8th World Congress on Structural and Multidisciplinary Optimization
June 1 - 5, 2009, Lisbon, Portugal

Reliability-Based Design Optimization with Confidence Level Using Copula under Input Uncertainty

Yoojeong Noh[1], K.K. Choi[2], Ikjin Lee[3], and David Gorsich[4]

[1,2,3]Department of Mechanical & Industrial Engineering
College of Engineering
The University of Iowa
Iowa City, IA 52242, U.S.A.
Email: liazhao@engineering.uiowa.edu
kkchoi@engineering.uiowa.edu
ilee@engineering.uiowa.edu

[4]US Army RDECOM/TARDEC
Warren, MI 48397-5000, U.S.A.
gorsichd@tacom.army.mil

1. Abstract

For obtaining correct reliability-based optimum design, an input model needs to be accurately estimated in identification of marginal and joint distribution types and quantification of their parameters. However, in most industrial applications, only limited data on input variables is available due to expensive experimental testing cost. The input model generated from the insufficient data will be inaccurate, which will lead to incorrect optimum design. In this paper, reliability-based design optimization (RBDO) with the confidence level is proposed to offset the inaccurate estimation of the input model due to limited data by using the upper bound of confidence interval of the standard deviation. Using the upper bound of confidence interval of the standard deviation, a confidence level on the input model can be assessed to obtain the confidence level of the output performance, i.e. a desired probability of failure, through the simulation-based design. For RBDO, the estimated input model with the associated confidence level is integrated with the most probable point (MPP)-based dimension reduction method (DRM), which improves accuracy over the first order reliability method (FORM). A mathematical example and a fatigue problem are used to illustrate how the input model with the confidence level yields a reliable optimum design by comparing it with the input model obtained using the estimated parameters.

2. Keywords: reliability-based design optimization, input model uncertainty (identification and quantification of marginal and joint CDFs), confidence level, MPP-based dimension reduction method.

3. Introduction

In RBDO, input model uncertainties consisting of input statistical uncertainty and input physical uncertainty exist due to limited available data and lack of information on input variables as shown in Fig. 1. Since in many engineering applications, input random variables such as fatigue material properties are correlated [1-4], the input statistical uncertainty needs to be modeled by identifying the joint cumulative distribution functions (CDFs) as well as marginal CDFs. The input physical uncertainty is modeled by quantifying the parameters of the identified marginal CDFs and joint CDFs such as mean, standard deviation, and correlation coefficient. However, in most industrial applications, only limited data on input variables is available due to expensive experimental testing cost, and the use of input model estimated from the insufficient data will yield incorrect reliability-based optimum design.

Research on the identification of marginal CDFs and a joint CDF for RBDO has been recently carried out [5,6], but the optimum design is obtained from the input model generated using given limited data without considering the input model uncertainty. In this paper, to offset the inexact identification and quantification of the input model, the confidence level on the input model is implemented for RBDO. In this method, instead of using the standard deviation of marginal CDFs estimated from given limited data, the upper bound of the confidence interval of the estimated standard deviation is used for the input model because a larger standard deviation provides a larger joint PDF contour for target reliability index β_t, which yields more reliable optimum design. Using the upper bound of the confidence interval on the standard deviation, a reliable optimum design with a specified confidence level could be obtained.

To check whether using the upper bound of the confidence interval of the estimated standard deviation provides the confidence level of the output performance, i.e., the desired target probability of failure, the joint PDF contour for β_t is

used to measure a confidence level of the input model. If the contour for β_t covers the true contour for the target reliability index β_t, then it will yield a reliable optimum design regardless of the location of the most probable point (MPP) on the contour. A mathematical example is tested to check whether the confidence level of the input model yields a desired confidence level of the output performance through a simulation study.

When the input variables are correlated with joint non-Gaussian distribution, the nonlinear transformation will convert the constraint functions to highly nonlinear functions in the standard Gaussian space where the MPP search is carried out. Thus, the FORM, which approximates constraint functions linearly, will not provide accurate results when the input variables are correlated with joint non-Gaussian distribution. Thus, RBDO using MPP-based DRM [7,8], which is more accurate than the FORM, is used in this paper. Using the input model obtained with the target confidence level and accurate inverse reliability analysis method (MPP-based DRM), RBDO results with the associate confidence level can be obtained as shown in Fig. 1. A mathematical example and a fatigue problem with correlated input variables show how the input model with the target confidence level yields reliable optimum designs, while the input model without the target confidence level yields unreliable optimum designs.

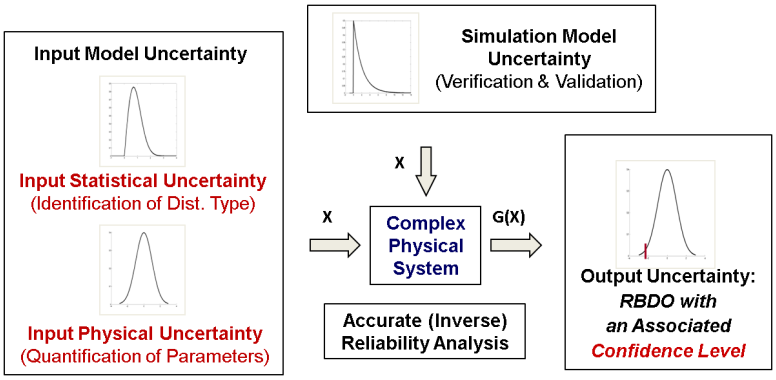


Figure 1. RBDO under Uncertainties

4. Estimation of Input Model

As shown in Fig 1, before carrying out the RBDO, the input model uncertainties such as marginal CDFs and joint CDF need to be identified and their parameters need to be quantified based on the experimental data. Section 4 presents how to estimate the input model with correlated input variables such as fatigue material properties. A copula is introduced to model the joint CDF of the correlated variables, and identification and quantification of the input model are explained.

4.1 Correlated Material Properties

In many structural RBDO problems, the input random variables such as the material properties and fatigue properties are correlated [1-4]. In fatigue problems, the strain-life relationship is expressed as

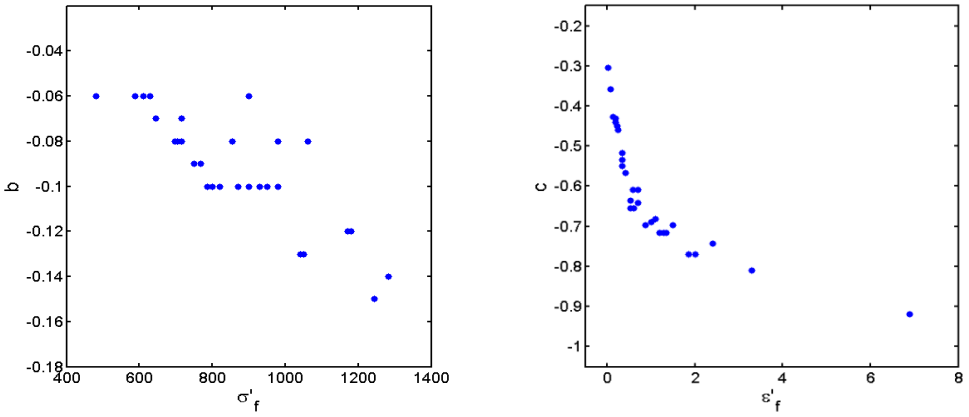
$$\frac{\Delta\varepsilon_f'}{2} = \frac{\sigma_f'}{E}\left(2N_f\right)^b + \varepsilon_f'\left(2N_f\right)^c \tag{1}$$

where $\Delta\varepsilon_f'$ is the strain amplitude, E is the Young's modulus, N_f is the fatigue life, σ_f' and b are the fatigue strength coefficient and exponent, and ε_f' and c are fatigue ductility coefficient and exponent, respectively.

Figure 2 shows 29 experimental data sets of the fatigue strength coefficient σ_f' versus fatigue strength exponent b; and fatigue ductility coefficient ε_f' versus fatigue ductility exponent c of the SAE 950X high strength low alloy [1]. As shown in Fig. 2 (a) and (b), these variables are highly negatively correlated where the correlation coefficients between σ_f' and b and between ε_f' and c are calculated as –0.828 and –0.906, respectively. It is known that σ_f' and ε_f' follow the lognormal distribution and b and c follow the Gaussian distribution [1]. Since these fatigue material properties are correlated, joint CDFs of the correlated variables need to be determined. However, commonly used multivariate distributions such as joint Gaussian distribution cannot be used because the marginal distribution types of two correlated variables are different, i.e., σ_f' has a lognormal distribution and b has a Gaussian distribution. For this case, a copula, which will be explained in Section 4.2, needs to be used.

