

Estimation of One-Sided Lower Tolerance Limits for a Weibull Distribution Using the Monte Carlo Pivotal Simulation Technique

Sandeep Makam, Yung-Li Lee and Pradeep Attibele
 Chrysler Group LLC

ABSTRACT

This paper introduces a methodology to calculate confidence bounds for a normal and Weibull distribution using Monte Carlo pivotal statistics. As an example, a ready-to-use lookup table to calculate one-sided lower confidence bounds is established and demonstrated for normal and Weibull distributions.

The concept of one-sided lower tolerance limits for a normal distribution was first introduced by G. J. Lieberman in 1958 (later modified by Link in 1985 and Wei in 2012), and has been widely used in the automotive industry because of the easy-to-use lookup tables. Monte Carlo simulation methods presented here are more accurate as they eliminate assumptions and approximations inherent in existing approaches by using random experiments. This developed methodology can be used to generate confidence bounds for any parametric distribution.

The ready-to-use table for the one-sided lower tolerance limits for a Weibull distribution is presented. The Weibull distribution is widely adopted in the automotive industry because of its wide range of engineering applications. The Weibull model covers many other parametric distributions (either exactly or approximately) like lognormal and exponential. Typically, the fatigue life of a product follows a Weibull distribution. Estimation of the confidence bounds for a Weibull distribution with limited experimental data is critical in making design and material selection decisions for any product to ensure it complies with durability targets.

CITATION: Makam, S., Lee, Y. and Attibele, P., "Estimation of One-Sided Lower Tolerance Limits for a Weibull Distribution Using the Monte Carlo Pivotal Simulation Technique," *SAE Int. J. Mater. Manf.* 6(3):2013, doi: 10.4271/2013-01-0329.

INTRODUCTION

In many applications such as quality control or material strength estimation, there is a need to estimate the lower tolerance limit of a distribution from a limited sample size. G. J. Lieberman introduced the idea of one-sided tolerance limits in 1958 and provided a closed form solution and a set of tables to calculate them for normal distributions using sample mean and standard deviation for a given reliability and confidence level [1]. This formula tends to underestimate the tolerance limits for small sample sizes because it is based on the asymptotic normality property of the tolerance limit [2]. This deficiency was addressed by Link in 1985 by providing a closed form solution that used the non-central Student's t-distribution as an approximation to solve analytically for the limits [2]. The same approach was used by Wei in 2012 for design curve constructions [3]. However, since both of these analytical approaches approximate the standard normal distribution to a non-central t-distribution and approximate

the normal quantiles for the t-distribution, the predicted tolerance limits are wider than in reality. In trying to address this issue, this study adopts a Monte Carlo pivotal statistics method which does not utilize approximations which in effect unnecessarily enlarge the confidence interval. Much like the earlier methods, a ready-to-use table is provided to calculate tolerance limits for the normal distribution based on sample data. As explained by Symynck and De Bal in 2011[4], the same method is applied for Weibull distributions to provide a ready-to-use table for applications where such a distribution is expected.

The two-parameter normal distribution, also known as the Gaussian function or the "Bell curve", is used to define a distribution of random variables that are clustered around a single valued mean (denoted usually by μ). The spread of this cluster is defined by the standard deviation, denoted by σ . The probability density function of a normal distribution is

given as:
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2\right)$$
 Therefore, the

cumulative density function of a normal distribution is

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

The two-parameter Weibull distribution is widely used for modeling the life of a product or component because it can account for either increasing or decreasing failure rates (or hazard functions) in a simple manner. It is also used to model the distribution for component properties such as strength (electrical or mechanical), elongation, resistance etc., in accelerated tests [5]. The Weibull model covers many other parametric distributions (either exactly or approximately) like Rayleigh and exponential. It is mathematically classified as belonging to the family of extreme value distributions. Weibull analysis also provides the engineer with a very simple, useful and convenient graphical plot of the failure data [6]. Typically the horizontal scale is a measure of life or aging in terms of cycles, number of events, mileage, operating time etc. The two parameters that define a Weibull line are its slope and the intercept vertical axis at 63.2%. They are generally referred to as the shape parameter or beta (β) and scale parameter or the characteristic life or theta (θ), respectively.

