Struct Multidisc Optim (2010) 41:205–217
DOI 10.1007/s00158-009-0419-8

RESEARCH PAPER

Application of bootstrap method in conservative estimation of reliability with limited samples

Victor Picheny · Nam Ho Kim · Raphael T. Haftka

Received: 8 September 2008 / Revised: 28 May 2009 / Accepted: 9 July 2009 / Published online: 7 August 2009

Abstract Accurate estimation of reliability of a system is a challenging task when only limited samples are available. This paper presents the use of the bootstrap method to safely estimate the reliability with the objective of obtaining a conservative but not overly conservative estimate. The performance of the bootstrap method is compared with alternative conservative estimation methods, based on biasing the distribution of system response. The relationship between accuracy and conservativeness of the estimates is explored for normal and lognormal distributions. In particular, detailed results are presented for the case when the goal has a 95% likelihood to be conservative. The bootstrap approach is found to be more accurate for this level of conservativeness. We explore the influence of sample size and target probability of failure on the quality of estimates, and show that for a given level of conservativeness, small sample sizes and low probabilities of failure can lead to a high likelihood of large overestimation. However, this likelihood can be reduced by increasing the sample size. Finally, the conservative approach is applied to the reliability-based optimization of a composite panel under thermal loading.

Keywords Bootstrap method · Conservative estimation · Reliability · Probability of failure

V. Picheny · N. H. Kim (✉) · R. T. Haftka
Department of Mechanical and Aerospace Engineering,
University of Florida, Gainesville, FL 32611, USA
e-mail: nkim@ufl.edu

V. Picheny
e-mail: vpicheny@ufl.edu

R. T. Haftka
e-mail: haftka@ufl.edu

Nomenclature

P_f	Probability of failure
$\hat{P}_f$	Estimate of probability of failure
β	Reliability index
MCS	Monte-Carlo simulations
CDF	Cumulative Distribution Function
G	Limit-state function
F_G	CDF of limit-state G
θ	Distribution parameters
$\hat{\theta}$	Estimate of distribution parameters
RMS	Root mean square
CSP	Conservative at sample point
CEC	Conservative to the empirical curve
CVaR	Conditional value-at-risk
PRS	Polynomial response surface

1 Introduction

When an engineering system has uncertainty in its input parameters and operating conditions, the safety of the system can be evaluated in terms of reliability. Many methods have been proposed to estimate the reliability of a system, such as Monte Carlo simulation (MCS) method (Haldar and Mahadevan 2000), First and Second-order Reliability Method (Enevoldsen and Sørensen 1994; Melchers 1999), importance sampling method (Engelund and Rackwitz 1993), tail modeling (Kim et al. 2006), and inverse methods (Qu and Haftka 2004). MCS is often used to estimate the reliability of the system that has many random inputs or multiple failure modes because its accuracy is independent of the complexity of the problem. In this paper, reliability analysis using MCS method is considered. The comparison between various reliability analysis methods is beyond the scope of the paper and can be found in the literature (Rackwitz 2000; Lee et al. 2002).

When the cost of simulation is high, engineers can afford to have only a limited number of samples, which is not sufficient to estimate the reliability with acceptable accuracy (Ben-Haim and Elishakoff 1990; Neal et al. 1991). In such cases, it is often required to compensate for the lack of accuracy with extra safety margins. For example, Starnes and Haftka (1979) replaced the linear Taylor series approximation with a tangent approximation biased to be conservative in order to reduce the chance of unconservative approximations to buckling loads. Many engineering applications have adopted conservative estimation. For example, the Federal Aviation Administration requires the use of conservative failure stress of the design of aircraft structures, using A- or B-basis stresses based on redundancy. The A-basis failure stress is the value of a failure stress exceeded by 99% (or 90% for B-basis) of the population with 95% confidence. In the same context, anti-optimization (Elishakoff 1990; Du et al. 2005) and possibility-based design (Du et al. 2006; Choi et al. 2005) are used to compensate for the lack of knowledge in the input distribution by seeking the worst case scenario for a given design. Such approaches have been found to lead to conservative designs (Nikolaidis et al. 2004). Bayesian reliability-based optimization (Youn and Wang 2007) uses Bayesian theory to ensure production of reliable designs when insufficient data is available for the inputs.

In this paper, we focus on the case when the probability of failure, P_f, of a system is estimated from a limited number of samples. The objective is to find a conservative estimate, $\hat{P}_f$, that is likely to be no lower than the true P_f. To provide such estimation, two alternatives are considered: the first method is based on biasing the process of fitting the distribution used to compute the estimator of P_f. The second is the use of the bootstrap method (Efron 1982; Chernick 1999) to quantify the uncertainty in probability of failure estimations, and defining conservative estimators based on bootstrapping. As the conservative estimations tend to overestimate the probability of failure, a trade-off analysis between accuracy and the level of conservativeness (i.e., chance of being conservative) is proposed with the help of numerical examples.

In the next section, we discuss how we use sampling techniques to estimate the probability of failure. Section 3 shows how to use constraints to obtain conservative estimators. Section 4 describes the bootstrap method and how to use it to define conservative estimators. The accuracy of such estimators is analyzed using a simple numerical example in Section 5, and an analysis of the effects of sample sizes and target probability of failure on the quality of conservative estimators is given in Section 6. Finally, the conservative approach is applied to an engineering problem in Section 7, followed by concluding remarks in Section 8.

2 Estimation of probability of failure from samples

2.1 Limit-state and probability of failure

Failure of a system can usually be determined through a criterion, called a limit-state, G. The limit-state is defined so that the system is considered safe if $G < 0$ and fails otherwise. For instance, the limit-state of a structure can be defined as the difference between response, R, (e.g., maximum stress or strain) and capacity, C, (e.g., maximum allowable stress or strain):

$$G = R - C \tag{1}$$

Due to uncertainties in material properties and loadings, the limit-state is random, and the safety of the system should be evaluated in terms of reliability or probability of failure. The probability of failure is defined as

$$P_f = \text{Prob}(G \geq 0) \tag{2}$$

There are many methods for calculating the failure probability of a system (Haldar and Mahadevan 2000; Enevoldsen and Sørensen 1994; Melchers 1999). Some of them use the relation between input random variables and the limit-state (e.g., first-order reliability method) and others consider the limit-state as a black-box output (e.g., MCS). When the number of input random variables is large, and the limit-state is complex and multi-modal, MCS has a particular advantage as its accuracy is independent of the problem dimension or complexity of G. MCS generates samples of the limit-state and counts the number of failed samples (Melchers 1999). The ratio between the numbers of failures and the total number of samples approximates the probability of failure of the system.

The variance of MCS estimates is inversely proportional to the square root of the number of samples times the probability of failure. Thus, accuracy is poor when the number of samples is small or when the probability of failure is low. For instance, if a probability of failure of 10^{-4} is estimated (which is a typical value in reliability based design), 10^6 samples are needed for 10% relative error.

When the cost of simulation is high, engineers can afford to have only a limited number of samples, and it is not good enough to estimate the reliability with acceptable accuracy (Ben-Haim and Elishakoff 1990; Neal et al. 1991). In such a case, it is possible to approximate the cumulative distribution function (CDF) F_G using a limited number of samples g_1, g_2,.., g_n of the limit-state, and estimate the probability of failure using:

$$P_f = 1 - F_G(0) \tag{3}$$

In this work, we consider the case where the number of samples is between 20 and 1,000.