For probabilistic life prediction, when variables are so closely related, another option is to let one variable be a function of the other variables, such as a linear fit. The problem with this approach is that the data in Fig. 2 cannot be properly fitted by linear functions. In addition, it is pointed out by Annis [2] that the result will be over corrected and thus underestimates the overall variability. According to Annis [2], the right way is "… to correctly modeling the joint

behavior to reduce greater than 700% error in the estimated of the standard deviation to about 1%" in his example.



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

Figure 2. Paired Data Obtained From SAE 950X [1]

4.2 Copula to Represent Joint CDFs of Correlated Input Variables

Copulas are multivariate distribution functions whose one-dimensional margins are uniform on the interval [0, 1]. If the random variables have marginal distributions, then by Sklar's theorem [9], there exists an *n*-dimensional copula *C* such that

$$F_{X_1,...,X_n}\left(x_1,...,x_n\right)=C\left(F_{X_1}\left(x_1\right),...,F_{X_n}\left(x_n\right)\middle|\boldsymbol{\theta}\right) \tag{2}$$

where $\boldsymbol{\theta}$ is the matrix of correlation parameters between $X_1,...,X_n$. If marginal distributions are all continuous, then *C* is unique. Conversely, if *C* is an *n*-dimensional copula and $F_{X_1}\left(x_1\right),...,F_{X_n}\left(x_n\right)$ are marginal CDFs, then the joint distribution is an *n*-dimensional function of marginal CDFs [9]. Some copula functions that are used in this paper and the domain of its parameter are shown in Table 1. The copula functions of other copulas are presented by Noh et al. [5,6].

Table 1. Copula Functions and Kendall's Tau

Copula	$C\left(u,v\middle\|\theta\right)$	$\theta\in\Omega^{\theta}$	$\tau=g\left(\theta\right)$	$\tau\in\Omega^{\tau}$
Clayton	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$	$\left(0,\infty\right)$	$1-\frac{2}{2+\theta}$	$\left(0,1\right]$
Frank	$-\frac{1}{\theta}\ln\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right)$	$\left(-\infty,\infty\right)$	$1-\frac{4}{\theta}\left(1-\frac{1}{\theta}\int_0^{\theta}\frac{t}{e^t-1}dt\right)$	$\left[-1,1\right]\backslash\left\{0\right\}$
Gaussian	$\int_{-\infty}^{\phi^{-1}(u)}\int_{-\infty}^{\phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}\exp\left(\frac{2\theta sw-s^2-w^2}{2\left(1-\theta^2\right)}\right)dsdw$	$\left[-1,1\right]$	$\frac{2}{\pi}\arcsin\theta$	$\left[-1,1\right]$

Since the joint CDF is expressed as a function of marginal CDFs, it is easy to model the joint CDF using marginal CDFs and correlation parameters that can be obtained from the experimental data. Moreover, since the copula decouples marginal CDFs and the joint CDF, the joint CDF modeled by the copula can be expressed in terms of any types of marginal CDFs. Even though the same copula is used, various types of joint distributions can be generated according to the marginal distribution types. To model the joint CDF using the copula, the matrix of correlation parameters $\boldsymbol{\theta}$ needs to be obtained from the experimental data. Since various types of copulas have their own correlation parameters, it is desirable to have a common correlation measure to obtain the correlation parameters from the experimental data.

Kendall's tau [10,11] is defined as the probability of concordance minus the probability of discordance between two random vectors $\left(X_1,Y_1\right)$ and $\left(X_2,Y_2\right)$ with the same margins $u=F\left(x\right)$ and $v=G\left(y\right)$, but with a common copula, $H\left(x,y\right)=C\left(F\left(x\right),G\left(y\right)\middle|\theta\right)$ of $\left(X_1,Y_1\right)$ and $\left(X_2,Y_2\right)$. The population version of Kendall's tau is expressed as

$$\tau=4\int\int_{I^2}C\left(u,v\middle|\theta\right)dC\left(u,v\middle|\theta\right)-1 \tag{3}$$

For some copulas, Eq. (3) can be simplified as shown in Table 1. The sample version of Kendall's tau is

$$t = \frac{c-d}{c+d} = (c-d)/\binom{ns}{2} \tag{4}$$

where *c* is the number of concordant pairs, *d* is the number of discordant pairs, and *ns* is the number of samples. Once the Kendall's tau is obtained from samples using Eq. (4), the correlation parameter θ can be obtained using Eq. (3) or explicit formulations given in Table 1.

4.3 Identification and Quantification of Input Model

The two most representative methods that determine a marginal CDF and copula for the given data are the goodness-of-fit test (GOF) [12-14] and the Bayesian method [5,6,15]. The GOF test has been developed and widely used to select a marginal CDF, but it relies on the parameters estimated from samples. Thus, if the parameters are incorrectly estimated, then a wrong marginal CDF might be selected. On the other hand, since the Bayesian method calculates weights to identify marginal CDFs by integrating the likelihood function over the domain of the parameter, it is less dependent of the choice of the parameter. Thus, the Bayesian method is preferred over the GOF test [5,6,15]. The numerical comparison of the GOF test and Bayesian method is presented by Noh et al. [5,6].

4.3.1 Identification of Input Model Using Bayesian Method

Consider a finite set $s_q \subset s$ consisting of candidate, marginal CDFs and copulas M_k, $k = 1, \cdots, q$, where s is a set of all candidates and *q* is the number of the candidates.

The Bayesian method consists of defining *q* hypotheses:

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

The probability of each hypothesis h_k given the data D is defined as [5,6,15]

$$\Pr(h_k|D,I) = \frac{\Pr(D|h_k,I)\Pr(h_k|I)}{\Pr(D|I)} \tag{5}$$

where $\Pr(D|h_k,I)$ is the likelihood function, $\Pr(h_k|I)$ is the prior on the candidate, and $\Pr(D|I)$ is the normalization constant with any relevant additional knowledge I.

Consider the likelihood function of the candidates. Under the hypothesis h_k that the data D come from the marginal CDF M_k, the probability of drawing the data D for the hypothesis on M_k is expressed as a likelihood function as

$$\Pr(D|h_k,\mu,\sigma,I) = \prod_{i=1}^{ns} f_k\left(x_i|a(\mu,\sigma),b(\mu,\sigma)\right) \tag{6}$$

where x_i is the i[th] sample value. Since each marginal PDF f_k has its own parameters *a* and *b*, a common parameter, mean or standard deviation, needs to be used as a nuisance variable to calculate Eq. (5) for candidate marginal CDFs. The parameters *a* and *b* of some marginal distributions, which are expressed in terms of mean and standard deviation, are presented by Noh et al [5]. In this paper, the mean is used as the parameter γ for integrating Eq. (6). Accordingly, Eq. (5) results in

$$\Pr(h_k|D,I) = \int_{-\infty}^{\infty} \Pr(h_k,\gamma,\sigma|D,I)\,d\gamma = \int_{-\infty}^{\infty} \frac{\Pr(D|h_k,\gamma,\sigma,I)\Pr(h_k|\gamma,I)\Pr(\gamma|I)}{\Pr(D|I)}\,d\gamma \tag{7}$$

where the mean is the variable for the integration and standard deviation is calculated from samples.

Under the hypothesis h_k that the data D come from the copula M_k, the probability of drawing the data D for the hypothesis on M_k is expressed as a likelihood function as

$$\Pr(D|h_k,\tau,I) = \prod_{i=1}^{ns} c_k\left(u_i,v_i|g_k^{-1}(\tau)\right) \tag{8}$$

where (u_i,v_i) are ns mutually independent pairs of the data and are calculated as $u_i = F_X(x_i)$ and $v_i = F_Y(y_i)$ where $F_X(x_i)$ and $F_Y(y_i)$ are the marginal CDF values evaluated at the given paired data (x_i,y_i). $g_k^{-1}(\tau)$ is the correlation parameter θ in the copula function, which is expressed as an inverse function of $\tau = g_k(\theta)$ for $k = 1, \cdots, q$ as shown in Table 1.

Using the Kendall's tau as a parameter γ for integrating the likelihood function, Eq. (8), Eq. (5) can be rewritten as [5,6,15]

$$\Pr(h_k|D,I) = \int_{-1}^{1} \Pr(h_k,\gamma|D,I)\,d\gamma = \int_{-1}^{1} \frac{\Pr(D|h_k,\gamma,I)\Pr(h_k|\gamma,I)\Pr(\gamma|I)}{\Pr(D|I)}\,d\gamma \tag{9}$$

In Eqs. (7) and (9), $\Pr(D|I)$ can be expressed as

$$\Pr(D|I) = \sum_{k=1}^{q} \Pr(D|h_k, I)\Pr(h_k|I) \tag{10}$$

Since $\Pr(D|I)$ is constant, it is not included for convenience in this paper.

