Non-Parametric Maximum Likelihood Estimation of Current Status Data with Competing Risks

Technical Report No. ASU/2016/6
Dated 22 December, 2016

Tamalika Koley
and
Anup Dewanji

Applied Statistics Unit
Indian Statistical Institute
Kolkata 700 108

Non-Parametric Maximum Likelihood Estimation of Current Status Data with Competing Risks

Tamalika Koley[1] and Anup Dewanji[2]
Applied Statistics Unit, Indian Statistical Institute
203, B.T. Road, Kolkata-700108
email[1] : tamalikakoley@gmail.com
email[2] : dewanjia@isical.ac.in

January 3, 2017

1 Introduction

In survival analysis, censoring is a very common phenomenon where the information about the survival time is incomplete. Current status data is a type of censoring in which each individual under study is observed only once and the current status of the individual is observed at the monitoring time recording whether the event of interest has occurred or not. The individuals experiencing the event of interest are exposed to risk of the event due to two or more causes called the competing risks (See Kalbfleisch and Prentice,2002). If the event of interest occurs for any individual, then the cause of the event is also observed. This type of data is called current status data with competing risks.

A great deal of recent attention has been focused on the estimation of sub-distribution functions based on current status data with competing risks (See Hudgens et al., 2001; Jewell et al., 2003; Jewell and Kalbfleisch, 2004; Maathuis, 2006; and many others). Consider the motivating example discussed in the thesis of Maathuis (2006) about the recent clinical trials of candidate vaccines for HIV/AIDS. Like many viruses, HIV exhibits various phenotypic and genotypic variations so that it can be distinguished into several subtypes. The main purpose of such trials is to determine the efficacy of a vaccine against each subtype. The variables of interest are the time of infection and the subtype of the virus causing the infection. However, the time of infection cannot be observed directly. It is only observed whether or not the infection has occurred before the time of the test along with the subtype of the infecting virus if the infection has occurred.

Jewell and Kalbfleisch (2004) have considered maximum likelihood estimation of ordered multinomial parameters, motivated by current status data with competing risks. They have suggested a modified pool adjacent violators algorithm along with an iterative algorithm to obtain the MLE of sub-distribution functions, when the monitoring times are discrete and prefixed. Two different algorithms have been suggested depending on the number of survivals observed at the last monitoring time point. The major difficulty in this method is that it fails to provide any asymptotic distribution of the MLEs and hence interval estimates for any quantity of interest. Maathuis (2006) have considered continuous, random monitoring times and used the reasoning of Turnbull (1976) that the estimators can only assign masses to a finite number of disjoint intervals, called the maximal intersections. Then the computation of MLE can be thought of as a two step process. First, in the reduction step, maximal intersections are obtained and the likelihood is written in terms of the probability of these intervals. Second is the optimization step, where the likelihood is optimized with respect to these probabilities following certain constraints. Thus, this becomes a constraint optimization problem. Maathuis and Hudgens (2011) have studied the current status data with competing risks for continuous or discrete monitoring times. The maximum likelihood estimator and a naive estimator have been developed along with their asymptotic properties. However, due to restrictions on the sub-distribution functions, they have done a constraint optimization to obtain the estimates of the sub-distribution functions.

In this paper, we consider the same set-up as that of Jewell and Kalbfleisch (2004) with a fixed number of monitoring times. We develop an alternative method for estimating the sub-distribution functions non-parametrically when both the monitoring times and the number of individuals observed at any such time are fixed. This method considers a re-parametrization in terms of interval hazards and then use the EM algorithm (Dempster, Rubin; 1977), as remarked at the beginning of Section 8 of Jewell and Kalbfleisch (2004). This turns out to be straightforward and easy to implement. The proposed EM algorithm computes the non-parametric maximum likelihood estimates of the quantities of interest and obtains some asymptotic results. Both the method of Jewell and Kalbfleisch and the proposed method have been implemented on some simulated data. A real life data from the National Center for Health Statistics' Health Examination Survey, discussed by Krailo and Pike (1983), has been analyzed for illustration. The proposed method clearly applies to the situation when the monitoring time point for an individual is random but taking finite, discrete values. Thus, the number of individuals observed at any time point is random.

In Section 2, we describe the data and construct the related likelihood function. Section 3 proposes the re-parametrization followed by the EM algorithm. While Section 4 provides the existence and uniqueness of the non-parametric maximum likelihood estimate (NPMLE), Section 5 derives the asymptotic results, along with the associated standard errors. Section 6 suggests a modification required for the case of random number of observations at each monitoring time. Section 7 presents a detailed simulation study to investigate the finite sample properties of the NPMLE and Section 8 illustrates the proposed method through analysis of a real data set. Section 9 ends with some concluding remarks.

2 The Data and The Likelihood

Let T be the random variable representing the incidence time of the event of interest (that is, failure). Assume that the individuals under study are exposed to the risk of the event due to m different causes. Let $J \in \{1, 2, .., m\}$ be the random variable denoting the cause of the event occurrence. The sub-distribution function for failure due to cause j is given by $F_j(t) = P\left[T \leq t, J = j\right]$, for $j = 1, .., m$. The overall distribution function of T is given by $F(t) = P\left[T \leq t\right] = \sum_{j=1}^{m} F_j(t)$. The survivor function is given by $S(t) = P\left[T > t\right] = 1 - F(t)$.

As already mentioned, the current status (survival/failure) of an individual is observed at a single monitoring time point and also the corresponding cause of failure is observed, if the event of interest has occurred. So, for each individual, we observe (X, Δ), where X is the fixed monitoring time point and Δ denotes the random variable representing the status of the individual at the monitoring time point X, defined as $\Delta = j$ if $T \leq X$ and $J = j$, for $j = 1, .., m$, and $\Delta = 0$ if $T > X$. Therefore, for n individuals under study, we observe (x_i, δ_i), in which x_i is fixed and δ_i is the current status of the i^{th} individual, for $i = 1, .., n$. In this set-up, the current status data takes the following simplified form. Let $0 < \tau_1 < \tau_2 < ... < \tau_k$ be the k distinct fixed monitoring time points. Suppose n_i individuals are observed at time point τ_i, of which d_{ji} are observed to have failed due to cause j, for $j = 1, .., m$, and $n_i - d_i$ are observed to have survived, where $d_i = \sum_{j=1}^{m} d_{ji}$. Write $n = \sum_{i=1}^{k} n_i$, total number of individuals under study. Note that this representation of data is derived from the observations $\{(x_i, \delta_i), i = 1, .., n\}$. Nevertheless, the likelihood of this data is given by

$$L_I = \prod_{i=1}^{k} \left\{ \prod_{j=1}^{m} F_j(\tau_i)^{d_{ji}} \right\} S(\tau_i)^{n_i - d_i}. \quad (1)$$

As in Jewell and Kalbfleisch (2004), writing $p_{ji} = F_j(\tau_i)$, for $j = 1, .., m$ and $i = 1, ..., k$, this likelihood can be written in terms of the p_{ji}'s as

$$L_I(\underset{\sim}{p}) = \prod_{i=1}^{k} \left[\left\{ \prod_{j=1}^{m} p_{ji}^{d_{ji}} \right\} (1 - \sum_{j=1}^{m} p_{ji})^{n_i - d_i} \right], \quad (2)$$

where $\underset{\sim}{p}$ denotes the vector of p_{ji}'s.

3 Estimation using EM algorithm

We wish to estimate these sub-distribution functions $F_j(.)$'s non-parametrically. It is clear from the likelihood (1) that these $F_j(.)$'s are estimable only at the monitoring time points $\tau_1 < \tau_2 < ... < \tau_k$. Using likelihood (2), therefore, these p_{ji}'s can be estimated using a modified Pooling-Adjacent Violators (PAV) algorithm. This algorithm, as presented in Jewell and Kalbfleisch (2004) for $m = 2$ and may be extendable for larger m, is very complicated because of the inherent isotonic constraints in the p_{ji}'s. Moreover, the derivation of asymptotic results including the variance formula remains unattempted because of this difficulty. In order to alleviate this difficulty, we reparametrize the model in terms of 'interval cause-specific hazards' by writing
$\lambda_{ji} = P\left[T \in I_i, J = j \mid T > \tau_{i-1}\right]$, where $I_i = (\tau_{i-1}, \tau_i]$, for $j = 1, .., m$, $i = 1, ..., k$, with $\tau_0 = 0$. Then, it is easy to verify that

$P\left[T > \tau_i\right] = \prod_{l=1}^{i}(1 - \sum_{j=1}^{m} \lambda_{jl}) = \prod_{l=1}^{i}(1 - \lambda_l)$, where $\lambda_l = \sum_{j=1}^{m} \lambda_{jl}$, and

$P\left[T \le \tau_i, J = j\right] = \sum_{l=1}^{i} \lambda_{jl} \prod_{l'=1}^{l-1}(1 - \lambda_{l'})$.

Here, we make an implicit assumption that $T > 0$ with probability 1. Therefore, the likelihood function (1) can be written in terms of the λ_{ji}'s as

$$L_I(\underset{\sim}{\lambda}) = \prod_{i=1}^{k} \left[\prod_{j=1}^{m} \left\{ \sum_{l=1}^{i} \lambda_{jl} \prod_{l'=1}^{l-1} (1 - \lambda_{jl'}) \right\}^{d_{ji}} \left\{ \prod_{l=1}^{i} (1 - \lambda_l) \right\}^{n_i - d_i} \right], \quad (3)$$

where $\underset{\sim}{\lambda}$ denotes the vector of λ_{ji}'s. Noting that $\lambda_{ji} = \frac{p_{ji} - p_{j,i-1}}{1 - \sum_{j=1}^{m} p_{ji}}$, with $p_{j0} =$

0, it is clear from (3) that the λ_{ji}'s are the estimable quantities. Since the λ_{ji}'s are unconstrained, the estimation is relatively easier, by using the Expectation- Maximization (EM) algorithm of Dempster et al (1977), as described in the following.

