A Cure rate model for exponentially distributed lifetimes with competing risks

Ayan Pal [*], Shuvashree Mondal [†], Debasis Kundu [‡]

Abstract

In real life, more often experimental units are susceptible to more than one risk factors. Moreover, some experimental units may not fail even if they are observed over a long period of time. In statistical analysis, competing risks models handle the first scenario while cure rate models have been introduced to analyze the long term survivors in the population. In this paper we consider a cure rate model when the failure of an item can be due to either of the two competing causes. To analyze the competing risk data in presence of long term survivors, we consider the latent failure times approach introduced by Cox [6]. The latent failure times are assumed to follow exponential distributions and they are independently distributed. Under this set up a random censoring scheme is applied and the observed data consist of either censored times or actual failure times along with the cause of failures. We derive the maximum likelihood estimators(MLEs) using the expectation maximization (EM) algorithm based on the missing value principle. As the over all survival function is not a proper survival function, the asymptotic behavior of the MLEs is not immediate. We provide the sufficient conditions for the existence, uniqueness, consistency and the asymptotic normality of the MLEs. Monte Carlo simulations are performed to support the theoretical validation numerically. For illustrative purposes, we have analyzed one real data set, and the results are quite satisfactory.

Key Words and Phrases: Cure Rate Model, Long term survivors, Competing Risk, EM algorithm, Asymptotic normality, Consistency, Maximum likelihood estimator.

AMS Subject Classifications: Primary 62F12; secondary 62N05.

[*]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India.

[†]Department of Mathematics and Computing, Indian Institute of Technology (ISM) Dhanbad, Police Line, Sardar Patel Nagar, Hirapur, Dhanbad, Jharkhand Pin 826004, India.

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

1 Introduction

In reliability and survival analysis, it is quite common that an item is exposed to more than one risk factors. Such a scenario naturally motivates an investigator to assess a specific risk in presence of the other risk factors. In the statistical framework, this problem is well known as the competing risk problem. Typically, the data in a competing risks problem is a paired observation (t, δ) where t denotes the failure time and δ is an indicator which denotes the cause of failure. Several models exist in the literature to analyze the competing risks data. Interested readers are referred to the monograph by Crowder [5] for a general overview on the analysis of different competing risk models. The two most well-known approaches to model the competing risks data are: i) the latent failure times model suggested by Cox [6], and ii) the cause specific hazard functions model suggested by Prentice et al. [20]. The distributions of the lifetimes of the competing risks responsible for the failure of the units, are assumed to be independent in the latent failure times model, whereas the assumption of independence is dropped in the cause specific hazard functions model. Kundu [16] showed that the likelihood functions of the observed data based on the above two approaches are exactly the same in case of the exponential and the Weibull distributions. Hence, the estimation techniques and the statistical properties of the different unknown parameters remain unchanged, although the interpretations of the different parameters might be different.

Again, in reliability and survival analysis, the most common assumption is that all the individuals will subsequently fail if they are observed over a long period of time. In real scenario, there might be a substantial proportion of individuals who will not fail over a long period of time and often called long term survivors. To address this issue, cure rate models are considered which embody the long term survivals in the population. Here, the underlined failure distribution is conditioned only on the individuals who will fail within a considerable period of time and a parameter is incorporated to measure the proportion of long term survivors in the overall population. Boag [3] was the first to propose a two-component mixture model for analyzing breast cancer data. The model referred to as "cure-rate" model introduces the fraction of immunes in the population and a latency distribution representing the survival experience of the susceptibles. Other references related to the application of cure rate model with long term survivors in cancer data are Berkson and Gage [2], Goldman [12], Farewell [8], Greenhouse and Wolfe [13]. In criminology study Maltz and McCleary [18], Maltz [19] assumed exponential distribution to study the time to recidivism as event of interest and incorporated a parameter to measure the long term success of prisoners. Other interesting applications include commercial fishing by Gulland [14], unemployment studies by Dunsmuir et al. [7] etc .

In real life scenario, it is quite likely that an individual who will fail over a considerable amount of time, may be exposed to more than one competing causes of failure. Again, no prior information is available about the particular cause leading to the occurrence of the event of interest. To study these kind of scenarios, many authors have proposed flexible cure rate models under the competing risks set up. Some of the recent references include Balakrishnan and Pal [1], Wiangnak and Pal [22], and Gallardo et al.[11].