To calculate the priors on the candidate, $\Pr(h_k|\gamma, I)$ and $\Pr(\gamma|I)$, let the additional information I be as follows:

I_1: A parameter γ belongs to the set Λ^γ , and the estimated parameter from samples is equally likely;

I_2: for a given parameter, all candidates satisfying $\gamma \in \Omega_k^\gamma$ are equally probable, where Ω_k^γ are domains of γ for M_k .

The set Λ^γ provides information on the interval of the parameter that the user might know. For example, if the user knows the specific domain of Λ^γ , the domain can be used to integrate the likelihood function for calculation of weights of each candidate. However, if any information on the interval of the parameter is not given, it might be assumed as $\Lambda^\gamma = (-\infty, \infty)$ for mean, and $\Lambda^\gamma = [-1,1]$ for Kendall's tau. For the mean, the infinite domain practically cannot be used to integrate the likelihood function, and thus the finite range of Λ^γ needs to be determined from samples such that Λ^γ covers the wide range of the parameter. Using the first additional information I_1, the prior on the parameter can be defined as

$$\Pr(\gamma|I_1) = \begin{cases} \dfrac{1}{\lambda(\Lambda^\gamma)}, & \gamma \in \Lambda^\gamma \\ 0, & \gamma \notin \Lambda^\gamma \end{cases} \tag{11}$$

where $\lambda(\cdot)$ is the Lebesgue measure, which is the width of the interval Λ^γ . Likewise, since all candidates are equally probable for $\gamma \in \Omega_k^\gamma$, the prior on the candidate is defined as

$$\Pr(h_k|\gamma, I_2) = \begin{cases} 1, & \gamma \in \Omega_k^\gamma \\ 0, & \gamma \notin \Omega_k^\gamma \end{cases} \tag{12}$$

In this paper, it is assumed that the prior follows a uniform distribution, which means there is no information on the distribution of the parameter, γ . If it is known that the prior distribution of γ follows a certain distribution such as the Gaussian, $\Pr(\gamma|I_1)$ might be expressed as a PDF and can be used as a prior distribution instead of Eq. (11). However, since the prior distribution of γ is usually unknown and the effect of the prior is negligible when the number of samples is enough (larger than 100 samples), Eq. (11) can be used in most cases.

By substituting Eqs. (11) and (12) into Eq. (7), Eq. (7) can be expressed as the computation of the weight for the marginal CDF:

$$W_k = \frac{1}{\lambda(\Lambda^\gamma)} \int_{\Omega_k^\gamma \cap \Lambda^\gamma} \prod_{i=1}^{ns} f_k\left(x_i | a(\gamma,\sigma), b(\gamma,\sigma)\right) d\gamma \tag{13}$$

Likewise, by substituting Eqs. (11) and (12) into Eq. (9), Eq. (9) is expressed as the calculation of the weight for the copulas:

$$W_k = \frac{1}{\lambda(\Lambda^\gamma)} \int_{\Omega_k^\gamma \cap \Lambda^\gamma} \prod_{i=1}^{ns} c_k\left(u_i, v_i | g_k^{-1}(\gamma)\right) d\gamma \tag{14}$$

The normalized weight of each candidate is calculated as

$$w_k = \frac{W_k}{\sum_{i=1}^{q} W_k} \tag{15}$$

The candidate with the highest normalized weight is identified as the one that best describes given data among candidates.

4.3.2 Quantification of Input Model

Once marginal and joint CDF types are identified using the Bayesian method, it is necessary to evaluate their parameters based on given data. To calculate the parameters of the marginal CDFs, the maximum likelihood estimate (MLE) can be used. However, the MLE might be biased [16], which means that the expected value of the parameter might not be the same as the parameter being estimated. For instance, the MLE is biased for estimating the variance σ^2 of a Gaussian distribution. On the other hand, a minimum variance unbiased estimator (MVUE), which is a commonly used estimator, is unbiased for estimating the variance σ^2 , and the variance of σ^2 , $Var[\sigma^2]$, is smaller than any other unbiased estimators.

This estimator is derived by Likes [17]. The mean and variance that are estimated from given samples are called the sample mean $\bar{x}$ and variance s^2, respectively.

The sample mean is calculated as

$$\bar{x} = \sum_{i=1}^{ns} \frac{x_i}{ns} \tag{16}$$

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

$$s^2 = \frac{1}{ns} \sum_{i=1}^{ns} \left(x_i - \bar{x} \right)^2 \tag{17}$$

and the sample variance for MVUE is calculated as

$$s^2 = \frac{1}{ns-1} \sum_{i=1}^{ns} \left(x_i - \bar{x} \right)^2 \tag{18}$$

respectively. Using the estimated mean and variance from the estimators, parameters *a* and *b* of the identified distribution can be estimated using explicit functions, which are expressed as mean and standard deviation presented by Noh et al. [5]. Second, to calculate the correlation parameter θ for the identified copula, Eq. (4) is first used to calculate the Kendall's tau from samples. Then, the correlation parameter θ is calculated using Eq. (3) or explicit formulations of θ expressed as Kendall's tau. Using the estimated parameters, the input model for carrying out the RBDO, is now obtained.

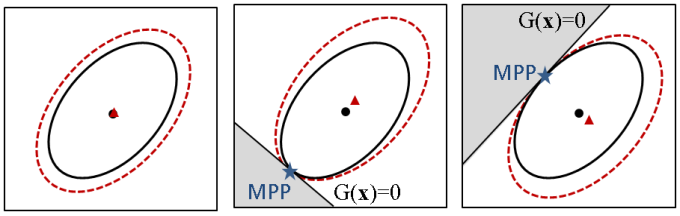
5. Assessment of Confidence Level of Input Model

Although the input model is identified and quantified as explained in Section 4, the input model uncertainties still exist due to the limited data. In this section, the confidence level on the input model is introduced to alleviate the inexact identification and quantification of the input model, and to obtain the confidence level of the output performance.

5.1. Confidence Level on Joint PDF Contour for Reliability Index

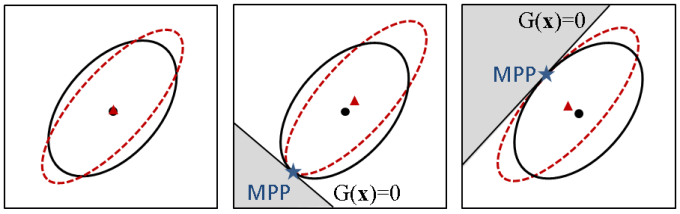
Even if the confidence level of the input model is obtained, the confidence level of the output performance might not be the same as the confidence level of the input model. The exact confidence level of the output performance is difficult to obtain because, for the given target reliability contour of the input model, the confidence level on the output performance can be different for different RBDO problems. Thus, a common measure for the calculation of the confidence level of the input model that can provide the confidence level of the output performance needs to be defined.

In this study, the joint PDF contour for the target reliability index β_t is used as the measure. If the joint PDF contour for β_t obtained from the estimated input model covers a true contour, then it will yield a reliable optimum design. Otherwise, it might yield either a reliable or unreliable optimum design depending on the active constraints and their MPP points. Figure 3 shows the former case in which the estimated joint PDF contour for β_t, indicated as a dotted contour, is larger than the true contour, indicated as a solid contour.



(a) Contours for β_t (b) Case 1 (c) Case 2

Figure 3. Large Estimated Joint PDF Contour and True Contour for β_t with Different Constraints



(a) Contours for β_t (b) Case 1 (c) Case 2

Figure 4. Estimated Joint PDF Contour and True Contour for β_t with Different Constraints

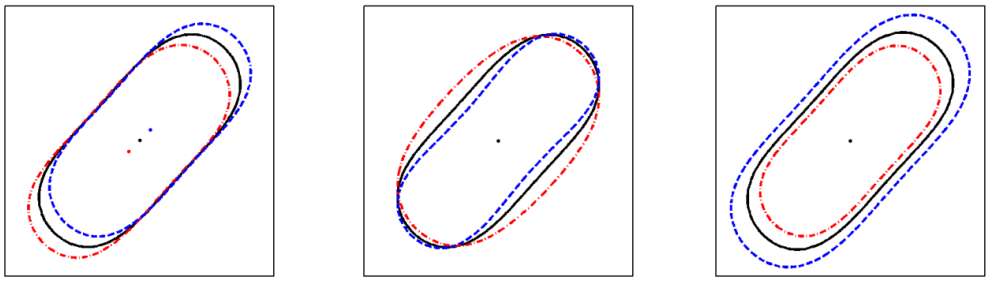
In Figs. 3 and 4, the gray region is where the constraint is failure, i.e., $G(\mathbf{x}) > 0$. In Fig. 3, the circular dot at the

center of the solid contour is the optimum design obtained from the true contour, and the triangular mark is the optimum design obtained from the estimated contour. Since the estimated optimum is farther away from the active constraint function than the true optimum, the estimated contour yields a reliable optimum designs as shown in Fig. 3 (b) and (c). However, if the estimated contour does not fully cover the true contour as shown in Fig. 4 (a), then it could yield a reliable optimum design as shown in Fig. 4 (b), or not as shown in Fig. 4 (c). Thus, even though it is a conservative measure, the confidence level of the output performance needs to be assured by using the probability that the estimated contour fully covers the true contour.