While applying the EM algorithm, the complete data version for an observation of the type $\{T \leq \tau_i, J = j\}$ assumes that the corresponding interval I_l, $l = 1, .., i$, at which the failure has occurred is known. Note that the individuals, who are observed to have failed at monitoring time point τ_i, can fail at any one of the intervals $I_1, ..., I_i$. So, the complete data corresponding to an observation $(X, \Delta = j)$ at monitoring time point X, for any $j = 1, .., m$, is of the form $(I, X, \Delta = j)$, where I denotes the interval in which the event of type j has occurred. Therefore, the complete data likelihood is given by

$$\begin{aligned} L_C(\underset{\sim}{\lambda}) &= \prod_{i=1}^{k} \left[\left\{ \prod_{j=1}^{m} \prod_{l=1}^{i} P(T \in I_l, J = j)^{d_{jil}} \right\} P(T > \tau_i)^{n_i - d_i} \right] \\ &= \prod_{i=1}^{k} \left[\prod_{j=1}^{m} \prod_{l=1}^{i} \left\{ \lambda_{jl} \prod_{l'=1}^{l-1} (1 - \lambda_{l'}) \right\}^{d_{jil}} \left\{ \prod_{l'=1}^{i} (1 - \lambda_{l'}) \right\}^{n_i - d_i} \right], \qquad (4) \end{aligned}$$

where d_{jil} $(l \leq i)$ is the number of individuals failed in I_l due to cause j and observed at the monitoring time point τ_i. Note that the d_{jil}'s are unobserved and $d_{ji} = \sum_{l=1}^{i} d_{jil}$ is observed.

The E-step requires the conditional expectation of the d_{jil}'s given the incomplete data and the initial estimates of the λ_{ji}'s. The M-step maximizes the complete data log likelihood involving these conditional expectations with respect to λ_{ji}'s to obtain the improved estimates. Let us denote the vector $\underset{\sim}{\lambda}$ of λ_{ji}'s with appropriate superscripts at different iterations. Then, at the $(t+1)^{\text{th}}$ iteration, the E-step calculates

$$\begin{aligned} \mathrm{E}[d_{jil} \mid \text{Incomplete data}, \underset{\sim}{\lambda}^{(t)}] &= d_{ji} \frac{\lambda_{jl}^{(t)} \prod_{l'=1}^{l-1} (1 - \lambda_{l'}^{(t)})}{\sum_{l'=1}^{i} \lambda_{jl'}^{(t)} \prod_{l''=1}^{l'-1} (1 - \lambda_{l'}^{(t)})} \\ &= d_{jil}^{(t+1)}, \text{ say}, \qquad (5) \end{aligned}$$

for $l = 1, .., i$, $j = 1, ..., m$ and $i = 1, .., k$. The M-step gives the next improved estimate as

$$\lambda_{jl}^{(t+1)} = \begin{cases} \dfrac{\sum_{i=l}^{k} d_{jil}^{(t)}}{\sum_{i=l}^{k} d_{+il}^{(t)} + \sum_{i=l+1}^{k} \sum_{l'=l+1}^{i} d_{+il'}^{(t)} + \sum_{i=l}^{k} (n_i - d_i)}, & \text{if } l = 1, .., (k-1) \\ \dfrac{d_{jkk}^{(t)}}{d_{+kk}^{(t+1)} + (n_k - d_k)}, & \text{if } l = k, \end{cases} \qquad (6)$$

for $j = 1, .., m$, where $d_{+il}^{(t+1)} = \sum_{j=1}^{m} d_{jil}^{(t+1)}$. Iterating E- and M-steps of the algorithm, we obtain the MLE of the λ_{ji}, denoted by $\hat{\lambda}_{ji}$, while the MLE of the vector $\underset{\sim}{\lambda}$ is denoted by $\hat{\underset{\sim}{\lambda}}$. Using the invariance property, the MLE of $F_j(\tau_i)$ is obtained as $\hat{F}_j(\tau_i)^{\text{EM}} = \sum_{l=1}^{i} \hat{\lambda}_{jl} \prod_{l'=1}^{l-1}(1 - \hat{\lambda}_{l'})$, where the superscript EM means the use of EM algorithm, for $j = 1, .., m$, $i = 1, .., k$.

4 Existence and Uniqueness

Following Theorem 1 of Dempster et al. (1977), we have $L_I(\underset{\sim}{\lambda}^{(t+1)}) \geq L_I(\underset{\sim}{\lambda}^{(t)})$, for $t = 0, 1, 2, ...$ (That is, the incomplete data likelihood increases at each iteration). Also, note that $L_I(\underset{\sim}{\lambda})$ is bounded above by 1. So, the sequence of the likelihood functions $\left\{ L_I(\underset{\sim}{\lambda}^{(t)}), t = 0, 1, 2, ... \right\}$ converges. Also the mapping $\underset{\sim}{\lambda}^{(t)} \mapsto \underset{\sim}{\lambda}^{(t+1)}$ is continuous as $E\left[\log L_C(\underset{\sim}{\lambda}) \mid \text{Incomplete Data}, \underset{\sim}{\lambda}^{(t)}\right]$ is continuous in both $\underset{\sim}{\lambda}$ and $\underset{\sim}{\lambda}^{(t)}$. Then, following Theorem 2 of Wu (1983), it is clear that $L_I(\underset{\sim}{\lambda}^{(t)})$ converges monotonically to $L_I(\underset{\sim}{\lambda}^*)$ for some stationary point $\underset{\sim}{\lambda}^*$. But this $\underset{\sim}{\lambda}^*$ can be a local maxima or a saddle point. We show that this is a local maxima.

Claim: The incomplete Hessian matrix based on $L_I(\underset{\sim}{\lambda})$ in (3) is positive definite.

Proof: It is clear from the definitions in Sections 2 and 3 that the vector $\underset{\sim}{\lambda}$ and the vector $\underset{\sim}{p}$ are one-one functions of each other. Therefore, it is equivalent to show that the hessian matrix based on the likelihood $L_I(\underset{\sim}{p})$ in (2) is positive definite. Note that the hessian matrix based on $L_I(\underset{\sim}{p})$ is given by the block-diagonal matrix

$$\begin{pmatrix} I_{11} & 0 & 0 & \cdots & 0 \\ 0 & I_{22} & 0 & \cdots & 0 \\ 0 & 0 & I_{33} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & I_{kk} \end{pmatrix},$$

where the i^{th} block I_{ii} of dimension $m \times m$ corresponds to the i^{th} monitoring time point τ_i, for $i = 1, .., k$. The $(j, j^{'})^{th}$ element of I_{ii} is given by

$$-\frac{\partial^2 log L_I(p)}{\partial p_{ji} \partial p_{j^{'}i}} = \begin{cases} \frac{d_{ji}}{(p_{ji})^2} + \frac{(n_i - d_i)}{(1-\sum_{j=1}^{m} p_{ji})^2}, & \text{if } j^{'} = j \\ \frac{(n_i - d_i)}{(1-\sum_{j=1}^{m} p_{ji})^2}, & \text{if } j^{'} \neq j, \end{cases}$$

for $j, j^{'} = 1, ..., m$. This hessian matrix is positive definite if and only if the I_{ii}'s are positive definite for all i.

Case 1 $(d_i = n_i)$: In this case, I_{ii} is a diagonal matrix with diagonal entries $\frac{d_{ji}}{p_{ji}^2}$, $j = 1, .., m$. So I_{ii} is positive definite if $d_{ji} > 0$ for all j. If $d_{ji} = 0$ for some j, then the likelihood is a decreasing function of p_{ji}. So, the p_{ji} is estimated as the minimum possible value. Since $p_{j,i-1} \leq p_{ji}$, the estimate of p_{ji} is the same as that of $p_{j,i-1}$ when $d_{ji} = 0$, with $p_{j0} = 0$. In other words, we take $\hat{F}_j(\tau_i) = \hat{F}_j(\tau_{i-1})$ with $\hat{F}_j(\tau_0) = 0$. We, therefore, reduce the problem by considering the likelihood only as a function of the remaining $p_{j^{'}i}$'s for $j^{'} \neq j$. As a function of these remaining $p_{j^{'}i}$'s, the likelihood is again strictly concave and there is unque MLE of these $p_{j^{'}i}$'s.

Case 2 $(d_i < n_i)$: In this case, I_{ii} is positive definite if at most one $d_{ji} = 0$. Otherwise, there will be at least two identical rows in I_{ii}. However, in this case also, if $d_{ji} = 0$ for some j, the likelihood is a decreasing function of p_{ji} resulting in the minimum possible value of p_{ji} as its MLE, that is $\hat{p}_{ji} = \hat{p}_{j,i-1} = \hat{F}_j(\tau_{i-1})$.

From the above claim, it is clear that the observed information matrix is positive definite. Thus, the EM algorithm converges to a local maxima. Also, since the incomplete information matrix is positive definite, the likelihood is strictly concave and hence has a unique maxima. It is well-known that any local extrema of a concave function is also a global extrema. Thus, the EM algorithm converges to a unique MLE.

5 Asymptotic Results

Note that, in this set-up, no matter how large n is, the sub-distribution functions $F_j(\cdot)$'s are estimable only at the monitoring times $\tau_1 < \tau_2 < .... < \tau_k$. Therefore, we can only study the limiting distributions, as $n \longrightarrow \infty$, of the $\hat{F}_j(\cdot)$'s at the τ_i's, or equivalently of the $\hat{p}_{ji}$'s. This, in turn, reduces to the study of the limiting distributions of the $\hat{\lambda}_{ji}$'s, or that of $\underset{\sim}{\hat{\lambda}}$. Since the dimension of $\underset{\sim}{\lambda}$ is fixed at mk, this becomes a special case of parametric likelihood analysis. Let us note that, at each monitoring time τ_i, there are n_i independent realizations of a random variable $\delta^{(i)}$ from the identical distribution given by

$$p^i(\delta^{(i)} \mid \underset{\sim}{\lambda}) = \begin{cases} \sum_{l=1}^{i} \lambda_{jl} \prod_{l'=1}^{l-1}(1-\lambda_{l'}), & \text{if } \delta^{(i)} = j, \text{ for } j = 1, ..., m, \\ \prod_{l'=1}^{l}(1-\lambda_{l'}), & \text{if } \delta^{(i)} = 0, \end{cases} \tag{7}$$

with the counting measure as the dominating probability measure. It is easy to check that $\sum_{\delta^{(i)}=0}^{m} p^i(\delta^{(i)} \mid \underset{\sim}{\lambda}) = 1$, for all $i = 1, ..., k$. Also, note that, for $i \neq i^{'}$, the two random variables $\delta^{(i)}$ and $\delta^{(i^{'})}$ are independent but not identically distributed. It is now easy to check that the observed incomplete data likelihood $L_I(\underset{\sim}{\lambda})$ of (3) can be written as a product of the densities given by

$$L_I(\underset{\sim}{\lambda}) = \prod_{i=1}^{k} \prod_{l=1}^{n_i} p^i(\delta_l^{(i)} \mid \underset{\sim}{\lambda}), \tag{8}$$

where $\delta_l^{(i)}$ denotes the l^{th} observed value of $\delta^{(i)}$. This representation of $L_I(\underset{\sim}{\lambda})$ is useful in establishing the asymptotic properties of $\hat{\underset{\sim}{\lambda}}$ using the results of Lehmann and Casella (1998, p463-465). Let us first note the following properties:

P1 : The true value of the parameter $\underset{\sim}{\lambda}$, denoted by $\underset{\sim}{\lambda}_0$, lies in an open set. In the present framework, we have $0 < \underset{\sim}{\lambda} < 1$ in the sense that $0 \leq \lambda_{ji} \leq 1$, for all j, i, with some restrictions. Note that $\underset{\sim}{\lambda} = 0$ means $\lambda_{ji} = 0$, for all j, i, which is not of interest. Similarly, $\underset{\sim}{\lambda} = 1$ meaning $\lambda_{ji} = 1$, for all j, i, which is also of no interest.

