RESEARCH PAPER

Conservative reliability-based design optimization method with insufficient input data

Hyunkyoo Cho[1] · K. K. Choi[1] · Nicholas J. Gaul[2] · Ikjin Lee[3] ·
David Lamb[4] · David Gorsich[4]

Abstract Reliability analysis and reliability-based design optimization (RBDO) require an exact input probabilistic model to obtain accurate probability of failure (PoF) and RBDO optimum design. However, often only limited input data is available to generate the input probabilistic model in practical engineering problems. The insufficient input data induces uncertainty in the input probabilistic model, and this uncertainty forces the PoF to be uncertain. Therefore, it is necessary to consider the PoF to follow a probability distribution. In this paper, the probability of the PoF is obtained with consecutive conditional probabilities of input distribution types and parameters using the Bayesian approach. The approximate conditional probabilities are obtained under reasonable assumptions, and Monte Carlo simulation is applied to calculate the probability of the PoF. The probability of the PoF at a user-specified target PoF is defined as the conservativeness level of the PoF. The conservativeness level, in addition to the target PoF, will be used as a probabilistic constraint in an RBDO process to obtain a conservative optimum design, for limited input data. Thus, the design sensitivity of the conservativeness level is derived to support an efficient optimization process. Using numerical examples, it is demonstrated that the conservativeness level should be involved in RBDO when input data is limited. The accuracy and efficiency of the proposed design sensitivity method is verified. Finally, conservative RBDO optimum designs are obtained using the developed methods for limited input data problems.

Nomenclature

CDF	Cumulative distribution function
PDF	Probability density function
PoF, p_F	Probability of failure
$G(\mathbf{x})$	Performance measure
Ω_F	Failure domain of a performance measure
$\mathbf{X}, \mathbf{x}$	Input random variable vector and its realization
N	Number (dimension) of input random variables
X_i	i-th input random variable
$f_{\mathbf{X}}(\mathbf{x}; \zeta, \psi)$	Joint PDF of $\mathbf{X}$
$f_{X_i}(x_i; \zeta_i, \mu_i, \sigma_i^2)$	Marginal PDF of X_i
$\mathbf{Z}, \zeta$	Input distribution types and their realizations
ζ_i	Marginal distribution type of X_i
$\mathbf{\Psi}, \psi$	Input distribution parameters and their realizations
M_i, μ_i	Input mean of X_i and its realization
Σ_i^2, σ_i^2	Input variance of X_i and its realization
$^*\mathbf{x}$	Input data set, $^*\mathbf{x} = \{^*\mathbf{x}_1, \ldots, ^*\mathbf{x}_N\}$
$^*x_i^{(j)}, ^*\mathbf{x}_i$	Input data and data set for X_i
ND	Number of input data

✉ K. K. Choi
kkchoi@engineering.uiowa.edu

1 Department of Mechanical and Industrial Engineering, The University of Iowa, Iowa City, IA 52242, USA

2 RAMDO Solutions, LLC, Iowa City, IA 52240, USA

3 Mechanical Engineering Department, Korea Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

4 US Army RDECOM/TARDEC, Warren, MI 48397-5000, USA

| $^*\overline{x}_i, {}^*\overline{\mathbf{x}}_i$ | Mean of input data and its vector form |
| $^*\tilde{x}_i, {}^*\tilde{\mathbf{x}}_i$ | Dispersion of input data and its vector form |

1 Introduction

Output variability, which is the variability of performance measures, impedes a design's ability to sustain performance during its life cycle. To obtain a safe and reliable design under output variability, reliability-based design optimization (RBDO) has been developed using the first-order reliability method (FORM) (Hasofer and Lind 1974; Ditlevsen and Madsen 1996; Tu et al. 1999; Haldar and Mahadevan 2000; Tu et al. 2001; Gumbert et al. 2003; Hou 2004), the second-order reliability method (SORM) (Hohenbichler and Rackwitz 1988; Breitung 1984; Lee et al. 2012; Lim et al. 2014), the dimension reduction method (DRM) (Rahman and Wei 2006; Rahman and Wei 2008; Lee et al. 2010), and Monte Carlo simulation (MCS) (Rubinstein and Kroese 2008; Lee et al. 2011a, b). Output variability is induced by input variability – i.e., the variability of the input random variable. In RBDO, an input probabilistic model, which is a statistical representation of the input variability, is used to obtain the output variability. Therefore, the accuracy of the input probabilistic model is necessary to obtain correct reliability of the RBDO optimum design.

The aforementioned RBDO methods require an accurate input probabilistic model – i.e., "true" input probabilistic model. Obtaining the true input probabilistic model is very difficult because it requires a very large number of test data for all subjects in the system. Unfortunately, due to cost and time constraints, it is highly probable that insufficient input data will be available for an input probabilistic model in a practical problem. Then, the input probabilistic model generated using the limited number of data becomes uncertain. As a result, the uncertainty in the input probabilistic model forces probability of failure (PoF), a measure of the reliability, to be uncertain. Consequently, new methods need to be developed to obtain more conservativeness when there is insufficient input data.

A safety factor approach could be an intuitive start to consider uncertainty in the input probabilistic model (Elishakoff 2004). P-boxes and probability bounds, which are essentially a new input probabilistic model at a certain confidence level based on the input data, have been 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 variability of the input random variables can be combined in a modified input probabilistic model by using intentionally enlarged input variances (Noh et al. 2011a, b). All of these methods adjust the input probabilistic model to reflect the uncertainty in it. However, the uncertainty in the input probabilistic model transfers to the PoF through performance measures. When the performance measure is nonlinear, it is hardly possible to estimate the uncertainty of the PoF accurately by altering the input probabilistic model. Moreover, modifying the input probabilistic model may mix the effect of input uncertainty (the uncertainty in the input probabilistic model due to insufficient data) and input variability (variability of the input random variables), which are essentially two different sources of output uncertainty and variability.

The Bayesian approach would be better for directly accessing the PoF and separating the effect of the input uncertainty and variability. In one study, the probability of fatigue failure 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 PoF was updated as more NDI data became available. The mean of the simulation output was qualified in the presence of the input uncertainty using the Bayesian model average (BMA) approach (Chick 2001), but the two sources were not clearly distinguished. Later, Gunawan and Papalambros successfully separated the two sources and assumed that the PoF follows beta distribution (Gunawan and Papalambros 2006). That is, the PoF, which quantifies the output variability induced by the input variability, also follows another probability distribution (the beta distribution, in this case) due to the input uncertainty. The cumulative distribution function (CDF) of the beta distribution at a certain PoF is the conservativeness level of the PoF. Using these observations, an RBDO problem of minimizing cost and maximizing conservativeness level was performed. Youn and Wang obtained an extreme case of the beta distribution using extreme distribution theory, and the median value of the extreme case was used as new probabilistic constraints for RBDO (Youn and Wang 2008). In addition, the design sensitivity of the probabilistic constraints was developed. However, the probability of the PoF still has not been fully utilized. Once the probability is obtained, the conservativeness level of the PoF is directly accessible. Then, the input uncertainty induced by insufficient input data is measured by the conservativeness level, while the input variability is captured by the PoF.

In this paper, a new method to estimate the conservativeness level of the PoF is presented. The new method directly

accesses the probability of the PoF using the Bayesian approach and distinguishes the input uncertainty and the input variability. By separating the two sources, users can specify separate target values for both the input uncertainty and variability in the new RBDO process. Moreover, a design sensitivity for the conservativeness level is developed to ensure the effectiveness and efficiency of the new RBDO process. In Section 2, the relationship between the PoF and insufficient input data is shown. In addition, how to deal with the input data in the new RBDO process is explained. In Section 3, the probability of the PoF is obtained using the Bayesian method, and the estimation method of conservativeness level is presented. In Section 4, the design sensitivity of the conservativeness level is derived. Then numerical examples are used to show the effectiveness and efficiency of the estimation of the conservativeness level, the design sensitivity method, and the new conservative RBDO process in Section 5. In Section 6, an 11-dimensional problem is tested to see the performance of developed method in high-dimensional applications. Finally, the conclusion is presented in Section 7.

2 Probability of failure and insufficient input data

Probability of failure is revisited in this section to associate it with input distribution types and parameters. Through this association, propagation of the input uncertainty due to insufficient input data to probability of the PoF is characterized in the following sections. Therefore, it is worth discussing the PoF before moving on to the main discoveries of this paper. In addition, how to treat input data in the RBDO process is explained in this section as well.

2.1 Probability of failure

The PoF p_F is defined using a multi-dimensional integral and an indicator function as

$$p_F(\boldsymbol{\zeta}, \boldsymbol{\psi}) = \int_{\mathbb{R}^N} I_{\Omega_F}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\zeta}, \boldsymbol{\psi}) \, d\mathbf{x} \quad (1)$$

where Ω_F is the failure domain such that a performance measure $G(\mathbf{x})$ is larger than zero (i.e., $G(\mathbf{x}) > 0$), $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\zeta}, \boldsymbol{\psi})$ is a joint probability density function (PDF) of input random variables $\mathbf{X}$ with input distribution types $\boldsymbol{\zeta}$ and input distribution parameters $\boldsymbol{\psi}$, $\mathbf{x}$ is realization of $\mathbf{X}$, N is the number of input random variables, and $I_{\Omega_F}(\mathbf{x})$ is an indicator function defined as

$$I_{\Omega_F}(\mathbf{x}) \equiv \begin{cases} 1, & \text{for } \mathbf{x} \in \Omega_F \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

In this paper, it is assumed that each input random variable X_i is statistically independent and has marginal distribution

with two parameters. Under the assumptions, the joint PDF in (1) can be expressed using marginal PDF as

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\zeta}, \boldsymbol{\psi}) = \prod_{i=1}^{N} f_{X_i}(x_i; \zeta_i, \mu_i, \sigma_i^2) \quad (3)$$

where f_{X_i}, ζ_i, μ_i, and σ_i^2 are marginal PDF, marginal distribution type, and mean and variance of input random variable X_i, respectively. Input distribution types and parameters can be defined as $\boldsymbol{\zeta} = \{\zeta_1, \ldots, \zeta_N\}$ and $\boldsymbol{\psi} = \{\mu_1, \sigma_1^2 \ldots, \mu_N, \sigma_N^2\}$, respectively. It is noted that mean (μ_i) and variance (σ_i^2) are used in (3) instead of the two parameters of the marginal distribution because they are invariants of the marginal distribution type and the two parameters can be uniquely determined using the mean and variance.

As a specific example, consider an input joint PDF with three input random variables of $\mathbf{X} = [X_1, X_2, X_3]^T$, which follow the Normal, Lognormal, and Gamma marginal distribution types, respectively. Each input random variable X_i has mean and variance of μ_i and σ_i^2. Then, the input joint PDF of the three input random variables can be expressed using (3) as

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\zeta}, \boldsymbol{\psi}) = f_{X_1}(x_1; \zeta_1, \mu_1, \sigma_1^2) \, f_{X_2}(x_2; \zeta_2, \mu_2, \sigma_2^2) f_{X_3}(x_3; \zeta_3, \mu_3, \sigma_3^2) \quad (4)$$

with input distribution types $\boldsymbol{\zeta} = \{\zeta_1, \zeta_2, \zeta_3\}$, input distribution parameters $\boldsymbol{\psi} = \{\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \mu_3, \sigma_3^2\}$, and ζ_1 =Normal, ζ_2 =Lognormal, ζ_3 =Gamma. As marginal distribution types and the distribution parameters are specified, the joint PDF in (4) is a specific PDF that can produce one value of the PoF in (1).

If the population data, which is a complete set of data for input random variables, is available, the true input distribution types $\boldsymbol{\zeta}$ and the true input distribution parameters $\boldsymbol{\psi}$ can be obtained. If this is the case, as explained earlier, (1) produces a fixed PoF value. However, in practical engineering problems, only limited data is available, which makes $\boldsymbol{\zeta}$ and $\boldsymbol{\psi}$ follow probability distributions instead of being fixed types or values. Therefore, capital characters $\mathbf{Z}$ and $\boldsymbol{\Psi}$ will be used to represent their randomness in the presence of limited data with $\boldsymbol{\zeta}$ and $\boldsymbol{\psi}$ being the corresponding realizations, respectively. $\mathbf{Z}$, $\boldsymbol{\Psi}$, and the amount of data cause the PoF to follow a probabilistic distribution.

The conventional RBDO methods that find optimum design using the fixed target PoF value based on a realization set $\boldsymbol{\zeta}$ and $\boldsymbol{\psi}$ can no longer produce reliable design when only limited data is available. Thus, a new RBDO method needs to be developed in order to produce a conservative optimum design that we can rely on, when only insufficient data is available, by considering the probability distributions of $\mathbf{Z}$ and $\boldsymbol{\Psi}$ and, eventually, the probability distribution of the PoF.