To obtain the contour that fully covers the true contour, it is necessary to know which parameter controls the size of the contour. First, the upper and lower bounds of the confidence interval of the mean are related to the position of the contour as shown in Fig. 5 (a) where the upper and lower bounds of the mean are indicated as dashed and dashed-dotted contours, respectively. Further, the mean is usually accurately estimated even for small number of samples and thus, it is not considered.

Second, the contour with the upper bound of the confidence interval of the correlation coefficient, indicated as a dashed contour in Fig. 5 (b), is the similar case with the dashed contour in Fig. 4. Even though the contour with the lower bound of the confidence interval of the correlation coefficient is used as shown in the dashed-dotted contour in Fig. 5 (b), it still does not cover the true contour. Therefore, the confidence interval on the correlation coefficient is not used to assess the confidence level of the input model.

Finally, the upper bound of the confidence interval of the standard deviation always yields a large contour, indicated as a dashed contour in Fig. 5 (c), which leads to a reliable design. Thus, the upper bound of standard deviation is used to estimate the confidence level on the input model.



(a) Mean (b) Correlation coefficient (c) Standard deviation

Figure 5. Joint PDF Contours for β_t Using Lower and Estimated, and Upper bound of Each Parameter

5.2. Confidence Interval of Standard Deviation

Suppose that X is a Gaussian random variable and *ns* samples, $x_1, x_2, \cdots, x_{ns}$, are collected to estimate the population variance σ^2, which is unknown constant. To estimate the population variance σ^2, it is assumed that the samples come from *ns* independent Gaussian random variables, i.e. X_1, X_2, $\cdots$, X_{ns}. That is, each sample is considered as a random variable. The sample variance can be calculated as [18]

$$S^2 = \frac{1}{ns-1}\sum_{i=1}^{ns}\left(X_i - \bar{X}\right)^2 \tag{19}$$

where $\bar{X}$ is the sample mean, which is a Gaussian variable.

Let μ be a population mean, which is unknown constant. To calculate the confidence interval of the standard deviation, the distribution of the variance S^2 needs to be determined when X is a Gaussian random variable. Equation (19) is rewritten as

$$(ns-1)S^2 = \sum_{i=1}^{ns}\left[\left(X_i-\mu\right)-\left(\bar{X}-\mu\right)\right]^2 = \sum_{i=1}^{ns}\left(X_i-\mu\right)^2 - ns\left(\bar{X}-\mu\right)^2 \tag{20}$$

Dividing both sides of Eq. (20) by σ^2, Eq. (20) is written as

$$\frac{(ns-1)S^2}{\sigma^2} = \sum_{i=1}^{ns}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{ns}}\right)^2 \tag{21}$$

Since μ and σ^2 are constant, and the first term on the right side of Eq. (21) is a sum of the *ns* squared independent Gaussian variables, and thus has a chi-square distribution with *ns* degrees of freedom [19], denoted as χ^2_{ns}.

The second term on the right side has only one squared Gaussian variable, and thus has a chi-square distribution with one degree of freedom. Since the sum of two chi-squared distributions with *i* and *j* degrees of freedom is also the

chi-square distribution with (i+j) degrees of freedom [20], the left side of Eq. (21) has a chi-square distribution with (ns–1) degree of freedom, denoted as χ^2_{ns-1}.

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

$$\Pr\left[c_{\alpha/2,ns-1} \le \frac{(ns-1)S^2}{\sigma^2} \le c_{1-\alpha/2,ns-1}\right] = 1-\alpha \tag{22}$$

where $c_{\alpha/2,ns-1}$ and $c_{1-\alpha/2,ns-1}$ are the critical values of the chi-square distribution evaluated at the probability levels of α/2 and (1–α/2) with (ns–1) degrees of freedom, respectively [18]. For 95% confidence level on the standard deviation, for example, set α = 0.05. Using the realization of S^2, denoted as s^2, Eq. (22) is rewritten as

$$\Pr\left[\frac{(ns-1)s^2}{c_{1-\alpha/2,ns-1}} \le \sigma^2 \le \frac{(ns-1)s^2}{c_{\alpha/2,ns-1}}\right] = 1-\alpha \tag{23}$$

Thus, for two-sided (1–α), the lower and upper bounds of the confidence interval for the standard deviation, σ^L and σ^U, respectively, are calculated as

$$\sigma^L = \sqrt{\frac{(ns-1)s^2}{c_{1-\alpha/2,ns-1}}} \text{ and } \sigma^U = \sqrt{\frac{(ns-1)s^2}{c_{\alpha/2,ns-1}}} \tag{24}$$

5.3. Confidence Level of Input Model

Once the upper bound of the confidence interval on the standard deviation is obtained for the target confidence level (1–α), the confidence level on the input model can be estimated. Figure 6 shows a flowchart for assessment of the confidence level on the input model. When a true input model is given – for example, the Frank copula with Gaussian marginal CDFs is the true joint CDF in this testing – different samples, ns=30, 100, and 300, can be generated from the true model for sufficient number of trials such as 300. Based on the generated samples, the marginal parameters and upper bound of the confidence intervals of the estimated standard deviation are quantified and the marginal distribution types of X_1 and X_2 are identified using the Bayesian method.

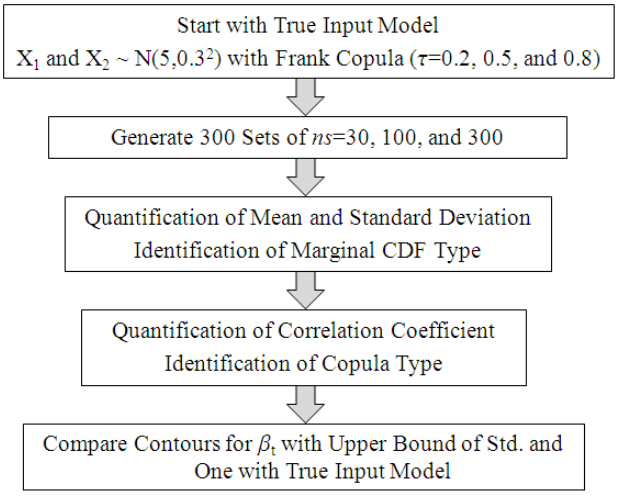


Figure 6. Flowchart of Assessment of Input Confidence Level

Once the marginal CDF types are identified, then the correlation coefficient is calculated and a copula that best describes the given data is identified using the Bayesian method. Using the identified input model with the upper bound of the confidence interval of the estimated standard deviation, the joint PDF contour for β_t can be drawn. By comparing the estimated contour with the target confidence level and the one with true model, the confidence level on the input model is assessed by calculating the probability that the obtained contour is larger than the true contour by carrying out 300 trials.

Figure 7 shows a true contour and 50 contours with estimated input models for the target confidence level of 95% from a different number of samples ns=30, 100, and 300 generated from a true input model, which is the Frank copula with X_1 and $X_2 \sim N(5, 0.3^2)$. To observe the effect of the correlation coefficient on the assessment of the confidence level, different correlation coefficients, τ=0.2, 0.5, and 0.8, are tested.

When the number of samples is small, *ns*=30, the obtained contour shapes are rather irregular and sometimes do not cover the true contour due to wrong estimation of the input model as shown in Fig. 7 (a), (d), and (g). As the number of samples is increased to *ns*=300, the obtained contours mostly cover the true contour as shown in Fig. 7 (c), (f), and (i). In addition, when the true correlation coefficient is small, τ=0.2, the obtained contour mostly covers the true contour even for the small number of samples as shown in Fig. 7 (a). On the other hand, when the true correlation coefficient is large, τ=0.8, the obtained contours cannot cover well the true contour especially when the correlation coefficient is estimated as higher than 0.8 as shown in Fig. 7 (g). However, for the large number of samples, since the correlation coefficient is accurately estimated and copulas are correctly identified in most cases, the obtained contours mostly cover the true contour as shown in Fig. 7 (i). Calculating the probability that the obtained contour is larger than the true contour over 300 data sets, the confidence level of the input model is obtained for a different number of samples and different correlation coefficients as shown in Table 2, where the target confidence level on the standard deviation is given as 95%.