In this paper we assume a fixed number of causes and study the competing risks model based on the latent failure times approach by Cox [6] in the presence of the long term survivors. For simplicity we consider two causes but the developments are along the similar lines for more than two number of causes. We assume two causes of failures with independent and exponentially distributed latent failure times. Under this set up when a random censoring scheme is applied, the data consist of observations either due to censoring or failure due to any of the two competing risks. If any observation is due to a failure, the indicator variable denoting the cause of failure, is also recorded. Based on the observed data, the exact form of the maximum likelihood estimators (MLEs) cannot be obtained explicitly from the likelihood equations. In this case, MLEs can be obtained by solving a three dimensional optimization problem. One needs to use some iterative methods to compute the MLEs. Any iterative procedure needs a very good set of initial values for its convergence. Otherwise, it may not even converge or converges to a local optimum rather than the global optimum. To avoid that we rely on the expectation maximization (EM) algorithm based on the missing value principle to compute the MLEs. In the 'M' step of the EM algorithm, all the estimators can be obtained in explicit forms. Therefore, the proposed EM algorithm works very efficiently in this context and provides an extra edge over the usual likelihood maximization technique. We also provide the observed Fisher information matrix which can be used to compute the asymptotic confidence intervals.

In presence of long term survivors, the survival function of the overall population is not a proper survival function. Therefore, deriving the asymptotic properties of the MLEs of the model parameters is not immediate. Ghitany and Maller [9] were the first to prove the existence, uniqueness, consistency and the asymptotic normality of the MLEs under a mixed-exponential survival model with a cured fraction. Ghitany, Maller and Zhou [10] proved similar results in the presence of explanatory vectors containing covariate information on each individual. The basic assumption in both Ghitany and Maller[9], and Ghitany, Maller and Zhou [10] is that only a particular cause is responsible for the occurrence of the event of interest. In this set up, we provide the asymptotic properties of the MLEs under the competing risks scenario in presence of long term survivors. We provide the sufficient conditions for the existence, uniqueness, consistency and asymptotic normality of the MLEs of the model parameters. These results may be extended even for more than two competing causes of failure or when the latent failure times follow Weibull distributions with a common known shape parameter.

Our paper is organized as follows. In Section 2, we formulate the cure rate model assuming that the latent lifetime distributions are exponential and provide the likelihood function. In Section 3, the maximum likelihood estimation of the unknown model parameters is carried out using the EM algorithm and the confidence intervals of the unknown model parameters are also provided. The existence, uniqueness, consistency and the asymptotic normality of the underlying model parameters are provided in Section 4. In Section 5, we perform some simulation experiments and analyze one real data set for illustrative purposes. Finally, we conclude the paper in Section 6.

2 Model description and likelihood function

2.1 Model description

We assume that the population consists of two groups of individuals, the individuals who are subject to the risk of failure or the susceptible group and the individuals who will never fail or the immune group. Let Y_i be an indicator variable with:

$$Y_i = \begin{cases} 1, & \text{if the } i\text{-th individual belongs to the susceptible group} \\ 0, & \text{if the } i\text{-th individual belongs to the immune group.} \end{cases}$$

Then $P\{Y_i = 1\} = p_0 \in (0, 1)$ is the proportion of individuals belonging to the susceptible group and $P\{Y_i = 0\} = 1 - p_0$ is the proportion of individuals belonging to the immune group.

We assume that the failure of an individual is subject to two competing causes and we model the competing risks data using Cox's [6] latent failure times model. Let T^*_{1i} and T^*_{2i} be the failure times due to cause 1 and cause 2, respectively, for the i-th individual. Based on Cox's [6] latent failure times assumption, T^*_{1i} and T^*_{2i} are independently distributed and they are not observable. If the i-th individual fails, we only observe T^*_i as the failure time where $T_i^* = \min(T_{1i}^*, T_{2i}^*)$ and an indicator Δ_i where

$$\Delta_i = \begin{cases} 1, & \text{if the } i\text{-th failure occurs due to first cause} \\ 2, & \text{if the } i\text{-th failure occurs due to second cause.} \end{cases}$$

We also assume $T_{1i}^* \sim \text{Exp}(\lambda_{10})$ with mean $\frac{1}{\lambda_{10}}$, $\lambda_{10} > 0$, $T_{2i}^* \sim \text{Exp}(\lambda_{20})$ with mean $\frac{1}{\lambda_{20}}$, $\lambda_{20} > 0$.

