Comparison of Chebyshev's Inequality and Non-parametric B-Basis to Estimate Failure Strength of Composite Open Hole Tension Tests

Charles Jekel[1*], Yiming Zhang[2], Bogdan Grechuk[3], Rafi Haftka[1]

[1]*Mechanical and Aerospace Engineering, University of Florida, 32603, Gainesville, FL USA*
** Corresponding author: cj@jekel.me*
[2]*General Electric Global Research, Niskayuna, NY USA*
[3]*Department of Mathematics, University of Leicester, Leicester UK*

Abstract B-basis failure strength represents the lower 10th percentile with 95% confidence level. In many risk averse applications the true statistical distribution is unknown, and the B-basis is calculated using a non-parametric formulation. Chebyshev's inequality makes no assumption about the statistical distribution and can be used to bound the 10th percentile. It is possible to improve these bounds by restricting Chebyshev's inequality to a class of statistical distributions. B-basis failure strengths are compared using these methods on a collection of composite open hole tension tests.

Keywords: *Uncertainty Quantification, B-basis failure strength, Chebyshev Inequality, Log-concave CDF*

1. Introduction

The failure strength of open hole tension (OHT) tests is an important factor in the design of composite structures for aircraft [1]. B-basis failure strengths are usually considered to conservatively design redundant structures [2]. The B-basis estimates the 10th percentile with 95% confidence and is usually calculated assuming a statistical distribution [3]. If the failure strength from OHT samples resembles a known statistical distribution, then estimating the B-basis allowable is a straightforward procedure. Some common distributions for calculating the B-basis include the Normal, Lognormal, and Weibull. Unfortunately, identifying the appropriate distribution can be a difficult task, especially when only small number of samples are available [4].

Non-parametric statistics can be used to estimate the B-basis calculation when the distribution form is unknown. The MIL-HDBK-17-1F standard [5] has been a popular non-parametric B-basis calculation based on ordered statistics [6]. This method estimates the percentile via a ratio of the lowest observation to some critical observation. Alternatively, it is also possible to bound the tail probability using the general Chebyshev inequality. In terms of the B-basis, the inequality would represent the worst possible 10th percentile from all possible distributions with a given mean and given variance. The underlying difference between Chebyshev's inequality is that it attempts to represent the worst possible 10th percentile, while the other B-basis methods attempt to estimate the percentile to 95% confidence.

The bound from Chebyshev's inequality is anticipated to be overly conservative, but restricting the inequality to a particular class of distributions may produce bounds that are not overly conservative. Grechuk et al. [7] developed a general methodology for deriving Chebyshev's inequality for a class of distributions. It is possible that Chebyshev's inequality restricted to a class of distributions may result in bounds that could rival traditional B-basis methods when the true distribution is unknown. The log-concave CDF class used in [8] appears to be a reasonable class, because it includes many distributions typically used in B-basis estimations and has a simple analytical expression.

The paper goes on to describe various methods of estimating the 10th percentile from a small sample. The methods are then applied to OHT test conducted by [1] on four different configurations. The tests were conducted in batches of three, where each batch consisted of about six OHT tests. The bounded estimates from Chebyshev's inequality are compared to the other B-basis estimates.

2. Methods

Composite open hole tension (OHT) tests are used to investigate the effect of the hole size on the tensile strength of composite laminates. The OHT tests were performed on a laminate with certain plate width (w) to hole diameter (D)

ratios [1,2]. The laminate consisted of 20 plies stacked in a sequence of $[0/90/0/90/45/-45/90/0/90/0]_s$. The 0 and 90 degree plies each accounted for 40% of the layup, while the 45 degree ply made up the remaining 20%. The methods describe how Chebyshev's inequality can be used to estimate 10th percentile. Additionally the Normal, Lognormal, and non-parametric B-basis methods are presented.

2.1 Chebyshev Inequality

The standard one-sided Chebyshev Inequality for estimating the probability that a random variable X lies beyond a threshold a is given as

$$P[X \leq \mu - a] \leq \frac{\sigma^2}{a^2 + \sigma^2} \tag{1}$$

where $a > 0, \mu$ is the true mean, and σ is the true standard deviation. The B-basis for the OHT tests can be estimated by finding the threshold which results in a probability of 10 percent. This is solved by setting the right hand side of the equation equal to 0.1, and solving for a. Chebyshev's inequality applies to any distribution with both finite mean and finite variance.