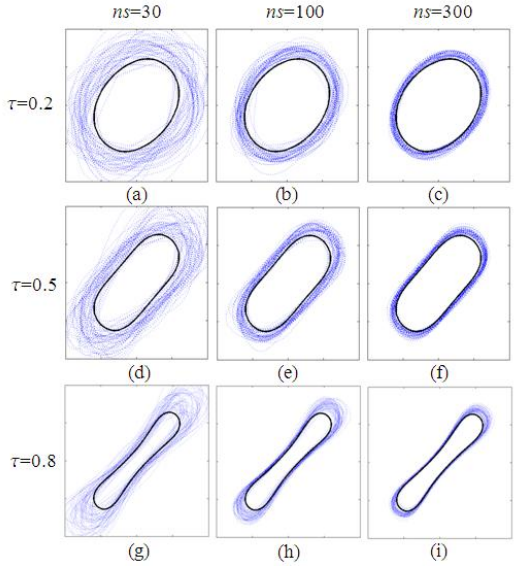


Figure 7. Joint PDF Contours for $\beta_t = 2$ Using Obtained Input Models

Table 2. Obtained Confidence Level of Input Model (%) for Frank Copula

ns	τ=0.2	τ=0.5	τ=0.8
30	88	77	52
100	92	83	75
300	95	94	93

As stated earlier, when the number of samples is small, the obtained confidence level for τ=0.8 is much lower than τ=0.2 because the contour shape with a high correlation coefficient is more sensitive to the estimated correlation coefficient than the one with the low correlation coefficient. Still, as the number of samples increases, the confidence level converges to the target confidence level, 95%. The Frank copula has a similar shape with some copulas such as Gaussian, whereas Clayton copula has a distinct shape. Thus, when the true model is the Clayton copula with X_1 and $X_2 \sim N\left(5, 0.3^2\right)$, the obtained confidence level for Clayton copula is higher than the one for Frank copula especially for *ns*=30 and 100 as shown in Table 3. As the number of samples increases, the obtained confidence level converges to the target confidence level 95%.

It is very important to note that the confidence level of the input model is a conservative measure from the RBDO point of view. That is, even if the obtained contour for the target reliability β_t does not cover the true contour for β_t, it does not necessarily mean that the reliability of the optimum design does not meet the target reliability β_t. It depends on the where the MPP points will be at the reliability-based optimum design. That is, for example, even if we use the input models with confidence levels of 52% in Table 3 or 63% in Table 4, the confidence level of the reliability-based optimum design meeting the target reliability β_t could be significantly higher than 52% or 63%, respectively. In Section 7, it is shown how the confidence level of the reliability-based optimum design meeting the target reliability β_t is obtained through a mathematical example.

Table 3. Obtained Confidence Level of Input Model (%) for Clayton Copula

ns	τ=0.2	τ=0.5	τ=0.8
30	84	80	63
100	88	88	82
300	95	94	94

6. Reliability-Based Design Optimization Using MPP-Based DRM

The FORM is commonly used for the inverse reliability analysis. However, when the constraint function is nonlinear or multi-dimensional, the inverse reliability analysis using the FORM could be erroneous because the FORM cannot handle the complexity of nonlinear functions or multi-dimensional functions. Accordingly, the correlation of input random variables, which increases the nonlinearity of constraint functions in the transformed standard Gaussian space, may cause significant error in the FORM-based inverse reliability analysis. In addition, it is shown in Ref. 8 that, to carry out the accurate reliability analysis and reduce the transformation ordering dependency of the inverse reliability analysis result, the MPP-based DRM [7] for the inverse reliability analysis is necessary.

In general, an RBDO problem can be formulated to

$$\begin{aligned} &\text{minimize} \quad \text{cost}(\mathbf{d}) \\ &\text{subject to} \quad P(G_i(\mathbf{X}) > 0) \le P_{F_i}^{Tar},\ i = 1, \cdots, nc \\ &\qquad \mathbf{d}_L \le \mathbf{d} \le \mathbf{d}_U,\ \ \mathbf{d} \in R^{ndv}\ \ and\ \ \mathbf{X} \in R^n \end{aligned} \tag{25}$$

where $\mathbf{X}$ is the vector of random variables; $\mathbf{d}$ is the vector of design variables, which is the mean value of the random variables $\mathbf{X}$, $\mathbf{d} = \mu(\mathbf{X})$; $G_i(\mathbf{X})$ represents the constraint functions; $P_{F_i}^{Tar}$ is the given target probability of failure for the i[th] constraint; and *nc*, *ndv*, and *n* are the number of probabilistic constraints, number of design variables, and number of random variables, respectively.

Using a performance measure approach (PMA) [21], the i[th] constraint can be rewritten, from Eq. (25), as

$$P\left(G_i(\mathbf{X}) > 0\right) - P_{F_i}^{\text{Tar}} \le 0 \Rightarrow G_i(\mathbf{x}^*) \le 0 \tag{26}$$

where $G_i(\mathbf{x}^*)$ is the i[th] constraint function evaluated at the MPP $\mathbf{x}^*$ in X-space.

To satisfy the feasibility of the constraint, the MPP needs to be estimated for each constraint by solving the following optimization problem:

$$\begin{aligned} &\text{maximize} \quad g_i(\mathbf{u}) \\ &\text{subject to} \quad \|\mathbf{u}\| = \beta_{t_i} \end{aligned} \tag{27}$$

where $g_i(\mathbf{u})$ is the i[th] constraint function that is transformed from the original space (**X**-space) into the standard Gaussian space (**U**-space), i.e., $g_i(\mathbf{u}) \equiv G_i(\mathbf{x}(\mathbf{u})) = G_i(\mathbf{x})$. β_{t_i} is the i[th] target reliability index, defined as $\beta_{t_i} = -\Phi^{-1}\left(P_{F_i}^{Tar}\right)$ where $\Phi^{-1}(\cdot)$ is the inverse of the Gaussian CDF. To further improve stability and efficiency of the PMA for RBDO, an enhanced performance measure approach (PMA+) was developed by Youn et al. [22], and is used in this paper.

The optimum point of Eq. (27) is denoted by the FORM-based MPP, $\mathbf{u}^*$ in **U**-space and correspondingly $\mathbf{x}^*$ in **X**-space. As shown in Eq. (26), if the constraint function at the MPP, $g_i(\mathbf{u}^*)$, is less than or equal to zero, then the i[th] constraint in Eq. (26) is satisfied for the given target reliability. If the MPP is updated through the calculated probability of failure using the MPP-based DRM, then the MPP is called a DRM-based MPP. The detailed algorithm of the MPP-based DRM is presented by Lee et al [7]. Thus, using Eqs. (25) and (26), the RBDO formulation using the MPP-based DRM can be rewritten as

$$\begin{aligned} &\text{minimize} \quad \text{cost}(\mathbf{d}) \\ &\text{subject to} \quad G_i(\mathbf{x}_{\text{DRM}}^*) \le 0,\ \ i = 1, \cdots, nc \\ &\qquad \mathbf{d}^L \le \mathbf{d} \le \mathbf{d}^U,\quad \mathbf{d} \in \mathrm{R}^{ndv} \text{ and } \mathbf{X} \in \mathrm{R}^n \end{aligned} \tag{28}$$

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

7. Numerical Examples

In this section, a mathematical example and a fatigue problem with correlated input variables are used to show how the input models with and without confidence level yield the confidence levels of RBDO results.

7.1. Mathematical Example

Consider a two-dimensional mathematical example where $X_1 \sim LN(1.608, 0.060)$ and $X_2 \sim N(5.0, 0.3^2)$ are correlated with the Frank copula ($\tau = 0.5$) and the parameters of the lognormal distribution, a = 1.608 and b = 0.060, are obtained from $\mu = 5.0$ and $\sigma = 0.3$ using

$$\mu = e^{a+b^2/2}, \text{ and } \sigma^2 = \left(e^{b^2} - 1\right)e^{2a+b^2} \tag{29}$$

An RBDO formulation is defined to

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

Carrying out the MPP-based DRM with the true input model, the optimum design is obtained as $\mathbf{d}^{\text{true}} = (2.005, 1.397)$, which is shown in Fig. 8. The probabilities of failures for two active constraints are estimated as $P_{F_1} = 2.243\%$ and $P_{F_2} = 2.044\%$ at the optimum design point using Monte Carlo simulation (MCS).

