EASA
European Aviation Safety Agency

Statistics of A and B basis

Elena Garcia

elena-beatriz.garcia-sanchez
@easa.europa.eu

Rotorcraft & VTOL Structures Workshop
19-20 February 2019

Your safety is our mission.

EASA is an agency of the European Union

Contents

- Intent of Basis values
- Statistics: Bayes theorem. Normal distribution example
- Other distributions
- Metals and composites
- Conclusions

A and B basis intent

- It provides a statistical robustness for design values
- CS X.613 Material strength properties and design values
  - SLP*: A basis, 99% probability with 95% confidence
  - MLP*: B basis, 90% probability with 95% confidence
- The properties are obtained in two steps:
  - **Probability**: 90 or 99% of the population will have better properties than the basis value
  - **Confidence**: Based on our limited sample, we ensure that 95% of future samples will have the expected probability above, or better.

    This means that data of higher reliability results in higher A or B basis values:
    - Sample conforming to a parametric distribution
    - Large sample size

*SLP: Single load path
*MLP: Multiple load path

20/02/2019 A & B basis statistics 3

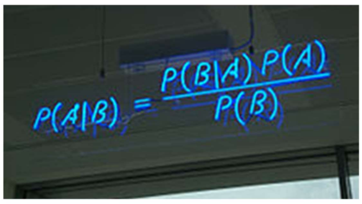
Materials can’t be controlled to be perfectly homogeneous. Even though the process is controlled, there will always be imperfections like impurities, distortions; interphase defects for composite; from the atomic to the macroscopic level. Material properties are affected by these variations and show a scatter of behaviour.

Design values have to be obtained from the limited information of a sample. This brings us into the domain of Uncertainty.

B basis: of 10 samples of the population, 9 will typically have better properties. This behaviour will be true or better for 95% of our future samples.

We will see what confidence means for the example in next pages: confidence on B basis assuming normality depends on confidence in the distribution mean and standard deviation.

Inference of population from samples: Bayes



- Posterior=Sampling*Prior/Normalising factor (see Ref. [4])
  - Posterior P(A/B): probability of the distribution parameters given the prior information and sampled data
  - Sampling distribution: reciprocal probability to obtain the sample, if the parameters are considered known
  - Prior P(A):Any information available before the test
  - Normalising factor P(B): Ensures total probability is 1

The Bayes theorem is the basic rule for inference., to determine the posterior probability of future data, by combining any prior information with the sampled data.

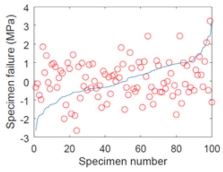
An example of prior information could be a probability distribution for the population mean, coming perhaps from other tests.

In the case of no particular prior information, for a certain model of the data, Bayes says that the probability of a combination of parameters is proportional to the sampling distribution. The sampling distribution is the product of probabilities of all sampling points. This seems simple but is quite powerful.

For example, if according to our information we can model the distribution as a Normal distribution, the parameters we are interested in are the mean and the standard deviation.

Example: Normal distribution B basis

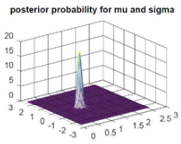
- Generate a sample of size n, from a Normal distribution of mean μ=0 and standard deviation σ=1
- Assume no prior information except normality.



$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Sampling distribution: Multiply probabilities of all the data points (x_1, $x_2 \ldots x_n$) for each combination of (μ_p,σ_p), and normalize
- Because of the Normal distribution exponential, the multiplications become additions and the posterior depends only on the sample mean x and standard deviation s (Ref.[4])

$$P(\mu_p, \sigma_p) = \frac{2^{\frac{1-n}{2}}}{\sqrt{\pi}} \cdot \frac{n^{\frac{1+n}{2}}}{\Gamma(\frac{n}{2})} \cdot \left(\frac{s}{\sigma_p}\right)^n \cdot \frac{1}{\sigma_p{}^2} \cdot e^{-\frac{n\cdot(s^2+(\overline{x}-\mu_p)^2)}{2\sigma_p{}^2}}$$



The normal distribution is very special and useful. Posterior closed form available.

Under mild conditions, the mean of a large number of random variables independently drawn from the same distribution is distributed approximately normally, irrespective of the form of the original distribution. Maximum entropy for given mean and variance. Maximum likelihood estimate=mean.

