The Beta Failure Rate Distribution

T. Moore and C. D. Lai
Department of Statistics
Massey University
Palmerston North
New Zealand
t.moore@massey.ac.nz/c.d.lai@massey.ac.nz

Abstract

The hazard-rate or failure-rate is used in reliability theory as a measure of the reliability of a component. It measures the failure rate at time, t, of those components that have not already failed before time, t. If failures are completely random, then the failure rate is constant. This corresponds to an exponential distribution of time to failure. As components wear out we can expect the failure rate to increase. On the other hand, new components often have defects which reduce their lifetimes. As these components fail we may expect the failure rate to decrease because the remaining components are more robust. These factors often combine to give a U-shaped (or bathtub shaped) hazard-rate. In practice, there will be a maximum theoretical life for any component, but we are unlikely to observe any component surviving anywhere near this maximum. Therefore, bounded life distributions might be of interest. The beta distribution can be U-shaped and we use this as a hazard-rate function rather than as a density. The use of this distribution is illustrated by fitting it to data from a human life table.

1. Introduction

In the study of reliability we are often interested in estimating the time to first (repairable) failure of a system or of one of its components. Or we might wish to estimate the total lifetime. In either case the distribution will range over the interval (0, θ) where θ is usually unknown, but could be estimated from a model of the wear on the system under the best possible conditions. As real systems are rarely operated under ideal conditions it is virtually impossible to observe lifetimes remotely near θ and it is usual to take θ to be ∞. This is akin to using the normal distribution for non-negative data such as heights of people. Without a precise physical model of wear processes, it would, in any case, be difficult to provide a reasonable estimate of θ.

Notwithstanding this difficulty, models with bounded lifetime might be useful additions to the range of models even if the maximum lifetime cannot be estimated well. Because most observed lifetimes will be much less than θ and the lifetime distribution is often highly positively skewed, it frequently occurs that an experimenter is disinclined to prolong an experiment until the last item fails. This results in censored data and may add further uncertainty as to the value of θ.
The reliability of a system is best described in terms of the failure rate. This describes the probability density of failure time for those systems (or components) which have survived to that time. This compensates for the fact that the few remaining items may be very unreliable although, as there are few components remaining, there will be a small density of failures. The failure rate for many items will be U-shaped. We shall discuss the precise definition of failure rate in the next section.

2. Definitions

Let $T$ be the failure time of an item and let it have a continuous distribution with density $f(t)$ and distribution function $F(t)$. The function $\bar{F}(t) = 1 - F(t)$ is known as the survival function. Consider the conditional probability distribution of failure given survival to time $u$,

$$g(t|u) = \frac{f(t)}{1 - F(u)} = \frac{f(t)}{\bar{F}(u)}.$$  

We are interested in evaluating this when $u = t$. This gives the failure rate or hazard rate

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{\bar{F}(t)}.$$  

For small $\delta t$, $h(t)\delta t$ is approximately the probability of failure during $(t, t + \delta t)$ given the item is functioning at time $t$. The hazard rate determines, and is determined by the density function.

Define the integrated hazard rate $H(t) = \int_0^t h(x)dx$. Then $H(t) = \int_0^t \frac{f(x)}{1 - F(x)} dx = -\ln(1 - F(t)) = -\ln \bar{F}(t)$ so $\bar{F}(t) = \exp(-H(t))$ and $f(t) = h(t) \exp(-H(t))$.

If failures are purely random, implying that items are equally reliable at any age and do not wear out, we will have constant hazard rate $h(t) = \lambda$. Then $H(t) = \lambda t$ and $f(t) = \lambda e^{-\lambda t}$, $\bar{F}(t) = e^{-\lambda t}$. This is the exponential distribution with mean $1/\lambda$. Realistically, items can be expected to wear out; this results in $h(t)$ being an increasing function. A distribution with this property is an increasing failure rate (IFR) distribution. In the opposite case, increasing reliability with age, $h(t)$ is a decreasing function and the distribution is said to be a decreasing failure rate (DFR) distribution.

In practice, we find that some items have manufacturing defects (or birth defects, in the case of humans), some serious resulting in immediate failure, and some resulting in reduced lifetimes. As these items fail the proportion of more reliable components increases. This results in a DFR initially. Then, as the remaining components start to wear, an IFR stage occurs. Thus the failure rate is U-shaped or bath-tub shaped. Often the flat middle region will be relatively long. As long as we are operating in this region the exponential distribution may be a reasonable approximation.

Rajarshi and Rajarshi [8] reviewed bath-tub failure rate distributions and their method of construction. In particular, they gave a list of 16 distributions including two with finite range. Of the two with finite range: (i) the power function density analysed by Mukherjee and Islam [7] and (ii) the distribution given by Govindaraju [3], Lai and Mukherjee [6] showed that the first is not bath-tub failure rate. The others are.

