Reliability Articles

Relex
A PTC InSight Product

Weibull Parameter Estimation Methods

By Suprasad Amari, PhD, CRE

Question

What are the different parameter estimation methods available for analyzing failure data? How do I select the right method?

The Short Answer

Estimation methods are used to find the parameters of the failure distribution from given failure data. Each method has a criterion, which yields estimates that are best in some situations. Different results are produced based on that criterion. The most commonly used methods can be broadly categorized into two types: MLE (Maximum Likelihood Estimation) and regression.

- MLE. From all possible values, MLE methods find the parameter values that maximize the likelihood of obtaining the data. MLE methods have several good statistical properties and are well suited for large samples. When there are several suspensions, MLE methods produce better results than regression methods.
- Regression. Using the concept of probability plotting, regression methods transform the scales of the probability of failure versus time plot to make it a straight line. The parameters found by this process mathematically correspond to the best straight line, which is the one with the least sum of error squares. Regression methods are best suited for small samples. If, however, there are several suspensions or interval data, they may not be suitable.

When data is limited, accuracy can be improved by making various modifications to the estimation method. Additionally, confidence intervals can determine the range of possible values for parameters and the uncertainty in these estimates.

The Details

Parameter estimation is a procedure to find the parameters of a model that are best suited to a given set of observations. In reliability analysis, it is used to find the parameters of the distribution associated with failure data. The process is known as failure data or life data analysis. The most popular and widely used distribution in failure data analysis is the Weibull distribution. Thus, this process is commonly known as Weibull analysis, even when the underlying failure distribution is not the Weibull distribution.

Failure data analysis involves identifying the form of distribution (best fit distribution), the parameters of the distribution (parameter estimation), and the uncertainties associated with the estimated parameters (confidence intervals). This article focuses on parameter estimation.

For example, if you have a failure data set but the parameters of this distribution are not known, you need to select the best estimation method to use to determine these parameters. To understand how to evaluate estimation methods, it is helpful to understand the criteria that statisticians employ for this purpose. The following criteria are often used when evaluating estimates:

- Invariance and Unbiasedness
- Consistency
- Efficiency

Consider the following random sample from the normal distribution: 15, 25, 25, 30, 34, 38, and 50. The parameters of this distribution are the mean (μ) and standard deviation (σ). In theory, the mean is equivalent to its mode and median. Thus, the mean, mode, and median should coincide, allowing any one of these functions to be used as the estimator for the parameter μ. However, for a specific sample, the values for the mean, mode, and median are likely to differ.

For the sample given, the mean, mode, and median are 31, 25, and 30, respectively. Which is the correct estimate of μ? Let μ^e be the estimator for μ. Here, μ^e is a function of the sample. Because the sample is random, μ^e is also random. If the expected value of μ^e coincides with the true value of μ, then it is an **unbiased** estimator. If the estimate of f(μ) is $f(\mu^e)$, then it is an **invariant** estimator under this transformation. For example, if λ^e is the failure rate estimate, and reliability at time t is $\exp(-\lambda^e[t])$, then λ^e is an invariant estimator. When the sample size n tends to infinity, if λ^e tends to the true parameter λ, then it is a **consistent** estimator. For the same sample size, the estimator λ^e is an **efficient** estimator if it has a least expected value of $(\lambda^e-\lambda)^2$.

MLE Methods

MLE methods are generally preferred for larger samples, especially when there are suspensions or interval data. Because the original MLE method produces biased results for small samples, various modifications have been made to it, resulting in the Modified MLE (MMLE) method. Modifications made to reduce the deviation due to uncertainty include:

- n/(n-1). This adjustment is applicable for normal and lognormal distributions. The sample variance is adjusted by multiplying it by n/(n-1), where n is the number of failures. While this reduces the bias, it does not eliminate it.
- Reduced Biased Adjustment (RBA). The standard deviation for the normal and lognormal distribution is adjusted by the square root of n/(n-1), and divided by an additional factor, C4. This method produces more accurate results for small samples. For normal and lognormal distributions, RBA is used to eliminate the mean bias, which is the difference between the true value and the mean value of the many estimates for the parameter. For the Weibull distribution, the RBA median bias adjustment is preferred.

Regression Methods

Regression methods are preferred for small samples, especially when there are no suspensions or interval data. Regression methods produce a straight line when failure probability, F(t), and time, t, are plotted using transformed scales. The straightforward way to compute F(t) is as the ratio of the number of failures (r) and the total number of components (n). However, this gives 100% at the last point, which is off scale. Rank regression methods correct this problem in different ways:

- Mean Rank Regression. Estimates the failure probability as r/(n+1).
- Hazen Rank Regression. Estimates the failure probability as (r-0.5)/n. It is generally better than mean rank regression.

- Median Rank Regression. Estimates the failure probability as the rth failure of n units that occurs at time t with a 50% confidence level. This results in the best estimate for F(t). Half of the time, the true value will be greater than the 50% confidence estimate. The other half of the time, the true value will be less than this estimate. This method, also known as Binomial rank regression, is preferred in engineering analysis.
- Benard Rank Regression. Provides a quick approximation to compute the median ranks as (r-0.3)/(n+0.4). This method is sufficiently accurate for most practical purposes and can be done by hand if necessary.

Conclusion

Almost all estimation methods produce accurate results for large samples. More care must be taken in the selection of the estimation methods of small samples due to the biased results that may be produced. For large samples, MLE is usually preferred. For small samples, MMLE or regression methods are preferred. If there are several suspensions or interval data, the MMLE method should be used. Otherwise, a regression method should be used. Typically, the median rank regression method is preferred in engineering analysis. To account for the statistical uncertainty in estimates, confidence intervals should be obtained, particularly for small samples. Relex Weibull supports all of these estimation methods. When you use the criteria presented in this article for selecting an MLE or regression method, the parameter estimates obtained are very likely to be the best suited to the given set of distributions. Relex Weibull also provides a **Find Best Fit Distribution** option to find the distribution that best fits a particular failure data set. For additional information, please visit www.relex.com/products/weibull.asp.