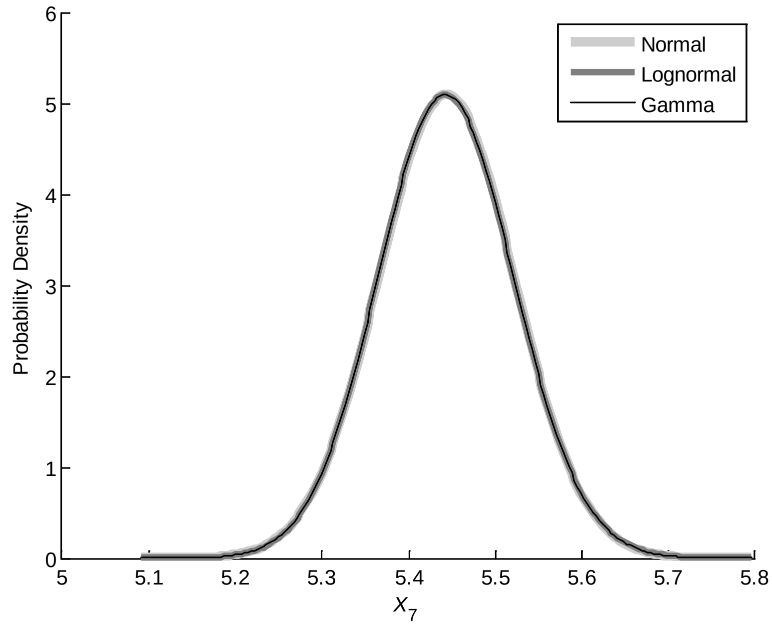
Table 6.15 Cumulative Probability of Input Distribution Type for Input Data Set 1

Type	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Normal	23.4%	7.8%	14.1%	**16.3%**	13.2%	**16.8%**	**24.5%**
Lognorm.	**24.1%**	6.6%	11.6%	14.1%	**14.6%**	16.4%	24.1%
Weibull	0.6%	38.5%	**32.9%**	**29.7%**	0.1%	**23.0%**	**12.1%**
Gumbel	**14.6%**	0.0%	0.0%	0.0%	**30.2%**	2.6%	2.6%
Gamma	23.9%	7.0%	12.4%	14.8%	14.2%	16.6%	24.2%
Extreme	0.4%	**40.1%**	29.0%	25.1%	0.0%	22.4%	10.6%
Ext. II	12.9%	0.0%	0.0%	0.0%	27.6%	2.2%	1.9%

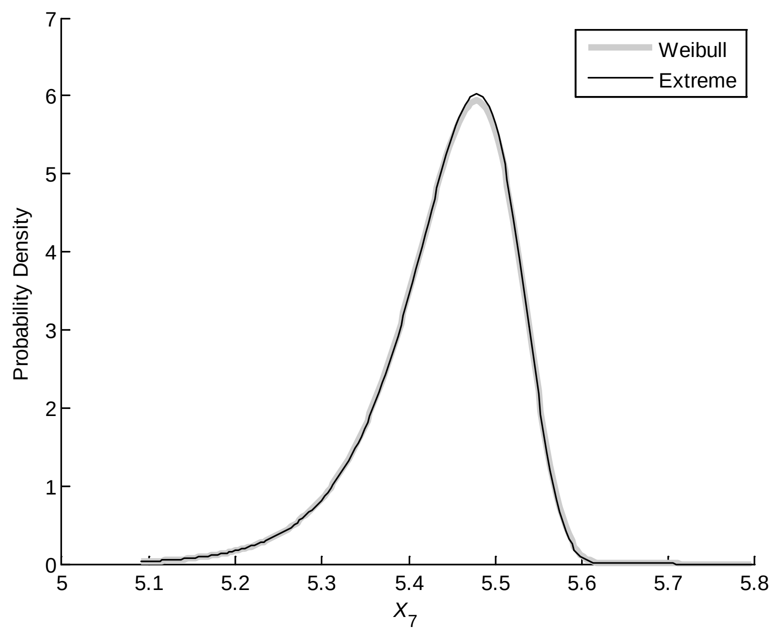
Table 6.16 Cumulative Probability of Input Distribution Type for Input Data Set 2

Type	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Normal	**22.4%**	23.5%	**31.8%**	**22.6%**	**32.3%**	**15.3%**	27.5%
Lognorm.	21.8%	**25.6%**	28.6%	21.4%	31.4%	14.5%	**28.6%**
Weibull	**15.6%**	0.2%	6.3%	**18.0%**	2.6%	**26.9%**	0.3%
Gumbel	2.4%	**14.7%**	0.0%	0.7%	0.3%	1.3%	**8.5%**
Gamma	22.0%	24.9%	29.7%	21.8%	31.8%	14.7%	28.3%
Extreme	13.9%	0.1%	3.7%	15.1%	1.5%	26.4%	0.2%
Ext. II	1.9%	11.0%	0.0%	0.4%	0.1%	0.9%	6.7%

