Analysis of Incomplete Data in Presence of Competing Risks Among Several Groups

Debasis Kundu[1] Ammar M. Sarhan[2]

Abstract

In reliability analysis an investigator is often interested in the assessment of a specific risk in presence of other risk factors. It is well known as the competing risks problem in the statistical literature. In this paper we consider the analysis of incomplete data in presence of competing risks among several groups. We mainly consider the latent failure times model formulation and it is assumed that the lifetime distributions of the different latent failure times of a particular group follow Weibull distributions with different scale parameters but the same shape parameter. Maximum likelihood estimators of the different parameters are obtained using a simple iterative procedure and also by EM algorithm. Asymptotic distributions of the maximum likelihood estimators of the different parameters are obtained and based on the asymptotic distributions, asymptotic confidence intervals are also proposed. Testing equality of the parameters among several groups is performed. One data set has been analyzed for illustrative purposes.

Key Words and Phrases: Competing risks; Incomplete data; Weibull distribution; Likelihood ratio test; Maximum likelihood estimators; Testing of hypotheses.

Corresponding Author: [1]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. e-mail: kundu@iitk.ac.in, Phone no. 91-512-2597141, Fax no. 91-512-2597500.

[2] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, 35516, Egypt. E-mail: asarhan0@yahoo.com

1 Introduction

This work was motivated from the paper of Park and Kulasekera (2004) and the data presented there, originally taken from Nelson (1970). The data can be briefly described as follows. There are three groups and each group contains the failure times and cause of failures of different electrical items. The data are incomplete in two ways, for some data points the failure times are known but the failure types are unknown and some of the data points are right truncated. Therefore, for each group there are three types of observations: (a) Item which has failed due to a particular cause and its failure time both are observed. (b) Item which has failed and its failure time has been observed but not the cause of failure. (c) Item has not failed up to a certain time.

Park and Kulasekera (2004) analyzed this data under the latent failure times model formulation as suggested by Cox (1959). In the latent failure times model, it is assumed that the different competing causes are independent. They assumed that the latent failure times of the various causes are independent exponentially distributed and considered estimations and testing procedures of the different parameters.

Since the exponential distribution has constant failure rate, it is well known that it has serious limitations in modeling lifetime data. In this paper, we consider the same latent failure times model formulation and it is assumed that the latent failure times are independent Weibull random variables with the same shape parameter within a particular group but different scale parameters. Since Weibull distribution has both increasing and decreasing failure rates, it is well known that the Weibull distribution can be used more effectively than the exponential distribution to analyze lifetimes data. We obtain the maximum likelihood estimators (MLEs) of the different parameters. We propose a very simple iterative scheme in one dimension to compute the MLEs. Alternatively, the EM algorithm also can be used to compute the MLEs and it is proved in the present setup that the EM algorithm converges with probability one.

Although, we could not obtain the exact distribution of the MLEs, we can obtain the asymptotic distributions of the MLEs. Based on the asymptotic distributions of the MLEs, confidence intervals of the different parameters can be constructed. Testing equality of the different parameters is also presented. For example, testing the equality of the shape parameters of the different groups and testing the equality of the scale parameters of the latent failure times across the groups and within a group are also performed. We reanalyze Nelson's data set using Weibull latent failure times distributions.

In Cox's latent failure time model formulation, it is assumed that the latent failure times are independent. Alternatively, the cause specific hazard functions model formulation of Prentice *et al.* (1978) can also be used to analyze competing risks data, where the distributions of the competing causes may not be independent. Unfortunately, without the presence of covariates, it is not possible to test the independence of the different causes from the failure times data. This is well known as the identifiability problem in competing risks. As it is assumed different parametric forms of the latent failure types in Cox's formulation, similarly, under the cause specific hazard functions model, parametric hazard functions are assumed for the different causes. Interestingly, in this case the likelihood functions of the observed data are the same for both formulations; see Kundu (2004). Therefore, the estimation procedures of the different parameters and their statistical properties remain unchanged, although, the interpretations of the different parameters might be different.

The rest of the paper is organized as follows. In section 2, we describe model assumptions and the maximum likelihood estimators are derived in section 3. Confidence intervals based on the asymptotic distributions are proposed in section 4. Testing procedures are presented in section 5. One data analysis results are presented in sections 6 and we conclude the paper in section 7.

2 Model Assumptions and Notations

Before proceeding any further, we describe different notations we are going to use in this paper. Prior to section 4, we will be considering only one group, and therefore, we do not use any notation for different groups. In section 5, we will use the notation for different groups. It is assumed that in each group there are K causes of failures. We will be using the following notations up to section 4.

X_i :	lifetime of the i-th unit
X_{ij} :	latent failure time of the i-th unit due to cause j
$F(.)$:	cumulative distribution function (CDF) of X_i
$f(.)$:	probability density function (PDF) of $F(.)$
$F_j(.)$:	CDF of X_{ij}
$f_j(.)$:	PDF of $F_j(.)$
$\bar{F}_j(.)$:	$1 - F_j(.)$
$h_j(.)$:	hazard function of X_{ij}
δ_i :	indicator variable denoting the cause of failure of unit i
$I[.]$:	indicator function of the event [.]
$exp(\lambda)$:	exponential random variable with PDF $\lambda e^{-\lambda x}$
$gamma(\alpha, \lambda)$:	gamma random variable with PDF $\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$
$Weibull(\alpha, \lambda)$:	Weibull random variable with PDF $\alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha}$
χ^2_k :	denotes the χ^2 distribution with k-degrees of freedom.
$w.r.t$:	with respect to
$r.h.s.$:	right hand side

