Some Method On Survival Analysis Via Weibull Model In the Present of Partly Interval Censored: A Short Review

Nurul Akma Mohd Fauzi, F. A. M. Elfaki, Yeakub Ali
Department of Manufacturing and Material Engineering, International Islamic University Malaysia
Department of Sciences, Faculty of Engineering, IIUM, P.O.Box 10, 50728 Kuala Lumpur, Malaysia

Abstract
Survival analysis or known as failure time analysis has its own achievement in the medical application. Thus, a few of researcher found that it also can be applied in the engineering application as it is one of the robust methods in the data analysis. Partly interval censoring is one of the censoring techniques that used in the survival analysis and it is developed by using Weibull distribution model. Weibull distribution model is one of the analysis techniques that commonly used as a lifetime distribution in the reliability applications. Thus, the estimation of parameter is needed in order to increase the flexibility of Weibull distribution so that it can suit in any condition of the treated data in engineering applications.
Keywords
Partly Interval Censored, Weibull Distribution, Maximum likelihood Estimation

1. Introduction

The analysis of data in engineering applications is quite important as it can discover many useful information, options and conclusion in the decision making. Statistical method is one way that is widely used by the researcher and engineer as it provides various kind of methods in dealing of the data. One of the methods that is used in analyzing data is the survival analysis method.
Recently the survival analysis method had been widely used in the engineering applications as well as used in the biomedical application. As mentioned by [1], one of the examples of engineering application that deal with the survival analysis method is in the life of testing the durability of mechanical or electrical component. The scientist applies this technique to track the products and material's life span for predicting the product reliability.
Censoring is a part of observation data's method in the survival analysis. There are consist of right censored, left censored and interval censored. In our study, we focus more on the partly interval censoring. It is one type of censoring under the interval censoring. According to [2], partly interval censored data consist of exact data and interval censored data. This means that some of the subject event of interest is exactly observed while for other it lays within an interval.
There are researchers who still continue using partly interval censored in their studies as cited by [2], [3], [4] and [5]. The examples are in [6] used a class of generalized log-rank test for partly interval censored failure data and as discussed the work by [7] in partly interval censored data where treating an exact observation as an interval censored observation with very short interval. [3] also maximized the used of Proportional Hazard Weibull Model in the study of parametric Cox's model for partly interval censored data of the application for the AIDS study.
In this research, partly interval censored will be mainly used in order to estimate the survivability of failure rate with using Weibull distribution model. The model is adapted from the model that developed by [4], [5] and [8] with some modification and different estimator methods. The estimation methods consist of analytical methods which is Maximum Likelihood Estimation and graphical methods which is Weibull Probability Plotting. Our objective is to study whether the analytical method and graphical method can perform the best analysis with giving better result especially when involve in the engineering application.

2. Maximum Likelihood Estimation For Weibull Distribution

The maximum likelihood is probably the most important method of estimation of parameters ([11], [12]). It is very popular and has a very wide application.
Maximum likelihood estimation has several major advantage. It can be used in most estimation problems likely to be met in practice. It's invariance property ensures that, having calculated the maximum likelihood estimation, any other function of parameter can be estimated directly without restarting the estimation process. Another advantage of the maximum likelihood estimators is that their variances may be approximated routinely by the inversion of the observed information matrix.
However, based on the model that developed by [4], the maximum likelihood estimation that developed by Weibull distribution model begin with the representation of probability density function (pdf) and cumulative distribution function (cdf):

Manuscript received April 5, 2015
Manuscript revised April 20, 2015

$$f(t,\alpha,\beta)=\frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}\exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\} \quad (1)$$

$$F(t,\alpha,\beta)=1-\exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\} \quad (2)$$

where for α and β which is represent the scale parameter and shape parameter .

The data that involve in this study are failure data, right censored data and interval censored data. The right censored data exist when the study did not show any sign of failure during the study period. The failure data arise when the study already stop malfunctioning during the study period and the partly interval data is consist of data that in between the interval of the failure data.Thus, the maximum likelihood is given as:

$$L(t_i,u_i,v_i,\alpha,\beta)=\prod_{i}^{k} f(t_i)\prod_{j=k+1}^{r}(1-F(t))\prod_{j=r+1}^{n}(F(v_i,\alpha,\beta)-F(u_i,\alpha,\beta)) \quad (3)$$

As implies with the equation (1) and (2) based on (3):

$$L(t_i,u_i,v_i,\alpha,\beta)=$$
$$\prod_{i}^{k}\left[\frac{\beta}{\alpha}\left(\frac{t_i}{\alpha}\right)^{\beta-1}\exp\left\{-\left(\frac{t_i}{\alpha}\right)^{\beta}\right\}\right]x\prod_{j=k+1}^{r}\left[\exp\left\{-\left(\frac{t_k}{\alpha}\right)^{\beta}\right\}\right]x\prod_{j=r+1}^{n}\left[\exp\left\{-\left(\frac{u_j}{\alpha}\right)^{\beta}\right\}-\exp\left\{-\left(\frac{v_j}{\alpha}\right)^{\beta}\right\}\right] \quad (4)$$