Conditioning on $Y_i = 1$, the failure time $T_i^* = \min(T_{1i}^*, T_{2i}^*)$, has the density

$$g(t_i^*; \lambda_{10}, \lambda_{20} | Y_i = 1) = (\lambda_{10} + \lambda_{20}) e^{-(\lambda_{10}+\lambda_{20})t_i^*}.$$

We denote $\boldsymbol{\theta} = (\lambda_1, \lambda_2, p)$, the set of model parameters to be estimated. The corresponding parameter space is $\Theta = (0, \infty) \times (0.\infty) \times (0, 1]$ and $\boldsymbol{\theta_0} = (\lambda_{10}, \lambda_{20}, p_0)$, the true value, is an interior point of Θ. We apply a random censoring scheme where C_i denotes the censored time for the i-th individual. The distribution of the censoring time is assumed to be non-informative, i.e it does not involve the parameter of interest $\boldsymbol{\theta_0} = (\lambda_{10}, \lambda_{20}, p_0)$. Furthermore, the sequences T_i^* and C_i, C_i and Y_i are assumed to be independent. Let us introduce another indicator variable Z_i as follows

$$Z_i = \begin{cases} 1, & \text{if } T_i^* \leq C_i \\ 0, & \text{otherwise.} \end{cases}$$

Here $T_i = \min(T_i^*, C_i)$ denotes the time to occurrence of the event of interest where the event is either failure or being censored. When $Z_i = 1$, individual must belong to the susceptible group and we observe the real failure time, along with the corresponding cause of failure δ_i and for that individual, data consist of $((t_i, \delta_i), 1)$. Here, δ_i is either 1 or 2, depending on the cause of failure. When $Z_i = 0$, we observe only the censored time and in this case the group identity of the individual is unknown and for that individual, data consist of $(t_i, 0)$.

2.2 Likelihood function

Suppose we have a sample of size n and for the i-th individual we observe $t_i = \min(t_i^*, c_i)$ and z_i. If the observed value of $z_i = 1$, we record the indicator δ_i denoting the cause of failure. An individual who fails at the time t_i contributes a likelihood factor

$$p(\lambda_1 e^{-(\lambda_1+\lambda_2)t_i})^{2-\delta_i}(\lambda_2 e^{-(\lambda_1+\lambda_2)t_i})^{\delta_i-1}.$$

An individual who survives at least till the time t_i contributes a likelihood factor

$$1-p+pe^{-(\lambda_1+\lambda_2)t_i},$$

which is the probability of being a long term survivor plus the probability of a failure occurring after time t_i. Thus, the total contribution of the i-th individual to the likelihood function is

$$\{p(\lambda_1 e^{-(\lambda_1+\lambda_2)t_i})^{2-\delta_i}(\lambda_2 e^{-(\lambda_1+\lambda_2)t_i})^{\delta_i-1}\}^{z_i}\{1-p+pe^{-(\lambda_1+\lambda_2)t_i}\}^{1-z_i}.$$

For a given data, thus the log-likelihood function is as follows:

$$\begin{aligned} \mathbf{l}_n(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\{z_i(\ln p + (\delta_i-1)\ln\lambda_2 + (2-\delta_i)\ln\lambda_1 - (\lambda_1+\lambda_2)t_i) + \\ &\quad (1-z_i)\ln(1-p+pe^{-(\lambda_1+\lambda_2)t_i})\}. \end{aligned} \tag{1}$$

Let us denote,

$$\mathbf{S}_n(\boldsymbol{\theta}) = \frac{\partial l_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}. \tag{2}$$

It can be observed that $\mathbf{S}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n}\mathbf{s}_i(\boldsymbol{\theta})$ where $\mathbf{s}_i(\boldsymbol{\theta}) = (s_{i1}(\boldsymbol{\theta}),\ s_{i2}(\boldsymbol{\theta}),\ s_{i3}(\boldsymbol{\theta}))^\top$ is a sequence of 3×1 random vectors with components :

$$s_{i1}(\boldsymbol{\theta}) = \frac{z_i(2-\delta_i)}{\lambda_1} - z_it_i - \frac{(1-z_i)pt_ie^{-(\lambda_1+\lambda_2)t_i}}{1-p+pe^{-(\lambda_1+\lambda_2)t_i}} \tag{3}$$

$$s_{i2}(\boldsymbol{\theta}) = \frac{z_i(\delta_i-1)}{\lambda_2} - z_it_i - \frac{(1-z_i)pt_ie^{-(\lambda_1+\lambda_2)t_i}}{1-p+pe^{-(\lambda_1+\lambda_2)t_i}} \tag{4}$$

$$s_{i3}(\boldsymbol{\theta}) = \frac{z_i}{p} - \frac{(1-z_i)(1-e^{-(\lambda_1+\lambda_2)t_i})}{1-p+pe^{-(\lambda_1+\lambda_2)t_i}}. \tag{5}$$

The MLE of $\boldsymbol{\theta}$ can be obtained by solving $\mathbf{S}_n(\boldsymbol{\theta}) = \mathbf{0}$ and it requires to solve a three dimensional optimization problem. It is not possible to obtain the solutions of the normal equations explicitly and we cannot obtain any closed form solutions for the MLEs of any of the model parameters.

