COVERAGE PROBABILITY OF TOLERANCE LEVELS

A Thesis by

Matthew Lann-Wolcott

Bachelor of Science, Wichita State University, 2016

Submitted to the Department of Mathematics and Statistics
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

December 2018

COVERAGE PROBABILITY OF TOLERANCE LEVELS

The following faculty members have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Master of Science with a major in Mathematics.

Xiaomi Hu, Committee Chair

Dharam Chopra, Committee Member

Adam Jarger, Committee Member

Hyuck Kwon, Committee Member

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Xaiomi Hu. Without his patience and motivation throughout this process this thesis would not have come to be. I would like to thank Dr. Dharam Chopra, Dr. Adam Jaeger, and Dr. Hyuck Kwon for taking time to be part of my committee. Finally I must thank my family for their support throughout.

ABSTRACT

In this thesis the ideas of the tolerance limit and coverage probabilities are explored in the context of the A-basis and B-basis for composite materials. The breaking strength of composite materials are random variables that require the percentiles of the random variable to estimate. The thesis will explore the use of point estimators and then one-sided confidence intervals as appropriate estimators. It will then look at the A-basis and B-Basis in an ANOVA setting. Finally, with a numerical example, it will explore the coverage probability of the A-basis and B-basis as a design method parameter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Composite material is now widely used in the aircraft industry and is being adapted to use in other areas as well. One of the qualities that engineers need to know when working with composites is its breaking strength. This design values takes on the form of a random variable, thus necessitating the need for a statisical approach to estimate the breaking strength's value. This thesis will explore tolerance levels and coverage probability in the context of the A-basis and B-basis as described in the Military Handbook 17F (2002). Ideally the A-basis provides and estimate for the breaking strength such that 99 percent of the material will have a higher breaking strength, while the B-basis provides and estimate such that 90 percent of the material will have a higher breaking strength. Chapter 2 will examine the use of point estimators for the A-basis and B-basis and their failure to provide the desired coverage probability for the A-basis and B-basis. Chapter 3 to will examine the use of the lower limit of a one sided confidence interval of the point estimators or what is often called a tolerance limit to achieve the desired coverage probabliity. Chapter 4 will concern itself with the pratical calculation of the A-basis and B-basis and then examine the A-basis and B-basis in the ANOVA setting. Finally Chapter 5 will then examine the coverage probability of the A-basis and B-basis, demonstrate the ability to use the coverage probability as a design perameter and explore the impact of different pooling methods for σ^2 in the ANOVA setting with a numerical example.

Chapter 2

Point Estimation

Since the breaking strength of composite materials is a random variable we need some way to provide an estimate that engineers can use. The mean is a common parameter that is often estimated, but in this situation is not the most usefull. Let $X \sim N(\mu, \sigma^2)$ be the breaking strength of a composite material, then the mean in this case would not be a very effective estimator as half the material would have a breaking strength below the mean and would break before expected. Another point estimator will be necessary, so we will consider the $100p$-th percentile x_p. Say we consider the first percentile, then 99 percent of the material would have a great breaking strength than that value, or say the 5th, then 95 percent of the material would have a greater breaking strength. Values like these provide an estimate the engineers can then design around.

So again let $X \sim N(\mu, \sigma^2)$, then we define the $100p$-th percentile x_p as a non-random number such that:

$$P(X \leq x_p) = p$$

Now since $X \sim N(\mu, \sigma^2)$, let us consider standardizing X

$$p = P(X \leq x_p) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x_p - \mu}{\sigma}\right)$$

Then

$$p = P\left(Z \leq \frac{x_p - \mu}{\sigma}\right)$$

Finally

$$\frac{x_p - \mu}{\sigma} = z_p \text{ and } x_p = \mu + z_p \sigma$$

So by standardizing X we see that x_p is a linear combination of both μ with a coeffiecent of one, and σ with a coeffiecent of z_P. This is very convienent since the value of z_p is easily found in z-tables or from statistical software.

Now since x_p is a parameter of the population an estimate based on a random sample is needed. So again let $X \sim N(\mu, \sigma^2)$ and let $X_1, X_2, ..., X_n$ be a random sample from the population X. From before, x_p is a linear combination of μ and σ, so it is not unreasonable to consider replacing μ with the sample mean $\bar{x}$ and sigma by the sample standard deviation s.