Grechuk et al. [7] developed a general methodology for deriving Chebyshev's inequality when the random variable X is restricted to belong to a class of statistical distributions, thereby reducing the conservativeness. Symmetric, log-concave, and unimodal are examples of statistical distributions that Chebyshev's inequality can be restricted to include distributions.

Faridafshin et al. [8] used the methodology from [7] to derive the tightest possible Chebyshev Inequality for lower tail probabilities when X has log-concave CDF. This is expressed as

$$P[X \leq \mu - a] \leq \beta \tag{2}$$

where $a > 0$, $\beta \in (0,1)$, and is the solution to the equation

$$\frac{\sqrt{1 + 2\beta \log \beta - \beta^2}}{\beta - \log \beta - 1} = \frac{\sigma}{a}. \tag{3}$$

To estimate the threshold for a 10^{th} percentile, set $\beta = 0.1$ and solve for a. The log-concave CDF class includes all distributions where the log of the CDF is a concave function. This includes all Normal, Exponential, Gumbel, Laplace, Logistic, Rayleigh, Maxwell, Uniform, Lognormal, and Pareto distributions. Additionally, the log-concave CDF class includes subsets (limited to certain parameters) of the Weibull, Gamma, Beta, Power Function, Chi-square, and Chi distributions. Figure 1 shows the log of the CDF for the standard normal distribution, and a 4 degrees of freedom t-distribution which does not have a log-concave CDF.

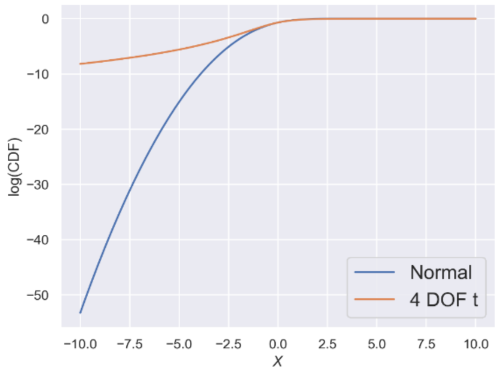


Figure 1. The standard normal distribution has a log-concave CDF while Student's t-distribution with 4 degrees of freedom (DOF) does not have a log-concave CDF.

It can be useful to express the threshold in terms of a knockdown factor f distance from the standard deviation as

$$\mu - a = \mu - f\sigma. \tag{4}$$

Using the general Chebyshev inequality to estimate the 10^{th} percentile results in a knockdown factor of $f = 3.0$, while selecting the log-concave CDF class of distributions results in a knockdown factor of $f = 1.928$. The knockdown factor for Chebyshev's inequality is constant, and does not depend upon the number of samples. However, it is based on a given mean and standard deviation, and their accuracy improves with increased number of samples.

2.2 Normal and Lognormal B-basis

The B-basis bound can be estimated from a Normal distribution using $\mu - f\sigma$, where f represents the knockdown factor for a normal distribution. This is calculated according to [9] as

$$f = \frac{1}{\sqrt{n}} t^*_{n-1;0.95}(1.2816\sqrt{n}) \tag{5}$$

where n is the sample size, $t^*_{d;1-\alpha}(\delta)$ is the $(1-\alpha)$-th quantile of a non-central t-distribution with d degrees of freedom and non-centrality parameter δ. This knockdown factor is also commonly expressed as k, and is tabulated in Table XII of [10]. Bhachu et al. [3] described that the Lognormal B-basis using the same knockdown factor from the Normal distribution as

$$x_b = \exp(\mu_l - f\sigma_l) \tag{6}$$

where μ_l and σ_l represent the mean and standard deviation of the log transformed sample.

