Life Data Analysis Reference
<table>
<thead>
<tr>
<th>Quick License Summary Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>You are Free to:</strong></td>
</tr>
<tr>
<td><strong>Share:</strong> Copy and redistribute the material in any medium or format</td>
</tr>
<tr>
<td><strong>Adapt:</strong> Remix, transform, and build upon the material</td>
</tr>
<tr>
<td><strong>Under the following terms:</strong></td>
</tr>
<tr>
<td><strong>Attribution:</strong> You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. See example at <a href="http://www.reliawiki.org/index.php/Attribution_Example">http://www.reliawiki.org/index.php/Attribution_Example</a></td>
</tr>
<tr>
<td><strong>NonCommercial:</strong> You may not use this material for commercial purposes (sell or distribute for profit). Commercial use is the distribution of the material for profit (selling a book based on the material) or adaptations of this material.</td>
</tr>
<tr>
<td><strong>ShareAlike:</strong> If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.</td>
</tr>
</tbody>
</table>

**Generation Date:** This document was generated on May 22, 2015 based on the current state of the online reference book posted on ReliaWiki.org. Information in this document is subject to change without notice and does not represent a commitment on the part of ReliaSoft Corporation. The content in the online reference book posted on ReliaWiki.org may be more up-to-date.

**Disclaimer:** Companies, names and data used herein are fictitious unless otherwise noted. This documentation and ReliaSoft’s software tools were developed at private expense; no portion was developed with U.S. government funds.

**Trademarks:** ReliaSoft, Synthesis Platform, Weibull++, ALTA, DOE++, RGA, BlockSim, RENO, Lambda Predict, Xfmea, RCM++ and XFRACAS are trademarks of ReliaSoft Corporation.

Other product names and services identified in this document are trademarks of their respective trademark holders, and are used for illustration purposes. Their use in no way conveys endorsement or other affiliation with ReliaSoft Corporation.
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License ("Public License"). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 – Definitions.

a. Adapted Material means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.

b. Adapter’s License means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.

c. BY-NC-SA Compatible License means a license listed at creativecommons.org/compatibllicenses, approved by Creative Commons as essentially the equivalent of this Public License.

d. Copyright and Similar Rights means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.

e. Effective Technological Measures means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.

f. Exceptions and Limitations means fair use, fair dealing, and/or any other exception or limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.

g. License Elements means the license attributes listed in the name of a Creative Commons Public License. The License Elements of this Public License are Attribution, NonCommercial, and ShareAlike.

h. Licensed Material means the artistic or literary work, database, or other material to which the Licensor applied this Public License.
i. **Licensed Rights** means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.

j. **Licensor** means ReliaSoft Corporation, 1450 Eastside Loop, Tucson, AZ 85710.

k. **NonCommercial** means not primarily intended for or directed towards commercial advantage or monetary compensation. For purposes of this Public License, the exchange of the Licensed Material for other material subject to Copyright and Similar Rights by digital file-sharing or similar means is NonCommercial provided there is no payment of monetary compensation in connection with the exchange.

l. **Share** means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.

m. **Sui Generis Database Rights** means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.

n. **You** means the individual or entity exercising the Licensed Rights under this Public License. **Your** has a corresponding meaning.

**Section 2 – Scope.**

a. **License grant.**
   1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:
      A. reproduce and Share the Licensed Material, in whole or in part, for NonCommercial purposes only; and
      B. produce, reproduce, and Share Adapted Material for NonCommercial purposes only.
   2. **Exceptions and Limitations.** For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.
   3. **Term.** The term of this Public License is specified in Section 6(a).
   4. **Media and formats; technical modifications allowed.** The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures. For purposes of this Public License, simply making modifications authorized by this Section 2(a)(4) never produces Adapted Material.
   5. **Downstream recipients.**
      A. **Offer from the Licensor – Licensed Material.** Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.
      B. **Additional offer from the Licensor – Adapted Material.** Every recipient of Adapted Material from You automatically receives an offer from the Licensor to exercise the Licensed Rights in the Adapted Material under the conditions of the Adapter’s License You apply.
C. **No downstream restrictions.** You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.

6. **No endorsement.** Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

b. **Other rights.**
   1. Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.
   2. Patent and trademark rights are not licensed under this Public License.
   3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties, including when the Licensed Material is used other than for NonCommercial purposes.

**Section 3 – License Conditions.**

Your exercise of the Licensed Rights is expressly made subject to the following conditions.

a. **Attribution.**
   1. If You Share the Licensed Material (including in modified form), You must:
      A. retain the following if it is supplied by the Licensor with the Licensed Material:
         i. identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);
         ii. a copyright notice;
         iii. a notice that refers to this Public License;
         iv. a notice that refers to the disclaimer of warranties;
         v. a URI or hyperlink to the Licensed Material to the extent reasonably practicable;
      B. indicate if You modified the Licensed Material and retain an indication of any previous modifications; and
      C. indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.
   2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.
   3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.
b. **ShareAlike.**

In addition to the conditions in Section 3(a), if You Share Adapted Material You produce, the following conditions also apply.

1. The Adapter’s License You apply must be a Creative Commons license with the same License Elements, this version or later, or a BY-NC-SA Compatible License.
2. You must include the text of, or the URI or hyperlink to, the Adapter’s License You apply. You may satisfy this condition in any reasonable manner based on the medium, means, and context in which You Share Adapted Material.
3. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, Adapted Material that restrict exercise of the rights granted under the Adapter’s License You apply.

**Section 4 – Sui Generis Database Rights.**

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

a. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database for NonCommercial purposes only;

b. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material, including for purposes of Section 3(b); and

c. You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

**Section 5 – Disclaimer of Warranties and Limitation of Liability.**

a. Unless otherwise separately undertaken by the Licensor, to the extent possible, the Licensor offers the Licensed Material as-is and as-available, and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors, whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.

b. To the extent possible, in no event will the Licensor be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this Public License or use of the Licensed Material, even if the Licensor has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.
c. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 – Term and Termination.

a. This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.

b. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:
   1. automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or
   2. upon express reinstatement by the Licensor.

For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.

c. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.

d. Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Section 7 – Other Terms and Conditions.

a. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed. 

b. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 – Interpretation.

a. For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully be made without permission under this Public License.

b. To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.

c. No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.

d. Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.
# Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1</td>
<td>1</td>
</tr>
<tr>
<td>Introduction to Life Data Analysis</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>10</td>
</tr>
<tr>
<td>Basic Statistical Background</td>
<td>10</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>16</td>
</tr>
<tr>
<td>Life Distributions</td>
<td>16</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>23</td>
</tr>
<tr>
<td>Parameter Estimation</td>
<td>23</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>41</td>
</tr>
<tr>
<td>Life Data Classification</td>
<td>41</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>48</td>
</tr>
<tr>
<td>Confidence Bounds</td>
<td>48</td>
</tr>
<tr>
<td>Chapter 7</td>
<td>71</td>
</tr>
<tr>
<td>The Exponential Distribution</td>
<td>71</td>
</tr>
<tr>
<td>Chapter 8</td>
<td>103</td>
</tr>
<tr>
<td>The Weibull Distribution</td>
<td>103</td>
</tr>
<tr>
<td>Chapter 9</td>
<td>155</td>
</tr>
<tr>
<td>The Normal Distribution</td>
<td>155</td>
</tr>
<tr>
<td>Chapter 10</td>
<td>187</td>
</tr>
<tr>
<td>The Lognormal Distribution</td>
<td>187</td>
</tr>
<tr>
<td>Chapter 11</td>
<td>214</td>
</tr>
<tr>
<td>The Mixed Weibull Distribution</td>
<td>214</td>
</tr>
<tr>
<td>Chapter 12</td>
<td>225</td>
</tr>
<tr>
<td>The Generalized Gamma Distribution</td>
<td>225</td>
</tr>
</tbody>
</table>
Chapter 13  
The Gamma Distribution  231

Chapter 14  
The Logistic Distribution  237

Chapter 15  
The Loglogistic Distribution  246

Chapter 16  
The Gumbel/SEV Distribution  251

Chapter 17  
Non-Parametric Life Data Analysis  257

Chapter 18  
Competing Failure Modes Analysis  264

Chapter 19  
Warranty Data Analysis  274

Chapter 20  
Recurrent Event Data Analysis  305

Chapter 21  
Degradation Data Analysis  319

Chapter 22  
Reliability Test Design  329

Chapter 23  
Stress-Strength Analysis  351
Comparing Life Data Sets  361
Risk Analysis and Probabilistic Design with Monte Carlo Simulation  365
Weibull++ SimuMatic  372
Target Reliability Tool  375
Event Log Data Analysis  383

Appendices  395
Least Squares/Rank Regression Equations  395
Introduction to Life Data Analysis

Reliability Life Data Analysis refers to the study and modeling of observed product lives. Life data can be lifetimes of products in the marketplace, such as the time the product operated successfully or the time the product operated before it failed. These lifetimes can be measured in hours, miles, cycles-to-failure, stress cycles or any other metric with which the life or exposure of a product can be measured. All such data of product lifetimes can be encompassed in the term life data or, more specifically, product life data. The subsequent analysis and prediction are described as life data analysis. For the purpose of this reference, we will limit our examples and discussions to lifetimes of inanimate objects, such as equipment, components and systems as they apply to reliability engineering, however the same concepts can be applied in other areas.

An Overview of Basic Concepts

When performing life data analysis (also commonly referred to as Weibull analysis), the practitioner attempts to make predictions about the life of all products in the population by fitting a statistical distribution (model) to life data from a representative sample of units. The parameterized distribution for the data set can then be used to estimate important life characteristics of the product such as reliability or probability of failure at a specific time, the mean life and the failure rate. Life data analysis requires the practitioner to:

1. Gather life data for the product.
2. Select a lifetime distribution that will fit the data and model the life of the product.
3. Estimate the parameters that will fit the distribution to the data.
4. Generate plots and results that estimate the life characteristics of the product, such as the reliability or mean life.

Lifetime Distributions (Life Data Models)

Statistical distributions have been formulated by statisticians, mathematicians and engineers to mathematically model or represent certain behavior. The probability density function (pdf) is a mathematical function that describes the distribution. The pdf can be represented mathematically or on a plot where the x-axis represents time, as shown next.

The 3-parameter Weibull pdf is given by:

\[ f(t) = \frac{\beta}{\eta} \left( t - \gamma \right)^{\beta-1} e^{-\left( \frac{t - \gamma}{\eta} \right)^\beta} \]

where:

- \( f(t) \geq 0, \ t \geq 0 \) or \( \gamma \)
- \( \beta > 0 \)
- \( \eta > 0 \)
- \( -\infty < \gamma < +\infty \)

and:

- \( \eta = \text{scale parameter, or characteristic life} \)
- \( \beta = \text{shape parameter (or slope)} \)
- \( \gamma = \text{location parameter (or failure free life)} \)
Some distributions, such as the Weibull and lognormal, tend to better represent life data and are commonly called "lifetime distributions" or "life distributions." In fact, life data analysis is sometimes called "Weibull analysis" because the Weibull distribution, formulated by Professor Waloddi Weibull, is a popular distribution for analyzing life data. The Weibull model can be applied in a variety of forms (including 1-parameter, 2-parameter, 3-parameter or mixed Weibull). Other commonly used life distributions include the exponential, lognormal and normal distributions. The analyst chooses the life distribution that is most appropriate to model each particular data set based on past experience and goodness-of-fit tests.

**Parameter Estimation**

In order to fit a statistical model to a life data set, the analyst estimates the parameters of the life distribution that will make the function most closely fit the data. The parameters control the scale, shape and location of the PDF function. For example, in the 3-parameter Weibull model (shown above), the scale parameter, $\eta$, defines where the bulk of the distribution lies. The shape parameter, $\beta$, defines the shape of the distribution and the location parameter, $\gamma$, defines the location of the distribution in time.

Several methods have been devised to estimate the parameters that will fit a lifetime distribution to a particular data set. Some available parameter estimation methods include probability plotting, rank regression on x (RRX), rank regression on y (RRY) and maximum likelihood estimation (MLE). The appropriate analysis method will vary depending on the data set and, in some cases, on the life distribution selected.

**Calculated Results and Plots**

Once you have calculated the parameters to fit a life distribution to a particular data set, you can obtain a variety of plots and calculated results from the analysis, including:

- **Reliability Given Time**: The probability that a unit will operate successfully at a particular point in time. For example, there is an 88% chance that the product will operate successfully after 3 years of operation.

- **Probability of Failure Given Time**: The probability that a unit will be failed at a particular point in time. Probability of failure is also known as "unreliability" and it is the reciprocal of the reliability. For example, there is a 12% chance that the unit will be failed after 3 years of operation (probability of failure or unreliability) and an 88% chance that it will operate successfully (reliability).

- **Mean Life**: The average time that the units in the population are expected to operate before failure. This metric is often referred to as "mean time to failure" (MTTF) or "mean time before failure" (MTBF).
• **Failure Rate:** The number of failures per unit time that can be expected to occur for the product.

• **Warranty Time:** The estimated time when the reliability will be equal to a specified goal. For example, the estimated time of operation is 4 years for a reliability of 90%.

• **B(X) Life:** The estimated time when the probability of failure will reach a specified point (X%). For example, if 10% of the products are expected to fail by 4 years of operation, then the B(10) life is 4 years. (Note that this is equivalent to a warranty time of 4 years for a 90% reliability.)

• **Probability Plot:** A plot of the probability of failure over time. (Note that probability plots are based on the linearization of a specific distribution. Consequently, the form of a probability plot for one distribution will be different than the form for another. For example, an exponential distribution probability plot has different axes than those of a normal distribution probability plot.)

• **Reliability vs. Time Plot:** A plot of the reliability over time.

• **pdf Plot:** A plot of the probability density function (pdf).

• **Failure Rate vs. Time Plot:** A plot of the failure rate over time.

• **Contour Plot:** A graphical representation of the possible solutions to the likelihood ratio equation. This is employed to make comparisons between two different data sets.

**Confidence Bounds**

Because life data analysis results are estimates based on the observed lifetimes of a sampling of units, there is uncertainty in the results due to the limited sample sizes. "Confidence bounds" (also called "confidence intervals") are used to quantify this uncertainty due to sampling error by expressing the confidence that a specific interval contains the quantity of interest. Whether or not a specific interval contains the quantity of interest is unknown.

Confidence bounds can be expressed as two-sided or one-sided. Two-sided bounds are used to indicate that the quantity of interest is contained within the bounds with a specific confidence. One-sided bounds are used to indicate that the quantity of interest is above the lower bound or below the upper bound with a specific confidence. The appropriate type of bounds depends on the application. For example, the analyst would use a one-sided lower bound on reliability, a one-sided upper bound for percent failing under warranty and two-sided bounds on the parameters of the distribution. (Note that one-sided and two-sided bounds are related. For example, the 90% lower two-sided bound is the 95% lower one-sided bound and the 90% upper two-sided bounds is the 95% upper one-sided bound.)

**Reliability Engineering**

Since the beginning of history, humanity has attempted to predict the future. Watching the flight of birds, the movement of the leaves on the trees and other methods were some of the practices used. Fortunately, today’s engineers do not have to depend on Pythia or a crystal ball in order to predict the future of their products. Through the use of life data analysis, reliability engineers use product life data to determine the probability and capability of parts, components, and systems to perform their required functions for desired periods of time without failure, in specified environments.
Life data can be lifetimes of products in the marketplace, such as the time the product operated successfully or the time the product operated before it failed. These lifetimes can be measured in hours, miles, cycles-to-failure, stress cycles or any other metric with which the life or exposure of a product can be measured. All such data of product lifetimes can be encompassed in the term *life data* or, more specifically, *product life data*. The subsequent analysis and prediction are described as *life data analysis*. For the purpose of this reference, we will limit our examples and discussions to lifetimes of inanimate objects, such as equipment, components and systems as they apply to reliability engineering. Before performing life data analysis, the failure mode and the life units (hours, cycles, miles, etc.) must be specified and clearly defined. Further, it is quite necessary to define exactly what constitutes a failure. In other words, before performing the analysis it must be clear when the product is considered to have actually failed. This may seem rather obvious, but it is not uncommon for problems with failure definitions or time unit discrepancies to completely invalidate the results of expensive and time consuming life testing and analysis.

**Estimation**

In life data analysis and reliability engineering, the output of the analysis is always an estimate. The true value of the probability of failure, the probability of success (or reliability), the mean life, the parameters of a distribution or any other applicable parameter is never known, and will almost certainly remain unknown to us for all practical purposes. Granted, once a product is no longer manufactured and all units that were ever produced have failed and all of that data has been collected and analyzed, one could claim to have learned the true value of the reliability of the product. Obviously, this is not a common occurrence. The objective of reliability engineering and life data analysis is to accurately estimate these true values. For example, let's assume that our job is to estimate the number of black marbles in a giant swimming pool filled with black and white marbles. One method is to pick out a small sample of marbles and count the black ones. Suppose we picked out ten marbles and counted four black marbles.
Based on this sampling, the estimate would be that 40% of the marbles are black. If we put the ten marbles back in the pool and repeated this step again, we might get five black marbles, changing the estimate to 50% black marbles. The range of our estimate for the percentage of black marbles in the pool is 40% to 50%. If we now repeat the experiment and pick out 1,000 marbles, we might get results for the number of black marbles such as 445 and 495 black marbles for each trial. In this case, we note that our estimate for the percentage of black marbles has a narrower range, or 44.5% to 49.5%. Using this, we can see that the larger the sample size, the narrower the estimate range and, presumably, the closer the estimate range is to the true value.

A Brief Introduction to Reliability

A Formal Definition

Reliability engineering provides the theoretical and practical tools whereby the probability and capability of parts, components, equipment, products and systems to perform their required functions for desired periods of time without failure, in specified environments and with a desired confidence, can be specified, designed in, predicted, tested and demonstrated, as discussed in Kececioglu [19].

Reliability Engineering and Business Plans

Reliability engineering assessment is based on the results of testing from in-house (or contracted) labs and data pertaining to the performance results of the product in the field. The data produced by these sources are utilized to accurately measure and improve the reliability of the products being produced. This is particularly important as market concerns drive a constant push for cost reduction. However, one must be able to keep a perspective on the big picture instead of merely looking for the quick fix. It is often the temptation to cut corners and save initial costs by using cheaper parts or cutting testing programs. Unfortunately, cheaper parts are usually less reliable and inadequate testing programs can allow products with undiscovered flaws to get out into the field. A quick savings in the short term by the use of cheaper components or small test sample sizes will usually result in higher long-term costs in the form of warranty costs or loss of customer confidence. The proper balance must be struck between reliability, customer satisfaction, time to market, sales and features. The figure below illustrates this concept. The polygon on the left represents a properly balanced project. The polygon on the right represents a project in which reliability and customer satisfaction have been sacrificed for the sake of sales and time to market.

Through proper testing and analysis in the in-house testing labs, as well as collection of adequate and meaningful data on a product's performance in the field, the reliability of any product can be measured, tracked and improved, leading to a balanced organization with a financially healthy outlook for the future.
Key Reasons for Reliability Engineering

1. For a company to succeed in today's highly competitive and technologically complex environment, it is “essential” that it knows the reliability of its product and is able to control it in order to produce products at an optimum reliability level. This yields the minimum life-cycle cost for the user and minimizes the manufacturer's costs of such a product without compromising the product's reliability and quality, as discussed in Kececioglu [19].

2. Our growing dependence on technology requires that the products that make up our daily lives successfully work for the desired or designed-in period of time. It is not sufficient that a product works for time shorter than its mission duration, but at the same time there is no need to design a product to operate much past its intended life, since this would impose additional costs on the manufacturer. In today's complex world where many important operations are performed with automated equipment, we are dependent on the successful operation of these equipment (i.e., their reliability) and, if they fail, on their quick restoration to function (i.e., their maintainability), as discussed in Kececioglu [19].

3. Product failures have varying effects, ranging from those that cause minor nuisances, such as the failure of a television's remote control (which can become a major nuisance, if not a catastrophe, depending on the football schedule of the day), to catastrophic failures involving loss of life and property, such as an aircraft accident. Reliability engineering was born out of the necessity to avoid such catastrophic events and, with them, the unnecessary loss of life and property. It is not surprising that Boeing was one of the first commercial companies to embrace and implement reliability engineering, the success of which can be seen in the safety of today's commercial air travel.

4. Today, reliability engineering can and should be applied to many products. The previous example of the failed remote control does not have any major life and death consequences to the consumer. However, it may pose a life and death risk to a non-biological entity: the company that produced it. Today's consumer is more intelligent and product-aware than the consumer of years past. The modern consumer will no longer tolerate products that do not perform in a reliable fashion, or as promised or advertised. Customer dissatisfaction with a product's reliability can have disastrous financial consequences to the manufacturer. Statistics show that when a customer is satisfied with a product he might tell eight other people; however, a dissatisfied customer will tell 22 people, on average.

5. The critical applications with which many modern products are entrusted make their reliability a factor of paramount importance. For example, the failure of a computer component will have more negative consequences today than it did twenty years ago. This is because twenty years ago the technology was relatively new and not very widespread, and one most likely had backup paper copies somewhere. Now, as computers are often the sole medium in which many clerical and computational functions are performed, the failure of a computer component will have a much greater effect.

Disciplines Covered by Reliability Engineering

Reliability engineering covers all aspects of a product's life, from its conception, subsequent design and production processes, through its practical use lifetime, with maintenance support and availability. Reliability engineering covers:
1. Reliability
2. Maintainability
3. Availability

All three of these areas can be numerically quantified with the use of reliability engineering principles and life data analysis. And the combination of these three areas introduces a new term, as defined in ISO-9000-4, "Dependability."

**A Few Common Sense Applications**

**The Reliability Bathtub Curve**

Most products (as well as humans) exhibit failure characteristics as shown in the bathtub curve of the following figure. (Do note, however, that this figure is somewhat idealized.)

![Bathtub Curve Diagram](image.png)

This curve is plotted with the product life on the x-axis and with the failure rate on the y-axis. The life can be in minutes, hours, years, cycles, actuations or any other quantifiable unit of time or use. The failure rate is given as failures among surviving units per time unit. As can be seen from this plot, many products will begin their lives with a higher failure rate (which can be due to manufacturing defects, poor workmanship, poor quality control of incoming parts, etc.) and exhibit a decreasing failure rate. The failure rate then usually stabilizes to an approximately constant rate in the useful life region, where the failures observed are chance failures. As the products experience more use and wear, the failure rate begins to rise as the population begins to experience failures related to wear-out. In the case of human mortality, the mortality rate (failure rate), is higher during the first year or so of life, then drops to a low constant level during our teens and early adult life and then rises as we progress in years.

**Burn-In**

Looking at this particular bathtub curve, it should be fairly obvious that it would be best to ship a product at the beginning of the useful life region, rather than right off the production line; thus preventing the customer from experiencing early failures. This practice is what is commonly referred to as "burn-in", and is frequently performed for electronic components. The determination of the correct burn-in time requires the use of reliability methodologies, as well as optimization of costs involved (i.e., costs of early failures vs. the cost of burn-in), to determine the optimum failure rate at shipment.
Minimizing the Manufacturer's Cost

The following shows the product reliability on the x-axis and the producer's cost on the y-axis.

If the producer increases the reliability of his product, he will increase the cost of the design and/or production of the product. However, a low production and design cost does not imply a low overall product cost. The overall product cost should not be calculated as merely the cost of the product when it leaves the shipping dock, but as the total cost of the product through its lifetime. This includes warranty and replacement costs for defective products, costs incurred by loss of customers due to defective products, loss of subsequent sales, etc. By increasing product reliability, one may increase the initial product costs, but decrease the support costs. An optimum minimal total product cost can be determined and implemented by calculating the optimum reliability for such a product. The figure depicts such a scenario. The total product cost is the sum of the production and design costs as well as the other post-shipment costs. It can be seen that at an optimum reliability level, the total product cost is at a minimum. The "optimum reliability level" is the one that coincides with the minimum total cost over the entire lifetime of the product.

Advantages of a Reliability Engineering Program

The following list presents some of the useful information that can be obtained with the implementation of a sound reliability program:

1. Optimum burn-in time or breaking-in period.
2. Optimum warranty period and estimated warranty costs.
3. Optimum preventive replacement time for components in a repairable system.
4. Spare parts requirements and production rate, resulting in improved inventory control through correct prediction of spare parts requirements.
5. Better information about the types of failures experienced by parts and systems that aid design, research and development efforts to minimize these failures.
6. Establishment of which failures occur at what time in the life of a product and better preparation to cope with them.
7. Studies of the effects of age, mission duration and application and operation stress levels on reliability.
8. A basis for comparing two or more designs and choosing the best design from the reliability point of view.
9. Evaluation of the amount of redundancy present in the design.
10. Estimations of the required redundancy to achieve the specified reliability.
11. Guidance regarding corrective action decisions to minimize failures and reduce maintenance and repair times, which will eliminate overdesign as well as underdesign.
12. Help provide guidelines for quality control practices.
13. Optimization of the reliability goal that should be designed into products and systems for minimum total cost to own, operate and maintain for their lifetime.
14. The ability to conduct trade-off studies among parameters such as reliability, maintainability, availability, cost, weight, volume, operability, serviceability and safety to obtain the optimum design.
15. Reduction of warranty costs or, for the same cost, increase in the length and the coverage of warranty.
16. Establishment of guidelines for evaluating suppliers from the point of view of their product reliability.
17. Promotion of sales on the basis of reliability indexes and metrics through sales and marketing departments.
18. Increase of customer satisfaction and an increase of sales as a result of customer satisfaction.
19. Increase of profits or, for the same profit, provision of even more reliable products and systems.
20. Promotion of positive image and company reputation.

Summary: Key Reasons for Implementing a Reliability Engineering Program

1. The typical manufacturer does not really know how satisfactorily its products are functioning. This is usually due to a lack of a reliability-wise viable failure reporting system. It is important to have a useful analysis, interpretation and feedback system in all company areas that deal with the product from its birth to its death.
2. If the manufacturer's products are functioning truly satisfactorily, it might be because they are unnecessarily over-designed, hence they are not designed optimally. Consequently, the products may be costing more than necessary and lowering profits.
3. Products are becoming more complex yearly, with the addition of more components and features to match competitors' products. This means that products with currently acceptable reliabilities need to be monitored constantly as the addition of features and components may degrade the product's overall reliability.
4. If the manufacturer does not design its products with reliability and quality in mind, SOMEONE ELSE WILL.

Reliability and Quality Control

Although the terms reliability and quality are often used interchangeably, there is a difference between these two disciplines. While reliability is concerned with the performance of a product over its entire lifetime, quality control is concerned with the performance of a product at one point in time, usually during the manufacturing process. As stated in the definition, reliability assures that components, equipment and systems function without failure for desired periods during their whole design life, from conception (birth) to junking (death). Quality control is a single, albeit vital, link in the total reliability process. Quality control assures conformance to specifications. This reduces manufacturing variance, which can degrade reliability. Quality control also checks that the incoming parts and components meet specifications, that products are inspected and tested correctly, and that the shipped products have a quality level equal to or greater than that specified. The specified quality level should be one that is acceptable to the users, the consumer and the public. No product can perform reliably without the inputs of quality control because quality parts and components are needed to go into the product so that its reliability is assured.
Chapter 2

Basic Statistical Background

This section provides a brief elementary introduction to the most common and fundamental statistical equations and definitions used in reliability engineering and life data analysis.

Random Variables

In general, most problems in reliability engineering deal with quantitative measures, such as the time-to-failure of a component, or qualitative measures, such as whether a component is defective or non-defective. We can then use a random variable $X$ to denote these possible measures.

In the case of times-to-failure, our random variable $X$ is the time-to-failure of the component and can take on an infinite number of possible values in a range from 0 to infinity (since we do not know the exact time a priori). Our component can be found failed at any time after time 0 (e.g., at 12 hours or at 100 hours and so forth), thus $X$ can take on any value in this range. In this case, our random variable $X$ is said to be a continuous random variable. In this reference, we will deal almost exclusively with continuous random variables.

In judging a component to be defective or non-defective, only two outcomes are possible. That is, $X$ is a random variable that can take on one of only two values (let's say defective = 0 and non-defective = 1). In this case, the variable is said to be a discrete random variable.

The Probability Density Function and the Cumulative Distribution Function

The probability density function (pdf) and cumulative distribution function (cdf) are two of the most important statistical functions in reliability and are very closely related. When these functions are known, almost any other reliability measure of interest can be derived or obtained. We will now take a closer look at these functions and how they relate to other reliability measures, such as the reliability function and failure rate.

From probability and statistics, given a continuous random variable $X$, we denote:

- The probability density function, pdf, as $f(x)$.
- The cumulative distribution function, cdf, as $F(x)$.

The pdf and cdf give a complete description of the probability distribution of a random variable. The following figure illustrates a pdf.
The next figures illustrate the pdf - cdf relationship.

If \( X \) is a continuous random variable, then the pdf of \( X \) is a function, \( f(x) \), such that for any two numbers, \( a \) and \( b \) with \( a \leq b \):

\[
P(a \leq X \leq b) = \int_a^b f(x) \, dx
\]

That is, the probability that \( X \) takes on a value in the interval \([a, b]\) is the area under the density function from \( a \) to \( b \) as shown above. The pdf represents the relative frequency of failure times as a function of time.

The cdf is a function, \( F(x) \), of a random variable \( X \), and is defined for a number \( x \) by:
That is, for a number \( x \), \( F(x) \) is the probability that the observed value of \( X \) will be at most \( x \). The \( cdf \) represents the cumulative values of the \( pdf \). That is, the value of a point on the curve of the \( cdf \) represents the area under the curve to the left of that point on the \( pdf \). In reliability, the \( cdf \) is used to measure the probability that the item in question will fail before the associated time value, \( t \), and is also called unreliability.

Note that depending on the density function, denoted by \( f(x) \), the limits will vary based on the region over which the distribution is defined. For example, for the life distributions considered in this reference, with the exception of the normal distribution, this range would be \([0, +\infty] \).

**Mathematical Relationship: pdf and cdf**

The mathematical relationship between the \( pdf \) and \( cdf \) is given by:

\[
F(x) = \int_0^x f(s) \, ds
\]

where \( s \) is a dummy integration variable.

Conversely:

\[
f(x) = \frac{d(F(x))}{dx}
\]

The \( cdf \) is the area under the probability density function up to a value of \( x \). The total area under the \( pdf \) is always equal to 1, or mathematically:

\[
\int_{-\infty}^{+\infty} f(x) \, dx = 1
\]

The well-known normal (or Gaussian) distribution is an example of a probability density function. The \( pdf \) for this distribution is given by:

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}
\]

where \( \mu \) is the mean and \( \sigma \) is the standard deviation. The normal distribution has two parameters, \( \mu \) and \( \sigma \).

Another is the lognormal distribution, whose \( pdf \) is given by:

\[
f(t) = \frac{1}{t \cdot \sigma' \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln t - \mu'}{\sigma'} \right)^2}
\]
where $\mu$ is the mean of the natural logarithms of the times-to-failure and $\sigma$ is the standard deviation of the natural logarithms of the times-to-failure. Again, this is a 2-parameter distribution.

**Reliability Function**

The reliability function can be derived using the previous definition of the cumulative distribution function, $F(x) = \int_0^x f(s)ds$. From our definition of the *cdf*, the probability of an event occurring by time $t$ is given by:

$$F(t) = \int_0^t f(s)ds$$

Or, one could equate this event to the probability of a unit failing by time $t$.

Since this function defines the probability of failure by a certain time, we could consider this the unreliability function. Subtracting this probability from 1 will give us the reliability function, one of the most important functions in life data analysis. The reliability function gives the probability of success of a unit undertaking a mission of a given time duration. The following figure illustrates this.

![Reliability and unreliability](image)

To show this mathematically, we first define the unreliability function, $Q(t)$, which is the probability of failure, or the probability that our time-to-failure is in the region of 0 and $t$. This is the same as the *cdf*. So from $F(t) = \int_0^t f(s)ds$:

$$Q(t) = 1 - F(t) = 1 - \int_0^t f(s)ds$$

Reliability and unreliability are the only two events being considered and they are mutually exclusive; hence, the sum of these probabilities is equal to unity.

Then:
Basic Statistical Background

Conversely:

\[ f(t) = -\frac{d[R(t)]}{dt} \]

**Conditional Reliability Function**

Conditional reliability is the probability of successfully completing another mission following the successful completion of a previous mission. The time of the previous mission and the time for the mission to be undertaken must be taken into account for conditional reliability calculations. The conditional reliability function is given by:

\[ R(T, t) = \frac{R(T+t)}{R(T)} \]

**Failure Rate Function**

The failure rate function enables the determination of the number of failures occurring per unit time. Omitting the derivation, the failure rate is mathematically given as:

\[ \lambda(t) = \frac{f(t)}{R(t)} \]

This gives the instantaneous failure rate, also known as the hazard function. It is useful in characterizing the failure behavior of a component, determining maintenance crew allocation, planning for spares provisioning, etc. Failure rate is denoted as failures per unit time.

**Mean Life (MTTF)**

The mean life function, which provides a measure of the average time of operation to failure, is given by:

\[ \bar{T} = m = \int_0^\infty t \cdot f(t) \, dt \]

This is the expected or average time-to-failure and is denoted as the MTTF (Mean Time To Failure).

The MTTF, even though an index of reliability performance, does not give any information on the failure distribution of the component in question when dealing with most lifetime distributions. Because vastly different distributions can have identical means, it is unwise to use the MTTF as the sole measure of the reliability of a component.
**Median Life**

Median life, $\hat{t}$, is the value of the random variable that has exactly one-half of the area under the pdf to its left and one-half to its right. It represents the centroid of the distribution. The median is obtained by solving the following equation for $\hat{t}$. (For individual data, the median is the midpoint value.)

$$\int_{-\infty}^{\hat{t}} f(t) \, dt = 0.5$$

**Modal Life (or Mode)**

The modal life (or mode), $\hat{t}$, is the value of $T$ that satisfies:

$$\frac{d[f(t)]}{dt} = 0$$

For a continuous distribution, the mode is that value of $t$ that corresponds to the maximum probability density (the value at which the pdf has its maximum value, or the peak of the curve).

**Lifetime Distributions**

A statistical distribution is fully described by its pdf. In the previous sections, we used the definition of the pdf to show how all other functions most commonly used in reliability engineering and life data analysis can be derived. The reliability function, failure rate function, mean time function, and median life function can be determined directly from the pdf definition, or $f(t)$. Different distributions exist, such as the normal (Gaussian), exponential, Weibull, etc., and each has a predefined form of $f(t)$ that can be found in many references. In fact, there are certain references that are devoted exclusively to different types of statistical distributions. These distributions were formulated by statisticians, mathematicians and engineers to mathematically model or represent certain behavior. For example, the Weibull distribution was formulated by Waloddi Weibull and thus it bears his name. Some distributions tend to better represent life data and are most commonly called "lifetime distributions".

A more detailed introduction to this topic is presented in Life Distributions.
Chapter 3

Life Distributions

We use the term life distributions to describe the collection of statistical probability distributions that we use in reliability engineering and life data analysis. A statistical distribution is fully described by its pdf (or probability density function). In the previous sections, we used the definition of the pdf to show how all other functions most commonly used in reliability engineering and life data analysis can be derived; namely, the reliability function, failure rate function, mean time function and median life function, etc. All of these can be determined directly from the pdf definition, or \( f(t) \). Different distributions exist, such as the normal, exponential, etc., and each one of them has a predefined form of \( f(t) \). These distribution definitions can be found in many references. In fact, entire texts have been dedicated to defining families of statistical distributions. These distributions were formulated by statisticians, mathematicians and engineers to mathematically model or represent certain behavior. For example, the Weibull distribution was formulated by Waloddi Weibull, and thus it bears his name. Some distributions tend to better represent life data and are commonly called lifetime distributions. One of the simplest and most commonly used distributions (and often erroneously overused due to its simplicity) is the exponential distribution. The pdf of the exponential distribution is mathematically defined as:

\[
f(t) = \lambda e^{-\lambda t}
\]

In this definition, note that \( t \) is our random variable, which represents time, and the Greek letter \( \lambda \) (lambda) represents what is commonly referred to as the parameter of the distribution. Depending on the value of \( \lambda \), \( f(t) \) will be scaled differently. For any distribution, the parameter or parameters of the distribution are estimated from the data. For example, the well-known normal (or Gaussian) distribution is given by:

\[
f(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t - \mu}{\sigma} \right)^2}
\]

\( \mu \), the mean, and \( \sigma \), the standard deviation, are its parameters. Both of these parameters are estimated from the data (i.e., the mean and standard deviation of the data). Once these parameters have been estimated, our function \( f(t) \) is fully defined and we can obtain any value for \( f(t) \) given any value of \( t \).

Given the mathematical representation of a distribution, we can also derive all of the functions needed for life data analysis, which again will depend only on the value of \( t \) after the value of the distribution parameter or parameters have been estimated from data. For example, we know that the exponential distribution pdf is given by:

\[
f(t) = \lambda e^{-\lambda t}
\]

Thus, the exponential reliability function can be derived as:

\[
R(t) = 1 - \int_0^t \lambda e^{-\lambda s} ds
\]

\[
= 1 - \left[ e^{-\lambda t} \right]
\]

\[
= e^{-\lambda t}
\]

The exponential failure rate function is:

\[
\lambda(t) = \frac{f(t)}{R(t)}
\]

\[
= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}
\]

\[
= \lambda
\]
The exponential mean-time-to-failure (MTTF) is given by:

\[
\mu = \int_0^\infty t \cdot f(t) \, dt = \int_0^\infty t \cdot \lambda \cdot e^{-\lambda t} \, dt = \frac{1}{\lambda}
\]

This exact same methodology can be applied to any distribution given its pdf, with various degrees of difficulty depending on the complexity of \(f(t)\).

**Parameter Types**

Distributions can have any number of parameters. Do note that as the number of parameters increases, so does the amount of data required for a proper fit. In general, the lifetime distributions used for reliability and life data analysis are usually limited to a maximum of three parameters. These three parameters are usually known as the scale parameter, the shape parameter and the location parameter.

- **Scale Parameter** The scale parameter is the most common type of parameter. All distributions in this reference have a scale parameter. In the case of one-parameter distributions, the sole parameter is the scale parameter. The scale parameter defines where the bulk of the distribution lies, or how stretched out the distribution is. In the case of the normal distribution, the scale parameter is the standard deviation.

- **Shape Parameter** The shape parameter, as the name implies, helps define the shape of a distribution. Some distributions, such as the exponential or normal, do not have a shape parameter since they have a predefined shape that does not change. In the case of the normal distribution, the shape is always the familiar bell shape. The effect of the shape parameter on a distribution is reflected in the shapes of the pdf, the reliability function and the failure rate function.

- **Location Parameter** The location parameter is used to shift a distribution in one direction or another. The location parameter, usually denoted as \(\gamma\), defines the location of the origin of a distribution and can be either positive or negative. In terms of lifetime distributions, the location parameter represents a time shift.

This means that the inclusion of a location parameter for a distribution whose domain is normally \([0, \infty)\) will change the domain to \([\gamma, \infty)\) where \(\gamma\) can either be positive or negative. This can have some profound effects in terms of reliability. For a positive location parameter, this indicates that the reliability for that particular distribution is always 100% up to that point. In other words, a failure cannot occur before this time \(\gamma\). Many engineers feel uncomfortable in saying that a failure will absolutely not happen before any given time. On the other hand, the argument can be
made that almost all life distributions have a location parameter, although many of them may be negligibly small. Similarly, many people are uncomfortable with the concept of a negative location parameter, which states that failures theoretically occur before time zero. Realistically, the calculation of a negative location parameter is indicative of quiescent failures (failures that occur before a product is used for the first time) or of problems with the manufacturing, packaging or shipping process. More attention will be given to the concept of the location parameter in subsequent discussions of the exponential and Weibull distributions, which are the lifetime distributions that most frequently employ the location parameter.

Most Commonly Used Distributions

There are many different lifetime distributions that can be used to model reliability data. Leemis [22] presents a good overview of many of these distributions. In this reference, we will concentrate on the most commonly used and most widely applicable distributions for life data analysis, as outlined in the following sections.

The Exponential Distribution

The exponential distribution is commonly used for components or systems exhibiting a constant failure rate. Due to its simplicity, it has been widely employed, even in cases where it doesn't apply. In its most general case, the 2-parameter exponential distribution is defined by:

\[
f(t) = \lambda e^{-\lambda (t-\gamma)}
\]

Where \( \lambda \) is the constant failure rate in failures per unit of measurement (e.g., failures per hour, per cycle, etc.) and \( \gamma \) is the location parameter. In addition, \( \lambda = \frac{1}{m} \) where \( m \) is the mean time between failures (or to failure).

If the location parameter, \( \gamma \), is assumed to be zero, then the distribution becomes the 1-parameter exponential or:

\[
f(t) = \lambda e^{-\lambda t}
\]

For a detailed discussion of this distribution, see The Exponential Distribution.

The Weibull Distribution

The Weibull distribution is a general purpose reliability distribution used to model material strength, times-to-failure of electronic and mechanical components, equipment or systems. In its most general case, the 3-parameter Weibull pdf is defined by:

\[
f(t) = \frac{\beta}{\eta} \left( \frac{t-\gamma}{\eta} \right)^{\beta-1} e^{-\left( \frac{t-\gamma}{\eta} \right)^{\beta}}
\]

where \( \beta \) = shape parameter, \( \eta \) = scale parameter and \( \gamma \) = location parameter.

If the location parameter, \( \gamma \), is assumed to be zero, then the distribution becomes the 2-parameter Weibull or:

\[
f(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta-1} e^{-\left( \frac{t}{\eta} \right)^{\beta}}
\]

One additional form is the 1-parameter Weibull distribution, which assumes that the location parameter, \( \gamma \) is zero, and the shape parameter is a known constant, or \( \beta \) = constant = \( C \), so:

\[
f(t) = \frac{C}{\eta} \left( \frac{t}{\eta} \right)^{C-1} e^{-\left( \frac{t}{\eta} \right)^{C}}
\]

For a detailed discussion of this distribution, see The Weibull Distribution.
Bayesian-Weibull Analysis

Another approach is the Weibull-Bayesian analysis method, which assumes that the analyst has some prior knowledge about the distribution of the shape parameter of the Weibull distribution (beta). There are many practical applications for this model, particularly when dealing with small sample sizes and/or when some prior knowledge for the shape parameter is available. For example, when a test is performed, there is often a good understanding about the behavior of the failure mode under investigation, primarily through historical data or physics-of-failure.

Note that this is not the same as the so called "WeiBayes model," which is really a one-parameter Weibull distribution that assumes a fixed value (constant) for the shape parameter and solves for the scale parameter. The Bayesian-Weibull feature in Weibull++ is actually a true Bayesian model and offers an alternative to the one-parameter Weibull by including the variation and uncertainty that is present in the prior estimation of the shape parameter.

This analysis method and its characteristics are presented in detail in Bayesian-Weibull Analysis.

The Normal Distribution

The normal distribution is commonly used for general reliability analysis, times-to-failure of simple electronic and mechanical components, equipment or systems. The pdf of the normal distribution is given by:

\[
f(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2}
\]

where \( \mu \) is the mean of the normal times to failure and \( \sigma \) is the standard deviation of the times to failure.

The normal distribution and its characteristics are presented in The Normal Distribution.

The Lognormal Distribution

The lognormal distribution is commonly used for general reliability analysis, cycles-to-failure in fatigue, material strengths and loading variables in probabilistic design. When the natural logarithms of the times-to-failure are normally distributed, then we say that the data follow the lognormal distribution.

The pdf of the lognormal distribution is given by:

\[
f(t) = \frac{1}{t\sigma' \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln(t')-\mu'}{\sigma'} \right)^2}
\]

\[
f(t) \geq 0, t > 0, \sigma' > 0
\]

\[t' = \ln(t)\]

where \( \mu' \) is the mean of the natural logarithms of the times-to-failure and \( \sigma' \) is the standard deviation of the natural logarithms of the times to failure.

For a detailed discussion of this distribution, see The Lognormal Distribution.

Other Distributions

In addition to the distributions mentioned above, which are more frequently used in life data analysis, the following distributions also have a variety of applications and can be found in many statistical references. They are included in Weibull++, as well as discussed in this reference.

The Mixed Weibull Distribution

The mixed Weibull distribution is commonly used for modeling the behavior of components or systems exhibiting multiple failure modes (mixed populations). It gives the global picture of the life of a product by mixing different Weibull distributions for different stages of the product’s life and is defined by:
\[ f_S(t) = \sum_{i=1}^{S} p_i \frac{\beta_i}{\eta_i} \left( \frac{t}{\eta_i} \right)^{\beta_i - 1} e^{-\left( \frac{t}{\eta_i} \right)^{\beta_i}} \]

where the value of \( S \) is equal to the number of subpopulations. Note that this results in a total of \( (3 \cdot S - 1) \) parameters. In other words, each population has a portion or mixing weight for the \( i^{th} \) population, a \( \beta_i \) or shape parameter for the \( i^{th} \) population and or scale parameter \( \eta_i \) for \( i^{th} \) population. Note that the parameters are reduced to \((3 \cdot S - 1)\), given the fact that the following condition can also be used:

\[ \sum_{i=1}^{S} p_i = 1 \]

The mixed Weibull distribution and its characteristics are presented in The Mixed Weibull Distribution.

The Generalized Gamma Distribution

Compared to the other distributions previously discussed, the generalized gamma distribution is not as frequently used for modeling life data; however, it has the ability to mimic the attributes of other distributions, such as the Weibull or lognormal, based on the values of the distribution’s parameters. This offers a compromise between two lifetime distributions. The generalized gamma function is a three-parameter distribution with parameters \( \mu \), \( \sigma \) and \( \lambda \). The pdf of the distribution is given by,

\[ f(x) = \begin{cases} \frac{|\lambda| \cdot 1}{\Gamma\left(\frac{\lambda}{\sigma}\right)} \cdot e^{-\lambda \frac{|\ln(x) - \mu|}{\sigma} - \frac{|\ln(x) - \mu|}{\sigma}} & \text{if } \lambda \neq 0 \\ \frac{1}{\Gamma(\lambda)} \cdot e^{-\frac{|\ln(x) - \mu|}{\sigma}^2} & \text{if } \lambda = 0 \end{cases} \]

where \( \Gamma(x) \) is the gamma function, defined by:

\[ \Gamma(x) = \int_0^\infty s^{x-1} e^{-s} ds \]

This distribution behaves as do other distributions based on the values of the parameters. For example, if \( \lambda = 1 \), then the distribution is identical to the Weibull distribution. If both \( \lambda = 1 \) and \( \sigma = 1 \), then the distribution is identical to the exponential distribution, and for \( \lambda = 0 \) it is identical to the lognormal distribution. While the generalized gamma distribution is not often used to model life data by itself, its ability to behave like other more commonly-used life distributions is sometimes used to determine which of those life distributions should be used to model a particular set of data.

The generalized gamma distribution and its characteristics are presented in The Generalized Gamma Distribution.

The Gamma Distribution

The gamma distribution is a flexible distribution that may offer a good fit to some sets of life data. Sometimes called the Erlang distribution, the gamma distribution has applications in Bayesian analysis as a prior distribution, and it is also commonly used in queueing theory.

The pdf of the distribution is given by:

\[ f(t) = \frac{e^{\frac{z - e^z}{k}}}{k \Gamma(k)} \]

where:

\[ z = \ln t - \mu \]

\[ k = \text{shape parameter} \]

\[ \mu = \text{scale parameter} \]

where \( 0 < t < \infty \), \( -\infty < \mu < \infty \) and \( k > 0 \)

The gamma distribution and its characteristics are presented in The Gamma Distribution.
**The Logistic Distribution**

The logistic distribution has a shape very similar to the normal distribution (i.e., bell shaped), but with heavier tails. Since the logistic distribution has closed form solutions for the reliability, \( cdf \) and failure rate functions, it is sometimes preferred over the normal distribution, where these functions can only be obtained numerically.

The \( pdf \) of the logistic distribution is given by:

\[
f(t) = \frac{e^z}{\sigma (1 + e^z)^2}
\]

where:

\[
\sigma \geq 0
\]

\[
z = \frac{t - \mu}{\sigma}
\]

\[
\mu = \text{location parameter (also denoted as } T)\]

\[
\sigma = \text{scale parameter}
\]

The logistic distribution and its characteristics are presented in The Logistic Distribution.

**The Loglogistic Distribution**

As may be surmised from the name, the loglogistic distribution is similar to the logistic distribution. Specifically, the data follows a loglogistic distribution when the natural logarithms of the times-to-failure follow a logistic distribution. Accordingly, the loglogistic and lognormal distributions also share many similarities.

The \( pdf \) of the loglogistic distribution is given by:

\[
f(t) = \frac{e^{t'}}{\sigma t (1 + e^{t'})^2}
\]

where:

\[
t' = \ln(t)
\]

\[
\mu = \text{scale parameter}
\]

\[
\sigma = \text{shape parameter}
\]

The loglogistic distribution and its characteristics are presented in The Loglogistic Distribution.

**The Gumbel Distribution**

The Gumbel distribution is also referred to as the Smallest Extreme Value (SEV) distribution or the Smallest Extreme Value (Type 1) distribution. The Gumbel distribution is appropriate for modeling strength, which is sometimes skewed to the left (e.g., few weak units fail under low stress, while the rest fail at higher stresses). The Gumbel distribution could also be appropriate for modeling the life of products that experience very quick wear out after reaching a certain age.

The \( pdf \) of the Gumbel distribution is given by:

\[
f(t) = \frac{1}{\sigma} e^{\frac{t - \mu}{\sigma} - e^{\frac{t - \mu}{\sigma}}}
\]

where:

\[
f(t) \geq 0, \sigma > 0
\]
\[ \mu = \text{location parameter} \]
\[ \sigma = \text{scale parameter} \]

The Gumbel distribution and its characteristics are presented in The Gumbel/SEV Distribution.
Chapter 4

Parameter Estimation

The term parameter estimation refers to the process of using sample data (in reliability engineering, usually times-to-failure or success data) to estimate the parameters of the selected distribution. Several parameter estimation methods are available. This section presents an overview of the available methods used in life data analysis. More specifically, we start with the relatively simple method of Probability Plotting and continue with the more sophisticated methods of Rank Regression (or Least Squares), Maximum Likelihood Estimation and Bayesian Estimation Methods.

Probability Plotting

The least mathematically intensive method for parameter estimation is the method of probability plotting. As the term implies, probability plotting involves a physical plot of the data on specially constructed probability plotting paper. This method is easily implemented by hand, given that one can obtain the appropriate probability plotting paper.

The method of probability plotting takes the cdf of the distribution and attempts to linearize it by employing a specially constructed paper. The following sections illustrate the steps in this method using the 2-parameter Weibull distribution as an example. This includes:

- Linearize the unreliability function
- Construct the probability plotting paper
- Determine the X and Y positions of the plot points

And then using the plot to read any particular time or reliability/unreliability value of interest.

Linearizing the Unreliability Function

In the case of the 2-parameter Weibull, the cdf (also the unreliability \( Q(t) \)) is given by:

\[
F(t) = Q(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^\beta}
\]

This function can then be linearized (i.e., put in the common form of \( y = m't + b \)) as follows:

\[
Q(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^\beta}
\]

\[
\ln(1 - Q(t)) = \ln\left[ e^{-\left(\frac{t}{\eta}\right)^\beta} \right]
\]

\[
\ln(1 - Q(t)) = -\left(\frac{t}{\eta}\right)^\beta
\]

\[
\ln(-\ln(1 - Q(t))) = \beta \ln\left(\frac{t}{\eta}\right)
\]

\[
\ln\left(\ln\left(\frac{1}{1 - Q(t)}\right)\right) = \beta \ln t - \beta(\eta)
\]

Then by setting:
The equation can then be rewritten as:

\[ y = \beta x - \beta \ln(\eta) \]

which is now a linear equation with a slope of:

\[ m = \beta \]

and an intercept of:

\[ b = -\beta \cdot \ln(\eta) \]

**Constructing the Paper**

The next task is to construct the Weibull probability plotting paper with the appropriate y and x axes. The x-axis transformation is simply logarithmic. The y-axis is a bit more complex, requiring a double log reciprocal transformation, or:

\[ y = \ln \left( \ln \left( \frac{1}{1 - Q(t)} \right) \right) \]

where \( Q(t) \) is the unreliability.

Such papers have been created by different vendors and are called *probability plotting papers*. ReliaSoft’s reliability engineering resource website at www.weibull.com has different plotting papers available for download \(^{[1]}\).

To illustrate, consider the following probability plot on a slightly different type of Weibull probability paper.
This paper is constructed based on the mentioned y and x transformations, where the y-axis represents unreliability and the x-axis represents time. Both of these values must be known for each time-to-failure point we want to plot.

Then, given the \( y \) and \( \beta \) value for each point, the points can easily be put on the plot. Once the points have been placed on the plot, the best possible straight line is drawn through these points. Once the line has been drawn, the slope of the line can be obtained (some probability papers include a slope indicator to simplify this calculation). This is the parameter \( \beta \), which is the value of the slope. To determine the scale parameter, \( \eta \) (also called the characteristic life), one reads the time from the x-axis corresponding to \( Q(t) = 63.2\% \).

Note that at:

\[
Q(t = \eta) = 1 - e^{-\left(\frac{t}{\eta}\right)^\beta} \\
= 1 - e^{-1} \\
= 0.632 \\
= 63.2\%
\]

Thus, if we enter the y axis at \( Q(t) = 63.2\% \), the corresponding value of \( t \) will be equal to \( \eta \). Thus, using this simple methodology, the parameters of the Weibull distribution can be estimated.
Determining the X and Y Position of the Plot Points

The points on the plot represent our data or, more specifically, our times-to-failure data. If, for example, we tested four units that failed at 10, 20, 30 and 40 hours, then we would use these times as our x values or time values.

Determining the appropriate y plotting positions, or the unreliability values, is a little more complex. To determine the y plotting positions, we must first determine a value indicating the corresponding unreliability for that failure. In other words, we need to obtain the cumulative percent failed for each time-to-failure. For example, the cumulative percent failed by 10 hours may be 25%, by 20 hours 50%, and so forth. This is a simple method illustrating the idea. The problem with this simple method is the fact that the 100% point is not defined on most probability plots; thus, an alternative and more robust approach must be used. The most widely used method of determining this value is the method of obtaining the median rank for each failure, as discussed next.

Median Ranks

The Median Ranks method is used to obtain an estimate of the unreliability for each failure. The median rank is the value that the true probability of failure, \( Q(T_j) \), should have at the \( j^{th} \) failure out of a sample of \( N \) units at the 50% confidence level.

The rank can be found for any percentage point, \( P \), greater than zero and less than one, by solving the cumulative binomial equation for \( Z \). This represents the rank, or unreliability estimate, for the \( j^{th} \) failure in the following equation for the cumulative binomial:

\[
P = \sum_{k=j}^{N} \binom{N}{k} Z^k (1 - Z)^{N-k}
\]

where \( N \) is the sample size and \( j \) the order number.

The median rank is obtained by solving this equation for \( Z \) at \( P = 0.50 \):

\[
0.50 = \sum_{k=j}^{N} \binom{N}{k} Z^k (1 - Z)^{N-k}
\]

For example, if \( N = 4 \) and we have four failures, we would solve the median rank equation for the value of \( Z \) four times; once for each failure with \( j = 1, 2, 3 \) and 4. This result can then be used as the unreliability estimate for each failure or the y plotting position. (See also The Weibull Distribution for a step-by-step example of this method.)

The solution of cumulative binomial equation for \( Z \) requires the use of numerical methods.

Beta and F Distributions Approach

A more straightforward and easier method of estimating median ranks is by applying two transformations to the cumulative binomial equation, first to the beta distribution and then to the F distribution, resulting in [12, 13]:

\[
MR = \frac{1}{\frac{N-j+1}{2}} \frac{F_{0.50;m,n}}{F_{0.50;m,n}}
\]

\[
m = 2(N - j + 1)
\]

\[
n = 2j
\]

where \( F_{0.50;m,n} \) denotes the \( F \) distribution at the 0.50 point, with \( m \) and \( n \) degrees of freedom, for failure \( j \) out of \( N \) units.
Benard's Approximation for Median Ranks

Another quick, and less accurate, approximation of the median ranks is also given by:

\[ MR = \frac{j - 0.3}{N + 0.4} \]

This approximation of the median ranks is also known as Benard's approximation.

Kaplan-Meier

The Kaplan-Meier estimator (also known as the product limit estimator) is used as an alternative to the median ranks method for calculating the estimates of the unreliability for probability plotting purposes. The equation of the estimator is given by:

\[ \hat{F}(t_i) = 1 - \prod_{j=1}^{i} \frac{n_j - r_j}{n_j}, \quad i = 1, ..., m \]

where:

- \( m \) = total number of data points
- \( n \) = the total number of units
- \( n_i = n - \sum_{j=0}^{i-1} s_j - \sum_{j=0}^{i-1} r_j, \quad i = 1, ..., m \)
- \( r_j \) = number of failures in the \( j^{th} \) data group, and
- \( s_j \) = number of surviving units in the \( j^{th} \) data group

Probability Plotting Example

This same methodology can be applied to other distributions with \( cdf \) equations that can be linearized. Different probability papers exist for each distribution, because different distributions have different \( cdf \) equations. ReliaSoft's software tools automatically create these plots for you. Special scales on these plots allow you to derive the parameter estimates directly from the plots, similar to the way \( \hat{\beta} \) and \( \hat{\eta} \) were obtained from the Weibull probability plot. The following example demonstrates the method again, this time using the 1-parameter exponential distribution.

Let's assume six identical units are reliability tested at the same application and operation stress levels. All of these units fail during the test after operating for the following times (in hours): 96, 257, 498, 763, 1051 and 1744.

The steps for using the probability plotting method to determine the parameters of the exponential pdf representing the data are as follows:

1. Rank the times-to-failure in ascending order as shown next.

<table>
<thead>
<tr>
<th>Failure Time (Hr)</th>
<th>Failure Order Number out of a Sample Size of 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>257</td>
<td>2</td>
</tr>
<tr>
<td>498</td>
<td>3</td>
</tr>
<tr>
<td>763</td>
<td>4</td>
</tr>
<tr>
<td>1,051</td>
<td>5</td>
</tr>
<tr>
<td>1,744</td>
<td>6</td>
</tr>
</tbody>
</table>

2. Obtain their median rank plotting positions. Median rank positions are used instead of other ranking methods because median ranks are at a specific confidence level (50%).

The times-to-failure, with their corresponding median ranks, are shown next:
On an exponential probability paper, plot the times on the x-axis and their corresponding rank value on the y-axis. The next figure displays an example of an exponential probability paper. The paper is simply a log-linear paper.

Draw the best possible straight line that goes through the $t = 0$ and $(t) = 100\%$ point and through the plotted points (as shown in the plot below).

At the $Q(t) = 63.2\%$ or $R(t) = 36.8\%$ ordinate point, draw a straight horizontal line until this line intersects the fitted straight line. Draw a vertical line through this intersection until it crosses the abscissa. The value at the intersection of the abscissa is the estimate of the mean. For this case, $\hat{\mu} = 833$ hours which means that $\lambda = \frac{1}{\mu} = 0.0012$ (This is always at $63.2\%$ because $T = 1 - e^{-\mu} = 1 - e^{-1} = 0.632 = 63.2\%$)
Now any reliability value for any mission time \( t \) can be obtained. For example, the reliability for a mission of 15 hours, or any other time, can now be obtained either from the plot or analytically.

To obtain the value from the plot, draw a vertical line from the abscissa, at \( t = 15 \) hours, to the fitted line. Draw a horizontal line from this intersection to the ordinate and read \( R(t) \). In this case, \( R(t = 15) = 98.15\% \). This can also be obtained analytically, from the exponential reliability function.

**Comments on the Probability Plotting Method**

Besides the most obvious drawback to probability plotting, which is the amount of effort required, manual probability plotting is not always consistent in the results. Two people plotting a straight line through a set of points will not always draw this line the same way, and thus will come up with slightly different results. This method was used primarily before the widespread use of computers that could easily perform the calculations for more complicated parameter estimation methods, such as the least squares and maximum likelihood methods.

**Least Squares (Rank Regression)**

Using the idea of probability plotting, regression analysis mathematically fits the best straight line to a set of points, in an attempt to estimate the parameters. Essentially, this is a mathematically based version of the probability plotting method discussed previously.

The method of linear least squares is used for all regression analysis performed by Weibull++, except for the cases of the 3-parameter Weibull, mixed Weibull, gamma and generalized gamma distributions, where a non-linear regression technique is employed. The terms *linear regression* and *least squares* are used synonymously in this reference. In Weibull++, the term *rank regression* is used instead of least squares, or linear regression, because the regression is performed on the rank values, more specifically, the median rank values (represented on the y-axis).
The method of least squares requires that a straight line be fitted to a set of data points, such that the sum of the squares of the distance of the points to the fitted line is minimized. This minimization can be performed in either the vertical or horizontal direction. If the regression is on \( X \), then the line is fitted so that the horizontal deviations from the points to the line are minimized. If the regression is on \( Y \), then this means that the distance of the vertical deviations from the points to the line is minimized. This is illustrated in the following figure.

**Rank Regression on \( Y \)**

Assume that a set of data pairs \((x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\) were obtained and plotted, and that the \( x \)-values are known exactly. Then, according to the least squares principle, which minimizes the vertical distance between the data points and the straight line fitted to the data, the best fitting straight line to these data is the straight line \( y = \hat{a} + \hat{b}x \) where the recently introduced (\( \hat{\phantom{a}} \)) symbol indicates that this value is an estimate) such that:

\[
\sum_{i=1}^{N} \left( \hat{a} + \hat{b}x_i - y_i \right)^2 = \min \sum_{i=1}^{N} (a + bx_i - y_i)^2
\]

and where \( \hat{a} \) and \( \hat{b} \) are the least squares estimates of \( a \) and \( b \), and \( N \) is the number of data points. These equations are minimized by estimates of \( \hat{a} \) and \( \hat{b} \) such that:

\[
\hat{a} = \frac{\sum_{i=1}^{N} y_i}{N} - \frac{\sum_{i=1}^{N} x_i}{N} \cdot \hat{b} = \bar{y} - \hat{b}\bar{x}
\]

and:

\[
\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \left( \sum_{i=1}^{N} x_i \right) \left( \sum_{i=1}^{N} y_i \right) / N}{\sum_{i=1}^{N} x_i^2 - \left( \sum_{i=1}^{N} x_i \right)^2 / N}
\]
Rank Regression on X

Assume that a set of data pairs \((x_1, y_1), \ldots, (x_N, y_N)\) were obtained and plotted, and that the y-values are known exactly. The same least squares principle is applied, but this time, minimizing the horizontal distance between the data points and the straight line fitted to the data. The best fitting straight line to these data is the straight line:

\[
x = \hat{a} + \hat{b}y
\]

such that:

\[
\sum_{i=1}^{N} (\hat{a} + \hat{b}y_i - x_i)^2 = \min(a, b) \sum_{i=1}^{N} (a + by_i - x_i)^2
\]

Again, \(\hat{a}\) and \(\hat{b}\) are the least squares estimates of \(a\) and \(b\), and \(N\) is the number of data points. These equations are minimized by estimates of \(\hat{a}\) and \(\hat{b}\) such that:

\[
\hat{a} = \frac{\sum_{i=1}^{N} x_i}{N} - \frac{\sum_{i=1}^{N} y_i}{N} = \bar{x} - \bar{y}
\]

and:

\[
\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sum_{i=1}^{N} y_i^2 - \left( \frac{\sum_{i=1}^{N} y_i}{N} \right)^2}
\]

The corresponding relations for determining the parameters for specific distributions (i.e., Weibull, exponential, etc.), are presented in the chapters covering that distribution.

Correlation Coefficient

The correlation coefficient is a measure of how well the linear regression model fits the data and is usually denoted by \(\rho\). In the case of life data analysis, it is a measure for the strength of the linear relation (correlation) between the median ranks and the data. The population correlation coefficient is defined as follows:

\[
\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}
\]

where \(\sigma_{xy}\) = covariance of \(x\) and \(y\), \(\sigma_x\) = standard deviation of \(x\), and \(\sigma_y\) = standard deviation of \(y\).

The estimator of \(\rho\) is the sample correlation coefficient, \(\hat{\rho}\), given by:

\[
\hat{\rho} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sqrt{\left( \sum_{i=1}^{N} x_i^2 - \left( \frac{\sum_{i=1}^{N} x_i}{N} \right)^2 \right) \left( \sum_{i=1}^{N} y_i^2 - \left( \frac{\sum_{i=1}^{N} y_i}{N} \right)^2 \right)}}
\]

The range of \(\hat{\rho}\) is \(-1 \leq \hat{\rho} \leq 1\).
The closer the value is to $\pm 1$, the better the linear fit. Note that +1 indicates a perfect fit (the paired values $(x_i, y_i)$ lie on a straight line) with a positive slope, while -1 indicates a perfect fit with a negative slope. A correlation coefficient value of zero would indicate that the data are randomly scattered and have no pattern or correlation in relation to the regression line model.

**Comments on the Least Squares Method**

The least squares estimation method is quite good for functions that can be linearized. For these distributions, the calculations are relatively easy and straightforward, having closed-form solutions that can readily yield an answer without having to resort to numerical techniques or tables. Furthermore, this technique provides a good measure of the goodness-of-fit of the chosen distribution in the correlation coefficient. Least squares is generally best used with data sets containing complete data, that is, data consisting only of single times-to-failure with no censored or interval data. (See Life Data Classification for information about the different data types, including complete, left censored, right censored (or suspended) and interval data.)

*See also:*
- Least Squares/Rank Regression Equations
- Grouped Data Analysis

**Rank Methods for Censored Data**

All available data should be considered in the analysis of times-to-failure data. This includes the case when a particular unit in a sample has been removed from the test prior to failure. An item, or unit, which is removed from a reliability test prior to failure, or a unit which is in the field and is still operating at the time the reliability of these units is to be determined, is called a suspended item or right censored observation or right censored data point.

Suspended items analysis would also be considered when:

1. We need to make an analysis of the available results before test completion.
2. The failure modes which are occurring are different than those anticipated and such units are withdrawn from the test.
3. We need to analyze a single mode and the actual data set comprises multiple modes.
4. A warranty analysis is to be made of all units in the field (non-failed and failed units). The non-failed units are considered to be suspended items (or right censored).

This section describes the rank methods that are used in both probability plotting and least squares (rank regression) to handle censored data. This includes:
- The rank adjustment method for right censored (suspension) data.
- ReliaSoft's alternative ranking method for censored data including left censored, right censored, and interval data.
**Rank Adjustment Method for Right Censored Data**

When using the probability plotting or least squares (rank regression) method for data sets where some of the units did not fail, or were suspended, we need to adjust their probability of failure, or unreliability. As discussed before, estimates of the unreliability for complete data are obtained using the median ranks approach. The following methodology illustrates how adjusted median ranks are computed to account for right censored data. To better illustrate the methodology, consider the following example in Kececioglu [20] where five items are tested resulting in three failures and two suspensions.

<table>
<thead>
<tr>
<th>Item Number (Position)</th>
<th>Failure (F) or Suspension (S)</th>
<th>Life of item, hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$F_1$</td>
<td>5,100</td>
</tr>
<tr>
<td>2</td>
<td>$S_1$</td>
<td>9,500</td>
</tr>
<tr>
<td>3</td>
<td>$F_2$</td>
<td>15,000</td>
</tr>
<tr>
<td>4</td>
<td>$S_2$</td>
<td>22,000</td>
</tr>
<tr>
<td>5</td>
<td>$F_3$</td>
<td>40,000</td>
</tr>
</tbody>
</table>

The methodology for plotting suspended items involves adjusting the rank positions and plotting the data based on new positions, determined by the location of the suspensions. If we consider these five units, the following methodology would be used: The first item must be the first failure; hence, it is assigned failure order number $j = 1$. The actual failure order number (or position) of the second failure, $F_2$, is in doubt. It could either be in position 2 or in position 3. Had not been withdrawn from the test at 9,500 hours, it could have operated successfully past 15,000 hours, thus placing $F_2$ in position 2. Alternatively, $S_1$ could also have failed before 15,000 hours, thus placing $F_2$ in position 3. In this case, the failure order number for $F_2$ will be some number between 2 and 3. To determine this number, consider the following:

We can find the number of ways the second failure can occur in either order number 2 (position 2) or order number 3 (position 3). The possible ways are listed next.

<table>
<thead>
<tr>
<th>$F_2$ in Position 2</th>
<th>OR</th>
<th>$F_2$ in Position 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6</td>
<td>1 2</td>
<td></td>
</tr>
<tr>
<td>$F_1$ $F_2$ $F_1$ $F_1$ $F_1$ $F_1$</td>
<td>$F_1$</td>
<td>$F_1$</td>
</tr>
<tr>
<td>$F_2$ $F_2$ $F_2$ $F_2$ $F_2$ $F_2$</td>
<td>$S_1$</td>
<td>$S_1$</td>
</tr>
<tr>
<td>$S_1$ $S_2$ $S_2$ $S_1$ $S_2$ $F_3$</td>
<td>$F_2$</td>
<td>$F_2$</td>
</tr>
<tr>
<td>$S_2$ $S_1$ $F_3$ $F_3$ $S_2$ $S_2$</td>
<td>$S_2$</td>
<td>$F_3$</td>
</tr>
<tr>
<td>$F_3$ $S_2$ $S_2$ $S_1$ $S_1$ $F_3$</td>
<td>$F_3$</td>
<td>$S_2$</td>
</tr>
</tbody>
</table>

It can be seen that $F_2$ can occur in the second position six ways and in the third position two ways. The most probable position is the average of these possible ways, or the *mean order number* (MON), given by:

$$F_2 = MON_2 = \frac{(6 \times 2) + (2 \times 3)}{6 + 2} = 2.25$$

Using the same logic on the third failure, it can be located in position numbers 3, 4 and 5 in the possible ways listed next.
Then, the mean order number for the third failure, (item 5) is:

\[
MON_3 = \frac{(2 \times 3) + (3 \times 4) + (3 \times 5)}{2 + 3 + 3} = 4.125
\]

Once the mean order number for each failure has been established, we obtain the median rank positions for these failures at their mean order number. Specifically, we obtain the median rank of the order numbers 1, 2.25 and 4.125 out of a sample size of 5, as given next.

<table>
<thead>
<tr>
<th>Failure Number</th>
<th>MON</th>
<th>Median Rank Position(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: ( F_1 )</td>
<td>1</td>
<td>13%</td>
</tr>
<tr>
<td>2: ( F_2 )</td>
<td>2.25</td>
<td>36%</td>
</tr>
<tr>
<td>3: ( F_3 )</td>
<td>4.125</td>
<td>71%</td>
</tr>
</tbody>
</table>

Once the median rank values have been obtained, the probability plotting analysis is identical to that presented before. As you might have noticed, this methodology is rather laborious. Other techniques and shortcuts have been developed over the years to streamline this procedure. For more details on this method, see Kececioglu [20]. Here, we will introduce one of these methods. This method calculates MON using an increment, \( I \), which is defined by:

\[
I_i = \frac{N + 1 - PMON}{1 + NIBPSS}
\]

Where
- \( N \) = the sample size, or total number of items in the test
- \( PMON \) = previous mean order number
- \( NIBPSS \) = the number of items beyond the present suspended set. It is the number of units (including all the failures and suspensions) at the current failure time.
- \( i \) = the \( i \)th failure item

MON is given as:

\[
MON_i = MON_{i-1} + I_i
\]

Let's calculate the previous example using the method.

For F1:

\[
MON_1 = MON_0 + I_1 = \frac{5 + 1 - 0}{1 + 5} = 1
\]

For F2:

\[
MON_2 = MON_1 + I_2 = \frac{5 + 1 - 1}{1 + 3} = 2.25
\]

For F3:
\[ MON_3 = MON_2 + I_3 = 2.25 + \frac{5 + 1 - 2.25}{1 + 1} = 4.125 \]

The MON obtained for each failure item via this method is same as from the first method, so the median rank values will also be the same.

For Grouped data, the increment \( I \) at each failure group will be multiplied by the number of failures in that group.

**Shortfalls of the Rank Adjustment Method**

Even though the rank adjustment method is the most widely used method for performing analysis for analysis of suspended items, we would like to point out the following shortcoming. As you may have noticed, only the position where the failure occurred is taken into account, and not the exact time-to-suspension. For example, this methodology would yield the exact same results for the next two cases.

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item Number</td>
<td>State*&quot;F&quot; or &quot;S&quot;</td>
</tr>
<tr>
<td>1</td>
<td>( F_1 )</td>
</tr>
<tr>
<td>2</td>
<td>( S_1 )</td>
</tr>
<tr>
<td>3</td>
<td>( S_2 )</td>
</tr>
<tr>
<td>4</td>
<td>( S_3 )</td>
</tr>
<tr>
<td>5</td>
<td>( F_2 )</td>
</tr>
</tbody>
</table>

* F - Failed, S - Suspended

This shortfall is significant when the number of failures is small and the number of suspensions is large and not spread uniformly between failures, as with these data. In cases like this, it is highly recommended to use maximum likelihood estimation (MLE) to estimate the parameters instead of using least squares, because MLE does not look at ranks or plotting positions, but rather considers each unique time-to-failure or suspension. For the data given above, the results are as follows. The estimated parameters using the method just described are the same for both cases (1 and 2):

\[ \hat{\beta} = 0.81 \]
\[ \hat{\eta} = 11,417 \text{ hr} \]

However, the MLE results for Case 1 are:

\[ \hat{\beta} = 1.33 \]
\[ \hat{\eta} = 6,900 \text{ hr} \]

And the MLE results for Case 2 are:

\[ \hat{\beta} = 0.9337 \]
\[ \hat{\eta} = 21,348 \text{ hr} \]

As we can see, there is a sizable difference in the results of the two sets calculated using MLE and the results using regression with the SRM. The results for both cases are identical when using the regression estimation technique with SRM, as SRM considers only the positions of the suspensions. The MLE results are quite different for the two cases, with the second case having a much larger value of \( \hat{\eta} \), which is due to the higher values of the suspension times in Case 2. This is because the maximum likelihood technique, unlike rank regression with SRM, considers the values of the suspensions when estimating the parameters. This is illustrated in the discussion of MLE given below.

One alternative to improve the regression method is to use the following ReliaSoft Ranking Method (RRM) to calculate the rank. RRM does consider the effect of the censoring time.
ReliaSoft's Ranking Method (RRM) for Interval Censored Data

When analyzing interval data, it is commonplace to assume that the actual failure time occurred at the midpoint of the interval. To be more conservative, you can use the starting point of the interval or you can use the end point of the interval to be most optimistic. Weibull++ allows you to employ ReliaSoft's ranking method (RRM) when analyzing interval data. Using an iterative process, this ranking method is an improvement over the standard ranking method (SRM).

When analyzing left or right censored data, RRM also considers the effect of the actual censoring time. Therefore, the resulted rank will be more accurate than the SRM where only the position not the exact time of censoring is used. For more details on this method see ReliaSoft's Ranking Method.

Maximum Likelihood Estimation (MLE)

From a statistical point of view, the method of maximum likelihood estimation method is, with some exceptions, considered to be the most robust of the parameter estimation techniques discussed here. The method presented in this section is for complete data (i.e., data consisting only of times-to-failure). The analysis for right censored (suspension) data, and for interval or left censored data, are then discussed in the following sections.

The basic idea behind MLE is to obtain the most likely values of the parameters, for a given distribution, that will best describe the data. As an example, consider the following data (-3, 0, 4) and assume that you are trying to estimate the mean of the data. Now, if you have to choose the most likely value for the mean from -5, 1 and 10, which one would you choose? In this case, the most likely value is 1 (given your limit on choices). Similarly, under MLE, one determines the most likely values for the parameters of the assumed distribution. It is mathematically formulated as follows.

If \( X \) is a continuous random variable with pdf:

\[
f(x; \theta_1, \theta_2, ..., \theta_k)
\]

where \( \theta_1, \theta_2, ..., \theta_k \) are unknown parameters which need to be estimated, with \( R \) independent observations, \( x_1, x_2, ..., x_R \) which correspond in the case of life data analysis to failure times. The likelihood function is given by:

\[
L(\theta_1, \theta_2, ..., \theta_k|x_1, x_2, ..., x_R) = L = \prod_{i=1}^{R} f(x_i; \theta_1, \theta_2, ..., \theta_k)
\]

\( i = 1, 2, ..., R \)

The logarithmic likelihood function is given by:

\[
\Lambda = \ln L = \sum_{i=1}^{R} \ln f(x_i; \theta_1, \theta_2, ..., \theta_k)
\]

The maximum likelihood estimators (or parameter values) of \( \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k \) are obtained by maximizing \( L \) or \( \Lambda \).

By maximizing \( \Lambda \) which is much easier to work with than \( L \), the maximum likelihood estimators (MLE) of \( \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k \) are the simultaneous solutions of \( k \) equations such that:

\[
\frac{\partial \Lambda}{\partial \theta_j} = 0, \quad j=1,2,...,k
\]

Even though it is common practice to plot the MLE solutions using median ranks (points are plotted according to median ranks and the line according to the MLE solutions), this is not completely representative. As can be seen from the equations above, the MLE method is independent of any kind of ranks. For this reason, the MLE solution often appears not to track the data on the probability plot. This is perfectly acceptable because the two methods are independent of each other, and in no way suggests that the solution is wrong.
MLE for Right Censored Data

When performing maximum likelihood analysis on data with suspended items, the likelihood function needs to be expanded to take into account the suspended items. The overall estimation technique does not change, but another term is added to the likelihood function to account for the suspended items. Beyond that, the method of solving for the parameter estimates remains the same. For example, consider a distribution where \( z \) is a continuous random variable with \( pdf \) and \( cdf \):

\[
\begin{align*}
    f(x; \theta_1, \theta_2, \ldots, \theta_k) \\
    F(x; \theta_1, \theta_2, \ldots, \theta_k)
\end{align*}
\]

where \( \theta_1, \theta_2, \ldots, \theta_k \) are the unknown parameters which need to be estimated from \( R \) observed failures at \( T_1, T_2, \ldots, T_R \) and \( M \) observed suspensions at \( S_1, S_2, \ldots, S_M \) then the likelihood function is formulated as follows:

\[
L(\theta_1, \ldots, \theta_k | T_1, \ldots, T_R, S_1, \ldots, S_M) = \prod_{i=1}^{R} f(T_i; \theta_1, \theta_2, \ldots, \theta_k) \\
\cdot \prod_{j=1}^{M} [1 - F(S_j; \theta_1, \theta_2, \ldots, \theta_k)]
\]

The parameters are solved by maximizing this equation. In most cases, no closed-form solution exists for this maximum or for the parameters. Solutions specific to each distribution utilizing MLE are presented in Appendix D.

MLE for Interval and Left Censored Data

The inclusion of left and interval censored data in an MLE solution for parameter estimates involves adding a term to the likelihood equation to account for the data types in question. When using interval data, it is assumed that the failures occurred in an interval; i.e., in the interval from time \( A \) to time \( B \) (or from time 0 to time \( B \) if left censored), where \( A < B \). In the case of interval data, and given \( P \) interval observations, the likelihood function is modified by multiplying the likelihood function with an additional term as follows:

\[
L(\theta_1, \theta_2, \ldots, \theta_k | x_1, x_2, \ldots, x_P) = \prod_{i=1}^{P} \left\{ F(x_i; \theta_1, \theta_2, \ldots, \theta_k) - F(x_{i-1}; \theta_1, \theta_2, \ldots, \theta_k) \right\}
\]

Note that if only interval data are present, this term will represent the entire likelihood function for the MLE solution. The next section gives a formulation of the complete likelihood function for all possible censoring schemes.

The Complete Likelihood Function

We have now seen that obtaining MLE parameter estimates for different types of data involves incorporating different terms in the likelihood function to account for complete data, right censored data, and left, interval censored data. After including the terms for the different types of data, the likelihood function can now be expressed in its complete form or:

\[
L = \prod_{i=1}^{R} f(T_i; \theta_1, \ldots, \theta_k) \cdot \prod_{j=1}^{M} [1 - F(S_j; \theta_1, \ldots, \theta_k)] \\
\cdot \prod_{i=1}^{P} \left\{ F(I_{i1}; \theta_1, \ldots, \theta_k) - F(I_{i2}; \theta_1, \ldots, \theta_k) \right\}
\]

where:

\[
L \rightarrow L(\theta_1, \ldots, \theta_k | T_1, \ldots, T_R, S_1, \ldots, S_M, I_1, \ldots I_P)
\]

and:

- \( R \) is the number of units with exact failures
Parameter Estimation

- $M$ is the number of suspended units
- $P$ is the number of units with left censored or interval times-to-failure
- $\theta$ are the parameters of the distribution
- $T_i$ is the $i^{th}$ time to failure
- $S_j$ is the $j^{th}$ time of suspension
- $I_{lj}$ is the ending of the time interval of the $l^{th}$ group
- $I_{lj}$ is the beginning of the time interval of the $l^{th}$ group

The total number of units is $N = R + M + P$. It should be noted that in this formulation, if either $R$, $M$ or $P$ is zero then the product term associated with them is assumed to be one and not zero.

Comments on the MLE Method

The MLE method has many large sample properties that make it attractive for use. It is asymptotically consistent, which means that as the sample size gets larger, the estimates converge to the right values. It is asymptotically efficient, which means that for large samples, it produces the most precise estimates. It is asymptotically unbiased, which means that for large samples, one expects to get the right value on average. The distribution of the estimates themselves is normal, if the sample is large enough, and this is the basis for the usual Fisher Matrix Confidence Bounds discussed later. These are all excellent large sample properties.

Unfortunately, the size of the sample necessary to achieve these properties can be quite large: thirty to fifty to more than a hundred exact failure times, depending on the application. With fewer points, the methods can be badly biased. It is known, for example, that MLE estimates of the shape parameter for the Weibull distribution are badly biased for small sample sizes, and the effect can be increased depending on the amount of censoring. This bias can cause major discrepancies in analysis. There are also pathological situations when the asymptotic properties of the MLE do not apply. One of these is estimating the location parameter for the three-parameter Weibull distribution when the shape parameter has a value close to 1. These problems, too, can cause major discrepancies.

However, MLE can handle suspensions and interval data better than rank regression, particularly when dealing with a heavily censored data set with few exact failure times or when the censoring times are unevenly distributed. It can also provide estimates with one or no observed failures, which rank regression cannot do. As a rule of thumb, our recommendation is to use rank regression techniques when the sample sizes are small and without heavy censoring (censoring is discussed in Life Data Classifications). When heavy or uneven censoring is present, when a high proportion of interval data is present and/or when the sample size is sufficient, MLE should be preferred.

See also:
- Maximum Likelihood Parameter Estimation Example
- Grouped Data Analysis

Bayesian Parameter Estimation Methods

Up to this point, we have dealt exclusively with what is commonly referred to as classical statistics. In this section, another school of thought in statistical analysis will be introduced, namely Bayesian statistics. The premise of Bayesian statistics (within the context of life data analysis) is to incorporate prior knowledge, along with a given set of current observations, in order to make statistical inferences. The prior information could come from operational or observational data, from previous comparable experiments or from engineering knowledge. This type of analysis can be particularly useful when there is limited test data for a given design or failure mode but there is a strong prior understanding of the failure rate behavior for that design or mode. By incorporating prior information about the parameter(s), a posterior distribution for the parameter(s) can be obtained and inferences on the model parameters and their functions can be made. This section is intended to give a quick and elementary overview of Bayesian methods, focused primarily on the material necessary for understanding the Bayesian analysis methods available in
Parameter Estimation

Weibull++. Extensive coverage of the subject can be found in numerous books dealing with Bayesian statistics.

Bayes’s Rule

Bayes’s rule provides the framework for combining prior information with sample data. In this reference, we apply Bayes’s rule for combining prior information on the assumed distribution’s parameter(s) with sample data in order to make inferences based on the model. The prior knowledge about the parameter(s) is expressed in terms of a $\phi(\theta)$, called the prior distribution. The posterior distribution of $\theta$ given the sample data, using Bayes’s rule, provides the updated information about the parameters. This is expressed with the following posterior pdf:

$$f(\theta|Data) = \frac{L(Data|\theta)\phi(\theta)}{\int_\zeta L(Data|\theta)\phi(\theta)d(\theta)}$$

where:

- $\theta$ is a vector of the parameters of the chosen distribution
- $\zeta$ is the range of $\theta$
- $L(Data|\theta)$ is the likelihood function based on the chosen distribution and data
- $\phi(\theta)$ is the prior distribution for each of the parameters

The integral in the Bayes’s rule equation is often referred to as the marginal probability, which is a constant number that can be interpreted as the probability of obtaining the sample data given a prior distribution. Generally, the integral in the Bayes’s rule equation does not have a closed form solution and numerical methods are needed for its solution.

As can be seen from the Bayes’s rule equation, there is a significant difference between classical and Bayesian statistics. First, the idea of prior information does not exist in classical statistics. All inferences in classical statistics are based on the sample data. On the other hand, in the Bayesian framework, prior information constitutes the basis of the theory. Another difference is in the overall approach of making inferences and their interpretation. For example, in Bayesian analysis, the parameters of the distribution to be fitted are the random variables. In reality, there is no distribution fitted to the data in the Bayesian case.

For instance, consider the case where data is obtained from a reliability test. Based on prior experience on a similar product, the analyst believes that the shape parameter of the Weibull distribution has a value between $\beta_1$ and $\beta_2$ and wants to utilize this information. This can be achieved by using the Bayes theorem. At this point, the analyst is automatically forcing the Weibull distribution as a model for the data and with a shape parameter between $\beta_1$ and $\beta_2$. In this example, the range of values for the shape parameter is the prior distribution, which in this case is Uniform. By applying Bayes’s rule, the posterior distribution of the shape parameter will be obtained. Thus, we end up with a distribution for the parameter rather than an estimate of the parameter, as in classical statistics.

To better illustrate the example, assume that a set of failure data was provided along with a distribution for the shape parameter (i.e., uniform prior) of the Weibull (automatically assuming that the data are Weibull distributed). Based on that, a new distribution (the posterior) for that parameter is then obtained using Bayes’s rule. This posterior distribution of the parameter may or may not resemble in form the assumed prior distribution. In other words, in this example the prior distribution of $\beta$ was assumed to be uniform but the posterior is most likely not a uniform distribution.

The question now becomes: what is the value of the shape parameter? What about the reliability and other results of interest? In order to answer these questions, we have to remember that in the Bayesian framework all of these metrics are random variables. Therefore, in order to obtain an estimate, a probability needs to be specified or we can use the expected value of the posterior distribution.

In order to demonstrate the procedure of obtaining results from the posterior distribution, we will rewrite the Bayes’s rule equation for a single parameter $\theta_1$: 
Parameter Estimation

\[ f(\theta|\text{Data}) = \frac{L(\text{Data}|\theta_1)\varphi(\theta_1)}{\int L(\text{Data}|\theta_1)\varphi(\theta_1)d\theta} \]

The expected value (or mean value) of the parameter \( \theta \) can be obtained using the equation for the mean and the Bayes's rule equation for single parameter:

\[ E(\theta_1) = m_{\theta_1} = \int \theta_1 \cdot f(\theta_1|\text{Data})d\theta_1 \]

An alternative result for \( \theta \) would be the median value. Using the equation for the median and the Bayes's rule equation for a single parameter:

\[ \int_{-\infty}^{\theta_{0.5}} f(\theta_1|\text{Data})d\theta_1 = 0.5 \]

The equation for the median is solved for the median value of \( \theta_1 \)

Similarly, any other percentile of the posterior pdf can be calculated and reported. For example, one could calculate the 90th percentile of \( \theta_1 \)’s posterior pdf:

\[ \int_{-\infty}^{\theta_{0.9}} f(\theta_1|\text{Data})d\theta_1 = 0.9 \]

This calculation will be used in Confidence Bounds and The Weibull Distribution for obtaining confidence bounds on the parameter(s).

The next step will be to make inferences on the reliability. Since the parameter \( \theta_1 \) is a random variable described by the posterior pdf, all subsequent functions of \( \theta_1 \) are distributed random variables as well and are entirely based on the posterior pdf of \( \theta_1 \). Therefore, expected value, median or other percentile values will also need to be calculated. For example, the expected reliability at time \( T \) is:

\[ E[R(T|\text{Data})] = \int R(T)f(\theta|\text{Data})d\theta \]

In other words, at a given time \( T \), there is a distribution that governs the reliability value at that time, \( T \), and by using Bayes's rule, the expected (or mean) value of the reliability is obtained. Other percentiles of this distribution can also be obtained. A similar procedure is followed for other functions of \( \theta_1 \) such as failure rate, reliable life, etc.

Prior Distributions

Prior distributions play a very important role in Bayesian Statistics. They are essentially the basis in Bayesian analysis. Different types of prior distributions exist, namely informative and non-informative. Non-informative prior distributions (a.k.a. vague, flat and diffuse) are distributions that have no population basis and play a minimal role in the posterior distribution. The idea behind the use of non-informative prior distributions is to make inferences that are not greatly affected by external information or when external information is not available. The uniform distribution is frequently used as a non-informative prior.