Thus

$$\hat{x}_p = \bar{x} + z_p \cdot s$$

Now by Hogg et al. (2013) $\begin{pmatrix} \bar{x} \\ s^2 \end{pmatrix}$ are sufficient and complete for $\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$ so $\hat{x}_p$ is a minimum-variance unbiases estimator for $E(\hat{x}_p)$. Now looking at $E(\hat{x}_p)$ we see

$$E(\hat{x}_p) = E(\bar{x} + z_p \cdot s) = \mu + z_p \cdot E(s)$$

and looking at $E(s)$ we see

$$\begin{aligned} E(s) &= \frac{\sigma}{\sqrt{n-1}} E\sqrt{\frac{(n-1)s^2}{\sigma^2}} = \frac{\sigma}{\sqrt{n-1}} E\sqrt{\chi^2(n-1)} \\ &= \frac{\sigma}{\sqrt{n-1}} \int_0^\infty \sqrt{x} \frac{1}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}}} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}} dx \\ &= \frac{\sigma}{\sqrt{n-1}} \frac{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}}} \int_0^\infty \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx \\ &= \frac{\sigma}{\sqrt{n-1}} \frac{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}}} = \frac{\Gamma\left(\frac{n}{2}\right)\sigma}{\sqrt{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} \end{aligned}$$

Now we see that $E(s) \neq \sigma$ and $E(\hat{x}_P) \neq x_p$. So our estimator $\hat{x}_P$ is a biased estimator. Calculating the mean squared error of this biased estimator can give us an indication as to its performance and whether it would be an appropriate estimator to use. The mean square error is defined as

$$\begin{aligned} MSE(\hat{x}_p) &= E(\hat{x}_p - x_p)^2 = E(\bar{x} + z_p \cdot s - \mu - z_p \cdot \sigma)^2 = E(\bar{x} - \mu)^2 + z_p^2 E(s - \sigma)^2 \\ &= \frac{\sigma^2}{n} + z_p^2 [E(s^2 + \sigma^2 - 2\sigma \cdot s)] = \frac{\sigma^2}{n} + 2z_p^2[\sigma^2 - \sigma E(s)] \\ &= \frac{\sigma^2}{n} + 2z_p^2 \sigma^2 \left[1 - \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\right] \end{aligned}$$

To provide a visualization of what happens to the MSE as n increase, Figure 1 is a plot of the MSE with $\sigma^2 = 1$ and $z_{0.05} = -1.645$

So it is obvious as n increases that the MSE is getting smaller and smaller thus $\hat{x}_p$ is becoming a better estimator as n increases. In fact the initial decrease is very rapid, so even a moderate increase in n provides a sizable improvment to the estimator.

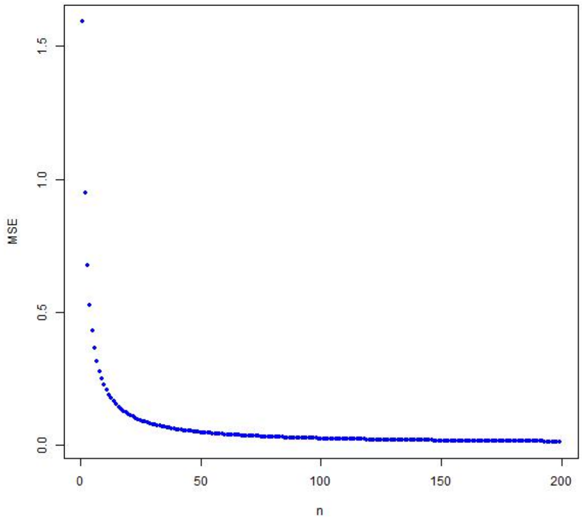

Figure 1: Mean Squared Error vs. n

The goal of $100p$-th percentile x_p point estimator is to provide a coverage probability of $1-p$. It is hoped that if the $100p$-th percentile x_p has the point estimator $\hat{x}_p$ and we let $p = 0.10$ then $\hat{x}_{0.10}$ we will have a coverage probablity of 90 percent. In other worlds, 90 percent of the material will have a breaking strength greater than or equal to $\hat{x}_{0.10}$. Since the estmator is based off a sample and we no longer have x_p. Instead x_p is being replaced by the estimator $\hat{x}_p$ so the coverage probability changes to