We are formulating using the latent failure times model, *i.e*, $X_i = \min\{X_{i1}, \ldots, X_{iK}\}$ and $X_{i1}, \ldots, X_{iK}$ are independently distributed. Further, it is assumed that X_{ij} follows $Weibull(\alpha, \lambda_j)$ for $j = 1, \ldots, K$. As we have mentioned in the previous section that we have the following three types of observations;

$$\text{(a)} \quad (x, \delta), \qquad \text{(b)} \quad (x, *), \qquad \text{(c)} \quad (x*, *), \tag{1}$$

where (x, δ) means the unit has failed at the point x due to cause δ, $(x, *)$ means the unit has failed at the point x but the cause of failure is unknown and $(x*, *)$ means the unit has been withdrawn at the point x and it has not failed till then. We further denote, I_1 = the set of Type (a) items, I_2 = the set of Type (b) items and I_3 = the set of Type (c) items. Also, $I_1 = I_{11} \cup \ldots \cup I_{1K}$, where I_{1j} = the set of items whose failure times are known and their failure types are also known to be j. Moreover, $|I_1| = r_1$, $|I_{1j}| = r_{1j}$, $|I_2| = r_2$, $|I_3| = r_3$, therefore, $r_1 = \sum_{j=1}^{K} r_{1j}$. The sample size is $n = r_1 + r_2 + r_3$. It is also assumed that $r_1 > 0$. Note that based on the above assumptions and notations, the likelihood contributions from the observation types (a), (b) and (c) are

$$h_j(x) \prod_{k=1}^{K} \bar{F}_k(x), \qquad f(x), \qquad \bar{F}(x) \tag{2}$$

respectively.

3 Maximum Likelihood Estimators

We have the following n observations

$$\{(x_i, \delta_i), i \in I_1\}, \quad \{(x_i, *), i \in I_2\}, \quad \{(x_i*, *), i \in I_3\}. \tag{3}$$

Based on the above observations, the likelihood function can be written as

$$l(\alpha, \boldsymbol{\lambda}) = \alpha^{r_1+r_2} \prod_{j=1}^{K} \lambda_j^{r_{1j}} \lambda^{r_2} \prod_{i \in I_1 \cup I_2} \left\{ x_i^{\alpha-1} e^{-\lambda x_i^{\alpha}} \right\} \prod_{i \in I_3} e^{-\lambda x_i^{\alpha}}, \tag{4}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$ and $\lambda = \sum_{j=1}^{K} \lambda_j$. Therefore, the log-likelihood function becomes

$$\ln(l(\alpha, \boldsymbol{\lambda})) = (r_1 + r_2)\ln\alpha + (\alpha - 1)\sum_{i \in I_1 \cup I_2} \ln x_i + r_2 \ln\lambda + \sum_{j=1}^{K} r_{1j}\ln\lambda_j - \lambda\sum_{i \in I} x_i^{\alpha}, \qquad (5)$$

where $I = I_1 \cup I_2 \cup I_3$, the complete data set.

Note that for fixed α, differentiating (5) *w.r.t* λ_j and equating it to zero, we obtain the MLE of λ_j, for fixed α as

$$\hat{\lambda}_j(\alpha) = \frac{r_1 + r_2}{r_1} \times \frac{r_{1j}}{\sum_{i \in I} x_i^{\alpha}}. \qquad (6)$$

Combining (5) in (6), we obtain the profile log-likelihood of α as,

$$\ln(l(\alpha, \hat{\lambda}_1(\alpha), \ldots, \hat{\lambda}_K(\alpha))) = c + (r_1 + r_2)\ln\alpha + (\alpha - 1)\sum_{i \in I_1 \cup I_2} \ln x_i - (r_1 + r_2)\ln\left(\sum_{i \in I} x_i^{\alpha}\right), \quad (7)$$

here c is a constant function. Note that if $x_i = 1$ for all i, then the right hand side of (7) is an increasing function of α. Therefore, in that case α cannot be estimated by maximizing the profile likelihood function. Since probability of than event is 0, we assume throughout that at least one of the $x_i \neq 1$.

Now we propose two different ways to maximize the profile likelihood function. We differentiate the *r.h.s.* of (7) *w.r.t* α and equate it to zero and obtain the following equation

$$w(\alpha) = \alpha, \qquad (8)$$

where

$$w(\alpha) = \left[\frac{\sum_{i \in I} x_i^{\alpha}\ln x_i}{\sum_{i \in I} x_i^{\alpha}} - \frac{1}{r_1 + r_2}\sum_{i \in I_1 \cup I_2} \ln x_i\right]^{-1}.$$

Therefore, a simple iterative scheme may be used to compute the fixed point solution of the equation (8). From the i-th iterate $\alpha_{(i)}$, the $(i+1)$-th iterate $\alpha_{(i+1)}$ can be obtained as $w(\alpha_{(i)})$. The iterative process should be stopped when the preassigned 'stopping criterion' is met. Once $\hat{\alpha}$, the MLE of α, is obtained, then $\hat{\lambda}_j$, the MLE of λ_j can be obtained as $\hat{\lambda}_j = \hat{\lambda}_j(\hat{\alpha})$.

Alternatively, we can use the EM algorithm of Dempster, Laird and Rubin (1977), see also Wu (1983), to compute the MLE of α. In the EM algorithm, the first step is the expectation (E) step and the second second step is the maximization (M) step. In this case, we consider the E-step as follows. For the E-step, the complete observations are left intact and 'pseudo observations' are formed from the incomplete observations. For example, in this case Type (a) and Type (c) observations are left intact and 'pseudo observations' are formed from the Type (b) observations. If observation x belongs to Type (b), we form 'pseudo observations' by fractioning x to K partially complete 'pseudo observations' of the form $\{(x, w_1(x, \boldsymbol{\gamma})), \ldots, (x, w_K(x, \boldsymbol{\gamma}))\}$, where $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\lambda})$. The fractional mass, $w_j(x, \boldsymbol{\gamma})$, assigned to this 'pseudo observation' x is the conditional probability that the item failed from risk j, given that it had failed at the time point x. Clearly,

$$w_1(x, \boldsymbol{\gamma}) = \frac{\lambda_1}{\lambda}, \ldots, w_K(x, \boldsymbol{\gamma}) = \frac{\lambda_K}{\lambda}.$$

