

A Practical Solution for Using Incomplete Warranty Data for Failure Analysis

Getting the Most out of Incomplete Warranty Data

By Karen Mohan, Application Engineer

Warranty data is a valuable source of information for analyzing the failure characteristics of your product. Knowing the failure trends of your products provides an array of benefits, including the ability to predict future returns, to estimate warranty-related claim costs, and to monitor product quality.

Typically, a reliability engineer uses *complete* warranty data in order to perform failure analyses. Obtaining complete warranty data means that a systematic and well-organized approach to warranty information collection is in place. In real-world scenarios this is often difficult or not possible. In these cases, you may only have partial warranty information, or an *incomplete* set of warranty data.

Recognizing that many companies have incomplete warranty data, Relex engineers developed a new approach to maximize the potential of this still valuable information. The methodology for this new technique was presented at the 2008 Reliability and Maintainability Symposium (RAMS). As the presenting engineer, I was pleased to receive the Evans-McElroy Best Paper award for this work on behalf of my co-authors and Relex Software Corporation.

Warranty Data Collection

In some cases, warranty data is systematically collected. Typically, product manufacturing data for each month of production is tracked and the corresponding monthly failure counts are derived from the warranty claims. When collected systematically, a failed product can be associated with its production period, and the product age at the time of failure can be derived. In this case, a *complete* warranty data set is available for failure analysis.

In other situations, depending on the methods used to collect warranty data, it may not be possible to know the failure ages of components. The information available each month might be limited to the volume of shipments and total claims or product returns. In these cases, the claims or returns are not associated with their specific production periods. The component age at the time of failure cannot be easily determined, resulting in a data set that is

incomplete. As a result of the incomplete data set, failure analysis using traditional predictive methods cannot be performed.

To show the differences in the treatments of complete and incomplete data, consider the following examples. The Acme Calculator Company and the Baker Calculator Company both collect warranty information. Acme systematically collects warranty data and has a complete warranty data set. Baker, however, has an incomplete data set where the age of the calculators at failure is unknown. (To demonstrate the effectiveness of the expectation-maximization (EM) algorithm, both Acme and Baker in this example have identical shipments and returns per month – something that would typically not occur in the real world!)

Warranty Information: Complete Data Sample

The Acme Calculator Company carefully records the number of calculators shipped in each month and the number of returns in each month corresponding to each shipment. With this complete data set, Acme is able to create a warranty returns chart, shown in Table 1. This style of data formatting is also referred as a “Nevada warranty chart” or “Idaho chart.”

Calculators Shipped		Calculators Returned					Total
Month	Volume	Jan	Feb	Mar	Apr	May	
Jan	1000	3	9	6	10	6	34
Feb	1100	--	4	10	6	12	32
Mar	1240	--	--	4	11	8	23
Apr	1250	--	--	--	4	11	15
May	1350	--	--	--	--	5	5
Total		3	13	20	31	42	

Table 1 – Acme Warranty Returns Chart

Since Acme has the warranty returns data available in this format, it is easy for them to convert the shipping and warranty returns data into a standard failure data form consisting of failures and suspensions, as shown in Table 2. The rows in Table 2 corresponding to the event type “Failure” indicate that the failures have occurred between the specified start and end times. For example, the first row indicates that 20 calculators have failed in the first month of operation (between zero months and one month). The rows corresponding to the event type “Suspension” indicate the number of calculators that have not failed by the last month in the analysis. The table acknowledges that the calculators could fail anywhere between the start time and a future, undefined end time, labeled as infinity. For example, the last row indicates that 966 calculators have not failed up to the end of the fifth month of operation.

Start Time	End Time	Number of Failed Calculators	Event Type
0	1	20	Failure
1	2	41	Failure
2	3	20	Failure
3	4	22	Failure
4	5	6	Failure
1	∞	1345	Suspension
2	∞	1235	Suspension
3	∞	1217	Suspension
4	∞	1068	Suspension
5	∞	966	Suspension

Table 2 – Standard Interval Format Failure Data Set

With the data set formatted as in Table 2, Acme can easily perform standard failure data analysis techniques. The Acme Calculator Company can then use a Weibull analysis tool, such as Relex Weibull, to analyze the data given in this format. Using Relex Weibull and the data from Table 2, the results for the two Weibull parameters (β - the slope, or shape parameter, and η - the characteristic life, or scale parameter) are $\eta = 53.2$ months and $\beta = 1.38$.