$$P(X \geq \hat{x}_p) = P(X \geq \bar{x} + z_p \cdot s)$$

In order to calculate the coverage probability consider

$$P(X \geq \bar{x} + z_p \cdot s) = P\left(\frac{X-\bar{x}}{s} \geq z_p\right)$$

Now we know that $X \sim N(\mu, \sigma^2)$ and $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ are idependent so it follows

$$X - \bar{x} \sim N\left(0, \sigma^2 + \frac{\sigma^2}{n}\right) = N\left(0, \frac{n+1}{n}\sigma^2\right).$$

$$\frac{X-\bar{x}}{\sqrt{\frac{n+1}{n}\sigma^2}} \sim N(0, 1^2)$$

Now since $X - \bar{x}/\sqrt{\frac{n+1}{n}\sigma^2}$ has a standard normal distribution and that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$ is independent to $X - \bar{x}/\sqrt{\frac{n+1}{n}\sigma^2}$, then using the definition of the t-distibution

$$\frac{X - \bar{x}}{\sqrt{\frac{n+1}{n}\sigma^2\frac{(n-1)s^2}{(n-1)\sigma^2}}} = \frac{X - \bar{x}}{\sqrt{\frac{n+1}{n}}s} \sim t(n-1)$$

Finally the coverage probability can be expressed as

$$P(X \geq \bar{x} + z_p \cdot s) = P\left(\frac{X - \bar{x}}{\sqrt{\frac{n+1}{n}}s} \geq \frac{z_p}{\sqrt{\frac{n+1}{n}}}\right) = P\left(t(n-1) \geq \sqrt{\frac{n}{n+1}}z_p\right) \tag{1}$$

In table 1 below some numerical examples are provided of the coverage probability for the point estimators of some percentiles. They where easily calculated using the pt function in R and equation(1). It is obvious from Table 1 that the coverage probability of the point estimator is less than the desired $1 - p$. Take $p = 0.01$ for example, even when $n = 100$, the coverage probability still falls short of the desired 0.99. This means we need an alternative to the point estimator to achieve the desired coverage probablity.

Table 1: Coverage Probabilites of Some Point Estimators of Percentiles

n	p=0.01	p=0.05	p=0.10
5	.9495371	.8961889	.8465022
10	.9731342	.9243728	.8736107
20	.9824916	.9375301	.8868803
50	.9872308	.9450995	.8947687
100	.9886555	.9475648	.8973871

Chapter 3

Tolerance Level

One way to increase the coverage probablity of is to consider the lower limit of a one sided confidence interval for x_p with the confidence coefficent $1-\alpha$. The idea is to replace the point estimator $\hat{x}_p$ with $L(p, 1-\alpha)$. $L(p, 1-\alpha)$ is a statistic satisfying the condition

$$P(L(p, 1-\alpha) \leq x_p) = 1-\alpha$$

This statistic was introduced by Fraser and Guttman (1959) and it is commoly called the tolerance limit for X with content $1-p$ and confidence coefficient $1-\alpha$. It was hoped that the point estimator of the percentile would have a coverage probability that contained $1-p$ of X, but the numerical example in table 1 demonistrated this was not the case. This new tolerance limit will, with a high degree of probability, provide us with the coverage probability we need. To increase the coverage probability it is natural to assume that $L(p, 1-\alpha) < \hat{x}_p$ and since $\hat{x}_p = \bar{x} + z_p \cdot s$ we can define $L(p, 1-\alpha) = \bar{x} - K \cdot s$, where $-K < z_p < 0$. Now we need a closed form for K. Consider

$$1-\alpha = P(L(p, 1-\alpha) \leq x_p) = P(\bar{x} - K \cdot s < \mu + z_p \cdot \sigma) = P\left(\frac{\bar{X} - \mu - z_p \cdot \sigma}{s} < K\right)$$

$$= P\left(\frac{\bar{X} - \mu - z_p \cdot \sigma}{\sqrt{\frac{s^2}{n}}} < \sqrt{n}K\right)$$