2.3 Non-parametric B-basis

Bhachu et al. [3] described the non-parametric method for calculating the B-basis value when a collection of samples cannot be modeled by a Normal, Lognormal, or Weibull distribution. The non-parametric B-basis is calculated using ordered statistics, which first requires the sample to be in ascending order as $x = \{x_1, x_2, \cdots, x_n\}$ for a sample size of n. The following equation is used to calculate the B-basis value

$$x_b = x_r \left[\frac{x_1}{x_r}\right]^k \tag{7}$$

where $\boldsymbol{x_1}$ is the lowest observed value in a sample, and $\boldsymbol{r}$ is the rank index. The parameters $\boldsymbol{r}$ and $\boldsymbol{k}$ depend upon the number of samples, and can be looked up in MIL-HDBK-17-1F Table 8.5.14 from [5] when $\boldsymbol{n \leq 28}$.

3. Results

Lognormal B-basis, Non-parametric B-basis values, general Chebyshev inequality, and log-concave CDF Chebyshev inequality were used to estimate the 10th percentile for OHT tests from [1]. Tab.1 shows the results from individual batches, which ranged from $n = 6$ to $n = 7$ samples. The coefficient of variation (CV) is the ratio of the mean and standard deviation as $CV = \sigma/\mu$. The non-parametric B-basis was more conservative than the general Chebyshev inequality in 11 of 12 cases. The one case which was less conservative, happens to also be less conservative than the Normal B-basis. The sample set where the Normal B-basis was more conservative than the non-parametric method was [70.113, 64.881, 66.797, 64.996, 64.522, 68.6, 64.59] (ksi), and the non-parametric B-basis was calculated using $r = 5$ and $k = 2.858$. The knockdown factor from the Lognormal distribution was always smaller than the Normal distribution. The Normal B-basis with six samples is nearly identical to the general Chebyshev inequality. Note, however, that the Chebyshev inequalities were calculated based on sample mean and standard deviation, which may not be accurate representations of the underlying statistics.

The failure strength from the pooled batches of OHT tests are presented in Tab.2. The number of samples here ranged from $n = 18$ to $n = 20$ samples. Here we see that the most conservative B-basis estimate comes from the general Chebyshev inequality, which was about 3 (ksi) smaller than the other B-basis estimates. The B-basis estimates from Normal, Lognormal, non-parametric, and log-concave CDF class were very similar. The knockdown factor of the Non-parametric B-basis ranged from $f = 1.91$ to $f = 2.19$, while the knockdown factor from the Lognormal distribution ranged from $f = 1.88$ to $f = 1.92$. With the larger number of samples, the Lognormal B-basis estimate was the least conservative B-basis estimator.

The knockdown factor changes for the Normal, Lognormal, and Non-parametric B-basis estimates depending on the sample size. A plot of the Normal knockdown factors, and the Chebyshev's bounds is shown in Fig.2. The Chebyshev bound crosses the Normal B-basis at 6 samples, while the Log-concave CDF bound crosses at 20 samples. The largest knockdown factor from the non-parametric B-basis method was $f = 5.99$ with 6 samples, which is on the order of having only three samples from a Normal distribution.

Table 1. B-basis and Chebyshev's inequality thresholds for batches of OHT tests.