3 Point and interval estimation

3.1 EM Algorithm

Based on the log likelihood function (1), we need to solve three non linear equations to compute the MLEs. The MLEs cannot be obtained in closed form. However, we can treat this problem as a missing data problem. We define two sets of indices as $I_1 = \{i : z_i = 1\}$ and $I_0 = \{i : z_i = 0\}$. If the subject's lifetime is actually observed, the indicator variable y_i takes the value 1. Otherwise, there can be two possibilities, the subject might belong to the cure group or it might belong to the susceptible group but the time to failure is greater than the censoring time. Thus for $i \in I_0$, y_i is unknown and these $y_i's$ are treated as missing. Incorporating the unobserved $y_i's$ with the observed data, the complete data set is constructed. The likelihood function based on the complete data is obtained as

$$L_c(\boldsymbol{\theta}) = \prod_{I_1} \left\{ p{\lambda_1}^{2-\delta_i}{\lambda_2}^{\delta_i - 1} e^{-(\lambda_1+\lambda_2)t_i} \right\}^{y_i} \prod_{I_0} (1-p)^{1-y_i} \prod_{I_0} \left\{ p e^{-(\lambda_1+\lambda_2)t_i} \right\}^{y_i}. \tag{6}$$

The corresponding log-likelihood function is given by

$$l_c(\boldsymbol{\theta}) = \sum_{I_1} y_i \ln \left\{ p{\lambda_1}^{2-\delta_i}{\lambda_2}^{\delta_i - 1} e^{-(\lambda_1+\lambda_2)t_i} \right\} + \sum_{I_0} (1-y_i)\ln(1-p) + \sum_{I_0} y_i \ln \left\{ p e^{-(\lambda_1+\lambda_2)t_i} \right\}. \tag{7}$$

Maximizing (7), we can obtain the MLEs of the model parameters in closed form as

$$\hat{p} = \frac{|I_1| + \sum\limits_{i \in I_0} y_i}{n}, \quad \hat{\lambda}_1 = \frac{\sum\limits_{i \in I_1} y_i(2-\delta_i)}{\sum\limits_{i \in I_0} y_i t_i + \sum\limits_{i \in I_1} y_i t_i}, \quad \hat{\lambda}_2 = \frac{\sum\limits_{i \in I_1} y_i(\delta_i - 1)}{\sum\limits_{i \in I_0} y_i t_i + \sum\limits_{i \in I_1} y_i t_i},$$

where $|I_1|$ denotes the cardinality of the set I_1. Hence in case of the complete data, the MLEs can be obtained explicitly. This is the main motivation of applying the proposed EM algorithm. We propose to use the following EM algorithm similar to Kannan et al. [15].

E-Step : In the E-Step, we compute the expectation of the complete log likelihood function with respect to the distribution of the unobserved $y_i's$, given the current values of the parameter and the observed data $\mathcal{O} = \{(t_i, z_i); i = 1, 2 \ldots, n\}$. It is clear that $Y_i; s$ are Bernoulli random variables in the complete log-likelihood and it is only required to compute $\pi_i^{(k)} = E(Y_i|\boldsymbol{\theta^{(k)}}, \mathcal{O}), i = 1, 2, \ldots, n$, where $\boldsymbol{\theta^{(k)}}$ denotes the parameter estimate at the k-th iteration step. Now for $i \in I_0$, we have

$$\begin{aligned}
\pi_i^{(k)} &= E(Y_i|\boldsymbol{\theta^{(k)}}, \mathcal{O}) \\
&= P(Y_i = 1|T_i > t_i; \boldsymbol{\theta^{(k)}}) \\
&= \left.\frac{P(T_i > t_i|Y_i = 1)P(Y_i = 1)}{P(T_i > t_i)}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta^{(k)}}} \\
&= \frac{p^{(k)} e^{-(\lambda_1^{(k)}+\lambda_2^{(k)})t_i}}{1 - p^{(k)} + p^{(k)} e^{-(\lambda_1^{(k)}+\lambda_2^{(k)})t_i}},
\end{aligned}$$

