International Journal of Reliability and Applications
Vol. 14, No. 2, pp. 107-114, 2013

Empirical modelling approaches to modelling failures

Jaiwook Baik*
Department of Information Statistics, Korea National Open University, Seoul, Korea

Jinnam Jo
Department of Statistics & Information Science, Dongduk Women's University
Seoul, Korea

Received 3 October 2013; revised 6 December 2013; accepted 9 December 2013

Abstract. Modelling of failures is an important element of reliability modelling. Empirical modelling approach suitable for complex item is explored in this paper. First step of the empirical modelling approach is to plot hazard function, density function, Weibull probability plot as well as cumulative intensity function to see which model fits best for the given data. Next step of the empirical modelling approach is select appropriate model for the data and fit the parametric model accordingly and estimate the parameters.

Key Words: *Empirical modelling, exploratory data analysis, Weibull probability plot, Weibull mixture*

1. INTRODUCTION

Reliability of a product conveys the concept of dependability and the absence of failures. Reliability theory deals with various aspects of product reliability and encompasses various reliability issues. These include reliability engineering to design and manufacture reliable products, reliability management to manage the activities during the design and manufacture of products and the operation of unreliable products, and reliability modelling to build models to obtain solutions to a variety of reliability related problems in predicting, estimating, and optimising the performance of unreliable products. The modelling of failures is an important element of reliability modelling. In one-dimensional failure modelling, failures are random points along a one-dimensional axis representing age or usage. For some products such as automobiles, however, failures depend on age and usage and, in this case, failures are random points in a two-dimensional plane with the two axes representing age and usage. Models play an important role in

* Corresponding Author.
E-mail address : jbaik@knou.ac.kr

decision-making. Many different types of models are used and these can be found in Lawless (1982), Blischke and Murthy (1994 and 2000) and Meeker and Escobar (1998). One-dimensional modelling has received considerable attention and so there is a vast literature covering this area. In contrast, two-dimensional failure modelling has received relatively little attention.

The outline of the chapter is as follows. In Section 2 we discuss two different approaches to modelling first and subsequent failures, namely empirical modelling approach and white-box approach, along with specific procedure in empirical modelling process. Section 3 deals with the first step of the procedure which is exploratory analysis of data, and Section 4 deals with the model selection. And finally how to estimate the parameters and how to validate the model are discussed in Section 5. We conclude with some comments and remarks in Section 6.

2. MODELLING FAILURES

2.1 Approaches to Modelling

Products can vary from simple to complex item. A product can be viewed as a system consisting of several parts and can be decomposed into a hierarchy of levels with the system at the top level and components at the lowest level and several levels such as sub-system and sub-assembly in between. The failure of a product is due to the failure of one or more of its components.

The occurrence of failure depends on several factors. These include decisions made during the design and manufacture of the product, usage intensity and operating environment, and the maintenance actions carried out during the operating life.

The approach to modelling depends on the kind of information available and the goal of the modelling. There are two basic approaches to modelling failures as indicated below.

(i) Empirical modelling Approach: Here the modelling is based solely on failure and censored data for similar items. This approach is used when there is very little understanding of the different mechanisms that lead to product failure or when the unit is too complex. This approach is also known as *data based or black-box approach*.

(ii) White-box Approach: Here the failure modelling at the component level is based on the different mechanisms that lead to failure. At the system level, the failure is done in terms of the failures of the different components. This approach is also known as physics *based modelling*.

2.2 First and Subsequent Failures

One needs to differentiate between the first failure and subsequent failures. The subsequent failures depend on the type of actions used to rectify the failures. In the case of a non-repairable item, the failed item needs to be replaced by either a new or used item to make the product functional. In the case of a repairable item, the product can be made operational through the repair of the failed item. Three types of repair are indicated below:

(i) Minimal repair, which restores the item to the condition just before failure

(ii) Perfect repair (which makes the item as good as new)

(iii) Imperfect repair that results in the item being better than what it was prior to failure but not as-good-as-new.

2.3 Empirical Modelling Process

In the empirical modelling approach to modelling failures, the data are the starting point that forms the basis for the model building. The data can be either item failure times or counts of item failures over an interval. In the former case, the data are continuous valued and in the latter case they are integer valued.

Lifetime data can be complete, censored or truncated. In the case of complete data, the data relate to the age at failure. With censored data, the lifetimes are only known to exceed some values. This could result from the item not having failed during the period of observation and hence still being operational for a certain length of time afterwards. When the data are the failures of an item over different disjoint time intervals we have grouped data. When failures of different components are pooled together, we have pooled data. In both cases, the data can be considered as categorical or, if they are in the form of counts, they are discrete valued.