Now we know $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ so we see that $\bar{x} - \mu - z_p \cdot \sigma^2 \sim N(-z_p\sigma^2, \frac{\sigma^2}{n})$ and finally we have $\frac{\bar{x}-\mu-z_p\cdot\sigma}{\sqrt{\sigma^2/n}} \sim N(-z_p\sqrt{n}, 1^2)$. From Hogg et. al (2013) we find that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$ is independent to $\frac{\bar{x}-\mu-z_p\cdot\sigma}{\sqrt{\sigma^2/n}} \sim N(-z_p\sqrt{n}, 1^2)$ therefore,

$$\frac{\bar{x} - \mu - z_p \cdot \sigma}{\sqrt{\sigma^2/n}} \cdot \frac{1}{\sqrt{\frac{(n-1)s^2}{(n-1)\sigma^2}}} = \frac{\bar{x} - \mu - z_p \cdot \sigma}{\sqrt{s^2/n}} \sim t(-z_p\sqrt{n}, n-1)$$

where $t(-z_p\sqrt{n}, n-1)$ is a t-distribution with the non-centrality parameter $-z_p\sqrt{n}$ and degrees of freedom $n-1$. So now subsituting back in we have
$1-\alpha = P(t(-z_p\sqrt{n}, n-1) < K\sqrt{n})$ and then $K\sqrt{n} = t_{1-\alpha}(-z_p\sqrt{n}, n-1)$. Now finally we

obatin K from Hu(2007)

$$K = \frac{t_{1-\alpha}(-z_p\sqrt{n}, n-1)}{\sqrt{n}}$$

Now we can define the lower limit of one-sided confidence interval for the 100pth percentile of $X \sim N(\mu, \sigma^2)$ with confidence coefficient $1 - \alpha$ as

$$L(p, 1-\alpha) = \bar{x} - K \cdot s, \text{ where } K = \frac{t_{1-\alpha}(-z_p\sqrt{(n)}, n-1)}{\sqrt{n}} \tag{2}$$

This statistic, as stated before, is called the tolerance limit for X with content $1 - p$ and confidence coefficient $1 - \alpha$. So now there is a statistic that will provide us the coverage probability that we need and since we found a closed form for K there is a way to calculate the statistic. We can now use this to calculate the A-basis and the B-basis as described in the Military Handbook 17F (2002).

Chapter 4

A-basis and B-basis

The Military Handbook 17F (2002) defines the A-basis as the lower limit of the one-sided confidence interval of the first percentile of the breaking strength of a composite material with confidence coefficent of 95 percent and the B-basis finds the lower limit of the one-sided confidence interval of the 10th percent of the breaking strength of a composite material with confidence coefficient of 95 percent. So the A-basis is $L(0.01, 0.95)$ and the B-basis is $L(0.10, 0.95)$. Then if we let $X \sim N(\mu, \sigma^2)$ be the breaking strength of the composite material then the A-basis and B-basis are as follows

$$A = \bar{x} - K_A s \tag{3}$$

and

$$B = \bar{x} - K_B s \tag{4}$$

Now we can use the closed form of K we found previously to calculate K_A and K_B with

$$K_A = \frac{t_{0.95}(2.3263\sqrt{n}, n-1)}{\sqrt{n}}$$

and

$$K_B = \frac{t_{0.90}(1.2816\sqrt{n}, n-1)}{\sqrt{n}}$$

Tomblin, Ng and Raju (2003) have provided the following formulas for practical computation. It should be noted that the pt function in R should not be used with the noncentral perameter option. This option uses a different algorithm and from Lenth (1989) it is known to be inaccurate in the tails. Now let $d = n$, and $f = n - 1$,

$$K_A(d, f) = \frac{t_{0.95}(2.3263\sqrt{d}, f)}{\sqrt{d}}$$

$$K_B(d, f) = \frac{t_{0.90}(1.2816\sqrt{d}, f)}{\sqrt{d}}$$

K_A and K_B can be approximated by

$$k_A(d,f) = \frac{2.3263}{\sqrt{q}} + \sqrt{\frac{1}{dc_A} + \left(\frac{b_A}{2c_A}\right)^2} - \frac{b_A}{2c_A}$$

$$k_B(d,f) = \frac{1.2816}{\sqrt{q}} + \sqrt{\frac{1}{dc_B} + \left(\frac{b_B}{2c_A}\right)^2} - \frac{b_B}{2c_B}$$

Let $x = 1/\sqrt{f}$ with the following equations for q, b_A, b_B, c_A, and c_B

$$q = 1 - 2.327x + 1.138x^2 + 0.6057x^3 - 0.3287x^4$$

$$b_A = 2.0643x - 0.95145x^2 + 0.51251x^3$$

$$b_B = 1.1372x - 0.49162x^2 + 0.18612x^3$$

$$c_A = 0.36961 + 0.0026958x - 0.65201x^2 + 0.011320x^3$$

$$c_B = 0.36961 + 0.0040342x - 0.71750x^2 + 0.16963x^3$$

The following table provides a comparison of some point estimators for the 1st and 10th percentiles and the associated A-basis and B-basis values for different values of the sample mean, sample standard deviation and sample size.