Marginal PDFs in each group for X_7 in set 1 are graphically shown in Figure 6.8. It is evident that the PDFs in the same group have very close shape. Hence, it can be seen that the similar cumulative probability indicates similar shapes of PDF in each group. Therefore, it is not necessary to consider all marginal distribution types in each group for the candidate distribution types. It would be better to consider one distribution type in each group because it can provide the same flexibility of distribution type to the C-RBDO process using a smaller number of candidate distribution types. In this example, three groups are determined according to the cumulative probabilities of each marginal distribution type. Then, the distribution that shows the largest cumulative probability in each group is selected as the representative of the group. If all the cumulative probabilities in a group are larger than 28.6%, which is twice the average cumulative probability (200% / 7), the group is treated as the dominant group, and only one marginal distribution in the dominant group is used for the corresponding random variable. For example, X_2 in the input data set 1 has only extreme distribution as its marginal candidate because the group of Weibull and extreme distributions has cumulative probability larger than 28.6%. So the group is treated as the dominant group, and only extreme distribution is used as the marginal candidate. If there is no group whose cumulative probabilities are larger than 28.6%, two groups are selected, and one distribution that has the largest cumulative probability in each group is selected. The group with lowest average cumulative probability is not considered in this example because it has small probability and therefore a small effect on the C-RBDO process. Following this procedure, 32 ($=2^5$) candidate distribution types are selected for both input data sets. The selected marginal distribution types are marked with bold font in Table 6.15 and Table 6.16.

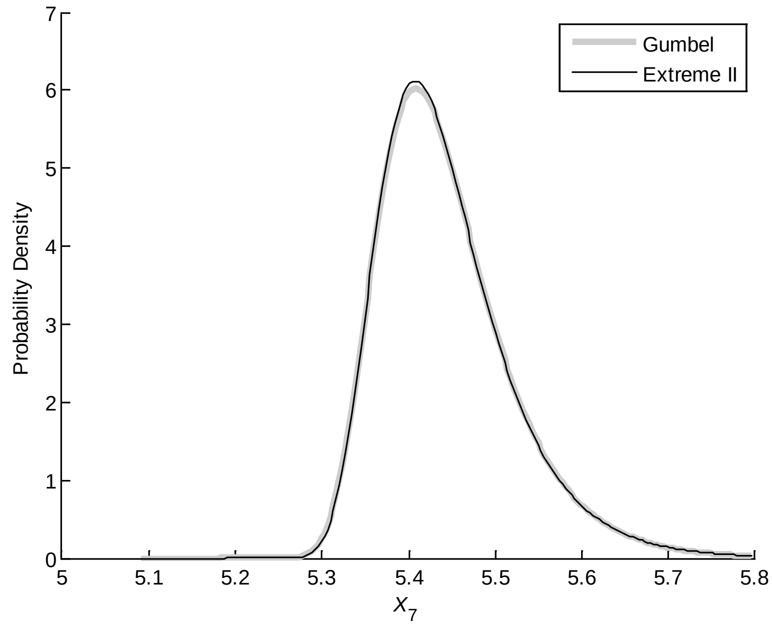


(a) PDFs of the first group



(b) PDFs of the second group

Figure 6.8 PDFs of Groups of Marginal Distribution for X_7 in Input Data Set 1



(c) PDFs of the third group

Figure 6.8 Continued

Using two input data sets in Table 6.13 and Table 6.14, and the selected 32 candidate distribution types, C-RBDO is performed. Both $NMCS_{\Psi}$ and $NMCS_M$ are set to 50,000. As explained earlier, DDO and conventional RBDO optimums are obtained in advance of C-RBDO. Then the C-RBDO is launched at the conventional RBDO optimum. Various target confidence level of 90%, 95%, and 97.5% are used to understand how high a confidence level is appropriate for engineering applications. Target reliability output is set to 2.275% as shown in Eq. (6.29). The DDO, conventional RBDO, and C-RBDO optimum designs are shown in Table 6.17. All C-RBDO optimum designs are found under 11 design iterations and 44 confidence level evaluations. d_1 hits the lower bound at the DDO optimum design, and it is not increased in conventional RBDO and C-RBDO. On the contrary, d_2 to d_7 show the lowest value at DDO optimum and tend to increase as more conservativeness is required. The reason is that constraints G_1 to G_6 are for strength (stress and deflection) of the system. More material will increase the strength, so d_2 to d_7 tend to increase as more conservative design is required. The percentage under each cost value shows an increment ratio from the cost at the DDO optimum design. From DDO to conventional RBDO optimums, 12.5% to 13.9% of the cost is increased to obtain a more reliable design considering input randomness in the data. Then, 18.4% to 22.1% more cost is necessary to secure a 97.5% confidence level of reliability output. Hence, 5.9% to 8.2% more cost from conventional RBDO is required to secure appropriate safety considering the uncertainty in the input probabilistic model due to the limited data.