The modelling process involves the following three steps.

Step 1: Exploratory Analysis of Data
Step 2: Model Selection
Step 3: Parameter Estimation and Model Validation

These are discussed further in Sections 3 - 5.

3. EXPLORATORY DATA ANALYSIS

The first step in constructing a model is to explore the data through plots of the data. By so doing, information can be extracted to assist in model selection. The plots can be either nonparametric or parametric and the plotting is different for perfect repair and imperfect repair situations. The data comprises both the failure times and the censored times.

3.1 Perfect Repair

(1) Plot of Hazard Function [Nonparametric]

The procedure (for complete or censored data) is as follows:

Divide the time axis into cells with cell i defined by $[t_i, t_{i+1}), i \geq 0$, $t_0 = 0$ and $t_i = i\delta$, where is the cell width. Let

N_i^f : Number of items with failure times in cell $i, i \geq 0$

N_i^c : Number of items with censoring times in cell $i, i \geq 0$

$N_i^{f|ri}$: Number of failures in cells i and beyond [= $\sum_{j=i}^{\infty} N_j^f$].

Similarly define $N_i^{c|ri}$ for censored data.

The estimator of the hazard function is given by

$$\hat{h}_i = \frac{N_i^f}{N_i^{f|ri} + N_i^{c|ri}},\ i \geq 0$$

(2) Plot of Density Function [Nonparametric]

The simplest form of nonparametric density estimator is the histogram. Assuming the data is complete, the procedure is to calculate the relative frequencies for each cell,

$$\hat{f}_i = \frac{N_i^f}{\sum_{j=0}^{\infty} N_j^f},$$

and then plot these against the cell midpoints. As histograms can be very unreliable for exploring the shape of the data, especially if the data set is not large, it is desirable to use more sophisticated density function estimators (see, for example, Silverman (1986)).

(3) Weibull Probability Plots [Parametric]

The Weibull Probability Plot (WPP) provides a systematic procedure to determine whether one of the Weibull based models is suitable for modelling a given data set or not, and is more reliable than considering just a simple histogram. It is based on the Weibull transformations

$$y = \ln(-\ln(1 - F(t))) \quad \text{and} \quad x = \ln(t).$$

The plot of *y* versus *x* gives a straight line if *F*(*t*) is a two-parameter Weibull distribution. Thus, if *F*(*t*) is estimated for (complete) data from a Weibull distribution and the equivalent transformations and plot obtained, then a “rough” linear relationship should be evident. To estimate *F*(*t*), we need an empirical estimate of $F(t_i)$ for each failure time t_i. Assuming the t_i’s are ordered, so that $t_1 \leq t_2 \leq \ldots \leq t_n$, a simple choice (in the case of complete data) is to take the empirical distribution function

$$\hat{F}\left(t_i\right) = i/(n+1). \tag{3.1}$$

We then plot $\hat{y}_i = \ln(-\ln(1 - \hat{F}(t_i)))$ versus $x_i = \ln(t_i)$ and assess visually whether a straight line could describe the points.

We illustrate by considering real data. The data refers to failure times and usage (defined through distance travelled between failures) for a component of an automobile engine over the warranty period given by three years and 60,000 miles. Here we only look at the failure times in the data set. Figure 1 shows a Weibull Probability Plot of the inter-failure times of a component that we shall call Component A. This clearly shows a curved relationship and so a simple Weibull model would not be appropriate.

Note that the plotting of the data depends on the type of data. So, for example, the presence of censored observations would necessitate a change in the empirical failure estimates. See Nelson (1982) for further details.

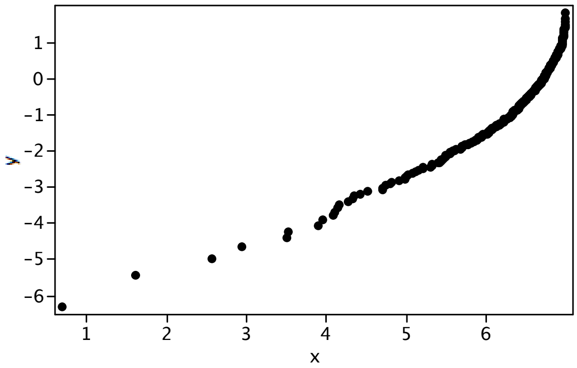

Figure 1. WPP of days to failure of Component A

3.2 Minimal Repair

(1) Plot of Cumulative Intensity Function [Non-parametric]