On the other hand, informative priors have a stronger influence on the posterior distribution. The influence of the prior distribution on the posterior is related to the sample size of the data and the form of the prior. Generally speaking, large sample sizes are required to modify strong priors, where weak priors are overwhelmed by even relatively small sample sizes. Informative priors are typically obtained from past data.

References

Life Data Classification

Statistical models rely extensively on data to make predictions. In life data analysis, the models are the statistical distributions and the data are the life data or times-to-failure data of our product. The accuracy of any prediction is directly proportional to the quality, accuracy and completeness of the supplied data. Good data, along with the appropriate model choice, usually results in good predictions. Bad or insufficient data will almost always result in bad predictions.

In the analysis of life data, we want to use all available data sets, which sometimes are incomplete or include uncertainty as to when a failure occurred. Life data can therefore be separated into two types: complete data (all information is available) or censored data (some of the information is missing). Each type is explained next.

Complete Data

Complete data means that the value of each sample unit is observed or known. For example, if we had to compute the average test score for a sample of ten students, complete data would consist of the known score for each student. Likewise in the case of life data analysis, our data set (if complete) would be composed of the times-to-failure of all units in our sample. For example, if we tested five units and they all failed (and their times-to-failure were recorded), we would then have complete information as to the time of each failure in the sample.
Censored Data

In many cases, all of the units in the sample may not have failed (i.e., the event of interest was not observed) or the exact times-to-failure of all the units are not known. This type of data is commonly called censored data. There are three types of possible censoring schemes, right censored (also called suspended data), interval censored and left censored.

Right Censored (Suspension) Data

The most common case of censoring is what is referred to as right censored data, or suspended data. In the case of life data, these data sets are composed of units that did not fail. For example, if we tested five units and only three had failed by the end of the test, we would have right censored data (or suspension data) for the two units that did not fail. The term right censored implies that the event of interest (i.e., the time-to-failure) is to the right of our data point. In other words, if the units were to keep on operating, the failure would occur at some time after our data point (or to the right on the time scale).

Interval Censored Data

The second type of censoring is commonly called interval censored data. Interval censored data reflects uncertainty as to the exact times the units failed within an interval. This type of data frequently comes from tests or situations where the objects of interest are not constantly monitored. For example, if we are running a test on five units and inspecting them every 100 hours, we only know that a unit failed or did not fail between inspections. Specifically, if we inspect a certain unit at 100 hours and find it operating, and then perform another inspection at 200 hours to find that the unit is no longer operating, then the only information we have is that the unit failed at some point in the interval between 100 and 200 hours. This type of censored data is also called inspection data by some authors.
It is generally recommended to avoid interval censored data because they are less informative compared to complete data. However, there are cases when interval data are unavoidable due to the nature of the product, the test and the test equipment. In those cases, caution must be taken to set the inspection intervals to be short enough to observe the spread of the failures. For example, if the inspection interval is too long, all the units in the test may fail within that interval, and thus no failure distribution could be obtained.

**Left Censored Data**

The third type of censoring is similar to the interval censoring and is called *left censored data*. In left censored data, a failure time is only known to be before a certain time. For instance, we may know that a certain unit failed sometime before 100 hours but not exactly when. In other words, it could have failed any time between 0 and 100 hours. This is identical to *interval censored data* in which the starting time for the interval is zero.
**Grouped Data Analysis**

In the standard folio, data can be entered individually or in groups. Grouped data analysis is used for tests in which groups of units possess the same time-to-failure or in which groups of units were suspended at the same time. We highly recommend entering redundant data in groups. Grouped data speeds data entry by the user and significantly speeds up the calculations.

**A Note about Complete and Suspension Data**

Depending on the event that we want to measure, data type classification (i.e., complete or suspension) can be open to interpretation. For example, under certain circumstances, and depending on the question one wishes to answer, a specimen that has failed might be classified as a suspension for analysis purposes. To illustrate this, consider the following times-to-failure data for a product that can fail due to modes A, B and C:

<table>
<thead>
<tr>
<th>Time-to-Failure, hr</th>
<th>Mode of Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>A</td>
</tr>
<tr>
<td>125</td>
<td>B</td>
</tr>
<tr>
<td>134</td>
<td>A</td>
</tr>
<tr>
<td>167</td>
<td>C</td>
</tr>
<tr>
<td>212</td>
<td>C</td>
</tr>
<tr>
<td>345</td>
<td>A</td>
</tr>
<tr>
<td>457</td>
<td>B</td>
</tr>
<tr>
<td>541</td>
<td>C</td>
</tr>
<tr>
<td>623</td>
<td>B</td>
</tr>
</tbody>
</table>

If the objective of the analysis is to determine the probability of failure of the product, regardless of the mode responsible for the failure, we would analyze the data with all data entries classified as failures (complete data). However, if the objective of the analysis is to determine the probability of failure of the product due to Mode A only, we would then choose to treat failures due to Modes B or C as suspension (right censored) data. Those data points would be treated as suspension data with respect to Mode A because the product operated until the recorded time without failure due to Mode A.

**Fractional Failures**

After the completion of a reliability test or after failures are observed in the field, a redesign can be implemented to improve a product’s reliability. After the redesign, and before new failure data become available, it is often times desirable to “adjust” the reliability that was calculated from the previous design and take “credit” for this theoretical improvement. This can be achieved with fractional failures. Using past experience to estimate the effectiveness of a corrective action or redesign, an analysis can take credit for this improvement by adjusting the failure count. Therefore, if a corrective action on a failure mode is believed to be 70% effective, then the failure count can be reduced from 1 to 0.3 to reflect the effectiveness of the corrective action.

For example, consider the following data set.
In this case, a design change has been implemented for the failure mode that occurred at 168 hours and is assumed to be 60% effective. In the background, Weibull++ converts this data set to:

<table>
<thead>
<tr>
<th>NumberinState</th>
<th>StateForS</th>
<th>StateEndTime(Hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>105</td>
</tr>
<tr>
<td>0.4</td>
<td>F</td>
<td>168</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>220</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>290</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>410</td>
</tr>
</tbody>
</table>

If Rank Regression is used to estimate distribution parameters, the median ranks for the previous data set are calculated as follows:

<table>
<thead>
<tr>
<th>NumberinState</th>
<th>StateForS</th>
<th>StateEndTime(Hr)</th>
<th>MON</th>
<th>MedianRank(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>105</td>
<td>1</td>
<td>12.945</td>
</tr>
<tr>
<td>0.4</td>
<td>F</td>
<td>168</td>
<td>1.4</td>
<td>20.267</td>
</tr>
<tr>
<td>0.6</td>
<td>S</td>
<td>168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>220</td>
<td>2.55</td>
<td>41.616</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>290</td>
<td>3.7</td>
<td>63.039</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>410</td>
<td>4.85</td>
<td>84.325</td>
</tr>
</tbody>
</table>

Given this information, the standard Rank Regression procedure is then followed to estimate parameters.

If Maximum Likelihood Estimation (MLE) is used to estimate distribution parameters, the grouped data likelihood function is used with the number in group being a non-integer value.
Example

A component underwent a reliability test. 12 samples were run to failure. The following figure shows the failures and the analysis in a Weibull++ standard folio.

The analysts believe that the planned design improvements will yield 50% effectiveness. To estimate the reliability of the product based on the assumptions about the repair effectiveness, they enter the data in groups, counting a 0.5 failure for each group. The following figure shows the adjusted data set and the calculated parameters.

The following overlay plot of unreliability vs. time shows that by using fractional failures the estimated unreliability of the component has decreased, while the B10 life has increased from 2,566 hours to 3,564 hours.
Chapter 6

Confidence Bounds

What Are Confidence Bounds?

One of the most confusing concepts to a novice reliability engineer is estimating the precision of an estimate. This is an important concept in the field of reliability engineering, leading to the use of confidence intervals (or bounds). In this section, we will try to briefly present the concept in relatively simple terms but based on solid common sense.

The Black and White Marbles

To illustrate, consider the case where there are millions of perfectly mixed black and white marbles in a rather large swimming pool and our job is to estimate the percentage of black marbles. The only way to be absolutely certain about the exact percentage of marbles in the pool is to accurately count every last marble and calculate the percentage. However, this is too time- and resource-intensive to be a viable option, so we need to come up with a way of estimating the percentage of black marbles in the pool. In order to do this, we would take a relatively small sample of marbles from the pool and then count how many black marbles are in the sample.

Taking a Small Sample of Marbles

First, pick out a small sample of marbles and count the black ones. Say you picked out ten marbles and counted four black marbles. Based on this, your estimate would be that 40% of the marbles are black.

If you put the ten marbles back in the pool and repeat this example again, you might get six black marbles, changing your estimate to 60% black marbles. Which of the two is correct? Both estimates are correct! As you repeat this experiment over and over again, you might find out that this estimate is usually between $X_1\%$ and $X_2\%$, and you can assign a percentage to the number of times your estimate falls between these limits. For example, you notice that 90% of the time this estimate is between $X_1\%$ and $X_2\%$.

Taking a Larger Sample of Marbles

If you now repeat the experiment and pick out 1,000 marbles, you might get results for the number of black marbles such as 545, 570, 530, etc., for each trial. The range of the estimates in this case will be much narrower than before. For example, you observe that 90% of the time, the number of black marbles will now be from $Y_1\%$ to $Y_2\%$, where $X_1\% < Y_1\%$ and $X_2\% > Y_2\%$, thus giving you a more narrow estimate interval. The same principle is true for confidence intervals; the larger the sample size, the more narrow the confidence intervals.

Back to Reliability
We will now look at how this phenomenon relates to reliability. Overall, the reliability engineer’s task is to determine the probability of failure, or reliability, of the population of units in question. However, one will never know the exact reliability value of the population unless one is able to obtain and analyze the failure data for every single unit in the population. Since this usually is not a realistic situation, the task then is to estimate the reliability based on a sample, much like estimating the number of black marbles in the pool. If we perform ten different reliability tests for our units, and analyze the results, we will obtain slightly different parameters for the distribution each time, and thus slightly different reliability results. However, by employing confidence bounds, we obtain a range within which these reliability values are likely to occur a certain percentage of the time. This helps us gauge the utility of the data and the accuracy of the resulting estimates. Plus, it is always useful to remember that each parameter is an estimate of the true parameter, one that is unknown to us. This range of plausible values is called a confidence interval.

**One-Sided and Two-Sided Confidence Bounds**

Confidence bounds are generally described as being one-sided or two-sided.

**Two-Sided Bounds**

![Two-sided confidence bounds diagram](image)

When we use two-sided confidence bounds (or intervals), we are looking at a closed interval where a certain percentage of the population is likely to lie. That is, we determine the values, or bounds, between which lies a specified percentage of the population. For example, when dealing with 90% two-sided confidence bounds of \((X, Y)\), we are saying that 90% of the population lies between \(X\) and \(Y\) with 5% less than \(X\) and 5% greater than \(Y\).

**One-Sided Bounds**

One-sided confidence bounds are essentially an open-ended version of two-sided bounds. A one-sided bound defines the point where a certain percentage of the population is either higher or lower than the defined point. This means that there are two types of one-sided bounds: upper and lower. An upper one-sided bound defines a point that a certain percentage of the population is less than. Conversely, a lower one-sided bound defines a point that a specified percentage of the population is greater than.
For example, if $\overline{X}$ is a 95% upper one-sided bound, this would imply that 95% of the population is less than $\overline{X}$. If $\overline{X}$ is a 95% lower one-sided bound, this would indicate that 95% of the population is greater than $\overline{X}$. Care must be taken to differentiate between one- and two-sided confidence bounds, as these bounds can take on identical values at different percentage levels. For example, in the figures above, we see bounds on a hypothetical distribution. Assuming that this is the same distribution in all of the figures, we see that $\overline{X}$ marks the spot below which 5% of the distribution's population lies. Similarly, $\overline{Y}$ represents the point above which 5% of the population lies. Therefore, $\overline{X}$ and $\overline{Y}$ represent the 90% two-sided bounds, since 90% of the population lies between the two points. However, $\overline{X}$ also represents the lower one-sided 95% confidence bound, since 95% of the population lies above that point; and $\overline{Y}$ represents the upper one-sided 95% confidence bound, since 95% of the population is below $\overline{Y}$. It is important to be sure of the type of bounds you are dealing with, particularly as both one-sided bounds can be displayed simultaneously in Weibull++. In Weibull++, we use upper to represent the higher limit and lower to represent the lower limit, regardless of their position, but based on the value of the results. So if obtaining the confidence bounds on the reliability, we would identify the lower value of reliability as the lower limit and the higher value of reliability as the higher limit. If obtaining the confidence bounds on probability of failure we will again identify the lower numeric value for the probability of failure as the lower limit and the higher value as the higher limit.

**Fisher Matrix Confidence Bounds**

This section presents an overview of the theory on obtaining approximate confidence bounds on suspended (multiple censored) data. The methodology used is the so-called Fisher matrix bounds (FM), described in Nelson [30] and Lloyd and Lipow [24]. These bounds are employed in most other commercial statistical applications. In general, these bounds tend to be more optimistic than the non-parametric rank based bounds. This may be a concern, particularly when dealing with small sample sizes. Some statisticians feel that the Fisher matrix bounds are too optimistic when dealing with small sample sizes and prefer to use other techniques for calculating confidence.
bounds, such as the likelihood ratio bounds.

**Approximate Estimates of the Mean and Variance of a Function**

In utilizing FM bounds for functions, one must first determine the mean and variance of the function in question (i.e., reliability function, failure rate function, etc.). An example of the methodology and assumptions for an arbitrary function $G$ as presented next.

**Single Parameter Case**

For simplicity, consider a one-parameter distribution represented by a general function $G$, which is a function of one parameter estimator, say $\hat{\theta}$. For example, the mean of the exponential distribution is a function of the parameter $\lambda : G(\lambda) = 1/\lambda = \mu$. Then, in general, the expected value of $G(\hat{\theta})$ can be found by:

$$E \left( G(\hat{\theta}) \right) = G(\theta) + O \left( \frac{1}{n} \right)$$

where $G(\hat{\theta})$ is some function of $\hat{\theta}$, such as the reliability function, and $\theta$ is the population parameter where $E(\hat{\theta}) = \theta$ as $n \to \infty$. The term $O \left( \frac{1}{n} \right)$ is a function of $n$, the sample size, and tends to zero, as fast as $n^{-1}$ as $n \to \infty$. For example, in the case of $\hat{\theta} = 1/\overline{x}$ and $G(x) = 1/x$, then $E(G(\hat{\theta})) = \overline{x} + O \left( \frac{1}{n} \right)$ where $O \left( \frac{1}{n} \right) = \sigma^2/n$. Thus as $n \to \infty$, $E(G(\hat{\theta})) = \mu$ where $\mu$ and $\sigma$ are the mean and standard deviation, respectively. Using the same one-parameter distribution, the variance of the function $G(\hat{\theta})$ can then be estimated by:

$$\text{Var} \left( G(\hat{\theta}) \right) = \left( \frac{\partial G}{\partial \theta} \right)_{\hat{\theta} = \theta}^2 + O \left( \frac{1}{n} \right)$$

**Two-Parameter Case**

Consider a Weibull distribution with two parameters $\beta$ and $\eta$. For a given value of $t$, $R(t) = G(\beta, \eta) = e^{-\left(\frac{t}{\eta}\right)^{\beta}}$. Repeating the previous method for the case of a two-parameter distribution, it is generally true that for a function $G$, which is a function of two parameter estimators, say $G(\hat{\theta}_1, \hat{\theta}_2)$, that:

$$E \left( G(\hat{\theta}_1, \hat{\theta}_2) \right) = G(\theta_1, \theta_2) + O \left( \frac{1}{n} \right)$$

and:

$$\text{Var} \left( G(\hat{\theta}_1, \hat{\theta}_2) \right) = \left( \frac{\partial G}{\partial \theta_1} \right)_{\hat{\theta}_1 = \theta_1}^2 \text{Var}(\hat{\theta}_1) + \left( \frac{\partial G}{\partial \theta_2} \right)_{\hat{\theta}_2 = \theta_2}^2 \text{Var}(\hat{\theta}_2)$$

$$+ 2 \left( \frac{\partial G}{\partial \theta_1} \right)_{\hat{\theta}_1 = \theta_1} \left( \frac{\partial G}{\partial \theta_2} \right)_{\hat{\theta}_2 = \theta_2} \text{Cov}(\hat{\theta}_1, \hat{\theta}_2) + O \left( \frac{1}{n^2} \right)$$

Note that the derivatives of the above equation are evaluated at $\hat{\theta}_1 = \theta_1$ and $\hat{\theta}_2 = \theta_2$, where $E(\hat{\theta}_1) \simeq \theta_1$ and $E(\hat{\theta}_2) \simeq \theta_2$.

**Parameter Variance and Covariance Determination**

The determination of the variance and covariance of the parameters is accomplished via the use of the Fisher information matrix. For a two-parameter distribution, and using maximum likelihood estimates (MLE), the log-likelihood function for censored data is given by:
\[ \ln[L] = \Lambda = \sum_{i=1}^{R} \ln[f(T_i; \theta_1, \theta_2)] + \sum_{j=1}^{M} \ln[1 - F(S_j; \theta_1, \theta_2)] + \sum_{l=1}^{P} \ln \{F(I_{l_0}; \theta_1, \theta_2) - F(I_{l_1}; \theta_1, \theta_2)\} \]

In the equation above, the first summation is for complete data, the second summation is for right censored data and the third summation is for interval or left censored data.

Then the Fisher information matrix is given by:

\[ F_0 = \begin{bmatrix}
E_0 \left[ -\frac{\partial^2 \Lambda}{\partial \theta_1^2} \right] & E_0 \left[ -\frac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_2} \right] \\
E_0 \left[ -\frac{\partial^2 \Lambda}{\partial \theta_2 \partial \theta_1} \right] & E_0 \left[ -\frac{\partial^2 \Lambda}{\partial \theta_2^2} \right]
\end{bmatrix}_{0} \]

The subscript 0 indicates that the quantity is evaluated at \( \theta_1 = \theta_{10} \) and \( \theta_2 = \theta_{20} \), the true values of the parameters.

So for a sample of \( N \) units where \( R \) units have failed, \( S \) have been suspended, and \( P \) have failed within a time interval, and \( N = R + M + P \), one could obtain the sample local information matrix by:

\[ F = \begin{bmatrix}
-\frac{\partial^2 \Lambda}{\partial \theta_1^2} & -\frac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_2} \\
-\frac{\partial^2 \Lambda}{\partial \theta_2 \partial \theta_1} & -\frac{\partial^2 \Lambda}{\partial \theta_2^2}
\end{bmatrix} \]

Substituting the values of the estimated parameters, in this case \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \), and then inverting the matrix, one can then obtain the local estimate of the covariance matrix or:

\[ \begin{bmatrix}
\widehat{Var}(\hat{\theta}_1) & \widehat{Cov}(\hat{\theta}_1, \hat{\theta}_2) \\
\widehat{Cov}(\hat{\theta}_1, \hat{\theta}_2) & \widehat{Var}(\hat{\theta}_2)
\end{bmatrix} = \begin{bmatrix}
-\frac{\partial^2 \Lambda}{\partial \theta_1^2} & -\frac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_2} \\
-\frac{\partial^2 \Lambda}{\partial \theta_2 \partial \theta_1} & -\frac{\partial^2 \Lambda}{\partial \theta_2^2}
\end{bmatrix}^{-1} \]

Then the variance of a function \( Var(G) \) can be estimated using equation for the variance. Values for the variance and covariance of the parameters are obtained from Fisher Matrix equation. Once they have been obtained, the approximate confidence bounds on the function are given as:

\[ CB_R = E(G) \pm z_\alpha \sqrt{Var(G)} \]

which is the estimated value plus or minus a certain number of standard deviations. We address finding \( z_\alpha \) next.

**Approximate Confidence Intervals on the Parameters**

In general, MLE estimates of the parameters are asymptotically normal, meaning that for large sample sizes, a distribution of parameter estimates from the same population would be very close to the normal distribution. Thus if \( \hat{\theta} \) is the MLE estimator for \( \theta \), in the case of a single parameter distribution estimated from a large sample of \( n \) units, then:

\[ z = \frac{\hat{\theta} - \theta}{\sqrt{Var(\hat{\theta})}} \]

follows an approximating normal distribution. That is

\[ P(x \leq z) \rightarrow \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt \]
for large \( n \). We now place confidence bounds on \( \theta \) at some confidence level \( \delta \), bounded by the two end points \( C_1 \) and \( C_2 \) where:

\[
P \left( C_1 < \theta < C_2 \right) = \delta
\]

From the above equation:

\[
P \left( \frac{-K_{1-\delta}}{2} < \frac{\hat{\theta} - \theta}{\sqrt{\text{Var}(\hat{\theta})}} < \frac{K_{1-\delta}}{2} \right) \simeq \delta
\]

where \( K_\alpha \) is defined by:

\[
\alpha = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 1 - \Phi(K_\alpha)
\]

Now by simplifying the equation for the confidence level, one can obtain the approximate two-sided confidence bounds on the parameter \( \theta \) at a confidence level \( \delta \) or:

\[
\left( \frac{\hat{\theta} - K_{1-\delta}}{2} \cdot \sqrt{\text{Var}(\hat{\theta})} < \theta < \frac{\hat{\theta} + K_{1-\delta}}{2} \cdot \sqrt{\text{Var}(\hat{\theta})} \right)
\]

The upper one-sided bounds are given by:

\[
\theta < \hat{\theta} + K_{1-\delta} \sqrt{\text{Var}(\hat{\theta})}
\]

while the lower one-sided bounds are given by:

\[
\theta > \hat{\theta} - K_{1-\delta} \sqrt{\text{Var}(\hat{\theta})}
\]

If \( \hat{\theta} \) must be positive, then \( \ln \hat{\theta} \) is treated as normally distributed. The two-sided approximate confidence bounds on the parameter \( \theta \) at confidence level \( \delta \), then become:

\[
\theta_U = \frac{\hat{\theta} \cdot e^{\frac{K_{1-\delta}}{2} \sqrt{\text{Var}(\hat{\theta})}}}{\hat{\theta}} \quad \text{(Two-sided upper)}
\]

\[
\theta_L = \frac{\hat{\theta} \cdot e^{-\frac{K_{1-\delta}}{2} \sqrt{\text{Var}(\hat{\theta})}}}{\hat{\theta}} \quad \text{(Two-sided lower)}
\]

The one-sided approximate confidence bounds on the parameter \( \theta \) at confidence level \( \delta \) can be found from:

\[
\theta_U = \frac{\hat{\theta} \cdot e^{\frac{K_{1-\delta}}{2} \sqrt{\text{Var}(\hat{\theta})}}}{\hat{\theta}} \quad \text{(One-sided upper)}
\]

\[
\theta_L = \frac{\hat{\theta} \cdot e^{-\frac{K_{1-\delta}}{2} \sqrt{\text{Var}(\hat{\theta})}}}{\hat{\theta}} \quad \text{(One-sided lower)}
\]

The same procedure can be extended for the case of a two or more parameter distribution. Lloyd and Lipow [24] further elaborate on this procedure.
Confidence Bounds on Time (Type 1)

Type 1 confidence bounds are confidence bounds around time for a given reliability. For example, when using the one-parameter exponential distribution, the corresponding time for a given exponential percentile (i.e., y-ordinate or unreliability, \( Q = 1 - R \)) is determined by solving the unreliability function for the time, \( T \), or:

\[
\hat{T}(Q) = -\frac{1}{\lambda} \ln(1 - Q) = -\frac{1}{\lambda} \ln(R)
\]

Bounds on time (Type 1) return the confidence bounds around this time value by determining the confidence intervals around \( \hat{\lambda} \) and substituting these values into the above equation. The bounds on \( \hat{\lambda} \) are determined using the method for the bounds on parameters, with its variance obtained from the Fisher Matrix. Note that the procedure is slightly more complicated for distributions with more than one parameter.

Confidence Bounds on Reliability (Type 2)

Type 2 confidence bounds are confidence bounds around reliability. For example, when using the two-parameter exponential distribution, the reliability function is:

\[
\hat{R}(T) = e^{-\lambda T}
\]

Reliability bounds (Type 2) return the confidence bounds by determining the confidence intervals around \( \hat{\lambda} \) and substituting these values into the above equation. The bounds on \( \hat{\lambda} \) are determined using the method for the bounds on parameters, with its variance obtained from the Fisher Matrix. Once again, the procedure is more complicated for distributions with more than one parameter.

Beta Binomial Confidence Bounds

Another less mathematically intensive method of calculating confidence bounds involves a procedure similar to that used in calculating median ranks (see Parameter Estimation). This is a non-parametric approach to confidence interval calculations that involves the use of rank tables and is commonly known as beta-binomial bounds (BB). By non-parametric, we mean that no underlying distribution is assumed. (Parametric implies that an underlying distribution, with parameters, is assumed.) In other words, this method can be used for any distribution, without having to make adjustments in the underlying equations based on the assumed distribution. Recall from the discussion on the median ranks that we used the binomial equation to compute the ranks at the 50% confidence level (or median ranks) by solving the cumulative binomial distribution for \( Z \) (rank for the \( j^{th} \) failure):

\[
P = \sum_{k=j}^{N} \binom{N}{k} Z^k (1 - Z)^{N-k}
\]

where \( N \) is the sample size and \( j \) is the order number.

The median rank was obtained by solving the following equation for \( Z \):

\[
0.50 = \sum_{k=j}^{N} \binom{N}{k} Z^k (1 - Z)^{N-k}
\]

The same methodology can then be repeated by changing \( P \) for 0.50 (50%) to our desired confidence level. For \( P = 0.90 \)%, one would formulate the equation as

\[
0.90 = \sum_{k=j}^{N} \binom{N}{k} Z^k (1 - Z)^{N-k}
\]

Keep in mind that one must be careful to select the appropriate values for \( P \) based on the type of confidence bounds desired. For example, if two-sided 80% confidence bounds are to be calculated, one must solve the equation twice (once with \( P = 0.1 \) and once with \( P = 0.9 \)) in order to place the bounds around 80% of the population.

Using this methodology, the appropriate ranks are obtained and plotted based on the desired confidence level. These
Confidence Bounds

points are then joined by a smooth curve to obtain the corresponding confidence bound.

In Weibull++, this non-parametric methodology is used only when plotting bounds on the mixed Weibull distribution. Full details on this methodology can be found in Kececioglu [20]. These binomial equations can again be transformed using the beta and F distributions, thus the name beta binomial confidence bounds.

Likelihood Ratio Confidence Bounds

Another method for calculating confidence bounds is the likelihood ratio bounds (LRB) method. Conceptually, this method is a great deal simpler than that of the Fisher matrix, although that does not mean that the results are of any less value. In fact, the LRB method is often preferred over the FM method in situations where there are smaller sample sizes.

Likelihood ratio confidence bounds are based on the following likelihood ratio equation:

$$-2 \cdot \ln \left( \frac{L(\hat{\theta}|k)}{L(\theta)} \right) \geq \chi_{\alpha,k}^2$$

where:

- $L(\hat{\theta}|k)$ is the likelihood function for the unknown parameter vector $\hat{\theta}$
- $L(\theta)$ is the likelihood function calculated at the estimated vector $\theta$
- $\chi_{\alpha,k}^2$ is the chi-squared statistic with probability $\alpha$ and $k$ degrees of freedom, where $k$ is the number of quantities jointly estimated

If $\delta$ is the confidence level, then $\alpha = \delta$ for two-sided bounds and $\alpha = (2\delta - 1)$ for one-sided. Recall from the Brief Statistical Background chapter that if $X$ is a continuous random variable with pdf:

$$f(x; \theta_1, \theta_2, ..., \theta_k)$$

where $\theta_1, \theta_2, ..., \theta_k$ are unknown constant parameters that need to be estimated, one can conduct an experiment and obtain $R$ independent observations, $x_1, x_2, ..., x_R$, which correspond in the case of life data analysis to failure times. The likelihood function is given by:

$$L(x_1, x_2, ..., x_R|\theta_1, \theta_2, ..., \theta_k) = L = \prod_{i=1}^{R} f(x_i; \theta_1, \theta_2, ..., \theta_k)$$

The maximum likelihood estimators (MLE) of $\theta_1, \theta_2, ..., \theta_k$ are obtained by maximizing $L$. These are represented by the $L(\hat{\theta}|k)$ term in the denominator of the ratio in the likelihood ratio equation. Since the values of the data points are known, and the values of the parameter estimates $\hat{\theta}$ have been calculated using MLE methods, the only unknown term in the likelihood ratio equation is the $L(\theta)$ term in the numerator of the ratio. It remains to find the values of the unknown parameter vector $\theta$ that satisfy the likelihood ratio equation. For distributions that have two parameters, the values of these two parameters can be varied in order to satisfy the likelihood ratio equation. The values of the parameters that satisfy this equation will change based on the desired confidence level $\delta$ but at a given value of $\delta$ there is only a certain region of values for $\theta$ and $\theta_k$ for which the likelihood ratio equation holds true.

This region can be represented graphically as a contour plot, an example of which is given in the following graphic.
The region of the contour plot essentially represents a cross-section of the likelihood function surface that satisfies the conditions of the likelihood ratio equation.

**Note on Contour Plots in Weibull++ and ALTA**

Contour plots can be used for comparing data sets. Consider two data sets, one for an old product design and another for a new design. The engineer would like to determine if the two designs are significantly different and at what confidence. By plotting the contour plots of each data set in an overlay plot (the same distribution must be fitted to each data set), one can determine the confidence at which the two sets are significantly different. If, for example, there is no overlap (i.e., the two plots do not intersect) between the two 90% contours, then the two data sets are significantly different with a 90% confidence. If the two 95% contours overlap, then the two designs are NOT significantly different at the 95% confidence level. An example of non-intersecting contours is shown next. For details on comparing data sets, see Comparing Life Data Sets.
Confidence Bounds on the Parameters

The bounds on the parameters are calculated by finding the extreme values of the contour plot on each axis for a given confidence level. Since each axis represents the possible values of a given parameter, the boundaries of the contour plot represent the extreme values of the parameters that satisfy the following:

\[-2 \cdot \ln \left( \frac{L(\theta_1, \theta_2)}{L(\hat{\theta}_1, \hat{\theta}_2)} \right) = \chi^2_{\alpha,1} \]

This equation can be rewritten as:

\[L(\theta_1, \theta_2) = L(\hat{\theta}_1, \hat{\theta}_2) \cdot e^{-\frac{1}{2} \chi^2_{0.1}} \]

The task now is to find the values of the parameters \( \theta_1 \) and \( \theta_2 \) so that the equality in the likelihood ratio equation shown above is satisfied. Unfortunately, there is no closed-form solution; therefore, these values must be arrived at numerically. One way to do this is to hold one parameter constant and iterate on the other until an acceptable solution is reached. This can prove to be rather tricky, since there will be two solutions for one parameter if the other is held constant. In situations such as these, it is best to begin the iterative calculations with values close to those of the MLE values, so as to ensure that one is not attempting to perform calculations outside of the region of the contour plot where no solution exists.

Example 1: Likelihood Ratio Bounds on Parameters

Five units were put on a reliability test and experienced failures at 10, 20, 30, 40 and 50 hours. Assuming a Weibull distribution, the MLE parameter estimates are calculated to be \( \hat{\beta} = 2.2938 \) and \( \hat{\gamma} = 33.9428 \). Calculate the 90% two-sided confidence bounds on these parameters using the likelihood ratio method.

Solution

The first step is to calculate the likelihood function for the parameter estimates:
Confidence Bounds

![Image](image1)

Confidence Bounds

\[ L(\hat{\beta}, \hat{\eta}) = \prod_{i=1}^{N} f(x_i; \hat{\beta}, \hat{\eta}) = \prod_{i=1}^{5} \frac{2.2938}{33.9428} \cdot \left(\frac{x_i}{33.9428}\right)^{1.2938} \cdot e^{-\frac{x_i}{33.9428}} \]

\[ L(\hat{\beta}, \hat{\eta}) = 1.714714 \times 10^{-9} \]

where \( x_i \) are the original time-to-failure data points. We can now rearrange the likelihood ratio equation to the form:

\[ L(\beta, \eta) - L(\hat{\beta}, \hat{\eta}) \cdot e^{-\frac{\chi^2_{0.1}}{2}} = 0 \]

Since our specified confidence level, \( \delta \), is 90%, we can calculate the value of the chi-squared statistic, \( \chi^2_{0.9;1} = 2.705543 \). We then substitute this information into the equation:

\[ L(\beta, \eta) - L(\hat{\beta}, \hat{\eta}) \cdot e^{-\frac{2.705543}{2}} = 0 \]

\[ L(\beta, \eta) - 1.714714 \times 10^{-9} \cdot e^{-2.705543} = 0 \]

or:

\[ L(\beta, \eta) = 4.432926 \times 10^{-10} = 0 \]

The next step is to find the set of values of \( \beta \) and \( \eta \) that satisfy this equation, or find the values of \( \beta \) and \( \eta \) such that

\[ L(\beta, \eta) = 4.432926 \times 10^{-10} \]

The solution is an iterative process that requires setting the value of \( \beta \) and finding the appropriate values of \( \eta \), and vice versa. The following table gives values of \( \beta \) based on given values of \( \eta \).

<table>
<thead>
<tr>
<th>( \eta )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \eta )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>1.321</td>
<td>1.742</td>
<td>37</td>
<td>1.183</td>
<td>3.950</td>
</tr>
<tr>
<td>24</td>
<td>1.241</td>
<td>1.985</td>
<td>38</td>
<td>1.195</td>
<td>3.943</td>
</tr>
<tr>
<td>25</td>
<td>1.201</td>
<td>2.193</td>
<td>39</td>
<td>1.210</td>
<td>3.897</td>
</tr>
<tr>
<td>26</td>
<td>1.177</td>
<td>2.390</td>
<td>40</td>
<td>1.226</td>
<td>3.814</td>
</tr>
<tr>
<td>27</td>
<td>1.161</td>
<td>2.582</td>
<td>41</td>
<td>1.245</td>
<td>3.701</td>
</tr>
<tr>
<td>28</td>
<td>1.151</td>
<td>2.770</td>
<td>42</td>
<td>1.267</td>
<td>3.564</td>
</tr>
<tr>
<td>29</td>
<td>1.145</td>
<td>2.955</td>
<td>43</td>
<td>1.291</td>
<td>3.409</td>
</tr>
<tr>
<td>30</td>
<td>1.142</td>
<td>3.135</td>
<td>44</td>
<td>1.319</td>
<td>3.241</td>
</tr>
<tr>
<td>31</td>
<td><strong>1.142</strong></td>
<td>3.308</td>
<td>45</td>
<td>1.352</td>
<td>3.065</td>
</tr>
<tr>
<td>32</td>
<td>1.145</td>
<td>3.471</td>
<td>46</td>
<td>1.391</td>
<td>2.883</td>
</tr>
<tr>
<td>33</td>
<td>1.149</td>
<td>3.619</td>
<td>47</td>
<td>1.439</td>
<td>2.695</td>
</tr>
<tr>
<td>34</td>
<td>1.155</td>
<td>3.746</td>
<td>48</td>
<td>1.502</td>
<td>2.495</td>
</tr>
<tr>
<td>35</td>
<td>1.162</td>
<td>3.848</td>
<td>49</td>
<td>1.594</td>
<td>2.272</td>
</tr>
<tr>
<td>36</td>
<td>1.172</td>
<td>3.917</td>
<td><strong>50</strong></td>
<td>1.871</td>
<td>1.871</td>
</tr>
</tbody>
</table>

These data are represented graphically in the following contour plot:
Confidence Bounds

Confidence Bounds on Time (Type 1)

The manner in which the bounds on the time estimate for a given reliability are calculated is much the same as the manner in which the bounds on the parameters are calculated. The difference lies in the form of the likelihood functions that comprise the likelihood ratio. In the preceding section, we used the standard form of the likelihood function, which was in terms of the parameters $\theta_1$ and $\theta_2$. In order to calculate the bounds on a time estimate, the likelihood function needs to be rewritten in terms of one parameter and time, so that the maximum and minimum values of the time can be observed as the parameter is varied. This process is best illustrated with an example.

Example 2: Likelihood Ratio Bounds on Time (Type I)

For the data given in Example 1, determine the 90% two-sided confidence bounds on the time estimate for a reliability of 50%. The ML estimate for the time at which $R(t) = 50\%$ is 28.930.

Solution

In this example, we are trying to determine the 90% two-sided confidence bounds on the time estimate of 28.930. As was mentioned, we need to rewrite the likelihood ratio equation so that it is in terms of $t$ and $\beta$. This is accomplished by using a form of the Weibull reliability equation,

$$R = e^{-\left(\frac{t}{\eta}\right)^\beta}$$

This can be rearranged in terms of $\eta$, with $R$
being considered a known variable or:

$$\eta = \frac{t}{(-\ln(R))^\frac{1}{\beta}}$$

This can then be substituted into the term in the likelihood ratio equation to form a likelihood equation in terms of \(t\) and \(\beta\) for:

$$L(\beta, t) = \prod_{i=1}^{N} f(x_i; \beta, t, R)$$

$$= \prod_{i=1}^{5} \left( \left( \frac{x_i}{(-\ln(R))^\frac{1}{\beta}} \right)^\beta \cdot \exp \left( - \left( \frac{x_i}{(-\ln(R))^\frac{1}{\beta}} \right) \right) \right)^{(\beta-1)}$$

where \(x\) are the original time-to-failure data points. We can now rearrange the likelihood ratio equation to the form:

$$L(\beta, t) - L(\hat{\beta}, \hat{\eta}) \cdot e^{-\frac{\chi^2_{0.90:1}}{2}} = 0$$

Since our specified confidence level, \(\delta\), is 90%, we can calculate the value of the chi-squared statistic, \(\chi^2_{0.90:1} = 2.705543\). We can now substitute this information into the equation:

$$L(\beta, t) - L(\hat{\beta}, \hat{\eta}) \cdot e^{-\frac{2.705543}{2}} = 0$$

$$L(\beta, t) - 1.714714 \times 10^{-9} \cdot e^{-\frac{2.705543}{2}} = 0$$

$$L(\beta, t) - 4.432926 \cdot 10^{-10} = 0$$

Note that the likelihood value for \(L(\hat{\beta}, \hat{\eta})\) is the same as it was for Example 1. This is because we are dealing with the same data and parameter estimates or, in other words, the maximum value of the likelihood function did not change. It now remains to find the values of \(\hat{\beta}\) and \(\hat{\eta}\) which satisfy this equation. This is an iterative process that requires setting the value of \(\hat{\beta}\) and finding the appropriate values of \(\hat{t}\). The following table gives the values of \(\hat{t}\) based on given values of \(\hat{\beta}\).
These points are represented graphically in the following contour plot:

![Weibull Contour Plot for Time vs. Beta](image)

As can be determined from the table, the lowest calculated value for $t$ is 17.389, while the highest is 41.714. These represent the 90% two-sided confidence limits on the time at which reliability is equal to 50%.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$\beta$</th>
<th>$t_1$</th>
<th>$t_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>18.448</td>
<td>28.244</td>
<td>2.6</td>
<td>23.533</td>
<td>41.241</td>
</tr>
<tr>
<td>1.3</td>
<td>17.489</td>
<td>32.694</td>
<td>2.7</td>
<td>24.119</td>
<td>41.009</td>
</tr>
<tr>
<td>1.4</td>
<td>17.389</td>
<td>35.568</td>
<td>2.8</td>
<td>24.705</td>
<td>40.747</td>
</tr>
<tr>
<td>1.5</td>
<td>17.602</td>
<td>37.576</td>
<td>2.9</td>
<td>25.292</td>
<td>40.459</td>
</tr>
<tr>
<td>1.6</td>
<td>17.971</td>
<td>39.005</td>
<td>3.0</td>
<td>25.883</td>
<td>40.146</td>
</tr>
<tr>
<td>1.7</td>
<td>18.428</td>
<td>40.019</td>
<td>3.1</td>
<td>26.478</td>
<td>39.809</td>
</tr>
<tr>
<td>1.8</td>
<td>18.937</td>
<td>40.726</td>
<td>3.2</td>
<td>27.082</td>
<td>39.450</td>
</tr>
<tr>
<td>1.9</td>
<td>19.478</td>
<td>41.201</td>
<td>3.3</td>
<td>27.696</td>
<td>39.066</td>
</tr>
<tr>
<td>2.0</td>
<td>20.039</td>
<td>41.499</td>
<td>3.4</td>
<td>28.329</td>
<td>38.654</td>
</tr>
<tr>
<td>2.1</td>
<td>20.612</td>
<td>41.660</td>
<td>3.5</td>
<td>28.987</td>
<td>38.205</td>
</tr>
<tr>
<td>2.2</td>
<td>21.192</td>
<td>41.714</td>
<td>3.6</td>
<td>29.684</td>
<td>37.709</td>
</tr>
<tr>
<td>2.3</td>
<td>21.776</td>
<td>41.684</td>
<td>3.7</td>
<td>30.444</td>
<td>37.142</td>
</tr>
<tr>
<td>2.4</td>
<td>22.361</td>
<td>41.587</td>
<td>3.8</td>
<td>31.321</td>
<td>36.452</td>
</tr>
<tr>
<td>2.5</td>
<td>22.947</td>
<td>41.436</td>
<td>3.9</td>
<td>32.496</td>
<td>35.457</td>
</tr>
</tbody>
</table>
Confidence Bounds on Reliability (Type 2)

The likelihood ratio bounds on a reliability estimate for a given time value are calculated in the same manner as were the bounds on time. The only difference is that the likelihood function must now be considered in terms of \( \hat{R} \) and \( \beta \).

The likelihood function is once again altered in the same way as before, only now \( \hat{R} \) is considered to be a parameter instead of \( \hat{t} \), since the value of \( \hat{t} \) must be specified in advance. Once again, this process is best illustrated with an example.

Example 3: Likelihood Ratio Bounds on Reliability (Type 2)

For the data given in Example 1, determine the 90% two-sided confidence bounds on the reliability estimate for \( \hat{t} = 45 \) The ML estimate for the reliability at \( \hat{t} = 45 \) is 14.816%.

Solution

In this example, we are trying to determine the 90% two-sided confidence bounds on the reliability estimate of 14.816%. As was mentioned, we need to rewrite the likelihood ratio equation so that it is in terms of \( \hat{R} \) and \( \beta \). This is again accomplished by substituting the Weibull reliability equation into the \( \eta \) term in the likelihood ratio equation to form a likelihood equation in terms of \( \hat{R} \) and \( \beta \):

\[
L(\beta, \hat{R}) = \prod_{i=1}^{N} f(x_i; \beta, \hat{t}, \hat{R})
\]

\[
= \prod_{i=1}^{5} \left( \frac{\beta}{(-\ln(\hat{R}))^\frac{1}{\beta}} \right) \cdot \left( \frac{x_i}{(-\ln(\hat{R}))^\frac{1}{\beta}} \right)^{\beta-1} \cdot \exp \left[ -\left( \frac{x_i}{(-\ln(\hat{R}))^\frac{1}{\beta}} \right) \right]^{\beta}
\]

where \( x_i \) are the original time-to-failure data points. We can now rearrange the likelihood ratio equation to the form:

\[
L(\beta, \hat{R}) - L(\hat{\beta}, \hat{\eta}) \cdot e^{-\frac{\chi^2_{0.1}}{2}} = 0
\]

Since our specified confidence level, \( \delta \), is 90%, we can calculate the value of the chi-squared statistic, \( \chi^2_{0.9,1} = 2.705543 \). We can now substitute this information into the equation:

\[
L(\beta, \hat{R}) - L(\hat{\beta}, \hat{\eta}) \cdot e^{-\frac{\chi^2_{0.1}}{2}} = 0
\]

\[
L(\beta, \hat{R}) = 1.714714 \times 10^{-9} \cdot e^{-\frac{2.705543}{2}} = 0
\]

\[
L(\beta, \hat{R}) = 4.432926 \times 10^{-10} = 0
\]

It now remains to find the values of \( \hat{\beta} \) and \( \hat{R} \) that satisfy this equation. This is an iterative process that requires setting the value of \( \hat{\beta} \) and finding the appropriate values of \( \hat{R} \). The following table gives the values of \( \hat{R} \) based on given values of \( \hat{\beta} \).
These points are represented graphically in the following contour plot:

As can be determined from the table, the lowest calculated value for $R_U$ is 2.38%, while the highest is 44.26%. These represent the 90% two-sided confidence limits on the reliability at $t = 45$. 

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$\beta$</th>
<th>$R_1$</th>
<th>$R_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>0.1325</td>
<td>0.2975</td>
<td>2.6</td>
<td>0.0238</td>
<td>0.4191</td>
</tr>
<tr>
<td>1.3</td>
<td>0.0936</td>
<td>0.3499</td>
<td>2.7</td>
<td>0.0239</td>
<td>0.4104</td>
</tr>
<tr>
<td>1.4</td>
<td>0.0725</td>
<td>0.3816</td>
<td>2.8</td>
<td>0.0243</td>
<td>0.4004</td>
</tr>
<tr>
<td>1.5</td>
<td>0.0588</td>
<td>0.4032</td>
<td>2.9</td>
<td>0.0251</td>
<td>0.3892</td>
</tr>
<tr>
<td>1.6</td>
<td>0.0493</td>
<td>0.4184</td>
<td>3.0</td>
<td>0.0262</td>
<td>0.3767</td>
</tr>
<tr>
<td>1.7</td>
<td>0.0423</td>
<td>0.4291</td>
<td>3.1</td>
<td>0.0277</td>
<td>0.3629</td>
</tr>
<tr>
<td>1.8</td>
<td>0.0372</td>
<td>0.4362</td>
<td>3.2</td>
<td>0.0296</td>
<td>0.3478</td>
</tr>
<tr>
<td>1.9</td>
<td>0.0333</td>
<td>0.4406</td>
<td>3.3</td>
<td>0.0321</td>
<td>0.3311</td>
</tr>
<tr>
<td>2.0</td>
<td>0.0303</td>
<td>0.4426</td>
<td>3.4</td>
<td>0.0353</td>
<td>0.3128</td>
</tr>
<tr>
<td>2.1</td>
<td>0.0281</td>
<td>0.4426</td>
<td>3.5</td>
<td>0.0395</td>
<td>0.2925</td>
</tr>
<tr>
<td>2.2</td>
<td>0.0264</td>
<td>0.4409</td>
<td>3.6</td>
<td>0.0451</td>
<td>0.2699</td>
</tr>
<tr>
<td>2.3</td>
<td>0.0252</td>
<td>0.4375</td>
<td>3.7</td>
<td>0.0527</td>
<td>0.2442</td>
</tr>
<tr>
<td>2.4</td>
<td>0.0244</td>
<td>0.4327</td>
<td>3.8</td>
<td>0.0641</td>
<td>0.2136</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0239</td>
<td>0.4266</td>
<td>3.9</td>
<td>0.0848</td>
<td>0.1727</td>
</tr>
</tbody>
</table>
**Bayesian Confidence Bounds**

A fourth method of estimating confidence bounds is based on the Bayes theorem. This type of confidence bounds relies on a different school of thought in statistical analysis, where prior information is combined with sample data in order to make inferences on model parameters and their functions. An introduction to Bayesian methods is given in the Parameter Estimation chapter. Bayesian confidence bounds are derived from Bayes's rule, which states that:

$$f(\hat{\theta}|\text{Data}) = \frac{L(Data|\theta)\varphi(\theta)}{\int \limits_{\theta} L(Data|\theta)\varphi(\theta)d\theta}$$

where:

- \( f(\hat{\theta}|\text{Data}) \) is the posterior pdf of \( \hat{\theta} \)
- \( \theta \) is the parameter vector of the chosen distribution (i.e., Weibull, lognormal, etc.)
- \( L(\bullet) \) is the likelihood function
- \( \varphi(\theta) \) is the prior pdf of the parameter vector \( \theta \)
- \( \Theta \) is the range of \( \theta \).

In other words, the prior knowledge is provided in the form of the prior pdf of the parameters, which in turn is combined with the sample data in order to obtain the posterior pdf. Different forms of prior information exist, such as past data, expert opinion or non-informative (refer to Parameter Estimation). It can be seen from the above Bayes's rule formula that we are now dealing with distributions of parameters rather than single value parameters. For example, consider a one-parameter distribution with a positive parameter \( \theta \). Given a set of sample data, and a prior distribution for \( \theta \), the above Bayes's rule formula can be written as:

$$f(\theta_1|\text{Data}) = \frac{L(Data|\theta_1)\varphi(\theta_1)}{\int \limits_{0}^{\infty} L(Data|\theta_1)\varphi(\theta_1)d\theta_1}$$

In other words, we now have the distribution of \( \theta \) and we can now make statistical inferences on this parameter, such as calculating probabilities. Specifically, the probability that \( \theta \) is less than or equal to a value \( x \), \( P(\theta_1 \leq x) \) can be obtained by integrating the posterior probability density function (pdf), or:

$$P(\theta_1 \leq x) = \int \limits_{0}^{x} f(\theta_1|\text{Data})d\theta_1$$

The above equation is the posterior cdf, which essentially calculates a confidence bound on the parameter, where \( P(\theta_1 \leq x) \) is the confidence level and \( x \) is the confidence bound. Substituting the posterior pdf into the above posterior cdf yields:

$$CL = \frac{\int \limits_{0}^{x} L(Data|\theta_1)\varphi(\theta_1)d\theta_1}{\int \limits_{0}^{\infty} L(Data|\theta_1)\varphi(\theta_1)d\theta_1}$$

The only question at this point is, what do we use as a prior distribution of \( \theta \)? For the confidence bounds calculation application, non-informative prior distributions are utilized. Non-informative prior distributions are distributions that have no population basis and play a minimal role in the posterior distribution. The idea behind the use of non-informative prior distributions is to make inferences that are not affected by external information, or when external information is not available. In the general case of calculating confidence bounds using Bayesian methods, the method should be independent of external information and it should only rely on the current data. Therefore, non-informative priors are used. Specifically, the uniform distribution is used as a prior distribution for the different parameters of the selected fitted distribution. For example, if the Weibull distribution is fitted to the data, the prior distributions for beta and eta are assumed to be uniform. The above equation can be generalized for any distribution having a vector of parameters \( \theta \) yielding the general equation for calculating Bayesian confidence bounds:
Confidence Bounds

\[
CL = \frac{\int_\xi^\zeta L(Data|\theta) \varphi(\theta) d\theta}{\int_\xi^\zeta L(Data|\theta) \varphi(\theta) d\theta}
\]

where:

- \(CL\) is the confidence level
- \(\theta\) is the parameter vector
- \(L(\cdot)\) is the likelihood function
- \(\varphi(\theta)\) is the prior pdf of the parameter vector \(\theta\)
- \(\xi\) is the range of \(\theta\)
- \(\zeta\) is the range in which \(\theta\) changes from \(\Psi(T, R)\) till \(\theta\)'s maximum value, or from \(\theta\)'s minimum value till \(\Psi(T, R)\)
- \(\Psi(T, R)\) is a function such that if \(T\) is given, then the bounds are calculated for \(R\). If \(R\) is given, then the bounds are calculated for \(T\).

If \(T\) is given, then from the above equation and \(\Psi\) and for a given \(CL\), the bounds on \(R\) are calculated. If \(R\) is given, then from the above equation and \(\Psi\) and for a given \(CL\), the bounds on \(T\) are calculated.

Confidence Bounds on Time (Type 1)

For a given failure time distribution and a given reliability \(R\), \(T(R)\) is a function of \(R\) and the distribution parameters. To illustrate the procedure for obtaining confidence bounds, the two-parameter Weibull distribution is used as an example. The bounds in other types of distributions can be obtained in similar fashion. For the two-parameter Weibull distribution:

\[
T(R) = \eta \exp\left(\frac{\ln(-\ln R)}{\beta}\right)
\]

For a given reliability, the Bayesian one-sided upper bound estimate for \(T(R)\) is:

\[
CL = \Pr\{T \leq T_U\} = \int_0^{T_U \Psi(R)} f(T|Data, R) dT
\]

where \(f(T|Data, R)\) is the posterior distribution of Time \(T\). Using the above equation, we have the following:

\[
CL = \Pr\{T \leq T_U\} = \Pr\left(\eta \exp\left(\frac{\ln(-\ln R)}{\beta}\right) \leq T_U\right)
\]

The above equation can be rewritten in terms of \(\eta\) as:

\[
CL = \Pr\left(\eta \leq T_U \exp\left(\frac{\ln(-\ln R)}{\beta}\right)\right)
\]

Applying the Bayes's rule by assuming that the priors of \(\beta\) and \(\eta\) are independent, we then obtain the following relationship:

\[
CL = \int_0^{\infty} \int_0^{T_U} \exp\left(-\frac{\ln\eta}{\beta}\right) L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta
\]

The above equation can be solved for \(T_U(R)\), where:

- \(CL\) is the confidence level,
- \(\varphi(\beta)\) is the prior pdf of the parameter \(\beta\). For non-informative prior distribution, \(\varphi(\beta) = \frac{1}{\beta}\),
- \(\varphi(\eta)\) is the prior pdf of the parameter \(\eta\). For non-informative prior distribution, \(\varphi(\eta) = \frac{1}{\eta}\),
- \(L(\cdot)\) is the likelihood function.

The same method can be used to get the one-sided lower bound of \(T(R)\) from:
Confidence Bounds

The above equation can be solved to get \( T_L(R) \).

The Bayesian two-sided bounds estimate for \( T(R) \) is:

\[
CL = \frac{\int_0^\infty \int_{T_L(R)}^{T_U(R)} f(T|Data, R) dT}{\int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta}
\]

which is equivalent to:

\[
(1 + CL)/2 = \int_0^{T_U(R)} f(T|Data, R) dT
\]

and:

\[
(1 - CL)/2 = \int_0^{T_L(R)} f(T|Data, R) dT
\]

Using the same method for the one-sided bounds, \( T_U(R) \) and \( T_L(R) \) can be solved.

Confidence Bounds on Reliability (Type 2)

For a given failure time distribution and a given time \( T \), \( R(T) \) is a function of \( T \) and the distribution parameters.

To illustrate the procedure for obtaining confidence bounds, the two-parameter Weibull distribution is used as an example. The bounds in other types of distributions can be obtained in similar fashion. For example, for two parameter Weibull distribution:

\[
R = \exp(-\left(\frac{T}{\eta}\right)^\beta)
\]

The Bayesian one-sided upper bound estimate for \( R(T) \) is:

\[
CL = \int_0^{R_U(T)} f(R|Data, T) dR
\]

Similar to the bounds on Time, the following is obtained:

\[
CL = \frac{\int_0^\infty \int_0^T \exp\left(-\frac{\ln(\frac{1}{R})}{\beta}\right) L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta}{\int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta}
\]

The above equation can be solved to get \( R_U(T) \).

The Bayesian one-sided lower bound estimate for \( R(T) \) is:

\[
1 - CL = \int_0^{R_L(T)} f(R|Data, T) dR
\]

Using the posterior distribution, the following is obtained:

\[
CL = \frac{\int_0^\infty \int_T^{\infty} \exp\left(-\frac{\ln(\frac{1}{R})}{\beta}\right) L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta}{\int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta}
\]

The above equation can be solved to get \( R_L(T) \).

The Bayesian two-sided bounds estimate for \( R(T) \) is:

\[
CL = \int_{R_L(T)}^{R_U(T)} f(R|Data, T) dR
\]

which is equivalent to:

\[
\int_0^{R_U(T)} f(R|Data, T) dR = (1 + CL)/2
\]

and
\[ \int_0^{R_L(T)} f(R|Data, T) dR = (1 - C.L)/2 \]

Using the same method for one-sided bounds, \( R_U(T) \) and \( R_L(T) \) can be solved.

**Simulation Based Bounds**

The SimuMatic tool in Weibull++ can be used to perform a large number of reliability analyses on data sets that have been created using Monte Carlo simulation. This utility can assist the analyst to a) better understand life data analysis concepts, b) experiment with the influences of sample sizes and censoring schemes on analysis methods, c) construct simulation-based confidence intervals, d) better understand the concepts behind confidence intervals and e) design reliability tests. This section describes how to use simulation for estimating confidence bounds.

SimuMatic generates confidence bounds and assists in visualizing and understanding them. In addition, it allows one to determine the adequacy of certain parameter estimation methods (such as rank regression on X, rank regression on Y and maximum likelihood estimation) and to visualize the effects of different data censoring schemes on the confidence bounds.

**Example: Comparing Parameter Estimation Methods Using Simulation Based Bounds**

The purpose of this example is to determine the best parameter estimation method for a sample of ten units with complete time-to-failure data for each unit (i.e., no censoring). The data set follows a Weibull distribution with \( \beta = 2 \) and \( \eta = 100 \) hours.

The confidence bounds for the data set could be obtained by using Weibull++’s SimuMatic utility. To obtain the results, use the following settings in SimuMatic.

1. On the Main tab, choose the 2P-Weibull distribution and enter the given parameters (i.e., \( \beta = 2 \) and \( \eta = 100 \) hours).
2. On the Censoring tab, select the **No censoring** option.
3. On the Settings tab, set the number of data sets to **1,000** and the number of data points to **10**.
4. On the Analysis tab, choose the **RRX** analysis method and set the confidence bounds to **90**.

The following plot shows the simulation-based confidence bounds for the RRX parameter estimation method, as well as the expected variation due to sampling error.
Create another SimuMatic folio and generate a second data using the same settings, but this time, select the RRY analysis method on the Analysis tab. The following plot shows the result.
The following plot shows the results using the MLE analysis method.
The results clearly demonstrate that the median RRX estimate provides the least deviation from the truth for this sample size and data type. However, the MLE outputs are grouped more closely together, as evidenced by the bounds.

This experiment can be repeated in SimuMatic using multiple censoring schemes (including Type I and Type II right censoring as well as random censoring) with various distributions. Multiple experiments can be performed with this utility to evaluate assumptions about the appropriate parameter estimation method to use for data sets.
Chapter 7

The Exponential Distribution

The exponential distribution is a commonly used distribution in reliability engineering. Mathematically, it is a fairly simple distribution, which many times leads to its use in inappropriate situations. It is, in fact, a special case of the Weibull distribution where \( \beta = 1 \). The exponential distribution is used to model the behavior of units that have a constant failure rate (or units that do not degrade with time or wear out).

Exponential Probability Density Function

The 2-Parameter Exponential Distribution

The 2-parameter exponential pdf is given by:

\[
f(t) = \lambda e^{-\lambda(t-\gamma)}, \quad f(t) \geq 0, \lambda > 0, t \geq 0 \text{ or } \gamma
\]

where \( \gamma \) is the location parameter. Some of the characteristics of the 2-parameter exponential distribution are discussed in Kececioglu [19]:

- The location parameter, \( \gamma \), if positive, shifts the beginning of the distribution by a distance of \( \gamma \) to the right of the origin, signifying that the chance failures start to occur only after \( \gamma \) hours of operation, and cannot occur before.
- The scale parameter is \( \lambda = \frac{1}{\lambda} = \bar{t} - \gamma = m - \gamma \).
- The exponential pdf has no shape parameter, as it has only one shape.
- The distribution starts at \( t = \gamma \) at the level of \( f(t = \gamma) = \lambda \) and decreases thereafter exponentially and monotonically as \( t \) increases beyond \( \gamma \) and is convex.
- As \( t \to \infty \), \( f(t) \to 0 \).

The 1-Parameter Exponential Distribution

The 1-parameter exponential pdf is obtained by setting \( \gamma = 0 \), and is given by:

\[
f(t) = \lambda e^{-\lambda t} = \frac{1}{m} e^{-\frac{1}{m} t}, \quad t \geq 0, \lambda > 0, m > 0
\]

where:

- \( \lambda \) = constant rate, in failures per unit of measurement, (e.g., failures per hour, per cycle, etc.)
- \( \lambda = \frac{1}{m} \)
- \( m \) = mean time between failures, or to failure
- \( \bar{t} \) = operating time, life, or age, in hours, cycles, miles, actuations, etc.

This distribution requires the knowledge of only one parameter, \( \lambda \), for its application. Some of the characteristics of the 1-parameter exponential distribution are discussed in Kececioglu [19]:

- The location parameter, \( \gamma \), is zero.
- The scale parameter is \( \frac{1}{\lambda} = m \).
- As \( \lambda \) is decreased in value, the distribution is stretched out to the right, and as \( \lambda \) is increased, the distribution is pushed toward the origin.
- This distribution has no shape parameter as it has only one shape, (i.e., the exponential, and the only parameter it has is the failure rate, \( \lambda \)).
• The distribution starts at \( t = 0 \) at the level of \( f(t = 0) = \lambda \) and decreases thereafter exponentially and monotonically as \( t \) increases, and is convex.
• As \( t \to \infty \), \( f(t) \to 0 \).
• The pdf can be thought of as a special case of the Weibull pdf with \( \gamma = 0 \) and \( \beta = 1 \).

Exponential Distribution Functions

The Mean or MTTF

The mean, \( \bar{T} \), or mean time to failure (MTTF) is given by:

\[
\bar{T} = \int_{0}^{\infty} t \cdot f(t) \, dt = \int_{0}^{\infty} t \cdot \lambda \cdot e^{-\lambda t} \, dt = \frac{1}{\lambda} + \gamma = m
\]

Note that when \( \gamma = 0 \), the MTTF is the inverse of the exponential distribution's constant failure rate. This is only true for the exponential distribution. Most other distributions do not have a constant failure rate. Consequently, the inverse relationship between failure rate and MTTF does not hold for these other distributions.

The Median

The median, \( \tilde{T} \), is:

\[
\tilde{T} = \gamma + \frac{1}{\lambda} \cdot 0.693
\]

The Mode

The mode, \( \hat{T} \), is:

\[
\hat{T} = \gamma
\]

The Standard Deviation

The standard deviation, \( \sigma_{T} \), is:

\[
\sigma_{T} = \frac{1}{\lambda} = m
\]

The Exponential Reliability Function

The equation for the 2-parameter exponential cumulative density function, or \( cdf \), is given by:

\[
F(t) = Q(t) = 1 - e^{-\lambda(t-\gamma)}
\]

Recalling that the reliability function of a distribution is simply one minus the \( cdf \), the reliability function of the 2-parameter exponential distribution is given by:

\[
R(t) = 1 - Q(t) = 1 - \int_{0}^{t-\gamma} f(x) \, dx
\]

\[
R(t) = 1 - \int_{0}^{t-\gamma} \lambda e^{-\lambda x} \, dx = e^{-\lambda(t-\gamma)}
\]

The 1-parameter exponential reliability function is given by:

\[
R(t) = e^{-\lambda t} = e^{-\frac{t}{m}}
\]
The Exponential Conditional Reliability Function

The exponential conditional reliability equation gives the reliability for a mission of duration, having already successfully accumulated operation up to the start of this new mission. The exponential conditional reliability function is:

\[
R(t|T) = \frac{R(T+t)}{R(T)} = \frac{e^{-\lambda(T+t-\gamma)}}{e^{-\lambda(T-\gamma)}} = e^{-\lambda t}
\]

which says that the reliability for a mission of duration undertaken after the component or equipment has already accumulated operation from age zero is only a function of the mission duration, and not a function of the age at the beginning of the mission. This is referred to as the memoryless property.

The Exponential Reliable Life Function

The reliable life, or the mission duration for a desired reliability goal, for the 1-parameter exponential distribution is:

\[
R(t_R) = e^{-\lambda(t_R-\gamma)} \\
\ln[R(t_R)] = -\lambda(t_R - \gamma)
\]

or:

\[
t_R = \gamma - \frac{\ln[R(t_R)]}{\lambda}
\]

The Exponential Failure Rate Function

The exponential failure rate function is:

\[
\lambda(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda(t-\gamma)}}{e^{-\lambda(t-\gamma)}} = \lambda = \text{constant}
\]

Once again, note that the constant failure rate is a characteristic of the exponential distribution, and special cases of other distributions only. Most other distributions have failure rates that are functions of time.

Characteristics of the Exponential Distribution

The primary trait of the exponential distribution is that it is used for modeling the behavior of items with a constant failure rate. It has a fairly simple mathematical form, which makes it fairly easy to manipulate. Unfortunately, this fact also leads to the use of this model in situations where it is not appropriate. For example, it would not be appropriate to use the exponential distribution to model the reliability of an automobile. The constant failure rate of the exponential distribution would require the assumption that the automobile would be just as likely to experience a breakdown during the first mile as it would during the one-hundred-thousandth mile. Clearly, this is not a valid assumption. However, some inexperienced practitioners of reliability engineering and life data analysis will overlook this fact, lured by the siren-call of the exponential distribution's relatively simple mathematical models.
The Exponential Distribution

The Effect of lambda and gamma on the Exponential pdf

- The exponential pdf has no shape parameter, as it has only one shape.
- The exponential pdf is always convex and is stretched to the right as $\lambda$ decreases in value.
- The value of the pdf function is always equal to the value of $\lambda$ at $t = 0$ (or $t = \gamma$).
- The location parameter, $\gamma$, if positive, shifts the beginning of the distribution by a distance of $\gamma$ to the right of the origin, signifying that the chance failures start to occur only after $\gamma$ hours of operation, and cannot occur before this time.
- The scale parameter is $\frac{1}{\lambda} = \bar{T} - \gamma = m - \gamma$.
- As $t \to \infty$, $f(t) \to 0$. 
The Exponential Distribution

The Effect of lambda and gamma on the Exponential Reliability Function

• The 1-parameter exponential reliability function starts at the value of 100% at \( t = 0 \), decreases thereafter monotonically and is convex.

• The 2-parameter exponential reliability function remains at the value of 100% for \( t = 0 \) up to \( t = \gamma \), and decreases thereafter monotonically and is convex.

• As \( t \to \infty \), \( R(t \to \infty) \to 0 \).

• The reliability for a mission duration of \( t = m = \frac{1}{\lambda} \) or of one MTTF duration, is always equal to 0.3679 or 36.79%. This means that the reliability for a mission which is as long as one MTTF is relatively low and is not recommended because only 36.8% of the missions will be completed successfully. In other words, of the equipment undertaking such a mission, only 36.8% will survive their mission.
The Effect of lambda and gamma on the Failure Rate Function

- The 1-parameter exponential failure rate function is constant and starts at \( t = 0 \).
- The 2-parameter exponential failure rate function remains at the value of 0 for \( t = 0 \) up to \( t = \gamma \), and then keeps at the constant value of \( \lambda \).

Estimation of the Exponential Parameters

Probability Plotting

Estimation of the parameters for the exponential distribution via probability plotting is very similar to the process used when dealing with the Weibull distribution. Recall, however, that the appearance of the probability plotting paper and the methods by which the parameters are estimated vary from distribution to distribution, so there will be some noticeable differences. In fact, due to the nature of the exponential cdf, the exponential probability plot is the only one with a negative slope. This is because the y-axis of the exponential probability plotting paper represents the reliability, whereas the y-axis for most of the other life distributions represents the unreliability.

This is illustrated in the process of linearizing the cdf, which is necessary to construct the exponential probability plotting paper. For the two-parameter exponential distribution the cumulative density function is given by:

\[
F(t) = 1 - e^{-\lambda(t-\gamma)}
\]

Taking the natural logarithm of both sides of the above equation yields:

\[
\ln [1 - F(t)] = -\lambda(t - \gamma)
\]

or:

\[
\ln[1 - F(t)] = \lambda \gamma - \lambda t
\]

Now, let:
\[ y = \ln[1 - F(t)] \]
\[ a = \lambda \gamma \]

and:
\[ b = -\lambda \]

which results in the linear equation of:
\[ y = a + bt \]

Note that with the exponential probability plotting paper, the y-axis scale is logarithmic and the x-axis scale is linear. This means that the zero value is present only on the x-axis. For \( t = 0 \), \( R = 1 \) and \( F(t) = 0 \). So if we were to use \( F(t) \) for the y-axis, we would have to plot the point \((0, 0)\). However, since the y-axis is logarithmic, there is no place to plot this on the exponential paper. Also, the failure rate, \( \lambda \), is the negative of the slope of the line, but there is an easier way to determine the value of \( \lambda \) from the probability plot, as will be illustrated in the following example.

**Plotting Example**

**1-Parameter Exponential Probability Plot Example**

6 units are put on a life test and tested to failure. The failure times are 7, 12, 19, 29, 41, and 67 hours. Estimate the failure rate for a 1-parameter exponential distribution using the probability plotting method.

In order to plot the points for the probability plot, the appropriate reliability estimate values must be obtained. These will be equivalent to \( 100\% - MR \) since the y-axis represents the reliability and the \( MR \) values represent unreliability estimates.

<table>
<thead>
<tr>
<th>Time-to-failure, hr</th>
<th>Reliability Estimate, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>100 - 10.91 = 89.09%</td>
</tr>
<tr>
<td>12</td>
<td>100 - 26.44 = 73.56%</td>
</tr>
<tr>
<td>19</td>
<td>100 - 42.14 = 57.86%</td>
</tr>
<tr>
<td>29</td>
<td>100 - 57.86 = 42.14%</td>
</tr>
<tr>
<td>41</td>
<td>100 - 73.56 = 26.44%</td>
</tr>
<tr>
<td>67</td>
<td>100 - 89.09 = 10.91%</td>
</tr>
</tbody>
</table>

Next, these points are plotted on an exponential probability plotting paper. A sample of this type of plotting paper is shown next, with the sample points in place. Notice how these points describe a line with a negative slope.
Once the points are plotted, draw the best possible straight line through these points. The time value at which this line intersects with a horizontal line drawn at the 36.8% reliability mark is the mean life, and the reciprocal of this is the failure rate $\lambda$. This is because at $t = m = \frac{1}{\lambda}$.

$$R(t) = e^{-\lambda t}$$

$$R(t) = e^{-\lambda \frac{1}{\lambda}} = e^{-1} = 0.368 = 36.8\%.$$ 

The following plot shows that the best-fit line through the data points crosses the $R = 36.8\%$ line at $t = 33$ hours. And because $\frac{1}{\lambda} = 33$ hours, $\lambda = 0.0303$ failures/hour.
Rank Regression on Y

Performing a rank regression on Y requires that a straight line be fitted to the set of available data points such that the sum of the squares of the vertical deviations from the points to the line is minimized. The least squares parameter estimation method (regression analysis) was discussed in Parameter Estimation, and the following equations for rank regression on Y (RRY) were derived:

\[ \hat{a} = \bar{y} - \hat{b} \bar{x} = \frac{\sum_{i=1}^{N} y_i}{N} - \frac{\sum_{i=1}^{N} x_i}{N} \]

and:

\[ \hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sum_{i=1}^{N} x_i^2 - \frac{\left( \sum_{i=1}^{N} x_i \right)^2}{N}} \]

In our case, the equations for \( y_i \) and \( x_i \) are:

\[ y_i = \ln[1 - F(t_i)] \]

and:

\[ x_i = t_i \]

and the \( F(t_i) \) is estimated from the median ranks. Once \( \hat{a} \) and \( \hat{b} \) are obtained, then \( \hat{\lambda} \) and \( \hat{\gamma} \) can easily be obtained from above equations. For the one-parameter exponential, equations for estimating \( a \) and \( b \) become:
\begin{align*}
\hat{\alpha} &= 0, \\
\hat{\beta} &= \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}
\end{align*}

The Correlation Coefficient

The estimator of $\rho$ is the sample correlation coefficient, $\hat{\rho}$, given by:

$$
\hat{\rho} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2}}
$$

RRY Example

2-Parameter Exponential RRY Example

14 units were being reliability tested and the following life test data were obtained. Assuming that the data follow a 2-parameter exponential distribution, estimate the parameters and determine the correlation coefficient, $\hat{\rho}$, using rank regression on Y (RRY).

<table>
<thead>
<tr>
<th>Data point index</th>
<th>Time-to-failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>11</td>
<td>70</td>
</tr>
<tr>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td>13</td>
<td>90</td>
</tr>
<tr>
<td>14</td>
<td>100</td>
</tr>
</tbody>
</table>

Solution

Construct the following table, as shown next.

Table- Least Squares Analysis
The Exponential Distribution

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>t_i</td>
<td>F(t_i)</td>
<td>y_i</td>
<td>t_i^2</td>
<td>y_i^2</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.0483</td>
<td>-0.0495</td>
<td>25</td>
<td>0.0025</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>0.1170</td>
<td>-0.1244</td>
<td>100</td>
<td>0.0155</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>0.1865</td>
<td>-0.2064</td>
<td>225</td>
<td>0.0426</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>0.2561</td>
<td>-0.2958</td>
<td>400</td>
<td>0.0875</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>0.3258</td>
<td>-0.3942</td>
<td>625</td>
<td>0.1554</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>0.3954</td>
<td>-0.5032</td>
<td>900</td>
<td>0.2532</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
<td>0.4651</td>
<td>-0.6257</td>
<td>1225</td>
<td>0.3915</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>0.5349</td>
<td>-0.7655</td>
<td>1600</td>
<td>0.5860</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>0.6046</td>
<td>-0.9279</td>
<td>2500</td>
<td>0.8609</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>0.6742</td>
<td>-1.1215</td>
<td>3600</td>
<td>1.2577</td>
</tr>
<tr>
<td>11</td>
<td>70</td>
<td>0.7439</td>
<td>-1.3622</td>
<td>4900</td>
<td>1.8456</td>
</tr>
<tr>
<td>12</td>
<td>80</td>
<td>0.8135</td>
<td>-1.6793</td>
<td>6400</td>
<td>2.8201</td>
</tr>
<tr>
<td>13</td>
<td>90</td>
<td>0.8830</td>
<td>-2.1456</td>
<td>8100</td>
<td>4.6035</td>
</tr>
<tr>
<td>14</td>
<td>100</td>
<td>0.9517</td>
<td>-3.0303</td>
<td>10000</td>
<td>9.1829</td>
</tr>
<tr>
<td></td>
<td></td>
<td>630</td>
<td></td>
<td>14</td>
<td>630</td>
</tr>
</tbody>
</table>

The median rank values \( F(t_i) \) can be found in rank tables or they can be estimated using the Quick Statistical Reference in Weibull++. Given the values in the table above, calculate \( \hat{a} \) and \( \hat{b} \):

\[
\hat{b} = \frac{\sum_{i=1}^{14} t_i y_i - \left( \sum_{i=1}^{14} t_i \right) \left( \sum_{i=1}^{14} y_i \right) / 14}{\sum_{i=1}^{14} t_i^2 - \left( \sum_{i=1}^{14} t_i \right)^2 / 14}
\]

\[
\hat{b} = \frac{-927.4899 - (630)(-13.2315)/14}{40,600 - (630)^2/14}
\]

or:

\[
\hat{b} = -0.02711
\]

and:

\[
\hat{a} = \bar{y} - \hat{b} \bar{t} = \frac{\sum_{i=1}^{14} y_i}{N} - \hat{b} \frac{\sum_{i=1}^{14} t_i}{N}
\]

or:

\[
\hat{a} = \frac{-13.2315}{14} - (-0.02711) \frac{630}{14} = 0.2748
\]

Therefore:

\[
\hat{\lambda} = -\hat{b} = -(-0.02711) = 0.02711 \text{ failures/hour}
\]

and:

\[
\hat{\gamma} = \frac{\hat{a}}{\hat{\lambda}} = \frac{0.2748}{0.02711}
\]

or:

\[
\hat{\gamma} = 10.1365 \text{ hours}
\]

Then:

\[
f(t) = (0.02711) \cdot e^{-0.02711(t-10.1365)}
\]

The correlation coefficient can be estimated using equation for calculating the correlation coefficient:

\[
\hat{\rho} = -0.9679
\]
This example can be repeated using Weibull++, choosing 2-parameter exponential and rank regression on Y (RRY), as shown next.

The estimated parameters and the correlation coefficient using Weibull++ were found to be:

\[ \hat{\lambda} = 0.0271 \text{ fr/hr}, \quad \hat{\gamma} = 10.1348 \text{ hr}, \quad \hat{\rho} = -0.9679 \]

Please note that the user must deselect the Reset if location parameter > T1 on Exponential option on the Calculations page of the Application Setup window.

The probability plot can be obtained simply by clicking the Plot icon.
Rank Regression on X

Similar to rank regression on Y, performing a rank regression on X requires that a straight line be fitted to a set of data points such that the sum of the squares of the horizontal deviations from the points to the line is minimized.

Again the first task is to bring our exponential cdf function into a linear form. This step is exactly the same as in regression on Y analysis. The deviation from the previous analysis begins on the least squares fit step, since in this case we treat \( x \) as the dependent variable and \( y \) as the independent variable. The best-fitting straight line to the data, for regression on X (see Parameter Estimation), is the straight line:

\[
x = \hat{a} + \hat{b}y
\]

The corresponding equations for \( \hat{a} \) and \( \hat{b} \) are:

\[
\hat{a} = \bar{x} - \hat{b}\bar{y} = \frac{\sum_{i=1}^{N} x_i}{N} - \hat{b}\frac{\sum_{i=1}^{N} y_i}{N}
\]

and:

\[
\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sum_{i=1}^{N} y_i^2 - \left(\frac{\sum_{i=1}^{N} y_i}{N}\right)^2}
\]

where:

\[
y_i = \ln[1 - F(t_i)]
\]

and:
The values of $F(t_i)$ are estimated from the median ranks. Once $\hat{\lambda}$ and $\hat{\mu}$ are obtained, solve for the unknown $\gamma$ value, which corresponds to:

$$y = -\frac{1}{\hat{\mu}} + \frac{1}{\hat{\lambda}} x$$

Solving for the parameters from above equations we get:

$$a = -\frac{\hat{\lambda}}{\hat{\mu}} = \lambda \gamma \Rightarrow \gamma = \frac{a}{\hat{\lambda}}$$

and:

$$b = \frac{1}{\hat{\mu}} = -\lambda \Rightarrow \lambda = -\frac{1}{b}$$

For the one-parameter exponential case, equations for estimating $a$ and $b$ become:

$$\hat{\lambda} = 0$$

$$\hat{\mu} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} y_i^2}$$

The correlation coefficient is evaluated as before.

**RRX Example**

**2-Parameter Exponential RRX Example**

Using the same data set from the RRY example above and assuming a 2-parameter exponential distribution, estimate the parameters and determine the correlation coefficient estimate, $\hat{\rho}$, using rank regression on $X$.

**Solution**

The table constructed for the RRY analysis applies to this example also. Using the values from this table, we get:

$$\hat{\mu} = \frac{\sum_{i=1}^{14} t_i y_i - \frac{14}{14} \sum_{i=1}^{14} t_i \sum_{i=1}^{14} y_i}{\sum_{i=1}^{14} y_i^2 - \left(\frac{1}{14} \sum_{i=1}^{14} y_i\right)^2}$$

$$\hat{\mu} = \frac{-927.4899 - (630)(-13.2315)/14}{22.1148 - (-13.2315)^2/14}$$

or:

$$\hat{\mu} = -34.5563$$

and:

$$\hat{\lambda} = \bar{x} - \hat{\mu} \bar{y} = \frac{\sum_{i=1}^{14} t_i}{14} - \hat{\mu} \frac{\sum_{i=1}^{14} y_i}{14}$$

or:

$$\hat{\lambda} = \frac{630}{14} - (-34.5563) \left(-13.2315\right) = 12.3406$$

Therefore:
The Exponential Distribution

\[ \hat{\lambda} = \frac{1}{\hat{b}} = \frac{1}{-34.5563} = 0.0289 \text{ failures/hour} \]

and:

\[ \hat{\gamma} = \hat{\alpha} = 12.3406 \]

The correlation coefficient is found to be:

\[ \hat{\rho} = -0.9679 \]

Note that the equation for regression on Y is not necessarily the same as that for the regression on X. The only time when the two regression methods yield identical results is when the data lie perfectly on a line. If this were the case, the correlation coefficient would be \(-1\). The negative value of the correlation coefficient is due to the fact that the slope of the exponential probability plot is negative.

This example can be repeated using Weibull++, choosing two-parameter exponential and rank regression on X (RRX) methods for analysis, as shown below. The estimated parameters and the correlation coefficient using Weibull++ were found to be:

\[ \hat{\lambda} = 0.0289 \text{ failures/hour} \]
\[ \hat{\gamma} = 12.3395 \text{ hours} \]
\[ \hat{\rho} = -0.9679 \]

The probability plot can be obtained simply by clicking the Plot icon.
Maximum Likelihood Estimation

As outlined in Parameter Estimation, maximum likelihood estimation works by developing a likelihood function based on the available data and finding the values of the parameter estimates that maximize the likelihood function. This can be achieved by using iterative methods to determine the parameter estimate values that maximize the likelihood function. This can be rather difficult and time-consuming, particularly when dealing with the three-parameter distribution. Another method of finding the parameter estimates involves taking the partial derivatives of the likelihood equation with respect to the parameters, setting the resulting equations equal to zero, and solving simultaneously to determine the values of the parameter estimates. The log-likelihood functions and associated partial derivatives used to determine maximum likelihood estimates for the exponential distribution are covered in Appendix D.

MLE Example

MLE for the Exponential Distribution

Using the same data set from the RRY and RRX examples above and assuming a 2-parameter exponential distribution, estimate the parameters using the MLE method.

Solution

In this example, we have complete data only. The partial derivative of the log-likelihood function, $\hat{\lambda}$, is given by:

$$\frac{\partial \Lambda}{\partial \lambda} = \sum_{i=1}^{F_{ex}} \left[ \frac{1}{\lambda} - (T_i - \gamma) \right] = \sum_{i=1}^{14} \left[ \frac{1}{\lambda} - (T_i - \gamma) \right] = 0$$

Complete descriptions of the partial derivatives can be found in Appendix D. Recall that when using the MLE method for the exponential distribution, the value of $\hat{\gamma}$ is equal to that of the first failure time. The first failure
occurred at 5 hours, thus $\gamma = 5\text{ hours}$.

Substituting the values for $T$ and $\gamma$, we get:

$$\frac{14}{\lambda} = 560$$

or:

$$\lambda = 0.025 \text{ failures/hour}$$

Using Weibull++:

The probability plot is:
Confidence Bounds

In this section, we present the methods used in the application to estimate the different types of confidence bounds for exponentially distributed data. The complete derivations were presented in detail (for a general function) in the chapter for Confidence Bounds. At this time we should point out that exact confidence bounds for the exponential distribution have been derived, and exist in a closed form, utilizing the $\chi^2$ distribution. These are described in detail in Kececioğlu [20], and are covered in the section in the test design chapter. For most exponential data analyses, Weibull++ will use the approximate confidence bounds, provided from the Fisher information matrix or the likelihood ratio, in order to stay consistent with all of the other available distributions in the application. The $\chi^2$ confidence bounds for the exponential distribution are discussed in more detail in the test design chapter.

Fisher Matrix Bounds

Bounds on the Parameters

For the failure rate $\hat{\lambda}$ the upper ($\lambda_U$) and lower ($\lambda_L$) bounds are estimated by Nelson [30]:

$$
\lambda_U = \hat{\lambda} \cdot e \left[ K_\alpha \sqrt{\text{Var}(\hat{\lambda})} \right] / \lambda
$$

$$
\lambda_L = \frac{\hat{\lambda}}{K_\alpha \sqrt{\text{Var}(\hat{\lambda})} \sqrt{e}}
$$

where $K_\alpha$ is defined by:
If \( \delta \) is the confidence level, then \( \alpha = \frac{1 - \delta}{2} \) for the two-sided bounds, and \( \alpha = 1 - \delta \) for the one-sided bounds.

The variance of \( \hat{\lambda}, \text{Var}(\hat{\lambda}) \) is estimated from the Fisher matrix, as follows:

\[
\text{Var}(\hat{\lambda}) = \left( -\frac{\partial^2 \Lambda}{\partial \lambda^2} \right)^{-1}
\]

where \( \Lambda \) is the log-likelihood function of the exponential distribution, described in Appendix D.

Note that no true MLE solution exists for the case of the two-parameter exponential distribution. The mathematics simply break down while trying to simultaneously solve the partial derivative equations for both the \( \gamma \) and \( \lambda \) parameters, resulting in unrealistic conditions. The way around this conundrum involves setting \( \gamma = \hat{t}_1 \) or the first time-to-failure, and calculating \( \lambda \) in the regular fashion for this methodology. Weibull++ treats \( \gamma \) as a constant when computing bounds, (i.e., \( \text{Var}(\hat{\gamma}) = 0 \)). (See the discussion in Appendix D for more information.)

**Bounds on Reliability**

The reliability of the two-parameter exponential distribution is:

\[
\hat{R}(t; \hat{\lambda}) = e^{-\hat{\lambda}(t-\hat{\gamma})}
\]

The corresponding confidence bounds are estimated from:

\[
\hat{R}_L = e^{-\lambda_L(t-\hat{\gamma})}
\]

\[
\hat{R}_U = e^{-\lambda_U(t-\hat{\gamma})}
\]

These equations hold true for the 1-parameter exponential distribution, with \( \gamma = 0 \).

**Bounds on Time**

The bounds around time for a given exponential percentile, or reliability value, are estimated by first solving the reliability equation with respect to time, or reliable life:

\[
\hat{t} = -\frac{1}{\hat{\lambda}} \cdot \ln(\hat{R}) + \hat{\gamma}
\]

The corresponding confidence bounds are estimated from:

\[
\hat{t}_U = -\frac{1}{\lambda_U} \cdot \ln(\hat{R}) + \hat{\gamma}
\]

\[
\hat{t}_L = -\frac{1}{\lambda_L} \cdot \ln(\hat{R}) + \hat{\gamma}
\]

The same equations apply for the one-parameter exponential with \( \gamma = 0 \).

**Likelihood Ratio Confidence Bounds**

**Bounds on Parameters**

For one-parameter distributions such as the exponential, the likelihood confidence bounds are calculated by finding values for \( \hat{\theta} \) that satisfy:

\[
-2 \cdot \ln \left( \frac{L(\theta)}{L(\hat{\theta})} \right) = \chi^2_{\alpha;1}
\]

This equation can be rewritten as:

\[
L(\theta) = L(\hat{\theta}) \cdot e^{-\frac{\chi^2_{\alpha;1}}{2}}
\]

For complete data, the likelihood function for the exponential distribution is given by:
The Exponential Distribution

where the \( t_i \) values represent the original time-to-failure data. For a given value of \( \alpha \), values for \( \lambda \) can be found which represent the maximum and minimum values that satisfy the above likelihood ratio equation. These represent the confidence bounds for the parameters at a confidence level \( \delta \) where \( \alpha = 2 \delta - 1 \) for two-sided bounds and \( \alpha = 2 \delta - 1 \) for one-sided.

**Example: LR Bounds for Lambda**

Five units are put on a reliability test and experience failures at 20, 40, 60, 100, and 150 hours. Assuming an exponential distribution, the MLE parameter estimate is calculated to be \( \hat{\lambda} = 0.013514 \). Calculate the 85\% two-sided confidence bounds on these parameters using the likelihood ratio method.

**Solution**

The first step is to calculate the likelihood function for the parameter estimates:

\[
L(\lambda) = \prod_{i=1}^{N} f(t_i; \lambda) = \prod_{i=1}^{N} \lambda \cdot e^{-\lambda t_i}
\]

\[
L(\hat{\lambda}) = \prod_{i=1}^{5} 0.013514 \cdot e^{-0.013514 t_i}
\]

\[
L(\hat{\lambda}) = 3.03647 \times 10^{-12}
\]

where \( x_i \) are the original time-to-failure data points. We can now rearrange the likelihood ratio equation to the form:

\[
L(\lambda) - L(\hat{\lambda}) \cdot e^{-\frac{\hat{\lambda}^2}{2}} = 0
\]

Since our specified confidence level, \( \delta \), is 85\%, we can calculate the value of the chi-squared statistic, \( \chi^2_{0.85,1} = 2.072251 \). We can now substitute this information into the equation:

\[
L(\lambda) - L(\hat{\lambda}) \cdot e^{-\frac{\hat{\lambda}^2}{2}} = 0,
\]

\[
L(\lambda) - 3.03647 \times 10^{-12} \cdot e^{-\frac{2.072251}{2}} = 0,
\]

\[
L(\lambda) - 1.07742 \times 10^{-12} = 0.
\]

It now remains to find the values of \( \lambda \) which satisfy this equation. Since there is only one parameter, there are only two values of \( \lambda \) that will satisfy the equation. These values represent the \( \delta = 85\% \) two-sided confidence limits of the parameter estimate \( \hat{\lambda} \). For our problem, the confidence limits are:

\[
\lambda_{0.85} = (0.006572, 0.024172)
\]

**Bounds on Time and Reliability**

In order to calculate the bounds on a time estimate for a given reliability, or on a reliability estimate for a given time, the likelihood function needs to be rewritten in terms of one parameter and time/reliability, so that the maximum and minimum values of the time can be observed as the parameter is varied. This can be accomplished by substituting a form of the exponential reliability equation into the likelihood function. The exponential reliability equation can be written as:

\[
R = e^{-\lambda t}
\]

This can be rearranged to the form:

\[
\lambda = \frac{-\ln(R)}{t}
\]
This equation can now be substituted into the likelihood ratio equation to produce a likelihood equation in terms of $t$ and $\tilde{R}$:

$$L(t/\tilde{R}) = \prod_{i=1}^{N} \left( \frac{-\ln(R_i)}{t} \right) \cdot e^{\left( \frac{-\ln(t)}{\tilde{R}} \right)_{-1}}$$

The unknown parameter $t/\tilde{R}$ depends on what type of bounds are being determined. If one is trying to determine the bounds on time for the equation for the mean and the Bayes's rule equation for single parameter given reliability, then $\tilde{R}$ is a known constant and $t$ is the unknown parameter. Conversely, if one is trying to determine the bounds on reliability for a given time, then $t$ is a known constant and $\tilde{R}$ is the unknown parameter. Either way, the likelihood ratio function can be solved for the values of interest.