2.2 Input data decomposition

Let *x be the given input data set. 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 *x could contain the following data subsets:

$$*\mathbf{x} = \{*\mathbf{x}_1, \ldots, *\mathbf{x}_N\}. \tag{5}$$

The data subset $*\mathbf{x}_i$ for the i-th input random variable X_i is a column vector of size ND as

$$*\mathbf{x}_i = \left[\begin{array}{cccc} *x_i^{(1)} & *x_i^{(2)} & \cdots & *x_i^{(ND)} \end{array} \right]^{\mathrm{T}} \tag{6}$$

where $*x_i^{(j)}$ is the j-th data for X_i. The data subset $*\mathbf{x}_i$ can be decomposed into two parts as

$$*\mathbf{x}_i = {}^*\overline{\mathbf{x}}_i + {}^*\tilde{\mathbf{x}}_i \tag{7}$$

where ${}^*\overline{\mathbf{x}}_i$ is a column vector of size ND, the entities of which are the sample mean of the i-th data subset $*\mathbf{x}_i$ as

$$
{}^*\overline{\mathbf{x}}_i = \left[\begin{array}{cccc} {}^*\overline{x}_i {}^*\overline{x}_i & \cdots & {}^*\overline{x}_i \end{array} \right]^{\mathrm{T}} \text{ such that } {}^*\overline{x}_i
$$
$$
= \frac{1}{ND} \sum_{j=1}^{ND} *x_i^{(j)}. \tag{8}
$$

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 assumed in the RBDO process as

$$
{}^*\mathbf{x}_i = \mathbf{d}_i + {}^*\tilde{\mathbf{x}}_i \tag{9}
$$

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

$$
\mathbf{d}_i = [d_i \quad d_i \quad \cdots \quad d_i]^T. \tag{10}
$$

In (9), the input data in the RBDO process are changed to be centered at the current design point $\mathbf{d}$. Hence, as the design optimization proceeds, $\mathbf{d}$ moves according to the optimization process, and the data *x follows the design movement. However, ${}^*\tilde{\mathbf{x}}_i$, which is the dispersion of the input data with respect to the design point, is maintained in the RBDO process. An example is shown in Fig. 1. A pair of input random variable has five data pairs. The mean of data pairs ${}^*\overline{\mathbf{x}} = \{{}^*\overline{\mathbf{x}}_1, {}^*\overline{\mathbf{x}}_2\}$ has been moved to a design point $\mathbf{d} = \{\mathbf{d}_1, \mathbf{d}_2\}$. However, the dispersion of data pairs ${}^*\tilde{\mathbf{x}} = \{{}^*\tilde{\mathbf{x}}_1, {}^*\tilde{\mathbf{x}}_2\}$ with respect to the center points ${}^*\overline{\mathbf{x}}$ and $\mathbf{d}$ is maintained. The data decomposition in (9) is a usual practice in the conventional RBDO process. In the process, the design variable, which is the mean of the corresponding input random variable, changes as design iteration proceeds, while the
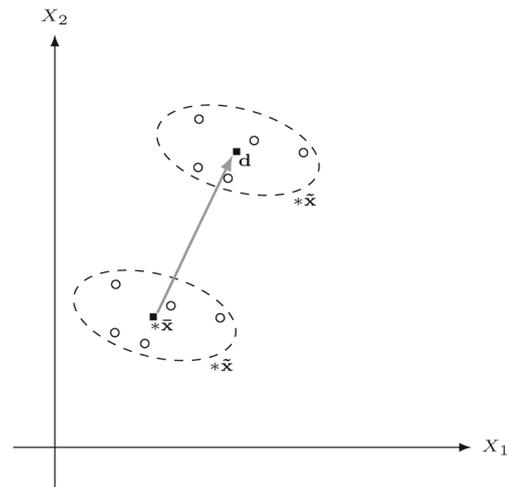
Fig. 1 Input data and design change

variance of the input random variable is constant. The same concept is applied to the RBDO with insufficient input data in this paper by decomposing input data and maintaining ${}^*\tilde{\mathbf{x}}_i$ in the RBDO process. It is noted that ${}^*\tilde{\mathbf{x}}_i$ contains the input uncertainty due to insufficient data and that it remains the same while the design is changed during the RBDO process.

3 Probability of PoF

In this section, the probability of the PoF is obtained using the given input data, the general expression of the joint PDF, and the Bayesian method. We can obtain a conservative RBDO optimal design, even with limited input data, by securing a certain probability of the PoF that is larger than a user-specified conservativeness level.

3.1 Probability of PoF

Consider a given input data set *x. As explained earlier, input distribution types $\mathbf{Z}$ and parameters $\mathbf{\Psi}$ follow probability distributions if the input data set *x, not true input distribution, is provided. In this paper, it is assumed that the probability distributions of $\mathbf{Z}$ and $\mathbf{\Psi}$ can be analogized from the *x. Using Bayes' theorem and the given *x, a joint PDF of the PoF P_F, input distribution types $\mathbf{Z}$, and input distribution parameters $\mathbf{\Psi}$ is obtained as

$$
f(p_F, \ \zeta, \ \psi|*\mathbf{x})
$$
$$
= f(p_F|\zeta, \psi, *\mathbf{x}) \, P(\zeta|\psi, *\mathbf{x}) \, f(\psi|*\mathbf{x}). \tag{11}
$$

In (11), the joint PDF is a product of three successive conditional probabilities. If all conditional probabilities on the

right side of (11) are available, the PDF of P_F can be obtained by integrating $\mathbf{Z}$ and $\boldsymbol{\Psi}$ in (11) as

$$
\begin{aligned}
& f_{P_F}(p_F|{}^*\mathbf{x}) \\
& = \sum_{\mathbf{Z}} \int_{\Omega_{\boldsymbol{\Psi}}} f(p_F,\ \boldsymbol{\zeta},\ \boldsymbol{\psi}|{}^*\mathbf{x})\, d\boldsymbol{\psi}, \quad p_F \in [0,\ 1]
\end{aligned}
\tag{12}
$$

Furthermore, the CDF of P_F is obtained by integrating (12) with respect to the PoF as

$$
\begin{aligned}
& F_{P_F}(p_F|{}^*\mathbf{x}) \\
& = \int_0^{p_F} \sum_{\mathbf{Z}} \int_{\Omega_{\boldsymbol{\Psi}}} f(\phi,\ \boldsymbol{\zeta},\ \boldsymbol{\psi}|{}^*\mathbf{x})\, d\boldsymbol{\psi}\, d\phi, \quad p_F \in [0,\ 1]
\end{aligned}
\tag{13}
$$

where ϕ is the variable that corresponds to P_F. The value of the CDF of P_F in (13) represents the probability that P_F of a design with the input data ${}^*\mathbf{x}$ is less than a specified value p_F. In other words, the CDF value is the probability that a design is more conservative and safer than p_F. Hence, in this paper, the CDF value of P_F at the specified p_F is designated as the "conservativeness level" of p_F.

Among the three conditional probabilities on the right side of (11), the first term $f(p_F|\boldsymbol{\zeta}, \boldsymbol{\psi}, {}^*\mathbf{x})$ is the probability of P_F with the given input distribution types $\boldsymbol{\zeta}$, parameters $\boldsymbol{\psi}$, and data ${}^*\mathbf{x}$. As explained earlier, the PoF in (1) is determined by $\boldsymbol{\zeta}$ and $\boldsymbol{\psi}$. Consequently, when $\boldsymbol{\zeta}$ and $\boldsymbol{\psi}$ are given, the PoF is a deterministic value, and the probability becomes a Dirac-delta measure as

$$
f(p_F|\boldsymbol{\zeta},\boldsymbol{\psi},{}^*\mathbf{x}) = \delta[p_F - p_F(\boldsymbol{\zeta},\boldsymbol{\psi})].
\tag{14}
$$

The second and third conditional probabilities are obtained in Sections 3.2 and 3.3.

3.2 Joint PDF of input distribution parameters

The last conditional probability in (11) is the joint PDF of input distribution parameters $\boldsymbol{\Psi}$ with given input data set ${}^*\mathbf{x}$. The exact joint PDF of $\boldsymbol{\Psi}$ is not known unless population data, which is the test data of all subjects, of the input random variable is provided. In fact, if input data set ${}^*\mathbf{x}$ is population data, we can obtain the exact value of $\boldsymbol{\Psi}$, so that the joint PDF becomes a Dirac delta measure. If not, all we can obtain is an approximated joint PDF of $\boldsymbol{\Psi}$. Hence, the approximated joint PDF of $\boldsymbol{\Psi}$ is obtained in this section. As indicated earlier, input mean μ_i and variance σ_i^2 are used as input distribution parameters. In addition, they are expressed using capital symbols M_i and Σ_i^2, respectively, because of their random features.

The joint PDF of input distribution parameters $\boldsymbol{\Psi}$ is a product of joint PDFs of input mean M_i and input variance Σ_i^2. M_i and Σ_i^2 of each input random variable have separate joint PDFs. Therefore, the joint PDF of $\boldsymbol{\Psi}$ can be expressed as

$$
f(\boldsymbol{\psi}|{}^*\mathbf{x}) = \prod_{i=1}^{N} f\left(\mu_i, \sigma_i^2 | {}^*\mathbf{x}_i\right).
\tag{15}
$$

The central limit theorem is a widely used method for obtaining the PDF of the input mean M_i with the given input data ${}^*\mathbf{x}$. Though the central limit theorem produces the PDF of the input mean under the assumption that the input data follow Normal distribution, it produces a well-approximated PDF of the input mean when the input data follow other distributions. In the same sense, the joint PDF of the input mean M_i and variance Σ_i^2 are obtained using Bayes' theorem under the assumption that the given input data ${}^*\mathbf{x}$ follows Normal distribution in this paper. This does not mean that the input distribution types $\mathbf{Z}$ are always Normal distributions; this is only an intermediate assumption to find the approximate joint PDF of the input mean M_i and variance Σ_i^2. Also, the non-informative prior, which means that there is no information except the given input data ${}^*\mathbf{x}$, is used for Bayes' theorem. It will be shown that the result of Bayes' theorem is the same as the one from the central limit theorem.

Under the Normality assumption described above and with the non-informative prior, the input variance Σ_i^2, 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)

$$
\begin{aligned}
\Sigma_i^2 \big|{}^*\mathbf{x}_i &\sim \text{Inv-}\chi^2\left(ND-1, s_i^2\right) \\
&= \text{IG}\left(\frac{ND-1}{2},\ \frac{(ND-1)s_i^2}{2}\right)
\end{aligned}
\tag{16}
$$

where the sample variance s_i^2 can be calculated as

$$
s_i^2 = \frac{1}{ND-1} \sum_{m=1}^{ND} \left({}^*\tilde{x}_i^{(m)}\right)^2 = \frac{1}{ND-1} {}^*\tilde{\mathbf{x}}_i^{\mathrm{T}} {}^*\tilde{\mathbf{x}}_i.
\tag{17}
$$

s_i^2 is constant in the RBDO process because the amount of data ND and the dispersion of input data ${}^*\tilde{\mathbf{x}}_i$ are invariant. Therefore, the inverse-gamma distribution of Σ_i^2 in (16) does not change during the RBDO process because the parameters for the distribution are s_i^2 and ND. The distribution has larger uncertainty when input data with smaller ND are provided. The larger uncertainty makes the PoF more uncertain. Eventually, the enlarged uncertainty of the PoF reduces the conservativeness level in (13).

The input mean M_i of the i-th independent variable X_i follows Normal distribution based on the non-informative

prior, given input variance σ_i^2 and data $^*\mathbf{x}_i$ as (Gelman et al. 2004)

$$M_i|\sigma_i^2, {}^*\mathbf{x}_i \sim \mathcal{N}\left({}^*\overline{x}_i, \sigma_i^2/ND\right) \qquad (18)$$

where $^*\overline{x}_i$ is mean of data subset $^*\mathbf{x}_i$ as defined in (8). In the distribution of M_i, the realization σ_i^2 of input variance Σ_i^2 is required, which means the distribution of Σ_i^2 in (16) is also used to derive the distribution of M_i in (18). The distribution of M_i is the same as the distribution from the central limit theorem; hence the distribution of M_i as well as the distribution of Σ_i^2, which affects the distribution of M_i, are reasonable and trustworthy. If X_i is related to a design variable d_i, (18) can be expressed as

$$M_i|\sigma_i^2, {}^*\mathbf{x}_i \sim \mathcal{N}\left(d_i, \sigma_i^2/ND\right) \qquad (19)$$

because the sample mean of data set $^*\overline{x}_i$ changes to d_i. It is noted that the design variable d_i is deterministic for the purpose of design optimization, and the input uncertainty is considered in M_i by treating it as a random variable. It can be seen that in (18) and (19), smaller ND makes the input mean M_i have larger variability, so the conservativeness level of the PoF decreases.