Table 2: Comparison of Point Estimators to the A-basis and B-basis

sample mean	20	50	35	80	120
sample st. dev.	2	3	2.5	4	3.8
sample size	20	15	25	40	17
$\hat{x}_{0.01}$	15.3473	43.0209	29.1841	70.6946	111.1599
A-basis	13.4091	39.4376	27.1049	68.2371	107.0235
$\hat{x}_{0.10}$	17.43690	46.1553	31.7961	74.8737	115.1301
B-Basis	16.1473	43.79267	30.4044	73.2122	112.3914

So when comparing point estimators of the 1st and 10th percentiles to their associated basis value we can see the the basis value is smaller, or more conservative, in each case. This is obviously what we expected as we defined the A-basis and B-basis as the lower limit of a one sided confidence interval with a confidence coefficent of 95 percent. This smaller value should now provide us with a better coverage probability than the point estimators where able to provide.

Computation of the A-basis and B-basis in ANOVA setting

One of the things that is desirable when collecting data on the breaking strength of the composites is to test them under different conditions. These different conditions, or factors, such as temperature, humidity, and other environmental conditions can have an impact on the breaking strength of the composite material. Now let us suppose that there are, say k, conditions or factors. Each one of these populations will have a sample size of n_i and a sample mean $\bar{x}_i$ and a sample variance s_i^2. Then if we assume the breaking strength for each environment is normally distributed with population mean μ_i and identical populations variances σ^2 a one-way ANOVA model can be formed. Since we assume an equal variance among the populations an common variance must be found to use with the different samples. Sahai and Ageel (2000) provide the pooled estimator,which is an unbiased estimator for the population variance, given by

$$s_p^2 = \frac{\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2 + ... + \sum_{i=1}^{n_k}(x_{ki} - \bar{x}_k)^2}{(n_1 - 1) + ... + (n_k - 1)}$$

or

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + ...(n_k - 1)s_k^2}{N - k}$$

where $N = n_1 + n_2 + ... + n_k$

Now if we consider $L(p, 1 - \alpha)$ when μ is estimated by $\hat{\mu} \sim N(\mu, \frac{\sigma^2}{d})$ and σ^2 is estimated by $\hat{\sigma}^2$ when $\frac{(f)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(f)$ is independent to $\hat{\mu}$ then by following the same proceedure that led to eqution (2) we see

$$L(p, 1 - \alpha) = \hat{\mu} - K \cdot \hat{\sigma} \text{ where } K = \frac{t_{1-\alpha}(-z_p\sqrt{d}, f)}{\sqrt{d}}$$

Now if we wish to calculate the A-basis and B-basis for the ith population, then we let $\hat{\mu} = \bar{x}_i$, $\hat{\sigma} = s_p$, $d = n_i$,and $f = N - k$. Then the A-basis and B-basis for the ith population become

$$A_i = \bar{x}_i - K_A(d, f) \cdot s_p$$

and

$$B_i = \bar{x}_i - K_B(d, f) \cdot s_p$$

Similar to equations (3) and (4). The estimations created by Tomblin, Ng and Raju (2003) can be used again to estimate $K_A(d, f)$ and $K_B(d, f)$.