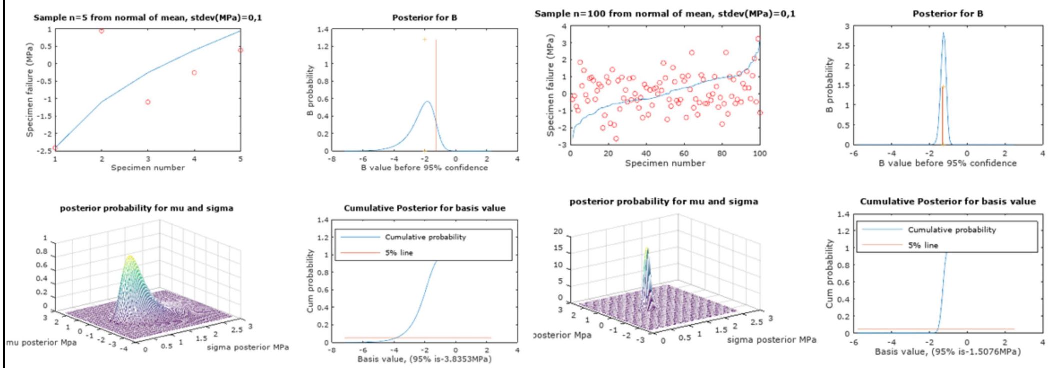
CMH17 prefers the Normal:

1.Experience has shown that large samples of composite material strength data tend to be modeled

well by the normal distribution.

B value for n=5 and n=100

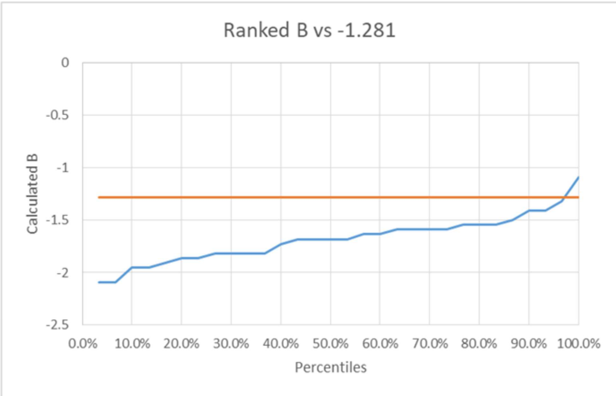
- To get the posterior for B values, substitute μ for (B+Z* σ), integrate sigma out and normalize.
- The B value is the B for which cumulative probability is 5%
- $Z_{\text{unlimited data}}$= inv Normal (.90, μ =0, σ =1) Z_B=1.281
- For n=5, B=-3.83. For n=100, B=-1.51



A Basis : Z=2.326

Check that the posterior works

- We calculate B for many samples and we see that in effect 95% of the times it lies below -1.281 (-Z for population B basis)

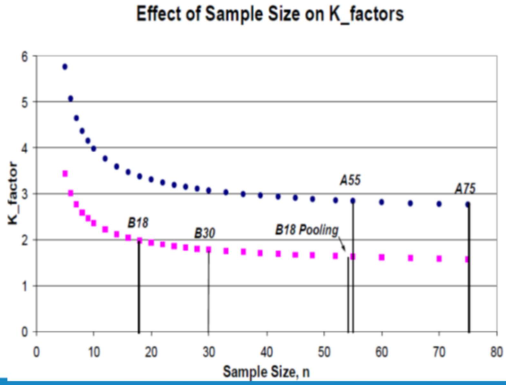


20/02/2019 A & B basis statistics 7

Bayes didn't lie.

Normal distributions. Takeaway.

- Used for S-basis (See Ref[1] table 9.10.9 and adjusted formula for n>30) and Ref [5]
- Ref. [1] (metal) s. 9.1.5 specifies variability of heats and lots. For A-basis direct computation (Ftu, Fty), this results in 100 data points needed for parametric distributions, and 299 for non-parametric.
- Outliers shouldn't be excluded unless explained.