OHT Test Description	n	μ (ksi)	CV %	Normal B-basis (ksi, f)	Lognormal B-basis (ksi, f)	Non-parametric B-basis (ksi, f)	Log-concave CDF Class Bound (ksi, f)	Chebyshev's Bound (ksi, f)
$\frac{w}{D} = 3$ Laminate batch 1	6	57.619	3.461	51.624, 3.006	51.897, 2.870	46.332, 5.660	53.775, 1.928	51.636, 3.000
$\frac{w}{D} = 3$ Laminate batch 2	6	63.31	3.408	56.824, 3.006	57.136, 2.862	52.440, 5.038	59.151, 1.928	56.837, 3.000
$\frac{w}{D} = 3$ Laminate batch 3	6	57.984	5.218	48.888, 3.006	49.557, 2.785	45.963, 3.973	52.152, 1.928	48.907, 3.000
$\frac{w}{D} = 4$ Laminate batch 1	6	63.77	4.033	56.038, 3.006	56.389, 2.870	48.368, 5.989	58.813, 1.928	56.054, 3.000
$\frac{w}{D} = 4$ Laminate batch 2	6	65.164	2.762	59.753, 3.006	59.972, 2.885	57.108, 4.476	61.695, 1.928	59.765, 3.000
$\frac{w}{D} = 4$ Laminate batch 3	6	60.731	3.411	54.503, 3.006	54.823, 2.852	51.157, 4.621	56.738, 1.928	54.516, 3.000
$\frac{w}{D} = 6$ Laminate batch 1	7	66.357	3.362	60.210, 2.755	60.521, 2.616	60.499, 2.626	62.057, 1.928	59.664, 3.000
$\frac{w}{D} = 6$ Laminate batch 2	7	68.656	4.828	59.523, 2.755	59.783, 2.677	49.256, 5.853	62.267, 1.928	58.712, 3.000
$\frac{w}{D} = 6$ Laminate batch 3	6	64.253	2.527	59.372, 3.006	59.522, 2.914	56.115, 5.012	61.123, 1.928	59.382, 3.000
$\frac{w}{D} = 8$ Laminate batch 1	6	67.254	3.202	60.780, 3.006	61.050, 2.881	56.111, 5.174	63.103, 1.928	60.794, 3.000
$\frac{w}{D} = 8$ Laminate batch 2	6	70.71	4.913	60.266, 3.006	61.009, 2.792	56.259, 4.160	64.014, 1.928	60.288, 3.000
$\frac{w}{D} = 8$ Laminate batch 3	6	69.835	3.769	61.922, 3.006	62.374, 2.835	56.691, 4.994	64.762, 1.928	61.939, 3.000

Table 2. B-basis and Chebyshev's inequality thresholds for pooled batches of OHT tests.

OHT Test Description	n	μ (ksi)	CV %	Normal B-basis (ksi, f)	Lognormal B-basis (ksi, f)	Non-parametric B-basis (ksi, f)	Log-concave CDF Class Bound (ksi, f)	Chebyshev's Bound (ksi, f)
$\frac{w}{D} = 3$ All three batches	18	59.638	5.903	52.689, 1.974	53.020, 1.880	52.745, 1.958	52.852, 1.928	49.077, 3.000
$\frac{w}{D} = 4$ All three batches	18	63.222	4.414	57.714, 1.974	57.867, 1.919	57.114, 2.189	57.843, 1.928	54.845, 3.000
$\frac{w}{D} = 6$ All three batches	20	66.530	4.521	60.874, 1.926	60.936, 1.905	60.916, 1.912	60.870, 1.928	57.720, 3.000
$\frac{w}{D} = 8$ All three batches	18	69.266	4.386	63.270, 1.974	63.523, 1.891	63.138, 2.017	63.410, 1.928	60.152, 3.000

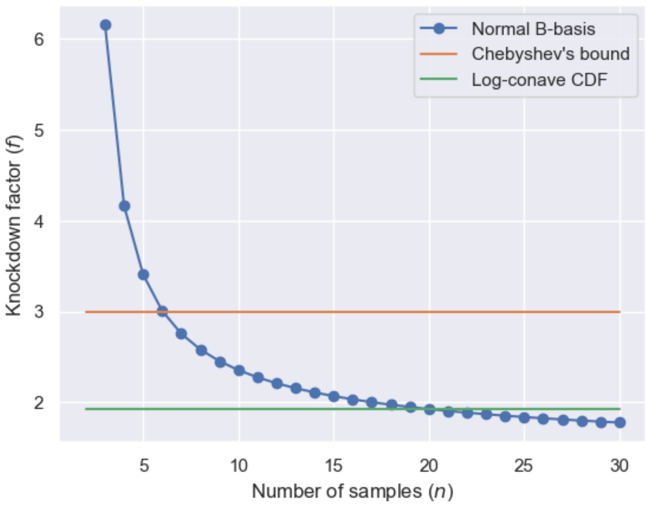
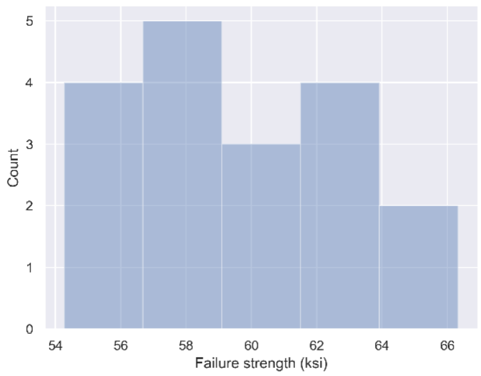
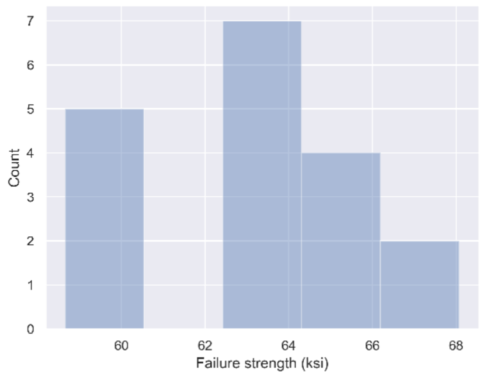