The samples of limit states can be obtained by generating input random variables and propagating through the system. However, the proposed method can be applied to the case when input random variables and uncertainty propagation are unknown. For example, the samples of limit states can be obtained from experiments.

In the following section, two methods of estimating distribution parameters from a set of samples are discussed.

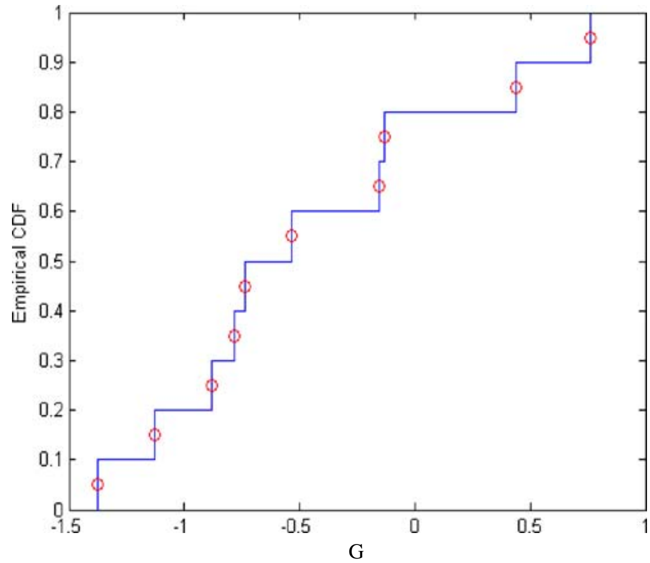

Fig. 1 Points (*circles*) chosen to fit an empirical CDF (*line*) obtained by ten samples from $N(0,1)$

2.2 Estimation of distribution parameters

In general, the CDF F_G in (3) is unknown and is often approximated by fitting empirical CDF. First, the limit-state is assumed to follow a parametric distribution F_θ, which is defined by distribution parameters, θ. Then, the parameters are estimated by minimizing differences between the empirical CDF and F_θ. Compared to other techniques, such as moment-based methods, this method is of interest because it can be modified to build conservative estimates.

Consider n limit-state samples arranged in increasing order: $(g_1 \le g_2 \le \cdots \le g_n)$. The empirical CDF F_n is defined as:

$$F_n(g) = \begin{cases} 0 & \text{for } g \le g_1 \\ k/n & \text{for } g_k \le g \le g_{k+1} \\ 1 & \text{for } g \ge g_n \end{cases} \tag{4}$$

It is then possible to estimate the parameters θ of the CDF that approximates F_n best. Two different approximation methods are discussed here: (1) minimizing the root-mean-square (RMS) error, and (2) minimizing the Kolmogorov–Smirnov distance (Kenney and Keeping 1951).

To minimize the RMS difference between the empirical and the estimated CDF, errors are calculated at the sample points. In order to have an unbiased estimation, the values of the empirical CDF are chosen at the middle of the two discrete data, as (see Fig. 1):

$$F'_n(g_k) = \frac{k - \frac{1}{2}}{n}, \quad 1 \le k \le n. \tag{5}$$

Then the estimated parameters $\hat{\theta}$ are chosen to minimize the following error:

$$\hat{\theta} = \underset{\theta}{\arg\min} \sqrt{\sum_{k=1}^{n} \left[F_\theta(g_k) - F'_n(g_k)\right]^2} \tag{6}$$

where 'arg min' is the values of parameters that minimize the error. The estimate based on (6) is called an 'RMS estimate'.

The Kolmogorov–Smirnov (K-S) distance is the classical way to test if a set of samples are representative of a distribution. The K-S distance is equal to the maximum distance between two CDFs (see Fig. 2). The optimal parameters for the K-S distance are:

$$\hat{\theta} = \underset{\theta}{\arg\min} \left[\max_{1 \le k \le n} \left(\left|F_\theta(g_k) - \frac{k}{n}\right|, \left|F_\theta(g_k) - \frac{k-1}{n}\right|\right)\right] \tag{7}$$

Figure 1 shows the empirical CDF from ten samples and the data points that are used in (5). Figure 2 shows the K-S distance between a normally distributed CDF and an empirical CDF from 10 samples.

For the cases we consider in this work, it is found that the two criteria give equivalent results. K-S distance links to classical statistical framework, but can be difficult to minimize because it is non-smooth. The RMS error minimization appears to be more robust.

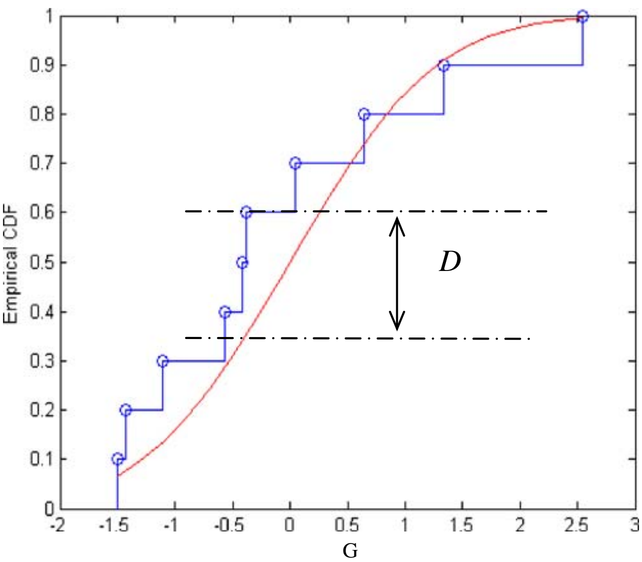

Fig. 2 K-S distance between an empirical CDF with ten samples and a normal CDF (continuous line)

Once the distribution parameters are estimated, the probability of failure in (3) can be estimated by

$$\hat{P}_f = 1 - F_{\hat{\theta}}(0) \tag{8}$$

The choice of the distribution F_θ is critical for accurate estimation of the probability of failure. Wrong assumption on the form of the distribution can lead to large bias in the estimate, for instance, if the distribution is assumed to be normal while it is heavy-tailed. Statistical techniques are available to test if a sample belongs to a given distribution type (goodness-of-fit tests), such as the Kolmogorov–Smirnov (for any distribution) or Lilliefors or Anderson–Darling tests (for normal distributions) (Kenney and Keeping 1951). Some statistical software also offers automated procedures to choose from a benchmark of which distributions best fit the data.

3 Conservative estimates using constraints

As shown in the previous section, fitting a distribution to a set of samples can be seen as an optimization problem. The key idea of this section is adding various constraints to this fitting problem so that the resulting estimate becomes conservative. As the process makes the distribution less accurate, the trade-off between conservativeness and accuracy becomes important.

A conservative estimate of the probability of failure should be equal or higher than the actual one. From the expression of the probability of failure given in (3), a conservative estimate can be obtained by constraining the estimated CDF to be lower than the true CDF when the parameters are found through the optimization problems in (6) or (7). Only small probabilities are considered, so the failure occurs in the upper tail region of the distribution. Hence, the constraints are applied to only the right half of the data. However, if the failure occurs in the lower tail region, the constraints should be applied on the left half of the data.