In Table 6.18, confidence levels at conventional RBDO and C-RBDO optimum designs are summarized. Confidence levels of active constraints are shown, and the other constraints of G_3, G_4, G_7, G_8, and G_9 have 100% confidence level of the reliability output. At the conventional RBDO optimum designs, very low confidence levels are found (see underlined values). The main reason is that the conventional RBDO considers only one input distribution type. Sometimes different distribution types could result in very different reliability output even though the same input distribution parameters are given. For this reason, the conventional RBDO could not secure even a moderate confidence level. At C-RBDO optimum designs, the target confidence level is met. That is, C-RBDO is successfully converged to the correct optimum designs.

To check the conservativeness of the optimum designs, reliability analysis is performed at the optimum designs using the benchmark distribution shown in Table 6.12. The result of the reliability analysis is summarized in Table 6.19 for active constraints. Other constraints show negligible reliability output. At the conventional RBDO optimum designs, a 2.275% target reliability output is not satisfied. As explained earlier, small sample standard deviations are estimated in both data sets, and thus input randomness is underestimated. Hence, reliability output values larger than 2.275% are estimated at the conventional RBDO optimum deigns when the benchmark distribution is applied. In the case of input data set 1, all C-RBDO optimum designs satisfy 2.275% target reliability output. Even though input randomness is underestimated, safe designs are successfully found considering the uncertainty in the input probabilistic model due to limited data. On the other hand, in set 2, the C-RBDO optimum with 90% target confidence level does not satisfy 2.275% in G_6. This indicates that the target confidence level should be set reasonably high so that the underestimated input randomness could be compensated by considering the uncertainty in the input probabilistic model. However, 2.6872% at 90% C-RBDO optimum is very close to 2.275% target reliability compared to 6.8555% at conventional RBDO optimum design in G_6. Hence, by considering uncertainty in the input probabilistic model, appropriate conservativeness is almost reached. C-RBDO optimum designs for set 2 with 95% and 97.5% target confidence levels satisfy target reliability output in all constraints.

In this example, it is shown that the C-RBDO formulation successfully finds its optimum design in a multi-dimensional engineering problem. It is also well shown that the limited data may cause small sample standard deviations so that input randomness could be underestimated. Then, conventional RBDO cannot find safe design, and even C-RBDO with a certain target confidence level may not secure target reliability output. Hence, in practical engineering problems, a reasonably high target confidence level, such as 97.5%, is recommended for C-RBDO. In practical situations, engineers will not have the benefit of the benchmark distribution (i.e., exact input distribution). Hence, it is impossible to estimate how much conservativeness the obtained optimum design yields. Therefore, a reasonably high target confidence level is necessary in practical engineering problems when only limited data are provided.

Table 6.17 Optimum Designs of Speed-Reducer 7-D Example

Case		Design iter.	Conf. eval.	Optimum design							Cost
				d_1	d_2	d_3	d_4	d_5	d_6	d_7	
DDO		-	-	4	0.6771	14.7222	6.9282	7.7159	3.3521	5.2872	2676
Set 1	RBDO	-	-	4	0.7051	15.3831	7.4649	8.1872	3.4520	5.4437	3047 (13.9%)
	C-RBDO (CL^{Tar} = 90%)	11	44	3.9998	0.7140	15.4839	7.5992	8.3907	3.4848	5.5327	3180 (18.8%)
	C-RBDO (CL^{Tar} = 95%)	5	7	4	0.7175	15.4916	7.6438	8.4571	3.4968	5.5643	3225 (20.5%)
	C-RBDO (CL^{Tar} = 97.5%)	6	9	3.9971	0.7211	15.5014	7.6938	8.5131	3.5102	5.5922	3267 (22.1%)
Set 2	RBDO	-	-	4	0.7043	15.1619	7.4452	8.2412	3.4978	5.4358	3011 (12.5%)
	C-RBDO (CL^{Tar} = 90%)	4	6	4	0.7098	15.2543	7.6341	8.3855	3.5434	5.4799	3099 (15.8%)
	C-RBDO (CL^{Tar} = 95%)	8	25	4	0.7134	15.2516	7.8013	8.4371	3.5636	5.5017	3138 (17.3%)
	C-RBDO (CL^{Tar} = 97.5%)	8	21	4	0.7168	15.2458	7.7626	8.4795	3.5826	5.5173	3168 (18.4%)