Figure 2. Knockdown factors for the Normal B-basis as a function of the number of samples.

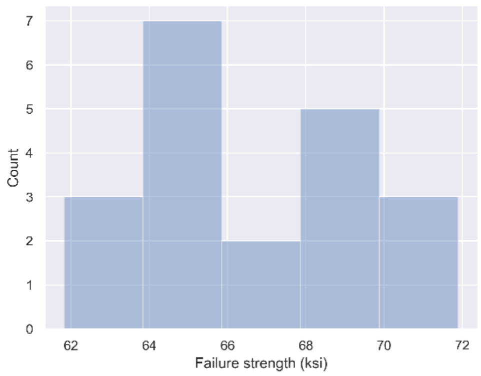
Histograms of the pooled OHT batches are presented in Fig.3. The pooled batches consist of 18 to 20 samples, and it's not particularly easy to identify the possible true distribution. This is perhaps most evident with the $w/d = 4$ test. It is even more difficult to estimate the distribution from a single batch of just 6 or 7 samples.



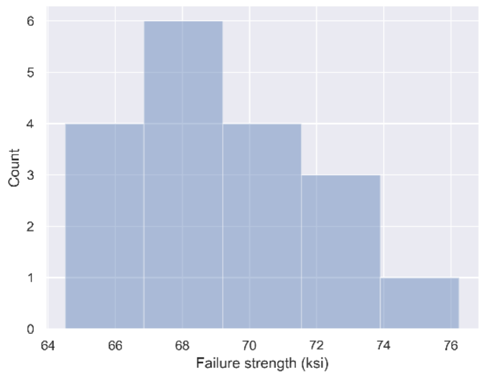
a) All three batches OHT tests $\frac{w}{D} = 3$



b) All three batches OHT tests $\frac{w}{D} = 4$



c) All three batches OHT tests $\frac{w}{D} = 6$



d) All three batches OHT tests $\frac{w}{D} = 8$

Figure 3. Histogram of the combined batches of OHT configurations.

4. Discussion

There is the potential to improve B-basis estimates with 6 or 7 samples, as there was a large penalty the non-parametric B-basis method over the Normal and Lognormal distributions. In the worst case the knockdown factor of the non-parametric distribution was nearly twice as large as the Normal distribution. Additionally, there was one case where the non-parametric B-basis was not as conservative as the Normal distribution.

Differences between the B-basis methods was hardly noticeable with samples totaling 18 to 20. The Normal, Lognormal, non-parametric, and Log-concave CDF bounds were basically identical, with the Lognormal resulting in the least conservative estimate. While the non-parametric B-basis could have a huge knockdown factor of $f = 5.99$ with 6 samples, the knockdown factor reduced to approximately $f = 2.0$ with 18 samples, which is on par to the other mentioned methods.