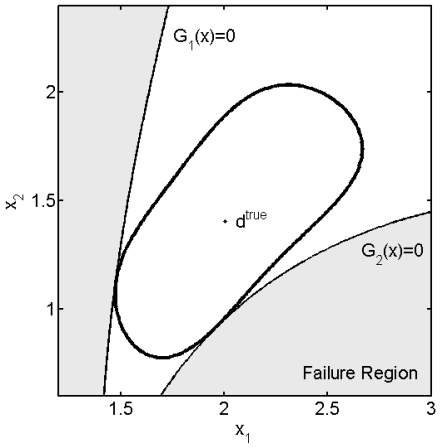


Figure 8. Contour for β_t=2 Using True Input Model

In Section 6, the confidence levels of the target contour β_t for the Frank copula with $\tau = 0.5$ and 0.8 are 77% and 52% using the upper bound of the 95% confidence interval of the standard deviation as shown in Table 2, which could lead to concern about the confidence level of the reliability-based optimum design meeting the target reliability β_t. In this example, the confidence level of the output performance is estimated using the Frank copula with $\tau = 0.5$ and 0.8; and 100 data sets of 30 samples generated from the true input model. Even though one mathematical example is not enough to test the confidence level of the output performance for RBDO problems, it might help us to understand that the output confidence level is larger than the confidence level of the input model.

First step is to estimate the mean μ^E, standard deviation σ^E, upper bound of the standard deviation σ^U, and Kendall's tau τ^E using 30 samples over 100 data sets. Table 4 shows the minimum, mean, and maximum values of the estimated parameters over 100 data sets where the true Kendall's tau is 0.5.

As shown in Table 4, since the mean μ^E can be more accurately estimated than other parameters, the variation of the estimated mean is smaller than the variation of the standard deviation and correlation coefficient. Thus, the effect of incorrectly estimated standard deviation and correlation coefficient on the optimum design is more significant than the effect of incorrectly estimated mean. However, even though the correlation coefficient has high variation, it cannot be effectively used to obtain the desirable confidence level of the output performance because its lower and upper bounds do not assure reliable optimum designs. On the other hand, the upper bound of the standard deviation is normally 0.1 higher than the estimated standard deviation, as shown in Table 4, so that the upper bound of the confidence interval of the standard deviation will yield more reliable design than the estimated standard deviation.

Table 4. Estimated Parameters Obtained from 100 Data Sets ($\tau = 0.5$)

Parameters		μ^E	σ^E	σ^U	τ^E
Min	X_1	4.892	0.227	0.305	0.315
	X_2	4.884	0.204	0.274	
Mean	X_1	5.015	0.304	0.408	0.513
	X_2	5.004	0.296	0.398	
Max	X_1	5.163	0.411	0.552	0.683
	X_2	5.127	0.415	0.557	

Second, the marginal and copula type needs to be identified from the data. Table 5 shows the number of identifications for marginal CDFs over 100 data sets. Since the Gaussian and Lognormal distribution shapes are very similar for $\mu = 5.0$ and $\sigma = 0.3$, those are mostly identified as correct marginal distributions for X_1 and X_2. Likewise, the Gaussian copula has a shape similar to that of the Frank copula; thus, it has the second largest number of identifications as shown in Table 6.

Table 5. Number of Identifications for Marginal CDFs (*ns*=30, 100 data sets)

Variables	Gaussian	Weibull	Gamma	Lognormal	Gumbel	Extreme	Extreme type-II
X_1	34	6	8	37	13	0	2
X_2	40	6	10	31	7	3	3

Table 6. Number of Identifications for Copula ($\tau = 0.5$, *ns*=30, 100 data sets)

Clayton	AMH	Gumbel	Frank	A12	A14	FGM	Gaussian	Independent
10	0	7	60	2	7	0	14	0

For each data set, an input model is obtained by identifying marginal CDF and copula types and quantifying parameters from the data set. Using the estimated standard deviation and its upper bound, two input models with and without the confidence level are implemented in RBDO where the target confidence level is given as 95%. Carrying out the MPP-based DRM for RBDO, the probabilities of failures for two active constraints are calculated using MCS with the true input model as shown in Table 7, where the target probability of failure P_F^{Tar} is 2.275%. Table 7 shows the minimum, maximum, and mean values of the probabilities of failures for two active constraints, P_{F_1} and P_{F_2}, over 100 data sets. The confidence level of the output performance is calculated by counting the number of cases that P_{F_1} and P_{F_2} are smaller than P_F^{Tar} over 100 data sets.

Table 7. Parameters of Probability of Failures Using Identified Marginal Distribution and Copula Types ($\tau = 0.5$, *ns*=30, 100 data sets)

	Input model without confidence level		Input model with confidence level	
	P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)
Min	0.125	0.523	0.001	0.079
Mean	2.265	3.244	0.484	1.572
Max	11.27	7.645	4.353	4.784
Estimated Confidence level	52%	28%	98%	81%

As seen in the minimum, mean, and maximum values of the probabilities of failures in Table 7, the input model with the confidence level indeed yields more reliable design than the one without confidence level. In this example, the first constraint is more mildly nonlinear than the second constraint, as shown in Fig. 8, so that the first constraint is less

dependent on the incorrect quantification and identification of the marginal and joint CDF than the second constraint. Thus, the confidence level for P_{F_1} is higher than the one for P_{F_2} for both input models. Depending on the constraint shapes near the MPPs, the confidence levels of the output performance are obtained as 98% and 81%, respectively, which are higher than the confidence level of the input model, 77% in Table 2. Even though 81% is not close to 95% of the target confidence level, note that the confidence level of P_{F_2} for the input model with the confidence level is much larger than the one without confidence level, 28%. Further, the output confidence level is obtained by counting the number of cases that the probability of failure is smaller than 2.275%, but among 10% of the other cases, P_{F_2} is less than 3.0%. That is, 90% of the estimated probabilities of failure are less than 3.0%. If the marginal CDF and copula types are known, the confidence level of the output performance becomes close to 95% as shown in Table 8.

Table 8. Parameters of Probability of Failures Using Correct Marginal Distribution and Copula Type ($\tau = 0.5$, *ns*=30, 100 data sets)

Parameters	Input model without confidence level		Input model with confidence level	
	P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)
Min	0.257	0.727	0.016	0.121
Max	6.855	6.402	2.171	2.597
Mean	2.536	2.560	0.531	0.962
Estimated Confidence level	56%	43%	99%	92%

Consider that two variables are highly correlated as $\tau = 0.8$. As shown in Table 2, the confidence level of the input model for small number of samples, *ns*=30, is very small, which is only 52%. Thus, it is also interesting to obtain the confidence level of the output performance. For this, 100 data sets of 30 samples for $\tau = 0.8$ are generated from the Frank copula with $X_1 \sim LN(1.608, 0.060)$ and $X_2 \sim N(5.0, 0.3^2)$. In this case, since the marginal CDF types and parameters are the same as the case for $\tau = 0.5$, the numbers of identification for marginal CDFs are similar to Table 5 as shown in Table 9. For the high correlation coefficient, because it is easy to identify the copula type due to its distinct shape, the correct copula (Frank) is more correctly identified than for $\tau = 0.5$, as shown in Table 10.

Table 9. Number of Identifications for Marginal CDF (*ns*=30, 100 data sets)

Variables	Gaussian	Weibull	Gamma	Lognormal	Gumbel	Extreme	Extreme type-II
X_1	33	5	12	39	8	1	2
X_2	40	12	10	29	5	0	4

Table 10. Number of Identifications for Copula ($\tau = 0.8$, *ns*=30, 100 data sets)

Clayton	AMH	Gumbel	Frank	A12	A14	FGM	Gaussian	Independent
3	0	3	82	5	3	0	4	0

Table 11 shows the minimum, mean, and maximum values of the estimated probability of failures for $\tau = 0.8$. As expected, the confidence levels of the output performance with the confidence level are much higher than those without confidence level. The output confidence levels for P_{F_1} and P_{F_2} are 99% and 70%, respectively, which are much higher than the confidence level of the input model, 52%, for the Frank copula. Even though 70% is not close to 95% of the target confidence level, the output confidence level for the input model with the confidence level is much larger than the one without confidence level, 51%. Further, 80% of the estimated probabilities of failure are less than 3.0%.