The procedure is as follows: With δ and the cells defined as before, let

M : Number of items at the start

N_i^f : Total number of failures over $[0, i\delta)$

M_i^c : Number of items censored in cell i

Λ_i : Cumulative intensity function till cell i

The estimator of the cumulative intensity is given by

$$\hat{\Lambda}_0 = \frac{N_0^f}{M} \text{ and } \hat{\Lambda}_i = \frac{N_i^f - \sum_{j=0}^{i-1} M_j^c \hat{\Lambda}_j}{[M - \sum_{j=0}^{i-1} M_j^c]}, \; i \geq 1$$

(2) Graphical Plot [Parametric]

When the failure distribution is a two-parameter Weibull distribution we see that a plot of $y = \ln(E[N(t)]/t)$ versus $x = \ln(t)$ is a straight line. Duane (1964) proposed plotting $y = \ln(N(t)/t)$ versus $x = \ln(t)$ to determine if a Weibull distribution is a suitable model or not to model a given data set. For a critical discussion of this approach, see Rigdon and Basu (2000).

4. MODEL SELECTION

We saw in Figure 1 that a simple Weibull model was clearly not adequate to model the failures of Component A. However, there are many extensions of the Weibull model that can fit a variety of shapes. Murthy et al. (2003) give a taxonomic guide to such models and give steps for model selection. This particular curve is suited to modelling with a mixture of two Weibull components. Figure 2 shows the WPP plot above with the transformed probability curve for this mixture. This seems to fit the pattern quite well, though it misses some of the curve present in the few small failure times at the left.

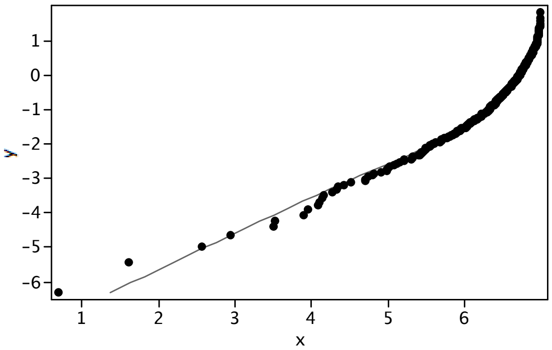

Figure 2. WPP of Component A failures with Weibull mixture

Figure 3 gives the empirical plot of the density function and the density function based on the mixture model. As can be seen, the model matches the data reasonably well. The plots illustrate the way in which the second Weibull component is being used. The nonparametric density estimate suggests that there is a small failure mode centered around 200 days. The second Weibull component, with a weight of 24.2%, captures these early failure times while the dominant component, with a weight of 75.8%, captures the bulk of the failures.

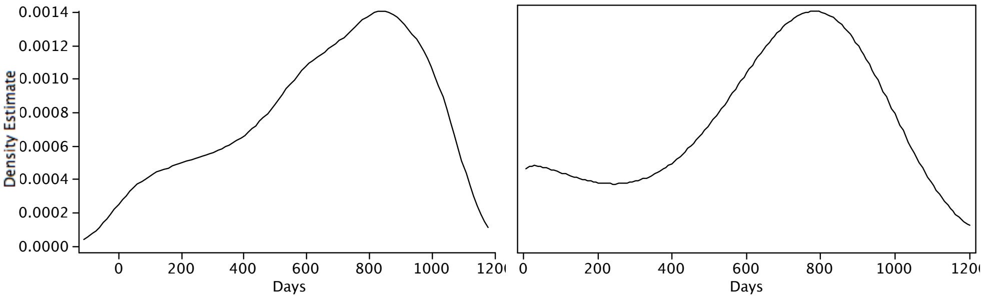

Figure 3. Empirical density (left) and Weibull mixture density (right) for Component A

5. PARAMETER ESTIMATION AND VALIDATION

The model parameters can be estimated either based on the graphical plots or by using statistical methods. Many different methods such as method of moments, method of maximum likelihood, least squares, Bayesian and so on have been proposed. The graphical methods yield crude estimates whereas the statistical methods are more refined and can be used to obtain confidence limits for the estimates.

The parameters for the Weibull mixture model in Figure 2 were estimated by minimizing the squared error between the points and the curve on the Weibull probability plot. The estimates are

$$\hat{p} = 0.242,\quad \hat{\beta}_1 = 1.07,\quad \hat{\beta}_2 = 4.32,\quad \hat{\eta}_1 = 381 \quad \text{and} \quad \hat{\eta}_2 = 839 .$$

Similar estimates can be obtained without computer software using the graphical methods given by Jiang & Murthy (1995).