One potential issue is that Chebyshev's inequality requires the true mean and true standard deviation, while bounded estimates were performed on the sample mean and standard deviation. It is possible to account for the potential uncertainty in the sample mean and standard deviation with Bootstrapping, which wouldn't make additional assumptions about the potential true distribution. This would result in a distribution of bounded estimates from the possible means and standard deviations. The estimated bound could then be selected for an appropriate confidence level. It is worthwhile to point out that in this case where the uncertainty in the sample means and standard deviations were neglected, the general Chebyshev and Log-concave CDF bounds were comparative to the other B-basis calculations.

5. Conclusion

Traditional B-basis methods attempt to estimate the 10th percentile to a 95% confidence level for a collection of samples. If the true distribution is unknown, then non-parametric statistics are used to estimate the B-basis. The non-parametric estimate will be very conservative for a small number of samples, as a cost of not knowing the true distribution. For instance, there was an OHT batch of 6 samples where the non-parametric B-basis was 6 standard deviations away from the mean. There may be better ways to approach non-knowing the true distribution with six or 7 samples than the non-parametric B-basis. An additional concern was that the non-parametric B-basis was not always more conservative than the B-basis from a Normal distribution. The general Chebyshev inequality can be applied to estimate a bounded 10th percentile for any distribution with a known mean and variance. With 6 or 7 samples in the OHT tests, the general Chebyshev's inequality was comparable to the Normal distribution. Chebyshev's inequality restricted to the log-concave CDF class was comparable to a Normal B-basis with about 20 samples. Although with 20 samples, B-basis estimates from the Normal, Lognormal, and non-parametric methods were nearly the same. It's possible to restrict Chebyshev's inequality to only apply to a particular class of statistical distributions. The log-concave CDF class applies to many distributions that are commonly used to estimate tail probabilities. Perhaps some restricted class (like the Log-concave CDF) could be used to improve B-basis estimates for unknown distributions with 6 or so samples.

References

1. Tomblin, J., & Seneviratne, W. (2009). Laminate Statistical Allowable Generation for Fiber-Reinforced Composite Materials: Lamina Variability Method. Office of Aviation Research and Development, Federal Aviation Administration.
2. Zhang, Y., Schutte, J., Meeker, J., Palliyaguru, U., Kim, N. H., & Haftka, R. T. (2017). Predicting B-basis allowable at untested points from experiments and simulations of plates with holes. In 12th World Congress on Structural and Multidisciplinary Optimization. Braunschweig, Germany.
3. Bhachu, K. S., Haftka, R. T., & Kim, N. H. (2016). Comparison of Methods for Calculating B-Basis Crack Growth Life Using Limited Tests. AIAA Journal, 54(4), 1287–1298. https://doi.org/10.2514/1.J054094
4. Romero, V. J., Schroeder, B. B., Dempsey, J. F., Breivik, N. L., Orient, G. E., Antoun, B. R., … Winokur, J. G. (2018). Simple Effective Conservative Treatment of Uncertainty From Sparse Samples of Random Variables and Functions. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering, 4(4), 41006–41017. Retrieved from http://dx.doi.org/10.1115/1.4039558
5. U.S. Department of Defense, MIL-HDBK-17-1F Polymer Matrix Composites, Volume 1: Guidelines for Characterization of Structural Materials. (2002).
6. Neal, D. M., & Vangel, M. G. (1990). Statistically Based Material Properties. A Military Handbook-17 Perspective. ARMY LAB COMMAND WATERTOWN MA MATERIAL TECHNOLOGY LAB.
7. Grechuk, B., Molyboha, A., & Zabarankin, M. (2010). Chebyshev inequalities with law-invariant deviation measures. Probability in the Engineering and Informational Sciences, 24, 145–170. https://doi.org/10.1017/S0269964809990192
8. Faridafshin, F., Grechuk, B., & Naess, A. (2017). Calculating exceedance probabilities using a distributionally robust method. Structural Safety, 67, 132–141. https://doi.org/10.1016/j.strusafe.2017.02.003
9. Young, D. S. (2010). tolerance: An R Package for Estimating Tolerance Intervals. Journal of Statistical Software; Vol 1, Issue 5 (2010). Retrieved from https://www.jstatsoft.org/v036/i05
10. Montgomery, D. C., & Runger, G. C. (2014). Applied statistics and probability for engineers (6th ed.). John Wiley & Sons, Inc.