The first conservative estimate of the CDF is obtained by constraining the estimate to pass below the sampling data points. A second can be obtained by constraining the estimated CDF below the entire empirical CDF. They will be called, respectively, CSP (Conservative at Sample Points) and CEC (Conservative to Experimental CDF). The latter is more conservative than the former. Obviously, both methods introduce bias, and the choice between the two constraints is a matter of balance between accuracy and conservativeness.

CSP constraints:

$$F_\theta(g_i) - \frac{i}{n} \le 0 \quad \text{for} \quad \frac{n}{2} \le i \le n \tag{9}$$

CEC constraints:

$$F_\theta(g_i) - \frac{i-1}{n} \le 0 \quad \text{for} \quad \frac{n}{2} \le i \le n \tag{10}$$

To illustrate these conservative estimators, ten samples are generated from a random variable G whose distribution is $N(-2.33, 1.0^2)$. The mean is chosen in such a way that the probability of failure is 1%. Assuming that the

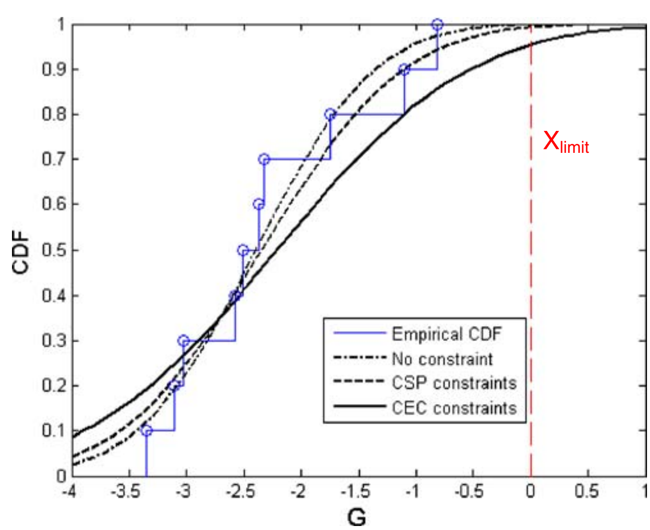

Fig. 3 Example of CDF estimators based on RMS error for a sample of size 10 generated from $N(-2.33, 1.0^2)$

Table 1 Comparison of the mean, standard deviation, and probability of failure of the three different CDF estimators for $N(-2.33, 1.0^2)$

	No constraint	CSP	CEC
$\hat{\mu}$	−2.29	−2.34	−2.21
$\hat{\sigma}$	0.84	0.97	1.31
$\hat{P}_f$	0.32%	0.77%	4.65%

Exact value of P_f is 1%

distribution parameters are unknown, the estimated CDF can be written as

$$F_{\theta}(g) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{g} \exp\left[\frac{-(u-\mu)^2}{2\sigma^2}\right] du \qquad (11)$$

where $\theta = \{\mu, \sigma\}$ is the vector of unknown distribution parameters. Using the form of estimated CDF, the distribution parameters can be found by solving the optimization problems in (6) and (7). In addition, conservative estimates can be found by solving the same optimization problems with constraints in (9) or (10). Figure 3 shows the empirical CDF along with the three estimates based on the minimum RMS error: (1) with no constraint, (2) with CSP constraints, and (3) with CEC constraints. Table 1 shows the parameters of the three estimated distributions and the corresponding probabilities of failure.

The effect of the constraints is clear from the graph. The CSP estimator is shifted down to the ninth data point; hence, the CDF at the tail is decreased. The CEC estimator is shifted even further down. Since the conservative estimators are unconstrained on the left half of the distribution, their CDF curves cross the empirical curve on that side.

For this illustration, we chose a sample realization that is unfavorable for conservativeness. As a consequence, the estimate with no constraint is strongly unconservative (0.32% compared to 1.0%) even though the estimation is unbiased. The CSP estimate is unconservative but substantially less than the unbiased estimate, while the CEC estimate is conservative. In order to generalize these results and derive reliable conclusions, statistical experiments based on large numbers of simulations will be performed in Section 5.

4 Conservative estimates using the bootstrap method

An alternative to biased fitting is to obtain confidence intervals for the probability of failure estimates in order to determine the margin needed to be conservative. However, analytical derivation of confidence intervals for the probability of failure is very challenging. To overcome this problem, we propose to obtain confidence intervals using numerical procedures, i.e. bootstrap method.

4.1 Bootstrap method

When only limited samples are available, the bootstrap method can provide an efficient way of estimating the distribution of statistical parameter θ using the re-sampling technique (Efron 1982; Chernick 1999). The idea is to create many sets of bootstrap samples by re-sampling with replacement from the original data.

This method only requires the initial set of samples. Figure 4 illustrates the procedure of the bootstrap method. The size of the initial samples is n and the number of bootstrap re-samplings is p. Each re-sampling can be performed by sampling with replacement n data out of the n initial samples (hence, the bootstrap samples contain repeated values from the initial samples and omit some of the initial values).

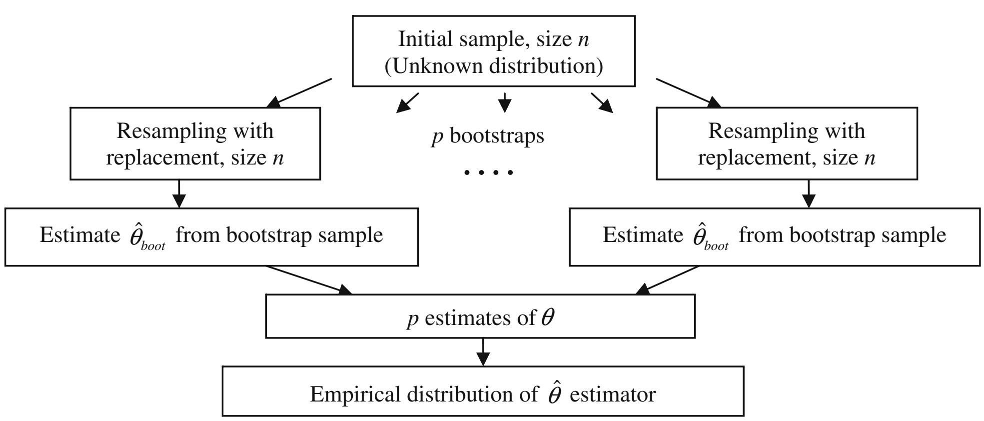

Fig. 4 Schematic representation of bootstrapping. Bootstrap distribution of θ (histogram or CDF of the p estimates) is obtained by multiple re-sampling (p times) from a single set of data

Since the re-sampling process draws samples from the existing set of samples, it does not require additional simulations. For example, if the initial set of samples consists of (g_1, g_2, g_3, g_4), the bootstrap method randomly selects four samples with replacement. Examples of bootstrap samples are (g_1, g_1, g_2, g_4), (g_4, g_2, g_2, g_3), etc. The parameter θ, such as the mean or standard deviation, is estimated for each set of bootstrap samples. Since the re-sampling procedure allows for selecting data with replacement, the statistical properties of the re-sampled data are different from those of the original data. Then, the set of p bootstrap estimates $\hat{\theta}_{boot}$ defines an empirical distribution of θ. This approach allows us to estimate the distribution of any statistical parameter without requiring additional data.