Warranty Information: Incomplete Data Sample

In many situations and for diverse reasons, information about the ages of component failures is not available, resulting in incomplete warranty data. Such is the case for the Baker Calculator Company. The information available each month is limited to the volume of calculators shipped (N_i , where i is the month shipped) and the total number of warranty claims or product returns (R_i). Baker's data is not able to be formatted into a Nevada style warranty chart. They are lacking the exact number of returned calculators corresponding to each month and shipment combination shown in the upper triangular matrix of the Acme Warranty Returns Chart (Table 1). Without the association of returns to shipments, Baker is only able to represent their data as shown in Table 3.

Month	Jan	Feb	Mar	Apr	May
MonthID	1	2	3	4	5
Number of Calculators Shipped	1000	1100	1240	1250	1350
Number of Calculators Returned	3	13	20	31	42

Table 3 – Baker's Incomplete Warranty Data

From the January column, it is easy to say that the three returned calculators actually belong to the January shipment (Shipment 1). However, in the February column, the Baker company cannot say how the 13 returns are distributed

between Shipments 1 and 2. Baker's returns could have occurred in any one of the following combinations: (0, 13), (1, 12), ..., (13, 0). Moving into March, Baker's uncertainty increases with the 20 returned calculators recorded for that month. These returns could have occurred with any one of the following combinations: (0, 0, 20), (0, 1, 19), ..., (20, 0, 0). The number of possible combinations increases from the hundreds to the thousands to far greater values with the number of intervals, and thus it becomes difficult to compute the Likelihood Function.

The challenge for Baker Calculator Company is obvious. Somewhere within the range of all possible combinations of values for each month of shipments is the data that they need to work with. Is there a way to estimate these hidden values and use them to generate the data they need to calculate failure characteristics?

Incomplete Data – Filling in the Blanks

Relex engineers proposed a valuable solution for this situation. By employing the expectation-maximization (EM) algorithm, the maximum likelihood estimates (MLE) of parameters in probability models, where the parameters are unobserved and hidden, can be determined. The EM algorithm alternates between performing an expectation step (E-step), which computes an expectation of the likelihood by including the hidden variables as if they were observed, and a maximization step (M-step), which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the E-step. The parameters found on the M-step are then used to begin another E-step, and the process is repeated.

Relex engineers found that the EM algorithm can be used to estimate the failure distribution parameters for an incomplete warranty data set. The flow diagram in Figure 1 shows the steps to apply the EM algorithm to the incomplete warranty data set. The explanation and examples that follow assume that the data is a two-parameter Weibull distribution, where β is the shape parameter, and η is the scale parameter.