We denote $w_j(x, \boldsymbol{\gamma}) = w_j$, for brevity. Using the above notation, the log-likelihood function of the 'pseudo observation' can be written as

$$\ln(L_s(\alpha, \boldsymbol{\lambda})) = (r_1 + r_2) \ln \alpha + (\alpha - 1) \sum_{i \in I_1 \cup I_2} \ln x_i + \sum_{j=1}^{K} (r_{1j} + w_j r_2) \ln \lambda_j - \lambda \sum_{i \in I} x_i^{\alpha}. \tag{9}$$

Therefore, the M-step involves the maximization of the 'pseudo log-likelihood function' (9) *w.r.t* α and $\boldsymbol{\lambda}$. Assuming w_j's are known we need to maximize $\ln(L_s(\alpha, \boldsymbol{\lambda}))$. Naturally, it has to be done iteratively. If $(\alpha^{(l)}, \boldsymbol{\lambda}^{(l)})$ is the l-th iterate, then $(l+1)$-th iterate can be obtained as

$$\lambda_1^{(l+1)} = \frac{r_{11} + w_1 r_2}{\sum_{i \in I} x_i^{\alpha^{(l)}}}, \ldots, \lambda_K^{(l+1)} = \frac{r_{1K} + w_K r_2}{\sum_{i \in I} x_i^{\alpha^{(l)}}}.$$

Finally, $\alpha^{(l+1)}$ can be obtained as

$$\alpha^{(l+1)} = \arg\max \quad g(\alpha),$$

where

$$g(\alpha) = (r_1 + r_2) \ln \alpha + (\alpha - 1) \sum_{i \in I_1 \cup I_2} \ln x_i + \sum_{j=1}^{K} (r_{1j} + w_j r_2) \ln \lambda_j^{(i)} - \lambda^{(i)} \sum_{i \in I} x_i^{\alpha}, \qquad \text{for} \quad \alpha > 0. \tag{10}$$

The following result will be useful for the maximization of $g(\alpha)$.

LEMMA 1: The function $g(\alpha)$ is unimodal for $\alpha \in (0, \infty)$.

PROOF: Note that

$$\frac{d^2}{d\alpha^2} g(\alpha) = -\frac{r_1 + r_2}{\alpha^2} - \lambda^{(i)} \sum_{i \in I} x_i^{\alpha} (\ln x_i)^2 < 0.$$

Therefore, $g(\alpha)$ is a concave function. Now the result follows by observing the fact that $g(\alpha) \to -\infty$ as $\alpha \to 0$ or ∞.

From Lemma 1, it follows that the EM algorithm will converge to the global maximum and it can be obtained as (8) by iterating the following equation

$$v(\alpha) = \alpha, \tag{11}$$

where

$$v(\alpha) = \frac{r_1 + r_2}{\lambda^{(i)} \sum_{i \in I} x_i^{\alpha} \ln x_i - \sum_{i \in I_1 \cup I_2} \ln x_i}.$$

Although, we could not prove theoretically that the iterative process proposed in (8) and the EM algorithm converge to the same point, but it is observed in our data analysis experiment that they converge to the same point.

4 Confidence Intervals:

Since the MLEs are not in a compact form, it is not possible to derive the exact distributions of the MLEs. Note that, when α is known, it is possible to obtain the exact distributions of the MLEs of λ_j's. We obtain the confidence intervals of the unknown parameters based on the asymptotic distributions of the MLEs. The asymptotic distribution of the MLEs can be written as follows;

$$\left(\hat{\lambda}_1 - \lambda_1, \ldots, \hat{\lambda}_K - \lambda_K, \hat{\alpha} - \alpha\right) \xrightarrow{d} N_{K+1}\left(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\lambda}, \alpha)\right), \tag{12}$$

here $\mathbf{I}(\boldsymbol{\lambda}, \alpha)$ is the Fisher information matrix for the unknown parameters $\lambda_1, \ldots, \lambda_K, \alpha$. The elements of the $(K+1) \times (K+1)$ matrix $\mathbf{I} = ((I_{ij}))$ are as follows;

$$\begin{aligned}
I_{ii}(\boldsymbol{\lambda}, \alpha) &= \frac{r_1\lambda + r_2\lambda_i}{\lambda^2\lambda_i}; \qquad \text{for} \quad i = 1, \ldots, K, \\
I_{ij}(\boldsymbol{\lambda}, \alpha) &= \frac{r_2}{\lambda^2} = I_{ji}(\boldsymbol{\lambda}, \alpha); \qquad \text{for} \quad i, j = 1, \ldots, K, i \neq j, \\
I_{K+1,K+1}(\boldsymbol{\lambda}, \alpha) &= \frac{r_1 + r_2}{\alpha^2} + \frac{\lambda n}{\alpha^2} V, \\
I_{i,K+1}(\boldsymbol{\lambda}, \alpha) &= \frac{n}{\alpha} U = I_{K+1,i}(\boldsymbol{\lambda}, \alpha) \qquad \text{for} \quad i = 1, \ldots K.
\end{aligned}$$

Here $V = E(Z(\ln Z)^2)$ and $U = E(Z \ln Z)$, where Z follows $exp(\lambda)$. Note that U and V can be written in terms of the digamma and trigamma functions as follows;

$$\begin{aligned}
U &= \frac{1}{\lambda^2}\left[\psi(2) - \ln\lambda\right] \qquad \text{and} \\
V &= \frac{1}{\lambda^2}\left[\psi'(2) + \psi(2)(\ln\lambda)^2 - 2\psi(2)\ln\lambda\right],
\end{aligned}$$

here $\psi(.)$ and $\psi'(.)$ are the digamma and trigamma functions respectively. The asymptotic distributions of the pivotal quantities may then be used to construct confidence intervals.