Table 6.18 Confidence Level at Optimum Designs

Case		Confidence level					
		G_1	G_2	G_5	G_6	G_{10}	G_{11}
Set 1	RBDO	42.71%	41.27%	62.82%	38.47%	61.43%	33.35%
	C-RBDO (CL^{Tar} = 90%)	89.63%	90.31%	90.98%	89.32%	92.05%	90.32%
	C-RBDO (CL^{Tar} = 95%)	95.00%	94.98%	95.15%	95.00%	94.92%	95.39%
	C-RBDO (CL^{Tar} = 97.5%)	97.50%	97.60%	97.42%	97.58%	97.65%	97.46%
Set 2	RBDO	58.76%	50.59%	61.83%	60.65%	29.49%	40.77%
	C-RBDO (CL^{Tar} = 90%)	89.84%	89.89%	90.11%	89.77%	89.86%	89.69%
	C-RBDO (CL^{Tar} = 95%)	94.94%	95.02%	94.92%	94.93%	99.76%	94.91%
	C-RBDO (CL^{Tar} = 97.5%)	97.37%	97.39%	97.53%	97.06%	97.98%	97.39%

Table 6.19 Reliability Analysis Result at Optimum Designs with Benchmark Distribution

Case		Probability of failure (%)					
		G_1	G_2	G_5	G_6	G_{10}	G_{11}
Set 1	RBDO	0.9475	0.8314	4.7818	5.8669	1.3552	5.4166
	C-RBDO (CL^{Tar} = 90%)	0.1581	0.1598	1.3475	0.7083	0.3474	1.4732
	C-RBDO (CL^{Tar} = 95%)	0.0769	0.0937	0.7914	0.2872	0.2202	0.9568
	C-RBDO (CL^{Tar} = 97.5%)	0.0416	0.0532	0.4218	0.1210	0.1243	0.6584
Set 2	RBDO	2.1393	2.6235	0.7542	6.8555	4.3886	2.6138
	C-RBDO (CL^{Tar} = 90%)	0.7322	0.9630	0.0730	2.6872	0.8441	0.7051
	C-RBDO (CL^{Tar} = 95%)	0.4131	0.6557	0.0213	1.6117	0.0730	0.4592
	C-RBDO (CL^{Tar} = 97.5%)	0.2407	0.4637	0.0059	1.0739	0.2619	0.3092

CHAPTER 7

CONCLUSIONS AND FUTURE RECOMMENDATIONS

In this chapter, the conclusions of this study and future recommendations are presented. Section 7.1 presents conclusions regarding the developed variable screening method and the confidence-based method for reliability-based design optimization (RBDO). In Section 7.2, possible improvements are recommended for the future.

7.1 Conclusions

A new efficient and effective variable screening method for RBDO is proposed in this study. For the proposed methods, the output variance is used as a measure that can identify important design variables. Thus, a partial output variance based on the univariate dimension reduction method (DRM) is proposed to approximate the output variance efficiently and to identify the design variables that affect output variance more significantly than others. The univariate DRM and partial output variance only require multiple 1-D surrogate models, which is much more efficient than the full-dimensional surrogate models. Hence, the proposed method has great merit in efficiency as well as effectiveness. To reduce computational time and maintain a user-specified statistical error level, hypothesis testing is used in the variable screening process. Also, a required minimum number of samples for calculating the correct output variance is proposed using the user-specified error level. In addition, the quadratic interpolation method is tailored to be applied for efficient partial output variance calculation.

Two analytical examples and a 44-D industrial example are used to verify the performance of the proposed variable screening method. Through the analytical examples, it is shown that at least quadratic approximation is required for the 1-D surrogate model and that partial output variance is a good measure that successfully identifies important variables. In the industrial example, 14 design variables out of 44 are selected by considering the output variances of 11 constraints. For comparison, another

14 design variables selected based on experience are used. In addition, 18 design variables are selected by adding four design variables, which affect the objective function significantly while not affecting the output variances much, to the 14 design variables previously selected with the proposed method. The selection based on experience shows a 7.6% reduced cost value, whereas the target probability of failure is violated by 77%. However, selection by the proposed method shows only a 12.3% disagreement of target value and a 3.6% reduced cost value. Moreover, the selection of 18 design variables shows 11.7% target disagreement as well as 9.4% reduced cost value. Therefore, the performance of the proposed variable screening method is verified.