Jaisingh et al [5] also considered a bath-tub failure rate distribution which is slightly more general than BTD10 listed in Rajarshi and Rajarshi [8].
3. Beta Failure Rate Distributions

Of the bathtub failure rate distributions given in the literature, only a few have finite range. We propose a distribution whose failure rate has a (shifted, truncated and rescaled) beta function of the form

\[ h(t) = c(t + p)^{a-1} (q - t)^{-1} \]  

where \( c > 0, p > 0, a > 0 \) if \( p = 0, \) and \( 0 \leq t \leq \theta \leq q. \) Here \( \theta \) is either a cut-off point representing censoring or \( \theta = q \) is the upper limit of the distribution. In the uncensored case we must have \( b \leq 0 \) in order that \( H(t) = \int_0^t h(x) \, dx \rightarrow 0 \) as \( t \rightarrow q. \)

Compare this with the beta distribution in which we must have \( a > 0, b > 0. \) Another difference is that \( c \) is an arbitrary positive constant rather than a normalising constant.

When \( a < 1, b < 1, \) \( h(t) \) is bathtub shaped.

It is worth remembering, however, that the distribution is defective if \( b > 0, i.e. \) \( F(t) \) does not tend to zero as \( t \) tends to \( q. \) Even when \( \theta < q \) so that the data are censored, it seems undesirable that the underlying model should have this property.

4. Fitting the Model to Data

Given a set of data, we wish to estimate the parameters of the best fitting model. We shall take \( \theta = q. \) Suppose the data consist of \( n \) failure times \( t_1, t_2, ..., t_n. \) For the maximum likelihood estimator, we wish to maximise

\[ L(a, b, c, p, q) = \prod_{i=1}^{n} h(t_j) \exp\{-H(t_j)\} \]  

where \( h(t) = c(t + p)^{a-1} (q - t)^{-1} \) and \( H(t) = \int_0^t h(x) \, dx. \) \( (4.1) \)

This will clearly have to be maximised numerically and numerical evaluation of the integrals will also be necessary.

If the data is only available in grouped form, with intervals \( (u_0, u_1), (u_1, u_2), ..., (u_{k-1}, u_k) \) where \( u_0 = 0, u_k = \infty \) and frequencies \( f_1, f_2, ..., f_k, \) then the likelihood function becomes

\[ L(a, b, c, p, q) = \prod_{i=1}^{h} \{\exp\{-H(u_{i-1})\} - \exp\{-H(u_i)\}\}^{f_i} \]  

This involves a slight loss of information because of the uncertainty of the exact position of the observations within the intervals. However, it is more accurate (and easier to program) than treating the observations as being at the mid points of the intervals.

If we use the Kolmogorov-Smirnov criterion we wish to minimise

\[ K = \max_{i=1, ..., n} \left[ \max \left( \left| \frac{1}{n} \sum_{j=1}^{i} f_j - 1 + \exp\{-H(u_i)\} \right|, \left| \frac{1}{n} \sum_{j=1}^{i-1} f_j - 1 + \exp\{-H(u_j)\} \right| \right) \]  

over \( a, b, c, p, q. \)

In the case of grouped data, we wish to maximise

\[ K = \max_{i=1, ..., n} \left[ \max \left( \left| F_i - 1 + \exp\{-H(u_i)\} \right|, \left| F_{i-1} - 1 + \exp\{-H(u_i)\} \right| \right) \]  

where \( F_i = \sum_{j=1}^{i} f_j, \) \( i = 0, 1, ..., k. \)

Both the maximum likelihood and Kolmogorov-Smirnov criteria can be given in grouped or ungrouped form. We can also try to fit the hazard rate, but empirical
estimates of this require grouping. The same would apply to directly fitting the density by some criterion such as minimum chi-squared.

The usual estimate for the hazard rate at time \( t = u_{i-1} \) is \( h_i = \frac{f_i}{1 - F_{i-1}} \) with \( F_i \) as in (4.6). However, actuaries prefer to estimate the hazard rate at 
\[
\frac{1}{2}(u_{i-1} + u_i) \]
by
\[
h_i = \frac{f_i}{1 - \frac{1}{2} (F_{i-1} + F_i)}
\]
which should usually be more accurate. Even this may not be sufficiently accurate when the hazard rate is changing quickly; and in our model, near to \( t = q \), it will be. We therefore recommend treating this as an estimate of an average hazard rate over the interval \((u_{i-1}, u_i)\) and comparing it to the corresponding model based estimate

\[
\hat{h}\left(\frac{1}{2}(u_{i-1} + u_i)\right) = \frac{\exp(-H(u_{i-1})) - \exp(-H(u_i))}{\frac{1}{2} \left[ \exp(-H(u_{i-1})) + \exp(-H(u_i)) \right]}
\]

(4.7)

Using the least squares criterion, we would minimise the sum of squares of the difference between this and the empirical estimate.

5.  An Example

In the paper by Jaisingh et al [5], a model is fitted to human life data taken from Halley [4]. This data shows a rapidly increasing empirical hazard rate near \( t = 85 \) which suggests that a bounded model such as the beta failure rate distribution might be useful.