It is observed in many cases (see Efron and Hinkley; 1978) that it is more appropriate to use the observed information matrix than the expected information matrix. It is very simple to obtain the observed information matrix when the EM algorithm is used (see Louis; 1982). Using the same notations as of Louis (1982), we obtain the observed information matrix $\hat{\mathbf{I}}$ as

$$\hat{\mathbf{I}} = \mathbf{B} - \mathbf{S}\mathbf{S}^T.$$

Here $\mathbf{B}$ is the $(K+1) \times (K+1)$ negative of the second derivative and $\mathbf{S}$ is the $(K+1) \times 1$

gradient vector of the log-likelihood function. They are as follows

$$
\begin{aligned}
B(i,i) &= -\frac{(r_{1i}+w_i r_2)}{\lambda_i^2}, \qquad \text{for} \qquad i=1,\ldots,K, \\
B(i,j) &= 0, \qquad \text{for} \qquad i,j=1,\ldots,K, i\neq j, \\
B(i,K+1) &= B(K+1,i) = -\sum_{i\in I} x_i^{\alpha}\ln x_i \\
B(K+1,K+1) &= -\frac{r_1+r_2}{\alpha^2} - \lambda\sum_{i\in I} x_i^{\alpha}(\ln x_i)^2 \\
S(1) &= \frac{r_{11}+w_1 r_2}{\lambda_1} - \sum_{i\in I} x_i^{\alpha}, \ldots, S(K) = \frac{r_{1K}+w_K r_2}{\lambda_K} - \sum_{i\in I} x_i^{\alpha}, \\
S(K+1) &= \frac{r_1+r_2}{\alpha} + \sum_{i\in I_1\cup I_2} \ln x_i - \lambda\sum_{i\in I} x_i^{\alpha}\ln x_i,
\end{aligned}
$$

here $\mathbf{B} = ((B(i,j)))$ and $\mathbf{S} = (S(1),\ldots,S(K+1))^T$.

5 More than One Group

In this section we consider the case when the data are coming from more than one group. We assume that there are M groups and the elements in each group can fail out of K causes exactly as before. We introduce the new notation in this section. We just put the super script '(m)' to indicate m-th group. For example, $X_i^{(m)}$ denotes the lifetime of the i-th item in the m-th group. Other notations are also similarly defined and it should be clear from the context. It is also assumed that $r_1^{(m)} > 0$ and at least one of the $x_i^{(m)} \neq 1$ for each $m = 1,\ldots,M$.

The log-likelihood function of the observed data without any restriction on the parameters is

$$
\begin{aligned}
L_1 &= \ln l(\alpha^{(1)},\ldots,\alpha^{(M)},\boldsymbol{\lambda}^{(1)},\ldots,\boldsymbol{\lambda}^{(M)}) = \sum_{m=1}^{M}\left(r_1^{(m)}+r_2^{(m)}\right)\ln\alpha^{(m)} - \sum_{m=1}^{M}\lambda^{(m)}\sum_{i\in I^{(m)}}(x_i^{(m)})^{\alpha^{(m)}} \\
&\quad + \sum_{m=1}^{M}(\alpha^{(m)}-1)\sum_{i\in I_1^{(m)}\cup I_2^{(m)}}\ln x_i^{(m)} + \sum_{m=1}^{M} r_2^{(m)}\ln\lambda^{(m)} + \sum_{m=1}^{M}\sum_{j=1}^{K} r_{1j}^{(m)}\ln\lambda_j^{(m)}.
\end{aligned}
$$

Note that we can obtain the MLEs of the different estimators separately for each group. Therefore, we can use the results for single group in each case.

In this section, we mainly consider different testing problems. First consider the following testing problem.

$$\text{Testing:} \quad H_{01} : \alpha^{(1)} = \ldots = \alpha^{(M)} = \alpha. \tag{13}$$

Under the null hypothesis the log-likelihood takes the form;

$$\begin{aligned}\ln(l_{01}(\alpha, \boldsymbol{\lambda}^{(1)}, \ldots, \boldsymbol{\lambda}^{(M)})) &= \ln \alpha \left(\sum_{m=1}^{M} (r_1^{(m)} + r_2^{(m)}) \right) + (\alpha - 1) \sum_{m=1}^{M} \sum_{i \in I_1^{(m)} \cup I_2^{(m)}} \ln x_i^{(m)} \\ &+ \sum_{m=1}^{M} r_2^{(m)} \ln \lambda^{(m)} + \sum_{m=1}^{M} \sum_{j=1}^{K} r_{1j}^{(m)} \ln \lambda_j^{(m)} - \sum_{m=1}^{M} \lambda^{(m)} \sum_{i \in I^{(m)}} x_i^{(m)\alpha}.\end{aligned}$$

The MLE of $\lambda_j^{(m)}$'s in terms of α can be obtained as

$$\hat{\lambda}_j^{(m)}(\alpha) = \frac{r_1^{(m)} + r_2^{(m)}}{r_1^{(m)}} \times \frac{r_{1j}^{(m)}}{\sum_{i \in I^{(m)}} x_i^{(m)\alpha}}. \tag{14}$$

Under null hypothesis the profile likelihood of α is

$$\begin{aligned}L_{01} &= \ln(l_{01}(\alpha)) = \ln \alpha \left(\sum_{m=1}^{M} (r_1^{(m)} + r_2^{(m)}) \right) + (\alpha - 1) \sum_{m=1}^{M} \sum_{i \in I_1^{(m)} \cup I_2^{(m)}} \ln x_i^{(m)} - \\ &- \sum_{m=1}^{M} (r_1^{(m)} + r_2^{(m)}) \ln \left(\sum_{i \in I^{(m)}} x_i^{(m)\alpha} \right).\end{aligned} \tag{15}$$