P2 : Let $\underset{\sim}{\lambda}^{'}$ and $\underset{\sim}{\lambda}^{''}$ be two values of the parameter vector $\underset{\sim}{\lambda}$ with $\underset{\sim}{\lambda}^{'} \neq \underset{\sim}{\lambda}^{''}$. Then, there is at least one component of these two vectors which are not equal, say, $\lambda_{ji}^{'} \neq \lambda_{ji}^{''}$, for some $i = 1, .., k$ and $j = 1, .., m$. Also, the densities $p^i(\delta^{(i)} \mid \underset{\sim}{\lambda})$'s are functions of $\lambda_{ji'}$, for $j = 1, 2, .., m$ and $i^{'} \leq i$, for $i = 1, ..., k$. Hence, $p^{i^{'}}(\delta^{(i^{'})} \mid \underset{\sim}{\lambda}^{'}) \neq p^{i^{'}}(\delta^{(i^{'})} \mid \underset{\sim}{\lambda}^{''})$, for $i^{'} \leq i$, and the likelihood $L(\underset{\sim}{\lambda})$, being the product of these densities, $L(\underset{\sim}{\lambda}^{'}) \neq L(\underset{\sim}{\lambda}^{''})$.

P3 : For each $i = 1,..,k$, $E_{i,\underset{\sim}{\lambda}}\left[\frac{\partial}{\partial\underset{\sim}{\lambda}}\log p^i(\delta^{(i)} \mid \underset{\sim}{\lambda})\right] = 0$, where $E_{i,\underset{\sim}{\lambda}}[\cdot]$ is the expectation with respect to the density $p^i(\delta^{(i)} \mid \underset{\sim}{\lambda})$. Again, differentiating with respect to $\underset{\sim}{\lambda}$, we have

$$E_{i,\underset{\sim}{\lambda}}\left[(\frac{\partial}{\partial\underset{\sim}{\lambda}}\log p^i(\delta^{(i)} \mid \underset{\sim}{\lambda}))^2\right] = E_{i,\underset{\sim}{\lambda}}\left[-\frac{\partial^2}{\partial\underset{\sim}{\lambda}\partial\underset{\sim}{\lambda}^{\mathrm{T}}}\log p^i(\delta^{(i)} \mid \underset{\sim}{\lambda})\right] = \mathscr{I}_i(\underset{\sim}{\lambda}), \text{ say,}$$

which is assumed to be non-negative definite.

P4 : The density function $p^i(\delta^{(i)} \mid \underset{\sim}{\lambda})$ is a polynomial in the λ_{ji}'s. Therefore, it is continuous in each λ_{ji} and admits all third order derivatives. Also, since each λ_{ji} is bounded above by 1, it can be shown that these third order derivatives are bounded by functions with finite expectations in a neighbourhood of the true value $\underset{\sim}{\lambda}_0$ of $\underset{\sim}{\lambda}$.

We now have the following theorem.

Theorem 1 *Under the assumption that* $\frac{n_i}{n} \longrightarrow w_i$, *for all* $i = 1, ..., k$, *such that* $\sum_{i=1}^k w_i = 1$, *we have*

(i) $\hat{\underset{\sim}{\lambda}} \xrightarrow{P} \underset{\sim}{\lambda}_0$, *and*

(ii) $\sqrt{n}(\hat{\underset{\sim}{\lambda}} - \underset{\sim}{\lambda}_0)$ *is asymptotically a mean zero normal random vector with variance-covariance matrix* $\left[\sum_{i=1}^k w_i \mathscr{I}_i(\underset{\sim}{\lambda}_0)\right]^{-1}$.

Proof: Let us consider a sphere Q_a with center at the true value $\underset{\sim}{\lambda}_0$ and radius a. We will show that, for sufficiently small value of a, $\log L_I(\underset{\sim}{\lambda}) < \log L_I(\underset{\sim}{\lambda}_0)$ with probability tending to 1 as $n \longrightarrow \infty$ for all $\underset{\sim}{\lambda}$ on the surface of Q_a. So, $L_I(\underset{\sim}{\lambda})$ has a local maximum in the interior of Q_a and, hence, the likelihood equations have a solution within Q_a. Note that, from (8), we have $\log L_I(\underset{\sim}{\lambda}) = \sum_{i=1}^{k}\sum_{l=1}^{n_i} p^i(\delta_l^{(i)} \mid \underset{\sim}{\lambda})$, so that

$$\frac{1}{n}\log L_I(\underset{\sim}{\lambda}) - \frac{1}{n}\log L_I(\underset{\sim}{\lambda}_0)$$

$$= \sum_{i=1}^{k} \frac{n_i}{n} \times \frac{1}{n_i} \sum_{l=1}^{n_i} \left[\log\, p^i(\delta_l^{(i)} \mid \utilde{\lambda}) - \log\, p^i(\delta_l^{(i)} \mid \utilde{\lambda}_0) \right]. \qquad (9)$$

Then, following the proof given in Lehmann and Casella (1998, p463-465) and using the properties $P1$-$P4$, it can be proved that the maximum of (9), over all $\utilde{\lambda}$ on the surface of Q_a, is less than zero. This completes the proof of (i). Using Taylor's series expansion on the likelihood equation, we have

$$0 = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^{k} \sum_{l=1}^{n_i} \frac{\partial}{\partial \utilde{\lambda}} \log\, p^i(\delta_l^{(i)} \mid \utilde{\lambda}) \mid_{\utilde{\lambda} = \hat{\utilde{\lambda}}}$$

$$= \textstyle\sum_{i=1}^{k} \sqrt{\frac{n_i}{n}} \times \frac{1}{\sqrt{n_i}} \sum_{l=1}^{n_i} \frac{\partial}{\partial \utilde{\lambda}} \log\, p^i(\delta_l^{(i)} \mid \utilde{\lambda}) \mid_{\utilde{\lambda} = \utilde{\lambda}_0} +$$

$$\sum_{i=1}^{k} \frac{n_i}{n} \times \frac{1}{n_i} \sum_{l=1}^{n_i} \frac{\partial^2}{\partial \utilde{\lambda} \partial \utilde{\lambda}^T} \log\, p^i(\delta_l^{(i)} \mid \utilde{\lambda}) \mid_{\utilde{\lambda} = \utilde{\lambda}_0} \times \sqrt{n}(\hat{\utilde{\lambda}} - \utilde{\lambda}_0) + o_p(1). \qquad (10)$$

For a particular i, by weak law of large numbers, we have

$$\tfrac{1}{n_i} \textstyle\sum_{l=1}^{n_i} \frac{\partial^2}{\partial \utilde{\lambda} \partial \utilde{\lambda}^T} \log\, p^i(\delta_l^{(i)} \mid \utilde{\lambda}) \mid_{\utilde{\lambda} = \utilde{\lambda}_0} \xrightarrow{P} E\left[\frac{\partial^2}{\partial \utilde{\lambda} \partial \utilde{\lambda}^T} \log\, p^i(\delta_l^{(i)} \mid \utilde{\lambda}) \mid_{\utilde{\lambda} = \utilde{\lambda}_0} \right]$$

$$= -\mathscr{I}_i(\utilde{\lambda}_0).$$

Also, using Central Limit Theorem, we have

$$\tfrac{1}{\sqrt{n_i}} \textstyle\sum_{l=1}^{n_i} \frac{\partial}{\partial \utilde{\lambda}} \log\, p^i(\delta_l^{(i)} \mid \utilde{\lambda}) \mid_{\utilde{\lambda} = \utilde{\lambda}_0} \xrightarrow{d} N(0, \mathscr{I}_i(\utilde{\lambda}_0)) \text{ and}$$

$$-\tfrac{1}{n} \frac{\partial^2}{\partial \utilde{\lambda} \partial \utilde{\lambda}^T} \log\, L_I(\utilde{\lambda}) = -\sum_{i=1}^{k} \frac{n_i}{n} \times \frac{1}{n_i} \sum_{l=1}^{n_i} \frac{\partial^2}{\partial \utilde{\lambda} \partial \utilde{\lambda}^T} \log\, p^i(\delta_l^{(i)} \mid \utilde{\lambda})$$

$$\xrightarrow{p} \textstyle\sum_{i=1}^{k} w_i \mathscr{I}_i(\utilde{\lambda}) = \mathscr{I}(\utilde{\lambda}), \text{ say.}$$

Using these results and Slutsky's theorem, we have from (10), $\sqrt{n}(\hat{\utilde{\lambda}} - \utilde{\lambda}_0) \xrightarrow{d} N(0, \mathscr{I}^{-1}(\utilde{\lambda}_0))$. This completes the proof of (ii).

Note that, since $\hat{\utilde{\lambda}}$ is a consistent estimate, using the weak law of large numbers, as before, $-\frac{1}{n} \frac{\partial^2}{\partial \utilde{\lambda} \partial \utilde{\lambda}^T} \log\, L_I(\utilde{\lambda})$ evaluated at $\utilde{\lambda} = \hat{\utilde{\lambda}}$ can be taken as a consistent estimate of $\mathscr{I}(\underset{\sim}{\lambda}_0) = \sum_{i=1}^{k} w_i \mathscr{I}_i(\underset{\sim}{\lambda})$.