**Example: LR Bounds on Time**

For the data given above for the LR Bounds on Lambda example (five failures at 20, 40, 60, 100 and 150 hours), determine the 85% two-sided confidence bounds on the time estimate for a reliability of 90%. The ML estimate for the time at $R(t) = 90\%$ is $\hat{t} = 7.797$.

**Solution**

In this example, we are trying to determine the 85% two-sided confidence bounds on the time estimate of 7.797. This is accomplished by substituting $\tilde{R} = 0.90$ and $\alpha = 0.85$ into the likelihood ratio bound equation. It now remains to find the values of $t$ which satisfy this equation. Since there is only one parameter, there are only two values of $t$ that will satisfy the equation. These values represent the $\delta = 85\%$ two-sided confidence limits of the time estimate $\hat{t}$. For our problem, the confidence limits are:

$$\hat{t}_{R=0.9} = (4.359, 16.033)$$

**Example: LR Bounds on Reliability**

Again using the data given above for the LR Bounds on Lambda example (five failures at 20, 40, 60, 100 and 150 hours), determine the 85% two-sided confidence bounds on the reliability estimate for a $t = 50$ hours. The ML estimate for the time at $t = 50$ is $\tilde{R} = 50.881\%$.

**Solution**

In this example, we are trying to determine the 85% two-sided confidence bounds on the reliability estimate of 50.881%. This is accomplished by substituting $t = 50$ and $\alpha = 0.85$ into the likelihood ratio bound equation. It now remains to find the values of $\tilde{R}$ which satisfy this equation. Since there is only one parameter, there are only two values of $\tilde{R}$ that will satisfy the equation. These values represent the $\delta = 85\%$ two-sided confidence limits of the reliability estimate $\tilde{R}$. For our problem, the confidence limits are:

$$\tilde{R}_{t=50} = (29.861\%, 71.794\%)$$

**Bayesian Confidence Bounds**

**Bounds on Parameters**

From Confidence Bounds, we know that the posterior distribution of $\lambda$ can be written as:

$$f(\lambda|Data) = \frac{L(Data|\lambda)\varphi(\lambda)}{\int_{0}^{\infty} L(Data|\lambda)\varphi(\lambda)d\lambda}$$

where $\varphi(\lambda) = \frac{1}{\lambda}$ is the non-informative prior of $\lambda$.

With the above prior distribution, $f(\lambda|Data)$ can be rewritten as:

$$f(\lambda|Data) = \frac{L(Data|\lambda)\frac{1}{\lambda}}{\int_{0}^{\infty} L(Data|\lambda)\frac{1}{\lambda}d\lambda}$$
The one-sided upper bound of $\lambda$ is:

$$CL = P(\lambda \leq \lambda_U) = \int_0^{\lambda_U} f(\lambda | Data) d\lambda$$

The one-sided lower bound of $\lambda$ is:

$$1 - CL = P(\lambda \leq \lambda_L) = \int_0^{\lambda_L} f(\lambda | Data) d\lambda$$

The two-sided bounds of $\lambda$ are:

$$CL = P(\lambda_L \leq \lambda \leq \lambda_U) = \int_{\lambda_L}^{\lambda_U} f(\lambda | Data) d\lambda$$

**Bounds on Time (Type 1)**

The reliable life equation is:

$$t = -\frac{\ln R}{\lambda}$$

For the one-sided upper bound on time we have:

$$CL = \Pr(t \leq T_U) = \Pr\left(\frac{-\ln R}{\lambda} \leq T_U\right)$$

The above equation can be rewritten in terms of $\lambda$ as:

$$CL = \Pr\left(\frac{-\ln R}{t_U} \leq \lambda\right)$$

From the above posterior distribution equation, we have:

$$CL = \frac{\int_{-\infty}^{\ln R} L(Data | \lambda) \frac{1}{\lambda} d\lambda}{\int_{-\infty}^{\ln R} L(Data | \lambda) \frac{1}{\lambda} d\lambda}$$

The above equation is solved w.r.t. $t_U$. The same method is applied for one-sided lower and two-sided bounds on time.

**Bounds on Reliability (Type 2)**

The one-sided upper bound on reliability is given by:

$$CL = \Pr(R \leq R_U) = \Pr\left(\exp(-\lambda t) \leq R_U\right)$$

The above equation can be rewritten in terms of $\lambda$ as:

$$CL = \Pr\left(\frac{-\ln R_U}{t} \leq \lambda\right)$$

From the equation for posterior distribution we have:

$$CL = \frac{\int_{-\infty}^{\ln R_U} L(Data | \lambda) \frac{1}{\lambda} d\lambda}{\int_{-\infty}^{\ln R_U} L(Data | \lambda) \frac{1}{\lambda} d\lambda}$$

The above equation is solved w.r.t. $R_U$. The same method can be used to calculate one-sided lower and two-sided bounds on reliability.
Exponential Distribution Examples

Grouped Data

20 units were reliability tested with the following results:

<table>
<thead>
<tr>
<th>Number of Units in Group</th>
<th>Time-to-Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
</tr>
<tr>
<td>1</td>
<td>500</td>
</tr>
<tr>
<td>2</td>
<td>600</td>
</tr>
</tbody>
</table>

1. Assuming a 2-parameter exponential distribution, estimate the parameters by hand using the MLE analysis method.

2. Repeat the above using Weibull++. (Enter the data as grouped data to duplicate the results.)

3. Show the Probability plot for the analysis results.

4. Show the Reliability vs. Time plot for the results.

5. Show the pdf plot for the results.

6. Show the Failure Rate vs. Time plot for the results.

7. Estimate the parameters using the rank regression on Y (RRY) analysis method (and using grouped ranks).

Solution

1. For the 2-parameter exponential distribution and for $\hat{T} = 100$ hours (first failure), the partial of the log-likelihood function, $\lambda$, becomes:

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{6} N_i \left[ \frac{1}{\lambda} - (T_i - 100) \right] = 0$$

$$\Rightarrow 7\left[ \frac{1}{\lambda} - (100 - 100) \right] + 5\left[ \frac{1}{\lambda} - (200 - 100) \right] + \ldots + 2\left[ \frac{1}{\lambda} - (600 - 100) \right] = 0$$

$$\Rightarrow \lambda = \frac{20}{3100} = 0.0065 \text{fr/hr}$$

2. Enter the data in a Weibull++ standard folio and calculate it as shown next.
3. On the Plot page of the folio, the exponential Probability plot will appear as shown next.
4. View the Reliability vs. Time plot.

5. View the pdf plot.
6. View the Failure Rate vs. Time plot.
Note that, as described at the beginning of this chapter, the failure rate for the exponential distribution is constant. Also note that the Failure Rate vs. Time plot does show values for times before the location parameter, $\gamma$, at 100 hours.

7. In the case of grouped data, one must be cautious when estimating the parameters using a rank regression method. This is because the median rank values are determined from the total number of failures observed by time $T$, where $i$ indicates the group number. In this example, the total number of groups is $N = 6$ and the total number of units is $N_T = 20$. Thus, the median rank values will be estimated for 20 units and for the total failed units ($N_F$) up to the $i^{th}$ group, for the $i^{th}$ rank value. The median ranks values can be found from rank tables or they can be estimated using ReliaSoft's Quick Statistical Reference tool.

For example, the median rank value of the fourth group will be the $17^{th}$ rank out of a sample size of twenty units (or 81.945%).

The following table is then constructed.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$N_F$</th>
<th>$N_{Fi}$</th>
<th>$T_i$</th>
<th>$F(T_i)$</th>
<th>$y_i$</th>
<th>$T_i^2$</th>
<th>$y_i^2$</th>
<th>$T_i y_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>7</td>
<td>100</td>
<td>0.32795</td>
<td>-0.3974</td>
<td>10000</td>
<td>0.1579</td>
<td>-39.7426</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>12</td>
<td>200</td>
<td>0.57374</td>
<td>-0.8527</td>
<td>40000</td>
<td>0.7271</td>
<td>-170.5402</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>15</td>
<td>300</td>
<td>0.72120</td>
<td>-1.2772</td>
<td>90000</td>
<td>1.6313</td>
<td>-383.1728</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>17</td>
<td>400</td>
<td>0.81945</td>
<td>-1.7117</td>
<td>160000</td>
<td>2.9301</td>
<td>-684.6990</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>18</td>
<td>500</td>
<td>0.86853</td>
<td>-2.0289</td>
<td>250000</td>
<td>4.1166</td>
<td>-1014.4731</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>20</td>
<td>600</td>
<td>0.96594</td>
<td>-3.3795</td>
<td>360000</td>
<td>11.4211</td>
<td>-2027.7085</td>
</tr>
<tr>
<td>$\sum$</td>
<td></td>
<td>2100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-4320.3362</td>
</tr>
</tbody>
</table>

Given the values in the table above, calculate $\hat{a}$ and $\hat{b}$.
The Exponential Distribution

\[
\hat{b} = \frac{\sum_{i=1}^{6} t_i y_i - \left(\sum_{i=1}^{6} t_i\right)\left(\sum_{i=1}^{6} y_i\right)/6}{\sum_{i=1}^{6} t_i^2 - \left(\sum_{i=1}^{6} t_i\right)^2/6}
\]

or:
\[
\hat{b} = \frac{-4320.3362 - (2100)(-9.6476)/6}{910,000 - (2100)^2/6}
\]

and:
\[
\hat{a} = \bar{y} - \hat{b}\bar{t} = \frac{\sum_{i=1}^{N} y_i}{N} - \hat{b}\frac{\sum_{i=1}^{N} t_i}{N}
\]

or:
\[
\hat{a} = \frac{-9.6476}{6} - \left(-0.005392\right)\frac{2100}{6} = 0.2793
\]

Therefore:
\[
\hat{\lambda} = -\hat{b} = -(0.005392) = 0.05392 \text{ failures/hour}
\]

and:
\[
\hat{\gamma} = \frac{\hat{a}}{\hat{\lambda}} = \frac{0.2793}{0.005392}
\]

or:
\[
\hat{\gamma} \approx 51.8 \text{ hours}
\]

Then:
\[
f(T) = (0.005392)e^{-0.005392(T-51.8)}
\]

Using Weibull++, the estimated parameters are:
\[
\hat{\lambda} = 0.0054 \text{ failures/hour}
\]
\[
\hat{\gamma} = 51.82 \text{ hours}
\]

The small difference in the values from Weibull++ is due to rounding. In the application, the calculations and the rank values are carried out up to the 15th decimal point.

**Using Auto Batch Run**

A number of leukemia patients were treated with either drug 6MP or a placebo, and the times in weeks until cancer symptoms returned were recorded. Analyze each treatment separately [21, p.175].
Create a new Weibull++ standard folio that's configured for grouped times-to-failure data with suspensions. In the first column, enter the number of patients. Whenever there are uncompleted tests, enter the number of patients who completed the test separately from the number of patients who did not (e.g., if 4 patients had symptoms return after 6 weeks and only 3 of them completed the test, then enter 1 in one row and 3 in another). In the second column enter F if the patients completed the test and S if they didn't. In the third column enter the time, and in the fourth column (Subset ID) specify whether the 6MP drug or a placebo was used.
Next, open the Batch Auto Run utility and select to separate the 6MP drug from the placebo, as shown next.
The software will create two data sheets, one for each subset ID, as shown next.

Calculate both data sheets using the 2-parameter exponential distribution and the MLE analysis method, then insert an additional plot and select to show the analysis results for both data sheets on that plot, which will appear as shown next.
The Exponential Distribution

Probability - Exponential

Reliability vs Time (Hr)

Time (Hr) vs Reliability (RU)

Example W/Step 1.00000. Failed at 1.00000. Failure of 1.00000.

Example W/Step 1.00000. Failed at 1.00000. Failure of 1.00000.

Example W/Step 1.00000. Failed at 1.00000. Failure of 1.00000.
Chapter 8

The Weibull Distribution

The Weibull distribution is one of the most widely used lifetime distributions in reliability engineering. It is a versatile distribution that can take on the characteristics of other types of distributions, based on the value of the shape parameter, $\beta$. This chapter provides a brief background on the Weibull distribution, presents and derives most of the applicable equations and presents examples calculated both manually and by using ReliaSoft's Weibull++.

Weibull Probability Density Function

The 3-Parameter Weibull

The 3-parameter Weibull pdf is given by:

$$f(t) = \frac{\beta}{\eta} \left( \frac{t - \gamma}{\eta} \right)^{\beta - 1} e^{-\left( \frac{t - \gamma}{\eta} \right)^{\beta}}$$

where:

- $f(t) \geq 0$, $t \geq 0$ or $\gamma$
- $\beta > 0$
- $\eta > 0$
- $-\infty < \gamma < +\infty$

and:

- $\eta =$ scale parameter, or characteristic life
- $\beta =$ shape parameter (or slope)
- $\gamma =$ location parameter (or failure free life)

The 2-Parameter Weibull

The 2-parameter Weibull pdf is obtained by setting $\gamma = 0$, and is given by:

$$f(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta - 1} e^{-\left( \frac{t}{\eta} \right)^{\beta}}$$

The 1-Parameter Weibull

The 1-parameter Weibull pdf is obtained by again setting $\gamma = 0$ and assuming $\beta = C =$ Constant assumed value or:

$$f(t) = \frac{C}{\eta} \left( \frac{t}{\eta} \right)^{C-1} e^{-\left( \frac{t}{\eta} \right)^C}$$

where the only unknown parameter is the scale parameter, $\eta$.

Note that in the formulation of the 1-parameter Weibull, we assume that the shape parameter $\beta$ is known a priori from past experience with identical or similar products. The advantage of doing this is that data sets with few or no failures can be analyzed.
Weibull Distribution Functions

The Mean or MTTF

The mean, $\overline{T}$, (also called MTTF) of the Weibull pdf is given by:

$$\overline{T} = \gamma + \eta \cdot \Gamma \left( \frac{1}{\beta} + 1 \right)$$

where

$$\Gamma \left( \frac{1}{\beta} + 1 \right)$$

is the gamma function evaluated at the value of:

$$\left( \frac{1}{\beta} + 1 \right)$$

The gamma function is defined as:

$$\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} \, dx$$

For the 2-parameter case, this can be reduced to:

$$\overline{T} = \eta \cdot \Gamma \left( \frac{1}{\beta} + 1 \right)$$

Note that some practitioners erroneously assume that $\eta$ is equal to the MTTF, $\overline{T}$. This is only true for the case of:

$$\beta = 1$$

or:

$$\overline{T} = \eta \cdot \Gamma \left( \frac{1}{1} + 1 \right)$$

$$= \eta \cdot \Gamma \left( \frac{1}{1} + 1 \right)$$

$$= \eta \cdot \Gamma (2)$$

$$= \eta \cdot 1$$

$$= \eta$$

The Median

The median, $\hat{T}$, of the Weibull distribution is given by:

$$\hat{T} = \gamma + \eta \left( \ln 2 \right)^{\frac{1}{\beta}}$$

The Mode

The mode, $\tilde{T}$, is given by:

$$\tilde{T} = \gamma + \eta \left( 1 - \frac{1}{\beta} \right)^{\frac{1}{\beta}}$$
The Weibull Distribution

The Standard Deviation

The standard deviation, $\sigma_T$, is given by:

$$\sigma_T = \eta \cdot \sqrt{\Gamma \left( \frac{2}{\beta} + 1 \right) - \Gamma \left( \frac{1}{\beta} + 1 \right)^2}$$

The Weibull Reliability Function

The equation for the 3-parameter Weibull cumulative density function, $cdf$, is given by:

$$F(t) = 1 - e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}$$

This is also referred to as unreliability and designated as $Q(t)$ by some authors.

Recalling that the reliability function of a distribution is simply one minus the $cdf$, the reliability function for the 3-parameter Weibull distribution is then given by:

$$R(t) = e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}$$

The Weibull Conditional Reliability Function

The 3-parameter Weibull conditional reliability function is given by:

$$R(t|T) = \frac{R(T + t)}{R(T)} = \frac{e^{-\left(\frac{T+t-\gamma}{\eta}\right)^\beta}}{e^{-\left(\frac{T-\gamma}{\eta}\right)^\beta}}$$

or:

$$R(t|T) = e^{-\left[\left(\frac{T+t-\gamma}{\eta}\right)^\beta - \left(\frac{T-\gamma}{\eta}\right)^\beta\right]}$$

These give the reliability for a new mission of $t$ duration, having already accumulated $T$ time of operation up to the start of this new mission, and the units are checked out to assure that they will start the next mission successfully. It is called conditional because you can calculate the reliability of a new mission based on the fact that the unit or units already accumulated hours of operation successfully.

The Weibull Reliable Life

The reliable life, $T_R$, of a unit for a specified reliability, $R$, starting the mission at age zero, is given by:

$$T_R = \gamma + \eta \cdot \{-\ln(R)\}^{\frac{1}{\beta}}$$

This is the life for which the unit/item will be functioning successfully with a reliability of $R$. If $R = 0.50$, then $T_R = \bar{T}$, the median life, or the life by which half of the units will survive.

The Weibull Failure Rate Function

The Weibull failure rate function, $\lambda(t)$, is given by:

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta}\right)^{\beta-1}$$
Characteristics of the Weibull Distribution

The Weibull distribution is widely used in reliability and life data analysis due to its versatility. Depending on the values of the parameters, the Weibull distribution can be used to model a variety of life behaviors. We will now examine how the values of the shape parameter, $\beta$, and the scale parameter, $\eta$, affect such distribution characteristics as the shape of the curve, the reliability and the failure rate. Note that in the rest of this section we will assume the most general form of the Weibull distribution, (i.e., the 3-parameter form). The appropriate substitutions to obtain the other forms, such as the 2-parameter form where $\gamma = 0$ or the 1-parameter form where $\beta = C = \text{constant}$, can easily be made.

Effects of the Shape Parameter, beta

The Weibull shape parameter, $\beta$, is also known as the slope. This is because the value of $\beta$ is equal to the slope of the regressed line in a probability plot. Different values of the shape parameter can have marked effects on the behavior of the distribution. In fact, some values of the shape parameter will cause the distribution equations to reduce to those of other distributions. For example, when $\beta = 1$, the pdf of the 3-parameter Weibull distribution reduces to that of the 2-parameter exponential distribution or:

$$f(t) = \frac{1}{\eta} e^{-\frac{t-\gamma}{\eta}}$$

where $\frac{1}{\eta} = \lambda$ = failure rate. The parameter $\beta$ is a pure number, (i.e., it is dimensionless). The following figure shows the effect of different values of the shape parameter, $\beta$, on the shape of the pdf. As you can see, the shape can take on a variety of forms based on the value of $\beta$.

For $0 < \beta < 1$:

- As $t \to 0$ or $\gamma$, $f(t) \to \infty$.
- As $t \to \infty$, $f(t) \to 0$. 

![Weibull pdf with 0 < \beta < 1, \beta = 1, and \beta < 1](image)
The Weibull Distribution

- \( f(t) \) decreases monotonically and is convex as it increases beyond the value of \( \gamma \).
- The mode is non-existent.

For \( \beta > 1 \):
- \( f(t) = 0 \) at \( t = 0 \) or \( \gamma \).
- \( f(t) \) increases as \( t \to \gamma \) (the mode) and decreases thereafter.
- For \( \beta < 2.6 \) the Weibull pdf is positively skewed (has a right tail), for \( 2.6 < \beta < 3.7 \) its coefficient of skewness approaches zero (no tail). Consequently, it may approximate the normal pdf, and for \( \beta > 3.7 \) it is negatively skewed (left tail). The way the value of \( \beta \) relates to the physical behavior of the items being modeled becomes more apparent when we observe how its different values affect the reliability and failure rate functions. Note that for \( \beta = 0.999 \), \( f(0) = \infty \), but for \( \beta = 1.001 \), \( f(0) = 0 \). This abrupt shift is what complicates MLE estimation when \( \beta \) is close to 1.

The Effect of beta on the cdf and Reliability Function

The above figure shows the effect of the value of \( \beta \) on the cdf, as manifested in the Weibull probability plot. It is easy to see why this parameter is sometimes referred to as the slope. Note that the models represented by the three lines all have the same value of \( \gamma \). The following figure shows the effects of these varied values of \( \beta \) on the reliability plot, which is a linear analog of the probability plot.
The Weibull Distribution

- $R(t)$ decreases sharply and monotonically for $0 < \beta < 1$ and is convex.
- For $\beta = 1$, $R(t)$ decreases monotonically but less sharply than for $0 < \beta < 1$ and is convex.
- For $\beta > 1$, $R(t)$ decreases as increases. As wear-out sets in, the curve goes through an inflection point and decreases sharply.

**The Effect of beta on the Weibull Failure Rate**

The value of $\hat{\beta}$ has a marked effect on the failure rate of the Weibull distribution and inferences can be drawn about a population’s failure characteristics just by considering whether the value of $\hat{\beta}$ is less than, equal to, or greater than one.
As indicated by the above figure, populations with \( \beta < 1 \) exhibit a failure rate that decreases with time, populations with \( \beta = 1 \) have a constant failure rate (consistent with the exponential distribution) and populations with \( \beta > 1 \) have a failure rate that increases with time. All three life stages of the bathtub curve can be modeled with the Weibull distribution and varying values of \( \beta \). The Weibull failure rate for \( 0 < \beta < 1 \) is unbounded at \( T = 0 \) or \( \gamma \). The failure rate, \( \lambda(t) \), decreases thereafter monotonically and is convex, approaching the value of zero as \( t \to \infty \) or \( \lambda(\infty) = 0 \). This behavior makes it suitable for representing the failure rate of units exhibiting early-type failures, for which the failure rate decreases with age. When encountering such behavior in a manufactured product, it may be indicative of problems in the production process, inadequate burn-in, substandard parts and components, or problems with packaging and shipping. For \( \beta = 1 \), \( \lambda(t) \) yields a constant value of \( \frac{1}{\eta} \) or:

\[
\lambda(t) = \lambda = \frac{1}{\eta}
\]

This makes it suitable for representing the failure rate of chance-type failures and the useful life period failure rate of units.

For \( \beta > 1 \), \( \lambda(t) \) increases as \( t \) increases and becomes suitable for representing the failure rate of units exhibiting wear-out type failures. For \( 1 < \beta < 2 \) the \( \lambda(t) \) curve is concave, consequently the failure rate increases at a decreasing rate as \( t \) increases.

For \( \beta = 2 \), there emerges a straight line relationship between \( \lambda(t) \) and \( t \), starting at a value of \( \lambda(t) = 0 \) at \( t = \gamma \), and increasing thereafter with a slope of \( \frac{2}{\eta^2} \). Consequently, the failure rate increases at a constant rate as \( t \) increases.

Furthermore, if \( \eta = 1 \) the slope becomes equal to 2, and when \( \gamma = 0 \), \( \lambda(t) \) becomes a straight line which passes through the origin with a slope of 2. Note that at \( \beta = 2 \), the Weibull distribution equations reduce to that of the Rayleigh distribution.

When \( \beta > 2 \), the \( \lambda(t) \) curve is convex, with its slope increasing as \( t \) increases. Consequently, the failure rate increases at an increasing rate as \( t \) increases, indicating wearout life.
Effects of the Scale Parameter, $\eta$

A change in the scale parameter $\eta$ has the same effect on the distribution as a change of the abscissa scale. Increasing the value of $\eta$ while holding $\beta$ constant has the effect of stretching out the pdf. Since the area under a pdf curve is a constant value of one, the “peak” of the pdf curve will also decrease with the increase of $\eta$, as indicated in the above figure.

- If $\eta$ is increased while $\beta$ and $\gamma$ are kept the same, the distribution gets stretched out to the right and its height decreases, while maintaining its shape and location.
- If $\eta$ is decreased while $\beta$ and $\gamma$ are kept the same, the distribution gets pushed in towards the left (i.e., towards its beginning or towards 0 or $\gamma$), and its height increases.
- $\eta$ has the same units as $t$, such as hours, miles, cycles, actuations, etc.
Effects of the Location Parameter, gamma

The location parameter, $\gamma$, as the name implies, locates the distribution along the abscissa. Changing the value of $\gamma$ has the effect of sliding the distribution and its associated function either to the right (if $\gamma > 0$) or to the left (if $\gamma < 0$).

- When $\gamma = 0$ the distribution starts at $t = 0$ or at the origin.
- If $\gamma > 0$ the distribution starts at the location $\gamma$ to the right of the origin.
- If $\gamma < 0$ the distribution starts at the location $\gamma$ to the left of the origin.
- $\gamma$ provides an estimate of the earliest time-to-failure of such units.
- The life period 0 to $+\gamma$ is a failure free operating period of such units.
- The parameter $\gamma$ may assume all values and provides an estimate of the earliest time a failure may be observed. A negative $\gamma$ may indicate that failures have occurred prior to the beginning of the test, namely during production, in storage, in transit, during checkout prior to the start of a mission, or prior to actual use.
- $\gamma$ has the same units as $t$, such as hours, miles, cycles, actuations, etc.
**Estimation of the Weibull Parameters**

The estimates of the parameters of the Weibull distribution can be found graphically via probability plotting paper, or analytically, using either least squares (rank regression) or maximum likelihood estimation (MLE).

**Probability Plotting**

One method of calculating the parameters of the Weibull distribution is by using probability plotting. To better illustrate this procedure, consider the following example from Kececioglu [20]. Assume that six identical units are being reliability tested at the same application and operation stress levels. All of these units fail during the test after operating the following number of hours: 93, 34, 16, 120, 53 and 75. Estimate the values of the parameters for a 2-parameter Weibull distribution and determine the reliability of the units at a time of 15 hours.

**Solution**

The steps for determining the parameters of the Weibull representing the data, using probability plotting, are outlined in the following instructions. First, rank the times-to-failure in ascending order as shown next.

<table>
<thead>
<tr>
<th>Time-to-failure, hours</th>
<th>Failure Order Number out of Sample Size of 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
</tr>
<tr>
<td>53</td>
<td>3</td>
</tr>
<tr>
<td>75</td>
<td>4</td>
</tr>
<tr>
<td>93</td>
<td>5</td>
</tr>
<tr>
<td>120</td>
<td>6</td>
</tr>
</tbody>
</table>

Obtain their median rank plotting positions. Median rank positions are used instead of other ranking methods because median ranks are at a specific confidence level (50%). Median ranks can be found tabulated in many reliability books. They can also be estimated using the following equation:

$$MR \sim \frac{i - 0.3}{\hat{N} + 0.4} \cdot 100$$

where \(i\) is the failure order number and \(\hat{N}\) is the total sample size. The exact median ranks are found in Weibull++ by solving:

$$\sum_{k=1}^{N} \left( \begin{array}{c} N \\ k \end{array} \right) MR^k (1 - MR)^{N-k} = 0.5 = 50\%$$

for \(MR\), where \(N\) is the sample size and \(i\) the order number. The times-to-failure, with their corresponding median ranks, are shown next.
<table>
<thead>
<tr>
<th>Time-to-failure, hours</th>
<th>Median Rank, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>10.91</td>
</tr>
<tr>
<td>34</td>
<td>26.44</td>
</tr>
<tr>
<td>53</td>
<td>42.14</td>
</tr>
<tr>
<td>75</td>
<td>57.86</td>
</tr>
<tr>
<td>93</td>
<td>73.56</td>
</tr>
<tr>
<td>120</td>
<td>89.1</td>
</tr>
</tbody>
</table>

On a Weibull probability paper, plot the times and their corresponding ranks. A sample of a Weibull probability paper is given in the following figure.

The points of the data in the example are shown in the figure below. Draw the best possible straight line through these points, as shown below, then obtain the slope of this line by drawing a line, parallel to the one just obtained, through the slope indicator. This value is the estimate of the shape parameter $\hat{\beta}$, in this case $\hat{\beta} = 1.4$. 
At the $Q(t) = 63.2\%$ ordinate point, draw a straight horizontal line until this line intersects the fitted straight line. Draw a vertical line through this intersection until it crosses the abscissa. The value at the intersection of the abscissa is the estimate of $\hat{\eta}$. For this case, $\hat{\eta} = 76$ hours. This is always at 63.2% since:

$$Q(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^\beta} = 1 - e^{-1} = 0.632 = 63.2\%$$

Now any reliability value for any mission time $t$ can be obtained. For example, the reliability for a mission of 15 hours, or any other time, can now be obtained either from the plot or analytically. To obtain the value from the plot, draw a vertical line from the abscissa, at hours, to the fitted line. Draw a horizontal line from this intersection to the ordinate and read $Q(t)$, in this case $Q(t) = 9.8\%$. Thus, $R(t) = 1 - Q(t) = 90.2\%$. This can also be obtained analytically from the Weibull reliability function since the estimates of both of the parameters are known or:

$$R(t = 15) = e^{-\left(\frac{15}{\eta}\right)^\beta} = e^{-\left(\frac{15}{76}\right)^{1.4}} = 90.2\%$$

**Probability Plotting for the Location Parameter, Gamma**

The third parameter of the Weibull distribution is utilized when the data do not fall on a straight line, but fall on either a concave up or down curve. The following statements can be made regarding the value of $\gamma$:

- **Case 1:** If the curve for MR versus $t_j$ is concave down and the curve for MR versus $(t_j - t_1)s$ concave up, then there exists a $\gamma$ such that $0 < \gamma < t_1$, or $\gamma$ has a positive value.

- **Case 2:** If the curves for MR versus $t_j$ and MR versus $(t_j - t_1)$ are both concave up, then there exists a negative $\gamma$ which will straighten out the curve of MR versus $t_j$.

- **Case 3:** If neither one of the previous two cases prevails, then either reject the Weibull as one capable of representing the data, or proceed with the multiple population (mixed Weibull) analysis. To obtain the location parameter, $\gamma$: 


• Subtract the same arbitrary value, $\gamma$, from all the times to failure and replot the data.
• If the initial curve is concave up, subtract a negative $\gamma$ from each failure time.
• If the initial curve is concave down, subtract a positive $\gamma$ from each failure time.
• Repeat until the data plots on an acceptable straight line.
• The value of $\gamma$ is the subtracted (positive or negative) value that places the points in an acceptable straight line.

The other two parameters are then obtained using the techniques previously described. Also, it is important to note that we used the term subtract a positive or negative gamma, where subtracting a negative gamma is equivalent to adding it. Note that when adjusting for gamma, the x-axis scale for the straight line becomes $(t - \gamma)$.

### Rank Regression on Y

Performing rank regression on Y requires that a straight line mathematically be fitted to a set of data points such that the sum of the squares of the vertical deviations from the points to the line is minimized. This is in essence the same methodology as the probability plotting method, except that we use the principle of least squares to determine the line through the points, as opposed to just eyeballing it. The first step is to bring our function into a linear form. For the two-parameter Weibull distribution, the (cumulative density function) is:

$$F(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^\beta}$$

Taking the natural logarithm of both sides of the equation yields:

$$\ln[1 - F(t)] = -\left(\frac{t}{\eta}\right)^\beta$$

$$\ln - \ln[1 - F(t)] = \beta \ln\left(\frac{t}{\eta}\right)$$

or:

$$\ln\{-\ln[1 - F(t)]\} = -\beta \ln(\eta) + \beta \ln(t)$$

Now let:

$$y = \ln\{-\ln[1 - F(t)]\}$$

and:

$$b = \beta$$

which results in the linear equation of:

$$y = a + bx$$

The least squares parameter estimation method (also known as regression analysis) was discussed in Parameter Estimation, and the following equations for regression on Y were derived:

$$\hat{a} = \bar{y} - \frac{\sum_{i=1}^{N} x_i \cdot \sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i^2} = \bar{y} - \hat{b} \bar{x}$$

and:

$$\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \left(\sum_{i=1}^{N} x_i\right) \cdot \left(\sum_{i=1}^{N} y_i\right)}{\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$$

In this case the equations for $y_i$ and $x_i$ are:

$$y_i = \ln\{-\ln[1 - F(t)]\}$$

$$x_i = t_i$$
The Weibull Distribution

\[ x_i = \ln(t_i) \]

The \( F(t_i) \) values are estimated from the median ranks.

Once \( \hat{\alpha} \) and \( \hat{\beta} \) are obtained, then \( \hat{\alpha} \) and \( \hat{\beta} \) can easily be obtained from previous equations.

**The Correlation Coefficient**

The correlation coefficient is defined as follows:

\[ \rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \]

where \( \sigma_{xy} \) = covariance of \( x \) and \( y \), \( \sigma_x \) = standard deviation of \( x \), and \( \sigma_y \) = standard deviation of \( y \). The estimator of \( \rho \) is the sample correlation coefficient, \( \tilde{\rho} \), given by:

\[ \tilde{\rho} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2}} \]

**RRY Example**

Consider the same data set from the probability plotting example given above (with six failures at 16, 34, 53, 75, 93 and 120 hours). Estimate the parameters and the correlation coefficient using rank regression on Y, assuming that the data follow the 2-parameter Weibull distribution.

**Solution**

Construct a table as shown next.

<table>
<thead>
<tr>
<th>( N )</th>
<th>( T_i )</th>
<th>( \ln(T_i) )</th>
<th>( F(T_i) )</th>
<th>( y_i )</th>
<th>( (\ln T_i)^2 )</th>
<th>( y_i^2 )</th>
<th>( (\ln T_i) y_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>2.7726</td>
<td>0.1091</td>
<td>-2.1583</td>
<td>7.6873</td>
<td>4.6582</td>
<td>-5.9840</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>3.5264</td>
<td>0.2645</td>
<td>-1.1802</td>
<td>12.4352</td>
<td>1.393</td>
<td>-4.1620</td>
</tr>
<tr>
<td>3</td>
<td>53</td>
<td>3.9703</td>
<td>0.4214</td>
<td>-0.6030</td>
<td>15.7632</td>
<td>0.3637</td>
<td>-2.3943</td>
</tr>
<tr>
<td>4</td>
<td>75</td>
<td>4.3175</td>
<td>0.5786</td>
<td>-0.146</td>
<td>18.6407</td>
<td>0.0213</td>
<td>-0.6303</td>
</tr>
<tr>
<td>5</td>
<td>93</td>
<td>4.5326</td>
<td>0.7355</td>
<td>0.2851</td>
<td>20.5445</td>
<td>0.0813</td>
<td>1.2923</td>
</tr>
<tr>
<td>6</td>
<td>120</td>
<td>4.7875</td>
<td>0.8909</td>
<td>0.7955</td>
<td>22.9201</td>
<td>0.6328</td>
<td>3.8083</td>
</tr>
<tr>
<td>( \sum )</td>
<td>23.9068</td>
<td>-3.007</td>
<td>97.9909</td>
<td>7.1502</td>
<td>-8.0699</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Utilizing the values from the table, calculate \( \hat{\alpha} \) and \( \hat{\beta} \) using the following equations:

\[ \hat{b} = \frac{\sum_{i=1}^{6} (\ln t_i)(y_i) - (\sum_{i=1}^{6} \ln t_i)(\sum_{i=1}^{6} y_i)/6}{\sum_{i=1}^{6} (\ln t_i)^2 - (\sum_{i=1}^{6} \ln t_i)^2/6} \]

\[ \hat{b} = \frac{-8.0699 - (23.9068)(-3.0070)/6}{97.9909 - (23.9068)^2/6} \]

or:

\[ \hat{b} = 1.4301 \]

and:
The Weibull Distribution

\[ \hat{a} = y - \hat{b}T = \frac{\sum_{i=1}^{N} y_i}{N} - \frac{\sum_{i=1}^{N} \ln t_i}{N} \]

or:

\[ \hat{a} = \frac{(-3.0070)}{6} - (1.4301) \frac{23.9068}{6} = -6.19935 \]

Therefore:

\[ \hat{\beta} = \hat{b} = 1.4301 \]

and:

\[ \hat{\eta} = e^{-\frac{\hat{b}}{\hat{\beta}}} = e^{\frac{(-6.19935)}{1.4301}} \]

or:

\[ \hat{\eta} = 76.318 \text{ hr} \]

The correlation coefficient can be estimated as:

\[ \hat{\rho} = 0.9956 \]

This example can be repeated in the Weibull++ software. The following plot shows the Weibull probability plot for the data set (with 90% two-sided confidence bounds).

If desired, the Weibull pdf representing the data set can be written as:

\[ f(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta-1} e^{-\left( \frac{t}{\eta} \right)^{\beta}} \]

or:
You can also plot this result in Weibull++, as shown next. From this point on, different results, reports and plots can be obtained.

### Rank Regression on X

Performing a rank regression on X is similar to the process for rank regression on Y, with the difference being that the horizontal deviations from the points to the line are minimized rather than the vertical. Again, the first task is to bring the reliability function into a linear form. This step is exactly the same as in the regression on Y analysis and all the equations apply in this case too. The derivation from the previous analysis begins on the least squares fit part, where in this case we treat as the dependent variable and as the independent variable. The best-fitting straight line to the data, for regression on X (see Parameter Estimation), is the straight line:

\[
x = \hat{a} + \hat{b}y
\]

The corresponding equations for \(\hat{a}\) and \(\hat{b}\) are:

\[
\hat{a} = x - \hat{b}y = \frac{\sum_{i=1}^{N} x_i}{N} - \hat{b} \frac{\sum_{i=1}^{N} y_i}{N}
\]

and:

\[
\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i}{N} \frac{\sum_{i=1}^{N} y_i}{N}}{\sum_{i=1}^{N} y_i^2 - \left( \frac{\sum_{i=1}^{N} y_i}{N} \right)^2}
\]
where:
\[ y_i = \ln \left\{ - \ln [1 - F(t_i)] \right\} \]
and:
\[ x_i = \ln(t_i) \]
and the \( F(t_i) \) values are again obtained from the median ranks.

Once \( \hat{\alpha} \) and \( \hat{\beta} \) are obtained, solve the linear equation for \( \beta \), which corresponds to:
\[ y = -\frac{\hat{\alpha}}{\hat{\beta}} + \frac{1}{\hat{\beta}} x \]
Solving for the parameters from above equations, we get:
\[ \alpha = -\frac{\hat{\alpha}}{\hat{\beta}} = -\beta \ln(\eta) \]
and
\[ b = \frac{1}{\hat{\beta}} = \beta \]
The correlation coefficient is evaluated as before.

**RRX Example**

Again using the same data set from the probability plotting and RRY examples (with six failures at 16, 34, 53, 75, 93 and 120 hours), calculate the parameters using rank regression on X.

**Solution**

The same table constructed above for the RRY example can also be applied for RRX.

Using the values from this table we get:
\[
\hat{\beta} = \frac{\sum_{i=1}^{6} (\ln T_i) y_i - \frac{6}{6} \ln T_i \cdot \frac{1}{6} \sum_{i=1}^{6} y_i}{\sum_{i=1}^{6} y_i^2 - \left( \frac{6}{6} \sum_{i=1}^{6} y_i \right)^2}
\]
\[
= \frac{-8.0699 - (23.9068)(-3.0070)/6}{7.1502 - (-3.0070)^2/6}
\]
or:
\[ \hat{\beta} = 0.6931 \]
and:
\[
\hat{\alpha} = \bar{x} - \hat{b} \bar{y} = \frac{\sum_{i=1}^{6} \ln T_i}{6} - \hat{b} \frac{1}{6} \sum_{i=1}^{6} y_i
\]
or:
\[ \hat{\alpha} = \frac{23.9068}{6} - (0.6931)(-3.0070)/6 = 4.3318 \]
Therefore:
\[ \hat{\beta} = \frac{1}{\hat{\beta}} = \frac{1}{0.6931} = 1.4428 \]
and:
\[ \hat{\eta} = e^{\frac{\hat{\alpha}}{\hat{\beta}}} = e^{\frac{4.3318}{1.4428}} = 76.0811 \text{ hr} \]
The correlation coefficient is:
\[ \hat{\rho} = 0.9956 \]

The results and the associated graph using Weibull++ are shown next. Note that the slight variation in the results is due to the number of significant figures used in the estimation of the median ranks. Weibull++ by default uses double precision accuracy when computing the median ranks.

### 3-Parameter Weibull Regression

When the MR versus data points plotted on the Weibull probability paper do not fall on a satisfactory straight line and the points fall on a curve, then a location parameter, \( \gamma \), might exist which may straighten out these points. The goal in this case is to fit a curve, instead of a line, through the data points using nonlinear regression. The Gauss-Newton method can be used to solve for the parameters, \( \hat{\beta} \), \( \hat{\eta} \) and \( \hat{\gamma} \), by performing a Taylor series expansion on \( F(t; \beta, \eta, \gamma) \). Then the nonlinear model is approximated with linear terms and ordinary least squares are employed to estimate the parameters. This procedure is iterated until a satisfactory solution is reached. (Note that other shapes, particularly \( S \) shapes, might suggest the existence of more than one population. In these cases, the multiple population mixed Weibull distribution, may be more appropriate.)

When you use the 3-parameter Weibull distribution, Weibull++ calculates the value of \( \gamma \) by utilizing an optimized Nelder-Mead algorithm and adjusts the points by this value of \( \gamma \) such that they fall on a straight line, and then plots both the adjusted and the original unadjusted points. To draw a curve through the original unadjusted points, if so desired, select Weibull 3P Line Unadjusted for Gamma from the Show Plot Line submenu under the Plot Options menu. The returned estimations of the parameters are the same when selecting RRX or RRY. To display the unadjusted data points and line along with the adjusted data points and line, select Show/Hide Items under the Plot Options menu and include the unadjusted data points and line as follows:
The results and the associated graph for the previous example using the 3-parameter Weibull case are shown next:
The Weibull Distribution

Maximum Likelihood Estimation

As outlined in Parameter Estimation, maximum likelihood estimation works by developing a likelihood function based on the available data and finding the values of the parameter estimates that maximize the likelihood function. This can be achieved by using iterative methods to determine the parameter estimate values that maximize the likelihood function, but this can be rather difficult and time-consuming, particularly when dealing with the three-parameter distribution. Another method of finding the parameter estimates involves taking the partial derivatives of the likelihood function with respect to the parameters, setting the resulting equations equal to zero and solving simultaneously to determine the values of the parameter estimates. (Note that MLE asymptotic properties do not hold when estimating using MLE, as discussed in Meeker and Escobar [27].) The log-likelihood functions and associated partial derivatives used to determine maximum likelihood estimates for the Weibull distribution are covered in Appendix D.

MLE Example

One last time, use the same data set from the probability plotting, RRY and RRX examples (with six failures at 16, 34, 53, 75, 93 and 120 hours) and calculate the parameters using MLE.

Solution

In this case, we have non-grouped data with no suspensions or intervals, (i.e., complete data). The equations for the partial derivatives of the log-likelihood function are derived in an appendix and given next:

\[
\frac{\partial \Lambda}{\partial \beta} = \frac{6}{\beta} + \sum_{i=1}^{6} \ln \left( \frac{T_i}{\eta} \right) - \sum_{i=1}^{6} \left( \frac{T_i}{\eta} \right)^{\beta} \ln \left( \frac{T_i}{\eta} \right) = 0
\]

And:
\[ \frac{\partial \Lambda}{\partial \eta} = -\frac{\beta}{\eta} \cdot 6 + \frac{\beta}{\eta} \sum_{i=1}^{6} \left( \frac{T_i}{\eta} \right)^{\beta} = 0 \]

Solving the above equations simultaneously we get:

\[ \hat{\beta} = 1.933, \hat{\eta} = 73.526 \]

The variance/covariance matrix is found to be:

\[ \begin{bmatrix} \text{Var}(\hat{\beta}) &=& 0.4211 & \\ \text{Cov}(\hat{\beta}, \hat{\eta}) &=& 3.272 \\ \text{Cov}(\hat{\beta}, \hat{\eta}) &=& 3.272 & \text{Var}(\hat{\eta}) &=& 266.646 \end{bmatrix} \]

The results and the associated plot using Weibull++ (MLE) are shown next.

You can view the variance/covariance matrix directly by clicking the Analysis Summary table in the control panel.
Note that the decimal accuracy displayed and used is based on your individual Application Setup.
Unbiased MLE $\hat{\beta}$

It is well known that the MLE $\hat{\beta}$ is biased. The biasness will affect the accuracy of reliability prediction, especially when the number of failures are small. Weibull++ provides a simple way to correct the bias of MLE $\hat{\beta}$.

When there are no right censored observations in the data, the following equation provided by Hirose [39] is used to calculated the unbiased $\hat{\beta}$.

$$\hat{\beta}_U = \frac{\beta}{1.0115 + \frac{1.278}{r} + \frac{1.001}{\sqrt{r}} + \frac{20.35}{r^2} - \frac{40.98}{r^4}}$$

where $r$ is the number of failures.

When there are right censored observations in the data, the following equation provided by Ross [40] is used to calculated the unbiased $\hat{\beta}$.

$$\hat{\beta}_U = \frac{\beta}{1 + \frac{1.37}{r - 1.92 \sqrt{\mu}} \sqrt{\mu}}$$

where $\mu$ is the number of observations.

The software will use the above equations only when there are more than two failures in the data set.

For an example on how you might correct biased estimates, see also:

Unbiasing Parameters in Weibull++[1]

Fisher Matrix Confidence Bounds

One of the methods used by the application in estimating the different types of confidence bounds for Weibull data, the Fisher matrix method, is presented in this section. The complete derivations were presented in detail (for a general function) in Confidence Bounds.
The Weibull Distribution

Bound on the Parameters

One of the properties of maximum likelihood estimators is that they are asymptotically normal, meaning that for large samples they are normally distributed. Additionally, since both the shape parameter estimate, $\hat{\beta}$, and the scale parameter estimate, $\hat{\eta}$, must be positive, thus $\ln\hat{\beta}$ and $\ln\hat{\eta}$ are treated as being normally distributed as well. The lower and upper bounds on the parameters are estimated from Nelson [30]:

$$\beta_U = \hat{\beta} \cdot e^{\frac{K_\alpha \sqrt{\text{Var}(\hat{\beta})}}{\hat{\beta}}}$$ (upper bound)

$$\beta_L = \frac{\hat{\beta}}{e^{\frac{K_\alpha \sqrt{\text{Var}(\hat{\beta})}}{\hat{\beta}}}}$$ (lower bound)

and:

$$\eta_U = \hat{\eta} \cdot e^{\frac{K_\alpha \sqrt{\text{Var}(\hat{\eta})}}{\hat{\eta}}}$$ (upper bound)

$$\eta_L = \frac{\hat{\eta}}{e^{\frac{K_\alpha \sqrt{\text{Var}(\hat{\eta})}}{\hat{\eta}}}}$$ (lower bound)

where $K_\alpha$ is defined by:

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{K_\alpha}^{\infty} e^{-\frac{t^2}{2}} dt = 1 - \Phi(K_\alpha)$$

If $d$ is the confidence level, then $\alpha = \frac{1 - d}{2}$ for the two-sided bounds and $\alpha = 1 - d$ for the one-sided bounds.

The variances and covariances of $\hat{\beta}$ and $\hat{\eta}$ are estimated from the inverse local Fisher matrix, as follows:

$$
\begin{pmatrix}
V\text{ar} (\hat{\beta}) & C\text{ov} (\hat{\beta}, \hat{\eta}) \\
C\text{ov} (\hat{\beta}, \hat{\eta}) & V\text{ar} (\hat{\eta})
\end{pmatrix}
= \left( \begin{pmatrix}
\frac{\partial^2 L}{\partial \hat{\beta}^2} & -\frac{\partial^2 L}{\partial \hat{\beta} \partial \hat{\eta}} \\
-\frac{\partial^2 L}{\partial \hat{\beta} \partial \hat{\eta}} & -\frac{\partial^2 L}{\partial \hat{\eta}^2}
\end{pmatrix} \right)^{-1}
$$

**Fisher Matrix Confidence Bounds and Regression Analysis**

Note that the variance and covariance of the parameters are obtained from the inverse Fisher information matrix as described in this section. The local Fisher information matrix is obtained from the second partials of the likelihood function, by substituting the solved parameter estimates into the particular functions. This method is based on maximum likelihood theory and is derived from the fact that the parameter estimates were computed using maximum likelihood estimation methods. When one uses least squares or regression analysis for the parameter estimates, this methodology is theoretically then not applicable. However, if one assumes that the variance and covariance of the parameters will be similar (One also assumes similar properties for both estimators.) regardless of the underlying solution method, then the above methodology can also be used in regression analysis.

The Fisher matrix is one of the methodologies that Weibull++ uses for both MLE and regression analysis. Specifically, Weibull++ uses the likelihood function and computes the local Fisher information matrix based on the estimates of the parameters and the current data. This gives consistent confidence bounds regardless of the underlying method of solution, (i.e., MLE or regression). In addition, Weibull++ checks this assumption and proceeds with it if it considers it to be acceptable. In some instances, Weibull++ will prompt you with an "Unable to Compute Confidence Bounds" message when using regression analysis. This is an indication that these assumptions were violated.
Bounds on Reliability

The bounds on reliability can easily be derived by first looking at the general extreme value distribution (EVD). Its reliability function is given by:

\[ R(t) = e^{-e^{\left(\frac{t-\mu}{\sigma}\right)}} \]

By transforming \( t = \ln t \) and converting \( p = \ln(\eta) \), \( p_2 = \frac{1}{\beta} \), the above equation becomes the Weibull reliability function:

\[ R(t) = e^{-e^{\ln(\eta) \beta}} = e^{-\left(\frac{t}{\eta}\right)^\beta} \]

with:

\[ R(T) = e^{-e^{\ln(T) \beta}} \]

set:

\[ u = \beta (\ln t - \ln \eta) \]

The reliability function now becomes:

\[ R(T) = e^{-e^u} \]

The next step is to find the upper and lower bounds on \( u \). Using the equations derived in Confidence Bounds, the bounds on are then estimated from Nelson [30]:

\[ u_U = \hat{u} + K_\alpha \sqrt{Var(\hat{u})} \]
\[ u_L = \hat{u} - K_\alpha \sqrt{Var(\hat{u})} \]

where:

\[ Var(\hat{u}) = \left(\frac{\partial u}{\partial \beta}\right)^2 Var(\hat{\beta}) + \left(\frac{\partial u}{\partial \eta}\right)^2 Var(\hat{\eta}) + 2 \left(\frac{\partial u}{\partial \beta}\right) \left(\frac{\partial u}{\partial \eta}\right) Cov(\hat{\beta}, \hat{\eta}) \]

or:

\[ Var(\hat{u}) = \frac{\hat{u}^2}{\beta^2} Var(\hat{\beta}) + \frac{\hat{\beta}^2}{\hat{\eta}^2} Var(\hat{\eta}) - \left(\frac{2u}{\hat{\eta}}\right) Cov(\hat{\beta}, \hat{\eta}) \]

The upper and lower bounds on reliability are:

\[ R_U = e^{-e^{u_U}} \text{ (upper bound)} \]
\[ R_L = e^{-e^{u_L}} \text{ (lower bound)} \]

Other Weibull Forms

Weibull++ makes the following assumptions/substitutions when using the three-parameter or one-parameter forms:

- For the 3-parameter case, substitute \( t = \ln(t - \gamma) \) and by definition \( \gamma < t \), instead of \( \ln t \). (Note that this is an approximation since it eliminates the third parameter and assumes that \( Var(\gamma) = 0 \))

- For the 1-parameter, \( Var(\hat{\beta}) = 0 \) thus:

\[ Var(\hat{u}) = \left(\frac{\partial u}{\partial \eta}\right)^2 Var(\hat{\eta}) = \left(\frac{\beta}{\eta}\right)^2 Var(\hat{\eta}) \]

Also note that the time axis (x-axis) in the three-parameter Weibull plot in Weibull++ is not but \( t - \gamma \). This means that one must be cautious when obtaining confidence bounds from the plot. If one desires to estimate the confidence bounds on reliability for a given time \( t_0 \) from the adjusted plotted line, then these bounds should be obtained for a \( t_0 - \gamma \) entry on the time axis.
Bounds on Time

The bounds around the time estimate or reliable life estimate, for a given Weibull percentile (unreliability), are estimated by first solving the reliability equation with respect to time, as discussed in Lloyd and Lipow [24] and in Nelson [30]:

\[ \ln R = -\left( \frac{t}{\eta} \right)^\beta \]
\[ \ln(- \ln R) = \beta \ln \left( \frac{t}{\eta} \right) \]
\[ \ln(- \ln R) = \beta(\ln t - \ln \eta) \]

or:

\[ u = \frac{1}{\beta} \ln(- \ln R) + \ln \eta \]

where \( u = \ln t \).

The upper and lower bounds on are estimated from:

\[ u_U = \hat{u} + K_{\alpha} \sqrt{Var(\hat{u})} \]
\[ u_L = \hat{u} - K_{\alpha} \sqrt{Var(\hat{u})} \]

where:

\[ Var(\hat{u}) = \left( \frac{\partial u}{\partial \beta} \right)^2 Var(\hat{\beta}) + \left( \frac{\partial u}{\partial \eta} \right)^2 Var(\hat{\eta}) + 2 \left( \frac{\partial u}{\partial \beta} \right) \left( \frac{\partial u}{\partial \eta} \right) Cov(\hat{\beta}, \hat{\eta}) \]

or:

\[ Var(\hat{u}) = \frac{1}{\beta^4} \left[ \ln(- \ln R) \right]^2 Var(\hat{\beta}) + \frac{1}{\eta^2} Var(\hat{\eta}) + 2 \left( -\frac{1}{\beta^2} \right) \left( \frac{\ln(- \ln R)}{\eta} \right) Cov(\hat{\beta}, \hat{\eta}) \]

The upper and lower bounds are then found by:

\[ T_U = e^{u_U} \text{ (upper bound)} \]
\[ T_L = e^{u_L} \text{ (lower bound)} \]

Likelihood Ratio Confidence Bounds

As covered in Confidence Bounds, the likelihood confidence bounds are calculated by finding values for \( \theta_1 \) and \( \theta_2 \) that satisfy:

\[ -2 \cdot \ln \left( \frac{L(\theta_1, \theta_2)}{L(\hat{\theta}_1, \hat{\theta}_2)} \right) = \chi^2_{\alpha,1} \]

This equation can be rewritten as:

\[ L(\theta_1, \theta_2) = L(\hat{\theta}_1, \hat{\theta}_2) \cdot e^{-\chi^2_{\alpha,1}/2} \]

For complete data, the likelihood function for the Weibull distribution is given by:

\[ L(\beta, \eta) = \prod_{i=1}^{N} f(x_i; \beta, \eta) = \prod_{i=1}^{N} \frac{\beta}{\eta} \cdot \left( \frac{x_i}{\eta} \right)^{\beta-1} \cdot e^{-\left( \frac{x_i}{\eta} \right)^\beta} \]

For a given value of \( \alpha \), values for \( \beta \) and \( \eta \) can be found which represent the maximum and minimum values that satisfy the above equation. These represent the confidence bounds for the parameters at a confidence level \( \delta \), where \( \alpha = \delta \) for two-sided bounds and \( \alpha = 2\delta - 1 \) for one-sided.
Similarly, the bounds on time and reliability can be found by substituting the Weibull reliability equation into the likelihood function so that it is in terms of $\hat{\beta}$ and time or reliability, as discussed in Confidence Bounds. The likelihood ratio equation used to solve for bounds on time (Type 1) is:

$$L(\beta, t) = \prod_{i=1}^{N} \left( \frac{\beta}{t} \cdot \left( \frac{x_i}{t^{\frac{1}{\beta}}} \right)^{\beta-1} \cdot \exp \left[ - \left( \frac{x_i}{t^{\frac{1}{\beta}}} \right)^{\beta} \right] \right)$$

The likelihood ratio equation used to solve for bounds on reliability (Type 2) is:

$$L(\beta, R) = \prod_{i=1}^{N} \left( \frac{\beta}{(-\ln(R))^{\beta}} \cdot \left( \frac{x_i}{(-\ln(R))^{\frac{1}{\beta}}} \right)^{\beta-1} \cdot \exp \left[ - \left( \frac{x_i}{(-\ln(R))^{\frac{1}{\beta}}} \right)^{\beta} \right] \right)$$

## Bayesian Confidence Bounds

### Bounds on Parameters

Bayesian Bounds use non-informative prior distributions for both parameters. From Confidence Bounds, we know that if the prior distribution of $\eta$ and $\beta$ are independent, the posterior joint distribution of $\eta$ and $\beta$ can be written as:

$$f(\eta, \beta | Data) = \frac{L(Data | \eta, \beta) \varphi(\eta) \varphi(\beta)}{\int_{0}^{\infty} \int_{0}^{\infty} L(Data | \eta, \beta) \varphi(\eta) \varphi(\beta) d\eta d\beta}$$

The marginal distribution of $\eta$ is:

$$f(\eta | Data) = \int_{0}^{\infty} f(\eta, \beta | Data) d\beta = \frac{\int_{0}^{\infty} \int_{0}^{\infty} L(Data | \eta, \beta) \varphi(\eta) \varphi(\beta) d\eta d\beta}{\int_{0}^{\infty} \int_{0}^{\infty} L(Data | \eta, \beta) \varphi(\eta) \varphi(\beta) d\eta d\beta}$$

where: $\varphi(\beta) = \frac{1}{\beta}$ is the non-informative prior of $\beta$. $\varphi(\eta) = \frac{1}{\eta}$ is the non-informative prior of $\eta$. Using these non-informative prior distributions, $f(\eta | Data)$ can be rewritten as:

$$f(\eta | Data) = \frac{\int_{0}^{\infty} L(Data | \eta, \beta) \frac{1}{\beta} d\beta}{\int_{0}^{\infty} \int_{0}^{\infty} L(Data | \eta, \beta) \frac{1}{\beta} \frac{1}{\eta} d\eta d\beta}$$

The one-sided upper bounds of $\eta$ is:

$$CL = P(\eta \leq \eta_U) = \int_{0}^{\eta_U} f(\eta | Data) d\eta$$

The one-sided lower bounds of $\eta$ is:

$$1 - CL = P(\eta \leq \eta_L) = \int_{0}^{\eta_L} f(\eta | Data) d\eta$$

The two-sided bounds of $\eta$ is:

$$CL = P(\eta_L \leq \eta \leq \eta_U) = \int_{\eta_L}^{\eta_U} f(\eta | Data) d\eta$$

Same method is used to obtain the bounds of $\beta$. 
### Bounds on Reliability

\[ CL = \Pr(R \leq R_U) = \Pr(\eta \leq T \exp\left(-\frac{\ln(-\ln R_U)}{\beta}\right)) \]

From the posterior distribution of \( \eta \), we have:

\[ CL = \frac{\int_0^\infty \int_0^\infty T \exp\left(-\frac{\ln(-\ln R_U)}{\beta}\right) L(\beta, \eta) \frac{1}{\beta} d\eta d\beta}{\int_0^\infty \int_0^\infty L(\beta, \eta) \frac{1}{\beta} d\eta d\beta} \]

The above equation is solved numerically for \( R_U \). The same method can be used to calculate the one-sided lower bounds and two-sided bounds on reliability.

### Bounds on Time

From Confidence Bounds, we know that:

\[ CL = \Pr(T \leq T_U) = \Pr(\eta \leq T_U \exp\left(-\frac{\ln(-\ln R)}{\beta}\right)) \]

From the posterior distribution of \( \eta \), we have:

\[ CL = \frac{\int_0^\infty \int_0^\infty T_U \exp\left(-\frac{\ln(-\ln R)}{\beta}\right) L(\beta, \eta) \frac{1}{\beta} d\eta d\beta}{\int_0^\infty \int_0^\infty L(\beta, \eta) \frac{1}{\beta} d\eta d\beta} \]

The above equation is solved numerically for \( T_U \). The same method can be applied to calculate one-sided lower bounds and two-sided bounds on time.

### Bayesian-Weibull Analysis

The Bayesian methods presented next are for the 2-parameter Weibull distribution. Bayesian concepts were introduced in Parameter Estimation. This model considers prior knowledge on the shape (\( \beta \)) parameter of the Weibull distribution when it is chosen to be fitted to a given set of data. There are many practical applications for this model, particularly when dealing with small sample sizes and some prior knowledge for the shape parameter is available. For example, when a test is performed, there is often a good understanding about the behavior of the failure mode under investigation, primarily through historical data. At the same time, most reliability tests are performed on a limited number of samples. Under these conditions, it would be very useful to use this prior knowledge with the goal of making more accurate predictions. A common approach for such scenarios is to use the 1-parameter Weibull distribution, but this approach is too deterministic, too absolute you may say (and you would be right). The Bayesian-Weibull model in Weibull++ (which is actually a true "WeiBayes" model, unlike the 1-parameter Weibull that is commonly referred to as such) offers an alternative to the 1-parameter Weibull, by including the variation and uncertainty that might have been observed in the past on the shape parameter. Applying Bayes's rule on the 2-parameter Weibull distribution and assuming the prior distributions of \( \beta \) and \( \eta \) are independent, we obtain the following posterior pdf:

\[ f(\beta, \eta|Data) = \frac{L(\beta, \eta)\varphi(\beta)\varphi(\eta)}{\int_0^\infty \int_0^\infty L(\beta, \eta)\varphi(\beta)\varphi(\eta) d\eta d\beta} \]

In this model, \( \eta \)'s assumed to follow a noninformative prior distribution with the density function \( \varphi(\eta) = \frac{1}{\eta} \). This is called Jeffrey's prior, and is obtained by performing a logarithmic transformation on \( \eta \). Specifically, since \( \eta \)'s always positive, we can assume that \( \ln(\eta) \) follows a uniform distribution, Applying Jeffrey's rule as given in Gelman et al. [9] which says "in general, an approximate non-informative prior is taken proportional to the square
The Weibull Distribution

root of Fisher’s information,” yields \( \varphi(\eta) = \frac{1}{\eta} \).

The prior distribution of \( \tilde{\beta} \), denoted as \( \varphi(\tilde{\beta}) \), can be selected from the following distributions: normal, lognormal, exponential and uniform. The procedure of performing a Bayesian-Weibull analysis is as follows:

- Collect the times-to-failure data.
- Specify a prior distribution for \( \tilde{\beta} \)(the prior for \( \tilde{\eta} \)s assumed to be \( 1/\tilde{\beta} \)).
- Obtain the posterior pdf from the above equation.

In other words, a distribution (the posterior pdf) is obtained, rather than a point estimate as in classical statistics (i.e., as in the parameter estimation methods described previously in this chapter). Therefore, if a point estimate needs to be reported, a point of the posterior pdf needs to be calculated. Typical points of the posterior distribution used are the mean (expected value) or median. In Weibull++, both options are available and can be chosen from the Analysis page, under the Results As area, as shown next.

The expected value of \( \tilde{\beta} \) is obtained by:

\[
E(\beta) = \int_0^\infty \int_0^\infty \beta \cdot f(\beta, \eta|Data) d\beta d\eta
\]

Similarly, the expected value of \( \tilde{\eta} \) is obtained by:

\[
E(\eta) = \int_0^\infty \int_0^\infty \eta \cdot f(\beta, \eta|Data) d\beta d\eta
\]

The median points are obtained by solving the following equations for \( \tilde{\beta} \) and \( \tilde{\eta} \) respectively:

\[
\int_0^{\tilde{\beta}} \int_0^{\infty} f(\beta, \eta|Data) d\beta d\eta = 0.5
\]

and:

\[
\int_0^{\tilde{\eta}} \int_0^{\infty} f(\beta, \eta|Data) d\beta d\eta = 0.5
\]
Of course, other points of the posterior distribution can be calculated as well. For example, one may want to calculate the 10th percentile of the joint posterior distribution (w.r.t. one of the parameters). The procedure for obtaining other points of the posterior distribution is similar to the one for obtaining the median values, where instead of 0.5 the percentage of interest is given. This procedure actually provides the confidence bounds on the parameters, which in the Bayesian framework are called “Credible Bounds.” However, since the engineering interpretation is the same, and to avoid confusion, we refer to them as confidence bounds in this reference and in Weibull++.

### Posterior Distributions for Functions of Parameters

As explained in Parameter Estimation, in Bayesian analysis, all the functions of the parameters are distributed. In other words, a posterior distribution is obtained for functions such as reliability and failure rate, instead of point estimate as in classical statistics. Therefore, in order to obtain a point estimate for these functions, a point on the posterior distributions needs to be calculated. Again, the expected value (mean) or median value are used. It is important to note that the Median value is preferable and is the default in Weibull++. This is because the Median value always corresponds to the 50th percentile of the distribution. On the other hand, the Mean is not a fixed point on the distribution, which could cause issues, especially when comparing results across different data sets.

#### pdf of the Times-to-Failure

The posterior distribution of the failure time $T$ is given by:

$$f(T|Data) = \int_0^\infty \int_0^\infty f(T, \beta, \eta) f(\beta, \eta|Data) d\eta d\beta$$

where:

$$f(T, \beta, \eta) = \frac{\beta}{\eta} \left( \frac{T}{\eta} \right)^{\beta-1} e^{-\left( \frac{T}{\eta} \right)^\beta}$$

For the pdf of the times-to-failure, only the expected value is calculated and reported in Weibull++.

#### Reliability

In order to calculate the median value of the reliability function, we first need to obtain posterior pdf of the reliability. Since $R(T)$ is a function of $\beta$, the density functions of $\beta$ and $R(T)$ have the following relationship:

$$f(R|Data, T)dR = f(\beta|Data)d\beta$$

$$= (\int_0^\infty f(\beta, \eta|Data)d\eta) d\beta$$

$$= \frac{1}{\int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta, \varphi(\eta)) d\eta d\beta} d\beta$$

The median value of the reliability is obtained by solving the following equation w.r.t. $R$:

$$\int_0^R f(R|Data, T)dR = 0.5$$

The expected value of the reliability at time $T$ is given by:

$$R(T|Data) = \int_0^\infty \int_0^\infty R(T, \beta, \eta) f(\beta, \eta|Data)d\eta d\beta$$

where:

$$R(T, \beta, \eta) = e^{-\left( \frac{T}{\eta} \right)^\beta}$$

#### Failure Rate

The failure rate at time is given by:
The Weibull Distribution

\[ \lambda(T | Data) = \frac{\int_0^\infty \int_0^\infty \lambda(T, \beta, \eta) L(\beta, \eta) \varphi(\eta) \varphi(\beta) d\eta d\beta}{\int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\eta) \varphi(\beta) d\eta d\beta} \]

where:

\[ \lambda(T, \beta, \eta) = \frac{\beta}{\eta} \left( \frac{T}{\eta} \right)^{\beta - 1} \]

Bounds on Reliability for Bayesian-Weibull

The confidence bounds calculation under the Bayesian-Weibull analysis is very similar to the Bayesian Confidence Bounds method described in the previous section, with the exception that in the case of the Bayesian-Weibull Analysis the specified prior of \( \beta \)s is considered instead of an non-informative prior. The Bayesian one-sided upper bound estimate for \( R(T) \) is given by:

\[ \int_0^{R_U(T)} f(R | Data, t) dR = CL \]

Using the posterior distribution, the following is obtained:

\[ \int_0^\infty \int_0^\infty \frac{\ln(-\ln R_U)}{\beta} L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta \]

\[ \left/ \int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta \right. = CL \]

The above equation can be solved for \( R_U(t) \). The Bayesian one-sided lower bound estimate for \( R(t) \) is given by:

\[ \int_0^{R_L(t)} f(R | Data, t) dR = 1 - CL \]

Using the posterior distribution, the following is obtained:

\[ \int_0^\infty \int_0^\infty \frac{T \exp(-\ln(-\ln R_L))}{\beta} L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta \]

\[ \left/ \int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta \right. = 1 - CL \]

The above equation can be solved for \( R_L(t) \). The Bayesian two-sided bounds estimate for \( R(t) \) is given by:

\[ \int_0^{R_U(t)} f(R | Data, t) dR = CL \]

which is equivalent to:

\[ \int_0^{R_L(t)} f(R | Data, t) dR = (1 + CL)/2 \]

and:

\[ \int_0^{R_U(t)} f(R | Data, T) dR = (1 - CL)/2 \]

Using the same method for one-sided bounds, \( R_U(t) \) and \( R_L(t) \) can be computed.
Bounds on Time for Bayesian-Weibull

Following the same procedure described for bounds on Reliability, the bounds of time $t$ can be calculated, given $R$.

The Bayesian one-sided upper bound estimate for $t(R)$ is given by:

$$\int_0^{T_U(R)} f(T | Data, R) dT = CL$$

Using the posterior distribution, the following is obtained:

$$\frac{\int_0^\infty \int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta}{\int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta} = CL$$

The above equation can be solved for $T_U(R)$. The Bayesian one-sided lower bound estimate for $T(R)$ is given by:

$$\int_0^{T_L(R)} f(T | Data, R) dT = 1 - CL$$

or:

$$\frac{\int_0^\infty \int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta}{\int_0^\infty \int_0^\infty L(\beta, \eta) \varphi(\beta) \varphi(\eta) d\eta d\beta} = CL$$

The above equation can be solved for $T_L(R)$. The Bayesian two-sided lower bounds estimate for $T(R)$ is:

$$\int_{T_{L}(R)}^{T_{U}(R)} f(T | Data, R) dT = CL$$

which is equivalent to:

$$\int_0^{T_U(R)} f(T | Data, R) dT = (1 + CL)/2$$

and:

$$\int_0^{T_L(R)} f(T | Data, R) dT = (1 - CL)/2$$

Bayesian-Weibull Example

A manufacturer has tested prototypes of a modified product. The test was terminated at 2,000 hours, with only 2 failures observed from a sample size of 18. The following table shows the data.

<table>
<thead>
<tr>
<th>Number of State</th>
<th>State of F or S</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>1180</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>1842</td>
</tr>
<tr>
<td>16</td>
<td>S</td>
<td>2000</td>
</tr>
</tbody>
</table>

Because of the lack of failure data in the prototype testing, the manufacturer decided to use information gathered from prior tests on this product to increase the confidence in the results of the prototype testing. This decision was made because failure analysis indicated that the failure mode of the two failures is the same as the one that was observed in previous tests. In other words, it is expected that the shape of the distribution (beta) hasn’t changed, but hopefully the scale (eta) has, indicating longer life. The 2-parameter Weibull distribution was used to model all prior tests results. The estimated beta ($\beta$) parameters of the prior test results are as follows:
The Weibull Distribution

Betas Obtained for Similar Mode

<table>
<thead>
<tr>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7</td>
</tr>
<tr>
<td>2.1</td>
</tr>
<tr>
<td>2.4</td>
</tr>
<tr>
<td>3.1</td>
</tr>
<tr>
<td>3.5</td>
</tr>
</tbody>
</table>

Solution

First, in order to fit the data to a Bayesian-Weibull model, a prior distribution for beta needs to be determined. Based on the beta values in the prior tests, the prior distribution for beta is found to be a lognormal distribution with $\mu = 0.9064$, $\sigma = 0.3325$ (The values of the parameters can be obtained by entering the beta values into a Weibull++ standard folio and analyzing it using the lognormal distribution and the RRX analysis method.)

Next, enter the data from the prototype testing into a standard folio. On the control panel, choose the Bayesian-Weibull > B-W Lognormal Prior distribution. Click Calculate and enter the parameters of the lognormal distribution, as shown next.

Click OK. The result is Beta (Median) = 2.361219 and Eta (Median) = 5321.631912 (by default Weibull++ returns the median values of the posterior distribution). Suppose that the reliability at 3,000 hours is the metric of interest in this example. Using the QCP, the reliability is calculated to be 76.97% at 3,000 hours. The following picture depicts the posterior pdf plot of the reliability at 3,000, with the corresponding median value as well as the 10th percentile value. The 10th percentile constitutes the 90% lower 1-sided bound on the reliability at 3,000 hours, which is calculated to be 50.77%.
The pdf of the times-to-failure data can be plotted in Weibull++, as shown next:
Weibull Distribution Examples

Median Rank Plot Example

In this example, we will determine the median rank value used for plotting the 6th failure from a sample size of 10. This example will use Weibull++’s Quick Statistical Reference (QSR) tool to show how the points in the plot of the following example are calculated.

First, open the Quick Statistical Reference tool and select the Inverse F-Distribution Values option.

In this example, \( n_1 = 10, j = 6, m = 2(10 - 6 + 1) = 10, \) and \( n_2 = 2 \times 6 = 12. \)

Thus, from the F-distribution rank equation:

\[
MR = \frac{1}{1 + \left(\frac{10 - 6 + 1}{6}\right)F_{0.5;10;12}}
\]

Use the QSR to calculate the value of \( F_{0.5;10;12} = 0.9886, \) as shown next:

Consequently:

\[
MR = \frac{1}{1 + \left(\frac{4}{6}\right) 	imes 0.9886} = 0.5483 = 54.83\%
\]

Another method is to use the Median Ranks option directly, which yields \( MR(%) = 54.8305\%, \) as shown next:
Complete Data Example
Assume that 10 identical units (N = 10) are being reliability tested at the same application and operation stress levels. 6 of these units fail during this test after operating the following numbers of hours, $T_i$: 150, 105, 83, 123, 64 and 46. The test is stopped at the 6th failure. Find the parameters of the Weibull pdf that represents these data.

Solution
Create a new Weibull++ standard folio that is configured for grouped times-to-failure data with suspensions.

Enter the data in the appropriate columns. Note that there are 4 suspensions, as only 6 of the 10 units were tested to failure (the next figure shows the data as entered). Use the 3-parameter Weibull and MLE for the calculations.
Plot the data.
Note that the original data points, on the curved line, were adjusted by subtracting 30.92 hours to yield a straight line as shown above.

**Suspension Data Example**

ACME company manufactures widgets, and it is currently engaged in reliability testing a new widget design. 19 units are being reliability tested, but due to the tremendous demand for widgets, units are removed from the test whenever the production cannot cover the demand. The test is terminated at the 67th day when the last widget is removed from the test. The following table contains the collected data.

**Widget Test Data**

<table>
<thead>
<tr>
<th>Data Point Index</th>
<th>State (F/S)</th>
<th>Time to Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>S</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>S</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>S</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>S</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>23</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>29</td>
</tr>
</tbody>
</table>
Solution

In this example, we see that the number of failures is less than the number of suspensions. This is a very common situation, since reliability tests are often terminated before all units fail due to financial or time constraints. Furthermore, some suspensions will be recorded when a failure occurs that is not due to a legitimate failure mode, such as operator error. In cases such as this, a suspension is recorded, since the unit under test cannot be said to have had a legitimate failure.

Enter the data into a Weibull++ standard folio that is configured for times-to-failure data with suspensions. The folio will appear as shown next:

![Weibull++ folio]

We will use the 2-parameter Weibull to solve this problem. The parameters using maximum likelihood are:

\[ \hat{\beta} = 1.145 \]
\[ \hat{\eta} = 65.97 \]
Using RRX:
\[ \hat{\beta} = 0.914 \]
\[ \hat{\eta} = 79.38 \]
Using RRY:
\[ \hat{\beta} = 0.895 \]
\[ \hat{\eta} = 82.02 \]

**Interval Data Example**

Suppose we have run an experiment with 8 units tested and the following is a table of their last inspection times and failure times:

<table>
<thead>
<tr>
<th>Data Point Index</th>
<th>Last Inspection</th>
<th>Failure Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>55</td>
<td>55</td>
</tr>
</tbody>
</table>

Analyze the data using several different parameter estimation techniques and compare the results.

**Solution**

Enter the data into a Weibull++ standard folio that is configured for interval data. The data is entered as follows:
The computed parameters using maximum likelihood are:

\[ \hat{\beta} = 5.76 \]
\[ \hat{\eta} = 44.68 \]

Using RRX or rank regression on X:

\[ \hat{\beta} = 5.70 \]
\[ \hat{\eta} = 44.54 \]

Using RRY or rank regression on Y:

\[ \hat{\beta} = 5.41 \]
\[ \hat{\eta} = 44.76 \]

The plot of the MLE solution with the two-sided 90% confidence bounds is:
Mixed Data Types Example

From Dimitri Kececioglu, Reliability & Life Testing Handbook, Page 406. [20].

Estimate the parameters for the 3-parameter Weibull, for a sample of 10 units that are all tested to failure. The recorded failure times are 200; 370; 500; 620; 730; 840; 950; 1,050; 1,160 and 1,400 hours.

Published Results:
Published results (using probability plotting):

$$\hat{\beta} = 3.0, \hat{\eta} = 1,220, \hat{\gamma} = -300$$

Computed Results in Weibull++

Weibull++ computed parameters for rank regression on X are:

$$\hat{\beta} = 2.9013, \hat{\eta} = 1195.5009, \hat{\gamma} = -279.000$$

The small difference between the published results and the ones obtained from Weibull++ are due to the difference in the estimation method. In the publication the parameters were estimated using probability plotting (i.e., the fitted line was "eye-balled"). In Weibull++, the parameters were estimated using non-linear regression (a more accurate, mathematically fitted line). Note that $\gamma$ in this example is negative. This means that the unadjusted for $\gamma$ line is concave up, as shown next.
Weibull Distribution RRX Example

Assume that 6 identical units are being tested. The failure times are: 93, 34, 16, 120, 53 and 75 hours.

1. What is the unreliability of the units for a mission duration of 30 hours, starting the mission at age zero?
2. What is the reliability for a mission duration of 10 hours, starting the new mission at the age of T = 30 hours?
3. What is the longest mission that this product should undertake for a reliability of 90%?

Solution

1. First, we use Weibull++ to obtain the parameters using RRX.

Then, we investigate several methods of solution for this problem. The first, and more laborious, method is to extract the information directly from the plot. You may do this with either the screen plot in RS Draw or the printed copy of the plot. (When extracting information from the screen plot in RS Draw, note that the translated axis position of your mouse is always shown on the bottom right corner.)
Using this first method, enter either the screen plot or the printed plot with \( T = 30 \) hours, go up vertically to the straight line fitted to the data, then go horizontally to the ordinate, and read off the result. A good estimate of the unreliability is 23\%. (Also, the reliability estimate is \( 1.0 - 0.23 = 0.77 \) or 77\%.)

The second method involves the use of the Quick Calculation Pad (QCP).

Select the **Prob. of Failure** calculation option and enter 30 hours in the **Mission End Time** field.
Note that the results in QCP vary according to the parameter estimation method used. The above results are obtained using RRX.

2. The conditional reliability is given by:

\[ R(t|T) = \frac{R(T+t)}{R(T)} \]

or:

\[ \hat{R}(10\text{hr}|30\text{hr}) = \frac{\hat{R}(10+30)}{\hat{R}(30)} = \frac{\hat{R}(40)}{\hat{R}(30)} \]

Again, the QCP can provide this result directly and more accurately than the plot.

3. To use the QCP to solve for the longest mission that this product should undertake for a reliability of 90%, choose Reliable Life and enter 0.9 for the required reliability. The result is 15.9933 hours.
Benchmark with Published Examples

The following examples compare published results to computed results obtained with Weibull++.

**Complete Data RRY Example**

From Dimitri Kececioglu, Reliability & Life Testing Handbook, Page 418 [20].

Sample of 10 units, all tested to failure. The failures were recorded at 16, 34, 53, 75, 93, 120, 150, 191, 240 and 339 hours.

**Published Results**

Published Results (using Rank Regression on Y):

\[ \hat{\beta} = 1.20 \]
\[ \hat{\eta} = 146.2 \]
\[ \hat{\rho} = 0.998703 \]

**Computed Results in Weibull++**

This same data set can be entered into a Weibull++ standard data sheet. Use RRY for the estimation method.

Weibull++ computed parameters for RRY are:

\[ \hat{\beta} = 1.1973 \]
\[ \hat{\eta} = 146.2545 \]
\[ \hat{\rho} = 0.9999 \]

The small difference between the published results and the ones obtained from Weibull++ is due to the difference in the median rank values between the two (in the publication, median ranks are obtained from tables to 3 decimal places, whereas in Weibull++ they are calculated and carried out up to the 15th decimal point).

You will also notice that in the examples that follow, a small difference may exist between the published results and the ones obtained from Weibull++. This can be attributed to the difference between the computer numerical precision employed by Weibull++ and the lower number of significant digits used by the original authors. In most of these publications, no information was given as to the numerical precision used.

**Suspension Data MLE Example**

70 diesel engine fans accumulated 344,440 hours in service and 12 of them failed. A table of their life data is shown next (+ denotes non-failed units or suspensions, using Dr. Nelson’s nomenclature). Evaluate the parameters with their two-sided 95% confidence bounds, using MLE for the 2-parameter Weibull distribution.

<table>
<thead>
<tr>
<th>Nelson’s Fan Failure Data (hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>450 1850+ 3000+ 4150+ 4850+ 6450+ 8500+</td>
</tr>
<tr>
<td>460+ 1850+ 3000+ 4150+ 5000+ 6700+ 8750+</td>
</tr>
<tr>
<td>1150 2030+ 3000+ 4300+ 5000+ 7450+ 8750</td>
</tr>
<tr>
<td>1150 2030+ 3100 4300+ 5000+ 7800+ 8750+</td>
</tr>
<tr>
<td>1560+ 2030+ 3200+ 4300+ 6100+ 7800+ 9400+</td>
</tr>
<tr>
<td>1600 2070 3450 4300+ 6100 8100+ 9900+</td>
</tr>
<tr>
<td>1660+ 2070 3750+ 4600 6100+ 8100+ 10100+</td>
</tr>
<tr>
<td>1850+ 2080 3750+ 4850+ 6100+ 8200+ 10100+</td>
</tr>
<tr>
<td>1850+ 2200+ 4150+ 4850+ 6300+ 8500+ 10100+</td>
</tr>
<tr>
<td>1850+ 3000+ 4150+ 4850+ 6450+ 8500+ 11500+</td>
</tr>
</tbody>
</table>

**Published Results:**

Weibull parameters (2P-Weibull, MLE):

\[
\hat{\beta} = 1.0584 \\
\hat{\eta} = 26.296
\]

Published 95% FM confidence limits on the parameters:

\[
\hat{\beta} = \{0.6441, 1.7394\}
\]

\[
\hat{\eta} = \{10.522, 65.532\}
\]

Published variance/covariance matrix:

\[
\begin{bmatrix}
\widehat{\text{Var}}(\hat{\beta}) = 0.0720 & \widehat{\text{Cov}}(\hat{\beta}, \hat{\eta}) = -2.664.40 \\
\widehat{\text{Cov}}(\hat{\beta}, \hat{\eta}) = -2.664.40 & \widehat{\text{Var}}(\hat{\eta}) = 15.009E+7
\end{bmatrix}
\]
Note that Nelson expresses the results as multiples of 1,000 (or = 26.297, etc.). The published results were adjusted by this factor to correlate with Weibull++ results.

**Computed Results in Weibull++**

This same data set can be entered into a Weibull++ standard folio, using 2-parameter Weibull and MLE to calculate the parameter estimates.

You can also enter the data as given in table without grouping them by opening a data sheet configured for suspension data. Then click the **Group Data** icon and chose **Group exactly identical values**.

The data will be automatically grouped and put into a new grouped data sheet.

Weibull++ computed parameters for maximum likelihood are:

\[ \hat{\beta} = 1.0584 \]
\[ \hat{\eta} = 26,297 \]

Weibull++ computed 95% FM confidence limits on the parameters:

\[ \hat{\beta} = \{0.6441, 1.7394\} \]
\[ \hat{\eta} = \{10, 522, 65,532\} \]

Weibull++ computed/variance covariance matrix:
The two-sided 95% bounds on the parameters can be determined from the QCP. Calculate and then click **Report** to see the results.

\[
\begin{bmatrix}
\hat{\text{Var}}(\hat{\beta}) = 0.071958 \\
\hat{\text{Cov}}(\hat{\beta}, \hat{\eta}) = -2,664.5 \\
\hat{\text{Cov}}(\hat{\beta}, \hat{\eta}) = -2,664.5 \\
\hat{\text{Var}}(\hat{\eta}) = 15.010E + 7
\end{bmatrix}
\]

**Interval Data MLE Example**

From Wayne Nelson, Applied Life Data Analysis, Page 415 [30]. 167 identical parts were inspected for cracks. The following is a table of their last inspection times and times-to-failure:

<table>
<thead>
<tr>
<th>Data point index</th>
<th>Number in State</th>
<th>Last Inspection</th>
<th>State (5 or 1)</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>F</td>
<td>6.12</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>6.12</td>
<td>F</td>
<td>19.02</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>19.92</td>
<td>F</td>
<td>29.64</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>29.64</td>
<td>F</td>
<td>35.4</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>35.4</td>
<td>F</td>
<td>39.72</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>39.72</td>
<td>F</td>
<td>45.24</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>45.24</td>
<td>F</td>
<td>63.32</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>52.32</td>
<td>F</td>
<td>63.48</td>
</tr>
<tr>
<td>9</td>
<td>73</td>
<td>63.48</td>
<td>F</td>
<td>63.48</td>
</tr>
</tbody>
</table>
Published Results:

Published results (using MLE):
\[ \hat{\beta} = 1.486 \]
\[ \hat{\eta} = 71.687 \]

Published 95% FM confidence limits on the parameters:
\[ \hat{\beta} = \{1.224, 1.802\} \]
\[ \hat{\eta} = \{61.962, 82.938\} \]

Published variance/covariance matrix:
\[
\begin{bmatrix}
\var{\hat{\beta}} &= 0.0215 & \cov{\hat{\beta}, \hat{\eta}} &= -0.2791 \\
\cov{\hat{\beta}, \hat{\eta}} &= -0.2791 & \var{\hat{\eta}} &= 28.432
\end{bmatrix}
\]

Computed Results in Weibull++

This same data set can be entered into a Weibull++ standard folio that’s configured for grouped times-to-failure data with suspensions and interval data.

Weibull++ computed parameters for maximum likelihood are:
\[ \hat{\beta} = 1.485 \]
\[ \hat{\eta} = 71.690 \]

Weibull++ computed 95% FM confidence limits on the parameters:
\[ \hat{\beta} = \{1.224, 1.802\} \]
\[ \hat{\eta} = \{61.961, 82.947\} \]

Weibull++ computed/variance covariance matrix:
\[
\begin{bmatrix}
\var{\hat{\beta}} &= 0.0215 & \cov{\hat{\beta}, \hat{\eta}} &= -0.2792 \\
\cov{\hat{\beta}, \hat{\eta}} &= -0.2792 & \var{\hat{\eta}} &= 28.4461
\end{bmatrix}
\]

Grouped Suspension MLE Example


Wingo uses the following times-to-failure: 37, 55, 64, 72, 74, 87, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 102, 102, 105, 105, 107, 113, 117, 120, 120, 122, 124, 126, 130, 135, 138, 182. In addition, the following suspensions are used: 4 at 70, 5 at 80, 4 at 99, 3 at 121 and 1 at 150.

Published Results (using MLE)
\[ \hat{\beta} = 3.7596935 \]
\[ \hat{\gamma} = 106.49758 \]
\[ \hat{\gamma} = 14.451684 \]

**Computed Results in Weibull++**

\[ \hat{\beta} = 3.7596935 \]
\[ \hat{\gamma} = 106.49758 \]
\[ \hat{\gamma} = 14.451684 \]

Note that you must select the **Use True 3-P MLE** option in the Weibull++ Application Setup to replicate these results.

**3-P Probability Plot Example**

Suppose we want to model a left censored, right censored, interval, and complete data set, consisting of 274 units under test of which 185 units fail. The following table contains the data.

### The Test Data

<table>
<thead>
<tr>
<th>Data Point Index</th>
<th>Number in State</th>
<th>Last Inspection</th>
<th>State (S or F)</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>5</td>
<td>F</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>5</td>
<td>S</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>0</td>
<td>F</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>10</td>
<td>F</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>15</td>
<td>F</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>20</td>
<td>F</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>29</td>
<td>20</td>
<td>S</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>0</td>
<td>F</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>25</td>
<td>F</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>27</td>
<td>F</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>30</td>
<td>F</td>
<td>35</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>30</td>
<td>F</td>
<td>40</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>27</td>
<td>F</td>
<td>45</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>25</td>
<td>F</td>
<td>50</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>20</td>
<td>F</td>
<td>55</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>15</td>
<td>F</td>
<td>60</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>10</td>
<td>F</td>
<td>65</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>5</td>
<td>F</td>
<td>70</td>
</tr>
<tr>
<td>19</td>
<td>37</td>
<td>100</td>
<td>S</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>48</td>
<td>0</td>
<td>F</td>
<td>102</td>
</tr>
</tbody>
</table>

**Solution**

Since standard ranking methods for dealing with these different data types are inadequate, we will want to use the ReliaSoft ranking method. This option is the default in Weibull++ when dealing with interval data. The filled-out standard folio is shown next:
The computed parameters using MLE are:

\[ \hat{\beta} = 0.748; \quad \hat{\eta} = 44.38 \]

Using RRX:

\[ \hat{\beta} = 1.057; \quad \hat{\eta} = 36.29 \]

Using RRY:

\[ \hat{\beta} = 0.998; \quad \hat{\eta} = 37.16 \]

The plot with the two-sided 90% confidence bounds for the rank regression on X solution is:
References

Chapter 9

The Normal Distribution

The normal distribution, also known as the Gaussian distribution, is the most widely-used general purpose distribution. It is for this reason that it is included among the lifetime distributions commonly used for reliability and life data analysis. There are some who argue that the normal distribution is inappropriate for modeling lifetime data because the left-hand limit of the distribution extends to negative infinity. This could conceivably result in modeling negative times-to-failure. However, provided that the distribution in question has a relatively high mean and a relatively small standard deviation, the issue of negative failure times should not present itself as a problem. Nevertheless, the normal distribution has been shown to be useful for modeling the lifetimes of consumable items, such as printer toner cartridges.