The standard error or confidence intervals of the statistical parameter can be estimated from the bootstrap distribution. However, the bootstrap method provides only an approximation of the true distribution because it depends on the values of the initial samples. In order to obtain reliable results, it is suggested that the size of the samples should be larger than 100 (i.e., n) (Efron 1982). The number of bootstrap re-samplings (i.e., p) is chosen to be large enough so that it does not affect the quality of the results (the major source of uncertainty being the initial sample). The value of p is typically taken from 500 to 5,000.

4.2 Estimation of probability of failure using the bootstrap method

To illustrate the process, the following case is considered: $n = 100$ and $p = 5{,}000$. That is, 100 samples of a random variable G are generated from the normal distribution $N(-2.33, 1.0^2)$. The mean of -2.33 is chosen in such a way that the true probability of failure $\mathrm{Prob}(G \geq 0)$ is 1.0%. Pretending that the statistical parameters (mean μ, standard deviation σ, or probability of failure P_f) are unknown, these parameters along with their confidence intervals will be estimated using the bootstrap method.

Using the given set of 100 initial samples, 5,000 bootstrap re-samplings are performed. Similar to the conservative estimations in the previous section, the distribution type is first assumed to be normal. Using each set of bootstrap re-samples, the mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ are estimated, from which the estimated probability of failure $\hat{P}_f$ is calculated. The 5,000 $\hat{P}_f$ values define the empirical bootstrap distribution of the estimator $\hat{P}_f$.

The empirical bootstrap distribution can be used to minimize the risk of yielding unconservative estimates. In other words, we want to find a procedure that calculates the following quantity:

$$\alpha = P\left(\hat{P}_f \geq P_f\right) \tag{12}$$

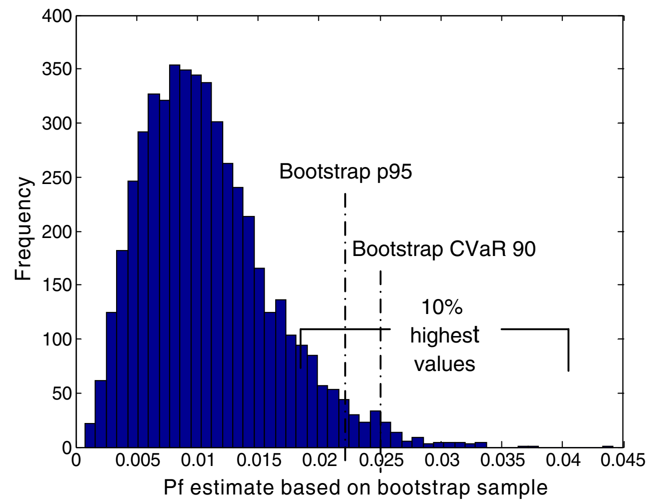

Fig. 5 Conservative estimators of P_f from bootstrap distribution: 95th percentile (*p95*) and mean of the 10% highest values (*CVaR*)

A procedure that satisfies (12) is called an α-conservative estimator of P_f. For example, if $\alpha = 0.95$ is desired, then $\hat{P}_f$ is selected at the 95th percentile of the bootstrap distribution of the estimated probability of failure. This estimator is referred as 'Bootstrap p95' (see Fig. 5). Due to the finite sample size, however, (12) is satisfied only approximately.

Besides this α-conservative estimator, the mean of the δ-highest bootstrap values (conditional value-at-risk CVaR, (Holton 2003)) is also used as a conservative estimate. Here $\delta = 10\%$ is used, so the estimator consists of the mean of the 10% highest bootstrap values. Since CVaR is a mean value, it is more stable than the α-conservative estimator. However, it is difficult to determine the value of δ that makes (12) satisfied precisely. This estimator is referred to 'Bootstrap CVaR 90' (see Fig. 5). Note that any bootstrap percentile higher than 50% is a conservative estimator. A very high α or low δ increases the value of $\hat{P}_f$ and yields over-conservative estimation.

5 Accuracy and conservativeness of conservative estimates for normal distribution

The goal of this section is to evaluate the accuracy and the conservativeness of the estimators presented in Sections 3 and 4 when the actual distribution and P_f is known to be normal. Statistical measures of the estimators are evaluated by estimating P_f a large number of times.

We also introduce here the reliability index, which is denoted by β and related to the probability of failure as:

$$\beta = -\Phi^{-1}(P_f) \tag{13}$$

where Φ is the CDF of the standard normal distribution.

The reliability index is often used instead of P_f in reliability-based design because the range of β (typically between one and five) is more convenient and its variability is lower than P_f. It is important to note that since $-\Phi^{-1}$ is a monotonically decreasing function, a low probability corresponds to a high reliability index. Thus, a conservative estimation of β should not overestimate the true β (since a conservative estimation should not underestimate the true P_f). In the following, we present the results for both probability of failure and reliability index.

First, 100 samples of G are randomly generated from the normal distribution with mean -2.33 and standard deviation 1.0. The failure is defined as $G \geq 0$, which corresponds to an actual probability of failure of 1.0%. For a given set of samples, different estimators are employed to estimate P_f. Five different estimators are compared: the unbiased fitting, CSP, CEC, Bootstrap p95, and Bootstrap CVaR90 estimators. This procedure is repeated 5,000 times in order to evaluate the accuracy and conservativeness of each estimator. For the unbiased, CSP and CEC estimators, we tested both RMS and Kolmogorov–Smirnov distance criteria and found that their performance was comparable but using K-S distance slightly increased variability. So, results are presented for RMS criterion only.

Most of the estimated values exceed the actual probability of failure, but it is desirable to maintain a certain level of accuracy. Thus, the objective is to compare each estimator in terms of accuracy and conservativeness. $N(-2.33, 1.0^2)$

Table 2 shows the statistical results from 5,000 repetitions. Results are presented in the form of the mean value and the 90% symmetric confidence interval [5%; 95%]. For the probability of failure estimates, the lower bound of the confidence interval shows the conservativeness of the estimator; the mean and the upper bound show the accuracy and the variability of the estimator. A high lower bound means a high level of conservativeness, but a high mean and upper bound mean poor accuracy and high variability. For the reliability index estimates, the upper bound shows the conservativeness and the mean and lower bound the accuracy and variability.

First, the confidence interval of the unbiased estimator illustrates the risk of unconservative prediction: indeed, the lower bound is 0.37%, which means there is a five per cent chance to underestimate P_f by a factor of at least 2.7 ($1.0/0.37 = 2.7$). This result provides an incentive for finding a way to improve the conservativeness of the probability estimate.

The CSP and CEC estimators are biased on the conservative side. As expected, the CEC is more conservative than the CSP. As a consequence, CEC is more biased and the risk of large overestimation is increased. The CEC confidence interval shows that there is a 5% chance of overestimating P_f by at least a factor of 5.5, while this value is 3.6 for the CSP estimator; on the other hand, the CEC leads to 94% conservative results, while the CSP estimator leads to only 82% conservative results. The choice between the CSP and CEC estimators is a choice between accuracy and conservativeness.