In the following, we describe derivation of the observed information matrix, $-\frac{\partial^2}{\partial \underset{\sim}{\lambda} \partial \underset{\sim}{\lambda}^T} \log L_I(\underset{\sim}{\lambda})$, evaluated at $\underset{\sim}{\lambda} = \hat{\underset{\sim}{\lambda}}$ and denoted by $I(\hat{\underset{\sim}{\lambda}})$. Following Louis (1982), the observed information matrix $I(\hat{\underset{\sim}{\lambda}})$ based on the likelihood $L_I(\underset{\sim}{\lambda})$ of (3) is given by

$$E\left[-\frac{\partial^2}{\partial \underset{\sim}{\lambda} \partial \underset{\sim}{\lambda}^T} \log L_C(\underset{\sim}{\lambda}) \mid \text{Incomplete Data}\right]_{\underset{\sim}{\lambda} = \hat{\underset{\sim}{\lambda}}}$$

$$-E\left[\left(\frac{\partial \log L_C(\underset{\sim}{\lambda})}{\partial \underset{\sim}{\lambda}}\right)\left(\frac{\partial \log L_C(\underset{\sim}{\lambda})}{\partial \underset{\sim}{\lambda}}\right)^T \mid \text{Incomplete Data}\right]_{\underset{\sim}{\lambda} = \hat{\underset{\sim}{\lambda}}}. \quad (11)$$

In order to evaluate (11), the following results, given the incomplete data, are useful.

Result 1 : For fixed j and i, given d_{ji},

$\{d_{jil}, l = 1, ..., i\} \sim \text{Multinomial}(d_{ji}; p_{ji1}, p_{ji2}, ..., p_{jii})$,

where $p_{jil} = \frac{\lambda_{jl} \prod_{l'=1}^{l-1}(1-\lambda_{l'})}{\sum_{l'=1}^{i} \lambda_{jl'} \prod_{l''=1}^{l-1}(1-\lambda_{l''})}$, for $l = 1, .., i$.

Proof: Let us consider those d_{ji} individuals with monitoring time point τ_i and are observed to be failed due to cause j, for a given j and i. Those individuals can fail at any one of the intervals $I_1, ..., I_i$. Given incomplete data (death due to cause j with the monitoring time point τ_i), the probability of failure in the interval I_l is given by

$p_{jil} = P\left[T \in I_l \mid T \le \tau_i, J = j\right] = \frac{P[T \in I_l, J=j]}{P[T \le \tau_i, J=j]} = \frac{\lambda_{jl} \prod_{l'=1}^{l-1}(1-\lambda_{l'})}{\sum_{l'=1}^{i} \lambda_{jl'} \prod_{l''=1}^{l-1}(1-\lambda_{l''})}$,

for $l = 1, .., i$. The result follows from the definition of d_{jil} and the fact that $\sum_{l=1}^{i} d_{jil} = d_{ji}$ is given.

Result 2 : d_{jil} and $d_{j'i'l'}$ are independent for $i \neq i'$ and for all j, j',with $l \le i, l' \le i'$.

Proof: Note that each individual is observed at exactly one monitoring time point out of $\tau_1, ..., \tau_k$. Therefore, those individuals who are observed as failed at τ_i are different from those observed as failed at $\tau_{i'}$, for $i^{'} \neq i$. Hence, the random variables d_{jil} and $d_{j'i'l'}$ are independent, for $i \neq i^{'}$ and for all $j, j^{'}, l \leq i, l^{'} \leq i^{'}$.

Result 3 : $\text{cov}(d_{jil}, d_{j'il'}) = 0$ for $j \neq j^{'}$ and for all $l, l^{'} \leq i$.

Proof: As argued in the proof of Result 2, given the incomplete data, the individuals observed as failed due to cause j at time τ_i are different from those observed as failed due to cause $j^{'}(\neq j)$ at the same time τ_i. Hence, the random variables d_{jil} and $d_{j'il'}$, for $j \neq j^{'}$, are independent.

Using these three results, calculation of observed information matrix $I(\hat{\underset{\sim}{\lambda}})$ from (11) is straightforward. The variance-covariance matrix of the MLE $\hat{\underset{\sim}{\lambda}}$, given by $\mathscr{I}^{-1}(\underset{\sim}{\lambda_0})/n$, is estimated by the inverse of the observed infromation matrix, $I^{-1}(\hat{\underset{\sim}{\lambda}})$. To estimate the variance of $\hat{F}_j(\tau_i)$, for $j = 1, .., m$ and $i = 1, .., k$, note that $p_{ji} = F_j(\tau_i) = \sum_{l=1}^{i} \lambda_{jl} \prod_{l'=1}^{l-1}(1 - \lambda_{l'}) = g_{ji}(\underset{\sim}{\lambda})$, say. Let us write A_{ji} to be the vector $\frac{\partial}{\partial \underset{\sim}{\lambda}} g_{ji}(\underset{\sim}{\lambda})$, evaluated at $\underset{\sim}{\lambda} = \hat{\underset{\sim}{\lambda}}$. Note that this $F_j(\tau_i) = g_{ji}(\underset{\sim}{\lambda})$ is a continuous function of $\underset{\sim}{\lambda}$. Therefore, using the delta method, the estimated variance of $\hat{F}_j(\tau_i)$ is given by $A_{ji}^T I^{-1}(\hat{\underset{\sim}{\lambda}}) A_{ji}$.

6 Random Monitoring Time

Now suppose the monitoring time X for an individual is a random variable taking discrete values $\tau_1, ..., \tau_k$. In this case, individuals are observed randomly at any one of the τ_i's. Such situation arises when, for example, set of possible monitoring times are fixed, but the individuals are observed at one of these monitoring times as per their availability. Assume that the random monitoring time X is independent of the lifetime T. Therefore, a typical observation on an individual is of the form (X, Δ), as in Section 2, but with X now being random. The observations on n individuals are independent realizations $\{(x_i, \delta_i), i = 1, .., n\}$ from the common density

$$p(\tau,\delta) = \begin{cases} \sum_{l=1}^{i} \lambda_{jl} \prod_{l'=1}^{l-1}(1-\lambda_{l'})P\left[X=\tau_i\right], & \text{if } \delta = j, \tau = \tau_i, \\ \prod_{l'=1}^{i}(1-\lambda_{l'})P\left[X=\tau_i\right], & \text{if } \delta = 0, \tau = \tau_i \end{cases}$$

for $j = 1,..,m$ and $i = 1,...,k$, with the kronecker product of discrete measure and counting measure being the dominating probability measure. It is now easy to check that the observed likelihood can be written as the product of densities given by $\prod_{i=1}^{n} p(x_i, \delta_i)$, which is proportional to the likelihood (3) with n_i being the random number of individuals monitored at τ_i. Therefore, the same estimation procedure, as that of Section 3, can be carried out to obtain the MLE $\hat{\underset{\sim}{\lambda}}$ of $\underset{\sim}{\lambda}$. In order to study the asymptotic properties, one can easily verify the conditions mentioned in the properties $P1 - P4$ for this particular case as well. Then, as in the previous section, we have the following theorem.

Theorem 2 *Assume that* $p_i = P\left[X = \tau_i\right] > 0$ *for all* $i = 1, 2, ..., k$. *Then,*

(i) $\hat{\underset{\sim}{\lambda}} \xrightarrow{P} \underset{\sim}{\lambda}_0$, *and*

(ii) $\sqrt{n}(\hat{\underset{\sim}{\lambda}} - \underset{\sim}{\lambda}_0)$ *is asymptotically a mean zero normal random vector with variance-covariance matrix* $\left[\sum_{i=1}^{k} p_i \mathscr{I}_i(\underset{\sim}{\lambda}_0)\right]^{-1}$

Proof: Note that $\frac{n_i}{n} \xrightarrow{P} p_i > 0$, for all $i = 1, 2, .., k$. Then, the proof follows that of Theorem 1.

This variance-covariance matrix of $\hat{\underset{\sim}{\lambda}}$ can be estimated, as before, by the inverse of the observed information matrix, that can be obtained by using the Louis' (1982) method.

7 Simulation Studies

In order to investigate the finite sample properties of the proposed estimates $\hat{F}_j^{EM}(\cdot), j = 1, ..., m$, we carry out an extensive simulation study. In the process, for $m = 2$, we also compare with the estimates $\hat{F}_j^{JK}(\cdot), j = 1, 2$, of Jewell and Kalbfleisch (2004). Since, as expected, the two sets of estimates give same values, the estimates $\hat{F}_j^{JK}(\cdot)$'s are not reported.

First we simulate n observations from the assumed life distribution $F(\cdot)$. Then, the first n_1 observations are considered to be monitored at time τ_1, the next n_2 observations at time τ_2, and so on. Once an observation at time τ_i is observed to have failed, the cause of failure is ascertained using the probability distribution $P[J = j \mid T = t]$, for $j = 1, .., m$, where t denotes the corresponding failure time. This probability is given by $\lambda_j(t)/\sum_{j=1}^{m} \lambda_j(t)$, where the $\lambda_j(t)$'s denote the cause specific hazards due to cause j. This leads to a simulated data set $\{(x_i, \delta_i), i = 1, .., n\}$, as described in Section 2. For each simulation study in this section, we consider 10000 such simulated data sets.

For each simulated data set, we obtain $\hat{F}_j^{EM}(\tau_i)$ using the method of Section 3, along with its standard error as described in Section 5, for $j = 1, .., m$, $i = 1, .., k$. Bias of these estimates are estimated by taking average of the differences $\hat{F}_j^{EM}(\tau_i) - F_j(\tau_i)$ over the 10000 simulations. Similarly, average of the standard errors (denoted by ASE) are also computed. Standard deviations of the estimates over the 10000 simulations are computed as the sample standard error (denoted by SSE). Cover percentage of the asymptotic 95% confidence interval of any $F_j(\tau_i)$, based on normal approximation, is estimated by the proportion of times these intervals contain the true value in the 10000 simulations (and is denoted by CP). Average length of these intervals are also computed, but not reported for space consideration. As expected, these lengths are decreasing with increasing n.

For all our simulation studies, we take $k = 5$ monitoring time points and $n_1 = ... = n_k$. We first consider exponential distribution for the failure time T with rate parameter $\lambda = 0.5$ and $m = 2$ types of failures occurring with $3 : 2$ rate ratio. That is, $F_1(t) = 0.6(1 - e^{-0.5t})$ while $F_2(t) = 0.4(1 - e^{-0.5t})$. The five monitoring points are chosen as $0.1, 1.57, 3.04, 4.51$ and 5.98, respectively. These are the $10^{th}, 25^{th}, 50^{th}, 75^{th}$ and 90^{th} quantiles of the true exponential failure time distribution with rate 0.5. The simulation is carried out for four different common n_i values given by $100, 150, 250$ and 300, respectively. The results are presented in Table 1. As expected, bias and standard error reduce with the sample size, while the average standard error (ASE) and sample standard error (SSE) are very similar, specially with increasing sample size. The estimated coverage probabilities are closer to 0.95, for larger n_i's, providing evidence in favour of asymptotic normality for individual $\hat{F}_j^{EM}(\tau_i)$'s.

