Conservative estimations of reliability with limited sampling

Victor Picheny, Nam-Ho Kim, Raphael T Haftka

CONSERVATIVE ESTIMATIONS OF RELIABILITY WITH LIMITED SAMPLING

Victor Picheny **Nam-Ho Kim** **Raphael T. Haftka**
University of Florida, Gainesville, FL 32611-6250
USA

ABSTRACT

The objective of this paper is to provide a method of safely estimating reliability based on small samples. First, it is shown that the commonly used estimators of the parameters of the normal distribution function are biased, and they tend to lead to unconservative estimates of reliability. Then, two ways of making this estimation conservative are proposed: (1) adding constraints when a distribution is fitted to the data to bias it to be conservative, and (2) using the bootstrap method to estimate the bias needed for a given level of conservativeness. The relationship between the accuracy and the conservativeness of the estimates is explored for a normal distribution. In particular, detailed results are presented for the case when the goal is 95% likelihood to be conservative. The bootstrap approach is found to be more accurate for this level of conservativeness. It is then applied to the reliability analysis of a composite panel under thermal loading. Finally, we explore the influence of sample sizes and target probability of failure on estimates quality, and show that for a constant level of conservativeness, small samples and low probabilities can lead to a high risk of large overestimation while this risk is limited to a very reasonable value for samples above.

Keywords: Reliability-based design, Probability of failure, Conservative estimation

1. INTRODUCTION

In analyzing mechanical systems, uncertainties in input parameters—such as material properties, geometric dimensions, or operating conditions—prevent us from taking the analysis results at face value. Quantification of the influence of these uncertainties on reliability is crucial. Engineering systems need to be designed so that the risk of failure should not exceed an acceptable value.

In the literature (e.g., Ref. [1]), many methods have been proposed to estimate the reliability of a system for design under uncertainties. When sampling-based methods are used for this purpose, randomness in reliability estimates is inevitable especially when a small sample is used. It has been shown that the errors in probability distributions due to insufficient information can have a large effect on probability calculation (e.g., Refs. [2] and [3]).

Accurate estimation of the reliability of a system requires a large number of samples. When the number of samples is limited due to computational or manufacturing costs, the available information is insufficient to accurately estimate the reliability of the system. However, it is possible to compensate for the lack of information by using reliability estimations that are biased to be safe. In this paper, this is called conservative estimation [4]. FAA defines conservative material property (A-basis and B-basis) as the value of a material property exceeded by 99% (for A-basis, 90% for B-basis) of the population with 95% confidence. As the conservative estimations tend to overestimate (or underestimate) the target values, each conservative estimator needs a trade-off analysis between accuracy and the level of conservativeness (i.e., chance of being conservative).

In this paper, we focus on the case when the probability of failure, P_f, of a system is estimated from a small 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 distribution fitting used to compute the estimator of P_f; we also explore the use of bootstrap method [5] [6] for probability of failure estimations, and defining conservative estimators based on bootstrapping. Finally, the relation between accuracy and the level of conservativeness is studied with the help of numerical examples.

In the next section, we discuss how we use the 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 estimator. The accuracy of such estimators is analyzed using a simple numerical example in Section 5, and the conservative estimators are applied to an engineering problem in Section 6. Finally, in Section 7 we analyze the effects of sample sizes and target probability of failure on the quality of the estimates, followed by concluding remarks in Section 8.

2. PROBABILITY OF FAILURE CALCULATION FROM SAMPLES

2.1. Limit-State Function and Probability of Failure

Failure of a system can usually be determined through a criterion, called a limit-state function. For instance, the limit-state function of a structure is often defined as the difference between its response, R, (e.g., maximum stress or strain) and its capacity, C, (e.g., maximum allowable stress or strain):

$$G(X) = R(X) - C(X) \tag{1}$$

Both response and capacity can be considered as a function of random variables, X.

When probability distributions of the random variables are given, Monte Carlo Simulations (MCS) can be used to generate samples of the limit-state [7]. The standard use of MCS is to calculate the number of time failure occurs. The ratio between the numbers of failures and the total number of samples approximates the probability of failure of the system. However, the accuracy of MCS strongly depends on the number of samples used. When low probabilities are considered, a limited number of samples cannot evaluate the probability of failure accurately enough. A popular solution is to evaluate low probabilities by fitting a distribution to the samples.

Given a set of samples of system response $g_1, g_2, \ldots, g_n$, estimating the probability of failure is equivalent to estimating the Cumulative Distribution Function (CDF) $F_G(g)$ of the limit-state at $g = 0$:

$$P_f = P(g \geq 0) = 1 - F_G(0) \tag{2}$$

In the following part, we discuss several alternatives of estimating CDF from a set of samples.

2.2. Various Estimates of CDF

When estimating a CDF from a set of samples, the first step is to choose the distribution type that fits the data best. Other alternatives are possible, such as using monotonic splines or composite distributions, which will not be discussed in this paper. In this paper, we consider only normally distributed data because it is the most common distribution and needs only two parameters to be defined. However, the approach presented here can be applied to other distributions.

a. Classical estimator: Assuming a certain distribution type, an analytical model of CDF can be fitted to the samples by adjusting its parameters. In the case of normal distribution, for example, we want to estimate the mean μ and the standard deviation σ. The classical estimations of the mean and the standard deviation from a sample of size n are, respectively:

Classical estimator:

$$\overline{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.1}$$

$$\overline{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{\mu})^2}\,, \tag{3.2}$$

where the estimator $\overline{\sigma}$ of the standard deviation is normalized by $(n-1)$ to make $\overline{\sigma}^2$ the best unbiased estimator of the variance σ^2 [8].