Therefore, similarly to (8) we can maximize (18) *w.r.t.* α by a simple iterative process as follows

$$w_{01}(\alpha) = \alpha, \tag{16}$$

where

$$w_{01}(\alpha) = \left(\sum_{m=1}^{M} r_1^{(m)} + r_2^{(m)} \right) \left[\sum_{m=1}^{M} (r_1^{(m)} + r_2^{(m)}) \frac{\sum_{i \in I^{(m)}} x_i^{(m)\alpha} \ln x_{(i)}^{(m)}}{\sum_{i \in I^{(m)}} x_i^{(m)\alpha}} - \sum_{m=1}^{M} \sum_{i \in I_1^{(m)} \cup I_2^{(m)}} \ln x_i^{(m)} \right]^{-1}.$$

Once we get the MLEs of the different parameters the likelihood ratio will take the form $(\hat{L}_1 - \hat{L}_{01})$, where $\hat{L}_{01}$ and $\hat{L}_1$ are the values of L_{01} and L_1 respectively by replacing the true parameter values by their estimates. Moreover, as the sample size in each group tends to ∞

$$2(\hat{L}_1 - \hat{L}_{01}) \sim \chi^2_{M-1}. \tag{17}$$

Here '$\sim$' denotes 'follows in distribution'. Note that (17) can be obtained from the general theory of the likelihood ratio test, see for example Lehmann and Romano (2005, page 513-517). The asymptotic distribution of $2(\hat{L}_1 - \hat{L}_{01})$ as provided in (17) can be used to construct the likelihood ratio test for testing (13). Reject H_{01} if $2(\hat{L}_1 - \hat{L}_{01}) > c$, where c is such that $P(\chi^2_{M-1} > c) =$ size of the test.

Now consider the following testing problem:

Testing: $H_{02} : \alpha^{(1)} = \ldots = \alpha^{(m)} = \alpha; \lambda_1^{(1)} = \ldots = \lambda_1^{(M)} = \lambda_1; \ldots, \lambda_K^{(1)} = \ldots = \lambda_K^{(M)} = \lambda_K.$
Under the null hypothesis the log-likelihood takes the form;

$$\begin{aligned}\ln(l_{02}(\alpha, \lambda_1, \ldots, \lambda_K)) &= \ln\alpha \left(\sum_{m=1}^{M} (r_1^{(m)} + r_2^{(m)})\right) + (\alpha - 1) \sum_{m=1}^{M} \sum_{i \in I_1^{(m)} \cup I_2^{(m)}} \ln x_i^{(m)} \\ &+ \ln\lambda \sum_{m=1}^{M} r_2^{(m)} + \sum_{m=1}^{M} \sum_{j=1}^{K} r_{1j}^{(m)} \ln\lambda_j - \lambda \sum_{m=1}^{M} \sum_{i \in I^{(m)}} x_i^{(m)\alpha}.\end{aligned}$$

The MLE of λ_j's in terms of α can be obtained as

$$\hat{\lambda}_j(\alpha) = \left(1 + \frac{\sum_{m=1}^{M} r_2^{(m)}}{\sum_{j=1}^{K} \sum_{m=1}^{M} r_{1j}^{(m)}}\right) \times \frac{\sum_{m=1}^{M} r_{1j}^{(m)}}{\sum_{m=1}^{M} \sum_{i \in I^{(m)}} \left(x_i^{(m)}\right)^{\alpha}}. \tag{18}$$

Therefore, the MLE of α can be obtained by maximizing $\ln(l_{02}(\alpha, \hat{\lambda}_1, \ldots, \hat{\lambda}_K))$, the profile log-likelihood of α, *w.r.t.* α. This can be achieved by a simple iterative procedure as follows;

$$w_{02}(\alpha) = \alpha, \tag{19}$$

where

$$w_{02}(\alpha) = \left[\frac{\sum_{m=1}^{M} \sum_{i \in I^{(m)}} \left(x_i^{(m)}\right)^{\alpha} \ln x_i^{(m)}}{\sum_{m=1}^{M} \sum_{i \in I^{(m)}} \left(x_i^{(m)}\right)^{\alpha}} - \frac{1}{\sum_{m=1}^{M} \left(r_1^{(m)} + r_2^{(m)}\right)} \times \sum_{m=1}^{M} \sum_{i \in I_1^{(m)} \cup I_2^{(m)}} \ln x_i^{(m)}\right]^{-1}.$$

In this case as the sample size in each group tends to ∞, the likelihood ratio test statistic;

$$2(\hat{L}_1 - \hat{L}_{02}) \sim \chi^2_{(M-1)(K-1)}, \tag{20}$$

where as before, $\hat{L}_{02}$ is the value of L_{02} replacing the true parameter values by their estimates. In this case also, (20) can be used for testing H_{02} with a given size of the test. Reject H_{02} if $2(\hat{L}_1 - \hat{L}_{02}) > c$, where c is such that $P(\chi^2_{(M-1)(K-1)} > c) =$ size of the test.