Normal Probability Density Function

The pdf of the normal distribution is given by:

\[
f(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t - \mu}{\sigma} \right)^2}
\]

where:
- \(\mu\) = mean of the normal times-to-failure, also noted as \(\bar{T}\),
- \(\sigma\) = standard deviation of the times-to-failure

It is a 2-parameter distribution with parameters \(\mu\) (or \(\bar{T}\)) and \(\sigma\) (i.e., the mean and the standard deviation, respectively).

Normal Statistical Properties

The Normal Mean, Median and Mode

The normal mean or MTTF is actually one of the parameters of the distribution, usually denoted as \(\mu\). Because the normal distribution is symmetrical, the median and the mode are always equal to the mean:

\[
\mu = \bar{T} = \hat{T}
\]

The Normal Standard Deviation

As with the mean, the standard deviation for the normal distribution is actually one of the parameters, usually denoted as \(\sigma T\).
The Normal Distribution

The Normal Reliability Function

The reliability for a mission of time $T$ for the normal distribution is determined by:

$$R(t) = \int_t^\infty f(x)\,dx = \int_t^\infty \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \,dx$$

There is no closed-form solution for the normal reliability function. Solutions can be obtained via the use of standard normal tables. Since the application automatically solves for the reliability, we will not discuss manual solution methods. For interested readers, full explanations can be found in the references.

The Normal Conditional Reliability Function

The normal conditional reliability function is given by:

$$R(t|T) = \frac{R(T + t)}{R(T)} = \frac{\int_{T+t}^\infty \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \,dx}{\int_T^\infty \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \,dx}$$

Once again, the use of standard normal tables for the calculation of the normal conditional reliability is necessary, as there is no closed-form solution.

The Normal Reliable Life

Since there is no closed-form solution for the normal reliability function, there will also be no closed-form solution for the normal reliable life. To determine the normal reliable life, one must solve:

$$R(T) = \int_T^\infty \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2} \,dt$$

for $T$.

The Normal Failure Rate Function

The instantaneous normal failure rate is given by:

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2}}{\int_T^\infty \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \,dx}$$
**Characteristics of the Normal Distribution**

Some of the specific characteristics of the normal distribution are the following:

- The normal pdf has a mean, $\bar{T}$, which is equal to the median, $\tilde{T}$, and also equal to the mode, $\check{T}$, or $\bar{T} = \tilde{T} = \check{T}$. This is because the normal distribution is symmetrical about its mean.

- The mean, $\mu$, or the mean life or the $MTTF$, is also the location parameter of the normal pdf, as it locates the pdf along the abscissa. It can assume values of $-\infty < \bar{T} < \infty$.

- The normal pdf has no shape parameter. This means that the normal pdf has only one shape, the bell shape, and this shape does not change.

- The standard deviation, $\sigma$, is the scale parameter of the normal pdf.

- As $\sigma$ decreases, the pdf gets pushed toward the mean, or it becomes narrower and taller.

- As $\sigma$ increases, the pdf spreads out away from the mean, or it becomes broader and shallower.

- The standard deviation can assume values of $0 < \sigma < \infty$.

- The greater the variability, the larger the value of $\sigma$, and vice versa.
• The standard deviation is also the distance between the mean and the point of inflection of the pdf, on each side of the mean. The point of inflection is that point of the pdf where the slope changes its value from a decreasing to an increasing one, or where the second derivative of the pdf has a value of zero.

• The normal pdf starts at \( t = -\infty \) with \( f(t) = 0 \). As \( t \) increases, \( f(t) \) also increases, goes through its point of inflection and reaches its maximum value at \( t = \bar{t} \). Thereafter, \( f(t) \) decreases, goes through its point of inflection, and assumes a value of \( f(t) = 0 \) at \( t = +\infty \).

**Weibull++ Notes on Negative Time Values**

One of the disadvantages of using the normal distribution for reliability calculations is the fact that the normal distribution starts at negative infinity. This can result in negative values for some of the results. Negative values for time are not accepted in most of the components of Weibull++, nor are they implemented. Certain components of the application reserve negative values for suspensions, or will not return negative results. For example, the Quick Calculation Pad will return a null value (zero) if the result is negative. Only the Free-Form (Probit) data sheet can accept negative values for the random variable (x-axis values).

**Estimation of the Parameters**

**Probability Plotting**

As described before, probability plotting involves plotting the failure times and associated unreliability estimates on specially constructed probability plotting paper. The form of this paper is based on a linearization of the cdf of the specific distribution. For the normal distribution, the cumulative density function can be written as:

\[
F(t) = \Phi \left( \frac{t - \mu}{\sigma} \right)
\]

or:

\[
\Phi^{-1}[F(t)] = -\frac{\mu}{\sigma} + \frac{1}{\sigma} t
\]

where:

\[
\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt
\]

Now, let:

\[
y = \Phi^{-1}[F(t)]
\]

\[
a = -\frac{\mu}{\sigma}
\]

and:

\[
b = \frac{1}{\sigma}
\]

which results in the linear equation of:

\[
y = a + bT
\]

The normal probability paper resulting from this linearized cdf function is shown next.
Since the normal distribution is symmetrical, the area under the pdf curve from $-\infty$ to $\mu$ is 0.5, as is the area from $\mu$ to $\infty$. Consequently, the value of $\mu$ is said to be the point where $R(t) = Q(t) = 50\%$. This means that the estimate of $\mu$ can be read from the point where the plotted line crosses the 50% unreliability line.

To determine the value of $\sigma$ from the probability plot, it is first necessary to understand that the area under the pdf curve that lies between one standard deviation in either direction from the mean (or two standard deviations total) represents 68.3% of the area under the curve. This is represented graphically in the following figure.

Consequently, the interval between $Q(t) = 84.15\%$ and $Q(t) = 15.85\%$ represents two standard deviations, since this is an interval of 68.3% ($84.15 - 15.85 = 68.3$), and is centered on the mean at 50%. As a result, the standard deviation can be estimated from:
The Normal Distribution

\[ \hat{\sigma} = \frac{t(Q = 84.15\%) - t(Q = 15.85\%)}{2} \]

That is: the value of \( \hat{\sigma} \) is obtained by subtracting the time value where the plotted line crosses the 84.15% unreliability line from the time value where the plotted line crosses the 15.85% unreliability line and dividing the result by two. This process is illustrated in the following example.

Normal Distribution Probability Plotting Example

7 units are put on a life test and run until failure. The failure times are 85, 90, 95, 100, 105, 110, and 115 hours. Assuming a normal distribution, estimate the parameters using probability plotting.

In order to plot the points for the probability plot, the appropriate estimates for the unreliability values must be obtained. These values will be estimated through the use of median ranks, which can be obtained from statistical tables or from the Quick Statistical Reference (QSR) tool in Weibull++. The following table shows the times-to-failure and the appropriate median rank values for this example:

<table>
<thead>
<tr>
<th>Time-to-Failure (hr)</th>
<th>Median Rank (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
<td>9.43%</td>
</tr>
<tr>
<td>90</td>
<td>22.85%</td>
</tr>
<tr>
<td>95</td>
<td>36.41%</td>
</tr>
<tr>
<td>100</td>
<td>50.00%</td>
</tr>
<tr>
<td>105</td>
<td>63.59%</td>
</tr>
<tr>
<td>110</td>
<td>77.15%</td>
</tr>
<tr>
<td>115</td>
<td>90.57%</td>
</tr>
</tbody>
</table>

These points can now be plotted on a normal probability plotting paper as shown in the next figure.
Draw the best possible line through the plot points. The time values where the line intersects the 15.85%, 50%, and 84.15% unreliability values should be projected down to the abscissa, as shown in the following plot.

The estimate of $\mu$ is determined from the time value at the 50% unreliability level, which in this case is 100 hours. The value of the estimator of $\sigma$ is determined as follows:

$$\hat{\sigma} = \frac{t(Q = 84.15\%) - t(Q = 15.85\%)}{2}$$

$$\hat{\sigma} = \frac{112 - 88}{2} = \frac{24}{2} = 12 \text{ hours}$$

Alternately, $\hat{\sigma}$ could be determined by measuring the distance from $t(Q = 15.85\%)$ to $t(Q = 50\%)$, or $t(Q = 50\%)$ to $t(Q = 84.15\%)$, as either of these two distances is equal to the value of one standard deviation.

**Rank Regression on Y**

Performing rank regression on Y requires that a straight line be fitted to a set of data points such that the sum of the squares of the vertical deviations from the points to the line is minimized.

The least squares parameter estimation method (regression analysis) was discussed in Parameter Estimation, and the following equations for regression on Y were derived:

$$\hat{\sigma} = \hat{\sigma} = \frac{\sum_{i=1}^{N} y_i - \hat{b} \sum_{i=1}^{N} x_i}{N}$$

and:
In the case of the normal distribution, the equations for $\hat{\mu}$ and $\hat{\sigma}$ are:

$$\hat{\mu} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sum_{i=1}^{N} x_i^2 - \left( \frac{\sum_{i=1}^{N} x_i}{N} \right)^2}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}}$$

where the values for $F(T_i)$ are estimated from the median ranks. Once $\hat{\mu}$ and $\hat{\sigma}$ are obtained, $\hat{\sigma}$ and $\hat{\mu}$ can easily be obtained from above equations.

**The Correlation Coefficient**

The estimator of the sample correlation coefficient, $\hat{\rho}$, is given by:

$$\hat{\rho} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2}}$$

**RRY Example**

**Normal Distribution RRY Example**

14 units were reliability tested and the following life test data were obtained. Assuming the data follow a normal distribution, estimate the parameters and determine the correlation coefficient, $\hat{\rho}$, using rank regression on Y.

<table>
<thead>
<tr>
<th>Data point index</th>
<th>Time-to-failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>11</td>
<td>70</td>
</tr>
<tr>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td>13</td>
<td>90</td>
</tr>
<tr>
<td>14</td>
<td>100</td>
</tr>
</tbody>
</table>

**Solution**

Construct a table like the one shown next.
The Normal Distribution

163

• The median rank values ($F(t_i)$) can be found in rank tables, available in many statistical texts, or they can be estimated by using the Quick Statistical Reference in Weibull++.

• The $y_i$ values were obtained from standardized normal distribution’s area tables by entering for $F(z)$ and getting the corresponding $z$ value ($y_i$). As with the median rank values, these standard normal values can be obtained with the Quick Statistical Reference.

Given the values in the table above, calculate $\widehat{\alpha}$ and $\widehat{\beta}$ using:

$$
\widehat{\beta} = \frac{\sum_{i=1}^{14} T_i y_i - (\sum_{i=1}^{14} T_i)(\sum_{i=1}^{14} y_i)/14}{\sum_{i=1}^{14} T_i^2 - (\sum_{i=1}^{14} T_i)^2/14} = 0.02982
$$

and:

$$
\widehat{\alpha} = \bar{y} - \widehat{\beta} \bar{T} = \frac{\sum_{i=1}^{N} y_i}{N} - \widehat{\beta} \frac{\sum_{i=1}^{N} T_i}{N}
$$
or:

$$
\widehat{\alpha} = \frac{0}{14} - (0.02982) \frac{630}{14} = -1.3419
$$

Therefore:

$$
\widehat{\sigma} = \frac{1}{\widehat{\beta}} = \frac{1}{0.02982} = 33.5367
$$

and:

$$
\widehat{\mu} = -\widehat{\alpha} \cdot \widehat{\sigma} = -( -1.3419) \cdot 33.5367 \approx 45
$$
or $\widehat{\mu} = 45$ hours.

The correlation coefficient can be estimated using:

$$
\widehat{\rho} = 0.979
$$
The preceding example can be repeated using Weibull++.

- Create a new folio for Times-to-Failure data, and enter the data given in this example.
- Choose Normal from the Distributions list.
- Go to the Analysis page and select Rank Regression on Y (RRY).
- Click the Calculate icon located on the Main page.

The probability plot is shown next.
**Rank Regression on X**

As was mentioned previously, performing a rank regression on X requires that a straight line be fitted to a set of data points such that the sum of the squares of the horizontal deviations from the points to the fitted line is minimized.

Again, the first task is to bring our function, the probability of failure function for normal distribution, into a linear form. This step is exactly the same as in regression on Y analysis. All other equations apply in this case as they did for the regression on Y. The deviation from the previous analysis begins on the least squares fit step where: in this case, we treat as the dependent variable and as the independent variable. The best-fitting straight line for the data, for regression on X, is the straight line:

\[ x = \hat{a} + \hat{b}y \]

The corresponding equations for \( \hat{a} \) and \( \hat{b} \) are:

\[
\hat{a} = \bar{x} - \bar{b}y = \frac{\sum_{i=1}^{N} x_i}{N} - \frac{\sum_{i=1}^{N} y_i}{N}
\]

and:

\[
\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \left( \frac{\sum_{i=1}^{N} x_i}{N} \right) \left( \sum_{i=1}^{N} y_i \right)}{\sum_{i=1}^{N} y_i^2 - \left( \frac{\sum_{i=1}^{N} y_i}{N} \right)^2}
\]

where:

\[
y_i = \Phi^{-1} [ F(t_i) ]
\]
and:
\[ x_i = t_i \]
and the \( F(t_i) \) values are estimated from the median ranks. Once \( \hat{a} \) and \( \hat{b} \) are obtained, solve the above linear equation for the unknown value of \( y \), which corresponds to:
\[ y = -\frac{\hat{a}}{\hat{b}} + \frac{x}{\hat{b}} \]
Solving for the parameters, we get:
\[ a = -\frac{\hat{a}}{\hat{b}} = -\frac{\mu}{\sigma} \Rightarrow \mu = \hat{a} \]
and:
\[ b = \frac{1}{\hat{b}} = \frac{1}{\sigma} \Rightarrow \sigma = \hat{b} \]
The correlation coefficient is evaluated as before.

**RRX Example**

**Normal Distribution RRX Example**

Using the same data set from the RRY example given above, and assuming a normal distribution, estimate the parameters and determine the correlation coefficient, \( \hat{\rho} \), using rank regression on X.

**Solution**

The table constructed for the RRY analysis applies to this example also. Using the values on this table, we get:
\[
\hat{b} = \frac{\sum_{i=1}^{14} t_i y_i - \frac{14}{14} \left( \sum_{i=1}^{14} t_i \right) \left( \sum_{i=1}^{14} y_i \right)}{\sum_{i=1}^{14} y_i^2 - \left( \frac{14}{14} \sum_{i=1}^{14} y_i \right)^2}
\]
\[
\hat{b} = \frac{365.2711 - (630)(0)/14}{11.3646 - (0)^2/14} = 32.1411
\]
and:
\[
\hat{a} = \bar{x} - \hat{b} \bar{y} = \frac{\sum_{i=1}^{14} t_i}{14} - \hat{b} \frac{\sum_{i=1}^{14} y_i}{14}
\]
or:
\[
\hat{a} = \frac{630}{14} - (32.1411)(0) = 45
\]
Therefore:
\[
\hat{\sigma} = \hat{b} = 32.1411
\]
and:
\[
\hat{\mu} = \hat{a} = 45 \text{ hours}
\]
The correlation coefficient is obtained as:
\[
\hat{\rho} = 0.979
\]
Note that the results for regression on X are not necessarily the same as the results for regression on Y. The only time when the two regressions are the same (i.e., will yield the same equation for a line) is when the data lie perfectly on a straight line.
The Normal Distribution

The plot of the Weibull++ solution for this example is shown next.

Maximum Likelihood Estimation

As it was outlined in Parameter Estimation, maximum likelihood estimation works by developing a likelihood function based on the available data and finding the values of the parameter estimates that maximize the likelihood function. This can be achieved by using iterative methods to determine the parameter estimate values that maximize the likelihood function. This can be rather difficult and time-consuming, particularly when dealing with the three-parameter distribution. Another method of finding the parameter estimates involves taking the partial derivatives of the likelihood function with respect to the parameters, setting the resulting equations equal to zero, and solving simultaneously to determine the values of the parameter estimates. The log-likelihood functions and associated partial derivatives used to determine maximum likelihood estimates for the normal distribution are covered in the Appendix.

Special Note About Bias

Estimators (i.e., parameter estimates) have properties such as unbiasedness, minimum variance, sufficiency, consistency, squared error constancy, efficiency and completeness, as discussed in Dudewicz and Mishra [7] and in Dietrich [5]. Numerous books and papers deal with these properties and this coverage is beyond the scope of this reference.

However, we would like to briefly address one of these properties, unbiasedness. An estimator is said to be unbiased if the estimator \( \hat{\theta} = d(X_1, X_2, \ldots, X_n) \) satisfies the condition \( E[\hat{\theta}] = \theta \) for all \( \theta \in \Omega \).

Note that \( E[X] \) denotes the expected value of \( X \) and is defined (for continuous distributions) by:
The Normal Distribution

\[ E[X] = \int_{\omega} x \cdot f(x) \, dx \]
\[ X \in \omega. \]

It can be shown in Dudewicz and Mishra [7] and in Dietrich [5] that the MLE estimator for the mean of the normal (and lognormal) distribution does satisfy the unbiasedness criteria, or \( E[\hat{\mu}] = \mu \). The same is not true for the estimate of the variance \( \hat{\sigma^2} \). The maximum likelihood estimate for the variance for the normal distribution is given by:

\[ \hat{\sigma^2} = \frac{1}{N} \sum_{i=1}^{N} (t_i - \bar{T})^2 \]

with a standard deviation of:

\[ \hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (t_i - \bar{T})^2} \]

These estimates, however, have been shown to be biased. It can be shown in Dudewicz and Mishra [7] and in Dietrich [5] that the unbiased estimate of the variance and standard deviation for complete data is given by:

\[ \hat{\sigma^2} = \left[ \frac{N}{N - 1} \right] \cdot \left[ \frac{1}{N} \sum_{i=1}^{N} (t_i - \bar{T})^2 \right] = \frac{1}{N - 1} \sum_{i=1}^{N} (t_i - \bar{T})^2 \]

\[ \hat{\sigma} = \sqrt{\left[ \frac{N}{N - 1} \right] \cdot \left[ \frac{1}{N} \sum_{i=1}^{N} (t_i - \bar{T})^2 \right]} \]

\[ = \sqrt{\frac{1}{N - 1} \sum_{i=1}^{N} (t_i - \bar{T})^2} \]

Note that for larger values of \( N \), \( \sqrt{N/(N - 1)} \) tends to 1.

The Use Unbiased Std on Normal Data option on the Calculations page of the User Setup allows biasing to be considered when estimating the parameters.

When this option is selected, Weibull++ returns the unbiased standard deviation as defined. This is only true for complete data sets. For all other data types, Weibull++ by default returns the biased standard deviation as defined above regardless of the selection status of this option. The next figure shows this setting in Weibull++.
MLE Example

Normal Distribution MLE Example

Using the same data set from the RRY and RRX examples given above and assuming a normal distribution, estimate the parameters using the MLE method.

Solution

In this example we have non-grouped data without suspensions and without interval data. The partial derivatives of the normal log-likelihood function, $L$, are given by:

\[
\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{14} (t_i - \mu) = 0
\]

\[
\frac{\partial L}{\partial \sigma} = \sum_{i=1}^{14} \left( \frac{t_i - \mu}{\sigma^3} - \frac{1}{\sigma} \right) = 0
\]

(The derivations of these equations are presented in the appendix.) Substituting the values of $t_i$ and solving the above system simultaneously, we get $\hat{\mu} = 29.5$ hours, $\hat{\sigma} = 45$ hours.

The Fisher matrix is:

\[
\begin{bmatrix}
\text{Var}(\hat{\mu}) & 62.5000 \\
\text{Cov}(\hat{\mu}, \hat{\sigma}) & 0.0000 \\
\text{Cov}(\hat{\mu}, \hat{\sigma}) & 0.0000 \\
\text{Var}(\hat{\sigma}) & 31.2500
\end{bmatrix}
\]

The plot of the Weibull++ solution for this example is shown next.
Confidence Bounds

The method used by the application in estimating the different types of confidence bounds for normally distributed data is presented in this section. The complete derivations were presented in detail (for a general function) in Confidence Bounds.

Exact Confidence Bounds

There are closed-form solutions for exact confidence bounds for both the normal and lognormal distributions. However, these closed-form solutions only apply to complete data. To achieve consistent application across all possible data types, Weibull++ always uses the Fisher matrix method or likelihood ratio method in computing confidence intervals.

Fisher Matrix Confidence Bounds

Bounds on the Parameters

The lower and upper bounds on the mean, \( \hat{\mu} \), are estimated from:

\[ \hat{\mu}_U = \hat{\mu} + K_\alpha \sqrt{Var(\hat{\mu})} \]  (upper bound)
\[ \hat{\mu}_L = \hat{\mu} - K_\alpha \sqrt{Var(\hat{\mu})} \]  (lower bound)

Since the standard deviation, \( \hat{\sigma} \), must be positive, \( \ln(\hat{\sigma}) \) is treated as normally distributed, and the bounds are estimated from:
The Normal Distribution

\[ \sigma_U = \bar{\sigma} \cdot e^{\frac{K_\alpha \sqrt{Var(\bar{\sigma})}}{\bar{\sigma}}} \text{ (upper bound)} \]

\[ \sigma_L = \frac{\bar{\sigma}}{e^{\frac{K_\alpha \sqrt{Var(\bar{\sigma})}}{\bar{\sigma}}}} \text{ (lower bound)} \]

where \( K_\alpha \) is defined by:

\[ \alpha = \frac{1}{\sqrt{2\pi}} \int_{K_\alpha}^{\infty} e^{-\frac{t^2}{2}} \, dt = 1 - \Phi(K_\alpha) \]

If \( \alpha \) is the confidence level, then \( \alpha = \frac{1-\beta}{2} \) for the two-sided bounds and \( \alpha = 1 - \beta \) for the one-sided bounds.

The variances and covariances of \( \hat{\mu} \) and \( \hat{\sigma} \) are estimated from the Fisher matrix, as follows:

\[ \begin{pmatrix} \text{Var}(\hat{\mu}) & \text{Cov}(\hat{\mu}, \hat{\sigma}) \\ \text{Cov}(\hat{\mu}, \hat{\sigma}) & \text{Var}(\hat{\sigma}) \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 \ell}{\partial \mu^2} & -\frac{\partial^2 \ell}{\partial \mu \partial \sigma} \\ -\frac{\partial^2 \ell}{\partial \mu \partial \sigma} & -\frac{\partial^2 \ell}{\partial \sigma^2} \end{pmatrix}^{-1} \]

is the log-likelihood function of the normal distribution, described in Parameter Estimation and Appendix D.

Bounds on Reliability

The reliability of the normal distribution is:

\[ \hat{R}(t; \hat{\mu}, \hat{\sigma}) = \int_{t}^{\infty} \frac{1}{\hat{\sigma} \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t - \hat{\mu}}{\hat{\sigma}} \right)^2} \, dt \]

Let \( \tilde{z} = \frac{t - \hat{\mu}}{\hat{\sigma}} \), the above equation then becomes:

\[ \hat{R}(\tilde{z}) = \int_{\tilde{z}(t)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \tilde{z}^2} \, d\tilde{z} \]

The bounds on \( z \) are estimated from:

\[ z_U = \tilde{z} + K_\alpha \sqrt{\text{Var}(\tilde{z})} \]

\[ z_L = \tilde{z} - K_\alpha \sqrt{\text{Var}(\tilde{z})} \]

where:

\[ \text{Var}(\tilde{z}) = \left( \frac{\partial \tilde{z}}{\partial \mu} \right)^2 \text{Var}(\hat{\mu}) + \left( \frac{\partial \tilde{z}}{\partial \sigma} \right)^2 \text{Var}(\hat{\sigma}) + 2 \left( \frac{\partial \tilde{z}}{\partial \mu} \right) \left( \frac{\partial \tilde{z}}{\partial \sigma} \right) \text{Cov}(\hat{\mu}, \hat{\sigma}) \]

or:

\[ \text{Var}(\tilde{z}) = \frac{1}{\hat{\sigma}^2} \left[ \text{Var}(\hat{\mu}) + \tilde{z}^2 \text{Var}(\hat{\sigma}) + 2 \cdot \tilde{z} \cdot \text{Cov}(\hat{\mu}, \hat{\sigma}) \right] \]

The upper and lower bounds on reliability are:

\[ R_U = \int_{z_L}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \tilde{z}^2} \, d\tilde{z} \text{ (upper bound)} \]

\[ R_L = \int_{z_U}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \tilde{z}^2} \, d\tilde{z} \text{ (lower bound)} \]
Bounds on Time

The bounds around time for a given normal percentile (unreliability) are estimated by first solving the reliability equation with respect to time, as follows:

\[ \hat{T}(\hat{\mu}, \hat{\sigma}) = \hat{\mu} + z \cdot \hat{\sigma} \]

where:

\[ z = \Phi^{-1}[F(T)] \]

and:

\[ \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}z^2} dz \]

The next step is to calculate the variance of \( \hat{T}(\hat{\mu}, \hat{\sigma}) \) or:

\[ \text{Var}(\hat{T}) = \left( \frac{\partial \hat{T}}{\partial \mu} \right)^2 \text{Var}(\hat{\mu}) + \left( \frac{\partial \hat{T}}{\partial \sigma} \right)^2 \text{Var}(\hat{\sigma}) + 2 \left( \frac{\partial \hat{T}}{\partial \mu} \right) \left( \frac{\partial \hat{T}}{\partial \sigma} \right) \text{Cov}(\hat{\mu}, \hat{\sigma}) \]

The upper and lower bounds are then found by:

\[ T_U = \hat{T} + K_\alpha \sqrt{\text{Var}(\hat{T})} \quad \text{(upper bound)} \]
\[ T_L = \hat{T} - K_\alpha \sqrt{\text{Var}(\hat{T})} \quad \text{(lower bound)} \]

Likelihood Ratio Confidence Bounds

Bounds on Parameters

As covered in Confidence Bounds, the likelihood confidence bounds are calculated by finding values for \( \theta_1 \) and \( \theta_2 \) that satisfy:

\[ -2 \cdot \ln \left( \frac{L(\theta_1, \theta_2)}{L(\hat{\theta_1}, \hat{\theta_2})} \right) = \chi^2_{\alpha, 1} \]

This equation can be rewritten as:

\[ L(\theta_1, \theta_2) = L(\hat{\theta_1}, \hat{\theta_2}) \cdot e^{-\frac{\chi^2_{\alpha, 1}}{2}} \]

For complete data, the likelihood formula for the normal distribution is given by:

\[ L(\mu, \sigma) = \prod_{i=1}^{N} f(t_i; \mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left( \frac{t_i - \mu}{\sigma} \right)^2} \]

where the \( t_i \) values represent the original time to failure data. For a given value of \( \alpha \), values for \( \mu \) and \( \sigma \) can be found which represent the maximum and minimum values that satisfy the above likelihood ratio equation. These represent the confidence bounds for the parameters at a confidence level \( \delta \), where \( \alpha = \delta \) for two-sided bounds and \( \alpha = 2\delta - 1 \) for one-sided.
Example: LR Bounds on Parameters

Five units are put on a reliability test and experience failures at 12, 24, 28, 34, and 46 hours. Assuming a normal distribution, the MLE parameter estimates are calculated to be $\hat{\mu} = 28.8$ and $\hat{\sigma} = 11.2143$. Calculate the two-sided 80% confidence bounds on these parameters using the likelihood ratio method.

Solution

The first step is to calculate the likelihood function for the parameter estimates:

$$L(\hat{\mu}, \hat{\sigma}) = \prod_{i=1}^{N} f(t_i; \hat{\mu}, \hat{\sigma}) = \prod_{i=1}^{5} \frac{1}{\hat{\sigma} \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left( \frac{t_i - \hat{\mu}}{\hat{\sigma}} \right)^2}$$

$$L(\hat{\mu}, \hat{\sigma}) = \prod_{i=1}^{5} \frac{1}{11.2143 \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left( \frac{t_i - 28.8}{11.2143} \right)^2}$$

$$L(\hat{\mu}, \hat{\sigma}) = 4.676897 \times 10^{-9}$$

where $t_i$ are the original time-to-failure data points. We can now rearrange the likelihood ratio equation to the form:

$$L(\mu, \sigma) - L(\hat{\mu}, \hat{\sigma}) \cdot e^{-\frac{\chi^2_{0.5,1}}{2}} = 0$$

Since our specified confidence level, $\delta$, is 80%, we can calculate the value of the chi-squared statistic, $\chi^2_{0.8,1} = 1.642374$. We can now substitute this information into the equation:

$$L(\mu, \sigma) - L(\hat{\mu}, \hat{\sigma}) \cdot e^{-\frac{1.642374}{2}} = 0,$$

$$L(\mu, \sigma) - 4.676897 \times 10^{-9} \cdot e^{-\frac{1.642374}{2}} = 0,$$

$$L(\mu, \sigma) - 2.057410 \times 10^{-9} = 0.$$

It now remains to find the values of $\mu$ and $\sigma$ which satisfy this equation. This is an iterative process that requires setting the value of $\mu$ and finding the appropriate values of $\sigma$, and vice versa.

The following table gives the values of $\sigma$ based on given values of $\mu$:

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\mu$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.0</td>
<td>12.045</td>
<td>14.354</td>
<td>29.0</td>
<td>7.849</td>
<td>19.909</td>
</tr>
<tr>
<td>22.5</td>
<td>11.004</td>
<td>15.310</td>
<td>29.5</td>
<td>7.876</td>
<td>17.889</td>
</tr>
<tr>
<td>23.0</td>
<td>10.341</td>
<td>15.894</td>
<td>30.0</td>
<td>7.935</td>
<td>17.844</td>
</tr>
<tr>
<td>23.5</td>
<td>9.832</td>
<td>16.328</td>
<td>30.5</td>
<td>8.025</td>
<td>17.776</td>
</tr>
<tr>
<td>24.0</td>
<td>9.418</td>
<td>16.673</td>
<td>31.0</td>
<td>8.147</td>
<td>17.683</td>
</tr>
<tr>
<td>24.5</td>
<td>9.074</td>
<td>16.954</td>
<td>31.5</td>
<td>8.304</td>
<td>17.562</td>
</tr>
<tr>
<td>25.0</td>
<td>8.784</td>
<td>17.186</td>
<td>32.0</td>
<td>8.498</td>
<td>17.411</td>
</tr>
<tr>
<td>25.5</td>
<td>8.542</td>
<td>17.377</td>
<td>32.5</td>
<td>8.732</td>
<td>17.227</td>
</tr>
<tr>
<td>26.0</td>
<td>8.340</td>
<td>17.534</td>
<td>33.0</td>
<td>9.012</td>
<td>17.004</td>
</tr>
<tr>
<td>26.5</td>
<td>8.176</td>
<td>17.661</td>
<td>33.5</td>
<td>9.344</td>
<td>16.734</td>
</tr>
<tr>
<td>27.0</td>
<td>8.047</td>
<td>17.760</td>
<td>34.0</td>
<td>9.742</td>
<td>16.403</td>
</tr>
<tr>
<td>27.5</td>
<td>7.950</td>
<td>17.833</td>
<td>34.5</td>
<td>10.229</td>
<td>15.990</td>
</tr>
<tr>
<td>28.0</td>
<td>7.885</td>
<td>17.882</td>
<td>35.0</td>
<td>10.854</td>
<td>15.444</td>
</tr>
<tr>
<td>28.5</td>
<td>7.852</td>
<td>17.907</td>
<td>35.5</td>
<td>11.772</td>
<td>14.609</td>
</tr>
</tbody>
</table>

This data set is represented graphically in the following contour plot:
(Note that this plot is generated with degrees of freedom \( k = 1 \), as we are only determining bounds on one parameter. The contour plots generated in Weibull++ are done with degrees of freedom \( k = 2 \), for use in comparing both parameters simultaneously.) As can be determined from the table, the lowest calculated value for \( \sigma \) is 7.849, while the highest is 17.909. These represent the two-sided 80% confidence limits on this parameter. Since solutions for the equation do not exist for values of \( \mu \) below 22 or above 35.5, these can be considered the two-sided 80% confidence limits for this parameter. In order to obtain more accurate values for the confidence limits on \( \mu \), we can perform the same procedure as before, but finding the two values of \( \mu \) that correspond with a given value of \( \sigma \). Using this method, we find that the two-sided 80% confidence limits on \( \mu \) are 21.807 and 35.793, which are close to the initial estimates of 22 and 35.5.

**Bounds on Time and Reliability**

In order to calculate the bounds on a time estimate for a given reliability, or on a reliability estimate for a given time, the likelihood function needs to be rewritten in terms of one parameter and time/reliability, so that the maximum and minimum values of the time can be observed as the parameter is varied. This can be accomplished by substituting a form of the normal reliability equation into the likelihood function. The normal reliability equation can be written as:

\[
R = 1 - \Phi \left( \frac{t - \mu}{\sigma} \right)
\]

This can be rearranged to the form:

\[
\mu = t - \sigma \cdot \Phi^{-1}(1 - R)
\]

where \( \Phi^{-1} \) is the inverse standard normal. This equation can now be substituted into the likelihood ratio equation to produce an equation in terms of \( \sigma \), \( t \) and \( R \):

\[
L(\sigma, t/R) = \prod_{i=1}^{N} \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left( \frac{t_i - \left( t - \sigma \cdot \Phi^{-1}(1 - R) \right)}{\sigma} \right)^2}
\]

The unknown parameter \( t/R \) depends on what type of bounds are being determined. If one is trying to determine the bounds on time for a given reliability, then \( R \) is a known constant and \( t \) is the unknown parameter. Conversely, if one is trying to determine the bounds on reliability for a given time, then \( t \) is a known constant and \( R \) is the unknown parameter. The likelihood ratio equation can be used to solve the values of interest.
Example: LR Bounds on Time

For the same data set given above in the parameter bounds example, determine the two-sided 80% confidence bounds on the time estimate for a reliability of 40%. The ML estimate for the time at $R(t) = 40\%$ is 31.637.

Solution

In this example, we are trying to determine the two-sided 80% confidence bounds on the time estimate of 31.637. This is accomplished by substituting $R = 0.40$ and $\alpha = 0.8$ into the likelihood ratio equation for the normal distribution, and varying $\sigma$ until the maximum and minimum values of $t$ are found. The following table gives the values of $t$ based on given values of $\sigma$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$t_1$</th>
<th>$t_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.0</td>
<td>29.248</td>
<td>32.406</td>
</tr>
<tr>
<td>8.5</td>
<td>27.749</td>
<td>34.158</td>
</tr>
<tr>
<td>9.0</td>
<td>26.899</td>
<td>35.261</td>
</tr>
<tr>
<td>9.5</td>
<td>26.300</td>
<td>36.114</td>
</tr>
<tr>
<td>10.0</td>
<td>25.856</td>
<td>36.811</td>
</tr>
<tr>
<td>10.5</td>
<td>25.526</td>
<td>37.394</td>
</tr>
<tr>
<td>11.0</td>
<td>25.135</td>
<td>37.884</td>
</tr>
<tr>
<td>11.5</td>
<td>25.135</td>
<td>38.292</td>
</tr>
<tr>
<td>12.0</td>
<td>25.055</td>
<td>38.625</td>
</tr>
<tr>
<td>12.5</td>
<td>25.046</td>
<td>38.887</td>
</tr>
</tbody>
</table>

This data set is represented graphically in the following contour plot:

![Normal Contour Plot for Time vs. Sigma](image)

As can be determined from the table, the lowest calculated value for $t$ is 25.046, while the highest is 39.250. These represent the 80% confidence limits on the time at which reliability is equal to 40%.
Example: LR Bounds on Reliability

For the same data set given above in the parameter bounds and time bounds examples, determine the two-sided 80% confidence bounds on the reliability estimate for $t = 30$. The ML estimate for the reliability at $t = 30$ is 45.739%.

Solution

In this example, we are trying to determine the two-sided 80% confidence bounds on the reliability estimate of 45.739%. This is accomplished by substituting $t = 30$ and $\alpha = 0.8$ into the likelihood ratio equation for normal distribution, and varying $\sigma$ until the maximum and minimum values of $R$ are found. The following table gives the values of $R$ based on given values of $\sigma$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$\sigma$</th>
<th>$R_1$</th>
<th>$R_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.0</td>
<td>36.415%</td>
<td>51.890%</td>
<td>13.0</td>
<td>26.444%</td>
<td>67.188%</td>
</tr>
<tr>
<td>8.5</td>
<td>30.218%</td>
<td>59.320%</td>
<td>13.5</td>
<td>27.225%</td>
<td>66.576%</td>
</tr>
<tr>
<td>9.0</td>
<td>27.496%</td>
<td>62.976%</td>
<td>14.0</td>
<td>28.137%</td>
<td>65.812%</td>
</tr>
<tr>
<td>9.5</td>
<td>26.016%</td>
<td>65.181%</td>
<td>14.5</td>
<td>29.186%</td>
<td>64.893%</td>
</tr>
<tr>
<td>10.0</td>
<td>25.213%</td>
<td>66.560%</td>
<td>15.0</td>
<td>30.384%</td>
<td>63.810%</td>
</tr>
<tr>
<td>10.5</td>
<td>24.842%</td>
<td>67.396%</td>
<td>15.5</td>
<td>31.755%</td>
<td>62.541%</td>
</tr>
<tr>
<td>11.0</td>
<td>24.776%</td>
<td>67.845%</td>
<td>16.0</td>
<td>33.339%</td>
<td>61.048%</td>
</tr>
<tr>
<td>11.5</td>
<td>24.940%</td>
<td>68.000%</td>
<td>16.5</td>
<td>35.210%</td>
<td>59.258%</td>
</tr>
<tr>
<td>12.0</td>
<td>25.289%</td>
<td>67.919%</td>
<td>17.0</td>
<td>37.520%</td>
<td>57.022%</td>
</tr>
<tr>
<td>12.5</td>
<td>25.796%</td>
<td>67.640%</td>
<td>17.5</td>
<td>40.698%</td>
<td>53.910%</td>
</tr>
</tbody>
</table>

This data set is represented graphically in the following contour plot:

As can be determined from the table, the lowest calculated value for $R$ is 24.776%, while the highest is 68.000%. These represent the 80% two-sided confidence limits on the reliability at $t = 30$. 
Bayesian Confidence Bounds

Bounds on Parameters

From Confidence Bounds, we know that the marginal posterior distribution of $\mu$ can be written as:

$$f(\mu|\text{Data}) = \int_0^\infty f(\mu, \sigma|\text{Data})d\sigma$$

$$= \frac{\int_0^\infty L(\text{Data}|\mu, \sigma)\varphi(\mu)\varphi(\sigma)d\sigma}{\int_0^\infty \int_{-\infty}^\infty L(\text{Data}|\mu, \sigma)\varphi(\mu)\varphi(\sigma)d\mu d\sigma}$$

where:

$\varphi(\sigma)$ is the non-informative prior of $\sigma$.

$\varphi(\mu)$ is a uniform distribution from $-\infty$ to $+\infty$, the non-informative prior of $\mu$.

Using the above prior distributions, $f(\mu|\text{Data})$ can be rewritten as:

$$f(\mu|\text{Data}) = \frac{\int_0^\infty L(\text{Data}|\mu, \sigma)d\sigma}{\int_0^\infty L(\text{Data}|\mu, \sigma)\frac{1}{\sigma}d\sigma}$$

The one-sided upper bound of $\mu$ is:

$$CL = P(\mu \leq \mu_U) = \int_{-\infty}^{\mu_U} f(\mu|\text{Data})d\mu$$

The one-sided lower bound of $\mu$ is:

$$1 - CL = P(\mu \leq \mu_L) = \int_{-\infty}^{\mu_L} f(\mu|\text{Data})d\mu$$

The two-sided bounds of $\mu$ are:

$$CL = P(\mu_L \leq \mu \leq \mu_U) = \int_{\mu_L}^{\mu_U} f(\mu|\text{Data})d\mu$$

The same method can be used to obtain the bounds of $\sigma$.

Bounds on Time (Type 1)

The reliable life for the normal distribution is:

$$T = \mu + \sigma \Phi^{-1}(1 - R)$$

The one-sided upper bound on time is:

$$CL = P(T \leq T_U) = P(\mu + \sigma \Phi^{-1}(1 - R) \leq T_U)$$

The above equation can be rewritten in terms of $\mu$ as:

$$CL = P(\mu \leq T_U - \sigma \Phi^{-1}(1 - R))$$

From the posterior distribution of $\mu$:

$$CL = \frac{\int_0^\infty \int_{-\infty}^{\Phi^{-1}(1 - R) - \sigma}\int_{\sigma}^{\infty} L(\sigma, \mu)d\mu d\sigma}{\int_0^\infty \int_{-\infty}^{\infty} L(\sigma, \mu)d\mu d\sigma}$$

The same method can be applied for one-sided lower bounds and two-sided bounds on time.
Bounds on Reliability (Type 2)

The one-sided upper bound on reliability is:

\[ CL = \Pr (R \leq R_U) = \Pr (\mu \leq T - \sigma \Phi^{-1}(1 - R_U)) \]

From the posterior distribution of \( \mu \):

\[ CL = \frac{\int_{0}^{\infty} \int_{-\infty}^{T - \sigma \Phi^{-1}(1 - R_U)} L(\sigma, \mu) \frac{1}{\sigma} d\mu d\sigma}{\int_{0}^{\infty} \int_{-\infty}^{\infty} L(\sigma, \mu) \frac{1}{\sigma} d\mu d\sigma} \]

The same method can be used to calculate the one-sided lower bounds and the two-sided bounds on reliability.

Normal Distribution Examples

The following examples illustrate the different types of life data that can be analyzed in Weibull++ using the normal distribution. For more information on the different types of life data, see Life Data Classification.

Complete Data Example

6 units are tested to failure. The following failure times data are obtained: 12125, 11260, 12080, 12825, 13550 and 14670 hours. Assuming that the data are normally distributed, do the following:

Objectives

1. Find the parameters for the data set, using the Rank Regression on X (RRX) parameter estimation method.
2. Obtain the probability plot for the data with 90%, two-sided Type 1 confidence bounds.
3. Obtain the pdf plot for the data.
4. Using the Quick Calculation Pad (QCP), determine the reliability for a mission of 11,000 hours, as well as the upper and lower two-sided 90% confidence limit on this reliability.
5. Using the QCP, determine the MTTF, as well as the upper and lower two-sided 90% confidence limit on this MTTF.
6. Obtain tabulated values for the failure rate for 10 different mission end times. The mission end times are 1,000 to 10,000 hours, using increments of 1,000 hours.

Solution

The following figure shows the data as entered in Weibull++, as well as the calculated parameters.
The Normal Distribution

The following figures show the probability plot with the 90% two-sided confidence bounds and the pdf plot.
The Normal Distribution

Probability - Normal

Unreliability, F(t) = 1 - R(t)

Time (Hr)

Harry Gau
ReliaSoft
3/1/2012
4:51:09 PM

Mean = 1.2752E+04, Std = 1348.1848, Rho = 0.9798
Both the reliability and MTTF can be easily obtained from the QCP. The QCP, with results, for both cases is shown in the next two figures.
To obtain tabulated values for the failure rate, use the Analysis Workbook or General Spreadsheet features that are included in Weibull++. (For more information on these features, please refer to the Weibull++ User's Guide. For a step-by-step example on creating Weibull++ reports, please see the Quick Start Guide \(^{(1)}\)). The following worksheet shows the mission times and the corresponding failure rates.
Suspension Data Example

19 units are being reliability tested and the following is a table of their times-to-failure and suspensions.

<table>
<thead>
<tr>
<th>Data point index</th>
<th>Last Inspected</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>S</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>S</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>S</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>S</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>23</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>29</td>
</tr>
<tr>
<td>11</td>
<td>S</td>
<td>31</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>37</td>
</tr>
<tr>
<td>13</td>
<td>S</td>
<td>41</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>43</td>
</tr>
<tr>
<td>15</td>
<td>S</td>
<td>47</td>
</tr>
<tr>
<td>16</td>
<td>S</td>
<td>53</td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>59</td>
</tr>
<tr>
<td>18</td>
<td>S</td>
<td>61</td>
</tr>
<tr>
<td>19</td>
<td>S</td>
<td>67</td>
</tr>
</tbody>
</table>

Using the normal distribution and the maximum likelihood (MLE) parameter estimation method, the computed parameters are:

$$\hat{\mu} = 48.07$$

$$\hat{\sigma}_T = 28.41.$$

If we analyze the data set with the rank regression on x (RRX) method, the computed parameters are:

$$\hat{\mu} = 46.40$$

$$\hat{\sigma}_T = 28.64.$$

For the rank regression on y (RRY) method, the parameters are:

$$\hat{\mu} = 47.34$$

$$\hat{\sigma}_T = 29.96.$$
Interval Censored Data Example

8 units are being reliability tested, and the following is a table of their failure times:

<table>
<thead>
<tr>
<th>Data point index</th>
<th>Last Inspected</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>55</td>
<td>55</td>
</tr>
</tbody>
</table>

This is a sequence of interval times-to-failure data. Using the normal distribution and the maximum likelihood (MLE) parameter estimation method, the computed parameters are:

\[
\hat{\mu} = 41.40 \\
\hat{\sigma}_T = 7.740.
\]

For rank regression on x:

\[
\hat{\mu} = 41.40 \\
\hat{\sigma}_T = 9.03.
\]

If we analyze the data set with the rank regression on y (RRY) parameter estimation method, the computed parameters are:

\[
\hat{\mu} = 41.39 \\
\hat{\sigma}_T = 9.25.
\]

The following plot shows the results if the data were analyzed using the rank regression on X (RRX) method.
Mixed Data Types Example

Suppose our data set includes left and right censored, interval censored and complete data, as shown in the following table.

<p>| Grouped Data Times-to-Failure with Suspensions and Intervals (Interval, Left and Right Censored) |
|-------------------------------------------------|---------------------------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Data point index</th>
<th>Number in State</th>
<th>Last Inspection</th>
<th>State (S or F)</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>10</td>
<td>F</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>20</td>
<td>S</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>F</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>40</td>
<td>F</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>50</td>
<td>F</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>60</td>
<td>S</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>70</td>
<td>F</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>80</td>
<td>F</td>
<td>80</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>10</td>
<td>F</td>
<td>85</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>100</td>
<td>F</td>
<td>100</td>
</tr>
</tbody>
</table>

Using the normal distribution and the maximum likelihood (MLE) parameter estimation method, the computed parameters are:

\[
\hat{\mu} = 48.11
\]

\[
\hat{\sigma}_T = 26.42
\]
If we analyze the data set with the rank regression on x (RRX) method, the computed parameters are:

\[
\hat{\mu} = 49.99 \\
\hat{\sigma}_T = 30.17
\]

For the rank regression on y (RRY) method, the parameters are:

\[
\hat{\mu} = 51.61 \\
\hat{\sigma}_T = 33.07
\]

**Comparison of Analysis Methods**

8 units are being reliability tested, and the following is a table of their failure times:

<table>
<thead>
<tr>
<th>Data point index</th>
<th>State F or S</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>59</td>
</tr>
</tbody>
</table>

Using the normal distribution and the maximum likelihood (MLE) parameter estimation method, the computed parameters are:

\[
\hat{\mu} = 26.13 \\
\hat{\sigma}_T = 18.57
\]

If we analyze the data set with the rank regression on x (RRX) method, the computed parameters are:

\[
\hat{\mu} = 26.13 \\
\hat{\sigma}_T = 21.64
\]

For the rank regression on y (RRY) method, the parameters are:

\[
\hat{\mu} = 26.13 \\
\hat{\sigma}_T = 22.28.
\]

**References**

Chapter 10

The Lognormal Distribution

The lognormal distribution is commonly used to model the lives of units whose failure modes are of a fatigue-stress nature. Since this includes most, if not all, mechanical systems, the lognormal distribution can have widespread application. Consequently, the lognormal distribution is a good companion to the Weibull distribution when attempting to model these types of units. As may be surmised by the name, the lognormal distribution has certain similarities to the normal distribution. A random variable is lognormally distributed if the logarithm of the random variable is normally distributed. Because of this, there are many mathematical similarities between the two distributions. For example, the mathematical reasoning for the construction of the probability plotting scales and the bias of parameter estimators is very similar for these two distributions.

Lognormal Probability Density Function

The lognormal distribution is a 2-parameter distribution with parameters $\mu'$ and $\sigma'$. The pdf for this distribution is given by:

$$f(t') = \frac{1}{\sigma' \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t' - \mu'}{\sigma'} \right)^2}$$

where:

- $t'$ values are the times-to-failure
- $\mu'$ = mean of the natural logarithms of the times-to-failure
- $\sigma'$ = standard deviation of the natural logarithms of the times-to-failure

The lognormal pdf can be obtained, realizing that for equal probabilities under the normal and lognormal pdfs, incremental areas should also be equal, or:

$$f(t) dt = f(t') dt'$$

Taking the derivative yields:

$$dt' = \frac{dt}{t}$$

Substitution yields:

$$f(t) = \frac{f(t')}{t}$$

$$f(t) = \frac{1}{t \cdot \sigma' \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln(t) - \mu'}{\sigma'} \right)^2}$$

where:

$$f(t) \geq 0, t > 0, -\infty < \mu' < \infty, \sigma' > 0$$
Lognormal Distribution Functions

The Mean or MTTF

The mean of the lognormal distribution, $\mu$, is discussed in Kececioglu [19]:

$$\mu = e^{\mu' + \frac{1}{2} \sigma'^2}$$

The mean of the natural logarithms of the times-to-failure, $\mu'$, in terms of $\bar{T}$ and $\sigma$ is given by:

$$\mu' = \ln(\bar{T}) - \frac{1}{2} \ln(\frac{\sigma^2}{\bar{T}^2} + 1)$$

The Median

The median of the lognormal distribution, $\hat{T}$, is discussed in Kececioglu [19]:

$$\hat{T} = e^{\mu'}$$

The Mode

The mode of the lognormal distribution, $\tilde{T}$, is discussed in Kececioglu [19]:

$$\tilde{T} = e^{\mu' - \frac{1}{2} \sigma'^2}$$

The Standard Deviation

The standard deviation of the lognormal distribution, $\sigma_T$, is discussed in Kececioglu [19]:

$$\sigma_T = \sqrt{(e^{2\mu' + \sigma'^2}) \left(e^{\sigma'^2} - 1\right)}$$

The standard deviation of the natural logarithms of the times-to-failure, $\sigma'$, in terms of $\bar{T}$ and $\sigma$ is given by:

$$\sigma' = \sqrt{\ln\left(\frac{\sigma^2}{\bar{T}^2} + 1\right)}$$

The Lognormal Reliability Function

The reliability for a mission of time $T$, starting at age 0, for the lognormal distribution is determined by:

$$R(t) = \int_t^{\infty} f(x) \, dx$$

or:

$$R(t) = \int_{\ln(t)}^{\infty} \frac{1}{\sigma'\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu'}{\sigma'}\right)^2} \, dx$$

As with the normal distribution, there is no closed-form solution for the lognormal reliability function. Solutions can be obtained via the use of standard normal tables. Since the application automatically solves for the reliability we will not discuss manual solution methods. For interested readers, full explanations can be found in the references.

The Lognormal Conditional Reliability Function

The lognormal conditional reliability function is given by:

$$R(t|T) = \frac{R(T + t)}{R(T)} = \int_{\ln(T + t)}^{\infty} \frac{1}{\sigma'\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu'}{\sigma'}\right)^2} \, dx$$

Once again, the use of standard normal tables is necessary to solve this equation, as no closed-form solution exists.
The Lognormal Reliable Life Function

As there is no closed-form solution for the lognormal reliability equation, no closed-form solution exists for the lognormal reliable life either. In order to determine this value, one must solve the following equation for $t$:

$$R_t = \frac{1}{\sigma' \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\mu - \mu'}{\sigma'} \right)^2} \int_{\ln(t)}^{\infty} \frac{1}{i \sigma' \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu'}{\sigma'} \right)^2} dx$$

The Lognormal Failure Rate Function

The lognormal failure rate is given by:

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{1}{\sigma' \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\mu - \mu'}{\sigma'} \right)^2} \int_{\ln(t)}^{\infty} \frac{1}{i \sigma' \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu'}{\sigma'} \right)^2} dx$$

As with the reliability equations, standard normal tables will be required to solve for this function.

Characteristics of the Lognormal Distribution

- The lognormal distribution is a distribution skewed to the right.
- The pdf starts at zero, increases to its mode, and decreases thereafter.
- The degree of skewness increases as $\sigma'$ increases, for a given $\mu'$. 

![Effect of $\sigma'$ on Lognormal pdf](image-url)
The Lognormal Distribution

- For the same $\sigma'$, the pdf's skewness increases as $\mu'$ increases.
- For $\sigma'$ values significantly greater than 1, the pdf rises very sharply in the beginning, (i.e., for very small values of $T$ near zero), and essentially follows the ordinate axis, peaks out early, and then decreases sharply like an exponential pdf or a Weibull pdf with $0 < \beta < 1$.
- The parameter, $\mu'$, in terms of the logarithm of the $T'$, is also the scale parameter, and not the location parameter as in the case of the normal pdf.
- The parameter $\sigma'$, or the standard deviation of the $T'$, in terms of their logarithm or of their $T'$, is also the shape parameter and not the scale parameter, as in the normal pdf, and assumes only positive values.

Lognormal Distribution Parameters in ReliaSoft's Software

In ReliaSoft's software, the parameters returned for the lognormal distribution are always logarithmic. That is: the parameter $\mu'$ represents the mean of the natural logarithms of the times-to-failure, while $\sigma'$ represents the standard deviation of these data point logarithms. Specifically, the returned $\sigma'$ is the square root of the variance of the natural logarithms of the data points. Even though the application denotes these values as mean and standard deviation, the user is reminded that these are given as the parameters of the distribution, and are thus the mean and standard deviation of the natural logarithms of the data. The mean value of the times-to-failure, not used as a parameter, as well as the standard deviation can be obtained through the QCP or the Function Wizard.

Estimation of the Parameters

Probability Plotting

As described before, probability plotting involves plotting the failure times and associated unreliability estimates on specially constructed probability plotting paper. The form of this paper is based on a linearization of the cdf of the specific distribution. For the lognormal distribution, the cumulative density function can be written as:

$$F(t') = \Phi \left( \frac{t' - \mu'}{\sigma'} \right)$$

or:
The Lognormal Distribution

\[ \Phi^{-1} [F(t')] = -\frac{\mu'}{\sigma'} + \frac{1}{\sigma'} \cdot t' \]

where:

\[ \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \]

Now, let:

\[ y = \Phi^{-1} [F(t')] \]

\[ a = -\frac{\mu'}{\sigma'} \]

and:

\[ b = \frac{1}{\sigma'} \]

which results in the linear equation of:

\[ y = a + bt' \]

The normal probability paper resulting from this linearized cdf function is shown next.

The process for reading the parameter estimate values from the lognormal probability plot is very similar to the method employed for the normal distribution (see The Normal Distribution). However, since the lognormal distribution models the natural logarithms of the times-to-failure, the values of the parameter estimates must be read and calculated based on a logarithmic scale, as opposed to the linear time scale as it was done with the normal distribution. This parameter scale appears at the top of the lognormal probability plot.

The process of lognormal probability plotting is illustrated in the following example.
Plotting Example

8 units are put on a life test and tested to failure. The failures occurred at 45, 140, 260, 500, 850, 1400, 3000, and 9000 hours. Estimate the parameters for the lognormal distribution using probability plotting.

Solution

In order to plot the points for the probability plot, the appropriate unreliability estimate values must be obtained. These will be estimated through the use of median ranks, which can be obtained from statistical tables or the Quick Statistical Reference in Weibull++. The following table shows the times-to-failure and the appropriate median rank values for this example:

<table>
<thead>
<tr>
<th>Time-to-Failure (hr.)</th>
<th>Median Rank (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>8.30%</td>
</tr>
<tr>
<td>140</td>
<td>20.11%</td>
</tr>
<tr>
<td>260</td>
<td>32.05%</td>
</tr>
<tr>
<td>500</td>
<td>44.02%</td>
</tr>
<tr>
<td>850</td>
<td>55.98%</td>
</tr>
<tr>
<td>1400</td>
<td>67.95%</td>
</tr>
<tr>
<td>3000</td>
<td>79.89%</td>
</tr>
<tr>
<td>9000</td>
<td>91.70%</td>
</tr>
</tbody>
</table>

These points may now be plotted on normal probability plotting paper as shown in the next figure.

Draw the best possible line through the plot points. The time values where this line intersects the 15.85% and 50% unreliability values should be projected up to the logarithmic scale, as shown in the following plot.
The natural logarithm of the time where the fitted line intersects is equivalent to \( \mu' \). In this case, \( \mu' = 6.45 \). The value for \( \sigma_T \) is equal to the difference between the natural logarithms of the times where the fitted line crosses \( Q(t) = 50\% \) and \( Q(t) = 15.85\% \). At \( Q(t) = 15.85\% \), \( \ln(t) = 4.55 \). Therefore, \( \sigma' = 6.45 - 4.55 = 1.9 \).

**Rank Regression on Y**

Performing a rank regression on Y requires that a straight line be fitted to a set of data points such that the sum of the squares of the vertical deviations from the points to the line is minimized.

The least squares parameter estimation method, or regression analysis, was discussed in Parameter Estimation and the following equations for regression on Y were derived, and are again applicable:

\[
\hat{a} = \bar{y} - \hat{b} \bar{x} = \frac{\sum_{i=1}^{N} y_i}{N} - \hat{b} \frac{\sum_{i=1}^{N} x_i}{N}
\]

and:

\[
\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \left( \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N} \right)}{\sum_{i=1}^{N} x_i^2 - \left( \frac{\sum_{i=1}^{N} x_i}{N} \right) \left( \frac{\sum_{i=1}^{N} x_i}{N} \right)}
\]

In our case the equations for \( \hat{y} \) and \( \hat{x} \) are:

\[
y_i = \Phi^{-1} \left[ F(t'_i) \right]
\]

and:
where the $F(t_i)$s is estimated from the median ranks. Once $\hat{\alpha}$ and $\hat{\beta}$ are obtained, then $\hat{\sigma}$ and $\hat{\mu}$ can easily be obtained from the above equations.

The Correlation Coefficient

The estimator of $\rho$ is the sample correlation coefficient, $\hat{\rho}$, given by:

$$\hat{\rho} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2}}$$

RRY Example

Lognormal Distribution RRY Example

14 units were reliability tested and the following life test data were obtained:

<table>
<thead>
<tr>
<th>Data point index</th>
<th>Time-to-failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>11</td>
<td>70</td>
</tr>
<tr>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td>13</td>
<td>90</td>
</tr>
<tr>
<td>14</td>
<td>100</td>
</tr>
</tbody>
</table>

Assuming the data follow a lognormal distribution, estimate the parameters and the correlation coefficient, $\hat{\rho}$, using rank regression on $Y$.

Solution

Construct a table like the one shown next.

Least Squares Analysis
The Lognormal Distribution

The median rank values \( \tilde{F}(t_i) \) can be found in rank tables or by using the Quick Statistical Reference in Weibull++.

The \( y_i \) values were obtained from the standardized normal distribution’s area tables by entering for \( F(z) \) and getting the corresponding \( z \) value (\( y_i \)).

Given the values in the table above, calculate \( \hat{a} \) and \( \hat{b} \):

\[
\hat{b} = \frac{\sum_{i=1}^{14} t_i^2 y_i - \left( \sum_{i=1}^{14} t_i \right) \left( \sum_{i=1}^{14} y_i \right) / 14}{\sum_{i=1}^{14} t_i^2 - \left( \sum_{i=1}^{14} t_i \right)^2 / 14}
\]

\[
\hat{b} = \frac{10.4473 - (49.2220)(0)/14}{183.1530 - (49.2220)^2 / 14}
\]

or:

\[ \hat{b} = 1.0349 \]

and:

\[
\hat{a} = \frac{\sum_{i=1}^{N} y_i - \hat{b} \sum_{i=1}^{N} t_i}{N}
\]

or:

\[ \hat{a} = 0 - (1.0349) \frac{49.2220}{14} = -3.6386 \]

Therefore:

\[ \sigma' = \frac{1}{\hat{b}} = \frac{1}{1.0349} = 0.9663 \]

and:

\[ \mu' = -\hat{a} \cdot \sigma' = -( -3.6386 ) \cdot 0.9663 \]

or:

\[ \mu' = 3.516 \]
The mean and the standard deviation of the lognormal distribution are obtained using equations in the Lognormal Distribution Functions section above:

\[ \bar{T} = \mu = e^{3.510 + \frac{1}{2}0.9663^2} = 53.6707 \text{ hours} \]

and:

\[ \sigma = \sqrt{(e^{2.3.510 + 0.9663^2})(e^{0.9663^2} - 1)} = 66.69 \text{ hours} \]

The correlation coefficient can be estimated as:

\[ \hat{\rho} = 0.9754 \]

The above example can be repeated using Weibull++ , using RRY.

The mean can be obtained from the QCP and both the mean and the standard deviation can be obtained from the Function Wizard.

**Rank Regression on X**

Performing a rank regression on X requires that a straight line be fitted to a set of data points such that the sum of the squares of the horizontal deviations from the points to the line is minimized.

Again, the first task is to bring our cdf function into a linear form. This step is exactly the same as in regression on Y analysis and all the equations apply in this case too. The deviation from the previous analysis begins on the least squares fit part, where in this case we treat \( T \) as the dependent variable and \( Y \) as the independent variable. The best-fitting straight line to the data, for regression on X (see Parameter Estimation), is the straight line:

\[ x = \hat{a} + \hat{b}y \]

The corresponding equations for \( \hat{a} \) and \( \hat{b} \) are:
The Lognormal Distribution

\[
\hat{a} = \bar{x} - \hat{b}\bar{y} = \frac{\sum_{i=1}^{N} x_i}{N} - \hat{b}\frac{\sum_{i=1}^{N} y_i}{N}
\]

and:

\[
\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N}}
\]

where:

\[y_i = \Phi^{-1}\left[F(t'_i)\right]\]

and:

\[x_i = t'_i\]

and the \(F(t'_i)\)s estimated from the median ranks. Once \(\hat{a}\) and \(\hat{b}\) are obtained, solve the linear equation for the unknown \(\hat{y}\), which corresponds to:

\[y = \frac{\hat{a}}{\hat{b}} + \frac{1}{\hat{b}}x\]

Solving for the parameters we get:

\[a = \frac{\hat{a}}{\hat{b}} = -\frac{\mu'}{\sigma'}\]

and:

\[b = \frac{1}{\hat{b}} = \frac{1}{\sigma'}\]

The correlation coefficient is evaluated as before using equation in the previous section.

**RRX Example**

**Lognormal Distribution RRX Example**

Using the same data set from the RRY example given above, and assuming a lognormal distribution, estimate the parameters and estimate the correlation coefficient, \(\hat{\rho}\), using rank regression on X.

**Solution**

The table constructed for the RRY example also applies to this example as well. Using the values in this table we get:

\[
\hat{b} = \frac{\sum_{i=1}^{14} t'_i y_i - \frac{\sum_{i=1}^{14} t'_i \sum_{i=1}^{14} y_i}{14}}{\sum_{i=1}^{14} y_i^2 - \frac{\left(\sum_{i=1}^{14} y_i\right)^2}{14}}
\]

\[
\hat{b} = \frac{10.4473 - (49.2220)(0)/14}{11.3646 - (0)^2/14}
\]

or:

\[\hat{b} = 0.9193\]

and:
The Lognormal Distribution

\[ \hat{a} = \bar{x} - b \bar{y} = \frac{\sum_{i=1}^{14} x_i}{14} - \frac{\sum_{i=1}^{14} y_i}{14} \]

or:

\[ \hat{a} = \frac{49.2220}{14} - (0.9193)(0) \cdot \frac{1}{14} = 3.5159 \]

Therefore:

\[ \sigma' = \hat{b} = 0.9193 \]

and:

\[ \mu' = \frac{\hat{a} \sigma'}{\hat{b}} = \frac{3.5159}{0.9193} \cdot 0.9193 = 3.5159 \]

Using for Mean and Standard Deviation we get:

\[ \bar{T} = \mu = 51.3398 \text{ hours} \]

and:

\[ \sigma' = 59.1682 \text{ hours.} \]

The correlation coefficient is found using the equation in previous section:

\[ \hat{\rho} = 0.9754. \]

Note that the regression on Y analysis is not necessarily the same as the regression on X. The only time when the results of the two regression types are the same (i.e., will yield the same equation for a line) is when the data lie perfectly on a line.

Using Weibull++ , with the Rank Regression on X option, the results are:
Maximum Likelihood Estimation

As it was outlined in Parameter Estimation, maximum likelihood estimation works by developing a likelihood function based on the available data and finding the values of the parameter estimates that maximize the likelihood function. This can be achieved by using iterative methods to determine the parameter estimate values that maximize the likelihood function. However, this can be rather difficult and time-consuming, particularly when dealing with the three-parameter distribution. Another method of finding the parameter estimates involves taking the partial derivatives of the likelihood equation with respect to the parameters, setting the resulting equations equal to zero, and solving simultaneously to determine the values of the parameter estimates. The log-likelihood functions and associated partial derivatives used to determine maximum likelihood estimates for the lognormal distribution are covered in Appendix D.

Note About Bias

See the discussion regarding bias with the normal distribution for information regarding parameter bias in the lognormal distribution.

MLE Example

Lognormal Distribution MLE Example

Using the same data set from the RRY and RRX examples given above and assuming a lognormal distribution, estimate the parameters using the MLE method.

Solution In this example we have only complete data. Thus, the partials reduce to:

\[
\frac{\partial \Lambda}{\partial \mu'} = \frac{1}{\sigma'^2} \sum_{i=1}^{14} \ln(t_i) - \mu' = 0
\]

\[
\frac{\partial \Lambda}{\partial \sigma'^2} = \sum_{i=1}^{14} \left( \frac{\ln(t_i) - \mu'}{\sigma'^3} - \frac{1}{\sigma'^4} \right) = 0
\]

Substituting the values of \(T\) and solving the above system simultaneously, we get:

\(\hat{\sigma}' = 0.849\)

\(\hat{\mu}' = 3.516\)

Using the equation for mean and standard deviation in the Lognormal Distribution Functions section above, we get:

\(\bar{T} = \mu = 48.25 \text{ hours}\)

and:

\(\hat{\sigma} = 49.61 \text{ hours}\).

The variance/covariance matrix is given by:

\[
\begin{bmatrix}
\mathrm{Var}(\hat{\mu}') &=& 0.0515 \\
\mathrm{Cov}(\hat{\mu}', \hat{\sigma}') &=& 0.0000 \\
\mathrm{Cov}(\hat{\mu}', \hat{\sigma}') &=& 0.0000 \\
\mathrm{Var}(\hat{\sigma}') &=& 0.0258
\end{bmatrix}
\]

Confidence Bounds

The method used by the application in estimating the different types of confidence bounds for lognormally distributed data is presented in this section. Note that there are closed-form solutions for both the normal and lognormal reliability that can be obtained without the use of the Fisher information matrix. However, these closed-form solutions only apply to complete data. To achieve consistent application across all possible data types, Weibull++ always uses the Fisher matrix in computing confidence intervals. The complete derivations were presented in detail for a general function in Confidence Bounds. For a discussion on exact confidence bounds for the
normal and lognormal, see The Normal Distribution.

**Fisher Matrix Bounds**

**Bounds on the Parameters**

The lower and upper bounds on the mean, \( \mu' \), are estimated from:

\[
\mu'_L = \hat{\mu'} - K_\alpha \sqrt{Var(\hat{\mu'})} \quad \text{(lower bound)},
\]

\[
\mu'_U = \hat{\mu'} + K_\alpha \sqrt{Var(\hat{\mu'})} \quad \text{(upper bound)},
\]

For the standard deviation, \( \sigma' \), \( \ln(\sigma') \) is treated as normally distributed, and the bounds are estimated from:

\[
\sigma'_L = \frac{\sigma'}{e^{K_\alpha \sqrt{Var(\sigma')}}} \quad \text{(lower bound)},
\]

\[
\sigma'_U = \sigma' \cdot e^{K_\alpha \sqrt{Var(\sigma')}} \quad \text{(upper bound)},
\]

where \( K_\alpha \) is defined by:

\[
\alpha = \frac{1}{\sqrt{2\pi}} \int_{K_\alpha}^\infty e^{-\frac{t^2}{2}} dt = 1 - \Phi(K_\alpha)
\]

If \( \delta \) is the confidence level, then \( \alpha = \frac{1-\delta}{2} \) for the two-sided bounds and \( \alpha = 1 - \delta \) for the one-sided bounds.

The variances and covariances of \( \mu' \) and \( \sigma' \) are estimated as follows:

\[
\begin{pmatrix}
\text{Var}(\hat{\mu'}) & \text{Cov}\left(\hat{\mu'}, \hat{\sigma'}\right) \\
\text{Cov}\left(\hat{\mu'}, \hat{\sigma'}\right) & \text{Var}(\hat{\sigma'})
\end{pmatrix} = \begin{pmatrix}
-\frac{\partial^2 \Lambda}{\partial (\mu')^2} & -\frac{\partial^2 \Lambda}{\partial (\mu') \partial (\sigma')} \\
-\frac{\partial^2 \Lambda}{\partial (\mu') \partial (\sigma')} & -\frac{\partial^2 \Lambda}{\partial (\sigma')^2}
\end{pmatrix}^{-1}
\]

where \( \Lambda \) is the log-likelihood function of the lognormal distribution.

**Bounds on Time (Type 1)**

The bounds around time for a given lognormal percentile, or unreliability, are estimated by first solving the reliability equation with respect to time, as follows:

\[
\hat{t}'(\hat{\mu'}, \hat{\sigma'}) = \hat{\mu'} + z \cdot \hat{\sigma'}
\]

where:

\[
z = \Phi^{-1} [F(t')]
\]

and:

\[
\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}z^2} dz
\]

The next step is to calculate the variance of \( T'(\hat{\mu'}, \hat{\sigma'}) \):

\[
Var(\hat{t}') = \left( \frac{\partial t'}{\partial \mu'} \right)^2 Var(\hat{\mu'}) + \left( \frac{\partial t'}{\partial \sigma'} \right)^2 Var(\hat{\sigma'})
\]

\[
+ 2 \left( \frac{\partial t'}{\partial \mu'} \right) \left( \frac{\partial t'}{\partial \sigma'} \right) Cov(\hat{\mu'}, \hat{\sigma'})
\]

\[
Var(\hat{t}') = Var(\hat{\mu'}) + z^2 Var(\hat{\sigma'}) + 2 \cdot z \cdot Cov(\hat{\mu'}, \hat{\sigma'})
\]

The upper and lower bounds are then found by:
Bounds on Reliability (Type 2)

The reliability of the lognormal distribution is:

\[ \hat{R}(t; \hat{\mu}', \hat{\sigma}') = \int_{t'}^{\infty} \frac{1}{\hat{\sigma}' \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \hat{\mu}'}{\hat{\sigma}'} \right)^2} dx \]

where \( t' = \ln(t) \). Let \( \hat{z}(x) = \frac{x - \hat{\mu}'}{\hat{\sigma}'} \), the above equation then becomes:

\[ \hat{R}(\hat{z}(t')) = \int_{\hat{z}(t')}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2} dz \]

The bounds on \( z \) are estimated from:

\[ z_U = \hat{z} + K_\alpha \sqrt{Var(\hat{z})} \]
\[ z_L = \hat{z} - K_\alpha \sqrt{Var(\hat{z})} \]

where:

\[ Var(\hat{z}) = \left( \frac{\partial \hat{z}}{\partial \hat{\mu}'} \right)^2 Var(\hat{\mu}') + \left( \frac{\partial \hat{z}}{\partial \hat{\sigma}'} \right)^2 Var(\hat{\sigma}') + 2 \left( \frac{\partial \hat{z}}{\partial \hat{\mu}'} \right) \left( \frac{\partial \hat{z}}{\partial \hat{\sigma}'} \right) Cov(\hat{\mu}', \hat{\sigma}') \]

or:

\[ Var(\hat{z}) = \frac{1}{\hat{\sigma}^2} \left[ Var(\hat{\mu}') + \hat{z}^2 Var(\hat{\sigma}') + 2 \cdot \hat{z} \cdot Cov(\hat{\mu}', \hat{\sigma}') \right] \]

The upper and lower bounds on reliability are:

\[ R_U = \int_{z_L}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2} dz \quad \text{(Upper bound)} \]
\[ R_L = \int_{z_U}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2} dz \quad \text{(Lower bound)} \]

Likelihood Ratio Confidence Bounds

Bounds on Parameters

As covered in Parameter Estimation, the likelihood confidence bounds are calculated by finding values for \( \theta_1 \) and \( \theta_2 \) that satisfy:

\[ -2 \cdot \ln \left( \frac{L(\hat{\theta}_1, \hat{\theta}_2)}{L(\hat{\theta}_1, \hat{\theta}_2)} \right) = \chi^2_{\alpha; 1} \]

This equation can be rewritten as:

\[ L(\hat{\theta}_1, \hat{\theta}_2) = L(\hat{\theta}_1, \hat{\theta}_2) \cdot e^{-\chi^2_{\alpha; 1} \cdot 2} \]

For complete data, the likelihood formula for the normal distribution is given by:
where the $x_i$ values represent the original time-to-failure data. For a given value of $\alpha$, values for $\mu'$ and $\sigma'$ can be found which represent the maximum and minimum values that satisfy likelihood ratio equation. These represent the confidence bounds for the parameters at a confidence level $\delta$, where $\alpha = 2\delta - 1$ for two-sided bounds and $\alpha = \delta$ for one-sided.

**Example: LR Bounds on Parameters**

**Lognormal Distribution Likelihood Ratio Bound Example (Parameters)**

Five units are put on a reliability test and experience failures at 45, 60, 75, 90, and 115 hours. Assuming a lognormal distribution, the MLE parameter estimates are calculated to be $\hat{\mu}' = 4.292$ and $\hat{\sigma}' = 0.32361$. Calculate the two-sided 75% confidence bounds on these parameters using the likelihood ratio method.

**Solution**

The first step is to calculate the likelihood function for the parameter estimates:

$$L(\hat{\mu}', \hat{\sigma}') = \prod_{i=1}^{N} f(x_i; \hat{\mu}', \hat{\sigma}'),$$

where $x_i$ are the original time-to-failure data points. We can now rearrange the likelihood ratio equation to the form:

$$L(\mu', \sigma') = L(\hat{\mu}', \hat{\sigma}') \cdot e^{-\frac{\chi^2}{2}} = 0$$

Since our specified confidence level, $\delta$, is 75%, we can calculate the value of the chi-squared statistic, $\chi^2_{0.75, 1} = 1.323303$. We can now substitute this information into the equation:

$$L(\mu', \sigma') - L(\hat{\mu}', \hat{\sigma}') \cdot e^{-\frac{\chi^2_{0.75, 1}}{2}} = 0$$

$$L(\mu', \sigma') - 1.115256 \times 10^{-10} \cdot e^{-1.323303} = 0$$

$$L(\mu', \sigma') - 5.754703 \times 10^{-11} = 0$$

It now remains to find the values of $\mu'$ and $\sigma'$ which satisfy this equation. This is an iterative process that requires setting the value of $\sigma'$ and finding the appropriate values of $\mu'$, and vice versa.

The following table gives the values of $\mu'$ based on given values of $\sigma'$.
The Lognormal Distribution

<table>
<thead>
<tr>
<th>$\sigma'$</th>
<th>$\mu_1'$</th>
<th>$\mu_2'$</th>
<th>$\sigma'$</th>
<th>$\mu_1'$</th>
<th>$\mu_2'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>4.2421</td>
<td>4.3432</td>
<td>0.37</td>
<td>4.1145</td>
<td>4.4708</td>
</tr>
<tr>
<td>0.25</td>
<td>4.2115</td>
<td>4.3738</td>
<td>0.38</td>
<td>4.1152</td>
<td>4.4701</td>
</tr>
<tr>
<td>0.26</td>
<td>4.1909</td>
<td>4.3944</td>
<td>0.39</td>
<td>4.1170</td>
<td>4.4683</td>
</tr>
<tr>
<td>0.27</td>
<td>4.1748</td>
<td>4.4105</td>
<td>0.40</td>
<td>4.1200</td>
<td>4.4653</td>
</tr>
<tr>
<td>0.28</td>
<td>4.1618</td>
<td>4.4235</td>
<td>0.41</td>
<td>4.1244</td>
<td>4.4609</td>
</tr>
<tr>
<td>0.29</td>
<td>4.1509</td>
<td>4.4344</td>
<td>0.42</td>
<td>4.1302</td>
<td>4.4551</td>
</tr>
<tr>
<td>0.30</td>
<td>4.1419</td>
<td>4.4434</td>
<td>0.43</td>
<td>4.1377</td>
<td>4.4476</td>
</tr>
<tr>
<td>0.31</td>
<td>4.1343</td>
<td>4.4510</td>
<td>0.44</td>
<td>4.1472</td>
<td>4.4381</td>
</tr>
<tr>
<td>0.32</td>
<td>4.1281</td>
<td>4.4572</td>
<td>0.45</td>
<td>4.1591</td>
<td>4.4262</td>
</tr>
<tr>
<td>0.33</td>
<td>4.1231</td>
<td>4.4622</td>
<td>0.46</td>
<td>4.1742</td>
<td>4.4111</td>
</tr>
<tr>
<td>0.34</td>
<td>4.1193</td>
<td>4.4660</td>
<td>0.47</td>
<td>4.1939</td>
<td>4.3914</td>
</tr>
<tr>
<td>0.35</td>
<td>4.1166</td>
<td>4.4687</td>
<td>0.48</td>
<td>4.2221</td>
<td>4.3632</td>
</tr>
<tr>
<td>0.36</td>
<td>4.1150</td>
<td>4.4703</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These points are represented graphically in the following contour plot:

(Note that this plot is generated with degrees of freedom $\hat{k} = 1$, as we are only determining bounds on one parameter. The contour plots generated in Weibull++ are done with degrees of freedom $\hat{k} = 2$, for use in comparing both parameters simultaneously.) As can be determined from the table the lowest calculated value for $\mu'$ is 4.1145, while the highest is 4.4708. These represent the two-sided 75% confidence limits on this parameter. Since solutions for the equation do not exist for values of $\sigma'$ below 0.24 or above 0.48, these can be considered the two-sided 75% confidence limits for this parameter. In order to obtain more accurate values for the confidence limits on $\sigma'$, we can perform the same procedure as before, but finding the two values of $\sigma'$ that correspond with a given value of $\mu'$. Using this method, we find that the 75% confidence limits on $\sigma'$ are 0.23405 and 0.48936, which are close to the initial estimates of 0.24 and 0.48.)
Bounds on Time and Reliability

In order to calculate the bounds on a time estimate for a given reliability, or on a reliability estimate for a given time, the likelihood function needs to be rewritten in terms of one parameter and time/reliability, so that the maximum and minimum values of the time can be observed as the parameter is varied. This can be accomplished by substituting a form of the normal reliability equation into the likelihood function. The normal reliability equation can be written as:

\[ R = 1 - \Phi \left( \frac{\ln(t) - \mu'}{\sigma'} \right) \]

This can be rearranged to the form:

\[ \mu' = \ln(t) - \sigma' \cdot \Phi^{-1}(1 - R) \]

where \( \Phi^{-1} \) is the inverse standard normal. This equation can now be substituted into likelihood function to produce a likelihood equation in terms of \( \sigma', t \) and \( R \):

\[ L(\sigma', t/R) = \prod_{i=1}^{N} \frac{1}{x_i \cdot \sigma' \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left( \frac{\ln(x_i) - (\ln(t) - \sigma' \cdot \Phi^{-1}(1 - R))}{\sigma'} \right)^2} \]

The unknown variable \( t/R \) depends on what type of bounds are being determined. If one is trying to determine the bounds on time for a given reliability, then \( R \) is a known constant and \( t \) is the unknown variable. Conversely, if one is trying to determine the bounds on reliability for a given time, then \( t \) is a known constant and \( R \) is the unknown variable. Either way, the above equation can be used to solve the likelihood ratio equation for the values of interest.

Example: LR Bounds on Time

Lognormal Distribution Likelihood Ratio Bound Example (Time)

For the same data set given for the parameter bounds example, determine the two-sided 75% confidence bounds on the time estimate for a reliability of 80%. The ML estimate for the time at \( R(t) = 80\% \) is 55.718.

Solution

In this example, we are trying to determine the two-sided 75% confidence bounds on the time estimate of 55.718. This is accomplished by substituting \( R = 0.8 \) and \( \sigma = 0.73 \) into the likelihood function, and varying \( \sigma' \) until the maximum and minimum values of \( t \) are found. The following table gives the values of \( t \) based on given values of \( \sigma' \):

<table>
<thead>
<tr>
<th>( \sigma' )</th>
<th>( t_1 )</th>
<th>( t_2 )</th>
<th>( \sigma' )</th>
<th>( t_1 )</th>
<th>( t_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>56.832</td>
<td>62.879</td>
<td>0.37</td>
<td>44.841</td>
<td>64.031</td>
</tr>
<tr>
<td>0.25</td>
<td>54.660</td>
<td>64.287</td>
<td>0.38</td>
<td>44.494</td>
<td>63.454</td>
</tr>
<tr>
<td>0.26</td>
<td>53.093</td>
<td>65.079</td>
<td>0.39</td>
<td>44.200</td>
<td>62.809</td>
</tr>
<tr>
<td>0.27</td>
<td>51.811</td>
<td>65.576</td>
<td>0.40</td>
<td>43.963</td>
<td>62.093</td>
</tr>
<tr>
<td>0.28</td>
<td>50.711</td>
<td>65.881</td>
<td>0.41</td>
<td>43.786</td>
<td>61.304</td>
</tr>
<tr>
<td>0.29</td>
<td>49.743</td>
<td>66.041</td>
<td>0.42</td>
<td>43.674</td>
<td>60.436</td>
</tr>
<tr>
<td>0.30</td>
<td>48.881</td>
<td>66.085</td>
<td>0.43</td>
<td>43.634</td>
<td>59.481</td>
</tr>
<tr>
<td>0.31</td>
<td>48.106</td>
<td>66.028</td>
<td>0.44</td>
<td>43.681</td>
<td>58.426</td>
</tr>
<tr>
<td>0.32</td>
<td>47.408</td>
<td>65.883</td>
<td>0.45</td>
<td>43.832</td>
<td>57.252</td>
</tr>
<tr>
<td>0.33</td>
<td>46.777</td>
<td>65.657</td>
<td>0.46</td>
<td>44.124</td>
<td>55.924</td>
</tr>
<tr>
<td>0.34</td>
<td>46.208</td>
<td>65.355</td>
<td>0.47</td>
<td>44.625</td>
<td>54.373</td>
</tr>
<tr>
<td>0.35</td>
<td>45.697</td>
<td>64.983</td>
<td>0.48</td>
<td>45.517</td>
<td>52.418</td>
</tr>
<tr>
<td>0.36</td>
<td>45.242</td>
<td>64.541</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This data set is represented graphically in the following contour plot:
As can be determined from the table, the lowest calculated value for $t$ is 43.634, while the highest is 66.085. These represent the two-sided 75% confidence limits on the time at which reliability is equal to 80%.

Example: LR Bounds on Reliability

Lognormal Distribution Likelihood Ratio Bound Example (Reliability)

For the same data set given above for the parameter bounds example, determine the two-sided 75% confidence bounds on the reliability estimate for $t = 65$. The ML estimate for the reliability at $t = 65$ is 64.261%.

Solution

In this example, we are trying to determine the two-sided 75% confidence bounds on the reliability estimate of 64.261%. This is accomplished by substituting $t = 65$ and $\alpha = 0.75$ into the likelihood function, and varying $\sigma'$ until the maximum and minimum values of $R$ are found. The following table gives the values of $R$ based on given values of $\sigma'$.

<table>
<thead>
<tr>
<th>$\sigma'$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$\sigma'$</th>
<th>$R_1$</th>
<th>$R_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>61.107%</td>
<td>75.910%</td>
<td>0.37</td>
<td>43.573%</td>
<td>78.845%</td>
</tr>
<tr>
<td>0.25</td>
<td>55.906%</td>
<td>78.742%</td>
<td>0.38</td>
<td>43.807%</td>
<td>78.180%</td>
</tr>
<tr>
<td>0.26</td>
<td>55.528%</td>
<td>80.131%</td>
<td>0.39</td>
<td>44.147%</td>
<td>77.448%</td>
</tr>
<tr>
<td>0.27</td>
<td>50.067%</td>
<td>80.903%</td>
<td>0.40</td>
<td>44.593%</td>
<td>76.646%</td>
</tr>
<tr>
<td>0.28</td>
<td>48.206%</td>
<td>81.319%</td>
<td>0.41</td>
<td>45.146%</td>
<td>75.767%</td>
</tr>
<tr>
<td>0.29</td>
<td>46.779%</td>
<td>81.499%</td>
<td>0.42</td>
<td>45.813%</td>
<td>74.802%</td>
</tr>
<tr>
<td>0.30</td>
<td>45.685%</td>
<td>81.508%</td>
<td>0.43</td>
<td>46.604%</td>
<td>73.737%</td>
</tr>
<tr>
<td>0.31</td>
<td>44.857%</td>
<td>81.387%</td>
<td>0.44</td>
<td>47.538%</td>
<td>72.551%</td>
</tr>
<tr>
<td>0.32</td>
<td>44.250%</td>
<td>81.159%</td>
<td>0.45</td>
<td>48.645%</td>
<td>71.212%</td>
</tr>
<tr>
<td>0.33</td>
<td>43.827%</td>
<td>80.842%</td>
<td>0.46</td>
<td>49.980%</td>
<td>69.661%</td>
</tr>
<tr>
<td>0.34</td>
<td>43.565%</td>
<td>80.446%</td>
<td>0.47</td>
<td>51.652%</td>
<td>67.789%</td>
</tr>
<tr>
<td>0.35</td>
<td>43.444%</td>
<td>79.979%</td>
<td>0.48</td>
<td>53.956%</td>
<td>65.299%</td>
</tr>
<tr>
<td>0.36</td>
<td>43.450%</td>
<td>79.444%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This data set is represented graphically in the following contour plot:
As can be determined from the table, the lowest calculated value for $R_1$ is 43.444%, while the highest is 81.508%. These represent the two-sided 75% confidence limits on the reliability at $t = 65$.

**Bayesian Confidence Bounds**

**Bounds on Parameters**

From Parameter Estimation, we know that the marginal distribution of parameter $\mu'$ is:

$$f(\mu'|\text{Data}) = \int_0^{\infty} f(\mu', \sigma'|\text{Data})d\sigma'$$

$$= \frac{\int_0^{\infty} L(\text{Data}|\mu', \sigma') \varphi(\mu') \varphi(\sigma')d\sigma'}{\int_0^{\infty} \int_{-\infty}^{\infty} L(\text{Data}|\mu', \sigma') \varphi(\mu') \varphi(\sigma')d\mu'd\sigma'}$$

where:

- $\varphi(\sigma')$ is $\frac{1}{\sigma'}$, non-informative prior of $\sigma'$.
- $\varphi(\mu')$ is an uniform distribution from $-\infty$ to $+\infty$, non-informative prior of $\mu'$. With the above prior distributions, $f(\mu'|\text{Data})$ can be rewritten as:

$$f(\mu'|\text{Data}) = \frac{\int_0^{\infty} L(\text{Data}|\mu', \sigma') \frac{1}{\sigma'}d\sigma'}{\int_0^{\infty} \int_{-\infty}^{\infty} L(\text{Data}|\mu', \sigma') \frac{1}{\sigma'}d\mu'd\sigma'}$$

The one-sided upper bound of $\mu'$ is:

$$CL = P(\mu' \leq \mu'_{U}) = \int_{-\infty}^{\mu'_{U}} f(\mu'|\text{Data})d\mu'$$

The one-sided lower bound of $\mu'$ is:

$$1 - CL = P(\mu' \leq \mu'_{L}) = \int_{-\infty}^{\mu'_{L}} f(\mu'|\text{Data})d\mu'$$

The two-sided bounds of $\mu'$ is:

$$CL = P(\mu'_{L} \leq \mu' \leq \mu'_{U}) = \int_{\mu'_{L}}^{\mu'_{U}} f(\mu'|\text{Data})d\mu'$$

The same method can be used to obtained the bounds of $\sigma'$. 