However, while $\overline{\sigma}^2$ is an unbiased estimator of σ^2, $\overline{\sigma}$ is a slightly biased estimate of σ [6]. The expected value of $\overline{\sigma}$ can be approximated by:

$$E(\overline{\sigma}) \approx \sigma\sqrt{\left(\frac{1+\sqrt{\frac{n-3}{n-1}}}{2}\right)} \tag{4}$$

(That is, 0.25% for n = 100, 3% for n = 10)

As a consequence, the standard deviation is likely to be underestimated; so the tail of the estimated CDF will be biased.

b. Estimators based on fitting the empirical CDF: Consider n samples are arranged in increasing order: $(x_1 \leq x_2 \leq \cdots \leq x_n)$. Then, the empirical CDF is defined as:

$$F_X(x) = \begin{cases} 0 & \text{for } x \leq x_1 \\ \dfrac{k}{n} & \text{for } x_k \leq x \leq x_{k+1} \\ 1 & \text{for } x_n \leq x \end{cases} \tag{5}$$

It is then possible to estimate the mean and standard deviation of the CDF that approximates the empirical CDF best. Two different ways of approximation are studied here: (1) minimizing the root-mean-square (RMS) error between the estimated CDF and the empirical CDF, and (2) minimizing the Kolmogorov-Smirnov distance [9].

To minimize the RMS error between the empirical and the estimated CDF, errors are calculated at 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 Figure 2):

$$F_X(x_k) = \frac{k - \frac{1}{2}}{n}, \quad 1 \leq k \leq n\,. \tag{6}$$

The parameters $(\overline{\mu}, \overline{\sigma})$ will then be calculated by solving the following optimization problem:

$$\underset{\overline{\mu},\overline{\sigma}}{\text{Minimize}} \frac{1}{n}\sum_{k=1}^{n}\left(F_{\overline{\mu},\overline{\sigma}}(x_k) - \frac{k-\frac{1}{2}}{n}\right)^2 \tag{7}$$

where $F_{\overline{\mu},\overline{\sigma}}$ is the value of the CDF of a normal distribution with parameters $(\overline{\mu}, \overline{\sigma})$:

$$F_{\overline{\mu},\overline{\sigma}}(x_k) = \frac{1}{\overline{\sigma}\sqrt{2\pi}}\int_{-\infty}^{x_k} \exp\left[\frac{-(t-\overline{\mu})^2}{2\overline{\sigma}^2}\right]dt \tag{8}$$

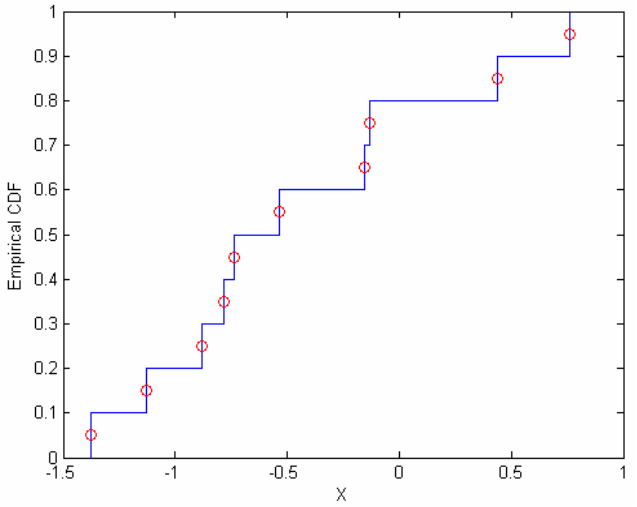

Figure 2 (a): Example of points (circles) chosen to fit an empirical CDF (line) obtained by sampling 10 points from N(0,1).

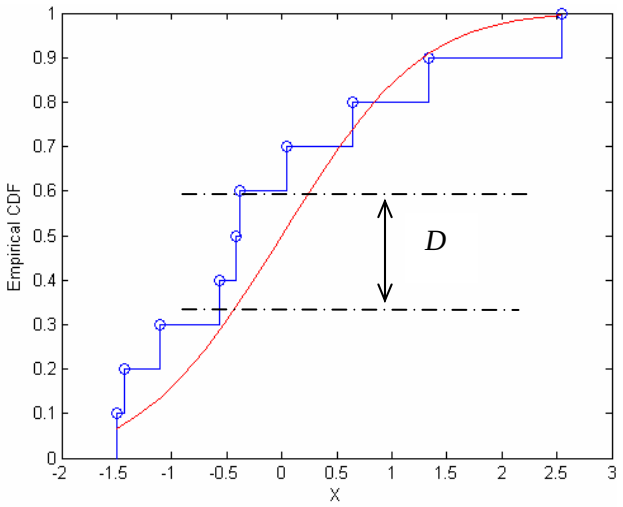

Figure 2 (b): Example of a K-S distance between an empirical CDF (staircase) and a normal CDF (continuous line).

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 Figure 2). The maximum distance occurs at one data point. The optimization problem for the K-S distance becomes:

$$\underset{\bar{\mu},\bar{\sigma}}{\text{Minimize}}\left[\max_{1\le k\le n}\left(\left|F_{\bar{\mu},\bar{\sigma}}(x_k)-\frac{k}{n}\right|,\left|F_{\bar{\mu},\bar{\sigma}}(x_k)-\frac{k-1}{n}\right|\right)\right] \quad (9)$$

3. CONSERVATIVE ESTIMATES USING BIASED FITTING

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 more conservative. We will also present the relationship between conservativeness and accuracy.