Finally we consider the following testing problem;

Testing: $H_{03} : \alpha^{(1)} = \ldots = \alpha^{(m)} = \alpha; \lambda_1^{(1)} = \ldots = \lambda_1^{(M)} = \ldots = \lambda_K^{(1)} = \ldots = \lambda_K^{(M)} = \mu.$

Under the null hypothesis the log-likelihood takes the form;

$$\begin{aligned} \ln(l_{03}(\alpha,\mu)) &= \ln\alpha\left(\sum_{m=1}^{M}(r_1^{(m)} + r_2^{(m)})\right) + (\alpha-1)\sum_{m=1}^{M}\sum_{i\in I_1^{(m)}\cup I_2^{(m)}} \ln x_i^{(m)} \\ &+ \ln\mu\sum_{m=1}^{M} r_2^{(m)} + \ln\mu\sum_{m=1}^{M}\sum_{j=1}^{K} r_{1j}^{(m)} - K\mu\sum_{m=1}^{M}\sum_{i\in I^{(m)}} x_i^{(m)\alpha}. \end{aligned}$$

The MLE of μ in terms of α can be obtained as

$$\hat{\mu}(\alpha) = \frac{\sum_{m=1}^{M}(r_1^{(m)} + r_2^{(m)})}{K\sum_{m=1}^{M}\sum_{i\in I^{(m)}}\left(x_i^{(m)}\right)^{\alpha}}. \tag{21}$$

In this case also, the MLE of α can be obtained by maximizing $\ln(l_{03}(\alpha,\hat{\mu}))$, the profile log-likelihood of α, *w.r.t.* α and that can be achieved by a simple iterative procedure as follows;

$$w_{03}(\alpha) = \alpha, \tag{22}$$

where

$$w_{03}(\alpha) = \left[\frac{\sum_{m=1}^{M}\sum_{i\in I^{(m)}}\left(x_i^{(m)}\right)^{\alpha}\ln x_i^{(m)}}{\sum_{m=1}^{M}\sum_{i\in I^{(m)}}\left(x_i^{(m)}\right)^{\alpha}} - \frac{1}{\sum_{m=1}^{M}\left(r_1^{(m)} + r_2^{(m)}\right)} \times \sum_{m=1}^{M}\sum_{i\in I_1^{(m)}\cup I_2^{(m)}} \ln x_i^{(m)}\right]^{-1}.$$

Therefore as the sample size in each group tends to ∞, the likelihood ratio test statistic;

$$2(\hat{L}_1 - \hat{L}_{03}) \sim \chi^2_{M(K+1)-2}, \tag{23}$$

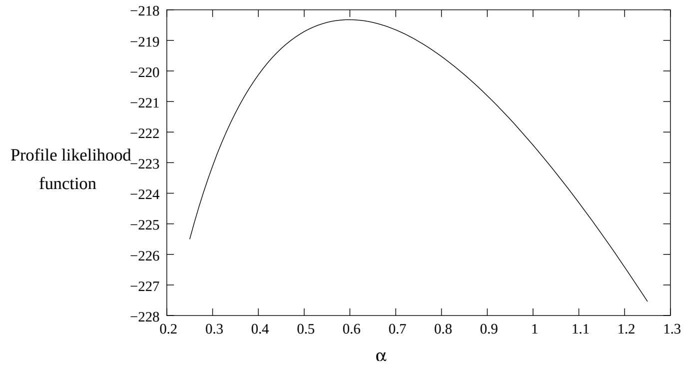

Figure 1: The profile log-likelihood function of α for Groups I.

where as before, $\hat{L}_{03}$ is the value of L_{03} replacing the true parameter values by their estimates. Again, we use (23) to construct the likelihood ratio test as before. Reject H_{03} if $2(\hat{L}_1 - \hat{L}_{03}) > c$, where c is such that $P(\chi^2_{M(K+1)-2} > c) =$ size of the test.

6 Data Analysis

In this section we analyze the data which was presented in Park and Kulasekera (2004). The data were originally presented in Nelson (1970). The data consist of failure times or censoring times for 139 appliances. They are divided into three groups and the elements in each group are subjected to a manual lifetime test. Failures were classified into 18 different modes. Among the observed failures only mode 11 occurs more than twice in all groups. We are interested mainly about the failure mode 11 while combining the rest of the failure modes into a single one. First we obtain the MLEs of the parameters for different groups. The profile log-likelihood functions of α for the three groups are presented in Figures 1, 2 and 3

From the figures it is clear that the profile log-likelihoods of all the groups are unimodal and therefore both the iterative procedures should work to compute the MLEs of α. We have used the iterative procedure (8) and also the EM algorithm with initial guess 0.2 in all

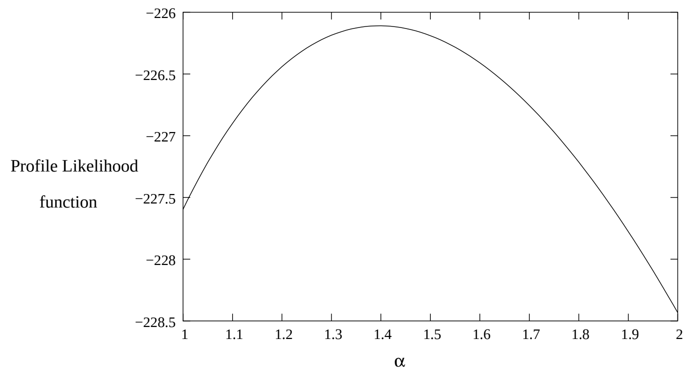

Figure 2: The profile log-likelihood function of α for Groups II.

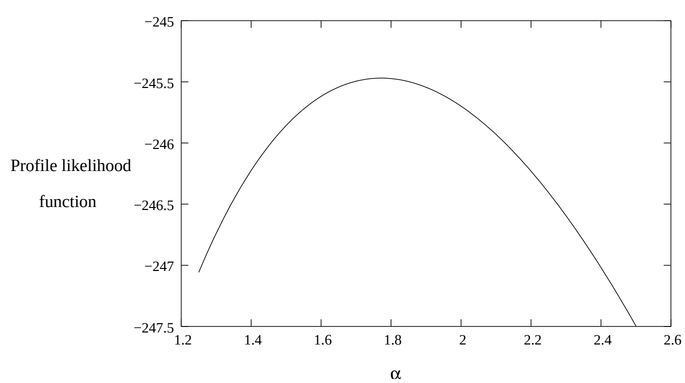

Figure 3: The profile log-likelihood function of α for Groups III.

cases. In all cases they converge to the same point. The results are reported in Table 1. We have also reported in Table 1 the MLEs of λ_1 and λ_2 when the common shape parameter is one. In Table 1, for each group $\hat{\alpha}$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ represent the MLEs of α, λ_1 and λ_2 respectively. Moreover, $\hat{\lambda}_{1E}$ and $\hat{\lambda}_{2E}$ represent the MLEs of λ_1 and λ_2 respectively, when $\alpha = 1$.