and for $i \in I_1$, we simply have $\pi_i^{(k)} = y_i = 1$. Thus, the E-Step replaces the $y_i's$ in the log likelihood by $\pi_i^{(k)}$ for $i \in I_0$ and by 1 for $i \in I_1$. Let us denote the conditional expectation of the complete log-likelihood function by $Q(\boldsymbol{\theta}, \boldsymbol{\pi^{(k)}})$, where $\boldsymbol{\pi^{(k)}}$ is the vector of $\pi_i^{(k)}$ values. Thus,

$$Q(\boldsymbol{\theta}, \boldsymbol{\pi^{(k)}}) = \sum_{I_1} \ln\left\{p{\lambda_1}^{2-\delta_i}{\lambda_2}^{\delta_i-1} e^{-(\lambda_1+\lambda_2)t_i}\right\} + \sum_{I_0}(1-\pi_i^{(k)})\ln(1-p) + \sum_{I_0}\pi_i^{(k)}\left\{pe^{-(\lambda_1+\lambda_2)t_i}\right\}.$$

M-Step : In the M-Step of the EM algorithm, we maximize $Q(\boldsymbol{\theta}, \boldsymbol{\pi^{(k)}})$ with respect to $\boldsymbol{\theta}$, given $\boldsymbol{\pi^{(k)}}$ over the parameter space Θ, to obtain an improved estimate of $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta^{(k+1)}} = \arg\max_{\theta\in\Theta} Q(\boldsymbol{\theta}, \boldsymbol{\pi^{(k)}})$$

The normal equations obtained after taking the first order derivative of $Q(\boldsymbol{\theta}, \boldsymbol{\pi^{(k)}})$ with respect to the model parameters are as follows:

$$\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\pi^{(k)}})}{\partial p} = \frac{|I_1|}{p} - \sum_{i\in I_0}\frac{1-\pi_i^{(k)}}{1-p} + \sum_{i\in I_0}\frac{\pi_i^{(k)}}{p} = 0, \tag{8}$$

$$\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\pi^{(k)}})}{\partial \lambda_1} = \sum_{i\in I_1}\left\{\frac{2-\delta_i}{\lambda_1} - t_i\right\} - \sum_{i\in I_0} t_i\pi_i^{(k)} = 0, \tag{9}$$

$$\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\pi^{(k)}})}{\partial \lambda_2} = \sum_{i \in I_1} \left\{ \frac{\delta_i - 1}{\lambda_2} - t_i \right\} - \sum_{i \in I_0} t_i \pi_i^{(k)} = 0. \tag{10}$$

To obtain the unique MLEs, we solve the above normal equations and obtain the closed form solutions as

$$p^{(k+1)} = \frac{|I_1| + \sum_{i \in I_0} \pi_i^{(k)}}{|I_1| + \sum_{i \in I_0} \pi_i^{(k)} + \sum_{i \in I_0} (1 - \pi_i^{(k)})} = \frac{|I_1| + \sum_{i \in I_0} \pi_i^{(k)}}{n}$$

$$\lambda_1^{(k+1)} = \frac{\sum_{i \in I_1} (2 - \delta_i)}{\sum_{i \in I_0} t_i \pi_i^{(k)} + \sum_{i \in I_1} t_i}, \quad \lambda_2^{(k+1)} = \frac{\sum_{i \in I_1} (\delta_i - 1)}{\sum_{i \in I_0} t_i \pi_i^{(k)} + \sum_{i \in I_1} t_i}.$$

3.2 Confidence intervals

Here, we provide the asymptotic confidence intervals of the unknown model parameters $\boldsymbol{\theta} = (\lambda_1, \lambda_2, p)$ using the observed Fisher information matrix $\mathbf{F}_n(\boldsymbol{\theta}) = -\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$. This method is useful for its computational ease and provides good coverage probabilities (close to nominal values). The explicit expressions for the elements of $\mathbf{F}_n(\boldsymbol{\theta})$ are provided in Appendix. The $100(1-\gamma)\%$ asymptotic confidence intervals for λ_1, λ_2, p are respectively

$$(\hat{\lambda}_1 \pm z_{1-\frac{\gamma}{2}} \sqrt{V_{11}}), \ (\hat{\lambda}_2 \pm z_{1-\frac{\gamma}{2}} \sqrt{V_{22}}), \ (\hat{p} \pm z_{1-\frac{\gamma}{2}} \sqrt{V_{33}}),$$

where V_{ij} is the $(i, j)-$ th element of the inverse of the observed Fisher information matrix, $\hat{\boldsymbol{\theta}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{p})$ is the MLE of $\boldsymbol{\theta}$, and $z_{1-\frac{\gamma}{2}}$ is the upper $(1 - \frac{\gamma}{2})-$ th percentile point of the standard normal distribution.