Finally, the joint PDF of input random variable and variance can be derived using the distributions obtained as

$$f\left(\mu_i, \sigma_i^2\big|{}^*\mathbf{x}_i\right) = f\left(\mu_i|\sigma_i^2, {}^*\mathbf{x}_i\right)f\left(\sigma_i^2\big|{}^*\mathbf{x}_i\right) \qquad (20)$$

where $f(\sigma_i^2|^*\mathbf{x}_i)$ and $f(\mu_i|\sigma_i^2, {}^*\mathbf{x}_i)$ are the PDF forms of (16) and (19), respectively. Finally, (20) can be used to obtain the joint PDF of input distribution parameters Ψ in (15).

3.3 Probability mass function of input distribution types

The probability mass function of an input distribution types $\mathbf{Z}$ with the given input data $^*\mathbf{x}$ and given parameters ψ is obtained using Bayes' theorem as

$$P\left(\zeta|\psi, {}^*\mathbf{x}\right) = \frac{P({}^*\mathbf{x}|\zeta, \psi)\,P(\zeta|\psi)}{\sum_{\mathbf{Z}}P({}^*\mathbf{x}|\zeta, \psi)\,P(\zeta|\psi)}$$
$$= \frac{L({}^*\mathbf{x};\zeta, \psi)\,P(\zeta|\psi)}{\sum_{\mathbf{Z}}L({}^*\mathbf{x};\zeta, \psi)\,P(\zeta|\psi)} \qquad (21)$$

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

$$L\left({}^*\mathbf{x};\zeta, \psi\right) = \prod_{i=1}^{N}\prod_{m=1}^{ND} f_{X_i}\left({}^*x_i^{(m)}\Big|\zeta_i; \mu_j; \sigma_j^2\right). \qquad (22)$$

The term $P(\zeta|\psi)$ in (21) is a constant under the assumption that there is no prior information. This assumption means that

all candidate distribution types are equally probable before the analysis using the given input data $^*\mathbf{x}$. Then, (21) can be simplified as

$$P\left(\zeta|\psi, {}^*\mathbf{x}\right) = \frac{L({}^*\mathbf{x};\zeta, \psi)}{\sum_{\mathbf{Z}}L({}^*\mathbf{x};\zeta, \psi)}. \qquad (23)$$

There are many marginal distributions; however, it is impossible to cover all the types in the evaluation of (23). Hence, it is reasonable to set combinations of marginal distribution types and then evaluate the probability of each combination. In this paper, seven marginal distribution types with two distribution parameters are used. Probability density functions of the selected types are listed in Appendix A.

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

3.4 Calculation of conservativeness level of PoF

Since all terms to evaluate the probability of the PoF in (11) are now available in (14), (15), and (23), the conservativeness level of the PoF in (13) at a PoF value p_F can be calculated. When p_F is given, (14) needs to be evaluated to calculate the conservativeness level. However, it is too complicated to solve (14) analytically because it involves the Dirac-delta measure. In addition, as the probability of the PoF is very likely not a standard distribution type, FORM, SORM or DRM is not applicable to the conservativeness level estimation. Therefore, the conservativeness level is calculated numerically using MCS as (Rubinstein and Kroese 2008)

$$\begin{aligned}
&F_{P_F}\left(p_F\big|{}^*\mathbf{x}\right)\\
&\cong \frac{1}{NMCS_{\mathbf{Z}}\,NMCS_{\Psi}}\int_0^{p_F}\sum_{n=1}^{NMCS_{\Psi}}\sum_{m=1}^{NMCS_{\mathbf{Z}}} f\left(\phi|\zeta^{(m)}, \psi^{(n)}, {}^*\mathbf{x}\right)d\phi\\
&= \frac{1}{NMCS_{\mathbf{Z}}\,NMCS_{\Psi}}\int_0^1\sum_{n=1}^{NMCS_{\Psi}}\sum_{m=1}^{NMCS_{\mathbf{Z}}} I_{[0,p_F]}(\phi)\,\delta\left[\phi-p_F\left(\zeta^{(m)}, \psi^{(n)}\right)\right]d\phi\\
&= \frac{1}{NMCS_{\mathbf{Z}}\,NMCS_{\Psi}}\sum_{n=1}^{NMCS_{\Psi}}\sum_{m=1}^{NMCS_{\mathbf{Z}}} I_{[0,p_F]}\left[p_F\left(\zeta^{(m)}, \psi^{(n)}\right)\right]
\end{aligned}$$
$$(24)$$

where $I_{[0,p_F]}(\phi)$ is an indicator function, the value of which is 1 when ϕ is between 0 and p_F, and 0 otherwise. $\zeta^{(m)}$ and $\psi^{(n)}$ are the m-th realization of $(\mathbf{Z}\big|\psi, {}^*\mathbf{x})$ and the n-th realization of $(\Psi|^*\mathbf{x})$, respectively. $NMCS_{\Psi}$ and $NMCS_{\mathbf{Z}}$ are the MCS sample sizes for Ψ and $\mathbf{Z}$, respectively. The overall procedure to evaluate (24) is shown in Fig. 2.

Fig. 2 box 1: (Insufficient) input data $^*\mathbf{x}$, design $\mathbf{d}$

Fig. 2 box 2: Draw $NMCS_\Psi$ sets of input variance
< Eq. (16) >

Fig. 2 box 3: Draw one set of input mean giving each set of input variance < Eq. (19) >

Fig. 2 box 4: Draw $NMCS_\mathbf{Z}$ sets of input distribution types giving each set of input parameters
< Eq. (23) >

Fig. 2 box 5: Calculate the PoF at each of drawn $(NMCS_\Psi \times NMCS_\mathbf{Z})$ sets
(Reliability analysis)

Fig. 2 box 6: Check the PoF is smaller than the given value of p_F
(Conservativeness level)

Fig. 2 Flowchart of conservativeness level calculation

4 Design sensitivity of conservativeness level

The conservativeness level calculated in Section 3.4 can be used as a constraint in the RBDO process. The RBDO process is called confidence-based RBDO (C-RBDO) because we can have confidence that its optimum design has a certain amount of conservativeness even when there is limited input data. The constraint can be expressed as

$$F_{P_F}\left(p_F^{Tar}\big|^*\mathbf{x}\right) \geq CL^{Tar} \tag{25}$$

where p_F^{Tar} and CL^{Tar} are the target PoF and the target conservativeness level, respectively, for the constraint. By using the two target values, C-RBDO is able to secure user-specified conservativeness even with a finite amount of data.

4.1 Design sensitivity

The design sensitivity of the conservativeness level is developed to provide an accurate and efficient direction for the design search in the C-RBDO process. The finite difference method (FDM) could be used to calculate the design sensitivity; however, it requires a great deal of computational time to calculate accurate design

sensitivity. Hence, an analytical design sensitivity is necessary to perform C-RBDO efficiently.

The derivative of (13) with respect to a design variable d_i yields

$$
\begin{aligned}
\frac{\partial}{\partial d_i} F_{P_F}\left(p_F\big|^*\mathbf{x}\right) &= \frac{\partial}{\partial d_i}\int_0^{p_F}\sum_{\mathbf{Z}}\int_{\Omega_\Psi} f(\phi,\,\zeta,\,\psi|^*\mathbf{x})\,d\psi\quad d\phi \\
&= \int_0^{p_F}\sum_{\mathbf{Z}}\left[\int_{\Omega_\Psi} f(\phi,\,\zeta,\,\psi|^*\mathbf{x})\frac{\partial}{\partial d_i}\{\ln P(\zeta|\psi,^*\mathbf{x}) + \ln f(\psi|^*\mathbf{x})\}\,d\psi\right]d\phi.
\end{aligned}
\tag{26}
$$

Compared with (13), there are two additional terms in (26). The first additional term is the log-derivative of the probability mass function of the input distribution types. This term is derived analytically in Section 4.2 and is defined for now as

$$\frac{\partial}{\partial d_i}\ln P\left(\zeta|\psi,^*\mathbf{x}\right)\equiv S_\mathbf{Z}\left(\zeta,\psi,^*\mathbf{x},d_i\right) \tag{27}$$

As discussed earlier, a design change in the optimization process does not affect the probability of the input variance; therefore, the probability is independent of the design variable d_i. Hence, the second additional term in (26), which is the log-derivative of the joint PDF of input distribution parameters, is derived when d_i is the design variable that corresponds to X_i:

$$
\begin{aligned}
\frac{\partial}{\partial d_i}\ln f\left(\psi|^*\mathbf{x}\right) &= \frac{\partial}{\partial d_i}\ln f\left(\mu_i|\sigma_i^2,^*\mathbf{x}\right) \\
&= \frac{ND(\mu_i - d_i)}{\sigma_i^2}\equiv S_\Psi\left(\mu_i,\sigma_i^2,d_i,ND\right).
\end{aligned}
\tag{28}
$$

Although the additional terms in (26) have analytical expression in (27) and (28), the design sensitivity cannot be calculated directly. The reason is the same as why (13) is evaluated using the MCS method in (24) of Section 3.4. Hence, the design sensitivity in (26) is calculated using the MCS method as well. The design sensitivity for the design variable d_i is

$$
\begin{aligned}
&\frac{\partial}{\partial d_i} F_{P_F}\left(p_F\big|^*\mathbf{x}\right) \\
&\cong \frac{1}{NMCS_\mathbf{Z}\,NMCS_\Psi}\sum_{n=1}^{NMCS_\Psi}\sum_{m=1}^{NMCS_\mathbf{Z}}\left\{I_{[0,p_F]}\left[p_F\left(\zeta^{(m)},\psi^{(n)}\right)\right]\right. \\
&\left.\times\left[S_\mathbf{Z}\left(\zeta^{(m)},\psi^{(n)},^*\mathbf{x},d_i\right) + S_\Psi\left(\mu_i^{(n)},\sigma_i^{2(n)},d_i,ND\right)\right]\right\}.
\end{aligned}
\tag{29}
$$

(29) is quite similar to the conservativeness level of the PoF in (24). Only the additional terms in (27) and (28) 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 conservativeness level of the PoF. 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 amount of data.

4.2 Log-derivative of probability mass function of input distribution types

In Section 4.1, (27), which is the log-derivative of the probability mass function of the input distribution types to the design variable d_i, is required for the design sensitivity of the conservativeness level in (29). The easiest way to calculate (27) is with the FDM because it only requires evaluations of the probability mass function of the input distribution types in (23) at the perturbed design and the current design. However, the FDM could be inaccurate when appropriate perturbation size is not provided. Moreover, there may be no unique perturbation size that is appropriate for all candidate distribution types. Hence, determining perturbation size could cause unnecessary difficulty and inaccuracy when calculating (27) using the FDM.

If analytical expressions of marginal PDFs are available, (27) could be derived analytically by taking the log-derivative of (23) with respect to the design variable d_i. First, the expression of data in (6), (9) and (10) are recalled:

$$^*\mathbf{x}_i = \mathbf{d}_i + {}^*\tilde{\mathbf{x}}_i \tag{9}$$

where

$$^*\mathbf{x}_i = \begin{bmatrix} {}^*x_i^{(1)} & {}^*x_i^{(2)} & \cdots & {}^*x_i^{(ND)} \end{bmatrix}^T, \tag{6}$$

$$\mathbf{d}_i = \begin{bmatrix} d_i & d_i & \cdots & d_i \end{bmatrix}^T. \tag{10}$$

(9) indicates that each data point $^*x_i^{(j)}$ is a function of the design variable d_i because the input data subset $^*\mathbf{x}_i$ moves exactly the same amount as d_i moves while ${}^*\tilde{\mathbf{x}}_i$ is invariant in an optimization process. Let $h(^*\mathbf{x})$ be a general function of the input data $^*\mathbf{x}$. Then, $h(^*\mathbf{x})$ contains the input data points $^*x_i^{(1)}, \ldots, {}^*x_i^{(ND)}$ in its expression. Therefore, the derivative of $h(^*\mathbf{x})$ with respect to d_i is the summation of the derivative of the function with respect to data $^*x_i^{(j)}$:

$$\frac{\partial}{\partial d_i} h(^*\mathbf{x}) = \sum_{j=1}^{ND} \frac{\partial}{\partial {}^*x_i^{(j)}} h(^*\mathbf{x}). \tag{30}$$

The log-derivative of the probability mass function of the input distribution types in (23) yields

$$\frac{\partial}{\partial d_i} \ln P(\zeta | \psi, {}^*\mathbf{x})$$
$$= \frac{\partial}{\partial d_i} \left[\ln L(^*\mathbf{x}; \zeta, \psi) - \ln \sum_{\mathbf{z}} L(^*\mathbf{x}; \zeta, \psi) \right]$$
$$= \frac{\partial}{\partial d_i} \ln L(^*\mathbf{x}; \zeta, \psi) - \frac{1}{\sum_{\mathbf{z}} L(^*\mathbf{x}; \zeta, \psi)} \sum_{\mathbf{z}} \left[L(^*\mathbf{x}; \zeta, \psi) \frac{\partial}{\partial d_i} \ln L(^*\mathbf{x}; \zeta, \psi) \right]. \tag{31}$$

Therefore, the log-derivative of the likelihood function is necessary for (31). The log-derivative is derived using the relationship in (30) as

$$\frac{\partial}{\partial d_i} \ln L(^*\mathbf{x}; \zeta, \psi) = \sum_{m=1}^{ND} \frac{\partial}{\partial {}^*x_i^{(m)}} \ln f_{X_i} \left({}^*x_i^{(m)} \Big| \zeta_i, \mu_j; , \sigma_j^2 \right)$$
$$= \sum_{m=1}^{ND} \frac{1}{f_{X_i} \left({}^*x_i^{(m)} \Big| \zeta_i, \mu_j; , \sigma_j^2 \right)} \frac{\partial f_{X_i} \left({}^*x_i^{(m)} \Big| \zeta_i, \mu_j; , \sigma_j^2 \right)}{\partial {}^*x_i^{(m)}}. \tag{32}$$

In (32), the derivatives of the marginal PDF f_{X_i} are required. The derivatives of commonly used marginal PDFs are derived in Table 1 using the original expressions of marginal PDFs in Appendix A, Table 18. Using the derivatives, (32), the log-derivative of the probability mass function of the input distribution types can be obtained.