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 Eq.2, we see that such an estimate can be obtained by constraining the estimated CDF to be less or equal than the true CDF when the parameters are found through the optimization problem in Eq. 7. Besides, failure occurs when the critical variable happens to be far from its mean value in one direction. Therefore, these constraints will be applied to the right half of the data.

One conservative estimate of the CDF can be obtained by constraining the estimate to pass below the data points. A second can be obtained by constraining it to pass below the entire empirical CDF. They will be called, respectively, CSP (Conservative at Sample Points) and CEC (Conservative to Experimental CDF). The latter has 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:

$$F_{\mu,\sigma}(x_i)-\frac{i}{n}\le 0 \quad \text{for} \quad \frac{n}{2}\le i\le n \quad (10)$$

CEC constraints:

$$F_{\mu,\sigma}(x_i)-\frac{i-1}{n}\le 0 \quad \text{for} \quad \frac{n}{2}\le i\le n \quad (11)$$

Example

To illustrate conservative estimators, a random variable X with 10 sample points generated from $N(0,1)$ is used. The probability of failure is defined as the probability that X is larger than x_{limit}:

$$G = X - x_{\text{limit}} \ge 0 \quad (12)$$

We chose $x_{\text{limit}} = 2.33$, corresponding to a probability of failure is 1 %. Figure 3 shows the empirical CDF along with the three estimates based on 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.

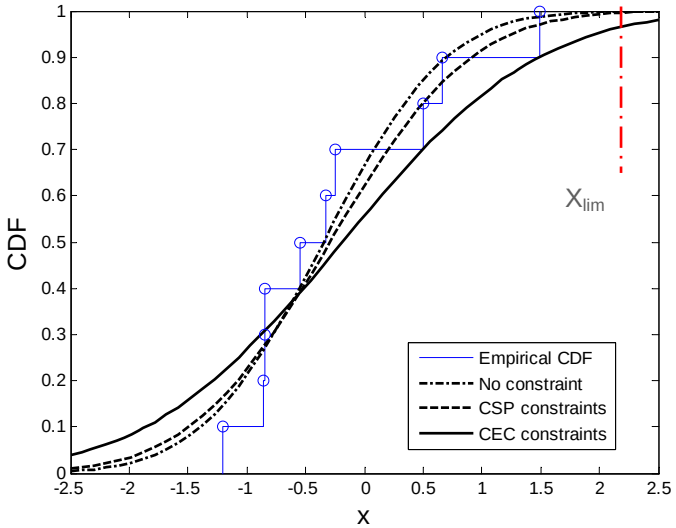

Figure 3: Example of CDF estimators based on RMS error for a sample of size 10 generated from N(0,1)

Table 1: Comparison of the mean, standard deviation, and probability of failure of the three different CDF estimators for $N(0,1)$. Exact values are $\mu = 0$, $\sigma = 1$, $P_f = 1.0\%$ (Failure if X > 2.33).

	No constraint	CSP	CEC
$\overline{\mu}$	–0.22	-0.29	–0.19
$\overline{\sigma}$	0.85	0.94	1.31
$\hat{P}_f$	0.14%	0.27%	2.74%

We can see on the graph the effect of the constraints: the CSP estimator is shifted down to be under the eighth data point; resulting the CDF at the tail is decreased. The CEC estimator is shifted even more by the combined effects of the eighth and the tenth points. Since the conservative estimators are unconstrained on the left half of the distribution, their CDF curves cross the empirical curve on this side.

In this example, the minimum RMS error with no constraint is strongly unconservative even if an unbiased estimation is used. The CSP estimate is unconservative, but substantially less than the unbiased estimate. The CEC estimate is conservative. In order to generalize these results and derive reliable conclusions, statistical experiments based on large number of simulations will be performed in Section 5.

4. CONSERVATIVE ESTIMATES USING THE BOOTSTRAP METHOD

4.1. Bootstrap Method

When only a small number of samples are available, the bootstrap method can provide an efficient way of estimating the distribution of a statistical parameter θ (for example, the mean of a population) using the re-sampling technique (Ref. [5] [6]). The idea is to create many sets of bootstrap samples by re-sampling with replacement from the original data. Then, the distribution of θ can be approximated by the empirical distribution of the parameter $\hat{\theta}$, estimate of θ computed from each set of the bootstrap samples. 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, while the number of bootstrap re-samplings is p. Each re-sampling can be performed by randomly selecting n data out of the n initial samples. Since the re-sampling procedure allows selecting data with replacement, the statistical properties of the re-sampled data are different from that of the original data. 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 must be larger than 100 [5]. A typical number of bootstrap re-samplings is typically from 500 to 5,000.

4.2. Estimation of Probability of Failure using the Bootstrap Method

For illustrating the process, we present the approach for $n = 100$ and $p = 5,000$. That is, 100 samples of a random variable X are generated from the standard normal distribution $N(0,1)$. The limit-state function is defined such that failure occurs when $G = X - 2.33 \geq 0$ (the actual probability of failure is 1.0%). Pretending that the statistical parameters (mean μ, standard deviation σ, or probability of failure P_f) are unknown, these parameters can be estimated from the samples. However, the confidence interval of the estimated parameters is unknown with one set of samples.