To simplify the equation (4), let $L(t_i,u_i,v_i,\alpha,\beta)$ = A Thus,

$$A=\left(\left(\frac{\beta}{\alpha}\right)^{k}\prod_{i=1}^{k}\left[\left(\frac{t_k}{\alpha}\right)^{\beta-1}\right]\exp\left\{-\left(\frac{t_i}{\alpha}\right)^{\beta}\right\}\prod_{j=k+1}^{r}\exp\left\{-\left(\frac{t_k}{\alpha}\right)^{\beta}\right\}x\prod_{j=r+1}^{r}\left[\exp\left\{-\left(\frac{u_j}{\alpha}\right)^{\beta}\right\}-\exp\left\{-\left(\frac{v_j}{\alpha}\right)^{\beta}\right\}\right]\right) \quad (5)$$

then,

$$A=\left(\left(\frac{\beta}{\alpha}\right)^{k}\prod_{i=1}^{k}\left[\left(\frac{t_k}{\alpha}\right)^{\beta-1}\right]\exp\left\{\prod_{i=1}^{k}\left\{-\left(\frac{t_i}{\alpha}\right)^{\beta}\right\}\prod_{j=+1}^{r}\left[-\left(\frac{t_k}{\alpha}\right)^{\beta}\right]^{r-k}\right\}x\prod_{j=r+1}^{r}\left[esp\left\{-\left(\frac{u_j}{\alpha}\right)^{\beta}\right\}-\exp\left\{-\left(\frac{v_j}{\alpha}\right)^{\beta}\right\}\right]\right) \quad (6)$$

The log-likelihood becomes:

$$\ln(A)=\left(k[\ln\beta-\beta\ln\alpha]+(\beta-1)\sum_{i=1}^{k}\ln(t_i)-\frac{1}{\alpha^{\beta}}\left[\sum_{i=1}^{k}(t_i)^{\beta}+(r-k)(t_k)^{\beta}\right]\right)$$
$$+\left(\sum_{j=r+1}^{r}\ln\left[\exp\left\{-\left(\frac{u_j}{\alpha}\right)^{\beta}\right\}-\exp\left\{-\left(\frac{v_j}{\alpha}\right)^{\beta}\right\}\right]\right) \quad (7)$$

In order to find the value for α and β which is represent the scale parameter and shape parameter. By differentiating equation (7) with respect to α and β, we obtain:

$$\frac{\partial\ln(A)}{\partial\alpha}=\left(-\frac{k\beta}{\alpha}+\frac{\beta}{\alpha}\left[\sum_{i=1}^{k}\left(\frac{t_i}{\alpha}\right)^{\beta}+(r-k)\left(\frac{t_k}{\alpha}\right)^{\beta}\right]\right)+\sum_{i=r+1}^{n}\left[\frac{\left(\frac{u_j}{\alpha}\right)^{\beta}\left(\frac{u_j}{\alpha}\right)\exp\left\{-\left(\frac{u_j}{\alpha}\right)^{\beta}\right\}-\left(\frac{v_j}{\alpha}\right)^{\beta}\left(\frac{v_j}{\alpha}\right)\exp\left\{-\left(\frac{v_j}{\alpha}\right)^{\beta}\right\}}{\exp\left\{-\left(\frac{u_j}{\alpha}\right)^{\beta}\right\}-\exp\left\{-\left(\frac{v_j}{\alpha}\right)^{\beta}\right\}}\right] \quad (8)$$

$$\frac{\partial\ln(A)}{\partial\beta}=\left(\frac{k}{\beta}-k\ln\alpha+\sum_{i=1}^{k}\ln(t_i)-\frac{1}{\alpha^{\beta}}\left[\sum_{i=1}^{k}(t_i)^{\beta}\ln\left(\frac{t_i}{\alpha}\right)+(r-k)(t_k)^{\beta}\ln\left(\frac{t_k}{\alpha}\right)\right]\right)+\sum_{i=r+1}^{n}\left[\frac{\left(\frac{u_j}{\alpha}\right)^{\beta}\ln\left(\frac{u_j}{\alpha}\right)\exp\left\{-\left(\frac{u_j}{\alpha}\right)^{\beta}\right\}-\left(\frac{v_j}{\alpha}\right)^{\beta}\ln\left(\frac{v_j}{\alpha}\right)\exp\left\{-\left(\frac{v_j}{\alpha}\right)^{\beta}\right\}}{\exp\left\{-\left(\frac{u_j}{\alpha}\right)^{\beta}\right\}-\exp\left\{-\left(\frac{v_j}{\alpha}\right)^{\beta}\right\}}\right] \quad (9)$$

By using Newton Raphson method, the parameter of α and β for the Weibull distribution is:

$$U(\alpha,\beta)=\frac{\partial \ln(A)}{\partial \alpha}=0 \quad \text{and} \quad U(\alpha,\beta)=\frac{\partial \ln(A)}{\partial \beta}=0$$
(10)

3. Weibull Probability Plotting

Another method of implementation of Weibull model in survival analysis is graphical method. Weibull probability plotting technique is one of the graphical method that had been apply by [8], [9] and [10] in their study. The graphical method will give view of the analysis although the accuracy is quite less than the analytical method which in our study is the maximum likelihood model. However, the graphical technique is very useful in the presenting of the data especially in the industrial area.