5 Numerical example: 2-dimensional mathematical example

In this section, the developed methods for estimation of conservativeness level and its design sensitivity are verified using a 2-D mathematical example. The conservativeness levels using two different data sets are compared to understand the effect of the amount of data. The developed design sensitivity method is then compared with FDM design sensitivity to check its accuracy. In addition, C-RBDO has been performed under

Table 1 Derivative of PDF

Distribution type	df_X/dx
Normal	$-\frac{x-a}{b^2} f_X(x)$
Lognormal	$-\frac{1}{x}\left(1 + \frac{\ln x - a}{b^2}\right) f_X(x)$
Weibull	$\left[\frac{b-1}{x} - \frac{b}{a}\left(\frac{x}{a}\right)^{b-1}\right] f_X(x)$
Gumbel	$\frac{1}{b}\left[\exp\left(-\frac{x-a}{b}\right) - 1\right] f_X(x)$
Gamma	$\left(\frac{a-1}{x} - \frac{1}{b}\right) f_X(x)$
Extreme	$\frac{1}{b}\left\{1 - \exp\left(\frac{x-a}{b}\right)\right\} f_X(x)$
Extreme type-II	$\frac{1}{x}\left\{a\left(\frac{b}{a}\right)^a - a - 1\right\} f_X(x)$

different conditions to understand the performance of the developed method.

For the mathematical example, three performance measures of the 2-D mathematical problem are considered:

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

(33)

where X_1 and X_2 are independent input random variables. The limit states ($G_i = 0$) of (33) are shown in Fig. 3, and $G_i < 0$ refers to the feasible area in this example. In Table 2, a benchmark distribution, which will be used as a true distribution, of X_1 and X_2 is shown. In addition, contours of 95.5 % probability density of the benchmark distribution at $\mathbf{d}^0 = [d_1, d_2]^T = [5, 5]^T$ and $\mathbf{d}^1 = [4.7, 1.6]^T$ are shown in Fig. 3.

5.1 Conservativeness level calculation

To verify the effectiveness of the proposed method for conservativeness level estimation, the conservativeness level of performance measures in (33) is calculated. Because the method is for a limited-data problem, 10 pairs of input data are randomly drawn from the benchmark distribution in Table 2 using the "normrnd"
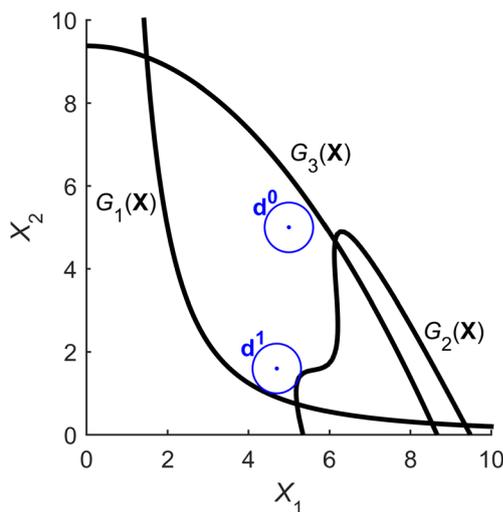
Fig. 3 Limit states of 2-D mathematical example and 95.5 % contour of benchmark distribution at $\mathbf{d}^0$ and $\mathbf{d}^1$

Table 2 Benchmark input distribution

Random variable	Marginal distribution	Mean	Variance
X_1	Normal	d_1	0.3^2
X_2	Normal	d_2	0.3^2

function in MATLAB. The $^*\tilde{\mathbf{x}}_1$ and $^*\tilde{\mathbf{x}}_2$ of the drawn data are shown as follows:

$$
^*\tilde{\mathbf{x}}_1 = \begin{bmatrix} -0.2705 \\ 0.1054 \\ 0.2906 \\ 0.8496 \\ -0.1261 \\ 0.1301 \\ 0.0492 \\ -0.5060 \\ -0.0578 \\ -0.4645 \end{bmatrix}, \quad
^*\tilde{\mathbf{x}}_2 = \begin{bmatrix} -0.2501 \\ -0.2259 \\ 0.2583 \\ 0.2341 \\ 0.2365 \\ 0.0128 \\ -0.5509 \\ 0.0266 \\ 0.3005 \\ -0.0419 \end{bmatrix}.
$$

(34)

As explained in Section 3.3, combinations or candidate distributions are necessary to evaluate the conservativeness level. In this example, for the input distribution type $\mathbf{Z}$, the 20 candidate types listed in Table 3 are

Table 3 Candidate input distribution types at $\mathbf{d}^1$ with 10 data pairs

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

used. Each input random variable X_i can have seven marginal distribution types (Normal, Lognormal, Weibull, Gumbel, Gamma, Extreme, and Extreme type-II). Hence, in this bivariate problem, there are 49 ($= 7 \times 7$) combinations. However, considering all 49 combinations is ineffective and inefficient because many of them have very small probability. Hence, it is reasonable to narrow the number of candidates down. Among the 49 combinations, the 20 most probable types (according to their likelihoods) are selected at design point $\mathbf{d}^1 = [4.7, \ 1.6]^T$ using the drawn data and the sample variances of the data. The most probable type (Gumbel – Extreme) has 9.45 % probability mass, and the 20th most probable type (Extreme type-II – Lognormal) has 1.89 %. Since they have meaningful probability mass values, those 20 are selected. Both numbers of MCS samples, $NMCS_\Psi$ and $NMCS_\mathbf{Z}$, are set to 20,000. Finally, following the procedure shown in Fig. 2, the conservativeness level of the PoF is calculated with the 10 drawn data pairs and the 20 candidate distribution types, and the obtained result is shown in Fig. 4.

Using the benchmark distribution in Table 2, the PoFs at $\mathbf{d}^1$ are 1.79 %, 1.49 % and 0 % for G_1, G_2 and G_3, respectively. However, when only 10 data pairs are available, the conservativeness levels at $p_F = 2.275\%$ are 23.9 % and 41.4 % for G_1 and G_2, respectively, as shown in Fig. 4. Even though $p_F = 2.275\%$ is larger than the benchmark PoFs (1.79 %, 1.49 % and 0 %), the conservativeness level is less than 50 %. This means that, at the design point $\mathbf{d}^1$, we have less than 50 % confidence that the design will meet the target PoF of 2.275 % due to the limited input data. Consequently, if only the 10 data pairs are available, conservativeness has to be applied to the design to assure that the target PoF is satisfied. For the third constraint G_3, the conservativeness level increases rapidly from zero at $p_F = 0$ to a 99.9 % conservativeness level at $p_F = 2.275\%$. This is a reasonable result because the limit state of G_3 is far enough from $\mathbf{d}^1$ and is thus a very conservative design for G_3, as shown in Fig. 3. The sample variances of X_1 and X_2 data in (34) are 0.3945^2 and 0.2764^2, respectively. Hence, there is more input uncertainty in the X_1 direction than in the X_2 direction. As shown in Fig. 3, G_2 is mainly affected by uncertainty in the X_1 (horizontal) direction, while G_1 is affected by uncertainty in both the X_1 and X_2 (vertical) directions. This is why the conservativeness level for G_1 (41.4 %) is larger than the one for G_2 (23.9 %). In addition, we can see that the statistical information in data as well as the number of data affects conservativeness level.

To understand how the amount of data affects the conservativeness level of the PoF, 100 pairs of data are again drawn from the benchmark distribution in Table 2. Using the drawn data, 20 new candidate distribution types are chosen according to the same procedure used for the 10 data pairs. Using the 100 data pairs and 20 candidate distribution types, the conservativeness level is evaluated as shown in Fig. 5. At the same design point $\mathbf{d}^1$, the conservativeness levels at $p_F = 2.275\%$ are 59.9 % and 63.3 % for G_1 and G_2, respectively. Therefore, the result agrees with the expectation that the conservativeness level is more assured when more data is available. For the inactive constraint G_3, the conservativeness level is 100 %, which has not changed much from the case of the 10 data pairs. This is reasonable because the conservativeness level is already maximized, even with the 10 data pairs. The conservativeness level for G_1 (59.9 %) is very similar to the one for G_2 (63.3 %). Because the sample variances
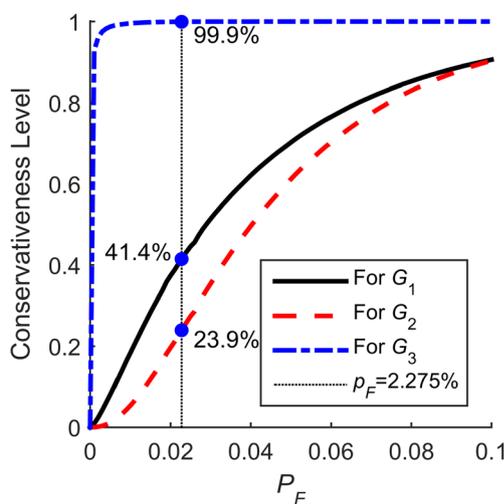
Fig. 4 Conservativeness level of PoF at $\mathbf{d}^1$ with 10 data pairs
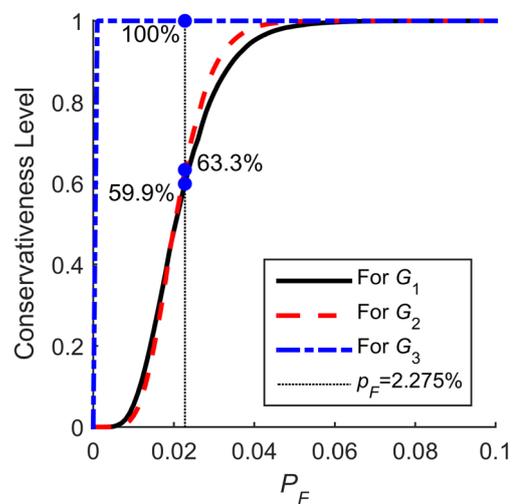
Fig. 5 Conservativeness level of PoF at $\mathbf{d}^1$ with 100 data pairs

of X_1 and X_2 data are 0.3202^2 and 0.3123^2, respectively, the uncertainty in both the X_1 and X_2 directions are similar. Therefore, the conservativeness levels are similar to each other. More data indicates that better information is in the data set. It can be seen that the sample variances of the 100 data pairs (0.3202^2 and 0.3123^2) are much closer to the ones of the benchmark distribution in Table 2 (0.3^2 and 0.3^2) than the ones of 10 data pairs (0.3945^2 and 0.2764^2). Therefore, it is anticipated that more data will increase the conservativeness level rapidly due to both larger number of data and better information in it. However, a design point with the conservativeness levels of 59.9 % and 63.3 % is still far from a safe and reliable design point.

Throughout the two examples with 10 and 100 data pairs, it is shown that the design point $\mathbf{d}^1$, which was safe and reliable with the benchmark distribution, cannot ensure 2.275 % target PoF when insufficient data or even when enough input data is provided. Therefore, the conservativeness level of the PoF should be incorporated in the RBDO for a limited amount of input data. It also can be seen that the developed method appropriately considers amount of provided data when estimating the conservativeness level.

5.2 Accuracy of design sensitivity of conservativeness level

In this section, the accuracy of the derived design sensitivity of the conservativeness level in (29) is tested using the performance measures in (33) and the same 10 input data pairs used in Section 5.1. The conservativeness level is calculated at $p_F = 10\%$ to obtain fast convergence of FDM design sensitivity. The design sensitivity is calculated at $\mathbf{d}^2 = [5, 1.5]^T$. Candidate distribution types are selected to cover several marginal distribution types as shown in Table 4.