Using the given set of the initial samples, 5,000 bootstrap re-samplings are performed. From the estimated mean and the standard deviation of each set of bootstrap re-samples, the probability of failure estimate $\hat{P}_f$ is computed. 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 estimate $\hat{P}_f$. In other words, we want to find a procedure that maximizes the quantity:

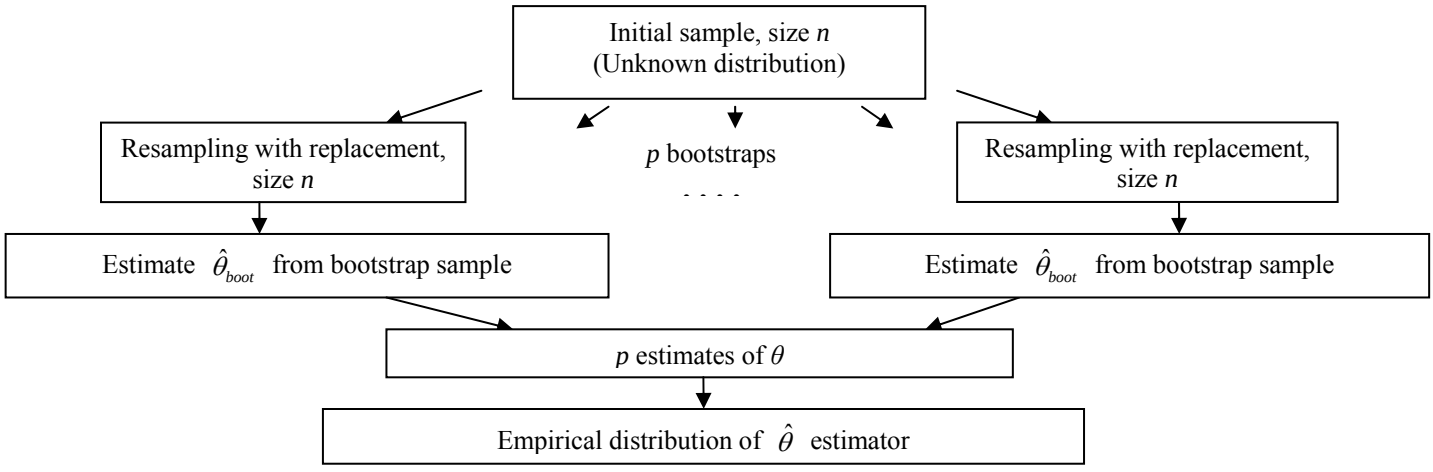

Figure 4: Schematic representation of bootstrapping. Bootstrap distribution of θ is obtained by multiple resampling (here p times) from a single set of data.

$$\alpha = P(\hat{P}_f \geq P_f) \quad (13)$$

A procedure that satisfies Eq. (13) is called an α-conservative estimator of P_f. For example, if α = 0.95 is desired, then $\hat{P}_f$ is selected at the 95th percentile of the bootstrap distribution of the probability of failure. Because of the finite sample, however, Eq. (13) will be satisfied only approximately.

Besides the α-percentile, we also use as conservative estimate the mean of the γ highest bootstrap values (CVaR, [10]). Since CVaR is a mean value, it is assumed to be more stable than a percentile. However, it is difficult to determine the value of γ that makes (13) satisfied precisely. Here we use $\gamma = 10\%$. These estimators are called, respectively, *Bootstrap p95* and *Bootstrap CVaR 90* (see *Figure 5*). Note that any bootstrap quantile higher than 50% is a conservative estimator. A very high α or low γ will increase the value of $\hat{P}_f$ and will yield over-conservative estimation.

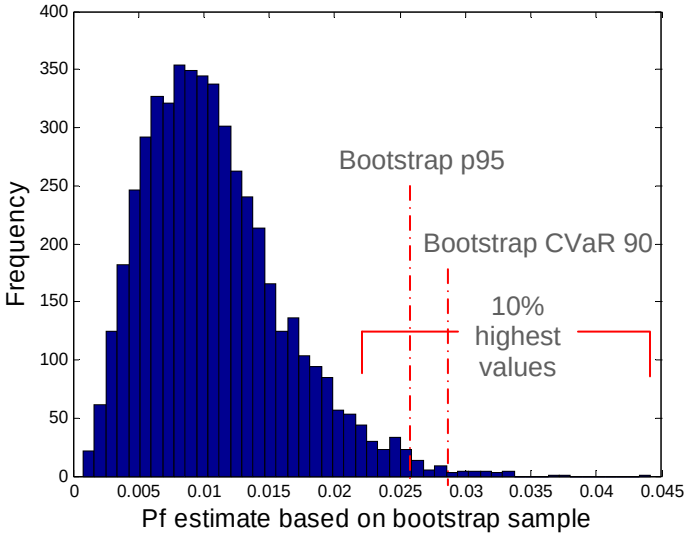

Figure 5: Conservative estimators of P_f from bootstrap distribution: 95th percentile (p95) and mean of the 10% highest values (CVaR).

5. STATISTICAL BEHAVIORS OF CONSERVATIVE ESTIMATES

The goal of this section is to evaluate the accuracy and the conservativeness of the estimators presented in Sections 3 and 4, using a simple numerical example, where the actual distribution and P_f are known. In addition, the 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:

$$\beta = -\Phi^{-1}\left(P_f\right) \quad (14)$$

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 values (typically between one and five) is more convenient and its variability lower than P_f's. It is important to notice that since $-\Phi^{-1}$ is a monotonically decreasing function, a low probability corresponds to a high reliability index. Thus, a conservative estimation $\hat{\beta}$ of β should not *overestimate* the true β (while conversely a conservative estimation $\hat{P}_f$ of P_f 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 X are randomly generated from the standard normal distribution $N(0,1)$. The failure is defined for $X \geq 2.33$, 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 classical, 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 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 increases variability. So, results are presented for RMS criterion only.