---

**Lognormal Contour Plot for Reliability vs. Sigma**

![Contour Plot](image)
Bounds on Time (Type 1)

The reliable life of the lognormal distribution is:

\[ \ln T = \mu' + \sigma'\Phi^{-1}(1 - R) \]

The one-sided upper on time bound is given by:

\[ CL = \Pr(\ln t \leq \ln t_U) = \Pr(\mu' + \sigma'\Phi^{-1}(1 - R) \leq \ln t_U) \]

The above equation can be rewritten in terms of \( \mu' \) as:

\[ CL = \Pr(\mu' \leq \ln t_U - \sigma'\Phi^{-1}(1 - R)) \]

From the posterior distribution of \( \mu' \), get:

The above equation is solved w.r.t. \( t_U \). The same method can be applied for one-sided lower bounds and two-sided bounds on Time.

Bounds on Reliability (Type 2)

The one-sided upper bound on reliability is given by:

\[ CL = \Pr(R \leq R_U) = \Pr(\mu' \leq \ln t - \sigma'\Phi^{-1}(1 - R_U)) \]

From the posterior distribution of \( \mu' \) is:

\[ CL = \frac{\int_{-\infty}^{\ln t - \sigma'\Phi^{-1}(1 - R_U)} \int_{-\infty}^{\infty} L(\sigma', \mu') \frac{1}{\sigma'} d\sigma' d\mu'}{\int_{0}^{\infty} \int_{-\infty}^{\infty} L(\sigma', \mu') \frac{1}{\sigma'} d\sigma' d\mu'} \]

The above equation is solved w.r.t. \( R_U \). The same method is used to calculate the one-sided lower bounds and two-sided bounds on Reliability.

Example: Bayesian Bounds

Lognormal Distribution Bayesian Bound Example (Parameters)

Determine the two-sided 90% Bayesian confidence bounds on the lognormal parameter estimates for the data given next:

<table>
<thead>
<tr>
<th>Data Point Index</th>
<th>State</th>
<th>End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>59</td>
<td></td>
</tr>
</tbody>
</table>

Solution

The data points are entered into a times-to-failure data sheet. The lognormal distribution is selected under Distributions. The Bayesian confidence bounds method only applies for the MLE analysis method, therefore, Maximum Likelihood (MLE) is selected under Analysis Method and Use Bayesian is selected under the Confidence Bounds Method in the Analysis tab.

The two-sided 90% Bayesian confidence bounds on the lognormal parameter are obtained using the QCP and clicking on the Calculate Bounds button in the Parameter Bounds tab as follows:
Lognormal Distribution Examples

Complete Data Example

Determine the lognormal parameter estimates for the data given in the following table.

<table>
<thead>
<tr>
<th>Data point index</th>
<th>State F or S</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>43</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>59</td>
</tr>
</tbody>
</table>

Solution

Using Weibull++, the computed parameters for maximum likelihood are:

\[ \hat{\mu}' = 2.83 \]

\[ \hat{\sigma}' = 1.10 \]

For rank regression on \( X \):

\[ \hat{\mu}' = 2.83 \]

\[ \hat{\sigma}' = 1.24 \]

For rank regression on \( Y \):

\[ \hat{\mu}' = 2.83 \]

\[ \hat{\sigma}' = 1.36 \]
Complete Data RRX Example

From Kececioglu [20, p. 347]. 15 identical units were tested to failure and following is a table of their failure times:

<table>
<thead>
<tr>
<th>Data Point Index</th>
<th>Failure Times (Hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62.5</td>
</tr>
<tr>
<td>2</td>
<td>91.9</td>
</tr>
<tr>
<td>3</td>
<td>100.3</td>
</tr>
<tr>
<td>4</td>
<td>117.4</td>
</tr>
<tr>
<td>5</td>
<td>141.1</td>
</tr>
<tr>
<td>6</td>
<td>146.8</td>
</tr>
<tr>
<td>7</td>
<td>172.7</td>
</tr>
<tr>
<td>8</td>
<td>192.5</td>
</tr>
<tr>
<td>9</td>
<td>201.6</td>
</tr>
<tr>
<td>10</td>
<td>235.8</td>
</tr>
<tr>
<td>11</td>
<td>249.2</td>
</tr>
<tr>
<td>12</td>
<td>297.5</td>
</tr>
<tr>
<td>13</td>
<td>318.3</td>
</tr>
<tr>
<td>14</td>
<td>410.6</td>
</tr>
<tr>
<td>15</td>
<td>550.5</td>
</tr>
</tbody>
</table>

Solution

Published results (using probability plotting):
\[
\hat{\mu} = 5.22575 \\
\hat{\sigma}' = 0.62048.
\]

Weibull++ computed parameters for rank regression on X are:
\[
\hat{\mu} = 5.2303 \\
\hat{\sigma}' = 0.6283.
\]

The small differences are due to the precision errors when fitting a line manually, whereas in Weibull++ the line was fitted mathematically.

Complete Data Unbiased MLE Example

From Kececioglu [19, p. 406]. 9 identical units are tested continuously to failure and failure times were recorded at 30.4, 36.7, 53.3, 58.5, 74.0, 99.3, 114.3, 140.1 and 257.9 hours.

Solution

The results published were obtained by using the unbiased model. Published Results (using MLE):
\[
\hat{\mu} = 4.3553 \\
\hat{\sigma}' = 0.67677
\]

This same data set can be entered into Weibull++ by creating a data sheet capable of handling non-grouped time-to-failure data. Since the results shown above are unbiased, the Use Unbiased Std on Normal Data option in the User Setup must be selected in order to duplicate these results. Weibull++ computed parameters for maximum likelihood are:
\[
\hat{\mu} = 4.3553 \\
\hat{\sigma}' = 0.6768
\]
**Suspension Data Example**

From Nelson [30, p. 324]. 96 locomotive controls were tested, 37 failed and 59 were suspended after running for 135,000 miles. The table below shows the failure and suspension times.

<table>
<thead>
<tr>
<th>Number in State</th>
<th>F or S</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>22.5</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>37.5</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>48.5</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>51.5</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>53</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>54.5</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>57.5</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>66.5</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>68</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>69.5</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>76.5</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>77</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>78.5</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>80</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>81.5</td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>82</td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>83</td>
</tr>
<tr>
<td>19</td>
<td>F</td>
<td>84</td>
</tr>
<tr>
<td>20</td>
<td>F</td>
<td>91.5</td>
</tr>
<tr>
<td>21</td>
<td>F</td>
<td>93.5</td>
</tr>
<tr>
<td>22</td>
<td>F</td>
<td>102.5</td>
</tr>
<tr>
<td>23</td>
<td>F</td>
<td>107</td>
</tr>
<tr>
<td>24</td>
<td>F</td>
<td>108.5</td>
</tr>
<tr>
<td>25</td>
<td>F</td>
<td>112.5</td>
</tr>
<tr>
<td>26</td>
<td>F</td>
<td>113.5</td>
</tr>
<tr>
<td>27</td>
<td>F</td>
<td>116</td>
</tr>
<tr>
<td>28</td>
<td>F</td>
<td>117</td>
</tr>
<tr>
<td>29</td>
<td>F</td>
<td>118.5</td>
</tr>
<tr>
<td>30</td>
<td>F</td>
<td>119</td>
</tr>
<tr>
<td>31</td>
<td>F</td>
<td>120</td>
</tr>
<tr>
<td>32</td>
<td>F</td>
<td>122.5</td>
</tr>
<tr>
<td>33</td>
<td>F</td>
<td>123</td>
</tr>
<tr>
<td>34</td>
<td>F</td>
<td>127.5</td>
</tr>
</tbody>
</table>
Solution

The distribution used in the publication was the base-10 lognormal. Published results (using MLE):

\[ \hat{\mu}' = 2.2223 \]
\[ \hat{\sigma}' = 0.3064 \]

Published 95% confidence limits on the parameters:

\[ \hat{\mu}' = \{2.1336, 2.3109\} \]
\[ \hat{\sigma}' = \{0.2365, 0.3970\} \]

Published variance/covariance matrix:

\[
\begin{bmatrix}
\widehat{\text{Var}}(\hat{\mu}') &= 0.0020 & \widehat{\text{Cov}}(\hat{\mu}', \hat{\sigma}') &= 0.001 \\
\widehat{\text{Cov}}(\hat{\mu}', \hat{\sigma}') &= 0.001 & \widehat{\text{Var}}(\hat{\sigma}') &= 0.0016
\end{bmatrix}
\]

To replicate the published results (since Weibull++ uses a lognormal to the base \( e \)), take the base-10 logarithm of the data and estimate the parameters using the normal distribution and MLE.

- Weibull++ computed parameters for maximum likelihood are:

\[ \hat{\mu}' = 2.2223 \]
\[ \hat{\sigma}' = 0.3064 \]

- Weibull++ computed 95% confidence limits on the parameters:

\[ \hat{\mu}' = \{2.1364, 2.3081\} \]
\[ \hat{\sigma}' = \{0.2395, 0.3920\} \]

- Weibull++ computed variance/covariance matrix:

\[
\begin{bmatrix}
\widehat{\text{Var}}(\hat{\mu}') &= 0.0019 & \widehat{\text{Cov}}(\hat{\mu}', \hat{\sigma}') &= 0.0009 \\
\widehat{\text{Cov}}(\hat{\mu}', \hat{\sigma}') &= 0.0009 & \widehat{\text{Var}}(\hat{\sigma}') &= 0.0015
\end{bmatrix}
\]

Interval Data Example

Determine the lognormal parameter estimates for the data given in the table below.

<table>
<thead>
<tr>
<th>Data point index</th>
<th>Last Inspected</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>55</td>
<td>55</td>
</tr>
</tbody>
</table>

The Lognormal Distribution
Solution

This is a sequence of interval times-to-failure where the intervals vary substantially in length. Using Weibull++, the computed parameters for maximum likelihood are calculated to be:

$$\hat{\mu}' = 3.64$$
$$\hat{\sigma}' = 0.18$$

For rank regression on $X$:

$$\hat{\mu}' = 3.64$$
$$\hat{\sigma}' = 0.17$$

For rank regression on $Y$:

$$\hat{\mu}' = 3.64$$
$$\hat{\sigma}' = 0.21$$
Chapter 11

The Mixed Weibull Distribution

The mixed Weibull distribution (also known as a multimodal Weibull) is used to model data that do not fall on a straight line on a Weibull probability plot. Data of this type, particularly if the data points follow an S-shape on the probability plot, may be indicative of more than one failure mode at work in the population of failure times. Field data from a given mixed population may frequently represent multiple failure modes. The necessity of determining the life regions where these failure modes occur is apparent when it is realized that the times-to-failure for each mode may follow a distinct Weibull distribution, thus requiring individual mathematical treatment. Another reason is that each failure mode may require a different design change to improve the component's reliability, as discussed in Kececioglu [19].

A decreasing failure rate is usually encountered during the early life period of components when the substandard components fail and are removed from the population. The failure rate continues to decrease until all such substandard components fail and are removed. This corresponds to a decreasing failure rate. The Weibull distribution having $\beta < 1$ is often used to depict this life characteristic.

A second type of failure prevails when the components fail by chance alone and their failure rate is nearly constant. This can be caused by sudden, unpredictable stress applications that have a stress level above those to which the product is designed. Such failures tend to occur throughout the life of a component. The distributions most often used to describe this failure rate characteristic are the exponential distribution and the Weibull distribution with $\beta \approx 1$.

A third type of failure is characterized by a failure rate that increases as operating hours are accumulated. Usually, wear has started to set in and this brings the component's performance out of specification. As age increases further, this wear-out process removes more and more components until all components fail. The normal distribution and the Weibull distribution with $\beta > 1$ have been successfully used to model the times-to-failure distribution during the wear-out period.

Several different failure modes may occur during the various life periods. A methodology is needed to identify these failure modes and determine their failure distributions and reliabilities. This section presents a procedure whereby the proportion of units failing in each mode is determined and their contribution to the reliability of the component is quantified. From this reliability expression, the remaining major reliability functions, the probability density, the failure rate and the conditional-reliability functions are calculated to complete the reliability analysis of such mixed populations.

Statistical Background

Consider a life test of identical components. The components were placed in a test at age $t = 0$ and were tested to failure, with their times-to-failure recorded. Further assume that the test covered the entire lifespan of the units, and different failure modes were observed over each region of life, namely early life (early failure mode), chance life (chance failure mode), and wear-out life (wear-out failure mode). Also, as items failed during the test, they were removed from the test, inspected and segregated into lots according to their failure mode. At the conclusion of the test, there will be $n$ subpopulations of $N_1, N_2, N_3, \ldots, N_n$ failed components. If the events of the test are now reconstructed, it may be theorized that at age $t = 0$ there were actually $n$ separate subpopulations in the test, each with a different times-to-failure distribution and failure mode, even though at $t = 0$ the subpopulations were not physically distinguishable. The mixed Weibull methodology accomplishes this segregation based on the results of
the life test.

If $N$ identical components from a mixed population undertake a mission of duration, starting the mission at age zero, then the number of components surviving this mission can be found from the following definition of reliability:

$$R_{1,2,...,n}(t) = \frac{N_{1,2,...,n}(t)}{N}$$

Then:

$$N_{1,2,...,n}(t) = N[R_{1,2,...,n}(t)]$$

$$N_{1}(t) = N_{1}R_{1}(t); N_{2}(t) = N_{2}R_{2}(t); \ldots; N_{n}(t) = N_{n}R_{n}(t)$$

The total number surviving by age $t$ in the mixed population is the sum of the number surviving in all subpopulations or:

$$N_{1,2,...,n}(t) = N_{1}(t) + N_{2}(t) + N_{3}(t) + \cdots + N_{n}(t)$$

Substituting into the reliability equation yields:

$$R_{1,2,...,n}(t) = \frac{1}{N}[N_{1}R_{1}(t) + N_{2}R_{2}(t) + N_{3}R_{3}(t) + \cdots + N_{n}R_{n}(t)]$$

or:

$$R_{1,2,...,n}(t) = \frac{N_{1}}{N}R_{1}(t) + \frac{N_{2}}{N}R_{2}(t) + \frac{N_{3}}{N}R_{3}(t) + \cdots + \frac{N_{n}}{N}R_{n}(t)$$

This expression can also be derived by applying Bayes's theorem, as discussed in Kececioglu [20], which says that the reliability of a component drawn at random from a mixed population composed of $n$ types of failure subpopulations is its reliability, $R_{1}(t)$, given that the component is from subpopulation 1, or $N_{1}/N$ plus its reliability, $R_{2}(t)$, given that the component is from subpopulation 2, or $N_{2}/N$ plus its reliability, $R_{3}(t)$, given that the component is from subpopulation 3, or $N_{3}/N$, and so on, plus its reliability, $R_{n}(t)$, given that the component is from subpopulation $n$, or $N_{n}/N$, and:

$$\sum_{i=1}^{n} \frac{N_{i}}{N} = 1$$

This may be written mathematically as:

$$R_{1,2,...,n}(t) = \frac{N_{1}}{N}R_{1}(t) + \frac{N_{2}}{N}R_{2}(t) + \frac{N_{3}}{N}R_{3}(t) + \cdots + \frac{N_{n}}{N}R_{n}(t)$$

Other functions of reliability engineering interest are found by applying the fundamentals to the above reliability equation. For example, the probability density function can be found from:

$$f_{1,2,...,n}(t) = -\frac{d}{dT}[R_{1,2,...,n}(t)]$$

$$f_{1,2,...,n}(t) = \frac{N_{1}}{N} \left( -\frac{d}{dT}[R_{1}(t)] \right) + \frac{N_{2}}{N} \left( -\frac{d}{dT}[R_{2}(t)] \right)$$

$$+ \frac{N_{3}}{N} \left( -\frac{d}{dT}[R_{3}(t)] \right) + \cdots + \frac{N_{n}}{N} \left( -\frac{d}{dT}[R_{n}(t)] \right)$$

$$f_{1,2,...,n}(t) = \frac{N_{1}}{N}f_{1}(t) + \frac{N_{2}}{N}f_{2}(t)$$

$$+ \frac{N_{3}}{N}f_{3}(t) + \cdots + \frac{N_{n}}{N}f_{n}(t)$$

Also, the failure rate function of a population is given by:
The conditional reliability for a new mission of duration $\breve{t}$, starting this mission at age $T$, or after having already operated a total of $T$ hours, is given by:

$$R_{1,2,\ldots,n}(T, t) = \frac{R_{1,2,\ldots,n}(T + t)}{R_{1,2,\ldots,n}(T)}$$

$$R_{1,2,\ldots,n}(T, t) = \frac{\frac{N_1}{N} R_1(T + t) + \frac{N_2}{N} R_2(T + t) + \cdots + \frac{N_n}{N} R_n(T + t)}{\frac{N_1}{N} R_1(T) + \frac{N_2}{N} R_2(T) + \cdots + \frac{N_n}{N} R_n(T)}$$

### The Mixed Weibull Equations

Depending on the number of subpopulations chosen, Weibull++ uses the following equations for the reliability and probability density functions:

$$R_{1,\ldots,S}(t) = \sum_{i=1}^{S} \frac{N_i}{N} e^{-\left(\frac{t}{\tilde{\eta}_i}\right)^{\tilde{\beta}_i}}$$

and:

$$f_{1,\ldots,S}(t) = \sum_{i=1}^{S} \frac{N_i}{N\tilde{\eta}_i^{\tilde{\beta}_i}} \left(\frac{t}{\tilde{\eta}_i}\right)^{\tilde{\beta}_i-1} e^{-\left(\frac{t}{\tilde{\eta}_i}\right)^{\tilde{\beta}_i}}$$

where $S = 2$, $S = 3$, and $S = 4$ for 2, 3 and 4 subpopulations respectively. Weibull++ uses a non-linear regression method or direct maximum likelihood methods to estimate the parameters.

### Mixed Weibull Parameter Estimation

#### Regression Solution

Weibull++ utilizes a modified Levenberg-Marquardt algorithm (non-linear regression) when performing regression analysis on a mixed Weibull distribution. The procedure is rather involved and is beyond the scope of this reference. It is sufficient to say that the algorithm fits a curved line of the form:

$$R_{1,\ldots,S}(t) = \sum_{i=1}^{S} \rho_i \cdot e^{-\left(\frac{t}{\tilde{\eta}_i}\right)^{\tilde{\beta}_i}}$$

where:

$$\sum_{i=1}^{S} \rho_i = 1$$

to the parameters $\tilde{\rho}_1$, $\tilde{\beta}_1$, $\tilde{\eta}_1$, $\tilde{\rho}_2$, $\tilde{\beta}_2$, $\tilde{\eta}_2$, ..., $\tilde{\rho}_S$, $\tilde{\beta}_S$, $\tilde{\eta}_S$ utilizing the times-to-failure and their respective plotting positions. It is important to note that in the case of regression analysis, using a mixed Weibull model, the choice of regression axis, (i.e., $RRX$ or $RRY$) is of no consequence since non-linear regression is utilized.
MLE

The same space of parameters, namely $\hat{\rho}, \hat{\beta}_1, \hat{\eta}_1, \hat{\rho}_2, \hat{\beta}_2, \hat{\eta}_2, \ldots, \hat{\rho}_S, \hat{\beta}_S, \hat{\eta}_S$, is also used under the MLE method, using the likelihood function as given in Appendix C of this reference. Weibull++ uses the EM algorithm, short for Expectation-Maximization algorithm, for the MLE analysis. Details on the numerical procedure are beyond the scope of this reference.

About the Calculated Parameters

Weibull++ uses the numbers 1, 2, 3 and 4 (or first, second, third and fourth subpopulation) to identify each subpopulation. These are just designations for each subpopulation, and they are ordered based on the value of the scale parameter, $\hat{\eta}$. Since the equation used is additive or:

$$R_{1,\ldots,S}(T) = \sum_{i=1}^{S} \frac{N_i}{N} e^{-\left(\frac{T}{\hat{\eta}_i}\right)^{\hat{\beta}_i}}$$

the order of the subpopulations which are given the designation 1, 2, 3, or 4 is of no consequence. For consistency, the application will always return the order of the results based on the magnitude of the scale parameter.

Mixed Weibull Confidence Bounds

In Weibull++, two methods are available for estimating the confidence bounds for the mixed Weibull distribution. The first method is the beta binomial, described in Confidence Bounds. The second method is the Fisher matrix confidence bounds. For the Fisher matrix bounds, the methodology is the same as described in Confidence Bounds. The variance/covariance matrix for the mixed Weibull is a $(3 \cdot S - 1) \times (3 \cdot S - 1)$ matrix, where $S$ is the number of subpopulations. Bounds on the parameters, reliability and time are estimated using the same transformations and methods that were used for the The Weibull Distribution. Note, however, that in addition to the Weibull parameters, the bounds on the subpopulation portions are obtained as well. The bounds on the portions are estimated by:

$$\rho_U = \frac{\hat{\rho}}{\hat{\rho} + (1 - \hat{\rho})e^{\frac{K_{\alpha} \sqrt{Var(\hat{\rho})}}{\hat{\beta}(1-\hat{\beta})}}}$$

$$\rho_L = \frac{\hat{\rho}}{\hat{\rho} + (1 - \hat{\rho})e^{\frac{K_{\alpha} \sqrt{Var(\hat{\rho})}}{\hat{\beta}(1-\hat{\beta})}}}$$

where $Var(\hat{\rho})$ is obtained from the variance/covariance matrix. When using the Fisher matrix bounds method, problems can occur on the transition points of the distribution, and in particular on the Type 1 confidence bounds (bounds on time). The problems (i.e., the departure from the expected monotonic behavior) occur when the transition region between two subpopulations becomes a “saddle” (i.e., the probability line is almost parallel to the time axis on a probability plot). In this case, the bounds on time approach infinity. This behavior is more frequently encountered with smaller sample sizes. The physical interpretation is that there is insufficient data to support any inferences when in this region.

This is graphically illustrated in the following figure. In this plot it can be seen that there are no data points between the last point of the first subpopulation and the first point of the second subpopulation, thus the uncertainty is high, as described by the mathematical model.
Beta binomial bounds can be used instead in these cases, especially when estimations are to be obtained close to these regions.
Other Uses for Mixed Weibull

Reliability Bathtub Curves

A reliability bathtub curve is nothing more than the graph of the failure rate versus time, over the life of the product. In general, the life stages of the product consist of early, chance and wear-out. Weibull++ allows you to plot this by simply selecting the failure rate plot, as shown next.

Determination of the Burn-in Period

If the failure rate goal is known, then the burn-in period can be found from the failure rate plot by drawing a horizontal line at the failure rate goal level and then finding the intersection with the failure rate curve. Next, drop vertically at the intersection, and read off the burn-in time from the time axis. This burn-in time helps insure that the population will have a failure rate that is at least equal to or lower than the goal after the burn-in period. The same could also be obtained using the Function Wizard and generating different failure rates based on time increments. Using these generated times and the corresponding failure rates, one can decide on the optimum burn-in time versus the corresponding desired failure rate.
Example

We will illustrate the mixed Weibull analysis using a Monte Carlo generated set of data. To repeat this example, generate data from a 2-parameter Weibull distribution using the Weibull++ Monte Carlo utility. The following figures illustrate the required steps, inputs and results.

In the Monte Carlo window, enter the values and select the options shown below for subpopulation 1.

Switch to subpopulation 2 and make the selection shown below. Click Generate.
The simulation settings are:

- **Distribution**: 2 Subpop-Mixed Weibull
- **Units**: Hour (Hr)
- **Parameters**:
  - Beta[2]: 2
  - Eta[2]: 1000
  - Portion[2]: 0.75
  - Subpop: 2
- **Seed**: 10
- **Precision**: 4
- **Number of points**: 100
- **Generate Data in Specified Folio and Data Sheet**
  - Folio: <New Folio>
  - Sheet: <New Sheet>
After the data set has been generated, choose the 2 Subpop-Mixed Weibull distribution. Click **Calculate**.
The results for subpopulation 1 are shown next. (Note that your results could be different due to the randomness of the simulation.)

![Image of data set and subpopulation analysis](image)

The results for subpopulation 2 are shown next. (Note that your results could be different due to the randomness of the simulation.)
The Weibull probability plot for this data is shown next. (Note that your results could be different due to the randomness of the simulation.)
Chapter 12

The Generalized Gamma Distribution

While not as frequently used for modeling life data as the previous distributions, the generalized gamma distribution does have the ability to mimic the attributes of other distributions such as the Weibull or lognormal, based on the values of the distribution's parameters. While the generalized gamma distribution is not often used to model life data by itself (mostly due to its mathematical complexity and its requirement of large sample sizes (>30) for convergence), its ability to behave like other more commonly-used life distributions is sometimes used to determine which of those life distributions should be used to model a particular set of data.

Generalized Gamma Probability Density Function

The generalized gamma function is a 3-parameter distribution. One version of the generalized gamma distribution uses the parameters $k$, $\beta$, and $\theta$. The pdf for this form of the generalized gamma distribution is given by:

$$ f(t) = \frac{\beta}{\Gamma(k) \cdot \theta} \left( \frac{t}{\theta} \right)^{k-1} e^{-\left( \frac{t}{\theta} \right)^\beta} $$

where $\theta$ is a scale parameter, $\beta > 0$ and $k > 0$ are shape parameters and $\Gamma(x)$ is the gamma function of $x$, which is defined by:

$$ \Gamma(x) = \int_0^\infty s^{x-1} \cdot e^{-s} \, ds $$

With this version of the distribution, however, convergence problems arise that severely limit its usefulness. Even with data sets containing 200 or more data points, the MLE methods may fail to converge. Further adding to the confusion is the fact that distributions with widely different values of $k$, $\beta$, and $\theta$ may appear almost identical, as discussed in Lawless [21]. In order to overcome these difficulties, Weibull++ uses a reparameterization with parameters $\tilde{k}$, $\tilde{\beta}$, and $\tilde{\theta}$, as shown in [21], where:

$$ \mu = \ln(\tilde{\theta}) + \frac{1}{\tilde{\beta}} \cdot \ln \left( \frac{1}{\tilde{k}} \right) $$

$$ \sigma = \frac{1}{\tilde{\beta} \sqrt{\tilde{k}}} $$

$$ \lambda = \frac{1}{\sqrt{\tilde{k}}} $$

where $-\infty < \mu < \infty$, $\sigma > 0$, $0 < \lambda$.

While this makes the distribution converge much more easily in computations, it does not facilitate manual manipulation of the equation. By allowing $\lambda$ to become negative, the pdf of the reparameterized distribution is given by:
The Generalized Gamma Distribution

\[ f(t) = \begin{cases} \frac{|\lambda|}{\sigma^2} \cdot \frac{1}{\Gamma\left(\frac{1}{\lambda^2}\right)} \cdot e^{\lambda \cdot \frac{\ln(t)-\mu}{\sigma}} \left(\frac{\ln(t)^2}{\lambda^2} - \frac{\ln(t)-\mu}{\lambda^2} \right) & \text{if } \lambda \neq 0 \\ \frac{1}{\lambda \cdot \sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\ln(t)-\mu}{\sigma}\right)^2} & \text{if } \lambda = 0 \end{cases} \]

Generalized Gamma Reliability Function

The reliability function for the generalized gamma distribution is given by:

\[ R(t) = \begin{cases} 1 - \Gamma_t\left(\frac{\ln(t)-\mu}{\sigma \cdot \lambda^2}; \frac{1}{\lambda^2}\right) & \text{if } \lambda > 0 \\ 1 - \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right) & \text{if } \lambda = 0 \\ \Gamma^t\left(\frac{\ln(t)-\mu}{\sigma \cdot \lambda^2}; \frac{1}{\lambda^2}\right) & \text{if } \lambda < 0 \end{cases} \]

where:

\[ \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx \]

and \( \Gamma_t(k; x) \) is the incomplete gamma function of \( k \) and \( x \), which is given by:

\[ \Gamma_t(k; x) = \frac{1}{\Gamma(k)} \int_{0}^{x} s^{k-1} e^{-s} ds \]

where \( \Gamma(x) \) is the gamma function of \( x \). Note that in Weibull++ the probability plot of the generalized gamma is created on lognormal probability paper. This means that the fitted line will not be straight unless \( \lambda = 0 \).

Generalized Gamma Failure Rate Function

As defined in Basic Statistical Background, the failure rate function is given by:

\[ \lambda(t) = \frac{f(t)}{R(t)} \]

Owing to the complexity of the equations involved, the function will not be displayed here, but the failure rate function for the generalized gamma distribution can be obtained merely by dividing the pdf function by the reliability function.
The Generalized Gamma Reliable Life

The reliable life, \( T_R \), of a unit for a specified reliability, starting the mission at age zero, is given by:

\[
T_R = \begin{cases} 
  e^{\mu + \frac{\sigma}{\lambda} \ln \left[ \frac{\lambda^3}{\Gamma(\frac{1}{\lambda})} \right]} & \text{if } \lambda > 0 \\
  \Phi^{-1}(1 - R) & \text{if } \lambda = 0 \\
  e^{\mu + \frac{\sigma}{\lambda} \ln \left[ \frac{\lambda^3}{\Gamma(\frac{1}{\lambda})} \right]} & \text{if } \lambda < 0 
\end{cases}
\]

Characteristics of the Generalized Gamma Distribution

As mentioned previously, the generalized gamma distribution includes other distributions as special cases based on the values of the parameters.

- The Weibull distribution is a special case when \( \lambda = 1 \) and:
  \[
  \beta = \frac{1}{\sigma}, \quad \eta = \exp(\mu)
  \]
- In this case, the generalized distribution has the same behavior as the Weibull for \( \sigma > 1, \sigma = 1, \sigma < 1 \) (\( \beta < 1, \beta = 1, \beta > 1 \) respectively).
- The exponential distribution is a special case when \( \lambda = 1 \) and \( \sigma = 1 \).
- The lognormal distribution is a special case when \( \lambda = 1 \).
- The gamma distribution is a special case when \( \lambda = 1 \).

By allowing \( \lambda \) to take negative values, the generalized gamma distribution can be further extended to include additional distributions as special cases. For example, the Fréchet distribution of maxima (also known as a reciprocal Weibull) is a special case when \( \lambda = -1 \).
Confidence Bounds

The only method available in Weibull++ for confidence bounds for the generalized gamma distribution is the Fisher matrix, which is described next.

Bounds on the Parameters

The lower and upper bounds on the parameter $\mu$ are estimated from:

$$
\mu_U = \hat{\mu} + K_\alpha \sqrt{\text{Var}(\hat{\mu})} \quad \text{(upper bound)}
$$

$$
\mu_L = \hat{\mu} - K_\alpha \sqrt{\text{Var}(\hat{\mu})} \quad \text{(lower bound)}
$$

For the parameter $\sigma$, $\ln(\hat{\sigma})$ is treated as normally distributed, and the bounds are estimated from:

$$
\sigma_U = \hat{\sigma} \cdot e^{K_\alpha \sqrt{\text{Var}(\hat{\sigma})}} \quad \text{(upper bound)}
$$

$$
\sigma_L = \frac{\hat{\sigma}}{e^{K_\alpha \sqrt{\text{Var}(\hat{\sigma})}}} \quad \text{(lower bound)}
$$

For the parameter $\lambda$, the bounds are estimated from:

$$
\lambda_U = \hat{\lambda} + K_\alpha \sqrt{\text{Var}(\hat{\lambda})} \quad \text{(upper bound)}
$$

$$
\lambda_L = \hat{\lambda} - K_\alpha \sqrt{\text{Var}(\hat{\lambda})} \quad \text{(lower bound)}
$$

where $K_\alpha$ is defined by:

$$
\alpha = \frac{1}{\sqrt{2\pi}} \int_{K_\alpha}^{\infty} e^{-\frac{t^2}{2}} dt = 1 - \Phi(K_\alpha)
$$

If $\delta$ is the confidence level, then $\alpha = \frac{1-\delta}{2}$ for the two-sided bounds, and $\alpha = 1 - \frac{\delta}{2}$ for the one-sided bounds.

The variances and covariances of $\hat{\mu}$ and $\hat{\sigma}$ are estimated as follows:

$$
\begin{pmatrix}
\text{Var}(\hat{\mu}) & \text{Cov}(\hat{\mu}, \hat{\sigma}) & \text{Cov}(\hat{\mu}, \hat{\lambda}) \\
\text{Cov}(\hat{\mu}, \hat{\sigma}) & \text{Var}(\hat{\sigma}) & \text{Cov}(\hat{\sigma}, \hat{\lambda}) \\
\text{Cov}(\hat{\mu}, \hat{\lambda}) & \text{Cov}(\hat{\sigma}, \hat{\lambda}) & \text{Var}(\hat{\lambda})
\end{pmatrix}
= \begin{pmatrix}
-\frac{\partial^2 L}{\partial \mu^2} & -\frac{\partial^2 L}{\partial \mu \partial \sigma} & -\frac{\partial^2 L}{\partial \mu \partial \lambda} \\
-\frac{\partial^2 L}{\partial \mu \partial \sigma} & -\frac{\partial^2 L}{\partial \sigma^2} & -\frac{\partial^2 L}{\partial \sigma \partial \lambda} \\
-\frac{\partial^2 L}{\partial \mu \partial \lambda} & -\frac{\partial^2 L}{\partial \sigma \partial \lambda} & -\frac{\partial^2 L}{\partial \lambda^2}
\end{pmatrix}_{\mu=\hat{\mu}, \sigma=\hat{\sigma}, \lambda=\hat{\lambda}}^{-1}
$$

Where $L$ is the log-likelihood function of the generalized gamma distribution.

Bounds on Reliability

The upper and lower bounds on reliability are given by:

$$
R_U = \frac{\hat{R}}{\hat{R} + (1 - \hat{R}) e^{-K_\alpha \sqrt{\text{Var}(\hat{R})} / R(1-R)}}
$$

$$
R_L = \frac{\hat{R}}{\hat{R} + (1 - \hat{R}) e^{-K_\alpha \sqrt{\text{Var}(\hat{R})} / R(1-R)}}
$$

where:
The Generalized Gamma Distribution

Bounds on Time

The bounds around time for a given percentile, or unreliability, are estimated by first solving the reliability equation with respect to time. Since $T$ is a positive variable, the transformed variable $	ilde{u} = \ln(T)$ is treated as normally distributed and the bounds are estimated from:

$$ u_u = \ln T_U = \tilde{u} + K_\alpha \sqrt{\text{Var}(\tilde{u})} $$

$$ u_L = \ln T_L = \tilde{u} - K_\alpha \sqrt{\text{Var}(\tilde{u})} $$

Solving for $T_U$ and $T_L$, we get:

$$ T_U = e^{\tilde{u}} \text{ (upper bound)} $$

$$ T_L = e^{\tilde{u}} \text{ (lower bound)} $$

The variance of $\tilde{u}$ is estimated from:

$$ \text{Var}(\tilde{u}) = \left( \frac{\partial u}{\partial \mu} \right)^2 \text{Var}(\tilde{\mu}) + \left( \frac{\partial u}{\partial \sigma} \right)^2 \text{Var}(\tilde{\sigma}) + \left( \frac{\partial u}{\partial \lambda} \right)^2 \text{Var}(\tilde{\lambda}) $$

$$ + 2 \left( \frac{\partial u}{\partial \mu} \right) \left( \frac{\partial u}{\partial \sigma} \right) \text{Cov}(\tilde{\mu}, \tilde{\sigma}) + 2 \left( \frac{\partial u}{\partial \mu} \right) \left( \frac{\partial u}{\partial \lambda} \right) \text{Cov}(\tilde{\mu}, \tilde{\lambda}) $$

$$ + 2 \left( \frac{\partial u}{\partial \sigma} \right) \left( \frac{\partial u}{\partial \lambda} \right) \text{Cov}(\tilde{\lambda}, \tilde{\sigma}) $$

Example

The following data set represents revolutions-to-failure (in millions) for 23 ball bearings in a fatigue test, as discussed in Lawless [21].

<table>
<thead>
<tr>
<th>Revolution</th>
<th>17.88</th>
<th>28.92</th>
<th>33</th>
<th>41.52</th>
<th>42.12</th>
<th>45.6</th>
<th>48.4</th>
<th>51.84</th>
<th>51.96</th>
<th>54.12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>55.56</td>
<td>67.8</td>
<td>68.64</td>
<td>68.64</td>
<td>68.88</td>
<td>84.12</td>
<td>93.12</td>
<td>98.64</td>
<td>105.12</td>
<td>105.84</td>
</tr>
<tr>
<td></td>
<td>127.92</td>
<td>128.04</td>
<td>173.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When the generalized gamma distribution is fitted to this data using MLE, the following values for parameters are obtained:

$$ \tilde{\mu} = 4.23064 $$

$$ \tilde{\sigma} = 0.509982 $$

$$ \tilde{\lambda} = 0.307639 $$

Note that for this data, the generalized gamma offers a compromise between the Weibull ($\lambda = 1$) and the lognormal ($\lambda = 0$) distributions. The value of $\lambda$ indicates that the lognormal distribution is better supported by the data. A better assessment, however, can be made by looking at the confidence bounds on $\lambda$. For example, the 90% two-sided confidence bounds are:

$$ \lambda_u = -0.592087 $$

$$ \lambda_l = 1.20736 $$

We can then conclude that both distributions (i.e., Weibull and lognormal) are well supported by the data, with the lognormal being the better supported of the two. In Weibull++ the generalized gamma probability is plotted on a gamma probability paper, as shown next.
It is also important to note that, as in the case of the mixed Weibull distribution, the choice of regression analysis (i.e., RRX or RRY) is of no consequence in the generalized gamma model because it uses non-linear regression.
Chapter 13

The Gamma Distribution

The Gamma distribution is a flexible life distribution model that may offer a good fit to some sets of failure data. It is not, however, widely used as a life distribution model for common failure mechanisms. The Gamma distribution does arise naturally as the time-to-first-fail distribution for a system with standby exponentially distributed backups, and is also a good fit for the sum of independent exponential random variables. The Gamma distribution is sometimes called the Erlang distribution, which is used frequently in queuing theory applications, as discussed in [32].

The Gamma Probability Density Function

The pdf of the gamma distribution is given by:

\[ f(t) = \frac{e^{k\ln(t) - \mu}}{t^{k-1} \Gamma(k)} \]

where:

\[ z = \ln(t) - \mu \]

and:

\[ e^\mu = \text{scale parameter} \]

\[ k = \text{shape parameter} \]

where \( 0 < t < \infty, -\infty < \mu < \infty \) and \( k > 0 \)

The Gamma Reliability Function

The reliability for a mission of time for the Gamma distribution is:

\[ R = 1 - \Gamma_1(k; e^z) \]

The Gamma Mean, Median and Mode

The gamma mean or MTTF is:

\[ \overline{T} = ke^\mu \]

The mode exists if \( k > 1 \) and is given by:

\[ \hat{T} = (k - 1)e^\mu \]

The median is:

\[ \hat{T} = e^{\mu + \ln(\Gamma^{-1}(0.5; k))} \]
The Gamma Standard Deviation

The standard deviation for the gamma distribution is:

$$\sigma_T = \sqrt{k}\mu$$

The Gamma Reliable Life

The gamma reliable life is:

$$T_R = e^{\mu + \ln(\Gamma^{-1}_1(1-R;k))}$$

The Gamma Failure Rate Function

The instantaneous gamma failure rate is given by:

$$\lambda = \frac{e^{kz-e^z}}{t\Gamma(k)(1-\Gamma_t(k;e^z))}$$

Characteristics of the Gamma Distribution

Some of the specific characteristics of the gamma distribution are the following:

For $k > 1$:

- As $t \to 0$, $\infty$, $f(t) \to 0$.
- $f(t)$ increases from 0 to the mode value and decreases thereafter.
- If $k \leq 2$ then pdf has one inflection point at $t = e^{\mu \sqrt{k-1}(\sqrt{k-1} + 1)}$.
- If $k > 2$ then pdf has two inflection points for $t = e^{\mu \sqrt{k-1}(\sqrt{k-1} \pm 1)}$.
- For a fixed $k$, as $\mu$ increases, the pdf starts to look more like a straight angle.
- As $t \to \infty$, $\lambda(t) \to \frac{1}{e^\mu}$.

For $k = 1$:

- Gamma becomes the exponential distribution.
- As $t \to 0$, $f(T) \to \frac{1}{e^\mu}$. 

![Probability Density Function](image-url)
• As $t \to \infty$, $f(t) \to 0$.
• The pdf decreases monotonically and is convex.
• $\lambda(t) \equiv \frac{1}{\theta \mu}$, $\lambda(t)$ is constant.
• The mode does not exist.

For $0 < k < 1$:
• As $t \to 0$, $f(t) \to \infty$.
• As $t \to \infty$, $f(t) \to 0$.
• As $t \to \infty$, $\lambda(t) \to \frac{1}{\theta \mu}$.
• The pdf decreases monotonically and is convex.
• As $\mu$ increases, the pdf gets stretched out to the right and its height decreases, while maintaining its shape.
• As $\mu$ decreases, the pdf shifts towards the left and its height increases.
• The mode does not exist.
Confidence Bounds

The only method available in Weibull++ for confidence bounds for the gamma distribution is the Fisher matrix, which is described next. The complete derivations were presented in detail (for a general function) in the Confidence Bounds chapter.

Bounds on the Parameters

The lower and upper bounds on the mean, \( \hat{\mu} \), are estimated from:

\[
\mu_U = \hat{\mu} + K_\alpha \sqrt{Var(\hat{\mu})} \quad \text{(upper bound)}
\]

\[
\mu_L = \hat{\mu} - K_\alpha \sqrt{Var(\hat{\mu})} \quad \text{(lower bound)}
\]

Since the standard deviation, \( \hat{\sigma} \), must be positive, \( \ln(\hat{\sigma}) \) is treated as normally distributed and the bounds are estimated from:

\[
k_{U} = \frac{\hat{k} \cdot e^{\frac{K_\alpha \sqrt{Var(k)}}{\hat{k}}}}{e^{\frac{K_\alpha \sqrt{Var(k)}}{\hat{k}}}} \quad \text{(upper bound)}
\]

\[
k_{L} = \frac{\hat{k}}{e^{\frac{K_\alpha \sqrt{Var(k)}}{\hat{k}}}} \quad \text{(lower bound)}
\]

where \( K_\alpha \) is defined by:

\[
\alpha = \frac{1}{\sqrt{2\pi}} \int_{K_\alpha}^{\infty} e^{-\frac{t^2}{2}} dt = 1 - \Phi(K_\alpha)
\]

If \( \hat{\alpha} \) is the confidence level, then \( \alpha = \frac{1 - \hat{\alpha}}{2} \) for the two-sided bounds and \( \alpha = 1 - \hat{\alpha} \) for the one-sided bounds.

The variances and covariances of \( \hat{\mu} \) and \( \hat{k} \) are estimated from the Fisher matrix, as follows:

\[
\begin{pmatrix}
\hat{Var}(\hat{\mu}) & \hat{Cov}(\hat{\mu}, \hat{k}) \\
\hat{Cov}(\hat{\mu}, \hat{k}) & \hat{Var}(\hat{k})
\end{pmatrix} = \begin{pmatrix}
\frac{\partial^2 \Lambda}{\partial \mu^2} & -\frac{\partial^2 \Lambda}{\partial \mu \partial k} \\
-\frac{\partial^2 \Lambda}{\partial \mu \partial k} & -\frac{\partial^2 \Lambda}{\partial k^2}
\end{pmatrix}_{\mu = \hat{\mu}, k = \hat{k}}^{-1}
\]

\( \Lambda \) is the log-likelihood function of the gamma distribution, described in Parameter Estimation and Appendix D.
**Bounds on Reliability**

The reliability of the gamma distribution is:

\[
\hat{R}(t; \hat{\mu}, \hat{k}) = 1 - \Gamma_1(\hat{k}; e^{\hat{\mu}})
\]

where:

\[
\hat{\mu} = \ln(t) - \hat{\mu}
\]

The upper and lower bounds on reliability are:

\[
R_U = \frac{\hat{R}}{R + (1 - \hat{R}) \exp\left(\frac{-K_0 \sqrt{\text{Var}(\hat{R})}}{R(1 - \hat{R})}\right)} \quad \text{(upper bound)}
\]

\[
R_L = \frac{\hat{R}}{R + (1 - \hat{R}) \exp\left(\frac{-K_0 \sqrt{\text{Var}(\hat{R})}}{R(1 - \hat{R})}\right)} \quad \text{(lower bound)}
\]

where:

\[
\text{Var}(\hat{R}) = \left(\frac{\partial R}{\partial \mu}\right)^2 \text{Var}(\hat{\mu}) + 2\left(\frac{\partial R}{\partial \mu}\right)\left(\frac{\partial R}{\partial k}\right)\text{Cov}(\hat{\mu}, \hat{k}) + \left(\frac{\partial R}{\partial k}\right)^2 \text{Var}(\hat{k})
\]

**Bounds on Time**

The bounds around time for a given gamma percentile (unreliability) are estimated by first solving the reliability equation with respect to time, as follows:

\[
\hat{T}(\hat{\mu}, \hat{\sigma}) = \hat{\mu} + \hat{\sigma} z
\]

where:

\[
z = \ln\left(-\ln(R)\right)
\]

\[
\text{Var}(\hat{T}) = \left(\frac{\partial T}{\partial \mu}\right)^2 \text{Var}(\hat{\mu}) + 2\left(\frac{\partial T}{\partial \mu}\right)\left(\frac{\partial T}{\partial \sigma}\right)\text{Cov}(\hat{\mu}, \hat{\sigma}) + \left(\frac{\partial T}{\partial \sigma}\right)^2 \text{Var}(\hat{\sigma})
\]

or:

\[
\text{Var}(\hat{T}) = \text{Var}(\hat{\mu}) + 2z\text{Cov}(\hat{\mu}, \hat{\sigma}) + z^2\text{Var}(\hat{\sigma})
\]

The upper and lower bounds are then found by:

\[
T_U = \hat{T} + K_0 \sqrt{\text{Var}(\hat{T})} \quad \text{(Upper bound)}
\]

\[
T_L = \hat{T} - K_0 \sqrt{\text{Var}(\hat{T})} \quad \text{(Lower bound)}
\]

**General Example**

24 units were reliability tested, and the following life test data were obtained:

<table>
<thead>
<tr>
<th>Data</th>
<th>61</th>
<th>50</th>
<th>67</th>
<th>49</th>
<th>53</th>
<th>62</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53</td>
<td>61</td>
<td>43</td>
<td>65</td>
<td>53</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>56</td>
<td>58</td>
<td>55</td>
<td>58</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>66</td>
<td>44</td>
<td>48</td>
<td>58</td>
<td>43</td>
<td>40</td>
</tr>
</tbody>
</table>

Fitting the gamma distribution to this data, using maximum likelihood as the analysis method, gives the following parameters:

\[
\hat{\mu} = 7.72E - 02
\]

\[
\hat{k} = 50.4908
\]

Using rank regression on \(X\), the estimated parameters are:
\[ \hat{\mu} = 0.2915 \]
\[ \hat{k} = 41.1726 \]

Using rank regression on \( \bar{Y} \), the estimated parameters are:

\[ \hat{\mu} = 0.2915 \]
\[ \hat{k} = 41.1726 \]
Chapter 14

The Logistic Distribution

The logistic distribution has been used for growth models, and is used in a certain type of regression known as the logistic regression. It has also applications in modeling life data. The shape of the logistic distribution and the normal distribution are very similar, as discussed in Meeker and Escobar [27]. There are some who argue that the logistic distribution is inappropriate for modeling lifetime data because the left-hand limit of the distribution extends to negative infinity. This could conceivably result in modeling negative times-to-failure. However, provided that the distribution in question has a relatively high mean and a relatively small location parameter, the issue of negative failure times should not present itself as a problem.

Logistic Probability Density Function

The logistic pdf is given by:

\[ f(t) = \frac{e^{z}}{\sigma (1+e^{z})^2} \]

where:

\[ z = \frac{t-\mu}{\sigma} \]

\[ -\infty < t < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0 \]

where:

\( \mu = \) location parameter (also denoted as \( \bar{T} \))

\( \sigma = \) scale parameter

The Logistic Mean, Median and Mode

The logistic mean or MTTF is actually one of the parameters of the distribution, usually denoted as \( \mu \). Since the logistic distribution is symmetrical, the median and the mode are always equal to the mean, \( \mu = \bar{T} = \tilde{T} \).
The Logistic Standard Deviation

The standard deviation of the logistic distribution, $\sigma T$, is given by:

$$\sigma T = \sigma \pi \frac{\sqrt{3}}{3}$$

The Logistic Reliability Function

The reliability for a mission of time $t$, starting at age 0, for the logistic distribution is determined by:

$$R(t) = \int_t^\infty f(t)dt$$

or:

$$R(t) = \frac{1}{1 + e^z}$$

The unreliability function is:

$$F = \frac{e^z}{1 + e^z}$$

where:

$$z = \frac{t - \mu}{\sigma}$$

The Logistic Conditional Reliability Function

The logistic conditional reliability function is given by:

$$R(t|T) = \frac{R(T + t)}{R(T)} = \frac{1 + e^{\frac{t - \mu}{\sigma}}}{1 + e^{\frac{T - \mu}{\sigma}}}$$

The Logistic Reliable Life

The logistic reliable life is given by:

$$T_R = \mu + \sigma [\ln(1 - R) - \ln(R)]$$

The Logistic Failure Rate Function

The logistic failure rate function is given by:

$$\lambda(t) = \frac{e^z}{\sigma(1 + e^z)}$$

Characteristics of the Logistic Distribution

- The logistic distribution has no shape parameter. This means that the logistic pdf has only one shape, the bell shape, and this shape does not change. The shape of the logistic distribution is very similar to that of the normal distribution.
- The mean, $\mu$, or the mean life or the $MTTF$, is also the location parameter of the logistic pdf, as it locates the pdf along the abscissa. It can assume values of $-\infty < \mu < \infty$.
- As $\mu$ decreases, the pdf is shifted to the left.
- As $\mu$ increases, the pdf is shifted to the right.
The Logistic Distribution

**Effect of \( \mu \) on Logistic Distribution pdf**

- As \( \sigma \) decreases, the pdf gets pushed toward the mean, or it becomes narrower and taller.
- As \( \sigma \) increases, the pdf spreads out away from the mean, or it becomes broader and shallower.
- The scale parameter can assume values of \( 0 < \sigma < \infty \).

**Effect of \( \sigma \) on Logistic Distribution pdf**

- The logistic pdf starts at \( t = -\infty \) with \( f(t) = 0 \). As \( t \) increases, \( f(t) \) also increases, goes through its point of inflection and reaches its maximum value at \( t = \hat{t} \). Thereafter, \( f(t) \) decreases, goes through its point of inflection and assumes a value of \( f(t) = 0 \) at \( t = +\infty \).
- For \( t = \pm \infty \) the pdf equals 0. The maximum value of the pdf occurs at \( t = \mu \) and equals \( \frac{1}{4\sigma^2} \).
- The point of inflection of the pdf plot is the point where the second derivative of the pdf equals zero. The inflection point occurs at \( t = \mu + \sigma \ln(2 \pm \sqrt{3}) \) or \( t \approx \mu \pm \sigma 1.31696 \).
• If the location parameter $\mu$ decreases, the reliability plot is shifted to the left. If $\mu$ increases, the reliability plot is shifted to the right.

• If $t = \mu$ then $R = 0.5$ is the inflection point. If $t < \mu$ then $R(t)$ is concave (concave down); if $t > \mu$ then $R(t)$ is convex (concave up). For $t < \mu$, $\lambda(t)$ is convex (concave up), for $t > \mu$, $\lambda(t)$ is concave (concave down).

• The main difference between the normal distribution and logistic distribution lies in the tails and in the behavior of the failure rate function. The logistic distribution has slightly longer tails compared to the normal distribution. Also, in the upper tail of the logistic distribution, the failure rate function levels out for large $t$, approaching $1/\delta$.

• If location parameter $\mu$ decreases, the failure rate plot is shifted to the left. Vice versa if $\mu$ increases, the failure rate plot is shifted to the right.

• $\lambda(t)$ always increases. For $t \to -\infty$ or $t \to \infty$, it is always $0 \leq \lambda(t) \leq \frac{1}{\sigma}$.

• If $\sigma$ increases, then $\lambda(t)$ increases more slowly and smoothly. The segment of time where $0 < \lambda(t) < \frac{1}{\sigma}$ increases, too, whereas the region where $\lambda(t)$ is close to 0 or $\frac{1}{\sigma}$ gets narrower. Conversely, if $\sigma$ decreases, then $\lambda(t)$ increases more quickly and sharply. The segment of time where $0 < \lambda(t) < \frac{1}{\sigma}$ decreases, too, whereas the region where $\lambda(t)$ is close to 0 or $\frac{1}{\sigma}$ gets broader.

**Weibull++ Notes on Negative Time Values**

One of the disadvantages of using the logistic distribution for reliability calculations is the fact that the logistic distribution starts at negative infinity. This can result in negative values for some of the results. Negative values for time are not accepted in most of the components of Weibull++, nor are they implemented. Certain components of the application reserve negative values for suspensions, or will not return negative results. For example, the Quick Calculation Pad will return a null value (zero) if the result is negative. Only the Free-Form (Probit) data sheet can accept negative values for the random variable (x-axis values).

**Logistic Distribution Probability Paper**

The form of the Logistic probability paper is based on linearizing the $cdf$. From unreliability equation, $z$ can be calculated as a function of the $cdf$ $F$ as follows:

$$z = \ln(F) - \ln(1 - F)$$

or using the equation for $z$

$$\frac{t - \mu}{\sigma} = \ln(F) - \ln(1 - F)$$

Then:

$$\ln(F) - \ln(1 - F) = -\frac{\mu}{\sigma} + \frac{1}{\sigma} t$$

Now let:

$$y = \ln(F) - \ln(1 - F)$$

$$x = t$$

and:

$$a = -\frac{\mu}{\sigma}$$

$$b = \frac{1}{\sigma}$$

which results in the following linear equation:

$$y = a + bx$$

The logistic probability paper resulting from this linearized $cdf$ function is shown next.
Since the logistic distribution is symmetrical, the area under the pdf curve from $-\infty$ to $\mu$ is 0.5, as is the area from $\mu$ to $+\infty$. Consequently, the value of $\mu$ is said to be the point where $R(t) = Q(t) = 50\%$. This means that the estimate of $\mu$ can be read from the point where the plotted line crosses the 50\% unreliability line.

For $z = 1$, $\sigma = t - \mu$ and $R(t) = \frac{1}{1+\exp(-t)} \approx 0.2689$. Therefore, $\sigma$ can be found by subtracting $\mu$ from the time value where the plotted probability line crosses the 73.10\% unreliability (26.89\% reliability) horizontal line.

**Confidence Bounds**

In this section, we present the methods used in the application to estimate the different types of confidence bounds for logistically distributed data. The complete derivations were presented in detail (for a general function) in Confidence Bounds.

**Bounds on the Parameters**

The lower and upper bounds on the location parameter $\hat{\mu}$ are estimated from:

$$
\mu_U = \hat{\mu} + K_\alpha \sqrt{Var(\hat{\mu})} \quad \text{(upper bound)}
$$

$$
\mu_L = \hat{\mu} - K_\alpha \sqrt{Var(\hat{\mu})} \quad \text{(lower bound)}
$$

The lower and upper bounds on the scale parameter $\hat{\sigma}$ are estimated from:

$$
\sigma_U = \hat{\sigma} e^{\frac{K_\alpha \sqrt{Var(\hat{\sigma})}}{\hat{\sigma}}} \quad \text{(upper bound)}
$$

$$
\sigma_L = \hat{\sigma} e^{-\frac{K_\alpha \sqrt{Var(\hat{\sigma})}}{\hat{\sigma}}} \quad \text{(lower bound)}
$$

where $K_\alpha$ is defined by:

$$
\alpha = \frac{1}{\sqrt{2\pi}} \int_{K_\alpha}^{\infty} e^{-\frac{t^2}{2}} dt = 1 - \Phi(K_\alpha)
$$

If $\delta$ is the confidence level, then $\alpha = 1 - \frac{\delta}{2}$ for the two-sided bounds, and $\alpha = 1 - \frac{\delta}{2}$ for the one-sided bounds.

The variances and covariances of $\hat{\mu}$ and $\hat{\sigma}$ are estimated from the Fisher matrix, as follows:
\[
\begin{pmatrix}
\hat{V}\bar{ar}(\hat{\mu}) & \hat{C}\bar{ov}(\hat{\mu}, \hat{\sigma}) \\
\hat{C}\bar{ov}(\hat{\mu}, \hat{\sigma}) & \hat{V}\bar{ar}(\hat{\sigma})
\end{pmatrix}
\begin{pmatrix}
- \frac{\partial^2 \Lambda}{\partial \mu^2} & - \frac{\partial^2 \Lambda}{\partial \mu \partial \sigma} \\
- \frac{\partial^2 \Lambda}{\partial \mu \partial \sigma} & - \frac{\partial^2 \Lambda}{\partial \sigma^2}
\end{pmatrix}^{-1}
\mu = \hat{\mu}, \sigma = \hat{\sigma}
\]

\(\Lambda\) is the log-likelihood function of the normal distribution, described in Parameter Estimation and Appendix D.

**Bounds on Reliability**

The reliability of the logistic distribution is:

\[
\hat{R} = \frac{1}{1 + e^{\hat{z}}}
\]

where:

\[
\hat{z} = \frac{t - \hat{\mu}}{\hat{\sigma}}
\]

Here \(-\infty < t < \infty, -\infty < \mu < \infty, 0 < \sigma < \infty\). Therefore, \(z\) also is changing from \(-\infty\) to \(+\infty\). Then the bounds on \(z\) are estimated from:

\[
\hat{z}_U = \hat{z} + K_\alpha \sqrt{\hat{V}ar(\hat{z})} \\
\hat{z}_L = \hat{z} - K_\alpha \sqrt{\hat{V}ar(\hat{z})}
\]

where:

\[
\hat{V}ar(\hat{z}) = \left(\frac{\partial \hat{z}}{\partial \hat{\mu}}\right)^2 \hat{V}ar(\hat{\mu}) + 2\left(\frac{\partial \hat{z}}{\partial \hat{\mu}}\right)\left(\frac{\partial \hat{z}}{\partial \hat{\sigma}}\right)\hat{C}ov(\hat{\mu}, \hat{\sigma}) + \left(\frac{\partial \hat{z}}{\partial \hat{\sigma}}\right)^2 \hat{V}ar(\hat{\sigma})
\]

or:

\[
\hat{V}ar(\hat{z}) = \frac{1}{\hat{\sigma}^2} \left[\hat{V}ar(\hat{\mu}) + 2\hat{z}\hat{C}ov(\hat{\mu}, \hat{\sigma}) + \hat{z}^2 \hat{V}ar(\hat{\sigma})\right]
\]

The upper and lower bounds on reliability are:

\[
\hat{R}_U = \frac{1}{1 + e^{\hat{z}_L}} \text{ (upper bound)} \\
\hat{R}_L = \frac{1}{1 + e^{\hat{z}_U}} \text{ (lower bound)}
\]

**Bounds on Time**

The bounds around time for a given logistic percentile (unreliability) are estimated by first solving the reliability equation with respect to time as follows:

\[
\hat{T}(\hat{\mu}, \hat{\sigma}) = \hat{\mu} + \hat{\sigma} z
\]

where:

\[
z = \ln(1 - \hat{R}) - \ln(\hat{R})
\]

\[
\hat{V}ar(\hat{T}) = \left(\frac{\partial \hat{T}}{\partial \hat{\mu}}\right)^2 \hat{V}ar(\hat{\mu}) + 2\left(\frac{\partial \hat{T}}{\partial \hat{\mu}}\right)\left(\frac{\partial \hat{T}}{\partial \hat{\sigma}}\right)\hat{C}ov(\hat{\mu}, \hat{\sigma}) + \left(\frac{\partial \hat{T}}{\partial \hat{\sigma}}\right)^2 \hat{V}ar(\hat{\sigma})
\]

or:

\[
\hat{V}ar(\hat{T}) = \hat{V}ar(\hat{\mu}) + 2\hat{z}\hat{C}ov(\hat{\mu}, \hat{\sigma}) + \hat{z}^2 \hat{V}ar(\hat{\sigma})
\]

The upper and lower bounds are then found by:

\[
\hat{T}_U = \hat{T} + K_\alpha \sqrt{\hat{V}ar(\hat{T})} \text{ (upper bound)} \\
\hat{T}_L = \hat{T} - K_\alpha \sqrt{\hat{V}ar(\hat{T})} \text{ (lower bound)}
\]
General Example

The lifetime of a mechanical valve is known to follow a logistic distribution. 10 units were tested for 28 months and the following months-to-failure data were collected.

Times-to-Failure Data with Suspensions

<table>
<thead>
<tr>
<th>Data Point Index</th>
<th>State F or S</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>27</td>
</tr>
<tr>
<td>8</td>
<td>S</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>S</td>
<td>28</td>
</tr>
<tr>
<td>10</td>
<td>S</td>
<td>28</td>
</tr>
</tbody>
</table>

• Determine the valve's design life if specifications call for a reliability goal of 0.90.
• The valve is to be used in a pumping device that requires 1 month of continuous operation. What is the probability of the pump failing due to the valve?

Enter the data set in a Weibull++ standard folio, as follows:

![Image of Weibull++ folio]

The computed parameters for maximum likelihood are:

\[
\hat{\mu} = 22.34 \\
\hat{\sigma} = 6.15
\]

The valve's design life, along with 90% two sided confidence bounds, can be obtained using the QCP as follows:
The probability, along with 90% two sided confidence bounds, that the pump fails due to a valve failure during the first month is obtained as follows:
Chapter 15

The Loglogistic Distribution

As may be indicated by the name, the loglogistic distribution has certain similarities to the logistic distribution. A random variable is loglogistically distributed if the logarithm of the random variable is logistically distributed. Because of this, there are many mathematical similarities between the two distributions, as discussed in Meeker and Escobar [27]. For example, the mathematical reasoning for the construction of the probability plotting scales is very similar for these two distributions.

Loglogistic Probability Density Function

The loglogistic distribution is a 2-parameter distribution with parameters \( \mu \) and \( \sigma \). The pdf for this distribution is given by:

\[
f(t) = \frac{e^z}{\sigma t (1 + e^z)^2}
\]

where:

\[
z = \frac{t' - \mu}{\sigma}
\]

\[
t' = \ln(t)
\]

and:

\[
\mu = \text{scale parameter}
\]

\[
\sigma = \text{shape parameter}
\]

where \( 0 < t < \infty, -\infty < \mu < \infty \) and \( 0 < \sigma < \infty \).

Mean, Median and Mode

The mean of the loglogistic distribution, \( \bar{T} \), is given by:

\[
\bar{T} = e^\mu \Gamma(1 + \sigma) \Gamma(1 - \sigma)
\]

Note that for \( \sigma \geq 1 \), \( \bar{T} \) does not exist.

The median of the loglogistic distribution, \( \hat{T} \), is given by:

\[
\hat{T} = e^\mu
\]

The mode of the loglogistic distribution, \( \hat{T} \), if \( \sigma < 1 \) is given by:

\[
\hat{T} = e^{\mu + \sigma \ln (\frac{1 - \frac{1}{\sigma}}{\gamma})}
\]
The Loglogistic Distribution

The standard deviation of the loglogistic distribution, \( \sigma_T \), is given by:

\[
\sigma_T = e^\mu \sqrt{\Gamma(1 + 2\sigma)\Gamma(1 - 2\sigma) - (\Gamma(1 + \sigma)\Gamma(1 - \sigma))^2}
\]

Note that for \( \sigma \geq 0.5 \), the standard deviation does not exist.

The Loglogistic Reliability Function

The reliability for a mission of time \( T \), starting at age 0, for the loglogistic distribution is determined by:

\[
R = \frac{1}{1 + e^{z}}
\]

where:

\[
\begin{align*}
z &= \frac{t' - \mu}{\sigma} \\
t' &= \ln(t)
\end{align*}
\]

The unreliability function is:

\[
F = \frac{e^{z}}{1 + e^{z}}
\]

The loglogistic Reliable Life

The logistic reliable life is:

\[
T_R = e^{\mu + \sigma[\ln(1-R) - \ln(R)]}
\]

The loglogistic Failure Rate Function

The loglogistic failure rate is given by:

\[
\lambda(t) = \frac{e^{z}}{\sigma t(1 + e^{z})}
\]

Distribution Characteristics

For \( \sigma > 1 \):

- \( f(t) \) decreases monotonically and is convex. Mode and mean do not exist.

For \( \sigma = 1 \):

- \( f(t) \) decreases monotonically and is convex. Mode and mean do not exist. As \( t \to \infty \), \( f(t) \to \frac{1}{e \sigma} \).

As \( t \to \infty \), \( \lambda(t) \to \frac{1}{e \sigma} \).

For \( 0 < \sigma < 1 \):

- The shape of the loglogistic distribution is very similar to that of the lognormal distribution and the Weibull distribution.
- The pdf starts at zero, increases to its mode, and decreases thereafter.
- As \( \mu \) increases, while \( \sigma \) is kept the same, the pdf gets stretched out to the right and its height decreases, while maintaining its shape.
- As \( \mu \) decreases, while \( \sigma \) is kept the same, the pdf gets pushed in towards the left and its height increases.
- \( \lambda(t) \) increases till \( t = e^{\mu + \sigma \ln(\frac{1-\sigma}{\sigma})} \) and decreases thereafter. \( \lambda(t) \) is concave at first, then becomes convex.
The Loglogistic Distribution

Confidence Bounds

The method used by the application in estimating the different types of confidence bounds for loglogistically distributed data is presented in this section. The complete derivations were presented in detail for a general function in Parameter Estimation.

Bounds on the Parameters

The lower and upper bounds $H$, are estimated from:

$$
\mu_U = \hat{\mu} + K_\alpha \sqrt{Var(\hat{\mu})} \quad \text{(upper bound)}
$$

$$
\mu_L = \hat{\mu} - K_\alpha \sqrt{Var(\hat{\mu})} \quad \text{(lower bound)}
$$

For parameter $\hat{\sigma}$, $\ln(\hat{\sigma})$ is treated as normally distributed, and the bounds are estimated from:

$$
\sigma_U = \hat{\sigma} \cdot e^{\frac{K_\alpha \sqrt{Var(\hat{\sigma})}}{\hat{\sigma}}} \quad \text{(upper bound)}
$$

$$
\sigma_L = \frac{\hat{\sigma}}{e^{\frac{K_\alpha \sqrt{Var(\hat{\sigma})}}{\hat{\sigma}}}} \quad \text{(lower bound)}
$$

where $K_\alpha$ is defined by:

$$
\alpha = \frac{1}{\sqrt{2\pi}} \int_{K_\alpha}^{\infty} e^{-\frac{t^2}{2}} dt = 1 - \Phi(K_\alpha)
$$

If $\delta$ is the confidence level, then $\alpha = \frac{1-\delta}{2}$ for the two-sided bounds, and $\alpha = 1 - \delta$ for the one-sided bounds.

The variances and covariances of $\hat{\mu}$ and $\hat{\sigma}$ are estimated as follows:

$$
\begin{pmatrix}
Var(\hat{\mu}) & Cov(\hat{\mu}, \hat{\sigma}) \\
Cov(\hat{\mu}, \hat{\sigma}) & Var(\hat{\sigma})
\end{pmatrix}
= \begin{pmatrix}
-\frac{\partial^2 \Lambda}{\partial \mu^2} & -\frac{\partial^2 \Lambda}{\partial \mu \partial \sigma} \\
-\frac{\partial^2 \Lambda}{\partial \mu \partial \sigma} & -\frac{\partial^2 \Lambda}{\partial \sigma^2}
\end{pmatrix}^{-1}
\mu = \hat{\mu}, \sigma = \hat{\sigma}
$$

where $\Lambda$ is the log-likelihood function of the loglogistic distribution.
Bounds on Reliability

The reliability of the logistic distribution is:

$$\hat{R} = \frac{1}{1 + \exp(\bar{z})}$$

where:

$$\bar{z} = \frac{t' - \hat{\mu}}{\hat{\sigma}}$$

Here $0 < t < \infty$, $-\infty < \mu < \infty$, $0 < \sigma < \infty$, therefore $0 < t' = \ln(t) < \infty$ and $\bar{z}$ also is changing from $-\infty$ till $+\infty$.

The bounds on $\bar{z}$ are estimated from:

$$z_U = \bar{z} + K_\alpha \sqrt{Var(\bar{z})}$$
$$z_L = \bar{z} - K_\alpha \sqrt{Var(\bar{z})}$$

where:

$$Var(\bar{z}) = \left(\frac{\partial^2 z}{\partial \mu} \right) Var(\bar{\mu}) + 2\left(\frac{\partial z}{\partial \mu} \right) C \left(\frac{\partial z}{\partial \sigma} \right) Cov(\bar{\mu}, \bar{\sigma}) + \left(\frac{\partial z}{\partial \sigma} \right) Var(\bar{\sigma})$$

or:

$$Var(\bar{z}) = \frac{1}{\sigma^2} \left(Var(\bar{\mu}) + 2\bar{z}C \left(\frac{\partial z}{\partial \sigma} \right) Cov(\bar{\mu}, \bar{\sigma}) + \bar{z}^2 Var(\bar{\sigma})\right)$$

The upper and lower bounds on reliability are:

$$R_U = \frac{1}{1 + e^{z_U}} \text{(Upper bound)}$$
$$R_L = \frac{1}{1 + e^{z_L}} \text{(Lower bound)}$$

Bounds on Time

The bounds around time for a given loglogistic percentile, or unreliability, are estimated by first solving the reliability equation with respect to time, as follows:

$$\hat{T}(\hat{\mu}, \hat{\sigma}) = e^{\hat{\mu} + \hat{\sigma} z}$$

where:

$$z = \ln(1 - R) - \ln(R)$$

or:

$$\ln(\hat{T}) = \hat{\mu} + \hat{\sigma} z$$

Let:

$$u = \ln(\hat{T}) = \hat{\mu} + \hat{\sigma} z$$

then:

$$u_U = \hat{u} + K_\alpha \sqrt{Var(\hat{u})}$$
$$u_L = \hat{u} - K_\alpha \sqrt{Var(\hat{u})}$$

where:

$$Var(\hat{u}) = \left(\frac{\partial u}{\partial \mu} \right) Var(\hat{\mu}) + 2\left(\frac{\partial u}{\partial \mu} \right) C \left(\frac{\partial u}{\partial \sigma} \right) Cov(\hat{\mu}, \hat{\sigma}) + \left(\frac{\partial u}{\partial \sigma} \right) Var(\hat{\sigma})$$

or:

$$Var(\hat{u}) = Var(\hat{\mu}) + 2\hat{z}C \left(\frac{\partial u}{\partial \sigma} \right) Cov(\hat{\mu}, \hat{\sigma}) + \hat{z}^2 Var(\hat{\sigma})$$
The upper and lower bounds are then found by:

\[ T_U = e^{u_U} \text{ (upper bound)} \]
\[ T_L = e^{u_L} \text{ (lower bound)} \]

**General Examples**

Determine the loglogistic parameter estimates for the data given in the following table.

<table>
<thead>
<tr>
<th>Data point index</th>
<th>Last Inspected</th>
<th>State</th>
<th>End time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>105</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>197</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>297</td>
<td>301</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>330</td>
<td>335</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>393</td>
<td>401</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>423</td>
<td>426</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>460</td>
<td>468</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>569</td>
<td>570</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>675</td>
<td>680</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>884</td>
<td>889</td>
<td></td>
</tr>
</tbody>
</table>

Set up the folio for times-to-failure data that includes interval and left censored data, then enter the data. The computed parameters for maximum likelihood are calculated to be:

\[ \hat{\mu}' = 5.9772 \]
\[ \hat{\sigma}_T = 0.3256 \]

For rank regression on \( X \):

\[ \hat{\mu} = 5.9281 \]
\[ \hat{\sigma} = 0.3821 \]

For rank regression on \( Y \):

\[ \hat{\mu} = 5.9772 \]
\[ \hat{\sigma} = 0.3256 \]
The Gumbel/SEV Distribution

The Gumbel distribution is also referred to as the Smallest Extreme Value (SEV) distribution or the Smallest Extreme Value (Type I) distribution. The Gumbel distribution's pdf is skewed to the left, unlike the Weibull distribution's pdf, which is skewed to the right. The Gumbel distribution is appropriate for modeling strength, which is sometimes skewed to the left (few weak units in the lower tail, most units in the upper tail of the strength population). The Gumbel distribution could also be appropriate for modeling the life of products that experience very quick wear-out after reaching a certain age. The distribution of logarithms of times can often be modeled with the Gumbel distribution (in addition to the more common lognormal distribution), as discussed in Meeker and Escobar [27].

Gumbel Probability Density Function

The pdf of the Gumbel distribution is given by:

\[ f(t) = \frac{1}{\sigma} e^{-z} e^{-e^{-z}} \]

where:

\[ z = \frac{t - \mu}{\sigma} \]

and:

\[ \mu = \text{location parameter} \]

\[ \sigma = \text{scale parameter} \]

Gumbel Mean, Median and Mode

The Gumbel mean or MTTF is:

\[ \overline{T} = \mu - \sigma \gamma \]

where \( \gamma \approx 0.5772 \) (Euler's constant).

The mode of the Gumbel distribution is:

\[ \hat{T} = \mu \]

The median of the Gumbel distribution is:

\[ \hat{T} = \mu + \sigma \ln(\ln(2)) \]
The Gumbel/SEV Distribution

**Gumbel Standard Deviation**

The standard deviation for the Gumbel distribution is given by:

\[ \sigma_T = \sigma \pi \sqrt{\frac{6}{\pi}} \]

**Gumbel Reliability Function**

The reliability for a mission of time \( t \) for the Gumbel distribution is given by:

\[ R(t) = e^{-e^{-t}} \]

The unreliability function is given by:

\[ F(t) = 1 - e^{-e^{-t}} \]

**Gumbel Reliable Life**

The Gumbel reliable life is given by:

\[ t_R = \mu + \sigma [\ln(-\ln(R))] \]

**Gumbel Failure Rate Function**

The instantaneous Gumbel failure rate is given by:

\[ \lambda = \frac{e^2}{\sigma} \]

**Characteristics of the Gumbel Distribution**

Some of the specific characteristics of the Gumbel distribution are the following:

- The shape of the Gumbel distribution is skewed to the left. The Gumbel pdf has no shape parameter. This means that the Gumbel pdf has only one shape, which does not change. 
- The Gumbel pdf has location parameter \( \mu \) which is equal to the mode \( \hat{T} \) but it differs from median and mean. This is because the Gumbel distribution is not symmetrical about its \( \mu \). 
- As \( \mu \) decreases, the pdf is shifted to the left. 
- As \( \mu \) increases, the pdf is shifted to the right. 

- As \( \sigma \) increases, the pdf spreads out and becomes shallower.
As $\sigma$ decreases, the pdf becomes taller and narrower.

For $T = \pm \infty$, pdf=0. For $T = \mu$, the pdf reaches its maximum point $\frac{1}{\sigma e}$

- The points of inflection of the pdf graph are $T = \mu \pm \sigma \ln\left(\frac{\beta}{2}\right)$ for $T \approx \mu \pm 0.96242$
- If times follow the Weibull distribution, then the logarithm of times follow a Gumbel distribution. If $t$ follows a Weibull distribution with $\beta$ and $\eta$, then the $\ln(t)$ follows a Gumbel distribution with $\mu = \ln(\eta)$ and $\sigma = \frac{1}{\beta}$ as discussed in [32].

**Probability Paper**

The form of the Gumbel probability paper is based on a linearization of the cdf. From the unreliability equation, we know:

$$z = \ln(-\ln(1 - F))$$

Using the equation for $z$, we get:

$$\frac{t - \mu}{\sigma} = \ln(-\ln(1 - F))$$

Then:

$$\ln(-\ln(1 - F)) = -\frac{\mu}{\sigma} + \frac{1}{\sigma}t$$

Now let:

$$y = \ln(-\ln(1 - F))$$

$$x = t$$

And:

$$a = -\frac{\mu}{\sigma}$$

$$b = \frac{1}{\sigma}$$

Which results in the linear equation of:

$$y = a + bx$$

The Gumbel probability paper resulting from this linearized cdf function is shown next.
For $z = 0$, $t = \mu$ and $R(t) = e^{-e^0} \approx 0.3678$ (63.21% unreliability). For $z = 1$, $\sigma = T - \mu$ and $R(t) = e^{-e^1} \approx 0.0659$. To read $\mu$ from the plot, find the time value that corresponds to the intersection of the probability plot with the 63.21% unreliability line. To read $\sigma$ from the plot, find the time value that corresponds to the intersection of the probability plot with the 93.40% unreliability line, then take the difference between this time value and the $\mu$ value.

**Confidence Bounds**

This section presents the method used by the application to estimate the different types of confidence bounds for data that follow the Gumbel distribution. The complete derivations were presented in detail (for a general function) in Confidence Bounds. Only Fisher Matrix confidence bounds are available for the Gumbel distribution.

**Bounds on the Parameters**

The lower and upper bounds on the mean, $\hat{\mu}$, are estimated from:

$$\mu_U = \hat{\mu} + K_\alpha \sqrt{\text{Var}(\hat{\mu})} \quad \text{(upper bound)}$$

$$\mu_L = \hat{\mu} - K_\alpha \sqrt{\text{Var}(\hat{\mu})} \quad \text{(lower bound)}$$

Since the standard deviation, $\hat{\sigma}$, must be positive, then $\ln(\hat{\sigma})$ is treated as normally distributed, and the bounds are estimated from:

$$\sigma_U = \hat{\sigma} \cdot e \frac{K_\alpha \sqrt{\text{Var}(\hat{\sigma})}}{\hat{\sigma}} \quad \text{(upper bound)}$$

$$\sigma_L = \hat{\sigma} \frac{K_\alpha \sqrt{\text{Var}(\hat{\sigma})}}{e} \quad \text{(lower bound)}$$

where $K_\alpha$ is defined by:

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 1 - \Phi(K_\alpha)$$

If $\delta$ is the confidence level, then $\alpha = \frac{1-\delta}{2}$ for the two-sided bounds, and $\alpha = 1 - \delta$ for the one-sided bounds.

The variances and covariances of $\hat{\mu}$ and $\hat{\sigma}$ are estimated from the Fisher matrix as follows:
The Gumbel/SEV Distribution

Bounds on Reliability

The reliability of the Gumbel distribution is given by:

\[ \hat{R}(t; \hat{\mu}, \hat{\sigma}) = e^{-e^{\frac{t}{\hat{\sigma}}}} \]

where:

\[ \hat{z} = \frac{t - \hat{\mu}}{\hat{\sigma}} \]

The bounds on \( \hat{z} \) are estimated from:

\[ z_U = \hat{z} + K_0 \sqrt{\text{Var}(\hat{z})} \]

\[ z_L = \hat{z} - K_0 \sqrt{\text{Var}(\hat{z})} \]

where:

\[ \text{Var}(\hat{z}) = \left( \frac{\partial \hat{z}}{\partial \hat{\mu}} \right)^2 \text{Var}(\hat{\mu}) + \left( \frac{\partial \hat{z}}{\partial \hat{\sigma}} \right)^2 \text{Var}(\hat{\sigma}) + 2 \left( \frac{\partial \hat{z}}{\partial \hat{\mu}} \right) \left( \frac{\partial \hat{z}}{\partial \hat{\sigma}} \right) \text{Cov}(\hat{\mu}, \hat{\sigma}) \]

or:

\[ \text{Var}(\hat{z}) = \frac{1}{\hat{\sigma}^2} \left[ \text{Var}(\hat{\mu}) + \hat{z}^2 \text{Var}(\hat{\sigma}) + 2 \cdot \hat{z} \cdot \text{Cov}(\hat{\mu}, \hat{\sigma}) \right] \]

The upper and lower bounds on reliability are:

\[ R_U = e^{-e^{z_U}} \quad \text{(upper bound)} \]
\[ R_L = e^{-e^{z_L}} \quad \text{(lower bound)} \]

Bounds on Time

The bounds around time for a given Gumbel percentile (unreliability) are estimated by first solving the reliability equation with respect to time, as follows:

\[ \hat{T}(\hat{\mu}, \hat{\sigma}) = \hat{\mu} + \hat{\sigma} z \]

where:

\[ z = \ln(-\ln(\hat{R})) \]

\[ \text{Var}(\hat{T}) = \left( \frac{\partial \hat{T}}{\partial \hat{\mu}} \right)^2 \text{Var}(\hat{\mu}) + 2 \left( \frac{\partial \hat{T}}{\partial \hat{\mu}} \right) \left( \frac{\partial \hat{T}}{\partial \hat{\sigma}} \right) \text{Cov}(\hat{\mu}, \hat{\sigma}) + \left( \frac{\partial \hat{T}}{\partial \hat{\sigma}} \right)^2 \text{Var}(\hat{\sigma}) \]

or:

\[ \text{Var}(\hat{T}) = \text{Var}(\hat{\mu}) + 2\hat{z} \text{Cov}(\hat{\mu}, \hat{\sigma}) + \hat{z}^2 \text{Var}(\hat{\sigma}) \]

The upper and lower bounds are then found by:

\[ T_U = \hat{T} + K_0 \sqrt{\text{Var}(\hat{T})} \quad \text{(Upper bound)} \]
\[ T_L = \hat{T} - K_0 \sqrt{\text{Var}(\hat{T})} \quad \text{(Lower bound)} \]
Example

Verify using Monte Carlo simulation that if \( t_i \) follows a Weibull distribution with \( \beta \) and \( \eta \), then the \( \ln(t_i) \) follows a Gumbel distribution with \( \mu = \ln(\eta) \) and \( \delta = 1/\beta \).

Let us assume that \( t_i \) follows a Weibull distribution with \( \beta = 0.5 \) and \( \eta = 10000 \). The Monte Carlo simulation tool in Weibull++ can be used to generate a set of random numbers that follow a Weibull distribution with the specified parameters. The following picture shows the Main tab of the Monte Carlo Data Generation utility.

On the Settings tab, set the number of points to 100 and click Generate. This creates a new data sheet in the folio that contains random time values \( t_i \).

Insert a new data sheet in the folio and enter the corresponding \( \ln(t_i) \) values of the time values generated by the Monte Carlo simulation. Delete any negative values, if there are any, because Weibull++ expects the time values to be positive. After obtaining the \( \ln(t_i) \) values, analyze the data sheet using the Gumbel distribution and the MLE parameter estimation method. The estimated parameters are (your results may vary due to the random numbers generated by simulation):

\[
\hat{\mu} = 9.3816 \quad \hat{\delta} = 1.9717
\]

Because \( \ln(\eta) = 9.2103 (\approx 9.3816) \) and \( 1/\beta = 2 (\approx 1.9717) \), then this simulation verifies that \( \ln(t_i) \) follows a Gumbel distribution with \( \mu = \ln(\eta) \) and \( \delta = 1/\beta \).

Note: This example illustrates a property of the Gumbel distribution; it is not meant to be a formal proof.
Non-Parametric Life Data Analysis

Non-parametric analysis allows the user to analyze data without assuming an underlying distribution. This can have certain advantages as well as disadvantages. The ability to analyze data without assuming an underlying life distribution avoids the potentially large errors brought about by making incorrect assumptions about the distribution. On the other hand, the confidence bounds associated with non-parametric analysis are usually much wider than those calculated via parametric analysis, and predictions outside the range of the observations are not possible. Some practitioners recommend that any set of life data should first be subjected to a non-parametric analysis before moving on to the assumption of an underlying distribution.

There are several methods for conducting a non-parametric analysis. In Weibull++, this includes the Kaplan-Meier, actuarial-simple and actuarial-standard methods. A method for attaching confidence bounds to the results of these non-parametric analysis techniques can also be developed. The basis of non-parametric life data analysis is the empirical cdf function, which is given by:

\[ \hat{F}(t) = \frac{\text{observations} \leq t}{n} \]

Note that this is similar to the Benard's approximation of the median ranks, as discussed in the Parameter Estimation chapter. The following non-parametric analysis methods are essentially variations of this concept.

**Kaplan-Meier Estimator**

The Kaplan-Meier estimator, also known as the product limit estimator, can be used to calculate values for non-parametric reliability for data sets with multiple failures and suspensions. The equation of the estimator is given by:

\[ \hat{R}(t_i) = \prod_{j=1}^{i} \frac{n_j - r_j}{n_j}, \quad i = 1, \ldots, m \]

where:
- \( m \) = the total number of data points
- \( n \) = the total number of units

The variable \( n_i \) is defined by:

\[ n_i = n - \sum_{j=0}^{i-1} s_j - \sum_{j=0}^{i-1} r_j, \quad i = 1, \ldots, m \]

where:
- \( r_j \) = the number of failures in the \( j^{th} \) data group
- \( s_j \) = the number of suspensions in the \( j^{th} \) data group

Note that the reliability estimate is only calculated for times at which one or more failures occurred. For the sake of calculating the value of \( \hat{R}(t_i) \) at time values that have failures and suspensions, it is assumed that the suspensions occur slightly after the failures, so that the suspended units are considered to be operating and included in the count of \( n_i \).
**Kaplan-Meier Example**

A group of 20 units are put on a life test with the following results.

<table>
<thead>
<tr>
<th>Number in State</th>
<th>State (For S)</th>
<th>State End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>F</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>17</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>21</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>22</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>24</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>26</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>28</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>35</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>39</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>41</td>
</tr>
</tbody>
</table>

Use the Kaplan-Meier estimator to determine the reliability estimates for each failure time.

**Solution**

Using the data and the reliability equation of the Kaplan-Meier estimator, the following table can be constructed:

<table>
<thead>
<tr>
<th>State End Time</th>
<th>Number of Failures, $r_i$</th>
<th>Number of Suspensions, $s_i$</th>
<th>Available Units, $n_i$</th>
<th>$\frac{n_i-r_i}{n_i}$</th>
<th>$\prod_{r_i}^{n_i-r_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>3</td>
<td>1</td>
<td>20</td>
<td>0.850</td>
<td>0.850</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td>0.938</td>
<td>0.797</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td>1.000</td>
<td>0.797</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>14</td>
<td>0.929</td>
<td>0.740</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>1.000</td>
<td>0.740</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>0.909</td>
<td>0.673</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>0.900</td>
<td>0.605</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>1.000</td>
<td>0.605</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>1.000</td>
<td>0.605</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>1.000</td>
<td>0.605</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0.833</td>
<td>0.505</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0.800</td>
<td>0.404</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1.000</td>
<td>0.404</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1.000</td>
<td>0.404</td>
</tr>
<tr>
<td>39</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1.000</td>
<td>0.404</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>0.404</td>
</tr>
</tbody>
</table>

As can be determined from the preceding table, the reliability estimates for the failure times are:
Actuarial-Simple Method

The actuarial-simple method is an easy-to-use form of non-parametric data analysis that can be used for multiple censored data that are arranged in intervals. This method is based on calculating the number of failures in a time interval, $T_i$, versus the number of operating units in that time period, $n_j$. The equation for the reliability estimator for the standard actuarial method is given by:

$$\hat{R}(t_i) = \prod_{j=1}^{i} \left( 1 - \frac{T_j}{n_j} \right), \quad i = 1, \ldots, m$$

where:

- $m$ = the total number of intervals
- $n$ = the total number of units

The variable $n_i$ is defined by:

$$n_i = n - \sum_{j=0}^{i-1} s_j - \sum_{j=0}^{i-1} r_j, \quad i = 1, \ldots, m$$

where:

- $r_j$ = the number of failures in interval $j$
- $s_j$ = the number of suspensions in interval $j$

Actuarial-Simple Example

A group of 55 units are put on a life test during which the units are evaluated every 50 hours. The results are:

<table>
<thead>
<tr>
<th>Start Time</th>
<th>End Time</th>
<th>Number of Failures, $r_i$</th>
<th>Number of Suspensions, $s_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>150</td>
<td>200</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>200</td>
<td>250</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>250</td>
<td>300</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>300</td>
<td>350</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>350</td>
<td>400</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>400</td>
<td>450</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>450</td>
<td>500</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>500</td>
<td>550</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>550</td>
<td>600</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>600</td>
<td>650</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Solution
The reliability estimates can be obtained by expanding the data table to include the calculations used in the actuarial-simple method:

<table>
<thead>
<tr>
<th>Start Time</th>
<th>End Time</th>
<th>Number of Failures, r_i</th>
<th>Number of Suspensions, s_i</th>
<th>Available Units, n_i</th>
<th>1 - ( \frac{r_i}{n_i} )</th>
<th>( \prod 1 - \frac{r_i}{n_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
<td>2</td>
<td>4</td>
<td>55</td>
<td>0.964</td>
<td>0.964</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0</td>
<td>5</td>
<td>49</td>
<td>1.000</td>
<td>0.964</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>2</td>
<td>2</td>
<td>44</td>
<td>0.955</td>
<td>0.920</td>
</tr>
<tr>
<td>150</td>
<td>200</td>
<td>3</td>
<td>5</td>
<td>40</td>
<td>0.925</td>
<td>0.851</td>
</tr>
<tr>
<td>200</td>
<td>250</td>
<td>2</td>
<td>1</td>
<td>32</td>
<td>0.938</td>
<td>0.798</td>
</tr>
<tr>
<td>250</td>
<td>300</td>
<td>1</td>
<td>2</td>
<td>29</td>
<td>0.966</td>
<td>0.770</td>
</tr>
<tr>
<td>300</td>
<td>350</td>
<td>2</td>
<td>1</td>
<td>26</td>
<td>0.923</td>
<td>0.711</td>
</tr>
<tr>
<td>350</td>
<td>400</td>
<td>3</td>
<td>3</td>
<td>23</td>
<td>0.870</td>
<td>0.618</td>
</tr>
<tr>
<td>400</td>
<td>450</td>
<td>3</td>
<td>4</td>
<td>17</td>
<td>0.824</td>
<td>0.509</td>
</tr>
<tr>
<td>450</td>
<td>500</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>0.900</td>
<td>0.458</td>
</tr>
<tr>
<td>500</td>
<td>550</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>0.714</td>
<td>0.327</td>
</tr>
<tr>
<td>550</td>
<td>600</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0.750</td>
<td>0.245</td>
</tr>
<tr>
<td>600</td>
<td>650</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0.333</td>
<td>0.082</td>
</tr>
</tbody>
</table>

As can be determined from the preceding table, the reliability estimates for the failure times are:

<table>
<thead>
<tr>
<th>Failure Period</th>
<th>Reliability Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>96.4%</td>
</tr>
<tr>
<td>150</td>
<td>92.0%</td>
</tr>
<tr>
<td>200</td>
<td>85.1%</td>
</tr>
<tr>
<td>250</td>
<td>79.8%</td>
</tr>
<tr>
<td>300</td>
<td>77.0%</td>
</tr>
<tr>
<td>350</td>
<td>71.1%</td>
</tr>
<tr>
<td>400</td>
<td>61.8%</td>
</tr>
<tr>
<td>450</td>
<td>50.9%</td>
</tr>
<tr>
<td>500</td>
<td>45.8%</td>
</tr>
<tr>
<td>550</td>
<td>32.7%</td>
</tr>
<tr>
<td>600</td>
<td>24.5%</td>
</tr>
<tr>
<td>650</td>
<td>8.2%</td>
</tr>
</tbody>
</table>

**Actuarial-Standard Method**

The actuarial-standard model is a variation of the actuarial-simple model. In the actuarial-simple method, the suspensions in a time period or interval are assumed to occur at the end of that interval, after the failures have occurred. The actuarial-standard model assumes that the suspensions occur in the middle of the interval, which has the effect of reducing the number of available units in the interval by half of the suspensions in that interval or:

\[
n_i' = n_i - \frac{s_i}{2}
\]

With this adjustment, the calculations are carried out just as they were for the actuarial-simple model or:

\[
\hat{R}(t_i) = \prod_{j=1}^{i} \left( 1 - \frac{r_j}{n_j'} \right), \quad i = 1, \ldots, m
\]
Actuarial-Standard Example

Use the data set from the Actuarial-Simple example and analyze it using the actuarial-standard method.