We know that, if a random vector $(\underset{\sim}{x})$ follows a multivariate normal distribution, then each of its component is univariate normal random variable. Also, for a large sample, the observed Mahalanobis distance $(\underset{\sim}{x} - \bar{x})' S^{-1} (\underset{\sim}{x} - \bar{x})$ has an approximate chi-square distribution with p degrees of freedom, if $\underset{\sim}{x}$ is a p-variate random vector with sample mean and sample covariance matrix being $\bar{x}$ and S, respectively (See Ververidis and Kotropoulos, 2008). This result can be used to check whether the observed multivariate data on $\underset{\sim}{x}$ may arise from a multivariate normal distribution. A Q-Q plot, in which the ordered observed Mahalanobis distances are plotted versus the estimated quantiles in a sample of same size from χ^2_p distribution, can be used to visually check for multivariate normality. If the data arises from a multivariate normal distribution, then this plot should resemble a 45^0 straight line. In order to check the asymptotic multivariate normality of $\left\{\hat{F}_j^{EM}(\tau_i), j = 1, .., m; i = 1, .., k\right\}$ graphically, we consider the Q-Q plot of the corresponding observed Mahalanobis distances in 10000 simulations comparing with the corresponding quantiles of the $\chi^2_{(10)}$ distributions (since $m = 2$ and $k = 5$). These plots for the four sample sizes ($n_i = 100, 150, 250$ and 300) are given in Figure 1 giving ample evidence in favour of asymptotic normality.

In the next two simulation studies, we consider Weibull distribution for the failure time T with shape parameter $p(> 0)$ and scale parameter $\theta = (\sum_{j=1}^m \lambda_j^p)^{-1}$, where the cause-specific hazard rates are given by $\lambda_j(t) = \lambda_j p(\lambda_j t)^{p-1}, \lambda_j > 0$, for $j = 1, 2$. We choose $p = 0.8$ and 1.2 to represent both increasing and decreasing failure rate. The failure types occur with 3: 2 rate ratio. That is, $F_j(t) = \frac{\lambda_j^p}{\theta^p}(1 - e^{-(\theta t)^p})$, for $j = 1, 2$, with $\lambda_1 = 0.6$ and $\lambda_2 = 0.4$. The five monitoring time points are similarly chosen to be the $10^{th}, 25^{th}, 50^{th}, 75^{th}$ and 90^{th} quantiles of the true Weibull failure time distribution with the above mentioned shape and scale parameters. The simulation is carried out with n_i's equal to $100, 150, 250$ and 300 for both $p = 0.8$ and 1.2, and the estimates are based on 10000 simulations, as before. The results for $p = 0.8$ and 1.2 are presented in Tables 2 and 3, respectively. Also, the four Q-Q plots for the four sample sizes are presented in Figures 2 and 3, for $p = 0.8$ and 1.2, respectively. The results are qualitatively similar.

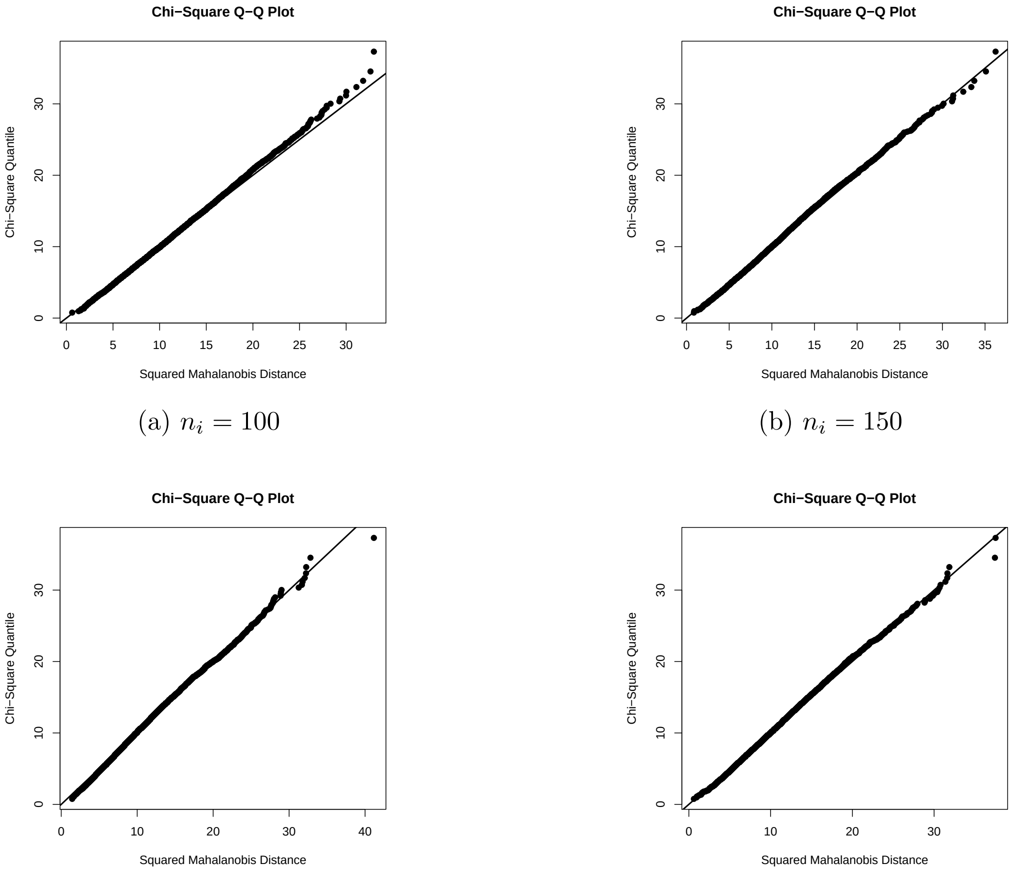
Next, we take $m = 3$ types of failures occurring with $3 : 2 : 1$ rate ratio. The failure time distribution is taken to be exponential with rate parameter $\lambda = 1.2$, so that $F_1(t) = \frac{1}{2}(1 - e^{-1.2t})$, $F_2(t) = \frac{1}{3}(1 - e^{-1.2t})$ and $F_3(t) = \frac{1}{6}(1 - e^{-1.2t})$. Here again the monitoring time points are chosen to be the $10^{th}, 25^{th}, 50^{th}, 75^{th}$ and 90^{th} quantiles of the true exponential life time distribution with rate parameter $\lambda = 1.2$ and the four sample sizes $200, 400, 600$ and 700. The results of this simulation study are presented in Table 4 and the Q-Q plots for multivariate normality are in Figure 4. The results are again similar in nature.

In the final simulation study, we consider the case when the monitoring time point for an individual is random taking finite discrete values (See Section 6). The failure time distribution is again taken to be exponential with rate $\lambda = 0.5$ and the two types of failures occur with $3 : 2$ rate ratio. The number of individuals observed at different time points is determined by assuming it to be a multinomial random vector with the probability vector

$(0.1, 0.2, 0.1, 0.3, 0.3)^T$ at time points 0.1, 1.57, 3.04, 4.51 and 5.98, as in Table 1. Four different sample sizes were taken as 250, 1250, 1750 and 2000. The results are similar as given in Table 5 and Figure 5.

Table 1: Simulation results on $\left\{\hat{F}_j^{EM}(t); j = 1, 2\right\}$ for exponential(0.5) failure time distribution and two types of failures with rate ratio 3 : 2.

		Cause 1				Cause 2			
n_i	τ_i	Bias	ASE	SSE	CP	Bias	ASE	SSE	CP
	0.1	6×10^{-5}	0.01870	0.01702	0.9062	0.00018	0.02053	0.01396	0.9082
	1.57	0.00164	0.04553	0.04683	0.9004	0.00072	0.04119	0.04025	0.9036
100	3.04	0.00115	0.04831	0.04387	0.9008	0.00255	0.04504	0.03853	0.8852
	4.51	0.00188	0.04595	0.03663	0.8983	0.00103	0.04424	0.03390	0.9033
	5.98	0.00237	0.04559	0.03685	0.8770	0.00366	0.04484	0.03621	0.8564
	0.1	4×10^{-5}	0.01392	0.01369	0.9371	9×10^{-05}	0.01365	0.01128	0.9339
	1.57	0.00134	0.03741	0.03844	0.9371	0.00039	0.03327	0.03331	0.9384
150	3.04	0.00093	0.03994	0.03702	0.9332	0.00209	0.03700	0.03325	0.9259
	4.51	0.00162	0.03882	0.03090	0.9248	0.00098	0.03720	0.02866	0.9175
	5.98	0.00163	0.03828	0.03081	0.8988	0.00215	0.03757	0.02973	0.9099
	0.1	1×10^{-5}	0.01058	0.01057	0.9455	3×10^{-5}	0.00889	0.00870	0.9498
	1.57	0.00127	0.02933	0.02971	0.9663	0.00031	0.02584	0.02591	0.9498
250	3.04	0.00072	0.03124	0.02968	0.9675	0.00134	0.02701	0.02889	0.9663
	4.51	0.00109	0.03069	0.02558	0.9328	0.00097	0.02933	0.02360	0.9257
	5.98	0.00156	0.03036	0.02540	0.9228	0.00196	0.02981	0.02476	0.9663
	0.1	1×10^{-5}	0.00967	0.00970	0.9498	0	0.00801	0.00796	0.9511
	1.57	9×10^{-04}	0.02686	0.02712	0.9598	1×10^{-5}	0.02363	0.02383	0.9528
300	3.04	0.00064	0.02860	0.02789	0.9621	0.00089	0.02646	0.02557	0.9640
	4.51	0.00088	0.02828	0.02373	0.9383	0.00077	0.02704	0.02222	0.9374
	5.98	0.00122	0.02802	0.02361	0.9253	0.00177	0.02748	0.02271	0.9613



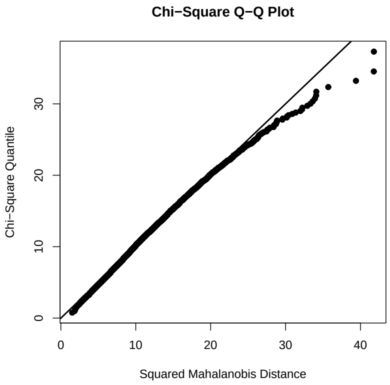
(a) $n_i = 100$ (b) $n_i = 150$

(c) $n_i = 250$ (d) $n_i = 300$

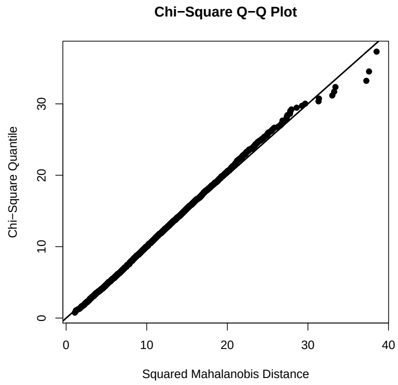
Figure 1: Multivariate QQ-plots of the Mahalanobis Squared Distance: Exponential(0.5) failure distribution and two failure types with rate ratio 3 : 2

Table 2: Simulation results on $\left\{\hat{F}_j^{EM}(t); j = 1, 2\right\}$ for Weibull failure time distribution with $p = 0.8(< 1)$ and two types of failures with rate ratio 3 : 2.