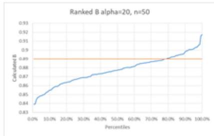


$$A - Basis\ value = \overline{x} - (K_A) \cdot s$$

$$B - Basis\ value = \overline{x} - (K_B) \cdot s$$

Other distributions

- The Normal distribution has no skew and the tail shape (kurtosis) is fixed. It may not be conservative.
- E.g assume Normal (e.g. S-basis) for Weibull of shape factor α=20 n=50 : confidence <95%



- Comparison of hypotheses using Bayes, or tests of the distribution shape (see Ref [1]) may suggest the most appropriate distribution(Normal, Weibull, Lognormal, Pearson…)
- Use the non-parametric distribution if there isn't enough confidence about the model (backup slide)

20/02/2019 A & B basis statistics 9

CMH17: 2. Without a significant quantity of data (roughly more than 60 data points), it is not possible to differentiate

between data sets following normal, Weibull, or lognormal distributions with a reasonable level of statistical power.

Weibull and lognormal

These distributions are useful in representing samples which are "skewed", that is one of the tails are not prominent. For materials controlled by material specifications which has a tendency of eliminating low test points

Weibull: weakest link model. Behaviour of brittle materials including composites

Lognormal: Distribution of a random variable whose logarithm is normally distributed. Can be thought of as a product of many independent random variables

Metals and composites

- For metallic **Single Load Paths**, A basis values provide robustness to material strength.
- **Composites** are designed as resistant to fatigue damage, but are susceptible to delamination. The strength demonstration relies on **combined assessment** of damage detectability (and likelihood), no growth, and residual strength, for the different damage categories.
- There are usually more composite properties, environmental conditions and directions under test, the total number of specimens in a **composite** testing program often **exceeds the total number of coupons** in an A/B-basis metals testing program
- Composite properties can't be taken directly from manuals.
- **Composites: Combination (Pooling) of data** extensively across conditions. Large variability between batches (checks of combinability). Pooling should only be for the same failure modes.

Damages on metal have the potential to develop into a growing crack, which won't stop until the part completely breaks, or will stop temporarily at the next arrest feature. This makes MLP the main means to ensure a robust design.

Focussing of use of A basis for composite at coupon level would be missing the point of more critical damage tolerance at higher assembly levels (except for slow growth)

Refer to shared database. Complex evaluation

Typical values of material scatter and shape factors

- Order of magnitude (just orientative) of static strength

Material	CV (σ / μ)	Weibull shape factor
Metal	>1%	~60
Unidirectional carbon	~10%	~20
Glass fibre	~>15%	~8

- A higher shape factor means less scatter and less skew
- Fatigue shape factors are of the order of ~1-4

Conclusions

- A, B basis apply to CS-27, CS-29 and SC-VTOL, for static strength from coupons.
- Complemented by the rest of the pyramid.
- Trade-off between sample size n and the design value.
- Interpretation of SLP for composite is subjected to discussion.

References

- [1] DOT/FAA/AR-MMPDS-01
- [2] CMH-17 Vol 1. Ch 8. Yeow Ng presentation.
- [3] Probability Theory, the Logic of Science. By E.T. Jaynes
- [4] Wikipedia: Conjugate prior
- [5] STAT17

Links:

Jaynes: Highly recommended to understand probability in a rational way as opposed to a collection of ad-hoc recipes

https://bayes.wustl.edu

Conjugate prior: Some postprocess of the Normal conjugates with unknown mean and variance is needed to arrive to the porterior for them

https://en.wikipedia.org/wiki/Conjugate_prior

EASA
European Aviation Safety Agency

Thank you for your attention!

Any questions….?

Your safety is our mission.

EASA is an agency of the European Union

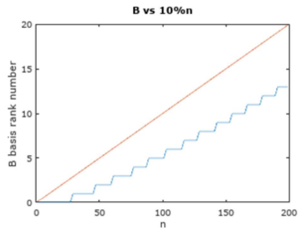
Non-parametric inference

- Non-Parametric: We draw a sample of size n. If the cumulative probability of the population B basis is 10%, what is the probability for the ranked sample to get just k draws below B? This is a binomial sampling distribution:

$$f(k,n,p) = \Pr(k;n,p) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Solve k for cumulative function of f=5%, with p=10% for B basis. The corresponding sample x_k is the B basis

- Generally produces lower allowables



- Rank value table
  in Ref [1] table 9.10.9