Group	$\hat{\alpha}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_{1E}$	$\hat{\lambda}_{2E}$
I	0.59869	0.00370	0.00601	2.00985E-04	3.26600E-04
II	1.39692	3.67058E-05	1.37647E-05	5.14304E-04	1.92864E-04
III	1.77172	3.43334E-06	1.14445E-06	5.52181E-04	1.84060E-04

Table 1: MLEs of the different parameters for the three groups.

The corresponding asymptotic 95% confidence intervals of the different parameters and for different groups are reported in Table 2.

Group	$\hat{\alpha}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
I	(0.54086, 0.65651)	(0.00114, 0.00626)	(0.00299, 0.00928)
II	(1.33451, 1.45933)	(1.87200E-05, 5.46916E-05)	(0.27501E-05, 2.47788E-05)
III	(1.71155, 1.83189)	(1.84721E-05, 5.01946E-05)	(0.22870E-06, 2.06020E-06)

Table 2: 95% Confidence intervals of the different parameters for the three groups.

From Table 2 it is clear that $H_0 : \alpha = 1$ will be rejected for all the three groups. Therefore, fitting the exponential models to the different groups may not be appropriate. We have also performed the following test: $H_{01} : \alpha_1 = \alpha_2 = \alpha_3 = 1$ and as expected it has also been rejected with 5% level of significance.

Now we want to see whether fitting Weibull distributions to the different groups are reasonable or not. We consider only the uncensored data in all the three groups and for comparison purposes we fitted both exponential and Weibull distributions to the uncensored observations. The plots of the empirical survival functions and the fitted Weibull and exponential survival functions are presented in Figures 4, 5 and 6.

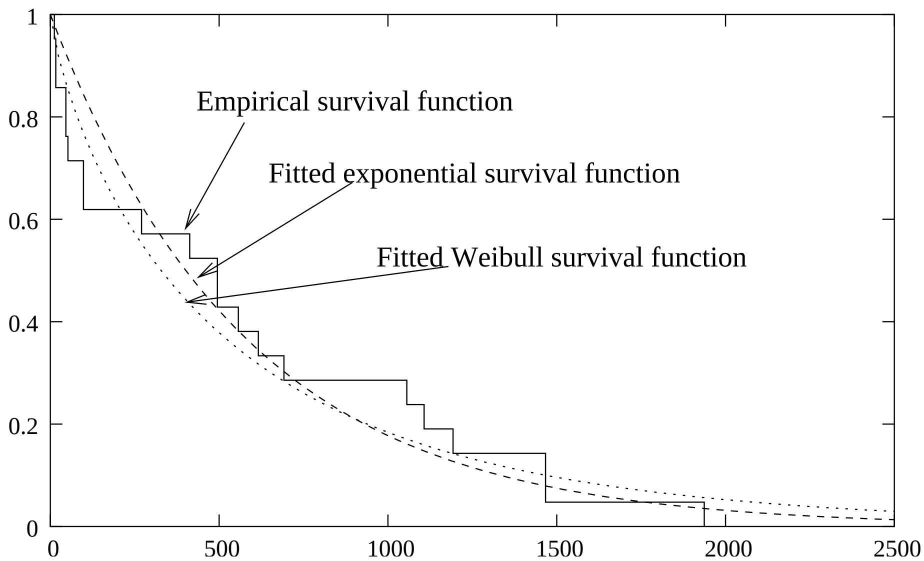

Figure 4: The empirical survival function and the fitted survival functions for Group I.

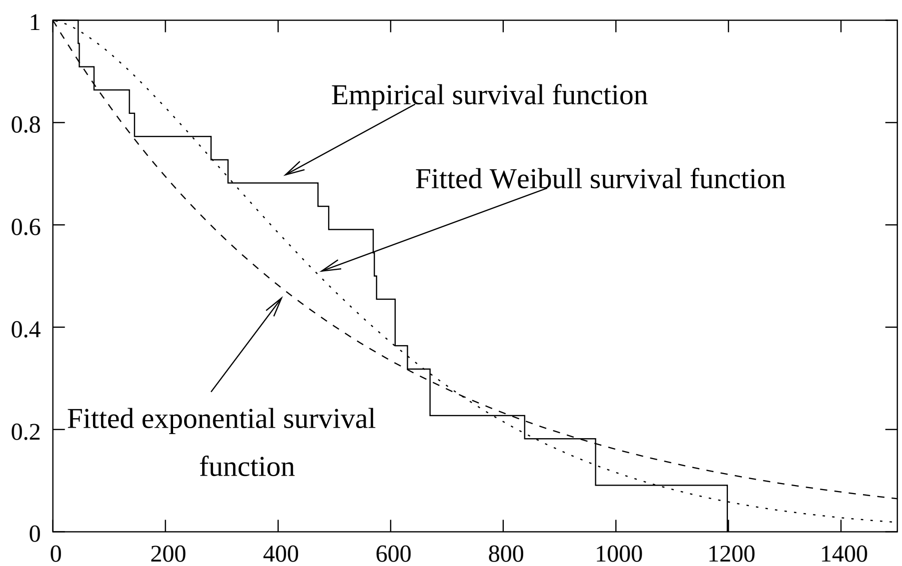

Figure 5: The empirical survival function and the fitted survival functions for Group II.