Alternatively, we can use the standard statistical approach of maximum likelihood estimation to get the parameter estimates. We find

$$\hat{p} = 0.303, \quad \hat{\beta}_1 = 1.46, \quad \hat{\beta}_2 = 5.38, \quad \hat{\eta}_1 = 383 \quad \text{and} \quad \hat{\eta}_2 = 870$$

These values are less affected by the short failures times at the left.

Validation of statistical models is highly dependent on the nature of the models being used. In many situations, it can simply involve an investigation of the shape of the data through plots such as quantile-quantile plots and through tests for goodness of fit. Many introductory statistics texts cover these plots and tests. In more complex situations, these approaches need to be used on residuals obtained after fitting a model involving explanatory variables. An alternative approach which can be taken when the data set is large, is to take a random sample from the data set, fit the model(s) to this sub-sample and then evaluate through plots and tests how well the model fits the sub-sample consisting of the remaining data.

To exemplify model validation, 80% of the data was randomly taken and the above mixed Weibull model fitted. The fitted model was then compared using a WPP to the remaining 20% of the data. The plot on the left in Figure 4 shows a Weibull plot of 80% of the failure data for Component A, together with the Weibull mixture fit to the data. The remaining 20% of failure data are plotted on the right. The Weibull mixture curve with the same parameters as in the plot on the left has been added here. Apart from the one short failure time, this curve seems to fit the test data quite well. This supports the use of the Weibull mixture for modelling the failures of this component.

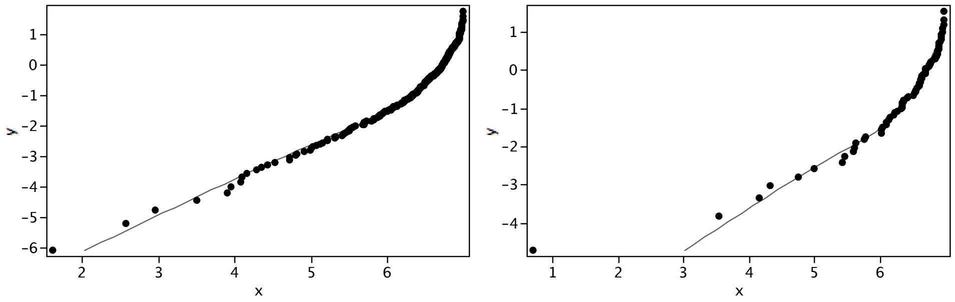

Figure 4. Weibull plots of fitting data (left) and test data (right) for Component A

7. CONCLUSIONS

The approach to modelling depends on the kind of information available and the goal of the modelling. There are two basic approaches to modelling failures; namely empirical modelling approach suitable for complex item, and white-box approach suitable for component that can lead to failure based on certain mechanism.

In this paper empirical modelling approach has been explored in depth. First step of the empirical modelling approach for perfect and minimal repair data is to plot hazard function, density function, Weibull probability plot as well as cumulative intensity function. Next step of the empirical modelling approach is select appropriate model for the data and fit the parametric model accordingly and estimate the parameters.

In this paper we have looked at the case where we are only concerned about one attribute, such as age only. But for some products such as automobiles failure depends both on age and usage. So in the near future we need to develop models that can be applied to two-dimensional data along with the empirical plots to help in the model selection and validation.

REFERENCES

Blischke, W. R., Murthy, D. N. P. (1994). *Warranty cost analysis*, Marcel Dekker, Inc., New York.

Blischke, W. R., Murthy, D. N. P. (2000). *Reliability*, Wiley, New York.

Duane, J. T. (1964). Learning curve approach to reliability monitoring, *IEEE Transactions on Aerospace*, **40**, 563-566.

Jiang, R. and Murthy, D. N. P. (1995). Modeling failure data by mixture of two Weibull distributions, *IEEE Transactions on Reliability*, **44**, 478-488.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, Inc., New York .

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical methods for reliability data*, John Wiley & Sons, Inc., New York.

Murthy, D. N. P., Xie, M. and Jiang, R. (2003). *Weibull Models*, Wiley & Sons, New York

Nelson, W. (1982). *Applied life data analysis*, John Wiley & Sons, Inc., New York.

Rigdon, S. E. and Basu, A. P. (2000). *Statistical methods for the reliability of repairable systems*, Wiley, New York.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, London.