Most of the estimated values will exceed the actual probability of failure, but it is desired to maintain a certain level of accuracy. Thus, the objective is to compare each estimator in terms of accuracy and conservativeness. Table 2 presents the results 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.

As shown in Eq. 4, the standard deviation of the classical estimator is biased. As a result, more than 50% of the $\hat{P}_f$ calculated from the classical estimator are less than the actual one. Moreover, the lower bound of the confidence interval is 0.37%, which means there is a five per cent chance to underestimate P_f by a factor of at least 2.7 (1.00/0.37=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 overestimate is increased. The CEC confidence interval shows that there is a five per cent chance to overestimate P_f by at least a factor of 5.5, while this value is 3.6 for the CSP estimator; on the other hand it leads to 94 percent conservative results, while the CSP estimator leads to only 82 percent conservative results. The choice between the CSP and CEC estimators will be a choice between accuracy and conservativeness.

The Bootstrap p95 estimator achieves 92 percent conservativeness and the Bootstrap CVaR90 93 percent conservativeness. From the upper bounds of both estimations,

Table 2: Means and confidence intervals of different estimates $\hat{P}_f$ of $X \geq 2.33$ and corresponding β values where X is the standard normal random variable

Estimators	Statistics obtained over 5000 simulations				
	Pf (%)		*Beta*		% of cons. results*
	90% C.I.	Mean	90% C.I.	Mean	
Classical	[0.37 ; 2.1]	1.05	[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 5000 simulation runs that resulted in a conservative estimate for each method.*

we see that the risk of overestimating P_f by at least a factor of 3.7 is five percent.

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 respectively 16% and 11%.

Bootstrap methods appear to be more efficient than the biased fitting (CSP and CEC) in terms of accuracy and conservativeness. For an equivalent level of conservativeness (92-94 percent), the level of bias is reduced and the risk of overestimations is lower. 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, the optimization based methods should be used instead.

6. APPLICATION TO A COMPOSITE PANEL UNDER THERMAL LAODING

In this section, the conservative estimates are applied to evaluate the probability of failure of a composite laminates panel under mechanical and thermal loadings. The panel is used for a hydrogen tank in aerospace structures. The 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) [11] performed the deterministic and probabilistic design optimizations of composite laminates under cryogenic temperatures, using response surface approximations for probability of failure calculations. Acar and Haftka (2005) [12] found that using CDF estimations for strains improves the accuracy of probability of failure calculation. In this paper, the optimization problem that is addressed by Qu et al. (2003) [11] is considered. The geometry, material parameters and the loading conditions are taken from that paper. The objective is to explore the possibilities to improve the estimation of the probability of failure calculations in a conservative way.

6.1. Problem Definition

The composite panel is subject to resultant stress caused by mechanical loading (N_x is 33 MPa and N_y is 16 MPa) and thermal loading due to the operating temperature 20K – 300K (Figure 6). The objective is to minimize the weight of the composite panel that is a symmetric balanced laminate with two ply angles (that means an eight-layer composite) $[\pm\theta_1, \pm\theta_2]$. The design variables are the ply angles and the ply thicknesses $[t_1, t_2]$. The geometry and loading condition are shown in Figure 6.

The material used in the laminates composite is IM600/133 graphite-epoxy, defined by the mechanical properties listed in Table 3.

Table 3: Mechanical properties of the laminate composite

Elastic properties	E_1, E_2, G_{12} and ν_{12}
Coefficients of thermal expansion	α_1 and α_2
Stress-free temperature	T_{zero}
Failure strains	$\varepsilon_1^L, \varepsilon_1^U, \varepsilon_2^L, \varepsilon_2^U$ and γ_{12}^U
Safety factor	S_F

The minimum thickness of each layer is taken as 0.127 mm, which is based on the manufacturing constraints as well as for 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:

$$\begin{aligned} &\textit{Minimize} && h = 4(t_1 + t_2) \\ &s.t. && t_1, t_2 \geq 0.127 \\ & && \varepsilon_1^L \leq S_F \varepsilon_1 \leq \varepsilon_1^U \\ & && \varepsilon_2^L \leq S_F \varepsilon_2 \leq \varepsilon_2^U \\ & && S_F \,|\gamma_{12}| \leq \gamma_{12}^U \end{aligned} \tag{15}$$

where S_F is chosen at 1.4.

The solutions for the deterministic optimization problem found by Qu *et al*. (2003)[11] are summarized in Table 4. Three optima are found with equal total thickness but different ply angles and ply thicknesses.

Table 4: Deterministic optima found by Qu *et al*. (2003)[11]