When the correct marginal CDFs and copula types are used as the input model, those yield the higher confidence levels of the output performance than the identified ones, as shown in Table 12, except the output confidence level for P_{F_1}. Since the identified models sometimes yield irregular contour shapes, which make the optimum designs be far away from the first constraint in this example, the output confidence levels using the incorrectly identified models are higher

than the one using the true model. On the other hand, when the upper bounds of the estimated standard deviation are used, those make the contours long and sharp with the positive correlation, and thus the optimum design becomes far away from the first constraint regardless of the identified models. Thus, the output confidence levels with true models are higher than the ones with identified models.

Table 11. Parameters of Probability of Failures Using Identified Marginal Distribution and Copula Type ($\tau = 0.8$, *ns*=30, 100 data sets)

Parameters	Input model without confidence level		Input model with confidence level	
	P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)
Min	0.075	0.578	0.000	0.270
Mean	2.048	2.711	0.364	1.953
Max	9.106	12.09	3.292	6.697
Estimated Confidence level	70%	51%	99%	70%

Table 12. Parameters of Probability of Failures Using Correct Marginal Distribution and Copula Type ($\tau = 0.8$, *ns*=30, 100 data sets)

Parameters	Input model without confidence level		Input model with confidence level	
	P_{F_1} (%)	P_{F_2} (%)	P_{F_1} (%)	P_{F_2} (%)
Min	0.196	0.368	0.011	0.106
Mean	2.366	2.646	0.463	1.519
Max	8.704	10.60	3.155	7.265
Estimated Confidence level	56%	54%	98%	81%

7.2. Roadarm Example

The roadarm in the M1A1 tank is modeled using 1572 eight-node isoparametric finite elements (SOLID45) and four beam elements (BEAM44) of a commercial program, ANSYS [24], as shown in Fig. 9. The material of the roadarm is S4340 steel with Young's modulus $E=3.0\times10^7$ psi and Poisson's ratio $\nu=0.3$. The durability analysis of the roadarm is carried out to obtain the fatigue life contour using Durability and Reliability Analysis Workspace (DRAW) [25,26]. The fatigue lives at the critical nodes are selected for design constraints of the RBDO in Fig. 10.

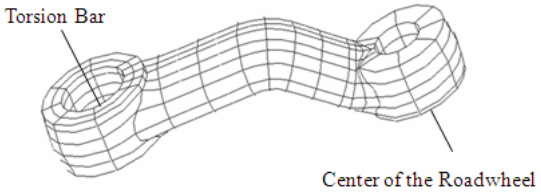


Figure 9. Finite Element Model of Roadarm

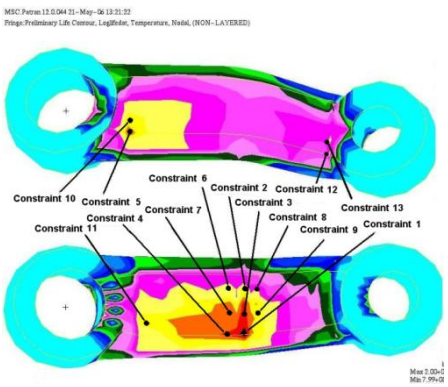


Figure 10. Fatigue Life Contour and Critical Nodes of Roadarm

In Fig. 11, the shape design variables consist of four cross-sectional shapes of the roadarm where the widths (x_1-direction) of the cross-sectional shapes are defined as design variables, d_1, d_3, d_5, and d_7, at intersections 1, 2, 3, and 4, respectively, and the heights (x_3-direction) of the cross-sectional shapes are defined as design variables, d_2, d_4, d_6, and d_8. Table 13 shows the initial design point, lower and upper bounds of eight design variables with their standard deviations and distribution types, and four material parameters with their means and standard deviations.

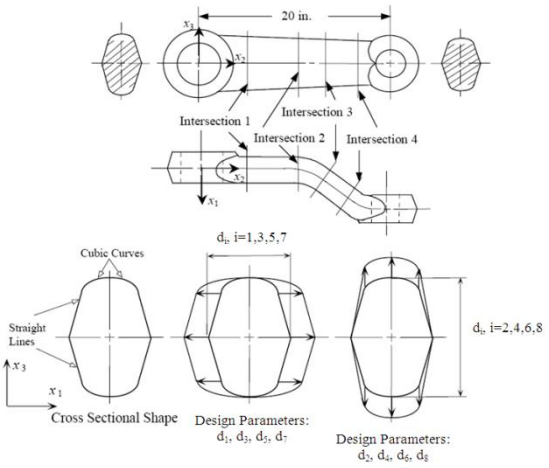


Figure 11. Shape Design Variables for Roadarm

Table 13. Random Variables and Fatigue Material Properties

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

Fatigue Material Properties

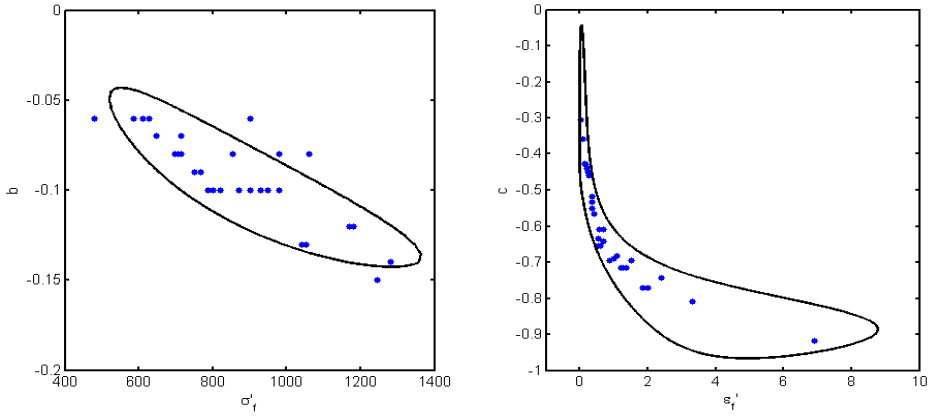
Non-design Uncertainties	Mean	Std.	Distribution Type
Fatigue Strength Coefficient, σ_f'	177000	44250	Lognormal
Fatigue Strength Exponent, b	-0.073	0.018	Gaussian
Fatigue Ductility Coefficient, ε_f'	0.410	0.205	Lognormal
Fatigue Ductility Exponent, c	-0.600	0.150	Gaussian

To test the input model with the confidence level, experimental data need to be used to obtain the upper bound of the standard deviation from the data. However, the experimental data of S4340, which is used in the roadarm, is not available. Thus, in this paper, 30 paired data are generated from an assumed true input model. First, it is assumed that Frank copula for $\sigma_f^{'}$ and b, and Gaussian copula for $\varepsilon_f^{'}$ and c, respectively, are the true copulas. As the two copulas well describe the experimental data of SAE 950X as shown in Fig. 12, it seems to be reasonable to select these two copulas to model joint CDFs of the four correlated random parameters of S4340. The marginal distribution types of S4340 are assumed to be the same as those of SAE 950X.

Second, once the copula and marginal distribution types are obtained, the mean and standard deviation of S4340 need to be determined. The mean values of four fatigue material properties of S4340 are known, but the standard deviations are unknown. Therefore, the standard deviations are assumed using the coefficient of variation (COV), which is the ratio of

the standard deviation to the mean, of SAE 950X. The coefficient of variation of SAE 950X is 115% for $\varepsilon_f^{'}$, and 25% for other material properties [1]. Since S4340 is stronger material than SAE 950X, in this paper, it is assumed that COV of S4340 is 50% for $\varepsilon_f^{'}$, and 25% for other material properties to estimate the standard deviation as shown in Table 13.

Assuming that a true input model has the above statistical information on S4340, 30 data are randomly generated, and RBDO is carried out using the estimated input model with and without confidence level. Table 14 shows the estimated parameters, and the target confidence level is specified as 95% in this roadarm example.