The design sensitivity of the conservativeness level is calculated and compared with the FDM result. The FDM result is carried out by using central finite difference and perturbing each design variable by 0.1 % forward and backward. Eight million MCS samples for the input distribution parameters and input distribution types are used ($NMCS_\Psi = NMCS_\mathbf{Z} = 8,000,000$). At the same time, the developed design sensitivity is calculated using only 20,000 MCS samples for both $NMCS_\Psi$ and $NMCS_\mathbf{Z}$. Because the FDM requires evaluations at both the forward and backward perturbed designs in each input random variable, a total of 32 million MCS samples is actually used to calculate the design sensitivity using the FDM in the 2-D mathematical problem. Thus, only 0.0625 % of the MCS samples for the FDM are used for the developed design sensitivity.

The accuracy check result is summarized in Table 5. The agreement of the developed design sensitivity compared to the FDM result varies from 94.2 % to 100.9 %. This indicates that the developed design sensitivity agrees with the FDM result. The conservativeness level values at the forward and backward perturbed designs have similar values; for example, conservativeness levels of G_1 at the forward and backward perturbed designs are -0.043815 and -0.041346, respectively. The finite difference (subtraction of the values) loses the first significant digit. Hence, the conservativeness level value at each perturbed design should have many significant digits for evaluating accurate FDM design sensitivity. This is why FDM design sensitivity uses more MCS samples than the developed design sensitivity method. The developed method calculates design sensitivity in a semi-analytical way, so that there is no loss of significant digits. Therefore, it uses fewer MCS samples than the FDM. Moreover, it does not require perturbation size, which may cause trouble in FDM design sensitivity when it is not selected reasonably. Hence, the developed design sensitivity method is as accurate as the FDM design sensitivity, does not require determination of perturbation size, and is much more efficient.

5.3 Confidence-based RBDO

In Sections 5.1 and 5.2, the developed estimation methods for conservativeness level and for its design sensitivity have been verified. In this section, the design optimization process (i.e., C-RBDO) has been performed using the developed methods.

Table 4 Candidate input distribution types for design sensitivity test

No.	X_1	X_2
1	Normal	Lognormal
2	Lognormal	Normal
3	Extreme	Lognormal
4	Extreme	Normal
5	Normal	Gumbel

Table 5 FDM and developed design sensitivity of conservativeness level

Design variable	Design sensitivity	G_1	G_2
d_1	FDM	-0.2469	2.0047
	Developed	-0.2325	2.0232
	Agreement	94.2 %	100.9 %
d_2	FDM	-0.5563	-1.5767
	Developed	-0.5478	-1.5545
	Agreement	98.5 %	98.6 %

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

$$\text{minimize} \quad cost\,(\mathbf{d}) = -\frac{(d_1 + d_2 - 10)^2}{30} - \frac{(d_1 - d_2 + 10)^2}{120}$$

$$\text{subject to} \quad F_{P_{F_j}}\left(p_F^{Tar} = 2.275\%\Big|^*\mathbf{x}\right) \geq 90\%, \quad j = 1, 2, 3$$

$$\mathbf{d}^L < \mathbf{d} < \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^2, \quad \text{and} \quad \mathbf{X} \in \mathbb{R}^2$$

$$(35)$$

where $F_{P_{F_j}}$ is the conservativeness level of the performance measure G_j in (33), $\mathbf{d}^L = [0, 0]^T$ and $\mathbf{d}^U = [10, 10]^T$. It is noted that the cost function in (35) is a deterministic function. For the C-RBDO process, the 10 data pairs in (34) are used. In addition, the optimization is also conducted for the 20 data pairs drawn from the benchmark distribution in Table 2. The $^*\tilde{\mathbf{x}}_1$ and $^*\tilde{\mathbf{x}}_2$ of the 20 data pairs are shown as follows.

$$^*\tilde{\mathbf{x}}_1 = \begin{bmatrix} -0.4224 \\ 0.1951 \\ -0.1442 \\ -0.4430 \\ 0.1225 \\ 0.4416 \\ 0.7250 \\ 0.0070 \\ -0.0581 \\ 0.1192 \\ -0.1191 \\ -0.6024 \\ 0.3532 \\ 0.1952 \\ -0.1171 \\ 0.4437 \\ -0.1119 \\ -0.9374 \\ 0.3611 \\ -0.0079 \end{bmatrix}, \quad ^*\tilde{\mathbf{x}}_2 = \begin{bmatrix} 0.1929 \\ 0.2856 \\ 0.5433 \\ -0.2927 \\ -0.1043 \\ 0.0144 \\ -0.2329 \\ -0.0825 \\ -0.2743 \\ 0.4301 \\ 0.0066 \\ 0.7394 \\ -0.3420 \\ -0.4818 \\ 0.4609 \\ 0.3256 \\ -0.2971 \\ -0.3731 \\ -0.2917 \\ -0.2263 \end{bmatrix}. \quad (36)$$

Confidence-based RBDO is computationally expensive compared to deterministic design optimization (DDO) and conventional RBDO. Therefore, for numerical efficiency, DDO and conventional RBDO are carried out in advance of C-RBDO. In this way, the computational effort for design iterations of the C-RBDO process is minimized. Deterministic design optimization is launched from the initial design $\mathbf{d}^0 = [5, 5]^T$. Because DDO does not consider the input uncertainty or variability, the optimum design of DDO is the same as $\mathbf{d}^{DDO} = [5.1969, 0.7405]^T$, regardless of given input data. From $\mathbf{d}^{DDO}$, the conventional RBDO is performed using the most likely distribution types and sample variances, which can be obtained using the given input data. It is noted that the means of input random variables are the design variables in the conventional RBDO process.

Finally, the C-RBDO is launched at the conventional RBDO optimum. A sequential quadratic programming (SQP) algorithm is used for the C-RBDO with convergence criteria 0.001 for first-optimality, constraint (conservativeness level) and design movement. Both $NMCS_{\Psi}$ and $NMCS_{\mathbf{Z}}$ are set to 20,000. Assuming that the conventional RBDO optimum design is close to the C-RBDO optimum design, 20 candidate input distribution types are determined among 49 combinations at the conventional RBDO optimum using the given input data. The C-RBDO process is carried out using an Intel Xeon E5-2690 processor with 16GB memory and the computational time is approximately 2.5 h. As a benchmark, the conventional RBDO optimum design based on the benchmark input distribution in Table 2 is obtained as well. The results of C-RBDO and the benchmark design are summarized in Table 6.

The C-RBDO optimum designs satisfy the given target confidence level of 90 % for both the 10 and 20 input data pairs. Both 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 optimum designs, the conservativeness levels for G_3 are almost 100 %, as shown in Table 6. Comparing both C-RBDO optimum designs with the benchmark design, the optimum cost is increased. For the case with 10 input data pairs, 12.4 % more cost (−1.6722 vs. −1.9089) is required to meet the target conservativeness level since significant input uncertainty arises due to the limited data. Hence, a more conservative design is obtained in C-RBDO using higher optimum cost. However, in the case with 20 data pairs, the optimum cost value increases only 8.3 % compared to the benchmark (−1.7500 vs. −1.9089) as it has more data. That is, less input uncertainty is induced by more data, so a less conservative design is obtained than for the case with 10 data pairs to satisfy 90 % target conservativeness level.

In real engineering applications with limited input data, the C-RBDO results in Table 6 are all that we can obtain. However, being this is a numerical example, a conventional reliability analysis can be performed at the C-RBDO optimum designs using the benchmark distribution in Table 2. The reliability analysis will reveal how conservative the C-RBDO optimum designs are. The calculated results are summarized in Table 7. The optimum design of the case with 10 data pairs has PoFs of 0.051 % and 0.023 % for G_1 and G_2, respectively, which are 1.0 % (0.023/2.219) of the result of the benchmark design. However, in the case with 20 data pairs, the PoFs are only 0.156 % and 0.245 % for G_1 and G_2, respectively, which are approximately 6.7 % (0.156/2.319) of the result of the benchmark design. Therefore, it can be concluded that the number of data is a crucial factor in the input uncertainty,

Table 6 Optimal design of C-RBDO result and benchmark design

Method	Data	Optimum design		Cost	Conservativeness level (%)		
		d_1	d_2		G_1	G_2	G_3
C-RBDO	10	4.5516	2.0851	−1.6722	90.0	90.4	99.9
	20	4.6658	1.9026	−1.7500	89.8	89.8	100.0
Benchmark design		4.7334	1.5498	−1.9089	–	–	–

especially when the data is limited. The trend will be shown further in Section 5.4

The result in Table 7 may seem to be overly conservative. However, the conservativeness is inevitable when the insufficient input data is given for the input probabilistic model. In Ref (Noh et al. 2011a), sample variance is calculated from data and is enlarged according to its confidence level to consider the input uncertainty due to limited number of data. For 10 and 20 data pairs, the sample variance is increased by 170.7 % and 87.8 %, respectively, for 90 % confidence level, which is the same level used for the target conservativeness level. Using the enlarged sample variance, the conventional RBDO is performed targeting 2.275 % PoF, which is the same target PoF in C-RBDO as well. The result is summarized in Table 8. It can be seen that cost is increased by 6.0 % (−1.6722 .vs. –1.5715) and 2.6 % (−1.7500 .vs. –1.7039) compared to the C-RBDO result. Conventional reliability analysis using the benchmark distribution is also shown in Table 8. All the PoF values in Table 8 are much smaller than the ones in Table 7. It is noted that the enlarged sample variance considers only uncertainty in the variance of the input random variable, while C-RBDO considers the uncertainty of both mean and variance. Therefore, we can see that the C-RBDO result is more appropriate and also produces a reliable design even with a limited number of input data.

5.4 Convergence test of C-RBDO

A large number of data ND would bring three aspects to the developed estimation method for the conservativeness level.

First, large ND would reduce the input uncertainty in input distribution parameters as shown in (16) and (19). Second, large ND would provide accurate estimation of sample variance s_i^2, which is used for distribution of the input variance in (16) that affects (19) as well. Third, large ND would suggest a correct input distribution type in (23) by providing distinctive likelihood value in (22). All three aspects eventually reduce the input uncertainty, so that the C-RBDO with large ND would be close to the benchmark design in Table 6. Therefore, in the C-RBDO process, the only parameter that determines whether or not the C-RBDO optimum converges to the benchmark design is ND. Changing the target conservativeness level may expedite the convergence, but the opposite is also possible. In this section, the number of data in the 2-D mathematical example is increased to 1000 to see if the C-RBDO optimum moves closer to the benchmark design.

For the convergence test, 50, 100, 200, 500 and 1000 pairs of data are drawn from the benchmark distribution in Table 2. The same procedure used in Section 5.3 has been followed, with the same target values for the large number of input data cases. First, the initial design is set to the DDO optimum design $\mathbf{d^{DDO}}$, as the DDO result is the same regardless of the number of data. Second, conventional RBDO is performed for each case with the most likely distribution types and the sample variances. Then, the 20 candidate distribution types are determined at the conventional RBDO optimum, and C-RBDO is performed from the conventional RBDO optimum using the candidate types.

The C-RBDO optimum designs using 50, 100, 200, 500 and 1000 pairs of data are shown graphically in Fig. 6. Because the optimum designs of 200, 500 and 1000 data pairs are all in gray area of Fig. 6a, the area is magnified in Fig. 6b. In addition, C-RBDO optimum designs of 10 and 20 data pairs and the benchmark design obtained in Section 5.3 are shown in Fig. 6. The benchmark design is the RBDO optimum design if there is no input uncertainty. Hence, we can expect that the C-RBDO optimum design with larger input data will approach closer to the benchmark design as smaller input uncertainty is in the input data. The optimum designs for 10 and 20 data pairs are the farthest from the benchmark optimum. Because they have smaller

Table 7 Reliability analysis result using benchmark distribution

Data	Probability of failure (%)		
	G_1	G_2	G_3
10	0.051	0.023	0
20	0.156	0.245	0
Benchmark	2.319	2.219	0

Table 8 RBDO with enlarged sample variance

Data	Optimum Design		cost	Benchmark PoF (%)		
	d_1	d_2		G_1	G_2	G_3
10	4.4243	2.3267	−1.5715	0.011	0.001	0.000
20	4.7002	2.0074	−1.7039	0.046	0.155	0.000

numbers of data compared to the other cases, the input uncertainty in the input data is larger. To compensate for the input uncertainty, the optimum designs are pushed further inside the feasible domain (upper left side in Fig. 6) to maintain a distance far from the limit states of G_1 and G_2. The optimum designs for 50 and 100 data pairs have moderate distance from the benchmark design as well as the limit states of G_1 and G_2. As the number of input data is increased, the input uncertainty is reduced, so that the designs could be closer to the benchmark design. Interestingly, the optimum design of 50 data pairs is closer to the benchmark design than the optimum design of 100 data pairs. The reason is that the sample variances of 50 data pairs are 0.2595^2 and 0.3094^2, while those of 100 data pairs are 0.3202^2 and 0.3123^2. The sample variance of 50 data pairs is smaller than 0.3^2, the variance of the benchmark distribution. The small sample variance of 50 data pairs indicates underestimation of input variability. The underestimation makes the optimum design close to the benchmark design; however, the conventional RBDO with 50 data pairs will violate the target PoF 2.275 % due to the underestimation. Therefore, it can be seen that even 50 data could induce unreliable conventional RBDO optimum design in this case. However, the C-RBDO method pushes the design to a feasible region to compensate for the underestimation, which is the input uncertainty. The same analogy can be applied to the

optimum designs of 200, 500 and 1000 data pair. As shown in Fig. 6b, the optimum design of 1000 data pair is farther from the benchmark design than the ones of 200 and 500 data; however, this is due to the statistical information in the data. The optimum designs of 200, 500 and 1000 data pairs are already very close to the benchmark design as they have a large number of data. In addition, they are reliable designs because they are further inside of the feasible domain than the benchmark design. Hence, we can conclude that C-RBDO optimum design converges to the benchmark design as more data are provided.