The Bootstrap p95 estimator achieves 92% conservativeness and the Bootstrap CVaR90 93% conservativeness. From the upper bounds of both estimations, we find that the risk of overestimating P_f by at least a factor of 3.7 is 5%.

The amplitude of error in the reliability index β is much lower than the amplitude in the probability of failure. For the CEC estimator, the lower bound of the confidence interval corresponds to 31% error ($(2.33 - 1.6)/2.33 = 0.31$). For the bootstrap estimators, this error is reduced to 23%. The mean errors are 16% and 11%, respectively.

Bootstrap methods appear to be more efficient than the biased fitting (CSP and CEC) in terms of accuracy and conservativeness. For the equivalent level of conservativeness (92–94%), the level of bias is reduced and the risk of

Table 2 Means and confidence intervals of different estimates of $P(G \geq 0)$ and corresponding β values where G is the normal random variable $N(-2.33, 1.0^2)$

Estimators	Statistics obtained over 5,000 simulations				
	P_f(%)		β		% of cons. results*
	90% C.I.	Mean	90% C.I.	Mean	
Unbiased	[0.37; 2.1]	1.02	[2.0; 2.7]	2.34	48
CSP	[0.63; 3.6]	1.86	[1.8; 2.5]	2.12	82
CEC	[0.95; 5.5]	2.97	[1.6; 2.3]	1.96	94
Boot. p95	[0.83; 3.7]	2.06	[1.8; 2.4]	2.07	92
Boot. CVaR90	[0.88; 3.8]	2.15	[1.8; 2.4]	2.05	93
Actual	1.00		2.33		

*Refers to the percent of the 5,000 simulation runs that resulted in a conservative estimate for each method

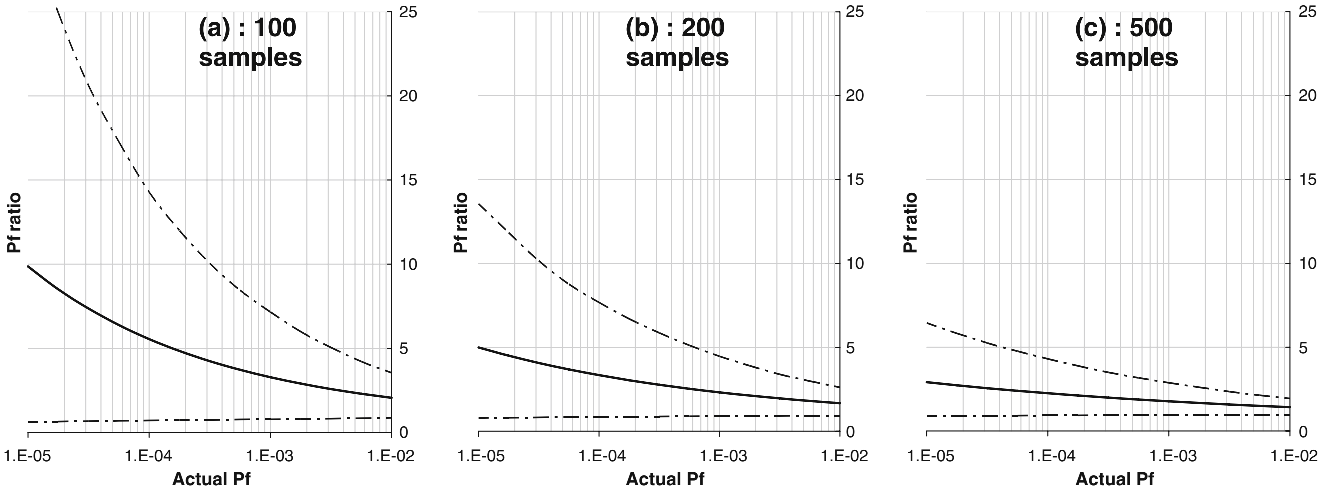

Fig. 6 Mean and confidence intervals of the bootstrap p95 conservative estimators for Normal distribution based on **a** 100, **b** 200, and **c** 500 samples. x-axis is the true probability (lognormal scale), and y-axis is the ratio between the estimate and the true probability. Variability increases when target probability or sample size are smaller

overestimation is lower when the bootstrap method is used. However, as mentioned earlier, the bootstrap method needs a minimum sample size to be used. It has been observed that when very small samples are available (10 to 50 data), the accuracy of the bootstrap method drops dramatically. In such a case, optimization-based methods should be used instead.

6 Effect of sample sizes and probability of failure on estimates quality

In the previous section, we showed that the bias in the conservative estimate can lead to large overestimations of the probability of failure. The magnitude of such an error mainly depends on two factors: the sample size and the value of the true probability. Indeed, increasing the sample size will reduce the variability of CDF fitting and, as a consequence, the upper bound of the confidence interval. Meanwhile, in order to estimate a lower value of the probability of failure, we need to use the tail of the CDF, which increases the variability of the estimation.

Controlling the level of uncertainty is crucial in optimization in order to avoid over-design. In this section, we quantify a measure of the uncertainty in the conservative estimate as a function of the sample size and the value of the actual P_f. Such a measure can help with deciding on the appropriate sample size to compute the estimate.

Bootstrap p95 performed well based on the previous example. Thus, in this section we consider only this estimator. We study two distribution cases: standard normal distribution and lognormal distribution with parameters $\mu = 0$ and $\sigma = 1$ (mean and standard deviation of the logarithm of the variable). Three sample sizes are considered: 100, 200, and 500 and seven probabilities of failure are estimated: (1×10^{-5}, 3×10^{-5}, 1×10^{-4}, 3×10^{-4}, 1×10^{-3}, 3×10^{-3} and 1×10^{-2}).[1]

For a given distribution, sample size, and P_f, the mean and 90% confidence intervals are calculated using 5,000 repetitions. Results are presented in Fig. 6 for normal distribution and in Fig. 7 for lognormal. The accuracy is measured in terms of ratios of the estimate over the true probability of failure.

As expected, the variability of $\hat{P}_f$ increases when the sample size and actual P_f decrease. Here, the most unfavorable case is when the sample size is equal to 100 and the actual P_f is equal to 10^{-5}. In such a case, for both distributions there is a 5% chance to overestimate P_f by more than 25 times its actual value! On the other hand, the case with 500 samples leads to a very reasonable variability. For the three sample sizes and all target probabilities, the lower bound of the ratio is equal to one, which means that the 95% conservativeness does not depend on these factors. The bootstrap estimate performances are remarkably similar for the two distributions, even though the distribution shapes are very different.

For any given reliability analysis problem, careful attention needs to be given to the accuracy of probability of failure estimates. The graphs in Figs. 6 and 7 address this issue. They show the confidence intervals and therefore define adequate sample sizes needed to compute reliable estimates. In terms of cost-effectiveness, the figures indicate

[1]For the normal distribution, the failures are defined for G greater, respectively, than 4.26, 4.01, 3.72, 3.43, 3.09, 2.75 and 2.33; for the lognormal case, the values are 71.2, 55.3, 41.2, 30.9, 22.0, 15.6 and 10.2.