By using the cumulative Weibull distribution function (2):

$$F(t,\alpha,\beta)=1-\exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}$$

Then

$$1-F(t,\alpha,\beta)=\exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\} \quad (11)$$

By taken the natural logarithm of (11), we find:

$$\ln\left[1-F(t,\alpha,\beta)\right]=-\left(\frac{t}{\alpha}\right)^{\beta} \quad (12)$$

Similarly, by taken the natural logarithm of (12), we obtain:

$$\ln\left[\ln\left[1-F(t,\alpha,\beta)\right]\right]=-\beta\ln(\alpha)+\beta\ln(t) \quad (13)$$

Equation (13) is referring to a straight line graph with a slope β. The intercept at y-axis is -β ln(α) while the intercept at x-axis is ln (t).

According to [8] and [9], some approaches need to use in order to estimate the F(t). F(t) is represent the cumulative failure probability. The approaches are:

$$F(t_i)=\frac{i}{n}$$

$F(t_i)=\dfrac{i}{n+1}$ Called the mean rank approach, and

$F(t_i)=\dfrac{i-0.5}{n}$ called the median rank estimator, then

$$F(t_i)=\frac{i-0.3}{n+0.4}$$

$$F(t_i)=\frac{i-0.375}{n+0.25}$$

The steps to apply this method are:

1. Rank the failure times in ascending order
2. Estimate the $F(t_i)$ of the ith failure by using approaches.
3. Then, plot the $F(t_i)$ versus t_i, where *n* is the sample of size and t_i is the failure time.

4. Simulation study

The study begins with the data collection of 240 pieces of manufacturing product that has gone through several processes in the manufacturing company. The manufacturing product is randomly picked from several manufacturing lot with identical type and batch number. The percentage of the defect for each process is recorded and the maximum defect percentage is considered as the guideline to create the interval analysis. Each interval survivability will be measured by using analytical and graphical method. The data also will be calculated by using statistical method such as Newton Rapson in order to find the differences between actual valued and estimated value that proposed by the Weibull distribution method.

5. Conclusion

By interpreting defect data through survival analysis method, it can give many predictions for the product's reliability and life span which is very helpful in the engineering application. Interpretation of data in term of analytical analysis and graphical analysis has its own pros and contras but it is very helpful in some situation of analysis. The quality, life span and the reliability of the product in the engineering applications can be improved if there are variety method of analysis especially involving the survivability of the product and applications.

References

[1] Liu, X. (2012). Survival Analysis Models and Applications. John Wiley &Sons. Ltd.

[2] Kim, J. S. (2003). Maximum Likehood Estimation for the Proportional Hazards Model with Partly Interval-Censored Data. J.R. Statist. Soc. B Vol 65 Part 2, 489-502.

[3] Elfaki, F.A.M, Azram, M., & Usman, M. (2012). Parametric Cox's Model for Partly Interval-Censored Data with Application to AIDS Studies. International Journal of Applied Physics and Mathematics Vol 2, No 5, 352-354.

[4] Guure, C. B., Ibrahim, N. A., & Adam, M. B. (2012). On Parlty Censored Data with the Weibull Distribution. ARPN Journal of Engineering and Applied Science, 1329-1334.

[5] Alharphy, A. M., & Ibrahim, N. A. (2013). Parametric Test for Partly Interval-censored Failure Time Data under Weibull Distribution via Multiple Imputation. Journal of Applied Sciences, 1-6.

[6] Zhang, L., M., X., & L.C, T. (2008). On weighted Least Square Estimation for the Parameters of Weibull Distribution. Springer-Verlag London Limited, 57-84.

[7] Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society, 135(2):185-207.
[8] Razali, A. M., Salih, A., & Mahdi, A. A. (2009). Estimation Accuracy of Weibull Distribution Parameters. Journal of Applied Science Research, 790-795.
[9] Al-Fawzan, M. (2000). Methods for Estimating the Parameters of the Weibull Distribution. King Abdulaziz City for Science and Technology.
[10] Zhen. Xu, (2012). Comparing Graphical Method and A Modified Method to Fit Weibull Distribution. Theses and Dissertations Leigh Preserve, 1149.
[11] Bain, L. J (1978). Statistical Analysis of Reliability and Life-Testing Models. New York and Basel: Marcel Dekker, Inc.
[12] LeCam, L. (1990). Maximum Likelihood: An Introduction. International Statistical Review. Vol 58(2):153-171.