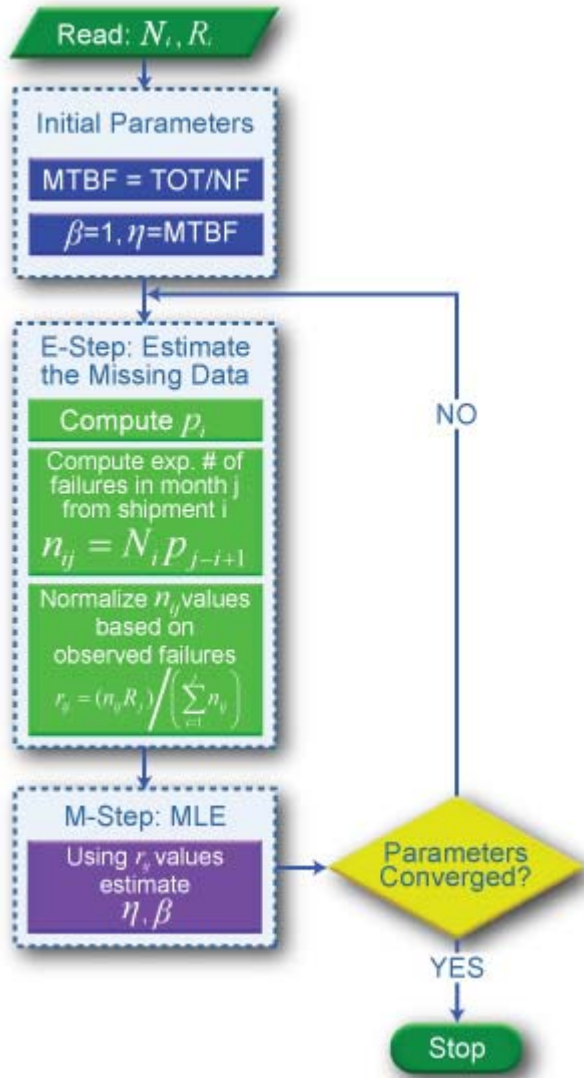


Figure 1- Flowchart of the Proposed Method

Here is a detailed explanation of the steps shown in Figure 1 for applying the EM algorithm to incomplete warranty data:

1. Assuming exponential distribution, compute the MTBF (Mean Time Between Failures).

$$MTBF = \frac{\text{Total Operating Time}}{\text{Total Number of Failures}} = \frac{TOT}{NF}$$

Where,

$$NF = \sum_{i=1}^M R_i$$

The number of warranty return periods is denoted as M . The items from shipment i are in the field for $(M-i+1)$ months. If there are no failures, they all together operate for $N_i(M-i+1)$ effective months. If there are R_j returns in month j , the

effective reduction in the operating time is $R_j(M-j+0.5)$. Therefore:

$$TOT = \sum_{i=1}^K N_i(M-i+1) - \sum_{j=1}^M R_j(M-j+0.5)$$

2. Set $\eta = \text{MTBF}$ and $\beta = 1$ as the initial parameters.
3. E-Step: Estimate the missing data.
Using the current η and β values, find the expected number of failures in month j from shipment i for all combinations of i and j values.

$$n_{ij} = N_i p_{j-i+1}$$

Parameter p_i is the probability of occurrence of a failure in the interval i . Assuming the failure distribution is a two-parameter Weibull, then p_i is calculated by:

$$p_i = \exp\left\{-\left(\frac{t_{i-1}}{\eta}\right)^\beta\right\} - \exp\left\{-\left(\frac{t_i}{\eta}\right)^\beta\right\}$$

If the time is measured in months, then $t_i = i$. Find the total number of failures that are expected to be failed in month j .

$$EF_j = \sum_{i=1}^M n_{ij} = \sum_{i=1}^j n_{ij}$$

Normalize the expected failures based on the observed failures. Normalization is done so that each column sums to the observed number of returns seen in that month.

$$r_{ij} = \frac{n_{ij}}{EF_j} \cdot R_j$$

In this equation, r_{ij} is the number of items returned in month j from shipment i . Now we have data in the format of Table 1.

4. M-Step: Find the maximum likelihood estimates (MLE) for η and β using the estimated failures in E-step, i.e., r_{ij} values. Relx Weibull software can be used to find the MLEs in this step.
5. Check the convergence. If the current η and β values match with the previous values within a prescribed accuracy, stop the iterations. Otherwise, go back to the E-step (Step 3).

Applying the Algorithm

Example 1: Baker Company's Incomplete Data Set

The Baker Calculator company, armed with their warranty information as shown in Table 3, can employ the Relx algorithm to get the most out of their incomplete data. The goal is to find estimates for the Weibull parameters η and β and to identify point of convergence where η and β remain stable to two decimal places.

Applying equations (1)-(3) to the data in Table 3 generates an MTBF = 154.0688. Thus, the first estimate is $\eta = 154.0688$ and $\beta = 1$. From equations (4)-(7) in the E-step, the expected number of failures are calculated and shown in Table 4.