5.5 Repeated test of C-RBDO

In this example, the theoretical meaning of 90 % conservativeness level is that there is a 90 % probability that the true PoF is less than the 2.275 % target PoF. However, this definition is not readily comprehensible. The practical meaning could be that at least 90 % of C-RBDO optimum designs satisfy the 2.275 % target PoF if we try C-RBDO with many sets of input data. Each C-RBDO trial will satisfy a 90 % target conservativeness level; however, to understand the meaning of a 90 % conservativeness level, it is necessary to try the same C-RBDO many times. Therefore, C-RBDO has been tested 1000 times with different sets of input data in this section.

There are three types of test cases: 20, 100 and 200 input data pairs. For each case, input data are drawn from the benchmark distribution in Table 2 1000 times. Starting from $\mathbf{d^{DDO}}$, conventional RBDO, C-RBDO and conventional reliability analysis using the benchmark distribution are performed consecutively. The same procedures used in Sections 5.3 and 5.4 are followed. Target conservativeness level and target PoF are set to 90 % and 2.275 %, respectively. This repeated test is performed on a high-performance computing system – Excalibur – in the U.S.

Fig. 6 C-RBDO optimum designs with large number of input data

(a) 10 ~ 100 data   (b) 200 ~ 1000 data
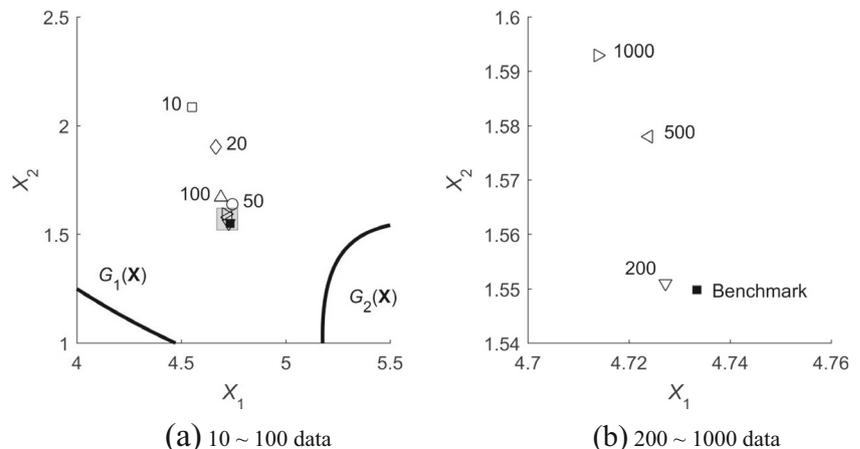
Army Research Laboratory due to its very high computational cost. The repeated test used approximately 50 nodes in parallel. Each node on the Excalibur has 32 cores and 128 GB memory, which result in excellent computation power.

The number of cases that have benchmark PoF larger than 2.275 % are summarized in Table 9. Those cases can be interpreted as safe and reliable cases in real situations. First, it can be seen that the results are in accordance with the practical meaning of conservativeness level. Among 1000 trials, the smallest number is 899, which indicates all cases satisfy a 90 % conservativeness level. Therefore, it can be seen that the developed C-RBDO method works as intended. The performance measure G_3 always has successful results because the optimum designs are far from the limit state of G_3. The other performance measure, G_1, has less success as the number of input data increases. As the number of input data increases, the joint PDF of input distribution parameter in (15) and the probability mass function of input distribution in (23) become more accurate. Therefore, the success rate could approach 90 % as more input data is provided. On the other hand, G_2 acquires more success as more input data is provided. The reason can be seen in Fig. 7, which shows limit state ($G_i = 0$) and two contours of $G_i = -0.8$ and 0.8 for $i = 1$, 2. G_2 has a narrower contour than G_1, which means output uncertainty in G_2 is smaller than in G_1. The developed C-RBDO tries to compensate for the output uncertainties in G_1 and G_2 in each trial using the conservativeness level and given set of input data. However, the compensatory amount of the output uncertainty is similar for G_1 and G_2. As G_1 has larger output uncertainty, it shows a success rate similar to the target conservativeness level, while G_2 has a higher success rate than G_1. This can be further explained using the result of the benchmark PoF.

The statistics of the benchmark PoF are shown in Table 10. The statistics of G_3 are not shown because all of them are close to 0 %. In Table 10, it can be seen that the average benchmark PoFs of G_1 and G_2 are close to each other, which indicates that C-RBDO tries to consider the same amount of output uncertainty in G_1 and G_2. However, the min-max interval of G_2 is smaller than that of G_1 in all data cases because G_2 has less output uncertainty, as explained earlier. Therefore, the success
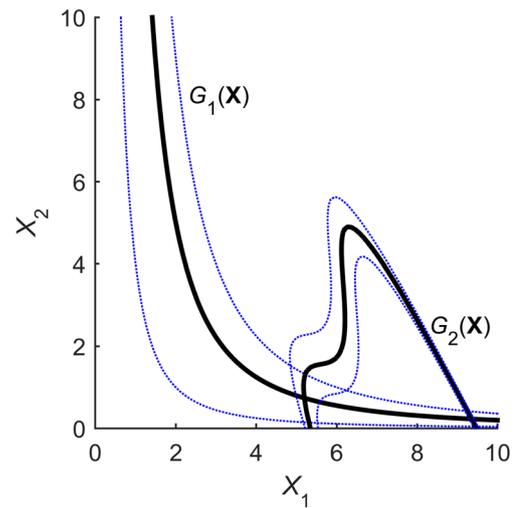
Fig. 7 Contour of G_1 and G_2 in 2-D mathematical example

rate of G_2 is larger than that of G_1 in Table 9. Moreover, as the number of data increases, the maximum benchmark PoF of G_2 decreases. This is why the success rate of G_2 increases as more data are provided. The average values are all increased toward 2.275 % as more data are given. This result coincides with the convergence test in Section 5.4. In addition, it can be seen that more data reduces the size of the min-max interval as less input uncertainty arises. The same feature can also be seen in Fig. 8, which shows 1000 optimum designs in each case. As more data are provided, the distribution of optimum designs is concentrated in a smaller area due to the reduced input uncertainty. It is also shown that most of the optimum designs are distributed well inside the feasible domain, leaving the benchmark design at a corner. Hence, it can be seen that C-RBDO finds conservative design compared to the benchmark design, which is not known during the C-RBDO process.

In this test, it is shown that the C-RBDO optimum designs satisfy the practical meaning of conservativeness level. Hence, it can be concluded that C-RBDO indeed finds reliable design in the presence of limited number of data. It is noted that the success rate is available when we have many sets of input data while the conservativeness level is from a data set. This is why we cannot match the two values exactly. Each C-RBDO trial finds a cost-effective as well as reliable design based on the given input data set. That is why the averages of the benchmark PoF of G_1 and G_2

Table 9 Number of successful results in repeated C-RBDO trials

Data	G_1	G_2	G_3
20	946	949	1000
100	913	958	1000
200	899	977	1000

Table 10 Statistics of benchmark PoF in repeated test

Data	Probability of failure of G_1 (%)			Probability of failure of G_2 (%)		
	Average	Min.	Max.	Average	Min.	Max.
20	0.772	0.001	6.842	0.904	0.022	5.372
100	1.368	0.131	4.229	1.434	0.499	3.498
200	1.589	0.437	3.806	1.576	0.741	3.211

Fig. 8 Optimum designs of repeated C-RBDO Test

(a) 20 data pairs

(b) 100 data pairs

(c) 200 data pairs
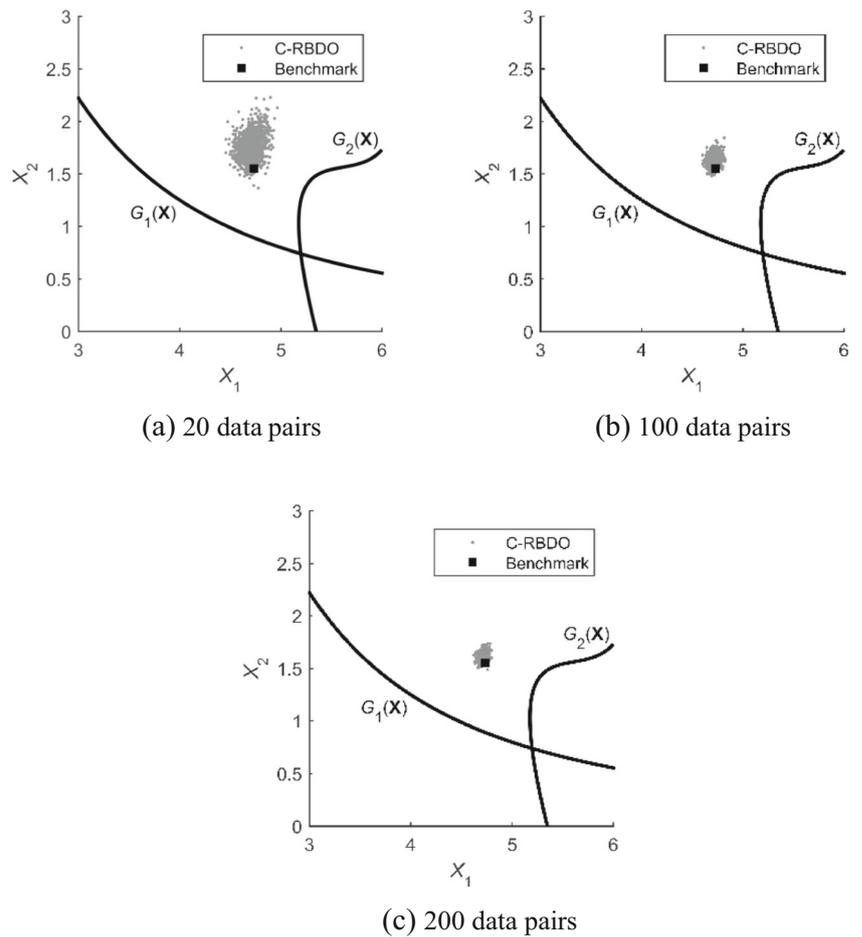
are similar to each other. The repeated test in this section requires the high-performance computing system, which is not usually available. The C-RBDO could be one of the best solutions for design optimization in the presence of a limited number of data because it provides a cost-effective and reliable optimum design based on the limited data.

6 Numerical example: 11-dimensional vehicle side impact problem

In this section, an 11-dimensional vehicle side impact problem (Youn and Choi 2004; Du and Choi 2008) has been tested to verify the performance of the C-RBDO in high-dimensional applications. The vehicle side impact problem has 11 input random variables: $X_1 \sim X_7$ are the thicknesses of structural members, X_8 and X_9 are the material properties of critical members, and X_{10} and X_{11} are the position of the impact test. Hence, X_{10} and X_{11} are random parameters and not related to the design variables. The benchmark distribution of the problem is shown in Table 11. It can be seen that all the input random variables follow Normal distribution.