In this study, the confidence level of the reliability output, which is the probability of failure, is estimated to quantify the uncertainty in the input probability model due to limited input data. The probability of the reliability output is decomposed into successive conditional probabilities of input distribution type and parameters. Then, the conditional probabilities are obtained using the Bayesian approach under reasonable assumptions. As the probability at a target reliability output (target probability of failure) is the confidence level for the target value, the confidence level of the reliability output is directly available and can be used for new probabilistic constraints in the confidence-based RBDO (C-RBDO) formulation. Consequently, the C-RBDO formulation can provide an optimum design that satisfies the target confidence level at the target reliability output. This is beneficial because appropriate conservativeness, which is necessary when only limited data are available, is included in the optimum design. The confidence levels at the selected design point, which is the conventional RBDO optimum design using a benchmark distribution, is computed for different numbers of data. As the results show, a greater confidence level is required when less data is provided, so it is necessary to adopt the developed C-RBDO.

For efficient and accurate evaluation of confidence-based RBDO, the analytical design sensitivity of the confidence level is derived. Compared to the finite difference method (FDM) sensitivity, the derived sensitivity shows accurate results in the numerical example. At the same time, it uses only 0.25% of the Monte Carlo simulation (MCS) samples that are used by the FDM, so it is efficient as well. To improve the efficiency of C-RBDO, two-step reliability analysis and reusable MCS are suggested. Using the developed design sensitivity and efficiency improvement methods, C-RBDO is performed for different numbers of data. Smooth convergence of the optimization process again verifies that the provided design sensitivity is accurate. At the obtained optimum designs, the target confidence level at the target reliability output is met for active constraints. Using the benchmark distribution, conventional RBDO is performed at the C-RBDO optimums. The result shows that the optimum design is more conservative as a smaller amount of data is provided. In addition, as more data are provided, the optimum design approaches the conventional RBDO optimum with the benchmark input distribution. Therefore, it can be seen that C-RBDO produces an appropriately conservative design according to the given input data. In addition, by using a 7-D engineering example, it is confirmed that C-RBDO finds optimum designs with reasonable conservativeness, even for high-dimensional problems.

7.2 Future Recommendations

The developed variable screening method selects important variables using the partial output variances, and hypothesis testing determines which variable has larger partial output variance than other variables. In the developed method, one variable will be selected by hypothesis testing even though two variables have very similar partial output variances. In the future, partial output skewness could be studied as a measure to estimate the importance of variables. If two random variables have very close partial output variances, the effect of partial output skewness on the probability of failure could be significant. Hence, the variable screening could calculate the partial output variances first and then decide which variables are important. If the hypothesis testing indicates that some partial output variances are statistically similar, the partial output skewness could decide which variable affects the probability of failure more significantly. Then the variable screening with two measures of partial output variances and skewness could be more effective than the present method.

The methodology of the C-RBDO has been developed, and its performance is verified as well in this study. To make the method more widely used, more tests could be performed to find potential applications. Adopting surrogate model methods could be a great leap toward improving the practicality of C-RBDO. Although C-RBDO requires many reliability analyses at a design, the analyses can share one surrogate model. Therefore, it would not require several surrogate models at a design. Finally, the consistency of design sensitivity for C-RBDO could be improved. As the design sensitivity uses MCS method, the sensitivity depends on how MCS samples are distributed (locations of samples), especially when a small number of samples is used. Hence, the design sensitivity method could be further investigated to provide consistent sensitivity even with small number of MCS samples.

REFERENCES

Arora, J. S., *Introduction to Optimum Design, 2nd Ed.*, Elsevier/Academic Press, San Diego, CA, 2004.

Aughenbaugh, J. M., and Paredis, C. J. J., "The Value of using Imprecise Probabilities in Engineering Design," *Journal of Mechanical Design*, Vol. 128, No. 4, pp. 969-979, 2006.

Bae, S., *Variable Screening Method using Statistical Sensitivity Analysis in RBDO*, Master's thesis, University of Iowa, 2012.

Barton, R. R., "Metamodeling: a state of the art review," *WSC '94: Proceedings of the 26th Conference on Winter Simulation*, Orlando, FL, 1994.

Breitung, K., "Asymptotic Approximations for Multinormal Integrals," *Journal of Engineering Mechanics*, Vol. 110, No. 3, pp. 357-366, 1984.

Chen, W., Jin, R., and Sudjianto, A., "Analytical Variance-Based Global Sensitivity Analysis in Simulation-Based Design under Uncertainty," *Journal of Mechanical Design*, Vol. 127, No. 5, pp. 875-886, 2005.

Chick, S. E., "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty," *Operations Research*, Vol. 49, No. 5, pp. pp. 744-758, 2001.