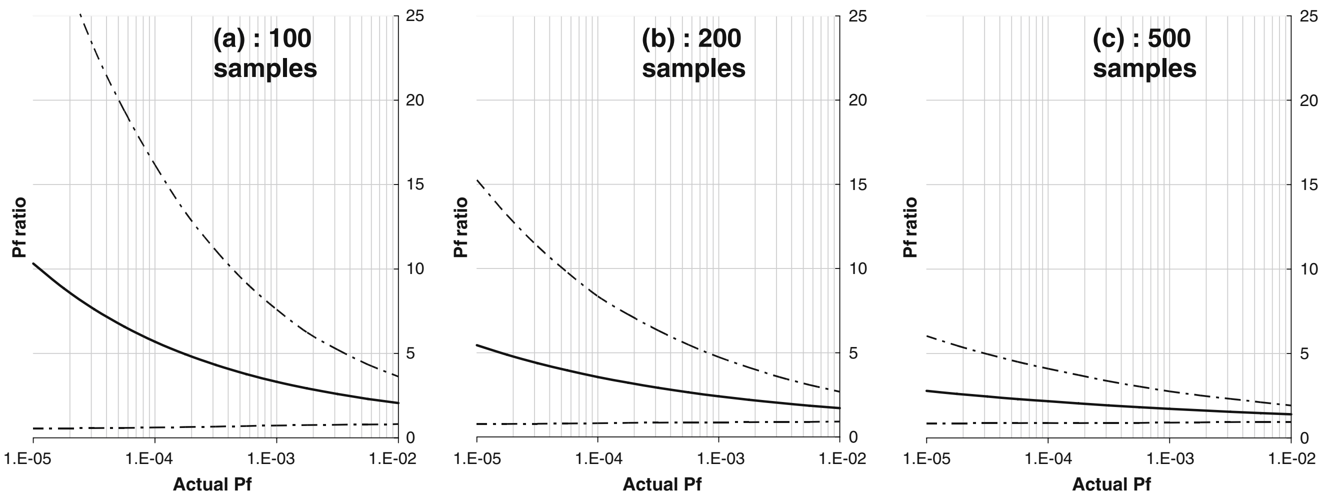

Fig. 7 Mean and confidence intervals of the bootstrap p95 conservative estimators for lognormal distribution based on **a** 100, **b** 200, and **c** 500 samples. Results are almost identical to the normal distribution case

that it may be smart to allocate greater number of simulations to low probability designs than to high probability design in order to get a constant level of relative accuracy.

7 Application to a composite panel under thermal loading

In this section, conservative estimates are obtained for the probability of failure of a composite laminated panel under mechanical and thermal loadings. The panel is used for a liquid hydrogen tank. Cryogenic operating temperatures are responsible for large residual strains due to the different coefficients of thermal expansion of the fiber and the matrix, which is challenging in design.

Qu et al. (2003) performed the deterministic (with safety factors) and probabilistic design optimizations of composite laminates under cryogenic temperatures using response surface approximations for probability of failure calculations. Acar and Haftka (2005) found that using CDF estimations for strains improves the accuracy of probability of failure calculation. In this paper, the optimization problem addressed by Qu et al. (2003) is considered. The geometry, material parameters, and the loading conditions are taken from their paper.

7.1 Problem definition

The composite panel is subject to resultant stress caused by internal pressure ($N_x = 8.4 \times 10^5$ N/m and $N_y = 4.2 \times 10^5$ N/m) and thermal loading due to the operating temperature in the range of 20 K–300 K. The objective is to minimize the weight of the composite panel that is made of a symmetric balanced laminate with two ply angles $[\pm\theta_1, \pm\theta_2]_s$ (that means an eight-layer composite). The design variables are the ply angles and the ply thicknesses $[t_1,\ t_2]$. The geometry and loading condition are shown in Fig. 8. The thermal loading is defined by a stress free temperature of 422 K, and working temperature of 300 K to 20 K. The material used in the laminates composite is IM600/133 graphite-epoxy, defined by the mechanical properties listed in Table 3.

The minimum thickness of each layer is taken as 0.127 mm, which is based on the manufacturing constraints as well as preventing hydrogen leakage. The failure is defined when the strain values of the first ply exceed failure strains. The deterministic optimization problem is formulated as:

$$
\begin{array}{ll}
\underset{t_1,t_2,\theta_1,\theta_2}{\text{Minimize}} & h = 4\,(t_1 + t_2) \\
\text{s.t.} & t_1, t_2 \geq 0.127 \\
 & \varepsilon_1^L \leq S_F \varepsilon_1 \leq \varepsilon_1^U \\
 & \varepsilon_2^L \leq S_F \varepsilon_2 \leq \varepsilon_2^U \\
 & S_F\,|\gamma_{12}| \leq \gamma_{12}^U
\end{array}
\qquad (14)
$$

where the safety factor S_F is chosen at 1.4.

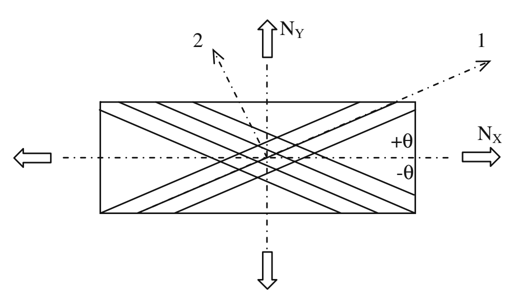

Fig. 8 Geometry and loading of the cryogenic laminate

Table 3 Mechanical properties of IM600/133 material

Elastic properties	E_1 (GPa)	147
	v_{12}	0.359
	E_2* (GPa)	[14 8]
	G_{12}* (GPa)	[8 4]
Coefficients of thermal expansion	α_1* (K^{-1})	$[-5 \times 10^{-7} \quad -1.5 \times 10^{-7}]$
	α_2* (K^{-1})	$[1 \times 10^{-5} \quad 3 \times 10^{-5}]$
Stress-free temperature	T_{zero} (K)	422
Failure strains	ε_1^U	0.0103
	ε_2^L	−0.013
	ε_2^U	0.0154
	γ_{12}^U	0.0138

*Temperature dependent; the numerical values in the bracket are the range for T going from 20 to 300 K

The analysis of the structural response is based on the classical lamination theory using temperature-dependent material properties. E_2, G_{12}, α_1 and α_2 are functions of temperature. Since the design must be feasible for the entire range of temperature, strain constraints are applied at 21 different temperatures, which are uniformly distributed from 20 K to 300 K. Details on the analysis and the temperature dependence of the properties are given in Qu et al. (2003). Their solutions for the deterministic optimization problem are summarized in Table 4. For those results, t_1 and t_2 were chosen only as multiples of 0.127 mm. Three optima are found with equal total thickness but different ply angles and ply thicknesses.

7.2 Reliability-based optimization problem

Given the material properties and design variables, the ply strains can be calculated using the classical lamination theory (Kwon and Berner 1997). Due to the manufacturing variability, the material properties and failure strains are considered random variables. All random variables are assumed to follow uncorrelated normal distributions. The coefficients of variation are given in Table 5. Since E_2, G_{12}, α_1 and α_2 are functions of the temperature, the mean values of the random variables are calculated for a given temperature, and then, a set of random samples are generated according to their distributions.