Calculators Shipped		Calculators Returned					Total
Month	Volume	Jan	Feb	Mar	Apr	May	
Jan	1000	3	6.17	5.95	6.68	6.97	28.77
Feb	1100	--	6.83	6.58	7.40	7.72	28.53
Mar	1240	--	--	7.47	8.40	8.76	24.63
Apr	1250	--	--	--	8.52	8.89	17.41
May	1350	--	--	--	--	9.66	9.66
Total		3	13	20	31	42	

Table 4 – Estimated Warranty Returns Chart (Iteration 1)

According to the EM algorithm flowchart, the iterations are repeated until the defined level of convergence is reached. From the first iteration data shown in Table 4, the MLE estimates in the M-step are $\eta = 120.763$ and $\beta = 1.067$. The η and β from the first iteration have changed more than 0.01 from the previous estimates of η and β , so the process goes back to the E-step (Step 3) for another iteration. Both the values of η and β stabilized to two decimal places after 38 iterations.

The results shown in Table 5 correspond to 50 iterations. The MLE estimates at iteration 50 are: $\eta = 53.88$ and $\beta = 1.3779$. The MLE estimates for the complete data shown for the Acme Company in Table 1 are: $\eta = 53.203$ and $\beta = 1.38$. Even though the estimated missing data for Baker Company in Table 5 are not exactly matched with the complete data shown for Acme in Table 1, the final estimated results for η and β are a close match. Keep in mind that the same set of incomplete data can be obtained from different instances of complete data.

Calculators Shipped		Calculators Returned					Total
Month	Volume	Jan	Feb	Mar	Apr	May	
Jan	1000	3	7.68	7.82	8.95	9.47	36.92
Feb	1100	--	5.31	7.13	8.73	9.56	30.74
Mar	1240	--	--	5.05	8.15	9.56	22.76
Apr	1250	--	--	--	5.16	7.98	13.15
May	1350	--	--	--	--	5.42	5.42
Total		3	13	20	31	42	

Table 5 – Baker Calculator Company Estimated Warranty Returns Chart (Iteration 50)

Example 2: Data over 10 Months

The Baker Company continues to collect warranty data as they have before, and the resulting information for ten months is shown in Table 6.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
Month ID	1	2	3	4	5	6	7	8	9	10
Number of Calculators Shipped	1000	1100	1240	1250	1350	1400	1500	1600	1650	1700
Number of Calculators Returned	3	13	20	31	42	50	63	75	93	109

Table 6 – Baker Calculator Company Incomplete Warranty Data for Ten Months

Using the Relex algorithm, the initial MTBF estimate from this data is 136.115. After the first EM iteration, $\eta = 121.001$ and $\beta = 1.039$. Both the values of η and β stabilize to two decimal places after 41 iterations. The MLE estimates at iteration 41 are: $\eta = 71.98$ and $\beta = 1.257$.

Conclusion

The Acme Calculator Data, with its systematic approach to recording warranty information, is in the best position for calculating failure characteristics based on their warranty data. But now a company like the Baker Calculator Company, which has warranty claims data with incomplete or missing information about component failure times, has a method available to estimate the field failure characteristics from the data that they do have on hand. With this innovative approach developed by Relex engineers, a method based on the EM algorithm can be used successfully to estimate the failure distribution parameters from the incomplete warranty data. The results from the EM algorithm, while not exact, are reasonably accurate in most cases. The accuracy increases with the number of months and the volume of shipment.

More importantly, once the parameters for fitting a life distribution to a warranty claims data set have been estimated, additional calculated results can include reliability over time, mean life, failure rate, and warranty time, providing companies like Baker Calculator Company with information useful in making reliability and quality decisions.

Relex Weibull provides full Weibull analysis capabilities in a powerful statistical tool wrapped in a friendly and flexible user interface. Relex Weibull has a broad range of applications and can be applied to any industry or process with ease. It moves the complexity of the statistical methodology into a powerful behind-the-scenes engine, allowing more accurate and efficient life data analysis. Relex Weibull goes beyond analyzing data sets, augmenting these fundamental activities with a broad range of product-enhancing functions, such as a built-in test planning calculator, an optimal replacement feature, and a Barringer process reliability analysis function. For more information about Relex Weibull, visit www.relex.com/products/weibull.asp.