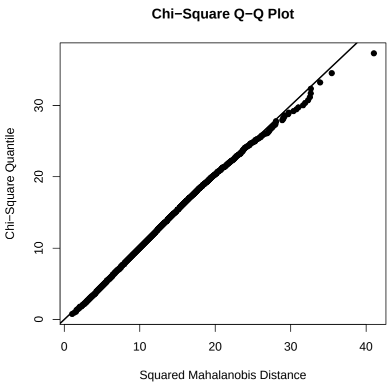
		Cause 1				Cause 2			
n_i	τ_i	Bias	ASE	SSE	CP	Bias	ASE	SSE	CP
	0.06	0.00039	0.01817	0.01694	0.8649	0.00014	0.01963	0.01415	0.8667
	2.39	7×10^{-04}	0.04583	0.04676	0.8666	0.00105	0.04174	0.04036	0.8674
100	4.72	0.0012	0.04694	0.04041	0.8614	0.00273	0.04384	0.03597	0.8545
	7.05	0.00163	0.04363	0.03428	0.7566	0.00091	0.04215	0.03247	0.7834
	9.38	0.00131	0.04358	0.03562	0.8362	0.00372	0.04284	0.03454	0.8646
	0.06	0.00015	0.01392	0.01411	0.8998	0.00013	0.01311	0.01157	0.8975
	2.39	0.00036	0.03790	0.03942	0.8942	9×10^{-04}	0.03411	0.03430	0.9007
150	4.72	0.00075	0.03889	0.03430	0.8922	0.00216	0.03633	0.03118	0.8939
	7.05	0.00069	0.03637	0.03004	0.8711	0.00071	0.03568	0.02897	0.8529
	9.38	0.0011	0.03637	0.03004	0.8736	0.00295	0.03568	0.02897	0.8762
	0.06	4×10^{-05}	0.01068	0.01093	0.9383	6×10^{-05}	0.00905	0.00908	0.9409
	2.39	3×10^{-04}	0.02998	0.03011	0.9329	0.00038	0.02687	0.02699	0.9341
250	4.72	0.00067	0.03095	0.02866	0.9255	0.00184	0.02893	0.02579	0.9339
	7.05	0.00041	0.02975	0.02373	0.8972	0.00058	0.02864	0.02236	0.9140
	9.38	0.00103	0.02940	0.02397	0.9117	0.00186	0.02882	0.02349	0.8845
	0.06	2×10^{-05}	0.00853	0.00870	0.9098	3×10^{-05}	0.00703	0.00702	0.8914
	2.39	3×10^{-04}	0.02401	0.02436	0.9306	0.00033	0.02150	0.02141	0.9302
300	4.72	0.00045	0.02472	0.02354	0.9456	0.00107	0.02313	0.02162	0.9484
	7.05	2×10^{-5}	0.02363	0.02071	0.9597	5×10^{-05}	0.02270	0.02053	0.9607
	9.38	0.00094	0.02346	0.02012	0.9560	0.00141	0.02301	0.02051	0.9570



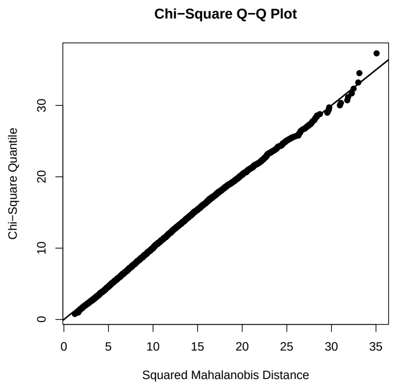
(a) $n_i = 100$



(b) $n_i = 150$



(c) $n_i = 250$

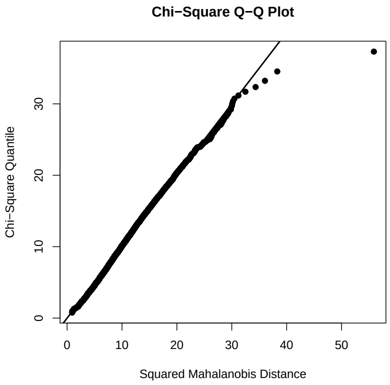


(d) $n_i = 300$

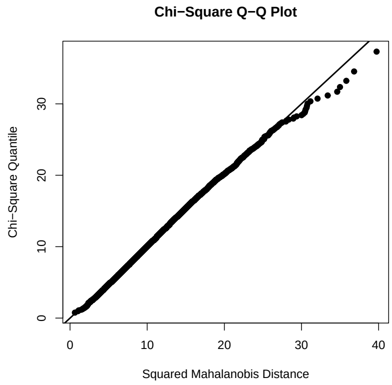
Figure 2: Multivariate QQ-plots of the Mahalanobis Squared Distance: Weibull failure time distribution ($p = 0.8$) and two failure types with rate ratio 3 : 2

Table 3: Simulation results on $\left\{\hat{F}_j^{EM}(t); j=1,2\right\}$ for Weibull failure time distribution with $p = 1.2(>1)$ and two types of failures with rate ratio 3 : 2.

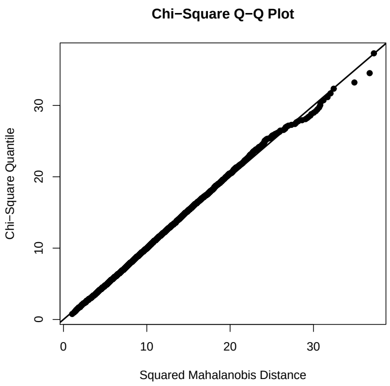
		Cause 1				Cause 2			
n_i	τ_i	Bias	ASE	SSE	CP	Bias	ASE	SSE	CP
	0.15	0.00027	0.01889	0.01691	0.8075	0.00014	0.02082	0.01416	0.9532
	1.22	0.00047	0.04437	0.04474	0.9053	0.00087	0.03991	0.03787	0.9053
100	2.29	0.00079	0.04896	0.04635	0.9266	0.00221	0.04534	0.04032	0.9266
	3.36	0.00041	0.04774	0.03820	0.9340	0.00164	0.04585	0.03573	0.9340
	4.43	9×10^{-4}	0.04722	0.03845	0.9726	0.00247	0.04648	0.03845	0.9726
	0.15	0.00018	0.01422	0.01398	0.9279	7×10^{-5}	0.01367	0.01146	0.8308
	1.22	0.00028	0.03631	0.03670	0.9157	0.00026	0.031996	0.03206	0.9157
150	2.29	0.00052	0.04018	0.03862	0.9381	0.00166	0.03689	0.03455	0.9381
	3.36	0.00018	0.03970	0.03297	0.9571	0.00152	0.03788	0.03034	0.9571
	4.43	0.00059	0.03927	0.03238	0.9270	0.00203	0.03927	0.03238	0.9270
	0.15	9×10^{-5}	0.01074	0.01066	0.9284	2×10^{-5}	0.00913	0.00894	0.8690
	1.22	0.00024	0.02831	0.02840	0.9311	4×10^{-5}	0.02464	0.02486	0.9311
250	2.29	0.00038	0.03127	0.03054	0.9548	0.00107	0.02867	0.02756	0.9548
	3.36	0.00016	0.03181	0.02672	0.9717	0.00101	0.03022	0.02475	0.9717
	4.43	0.00155	0.03158	0.02635	0.9698	0.00168	0.03158	0.02635	0.9698
	0.15	0	0.00980	0.00989	0.9318	0	0.00814	0.00802	0.9322
	1.22	0.00014	0.02588	0.02606	0.9400	1×10^{-5}	0.02248	0.02248	0.9400
300	2.29	0.00016	0.02859	0.02841	0.9520	0.00041	0.02622	0.02599	0.9520
	3.36	1×10^{-5}	0.02860	0.02480	0.9715	9×10^{-4}	0.02724	0.02400	0.9715
	4.43	0.00114	0.02832	0.02431	0.9552	0.00144	0.02778	0.02370	0.9552



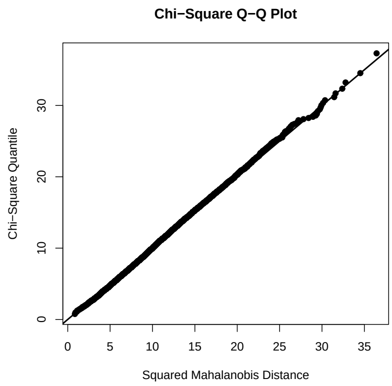
(a) $n_i = 100$



(b) $n_i = 150$



(c) $n_i = 250$

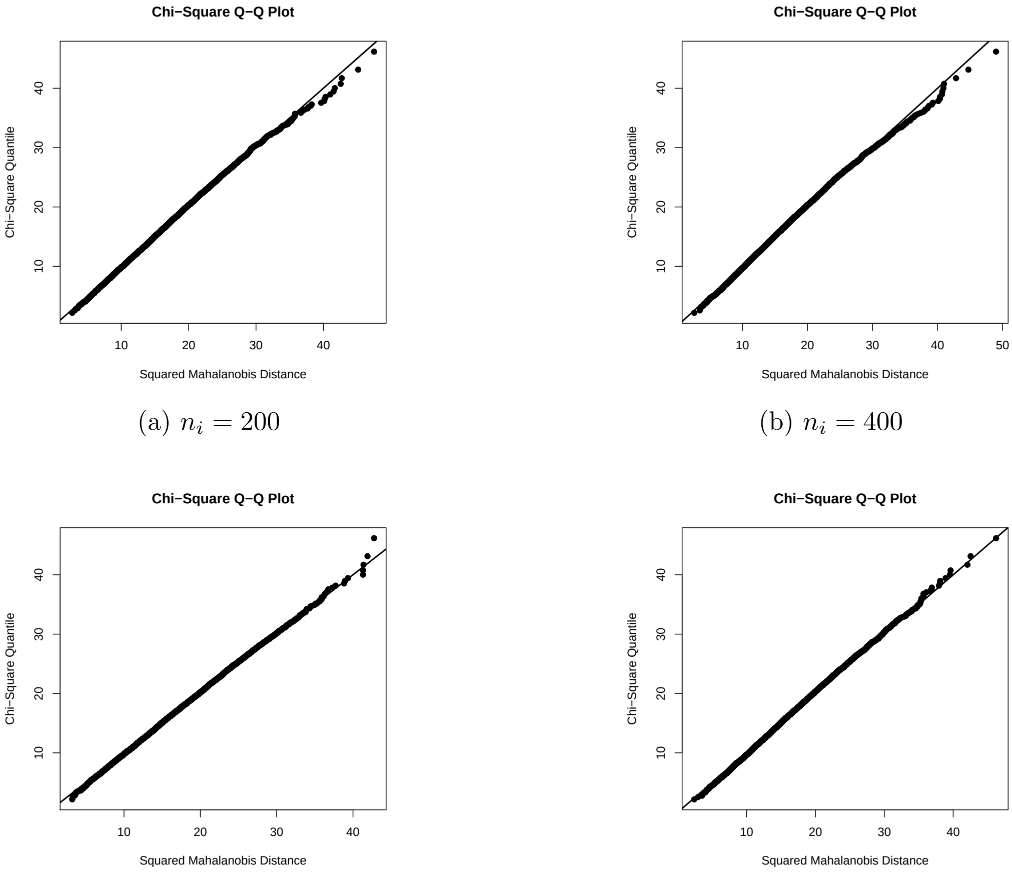


(d) $n_i = 300$

Figure 3: Multivariate QQ-plots of the Mahalanobis Squared Distance: Weibull failure time distribution ($p = 1.2$) and two failure types with rate ratio 3 : 2