Chapter 5

Coverage Probability in an ANOVA Setting

To consider the coverage probability in an ANOVA setting let $L(p, 1-\alpha)$ be the lower limit of a one-sided confidence interval for the 100pth percentile of $N(\mu, \sigma^2)$ with the confidence coefficient $1-\alpha$, and following the assumption of the existence of independent $\hat{\mu} \sim N(\mu, \frac{\sigma^2}{d})$ and $\frac{f \cdot \hat{\sigma}^2}{\sigma^2} \sim \chi^2(f)$. We then call $P(X \geq L(p, 1-\alpha))$ the coverage probability of $L(p, 1-\alpha)$. Now to calculate the coverage probability, following a similar process to what was used for point estimators in chaper 1, we know $L(p, 1-\alpha) = \hat{\mu} - K \cdot \hat{\sigma}$. Then $P(X \geq L(p, 1-\alpha)) = P(\frac{X-\hat{\mu}}{\hat{\sigma}} \geq -K)$. So $X - \hat{\mu} \sim N(0, (1+\frac{1}{d})\sigma^2)$ since both $X \sim N(\mu, \sigma^2)$ and $\hat{\mu} \sim N(\mu, \frac{\sigma^2}{d})$ are normally distributed. Therefore $\frac{X-\hat{\mu}}{\sqrt{(1+\frac{1}{d})\cdot\sigma^2}}$ has a standard normal distribution and is independent of $\frac{f \cdot \hat{\sigma}^2}{\sigma^2} \sim \chi^2(f)$, so finally we see that $\frac{X-\hat{\mu}}{\sqrt{(1+\frac{1}{d})\cdot\hat{\sigma}^2}} \sim t(f)$. Now the coverage probability is

$$P\left(\frac{X-\hat{\mu}}{\hat{\sigma}} \geq -K\right) = P\left(t(f) \geq -\frac{K}{\sqrt{(1+\frac{1}{d})}}\right) \text{ where } K = \frac{t_{1-\alpha}(-z_p\sqrt{d}, f)}{\sqrt{d}}$$

The coverage probabilities for the A-basis and B-basis are then give by

$$P(X \geq A) = P\left(t(f) \geq -\frac{K_A(d,f)}{\sqrt{1+\frac{1}{d}}}\right)$$

$$P(X \geq B) = P\left(t(f) \geq -\frac{K_B(d,f)}{\sqrt{1+\frac{1}{d}}}\right)$$

One of the first things to note here is the fact that the coverage probability in the ANOVA setting doesn't actually depend on the sample, only the design of the ANOVA model. Only d and f impact the value of the coverage probability, and since $d = n$ and $f = N - k$, it is the sample size and the number of factors that impact the coverage probability. In other words, it is the choice of pooling method that is creating different, but valid basis values. Since this is the case it is possible for the coverage probability to be used as a method parameter to differ the valid but different pooling methods. So as proposed by Hu (2007) the coverage probability can be calculated before for use with pooling method selection or reported after with the basis value as a method parameter.

Numerical Study

In order to provide a numerical example of the A-basis, B-basis and coverage probabilities in an ANOVA setting and the effects of pooling strategy on the values a R-script,from the appendix, was written. A user inputs the values of the sample mean, pooled sample variance, d which is sample size for a given factor, and f which is $N-k$. The script ,when run, will produce a table with the 1st percentile, 10th percentile, the A-basis, the B-basis, and their associated coverage probabilities. The R language and environment was chosen mainly due to its built in pt function, which from Rizzo (2008) we find calculates $P(t(f) \geq x)$. This made for a very simple and concise script, but again it should be noted that the pt function should not be used with the noncentral parameter option. This option uses a different algorithm that is known to be inaccurate in the tails. This is why the estimations for K_A and K_B are still calculated by the equations given by Tomblin, Ng and Raju (2003).

Table 3 provides the an example of the differences in the A-basis and B-basis under various pooling methods given a sample mean of 103.1, and a pooled sample variance of 6.175^2. While table 4 gives the coverage probability of various pooling methods given different values for d and f.

Table 3: A-basis and B-basis for $\bar{x} = 103.1$ and $s_p = 6.175$

d	Stat	f=d-1	f=2d-2	f=3d-3	f=4d-4	f=5d-5
5	A-basis	67.51023	76.77821	79.36736	80.60060	81.32675
	B-basis	81.987328	86.93104	88.27514	88.90347	89.26837
10	A-basis	78.50345	81.99180	83.13510	83.71320	84.06446
	B-basis	88.54940	90.33121	90.89116	91.16686	91.33126
15	A-basis	81.35906	83.63362	84.41699	84.82062	85.06832
	B-basis	90.32325	91.45541	91.82677	92.01241	92.12395
20	A-basis	82.75081	84.49272	85.10793	85.42801	85.62545
	B-basis	91.20491	92.05679	92.34214	92.48584	92.57250
25	A-basis	83.59929	85.03758	85.55357	85.82365	85.99079
	B-basis	91.74907	92.44332	92.67884	92.79797	92.86997
30	A-basis	84.18086	85.42085	85.87054	86.10692	86.25353
	B-basis	92.12529	92.71772	92.92043	93.02327	93.08552