The two-parameter Weibull distribution is defined with the following probability density function:

$$f(x) = \frac{\beta}{\theta} \left(\frac{x}{\theta}\right)^{\beta-1} \exp \left[-\left(\frac{x}{\theta}\right)^{\beta} \right] \quad (1)$$

And the cumulative density function is therefore

$$F(x) = 1 - \exp \left[-\left(\frac{x}{\theta}\right)^{\beta} \right] \quad (2)$$

For the purpose of any sort of statistical analysis, one would always prefer as large a sample size as can be afforded given the time, cost and safety constraints. The ability to provide reasonably accurate failure analysis and forecasts with extremely small sample sizes is one of the key advantages of a Weibull analysis. The size of sample is always tied with the amount of uncertainty one would have with the results of the analysis. A small sample size generally indicates large uncertainties. To evaluate these uncertainties for a Weibull distribution, many methods have been established to calculate the confidence bounds such as beta-binomial, likelihood ratio, Fisher matrix methods, Monte Carlo bounds and pivotal bounds [6]. In 1974, Mann introduced a method for calculating bounds using maximum likelihood and best linear invariant estimators via Monte Carlo simulations [7]. Symynck and De Bal [4] introduced a method to calculate bounds using Monte Carlo pivotal statistics in 2011. They compared it with regular (or Straight) Monte Carlo methods and concluded that using pivotal statistics in conjunction with Monte Carlo methods gave the best results. However, their approach involves extensive Monte Carlo simulations for each sample set. In this study,

this general approach has been used to simplify calculation of confidence bounds by generating a set of tables for ready usage.

PIVOTAL STATISTICS

A pivotal statistic is a quantity that is parameter-free i.e. it is independent of the properties/characteristics (such as mean, standard deviation or slope, shape parameters) of the distribution while representing the data points of the sample. It is essentially a function of the sample data but does not depend on the unknown parameters of the distribution. [8]

Given n independent, identically distributed observations $X = \{X_1, X_2, X_3, \dots, X_n\}$ from a known distribution with k parameters $\theta = \{\theta_1, \theta_2, \dots, \theta_k\}$, a function g is a pivotal statistic if the distribution of $g(X, \theta)$ is independent of θ .

Pivotal Statistics for Normal Distributions

For a normal distribution, with mean μ , standard deviation σ , and an observation x , a z -score can be defined as:

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

It is noted that the z -score follows a normal distribution with a mean of 0 and a standard deviation of 1, irrespective of the values of μ and σ calculated from the sample. This is said to be following a "Standard Normal Distribution". Thus, the z -score is effectively a pivotal statistic for any data that fits a normal distribution. It is noted that the pivot was essentially derived by calculating the ratio of the difference between the observation and the location parameter (the mean, in this case) and the scale parameter (the standard deviation, in this case).

Pivotal Statistics for Weibull Distribution

According to [8, 4], extending the idea of pivotal quantities to a Weibull distribution, a z -score can be defined as:

$$z_p = \frac{\hat{u} - y_p}{\hat{b}} = \frac{\log(\hat{\theta}) - \log(t_p)}{(1/\hat{\beta})} \quad (4)$$

where $\hat{u} = \log(\hat{\theta})$, $y_p = \log(t_p)$, $\hat{b} = (1/\hat{\beta})$, $\hat{\theta}$ is the sample characteristic life, $\hat{\beta}$ is the sample shape parameter and t_p is the time/age/cycles at which the unreliability is $p = F(t; \beta, \theta)$ or the Weibull CDF function. That is, $t_p = F^{-1}(p; \beta, \theta)$.

By definition, z_p does not depend on unknown parameters, meaning that its distribution form stays the same regardless of the actual values of β , and θ . The pivotal can therefore be calculated by arbitrarily setting $\beta, \theta = 1$, for further calculations.