Table 4: Simulation results on $\left\{\hat{F}_j^{EM}(t); j = 1, 2\right\}$ for exponential(1.2) failure time distribution with $m = 3$ and $3 : 2 : 1$ rate ratio.

		Cause 1			Cause 2			Cause 3		
n_i	τ_i	Bias	SSE	CP	Bias	SSE	CP	Bias	SSE	CP
200	0.04	0.00016	0.01067	0.8207	0.00011	0.00882	0.8216	7×10^{-5}	0.00622	0.8428
	0.65	0.00027	0.03130	0.8032	0.00017	0.026999	0.7988	0.00108	0.01918	0.8043
	1.26	0.00137	0.03178	0.8016	0.00096	0.02742	0.7250	0.00249	0.01918	0.8127
	1.87	0.00096	0.02664	0.8171	0.00074	0.02402	0.8146	0.00092	0.01792	0.8218
	2.48	0.00249	0.02717	0.8104	0.0006	0.02521	0.8116	0.00348	0.01992	0.8141
400	0.04	0.00011	0.00755	0.8392	4×10^{-5}	0.00620	0.8790	7×10^{-5}	0.00442	0.7946
	0.65	0.00014	0.02235	0.8787	0.00025	0.01910	0.8773	8×10^{-5}	0.01429	0.8740
	1.26	0.00061	0.02352	0.8847	0.00013	0.02048	0.8900	0.00115	0.01476	0.8928
	1.87	0.00049	0.02028	0.9040	0.00054	0.01826	0.9038	0.00091	0.01361	0.9044
	2.48	0.00203	0.02043	0.9031	0.00057	0.01913	0.8990	0.00217	0.01489	0.8970
600	0.04	4×10^{-5}	0.00618	0.8999	4×10^{-5}	0.00506	0.8691	2×10^{-5}	0.00361	0.9255
	0.65	0.00012	0.01806	0.9094	0.00025	0.01551	0.9107	0.00019	0.01158	0.9204
	1.26	0.00036	0.01979	0.9130	4×10^{-5}	0.01731	0.9220	0.00105	0.01258	0.9224
	1.87	0.0002	0.01736	0.9342	0.00048	0.01572	0.9380	0.00088	0.01140	0.9389
	2.48	0.00131	0.01737	0.9399	0.00048	0.01606	0.9397	0.00157	0.01253	0.9304
700	0.04	0	0.00570	0.9043	4×10^{-5}	0.00466	0.8878	0	0.00299	0.8865
	0.65	6×10^{-5}	0.01672	0.9218	5×10^{-5}	0.01458	0.9195	0.00019	0.01158	0.9240
	1.26	0.00033	0.01816	0.9223	2×10^{-5}	0.01607	0.9289	0.00048	0.01201	0.9396
	1.87	0.0002	0.01637	0.9434	0.00046	0.01464	0.9471	0.00081	0.01198	0.9484
	2.48	0.0013	0.01644	0.9423	0.0002	0.01502	0.9488	0.00151	0.01246	0.9408



(a) $n_i = 200$

(b) $n_i = 400$

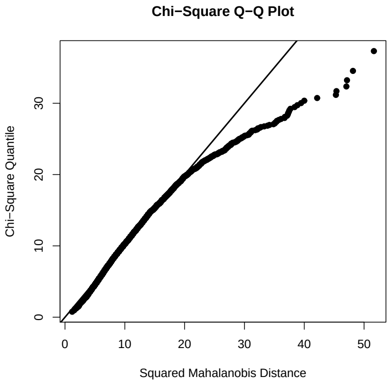
(c) $n_i = 600$

(d) $n_i = 700$

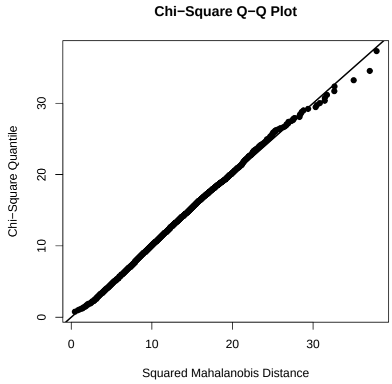
Figure 4: Multivariate QQ-plots of the Mahalanobis Squared Distance: Exponential(1.2) failure time distribution and three failure types with rate ratio 3 : 2 : 1

Table 5: Simulation results on $\left\{\hat{F}_j^{EM}(t); j = 1, 2\right\}$ for exponential(0.5) failure time distribution and two types of failures with rate ratio 3 : 2 and random monitoring time.

		Cause 1				Cause 2			
n	τ_i	Bias	ASE	SSE	CP	Bias	ASE	SSE	CP
	0.05	2×10^{-04}	0.04251	0.03462	0.8355	8×10^{-05}	0.04409	0.02839	0.8363
	0.79	0.00047	0.06390	0.06381	0.8019	0.00284	0.05935	0.05525	0.8056
250	1.53	0.00252	0.09417	0.06965	0.8207	0.00651	0.08671	0.05784	0.8206
	2.27	0.00046	0.05469	0.04240	0.8197	0.00188	0.05086	0.04042	0.8212
	3.00	0.00087	0.05429	0.04232	0.8174	0.00348	0.05678	0.04067	0.8182
	0.05	1×10^{-04}	0.01548	0.01505	0.8788	4×10^{-05}	0.01460	0.01241	0.9677
	0.79	0.00032	0.02948	0.02993	0.9198	.0003	0.02646	0.02597	0.9230
1250	1.53	0.00085	0.04424	0.03960	0.9476	0.00271	0.04089	0.03549	0.9458
	2.27	0.00011	0.02516	0.02150	0.9456	0.00083	0.02426	0.02004	0.9488
	3.00	0.00085	0.02493	0.02142	0.9402	0.00141	0.02450	0.02083	0.9438
	0.05	5×10^{-05}	0.01274	0.01266	0.8895	3×10^{-05}	0.01119	0.01046	
	0.79	3×10^{-04}	0.02492	0.02538	0.9292	9×10^{-05}	0.02212	0.02208	0.9342
1750	1.53	0.00074	0.03747	0.03540	0.9489	0.00178	0.03467	0.03145	0.9521
	2.27	3×10^{-05}	0.02138	0.01906	0.9530	0.00078	0.02054	0.01779	0.9554
	3.00	0.00074	0.02117	0.01852	0.9539	0.00116	0.02078	0.01811	0.9524
	0.05	5×10^{-05}	0.01190	0.01191	0.9098	0	0.01022	0.00980	
	0.79	.00027	0.02330	0.02330	0.9342	6×10^{-05}	0.02064	0.02067	0.9343
2000	1.53	0.00053	0.03504	0.03341	0.9457	0.00157	0.03235	0.03078	0.9515
	2.27	0	0.01989	0.01788	0.9522	0.00062	0.01911	0.01687	0.9543
	3.00	0.00066	0.01968	0.01763	0.9524	0.00105	0.01933	0.01716	0.9521

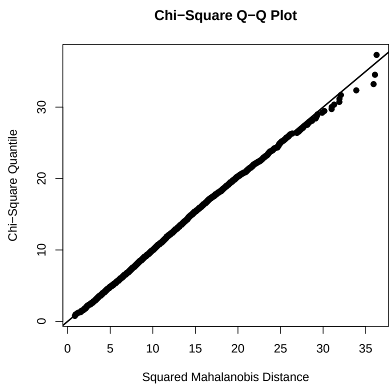


(a) $n = 250$



(b) $n = 1250$

(c) $n = 1750$



(d) $n = 2000$

Figure 5: Multivariate QQ-plots of the Mahalanobis Squared Distance: Exponential(0.5) failure time distribution and two failure types with rate ratio 3 : 2 and random monitoring time.

8 An Example

Let us consider the data on onset of menopause (See Table 6) from the National Center for Health Statistics' Examination survey, which was discussed by Krailo and Pike (1983) and analyzed originally by MacMahon and Worcestor (1966). This data consists of the menopausal history of 3581 female respondents providing information on their age and menopausal status. Here menopausal status refers to whether or not the respondents experienced menopause. For those with menopause, further observations are made on whether the menopause is natural or operative with $J = 1$ or 2, respectively. At any monitoring time τ_i, d_{1i} denotes the number of natural menopause, while d_{2i} the number of operative menopause. We analyze this dataset and obtain the non-parametric MLEs of the sub-distribution functions using the proposed method with EM algorithm and the constrained maximization method of Jewell and Kalbfleisch (2004) as well. Table 6 provides the MLEs of the corresponding sub-distribution functions $\hat{F}_j^{EM}$ and $\hat{F}_j^{JK}$, for $j = 1, 2$, denoting the two sets of estimates, obtained by using the EM algorithm and the method of Jewell and Kalbfleisch, respectively. The standard errors of the first set of estimators, namely, $\hat{F}_j^{EM}(.)$, for $j = 1, 2$, are also obtained using the derivation in Section 5, and are denoted by $SE(\hat{F}_j^{EM}(.))$, for $j = 1, 2$. We have plotted the estimated sub-distribution functions obtained from both the methods (See Figure 6). As expected, both the methods give the same estimates. Also, the standard errors tend to be larger towards the later time points.