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

Figure 12. PDF Contours of Gaussian and Frank Copula Identified from 29 Paired Data of SAE 950X Steel

Table 14. Estimated Parameters and Identified Copulas

	σ_f'	b	ε_f'	c
μ^E	176738	–0.073	0.395	–0.594
σ^E	34939	0.015	0.166	0.125
σ^U	46969	0.020	0.223	0.168
τ^E	-0.830		-0.908	
Copula	Gaussian		Frank	

The RBDO formulation of the roadarm is defined to

$$\begin{aligned}
&\text{minimize} \quad \text{cost}(\mathbf{d}) \\
&\text{subject to} \quad P(G_i(\mathbf{X}) \geq 0) \leq P_{F_i}^{Tar},\ i = 1 \sim 13 \\
&\mathbf{d} = \boldsymbol{\mu}(\mathbf{X}),\ \mathbf{d}_L \leq \mathbf{d} \leq \mathbf{d}_U,\ \mathbf{d} \in R^n, P_{F_i}^{\text{Tar}} = 2.275\% \\
&G_i(\mathbf{X}) = 1 - \frac{L(\mathbf{X})}{L_t},\ i = 1 \sim 13 \qquad (31) \\
&\text{cost}(\mathbf{d}) : \text{Weight of Roadarm} \\
&L(\mathbf{X}) : \text{Crack Initiation Fatigue Life,} \\
&L_t : \text{Crack Initiation Target Fatigue Life (=5 years)}
\end{aligned}$$

Table 15 shows a comparison of RBDO results using the MPP-based DRM for various input models. First, the RBDO results are compared for the independent and correlated input fatigue material properties. As shown in the table, when the correlation between material properties is considered, the optimized weight of the roadarm is significantly reduced from 592.22 to 514.02 for the same target reliability because the material properties are highly correlated. Thus, it is very important to correctly model the correlation between material properties to carry out the RBDO. Second, when the input model with the estimated standard deviation σ^E is used, the underestimated standard deviations (see Tables 13 and 14) yield an unreliable optimum design with the optimum cost that is smaller than the optimum cost obtained using the true input model (511.34 vs. 514.02). Since the MCS cannot be used for the benchmark test for this problem due to computational cost, the comparison of costs is used as a measure to check whether the obtained optimum design is reliable or not. On the other hand, when the input model with the upper bound of the standard deviation σ^U is used, the obtained

optimum cost is higher than the optimum cost obtained from the true input model (523.85 vs. 514.02), which indicates the obtained optimum design is reliable. Accordingly, the input model with a confidence level is indeed necessary to obtain a reliable optimum design.

Table 15. DRM-Based RBDO Comparison

	Initial	Independent	Correlated		
			True	σ^E	σ^U
d_1	1.750	2.194	1.928	1.959	2.013
d_2	3.250	2.650	2.650	2.650	2.650
d_3	1.750	2.602	2.067	2.030	2.047
d_4	3.170	3.010	2.577	2.615	2.616
d_5	1.756	2.656	1.776	1.705	1.770
d_6	3.038	2.538	3.535	3.537	3.538
d_7	1.752	2.422	2.075	2.060	2.152
d_8	2.908	2.895	2.512	2.509	2.645
Cost	515.09	592.22	514.02	511.34	523.85

8. Conclusion

To obtain reliable RBDO results, it is important to obtain an accurate input model from given experimental data for RBDO. The input model can be generated by identifying marginal CDFs and copula using the Bayesian method and by quantifying marginal and correlation parameters based on the given data. However, in practical applications, the experimental data is often insufficient, and thus it is difficult to obtain an accurate input model.

To offset inaccurate estimation of the input model, confidence level is incorporated in the input model for RBDO by using the upper bound of the confidence interval of the standard deviation. Since the confidence level of the input model is not the same as the confidence level of the output performance, a conservative measure of the confidence level of the input model is assessed by calculating the probability that the estimated joint PDF contour for the target reliability index β_t fully covers the true joint PDF contour for β_t. Simulation results show that, even though the upper bound of the 95% confidence interval of the standard deviation is used, the confidence level of the input model is significantly lower (as shown in Tables 2 and 3) when the sample size is small. On the other hand, since the measure of the confidence level used of the input model is conservative, the confidence level of the optimum design meeting the target reliability is higher as shown in Table 7 and 11 for the mathematical example. However, the results of Table 7 and 11 still do not reach the target confidence level of 95%. To achieve the target confidence level of the output performance especially for small number of samples with the high correlation, new options, such as the mean or correlation coefficient to enlarge the contour for β_t, is currently being investigated. In addition, the upper bound of the prediction interval of the standard deviation is another option to investigate.

9. Acknowledgments

This research is supported by the Automotive Research Center, which is sponsored by the U.S. Army Tank Automotive Research, Development and Engineering Center (TARDEC).

10. References

[1] Socie, D. F., Seminar notes: "Probabilistic Aspects of Fatigue", 2003, URL: http://www.fatiguecaculator.com [cited May 8 2008].

[2] Annis, C., "Probabilistic Life Prediction Isn't as Easy as It Looks," *Journal of ASTM International*, Vol. 1, No.2, pp. 3-14, 2004.

[3] Efstratios, N., Ghiocel, D., and Singhal, S., *Engineering design reliability handbook*, CRC press, New York, 2004.

[4] Pham, H., *Springer Handbook of Engineering Statistics*, Springer, London, 2006.

[5] Noh, Y., Choi, K.K., and Du, L., "Selection of Copulas to Generate Joint Distributions of Correlated Input Variables for RBDO," *ASME 34th Design Automation Conference*, August 3-6, New York, New York, 2008.

[6] Noh, Y., Choi, K.K., and Lee, I., "Identification of Marginal and Joint CDFs Using the Bayesian Method for RBDO," *Structural and Multidisciplinary Optimization*, Submitted, 2008.

[7] Lee, I., Choi, K.K., Du, L., and Gorsich, D., "Dimension Reduction Method-Based MPP for Reliability-Based Design Optimization of Highly Nonlinear Multi-Dimensional Systems," *Special Issue of Computer Methods in Applied Mechanics and Engineering: Computational Methods in Optimization Considering Uncertainties*, Vol. 198, No. 9-12, pp. 14-27, 2008.

[8] Noh, Y., Choi, K.K., and Lee, I., "Reduction of Ordering Effect in Reliability-Based Design Optimization Using Dimension Reduction Method," *AIAA Journal*, to appear, 2009.

[9] Nelsen, R.B., *An Introduction to Copulas*, Springer, New York, 1999.

[10] Kendall, M., "A New Measure of Rank Correlation," *Biometrika*, Vol. 30, pp. 81-89, 1938.

[11] Kruskal, W.H., "Ordinal Measures of Associations," *American Statistical Association Journal*, Vol. 53, No. 284, pp. 814-861, 1958.

[12] Cirrone, G. A.P., Donadio, S., Guatelli, S., and et al., "A Goodness-of-Fit Statistical Toolkit," *IEEE Transactions on Nuclear Science*, Vol. 51, No. 5, pp. 2056–2063, 2004.

[13] Genest, C., and Rémillard, B., "Validity of the Parametric Bootstrap for Goodness-of-fit Testing in Semiparametric Models," *Technical Rep. G-2005-51*, Group d'Études et de Recherche en Analyse des Décision, 2005.

[14] Genest, C., and Favre, A.-C., "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask," *Journal of Hydrologic Engineering*, Vol. 12, No. 4, pp. 347-368, 2007.

[15] Huard, D., Évin, G., and Favre, A.-C., "Bayesian Copula Selection," *Computational Statistics & Data Analysis*, Vol. 51, pp. 809-822, 2006.

[16] Baxter, M., "Minimum Variance Unbiased Estimation of the parameters of the Pareto Distribution," *Metrika*, Springer, Vol. 29. No. 1, pp. 133-138, 1980.

[17] Likes, J., "Minimum Variqance Unbiased Estimates of the Parameters of Power-Function and Pareto's Distribution," *Statistische Hefte*, Vol. 10, pp. 104-110, 1969.

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

[19] Ang, A.H-S., and Tang, W.H., *Probability Concepts in Engineering Design*, Vol. I: Decision, Risk and Reliability, Wiley, New York, 1984.

[20] Hoel, P.G. *Introduction to Mathematical Statistics (3rd Edition)*, Wiley, New York, 1962.

[21] Tu, J., and Choi, K.K., "A New Study on Reliability-Based Design Optimization," *ASME Journal of Mechanical Design*, Vo. 121, No. 4, pp. 557-564, 1999.

[22] Youn, B.D., Choi, K.K., and Du, L., "Enriched Performance Measure Approach for Reliability-Based Design Optimization," *AIAA Journal*, Vol. 43, No. 4, pp. 874-884, 2005.

[23] Lee, I., Choi, K.K., Noh, Y., Comparison Study between Probabilistic and Possibilistic Approach for Problems with Correlated Input and Lack of Input Statistical Information, *ASME 35th Design Automation Conference*, Aug. 30-Sep. 2, San Diego, CA, submitted, 2009.

[24] Swanson Analysis System Inc., *ANSYS Engineering Analysis System User's Manual*, Vol. I & II, Houston, PA, 1989.

[25] Center for Computer-Aided Design, College of Engineering, *DRAW Concept Manual*, The University of Iowa, Iowa City, IA, 1999a.

[26] Center for Computer-Aided Design, College of Engineering, *DRAW User Reference*, The University of Iowa, Iowa City, IA, 1999b.