Cho, H., Choi, K. K., Lee, I., "Confidence Level Estimation and Design Sensitivity Analysis for Confidence-based RBDO," *38th ASME Design Automation Conference*, Chicago, IL, August 12-15, 2012.

Choi, K. K., Tu, J., and Park, Y., "Extensions of Design Potential Concept for Reliability-Based Design Optimization to Nonsmooth and Extreme Cases," *Structural and Multidisciplinary Optimization*, Vol. 22, No. 5, pp. 335-350, 2001.

Choi, K. K., and Kim, N. H., *Structural Sensitivity Analysis and Optimization 1: Linear Systems*, Springer, New York, NY, 2005a.

Choi, K. K., and Kim, N. H., *Structural Sensitivity Analysis and Optimization 2: Nonlinear Systems and Applications*, Springer, New York, NY, 2005b.

Choi, K. K., Gaul, N., Song, H., Cho, H., Lee, I., Zhao, L., Noh, Y., Lamb, D., Sheng, J., Gorsich, D., Yang, R., Shi, L., Epureanu, B., and Hong, S., "Iowa Reliability-Based Design Optimization (I-RBDO) Code and Its Applications," *ARC Conference*, Ann Arbor, MI, May 21-22, 2012.

Craig, K. J., Stander, N., Dooge, D. A., and Varadappa, S., "Automotive Crashworthiness Design using Response Surface-Based Variable Screening and Optimization," *Engineering Computations*, Vol. 22, No. 1, pp. 38-61, 2005.

Cressie, N. A. C., *Statistics for Spatial Data*, John Wiley & Sons, New York, NY, 1991.

Ditlevsen, O., and Madsen, H. O., *Structural Reliability Methods*, John Wiley & Sons, Chichester, UK, 1996.

Dixon, L. C. W., and Szegö, G. P., *Towards Global Optimisation 2*, North-Holland Publishing Company, Amsterdam, Netherlands, 1978.

Duarte Silva, A. P., "Efficient Variable Screening for Multivariate Analysis," *Journal of Multivariate Analysis*, Vol. 76, No. 1, pp. 35-62, 2001.

Eichler, J., and Wille, U., "Variable-Screening Model for the Quasimolecular Treatment of Ion-Atom Collisions," *Physical Review A*, Vol. 11, No. 6, pp. 1973-1982, 1975.

Elishakoff, I., *Safety Factors and Reliability: Friends Or Foes?*, Kluwer Academic Publishers, Dordrecht, Netherlands, 2004.

Forrester, A. I. J., Sóbester, A., and Keane, A. J., *Engineering Design Via Surrogate Modelling : A Practical Guide*, John Wiley & Sons, Chichester, UK, 2008.

Forrester, A. I. J., and Keane, A. J., "Recent Advances in Surrogate-Based Optimization," *Progress in Aerospace Sciences*, Vol. 45, No. 1–3, pp. 50-79, 2009.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., *Bayesian Data Analysis, Second Edition*, Chapman & Hall/CRC, Boca Raton, FL, 2004.

Genest, C., and Favre, A., "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.

Golinski, J., "Optimal Synthesis Problems Solved by Means of Nonlinear Programming and Random Methods," *Journal of Mechanisms*, Vol. 5, No. 3, pp. 287-309, 1970.

Gumbert, C. R., Hou, G. J.-W., and Newman, P. A., "Reliability Assessment of a Robust Design under Uncertainty for a 3-D Flexible Wing," *16th AIAA Coputational Fluid Dynamics Conference*, Orlando, FL, June 23-26, 2003.

Gunawan, S., and Papalambros, P. Y., "A Bayesian Approach to Reliability-Based Optimization with Incomplete Information," *Journal of Mechanical Design*, Vol. 128, No. 4, pp. 909-918, 2006.

Guyon, I., and Elisseeff, A., "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, 2003.

Haldar, A., and Mahadevan, S., *Probability, Reliability and Statistical Methods in Engineering Design*, John Wiley & Sons, New York, NY, 2000.

Hasofer, A. M., and Lind, N. C., "Exact and Invariant Second-Moment Code Format," *Journal of the Engineering Mechanics Division, ASCE*, Vol. 100, No. 1, pp. 111-121, 1974.

Hohenbichler, M., and Rackwitz, R., "Improvement of Second-Order Reliability Estimates by Importance Sampling," *Journal of Engineering Mechanics*, Vol. 114, No. 12, pp. 2195-2199, 1988.

Hou, G. J.-W., "A Most Probable Point-Based Method for Reliability Analysis, Sensitivity Analysis, and Design Optimization," *NASA*, NASA/CR-2004-213002, 2004.