Solution

The solution to this example is similar to that in the Actuarial-Simple example, with the exception of the inclusion of the $n'_i$ term, which is used in the equation for the actuarial-standard method. Applying this equation to the data, we can generate the following table:

<table>
<thead>
<tr>
<th>Start Time</th>
<th>End Time</th>
<th>Number of Failures, $r_i$</th>
<th>Number of Suspensions, $s_i$</th>
<th>Adjusted Units, $n'_i$</th>
<th>$1 - \frac{r_i}{n'_i}$</th>
<th>$\prod 1 - \frac{r_j}{n'_j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
<td>2</td>
<td>4</td>
<td>53</td>
<td>0.962</td>
<td>0.962</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0</td>
<td>5</td>
<td>46.5</td>
<td>1.000</td>
<td>0.962</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>2</td>
<td>2</td>
<td>43</td>
<td>0.953</td>
<td>0.918</td>
</tr>
<tr>
<td>150</td>
<td>200</td>
<td>3</td>
<td>5</td>
<td>37.5</td>
<td>0.920</td>
<td>0.844</td>
</tr>
<tr>
<td>200</td>
<td>250</td>
<td>2</td>
<td>1</td>
<td>31.5</td>
<td>0.937</td>
<td>0.791</td>
</tr>
<tr>
<td>250</td>
<td>300</td>
<td>1</td>
<td>2</td>
<td>28</td>
<td>0.964</td>
<td>0.762</td>
</tr>
<tr>
<td>300</td>
<td>350</td>
<td>2</td>
<td>1</td>
<td>25.5</td>
<td>0.922</td>
<td>0.702</td>
</tr>
<tr>
<td>350</td>
<td>400</td>
<td>3</td>
<td>3</td>
<td>21.5</td>
<td>0.860</td>
<td>0.604</td>
</tr>
<tr>
<td>400</td>
<td>450</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>0.800</td>
<td>0.484</td>
</tr>
<tr>
<td>450</td>
<td>500</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>0.889</td>
<td>0.430</td>
</tr>
<tr>
<td>500</td>
<td>550</td>
<td>2</td>
<td>1</td>
<td>6.5</td>
<td>0.692</td>
<td>0.298</td>
</tr>
<tr>
<td>550</td>
<td>600</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0.750</td>
<td>0.223</td>
</tr>
<tr>
<td>600</td>
<td>650</td>
<td>2</td>
<td>1</td>
<td>2.5</td>
<td>0.200</td>
<td>0.045</td>
</tr>
</tbody>
</table>

As can be determined from the preceding table, the reliability estimates for the failure times are:

<table>
<thead>
<tr>
<th>Failure Period</th>
<th>Reliability Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>96.2%</td>
</tr>
<tr>
<td>150</td>
<td>91.8%</td>
</tr>
<tr>
<td>200</td>
<td>84.4%</td>
</tr>
<tr>
<td>250</td>
<td>79.1%</td>
</tr>
<tr>
<td>300</td>
<td>76.2%</td>
</tr>
<tr>
<td>350</td>
<td>70.2%</td>
</tr>
<tr>
<td>400</td>
<td>60.4%</td>
</tr>
<tr>
<td>450</td>
<td>48.4%</td>
</tr>
<tr>
<td>500</td>
<td>43.0%</td>
</tr>
<tr>
<td>550</td>
<td>29.8%</td>
</tr>
<tr>
<td>600</td>
<td>22.3%</td>
</tr>
<tr>
<td>650</td>
<td>4.5%</td>
</tr>
</tbody>
</table>

Non-Parametric Confidence Bounds

Confidence bounds for non-parametric reliability estimates can be calculated using a method similar to that of parametric confidence bounds. The difficulty in dealing with nonparametric data lies in the estimation of the variance. To estimate the variance for non-parametric data, Weibull++ uses Greenwood’s formula [27]:

$$\overline{Var}(\hat{R}(t_i)) = \left[ \hat{R}(t_i) \right]^2 \cdot \sum_{j=1}^{i} \frac{r_j}{n_j} \cdot \left(1 - \frac{r_j}{n_j} \right)$$

where:

$n = \text{the total number of intervals}$

$m = \text{the total number of units}$
The variable $n_i$ is defined by:

$$n_i = n - \sum_{j=0}^{i-1} s_j - \sum_{j=0}^{i-1} r_j, \quad i = 1, \ldots, m$$

where:

- $r_j =$ the number of failures in interval $j$
- $s_j =$ the number of suspensions in interval $j$

Once the variance has been calculated, the standard error can be determined by taking the square root of the variance:

$$\hat{se}_R = \sqrt{\text{Var}(\hat{R}(t))}$$

This information can then be applied to determine the confidence bounds:

$$\left[ \text{LCB}_R, \text{UCB}_R \right] = \left[ \frac{\hat{R}}{\hat{R} + (1 - \hat{R}) \cdot w'}, \frac{\hat{R}}{\hat{R} + (1 - \hat{R}) / w} \right]$$

where:

$$w' = \frac{z_{\alpha/2} \cdot \hat{se}_R}{\hat{R} (1 - \hat{R})}$$

and $\alpha$ is the desired confidence level for the 1-sided confidence bounds.

**Confidence Bounds Example**

Determine the 1-sided confidence bounds for the reliability estimates in the Actuarial-Simple example, with a 95% confidence level.

**Solution**

Once again, this type of problem is most readily solved by constructing a table similar to the following:

<table>
<thead>
<tr>
<th>Failure Time $t_i$</th>
<th>Reliability Est. $\hat{R}(t_i)$</th>
<th>Number of Failures $r_i$</th>
<th>Adjusted Units $n'_i$</th>
<th>$r_i / n'_i$</th>
<th>Variance $\text{Var}(r_i)$</th>
<th>Error $\hat{se}_R$</th>
<th>$w'$</th>
<th>Lower Cl</th>
<th>Upper Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.962</td>
<td>2</td>
<td>53.0</td>
<td>0.0377</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.007</td>
<td>0.861</td>
<td>0.991</td>
</tr>
<tr>
<td>150</td>
<td>0.918</td>
<td>2</td>
<td>43.0</td>
<td>0.0455</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.795</td>
<td>0.869</td>
</tr>
<tr>
<td>200</td>
<td>0.844</td>
<td>3</td>
<td>37.5</td>
<td>0.0800</td>
<td>0.0030</td>
<td>0.0030</td>
<td>0.0030</td>
<td>0.706</td>
<td>0.824</td>
</tr>
<tr>
<td>250</td>
<td>0.791</td>
<td>2</td>
<td>31.5</td>
<td>0.0635</td>
<td>0.0040</td>
<td>0.0040</td>
<td>0.0040</td>
<td>0.642</td>
<td>0.888</td>
</tr>
<tr>
<td>300</td>
<td>0.762</td>
<td>1</td>
<td>28.0</td>
<td>0.0557</td>
<td>0.0045</td>
<td>0.0045</td>
<td>0.0045</td>
<td>0.605</td>
<td>0.868</td>
</tr>
<tr>
<td>350</td>
<td>0.702</td>
<td>2</td>
<td>25.5</td>
<td>0.0784</td>
<td>0.0054</td>
<td>0.0054</td>
<td>0.0054</td>
<td>0.542</td>
<td>0.825</td>
</tr>
<tr>
<td>400</td>
<td>0.604</td>
<td>3</td>
<td>21.5</td>
<td>0.1395</td>
<td>0.0068</td>
<td>0.0068</td>
<td>0.0068</td>
<td>0.438</td>
<td>0.750</td>
</tr>
<tr>
<td>450</td>
<td>0.484</td>
<td>3</td>
<td>15.0</td>
<td>0.2000</td>
<td>0.0082</td>
<td>0.0082</td>
<td>0.0082</td>
<td>0.315</td>
<td>0.656</td>
</tr>
<tr>
<td>500</td>
<td>0.430</td>
<td>1</td>
<td>9.0</td>
<td>0.1111</td>
<td>0.0091</td>
<td>0.0091</td>
<td>0.0091</td>
<td>0.260</td>
<td>0.618</td>
</tr>
<tr>
<td>550</td>
<td>0.298</td>
<td>2</td>
<td>6.5</td>
<td>0.3077</td>
<td>0.0104</td>
<td>0.0104</td>
<td>0.0104</td>
<td>0.140</td>
<td>0.524</td>
</tr>
<tr>
<td>600</td>
<td>0.223</td>
<td>1</td>
<td>4.0</td>
<td>0.2500</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.085</td>
<td>0.471</td>
</tr>
<tr>
<td>650</td>
<td>0.045</td>
<td>2</td>
<td>2.5</td>
<td>0.8000</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.003</td>
<td>0.423</td>
</tr>
</tbody>
</table>

The following plot illustrates these results graphically:
Nonparametric Reliability Plot

- **Estimated Reliability**
  - 100%
  - 90%
  - 80%
  - 70%
  - 60%
  - 50%
  - 40%
  - 30%
  - 20%
  - 10%
  - 0%

- **Time**
  - 0
  - 100
  - 200
  - 300
  - 400
  - 500
  - 600
  - 700

- **Graph Legend**
  - Rhat
  - LCL
  - UCL
Chapter 18

Competing Failure Modes Analysis

Often, a group of products will fail due to more than one failure mode. One can take the view that the products could have failed due to any one of the possible failure modes, but since an item cannot fail more than one time, there can only be one failure mode for each failed product. In this view, the failure modes compete as to which causes the failure for each particular item. This can be viewed as a series system reliability model, with each failure mode composing a block of the series system. Competing failure modes (CFM) analysis segregates the analyses of failure modes and then combines the results to provide an overall model for the product in question.

CFM Analysis Approach

In order to begin analyzing data sets with more than one competing failure mode, one must perform a separate analysis for each failure mode. During each of these analyses, the failure times for all other failure modes not being analyzed are considered to be suspensions. This is because the units under test would have failed at some time in the future due to the failure mode being analyzed, had the unrelated (not analyzed) mode not occurred. Thus, in this case, the information available is that the mode under consideration did not occur and the unit under consideration accumulated test time without a failure due to the mode under consideration (or a suspension due to that mode).

Once the analysis for each separate failure mode has been completed (using the same principles as before), the resulting reliability equation for all modes is the product of the reliability equation for each mode, or:

\[ R(t) = R_1(t) \cdot R_2(t) \cdot \ldots \cdot R_n(t) \]

where \( n \) is the total number of failure modes considered. This is the product rule for the reliability of series systems with statistically independent components, which states that the reliability for a series system is equal to the product of the reliability values of the components comprising the system. Do note that the above equation is the reliability function based on any assumed life distribution. In Weibull++ this life distribution can be either the 2-parameter Weibull, lognormal, normal or the 1-parameter exponential.

CFM Example

The following example demonstrates how you can use the reliability equation to determine the overall reliability of a component. (This example has been abstracted from Example 15.6 from the Meeker and Escobar textbook Statistical Methods for Reliability Data [27].)

An electronic component has two competing failure modes. One failure mode is due to random voltage spikes, which cause failure by overloading the system. The other failure mode is due to wearout failures, which usually happen only after the system has run for many cycles. The objective is to determine the overall reliability for the component at 100,000 cycles.

30 units are tested, and the failure times are recorded in the following table. The failures that are due to the random voltage spikes are denoted by a V. The failures that are due to wearout failures are denoted by a W.
Solution

To obtain the overall reliability of the component, we will first need to analyze the data set due to each failure mode. For example, to obtain the reliability of the component due to voltage spikes, we must consider all of the failures for the wear-out mode to be suspensions. We do the same for analyzing the wear-out failure mode, counting only the wear-out data as failures and assuming that the voltage spike failures are suspensions. Once we have obtained the reliability of the component due to each mode, we can use the system Reliability Equation to determine the overall component reliability.

The following analysis shows the data set for the voltage spikes. Using the Weibull distribution and the MLE analysis method (recommended due to the number of suspensions in the data), the parameters are $\beta_V = 0.671072$ and $\eta_V = 449.427230$. The reliability for this failure mode at $t = 100\,\text{ks}$, $R_V(100) = 0.694357$. 

<table>
<thead>
<tr>
<th>Number in State</th>
<th>State F or S</th>
<th>State End</th>
<th>State Time (hr)</th>
<th>Subset ID 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>2</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>10</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>13</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>23</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>28</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>30</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>65</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>80</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>88</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>106</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>143</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>S</td>
<td>147</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>173</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>S</td>
<td>181</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>S</td>
<td>212</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>S</td>
<td>245</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>247</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>261</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>S</td>
<td>266</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>S</td>
<td>275</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>S</td>
<td>293</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>S</td>
<td>300</td>
<td>W</td>
<td></td>
</tr>
</tbody>
</table>

*Failure times given are in thousands of cycles.*
The following analysis shows the data set for the wearout failure mode. Using the same analysis settings (i.e., Weibull distribution and MLE analysis method), the parameters are $\hat{\beta}_W = 4.33727$ and $\eta_W = 340.384242$. The reliability for this failure mode at $t = 100$ is $R_W(100) = 0.995084$.

Using the Reliability Equation to obtain the overall component reliability at 100,000 cycles, we get:

$$R_{sys}(100) = R_V(100) \cdot R_W(100)$$

$$= 0.694357 \cdot 0.995084$$

$$= 0.690943$$

Or the reliability of the unit (or system) under both modes is $R_{sys}(100) = 69.094\%$.

You can also perform this analysis using Weibull++'s built-in CFM analysis options, which allow you to generate a probability plot that contains the combined mode line as well as the individual mode lines.

See It In action...
Confidence Bounds for CFM Analysis

The method available in Weibull++ for estimating the different types of confidence bounds, for competing failure modes analysis, is the Fisher matrix method, and is presented in this section.

Variance/Covariance Matrix

The variances and covariances of the parameters are estimated from the inverse local Fisher matrix, as follows:

\[
\begin{pmatrix}
\text{Var}(\hat{a}_1) & \text{Cov}(\hat{a}_1, \hat{b}_1) & 0 & 0 & 0 & 0 & 0 \\
\text{Cov}(\hat{a}_1, \hat{b}_1) & \text{Var}(\hat{b}_1) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

\[
= \left( \begin{array}{cccccc}
-\frac{\partial^2 \Lambda}{\partial a_1 \partial a_1} & -\frac{\partial^2 \Lambda}{\partial a_1 \partial b_1} & 0 & 0 & 0 & 0 \\
-\frac{\partial^2 \Lambda}{\partial b_1 \partial a_1} & -\frac{\partial^2 \Lambda}{\partial b_1 \partial b_1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{array} \right)
\]

where \( \Lambda \) is the log-likelihood function of the failure distribution, described in Parameter Estimation.

Bounds on Reliability

The competing failure modes reliability function is given by:

\[
\hat{R} = \prod_{i=1}^{n} \hat{R}_i
\]

where:

- \( \hat{R}_i \) is the reliability of the \( i \)th mode.
- \( n \) is the number of failure modes.

The upper and lower bounds on reliability are estimated using the logit transformation:

\[
\hat{R}_U = \frac{\hat{R}}{\hat{R} + (1 - \hat{R})e^{-\frac{K_\alpha \sqrt{\text{Var}(\hat{R})}}{\hat{R}(1 - \hat{R})}}}
\]

\[
\hat{R}_L = \frac{\hat{R}}{\hat{R} + (1 - \hat{R})e^{-\frac{K_\alpha \sqrt{\text{Var}(\hat{R})}}{\hat{R}(1 - \hat{R})}}}
\]

where \( \hat{R} \) is calculated using the reliability equation for competing failure modes. \( K_\alpha \) is defined by:
\[
\alpha = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} \, dt = 1 - \Phi(K_\alpha)
\]

(If is the confidence level, then \(\alpha = \frac{1-\delta}{2}\) for the two-sided bounds, and \(\alpha = 1 - \delta\) for the one-sided bounds.)

The variance of \(R\) is estimated by:

\[
\text{Var}(\hat{R}) = \sum_{i=1}^{n} \left( \frac{\partial R}{\partial R_i} \right)^2 \text{Var}(\hat{R}_i)
\]

\[
\frac{\partial R}{\partial R_i} = \prod_{j=1, j \neq i}^{n} \hat{R}_j
\]

Thus:

\[
\text{Var}(\hat{R}) = \sum_{i=1}^{n} \left( \prod_{j=1, j \neq i}^{n} \hat{R}_j^2 \right) \text{Var}(\hat{R}_i)
\]

\[
\text{Var}(\hat{\alpha}_i) = \sum_{i=1}^{n} \left( \frac{\partial R_i}{\partial \alpha_i} \right)^2 \text{Var}(\hat{\alpha}_i)
\]

where \(\hat{\alpha}_i\) is an element of the model parameter vector.

Therefore, the value of \(\text{Var}(\hat{R}_i)\) is dependent on the underlying distribution.

For the Weibull distribution:

\[
\text{Var}(\hat{R}_i) = \left( \hat{R}_i e^{\hat{\mu}_i} \right)^2 \text{Var}(\hat{\mu}_i)
\]

where:

\[
\hat{\mu}_i = \hat{\beta}_i (\ln(t - \hat{\gamma}_i) - \ln \hat{\eta}_i)
\]

and \(\text{Var}(\hat{\mu}_i)\) is given in The Weibull Distribution.

For the exponential distribution:

\[
\text{Var}(\hat{\lambda}_i) = \left( \hat{R}_i (t - \gamma_i) \right)^2 \text{Var}(\hat{\lambda}_i)
\]

where \(\text{Var}(\hat{\lambda}_i)\) is given in The Exponential Distribution.

For the normal distribution:

\[
\text{Var}(\hat{z}_i) = (f(\hat{z}_i)) \text{Var}(\hat{z}_i)
\]

\[
\hat{z}_i = \frac{t - \hat{\mu}_i}{\hat{\sigma}_i}
\]

where \(\text{Var}(\hat{z}_i)\) is given in The Normal Distribution.

For the lognormal distribution:

\[
\text{Var}(\hat{z}_i) = (f(\hat{z}_i) \cdot \hat{\sigma}_i) \text{Var}(\hat{z}_i)
\]

\[
\hat{z}_i = \frac{\ln(i) - \hat{\mu}_i}{\hat{\sigma}_i}
\]

where \(\text{Var}(\hat{z}_i)\) is given in The Lognormal Distribution.
**Bounds on Time**

The bounds on time are estimated by solving the reliability equation with respect to time. From the reliability equation for competing failure modes, we have that:

\[ t = \varphi(R, \hat{a}_i, \hat{b}_i) \]

where:

- \( \varphi \) is the inverse function for the reliability equation for competing failure modes.
- For the Weibull distribution, \( \hat{a}_i \) is \( \hat{\beta}_i \) and \( \hat{b}_i \) is \( \hat{\eta}_i \).
- For the exponential distribution, \( \hat{a}_i \) is \( \hat{\lambda}_i \) and \( \hat{b}_i = 0 \).
- For the normal distribution, \( \hat{a}_i \) is \( \hat{\mu}_i \) and \( \hat{b}_i \) is \( \hat{\sigma}_i \).
- For the lognormal distribution, \( \hat{a}_i \) is \( \hat{\mu}'_i \) and \( \hat{b}_i \) is \( \hat{\sigma}'_i \).

Set:

\[ u = \ln(t) \]

The bounds on \( u \) are estimated from:

\[ u_U = \hat{u} + K \alpha \sqrt{\text{Var}(\hat{u})} \]

and:

\[ u_L = \hat{u} - K \alpha \sqrt{\text{Var}(\hat{u})} \]

Then the upper and lower bounds on time are found by using the equations:

\[ t_U = e^{u_U} \]

and:

\[ t_L = e^{u_L} \]

\( K \alpha \) is calculated using the inverse standard normal distribution and \( \text{Var}(\hat{u}) \) is computed as:

\[ \text{Var}(\hat{u}) = \sum_{i=1}^{n} \left( \frac{\partial u}{\partial a_i} \right)^2 \text{Var}(\hat{a}_i) + \left( \frac{\partial u}{\partial b_i} \right)^2 \text{Var}(\hat{b}_i) + 2 \frac{\partial u}{\partial a_i} \frac{\partial u}{\partial b_i} \text{Cov}(\hat{a}_i, \hat{b}_i) \]

**Complex Failure Modes Analysis**

In addition to being viewed as a series system, the relationship between the different competing failure modes can be more complex. After performing separate analysis for each failure mode, a diagram that describes how each failure mode can result in a product failure can be used to perform analysis for the item in question. Such diagrams are usually referred to as Reliability Block Diagrams (RBD) (for more on RBDs see ReliaSoft's System Analysis Reference and ReliaSoft's BlockSim software[^3]).

A reliability block diagram is made of blocks that represent the failure modes and arrows and connects the blocks in different configurations. Note that the blocks can also be used to represent different components or subsystems that make up the product. Weibull ++ provides the capability to use a diagram to model, series, parallel, k-out-of-n configurations in addition to any complex combinations of these configurations.

In this analysis, the failure modes are assumed to be statistically independent. (Note: In the context of this reference, statistically independent implies that failure information for one failure mode provides no information about, i.e. does not affect, other failure mode). Analysis of dependent modes is more complex. Advanced RBD software such as ReliaSoft's BlockSim can handle and analyze such dependencies, as well as provide more advanced constructs and analyses (see http://www.ReliaSoft.com/BlockSim).
Failure Modes Configurations

Series Configuration
The basic competing failure modes configuration, which has already been discussed, is a series configuration. In a series configuration, the occurrence of any failure mode results in failure for the product.

The equation that describes series configuration is:

\[ R(t) = R_1(t) \cdot R_2(t) \cdot \ldots \cdot R_n(t) \]

where \( n \) is the total number of failure modes considered.

Parallel
In a simple parallel configuration, at least one of the failure modes must not occur for the product to continue operation.

The equation that describes the parallel configuration is:

\[ R(t) = 1 - \prod_{i=1}^{n} (1 - R_i(t)) \]

where \( n \) is the total number of failure modes considered.

Combination of Series and Parallel
While many smaller products can be accurately represented by either a simple series or parallel configuration, there may be larger products that involve both series and parallel configurations in the overall model of the product. Such products can be analyzed by calculating the reliabilities for the individual series and parallel sections and then combining them in the appropriate manner.

\[ k\text{-out-of-}n \text{ Parallel Configuration=} \]
The k-out-of-n configuration is a special case of parallel redundancy. This type of configuration requires that at least \( k \) failure modes do not happen out of the total \( n \) parallel failure modes for the product to succeed. The simplest case of a k-out-of-n configuration is when the failure modes are independent and identical and have the same failure distribution and uncertainties about the parameters (in other words they are derived from the same test data). In this case, the reliability of the product with such a configuration can be evaluated using the binomial distribution, or:

\[
R(t) = \sum_{r=k}^{n} \binom{n}{k} R^r(t)(1 - R(t))^{n-r}
\]

In the case where the k-out-of-n failure modes are not identical, other approaches for calculating the reliability must be used (e.g. the event space method). Discussion of these is beyond the scope of this reference. Interested readers can consult the System Analysis Reference book.

**Complex Systems**

In many cases, it is not easy to recognize which components are in series and which are in parallel in a complex system.

The previous configuration cannot be broken down into a group of series and parallel configurations. This is primarily due to the fact that failure mode C has two paths leading away from it, whereas B and D have only one. Several methods exist for obtaining the reliability of a complex configuration including the decomposition method, the event space method and the path-tracing method. Discussion of these is beyond the scope of this reference. Interested readers can consult the System Analysis Reference book.

**Complex Failure Modes Example**

Assume that a product has five independent failure modes: A, B, C, D and E. Furthermore, assume that failure of the product will occur if mode A occurs, modes B and C occur simultaneously or if modes D and E occur simultaneously. The objective is to estimate the reliability of the product at 100 hours, with 90% two-sided confidence bounds.

The product is tested to failure, and the failure times due to each mode are recorded in the following table.

<table>
<thead>
<tr>
<th>TTF for A</th>
<th>TTF for B</th>
<th>TTF for C</th>
<th>TTF for D</th>
<th>TTF for E</th>
</tr>
</thead>
<tbody>
<tr>
<td>276</td>
<td>23</td>
<td>499</td>
<td>467</td>
<td>67</td>
</tr>
<tr>
<td>320</td>
<td>36</td>
<td>545</td>
<td>540</td>
<td>72</td>
</tr>
<tr>
<td>323</td>
<td>57</td>
<td>661</td>
<td>716</td>
<td>81</td>
</tr>
<tr>
<td>558</td>
<td>89</td>
<td>738</td>
<td>737</td>
<td>108</td>
</tr>
<tr>
<td>674</td>
<td>99</td>
<td>987</td>
<td>761</td>
<td>110</td>
</tr>
<tr>
<td>829</td>
<td>154</td>
<td>1165</td>
<td>1093</td>
<td>127</td>
</tr>
<tr>
<td>878</td>
<td>200</td>
<td>1337</td>
<td>1283</td>
<td>148</td>
</tr>
</tbody>
</table>

**Solution**

The reliability block diagram (RBD) approach can be used to analyze the reliability of the product. But before creating a diagram, the data sets of the failure modes need to be segregated so that each mode can be represented by a single block in the diagram. Recall that when you analyze a particular mode, the failure times for all other competing modes are considered to be suspensions. This captures the fact that those units operated for a period of
time without experiencing the failure mode of interest before they were removed from observation when they failed due to another mode. We can easily perform this step via Weibull++’s Batch Auto Run utility. To do this, enter the data from the table into a single data sheet. Choose the 2P-Weibull distribution and the MLE analysis method, and then click the Batch Auto Run icon on the control panel. When prompted to select the subset IDs, select them all. Click the Processing Preferences tab. In the Extraction Options area, select the second option, as shown next.

This will extract the data sets that are required for the analysis. Select the check box in the Calculation Options area and click OK. The data sets are extracted into separate data sheets in the folio and automatically calculated.

Next, create a diagram by choosing Insert > Tools > Diagram. Add blocks by right-clicking the diagram and choosing Add Block on the shortcut menu. When prompted to select the data sheet of the failure mode that the block will represent, select the data sheet for mode A. Use the same approach to add the blocks that will represent failure modes B, C, D and E. Add a connector by right-clicking the diagram sheet and choosing Connect Blocks, and then connect the blocks in an appropriate configuration to describe the relationships between the failure modes. To insert a node, which acts as a switch that the diagram paths move through, right-click the diagram and choose Add Node. Specify the number of required paths in the node by double-clicking the node and entering the appropriate number (use 2 in both nodes).

The following figure shows the completed diagram.
Click **Analyze** to analyze the diagram, and then use the Quick Calculation Pad (QCP) to estimate the reliability. The estimated $R(100)$ and the 90% two-sided confidence bounds are:

\[
\hat{R}_{100} = 0.895940 \\
\hat{R}(100) = 0.824397 \\
\hat{R}_{L}(100) = 0.719090
\]

**References**

Warranty Data Analysis

The Weibull++ warranty analysis folio provides four different data entry formats for warranty claims data. It allows the user to automatically perform life data analysis, predict future failures (through the use of conditional probability analysis), and provides a method for detecting outliers. The four data-entry formats for storing sales and returns information are:

1) Nevada Chart Format
2) Time-to-Failure Format
3) Dates of Failure Format
4) Usage Format

These formats are explained in the next sections. We will also discuss some specific warranty analysis calculations, including warranty predictions, analysis of non-homogeneous warranty data and using statistical process control (SPC) to monitor warranty returns.

Nevada Chart Format

The Nevada format allows the user to convert shipping and warranty return data into the standard reliability data form of failures and suspensions so that it can easily be analyzed with traditional life data analysis methods. For each time period in which a number of products are shipped, there will be a certain number of returns or failures in subsequent time periods, while the rest of the population that was shipped will continue to operate in the following time periods. For example, if 500 units are shipped in May, and 10 of those units are warranty returns in June, that is equivalent to 10 failures at a time of one month. The other 490 units will go on to operate and possibly fail in the months that follow. This information can be arranged in a diagonal chart, as shown in the following figure.
At the end of the analysis period, all of the units that were shipped and have not failed in the time since shipment are considered to be suspensions. This process is repeated for each shipment and the results tabulated for each particular failure and suspension time prior to reliability analysis. This process may sound confusing, but it is actually just a matter of careful bookkeeping. The following example illustrates this process.

**Example**

**Nevada Chart Format Calculations Example**

A company keeps track of its shipments and warranty returns on a month-by-month basis. The following table records the shipments in June, July and August, and the warranty returns through September:

<table>
<thead>
<tr>
<th>Shipped</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1623</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>12</td>
<td>15</td>
<td>17</td>
<td>18</td>
<td>20</td>
<td>21</td>
<td>23</td>
<td>25</td>
<td>27</td>
<td>29</td>
<td>29</td>
<td>30</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>2 3723</td>
<td>2</td>
<td>7</td>
<td>11</td>
<td>17</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>33</td>
<td>38</td>
<td>42</td>
<td>46</td>
<td>49</td>
<td>53</td>
<td>56</td>
<td>59</td>
<td>62</td>
<td>65</td>
<td>67</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>3 1319</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>14</td>
<td>17</td>
<td>17</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>4 3600</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>28</td>
<td>33</td>
<td>36</td>
<td>41</td>
<td>44</td>
<td>47</td>
<td>51</td>
<td>55</td>
<td>57</td>
<td>60</td>
<td>62</td>
<td>60</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>5 3298</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>14</td>
<td>19</td>
<td>22</td>
<td>26</td>
<td>30</td>
<td>33</td>
<td>37</td>
<td>41</td>
<td>43</td>
<td>47</td>
<td>50</td>
<td>52</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 1333</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>15</td>
<td>17</td>
<td>17</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 1584</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>12</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 4508</td>
<td>2</td>
<td>9</td>
<td>14</td>
<td>19</td>
<td>26</td>
<td>31</td>
<td>35</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>55</td>
<td>59</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 4463</td>
<td>2</td>
<td>9</td>
<td>14</td>
<td>19</td>
<td>26</td>
<td>31</td>
<td>35</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>55</td>
<td>59</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 3176</td>
<td>1</td>
<td>6</td>
<td>10</td>
<td>14</td>
<td>18</td>
<td>21</td>
<td>25</td>
<td>29</td>
<td>32</td>
<td>36</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 1047</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 4151</td>
<td>2</td>
<td>8</td>
<td>13</td>
<td>18</td>
<td>23</td>
<td>28</td>
<td>33</td>
<td>37</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 4436</td>
<td>2</td>
<td>9</td>
<td>13</td>
<td>20</td>
<td>24</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 3549</td>
<td>2</td>
<td>6</td>
<td>11</td>
<td>16</td>
<td>20</td>
<td>23</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 4058</td>
<td>2</td>
<td>8</td>
<td>12</td>
<td>18</td>
<td>22</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 3731</td>
<td>2</td>
<td>7</td>
<td>11</td>
<td>17</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 4864</td>
<td>3</td>
<td>9</td>
<td>15</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 1757</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 3953</td>
<td>2</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 1483</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

At the end of the analysis period, all of the units that were shipped and have not failed in the time since shipment are considered to be suspensions. This process is repeated for each shipment and the results tabulated for each particular failure and suspension time prior to reliability analysis. This process may sound confusing, but it is actually just a matter of careful bookkeeping. The following example illustrates this process.

**Example**

**Nevada Chart Format Calculations Example**

A company keeps track of its shipments and warranty returns on a month-by-month basis. The following table records the shipments in June, July and August, and the warranty returns through September:

<table>
<thead>
<tr>
<th>Ship</th>
<th>RETURNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun. 2010</td>
<td>100</td>
</tr>
<tr>
<td>Jul. 2010</td>
<td>140</td>
</tr>
<tr>
<td>Aug. 2010</td>
<td>150</td>
</tr>
</tbody>
</table>

We will examine the data month by month. In June 100 units were sold, and in July 3 of these units were returned. This gives 3 failures at one month for the June shipment, which we will denote as $F_{JUN,1} = 3$. Likewise, 3 failures occurred in August and 5 occurred in September for this shipment, or $F_{JUN,2} = 3$ and $F_{JUN,3} = 5$. Consequently, at the end of our three-month analysis period, there were a total of 11 failures for the 100 units shipped in June. This means that 89 units are presumably still operating, and can be considered suspensions at three months, or $S_{JUN,3} = 89$. For the shipment of 140 in July, 2 were returned the following month, or $F_{JUL,1} = 2$, and 4 more were returned the month after that, or $F_{JUL,2} = 4$. After two months, there are 134 ($140 - 2 - 4 = 134$) units from the July shipment still operating, or $S_{JUL,2} = 134$. For the final shipment of 150 in August, 4 fail in September, or $F_{AUG,1} = 4$, with the remaining 146 units being suspensions at one month, or $S_{AUG,1} = 146$. 
It is now a simple matter to add up the number of failures for 1, 2, and 3 months, then add the suspensions to get our reliability data set:

- **Failures at 1 month:** \( F_1 = F_{JUN,1} + F_{JUL,1} + F_{AUG,1} = 3 + 2 + 4 = 9 \)
- **Suspensions at 1 month:** \( S_1 = S_{AUG,1} = 146 \)
- **Failures at 2 months:** \( F_2 = F_{JUN,2} + F_{JUL,2} = 3 + 4 = 7 \)
- **Suspensions at 2 months:** \( S_2 = S_{JUL,2} = 134 \)
- **Failures at 3 months:** \( F_3 = F_{JUN,3} = 5 \)
- **Suspensions at 3 months:** \( S_{JUN,3} = 89 \)

These calculations can be performed automatically in Weibull++.

---

**Time-to-Failure Format**

This format is similar to the standard folio data entry format (all number of units, failure times and suspension times are entered by the user). The difference is that when the data is used within the context of warranty analysis, the ability to generate forecasts is available to the user.

**Example**

**Times-to-Failure Format Warranty Analysis**

Assume that we have the following information for a given product.

### Times-to-Failure Data

<table>
<thead>
<tr>
<th>Number in State</th>
<th>State F or S</th>
<th>State End Time (Hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>F</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>125</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>175</td>
</tr>
<tr>
<td>1500</td>
<td>S</td>
<td>200</td>
</tr>
</tbody>
</table>

### Future Sales

<table>
<thead>
<tr>
<th>Quantity In-Service</th>
<th>Time (Hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>200</td>
</tr>
<tr>
<td>400</td>
<td>300</td>
</tr>
<tr>
<td>100</td>
<td>500</td>
</tr>
</tbody>
</table>

Use the time-to-failure warranty analysis folio to analyze the data and generate a forecast for future returns.

**Solution**

Create a warranty analysis folio and select the times-to-failure format. Enter the data from the tables in the **Data** and **Future Sales** sheets, and then analyze the data using the 2P-Weibull distribution and RRX analysis method. The parameters are estimated to be \( \beta = 3.199832 \) and \( \eta = 814.293442 \).
Click the **Forecast** icon on the control panel. In the Forecast Setup window, set the forecast to start on the 100th hour and set the number of forecast periods to 5. Set the increment (length of each period) to 100, as shown next.

Click **OK**. A Forecast sheet will be created, with the following predicted future returns.

We will use the first row to explain how the forecast for each cell is calculated. For example, there are 1,500 units with a current age of 200 hours. The probability of failure in the next 100 hours can be calculated in the QCP, as follows.
Therefore, the predicted number of failures for the first 100 hours is:

\[ 1500 \times 0.02932968 = 43.99452 \]

This is identical to the result given in the Forecast sheet (shown in the 3rd cell in the first row) of the analysis. The bounds and the values in other cells can be calculated similarly.

All the plots that are available for the standard folio are also available in the warranty analysis, such as the Probability plot, Reliability plot, etc. One additional plot in warranty analysis is the Expected Failures plot, which shows the expected number of failures over time. The following figure shows the Expected Failures plot of the example, with confidence bounds.
Dates of Failure Format

Another common way for reporting field information is to enter a date and quantity of sales or shipments (Quantity In-Service data) and the date and quantity of returns (Quantity Returned data). In order to identify which lot the unit comes from, a failure is identified by a return date and the date of when it was put in service. The date that the unit went into service is then associated with the lot going into service during that time period. You can use the optional Subset ID column in the data sheet to record any information to identify the lots.

Example

Dates of Failure Warranty Analysis

Assume that a company has the following information for a product.

Sales

<table>
<thead>
<tr>
<th>Quantity In-Service</th>
<th>Date In-Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>6316</td>
<td>1/1/2010</td>
</tr>
<tr>
<td>8447</td>
<td>2/1/2010</td>
</tr>
<tr>
<td>5892</td>
<td>3/1/2010</td>
</tr>
<tr>
<td>596</td>
<td>4/1/2010</td>
</tr>
<tr>
<td>996</td>
<td>5/1/2010</td>
</tr>
<tr>
<td>8977</td>
<td>6/1/2010</td>
</tr>
<tr>
<td>2578</td>
<td>7/1/2010</td>
</tr>
<tr>
<td>Quantity Returned</td>
<td>Date of Return</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>2</td>
<td>10/29/2010</td>
</tr>
<tr>
<td>1</td>
<td>11/13/2010</td>
</tr>
<tr>
<td>2</td>
<td>3/15/2011</td>
</tr>
<tr>
<td>5</td>
<td>4/10/2011</td>
</tr>
<tr>
<td>1</td>
<td>11/13/2010</td>
</tr>
<tr>
<td>2</td>
<td>2/19/2011</td>
</tr>
<tr>
<td>2</td>
<td>5/18/2011</td>
</tr>
<tr>
<td>1</td>
<td>1/9/2011</td>
</tr>
<tr>
<td>2</td>
<td>2/13/2011</td>
</tr>
<tr>
<td>1</td>
<td>3/2/2011</td>
</tr>
<tr>
<td>1</td>
<td>6/7/2011</td>
</tr>
<tr>
<td>1</td>
<td>4/28/2011</td>
</tr>
<tr>
<td>3</td>
<td>7/15/2011</td>
</tr>
<tr>
<td>1</td>
<td>8/10/2011</td>
</tr>
<tr>
<td>1</td>
<td>8/12/2011</td>
</tr>
<tr>
<td>1</td>
<td>8/14/2011</td>
</tr>
</tbody>
</table>
Using the given information to estimate the failure distribution of the product and forecast warranty returns.

**Solution**

Create a warranty analysis folio using the dates of failure format. Enter the data from the tables in the Sales, Returns and Future Sales sheets. On the control panel, click the Auto-Set button to automatically set the end date to the last day the warranty data were collected (September 14, 2011). Analyze the data using the 2P-Weibull distribution and RRX analysis method. The parameters are estimated to be beta = 1.315379 and eta = 102,381.486165.

The warranty folio automatically converts the warranty data into a format that can be used in a Weibull++ standard folio. To see this result, click anywhere within the Analysis Summary area of the control panel to open a report, as shown next (showing only the first 35 rows of data). In this example, rows 23 to 60 show the time-to-failure data that resulted from the conversion.

To generate a forecast, click the Forecast icon on the control panel. In the Forecast Setup window, set the forecast to start on September 2011 and set the number of forecast periods to 6. Set the increment (length of each period) to 1 Month, as shown next.
Click **OK**. A Forecast sheet will be created, with the predicted future returns. Note that the first forecast will start on September 15, 2011 because the end of observation period was set to September 14, 2011.

Click the **Plot** icon and choose the **Expected Failures** plot. The plot displays the predicted number of returns for each month, as shown next.
Usage Format

Often, the driving factor for reliability is usage rather than time. For example, in the automotive industry, the failure behavior in the majority of the products is mileage-dependent rather than time-dependent. The usage format allows the user to convert shipping and warranty return data into the standard reliability data for failures and suspensions when the return information is based on usage rather than return dates or periods. Similar to the dates of failure format, a failure is identified by the return number and the date of when it was put in service in order to identify which lot the unit comes from. The date that the returned unit went into service associates the returned unit with the lot it belonged to when it started operation. However, the return data is in terms of usage and not date of return. Therefore the usage of the units needs to be specified as a constant usage per unit time or as a distribution. This allows for determining the expected usage of the surviving units.

Suppose that you have been collecting sales (units in service) and returns data. For the returns data, you can determine the number of failures and their usage by reading the odometer value, for example. Determining the number of surviving units (suspending) and their ages is a straightforward step. By taking the difference between the analysis date and the date when a unit was put in service, you can determine the age of the surviving units.

What is unknown, however, is the exact usage accumulated by each surviving unit. The key part of the usage-based warranty analysis is the determination of the usage of the surviving units based on their age. Therefore, the analyst needs to have an idea about the usage of the product. This can be obtained, for example, from customer surveys or by designing the products to collect usage data. For example, in automotive applications, engineers often use 12,000 miles/year as the average usage. Based on this average, the usage of an item that has been in the field for 6 months and has not yet failed would be 6,000 miles. So to obtain the usage of a suspension based on an average usage, one could take the time of each suspension and multiply it by this average usage. In this situation, the analysis becomes straightforward. With the usage values and the quantities of the returned units, a failure distribution can be constructed and subsequent warranty analysis becomes possible.

Alternatively, and more realistically, instead of using an average usage, an actual distribution that reflects the variation in usage and customer behavior can be used. This distribution describes the usage of a unit over a certain time period (e.g., 1 year, 1 month, etc). This probabilistic model can be used to estimate the usage for all surviving components in service and the percentage of users running the product at different usage rates. In the automotive example, for instance, such a distribution can be used to calculate the percentage of customers that drive 0-200 miles/month, 200-400 miles/month, etc. We can take these percentages and multiply them by the number of suspensions to find the number of items that have been accumulating usage values in these ranges.

To proceed with applying a usage distribution, the usage distribution is divided into increments based on a specified interval width denoted as $Z$. The usage distribution, $Q$, is divided into intervals of $[0 + Z, Z + Z, 2Z + Z, \text{etc.}]$, or $x_i = x_{i-1} + Z$, as shown in the next figure.
The interval width should be selected such that it creates segments that are large enough to contain adequate numbers of suspensions within the intervals. The percentage of suspensions that belong to each usage interval is calculated as follows:

\[ F(x_i) = Q(x_i) - Q(x_i - 1) \]

where:

- \( Q() \) is the usage distribution Cumulative Density Function, \( cdf \).
- \( x_i \) represents the intervals used in apportioning the suspended population.

A suspension group is a collection of suspensions that have the same age. The percentage of suspensions can be translated to numbers of suspensions within each interval, \( x_i \). This is done by taking each group of suspensions and multiplying it by each \( F(x_i) \), or:

\[
N_{1,j} = F(x_1) \times NS_j \\
N_{2,j} = F(x_2) \times NS_j \\
\vdots
\]

\[
N_{n,j} = F(x_n) \times NS_j
\]

where:

- \( N_{n,j} \) is the number of suspensions that belong to each interval.
- \( NS_j \) is the jth group of suspensions from the data set.

This is repeated for all the groups of suspensions.

The age of the suspensions is calculated by subtracting the Date In-Service \( DIS \), which is the date at which the unit started operation, from the end of observation period date or End Date \( ED \). This is the Time In-Service \( TIS \) value that describes the age of the surviving unit.

\[
TIS = ED - DIS
\]

Note: \( TIS \) is in the same time units as the period in which the usage distribution is defined.

For each \( N_{h,j} \), the usage is calculated as:

\[
U_{k,j} = xi \times TIS_j
\]
After this step, the usage of each suspension group is estimated. This data can be combined with the failures data set, and a failure distribution can be fitted to the combined data.

**Example**

**Warranty Analysis Usage Format Example**

Suppose that an automotive manufacturer collects the warranty returns and sales data given in the following tables. Convert this information to life data and analyze it using the lognormal distribution.

**Quality In-Service Data**

<table>
<thead>
<tr>
<th>Quantity In-Service</th>
<th>Date In-Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Dec-09</td>
</tr>
<tr>
<td>13</td>
<td>Jan-10</td>
</tr>
<tr>
<td>15</td>
<td>Feb-10</td>
</tr>
<tr>
<td>20</td>
<td>Mar-10</td>
</tr>
<tr>
<td>15</td>
<td>Apr-10</td>
</tr>
<tr>
<td>25</td>
<td>May-10</td>
</tr>
<tr>
<td>19</td>
<td>Jun-10</td>
</tr>
<tr>
<td>16</td>
<td>Jul-10</td>
</tr>
<tr>
<td>20</td>
<td>Aug-10</td>
</tr>
<tr>
<td>19</td>
<td>Sep-10</td>
</tr>
<tr>
<td>25</td>
<td>Oct-10</td>
</tr>
<tr>
<td>30</td>
<td>Nov-10</td>
</tr>
</tbody>
</table>

**Quality Return Data**

<table>
<thead>
<tr>
<th>Quantity Returned</th>
<th>Usage at Return Date</th>
<th>Return Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9072</td>
<td>Dec-09</td>
</tr>
<tr>
<td>1</td>
<td>9743</td>
<td>Jan-10</td>
</tr>
<tr>
<td>1</td>
<td>6857</td>
<td>Feb-10</td>
</tr>
<tr>
<td>1</td>
<td>7651</td>
<td>Mar-10</td>
</tr>
<tr>
<td>1</td>
<td>5083</td>
<td>May-10</td>
</tr>
<tr>
<td>1</td>
<td>5990</td>
<td>May-10</td>
</tr>
<tr>
<td>1</td>
<td>7432</td>
<td>May-10</td>
</tr>
<tr>
<td>1</td>
<td>8739</td>
<td>May-10</td>
</tr>
<tr>
<td>1</td>
<td>3158</td>
<td>Jun-10</td>
</tr>
<tr>
<td>1</td>
<td>1136</td>
<td>Jul-10</td>
</tr>
<tr>
<td>1</td>
<td>4646</td>
<td>Aug-10</td>
</tr>
<tr>
<td>1</td>
<td>3965</td>
<td>Sep-10</td>
</tr>
<tr>
<td>1</td>
<td>3117</td>
<td>Oct-10</td>
</tr>
<tr>
<td>1</td>
<td>3250</td>
<td>Nov-10</td>
</tr>
</tbody>
</table>
Solution

Create a warranty analysis folio and select the usage format. Enter the data from the tables in the Sales, Returns and Future Sales sheets. The warranty data were collected until 12/1/2010; therefore, on the control panel, set the End of Observation Period to that date. Set the failure distribution to Lognormal, as shown next.

In this example, the manufacturer has been documenting the mileage accumulation per year for this type of product across the customer base in comparable regions for many years. The yearly usage has been determined to follow a lognormal distribution with \( \mu = 9.38, \sigma = 0.085 \). The Interval Width is defined to be 1,000 miles. Enter the information about the usage distribution on the Suspensions page of the control panel, as shown next.
Click **Calculate** to analyze the data set. The parameters are estimated to be:

\[
\mu_T = 10.528098 \\
\sigma_T = 1.135150
\]

The reliability plot (with mileage being the random variable driving reliability), along with the 90% confidence bounds on reliability, is shown next.
In this example, the life data set contains 14 failures and 212 suspensions spread according to the defined usage distribution. You can display this data in a standard folio by choosing Warranty > Transfer Life Data > Transfer Life Data to New Folio. The failures and suspensions data set, as presented in the standard folio, is shown next (showing only the first 30 rows of data).
To illustrate the calculations behind the results of this example, consider the 9 units that went in service on December 2009. 1 unit failed from that group; therefore, 8 suspensions have survived from December 2009 until the beginning of December 2010, a total of 12 months. The calculations are summarized as follows.
The two columns on the right constitute the calculated suspension data (number of suspensions and their usage) for the group. The calculation is then repeated for each of the remaining groups in the data set. These data are then combined with the data about the failures to form the life data set that is used to estimate the failure distribution model.

### Warranty Prediction

Once a life data analysis has been performed on warranty data, this information can be used to predict how many warranty returns there will be in subsequent time periods. This methodology uses the concept of conditional reliability (see Basic Statistical Background) to calculate the probability of failure for the remaining units for each shipment time period. This conditional probability of failure is then multiplied by the number of units at risk from that particular shipment period that are still in the field (i.e., the suspensions) in order to predict the number of failures or warranty returns expected for this time period. The next example illustrates this.

### Example

Using the data in the following table, predict the number of warranty returns for October for each of the three shipment periods. Use the following Weibull parameters, beta = 2.4928 and eta = 6.6951.

<table>
<thead>
<tr>
<th>i</th>
<th>Interval (xi)</th>
<th>Percentage of Suspensions in xi</th>
<th>Number of Suspensions in xi</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>4000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>6000</td>
<td>5.55E-16</td>
<td>4.44089E-15</td>
<td>6000</td>
</tr>
<tr>
<td>8</td>
<td>7000</td>
<td>2.97E-10</td>
<td>2.37363E-09</td>
<td>7000</td>
</tr>
<tr>
<td>9</td>
<td>8000</td>
<td>1.91E-06</td>
<td>1.52577E-05</td>
<td>8000</td>
</tr>
<tr>
<td>10</td>
<td>9000</td>
<td>0.000605176</td>
<td>0.004841406</td>
<td>9000</td>
</tr>
<tr>
<td>11</td>
<td>10000</td>
<td>0.022360116</td>
<td>0.1788093</td>
<td>10000</td>
</tr>
<tr>
<td>12</td>
<td>11000</td>
<td>0.167901546</td>
<td>1.34321237</td>
<td>11000</td>
</tr>
<tr>
<td>13</td>
<td>12000</td>
<td>0.368340182</td>
<td>2.946721457</td>
<td>12000</td>
</tr>
<tr>
<td>14</td>
<td>13000</td>
<td>0.930760224</td>
<td>2.424608189</td>
<td>13000</td>
</tr>
<tr>
<td>15</td>
<td>14000</td>
<td>0.112862973</td>
<td>0.92930378</td>
<td>14000</td>
</tr>
<tr>
<td>16</td>
<td>15000</td>
<td>0.022085042</td>
<td>0.176680333</td>
<td>15000</td>
</tr>
<tr>
<td>17</td>
<td>16000</td>
<td>0.002561954</td>
<td>0.020495636</td>
<td>16000</td>
</tr>
<tr>
<td>18</td>
<td>17000</td>
<td>0.000194232</td>
<td>0.001553859</td>
<td>17000</td>
</tr>
<tr>
<td>19</td>
<td>18000</td>
<td>1.04E-05</td>
<td>8.333335E-05</td>
<td>18000</td>
</tr>
<tr>
<td>20</td>
<td>19000</td>
<td>4.21E-07</td>
<td>3.36608E-06</td>
<td>19000</td>
</tr>
<tr>
<td>21</td>
<td>20000</td>
<td>1.35E-08</td>
<td>1.07967E-07</td>
<td>20000</td>
</tr>
<tr>
<td>22</td>
<td>21000</td>
<td>3.58E-10</td>
<td>2.86575E-09</td>
<td>21000</td>
</tr>
<tr>
<td>23</td>
<td>22000</td>
<td>8.16E-12</td>
<td>6.52914E-11</td>
<td>22000</td>
</tr>
<tr>
<td>24</td>
<td>23000</td>
<td>1.64E-13</td>
<td>1.31135E-12</td>
<td>23000</td>
</tr>
<tr>
<td>25</td>
<td>24000</td>
<td>2.97E-15</td>
<td>2.37896E-14</td>
<td>24000</td>
</tr>
<tr>
<td>26</td>
<td>25000</td>
<td>4.97E-17</td>
<td>3.97252E-16</td>
<td>25000</td>
</tr>
<tr>
<td>27</td>
<td>26000</td>
<td>8.13E-19</td>
<td>6.50521E-18</td>
<td>26000</td>
</tr>
<tr>
<td>28</td>
<td>27000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>28000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>29000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>30000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Sum 1 8
### Solution

Use the Weibull parameter estimates to determine the conditional probability of failure for each shipment time period, and then multiply that probability with the number of units that are at risk for that period as follows. The equation for the conditional probability of failure is given by:

\[ Q(t|T) = 1 - R(t|T) = 1 - \frac{R(T + t)}{R(T)} \]

For the June shipment, there are 89 units that have successfully operated until the end of September (\( T = 3 \) months). The probability of one of these units failing in the next month (\( t = 1 \) month) is then given by:

\[ Q(1|3) = 1 - \frac{R(4)}{R(3)} = 1 - \frac{e^{-\left(\frac{4}{6.70}\right)^{2.40}}}{e^{-\left(\frac{3}{6.70}\right)^{2.40}}} = 1 - \frac{0.7582}{0.8735} = 0.132 \]

Once the probability of failure for an additional month of operation is determined, the expected number of failed units during the next month, from the June shipment, is the product of this probability and the number of units at risk (\( \hat{S}_{JUN,3} = 89 \)):

\[ \hat{F}_{JUN,A} = 89 \cdot 0.132 = 11.748, \text{ or } 12 \text{ units} \]

This is then repeated for the July shipment, where there were 134 units operating at the end of September, with an exposure time of two months. The probability of failure in the next month is:

\[ Q(1|2) = 1 - \frac{R(3)}{R(2)} = 1 - \frac{0.8735}{0.9519} = 0.0824 \]

This value is multiplied by \( S_{JUL,2} = 134 \) to determine the number of failures, or:

\[ \hat{F}_{JUL,3} = 134 \cdot 0.0824 = 11.035, \text{ or } 11 \text{ units} \]

For the August shipment, there were 146 units operating at the end of September, with an exposure time of one month. The probability of failure in the next month is:

\[ Q(1|1) = 1 - \frac{R(2)}{R(1)} = 1 - \frac{0.9519}{0.9913} = 0.0397 \]

This value is multiplied by \( S_{AUG,1} = 146 \) to determine the number of failures, or:

\[ \hat{F}_{AUG,2} = 146 \cdot 0.0397 = 5.796, \text{ or } 6 \text{ units} \]

Thus, the total expected returns from all shipments for the next month is the sum of the above, or 29 units. This method can be easily repeated for different future sales periods, and utilizing projected shipments. If the user lists the number of units that are expected to be sold or shipped during future periods, then these units are added to the number of units at risk whenever they are introduced into the field. The **Generate Forecast** functionality in the Weibull++ warranty analysis folio can automate this process for you.
Non-Homogeneous Warranty Data

In the previous sections and examples, it is important to note that the underlying assumption was that the population was homogeneous. In other words, all sold and returned units were exactly the same (i.e., the same population with no design changes and/or modifications). In many situations, as the product matures, design changes are made to enhance and/or improve the reliability of the product. Obviously, an improved product will exhibit different failure characteristics than its predecessor. To analyze such cases, where the population is non-homogeneous, one needs to extract each homogenous group, fit a life model to each group and then project the expected returns for each group based on the number of units at risk for each specific group.

Using Subset IDs in Weibull++

Weibull++ includes an optional Subset ID column that allows to differentiate between product versions or different designs (lots). Based on the entries, the software will separately analyze (i.e., obtain parameters and failure projections for) each subset of data. Note that it is important to realize that the same limitations with regards to the number of failures that are needed are also applicable here. In other words, distributions can be automatically fitted to lots that have return (failure) data, whereas if no returns have been experienced yet (either because the units are going to be introduced in the future or because no failures happened yet), the user will be asked to specify the parameters, since they can not be computed. Consequently, subsequent estimation/predictions related to these lots would be based on the user specified parameters. The following example illustrates the use of Subset IDs.

Example

Warranty Analysis Non-Homogeneous Data Example

A company keeps track of its production and returns. The company uses the dates of failure format to record the data. For the product in question, three versions (A, B and C) have been produced and put in service. The in-service data is as follows (using the Month/Day/Year date format):

<table>
<thead>
<tr>
<th>Quantity In – Service</th>
<th>Date of In – Service</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>1/1/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>500</td>
<td>1/31/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>500</td>
<td>5/1/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>600</td>
<td>5/31/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>550</td>
<td>6/30/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>600</td>
<td>7/30/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>800</td>
<td>9/28/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>200</td>
<td>1/1/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>350</td>
<td>3/2/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>450</td>
<td>4/1/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>300</td>
<td>6/30/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>200</td>
<td>8/29/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>350</td>
<td>10/28/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>1100</td>
<td>2/1/2005</td>
<td>Model C</td>
</tr>
<tr>
<td>1200</td>
<td>3/27/2005</td>
<td>Model C</td>
</tr>
<tr>
<td>1200</td>
<td>4/25/2005</td>
<td>Model C</td>
</tr>
<tr>
<td>1300</td>
<td>6/1/2005</td>
<td>Model C</td>
</tr>
<tr>
<td>1400</td>
<td>8/26/2005</td>
<td>Model C</td>
</tr>
</tbody>
</table>

Furthermore, the following sales are forecast:
The return data are as follows. Note that in order to identify which lot each unit comes from, and to be able to compute its time-in-service, each return (failure) includes a return date, the date of when it was put in service and the model ID.

<table>
<thead>
<tr>
<th>Quantity Returned</th>
<th>Date of Return</th>
<th>Date In – Service</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1/31/2005</td>
<td>1/1/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>11</td>
<td>4/1/2005</td>
<td>1/31/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>7</td>
<td>7/22/2005</td>
<td>5/1/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>8</td>
<td>8/27/2005</td>
<td>5/31/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>12</td>
<td>12/27/2005</td>
<td>5/31/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>13</td>
<td>1/26/2006</td>
<td>6/30/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>12</td>
<td>1/26/2006</td>
<td>7/30/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>14</td>
<td>1/11/2006</td>
<td>9/28/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>15</td>
<td>1/18/2006</td>
<td>9/28/2005</td>
<td>Model A</td>
</tr>
<tr>
<td>23</td>
<td>1/26/2005</td>
<td>1/1/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>16</td>
<td>1/26/2005</td>
<td>1/1/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>18</td>
<td>3/17/2005</td>
<td>1/1/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>20</td>
<td>5/31/2005</td>
<td>3/2/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>21</td>
<td>6/30/2005</td>
<td>3/2/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>18</td>
<td>7/30/2005</td>
<td>4/1/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>11</td>
<td>2/7/2006</td>
<td>10/28/2005</td>
<td>Model B</td>
</tr>
<tr>
<td>34</td>
<td>8/14/2005</td>
<td>3/27/2005</td>
<td>Model C</td>
</tr>
<tr>
<td>44</td>
<td>1/26/2006</td>
<td>6/1/2005</td>
<td>Model C</td>
</tr>
<tr>
<td>26</td>
<td>1/26/2006</td>
<td>8/26/2005</td>
<td>Model C</td>
</tr>
</tbody>
</table>

Assuming that the given information is current as of 5/1/2006, analyze the data using the lognormal distribution and MLE analysis method for all models (Model A, Model B, Model C), and provide a return forecast for the next ten months.

**Solution**

Create a warranty analysis folio and select the dates of failure format. Enter the data from the tables in the Sales, Returns and Future Sales sheets. On the control panel, select the Use Subsets check box, as shown next. This allows the software to separately analyze each subset of data. Use the drop-down list to switch between subset IDs and alter the analysis settings (use the lognormal distribution and MLE analysis method for all models).
In the End of Observation Period field, enter 5/1/2006, and then calculate the parameters. The results are:

\[
\begin{align*}
\hat{\mu}' & = 11.28 \\
\hat{\mu}' & = 8.11 \\
\hat{\mu}' & = 9.79 \\
\hat{\sigma}_T & = 2.83 \\
\hat{\sigma}_T & = 2.30 \\
\hat{\sigma}_T & = 1.92
\end{align*}
\]

Note that in this example, the same distribution and analysis method were assumed for each of the product models. If desired, different distribution types, analysis methods, confidence bounds methods, etc., can be assumed for each IDs.

To obtain the expected failures for the next 10 months, click the Generate Forecast icon. In the Forecast Setup window, set the forecast to start on May 2, 2006 and set the number of forecast periods to 10. Set the increment (length of each period) to 1 Month, as shown next.
Click OK. A Forecast sheet will be created, with the predicted future returns. The following figure shows part of the Forecast sheet.

To view a summary of the analysis, click the Show Analysis Summary (...) button. The following figure shows the summary of the forecasted returns.
Click the **Plot** icon and choose the **Expected Failures** plot. The plot displays the predicted number of returns for each month, as shown next.
Monitoring Warranty Returns Using Statistical Process Control (SPC)

By monitoring and analyzing warranty return data, one can detect specific return periods and/or batches of sales or shipments that may deviate (differ) from the assumed model. This provides the analyst (and the organization) the advantage of early notification of possible deviations in manufacturing, use conditions and/or any other factor that may adversely affect the reliability of the fielded product. Obviously, the motivation for performing such analysis is to allow for faster intervention to avoid increased costs due to increased warranty returns or more serious repercussions. Additionally, this analysis can also be used to uncover different sub-populations that may exist within the population.

Basic Analysis Method

For each sales period \( i \) and return period \( j \), the prediction error can be calculated as follows:

\[
e_{i,j} = \hat{F}_{i,j} - F_{i,j}
\]

where \( \hat{F}_{i,j} \) is the estimated number of failures based on the estimated distribution parameters for the sales period \( i \) and the return period \( j \), which is calculated using the equation for the conditional probability, and \( F_{i,j} \) is the actual number of failure for the sales period \( i \) and the return period \( j \).

Since we are assuming that the model is accurate, \( e_{i,j} \) should follow a normal distribution with mean value of zero and a standard deviation \( \sigma \), where:

\[
\bar{e}_{i,j} = \frac{\sum_{i} \sum_{j} e_{i,j}}{n} = 0
\]

and \( n \) is the total number of return data (total number of residuals).
The estimated standard deviation of the prediction errors can then be calculated by:

\[ s = \sqrt{\frac{1}{n - 1} \sum_{i} \sum_{j} e_{i,j}^2} \]

and \( e_{i,j} \) can be normalized as follows:

\[ z_{i,j} = \frac{e_{i,j}}{s} \]

where \( z_{i,j} \) is the standardized error. \( z_{i,j} \) follows a normal distribution with \( \mu = 0 \) and \( \sigma = 1 \).

It is known that the square of a random variable with standard normal distribution follows the \( \chi^2 \) (Chi Square) distribution with 1 degree of freedom and that the sum of the squares of \( m \) random variables with standard normal distribution follows the \( \chi^2 \) distribution with \( m \) degrees of freedom. This then can be used to help detect the abnormal returns for a given sales period, return period or just a specific cell (combination of a return and a sales period).

- For a cell, abnormality is detected if \( z_{i,j}^2 = \chi^2_{1,\alpha} \).
- For an entire sales period \( i \), abnormality is detected if \( \sum_{j} z_{i,j}^2 = \chi^2_I \geq \chi^2_{\alpha,I} \) where \( I \) is the total number of return period for a sales period \( i \).
- For an entire return period \( j \), abnormality is detected if \( \sum_{i} z_{i,j}^2 = \chi^2_I \geq \chi^2_{\alpha,I} \) where \( I \) is the total number of sales period for a return period \( j \).

Here \( \alpha \) is the criticality value of the \( \chi^2 \) distribution, which can be set at critical value or caution value. It describes the level of sensitivity to outliers (returns that deviate significantly from the predictions based on the fitted model). Increasing the value of \( \alpha \) increases the power of detection, but this could lead to more false alarms.

Example

Example Using SPC for Warranty Analysis Data

Using the data from the following table, the expected returns for each sales period can be obtained using conditional reliability concepts, as given in the conditional probability equation.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun. 2010</td>
<td>100</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Jul. 2010</td>
<td>140</td>
<td>-</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Aug. 2010</td>
<td>150</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
</tbody>
</table>

For example, for the month of September, the expected return number is given by:

\[ \hat{F}_{Jun,3} = (100 - 6) \cdot \left( 1 - \frac{R(3)}{R(2)} \right) = 94 \cdot 0.08239 = 7.7447 \]

The actual number of returns during this period is five; thus, the prediction error for this period is:

\[ e_{Jun,3} = \hat{F}_{Jun,3} - F_{Jun,3} = 7.7447 - 5 = 2.7447. \]

This can then be repeated for each cell, yielding the following table for \( e_{i,j} \):

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun. 2005</td>
<td>100</td>
<td>-2.1297</td>
<td>0.8462</td>
<td>2.7447</td>
</tr>
<tr>
<td>Jul. 2005</td>
<td>140</td>
<td>-</td>
<td>-0.7816</td>
<td>1.4719</td>
</tr>
<tr>
<td>Aug. 2005</td>
<td>150</td>
<td>-</td>
<td>-</td>
<td>-2.6946</td>
</tr>
</tbody>
</table>
Now, for this example, \( n = 6 \), \( \bar{c}_{i,j} = -0.5432 \) and \( s = 1.6890 \).

Thus the \( z_{i,j} \) values are:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun. 2005</td>
<td>100</td>
<td>-0.9968</td>
<td>0.3960</td>
<td>1.2846</td>
</tr>
<tr>
<td>Jul. 2005</td>
<td>140</td>
<td>-</td>
<td>-0.3658</td>
<td>0.6889</td>
</tr>
<tr>
<td>Aug. 2005</td>
<td>150</td>
<td>-</td>
<td>-</td>
<td>-1.2612</td>
</tr>
</tbody>
</table>

The \( z_{i,j}^2 \) values, for each cell, are given in the following table.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun. 2005</td>
<td>100</td>
<td>0.9936</td>
<td>0.1569</td>
<td>1.6505</td>
<td>2.8010</td>
</tr>
<tr>
<td>Jul. 2005</td>
<td>140</td>
<td>-</td>
<td>0.1338</td>
<td>0.4747</td>
<td>0.6085</td>
</tr>
<tr>
<td>Aug. 2005</td>
<td>150</td>
<td>-</td>
<td>-</td>
<td>1.5905</td>
<td>1.5905</td>
</tr>
<tr>
<td>Sum</td>
<td>0.9936</td>
<td>0.2907</td>
<td>3.7157</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If the critical value is set at \( \alpha = 0.01 \) and the caution value is set at \( \alpha = 0.1 \), then the critical and caution \( \chi^2 \) values will be:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi^2 ) Critical</td>
<td>6.6349</td>
<td>9.2103</td>
<td>11.3449</td>
</tr>
<tr>
<td>( \chi^2 ) Caution</td>
<td>2.7055</td>
<td>4.6052</td>
<td>6.2514</td>
</tr>
</tbody>
</table>

If we consider the sales periods as the basis for outlier detection, then after comparing the above table to the sum of \( z_{i,j}^2(\chi^2_i) \) values for each sales period, we find that all the sales values do not exceed the critical and caution limits.

For example, the total \( \chi^2 \) value of the sale month of July is 0.6085. Its degrees of freedom is 2, so the corresponding caution and critical values are 4.6052 and 9.2103 respectively. Both values are larger than 0.6085, so the return numbers of the July sales period do not deviate (based on the chosen significance) from the model’s predictions.

If we consider returns periods as the basis for outliers detection, then after comparing the above table to the sum of \( z_{i,j}(\chi^2_i) \) values for each return period, we find that all the return values do not exceed the critical and caution limits.

For example, the total \( \chi^2 \) value of the sale month of August is 3.7157. Its degree of freedom is 3, so the corresponding caution and critical values are 6.2514 and 11.3449 respectively. Both values are larger than 3.7157, so the return numbers for the June return period do not deviate from the model’s predictions.

This analysis can be automatically performed in Weibull++ by entering the alpha values in the Statistical Process Control page of the control panel and selecting which period to color code, as shown next.
To view the table of chi-squared values (\( \chi^2 \) values), click the **Show Results (...)** button.

Weibull++ automatically color codes SPC results for easy visualization in the returns data sheet. By default, the green color means that the return number is normal; the yellow color indicates that the return number is larger than the caution threshold but smaller than the critical value; the red color means that the return is abnormal, meaning that the return number is either too big or too small compared to the predicted value.
In this example, all the cells are coded in green for both analyses (i.e., by sales periods or by return periods), indicating that all returns fall within the caution and critical limits (i.e., nothing abnormal). Another way to visualize this is by using a Chi-Squared plot for the sales period and return period, as shown next.
Using Subset IDs with SPC for Warranty Data

The warranty monitoring methodology explained in this section can also be used to detect different subpopulations in a data set. The different subpopulations can reflect different use conditions, different material, etc. In this methodology, one can use different subset IDs to differentiate between subpopulations, and obtain models that are distinct to each subpopulation. The following example illustrates this concept.

Example

Using Subset IDs with Statistical Process Control

A manufacturer wants to monitor and analyze the warranty returns for a particular product. They collected the following sales and return data.

<table>
<thead>
<tr>
<th>Period</th>
<th>QuantityIn – Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sep 04</td>
<td>1150</td>
</tr>
<tr>
<td>Oct 04</td>
<td>1100</td>
</tr>
<tr>
<td>Nov 04</td>
<td>1200</td>
</tr>
<tr>
<td>Dec 04</td>
<td>1155</td>
</tr>
<tr>
<td>Jan 05</td>
<td>1255</td>
</tr>
<tr>
<td>Feb 05</td>
<td>1150</td>
</tr>
<tr>
<td>Mar 05</td>
<td>1105</td>
</tr>
<tr>
<td>Apr 05</td>
<td>1110</td>
</tr>
</tbody>
</table>
Solution

Analyze the data using the two-parameter Weibull distribution and the MLE analysis method. The parameters are estimated to be:

\[ \beta = 2.318144 \]
\[ \eta = 25.071878 \]

To analyze the warranty returns, select the check box in the Statistical Process Control page of the control panel and set the alpha values to 0.01 for the Critical Value and 0.1 for the Caution Value. Select to color code the results By sales period. The following figure shows the analysis settings and results of the analysis.

As you can see, the November 04 and March 05 sales periods are colored in yellow indicating that they are outlier sales periods, while the rest are green. One suspected reason for the variation may be the material used in production during these periods. Further analysis confirmed that for these periods, the material was acquired from a different supplier. This implies that the units are not homogenous, and that there are different sub-populations present in the field population.

Categorized each shipment (using the Subset ID column) based on their material supplier, as shown next. On the control panel, select the Use Subsets check box. Perform the analysis again using the two-parameter Weibull distribution and the MLE analysis method for both sub-populations.
The new models that describe the data are:

\[
\begin{align*}
\text{Supplier 1} & : \beta = 2.381905 \quad \eta = 25.397633 \\
\text{Supplier 2} & : \beta = 2.320696 \quad \eta = 21.282926
\end{align*}
\]

This analysis uncovered different sub-populations in the data set. Note that if the analysis were performed on the failure and suspension times in a regular standard folio using the mixed Weibull distribution, one would not be able to detect which units fall into which sub-population.

References

Chapter 20

Recurrent Event Data Analysis

Recurrent Event Data Analysis (RDA) is used in various applied fields such as reliability, medicine, social sciences, economics, business and criminology. Whereas in life data analysis (LDA) it was assumed that events (failures) were independent and identically distributed (iid), there are many cases where events are dependent and not identically distributed (such as repairable system data) or where the analyst is interested in modeling the number of occurrences of events over time rather than the length of time prior to the first event, as in LDA.

Weibull++ provides both parametric and non-parametric approaches to analyze such data.

• The non-parametric approach is based on the well-known Mean Cumulative Function (MCF). The Weibull++ module for this type of analysis builds upon the work of Dr. Wayne Nelson, who has written extensively on the calculation and applications of MCF [31].

• The parametric approach is based on the General Renewal Process (GRP) model, which is particularly useful in understanding the effects of the repairs on the age of a system. Traditionally, the commonly used models for analyzing repairable systems data are the perfect renewal processes (PRP), which corresponds to perfect repairs, and the nonhomogeneous Poisson processes (NHPP), which corresponds to minimal repairs. However, most repair activities may realistically not result in such extreme situations but in a complicated intermediate one (general repair or imperfect repair/maintenance), which are well treated with the GRP model.

Non-Parametric Recurrent Event Data Analysis

Non-parametric RDA provides a non-parametric graphical estimate of the mean cumulative number or cost of recurrence per unit versus age. As discussed in Nelson [31], in the reliability field, the Mean Cumulative Function (MCF) can be used to:

• Evaluate whether the population repair (or cost) rate increases or decreases with age (this is useful for product retirement and burn-in decisions).
• Estimate the average number or cost of repairs per unit during warranty or some time period.
• Compare two or more sets of data from different designs, production periods, maintenance policies, environments, operating conditions, etc.
• Predict future numbers and costs of repairs, such as the expected number of failures next month, quarter, or year.
• Reveal unexpected information and insight.
The Mean Cumulative Function (MCF)

In a non-parametric analysis of recurrent event data, each population unit can be described by a cumulative history function for the cumulative number of recurrences. It is a staircase function that depicts the cumulative number of recurrences of a particular event, such as repairs over time. The figure below depicts a unit's cumulative history function.

The non-parametric model for a population of units is described as the population of cumulative history functions (curves). It is the population of all staircase functions of every unit in the population. At age $t$, the units have a distribution of their cumulative number of events. That is, a fraction of the population has accumulated 0 recurrences, another fraction has accumulated 1 recurrence, another fraction has accumulated 2 recurrences, etc. This distribution differs at different ages $t$, and has a mean $\bar{M}(t)$ called the mean cumulative function (MCF). The $\bar{M}(t)$ is the point-wise average of all population cumulative history functions (see figure below).

For the case of uncensored data, the mean cumulative function $\bar{M}(t)$ values at different recurrence ages $t$ are estimated by calculating the average of the cumulative number of recurrences of events for each unit in the
population at $t_i$. When the histories are censored, the following steps are applied.

**1st Step - Order all ages:**

Order all recurrence and censoring ages from smallest to largest. If a recurrence age for a unit is the same as its censoring (suspension) age, then the recurrence age goes first. If multiple units have a common recurrence or censoring age, then these units could be put in a certain order or be sorted randomly.

**2nd Step - Calculate the number, $r_i$, of units that passed through age $t_i$:**

$$r_i = r_{i-1} - 1$$ if $t_i$ is a recurrence age

$$r_i = r_{i-1}$$ if $t_i$ is a censoring age

$N$ is the total number of units and $r_1 = N$ at the first observed age which could be a recurrence or suspension.

**3rd Step - Calculate the MCF estimate, $M^*(t)$:**

For each sample recurrence age $t_i$, calculate the mean cumulative function estimate as follows

$$M^*(t_i) = \frac{1}{r_i} + M^*(t_{i-1})$$

where $M^*(t_1) = \frac{1}{r_1}$ at the earliest observed recurrence age, $t_1$.

**Confidence Limits for the MCF**

Upper and lower confidence limits for $M(t_i)$ are:

$$M_U(t_i) = M^*(t_i) \cdot \frac{K_{\alpha} \cdot \sqrt{\text{Var}[M^*(t_i)]}}{M^*(t_i)}$$

$$M_L(t_i) = M^*(t_i) \cdot \frac{1}{M^*(t_i)}$$

where $0 < \alpha < 100\%$ is confidence level, $K_{\alpha}$ is the standard normal percentile and $\text{Var}[M^*(t_i)]$ is the variance of the MCF estimate at recurrence age $t_i$. The variance is calculated as follows:

$$\text{Var}[M^*(t_i)] = \text{Var}[M^*(t_{i-1})] + \frac{1}{r_i^2} \left( \sum_{j \in R_i} d_{ji} - \frac{1}{r_i} \right)^2$$

where $R_i$ is defined in the equation of the survivals, $R_i$ is the set of the units that have not been suspended by $t_i$ and $d_{ji}$ is defined as follows:

$$d_{ji} = 1$$ if the $j$th unit had an event recurrence at age $t_i$

$$d_{ji} = 0$$ if the $j$th unit did not have an event reoccur at age $t_i$

In case there are multiple events at the same time $t_i$, $d_{ji}$ is calculated sequentially for each event. For each event, only one $d_{ji}$ can take value of 1. Once all the events at $t_i$ are calculated, the final calculated MCF and its variance are the values for time $t_i$. This is illustrated in the following example.

**Example: Mean Cumulative Function**

A health care company maintains five identical pieces of equipment used by a hospital. When a piece of equipment fails, the company sends a crew to repair it. The following table gives the failure and censoring ages for each machine, where the + sign indicates a censoring age.

<table>
<thead>
<tr>
<th>EquipmentID</th>
<th>Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5, 10, 15, 17+</td>
</tr>
<tr>
<td>2</td>
<td>6, 13, 17, 19+</td>
</tr>
<tr>
<td>3</td>
<td>12, 20, 25, 26+</td>
</tr>
<tr>
<td>4</td>
<td>13, 15, 24+</td>
</tr>
<tr>
<td>5</td>
<td>16, 22, 25, 28+</td>
</tr>
</tbody>
</table>
Estimate the MCF values, with 95% confidence bounds.

Solution

The MCF estimates are obtained as follows:

<table>
<thead>
<tr>
<th>ID</th>
<th>Months($t_i$)</th>
<th>State</th>
<th>$r_i$</th>
<th>$1/r_i$</th>
<th>$M^*(t_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>0.20 + 0.20 = 0.40</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>0.40 + 0.20 = 0.60</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>0.60 + 0.20 = 0.80</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>0.80 + 0.20 = 1.00</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>1.00 + 0.20 = 1.20</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>1.20 + 0.20 = 1.40</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>1.40 + 0.20 = 1.60</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>1.60 + 0.20 = 1.80</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>F</td>
<td>5</td>
<td>0.20</td>
<td>1.80 + 0.20 = 2.00</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>S</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>S</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>F</td>
<td>3</td>
<td>0.33</td>
<td>2.00 + 0.33 = 2.33</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>F</td>
<td>3</td>
<td>0.33</td>
<td>2.33 + 0.33 = 2.66</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>S</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>F</td>
<td>2</td>
<td>0.50</td>
<td>2.66 + 0.50 = 3.16</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>F</td>
<td>2</td>
<td>0.50</td>
<td>3.16 + 0.50 = 3.66</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>S</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>S</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using the MCF variance equation, the following table of variance values can be obtained:

<table>
<thead>
<tr>
<th>ID</th>
<th>Months</th>
<th>State</th>
<th>$r_i$</th>
<th>$Var_{r_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>F</td>
<td>5</td>
<td>$(\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.032$</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>F</td>
<td>5</td>
<td>$0.032 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.064$</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>F</td>
<td>5</td>
<td>$0.064 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.096$</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>F</td>
<td>5</td>
<td>$0.096 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.128$</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>F</td>
<td>5</td>
<td>$0.128 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.160$</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>F</td>
<td>5</td>
<td>$0.160 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.192$</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>F</td>
<td>5</td>
<td>$0.192 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.224$</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>F</td>
<td>5</td>
<td>$0.224 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.256$</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>F</td>
<td>5</td>
<td>$0.256 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.288$</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>F</td>
<td>5</td>
<td>$0.288 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 4(0 - \frac{1}{5})^2] = 0.320$</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>S</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>S</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>F</td>
<td>3</td>
<td>$0.320 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 2(0 - \frac{1}{5})^2] = 0.394$</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>F</td>
<td>3</td>
<td>$0.394 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + 2(0 - \frac{1}{5})^2] = 0.468$</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>S</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>F</td>
<td>2</td>
<td>$0.468 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + (0 - \frac{1}{5})^2] = 0.593$</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>F</td>
<td>2</td>
<td>$0.593 + (\frac{1}{5})^2[(1 - \frac{1}{5})^2 + (0 - \frac{1}{5})^2] = 0.718$</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>S</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Using the equation for the MCF bounds and $K_5 = 1.644$ for a 95% confidence level, the confidence bounds can be obtained as follows:

<table>
<thead>
<tr>
<th>ID</th>
<th>Months</th>
<th>State</th>
<th>MCF_i</th>
<th>Var_i</th>
<th>MCF_L</th>
<th>MCF_U</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>F</td>
<td>0.20</td>
<td>0.032</td>
<td>0.0459</td>
<td>0.8709</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>F</td>
<td>0.40</td>
<td>0.064</td>
<td>0.1413</td>
<td>1.1320</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>F</td>
<td>0.60</td>
<td>0.096</td>
<td>0.2566</td>
<td>1.4029</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>F</td>
<td>0.80</td>
<td>0.128</td>
<td>0.3834</td>
<td>1.6694</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>F</td>
<td>1.00</td>
<td>0.160</td>
<td>0.5179</td>
<td>1.9308</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>F</td>
<td>1.20</td>
<td>0.192</td>
<td>0.6582</td>
<td>2.1879</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>F</td>
<td>1.40</td>
<td>0.224</td>
<td>0.8028</td>
<td>2.4413</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>F</td>
<td>1.60</td>
<td>0.256</td>
<td>0.9511</td>
<td>2.6916</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>F</td>
<td>1.80</td>
<td>0.288</td>
<td>1.1023</td>
<td>2.9393</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>F</td>
<td>2.00</td>
<td>0.320</td>
<td>1.2560</td>
<td>3.1848</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>F</td>
<td>2.33</td>
<td>0.394</td>
<td>1.4990</td>
<td>3.6321</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>F</td>
<td>2.66</td>
<td>0.468</td>
<td>1.7486</td>
<td>4.0668</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>F</td>
<td>3.16</td>
<td>0.593</td>
<td>2.1226</td>
<td>4.7243</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>F</td>
<td>3.66</td>
<td>0.718</td>
<td>2.5071</td>
<td>5.3626</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The analysis presented in this example can be performed automatically in Weibull++'s non-parametric RDA folio, as shown next.
Note: In the folio above, the \( F \) refers to failures and \( E \) refers to suspensions (or censoring ages). The results, with calculated MCF values and upper and lower 95% confidence limits, are shown next along with the graphical plot.
Parametric Recurrent Event Data Analysis

Weibull++’s parametric RDA folio is a tool for modeling recurrent event data. It can capture the trend, estimate the rate and predict the total number of recurrences. The failure and repair data of a repairable system can be treated as one type of recurrence data. Past and current repairs may affect the future failure process. For most recurrent events, time (distance, cycles, etc.) is a key factor. With time, the recurrence rate may remain constant, increase or decrease. For other recurrent events, not only the time, but also the number of events can affect the recurrence process (e.g., the debugging process in software development).

The parametric analysis approach utilizes the General Renewal Process (GRP) model, as discussed in Mettas and Zhao [28]. In this model, the repair time is assumed to be negligible so that the processes can be viewed as point processes. This model provides a way to describe the rate of occurrence of events over time, such as in the case of data obtained from a repairable system. This model is particularly useful in modeling the failure behavior of a specific system and understanding the effects of the repairs on the age of that system. For example, consider a system that is repaired after a failure, where the repair does not bring the system to an as-good-as-new or an as-bad-as-old condition. In other words, the system is partially rejuvenated after the repair. Traditionally, in as-bad-as-old repairs, also known as minimal repairs, the failure data from such a system would have been modeled using a homogeneous or non-homogeneous Poisson process (NHPP). On rare occasions, a Weibull distribution has been used as well in cases where the system is almost as-good-as-new after the repair, also known as a perfect renewal process (PRP). However, for the intermediate states after the repair, there has not been a commercially available model, even though many models have been proposed in literature. In Weibull++, the GRP model provides the capability to model systems with partial renewal (general repair or imperfect repair/maintenance) and allows for a variety of predictions such as reliability, expected failures, etc.

The GRP Model

In this model, the concept of virtual age is introduced. Let \( t_1, t_2, \ldots, t_n \) represent the successive failure times and let \( x_1, x_2, \ldots, x_n \) represent the time between failures (\( t_i = \sum_{j=1}^{i-1} x_j \)). Assume that after each event, actions are taken to improve the system performance. Let \( q \) be the action effectiveness factor. There are two GRP models:

Type I:

\[
v_i = v_{i-1} + qx_i = q^i\]

Type II:

\[
v_i = q(v_{i-1} + x_i) = q^i x_1 + q^{i-1} x_2 + \cdots + qx_i\]

where \( v_i \) is the virtual age of the system right after \( i \)th repair. The Type I model assumes that the \( i \)th repair cannot remove the damage incurred before the \( (i - 1) \)th repair. It can only reduce the additional age to \( x_i \). The Type II model assumes that at the \( i \)th repair, the virtual age has been accumulated to \( v_{i-1} + x_i \). The \( i \)th repair will remove the cumulative damage from both current and previous failures by reducing the virtual age to \( q(v_{i-1} + x_i) \).

The power law function is used to model the rate of recurrence, which is:

\[
\lambda(t) = \lambda_0 t^{\beta-1}
\]
The conditional pdf is:

$$f(t_i|t_{i-1}) = \lambda \beta (x_i + v_{i-1})^{\beta-1} e^{-\lambda (x_i + v_{i-1})^\beta - v_{i-1}^\beta}$$

MLE method is used to estimate the model parameters. The log likelihood function is discussed in Mettas and Zhao [28]:

$$\ln(L) = n(\ln \lambda + \ln \beta) - \lambda \left[(T - t_n + v_n)^\beta - v_n^\beta\right]$$

$$- \lambda \sum_{i=1}^n \left[(x_i + v_{i-1})^\beta - v_i^\beta\right] + (\beta - 1) \sum_{i=1}^n \ln(x_i + v_{i-1})$$

where $n$ is the total number of events during the entire observation period. $T$ is the stop time of the observation.

Confidence Bounds

In general, in order to obtain the virtual age, the exact occurrence time of each event (failure) should be available (see equations for Type I and Type II models). However, the times are unknown until the corresponding events occur. For this reason, there are no closed-form expressions for total failure number and failure intensity, which are functions of failure times and virtual age. Therefore, in Weibull++, a Monte Carlo simulation is used to predict values of virtual time, failure number, MTBF and failure rate. The approximate confidence bounds obtained from simulation are provided. The uncertainty of model parameters is also considered in the bounds.

Bounds on Cumulative Failure (Event) Numbers

The variance of the cumulative failure number $N(t)$ is:

$$Var[N(t)] = Var[E(N(t)|\lambda, \beta, q)] + E[Var(N(t)|\lambda, \beta, q)]$$

The first term accounts for the uncertainty of the parameter estimation. The second term considers the uncertainty caused by the renewal process even when model parameters are fixed. However, unless $q = 1$, $Var[E(N(t)|\lambda, \beta, q)]$ cannot be calculated because $E(N(t))$ cannot be expressed as a closed-form function of $\lambda, \beta$, and $q$. In order to consider the uncertainty of the parameter estimation, $Var[N(t)|\lambda, \beta, q]$ is approximated by:

$$Var[E(N(t)|\lambda, \beta, q)] = Var[E(N(t)|\lambda, \beta)] = Var[\lambda v_t^\beta]$$

where $v_t$ is the expected virtual age at time $t$ and $Var[\lambda v_t^\beta]$ is:

$$Var[\lambda v_t^\beta] = \frac{\partial^2 (\lambda v_t^\beta)}{\partial \beta^2} Var(\beta) + \frac{\partial^2 (\lambda v_t^\beta)}{\partial \lambda^2} Var(\lambda)$$

$$+ 2 \frac{\partial (\lambda v_t^\beta)}{\partial \beta} \frac{\partial (\lambda v_t^\beta)}{\partial \lambda} Cov(\beta, \lambda)$$

By conducting this approximation, the uncertainty of $\lambda$ and $\beta$ are considered. The value of $v_t$ and the value of the second term in the equation for the variance of number of failures are obtained through the Monte Carlo simulation using parameters $\hat{\lambda}, \hat{\beta}, \hat{q}$, which are the ML estimators. The same simulation is used to estimate the cumulative number of failures $\hat{N}(t) = E(N(t)|\hat{\lambda}, \hat{\beta}, \hat{q})$.