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

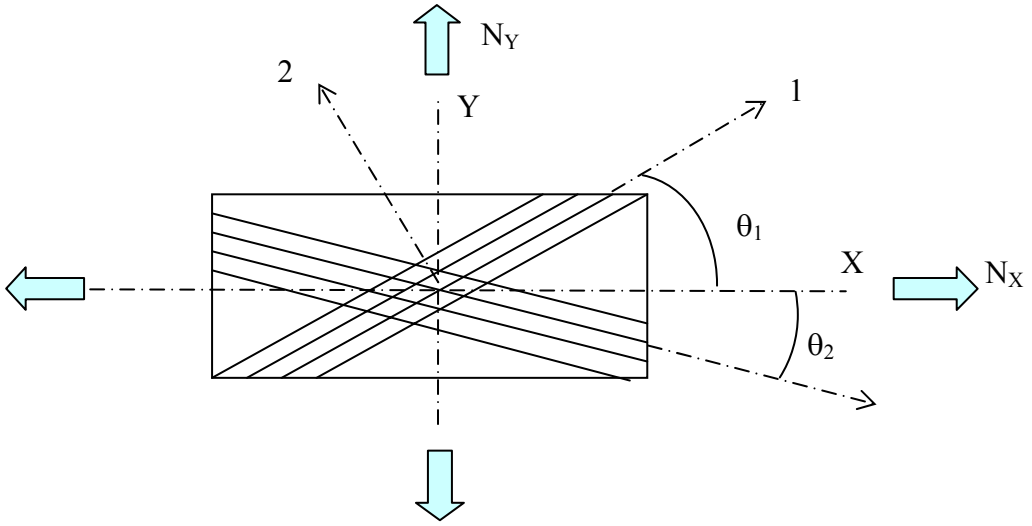

Figure 6: Geometry and loading of the cryogenic laminate

6.2. Calculation of the Probability of Failure

Given the material properties and the design variables at the optimum design, the ply strains can be calculated using Classical Lamination Theory. Due to the manufacturing variability, the material properties and failure strains are considered as random variables. All random variables are assumed to follow uncorrelated normal distributions. The coefficients of variation are given in Table 5. $E_2, G_{12}, \alpha_1, \text{and } \alpha_2$ are a function of the 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 20K to 300K. First, 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. The mean of the other parameters are given in Table 6.

Table 5: Coefficients of variation of the random variables

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

Table 6: Mean of random parameters

E_1	ν_{12}	T_{zero}	ε_1^L
21.5x10^6	0.359	300	-0.0109
ε_1^U	ε_2^L	ε_2^U	γ_{12}^U
0.0103	-0.013	0.0154	0.0138

The critical strain is the transverse strain on the first ply (direction 2 in Figure 6). The limit-state is defined as the difference between the critical strain and the failure strain:

$$G = \varepsilon_2 - \varepsilon_2^U \tag{16}$$

Then, the probability of failure is given as:

$$P_f = 1 - F_G(0), \tag{17}$$

where F_G is the CDF of the limit-state.

The probabilistic distribution of the limit-state function is in general unknown. In a real-world problem, one could generate a large sample prior to the optimization to determine the distribution type, for instance in the middle of the optimization domain, and then assumes that the limit state follows that distribution everywhere inside the domain. In this paper, we generate 1,000 samples at the first optimum design $(\theta_1 = \theta_2 = 27.04)$ to determine which distribution type fits the best the critical strain data. First, we standardize the data by subtracting the mean of the sample and divide by the standard deviation of the sample; then, we perform a Kolmogorov-Smirnov test to determine whether or not the standardized sample belongs to a standard normal distribution.

The null hypothesis H_0 tested is that the data has a standard normal distribution. The test statistic is the maximum distance *KS* between the sample empirical distribution and standard normal distribution. We reject the null hypothesis if *KS* is greater than a certain value.
Here we found for 1,000 samples:
KS = 0.0163; Rejection region: reject H_0 if: $KS \geq 0.0428$
We cannot reject the null hypothesis, so we conclude that the data is normal. The power of the test is 0.9519.

Since both critical strain and failure strain are normally distributed, the limit-state is also normally distributed. Thus, the probability of failure can be estimated by using the methods described in previous sections.

6.3 Results

At the first optimum design $(\theta_1 = \theta_2 = 27.04)$, 100 samples of critical strain are generated using the distributions of the input random variables and MCS. Using five different estimators, the mean and the standard deviation are estimated, from which the probability of failure $\hat{P}_f$ is calculated. This procedure is repeated 5,000 times in order to evaluate the statistical properties of the estimates. Since the exact probability of failure is unavailable, a large sample size (10^7) is used to compute P_f and compare it with the five different estimates.

Table 7: Means and confidence intervals of different estimates $\hat{P}_f$ ($\times 10^{-4}$) and corresponding β values of composite panel. The actual P_f is approximated using 10^7 samples.

Estimators	Statistics obtained over 5000 simulations				
	Pf (x10^{-4})		*Beta*		% of cons. results
	90% C.I.	Mean	90% C.I.	Mean	
Classical	[0.7 ; 15.4]	5.7	[3.0 ; 3.8]	3.36	48
CSP	[2.1 ; 47.5]	17.3	[2.6 ; 3.5]	3.03	85
CEC	[4.3 ; 78.5]	29.6	[2.4 ; 3.3]	2.85	95
Boot. p95	[3.2 ; 43.9]	17.6	[2.6 ; 3.4]	2.99	92
Boot. CVaR90	[3.5 ; 45.2]	18.5	[2.6 ; 3.4]	2.98	93
Actual	4.4		3.33		

Table 7 summarizes the results of five different estimators for the probability of failure and reliability index of the composite panel. The use of classical estimators of μ and σ leads to a five percent chance of underestimating P_f by at least a factor of 6 (4.4/0.7=6), which is strongly unconservative, whereas the CEC estimator is 95% conservative (the 5% percentile is equal to the actual P_f). However, the right tail of its distribution and the bias are very large. The CSP results conform to the bootstrap results in terms of mean and variability, but it is less conservative.

Bootstrap p95 is 92% conservative and Bootstrap CVaR90 is 93% conservative. The upper bounds of their confidence interval are almost two times lower than the CEC estimator, for an equivalent level of conservativeness. We can conclude that bootstrap estimators outperform the estimators based on biased fitting.

We see here that the error in P_f estimation can be very large. However, this error is reasonable in terms of reliability index: for the Bootstrap p95, the confidence interval shows that there is 95% chance that the error remains less than 22%. For the unbiased reliability index, this value is equal to 10%. Thus, we can consider in first approximation that a 95% conservativeness level doubles the error in the reliability index estimate compared to the unbiased estimate.