Table 4 Deterministic optima found by Qu et al. (2003)

θ_1 (deg)	θ_2 (deg)	t_1 (mm)	t_2 (mm)	h (mm)
27.04	27.04	0.254	0.381	2.540
0	28.16	0.127	0.508	2.540
25.16	27.31	0.127	0.508	2.540

Table 5 Coefficients of variation of the random variables

E_1, E_2, G_{12}, v_{12}	α_1, α_2	T_{zero}	$\varepsilon_1^L, \varepsilon_1^U$	$\varepsilon_2^L, \varepsilon_1^U, \gamma_{12}^U$
0.035	0.035	0.03	0.06	0.09

The transverse strain on the first ply (direction 2 in Fig. 8) turns out to be the most critical, and the effect of other strains on the probability of failure is negligible. Hence, the limit-state is defined as the difference between the critical strain and the failure strain:

$$G = \varepsilon_2 - \varepsilon_2^U \tag{15}$$

Note that the safety factor is not applied in the definition of limit state. The probability of failure is defined in (3) using the distribution F_G of the limit-state.

In order to determine which distribution type fits the best the limit-state G, we generated 1,000 samples at each of the three first optimum designs. Using a Lilliefors test (Kenney and Keeping 1951), we found that all the samples belong to a normal distribution. Hence, we assumed that the limit-state G is normally distributed for any design.

One might prefer not to assume a single distribution type over the design domain and test several distributions to find the one that fits best the data. Such a procedure is necessary for instance when the limit-state distribution varies from normal to heavy-tail within the design domain; assuming a single distribution can lead to large error in the reliability estimation. More generally, one has to keep in mind that the method proposed here can become hazardous if the determination of the distribution type is uncertain.

In our case, since the three designs considered are quite different from one another and have the same limit-state distribution, it is reasonable to assume that the limit-state distribution is always normal, hence reducing the computational burden.

The reliability-based optimization replaces the constraints on the strains in (14) by a constraint on the probability of failure. The target reliability of the cryogenic tank is chosen as 10^{-4}. Since the probability of failure can have a variation of several orders of magnitude from one design

Table 6 Variable range for response surface

Variables	Range
θ_1, θ_2 (deg)	[20 30]
t_1, t_2 (mm)	[0.127 0.800]

Table 7 Statistics of the PRS for the reliability index based on the unbiased and conservative data sets

Data sets	R^2	F	p-value
Unbiased estimates $\left\{\hat{\beta}_{unb}^{(1)}, \hat{\beta}_{unb}^{(2)}, \ldots, \hat{\beta}_{unb}^{(m)}\right\}$	0.96	138	$<10^{-6}$
Bootstrap conservative estimates $\left\{\hat{\beta}_{cons}^{(1)}, \hat{\beta}_{cons}^{(2)}, \ldots, \hat{\beta}_{cons}^{(m)}\right\}$	0.96	136	$<10^{-6}$

to another, it is preferable to solve the problem based on the reliability index:

$$\begin{aligned} &\underset{t_1,t_2,\theta_1,\theta_2}{\text{Minimize}} && h = 4\left(t_1 + t_2\right) \\ &\text{s.t.} && t_1, t_2 \geq 0.127 \\ &&& \beta\left(t_1, t_2, \theta_1, \theta_2\right) \geq -\Phi^{-1}\left(10^{-4}\right) = 3.719 \end{aligned} \tag{16}$$

7.3 Reliability-based optimization using conservative estimates

By solving (16) with a sampling-based estimate of reliability, we face the problem of having noise in the constraint evaluation, which can severely harm the optimization process. To address this issue, we chose to fit a polynomial response surface (PRS) to approximate the reliability index everywhere on a region of the design space, and solve the optimization with the response surface. The response surface is based on the estimation of the reliability index for a selected number of designs, called a design of experiments:$\{\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \ldots, \hat{\beta}^{(m)}\}$. All the estimates are done before the optimization process.

The range of the response surface is given in Table 6. We find that a fourth order polynomial provides a good approximation for the reliability index. The number of training points is taken as 500; the points are generated using Latin hypercube sampling to ensure a good space-filling property. Each estimate of the reliability index is based on the same $n = 200$ samples, which gives us a total number of 100,000 simulations to fit the response surface.

With the 3,000 simulations used to determine the distribution of the limit state, the total number of simulations is 103,000. This number is a reasonable total budget for solving the RBDO problem; indeed, there is no further simulation run during the optimization, the reliability being estimated by the PRS. To compare this number to classical MCS estimates, in order to achieve reasonable noise for a probability of failure of 10^{-4}, at least 10^6 samples are needed for a single evaluation, and the reliability is re-estimated at each optimization step. It is more difficult to compare to FORM and SORM methods, but due to the large number of random parameters in the problem, they are not efficient.

The reliability indexes are calculated using two methods: the unbiased fitting and the bootstrap conservative estimation with 95% confidence. The estimates are denoted, respectively, $\hat{\beta}_{unb}^{(i)}$ and $\hat{\beta}_{cons}^{(i)}$. Figure 6b shows that with 200 samples, the confidence interval of the conservative estimator for a probability of 10^{-4} is $[10^{-4}; 8 \times 10^{-4}]$, which corresponds to an error in β between 0 and 20%. Since the number of observations is much larger than the number of coefficients (500 compared to 70), the noise is filtered by the PRS.

A different response surface is fitted to each set of data. For both, the response surface is found to fit accurately the data. Table 7 shows the statistics of each response surface. Both R^2 values (percentage of variance explained) are very close to one, and p-values show that both models are very significant.

In addition, we computed accurate estimates of the reliability index at 22 designs uniformly chosen in the design space using separable Monte-Carlo method (SMC) (Smarslok et al. 2008) with 40,000 samples for each design. Table 8 shows the statistics based on the test points: the root mean square error (RMSE) between the 22 accurate responses and the response surface, error mean and number of unconservative predictions. Based on the test points, we can see that the first PRS is unbiased: the error mean is

Table 8 Statistics of the PRS based on 22 test points

Data sets	RMSE	Error mean	No. of unconservative predictions
Unbiased estimates $\left\{\hat{\beta}_{unb}^{(1)}, \hat{\beta}_{unb}^{(2)}, ..., \hat{\beta}_{unb}^{(m)}\right\}$	0.081	−0.01	10
Bootstrap conservative estimates $\left\{\hat{\beta}_{cons}^{(1)}, \hat{\beta}_{cons}^{(2)}, ..., \hat{\beta}_{cons}^{(m)}\right\}$	0.354	−0.34	0

Table 9 Optimal designs of the deterministic and probabilistic problems with unbiased and conservative data sets

	t_1	t_2	θ_1	θ_2	h	β from PRS	Actual β	P_f from PRS	Actual P_f (SD)*
Unbiased data set	0.127	0.507	22.2	30.0	2.536	3.72	3.61	10^{-4}	1.51e–4 (2.28e–6)
95% cons. data set	0.127	0.582	21.9	30.0	2.835	3.72	3.98	10^{-4}	3.50e–5 (6.3e–7)
Deterministic optima	0.127	0.416	20.0	30.0	2.170	X	2.97	X	15.0e–4 (9.6e–6)

*The standard deviation of the accurate P_f is computed using formula given in Smarslok et al. (2008)

approximately zero and there are ten unconservative errors for 12 conservative. On the other hand, the second PRS has large bias since the error mean is 0.34 and all the predictions at test points are conservative.