Once the variance and the expected value of $\hat{N}(t)$ have been obtained, the bounds can be calculated by assuming that $\hat{N}(t)$ is lognormally distributed as:

$$\frac{\ln N(t) - \ln \hat{N}(t)}{\sqrt{Var(\ln N(t))}} \sim N(0, 1)$$

The upper and lower bounds for a given confidence level $\alpha$ can be calculated by:
Recurrent Event Data Analysis

\[ N(t)_{U,L} = \hat{N}(t)e^{\pm z_a\sqrt{Var(N(t))}/\hat{N}(t)} \]

where \( z_a \) is the standard normal distribution.

If \( \hat{N}(t) \) is assumed to be normally distributed, the bounds can be calculated by:

\[ N(t)_U = \hat{N}(t) + z_a\sqrt{Var(N(t))} \]
\[ N(t)_L = \hat{N}(t) - z_a\sqrt{Var(N(t))} \]

In Weibull++, the \( \hat{N}(t)_{U,L} \) is the smaller of the upper bounds obtained from lognormal and normal distribution approximation. The \( \hat{N}(t)_{L,U} \) is the set to the largest of the lower bounds obtained from lognormal and normal distribution approximation. This combined method can prevent the out-of-range values of bounds for some small \( t \) values.

**Bounds of Cumulative Failure Intensity and MTBF**

For a given time \( t \), the expected value of cumulative MTBF \( m_c(t) \) and cumulative failure intensity \( \lambda_c(t) \) can be calculated using the following equations:

\[ \hat{\lambda}_c(t) = \frac{\hat{N}(t)}{t}; \hat{m}_c(t) = \frac{t}{\hat{N}(t)} \]

The bounds can be easily obtained from the corresponding bounds of \( \hat{N}(t) \):

\[ \hat{\lambda}_c(t)_L = \frac{\hat{N}(t)_L}{t}; \hat{\lambda}_c(t)_U = \frac{\hat{N}(t)_U}{t}; \]
\[ \hat{m}_c(t)_L = \frac{t}{\hat{N}(t)_U}; \hat{m}_c(t)_U = \frac{t}{\hat{N}(t)_L} \]

**Bounds on Instantaneous Failure Intensity and MTBF**

The instantaneous failure intensity is given by:

\[ \lambda_i(t) = \lambda \beta v_i^{\beta-1} \]

where \( v_i \) is the virtual age at time \( t \). When \( q \neq 1 \), it is obtained from simulation. When \( q = 1 \), \( v_i = t \) from model Type I and Type II.

The variance of instantaneous failure intensity can be calculated by:

\[ Var(\lambda_i(t)) = \left( \frac{\partial \lambda_i(t)}{\partial \beta} \right)^2 \cdot Var(\hat{\beta}) + \left( \frac{\partial \lambda_i(t)}{\partial \lambda} \right)^2 \cdot Var(\hat{\lambda}) + 2 \cdot \frac{\partial \lambda_i(t)}{\partial \beta} \cdot \frac{\partial \lambda_i(t)}{\partial \lambda} \cdot Cov(\hat{\beta}, \hat{\lambda}) + \left( \frac{\partial \lambda_i(t)}{\partial v(t)} \right)^2 \cdot Var(\hat{v}_t) \]

The expected value and variance of \( v_t \) are obtained from the Monte Carlo simulation with parameters \( \hat{\lambda}, \hat{\beta}, \hat{\gamma} \). Because of the simulation accuracy and the convergence problem in calculation of \( Var(\hat{\beta}), Var(\hat{\lambda}) \) and \( Cov(\hat{\beta}, \hat{\lambda}) \), \( Var(\lambda_i(t)) \) can be a negative value at some time points. When this case happens, the bounds of instantaneous failure intensity are not provided.

Once the variance and the expected value of \( \lambda_i(t) \) are obtained, the bounds can be calculated by assuming that \( \lambda_i(t) \) is lognormally distributed as:

\[ \frac{\ln \lambda_i(t) - \ln \hat{\lambda}_i(t)}{\sqrt{Var(\ln \lambda_i(t))}} \sim N(0,1) \]

The upper and lower bounds for a given confidence level \( \alpha \) can be calculated by:

\[ \lambda_i(t) = \hat{\lambda}_i(t)e^{\pm z_a\sqrt{Var(\lambda_i(t))}/\hat{\lambda}_i(t)} \]

where \( z_a \) is the standard normal distribution.
If \( \lambda_i(t) \) is assumed to be normally distributed, the bounds can be calculated by:

\[
\lambda_i(t)_U = \tilde{\lambda}_i(t) + z_a \sqrt{\text{Var}(N(t))} \\
\lambda_i(t)_L = \tilde{\lambda}_i(t) - z_a \sqrt{\text{Var}(N(t))}
\]

In Weibull++, \( \lambda_i(t) \) is set to the smaller of the two upper bounds obtained from the above lognormal and normal distribution approximation. \( \lambda_i(t)_L \) is set to the largest of the two lower bounds obtained from the above lognormal and normal distribution approximation. This combination method can prevent the out of range values of bounds when \( b \) values are small.

For a given time \( t \), the expected value of cumulative MTBF \( \hat{m}_i(t) \) is:

\[
\hat{m}_i(t) = \frac{1}{\tilde{\lambda}_i(t)}
\]

The upper and lower bounds can be easily obtained from the corresponding bounds of \( \tilde{\lambda}_i(t) \):

\[
\hat{m}_i(t)_U = \frac{1}{\lambda_i(t)_L} \\
\hat{m}_i(t)_L = \frac{1}{\lambda_i(t)_U}
\]

**Bounds on Conditional Reliability**

Given mission start time \( t_0 \) and mission time \( T \), the conditional reliability can be calculated by:

\[
R(T|t_0) = \frac{R(T + \hat{v}_0)}{R(\hat{v}_0)} = e^{-\lambda(\hat{v}_0 + T)\beta - \hat{v}_0}
\]

\( \hat{v}_0 \) is the virtual age corresponding to time \( t_0 \). The expected value and the variance of \( \hat{v}_0 \) are obtained from Monte Carlo simulation. The variance of the conditional reliability \( R(T|t_0) \) is:

\[
\text{Var}(R) = \left( \frac{\partial R}{\partial \beta} \right)^2 \text{Var}(\hat{\beta}) + \left( \frac{\partial R}{\partial \lambda} \right)^2 \text{Var}(\hat{\lambda})
\]

\[
+ 2 \frac{\partial R}{\partial \beta} \frac{\partial R}{\partial \lambda} \text{Cov}(\hat{\beta}, \hat{\lambda}) + \left( \frac{\partial R}{\partial \hat{v}_0} \right)^2 \text{Var}(\hat{v}_0)
\]

Because of the simulation accuracy and the convergence problem in calculation of \( \text{Var}(\hat{\beta}), \text{Var}(\hat{\lambda}) \) and \( \text{Cov}(\hat{\beta}, \hat{\lambda}) \), \( \text{Var}(R) \) can be a negative value at some time points. When this case happens, the bounds are not provided.

The bounds are based on:

\[
\log \text{it}(\hat{R}(T)) = N(0, 1)
\]

\[
\log \text{it}(\hat{R}(T)) = \ln \left\{ \frac{\hat{R}(T)}{1 - \hat{R}(T)} \right\}
\]

The confidence bounds on reliability are given by:

\[
R = \frac{\hat{R}}{\hat{R} + (1 - \hat{R}) e^{\pm \sqrt{\text{Var}(R)}/[\hat{R}(1 - \hat{R})]}}
\]

It will be compared with the bounds obtained from:

\[
R = \hat{R} e^{\pm z_a \sqrt{\text{Var}(R)}/\hat{R}}
\]

The smaller of the two upper bounds will be the final upper bound and the larger of the two lower bounds will be the final lower bound.
**Example: Air Condition Unit**

The following table gives the failure times for the air conditioning unit of an aircraft. The observation ended by the time the last failure occurred, as discussed in Cox [3].

<table>
<thead>
<tr>
<th>Failure Time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
</tr>
<tr>
<td>292</td>
</tr>
<tr>
<td>811</td>
</tr>
<tr>
<td>991</td>
</tr>
<tr>
<td>1489</td>
</tr>
<tr>
<td>94</td>
</tr>
<tr>
<td>332</td>
</tr>
<tr>
<td>899</td>
</tr>
<tr>
<td>1013</td>
</tr>
<tr>
<td>1512</td>
</tr>
<tr>
<td>196</td>
</tr>
<tr>
<td>347</td>
</tr>
<tr>
<td>945</td>
</tr>
<tr>
<td>1152</td>
</tr>
<tr>
<td>1525</td>
</tr>
<tr>
<td>268</td>
</tr>
<tr>
<td>544</td>
</tr>
<tr>
<td>950</td>
</tr>
<tr>
<td>1362</td>
</tr>
<tr>
<td>1539</td>
</tr>
<tr>
<td>290</td>
</tr>
<tr>
<td>732</td>
</tr>
<tr>
<td>955</td>
</tr>
<tr>
<td>1459</td>
</tr>
</tbody>
</table>

1. Estimate the GRP model parameters using the Type I virtual age option.
2. Plot the failure number and instantaneous failure intensity vs. time with 90% two-sided confidence bounds.
3. Plot the conditional reliability vs. time with 90% two-sided confidence bounds. The mission start time is 40 and mission time is varying.
4. Using the QCP, calculate the expected failure number and expected instantaneous failure intensity by time 1800.

**Solution**

Enter the data into a parametric RDA folio in Weibull++. On the control panel, select the 3 parameters option and the Type I setting. Keep the default simulation settings. Click **Calculate**.

1. The estimated parameters are \( \hat{\beta} = 1.1976, \hat{\lambda} = 4.94 \times 10^{-3}, \hat{q} = 0.1344 \).
2. The following plots show the cumulative number of failures and instantaneous failure intensity, respectively.
3. The following plot shows the conditional reliability.
4. Using the QCP, the failure number and instantaneous failure intensity are:
References

Chapter 21

Degradation Data Analysis

Given that products are more frequently being designed with higher reliability and developed in a shorter amount of time, it is often not possible to test new designs to failure under normal operating conditions. In some cases, it is possible to infer the reliability behavior of unfailed test samples with only the accumulated test time information and assumptions about the distribution. However, this generally leads to a great deal of uncertainty in the results. Another option in this situation is the use of degradation analysis. Degradation analysis involves the measurement of performance data that can be directly related to the presumed failure of the product in question. Many failure mechanisms can be directly linked to the degradation of part of the product, and degradation analysis allows the analyst to extrapolate to an assumed failure time based on the measurements of degradation over time.

In some cases, it is possible to directly measure the degradation of a physical characteristic over time, as with the wear of brake pads, the propagation of crack size, or the degradation of a performance characteristic over time such as the voltage of a battery or the luminous flux of an LED bulb. These cases belong to the Non-Destructive Degradation Analysis category. In other cases, direct measurement of degradation might not be possible without invasive or destructive measurement techniques that would directly affect the subsequent performance of the product; therefore, only one degradation measurement is possible. Examples are the measurement of corrosion in a chemical container or the strength measurement of an adhesive bond. These cases belong to the Destructive Degradation Analysis Category. In either case, however, it is necessary to be able to define a level of degradation or performance at which a failure is said to have occurred.

With this failure level defined, it is a relatively simple matter to use basic mathematical models to extrapolate the measurements over time to the point where the failure is said to occur. Once these have been determined, it is merely a matter of analyzing the extrapolated failure times in the same manner as conventional time-to-failure data.

Non-Destructive Degradation Analysis

The Non-Destructive Degradation Analysis applies to cases where multiple degradation measurements over time can be obtained for each sample in the test. Given a defined level of failure (or the degradation level that would constitute a failure), basic mathematical models are used to extrapolate the degradation measurements over time of each sample to the point in time where the failure will occur. Once these extrapolated failure times are obtained, it is merely a matter of analyzing the extrapolated failure times in the same manner as conventional time-to-failure data. As with conventional life data analysis, the amount of certainty in the results is directly related to the number of samples being tested. The following figure combines the steps of the analysis by showing the extrapolation of the degradation measurements to a failure time and the subsequent distribution analysis of these failure times.
Non-Destructive Degradation Models

Once the degradation information has been recorded, the next task is to extrapolate the measurements to the defined failure level in order to estimate the failure time. Weibull++ allows the user to perform such extrapolation using a linear, exponential, power or logarithmic model. These models have the following forms:

- Linear: \( y = a \cdot x + b \)
- Exponential: \( y = b \cdot e^{a \cdot x} \)
- Power: \( y = b \cdot x^a \)
- Logarithmic: \( y = a \cdot \ln(x) + b \)
- Gompertz: \( y = a \cdot b^x \)
- Lloyd-Lipow: \( y = a - \frac{b}{x} \)

where \( y \) represents the performance, \( x \) represents time, and \( a, b \) and \( c \) are model parameters to be solved for.

Once the model parameters \( a_i, b_i \) and \( c_i \) are estimated for each sample \( i \), a time \( x_i \) can be extrapolated, which corresponds to the defined level of failure \( y \). The computed \( x_i \) values can now be used as our times-to-failure for subsequent life data analysis. As with any sort of extrapolation, one must be careful not to extrapolate too far beyond the actual range of data in order to avoid large inaccuracies (modeling errors).
Example

Crack Propagation Example (Point Estimation)

Five turbine blades are tested for crack propagation. The test units are cyclically stressed and inspected every 100,000 cycles for crack length. Failure is defined as a crack of length 30mm or greater. The following table shows the test results for the five units at each cycle:

<table>
<thead>
<tr>
<th>Cycles (x1000)</th>
<th>Unit A (mm)</th>
<th>Unit B (mm)</th>
<th>Unit C (mm)</th>
<th>Unit D (mm)</th>
<th>Unit E (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>15</td>
<td>10</td>
<td>17</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>15</td>
<td>25</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>300</td>
<td>22</td>
<td>20</td>
<td>26</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>400</td>
<td>26</td>
<td>25</td>
<td>27</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td>500</td>
<td>29</td>
<td>30</td>
<td>33</td>
<td>26</td>
<td>33</td>
</tr>
</tbody>
</table>

Use the exponential degradation model to extrapolate the times-to-failure data.

Solution

The first step is to solve the equation \( y = b \cdot e^{a_x} \) for \( a \) and \( b \) for each of the test units. Using regression analysis, the values for each of the test units are:

- Unit A: \( a = 0.00158 \), \( b = 13.596 \)
- Unit B: \( a = 0.00271 \), \( b = 8.272 \)
- Unit C: \( a = 0.00140 \), \( b = 16.435 \)
- Unit D: \( a = 0.00177 \), \( b = 10.361 \)
- Unit E: \( a = 0.00294 \), \( b = 7.931 \)

Substituting the values into the underlying exponential model, solve for \( x \):

\[
x = \frac{\ln(y) - \ln(b)}{a}
\]

Using the values of \( a \) and \( b \), with \( y = 30 \), the resulting time at which the crack length reaches 30mm can then be found for each sample:

<table>
<thead>
<tr>
<th>Cycles – to – Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit A: 500,622</td>
</tr>
<tr>
<td>Unit B: 475,739</td>
</tr>
<tr>
<td>Unit C: 428,739</td>
</tr>
<tr>
<td>Unit D: 600,810</td>
</tr>
<tr>
<td>Unit E: 452,832</td>
</tr>
</tbody>
</table>

These times-to-failure can now be analyzed using traditional life data analysis to obtain metrics such as the probability of failure, B10 life, mean life, etc. This analysis can be automatically performed in the Weibull++ degradation analysis folio.

More degradation analysis examples are available! See also:

- Degradation Analysis \[1\]
- Watch the video... \[2\]

Using Extrapolated Intervals

The parameters in a degradation model are estimated using available degradation data. If the data is large, the uncertainty of the estimated parameters will be small; otherwise, the uncertainty will be large. Since the failure time for a test unit is predicted based on the estimated model, we sometimes would like to see how the parameter uncertainty affects the failure time prediction. Let’s use the exponential model as an example. Assume that the critical degradation value is \( y_{\text{crit}} \). The predicted failure time will be:
\[ \hat{x} = \frac{\ln(y_{\text{crit}}) - \ln(\hat{b})}{\hat{a}} \]

The variance of the predicted failure time will be:

\[ \text{Var}(\hat{x}) = \left( \frac{\partial x}{\partial \hat{a}} \right)^2 \text{Var}(\hat{a}) + 2 \left( \frac{\partial x}{\partial \hat{a}} \right) \left( \frac{\partial x}{\partial \hat{b}} \right) \text{Cov}(\hat{a}, \hat{b}) + \left( \frac{\partial x}{\partial \hat{b}} \right)^2 \text{Var}(\hat{b}) \]

The variance and covariance of the model parameters are calculated from using Least Squares Estimation. The details of the calculation are not given here.

The 2-sided upper and lower bounds for the predicted failure time, with a confidence level of \( 1 - \alpha \) are:

\[ x_U = \hat{x} + K_{1-\alpha/2} \sqrt{\text{Var}(\hat{x})} \]
\[ x_L = \hat{x} - K_{1-\alpha/2} \sqrt{\text{Var}(\hat{x})} \]

In Weibull++, the default confidence level is 90%.

**Example**

**Crack Propagation Example (Extrapolated Intervals)**

Using the same data set from the previous example, predict the interval failure times for the turbine blades.

**Solution**

In the Weibull++ degradation analysis folio, select the **Use extrapolated intervals** check box, as shown next.

![Image of Weibull++ interface showing extrapolated intervals settings](image-url)

Use the exponential degradation model for the degradation analysis, and the Weibull distribution parameters with MLE for the life data analysis. The following report shows the estimated degradation model parameters.
The following report shows the extrapolated failure time intervals.

**Destructive Degradation Analysis**

The Destructive Degradation Analysis applies to cases where the sample has to be destroyed in order to obtain a degradation measurement. As a result, degradation measurements for multiple samples are required at different points in time. The analysis performed is very similar to Accelerated Life Testing Analysis (ALTA). In this case, the "stress" used in ALTA becomes time, while the random variable becomes the degradation measurement instead of the time-to-failure. Given a defined level of failure (or the degradation level that would constitute a failure), the probability that the degradation measurement will be beyond that level at a given time can be obtained. The following plot shows the relationship between the distribution of the degradation measurement and time. The red shaded area of the last two pdfs represents the probability that the degradation measurement will be less than the critical degradation level at the corresponding times.
Destructive Degradation Models

The first step of destructive degradation analysis involves using a statistical distribution to represent the variability of a degradation measurement at a given time. The following distributions can be used:

- Weibull
- Exponential
- Normal
- Lognormal
- Gumbel

Similar to accelerated life testing analysis, the assumption is that the location or log-location parameter of the degradation measurement distribution will change with time while the shape parameter will remain constant. For each distribution:

- Weibull: \( \ln(\eta) \) is set as a function of time while \( \beta \) remains constant.
- Exponential: \( \ln(\text{MTTF}) \) is set as a function of time.
- Normal: \( \mu \) is set as a function of time while \( \sigma \) remains constant.
- Lognormal: \( \mu \) is set as a function of time while \( \sigma \) remains constant.
- Gumbel: \( \mu \) is set as a function of time while \( \sigma \) remains constant.

Finally, given the selected distribution, a degradation model is used to represent the change of the location (or log-location) parameter with time. The following degradation models can be used:

- Linear: \( \mu(t) = b + a \times t \)
- Exponential: \( \mu(t) = b \times e^{a \times t} \)
- Power: \( \mu(t) = b \times t^a \)
- Logarithm: \( \mu(t) = a \times \ln(t) + b \)
- Lloyd-Lipow: \( \mu(t) = a - \frac{b}{t} \)
The distribution and degradation models parameters are then calculated using Maximum Likelihood Estimation (MLE).

For example, if a normal distribution is used to represent the degradation measurement and a linear degradation model is assumed, then the standard deviation, $\sigma$, will be assumed constant with time, and the mean, $\mu$, will be:

$$\mu(t) = b + a \times t$$

The CDF of the degradation measurement $x(t)$ is:

$$\Pr(x(t) < X) = \Phi \left( \frac{X - \mu(t)}{\sigma} \right)$$

Given the CDF, the parameters $\sigma$, $b$ and $a$ are estimated using Maximum Likelihood Estimation.

Assuming that the requirement is that the measurement needs to be greater than a critical degradation value $D_{\text{crit}}$ for the product to fail, the probability of failure at time $t$ will be:

$$F(t) = \Pr(x(t) > D_{\text{crit}}) = 1 - \Phi \left( \frac{D_{\text{crit}} - \mu(t)}{\sigma} \right)$$

It should be noted that the failure threshold could be specified as a degradation measurement less than (for the case of decreasing degradation) or greater than (for the case of increasing degradation) a critical degradation value.

The relationship between the distribution of degradation measurement and the distribution of failure time is illustrated in the following plot.
Example

A company has been collecting degradation data over a period of 4 years with the purpose of calculating reliability after 5 years. With the degradation measurement decreasing with time, failure is defined as a measurement of 150 or below. In order to obtain such measurements, the unit has to be destroyed and therefore removed from the population.

The following table gives the degradation measurements over 4 years.

<table>
<thead>
<tr>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>437</td>
<td>412</td>
<td>246</td>
<td>125</td>
</tr>
<tr>
<td>446</td>
<td>420</td>
<td>324</td>
<td>208</td>
</tr>
<tr>
<td>497</td>
<td>451</td>
<td>330</td>
<td>229</td>
</tr>
<tr>
<td>503</td>
<td>454</td>
<td>426</td>
<td>242</td>
</tr>
<tr>
<td>705</td>
<td>554</td>
<td>499</td>
<td>273</td>
</tr>
<tr>
<td>737</td>
<td>580</td>
<td>546</td>
<td>297</td>
</tr>
<tr>
<td>748</td>
<td>608</td>
<td>554</td>
<td>311</td>
</tr>
<tr>
<td>788</td>
<td>610</td>
<td>559</td>
<td>318</td>
</tr>
<tr>
<td>818</td>
<td>727</td>
<td>625</td>
<td>393</td>
</tr>
<tr>
<td>860</td>
<td>825</td>
<td>403</td>
<td></td>
</tr>
<tr>
<td>875</td>
<td>925</td>
<td>470</td>
<td></td>
</tr>
<tr>
<td>934</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1350</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do the following:

1. Estimate parameters using the Linear degradation model and the 2-Parameter Weibull distribution
2. Plot the degradation curve vs. time
3. Calculate the reliability at 5 years.

Solution

Enter the data into a destructive degradation folio in Weibull++. Select Linear under the Degradation Model and 2P-Weibull under the Measurement Distribution. Set the critical degradation to 150. Click Calculate.

1. The estimated parameters are $\hat{\beta} = 3.618336$, $\hat{\alpha} = -0.331695$ and $\hat{b} = 7.154643$.
2. The following plot shows the degradation curve vs. time
3. Using the QCP, the reliability at 5 years, or in other words the probability that the degradation measurement will be less than 150 at 5 years, is: R(t=5) = 0.841481
References

Chapter 22

Reliability Test Design

This chapter discusses several methods for designing reliability tests. This includes:

- **Reliability Demonstration Tests (RDT):** Often used to demonstrate if the product reliability can meet the requirement. For this type of test design, four methods are supported in Weibull++:
  - **Parametric Binomial:** Used when the test duration is different from the time of the required reliability. An underlying distribution should be assumed.
  - **Non-Parametric Binomial:** No distribution assumption is needed for this test design method. It can be used for one shot devices.
  - **Exponential Chi-Squared:** Designed for exponential failure time.
  - **Non-Parametric Bayesian:** Integrated Bayesian theory with the traditional non-parametric binomial method.
- **Expected Failure Times Plot:** Can help the engineer determine the expected test duration when the total sample size is known and the allowed number of failures is given.
- **Difference Detection Matrix:** Can help the engineer design a test to compare the BX% life or mean life of two different designs/products.
- **Simulation:** Simulation can be used to help the engineer determine the sample size, test duration or expected number of failures in a test. To determine these variables, analytical methods need to make assumptions such as the distribution of model parameters. The simulation method does not need any assumptions. Therefore, it is more accurate than the analytical method, especially when the sample size is small.

Readers may also be interested in test design methods for quantitative accelerated life tests. That topic is discussed in the Accelerated Life Testing Reference.

**Reliability Demonstration Tests**

Frequently, a manufacturer will have to demonstrate that a certain product has met a goal of a certain reliability at a given time with a specific confidence. Several methods have been designed to help engineers: Cumulative Binomial, Non-Parametric Binomial, Exponential Chi-Squared and Non-Parametric Bayesian. They are discussed in the following sections.

**Cumulative Binomial**

This methodology requires the use of the cumulative binomial distribution in addition to the assumed distribution of the product's lifetimes. Not only does the life distribution of the product need to be assumed beforehand, but a reasonable assumption of the distribution's shape parameter must be provided as well. Additional information that must be supplied includes: a) the reliability to be demonstrated, b) the confidence level at which the demonstration takes place, c) the acceptable number of failures and d) either the number of available units or the amount of available test time. The output of this analysis can be the amount of time required to test the available units or the required number of units that need to be tested during the available test time. Usually the engineer designing the test will have to study the financial trade-offs between the number of units and the amount of test time needed to demonstrate the desired goal. In cases like this, it is useful to have a "carpet plot" that shows the possibilities of how a certain specification can be met.
Test to Demonstrate Reliability

Frequently, the entire purpose of designing a test with few or no failures is to demonstrate a certain reliability, \( R_{DEMO} \), at a certain time. With the exception of the exponential distribution (and ignoring the location parameter for the time being), this reliability is going to be a function of time, a shape parameter and a scale parameter.

\[ R_{DEMO} = g(t_{DEMO}; \theta, \phi) \]

where:

- \( t_{DEMO} \) is the time at which the demonstrated reliability is specified.
- \( \theta \) is the shape parameter.
- \( \phi \) is the scale parameter.

Since required inputs to the process include \( R_{DEMO} \), \( t_{DEMO} \), and \( \theta \), the value of the scale parameter can be backed out of the reliability equation of the assumed distribution, and will be used in the calculation of another reliability value, \( R_{TEST} \), which is the reliability that is going to be incorporated into the actual test calculation. How this calculation is performed depends on whether one is attempting to solve for the number of units to be tested in an available amount of time, or attempting to determine how long to test an available number of test units.

Determining Units for Available Test Time

If one knows that the test is to last a certain amount of time, \( t_{TEST} \), the number of units that must be tested to demonstrate the specification must be determined. The first step in accomplishing this involves calculating the \( R_{TEST} \) value.

This should be a simple procedure since:

\[ R_{TEST} = g(t_{TEST}; \theta, \phi) \]

and \( t_{DEMO} \), \( \theta \), and \( \phi \) are already known, and it is just a matter of plugging these values into the appropriate reliability equation.

We now incorporate a form of the cumulative binomial distribution in order to solve for the required number of units. This form of the cumulative binomial appears as:

\[ 1 - CL = \sum_{i=0}^{f} \frac{n!}{i!(n-i)!} \cdot (1 - R_{TEST})^i \cdot R_{TEST}^{(n-i)} \]

where:

- \( CL \) = the required confidence level
- \( f \) = the allowable number of failures
- \( n \) = the total number of units on test
- \( R_{TEST} \) = the reliability on test

Since \( CL \) and \( f \) are required inputs to the process and \( R_{TEST} \) has already been calculated, it merely remains to solve the cumulative binomial equation for \( n \), the number of units that need to be tested.

Determining Test Time for Available Units

The way that one determines the test time for the available number of test units is quite similar to the process described previously. In this case, one knows beforehand the number of units, \( n \), the number of allowable failures, \( f \), and the confidence level, \( CL \). With this information, the next step involves solving the binomial equation for \( R_{TEST} \). With this value known, one can use the appropriate reliability equation to back out the value of \( t_{TEST} \), since \( R_{TEST} = g(t_{TEST}; \theta, \phi) \), and \( R_{TEST} \), \( \theta \) and \( \phi \) have already been calculated or specified.
Example

In this example, we will use the parametric binomial method to design a test to demonstrate a reliability of 90% at $t_{DEMO} = 100\text{ hours}$ with a 95% confidence if no failure occur during the test. We will assume a Weibull distribution with a shape parameter $\beta = 1.5$.

Determining Units for Available Test Time

In the above scenario, we know that we have the testing facilities available for $t = 48\text{ hours}$. We must now determine the number of units to test for this amount of time with no failures in order to have demonstrated our reliability goal. The first step is to determine the Weibull scale parameter, $\eta$. The Weibull reliability equation is:

$$R = e^{-(t/\eta)^\beta}$$

This can be rewritten as:

$$\eta = \frac{t_{DEMO}}{(-\ln(R_{DEMO}))^{\frac{1}{\beta}}}$$

Since we know the values of $t_{DEMO}$, $R_{DEMO}$ and $\beta$, we can substitute these in the equation and solve for $\eta$:

$$\eta = \frac{100}{(-\ln(0.9))^{\frac{1}{1.5}}} = 448.3$$

Next, the value of $R_{TEST}$ is calculated by:

$$R_{TEST} = e^{-(t_{TEST}/\eta)^\beta} = e^{-(48/448.3)^1.5} = 0.966 = 96.6\%$$

The last step is to substitute the appropriate values into the cumulative binomial equation, which for the Weibull distribution appears as:

$$1 - CL = \sum_{i=0}^{f} \frac{n!}{i! \cdot (n-i)!} \cdot (1 - e^{-(t_{TEST}/\eta)^\beta})^i \cdot (e^{-(t_{TEST}/\eta)^\beta})^{(n-i)}$$

The values of $CL$, $t_{TEST}$, $\beta$, $f$ and $\eta$ have already been calculated or specified, so it merely remains to solve the equation for $n$. This value is $n = 85.4994$, or $n = 86\text{ units}$, since the fractional value must be rounded up to the next integer value. This example solved in Weibull++ is shown next.
Determining Time for Available Units

In this case, we will assume that we have 20 units to test, $n = 20$, and must determine the test time, $t_{TEST}$. We have already determined the value of the scale parameter, $\eta$, in the previous example. Since we know the values of $n, \alpha, f, \eta$, and $\beta$, it remains to solve the binomial equation with the Weibull distribution for $t_{TEST}$. This value is $t_{TEST} = 126.433$ hours. This example solved in Weibull++ is shown next.

Test to Demonstrate MTTF

Designing a test to demonstrate a certain value of the $MTTF$ is identical to designing a reliability demonstration test, with the exception of how the value of the scale parameter $\theta$ is determined. Given the value of the $MTTF$ and the value of the shape parameter $\beta$, the value of the scale parameter $\theta$ can be calculated. With this, the analysis can proceed as with the reliability demonstration methodology.

Example

In this example, we will use the parametric binomial method to design a test that will demonstrate $MTTF = 75$ hours with a 95% confidence if no failure occur during the test $f = 0$. We will assume a Weibull distribution with a shape parameter $\beta = 1.5$. We want to determine the number of units to test for $t_{TEST} = 60$ hours to demonstrate this goal.

The first step in this case involves determining the value of the scale parameter $\eta$ from the $MTTF$ equation. The equation for the $MTTF$ for the Weibull distribution is:

$$MTTF = \eta \cdot \Gamma\left(1 + \frac{1}{\beta}\right)$$

where $\Gamma(x)$ is the gamma function of $x$. This can be rearranged in terms of $\eta$:

$$\eta = \frac{MTTF}{\Gamma\left(1 + \frac{1}{\beta}\right)}$$

Since $MTTF$ and $\beta$ have been specified, it is a relatively simple matter to calculate $\eta = 83.1$. From this point on, the procedure is the same as the reliability demonstration example. Next, the value of $R_{TEST}$ is calculated as:
The last step is to substitute the appropriate values into the cumulative binomial equation. The values of $CL$, $t_{\text{TEST}}$, $\beta$, $f$, and $n$ have already been calculated or specified, so it merely remains to solve the binomial equation for $n$. The value is calculated as $n = 4.8811$ or $n = 5$ units, since the fractional value must be rounded up to the next integer value. This example solved in Weibull++ is shown next.

The procedure for determining the required test time proceeds in the same manner, determining $n$ from the $MTTF$ equation, and following the previously described methodology to determine $t_{\text{TEST}}$ from the binomial equation with Weibull distribution.

**Non-Parametric Binomial**

The binomial equation can also be used for non-parametric demonstration test design. There is no time value associated with this methodology, so one must assume that the value of $R_{\text{TEST}}$ is associated with the amount of time for which the units were tested.

In other words, in cases where the available test time is equal to the demonstration time, the following non-parametric binomial equation is widely used in practice:

$$1 - CL = \sum_{i=0}^{f} \binom{n}{i} (1 - R_{\text{TEST}})^i R_{\text{TEST}}^{n-i}$$

where $CL$ is the confidence level, $f$ is the number of failures, $n$ is the sample size and $R_{\text{TEST}}$ is the demonstrated reliability. Given any three of them, the remaining one can be solved for.

Non-parametric demonstration test design is also often used for one shot devices where the reliability is not related to time. In this case, $R_{\text{TEST}}$ can be simply written as $R$. 

$$R_{\text{TEST}} = e^{-\left(\frac{t_{\text{TEST}}}{\eta}\right)^{\beta}} = e^{-\left(\frac{60}{83.1}\right)^{1.5}} = 0.541 = 54.1\%$$
Example

A reliability engineer wants to design a zero-failure demonstration test in order to demonstrate a reliability of 80% at a 90% confidence level. Use the non-parametric binomial method to determine the required sample size.

Solution

By substituting \( f = 0 \) (since it is a zero-failure test) the non-parametric binomial equation becomes:

\[
1 - CL = R^n
\]

So now the required sample size can be easily solved for any required reliability and confidence level. The result of this test design was obtained using Weibull++ and is:

![Test Design Result](image)

The result shows that 11 samples are needed. Note that the time value shown in the above figure is chance indicative and not part of the test design (the "Test time per unit" that was input will be the same as the "Demonstrated at time" value for the results). If those 11 samples are run for the required demonstration time and no failures are observed, then a reliability of 80% with a 90% confidence level has been demonstrated. If the reliability of the system is less than or equal to 80%, the chance of passing this test is \( 1 - CL = 0.1 \), which is the Type II error. Therefore, the non-parametric binomial equation determines the sample size by controlling for the Type II error.

If 11 samples are used and one failure is observed by the end of the test, then the demonstrated reliability will be less than required. The demonstrated reliability is 68.98% as shown below.
Constant Failure Rate/Chi-Squared

Another method for designing tests for products that have an assumed constant failure rate, or exponential life distribution, draws on the chi-squared distribution. These represent the true exponential distribution confidence bounds referred to in The Exponential Distribution. This method only returns the necessary accumulated test time for a demonstrated reliability or $MTTF$, not a specific time/test unit combination that is obtained using the cumulative binomial method described above. The accumulated test time is equal to the total amount of time experienced by all of the units on test. Assuming that the units undergo the same amount of test time, this works out to be:

$$T_a = n \cdot t_{TEST}$$

where $n$ is the number of units on test and $t_{TEST}$ is the test time. The chi-squared equation for test time is:

$$T_a = \frac{MTTF \cdot \chi^2_{1-CL;2f+2}}{2}$$

where:

- $\chi^2_{1-CL;2f+2}$ is the chi-squared distribution
- $T_a$ = the necessary accumulated test time
- $CL$ = the confidence level
- $f$ = the number of failures

Since this methodology only applies to the exponential distribution, the exponential reliability equation can be rewritten as:

$$MTTF = \frac{t}{-\ln(R)}$$

and substituted into the chi-squared equation for developing a test that demonstrates reliability at a given time, rather than $MTTF$:

$$T_a = \frac{t_{DEMO} \cdot \chi^2_{1-CL;2f+2}}{2}$$
Example
In this example, we will use the exponential chi-squared method to design a test that will demonstrate a reliability of 85\% at \( t_{DEMO} = 500 \) hours with a 90\% confidence (or \( CL = 0.9 \)) if no more than 2 failures occur during the test (\( f = 2 \)). The chi-squared value can be determined from tables or the Quick Statistical Reference (QSR) tool in Weibull++. In this example, the value is calculated as:

\[
\chi^2_{1-CL,2f+2} = \chi^2_{0.1,6} = 10.6446
\]

Substituting this into the chi-squared equation, we obtain:

\[
T_a = \frac{500 \cdot 10.6446}{2} = 16,374 \text{ hours}
\]

This means that 16,374 hours of total test time needs to be accumulated with no more than two failures in order to demonstrate the specified reliability.

This example solved in Weibull++ is shown next.

Given the test time, one can now solve for the number of units using the chi-squared equation. Similarly, if the number of units is given, one can determine the test time from the chi-squared equation for exponential test design.

Bayesian Non-Parametric Test Design
The regular non-parametric analyses performed based on either the binomial or the chi-squared equation were performed with only the direct system test data. However, if prior information regarding system performance is available, it can be incorporated into a Bayesian non-parametric analysis. This subsection will demonstrate how to incorporate prior information about system reliability and also how to incorporate prior information from subsystem tests into system test design.

If we assume the system reliability follows a beta distribution, the values of system reliability, \( R \), confidence level, \( CL \), number of units tested, \( n \), and number of failures, \( r \), are related by the following equation:

\[
1 - CL = \text{Beta}(R, \alpha, \beta) = \text{Beta}(R, n - r + \alpha_0, r + \beta_0)
\]

where \( \text{Beta} \) is the incomplete beta function. If \( \alpha_0 > 0 \) and \( \beta_0 > 0 \) are known, then any quantity of interest can be calculated using the remaining three. The next two examples demonstrate how to calculate \( \alpha_0 > 0 \) and \( \beta_0 > 0 \) depending on the type of prior information available.
Use Prior Expert Opinion on System Reliability

Prior information on system reliability can be exploited to determine $\alpha_0$ and $\beta_0$. To do so, first approximate the expected value and variance of prior system reliability $R_0$. This requires knowledge of the lowest possible reliability, the most likely possible reliability and the highest possible reliability of the system. These quantities will be referred to as $a$, $b$ and $c$, respectively. The expected value of the prior system reliability is approximately given as:

$$E(R_0) = \frac{a + 4b + c}{6}$$

and the variance is approximately given by:

$$Var(R_0) = \left(\frac{c - a}{6}\right)^2$$

These approximate values of the expected value and variance of the prior system reliability can then be used to estimate the values of $\alpha_0$ and $\beta_0$, assuming that the prior reliability is a beta-distributed random variable. The values of $\alpha_0$ and $\beta_0$ are calculated as:

$$\alpha_0 = E(R_0) \left[\frac{E(R_0) - E^2(R_0)}{Var(R_0)} - 1\right]$$

$$\beta_0 = (1 - E(R_0)) \left[\frac{E(R_0) - E^2(R_0)}{Var(R_0)} - 1\right]$$

With $\alpha_0$ and $\beta_0$ known, the above beta distribution equation can now be used to calculate a quantity of interest.

Example

You can use the non-parametric Bayesian method to design a test using prior knowledge about a system’s reliability. For example, suppose you wanted to know the reliability of a system and you had the following prior knowledge of the system:

- Lowest possible reliability: $a = 0.8$
- Most likely reliability: $b = 0.85$
- Highest possible reliability: $c = 0.97$

This information can be used to approximate the expected value and the variance of the prior system reliability.

$$E(R_0) = \frac{a + 4b + c}{6} = 0.861667$$

$$Var(R_0) = \left(\frac{c - a}{6}\right)^2 = 0.000803$$

These approximations of the expected value and variance of the prior system reliability can then be used to estimate $\alpha_0$ and $\beta_0$ used in the beta distribution for the system reliability, as given next:

$$\alpha_0 = E(R_0) \left[\frac{E(R_0) - E^2(R_0)}{Var(R_0)} - 1\right] = 127.0794$$

$$\beta_0 = (1 - E(R_0)) \left[\frac{E(R_0) - E^2(R_0)}{Var(R_0)} - 1\right] = 20.40153$$

With $\alpha_0$ and $\beta_0$ known, any single value of the four quantities system reliability $R$, confidence level $CL$, number of units $n$, or number of failures $r$ can be calculated from the other three using the beta distribution function:

$$1 - CL = Beta(R, \alpha, \beta) = Beta(R, n - r + \alpha_0, r + \beta_0)$$

Solve for System Reliability $R$

Given $CL = 0.9$, $n = 20$, and $r = 1$, using the above prior information to solve $R$.

First, we get the number of successes: $s = n - r = 19$. Then the parameters in the posterior beta distribution for $R$ are calculated as:

$$\alpha = \alpha_0 + s = 146.0794$$

$$\beta = \beta_0 + r = 21.40153$$
Finally, from this posterior distribution, the system reliability $R$ at a confidence level of $CL=0.9$ is solved as:

$$R = \text{BetaINV}(1 - CL, \alpha, \beta) = 0.838374$$

**Solve for Confidence Level $CL$**

Given $R = 0.85$, $n = 20$, and $r = 1$, using the above prior information on system reliability to solve for $CL$.

First, we get the number of successes: $s = n - r = 19$. Then the parameters in the posterior beta distribution for $R$ are calculated as:

$$\alpha = \alpha_0 + s = 146.07943$$
$$\beta = \beta_0 + r = 21.40153$$

Finally, from this posterior distribution, the corresponding confidence level for reliability $R=0.85$ is:

$$CL = \text{Beta}(R, \alpha, \beta) = 0.81011$$

**Solve for Sample Size $n$**

Given $R = 0.9$, $CL = 0.8$, and $r = 1$, using the above prior information on system reliability to solve the required sample size in the demonstration test.

Again, the above beta distribution equation for the system reliability can be utilized. The figure below shows the result from Weibull++. The results show that the required sample size is 103. Weibull++ always displays the sample size as an integer.
Use Prior Information from Subsystem Tests

Prior information from subsystem tests can also be used to determine values of alpha and beta. Information from subsystem tests can be used to calculate the expected value and variance of the reliability of individual components, which can then be used to calculate the expected value and variance of the reliability of the entire system. \( \alpha_0 \) and \( \beta_0 \) are then calculated as before:

\[
\alpha_0 = E (R_0) \left[ \frac{E (R_0) - E^2 (R_0)}{Var (R_0)} - 1 \right]
\]

\[
\beta_0 = (1 - E (R_0)) \left[ \frac{E (R_0) - E^2 (R_0)}{Var (R_0)} - 1 \right]
\]

For each subsystem \( i \), from the beta distribution, we can calculate the expected value and the variance of the subsystem’s reliability \( R_i \), as discussed in Guo [38]:

\[
E (R_i) = \frac{s_i}{n_i + 1}
\]

\[
Var (R_i) = \frac{s_i (n_i + 1 - s_i)}{(n_i + 1)^2 (n_i + 2)}
\]

Assuming that all the subsystems are in a series reliability-wise configuration, the expected value and variance of the system’s reliability \( R \) can then be calculated as per Guo [38]:

\[
E (R_0) = (i = 1)^k E (R_i) = E (R_1) \times E (R_2) \ldots E (R_k)
\]

\[
Var (R_0) = \prod_{i=1}^{k} \left[ E^2 (R_i) + Var (R_i) \right] - \prod_{i=1}^{k} \left[ E^2 (R_i) \right]
\]

With the above prior information on the expected value and variance of the system reliability, all the calculations can now be calculated as before.

Example

You can use the non-parametric Bayesian method to design a test for a system using information from tests on its subsystems. For example, suppose a system of interest is composed of three subsystems A, B and C -- with prior information from tests of these subsystems given in the table below.

<table>
<thead>
<tr>
<th>Subsystem</th>
<th>Number of Units (n)</th>
<th>Number of Failures (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>100</td>
<td>4</td>
</tr>
</tbody>
</table>

This data can be used to calculate the expected value and variance of the reliability for each subsystem.

\[
E (R_i) = \frac{n_i - r_i}{n_i + 1}
\]

\[
Var (R_i) = \frac{(n_i - r_i) (r_i + 1)}{(n_i + 1)^2 (n_i + 2)}
\]

The results of these calculations are given in the table below.
These values can then be used to find the prior system reliability and its variance:

\[ E(R_0) = 0.846831227 \]
\[ \text{Var}(R_0) = 0.003546663 \]

From the above two values, the parameters of the prior distribution of the system reliability can be calculated by:

\[ \alpha_0 = E(R_0) \left[ \frac{E(R_0) - E^2(R_0)}{\text{Var}(R_0)} - 1 \right] = 30.12337003 \]
\[ \beta_0 = (1 - E(R_0)) \left[ \frac{E(R_0) - E^2(R_0)}{\text{Var}(R_0)} - 1 \right] = 5.448499634 \]

With this prior distribution, we now can design a system reliability demonstration test by calculating system reliability \( R \), confidence level \( CL \), number of units \( n \) or number of failures \( r \), as needed.

**Solve for Sample Size**

Given the above subsystem test information, in order to demonstrate the system reliability of 0.9 at a confidence level of 0.8, how many samples are needed in the test? Assume the allowed number of failures is 1.

Using Weibull++, the results are given in the figure below. The result shows that at least 49 test units are needed.
Expected Failure Times Plots

Test duration is one of the key factors that should be considered in designing a test. If the expected test duration can be estimated prior to the test, test resources can be better allocated. In this section, we will explain how to estimate the expected test time based on test sample size and the assumed underlying failure distribution.

The binomial equation used in non-parametric demonstration test design is the base for predicting expected failure times. The equation is:

\[ 1 - CL = \sum_{i=0}^{r} \frac{n!}{i! \cdot (n-i)!} \cdot (1 - R_{\text{TEST}})^i \cdot R_{\text{TEST}}^{n-i} \]

where:
- \( CL \) = the required confidence level
- \( r \) = the number of failures
- \( n \) = the total number of units on test
- \( R_{\text{TEST}} \) = the reliability on test

If \( CL, r \) and \( n \) are given, the \( R \) value can be solved from the above equation. When \( CL=0.5 \), the solved \( R \) (or \( Q \), the probability of failure whose value is \( 1-R \)) is the so called median rank for the corresponding failure. (For more information on median ranks, please see Parameter Estimation).

For example, given \( n = 4, r = 2 \) and \( CL = 0.5 \), the calculated \( Q \) is 0.385728. This means, at the time when the second failure occurs, the estimated system probability of failure is 0.385728. The median rank can be calculated in Weibull++ using the Quick Statistical Reference, as shown below:

Similarly, if we set \( r = 3 \) for the above example, we can get the probability of failure at the time when the third failure occurs. Using the estimated median rank for each failure and the assumed underlying failure distribution, we can calculate the expected time for each failure. Assume the failure distribution is Weibull, then we know:
\[ Q = 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}} \]

where:

- \( \beta \) is the shape parameter
- \( \eta \) is the scale parameter

Using the above equation, for a given \( Q \), we can get the corresponding time \( t \). The above calculation gives the median of each failure time for \( CL = 0.5 \). If we set \( CL \) at different values, the confidence bounds of each failure time can be obtained. For the above example, if we set \( CL = 0.9 \), from the calculated \( Q \) we can get the upper bound of the time for each failure. The calculated \( Q \) is given in the next figure:

If we set \( CL = 0.1 \), from the calculated \( Q \) we can get the lower bound of the time for each failure. The calculated \( Q \) is given in the figure below:
Example

In this example you will use the Expected Failure Times plot to estimate the duration of a planned reliability test. 4 units were allocated for the test, and the test engineers want to know how long the test will last if all the units are tested to failure. Based on previous experiments, they assume the underlying failure distribution is a Weibull distribution with $\beta = 2$ and $\eta = 500$.

Solution

Using Weibull++’s Expected Failure Times plot, the expected failure times with 80% 2-sided confidence bounds are given below.
From the above results, we can see the upper bound of the last failure is about 955 hours. Therefore, the test probably will last for around 955 hours.

As we know, with 4 samples, the median rank for the second failure is 0.385728. Using this value and the assumed Weibull distribution, the median value of the failure time of the second failure is calculated as:

\[ Q = 1 - e^{-\left(\frac{t}{\eta}\right)^\beta} \]

\[ \ln(1 - Q) = \left(\frac{t}{\eta}\right)^\beta \]

\[ \Rightarrow t = 349.04 \]

Its bounds and other failure times can be calculated in a similar way.

**Life Difference Detection Matrix**

Engineers often need to design tests for detecting life differences between two or more product designs. The questions are how many samples and how long should the test be conducted in order to detect a certain amount of difference. There are no simple answers. Usually, advanced design of experiments (DOE) techniques should be utilized. For a simple case, such as comparing two designs, the Difference Detection Matrix in Weibull++ can be used. The Difference Detection Matrix graphically indicates the amount of test time required to detect a statistical difference in the lives of two populations.

As discussed in the test design using Expected Failure Times plot, if the sample size is known, the expected failure time of each test unit can be obtained based on the assumed failure distribution. Now let's go one step further. With these failure times, we can then estimate the failure distribution and calculate any reliability metrics. This process is similar to the simulation used in SimuMatic where random failure times are generated from simulation and then used to estimate the failure distribution. This approach is also used by the Difference Detection Matrix.

Assume we want to compare the B10 lives (or mean lives) of two designs. The test is time-terminated and the termination time is set to \( T \). Using the method given in Expected Failure Times Plots, we can generate the failure times. For any failure time greater than \( T \), it is a suspension and the suspension time is \( T \). For each design, its B10 life and confidence bounds can be estimated from the generated failure/suspension times. If the two estimated confidence intervals overlap with each other, it means the difference of the two B10 lives cannot be detected from this test. We have to either increase the sample size or the test duration.

**Example**

In this example, you will use the Difference Detection Matrix to choose the suitable sample size and duration for a reliability test. Assume that there are two design options for a new product. The engineers need to design a test that compares the reliability performance of these two options. The reliability for both designs is assumed to follow a Weibull distribution. For Design 1, its shape parameter \( \beta = 3 \); for Design 2, its \( \beta = 2 \). Their B10 lives may range from 500 to 3,000 hours.

**Solution**

For the initial setup, set the sample size for each design to 20, and use two test durations of 3,000 and 5,000 hours. The following picture shows the complete control panel setup and the results of the analysis.
The columns in the matrix show the range of the assumed B10 life for design 1, while the rows show the range for design 2. A value of 0 means the difference cannot be detected through the test, 1 means the difference can be detected if the test duration is 5,000 hours, and 2 means the difference can be detected if the test duration is 3,000 hours. For example, the number is 2 for cell (1000, 2000). This means that if the B10 life for Design 1 is 1,000 hours and the B10 life for Design 2 is 2,000 hours, the difference can be detected if the test duration is at least 5,000 hours.

Click inside the cell to show the estimated confidence intervals, as shown next. By testing 20 samples each for 3,000 hours, the difference of their B10 lives probably can be detected. This is because, at a confidence level of 90%, the estimated confidence intervals on the B10 life do not overlap.
We will use Design 1 to illustrate how the interval is calculated. For cell (1000, 2000), Design 1’s B10 life is 1,000 and the assumed $\beta$ is 3. We can calculate the $\eta$ for the Weibull distribution using the Quick Parameter Estimator tool, as shown next.
The estimated $T$ is 2117.2592 hours. We can then use these distribution parameters and the sample size of 20 to get the expected failure times by using Weibull's Expected Failure Times Plot. The following report shows the result from that utility.

![Test Results](image)

The median failure times are used to estimate the failure distribution. Note that since the test duration is set to 3,000 hours, any failures that occur after 3,000 are treated as suspensions. In this case, the last failure is a suspension with a suspension time of 3,000 hours. We can enter the median failure times data set into a standard Weibull++ folio as given in the next figure.
After analyzing the data set with the MLE and FM analysis options, we can now calculate the B10 life and its interval in the QCP, as shown next.

From this result, we can see that the estimated B10 life and its confidence intervals are the same as the results displayed in the Difference Detection Matrix.

The above procedure can be repeated to get the results for the other cells and for Design 2. Therefore, by adjusting the sample size and test duration, a suitable test time can be identified for detecting a certain amount of difference between two designs/populations.
Simulation

Monte Carlo simulation provides another useful tool for test design. The SimuMatic utility in Weibull++ can be used for this purpose. SimuMatic is simulating the outcome from a particular test design that is intended to demonstrate a target reliability. You can specify various factors of the design, such as the test duration (for a time-terminated test), number of failures (for a failure-terminated test) and sample size. By running the simulations you can assess whether the planned test design can achieve the reliability target. Depending on the results, you can modify the design by adjusting these factors and repeating the simulation process—in effect, simulating a modified test design—until you arrive at a modified design that is capable of demonstrating the target reliability within the available time and sample size constraints.

Of course, all the design factors mentioned in SimuMatic also can be calculated using analytical methods as discussed in previous sections. However, all of the analytical methods need assumptions. When sample size is small or test duration is short, these assumptions may not be accurate enough. The simulation method usually does not require any assumptions. For example, the confidence bounds of reliability from SimuMatic are purely based on simulation results. In analytical methods, both Fisher bounds and likelihood ratio bounds need to use assumptions. Another advantage of using the simulation method is that it is straightforward and results can be visually displayed in SimuMatic.

For details, see the Weibull++ SimuMatic chapter.
Chapter 23

Stress-Strength Analysis

Stress-strength analysis has been used in mechanical component design. The probability of failure is based on the probability of stress exceeding strength. The following equation is used to calculate the expected probability of failure, $F$:

$$F = P[\text{Stress} \geq \text{Strength}] = \int_0^\infty f_{\text{Strength}}(x) \cdot R_{\text{Strength}}(x) \, dx$$

The expected probability of success or the expected reliability, $R$, is calculated as:

$$R = P[\text{Stress} \leq \text{Strength}] = \int_0^\infty f_{\text{Strength}}(x) \cdot R_{\text{Strength}}(x) \, dx$$

The equations above assume that both stress and strength are in the positive domain. For general cases, the expected reliability can be calculated using the following equation:

$$R = P[X_1 \leq X_2] = \frac{1}{F_1(U) - F_1(L)} \int_L^U f_1(x) \cdot R_2(x) \, dx$$

where:

$L \leq X_1 \leq U$

$X_1$: Stress

$X_2$: Strength

When $U = \infty$ and $L = 0$, this equation becomes equal to previous equation (i.e., the equation for the expected reliability $R$).

Confidence Intervals on the Probability

Both the stress and strength distributions can be estimated from actual data or specified by engineers based on engineering knowledge or existing references. Based on the source of the distribution, there are two types of variation associated with the calculated probability: variation in the model parameters and variation in the probability values. Both are described next.

Variation in Model Parameters

If both the stress and strength distributions are estimated from data sets, then there are uncertainties associated with the estimated distribution parameters. These uncertainties will cause some degree of variation of the probability calculated from the stress-strength analysis. Therefore, we can use these uncertainties to estimate the confidence intervals on the calculated probability. To get the confidence intervals, we first calculate the variance of the reliability based on Taylor expansion by ignoring the 2nd order term. The approximation for the variance is:

$$\text{Var} \left[ R \right] = \int_0^\infty \text{Var} \left[ f_1(x) \right] \left[ R_2(x) \right]^2 \, dx + \int_0^\infty \left[ f_1(x) \right]^2 \text{Var} \left[ R_2(x) \right] \, dx$$

Variance of $f_1(x)$ and $R_2(x)$ can be estimated from the Fisher Information Matrix. For details, please see Confidence Bounds.

Once the variance of the expected reliability is obtained, the two-sided confidence intervals can be calculated using:

$$\left[ \frac{R}{R + (1 - R) \bar{w}}, \frac{R}{R + (1 - R) / \bar{w}} \right]$$
where:

- \(CL\) is the confidence level
- \(\alpha\) is \(1-CL\)

\[
w = \exp\left\{ z_{1-\alpha/2} \sqrt{\text{Var}(R)}/[R(1-R)] \right\}
\]

\(Z_{1-\alpha/2}\) is the \(1-\alpha/2\) percentile of a standard normal distribution.

If the upper bound (U) and lower bound (L) are not infinite and 0, respectively, then the calculated variance of \(R\) is adjusted by \([1/(F_1(U) - F_1(L))]^2\).

### Variation in Probability Values

Assume the distributions for stress and strength are known. From the stress-strength equation

\[
R = P[\text{Stress} \leq \text{Strength}] = \int_0^\infty f_{\text{Stress}}(x) \cdot R_{\text{Strength}}(x) \, dx
\]

we know that the calculated reliability is the expected value of the probability that a strength value is larger than a stress value. Since both strength and stress are random variables from their distributions, the reliability is also a random variable. This can be explained using the following example. Let's first assume that stress is a fixed value of 567. The reliability then is:

\[
R(567) = \Pr(\text{Strength} > 567) = R_2(567)
\]

This is the reliability when the stress value is 567 and when the strength distribution is given. If stress is not a fixed value (i.e., it follows a distribution instead), then it can take values other than 567. For instance, if stress takes a value of 700, then we get another reliability value of \(R(700)\). Since stress is a random variable, for any stress value \(x_i\), there is a reliability value of \(R(x_i)\) calculated from the strength distribution. We will end up with many \(R(x_i)\)s or \(R_2(x_i)\)s. From these \(R(x_i)\)s, we can get the mean and variance of the reliability. In fact, its mean is the result from:

\[
R = \int_0^\infty f_{\text{Stress}}(x) \cdot R_{\text{Strength}}(x) \, dx
\]

and its variance is:

\[
\text{Var}[R] = \text{Var}[R_2(X_1)] = E\left[ (R_2(X_1))^2 \right] - (E[R_2(X_1)])^2
\]

\[
= \int_0^\infty f_1(x) [R_2(x)]^2 \, dx - (E[R_2(X_1)])^2
\]

\[
= \int_0^\infty f_1(x) [R_2(x)]^2 \, dx - (R)^2
\]

where \(R\) is the expected value of the reliability.

Once the variance of the expected reliability is obtained, the two-sided confidence intervals can be calculated using:

\[
\frac{R}{R + (1-R)w} \leq R \leq \frac{R}{R + (1-R)/w}
\]

where:

- \(CL\) is the confidence level
- \(\alpha\) is \(1-CL\)

\[
w = \exp\left\{ z_{1-\alpha/2} \sqrt{\text{Var}(R)}/[R(1-R)] \right\}
\]

\(Z_{1-\alpha/2}\) is the \(1-\alpha/2\) percentile of a standard normal distribution.

If the upper bound (U) and lower bound (L) are not infinite and 0, the above calculated variance of \(R\) is adjusted by \([1/(F_1(U) - F_1(L))]^2\).
Example 1

Assume that we are going to use stress-strength analysis to estimate the reliability of a component used in a vehicle. The stress is the usage mileage distribution and the strength is the miles-to-failure distribution of the component. The warranty is 1 year or 15,000 miles, whichever is earlier. The goal is to estimate the reliability of the component within the warranty period (1 year/15,000 miles).

The following table gives the data for the mileage distribution per year (stress):

<table>
<thead>
<tr>
<th>Stress: Usage Mileage Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>10096</td>
</tr>
<tr>
<td>10469</td>
</tr>
<tr>
<td>10955</td>
</tr>
<tr>
<td>11183</td>
</tr>
<tr>
<td>11391</td>
</tr>
<tr>
<td>11486</td>
</tr>
<tr>
<td>11534</td>
</tr>
<tr>
<td>11919</td>
</tr>
<tr>
<td>12105</td>
</tr>
<tr>
<td>12141</td>
</tr>
</tbody>
</table>

The following table gives the data for the miles-to-failure distribution (strength):

<table>
<thead>
<tr>
<th>Strength: Failure Mileage Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>13507</td>
</tr>
<tr>
<td>13793</td>
</tr>
<tr>
<td>13943</td>
</tr>
<tr>
<td>14017</td>
</tr>
<tr>
<td>14147</td>
</tr>
<tr>
<td>14351</td>
</tr>
<tr>
<td>14376</td>
</tr>
<tr>
<td>14595</td>
</tr>
<tr>
<td>14746</td>
</tr>
<tr>
<td>14810</td>
</tr>
<tr>
<td>14940</td>
</tr>
<tr>
<td>14951</td>
</tr>
<tr>
<td>15104</td>
</tr>
<tr>
<td>15218</td>
</tr>
<tr>
<td>15303</td>
</tr>
<tr>
<td>15311</td>
</tr>
<tr>
<td>15480</td>
</tr>
<tr>
<td>15496</td>
</tr>
<tr>
<td>15522</td>
</tr>
<tr>
<td>15547</td>
</tr>
</tbody>
</table>
Solution

First, estimate the stress and strength distributions using the given data. Enter the stress and strength data into two separate data sheets and analyze each data sheet using the lognormal distribution and MLE analysis method. The parameters of the stress distribution are estimated to be log-mean = 9.411844 and log-std = 0.098741.

The parameters of the strength distribution are estimated to be log-mean = 9.681503 and log-std = 0.083494.
Next, open the Stress-Strength tool and choose to compare the two data sheets. The following picture shows the pdf curves of the two data sets:
Since the warranty is 1 year/15,000 miles, all the vehicles with mileage larger than 15,000 should not be considered in the calculation. To do this, go to the Setup page of the control panel and select the **Override auto-calculated limits** check box. Set the value of the upper limit to **15,000** as shown next.
Recalculate the results. The estimated reliability for vehicles less than 15,000 miles per year is 98.84%. The associated confidence bounds are estimated from the variance of the distribution parameters. With larger samples for the stress and strength data, the width of the bounds will be narrower.
Stress-Strength Analysis in Design for Reliability

As we know, the expected reliability is called from the following stress-strength calculation:

\[ R = P[Stress \leq Strength] = \int_{0}^{\infty} f_{Stress}(x) \cdot R_{Strength}(x) \, dx \]

The stress distribution is usually estimated from customer usage data, such as the mileage per year of a passenger car or the load distribution for a beam. The strength distribution, on the other hand, is affected by the material used in the component, the geometric dimensions and the manufacturing process.

Because the stress distribution can be estimated from customer usage data, we can assume that \( f_{Stress} \) is known. Therefore, for a given reliability goal, the strength distribution \( R_{Strength} \) is the only unknown in the given equation. The factors that affect the strength distribution can be adjusted to obtain a strength distribution that meets the reliability goal. The following example shows how to use the Target Reliability Parameter Estimator tool in the stress-strength folio to obtain the parameters for a strength distribution that will meet a specified reliability goal.
Example 2

Assume that the stress distribution for a component is known to be a Weibull distribution with beta = 3 and eta = 2000. For the current design, the strength distribution is also a Weibull distribution with beta = 1.5 and eta = 4000. Evaluate the current reliability of the component. If the reliability does not meet the target reliability of 90%, determine what parameters would be required for the strength distribution in order to meet the specified target.

Solution

The following picture shows the stress-strength tool and the calculated reliability of the current design.

The result shows that the current reliability is about 74.0543%, which is below the target value of 90%. We need to use the Target Reliability Parameter Estimator to determine the parameters for the strength distribution that, when compared against the stress distribution, would result in the target reliability.

The following picture shows the Target Reliability Parameter Estimator window. In the Strength Parameters area, select eta. Set the Target Reliability to 90% and click Calculate. The calculated eta is 8192.2385 hours.
Click **Update** to perform the stress-strength analysis again using the altered parameters for the strength distribution. The following plot shows that the calculated reliability is 90%. Therefore, in order to meet the reliability requirement, the component must be redesigned such that the eta parameter of the strength distribution is at least 8192.2385 hours.
Comparing Life Data Sets

It is often desirable to be able to compare two sets of reliability or life data in order to determine which of the data sets has a more favorable life distribution. The data sets could be from two alternate designs, manufacturers, lots, assembly lines, etc. Many methods are available in statistical literature for doing this when the units come from a complete sample (i.e., a sample with no censoring). This process becomes a little more difficult when dealing with data sets that have censoring, or when trying to compare two data sets that have different distributions. In general, the problem boils down to that of being able to determine any statistically significant difference between the two samples of potentially censored data from two possibly different populations. This section discusses some of the methods available in Weibull++ that are applicable to censored data.

Simple Plotting

One popular graphical method for making this determination involves plotting the data with confidence bounds and seeing whether the bounds overlap or separate at the point of interest. This can be easily done using the Overlay Plot feature in Weibull++. This approach can be effective for comparisons at a given point in time or a given reliability level, but it is difficult to assess the overall behavior of the two distributions because the confidence bounds may overlap at some points and be far apart at others.

Contour Plots

To determine whether two data sets are significantly different and at what confidence level, one can utilize the contour plots provided in Weibull++. By overlaying two contour plots from two different data sets at the same confidence level, one can visually assess whether the data sets are significantly different at that confidence level if there is no overlap on the contours. The disadvantage of this method is that the same distribution must be fitted to both data sets.

Example: Using a Contour Plot to Compare Two Designs

The design of a product was modified to improve its reliability. The reliability engineers want to determine whether the improvements to the design have significantly improved the product's reliability. The following data sets represent the times-to-failure for the product. At what significance level can the engineers claim that the two designs are different?

<table>
<thead>
<tr>
<th>Old Design</th>
<th>2</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>11</td>
<td>17</td>
<td>17</td>
<td>19</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>28</td>
<td>33</td>
<td>34</td>
<td>34</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>40</td>
<td>45</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>76</td>
<td>90</td>
<td>115</td>
<td>126</td>
<td>197</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>New Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
</tr>
<tr>
<td>116</td>
</tr>
</tbody>
</table>

The data sets are entered into separate Weibull++ standard folio data sheets, and then analyzed with the two-parameter Weibull distribution and the maximum likelihood estimation (MLE) method. The following figure shows the contour plots of the data sets superimposed in an overlay plot. This plot is configured to show the contour lines that represent the 90% and 95% confidence levels.
As you can see, the contours overlap at the 95% confidence level (outer rings), but there is no overlap at the 90% confidence level (inner rings). We can then conclude that there is a statistically significant difference between the data sets at the 90% confidence level. If we wanted to know the exact confidence level (i.e., critical confidence level) at which the two contour plots meet, we would have to incrementally raise the confidence level from 90% until the two contour lines meet.

Weibull++ includes a utility for automatically obtaining the critical confidence level. For two contour plots that are superimposed in an overlay plot, the **Plot Critical Level** check box will be available in the Contours Setup window, as shown next.
Comparing Life Data Sets

The plot critical level is the confidence level at which the contour plots of the two data sets meet at a single point. This is the minimum confidence level at which the contour lines of the two different data sets overlap. At any confidence level below this minimum confidence level, the contour lines of the two data sets will not overlap and there will be a statistically significant difference between the two populations at that level. For the two data sets in this example, the critical confidence level is 94.243%. This value will be displayed in the Legend area of the plot.

Note that due to the calculation resolution and plot precision, the contour lines at the calculated critical level may appear to overlap or have a gap.

Life Comparison Tool

Another methodology, suggested by Gerald G. Brown and Herbert C. Rutemiller, is to estimate the probability of whether the times-to-failure of one population are better or worse than the times-to-failure of the second. The equation used to estimate this probability is given by:

\[
P [t_2 \geq t_1] = \int_0^\infty f_1(t) \cdot R_2(t) \cdot dt
\]

where \( f_1(t) \) is the pdf of the first distribution and \( R_2(t) \) is the reliability function of the second distribution. The evaluation of the superior data set is based on whether this probability is smaller or greater than 0.5. If the probability is equal to 0.5, then it is equivalent to saying that the two distributions are identical.

Sometimes we may need to compare the life when one of the distributions is truncated. For example, if the random variable from the first distribution is truncated with a range of \([L, U]\), then the comparison with the truncated distribution should be used. For details, please see Stress-Strength Analysis.

Consider two product designs where \( X \) and \( Y \) represent the life test data from two different populations. If we simply wanted to determine which component has a higher reliability, we would simply compare the reliability estimates of both components at a time \( t \). But if we wanted to determine which product will have a longer life, we would want to calculate the probability that the distribution of one product is better than the other. Using the equation given above, the probability that \( X \) is greater than or equal to \( Y \) can be interpreted as follows:

- If \( P = 0.50 \), then lives of both \( X \) and \( Y \) are equal.
Comparing Life Data Sets

- If \( P < 0.5 \) or, for example, \( P = 0.1 \), then \( P = 1 - 0.1 = 0.9 \), or \( Y \) is better than \( X \) with a 90\% probability.

**Example: Using the Life Comparison Tool to Compare Two Designs**

Using the same data set from the contour plot example, use Weibull++’s Life Comparison tool to estimate the probability that the units from the new design will outlast the units from the old design.

First, enter the data sets into two separate Weibull++ standard folios (or two separate data sheets within the same folio) and analyze the data sets using the two-parameter Weibull distribution and maximum likelihood estimation (MLE) method. Next, open the Life Comparison tool and select to compare the two data sets. The next figure shows the pdf curves and the result of the comparison.

The comparison summary is given in the Results Panel window.
Risk Analysis and Probabilistic Design with Monte Carlo Simulation

Monte Carlo simulation can be used to perform simple relationship-based simulations. This type of simulation has many applications in probabilistic design, risk analysis, quality control, etc. The Monte Carlo utility includes a User Defined distribution feature that allows you to specify an equation relating different random variables. The following example uses the Life Comparison tool to compare the pdf of two user-defined distributions. A variation of the example that demonstrates how to obtain the joint pdf of random variables is available in the Weibull++/ALTA Help file [1]. A demonstration on how to perform the example using ReliaSoft’s advanced stochastic event simulation software, RENO [2], is also available (view it in html [3]).

Monte Carlo Simulation: A Hinge Length Example

A hinge is made up of four components A, B, C, D, as shown next. Seven units of each component were taken from the assembly line and measurements (in cm) were recorded.
The following table shows the measurements. Determine the probability that D will fall out of specifications.

<table>
<thead>
<tr>
<th>Dimensions for A</th>
<th>Dimensions for B</th>
<th>Dimensions for C</th>
<th>Dimensions for D</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0187</td>
<td>1.9795</td>
<td>30.4216</td>
<td>33.6573</td>
</tr>
<tr>
<td>1.9996</td>
<td>2.0288</td>
<td>29.9818</td>
<td>34.5432</td>
</tr>
<tr>
<td>2.0167</td>
<td>1.9883</td>
<td>29.9724</td>
<td>34.6218</td>
</tr>
<tr>
<td>2.0329</td>
<td>2.0327</td>
<td>30.192</td>
<td>34.7538</td>
</tr>
<tr>
<td>2.0233</td>
<td>2.0119</td>
<td>29.9421</td>
<td>35.1508</td>
</tr>
<tr>
<td>2.0273</td>
<td>2.0354</td>
<td>30.1343</td>
<td>35.2666</td>
</tr>
<tr>
<td>1.984</td>
<td>1.9908</td>
<td>30.0423</td>
<td>35.7111</td>
</tr>
</tbody>
</table>

Solution

In a Weibull++ standard folio, enter the parts dimensions measurements of each component into separate data sheets. Analyze each data sheet using the normal distribution and the RRX analysis method. The parameters are:

\[
\begin{align*}
\hat{\mu}_A &= 2.0146 \\
\hat{\mu}_B &= 2.0096 \\
\hat{\mu}_C &= 30.0981 \\
\hat{\mu}_D &= 34.8149 \\
\hat{\sigma}_A &= 0.0181 \\
\hat{\sigma}_B &= 0.0249 \\
\hat{\sigma}_C &= 0.1762 \\
\hat{\sigma}_D &= 0.7121
\end{align*}
\]

Next, perform a Monte Carlo simulation to estimate the probability that (A+B+C) will be greater than D. To do this, choose the User Defined distribution and enter its equation as follows. (Click the Insert Data Source button to insert the data sheets that contain the measurements for the components.)
On the Settings tab, set the number of data points to 100, as shown next.
Click **Generate** to create a data sheet that contains the generated data points. Rename the new data sheet to "Simulated A+B+C."

Follow the same procedure to generate 100 data points to represent the D measurements. Rename the new data sheet to "Simulated D."
Analyze the two data sets, "Simulated A+B+C" and "Simulated D," using the normal distribution and the RRX analysis method.

Next, open the Life Comparison tool and choose to compare the two data sheets. The following picture shows the pdf curves of the two data sets.
The following report shows that the probability that "Simulated A+B+C" will be greater than "Simulated D" is 16.033%. (Note that the results may vary because of the randomness in the simulation.)
References

Weibull++ SimuMatic

Reliability analysis using simulation, in which reliability analyses are performed a large number of times on data sets that have been created using Monte Carlo simulation, can be a valuable tool for reliability practitioners. Such simulation analyses can assist the analyst to a) better understand life data analysis concepts, b) experiment with the influences of sample sizes and censoring schemes on analysis methods, c) construct simulation-based confidence intervals, d) better understand the concepts behind confidence intervals and e) design reliability tests. This section explores some of the results that can be obtained from simulation analyses using the Weibull++ SimuMatic tool.

Parameter Estimation and Confidence Bounds Techniques

In life data analysis, we use data (usually times-to-failure or times-to-success data) obtained from a sample of units to make predictions for the entire population of units. Depending on the sample size, the data censoring scheme and the parameter estimation method, the amount of error in the results can vary widely. To quantify this sampling error, or uncertainty, confidence bounds are widely used. In addition to the analytical calculation methods that are available, simulation can also be used. SimuMatic generates these confidence bounds and assists the practitioner (or the teacher) to visualize and understand them. In addition, it allows the analyst to determine the adequacy of certain parameter estimation methods (such as rank regression on X, rank regression on Y and maximum likelihood estimation) and to visualize the effects of different data censoring schemes on the confidence bounds.

As an example, we will attempt to determine the best parameter estimation method for a sample of ten units following a Weibull distribution with \( \beta = 2 \) and \( \eta = 100 \) and with complete time-to-failure data for each unit (i.e., no censoring). Using SimuMatic, 10,000 data sets are generated (using Monte Carlo methods based on the Weibull distribution) and we estimate their parameters using RRX, RRY and MLE. The plotted results generated by SimuMatic are shown next.
The results clearly demonstrate that the median RRX estimate provides the least deviation from the truth for this sample size and data type. However, the MLE outputs are grouped more closely together, as evidenced by the confidence bounds. The same figures also show the simulation-based bounds, as well as the expected variation due to sampling error.

This experiment can be repeated in SimuMatic using multiple censoring schemes (including Type I and Type II right censoring as well as random censoring) with the included distributions. We can perform multiple experiments with this utility to evaluate our assumptions about the appropriate parameter estimation method to use for the data set.

Using Simulation to Design Reliability Tests

Good reliability specifications include requirements for reliability and an associated lower one-sided confidence interval. When designing a test, we must determine the sample size to test as well as the expected test duration. The next simple example illustrates the methods available in SimuMatic.

Let us assume that a specific reliability specification states that at $T = 10$ hr the reliability must be 99%, or $R(T = 10) = 99\%$ (unreliability = 1%), and at $T = 20$ hr the reliability must be 90%, or $R(T = 20) = 90\%$, at an 80% lower one-sided confidence level ($L1S = 80\%$).

One way to meet this specification is to design a test that will demonstrate either of these requirements at $L1S = 80\%$ with the required parameters (for this example we will use the $R(T = 10) = 99\% @ L1S = 80\%$ requirement). With SimuMatic, we can specify the underlying distribution, distribution parameters (the Quick Parameter Estimator tool can be utilized), sample size on test, censoring scheme, required reliability and associated confidence level. From these inputs, SimuMatic will solve (via simulation) for the time demonstrated at the specified reliability and confidence level (i.e., $X$ in the $R(T = X) = 99\% @ L1S = 80\%$ formulation), as well as the expected test duration. If the demonstrated time is greater than the time requirement, this indicates that the test design would accomplish its required objective. Since there are multiple test designs that may accomplish the objective, multiple experiments should be performed until we arrive at an acceptable test design (i.e., number of units and test duration).

We start with a test design using a sample size of ten, with no censoring (i.e., all units to be tested to failure). We perform the analysis using RRX and 10,000 simulated data sets. The outcome is an expected test duration of 217 hr and a demonstrated time of 25 hr. This result is well above the stated requirement of 10 hr (note that in this case, the true value of $T$ at a 50% CL, for $R = 99\%$, is 40 hrs which gives us a ratio of 1.6 between true and demonstrated).

Since this would demonstrate the requirement, we can then attempt to reduce the number of units or test time. Suppose that we need to bring the test time down to 100 hr (instead of the expected 217 hr). The test could then be designed using Type II censoring (i.e., any unit that has not failed by 100 hr is right censored) assuring completion...
by 100 hr. Again, we specify Type II censoring at 100 hr in SimuMatic, and we repeat the simulation with the same parameters as before. The simulation results in this case yield an expected test duration of 100 hr and a demonstrated time of 17 hr at the stated requirements. This result is also above our requirement. The next figure graphically shows the results of this experiment. This process can then be repeated using different sample sizes and censoring schemes until we arrive at a desirable test plan.
Product reliability affects total product costs in multiple ways. Increasing reliability increases the initial cost of production but decreases other costs incurred over the life of the product. For example, increased reliability results in lower warranty and replacement costs for defective products. Increased reliability also results in greater market share because satisfied customers typically become repeat customers and recommend reliable products to others. A minimal total product cost can be determined by calculating the optimum reliability for such a product. The Target Reliability tool in Weibull++ does this by minimizing the sum of lost sales costs, warranty costs and manufacturing costs.

Cost Factors in Determining Target Reliability

Lost Sales Cost
The lost sales cost is caused due to lost market share. It is caused by customers choosing to go elsewhere for goods and services. The lost sales cost depends on the total market value for a product and the actual sales revenue of a product.

\[ \text{Lost Sales Cost} = \max \{ 0, \text{Total Market Value} - \text{Sales Revenue} \} \]

In Weibull++, we assume the total potential market value is the product of maximum market potential (number of units that could be sold) and the best unit sale price.

\[ \text{Total Market Value} = \text{Maximum Market Potential} \times \text{Best Market Unit Sale Price} \]

For example, if the maximum number of units demanded by the market is 100,000 and the best market unit sale price is $12.00, then the total market value would be:

\[ 100,000 \times 12.00 = 1,200,000.00 \]

Calculating sales revenue requires knowledge of market share and unit sale price. The function for market share is given by the equation:

\[ f_{\text{Market Share}}(R) = 1 - e^{-(R/a)^b} \]

where \( a \) and \( b \) are parameters fitted to market share data, and \( R \) is the product reliability.

The function for unit sale price is given by:

\[ f_{\text{Sale Price}}(R) = b \times e^{aR} \]

where \( a \) and \( b \) are parameters fitted to data, and \( R \) is the product reliability.

As a function of reliability, \( R \), sales revenue is then calculated as:

\[ \text{Sales Revenue} (R) = \text{Maximum Market Potential} \times \text{Market Share} (R) \times \text{Unit Price} (R) \]

Once the total market value and the sales revenue are obtained, they can then be used to calculate the lost sales cost using the formula given at the beginning of this section.
Production Cost

Production cost is a function of total market value, market share and manufacturing cost per unit. The function for production cost per unit is given by:

$$f_{\text{Production Cost}}(R) = b \times e^{\frac{a}{1-R}}$$

where $a$ and $b$ are parameters fitted to data, and $R$ is the product reliability.

Using the substitution of variable $R' = \frac{1}{1-R}$ results in the equation:

$$f_{\text{Production Cost}}(R') = b \times e^{aR'}$$

for which the parameters $a$ and $b$ can be determined using simple regression tools such as the functions in the Degradation Data Analysis in Weibull++.

Warranty Cost

Warranty cost is a function of total market value, market share, reliability and cost per failure. The function of cost per failure is given by:

$$f_{\text{Failure Cost}}(R) = b \times e^{aR}$$

where $a$ and $b$ are parameters fitted to data. For a given reliability value, $R$, the warranty cost is given by:

$$\text{Warranty Cost} (R) = \text{Maximum Market Potential} \times \text{Market Share} (R) \times (1 - R) \times \text{Cost Per Failure} (R)$$

Unreliability Cost

The sum of the Lost Sales Cost and Warranty Cost is called Unreliability Cost.

Total Cost

For a given reliability, $R$, the expected total cost is given by:

$$\text{Total Cost} (R) = \text{Lost Sales Cost} (R) + \text{Warranty Cost} (R) + \text{Production Cost} (R) = \text{Unreliability Cost} (R) + \text{Production Cost} (R)$$

The production cost is a pre-shipping cost, whereas the warranty and lost sales costs are incurred after a product is shipped. These pre- and post-shipping costs can be seen in the figure below.

The reliability value resulting in the lowest total cost will be the target reliability for the product.
**Profit and Return at Target Reliability**

With all of the costs described above, the *profit at a given reliability* can be calculated as:

\[
\text{Profit} (R) = \text{Sales Revenue} - \text{Warranty Cost} (R) - \text{Production Cost} (R)
\]

**Traditional ROI**

First, consider that traditional Return On Investment (ROI) is a performance measure used to evaluate the efficiency of an investment, or to compare the efficiency of a number of different investments. In general to calculate ROI, the benefit (return) of an investment is divided by the cost of the investment; and the result is expressed as a percentage or a ratio. The following equation illustrates this.

\[
\text{ROI} = \frac{\text{Gain from Investment} - \text{Cost of Investment}}{\text{Cost of Investment}}
\]

In this formula, *Gain from Investment* refers to the revenue or proceeds obtained due to the investment of interest.

Return on investment is a very popular metric because of its versatility and simplicity. That is, if an investment does not have a positive ROI, or if there are other opportunities with a higher ROI, then the investment should not be undertaken. Reliability ROI is computed in a similar manner by looking at the investment as the the investment in improving the reliability.

**ReliaSoft's Reliability Return on Investment (R3OI)**

R3OI considers the cost and return due to the product reliability. As we discussed before, high reliability will reduce the unreliability cost, but will increase the sales revenue and production cost. A balanced reliability target should be determined based on all of the costs involved. For a given initial investment value, the R3OI is calculated by:

\[
\text{R3OI} = \frac{\text{Profit}(R)-\text{Initial Investment}}{\text{Initial Investment}}
\]

**The Weibull++ Target Reliability Tool**

The purpose of this tool is to qualitatively explore different options with regards to a target reliability for a component, subsystem or system. All the costs are calculated using the equations given above.

There are five inputs. Specifically:

<table>
<thead>
<tr>
<th>Input Title</th>
<th>Input Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected failures/returns per period (as % of sales)</td>
<td>(Q% \ where \ 0 \leq Q \leq 100)</td>
</tr>
<tr>
<td>% of market share you expect to capture</td>
<td>(S% \ where \ 0 \leq S \leq 100)</td>
</tr>
<tr>
<td>Average unit sales price</td>
<td>(P \ where \ 0 &lt; P)</td>
</tr>
<tr>
<td>Average cost per unit to produce</td>
<td>(C \ where \ 0 &lt; C &lt; P + O)</td>
</tr>
<tr>
<td>Other costs per failure (in addition to replacement costs)</td>
<td>(O \ where \ 0 &lt; O &lt; C + O)</td>
</tr>
</tbody>
</table>

These five inputs are then repeated for three specific cases: Best Case, Most Likely and Worst Case.
Based on the above inputs, four models are then fitted as functions of reliability, \( R = (1 - Q) \) or:

\[
\begin{align*}
\hat{f}_{\text{Market Share}}(R) &= 1 - e^{(-\frac{R}{\alpha})^b} \\
\hat{f}_{\text{Sale Price}}(R) &= b \cdot e^{(a-R)} \\
\hat{f}_{\text{Production Cost}}(R) &= b \cdot e^{(a \left( \frac{1}{1-R} \right))} \\
\hat{f}_{\text{Failure Cost}}(R) &= b \cdot e^{(a-R)}
\end{align*}
\]

An additional variable needed is maximum market potential, \( M \). It is defined by users in the following input box:

All the related costs are defined as given in the previous section and calculated as a function of reliability. The value giving the lowest total cost is the optimal (target) reliability.

**Example**

**Determining Reliability Based on Costs**

The following table provides information for a particular product regarding market share, sales prices, cost of production and costs due to failure.

<table>
<thead>
<tr>
<th>Target Estimation Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate each scenario with the below factors:</td>
</tr>
<tr>
<td>Expected failures/returns per period (as % of sales)</td>
</tr>
<tr>
<td>% of market share you expect to capture</td>
</tr>
<tr>
<td>Average unit sales price</td>
</tr>
<tr>
<td>Average cost per unit to produce</td>
</tr>
<tr>
<td>Other costs per failure (in addition to replacement cost)</td>
</tr>
</tbody>
</table>

The first row in the table indicates the probability of failure during the warranty period. For example, in the best case scenario, the expected probability of failure will be 1% (i.e., the reliability will be 99%). Under this reliability, the expected market share is 80%, average unit sale price is $2.10, average cost per unit to produce is $1.50 and other costs per failure is $0.50.

The assumed maximum market potential is 1,000,000 units and the initial investment is $10,000.

**Solution**

Enter the given information in the Target Reliability tool in Weibull++ and click the Plot icon on the control panel. Next, click the Analysis Details button on the control panel to generate a report of the analysis. The following report
The following figure shows the Cost vs. Reliability plot of the cost models. The green vertical line on the plot represents the estimated reliability value that will minimize costs. In this example, this reliability value is estimated to be 96.7% at the end of the warranty period.
The following figures show the Profit vs. Reliability plot and the R3OI vs. Reliability plot. In the R3OI plot, the initial investment is set to $10,000.
Event Log Data Analysis

Event logs, or maintenance logs, store information about a piece of equipment's failures and repairs. They provide useful information that can help companies achieve their productivity goals by giving insight about the failure modes, frequency of outages, repair duration, uptime/downtime and availability of the equipment. Some event logs contain more information than others, but essentially event logs capture data in a format that includes the type of event, the date/time when the event occurred and the date/time when the system was restored to operation.

The data from event logs can be used to extract failure times and repair times information. Once the times-to-failure data and times-to-repair data have been obtained, a life distribution can be fitted to each data set. The principles and theory for fitting a life distribution is presented in detail in Life Distributions. The process of data extraction and model fitting can be automated using the Weibull++ event log folio.

Converting Event Logs to Failure/Repair Data

For \( n \) number of failures and repair actions that took place during the event logging period, the times-to-failure of every unique occurrence of an event are obtained by calculating the time between the last repair and the time the new failure occurred, or:

\[
\text{Time-to-Failure}_i = t_i - r_{i-1}
\]

where:

- \( i = 1, \ldots , n \)
- \( t_i \) is the date/time of occurrence of \( i \).
- \( r_{i-1} \) is the date/time of restoration of the previous occurrence \( (i - 1) \).

For systems that were new when the collection of the event log data started, the times to first occurrence of every unique event is equivalent to the date/time of the occurrence of the event minus the time the system monitoring started. That is:

\[
\text{Time-to-Failure}_1 = t_1 - \text{System Start Time}
\]

For systems that were not new when the collection of event log data started, the times to first occurrence of every unique event are considered to be suspensions (right censored) because the system is assumed to have accumulated more hours before the data collection period started (i.e., the time between the start date/time and the first occurrence of an event is not the entire operating time). In this case:

\[
\text{Suspension}_1 = t_1 - \text{System Start Time}
\]

When monitoring on the system is stopped or when the system is no longer being used, all events that have not occurred by this time are considered to be suspensions.

\[
\text{Last Suspension} = \text{System End Time} - r_n
\]

The four equations given above are valid for cases in which a component operates through the failure of other components. When the component does not operate through the failures, the assumptions must include the downtime of the system due to the other failures. In other words, the first four equations become:

\[
\text{Time-to-Failure}_i = t_i - r_{i-1} - (\text{System Downtime since } r_{i-1})
\]

\[
\text{Time-to-Failure}_1 = t_1 - (\text{System Start Time} - \text{System Downtime since System Start Time})
\]

\[
\text{Suspension}_1 = t_1 - (\text{System Start Time} - \text{System Downtime since System Start Time})
\]

\[
\text{Last Suspension} = \text{System End Time} - r_n - \text{System Downtime since } r_n
\]

Repair times are obtained by calculating the difference between the date/time of event occurrence and the date/time of restoration, or:

\[
\text{Time-to-repair}_i = r_i - t_i
\]
All these equations should also take into consideration the periods when the system is not operating or not in use, as in the case of operations that do not run on a 24/7 basis.

**Example: Simple System**

Consider a very simple system composed of only two components, A and B. The system runs from 8 AM to 5 PM, Monday through Friday. When a failure is observed, the system undergoes repair and the failed component is replaced. The date and time of each failure is recorded in an equipment downtime log, along with an indication of the component that caused the failure. The date and time when the system was restored is also recorded. The downtime log for this simple system is given next.