4 Main results

In this section, we provide the asymptotic properties of the MLEs of the model parameters under a random censoring scheme. Type-I censoring scheme is a particular case of such a random censoring scheme. We provide sufficient conditions for the existence, uniqueness along with the consistency and asymptotic normality of the MLEs of the unknown model parameters.

The following results state the asymptotic properties of the MLEs of the model parameters when a simple distributional assumption is imposed on the censored times.

Theorem 4.1 (Existence and uniqueness)
If the censored times C_i's for $i = 1, \ldots, n$, are iid (independent and identically distributed) random variables whose common distribution does not degenerate at the origin, then, for $\lambda_{10} > 0$, $\lambda_{20} > 0$ and $p_0 \in (0, 1)$, the MLE of $\boldsymbol{\theta_0}$, say, $\widehat{\boldsymbol{\theta_n}}$, exists and is unique.

Proof: See in Appendix. ∎

Theorem 4.2 (Consistency) *Under the same conditions on C_i's for $i = 1, \ldots, n$, as in Theorem (4.1), $\widehat{\boldsymbol{\theta_n}}$ is a consistent estimator of $\boldsymbol{\theta}$.*

Proof: See in Appendix. ∎

Theorem 4.3 (Asymptotic normality) *Under the same conditions on C_i's for $i = 1, \ldots, n$, as in Theorem (4.1), $\mathbf{D}_n^{1/2}(\boldsymbol{\theta}_o)(\widehat{\boldsymbol{\theta_n}} - \boldsymbol{\theta}_o)$ is asymptotically normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}$. Here $\mathbf{I}$ is the 3 × 3 identity matrix and $\mathbf{D}_n(\boldsymbol{\theta}_o) = E(\mathbf{S}_n(\boldsymbol{\theta}_o)\mathbf{S}_n(\boldsymbol{\theta}_o)^\top)$.*

Proof: See in Appendix. ∎

5 Simulation experiment and data analysis

5.1 Simulation experiment

In this section we perform simulation experiment to assess how the EM algorithm works and the MLEs behave for different sample sizes and for different parameter values. We consider $n = 20, 25, 30, 40, 50$ and $p = 0.25, 0.5, 0.75$, $\lambda_1 = 0.5$ and $\lambda_2 = 1$. Here we assume that the censored time for each of the n individuals follows independently a uniform distribution with lower limit 0 and upper limit 25. For each combination of n and the parameter set, we compute average estimates (AEs) and the mean squared errors (MSEs) of the MLEs based on 1000 replications. Based on the observed information matrix $\mathbf{F}_n(\boldsymbol{\theta})$ we construct the 90% asymptotic confidence intervals of the unknown parameters. The average length (AL) and the coverage percentages (CP) of these confidence intervals are computed based on 1000 replications. All these results are reported in Tables 1, 2 and 3.

It is clear that the average biases and the MSEs of the estimators decrease as the sample size increases which clearly indicates the consistency property of the estimators. We notice that as the true value of p increases, the number of observed failure increases and the MLEs of λ_1 and λ_2 provide better estimates in terms of biases and MSEs. Hence it is clear that the EM algorithm is performing quite well.

In interval estimation it is observed that the average length of the confidence intervals become shorter as the sample size increases. Also the average lengths of the confidence intervals of λ_1 and λ_2 decrease as p increases. Moreover, in all the cases the coverage probabilities are very close to the nominal level.

Table 1: $\lambda_1 = 0.5,\ \lambda_2 = 1,\ \mathsf{p} = 0.25$

n		AE	MSE	AL	CP
20	p	0.2675	0.0081	0.3223	0.922
	λ_1	0.7782	0.8316	1.7060	0.949
	λ_2	1.1546	0.8325	2.1790	0.899
25	p	0.2632	0.0069	0.2874	0.905
	λ_1	0.7080	0.6787	1.4510	0.923
	λ_2	1.1775	0.6317	1.9916	0.883
30	p	0.2529	0.0058	0.2599	0.873
	λ_1	0.6387	0.2239	1.2549	0.931
	λ_2	1.1309	0.3796	1.7769	0.903
40	p	0.2517	0.0045	0.2257	0.880
	λ_1	0.5844	0.1565	1.0437	0.894
	λ_2	1.1347	0.3228	1.5353	0.903
50	p	0.2539	0.0040	0.2029	0.886
	λ_1	0.5476	0.1093	0.8855	0.892
	λ_2	1.0758	0.1973	1.2907	0.898