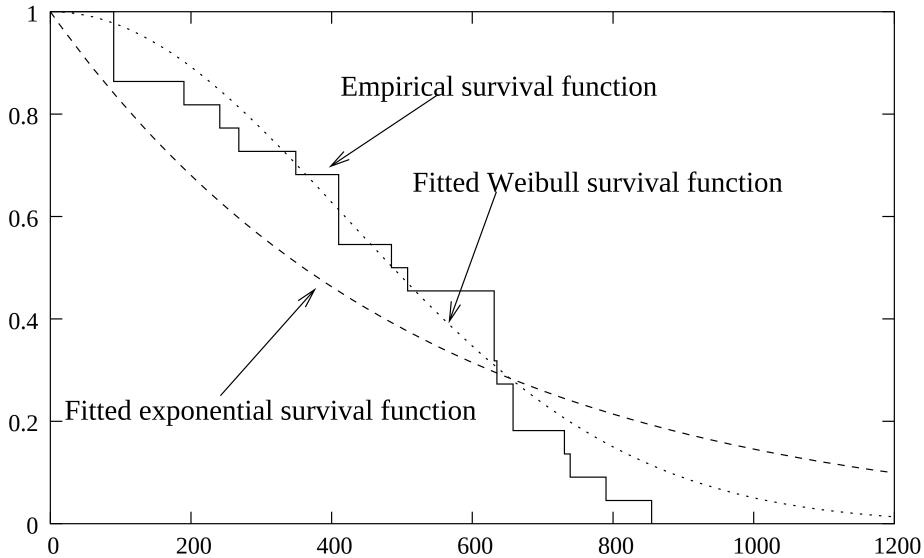

Figure 6: The empirical survival function and the fitted survival functions for Group III.

Group	Weibull		Exponential	
	K-S distance	p-value	K-S distance	p-value
I	0.1505	0.7284	0.2251	0.2376
II	0.1907	0.4002	0.2583	0.1063
III	0.1903	0.3495	0.2544	0.0896

Table 3: K-S distances between the empirical distribution functions and the fitted Weibull and fitted exponential distribution functions and the corresponding p-values for the three groups.

The Kolmogorov-Smirnov distances of the empirical distribution function and the fitted distribution for the whole data set and the corresponding p value for the three groups are presented in Table 3. From the p values of Table 3, it is quite clear that the Weibull model provides much better fit than the exponential distribution in all the groups. Based on the K-S distances and p-values, it can be said that although exponential distribution can be used for Group-I but it should not be used for Group-II and Group-III.

Finally we present the log-likelihood values for the fitted Weibull and fitted exponential distributions and the corresponding p values for the different groups in the Table 4. Now comparing $2 \times (LL_W - LL_E)$ with the χ_1^2 upper tail it is observed that in all cases $p < 0.1$.

Group	Weibull Log-likelihood (LL_W)	Exponential Log-likelihood (LL_E)	$2 \times (LL_W - LL_E)$	p-value
I	-189.3438	-193.4463	8.2050	< 0.005
II	-192.9975	-194.4843	2.9736	< 0.10
III	-206.6923	-210.6309	7.8772	< 0.01

Table 4: The log-likelihood values for the fitted Weibull, fitted exponential distributions and their differences for the three groups.

Therefore, the log-likelihood values indicate that Weibull distribution should be fitted rather than exponential distribution to the different groups.

7 Conclusions

This work is mainly the continuation of the work of Park and Kulasekera (2004). Park and Kulasekera (2004) analyzed Nelson's (1970) data assuming the lifetime distributions to be exponential for different groups. We have reanalyzed that data set using both exponential and Weibull models. We have used different criteria, namely K-S distance and log-likelihood value and it is observed that the assumptions of exponential distributions may not be appropriate for all the three groups. It is observed that instead of exponential distribution, the Weibull distribution may be used in this case. We have developed the complete classical inferential procedures for several groups when the lifetime distributions of the items of each group are Weibull. It may be mentioned that the corresponding Bayesian inferential procedure can also be developed in this case under the assumptions of suitable priors of the unknown parameters. Work is in progress and it will be reported elsewhere.

Acknowledgements

The authors would like to thank the referees for carefully reading the manuscript and for their valuable suggestions. The authors would also like to thank the associate editor Professor Caroline Smith for her encouragements.

References

Cox, D. R. (1959), "The analysis of exponentially distributed lifetimes with two types of failures", *Journal of the Royal Statistical Society, Series B*, vol. 21, 411 - 421.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), "Maximum likelihood estimation from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society, Ser. B*, vol. 39, 1 - 22.

Efron, B. and Hinkley, D. V. (1978), "The observed versus expected information", *Biometrika*, vol. 65, 457 - 487.

Kundu, D. (2004), "Parameter estimation of the partially complete time and type of failure data", *Biometrical Journal*, vol. 46, no. 2, 165 - 179.

Lehmann, E.L. and Romano, J.P. (2005), *Testing Statistical Hypotheses*, 3rd. Edition, Springer, New York.

Louis, T. A. (1982), "Fitting the observed information matrix when using the EM algorithm", *Journal of the Royal Statistical Society, Series B*, vol. 44, 226 - 233.

Nelson, W. (1970), "Hazard plotting methods for analysis of life data with different failure modes", *Journal of Quality Technology*, vol. 2, 126 - 149.

Park, C. and Kulasekera, K. B. (2004), "Parametric inference of incomplete data with competing risks among several groups", *IEEE Transactions on Reliability*, vol. 53, 11 - 21.

Prentice, R. L., Kalbfleish, J. D., Peterson, Jr. A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978), "The analysis of failure times in presence of competing risks", *Biometrics*, vol. 34, 541 - 554.

Wu, C. F. J. (1983), "On the convergence properties of the EM algorithm", *Annals of Statistics*, vol. 11, 95 - 103.