<table>
<thead>
<tr>
<th>CL</th>
<th>Date</th>
<th>Time</th>
<th>Date</th>
<th>Time</th>
<th>Component Responsible</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan-02-1997</td>
<td>4:00 PM</td>
<td>Jan-02-1997</td>
<td>7:49 PM</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>Jan-09-1997</td>
<td>8:30 AM</td>
<td>Jan-09-1997</td>
<td>10:43 AM</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>Jan-10-1997</td>
<td>9:13 AM</td>
<td>Jan-10-1997</td>
<td>7:48 PM</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>Jan-12-1997</td>
<td>3:26 PM</td>
<td>Jan-12-1997</td>
<td>6:46 PM</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>Jan-13-1997</td>
<td>4:56 PM</td>
<td>Jan-13-1997</td>
<td>5:21 PM</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>Jan-20-1997</td>
<td>1:38 PM</td>
<td>Jan-21-1997</td>
<td>7:15 PM</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>Jan-25-1997</td>
<td>10:32 AM</td>
<td>Jan-27-1997</td>
<td>10:47 PM</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>Jan-28-1997</td>
<td>11:31 AM</td>
<td>Jan-28-1997</td>
<td>12:00 PM</td>
<td>A</td>
</tr>
<tr>
<td>10</td>
<td>Feb-02-1997</td>
<td>2:38 PM</td>
<td>Feb-02-1997</td>
<td>7:11 PM</td>
<td>A</td>
</tr>
<tr>
<td>11</td>
<td>Feb-08-1997</td>
<td>3:51 PM</td>
<td>Feb-08-1997</td>
<td>8:22 PM</td>
<td>A</td>
</tr>
<tr>
<td>12</td>
<td>Feb-12-1997</td>
<td>4:42 PM</td>
<td>Feb-13-1997</td>
<td>9:59 AM</td>
<td>A</td>
</tr>
<tr>
<td>13</td>
<td>Feb-17-1997</td>
<td>2:47 PM</td>
<td>Feb-17-1997</td>
<td>7:13 PM</td>
<td>B</td>
</tr>
<tr>
<td>14</td>
<td>Feb-25-1997</td>
<td>4:31 PM</td>
<td>Feb-25-1997</td>
<td>5:00 PM</td>
<td>A</td>
</tr>
<tr>
<td>15</td>
<td>Feb-28-1997</td>
<td>9:00 AM</td>
<td>Feb-28-1997</td>
<td>3:10 PM</td>
<td>B</td>
</tr>
<tr>
<td>16</td>
<td>Mar-01-1997</td>
<td>10:16 AM</td>
<td>Mar-01-1997</td>
<td>10:43 AM</td>
<td>A</td>
</tr>
<tr>
<td>17</td>
<td>Mar-02-1997</td>
<td>3:41 PM</td>
<td>Mar-02-1997</td>
<td>9:11 PM</td>
<td>A</td>
</tr>
<tr>
<td>18</td>
<td>Mar-12-1997</td>
<td>8:46 AM</td>
<td>Mar-12-1997</td>
<td>9:20 PM</td>
<td>B</td>
</tr>
</tbody>
</table>

Note that:
- The date and time of each failure is recorded.
- The date and time of repair completion for each failure is recorded.
- The repair involves replacement of the responsible component.
- The responsible component for each failure is recorded.

For this example, we will assume that an engineer began recording these events on January 1, 1997 at 12 PM and stopped recording on March 18, 1997 at 1 PM, at which time the analysis was performed. Information for events prior to January 1 is unknown. The objective of the analysis is to obtain the failure and repair distributions for each component.

**Solution**

We begin the analysis by looking at component A. The first time that component A is known to have failed is recorded in row 1 of the data sheet; thus, the first age (or time-to-failure) for A is the difference between the time we began recording the data and the time when this failure event happened. Also, the component does not age when the
system is down due to the failure of another component. Therefore, this time must be taken into account.

1. The First Time-To-Failure for Component A, $TTF_A^{[1]}$

The first time-to-failure of component A, $TTF_A^{[1]}$, is the sum of the hours of operation for each day, starting on the start date (and time) and ending with the failure date (and time). This is graphically shown next. The boxes with the green background indicate the operating periods. Thus, $TTF_A^{[1]} = 5 + 8 = 13$ hours.

![Graph showing operating periods for Component A](image)

2. The First Time-To-Repair for Component A, $TTR_A^{[1]}$

The time-to-repair for component A for this failure, $TTR_A^{[1]}$, is $[\text{Date/Time Restored - Date/Time Occurred}]$ or:

$$TTR_A^{[1]} = (\text{Jan 02 1997/7:49 PM}) - (\text{Jan 02 1997/4:00 PM}) = 3:49 = 3.8166 \text{ hours}$$

(Note that in the case of repair actions, shifts are not taken into account since it is assumed that repair actions will be performed as needed to bring the system up.)

3. The Second Time-To-Failure for Component A, $TTF_A^{[2]}$

Continuing with component A, the second system failure due to component A is found in row 4, on January 12, 1997 at 3:26 PM. Thus, to compute $TTF_A^{[2]}$, you must look at the age the component accumulated from the last repair time, taking shifts into account as before, but with the added complexity of accounting for the times that the system was down due to failures of other components (i.e., component A was not aging when the system was down for repair due to a component B failure).

This is graphically shown next using green to show the operating times of A and orange to show the downtimes of the system for reasons other than the failure of A (to the closest hour).

![Graph showing operating and downtime periods for Component A](image)

To illustrate this mathematically, we will use a function, $\mathcal{T}$, which, given a range of times, returns the shift hours worked during that period. In other words, for this example $\mathcal{T}(1/1/97 \text{ 3:00 AM} - 1/1/97 \text{ 6:00 PM}) = 9$ hours given an 8 AM to 5 PM shift. Furthermore, we will show the date and time a failure occurred as DTO and the date and time a repair was completed at DTR with a numerical subscript indicating the row that this entry is in (e.g., DTO$_4$ for the date and time a failure occurred in row 4).

Then the total possible hours (TPH) that component A could have operated from the time it was repaired to the time it failed the second time is:

$$TPH = \mathcal{T}(\text{DTO}_4 - \text{DTR}_1),$$

$$TPH = \mathcal{T}(\text{DTO}_4 - \text{DTR}_1) = 9 \text{ Days} \ast 9 \text{ hours} + 7:26 \text{ hours} = 88:26 \text{ hours} = 88.433 \text{ hours}$$

The time that component A was not operating (NOP) during normal hours of operation is the time that the system was down due to failure of component B, or:
NOP = \tau(DTO_2 - DTR_2) + \tau(DTO_3 - DTR_3) \\
NOP = \tau(DTO_2 - DTR_2) + \tau(DTO_3 - DTR_3) = 2:13 \text{ hours} + 7:47 \text{ hours} = 10:00 \text{ hours}

Thus, the second time-to-failure for component A, \( TTF_A[2] \), is:

\[ TTF_A[2] = TPH - NOP \]

\[ TTF_A[2] = 88:26 \text{ hours} - 10:00 \text{ hours} = 78:26 \text{ hours} = 78.433 \text{ hours} \]


To compute the time-to-repair for this failure:

\[ TTR_A[2] = \tau(DTO_4 - DTR_4) = (3 \text{ h}, 49 \text{ m}) = 3.8166 \text{ hours} \]

5. Computing the Rest of the Observed Failures

This same process can be repeated for the rest of the observed failures, yielding:

\[ TTF_A[3] = 8.9333 \]
\[ TTF_A[4] = 56.25 \]
\[ TTF_A[5] = 33.05 \]
\[ TTF_A[6] = 100.8433 \]
\[ TTF_A[7] = 35.7 \]
\[ TTF_A[8] = 112.3166 \]
\[ TTF_A[9] = 23.1 \]
\[ TTF_A[10] = 13.9666 \]

and

\[ TTR_A[3] = 0.4166 \]
\[ TTR_A[4] = 29.6166 \]
\[ TTR_A[5] = 0.4833 \]
\[ TTR_A[6] = 4.5166 \]
\[ TTR_A[7] = 17.2833 \]
\[ TTR_A[8] = 0.4833 \]
\[ TTR_A[9] = 0.45 \]
\[ TTR_A[10] = 5.5 \]
\[ TTR_A[11] = 0.4666 \]

6. Creating the Data Sets

When the computations shown above are complete, we can create the data set needed to obtain the life distributions for the failure and repair times for component A. To accomplish this, modifications will need to be performed on the TTF data, given the original assumptions, as follows:

- \( TTF_A[1] \) will be designated as a right censored data point (or suspension, S). This is because when we started collecting data, component A was operating for an unknown period of time \( X \), so the true time-to-failure for component A is the operating time observed (in this case, \( TTF_A[1] = 13 \) hours) plus the unknown operating time \( X \). Thus, what we know is that the true time-to-failure for A is some time greater than the observed \( TTF_A[1] \) (i.e., a right censored data point).
- An additional right censored observation (suspension) will be added to the data set to reflect the time that component A operated without failure from its last repair time to the end of the observation period. This is
presented next.

Since our analysis time ends on March 18, 1997 at 1:00 PM and component A has operated successfully from the last time it was replaced on March 13, 1997 at 5:13 PM, the additional time of successful operation is:

\[
TPH = \tau(\text{End Time} - \text{DTR}_{19}) = (4 \text{ days} \times 9 \text{ hours/day} + 5:00 \text{ hours}) = 41:00 \text{ hours}
\]

\[
\text{NOP} = \tau(\text{DTO}_{20} - \text{DTR}_{20}) = 7:24 \text{ hours}
\]

Thus, the remaining time that component A operated without failure is:

\[
\text{TTS} = TPH - \text{NOP} = 33:36 = 33.6 \text{ hours}
\]

The next two tables show component A's failure and repair data. The entire analysis can be repeated to obtain the failure and repair times for component B.

### Failure Data for A

<table>
<thead>
<tr>
<th>#</th>
<th>F/S</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>76.43333</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>8.933333</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>56.25</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>33.05</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>100.4833</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>35.7</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>112.3167</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>23.1</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>13.96667</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>90.51667</td>
</tr>
<tr>
<td>12</td>
<td>S</td>
<td>33.6</td>
</tr>
</tbody>
</table>

### Repair Data for A

<table>
<thead>
<tr>
<th>#</th>
<th>F/S</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>3.8167</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>3.3333</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>0.4167</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>29.6167</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>0.4833</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>4.5167</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>17.2833</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>0.4833</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>0.4900</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>5.5000</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>0.4667</td>
</tr>
</tbody>
</table>

7. Automated Analysis in Weibull++8

The analysis can be automatically performed in the Weibull++ 8 event log folio. Simply enter the data from the equipment downtime log into the folio, as shown next.
Use the Shift Pattern window (Event Log > Action and Settings > Set Shift Pattern) to specify the 8:00 AM to 5:00 PM shifts that occur seven days a week, as shown next.

The folio will automatically convert the equipment downtime log data to time-to-failure and time-to-repair data and fit failure and repair distributions to the data set. To view the failure and repair results, click the Show Analysis Summary (...) button on the control panel. The Results window shows the calculated values.
Example: System with Failure and Non-Failure Events

This example is similar to the previous example; however, we will now classify each occurrence as a failure or event, thus adding another level of complexity to the analysis.

Consider the same data set from the previous example, but with the addition of a column indicating whether the occurrence was a failure (F) or an event (E). (A general event represents an activity that brings the system down but is not directly relevant to the reliability of the equipment, such as preventive maintenance, routine inspections and the like.)
Allowing for the inclusion of an F/E identifier increases the number of items that we will be considering. In other words, the objective now will be to determine a failure and repair distribution for component A due to failure, component B due to failure, component A due to an event and component B due to an event. This is mathematically equivalent to increasing the number of components from two (A and B) to four (FA, EA, FB and EB); thus, the analysis steps are identical to the ones performed in the first example, but for four components instead of two.

In the Weibull++ event log folio, to consider the F/E in the analysis, the setting in the Analyze Failures and Events area of the control panel must be set to Separately. Selecting the Combined option will result in ignoring the F and E distinction and treating all entries as Fs (i.e., perform the same analysis as in the first example).

<table>
<thead>
<tr>
<th>Row</th>
<th>F/E</th>
<th>System Failed Date</th>
<th>System Failed Time</th>
<th>System Repaired Date</th>
<th>System Repaired Time</th>
<th>Component Responsible</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>Jan-02-1997 4:00 PM</td>
<td></td>
<td>Jan-02-1997 7:49 PM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td>Jan-09-1997 8:30 AM</td>
<td></td>
<td>Jan-09-1997 10:43 AM</td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>Jan-10-1997 9:13 AM</td>
<td></td>
<td>Jan-10-1997 7:48 PM</td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>Jan-12-1997 3:26 PM</td>
<td></td>
<td>Jan-12-1997 6:45 PM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>Jan-13-1997 4:56 PM</td>
<td></td>
<td>Jan-13-1997 5:21 PM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>Jan-20-1997 1:36 PM</td>
<td></td>
<td>Jan-21-1997 7:15 PM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>E</td>
<td>Jan-28-1997 11:31 AM</td>
<td></td>
<td>Jan-28-1997 12:00 PM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>Feb-02-1997 2:36 PM</td>
<td></td>
<td>Feb-02-1997 7:11 PM</td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>11</td>
<td>E</td>
<td>Feb-08-1997 3:51 PM</td>
<td></td>
<td>Feb-08-1997 8:22 PM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>Feb-12-1997 4:42 PM</td>
<td></td>
<td>Feb-13-1997 9:59 AM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>13</td>
<td>E</td>
<td>Feb-17-1997 2:47 PM</td>
<td></td>
<td>Feb-17-1997 7:13 PM</td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>14</td>
<td>E</td>
<td>Feb-25-1997 4:31 PM</td>
<td></td>
<td>Feb-25-1997 5:00 PM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>Feb-28-1997 9:00 AM</td>
<td></td>
<td>Feb-28-1997 3:10 PM</td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>Mar-01-1997 10:16 AM</td>
<td></td>
<td>Mar-01-1997 10:43 AM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>17</td>
<td>E</td>
<td>Mar-02-1997 3:41 PM</td>
<td></td>
<td>Mar-02-1997 9:11 PM</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>Mar-12-1997 8:46 AM</td>
<td></td>
<td>Mar-12-1997 9:20 PM</td>
<td></td>
<td>B</td>
</tr>
</tbody>
</table>
Example: System with Components Operating Through System Failures

We will repeat the first example, but with an interesting twist. In the first example, the assumption was that when the system was down due to a failure, none of the components aged. We will now allow some of these components to age even when the system is down.

Using the same data as in the first example, we will assume that component B accumulates age when the system is down for any reason other than the failure of component B. Component A will not accumulate age when the system is down. We indicate this in the data sheet in the OTF (Operate Through other Failures) column, where Y = yes and N = no. Note that this entry must be consistent with all other events in the data set associated with the component (i.e., all component A records must show N and all component B records must show Y).
Because component A does not operate through other failures, the analysis steps (and results) are identical to the ones in the first example. However, the analysis for the times-to-failure for component B will be different (and is actually less cumbersome than in the first example because we do not need to subtract the time that the system was down due to the failure of other components).

Specifically:

\[ TTF_{B1} = \tau(\text{Start Time} - DTO_2) = 68:30 = 68.5 \text{ hours} \]
\[ TTF_{B2} = \tau(DTR_2 - DTO_3) = 7:30 = 7.5 \text{ hours} \]
\[ TTF_{B3} = \tau(DTR_3 - DTO_6) = 41.26 \text{ hours} \]

The TTR analysis is not affected by this setting and is identical to the result in the first example.
Example: System with Sub-Components

In this example, we will repeat the first example, but use two component levels instead of a single level. Using the same data as in the first example, we will assume that component A has two sub-components, A1 and A2; and component B also has two sub-components, B1 and B2, as shown in the following table.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Date</td>
<td>Time</td>
<td>Date</td>
<td>Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan-02-1997</td>
<td>4:00 PM</td>
<td>Jan-02-1997</td>
<td>7:49 PM</td>
<td>A</td>
<td>A1</td>
</tr>
<tr>
<td>Jan-09-1997</td>
<td>8:30 AM</td>
<td>Jan-09-1997</td>
<td>10:43 AM</td>
<td>B</td>
<td>B1</td>
</tr>
<tr>
<td>Jan-12-1997</td>
<td>3:26 PM</td>
<td>Jan-12-1997</td>
<td>6:46 PM</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>Jan-20-1997</td>
<td>1:38 PM</td>
<td>Jan-21-1997</td>
<td>7:15 PM</td>
<td>A</td>
<td>A1</td>
</tr>
<tr>
<td>Jan-28-1997</td>
<td>11:31 AM</td>
<td>Jan-28-1997</td>
<td>12:00 PM</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>Feb-02-1997</td>
<td>2:38 PM</td>
<td>Feb-02-1997</td>
<td>7:11 PM</td>
<td>B</td>
<td>B1</td>
</tr>
<tr>
<td>Feb-08-1997</td>
<td>3:51 PM</td>
<td>Feb-08-1997</td>
<td>8:22 PM</td>
<td>A</td>
<td>A1</td>
</tr>
<tr>
<td>Feb-12-1997</td>
<td>4:42 PM</td>
<td>Feb-13-1997</td>
<td>9:59 AM</td>
<td>A</td>
<td>A2</td>
</tr>
<tr>
<td>Feb-25-1997</td>
<td>4:31 PM</td>
<td>Feb-25-1997</td>
<td>5:00 PM</td>
<td>A</td>
<td>A1</td>
</tr>
<tr>
<td>Feb-28-1997</td>
<td>9:00 AM</td>
<td>Feb-28-1997</td>
<td>3:10 PM</td>
<td>B</td>
<td>B1</td>
</tr>
</tbody>
</table>

Analyzing the data for only level 1 will ignore all entries for level 2. However, if performing a level 2 analysis, both entries in level 1 and 2 are taken into account, creating a unique item. More specifically, the analysis for level 2 will be done at the sub-component level and depend on the component that the sub-component belongs to. Note that this is equivalent to reducing the analysis to a single level but looking at each component and sub-component combination individually through the process. In other words, for level 2 analyses, the data set shown above can be reduced to a single level analysis for four items: A-A1, A-A2, B-B1 and B-B2.

This component and sub-component combination will yield a unique item that must be treated separately, much as components A and B were in the first example. The same concept applies when more levels are present. When specifying OTF (Operate Through other Failures) characteristics at this level, each component and sub-component combination is treated differently. In other words, it is possible that A-A1 and A-A2 have different OTF designations, even though they belong to the same component; however, all A-A2s must have the same OTF designation.
More event log folio examples are available! See also:

🔗 Factory Equipment Failure Log[^1] or 🎬 Watch the video...[^2]

---

**References**

Appendices

Least Squares/Rank Regression Equations

Rank Regression on Y

Assume that a set of data pairs \((x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\), were obtained and plotted. Then, according to the least squares principle, which minimizes the vertical distance between the data points and the straight line fitted to the data, the best fitting straight line to these data is the straight line \(y = \hat{a} + \hat{b}x\) such that:

\[
\sum_{i=1}^{N}(\hat{a} + \hat{b}x_i - y_i)^2 = \min(a, b) \sum_{i=1}^{N}(a + bx_i - y_i)^2
\]

and where \(\hat{a}\) and \(\hat{b}\) are the least squares estimates of \(a\) and \(b\), and \(N\) is the number of data points.

To obtain \(\hat{a}\) and \(\hat{b}\), let:

\[
F = \sum_{i=1}^{N}(a + bx_i - y_i)^2
\]

Differentiating \(F\) with respect to \(a\) and \(b\) yields:

\[
\frac{\partial F}{\partial a} = 2 \sum_{i=1}^{N}(a + bx_i - y_i)\quad (1)
\]

and:

\[
\frac{\partial F}{\partial b} = 2 \sum_{i=1}^{N}(a + bx_i - y_i)x_i\quad (2)
\]

Setting Eqns. (1) and (2) equal to zero yields:

\[
\sum_{i=1}^{N}(a + bx_i - y_i)x_i = \sum_{i=1}^{N}(\hat{y}_i - y_i)x_i = -\sum_{i=1}^{N}(y_i - \hat{y}_i)x_i = 0
\]

and:

\[
\sum_{i=1}^{N}(a + bx_i - y_i)x_i = \sum_{i=1}^{N}(\hat{y}_i - y_i)x_i = -\sum_{i=1}^{N}(y_i - \hat{y}_i)x_i = 0
\]

Solving the equations simultaneously yields:

\[
\hat{a} = \frac{\sum_{i=1}^{N} y_i}{N} - \hat{b}\frac{\sum_{i=1}^{N} x_i}{N} = \bar{y} - \bar{x}\hat{b}\quad (3)
\]

and:

\[
\hat{b} = \frac{\sum_{i=1}^{N} i \cdot x_i \cdot y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sum_{i=1}^{N} x_i^2 - \frac{(\sum_{i=1}^{N} x_i)^2}{N}}\quad (4)
\]
Rank Regression on X

Assume that a set of data pairs \((x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\) were obtained and plotted. Then, according to the least squares principle, which minimizes the horizontal distance between the data points and the straight line fitted to the data, the best fitting straight line to these data is the straight line \(x = y\) such that:

\[
\sum_{i=1}^{N} (\hat{a} + \hat{b}y_i - x_i)^2 = \min(a, b) \sum_{i=1}^{N} (a + b y_i - x_i)^2
\]

Again, \(\hat{a}\) and \(\hat{b}\) are the least squares estimates of \(a\) and \(b\), and \(N\) is the number of data points.

To obtain \(\hat{a}\) and \(\hat{b}\), let:

\[
F = \sum_{i=1}^{N} (a + b y_i - x_i)^2
\]

Differentiating \(F\) with respect to \(a\) and \(b\) yields:

\[
\frac{\partial F}{\partial a} = 2 \sum_{i=1}^{N} (a + b y_i - x_i) = 0
\]

and:

\[
\frac{\partial F}{\partial b} = 2 \sum_{i=1}^{N} (a + b y_i - x_i) y_i = 0
\]

Setting Eqns. (5) and (6) equal to zero yields:

\[
\sum_{i=1}^{N} (a + b y_i - x_i) = \sum_{i=1}^{N} (\hat{x}_i - x_i) = 0
\]

and:

\[
\sum_{i=1}^{N} (a + b y_i - x_i) y_i = \sum_{i=1}^{N} (\hat{x}_i - x_i) y_i = 0
\]

Solving the above equations simultaneously yields:

\[
\hat{a} = \frac{\sum_{i=1}^{N} x_i}{N} - \hat{b} \frac{\sum_{i=1}^{N} y_i}{N} = \bar{x} - \hat{b}\bar{y}
\]

and:

\[
\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \left(\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i\right)}{N} \quad (8)
\]

Solving the equation of the line for \(y\) yields:

\[
y = \frac{\hat{a}}{\hat{b}} + \frac{1}{\hat{b}} x
\]
Example

Fit a least squares straight line using regression on X and regression on Y to the following data:

\[ \begin{array}{cccccccc}
  x & 1 & 2.5 & 4 & 6 & 8 & 9 & 11 & 15 \\
  y & 1.5 & 2 & 4 & 4 & 5 & 7 & 8 & 10 \\
\end{array} \]

The first step is to generate the following table:

<table>
<thead>
<tr>
<th>i</th>
<th>( x_i )</th>
<th>( y_i )</th>
<th>( x_i^2 )</th>
<th>( x_i y_i )</th>
<th>( y_i^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1.5</td>
<td>1</td>
<td>1.5</td>
<td>2.25</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>2</td>
<td>6.25</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>4</td>
<td>36</td>
<td>24</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>5</td>
<td>64</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>7</td>
<td>81</td>
<td>63</td>
<td>49</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>8</td>
<td>121</td>
<td>88</td>
<td>64</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>10</td>
<td>225</td>
<td>150</td>
<td>100</td>
</tr>
<tr>
<td>( \Sigma )</td>
<td>56.5</td>
<td>41.5</td>
<td>550.25</td>
<td>387.5</td>
<td>276.25</td>
</tr>
</tbody>
</table>

From the table then, and for rank regression on Y (RRY):

\[ \hat{b} = \frac{387.5 - (56.5)(41.5)/8}{550.25 - (56.5)^2/8} \]

\[ \hat{b} = 0.6243 \]

and:

\[ \hat{a} = \frac{41.5}{8} - 0.6243 \frac{56.5}{8} \]

\[ \hat{a} = 0.77836 \]

The least squares line is given by:

\[ y = 0.77836 + 0.6243x \]

The plotted line is shown in the next figure.
For rank regression on X (RRX) using the same table yields:

\[
\hat{b} = \frac{387.5 - (56.5)(41.5)/8}{276.25 - (41.5)^2/8}
\]
\[
\hat{a} = -0.97002
\]

and:

\[
\hat{a} = \frac{56.5}{8} - 1.5484 \frac{41.5}{8}
\]
\[
\hat{a} = -0.97002
\]

The least squares line is given by:

\[
y = \frac{(-0.97002)}{1.5484} + \frac{1}{1.5484} \cdot x
\]
\[
y = 0.62645 + 0.64581 \cdot x
\]

The plotted line is shown in the next figure.
Note that the regression on \( Y \) is not necessarily the same as the regression on \( X \). The only time when the two regressions are the same (i.e., will yield the same equation for a line) is when the data lie perfectly on a line.

The correlation coefficient is given by:

\[
\hat{\rho} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{N}{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\left(\sum_{i=1}^{N} x_i^2 - \frac{(\sum_{i=1}^{N} x_i)^2}{N}\right) \left(\sum_{i=1}^{N} y_i^2 - \frac{(\sum_{i=1}^{N} y_i)^2}{N}\right)}
\]

\[
\hat{\rho} = \frac{387.5 - (56.5)(41.5)/8}{\left((550.25 - (56.5)^2/8)(276.25 - (41.5)^2/8)\right)^{\frac{1}{2}}}
\]

\[
\hat{\rho} = 0.98321
\]
Appendix: Maximum Likelihood Estimation Example

MLE Statistical Background

If \( x \) is a continuous random variable with pdf:

\[
f(x; \theta_1, \theta_2, \ldots, \theta_k),
\]

where \( \theta_1, \theta_2, \ldots, \theta_k \) are unknown constant parameters that need to be estimated, conduct an experiment and obtain \( N \) independent observations, \( x_1, x_2, \ldots, x_N \), which correspond in the case of life data analysis to failure times. The likelihood function (for complete data) is given by:

\[
L(x_1, x_2, \ldots, x_N | \theta_1, \theta_2, \ldots, \theta_k) = L = \prod_{i=1}^{N} f(x_i; \theta_1, \theta_2, \ldots, \theta_k)
\]

\( i = 1, 2, \ldots, N \)

The logarithmic likelihood function is:

\[
\Lambda = \ln L = \sum_{i=1}^{N} \ln f(x_i; \theta_1, \theta_2, \ldots, \theta_k)
\]

The maximum likelihood estimators (MLE) of \( \theta_1, \theta_2, \ldots, \theta_k \) are obtained by maximizing \( L \) or \( \Lambda \). By maximizing \( \Lambda \), which is much easier to work with than \( L \), the maximum likelihood estimators (MLE) of \( \theta_1, \theta_2, \ldots, \theta_k \) are the simultaneous solutions of \( k \) equations such that:

\[
\frac{\partial \Lambda}{\partial \theta_j} = 0; \quad j = 1, 2, \ldots, k
\]

Even though it is common practice to plot the MLE solutions using median ranks (points are plotted according to median ranks and the line according to the MLE solutions), this is not completely accurate. As it can be seen from the equations above, the MLE method is independent of any kind of ranks. For this reason, many times the MLE solution appears not to track the data on the probability plot. This is perfectly acceptable since the two methods are independent of each other, and in no way suggests that the solution is wrong.

Illustrating the MLE Method Using the Exponential Distribution

- To estimate \( \lambda \) for a sample of \( n \) units (all tested to failure), first obtain the likelihood function:

\[
L(\lambda | t_1, t_2, \ldots, t_n) = \prod_{i=1}^{n} f(t_i)
\]

\[
= \prod_{i=1}^{n} \lambda e^{-\lambda t_i}
\]

\[
= \lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} t_i}
\]

- Take the natural log of both sides:

\[
\Lambda = \ln(L) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} t_i.
\]

- Obtain \( \frac{\partial \Lambda}{\partial \lambda} \), and set it equal to zero:

\[
\frac{\partial \Lambda}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i = 0
\]
• Solve for \( \hat{\lambda} \) or:
\[
\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}
\]

Notes About \( \hat{\lambda} \)

Note that the value of \( \hat{\lambda} \) is an estimate because if we obtain another sample from the same population and re-estimate \( \lambda \), the new value would differ from the one previously calculated. In plain language, \( \hat{\lambda} \) is an estimate of the true value of \( \lambda \). How close is the value of our estimate to the true value? To answer this question, one must first determine the distribution of the parameter, in this case \( \lambda \). This methodology introduces a new term, confidence bound, which allows us to specify a range for our estimate with a certain confidence level. The treatment of confidence bounds is integral to reliability engineering, and to all of statistics. (Confidence bounds are covered in Confidence Bounds.)

Illustrating the MLE Method Using Normal Distribution

To obtain the MLE estimates for the mean, \( \bar{T} \), and standard deviation, \( \sigma_T \) for the normal distribution, start with the pdf of the normal distribution which is given by:
\[
f(T) = \frac{1}{\sigma_T \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{T - \bar{T}}{\sigma_T} \right)^2}
\]

If \( T_1, T_2, \ldots, T_N \) are known times-to-failure (and with no suspensions), then the likelihood function is given by:
\[
L(T_1, T_2, \ldots, T_N | \bar{T}, \sigma_T) = L = \prod_{i=1}^{N} \left[ \frac{1}{\sigma_T \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{T_i - \bar{T}}{\sigma_T} \right)^2} \right]
\]
\[
L = \frac{1}{(\sigma_T \sqrt{2\pi})^N} e^{-\frac{1}{2} \sum_{i=1}^{N} \left( \frac{T_i - \bar{T}}{\sigma_T} \right)^2}
\]

then:
\[
\Lambda = \ln L = -\frac{N}{2} \ln(2\pi) - N \ln \sigma_T - \frac{1}{2} \sum_{i=1}^{N} \left( \frac{T_i - \bar{T}}{\sigma_T} \right)^2
\]

Then taking the partial derivatives of \( \Lambda \) with respect to each one of the parameters and setting them equal to zero yields:
\[
\frac{\partial \Lambda}{\partial \bar{T}} = \frac{1}{\sigma_T^2} \sum_{i=1}^{N} (T_i - \bar{T}) = 0
\]
and:
\[
\frac{\partial \Lambda}{\partial \sigma_T} = -\frac{N}{\sigma_T} + \frac{1}{\sigma_T^2} \sum_{i=1}^{N} (T_i - \bar{T})^2 = 0
\]

Solving the above two derivative equations simultaneously yields:
\[
\bar{T} = \frac{1}{N} \sum_{i=1}^{N} T_i
\]
and:
\[
\hat{o}_T^2 = \frac{1}{N} \sum_{i=1}^{N} (T_i - \bar{T})^2
\]

\[
\hat{o}_T = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (T_i - \bar{T})^2}
\]

It should be noted that these solutions are valid only for data with no suspensions, (i.e., all units are tested to failure). In the case where suspensions are present or all units are not tested to failure, the methodology changes and the problem becomes much more complicated.

**Illustrating with an Example of the Normal Distribution**

If we had five units that failed at 10, 20, 30, 40 and 50 hours, the mean would be:

\[
\bar{T} = \frac{1}{N} \sum_{i=1}^{N} T_i
\]

\[
= \frac{10 + 20 + 30 + 40 + 50}{5}
\]

\[
= 30
\]

The standard deviation estimate then would be:

\[
\hat{o}_T = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (T_i - \bar{T})^2}
\]

\[
= \sqrt{\frac{(10 - 30)^2 + (20 - 30)^2 + (30 - 30)^2 + (40 - 30)^2 + (50 - 30)^2}{5}}
\]

\[
= 14.1421
\]

A look at the likelihood function surface plot in the figure below reveals that both of these values are the maximum values of the function.
This three-dimensional plot represents the likelihood function. As can be seen from the plot, the maximum likelihood estimates for the two parameters correspond with the peak or maximum of the likelihood function surface.
Appendix: Special Analysis Methods

Grouped Data Analysis
The grouped data type in Weibull++ is used for tests where there are groups of units having the same time-to-failure, or units are grouped together in intervals, or there are groups of units suspended at the same time. However, you must be cautious in using the different parameter estimation methods because different methods treat grouped data in different ways. ReliaSoft designed Weibull++ to treat grouped data in different ways to maximize the options available to you.

When Using Rank Regression (Least Squares)
When using grouped data, Weibull++ plots the data point corresponding to the highest rank position in each group. For example, given 3 groups of 10 units, each failing at 100, 200 and 300 hours respectively, the three plotted points will be the end point of each group, or the 10th rank position out of 30, the 20th rank position out of 30 and the 30th rank position out of 30. This procedure is identical to standard procedures for using grouped data, as discussed in Kececioglu [19]. In cases where the grouped data are interval censored, it is assumed that the failures occurred at some time in the interval between the previous and current time to failure. In our example, this would be the same as saying that 10 units have failed in the interval between zero and 100 hours, another 10 units failed in the interval between 100 and 200 hours, and 10 more units failed in the interval from 200 to 300 hours. The rank regression analysis automatically takes this into account. If this assumption of interval failure is incorrect (i.e., 10 units failed at exactly 100 hours, 10 failed at exactly 200 hours and 10 failed at exactly 300 hours), then it is recommended that you enter the data as non-grouped when using rank regression, or select the Ungroup on Regression check box on the Analysis page of the folio’s control panel.

The Mathematics
Median ranks are used to obtain an estimate of the unreliability, \( Q(T_j) \) for each failure at a 50% confidence level. In the case of grouped data, the ranks are estimated for each group of failures, instead of each failure. For example, consider a group of 10 failures at 100 hours, 10 at 200 hours and 10 at 300 hours. Weibull++ estimates the median ranks (Z values) by solving the cumulative binomial equation with the appropriate values for order number and total number of test units. For 10 failures at 100 hours, the median rank, \( Z \), is estimated by using:

\[
0.50 = \sum_{k=j}^{N} \binom{N}{k} Z^k (1 - Z)^{N-k}
\]

with:

\[ N = 30, \quad J = 10 \]

One \( Z \) is obtained for the group, to represent the probability of 10 failures occurring out of 30. For 10 failures at 200 hours, \( Z \) is estimated by using:

\[
0.50 = \sum_{k=j}^{N} \binom{N}{k} Z^k (1 - Z)^{N-k}
\]

where:

\[ N = 30, \quad J = 20 \]

This represents the probability of 20 failures out of 30. For 10 failures at 300 hours, \( Z \) is estimated by using:
\[
0.50 = \sum_{k=j}^{N} \binom{N}{k} Z^k (1 - Z)^{N-k}
\]
where:
\[N = 30, \quad J = 30\]
This represents the probability of 30 failures out of 30.

**When Using Maximum Likelihood**

When using maximum likelihood methods, each individual time is explicitly used in the calculation of the parameters. Theoretically, there is no difference in the entry of a group of 10 units failing at 100 hours and 10 individual entries of 100 hours. This is inherent in the standard MLE method. In other words, no matter how the data were entered (i.e., as grouped or non-grouped) the results will be identical. However, due to the precision issues during the computation, the grouped and ungrouped data may give slightly different results. When using maximum likelihood, we highly recommend entering redundant data in groups, as this significantly speeds up the calculations.

**ReliaSoft Ranking Method**

In probability plotting or rank regression analysis of interval or left censored data, difficulties arise when attempting to estimate the exact time within the interval when the failure actually occurs, especially when an overlap on the intervals is present. In this case, the standard ranking method (SRM) is not applicable when dealing with interval data; thus, ReliaSoft has formulated a more sophisticated methodology to allow for more accurate probability plotting and regression analysis of data sets with interval or left censored data. This method utilizes the traditional rank regression method and iteratively improves upon the computed ranks by parametrically recomputing new ranks and the most probable failure time for interval data.

In the traditional method for plotting or rank regression analysis of right censored data, the effect of the exact censoring time is not considered. One example of this can be seen at the parameter estimation chapter. The ReliaSoft ranking method also can be used to overcome this shortfall of the standard ranking method.

The following step-by-step example illustrates the ReliaSoft ranking method (RRM), which is an iterative improvement on the standard ranking method (SRM). Although this method is illustrated by the use of the two-parameter Weibull distribution, it can be easily generalized for other models.

Consider the following test data:

<table>
<thead>
<tr>
<th>Table B.1 - The Test Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Items</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
Initial Parameter Estimation

As a preliminary step, we need to provide a crude estimate of the Weibull parameters for this data. To begin, we will extract the exact times-to-failure (10, 40, and 50) and the midpoints of the interval failures. The midpoints are 50 (for the interval of 20 to 80) and 47.5 (for the interval of 10 to 85). Next, we group together the items that have the same failure times, as shown in Table B.2.

Using the traditional rank regression, we obtain the first initial estimates:

\[
\hat{\beta}_0 = 1.91367089 \quad \hat{\eta}_0 = 43.91657736
\]

<table>
<thead>
<tr>
<th>Number of Items</th>
<th>Type</th>
<th>Last Inspection</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exact Failure</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Exact Failure</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Exact Failure</td>
<td>47.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Exact Failure</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

**Step 1**

For all intervals, we obtain a weighted midpoint using:

\[
i_m\left(\beta, \eta\right) = \frac{\int_{L}^{T} t f(t; \beta, \eta) dt}{\int_{L}^{T} f(t; \beta, \eta) dt}
\]

\[
= \frac{\int_{L}^{T} \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} dt}{\int_{L}^{T} \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} dt}
\]

This transforms our data into the format in Table B.3.

<table>
<thead>
<tr>
<th>Number of Items</th>
<th>Type</th>
<th>Last Inspection</th>
<th>Time</th>
<th>Weighted &quot;Midpoint&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exact Failure</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Exact Failure</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Exact Failure</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Interval Failure</td>
<td>20</td>
<td>80</td>
<td>42.837</td>
</tr>
<tr>
<td>1</td>
<td>Interval Failure</td>
<td>10</td>
<td>85</td>
<td>39.169</td>
</tr>
</tbody>
</table>

**Step 2**

Now we arrange the data as in Table B.4.
Appendix: Special Analysis Methods

Table B.4 - The Union of Exact Times-to-Failure with the "Midpoint" of the Interval Failures, in Ascending Order.

<table>
<thead>
<tr>
<th>Number of Items</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>39.169</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>42.837</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
</tr>
</tbody>
</table>

Step 3

We now consider the left and right censored data, as in Table B.5.

Table B.5 - Computation of Increments in a Matrix Format for Computing a Revised Mean Order Number.

<table>
<thead>
<tr>
<th>Number of Items</th>
<th>Time of Failure</th>
<th>2 Left Censored $t = 30$</th>
<th>1 Left Censored $t = 70$</th>
<th>1 Left Censored $t = 100$</th>
<th>1 Right Censored $t = 20$</th>
<th>1 Right Censored $t = 60$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>$2 \int_{0}^{10} f_0(t) , dt \frac{F_0(30) - F_0(0)}{F_0(70) - F_0(0)}$</td>
<td>$\int_{0}^{10} f_0(t) , dt \frac{F_0(100) - F_0(0)}{F_0(70) - F_0(0)}$</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>39.169</td>
<td>$\int_{0}^{39.169} f_0(t) , dt \frac{F_0(70) - F_0(0)}{F_0(100) - F_0(0)}$</td>
<td>$\int_{0}^{39.169} f_0(t) , dt \frac{F_0(100) - F_0(0)}{F_0(100) - F_0(0)}$</td>
<td>$\int_{0}^{39.169} f_0(t) , dt \frac{F_0(\infty) - F_0(20)}{F_0(\infty) - F_0(0)}$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>0</td>
<td>$\int_{30}^{40} f_0(t) , dt \frac{F_0(70) - F_0(0)}{F_0(100) - F_0(0)}$</td>
<td>$\int_{30}^{40} f_0(t) , dt \frac{F_0(100) - F_0(0)}{F_0(100) - F_0(0)}$</td>
<td>$\int_{30}^{40} f_0(t) , dt \frac{F_0(\infty) - F_0(20)}{F_0(\infty) - F_0(0)}$</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>42.837</td>
<td>0</td>
<td>$\int_{40}^{42.837} f_0(t) , dt \frac{F_0(70) - F_0(0)}{F_0(100) - F_0(0)}$</td>
<td>$\int_{40}^{42.837} f_0(t) , dt \frac{F_0(100) - F_0(0)}{F_0(100) - F_0(0)}$</td>
<td>$\int_{40}^{42.837} f_0(t) , dt \frac{F_0(\infty) - F_0(20)}{F_0(\infty) - F_0(0)}$</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>0</td>
<td>$\int_{42.837}^{50} f_0(t) , dt \frac{F_0(70) - F_0(0)}{F_0(100) - F_0(0)}$</td>
<td>$\int_{42.837}^{50} f_0(t) , dt \frac{F_0(100) - F_0(0)}{F_0(100) - F_0(0)}$</td>
<td>$\int_{42.837}^{50} f_0(t) , dt \frac{F_0(\infty) - F_0(20)}{F_0(\infty) - F_0(0)}$</td>
<td>0</td>
</tr>
</tbody>
</table>

In general, for left censored data:

- The increment term for $n$ left censored items at time $t_0$ with a time-to-failure of $t_i$ when $t_0 \leq t_i$ is zero.
- When $t_0 > t_i$ the contribution is:

$$\frac{n}{F_0(t_0) - F_0(0)} \int_{t_i}^{t_0} f_0(t) \, dt$$

or:

$$\frac{n}{F_0(MIN(t_i, t_0)) - F_0(t_i)}$$

where $t_i$ is the time-to-failure previous to the $t_0$ time-to-failure and $n$ is the number of units associated with that time-to-failure (or units in the group).

In general, for right censored data:

- The increment term for $n$ right censored at time $t_0$ with a time-to-failure of $t_i$ when $t_0 \geq t_i$ is zero.
- When $t_0 < t_i$ the contribution is:

$$\frac{n}{F_0(\infty) - F_0(t_0)} \int_{t_i}^{MAX(t_i, t_0)} f_0(t) \, dt$$

or:

$$\frac{n}{F_0(\infty) - F_0(t_0)}$$
where \( t_i \) is the time-to-failure previous to the \( t_i \) time-to-failure and \( n \) is the number of units associated with that time-to-failure (or units in the group).

**Step 4**
Sum up the increments (horizontally in rows), as in Table B.6.

<table>
<thead>
<tr>
<th>Number of</th>
<th>Time of Failure</th>
<th>2 Left Censored ( t=30 )</th>
<th>1 Left Censored ( t=70 )</th>
<th>1 Left Censored ( t=100 )</th>
<th>1 Right Censored ( t=20 )</th>
<th>1 Right Censored ( t=60 )</th>
<th>Sum of row(increment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>0.299065</td>
<td>0.062673</td>
<td>0.057673</td>
<td>0</td>
<td>0</td>
<td>0.419411</td>
</tr>
<tr>
<td>1</td>
<td>39.169</td>
<td>1.700935</td>
<td>0.542213</td>
<td>0.498959</td>
<td>0.440887</td>
<td>0</td>
<td>3.182994</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>0</td>
<td>0.015892</td>
<td>0.014625</td>
<td>0.018113</td>
<td>0</td>
<td>0.048630</td>
</tr>
<tr>
<td>2</td>
<td>42.831</td>
<td>0</td>
<td>0.052486</td>
<td>0.048299</td>
<td>0.059821</td>
<td>0</td>
<td>0.160606</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>0</td>
<td>0.118151</td>
<td>0.108726</td>
<td>0.134663</td>
<td>0</td>
<td>0.361540</td>
</tr>
</tbody>
</table>

**Step 5**
Compute new mean order numbers (MON), as shown Table B.7, utilizing the increments obtained in Table B.6, by adding the number of items plus the previous MON plus the current increment.

<table>
<thead>
<tr>
<th>Number of items</th>
<th>Time of Failure</th>
<th>Sum of row(increment)</th>
<th>Mean Order Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>0.419411</td>
<td>1.419411</td>
</tr>
<tr>
<td>1</td>
<td>39.169</td>
<td>3.182994</td>
<td>5.602405</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>0.048630</td>
<td>7.651035</td>
</tr>
<tr>
<td>2</td>
<td>42.837</td>
<td>0.160606</td>
<td>9.811641</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>0.361540</td>
<td>11.173181</td>
</tr>
</tbody>
</table>

**Step 6**
Compute the median ranks based on these new MONs as shown in Table B.8.

<table>
<thead>
<tr>
<th>Time</th>
<th>MON</th>
<th>Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.419411</td>
<td>0.0825889</td>
</tr>
<tr>
<td>39.169</td>
<td>5.602405</td>
<td>0.3952894</td>
</tr>
<tr>
<td>40</td>
<td>7.651035</td>
<td>0.5487781</td>
</tr>
<tr>
<td>42.837</td>
<td>9.811641</td>
<td>0.7106217</td>
</tr>
<tr>
<td>50</td>
<td>11.173181</td>
<td>0.8124983</td>
</tr>
</tbody>
</table>

**Step 7**
Compute new \( \hat{\beta} \) and \( \hat{n} \) using standard rank regression and based upon the data as shown in Table B.9.
Step 8 Return and repeat the process from Step 1 until an acceptable convergence is reached on the parameters (i.e., the parameter values stabilize).

Results
The results of the first five iterations are shown in Table B.10. Using Weibull++ with rank regression on X yields:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\eta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.845638</td>
<td>42.576422</td>
</tr>
<tr>
<td>2</td>
<td>1.830621</td>
<td>42.039743</td>
</tr>
<tr>
<td>3</td>
<td>1.828010</td>
<td>41.830615</td>
</tr>
<tr>
<td>4</td>
<td>1.828030</td>
<td>41.749708</td>
</tr>
<tr>
<td>5</td>
<td>1.828383</td>
<td>41.717990</td>
</tr>
</tbody>
</table>

$\hat{\beta}_{RRX} = 1.82890$, $\hat{\eta}_{RRX} = 41.69774$

The direct MLE solution yields:

$\hat{\beta}_{MLE} = 2.10432$, $\hat{\eta}_{MLE} = 42.31535$
Appendix: Log-Likelihood Equations

This appendix covers the log-likelihood functions and their associated partial derivatives for most of the distributions available in Weibull++. These distributions are discussed in more detail in the chapter for each distribution.

Weibull Log-Likelihood Functions and their Partials

The Two-Parameter Weibull

This log-likelihood function is composed of three summation portions:

\[
\ln(L) = \Lambda = \sum_{i=1}^{F_e} N_i \ln \left( \frac{T_i}{\eta} \right) - \sum_{i=1}^{S} N_i' \left( \frac{T_i'}{\eta} \right)^{\beta} + \sum_{i=1}^{F_I} N_i'' \ln \left[ \frac{e^{-\left(\frac{T_i''}{\eta}\right)^{\beta}} - e^{-\left(\frac{T_i''}{\eta}\right)}}{e^{-\left(\frac{T_i''}{\eta}\right)^{\beta}}} \right]
\]

where:
- \(F_e\) is the number of groups of times-to-failure data points
- \(N_i\) is the number of times-to-failure in the \(i^{th}\) time-to-failure data group
- \(\beta\) is the Weibull shape parameter (unknown a priori, the first of two parameters to be found)
- \(\eta\) is the Weibull scale parameter (unknown a priori, the second of two parameters to be found)
- \(T_i\) is the time of the \(i^{th}\) group of time-to-failure data
- \(S\) is the number of groups of suspension data points
- \(N_i'\) is the number of suspensions in the \(i^{th}\) group of suspension data points
- \(T_i'\) is the time of the \(i^{th}\) suspension data group
- \(F_I\) is the number of interval failure data groups
- \(N_i''\) is the number of intervals in the \(i^{th}\) group of data intervals
- \(T_i''\) is the beginning of the \(i^{th}\) interval
- \(T_i'''\) is the ending of the \(i^{th}\) interval

For the purposes of MLE, left censored data will be considered to be intervals with \(T_i''' = 0\).

The solution will be found by solving for a pair of parameters \((\hat{\beta}, \hat{\eta})\) so that \(\frac{\partial \Lambda}{\partial \beta} = 0\) and \(\frac{\partial \Lambda}{\partial \eta} = 0\). It should be noted that other methods can also be used, such as direct maximization of the likelihood function, without having to compute the derivatives.

\[
\frac{\partial \Lambda}{\partial \beta} = \frac{1}{\beta} \sum_{i=1}^{F_e} N_i + \beta \sum_{i=1}^{S} N_i \ln \left( \frac{T_i}{\eta} \right) - \sum_{i=1}^{F_e} N_i \left( \frac{T_i}{\eta} \right)^{\beta} \ln \left( \frac{T_i}{\eta} \right) - \sum_{i=1}^{S} N_i' \left( \frac{T_i'}{\eta} \right)^{\beta} \ln \left( \frac{T_i'}{\eta} \right) + \sum_{i=1}^{F_I} N_i'' \ln \left[ \frac{e^{-\left(\frac{T_i''}{\eta}\right)^{\beta}} - e^{-\left(\frac{T_i''}{\eta}\right)}}{e^{-\left(\frac{T_i''}{\eta}\right)^{\beta}}} \right]
\]

\[
\frac{\partial \Lambda}{\partial \eta} = \sum_{i=1}^{F_e} \frac{N_i}{\eta} - \beta \sum_{i=1}^{S} N_i \left( \frac{T_i}{\eta} \right) \ln \left( \frac{T_i}{\eta} \right) - \sum_{i=1}^{F_e} \frac{N_i (T_i)^{\beta}}{\eta} + \sum_{i=1}^{S} N_i' \left( \frac{T_i'}{\eta} \right)^{\beta} \ln \left( \frac{T_i'}{\eta} \right) e^{-\left(\frac{T_i'}{\eta}\right)} + \sum_{i=1}^{F_I} N_i'' \ln \left[ \frac{e^{-\left(\frac{T_i''}{\eta}\right)^{\beta}} - e^{-\left(\frac{T_i''}{\eta}\right)}}{e^{-\left(\frac{T_i''}{\eta}\right)^{\beta}}} \right]
\]

\[
- \frac{\left( \frac{T_i''}{\eta} \right)^{\beta} - e^{-\left(\frac{T_i''}{\eta}\right)}}{e^{-\left(\frac{T_i''}{\eta}\right)^{\beta}} - e^{-\left(\frac{T_i''}{\eta}\right)}} \frac{\left( \frac{T_i''}{\eta} \right)^{\beta}}{\eta}
\]

\[
- e^{-\left(\frac{T_i''}{\eta}\right)} - e^{-\left(\frac{T_i''}{\eta}\right)^{\beta}}
\]
The Three-Parameter Weibull

This log-likelihood function is again composed of three summation portions:

\[
\ln(L) = \Lambda = \sum_{i=1}^{F} N_i \ln \left[ \frac{\beta}{\eta} \left( \frac{T_i - \gamma}{\eta} \right)^{\beta - 1} e^{-\left( \frac{T_i - \gamma}{\eta} \right)^{\beta}} \right] - \sum_{i=1}^{S} N_i' \left( \frac{T_i' - \gamma}{\eta} \right)^{\beta}
\]

\[
+ \sum_{i=1}^{FI} N_i'' \ln \left[ e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)} - e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^{\beta}} \right]
\]

where:

- \( F \) is the number of groups of times-to-failure data points
- \( N \) is the number of times-to-failure in the \( i \)th time-to-failure data group
- \( \beta \) is the Weibull shape parameter (unknown a priori, the first of three parameters to be found)
- \( \eta \) is the Weibull scale parameter (unknown a priori, the second of three parameters to be found)
- \( T_i \) is the time of the \( i \)th group of time-to-failure data
- \( \gamma \) is the Weibull location parameter (unknown a priori, the third of three parameters to be found)
- \( S \) is the number of groups of suspension data points
- \( N_i' \) is the number of suspensions in \( i \)th group of suspension data points
- \( T_i' \) is the time of the \( i \)th suspension data group
- \( FI \) is the number of interval data groups
- \( N_i'' \) is the number of intervals in the \( i \)th group of data intervals
- \( T_{iB} \) is the beginning of the \( i \)th interval
- \( T_{iE} \) is the ending of the \( i \)th interval

The solution is found by solving for \( \beta, \eta, \gamma \) so that \( \frac{\partial \Lambda}{\partial \beta} = 0, \frac{\partial \Lambda}{\partial \eta} = 0 \) and \( \frac{\partial \Lambda}{\partial \gamma} = 0 \).
Appendix: Log-Likelihood Equations

\[
\frac{\partial \Lambda}{\partial \beta} = \frac{1}{\beta} \left[ \sum_{i=1}^{F_x} N_i + \sum_{i=1}^{F_x} N_i \ln \left( \frac{T_i - \gamma}{\eta} \right) - \sum_{i=1}^{F_x} N_i \left( \frac{T_i - \gamma}{\eta} \right)^\beta \ln \left( \frac{T_i - \gamma}{\eta} \right) \right] - \sum_{i=1}^{S} N_i' \left( \frac{T_i' - \gamma}{\eta} \right)^\beta \ln \left( \frac{T_i' - \gamma}{\eta} \right)
\]
\[
+ \sum_{i=1}^{F_1} N_i'' \left( \frac{T_i'' - \gamma}{\eta} \right)^\beta \ln \left( \frac{T_i'' - \gamma}{\eta} \right) e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta} - e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta}
\]
\[
+ \sum_{i=1}^{F_1} N_i'' \left( \frac{T_i'' - \gamma}{\eta} \right)^\beta \ln \left( \frac{T_i'' - \gamma}{\eta} \right) e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta} - e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta}
\]
\[
\frac{\partial \Lambda}{\partial \eta} = -\beta \left[ \sum_{i=1}^{F_x} N_i + \beta \sum_{i=1}^{F_x} N_i \left( \frac{T_i - \gamma}{\eta} \right)^\beta \right] + \sum_{i=1}^{S} N_i' \left( \frac{T_i' - \gamma}{\eta} \right)^\beta \left( \frac{\beta}{\eta} \right)
\]
\[
+ \sum_{i=1}^{F_1} N_i'' \left( \frac{T_i'' - \gamma}{\eta} \right)^\beta \ln \left( \frac{T_i'' - \gamma}{\eta} \right) e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta} - e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta}
\]
\[
- \sum_{i=1}^{F_1} N_i'' \left( \frac{T_i'' - \gamma}{\eta} \right)^\beta \ln \left( \frac{T_i'' - \gamma}{\eta} \right) e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta} - e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta}
\]
\[
\frac{\partial \Lambda}{\partial \gamma} = (1 - \beta) \sum_{i=1}^{F_x} \left( \frac{N_i}{T_i - \gamma} \right) + \sum_{i=1}^{F_x} N_i \left( \frac{T_i - \gamma}{\eta} \right)^\beta \left( \frac{\beta}{T_i - \gamma} \right)
\]
\[
+ \sum_{i=1}^{S} N_i' \left( \frac{T_i' - \gamma}{\eta} \right)^\beta \left( \frac{\beta}{T_i' - \gamma} \right)
\]
\[
+ \sum_{i=1}^{F_1} N_i'' \left( \frac{T_i'' - \gamma}{\eta} \right)^\beta e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta} - e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta}
\]
\[
- \sum_{i=1}^{F_1} N_i'' \left( \frac{T_i'' - \gamma}{\eta} \right)^\beta e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta} - e^{-\left( \frac{T_i'' - \gamma}{\eta} \right)^\beta}
\]

It should be pointed out that the solution to the three-parameter Weibull via MLE is not always stable and can collapse if \( \beta \sim 1 \). In estimating the true MLE of the three-parameter Weibull distribution, two difficulties arise. The first is a problem of non-regularity and the second is the parameter divergence problem, as discussed in Hirose [14].

Non-regularity occurs when \( \beta \leq 2 \). In general, there are no MLE solutions in the region of \( 0 < \beta < 1 \). When \( 1 < \beta < 2 \), MLE solutions exist but are not asymptotically normal, as discussed in Hirose [14]. In the case of non-regularity, the solution is treated anomalously.

Weibull++ attempts to find a solution in all of the regions using a variety of methods, but the user should be forewarned that not all possible data can be addressed. Thus, some solutions using MLE for the three-parameter Weibull will fail when the algorithm has reached predefined limits or fails to converge. In these cases, the user can change to the non-true MLE approach (in Weibull++ Application Setup), where \( \hat{\gamma} \) is estimated using non-linear regression. Once \( \hat{\gamma} \) is obtained, the MLE estimates of \( \hat{\beta} \) and \( \hat{\eta} \) are computed using the transformation \( T_i' = (T_i - \gamma) \).
Exponential Log-Likelihood Functions and their Partials

The One-Parameter Exponential

This log-likelihood function is composed of three summation portions:

\[
\ln(L) = \Lambda = \sum_{i=1}^{F_e} N_i \ln \left[ \lambda e^{-\lambda T_i} \right] - \sum_{i=1}^{S} N'_i \lambda T'_i + \sum_{i=1}^{FI} N''_i \ln \left[ e^{-\lambda T''_i} - e^{-\lambda T''_i} \right]
\]

where:
- \(F_e\) is the number of groups of times-to-failure data points
- \(N_i\) is the number of times-to-failure in the \(i^{th}\) time-to-failure data group
- \(\lambda\) is the failure rate parameter (unknown a priori, the only parameter to be found)
- \(T_i\) is the time of the \(i^{th}\) group of time-to-failure data
- \(S\) is the number of groups of suspension data points
- \(N'_i\) is the number of suspensions in the \(i^{th}\) group of suspension data points
- \(T'_i\) is the time of the \(i^{th}\) suspension data group
- \(FI\) is the number of interval data groups
- \(N''_i\) is the number of intervals in the \(i^{th}\) group of data intervals
- \(T''_i\) is the beginning of the \(i^{th}\) interval
- \(T''_i\) is the ending of the \(i^{th}\) interval

The solution will be found by solving for a parameter \(\hat{\lambda}\) so that \(\frac{\partial \Lambda}{\partial \lambda} = 0\). Note that for \(FI = 0\) there exists a closed form solution.

\[
\frac{\partial \Lambda}{\partial \lambda} = \sum_{i=1}^{F_e} N_i \left( \frac{1}{\lambda} - T_i \right) - \sum_{i=1}^{S} N'_i T'_i
\]

\[
- \sum_{i=1}^{FI} N''_i \left[ \frac{T''_i e^{-\lambda T''_i} - T''_i e^{-\lambda T''_i}}{e^{-\lambda T''_i} - e^{-\lambda T''_i}} \right]
\]

The Two-Parameter Exponential

This log-likelihood function for the two-parameter exponential distribution is very similar to that of the one-parameter distribution and is composed of three summation portions:

\[
\ln(L) = \Lambda = \sum_{i=1}^{F_e} N_i \ln \left[ \lambda e^{-\lambda T_i} (T_i - \gamma) \right] - \sum_{i=1}^{S} N'_i \lambda (T'_i - \gamma)
\]

\[
+ \sum_{i=1}^{FI} N''_i \ln \left[ e^{-\lambda (T''_i - \gamma)} - e^{-\lambda (T''_i - \gamma)} \right]
\]

where:
- \(F_e\) is the number of groups of times-to-failure data points
- \(N_i\) is the number of times-to-failure in the \(i^{th}\) time-to-failure data group
- \(\lambda\) is the failure rate parameter (unknown a priori, the first of two parameters to be found)
- \(\gamma\) is the location parameter (unknown a priori, the second of two parameters to be found)
- \(T_i\) is the time of the \(i^{th}\) group of time-to-failure data
- \(S\) is the number of groups of suspension data points
- \(N'_i\) is the number of suspensions in the \(i^{th}\) group of suspension data points
- \(T'_i\) is the time of the \(i^{th}\) suspension data group
- \(FI\) is the number of interval data groups
- \(N''_i\) is the number of intervals in the \(i^{th}\) group of data intervals
Appendix: Log-Likelihood Equations

- \( T_{L_i} \) is the beginning of the \( i^{th} \) interval
- \( T_{R_i} \) is the ending of the \( i^{th} \) interval

The two-parameter solution will be found by solving for a pair of parameters \((\hat{\lambda}, \hat{\gamma})\) such that \( \frac{\partial \Lambda}{\partial \lambda} = 0, \frac{\partial \Lambda}{\partial \gamma} = 0 \).

For the one-parameter case, solve for \( \frac{\partial \Lambda}{\partial \lambda} = 0 \).

\[
\frac{\partial \Lambda}{\partial \lambda} = \sum_{i=1}^{F_r} N_i \left[ \frac{1}{\lambda} - (T_i - \gamma) \right] - \sum_{i=1}^{S} N_i' (T_i' - \gamma) - \sum_{i=1}^{FI} N_i'' \left[ \frac{(T_{L_i}'' - \gamma) e^{-\lambda(T_{L_i}'' - \gamma)}}{e^{-\lambda(T_{L_i}'' - \gamma)}} - (T_{R_i}'' - \gamma) e^{-\lambda(T_{R_i}'' - \gamma)} \right]
\]

and:

\[
\frac{\partial \Lambda}{\partial \gamma} = \sum_{i=1}^{F_r} N_i \lambda + \sum_{i=1}^{S} N_i' \lambda + \sum_{i=1}^{FI} N_i'' \lambda
\]

Examination the derivative for \( \gamma \) will reveal that:

\[
\frac{\partial \Lambda}{\partial \gamma} = \left( \sum_{i=1}^{F_r} N_i + \sum_{i=1}^{S} N_i' + \sum_{i=1}^{FI} N_i'' \right) \lambda \equiv 0
\]

The above equation will be equal to zero only if either:

\[
\lambda = 0
\]

or:

\[
\left( \sum_{i=1}^{F_r} N_i + \sum_{i=1}^{S} N_i' + \sum_{i=1}^{FI} N_i'' \right) = 0
\]

This is an unwelcome fact, alluded to earlier in the chapter, that essentially indicates that there is no realistic solution for the two-parameter MLE for exponential. The above equations indicate that there is no non-trivial MLE solution that satisfies both \( \frac{\partial \Lambda}{\partial \lambda} = 0, \frac{\partial \Lambda}{\partial \gamma} = 0 \). It can be shown that the best solution for \( \gamma \) satisfying the constraint that \( \gamma \leq T_{R_i} \) is \( \gamma = T_1 \). To then solve for the two-parameter exponential distribution via MLE, one can set equal to the first time-to-failure, and then find a \( \lambda \) such that \( \frac{\partial \Lambda}{\partial \lambda} = 0 \).

Using this methodology, a maximum can be achieved along the \( \lambda \)-axis, and a local maximum along the \( \gamma \)-axis at \( \gamma = T_1 \), constrained by the fact that \( \gamma \leq T_1 \). The 3D Plot utility in Weibull++ illustrates this behavior of the log-likelihood function, as shown next:
Normal Log-Likelihood Functions and their Partials

The complete normal likelihood function (without the constant) is composed of three summation portions:

$$
\ln(L) = \Lambda = \sum_{i=1}^{F_s} N_i \ln \left[ \frac{1}{\sigma} \phi \left( \frac{T_i - \mu}{\sigma} \right) \right] \\
+ \sum_{i=1}^{S} N'_i \ln \left[ 1 - \Phi \left( \frac{T'_i - \mu}{\sigma} \right) \right] \\
+ \sum_{i=1}^{F_l} N''_i \ln \left[ \Phi \left( \frac{T''_i - \mu}{\sigma} \right) - \Phi \left( \frac{T''_i - \mu}{\sigma} \right) \right]
$$

where:

- $F_s$ is the number of groups of times-to-failure data points
- $N_i$ is the number of times-to-failure in the $i^{th}$ time-to-failure data group
- $\mu$ is the mean parameter (unknown a priori, the first of two parameters to be found)
- $\sigma$ is the standard deviation parameter (unknown a priori, the second of two parameters to be found)
- $T_i$ is the time of the $i^{th}$ group of time-to-failure data
- $S$ is the number of groups of suspension data points
- $N'_i$ is the number of suspensions in the $i^{th}$ group of suspension data points
- $T'_i$ is the time of the $i^{th}$ suspension data group
- $F_l$ is the number of interval data groups
- $N''_i$ is the number of intervals in the $i^{th}$ group of data intervals
- $T''_i$ is the beginning of the $i^{th}$ interval
- $T''_i$ is the ending of the $i^{th}$ interval

The solution will be found by solving for a pair of parameters $(\mu_0, \sigma_0)$ so that $\frac{\partial \Lambda}{\partial \mu} = 0$ and $\frac{\partial \Lambda}{\partial \sigma} = 0$. 
where:

\[
\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}
\]

and:

\[
\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt
\]

**Complete Data**

Note that for the normal distribution, and in the case of complete data only (as was shown in Basic Statistical Background), there exists a closed-form solution for both of the parameters or:

\[
\hat{\mu} = \bar{T} = \frac{1}{N} \sum_{i=1}^{N} T_i
\]

and:

\[
\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (T_i - \bar{T})^2
\]

\[
\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (T_i - \bar{T})^2}
\]
Lognormal Log-Likelihood Functions and their Partials

The general log-likelihood function (without the constant) for the lognormal distribution is composed of three summation portions:

\[
\ln(L) = \Lambda = \sum_{i=1}^{F} N_i \ln \left[ \frac{1}{T_i \sigma_T} \phi \left( \frac{\ln(T_i) - \mu'}{\sigma_T} \right) \right] \\
+ \sum_{i=1}^{S} N_i' \ln \left[ 1 - \Phi \left( \frac{\ln(T_i') - \mu'}{\sigma_T} \right) \right] \\
+ \sum_{i=1}^{F'} N_i'' \ln \left[ \Phi \left( \frac{\ln(T_i'') - \mu'}{\sigma_T} \right) - \Phi \left( \frac{\ln(T_i''') - \mu'}{\sigma_T} \right) \right]
\]

where:

- \( F \) is the number of groups of times-to-failure data points
- \( N_i \) is the number of times-to-failure in the \( i^{th} \) time-to-failure data group
- \( \mu' \) is the mean of the natural logarithms of the times-to-failure (unknown a priori, the first of two parameters to be found)
- \( \sigma_T \) is the standard deviation of the natural logarithms of the times-to-failure (unknown a priori, the second of two parameters to be found)
- \( T_i \) is the time of the \( i^{th} \) group of time-to-failure data
- \( S \) is the number of groups of suspension data points
- \( N_i' \) is the number of suspensions in the \( i^{th} \) group of suspension data points
- \( T_i' \) is the time of the \( i^{th} \) suspension data group
- \( F' \) is the number of interval data groups
- \( N_i'' \) is the number of intervals in the \( i^{th} \) group of data intervals
- \( T_i'' \) is the beginning of the \( i^{th} \) interval
- \( T_i''' \) is the ending of the \( i^{th} \) interval

The solution will be found by solving for a pair of parameters \((\mu', \sigma_T)\) so that \( \frac{\partial \Lambda}{\partial \mu'} = 0 \) and \( \frac{\partial \Lambda}{\partial \sigma_T} = 0 \):

\[
\frac{\partial \Lambda}{\partial \mu'} = \frac{1}{\sigma_T^2} \sum_{i=1}^{F} N_i (\ln(T_i) - \mu') \\
+ \frac{1}{\sigma_T} \sum_{i=1}^{S} N_i' \phi \left( \frac{\ln(T_i') - \mu'}{\sigma_T} \right) - \phi \left( \frac{\ln(T_i'') - \mu'}{\sigma_T} \right) \\
+ \frac{\sigma}{\sigma_T} \sum_{i=1}^{F'} N_i'' \phi \left( \frac{\ln(T_i'') - \mu'}{\sigma_T} \right) - \Phi \left( \frac{\ln(T_i''') - \mu'}{\sigma_T} \right)
\]
Appendix: Log-Likelihood Equations

\[
\begin{align*}
\frac{\partial \Lambda}{\partial \sigma_{T'}} &= \sum_{i=1}^{F_i} N_i \left( \frac{(\ln(T_i') - \mu')^2}{\sigma_{T'}^2} - \frac{1}{\sigma_{T'}} \right) \\
&+ \frac{1}{\sigma_{T'}} \sum_{i=1}^{S} N_i' \left( \frac{\ln(T_i') - \mu'}{\sigma_{T'}} \right) \phi \left( \frac{\ln(T_i') - \mu'}{\sigma_{T'}} \right) \\
&- \frac{1}{\sigma_{T'}} \sum_{i=1}^{FL_i} N_i'' \left( \frac{\ln(T_i''') - \mu'}{\sigma_{T'}} \right) \phi \left( \frac{\ln(T_i''') - \mu'}{\sigma_{T'}} \right) \Phi \left( \frac{\ln(T_i''') - \mu'}{\sigma_{T'}} \right) \\
&- \frac{1}{\sigma_{T'}} \sum_{i=1}^{FL_i} N_i'' \left( \frac{\ln(T_i''') - \mu'}{\sigma_{T'}} \right) \phi \left( \frac{\ln(T_i''') - \mu'}{\sigma_{T'}} \right) - \frac{1}{\sigma_{T'}} \sum_{i=1}^{FL_i} N_i'' \left( \frac{\ln(T_i''') - \mu'}{\sigma_{T'}} \right) \phi \left( \frac{\ln(T_i''') - \mu'}{\sigma_{T'}} \right)
\end{align*}
\]

where:

\[
\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} x^2}
\]

and:

\[
\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt
\]

Mixed Weibull Log-Likelihood Functions and their Partials

The log-likelihood function (without the constant) is composed of three summation portions:

\[
\ln(L) = \Lambda = \sum_{i=1}^{F_i} N_i \ln \left[ \sum_{k=1}^{Q_i} \beta_k \left( \frac{T_i}{\eta_k} \right)^{\beta_k-1} e^{-\left( \frac{T_i}{\eta_k} \right)^{\beta_k}} \right] \\
+ \sum_{i=1}^{S} N_i' \ln \left[ \sum_{k=1}^{Q} \rho_k e^{-\left( \frac{T_i'}{\eta_k} \right)^{\beta_k}} \right] \\
+ \sum_{i=1}^{FL_i} N_i'' \ln \left[ \sum_{k=1}^{Q} \rho_k e^{-\left( \frac{T_i''}{2\eta_k} \right)^{\beta_k}} \right]
\]

where:
- \( F_i \) is the number of groups of times-to-failure data points
- \( N_i \) is the number of times-to-failure in the \( i^{th} \) time-to-failure data group
- \( Q_i \) is the number of subpopulations
- \( \rho_k \) is the proportionality of the \( k^{th} \) subpopulation (unknown a priori, the first set of three sets of parameters to be found)
- \( \beta_k \) is the Weibull shape parameter of the \( k^{th} \) subpopulation (unknown a priori, the second set of three sets of parameters to be found)
- \( \eta_k \) is the Weibull scale parameter (unknown a priori, the third set of three sets of parameters to be found)
- \( T_i \) is the time of the \( i^{th} \) group of time-to-failure data
- \( S \) is the number of groups of suspension data points
- \( N_i' \) is the number of suspensions in the \( i^{th} \) group of suspension data points
- \( T_i' \) is the time of the \( i^{th} \) suspension data group
- \( FL_i \) is the number of groups of interval data points
- \( N_i'' \) is the number of intervals in the \( i^{th} \) group of data intervals
- \( T_{L_i}'' \) is the beginning of the \( i^{th} \) interval
- \( T_{R_i}'' \) is the ending of the \( i^{th} \) interval.
The solution will be found by solving for a group of parameters:

\( \left( \hat{\mu}, \hat{\sigma}, \hat{\eta}_1, \hat{\rho}_1, \hat{\beta}_1, \hat{\eta}_2, \hat{\rho}_2, \hat{\beta}_2, \ldots, \hat{\rho}_Q, \hat{\beta}_Q, \hat{\eta}_Q \right) \)

so that:

\[
\frac{\partial \Lambda}{\partial \mu} = 0, \quad \frac{\partial \Lambda}{\partial \sigma} = 0, \quad \frac{\partial \Lambda}{\partial \eta_1} = 0 \]
\[
\frac{\partial \Lambda}{\partial \rho_1} = 0, \quad \frac{\partial \Lambda}{\partial \beta_1} = 0, \quad \frac{\partial \Lambda}{\partial \eta_2} = 0 \]
\[
\vdots
\]
\[
\frac{\partial \Lambda}{\partial \rho_{Q-1}} = 0, \quad \frac{\partial \Lambda}{\partial \beta_{Q-1}} = 0, \quad \frac{\partial \Lambda}{\partial \eta_{Q-1}} = 0 \]
\[
\frac{\partial \Lambda}{\partial \beta_Q} = 0, \quad \text{and} \quad \frac{\partial \Lambda}{\partial \eta_Q} = 0
\]

**Logistic Log-Likelihood Functions and their Partialis**

This log-likelihood function is composed of three summation portions:

\[
\ln(L) = \Lambda = \sum_{i=1}^{F_i} N_i \ln \left( \frac{T_i - \mu}{\sigma \left( 1 + e^{\frac{T_i - \mu}{\sigma}} \right)^2} \right) - \sum_{i=1}^{S} N_i' \ln \left( 1 + e^\frac{\mu - T_i'}{\sigma} \right)
\]
\[
+ \sum_{i=1}^{F_i} N_i'' \ln \left( \frac{1 - e^\frac{T_i'' - \sigma}{\sigma}}{1 + e^\frac{T_i'' - \sigma}{\sigma}} \right)
\]

where:

- \( F_i \) is the number of groups of times-to-failure data points
- \( N_i \) is the number of times-to-failure in the \( i \)th time-to-failure data group
- \( \mu \) is the logistic shape parameter (unknown a priori, the first of two parameters to be found)
- \( \sigma \) is the logistic scale parameter (unknown a priori, the second of two parameters to be found)
- \( T_i \) is the time of the \( i \)th group of time-to-failure data
- \( S \) is the number of groups of suspension data points
- \( N_i' \) is the number of suspensions in \( i \)th group of suspension data points
- \( T_i' \) is the time of the \( i \)th suspension data group
- \( F_i \) is the number of interval failure data group
- \( N_i'' \) is the number of intervals in \( i \)th group of data intervals
- \( T_i'' \) is the beginning of the \( i \)th interval
- \( T_i'' \) is the ending of the \( i \)th interval

For the purposes of MLE, left censored data will be considered to be intervals with \( T_i = 0 \).

The solution of the maximum log-likelihood function is found by solving for \(( \hat{\mu}, \hat{\sigma} )\) so that \( \frac{\partial \Lambda}{\partial \mu} = 0 \) and \( \frac{\partial \Lambda}{\partial \sigma} = 0 \).

\[
\frac{\partial \Lambda}{\partial \mu} = -\frac{1}{\sigma} \sum_{i=1}^{F_i} N_i + 2 \sum_{i=1}^{F_i} N_i \frac{e^{\frac{T_i - \mu}{\sigma}}}{1 + e^{\frac{T_i - \mu}{\sigma}}} + \sum_{i=1}^{S} N_i' \frac{e^{\frac{\mu - T_i'}{\sigma}}}{1 + e^{\frac{\mu - T_i'}{\sigma}}}
\]
\[
- \sum_{i=1}^{F_i} N_i'' \left( \frac{e^{\frac{T_i'' - \sigma}{\sigma}}}{1 + e^{\frac{T_i'' - \sigma}{\sigma}}} + \frac{e^{\frac{T_i'' - \sigma}{\sigma}}}{1 + e^{\frac{T_i'' - \sigma}{\sigma}}} \right)
\]
The Loglogistic Log-Likelihood Functions and their Partials

This log-likelihood function is composed of three summation portions:

\[ \ln(L) = \Lambda = \sum_{i=1}^{I} N_i \ln \left( \frac{\ln(T_i) - \mu}{\sigma} \right) \]

\[ - \sum_{i=1}^{I} N_i' \ln \left( 1 + e^{-\ln(T_i') - \mu} \right) \]

\[ + \sum_{i=1}^{F_i} N_i'' \ln \left( \frac{1}{1 + e^{-\ln(T_i'') - \mu}} - \frac{1}{1 + e^{-\ln(T_i'') - \mu}} \right) \]

where:
- \( F_e \) is the number of groups of times-to-failure data points
- \( N_i \) is the number of times-to-failure in the \( i \)th time-to-failure data group
- \( \mu \) is the loglogistic shape parameter (unknown a priori, the first of two parameters to be found)
- \( \sigma \) is the loglogistic scale parameter (unknown a priori, the second of two parameters to be found)
- \( T_i \) is the time of the \( i \)th group of time-to-failure data
- \( S \) is the number of groups of suspension data points
- \( N_i' \) is the number of suspensions in \( i \)th group of suspension data points
- \( T_i' \) is the time of the \( i \)th suspension data group
- \( F_i \) is the number of interval failure data groups,
- \( N_i'' \) is the number of intervals in \( i \)th group of data intervals
- \( T_i'' \) is the beginning of the \( i \)th interval
- \( T_i''' \) is the ending of the \( i \)th interval

For the purposes of MLE, left censored data will be considered to be intervals with \( T_L'' = 0 \).

The solution of the maximum log-likelihood function is found by solving for \( \hat{\mu}, \hat{\sigma} \) so that \( \frac{\partial \Lambda}{\partial \mu} = 0, \frac{\partial \Lambda}{\partial \sigma} = 0. \)
Appendix: Log-Likelihood Equations

\[
\frac{\partial \Lambda}{\partial \mu} = -\frac{F_e}{\sigma} \sum_{i=1}^{F_e} N_i + \frac{2}{\sigma} \sum_{i=1}^{F_e} N_i \ln(T_i) - \mu \frac{\lambda(T_i) - \mu}{1 + e^{\frac{\lambda(T_i) - \mu}{\sigma}}}
\]

\[
+ \frac{1}{\sigma} \sum_{i=1}^{S} N_i^I \ln(T_i^I) - \mu \frac{e^{\frac{\ln(T_i^I) - \mu}{\sigma}}}{1 + e^{\frac{\ln(T_i^I) - \mu}{\sigma}}} - \frac{F_I}{\sigma}
\]

\[
+ \frac{1}{\sigma} \sum_{i=1}^{F_I} N_i'' \left( \frac{\ln(T_i'') - \mu}{\sigma} \frac{e^{\frac{\ln(T_i''') - \mu}{\sigma}}}{1 + e^{\frac{\ln(T_i''') - \mu}{\sigma}}} + \frac{\ln(T_i''') - \mu}{\sigma} \frac{e^{\frac{\ln(T_i''') - \mu}{\sigma}}}{1 + e^{\frac{\ln(T_i''') - \mu}{\sigma}}} \right)
\]

\[
\frac{\partial \Lambda}{\partial \sigma} = -\sum_{i=1}^{F_e} N_i \ln(T_i) - \mu \frac{\lambda(T_i) - \mu}{\sigma^2} - \frac{1}{\sigma} \sum_{i=1}^{F_e} N_i + \frac{2}{\sigma} \sum_{i=1}^{F_e} N_i \ln(T_i) - \mu \frac{\lambda(T_i) - \mu}{\sigma^2} \frac{e^{\frac{\ln(T_i) - \mu}{\sigma}}}{1 + e^{\frac{\ln(T_i) - \mu}{\sigma}}}
\]

\[
+ \frac{1}{\sigma} \sum_{i=1}^{S} N_i^I \ln(T_i^I) - \mu \frac{e^{\frac{\ln(T_i^I) - \mu}{\sigma}}}{1 + e^{\frac{\ln(T_i^I) - \mu}{\sigma}}}
\]

\[
+ \frac{1}{\sigma} \sum_{i=1}^{F_I} N_i'' \left( \frac{\ln(T_i'') - \mu}{\sigma} \frac{e^{\frac{\ln(T_i''') - \mu}{\sigma}}}{1 + e^{\frac{\ln(T_i''') - \mu}{\sigma}}} + \frac{\ln(T_i''') - \mu}{\sigma} \frac{e^{\frac{\ln(T_i''') - \mu}{\sigma}}}{1 + e^{\frac{\ln(T_i''') - \mu}{\sigma}}} \right)
\]

\[
- \frac{\ln(T_i''') - \mu}{\sigma} \frac{e^{\frac{\ln(T_i''') - \mu}{\sigma}}}{1 + e^{\frac{\ln(T_i''') - \mu}{\sigma}}} + \frac{\ln(T_i''') - \mu}{\sigma} \frac{e^{\frac{\ln(T_i''') - \mu}{\sigma}}}{1 + e^{\frac{\ln(T_i''') - \mu}{\sigma}}} \right)
\]

The Gumbel Log-Likelihood Functions and their Partials

This log-likelihood function is composed of three summation portions:

\[
\ln(L) = \Lambda = \sum_{i=1}^{F_e} N_i \ln \left( \frac{T_i - \mu}{e^{\frac{T_i - \mu}{\sigma}} - e^{\frac{T_i - \mu}{\sigma}}} \right)
\]

\[
- \sum_{i=1}^{S} N_i^I \ln \left( e^{\frac{T_i^I - \mu}{\sigma}} \right)
\]

\[
+ \sum_{i=1}^{F_I} N_i'' \ln \left( e^{\frac{T_i'' - \mu}{\sigma}} - e^{\frac{T_i'' - \mu}{\sigma}} \right)
\]

or:
Appendix: Log-Likelihood Equations

\[ \Lambda = \sum_{i=1}^{F_e} N_i \left( \frac{T_i - \mu}{\sigma} - \frac{T_i - \mu}{e^{\frac{T_i - \mu}{\sigma}}} \right) - \ln(\sigma) \sum_{i=1}^{F_e} N_i \]
\[ + \sum_{i=1}^{S} N'_i e^{\frac{T'_i - \mu}{\sigma}} \]
\[ + \sum_{i=1}^{F_I} N''_i \ln \left( \frac{\frac{T''_i - \mu}{\sigma} - \frac{T''_i - \mu}{e^{\frac{T''_i - \mu}{\sigma}}}}{e^{-\frac{T''_i - \mu}{\sigma}} - e^{-\frac{T''_i - \mu}{\sigma}}} \right) \]

where:

- \( F_e \) is the number of groups of times-to-failure data points
- \( N_i \) is the number of times-to-failure in the \( i^{th} \) time-to-failure data group
- \( \mu \) is the Gumbel shape parameter (unknown a priori, the first of two parameters to be found)
- \( \sigma \) is the Gumbel scale parameter (unknown a priori, the second of two parameters to be found)
- \( T_i \) is the time of the \( i^{th} \) group of time-to-failure data
- \( S \) is the number of groups of suspension data points
- \( N'_i \) is the number of suspensions in \( i^{th} \) group of suspension data points
- \( T'_i \) is the time of the \( i^{th} \) suspension data group
- \( F_I \) is the number of interval failure data groups
- \( N''_i \) is the number of intervals in \( i^{th} \) group of data intervals
- \( T''_i \) is the beginning of the \( i^{th} \) interval
- \( T''_{i'} \) is the ending of the \( i^{th} \) interval

For the purposes of MLE, left censored data will be considered to be intervals with \( T''_{i'} = 0 \).

The solution of the maximum log-likelihood function is found by solving for \((\hat{\mu}, \hat{\sigma})\) so that:

\[ \frac{\partial \Lambda}{\partial \mu} = 0, \frac{\partial \Lambda}{\partial \sigma} = 0, \]
\[ \frac{\partial \Lambda}{\partial \mu} = -\frac{1}{\sigma} \sum_{i=1}^{F_e} N_i + \frac{1}{\sigma} \sum_{i=1}^{F_e} N_i e^{\frac{T_i - \mu}{\sigma}} - \frac{1}{\sigma} \sum_{i=1}^{S} N'_i e^{\frac{T'_i - \mu}{\sigma}} \]
\[ + \frac{1}{\sigma} \sum_{i=1}^{F_I} N''_i \left( \frac{\frac{T''_i - \mu}{\sigma} - \frac{T''_i - \mu}{e^{\frac{T''_i - \mu}{\sigma}}}}{e^{-\frac{T''_i - \mu}{\sigma}} - e^{-\frac{T''_i - \mu}{\sigma}}} \right) \]
\[ \frac{\partial \Lambda}{\partial \sigma} = -\sum_{i=1}^{F_e} N_i \frac{T_i - \mu}{\sigma^2} - \frac{1}{\sigma} \sum_{i=1}^{F_e} N_i e^{\frac{T_i - \mu}{\sigma}} + \frac{1}{\sigma} \sum_{i=1}^{S} N_i \frac{T_i - \mu}{\sigma} e^{\frac{T_i - \mu}{\sigma}} \]
\[ - \frac{1}{\sigma} \sum_{i=1}^{S} N'_i e^{\frac{T'_i - \mu}{\sigma}} + \frac{1}{\sigma} \sum_{i=1}^{F_I} N''_i \left( \frac{\frac{T''_i - \mu}{\sigma} - \frac{T''_i - \mu}{e^{\frac{T''_i - \mu}{\sigma}}}}{e^{-\frac{T''_i - \mu}{\sigma}} - e^{-\frac{T''_i - \mu}{\sigma}}} \right) \]
The Gamma Log-Likelihood Functions and their Partialis

This log-likelihood function is composed of three summation portions:

\[
\ln(L) = \Lambda = \sum_{i=1}^{F_e} N_i \ln \left( \frac{e^{k_i (\ln(T_i) - \mu_i)} - e^{\ln(T_i) - \mu_i}}{T_i \Gamma(k_i)} \right) \\
+ \sum_{i=1}^{S} N_i' \ln \left( 1 - \Gamma \left( k_i, e^{\ln(T_i') - \mu_i} \right) \right) \\
+ \sum_{i=1}^{N_i''} \ln \left( \Gamma_1 \left( k_i, e^{\ln(T_i'') - \mu_i} \right) - \Gamma_1 \left( k_i, e^{\ln(T_i'') - \mu_i} \right) \right)
\]

or:

\[
\Lambda = -\sum_{i=1}^{F_e} N_i \ln(T_i) - k \sum_{i=1}^{F_e} N_i (\ln(T_i) - \mu) \\
- \sum_{i=1}^{S} N_i e^{\ln(T_i') - \mu} \\
+ \sum_{i=1}^{F_e} N_i' \ln \left( 1 - \Gamma_1 \left( k_i, e^{\ln(T_i') - \mu} \right) \right) \\
+ \sum_{i=1}^{N_i''} \ln \left( \Gamma_1 \left( k_i, e^{\ln(T_i'') - \mu} \right) - \Gamma_1 \left( k_i, e^{\ln(T_i'') - \mu} \right) \right)
\]

where:

- \(F_e\) is the number of groups of times-to-failure data points
- \(N_i\) is the number of times-to-failure in the \(i^{th}\) time-to-failure data group
- \(k\) is the gamma shape parameter (unknown a priori, the first of two parameters to be found)
- \(\mu\) is the gamma scale parameter (unknown a priori, the second of two parameters to be found)
- \(T_i\) is the time of the \(i^{th}\) group of time-to-failure data
- \(S\) is the number of groups of suspension data points
- \(N_i'\) is the number of suspensions in the \(i^{th}\) group of suspension data points
- \(T_i'\) is the time of the \(i^{th}\) suspension data group
- \(F_i\) is the number of interval failure data groups
- \(N_i''\) is the number of intervals in the \(i^{th}\) group of data intervals
- \(T_i''\) is the beginning of the \(i^{th}\) interval
- \(T_i''\) is the ending of the \(i^{th}\) interval

For the purposes of MLE, left censored data will be considered to be intervals with \(T_{L_i}'' = 0\).

The solution of the maximum log-likelihood function is found by solving for \((\hat{\mu}, \hat{\sigma})\) so that \(\frac{\partial \Lambda}{\partial \mu} = 0, \frac{\partial \Lambda}{\partial k} = 0\).
\[
\frac{\partial A}{\partial \mu} = -k \sum_{i=1}^{F_e} N_i + \sum_{i=1}^{F_e} N_i e^{\ln(T_i) - \mu} \\
+ \frac{1}{\Gamma(k)} \sum_{i=1}^{S} \frac{N_i e^{k \left( \ln(T_i') - \mu \right) - e^{\ln(T_i') - \mu}}}{1 - \Gamma_1 \left( k; e^{\ln(T_i') - \mu} \right)} \\
+ \frac{1}{\Gamma(k)} \sum_{i=1}^{F_i} N_i'' \left\{ \frac{e^{k e^{\ln(T_i'') - \mu} - e^{\ln(T_i'') - \mu}}}{\Gamma_1 \left( k; e^{\ln(T_i'') - \mu} \right)} - \Gamma_1 \left( k; e^{\ln(T_i'') - \mu} \right) \right\} \\
- \frac{F_e}{\Gamma(k)} \sum_{i=1}^{S} N_i \left( \ln(T_i) - \mu \right) - \frac{\Gamma'(k)}{\Gamma(k)} \sum_{i=1}^{F_e} N_i \\
- \sum_{i=1}^{S} N_i' \frac{\frac{\partial \Gamma_1 \left( k; e^{\ln(T_i') - \mu} \right)}{\partial k}}{1 - \Gamma_1 \left( k; e^{\ln(T_i') - \mu} \right)} \\
+ \sum_{i=1}^{F_i} N_i'' \left( \frac{\frac{\partial \Gamma_1 \left( k; e^{\ln(T_i'') - \mu} \right)}{\partial k}}{\Gamma_1 \left( k; e^{\ln(T_i'') - \mu} \right)} - \frac{\frac{\partial \Gamma_1 \left( k; e^{\ln(T_i'') - \mu} \right)}{\partial k}}{\Gamma_1 \left( k; e^{\ln(T_i'') - \mu} \right)} \right) \right) 
\]
Appendix: Life Data Analysis References