Table 4: Coverage Probability

d	Stat	f=d-1	f=2d-2	f=3d-3	f=4d-4	f=5d-5
5	$\hat{x}_{0.01}$	0.9495371	0.9667770	0.9724126	0.9751790	0.9768171
	A-basis	0.9968756	0.9976994	0.9978430	0.9978615	0.9978464
	$\hat{x}_{0.10}$	0.8465022	0.8621423	0.8676219	0.8704115	0.8721011
	B-basis	0.9822566	0.9780827	0.9755667	0.9739662	0.9728684
10	$\hat{x}_{0.01}$	0.9731342	0.9801743	0.9824249	0.9835272	0.9841804
	A-basis	0.9978844	0.9978221	0.9976567	0.9975193	0.9974116
	$\hat{x}_{0.10}$	0.8736107	0.8812488	0.8838489	0.8851590	0.8859483
	B-basis	0.9743610	0.9678867	0.9648912	0.9631795	0.9620734
15	$\hat{x}_{0.01}$	0.9795687	0.9838410	0.9852114	0.9858849	0.9862851
	A-basis	0.9978818	0.9975335	0.9972663	0.9970790	0.9969428
	$\hat{x}_{0.10}$	0.8824749	0.8875225	0.8892274	0.8900841	0.8905994
	B-basis	0.9675582	0.9607246	0.9578082	0.9561948	0.9551709
20	$\hat{x}_{0.01}$	0.9824916	0.9855272	0.9865056	0.9869879	0.9872749
	A-basis	0.9977260	0.9972257	0.9969076	0.9966970	0.9965483
	$\hat{x}_{0.10}$	0.8868803	0.8906487	0.8919170	0.8925535	0.8929361
	B-basis	0.9622278	0.9554901	0.9527258	0.9512202	0.9502732
25	$\hat{x}_{0.01}$	0.9841473	0.9864919	0.9872509	0.9876257	0.9878491
	A-basis	0.9975367	0.9969420	0.9965950	0.9963724	0.9962179
	$\hat{x}_{0.10}$	0.8895155	0.8925218	0.8935316	0.8940379	0.8943421
	B-basis	0.9579807	0.9514609	0.9488475	0.9474373	0.9465549
30	$\hat{x}_{0.01}$	0.9852093	0.9871156	0.9877348	0.9880411	0.9882238
	A-basis	0.9973449	0.9966872	0.9963230	0.9960941	0.9959371
	$\hat{x}_{0.10}$	0.8912691	0.8937696	0.8946084	0.8950287	0.8952812
	B-basis	0.9545107	0.9482348	0.9457575	0.9444291	0.9436007

From Table 3 we can see that moving down the table with the increase in d, the sample size for each factor, that both the basis are increasing in value, thus producing a less conservative value. The same happens moving arcoss the columns with the increase in factors. The basis value again is increasing producing a less conservative value. This is as we would expect because moving down and across the table increases the total sample size, improving our estimator $\hat{s}_p^2$, which in turn decreases the interval in the tolerance limit making for a less conservative estimate. Now we need to check to while the A-basis and B-basis values are increaseing and becoming less conservative that we do not drop below are desired coverage probability. Looking to Table 4 we see that at no point does the coverage probability of the A-basis and B-basis go below the desired level. While there is a decrease in the coverage probability, it is to be expected as the basis value becomes less conservative. It also worth noting that while the increase in sample size and factor size do increase the coverage probability of the 1st and 10th percentile point estimators, they never actually achieve the desired coverage probability.

Chapter 6

Conclusion

The use of composite materials in aviation construction is only increasing and the need for a reliable estimator for the breaking strength is important. Under the assumption of $X \sim N(\mu, \sigma^2)$, the use of a percentile estimator was explored. While a biased estimator was found, it was shown that as n increased the mean squared error became small. Unfortunately as the coverage probability was considered it became apparent that a point estimator would not produce the desired level of coverage probability and another estimator would be needed. The use of the lower limit of a one-sided confidence interval for the percentile point estimator with confiedence coefficient $1 - \alpha$ to increase the coverage probability, commonly called the tolerance level, is suggested. It was then shown that when defining $(L(p, 1 - \alpha) = \bar{x} - K \cdot s$, that there was closed form for K and thus the tolerance level is calculable. With the use of estimations from the Military Handbook 17F (2002) we where able to find K_A and K_B and then the A-basis and B-basis. Table 4 also provided a numerical example demonstrating that the use of a tolerance level was able to provide use with the deisred coverage probability where as the percentile estimators fell short of the desired coverage probability