The overall performance of the conservative estimators is not as good as the numerical example in Section 4. Indeed, the actual probability of failure is of the order of 10^{-4} instead of 10^{-2} previously. Since we estimate the value of the CDF at a farther point in the tail, the variability is logically increased.

7. EFFECT OF SAMPLE SIZES AND TARGET PROBABILITY OF FAILURE ON ESTIMATES QUALITY

In Section 5, it is shown that the upper bound of the confidence interval using the Bootstrap estimators is 3.7 times the actual probability of failure (that means, there is five per cent chance to overestimate the true probability of failure by a factor of minimum 3.7). In Section 6, however, this ratio rises to the value of 10. Such a large difference is due to the values of the target probability: of the order of 10^{-2} and 10^{-4} for Tables 2 and 7, respectively. Indeed, 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. Another critical factor in the accuracy of the estimation is the sample size. Increasing the sample size will reduce the variability of CDF fitting and, as a consequence, the upper bound of the confidence interval.

Controlling the level of uncertainty is crucial in optimization in order to avoid over-design. We want to quantify a measure of the uncertainty in $\hat{P}_f$ as a function of the sample size and the value of the actual P_f. Such a measure can help deciding on the appropriate sample size to compute the estimate.

It turns out that Bootstrap p95 performs well based on the previous two examples. Thus, in this section we consider only this estimator. The estimator achieved a 92% conservativeness level with 100 samples on both cases; since the accuracy of the bootstrap method increases with sample size, the chance of being conservative will slightly increase with higher sample sizes. Thus, the focus in this section is on the risk of large overestimation; i.e., the 95% unilateral confidence interval of the conservative estimate.

To obtain a measure of the uncertainty in $\hat{P}_f$ as a function of the sample size and the actual P_f, three different sample sizes are used: 100, 200 and 500. In addition, seven different 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}). Since the samples are generated from standard normal distribution, the seven failures are defined for X greater, respectively, than 4.26, 4.01, 3.72, 3.43, 3.09, 2.75 and 2.33. For a given sample size and P_f, the upper bound of the confidence interval of the p95 estimator is calculated using 5000 repetitions. Results are presented in Figure 7 for probability of failure, and in Figure 8 for reliability index. The error is measured in terms of ratios for probabilities of failure, and in terms of relative error for reliability index.

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, there is a five percent chance to overestimate P_f by more than 30 times its actual value! On the other hand, the case with 500 samples leads to a very reasonable variability.

The relative error in β appears to be almost independent of the actual probability of failure. For a sample size of 100,

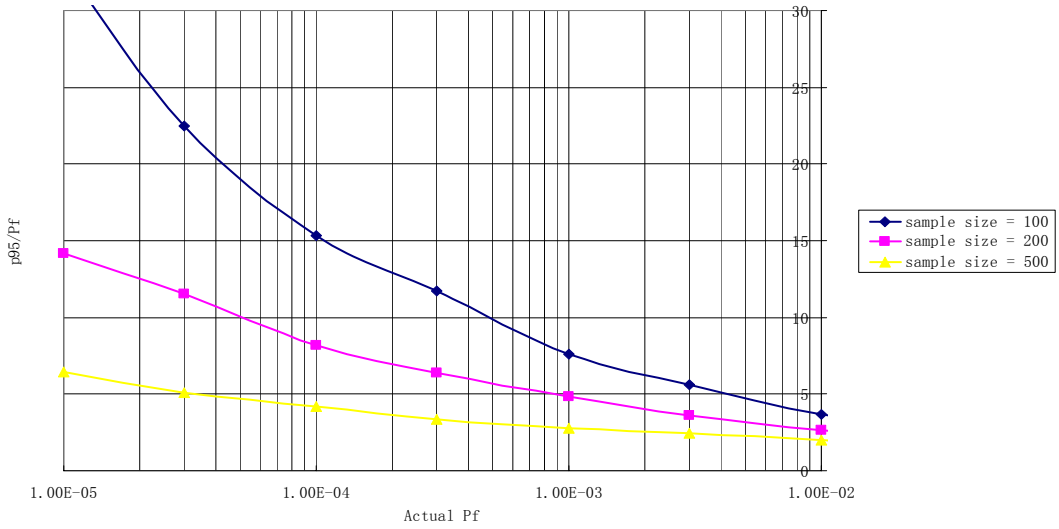

Figure 7: Evolution of the unilateral confidence interval of Bootstrap p95 with respect to the actual probability of failure for sample with N(0,1) distributions. Variability of Bootstrap p95 is greater for small sample sizes and low P_f.