7.4 Optimization results

We present the results for the two probabilistic optimizations based on the response surfaces fitted on unbiased and conservative estimates. To compare deterministic and probabilistic approaches, the deterministic optimization as stated in (14) is also performed for the same range of the ply angles ([20 30] degrees). The optimization is performed using MATLAB's function fmincon repeated 20 times with different initial points to ensure global convergence. The optimal designs (best over the 20 optimizations) are given in Table 9. For these designs, an accurate estimate of the probability of failure is computed using SMC (Smarslok et al. 2008) with 40,000 samples.

The three optima are similar in terms of ply angles, and for all the first ply thickness t_1 is the lower bound; the significant difference is in the second ply thickness t_2. Both probabilistic designs are heavier than the deterministic optimum. The optimum found using unbiased dataset is substantially lighter than the other ($h = 2.54$ compared to 2.84). However, the accurate estimate of reliability shows that the optimum design using unbiased dataset violates the constraint. On the other hand, the design found using the conservative dataset is conservative; the actual probability is three times smaller than the target probability of failure. The deterministic design is very unconservative, its probability of failure being 15 times the target.

The fact that an unbiased strategy leads to an unconservative design is not surprising. Indeed, optimization is biased to explore regions where the error is 'favorable'; that is, where the constraint is underestimated. Using the conservative approach, the level of bias is sufficient to overcome this problem, but at the price of overdesign: since the probability of failure is three times the target, the actual optimum is lighter than the one we found.

We have shown that despite a very limited computational budget (103,000 MCS to solve the RBDO problem), it was possible to obtain a reasonable design by compensating the lack of information by taking conservative estimates.

8 Concluding remarks

The estimation of the probability of failure of a system is crucial in reliability analysis and design. In the context of expensive numerical experiments, or when a limited number of data samples are available, the direct use of Monte Carlo Simulation is not practical, and estimation of continuous distributions is necessary. However, classical techniques of estimation of distribution do not prevent dangerous underestimates of the probability of failure.

In this paper, several methods of estimating safely the probability of failure based on the limited number of samples are tested, when the sample distribution type is known. The first method constrains distribution fitting in order to bias the probability of failure estimate. The second method uses the bootstrap technique to obtain distributions of probability of failure estimators, and these distributions define conservative estimators.

In the case of samples generated from normal distributions, the numerical test case shows that both methods improve the chance of the estimation to be conservative. Bootstrap-based estimators outperformed constrained fits to the experimental CDF. That is, for the same confidence in the conservativeness of the probability estimate, the penalty in the accuracy of the estimate was substantially smaller. However, optimization based methods can be used when the sample size is very small, where the bootstrap method cannot be used.

We also explored the influence of sample sizes and target probability of failure on the quality of estimates. We found that larger sample sizes are required to avoid large variability in probability of failure estimates when that probability is small. The results indicate that when sampling at different points in design space, it may be more cost effective to have different numbers of samples at different points. Such approaches will be explored in a future work.

Finally, we have applied the conservative estimation procedures to perform the optimization of composite laminates at cryogenic temperatures. We compared the optimization results found where response surfaces are fitted to unbiased and conservative estimates respectively. We found that the unbiased response surfaces led to unsafe designs, while the conservative approach returned an acceptable design.

Acknowledgments This research was partly supported by National Science Foundation (Grant # 0423280). The authors gratefully acknowledge this support.

References

Acar E, Haftka RT (2005) The effects of uncertainty control measures on the weight saving from composite laminates at cryogenic temperatures. In: ASME-IDECT-2005, Long Beach, California, 24–28 September

Ben-Haim Y, Elishakoff I (1990) Convex models of uncertainty in applied mechanics. Elsevier, Amsterdam

Chernick MR (1999) Bootstrap methods, a practitioner's guide (Wiley series in probability and statistics). A Wiley-Interscience publication

Choi KK, Du L, Youn BD, Gorsich D (2005) Possibility-based design optimization method for design problems with both statistical and fuzzy input data. In: 6th world congress on structural and multidisciplinary optimization, 30 May–3 June

Du X, Sudjianto A, Huang B (2005) Reliability-based design under the mixture of random variables. ASME J Mech Des 127(6):1068–1076

Du L, Choi KK, Youn BD, Gorsich D (2006) Possibility-based design optimization method for design problems with both statistical and fuzzy input data. ASME J Mech Des 128(4):928–935

Efron B (1982) The jackknife, the bootstrap, and other resampling plans. SIAM, Philadelphia

Enevoldsen I, Sørensen J (1994) Reliability-based optimization in structural engineering. Struct Saf 15(3):169–196

Engelund S, Rackwitz R (1993) A benchmark study on importance sampling techniques in structural reliability. Struct Saf 12:255–276

Elishakoff IE (1990) An idea on the uncertainty triangle. Shock Vibr Dig 22(1):1–2

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

Holton GA (2003) Value-at-risk: theory and practice. Academic, New York

Kenney JF, Keeping ES (1951) Mathematics of statistics, 2nd edn. Van Nostrand, Princeton

Kim NH, Ramu P, Queipo NV (2006) Tail modeling in reliability-based design optimization for highly safe structural systems. In: 47th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics & materials conference

Kwon YW, Berner JM (1997) Matrix damage of fibrous composites: effect of thermal residual stresses and layer sequences. Comput Struct 64(1–4):375–382

Lee JO, Yang YS, Ruy WS (2002) A comparative study on reliability-index and target-performance-based probabilistic structural design optimization. Comput Struct 80(3–4):257–269

Melchers RE (1999) Structural reliability analysis and prediction, 2nd edn. Wiley, New York

Neal DM, Matthews WT, Vangel MG (1991) Uncertainty in obtaining high reliability fro stress-strength models. In: 9th DOD-/NASA/FAA conference on fibrous composites in structural design, vol 1. Dept. of Defense, Lake Tahoe, NV, pp 503–521

Nikolaidis E, Cudney HH, Chen S, Haftka RT, Rosca R (2004) Comparison of probability and possibility for design against catastrophic failure under uncertainty. ASME J Mech Des 126: 386–394

Qu X, Haftka RT (2004) Reliability-based design optimization using probability sufficiency factor. Struct Multidisc Optim 27(5):314–325

Qu X, Haftka RT, Venkataraman S, Johnson T (2003) Deterministic and reliability-based optimization of composite laminates for propeller tanks. AIAA J 41(10):2029–2036

Rackwitz R (2000) Reliability analysis—past, present and future. In: 8th ASCE specialty conference on probabilistic mechanics and structural reliability, South Bend, IN, USA, PMC 2000-RRR

Smarslok BP, Alexander D, Haftka RT, Carraro L, Ginsbourger D (2008) Separable Monte Carlo applied to laminated composite plates reliability. In: 49th AIAA/ASME/ ASCE/AHS/ASC structures, structural dynamics, and materials conference, Schaumburg, Illinois

Starnes JH Jr, Haftka RT (1979) Preliminary design of composite wings for bucking, stress and displacement constraints. J Aircr 16:564–570

Youn BD, Wang P (2007) Bayesian reliability-based design optimization using eigenvector dimension reduction (EDR) method. Structural multidisciplinary optimization. Springer, Heidelberg