Izenman, A. J., *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer, New York, NY, 2008.

Jin, R., Chen, W., and Simpson, T. W., "Comparative Studies of Metamodelling Techniques under Multiple Modelling Criteria," *Structural and Multidisciplinary Optimization*, Vol. 23, No. 1, pp. 1-13, 2001.

Lee, I., Choi, K. K., and Gorsich, D., "System reliability-based design optimization using the MPP-based dimension reduction method," *Structural and Multidisciplinary Optimization*, Vol. 41, No. 6, pp. 823-839, 2010a.

Lee, I., Choi, K. K., and Zhao, L., "Sampling-Based RBDO using the Stochastic Sensitivity Analysis and Dynamic Kriging Method," *Structural and Multidisciplinary Optimization*, Vol. 44, No. 3, pp. 299-317, 2011a.

Lee, I., Choi, K. K., and Gorsich, D., "Sensitivity Analyses of FORM-Based and DRM-Based Performance Measure Approach (PMA) for Reliability-Based Design Optimization (RBDO)," *International Journal for Numerical Methods in Engineering*, Vol. 82, No. 1, pp. 26-46, 2010b.

Lee, I., Choi, K. K., Noh, Y., Zhao, L., and Gorsich, D., "Sampling-Based Stochastic Sensitivity Analysis using Score Functions for RBDO Problems with Correlated Random Variables," *Journal of Mechanical Design*, Vol. 133, No. 2, pp. 021003, 2011b.

Lee, J., "A Novel Three-Phase Trajectory Informed Search Methodology for Global Optimization," *Journal of Global Optimization*, Vol. 38, No. 1, pp. 61-77, 2007.

Mack, Y., Goel, T., Shyy, W., and Haftka, R., "Surrogate Model-Based Optimization Framework: A Case Study in Aerospace Design," *Studies in Computational Intelligence*, Vol. 51, pp. 323-342, 2007.

Madsen, H. O., Krenk, S., and Lind, N. C., *Methods of Structural Safety*, Prentice-Hall, Englewood Cliffs, NJ, 1986.

McKay, M. D., Morrison, J. D., and Upton, S. C., "Evaluating Prediction Uncertainty in Simulation Models," *Computer Physics Communications*, Vol. 117, No. 1–2, pp. 44-51, 1999.

Nelsen, R. B., *An Introduction to Copulas*, Springer, New York, NY, 2006.

Noh, Y., Choi, K. K., and Lee, I., "Identification of Marginal and Joint CDFs Using Bayesian Method for RBDO," *Structural and Multidisciplinary Optimization*, Vol. 40, No. 1, pp. 35-51, 2010.

Noh, Y., Choi, K. K., Lee, I., Gorsich, D., and Lamb, D., "Reliability-based design optimization with confidence level under input model uncertainty due to limited test data," *Structural and Multidisciplinary Optimization*, Vol. 43, No. 4, pp. 443-458, 2011a.

Noh, Y., Choi, K. K., and Du, L., "Reliability-Based Design Optimization of Problems with Correlated Input Variables using a Gaussian Copula," *Structural and Multidisciplinary Optimization*, Vol. 38, No. 1, pp. 1-16, 2009.

Noh, Y., Choi, K. K., Lee, I., Gorsich, D., and Lamb, D., "Reliability-Based Design Optimization with Confidence Level for Non-Gaussian Distributions using Bootstrap Method," *Journal of Mechanical Design*, Vol. 133, No. 9, pp. 091001, 2011b.

Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Kevin Tucker, P., "Surrogate-Based Analysis and Optimization," *Progress in Aerospace Sciences*, Vol. 41, No. 1, pp. 1-28, 2005.

Rahman, S., and Wei, D., "Design Sensitivity and Reliability-Based Structural Optimization by Univariate Decomposition," *Structural and Multidisciplinary Optimization*, Vol. 35, No. 3, pp. 245-261, 2008.

Rahman, S., and Wei, D., "A Univariate Approximation at most Probable Point for Higher-Order Reliability Analysis," *International Journal of Solids and Structures*, Vol. 43, No. 9, pp. 2820-2839, 2006.

Rahman, S., and Xu, H., "A Univariate Dimension-Reduction Method for Multi-Dimensional Integration in Stochastic Mechanics," *Probabilistic Engineering Mechanics*, Vol. 19, No. 4, pp. 393-408, 2004.

Rahman, S., "Stochastic Sensitivity Analysis by Dimensional Decomposition and Score Functions," *Probabilistic Engineering Mechanics*, Vol. 24, No. 3, pp. 278-287, 2009.

Rao, S. S., *Applied Numerical Methods for Engineers and Scientists*, Prentice Hall, Upper Saddle River, NJ, 2002.