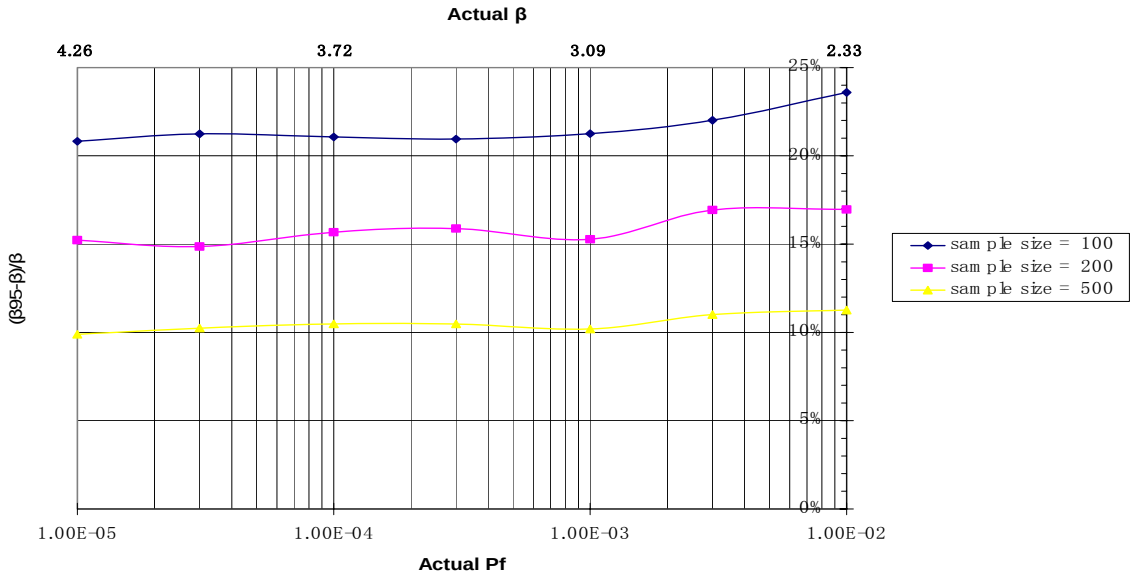

Figure 8: Evolution of the unilateral confidence interval of the reliability index of the Bootstrap P95 with respect to the actual probability of failure for sample with N(0,1) distributions. The relative error in β depends only on sample size and do not exceeds 25% with 100 samples.

there is 95% chance that the error remains below 25%. For a sample size of 500, the upper bound of the confidence interval is 11%.

For any given reliability analysis problem, careful attention needs to be given to the level of accuracy that is reached by probability of failure estimates. The graph in Figure 7 can address this issue by providing the adequate sample size to compute reliable estimates. In cost-effectiveness approach, it may also help deciding on allocating greater number of simulations to low probability designs than to high probability design in order to get constant level of accuracy.

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, it is shown that the classical ways to estimate a CDF may lead to dangerous underestimates of the probability of failure.

In this paper, several methods of estimating the probability of failure based on finite samples are tested. The first method constrains distribution fitting in order to bias the probability of failure estimate. Then, it is also shown how to use the bootstrap method to obtain distributions of probability of failure estimators, and how to use this bootstrap distribution to define conservative estimators.

In the case of samples generated from standard normal distribution, the numerical test case shows that both methods improve the chance of the estimation to be conservative. Bootstrap based estimators appear to provide much better results than optimization based methods. However, optimization based methods can be used when the sample size is very small, where the bootstrap method cannot be used.

We have also applied these procedures to estimate the probability of failure of composite laminates at cryogenic temperatures. We found that estimating the probability of failure from the mean and standard deviation of a sample lead to a five percent chance of underestimating the probability of failure by a factor of four. Using conservative estimations allows us to provide safe estimations with confidence levels, but at a price of accuracy.

For both the analytical example and the composite laminates, it is found that the conservative estimates based on the bootstrap approach outperform one-sided 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 is substantially smaller.

Controlling the uncertainty of the conservative estimation is crucial to limit the risks of over-design. To address this issue, we explored the influence of sample sizes and target probability of failure on estimates quality. We showed that larger sample sizes are required to avoid large variability in probability of failure estimates when that probability is small. However, the variability in the reliability index is independent of its value. A cost-effectiveness approach might be used to determine where to allocate computational resources (i.e., sample sizes) in order to get acceptable levels of accuracy. Such approach will be explored in a future work.

AKNOWLEDGEMENTS

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

REFERENCES

[1] *Integration of Probabilistic Methods into the Design Process*, Society of Automotive Engineers, Aerospace Information Rept. 5080, 1997

[2] Ben-Haim, Y., and Elishakoff, I., *Convex models of Uncertainty in Applied Mechanics*, Elsevier, Amsterdam, 1990

[3] Neal, D.M., Matthews, W. T. and Vangel, M.G., 1992, *Uncertainty in Obtaining High Reliability fro Stress-Strength Models*, Proceedings of the 9th DOD-/NASA/FAA Conference on Fibrious Composites in Structural Design, Vol. 1, Dept. of Defense, Lake Tahoe, NV, 1991, pp. 503-521

[4] Starnes, J. H. Jr., Haftka, R. T., *Preliminary design of composite wings for bucking, stress and displacement constraints*, Journal of Aircraft, 16 (1979) 564-570.

[5] Efron, B., 1982, *The Jackknife, the Bootstrap, and other Resampling Plans*, SIAM, Philadelphia

[6] Chernick, M.R., 1999, *Bootstrap Methods, a Practitioner's guide*, (Wiley series in probability and statistics) A Wiley-Interscience publication

[7] Melchers, R.E., *Structural Reliability Analysis and Prediction*, New York: Wiley, 1999

[8] Weisstein, E.W., *Sample Variance Distribution*, From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/SampleVarianceDistribution.html

[9] Kenney, J. F. and Keeping, E. S., *Mathematics of Statistics, 2nd ed.*, Princeton, NJ: Van Nostrand, 1951.

[10] Holton, Glyn A., *Value-at-Risk: Theory and Practice*, Academic Press, 2003

[11] Qu, X., Haftka, R.T., Venkataraman, S., and Johnson, T., *Deterministic and Reliability -Based Optimization of Composite Laminates for Propeller Tanks*, AIAA Journal, 41 (10), pp. 2029-2036.

[12] Acar, E., Haftka, R.T., *The Effects of uncertainty control measures on the weight saving from composite laminates at cryogenic temperatures*, Proceedings of ASME-IDECT-2005, September 24-28, 2005 Long Beach, California, USA