Table 6: Data on menopausal history with the MLEs from the proposed method and the constrained maximization method with the corresponding standard errors for the proposed estimators

τ_i	n_i	d_{1i}	d_{2i}	$n_i - d_i$	$\hat{F}_1^{JK}(t)$	$\hat{F}_1^{EM}(t)$	$\hat{F}_2^{JK}(t)$	$\hat{F}_2^{EM}(t)$	SE($\hat{F}_1^{EM}(t)$)	SE($\hat{F}_2^{EM}(t)$)
27.5	380	4	0	376	0.01053	0.01053	0	0	0.00540	0.02032
32.5	359	21	0	338	0.05849	0.05849	0	0	0.01482	0.03053
35.5	89	7	0	82	0.06818	0.06818	0	0	0.01545	0.04207
36.5	87	5	0	82	0.06818	0.06818	0	0	0.10111	0.04207
37.5	61	5	1	55	0.08197	0.081967	0.01639	0.01639	0.10159	0.04436
38.5	83	11	2	70	0.11350	0.11350	0.01840	0.01840	0.09738	0.04333
39.5	99	11	2	86	0.11350	0.11350	0.01840	0.01840	0.09728	0.04244
40.5	78	8	1	69	0.11350	0.11350	0.01840	0.01840	0.09922	0.04325
41.5	66	7	1	58	0.11350	0.11350	0.01840	0.01840	0.10040	0.04610
42.5	80	16	4	60	0.16742	0.16742	0.05204	0.05204	0.09420	0.04814
43.5	74	11	5	58	0.16742	0.16742	0.05551	0.05551	0.09721	0.04737
44.5	67	10	3	54	0.16742	0.16742	0.05551	0.05551	0.09868	0.05120
45.5	99	20	12	67	0.20201	0.20202	0.12121	0.12121	0.09210	0.050999
46.5	76	16	11	49	0.21054	0.21053	0.14473	0.14474	0.09432	0.05596
47.5	75	18	16	41	0.23676	0.23677	0.21425	0.21424	0.09299	0.06151
48.5	80	19	18	43	0.23677	0.23677	0.22521	0.22521	0.09226	0.06108
49.5	66	20	19	27	0.23678	0.23678	0.31525	0.31525	0.08335	0.07008
50.5	72	18	32	22	0.23677	0.23677	0.45228	0.45228	0.08714	0.07336
51.5	66	10	38	18	0.23677	0.23677	0.51790	0.51790	0.08856	0.07488
52.5	54	16	30	8	0.28099	0.28099	0.56764	0.56764	0.09972	0.09250
53.5	67	18	40	9	0.28099	0.28099	0.58695	0.58695	0.09778	0.08764
54.5	50	18	28	4	0.31020	0.31020	0.60357	0.60357	0.08899	0.08638
55.5	45	19	25	1	0.31020	0.31020	0.66326	0.66326	0.08996	0.08843
56.5	50	13	36	1	0.31020	0.31020	0.67115	0.67115	0.08968	0.08747
57.5	54	13	40	1	0.31020	0.31020	0.67297	0.67297	0.08829	0.08525
58.5	46	13	33	0	0.31020	0.31020	0.68980	0.68980	0.18438	0.16692

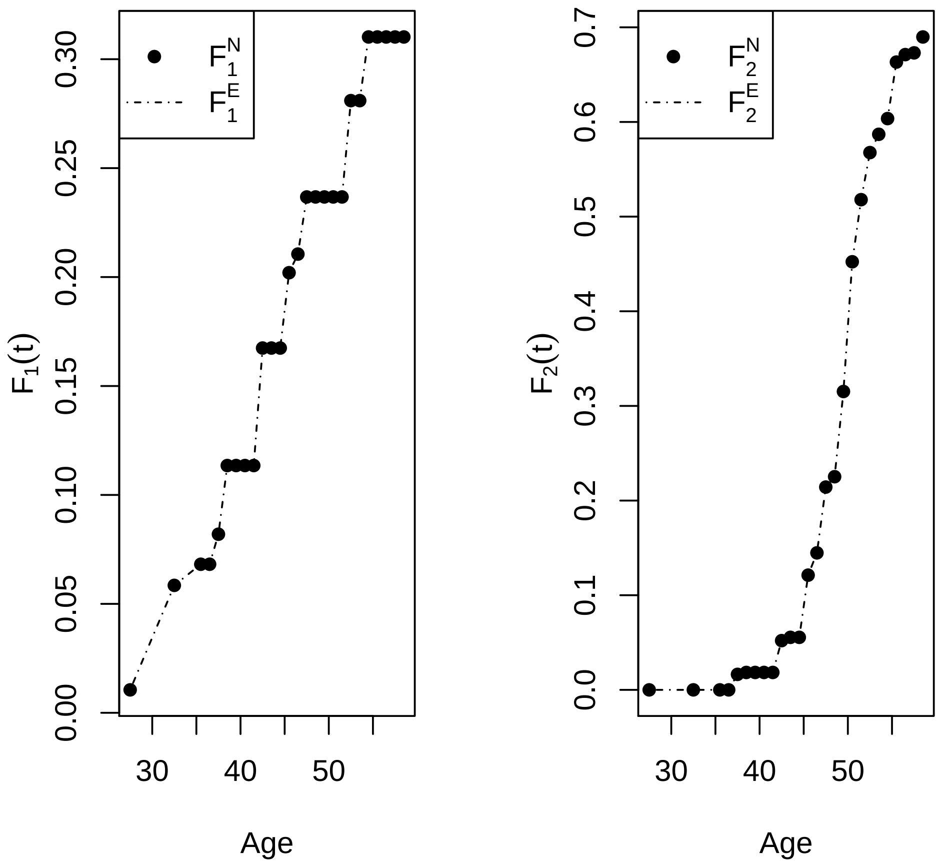


Figure 6: Plots of the MLEs of the subdistribution functions obtained from both the methods against age for menopausal history data

9 Concluding Remarks

For a particular cause j, the corresponding sub-distribution function $F_j(\cdot)$ satisfies certain properties (right continunous, non-decreasing, $F_j(\infty) \leq 1$). Also, the sum of sub-distribution functions over the causes is a distribution function satisfying its basic properties. Thus, estimation of sub-distribution functions directly is a constrained optimization problem. Pool adjacent violators algorithm (Both the usual one and the modified one) are used by many to estimate these quantities. Our aim is to propose an alternative method that will averts constrained maximization thereby providing simplicity and ease of programming. The proposed method is also flexible to be applicable in some other design scenario (See Section 6, for example). Though the use of EM algorithm had already been indicated in Section 8 of Jewell and Kalbfleisch (2004), we have tried to derive the necessary results by explicitly using the idea. The concept of re-parametrization of the model in terms of the interval hazards convert it into an unconstrained optimization problem thus making the estimation procedure simple and easy to deal with, including testing and interval estimation. Computation of standard error and derivation of the asymptotic results are carried out using the basic knowledge of MLE and their asymptotic properties.

Note that the sub-distribution functions $F_j(\cdot)$ is estimable only at the τ_i's, as indicated at the end of Section 3. In particular, one can estimate $F_j(\tau_k)$ as $\sum_{l=1}^{k} \hat{\lambda}_{jl} \prod_{l'=1}^{l-1}(1 - \hat{\lambda}_{l'})$, for $j = 1, .., m$. Therefore, $F(\tau_k) = \sum_{j=1}^{m} F_j(\tau_k)$ is also estimable along with $\bar{F}(\tau_k) = P(T > \tau_k)$. But, the sub-distribution functions beyond τ_k are naturally not estimable. Hence, one is also not able to estimate $P[J = j] = F_j(\infty)$, for $j = 1, 2, .., m$. However, when $d_k = n_k$, then $\hat{F}_j^{EM}(\tau_k)$ also estimates $F_j(\infty)$, as in case of Kaplan-Meier estimator.

In this work, we have considered fixed monitoring times, although the number of individuals observed at each monitoring time is allowed to be random. In some examples, different individuals are observed at different times so that the monitoring times for the individuals under study may be assumed to be independent realizations from a common distribution. In such cases, following Turnbull(1976), one can construct intervals in which the mass is concentrated. The existing works (See Hudgens et al., 2001) also exploit this fact and suggest EM algorithm to estimate these mass distributions. Clearly, the proposed reparametrization in terms of the interval cause-specific hazards make this much simpler and easier to program. More generally, our proposed reparametrization is useful for analysing arbitrarily censored failure time data with or without competing risks.

References

Dempster, A.P, Laird, N.M, Rubin, D.B.(1977).Maximum Likelihood from Incomplete Data via the EM Algorithm.*Journal of the Royal Statistical Society.Series B (Methodological)* 39, 1-38.

Hudgens, M.G., Satten, G.A. and Longini, I.M.(2001).Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation.*Biometrics* 57,74-80.

Jewell, N.P.,Van Der Laan, M. and Henneman, T.(2003).Nonparametric estimation from current status data with competing risks.*Biometrika* 90,183-197.

Jewell, N.P., Kalbfleisch, J.D.(2004).Maximum Likelihood estimation of ordered multinomial parameters.*Biostatistics* $5; 2, 291 - 306$.

Jewell, N.P., Van Der Laan and Lei, X.(2005).Bivariate current status data with univariate monitoring times.*Biometrika*,92,4,847-862.

Krailo, M.D. and Pike, M. C.(1983).Estimation of the distribution of age at natural menopause from prevalence data.*American Journal of Epidemiology* 117,356-361.

Kalbfleisch, J.D. and Prentice, R. L.(2002).The Statistical Analysis of Failure Time Data, 2nd edition.*New York:Wiley.*

Lehmann, E.L., Casella, G.(1998).Theory of Point Estimation,Second edition.*Springer.*

Louis T.A.(1982).Finding the Observed Information Matrix when using the EM Algorithm.*J.R.Statist.Soc.B.*44,No.2,226-233.

Maathuis, M.H.(2006). Nonparametric estimation for current status data with competing risks. Diss. University of Washington

Maathuis, M.H., Hudgens, M.G.(2011).Nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times.*Biometrika*,1-16.

Turnbull, B.W.(1976).The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data.*Journal of the Royal Statistical Society.Series B (Methodological)*38, 290-295

Ververidis, D. and Kotropoulos, C.(2008). Gaussian Mixture Modeling by Exploiting the Mahalanobis Distance.*IEEE Transactions On Signal Processing* 56,7

WU, C.F.J.(1983).On the convergence Properties of the EM Algorithm.*The Annals of Statistics* 11,95-103.