Table 2: $\lambda_1 = 0.5,\ \lambda_2 = 1,\ \mathsf{p} = 0.5$

n		AE	MSE	A.L	C.P
20	p	0.4984	0.0125	0.3658	0.886
	λ_1	0.5866	0.1492	1.0321	0.897
	λ_2	1.1100	0.2674	1.4972	0.914
25	p	0.5013	0.0109	0.3284	0.864
	λ_1	0.5520	0.0865	0.8920	0.886
	λ_2	1.0874	0.1892	1.2976	0.913
30	p	0.5007	0.0082	0.3012	0.896
	λ_1	0.5306	0.0628	0.7884	0.876
	λ_2	1.0634	0.1333	1.1498	0.920
40	p	0.5034	0.0068	0.2617	0.884
	λ_1	0.5290	0.0496	0.6840	0.878
	λ_2	1.0576	0.0967	0.9871	0.906
50	p	0.5013	0.0049	0.2349	0.901
	λ_1	0.5096	0.0339	0.5969	0.888
	λ_2	1.0502	0.0787	0.8721	0.898

Table 3: $\lambda_1 = 0.5,\ \lambda_2 = 1,\ \mathsf{p} = 0.75$

n		AE	MSE	A.L	C.P
20	p	0.7552	0.0095	0.3152	0.861
	λ_1	0.5383	0.0667	0.7961	0.887
	λ_2	1.0896	0.1575	1.1647	0.898
25	p	0.7492	0.0081	0.2852	0.858
	λ_1	0.5274	0.0488	0.7039	0.878
	λ_2	1.0652	0.1145	1.0218	0.909
30	p	0.7506	0.0067	0.2606	0.876
	λ_1	0.5192	0.0384	0.6327	0.891
	λ_2	1.0461	0.0974	0.9159	0.890
40	p	0.7484	0.0051	0.2275	0.880
	λ_1	0.5255	0.0293	0.5532	0.902
	λ_2	1.0320	0.0643	0.7858	0.890
50	p	0.7504	0.0043	0.2036	0.871
	λ_1	0.5204	0.0233	0.4912	0.904
	λ_2	1.0290	0.0492	0.6995	0.894

5.2 Data Analysis: Melanoma data

In this subsection, we analyze one real data set for illustrative purposes. This data set is about the classic Melanoma dataset included in the *riskRegression* package of R. The data consist of measurements made on 205 patients with malignant melanoma (cancer of the skin). Each patient had their tumor removed by radical operation at the Department of Plastic Surgery, Odense University Hospital, Denmark during the period 1962 - 1977. All patients were followed up until the end of 1977 by which time 134 patients were still alive (censored) while 71 had died (of which 57 patients had died from cancer and 14 patients from other causes). Here $\delta = 1$ if failure is subject to malignant melanoma and $\delta = 2$ if the failure is subject to other causes. The time is recorded as the number of days from surgery until either the occurrence of the event of interest (death) or the last time the patient was known to be alive.

Before analyzing, we divide the data points by 1000 and then exponentiate by 1.4153.

From now onwards we work on the modified data only.

Assuming the failure times to be the minimum failure time due to the two causes, an exponential distribution is fitted on the observed failure times. As a measure of goodness of fit, we have computed the KS distance and the associated p-value. The KS distance between the empirical distribution function and the fitted distribution function is 0.0552 and the p value is 0.9786. Thus, it is reasonable to assume that the failure time due to two different causes are exponentially distributed.

The MLEs of the model parameters are computed based on the EM algorithm and hence initial guesses of each of the model parameters is required. The MLE of λ_1 is computed based on the observed failures only due to the first cause which is 0.0868 and it is used as initial value of λ_1. Similarly, we compute MLE of λ_2 based on observed failure times due to the second cause and the MLE is obtained as 0.0213. It is used an initial guess for λ_2. The initial choice of p is taken as $\frac{71}{205} = 0.3463$. We start the EM algorithm with these initial guesses and continue until the absolute difference of the estimates in two consecutive iterations is less than 10^{-4}. The EM algorithm stops after 29 iterations.

In Table 5, we record the MLEs derived through EM algorithm and the 90% asymptotic confidence intervals.