The optimization problem is to find $\mathbf{d} = [d_1, \ldots, d_9]^T$ to

minimize $\quad cost\,(\mathbf{d})$

subject to $\quad F_{P_{F_j}}\left(p_F^{Tar} = 10\% |^*\mathbf{x}\right) \geq 90\%, \quad j = 1, \ldots, 10$

$\quad\quad\quad \mathbf{d}^L < \mathbf{d} < \mathbf{d}^U, \quad \mathbf{d} \in \mathbb{R}^9, \quad \text{and} \quad \mathbf{X} \in \mathbb{R}^{11}$

$$(37)$$

Table 11 Benchmark input distribution

Random variable	Marginal distribution	Mean	Variance
X_1	Normal	d_1	0.03^2
X_2	Normal	d_2	0.03^2
X_3	Normal	d_3	0.03^2
X_4	Normal	d_4	0.03^2
X_5	Normal	d_5	0.03^2
X_6	Normal	d_6	0.03^2
X_7	Normal	d_7	0.03^2
X_8	Normal	d_8	0.006^2
X_9	Normal	d_9	0.006^2
X_{10}	Normal	0	10^2
X_{11}	Normal	0	10^2

where $\mathbf{d}^L = [0.5,\ 0.5,\ 0.5,\ 0.5,\ 0.5,\ 0.5,\ 0.5, 0.192,$ 0.192]^T and $\mathbf{d}^U = [1.5,\ 1.5,\ 1.5,\ 1.5,\ 1.5,\ 1.5,\ 1.5,$ $0.345,\ 0.345]^T$. The cost function and 10 constraint functions are defined by

$$cost(\mathbf{d}) = 1.98 + 4.9\,d_1 + 6.67d_2 + 6.98\,d_3 + 4.01d_4 + 1.78\,d_5 + 2.73d_7$$

$$G_1(\mathbf{X}) = 14.36 + (-9.9X_2 - 12.9X_1X_8 + 0.1107X_3X_{10})$$

$$G_2(\mathbf{X}) = 1.86 + (2.95X_3 + 0.1792X_{10} - 5.057X_1X_2 - 11X_2X_8 - 0.0215X_5X_{10} \\ - 9.98X_7X_8 + 22X_8X_9)$$

$$G_3(\mathbf{X}) = -3.02 + (3.818X_3 - 4.2X_1X_2 + 0.0207X_5X_{10} + 6.63X_6X_9 - 7.7X_7X_8 \\ + 0.32X_9X_{10})$$

$$G_4(\mathbf{X}) = -0.059 + (-0.0159X_1X_2 - 0.188X_1X_8 - 0.019X_2X_7 + 0.0144X_3X_5 \\ + 0.0008757X_5X_{10} + 0.08045X_6X_9 + 0.00139X_8X_{11} \\ + 0.00001575X_{10}X_{11})$$

$$G_5(\mathbf{X}) = -0.106 + (0.00817X_5 - 0.131X_1X_8 - 0.0704X_1X_9 + 0.03099X_2X_6 - 0.018X_2X_7 \\ + 0.0208X_3X_8 + 0.121X_3X_9 - 0.00364X_5X_6 + 0.0007715X_5X_{10} \\ - 0.0005354X_6X_{10} + 0.00121X_8X_{11} + 0.00184X_9X_{10} - 0.018X_2^2)$$ \hfill (38)

$$G_6(\mathbf{X}) = 0.42 + (-0.61X_2 - 0.163X_3X_8 + 0.001232X_3X_{10} - 0.166X_7X_9 + 0.227X_2^2)$$

$$G_7(\mathbf{X}) = 0.72 + (-0.5X_4 - 0.19X_2X_3 - 0.0122X_4X_{10} + 0.009325X_6X_{10} \\ + 0.000191X_{11}^2)$$

$$G_8(\mathbf{X}) = 0.68 + (0.674X_1X_2 - 1.95X_2X_8 + 0.02054X_3X_{10} - 0.0198X_4X_{10} \\ + 0.028X_6X_{10})$$

$$G_9(\mathbf{X}) = 1.35 + (-0.489X_3X_7 - 0.843X_5X_6 + 0.0432X_9X_{10} - 0.0556X_9X_{11} \\ - 0.000786X_{11}^2)$$

$$G_{10}(\mathbf{X}) = 0.16 + (-0.3717X_2X_4 - 0.00931X_2X_{10} - 0.484X_3X_9 + 0.01343X_6X_{10})$$

The target PoF is 10 % and the target conservativeness level is 90 % in this example as shown in (37). It is difficult to carry out the C-RBDO for high-dimensional applications like this problem due to the huge number of candidate distribution types. If seven marginal distribution types are used for each random variable of the vehicle side impact problem, there are nearly two billion ($7^1 = 1{,}977{,}326{,}743$) candidate distribution types. Two billion types cannot be considered computationally; therefore, we need a method to break the number of types down effectively. Section 6.2 is dedicated to explain and resolve this challenge in detail.

6.1 Test cases

Similar to the 2-D mathematical example, sets of limited input data have been drawn from the benchmark distribution in Table 11. To see the effect of the number of data, three cases with different data sizes are considered as shown in Table 12. Moreover, in each case, the number of data for random variables is different. Sample variance, which is the variance of the data, is summarized in Table 12 as well. By comparing Tables 11 and 12, it can be seen that the more data we have, the more accurate the information (sample variance) we can obtain.

Table 12 Test cases

Random variable	Case 1		Case 2		Case 3	
	Number of data	Sample variance	Number of data	Sample variance	Number of data	Sample variance
X_1	15	0.02439^2	50	0.02736^2	150	0.02801^2
X_2		0.03748^2		0.02672^2		0.02863^2
X_3		0.02558^2		0.02487^2		0.02693^2
X_4		0.02917^2		0.03165^2		0.02774^2
X_5		0.02490^2		0.02886^2		0.02761^2
X_6		0.03714^2		0.03283^2		0.03378^2
X_7		0.03127^2		0.03202^2		0.03029^2
X_8	20	0.004683^2	60	0.005643^2	200	0.006108^2
X_9		0.005663^2		0.005683^2		0.005687^2
X_{10}	10	7.0064^2	30	10.0208^2	100	9.4292^2
X_{11}		10.5228^2		11.6460^2		10.0290^2

Table 13 Optimum designs of DDO and conventional RBDO

Case		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
DDO		0.5	1.2257	0.5	1.2071	0.9709	1.5	0.5	0.345	0.3
RBDO	Case 1	0.5	1.2962	0.5	1.3253	1.0398	1.5	0.5	0.345	0.192
	Case 2	0.5	1.3074	0.5	1.3560	1.0532	1.5	0.5	0.345	0.192
	Case 3	0.5	1.3042	0.5	1.3185	1.0526	1.5	0.5	0.345	0.192

To pursue an efficient process of the C-RBDO, the DDO and the conventional RBDO have been performed a priori. The same procedure with the 2-D mathematical example has been followed for the conventional RBDO. The optimum design of DDO is the same for the three cases, whereas different RBDO optimum design is obtained due to the different input data. The optimum designs are summarized in Table 13. It can be seen that d_2, d_4 and d_5 are changed while the other design variables are on their design bounds. d_9 does not move in the DDO process because the design sensitivities of the cost function and the active constraints with respect to d_9 are zero. However, in the conventional RBDO process, d_9 is moved to the lower bound for all three cases.

The cost function values at the DDO and conventional RBDO optimum designs are shown in Table 14. All three conventional RBDO optimum designs have larger cost value than the one at the DDO optimum design because they secure more reliability to compensate for the input variability. However, in the conventional RBDO, the input variability cannot be considered correctly because it is estimated using limited input data, as shown in Table 12. To understand the effect of the limited input data, the reliability analysis has been performed at these optimum designs using the benchmark distribution in Table 11 and the benchmark PoF of active constraints are listed in Table 14. In Case 1, the 10 % target PoF is almost satisfied for G_1 and G_9; however, this result cannot be always expected. Because input uncertainty is not considered in the conventional RBDO, the result may not satisfy the target PoF with different sets of input data. Case 2 has more data than Case 1, but it fails to satisfy the 10 % target PoF for G_1 and G_9 due to the input uncertainty. Even though the sample variances of Case 2 are closer to the ones of the benchmark distribution compared to Case 1 (see Tables 11 and 12), Case 2 violates the target PoF more. The benchmark PoF of Case 3 is close to the 10 % target PoF even for G_7. Since a large number of data is given in Case 3, the input uncertainty is substantially small. Hence, the conventional RBDO optimum design of Case 3 is more trustworthy compared to that of the other two cases.

6.2 Candidate distribution type selection

As explained earlier, the number of candidate distribution types is enormous (2 billion) if we try to consider all marginal distribution types for each input random variable. Hence, a method to effectively reduce the number of candidates is introduced in this section. At the conventional RBDO optimum design, which is the initial design of the C-RBDO, all marginal PDFs for X_1 are depicted using the sample variance (see Table 12) as shown in Fig. 9. It is shown that the marginal distribution types can be divided into three groups as symmetric, left-skewed and right-skewed PDFs and that we can select one representative type from each group because they have almost identical shape. Moreover, the slight shape difference can be covered by the probability of input distribution parameters. The other input random variables follow the same PDF grouping as X_1 except for X_{10} and X_{11}. Because X_{10} and X_{11} can have negative values, the applicable marginal distribution types are Normal, Gumbel and extreme. These three types also represent the PDF groups for $X_1 \sim X_9$. Hence, Normal, Gumbel and extreme are used as the marginal distribution types for each input random variable.

Thus, the new number of candidate distribution types is 3^{11} (=177,147), which is still a very large number to be considered computationally. At the conventional RBDO optimum design, the probability of 177,147 candidate distribution types is calculated using (23) and setting ψ to be the conventional RBDO optimum design (input mean) and the sample variance (input variance). These probability values are very small due to the large number of combinations; for this reason, the probability values are accumulated. In Table 15, the accumulated probability of distribution type of Case 1 is shown. Among the 177,147 probability values, the probability that X_1 would be Normal distribution is accumulated as 50.6 %, as shown in Table 15. In other words, the 50.6 % is the sum of all

Table 14 Cost function values and reliability analysis result using benchmark distribution at DDO and conventional RBDO optimum designs

Case		Cost	Probability of failure (%)		
			G_1	G_7	G_9
DDO		24.029	49.79	73.50	41.85
RBDO	Case 1	25.318	10.42	6.19	10.74
	Case 2	25.096	13.90	9.81	13.70
	Case 3	25.145	11.41	10.64	10.91

Fig. 9 Marginal PDFs of X_1 using sample variance at conventional RBDO optimum design

(a) Symmetric PDFs

(b) Left-skewed PDFs

(c) Right-skewed PDFs
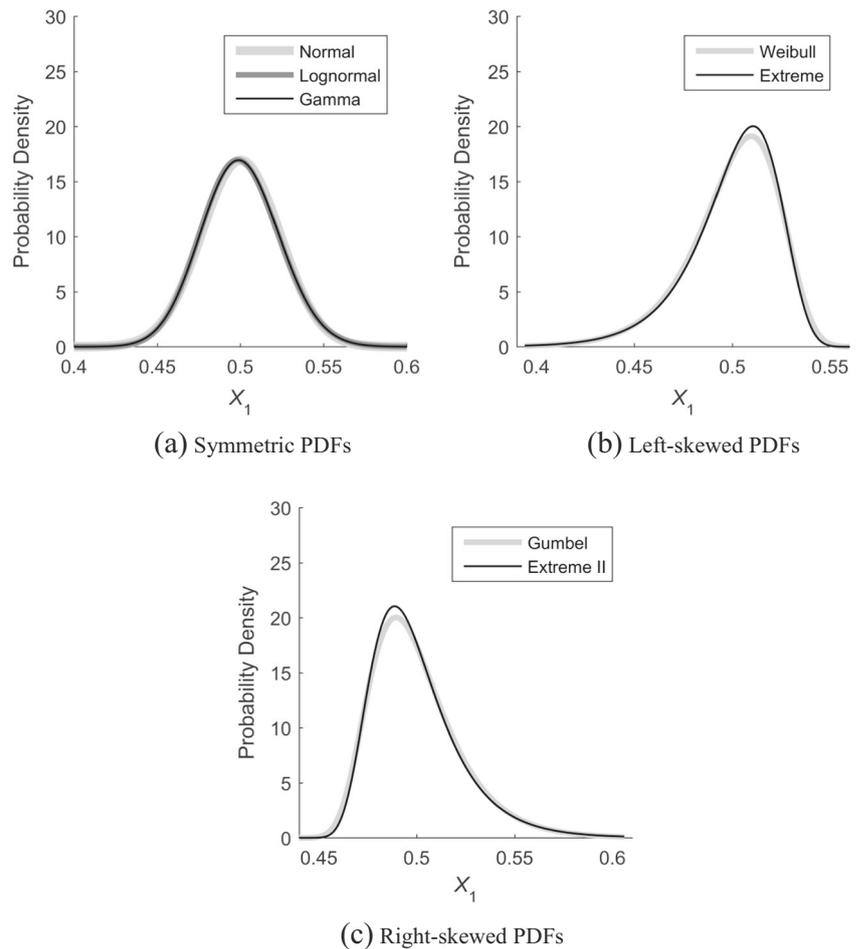
probabilities that X_1 is Normal distribution. In this way, all the accumulated probability is calculated. From Table 15, the dominant marginal distribution type for each random variable is selected if the probability is over 66.7 % (=200 %/3). The threshold value could be different; in this paper, we give double weight (200 %) to determine dominancy and divide it by three – the number of marginal distribution types. Once, a dominant type is selected, it is the only marginal type for the random variable. In this way, $X_3 \sim X_6$ and X_8 have one dominant type as marked with bold font in Table 15. If there is no dominant type, multiple types are considered for the variable. The marginal distribution types with a probability less than 16.7 % (=50 %/3) have been opted out from the candidate types. Again, we give half

weight to declare opt-out. For example, Gumbel has not been used for the marginal distribution type for X_1 because its probability of 8.4 % is less than 16.7 %. The selected marginal distribution types are marked with bold fonts in Table 15, and there is a total of 144 (=2×3×3×2×2×2) candidate distribution types for Case 1. The same strategy is used for Cases 2 and 3. Because these cases have more data, a smaller number of candidate distribution types is selected. For Case 2, $X_1 \sim X_{10}$ have Normal distribution for the marginal distribution type, while X_{11} could have Normal and Gumbel. Hence, there are two candidate distribution types for Case 2. For Case 3, Normal distribution is dominant for all random variables since there is enough data. Hence, there is only one candidate distribution type in Case 3.