Finally the A-basis and B-basis where explored in an ANOVA setting. This is important as composite materials are often tested under a variety of different environmental conditions and ability to combine these different factors is both cost effective and beneficial to the estimation of the A-basis and B-basis. It was shown that the coverage probability in an ANOVA setting does not depend on the actual sample, but on d and f, the size of sample for each factor and the number of factors. Thus we believe that the coverage probability of can be of use as either a method parameter or in the consideration of pooling method selection.

REFERENCES

REFERENCES

[1] Military Handbook - MIL-HDBK-17-1F: Composite Materials Handbook, Volume 1 - Polymer Matrix Composites Guidelines for Characterization of Structural Materials (2002), U.S. Department of Defense.

[2] Hogg, Mckean, Craig, Introduction to Mathematical Statistics, 7th ed. (2013), Pearson

[3] Fraser, D. A. S. and Guttman, I., Tolerance regions. Annals of Mathematical Statistcs (1959), 27 162-179,

[4] John S. Tomblin, Yeow C. NG and K. Sresh Raju, Material qualification and equivalency for polymer matrix composite materials systems: Updated procedure, DOT/FAA/AR-03/19 (2003), Federal Aviation Administration William J. Hughes Technical Center.

[5] Hardeo Sahai and Mohammed I. Ageel, The Analysis of Variance, Fixed, Random, and Mixed Models (2000), Birkhauser

[6] Xiaomi Hu, Asymptotical distributions, parameters and coverage probabilities of tolerance limits, Computational Statistics and Data Analysis (2007), 51(9) 4753 - 4760.

[7] Maria L. Rizzo, Statistical Computing with R (2008), Chapman and Hall/CRC

[8] Lenth, R.V. Algorithm AS 243 - Cumulative distribution function of the non-central t distribution, Applied Statistics (1989), 38 185-189.

APPENDIX

Appendix

```
xbar=0 #input sample mean
s=1 #input sample standard deviation
d=30 #input value of d
f=145 #input value off
#estimation of the K values for the A and B basis
x=1/sqrt(f)
q= 1-2.327*x + 1.138*x2̂ + 0.6057*x3̂ - 0.3287*x4̂
b_A= 2.0643*x - 0.95145*x2̂ + 0.51251*x3̂
c_A= 0.36961 + 0.0026958*x - 0.65201*x2̂ + 0.011320*x3̂
b_B= 1.1372*x - 0.49162*x2̂ + 0.18612*x3̂
c_B= 0.36961 + 0.0040342*x - 0.71750*x2̂ + 0.16963*x3̂
K_A = 2.3263/sqrt(q) + sqrt((1/(d*c_A))+(b_A/(2*c_A))2̂)-(b_A/(2*c_A))
K_B = 1.2816/sqrt(q) + sqrt((1/(d*c_B))+(b_B/(2*c_B))2̂)-(b_B/(2*c_B))
A_basis = xbar-(K_A)*s #calculating the A-basis
B_basis = xbar-(K_B)*s #calculating the B-basis
P01= xbar-qnorm(.99)*s #calculating the .01 point estimator
P10= xbar-qnorm(.90)*s #calculating the .10 point estimator
#coverage probability
coverage_P01 = pt(qnorm(.01)/(sqrt(1+(1/d))),f,lower.tail=FALSE)
coverage_P10 = pt(qnorm(.1)/(sqrt(1+(1/d))),f,lower.tail=FALSE)
coverage_A_basis = pt(-K_A/(sqrt(1+(1/d))),f,lower.tail=FALSE)
coverage_B_basis = pt(-K_B/(sqrt(1+(1/d))),f,lower.tail=FALSE)
estimators=c(P01,P10,A_basis,B_basis)
coverage_probability=c(coverage_P01, coverage_P10, coverage_A_basis, coverage_B_basis)
table=cbind(estimators, coverage_probability)
rownames(table)=c(”P01”,”P10”,”A-basis”,”B-basis”)
colnames(table)=c(”Estimators”,”Coverage Probability”)
table
```