The data consists of the survival function for the city of Breslau in Germany (now Wroclaw in Poland) in one year age ranges. Halley also had data for London and Dublin but there was too much migration to and from these cities to produce reliable actuarial data. According to Todhunter [9], it is not clear whether the 1000 in category 1 refers to 1000 births or 1000 at age 1. Bernoulli [1] takes the latter interpretation and estimates the number of births to be 1300. He uses the data in a model of the effects of smallpox vaccination and estimates an increase in mean life expectancy of three years.

Incidently, D’Alembert [2] notes that the mean life expectancy is 26 years, while the median is 8. The disparity leads him to reject statistics as meaningless.

Unfortunately, we have not obtained the original data and Jaisingh et al only give the data grouped in five year intervals. This is reproduced below.

<table>
<thead>
<tr>
<th>Age group</th>
<th>0-4</th>
<th>5-9</th>
<th>10-14</th>
<th>15-19</th>
<th>20-24</th>
<th>25-29</th>
<th>30-34</th>
<th>35-39</th>
<th>40-44</th>
</tr>
</thead>
<tbody>
<tr>
<td>deaths fi</td>
<td>290</td>
<td>57</td>
<td>31</td>
<td>30</td>
<td>32</td>
<td>37</td>
<td>42</td>
<td>45</td>
<td>49</td>
</tr>
<tr>
<td>survivors Fi</td>
<td>1000</td>
<td>710</td>
<td>653</td>
<td>622</td>
<td>592</td>
<td>560</td>
<td>523</td>
<td>481</td>
<td>436</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age group</th>
<th>45-49</th>
<th>50-54</th>
<th>55-59</th>
<th>60-64</th>
<th>65-69</th>
<th>70-74</th>
<th>75-79</th>
<th>80-84</th>
</tr>
</thead>
<tbody>
<tr>
<td>deaths fi</td>
<td>52</td>
<td>53</td>
<td>50</td>
<td>50</td>
<td>51</td>
<td>53</td>
<td>44</td>
<td>34</td>
</tr>
<tr>
<td>survivors Fi</td>
<td>387</td>
<td>335</td>
<td>282</td>
<td>232</td>
<td>182</td>
<td>131</td>
<td>78</td>
<td>34</td>
</tr>
</tbody>
</table>

It seems reasonable that the data refer to numbers per thousand and that the population size is not given. It turns out that we can estimate the population size from this data. There are seven adjacent cells with almost identical frequencies. Equating the variance of these cells with the variance of a binomial distribution we estimate the number of births to be 19598. Combining this with an average age of 26 gives a
population of 509,536 if it was in a steady state. This is a little conservative as it treats the similar cells as identical (with \( p = 0.0511 \)). The present population of Wroclaw city is a similar size; for the Wroclaw province the population is about 2 million.

Jaisingh et al fitted their model to this data using both maximum likelihood and a least squares fit to the hazard rate. However, they did not explain their procedure in detail. They measured the goodness of fit of their models by \( R^2 \) which is of limited value for non-linear models. Now that we have a rough estimate of the annual number of births, it is possible to estimate the minimum sum of squared errors which we could reasonably expect to achieve. This value is about 0.006 and we should not try to obtain a smaller sum of squared errors.

Fitting the beta failure rate distribution by six different criteria all converged very slowly by both the Nelder-Mead simplex algorithm and Brent's modification of the Fletcher-Powell quasi-Newton method.

The criteria used were maximum likelihood, Kolmogorov-Smirnov and two modifications based on sums of squares and absolute deviations from the distribution function, and sums of squares and absolute values of deviations from the hazard rate. These functions seemed to have many local maxima and minima or flat regions. Global optima were achieved by using different criteria sequentially so as to have a variety of starting points. Estimating the maximum, \( \theta = q \), proved to be unstable and resulted in a value which sometimes reached 104 but was usually nearer to 93 with about the same degree of fit. This agrees with intuition that a maximum cannot be estimated very accurately.

The graphs show the best fit attained by the maximum likelihood method. The parameters were as follows:

<table>
<thead>
<tr>
<th>( p )</th>
<th>( q )</th>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.0000056</td>
<td>93.3</td>
<td>0.298</td>
<td>-1.42</td>
<td>3524</td>
</tr>
</tbody>
</table>

KS = Kolmogorov-Smirnov, InL = log likelihood, SS = sum of squares for hazard rate, GoF = goodness of fit for density.

Note that the sum of squared errors is larger than the ideal value of 0.006. The age group 0-4 is hard to fit with this model. The model might fit modern survival data better.

The fit is comparable with that of Jaisingh et al.
The Kolmogorov-Smirnov statistic is halved for the minimum Kolmogorov-Smirnov method because of the big jump in the first interval. However, using a modified K-S method first resulted in a better fit elsewhere.

6. Conclusions

The beta failure rate model is not easy to fit but seems to be a useful model in some cases. Although there is no simple closed form for the density or the distribution function, this is not really a disadvantage, after all, even the normal distribution has no closed form for the distribution function. It is unfortunate that there is no closed form for the parameter estimates, but this is not an uncommon situation.

Another approach, fitting a beta shape to $H(t)$ gives a simpler distribution function and will be discussed elsewhere.

References