Table 15 Accumulate probability of distribution types at conventional RBDO optimum design for Case 1

Types	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
Normal	**50.6 %**	**46.0 %**	28.3 %	30.6 %	9.0 %	**87.1 %**	**40.5 %**	**70.2 %**	**42.8 %**	**55.2 %**	**47.9 %**
Gumbel	8.4 %	**18.0 %**	0.2 %	0.5 %	0.0 %	10.2 %	**20.6 %**	22.7 %	0.2 %	**28.3 %**	**44.8 %**
Extreme	**40.9 %**	**36.0 %**	**71.5 %**	**68.9 %**	**91.0 %**	2.7 %	**38.8 %**	7.1 %	**57.0 %**	16.6 %	7.3 %

Table 16 Optimum designs of C-RBDO

Case	Optimum design					Cost	Conservativeness level (%)		
	d_2	d_4	d_5	d_8	d_9		G_1	G_7	G_9
Case 1	1.3323	1.4692	1.0768	0.3442	0.1945	25.979	90.3	90.1	90.5
Case 2	1.3281	1.4192	1.0797	0.3450	0.1920	25.756	90.0	89.7	90.2
Case 3	1.3143	1.3407	1.0642	0.3450	0.1920	25.322	89.3	88.8	89.4
Benchmark	1.3088	1.3215	1.0568	0.3450	0.1920	25.195	–	–	–

6.3 Confidence-based RBDO result

Using the selected candidate distribution types, C-RBDO has been performed for the 11-D vehicle side impact problem. The C-RBDO is initiated at the conventional RBDO optimum design in Table 13. Both $NMCS_\Psi$ and $NMCS_Z$ are set to 20,000 in this example. The SQP is used for the optimization process, and the convergence criteria are set to 0.001 for first-optimality, conservativeness level and design movement. The optimization uses an Intel Xeon E5-2690 processor with 16GB memory, and the computation time is 36.3, 0.5 and 0.8 h for Cases 1, 2 and 3, respectively. As more data is given, the computation time is much smaller because there is a smaller number of candidate distribution types and the initial design is closer to the C-RBDO optimum design. The obtained optimum designs are listed in Table 16. The optimum design variables d_2, d_4 and d_5 move to secure more conservativeness in the C-RBDO process. In Case 1, d_8 and d_9 move a little away from their bounds, but they stay on their bounds in other cases. The other design variables remain at the conventional RBDO optimum design. The cost function values are all increased to satisfy the 90 % target conservativeness level for the active constraints as shown in Table 16. The other constraints are inactive, which means their conservativeness level is almost 100 %. Case 1 uses 17 design iterations and 28 conservativeness level estimations, Case 2 uses four design iterations and five conservativeness level estimations, and Case 3 uses three design iterations and eight

conservativeness level estimations. Case 1 uses more design iterations and conservativeness level estimations to compensate larger input uncertainty than other cases. Considering this is a high-dimensional optimization problem, the optimization processes are performed efficiently and effectively by providing accurate design sensitivity. The conventional RBDO optimum design using the benchmark input distribution is also shown in Table 16. It can be seen that the C-RBDO optimum design approaches the benchmark design as more input data is provided. Therefore, it is verified that the C-RBDO performs effectively in this vehicle side impact problem.

Reliability analysis is again carried out at the C-RBDO optimum designs using the benchmark input distribution in Table 11. The benchmark PoF of active constraints is shown in Table 17. It can be seen the all PoFs are smaller than the 10 % target PoF. Hence, C-RBDO secures more conservativeness in the optimum design to compensate for the input uncertainty. At the same time, the optimum designs show convergence to the benchmark optimum in the sense of benchmark PoF as well. Therefore, it can be confirmed that the C-RBDO considers the limited input data correctly.

7 Conclusion

This paper presents a new method that takes insufficient input data into consideration for RBDO. Probability of failure is defined as a function of input distribution parameters and types. When the amount of input data is limited, there is no fixed input distribution parameter and type, and the PoF becomes uncertain. Using the Bayesian method, joint PDFs of input distribution parameters and types are obtained based on the input data. Then, MCS is used to obtain the probability of the PoF, and the conservativeness level is defined as the probability at a user-specified PoF value. The conservativeness level can be used as a new constraint for RBDO to acquire reasonable conservativeness in the optimum design. Moreover, a new design sensitivity of the

Table 17 Reliability analysis result using benchmark distribution

Case	Probability of failure (%)		
	G_1	G_7	G_9
Case 1	5.085	1.523	6.761
Case 2	5.742	2.649	5.992
Case 3	8.630	7.510	8.629
Benchmark	9.877	9.918	9.908

conservativeness level is derived for an efficient and effective optimization process.

Through a 2-D mathematical example, it is shown that a safe design assuming the true input distribution is no longer reliable when only limited input data is provided. In addition, it is shown that more data provides a greater conservativeness level at the design, which is a good indication that the developed method correctly considers amount of input data. The developed design sensitivity of the conservativeness level is compared with FDM design sensitivity. The accuracy of the developed design sensitivity agrees with the FDM design sensitivity well while using only 0.0625 % MCS samples of the FDM. Therefore, it is proven that the developed design sensitivity method is accurate as well as efficient. Using the developed method for estimation of conservativeness level and design sensitivity, an RBDO process is performed to acquire a more reliable optimum design even with limited data. An optimum design closer to the benchmark optimum design is obtained as more data is provided. At the same time, the C-RBDO design optimum is not overly conservative compared to the previously developed method, which uses confidence level of input variance. Finally, C-RBDO is tested repeatedly to see the performance of

conservativeness level estimation. It is shown that the obtained C-RBDO optimum design correctly represents the target conservativeness level. Hence, the C-RBDO method can be recommended for a design problem with insufficient input data. To verify the scalability of the developed C-RBDO method, an 11-D vehicle side impact problem is tested. A procedure to effectively select candidate distribution type is proposed for high-dimensional applications. The C-RBDO result is in accordance with the 2-D mathematical example, which verifies the scalability of the C-RBDO method.

In the future, the C-RBDO method could be expanded to consider correlated input data. Ultimately, an engineering application could be carried out for realistic applications.

Appendix A

Table 18 Marginal PDFs

Distribution type	PDF $f_X(x; a, b)$	Parameters
Normal	$\frac{1}{\sqrt{2\pi}b}e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2}$	$\mu = a, \sigma = b$
Lognormal	$\frac{1}{\sqrt{2\pi}xb}e^{-\frac{1}{2}\left(\frac{\ln x - a}{b}\right)^2}$	$\mu = e^{a+b^2/2}$, $\sigma^2 = \left(e^{b^2}-1\right)e^{2a+b^2}$
Weibull	$\frac{b}{a}\left(\frac{x}{a}\right)^{b-1}e^{-\left(\frac{x}{a}\right)^b}$	$\mu = a\Gamma(1+1/b)$, $\sigma^2 = a^2\left[\Gamma(1+2/b)-\Gamma^2(1+1/b)\right]$
Gumbel	$\frac{1}{b}\exp\left[-\frac{x-a}{b}-\exp\left(-\frac{x-a}{b}\right)\right]$	$\mu = a + 0.5772b$, $\sigma^2 = b^2\pi^2/6$
Gamma	$x^{a-1}\frac{e^{-x/b}}{\Gamma(a)b^a}$	$\mu = ab$, $\sigma^2 = ab^2$
Extreme	$\frac{1}{b}\exp\left[\frac{x-a}{b}-\exp\left(\frac{x-a}{b}\right)\right]$	$\mu = a - 0.5772b$, $\sigma^2 = b^2\pi^2/6$
Extreme Type II	$\frac{a}{b}\left(\frac{b}{x}\right)^{a+1}e^{-\left(\frac{b}{x}\right)^a}$	$\mu = b\,\Gamma(1-1/a)$, $\sigma^2 = b^2\left[\Gamma(1-2/a)-\Gamma^2(1-1/a)\right]$

References

Aughenbaugh JM, Paredis CJJ (2006) The value of using imprecise probabilities in engineering design. J Mech Des 128(4):969–979

Breitung K (1984) Asymptotic approximations for multinormal integrals. J Eng Mech 110(3):357–366

Chick SE (2001) Input distribution selection for simulation experiments: accounting for input uncertainty. Oper Res 49(5):744–758

Ditlevsen O, Madsen HO (1996) Structural reliability methods. Wiley, Chichester

Du L, Choi KK (2008) An inverse analysis method for design optimization with both statistical and fuzzy uncertainties. Struct Multidiscip Optim 37(2):107–119

Elishakoff I (2004) Safety factors and reliability: friends or foes? Kluwer Academic Publishers, Dordrecht

Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian data analysis, 2nd edn. Chapman & Hall/CRC, Boca Raton

Gumbert CR, Hou GJW, Newman PA (2003) Reliability assessment of a robust design under uncertainty for a 3-D flexible wing. Proc. 16th AIAA Computational Fluid Dynamics Conference, Orlando

Gunawan S, Papalambros PY (2006) A Bayesian approach to reliability-based optimization with incomplete information. J Mech Des 128(4):909–918

Haldar A, Mahadevan S (2000) Probability, reliability and statistical methods in engineering design. Wiley, New York

Hasofer AM, Lind NC (1974) Exact and invariant second-moment code format. J Eng Mech Div ASCE 100(1):111–121

Hohenbichler M, Rackwitz R (1988) Improvement of second-order reliability estimates by importance sampling. J Eng Mech 114(12):2195–2199

Hou GJW (2004) A most probable point-based method for reliability analysis, sensitivity analysis, and design optimization. NASA/CR-2004-213002, NASA

Lee I, Choi KK, Gorsich D (2010) System reliability-based design optimization using the MPP-based dimension reduction method. Struct Multidiscip Optim 41(6):823–839

Lee I, Choi KK, Noh Y, Zhao L, Gorsich D (2011a) Sampling-based stochastic sensitivity analysis using score functions for RBDO problems with correlated random variables. J Mech Des 133(2):021003

Lee I, Choi KK, Zhao L (2011b) Sampling-based RBDO using the stochastic sensitivity analysis and dynamic kriging method. Struct Multidiscip Optim 44(3):299–317

Lee I, Noh Y, Yoo D (2012) A novel Second-Order Reliability Method (SORM) using non-central or generalized chi-squared distributions. J Mech Des 134(10):100912

Lim J, Lee B, Lee I (2014) Second-order reliability method-based inverse reliability analysis using hessian update for accurate and efficient reliability-based design optimization. Int J Numer Methods Eng 100(10):773–792

Noh Y, Choi KK, Lee I, Gorsich D, Lamb D (2011a) Reliability-based design optimization with confidence level under input model uncertainty due to limited test data. Struct Multidiscip Optim 43(4):443–458

Noh Y, Choi KK, Lee I, Gorsich D, Lamb D (2011b) Reliability-based design optimization with confidence level for non-gaussian distributions using bootstrap method. J Mech Des 133(9):091001

Rahman S, Wei D (2006) A univariate approximation at most probable point for higher-order reliability analysis. Int J Solids Struct 43(9):2820–2839

Rahman S, Wei D (2008) Design sensitivity and reliability-based structural optimization by univariate decomposition. Struct Multidiscip Optim 35(3):245–261

Rubinstein RY, Kroese DP (2008) Simulation and the Monte Carlo method, 2nd edn. Wiley, Hoboken

Tu J, Choi KK, Park YH (1999) A new study on reliability-based design optimization. J Mech Des 121(4):557–564

Tu J, Choi KK, Park YH (2001) Design potential method for robust system parameter design. AIAA J 39(4):667–677

Tucker WT, Ferson S (2003) Probability bounds analysis in environmental risk assessment. Applied Biomathematics, Setauket

Utkin L, Destercke S (2009) Computing expectations with continuous P-Boxes: univariate case. Int J Approx Reason 50(5):778–798

Youn BD, Choi KK (2004) A new response surface methodology for reliability-based design optimization. Comput Struct 82(2–3):241–256

Youn B, Wang P (2008) Bayesian reliability-based design optimization using Eigenvector Dimension Reduction (EDR) method. Struct Multidiscip Optim 36(2):107–123

Zhang R, Mahadevan S (2000) Model uncertainty and bayesian updating in reliability-based inspection. Struct Saf 22(2):145–160