Reedijk, C. I., *Sensitivity Analysis of Model Output: Performance of various and Global Sensitivity Measures on Reliability Problems*, Master’s thesis, Delft University of Technology, 2000.

Rosenblatt, M., "Remarks on a Multivariate Transformation," *The Annals of Mathematical Statistics*, Vol. 23, No. 3, pp. pp. 470-472, 1952.

Rosner, B., *Fundamentals of Biostatistics, 6th Ed.*, Thomson Brooks/Cole, Belmont, CA, 2006.

Rubinstein, R. Y., and Kroese, D. P., *Simulation and the Monte Carlo Method, 2nd Ed.*, John Wiley & Sons, Hoboken, NJ, 2008.

Rubinstein, R. Y., and Shapiro, A., *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, Wiley, New York, NY, 1993.

Simpson, T. W., Poplinski, J. D., Koch, P. N., and Allen, J. K., "Metamodels for Computer-Based Engineering Design: Survey and Recommendations," *Engineering with Computers*, Vol. 17, No. 2, pp. 129-150, 2001.

Sobol, I. M., "Global Sensitivity Indices for Nonlinear Mathematical Models and their Monte Carlo Estimates," *Mathematics and Computers in Simulation*, Vol. 55, No. 1–3, pp. 271-280, 2001.

Tu, J., Choi, K. K., and Park, Y. H., "Design Potential Method for Robust System Parameter Design," *AIAA Journal*, Vol. 39, No. 4, pp. 667-677, 2001.

Tu, J., Choi, K. K., and Park, Y. H., "A New Study on Reliability-Based Design Optimization," *Journal of Mechanical Design*, Vol. 121, No. 4, pp. 557-564, 1999.

Tucker, W. T., and Ferson, S., *Probability Bounds Analysis in Environmental Risk Assessment*, Applied Biomathematics, Setauket, NY, 2003.

Utkin, L., and Destercke, S., "Computing Expectations with Continuous p-Boxes: Univariate Case," *International Journal of Approximate Reasoning*, Vol. 50, No. 5, pp. 778-798, 2009.

Wang, G. G., and Shan, S., "Review of Metamodeling Techniques in Support of Engineering Design Optimization," *Journal of Mechanical Design*, Vol. 129, No. 4, pp. 370-380, 2007.

Wang, H., "Forward Regression for Ultra-High Dimensional Variable Screening," *Journal of the American Statistical Association*, Vol. 104, No. 488, pp. 1512-1524, 2009.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D., "Screening, Predicting, and Computer Experiments," *Technometrics*, Vol. 34, No. 1, pp. 15-25, 1992.

Wu, Y., "Computational Methods for Efficient Structural Reliability and Reliability Sensitivity Analysis," *AIAA Journal*, Vol. 32, No. 8, pp. 1717-1723, 1994.

Wu, Y., Millwater, H., and Cruse, T., "Advanced Probabilistic Structural Analysis Method for Implicit Performance Functions," *AIAA Journal*, Vol. 28, No. 9, pp. 1663-1669, 1990.

Wu, Y., and Mohanty, S., "Variable Screening and Ranking using Sampling-Based Sensitivity Measures," *Reliability Engineering & System Safety*, Vol. 91, No. 6, pp. 634-647, 2006.

Xu, H., and Rahman, S., "A Generalized Dimension-Reduction Method for Multidimensional Integration in Stochastic Mechanics," *International Journal for Numerical Methods in Engineering*, Vol. 61, No. 12, pp. 1992-2019, 2004.

Youn, B. D., Choi, K. K., and Du, L., "Adaptive Probability Analysis using an Enhanced Hybrid Mean Value Method," *Structural and Multidisciplinary Optimization*, Vol. 29, No. 2, pp. 134-148, 2005.

Youn, B. D., Choi, K. K., and Park, Y. H., "Hybrid Analysis Method for Reliability-Based Design Optimization," *Journal of Mechanical Design*, Vol. 125, pp. 221, 2003.

Youn, B. D., and Wang, P., "Bayesian reliability-based design optimization using eigenvector dimension reduction (EDR) method," *Structural and Multidisciplinary Optimization*, Vol. 36, No. 2, pp. 107-123, 2008.

Zhang, R., and Mahadevan, S., "Model Uncertainty and Bayesian Updating in Reliability-Based Inspection," *Structural Safety*, Vol. 22, No. 2, pp. 145-160, 2000.

Zhao, L., Choi, K. K., and Lee, I., "Metamodeling Method using Dynamic Kriging for Design Optimization," *AIAA Journal*, Vol. 49, No. 9, pp. 2034-2046, 2011.