Table 4: MLEs and CIs for Melanoma data

Parameter	MLE	90% Asymptotic CI
λ_1	0.3014	(0.2125, 0.3902)
λ_2	0.0740	(0.0430, 0.1050)
p	0.4670	(0.3908, 0.5432)

Testing of the presence of cure fraction :

One may be interested to test whether the underlying population consists of cure fraction. Thus, the hypothesis testing problem turns out to be

$$H_o : p = 1 \quad \textbf{vs} \quad H_1 : p < 1.$$

Since the range of p contains the boundary point 1, applying Theorem 3 of [21], it follows that $-2(l_0 - l_1) \sim \frac{1}{2} + \frac{1}{2}\chi^2_1$. Here, l_0 and l_1 denotes the maximized log-likelihood (MLL) values under H_o and H_1 respectively. Under the null hypothesis H_o, the MLEs of the model parameters are $\widehat{\lambda}_1 = 0.0868$, $\widehat{\lambda}_2 = 0.0213$, and the corresponding MLL (l_0) is -264.1752. The MLEs of the model parameters under the alternate hypothesis H_1, are provided in Table 5 and the corresponding MLL (l_1) is -257.3717. The test statistic value is obtained as 13.6070 and the associated p-value comes out to be less than 0.00001. The p value leads to the rejection of the null hypothesis. It means that cure fraction is present in the underlying population.

6 Conclusion

In this article we have proposed a cure rate model in presence of competing risks. The lifetime of the susceptible unit is subject to two competing causes and we assume latent failure times approach by Cox [6]. The latent failure times are assumed to follow exponential distributions independently. Under this set up when a random censoring scheme is applied, we provide the maximum likelihood estimators by EM algorithm. Through simulation study, it is observed that the EM algorithm works very efficiently. We have studied the sufficient condition for the existence, uniqueness, consistency and asymptotic normality of the maximum likelihood estimators. When the censored times are independent and identical random variables, this sufficient condition boils down to non-degenerate at origin. Also these results are valid for more than two competing causes and latent failure times follow Weibull distributions with a known common shape parameter.

Acknowledgements:

The authors would like to thank the unknown reviewers for their constructive comments which have helped to improve the manuscript significantly.

Appendix

In the Appendix, we provide the proofs of Theorems 4.1, 4.2 and 4.3. To prove these Theorems, we need a series of lemmas and some preliminaries. We provide a flowchart to show how these Lemmas have been used to prove the Theorems.

Result	Prerequisite Results
Fisher information matrix	Lemma 6.1 and Lemma 6.2.
Lemma 6.3	Fisher information matrix.
Lemma 6.6	Lemma 6.4, Lemma 6.5.
Lemma 6.7	Lemma 6.6.
Lemma 6.8	Lemma 6.7.
Lemma 6.10	Lemma 6.6, Lemma 6.9.
Theorem 4.1	Lemma 6.3, Lemma 6.7.
Theorem 4.2	Lemma 6.3, Lemma 6.8.
Theorem 4.3	Lemma 6.3, Lemma 6.10.

The following notations have been used throughout this section.

Let $\{X_n\}$ be a sequence of random variables.

- $X_n = o_p(1)$ implies $\lim\limits_{n\to\infty} P(|X_n| > \epsilon) = 0 \ \forall \epsilon > 0$.
- $X_n = O_p(1)$ implies, for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $N > 0$ such that, $P(|X_n| > M) < \epsilon, \ \forall n > N$.
- $\lambda_{min}\{\mathbf{A}\}$ is the minimum eigen value of the matrix $\mathbf{A}$.

- The observed Fisher information matrix is

$$\mathbf{F}_n(\boldsymbol{\theta}) = -\frac{\partial \mathbf{S}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{11}$$

where $\mathbf{S}_n(\boldsymbol{\theta})$ is defined in (2).

- The expected Fisher information matrix about $\boldsymbol{\theta_0}$ is

$$\mathbf{D}_n(\boldsymbol{\theta_0}) = E[\mathbf{F}_n(\boldsymbol{\theta_0})] = E[\mathbf{S}_n(\boldsymbol{\theta_0})\mathbf{S}_n(\boldsymbol{\theta_0})^\top] = Cov(\mathbf{S}_n(\boldsymbol{\theta_0})). \tag{12}$$

- We define the region $\mathbf{N}_n(\boldsymbol{\theta_0})$ for fixed $A \geq 1$, as

$$\mathbf{N}_n(\boldsymbol{\theta_0}) = \{\boldsymbol{\theta} : \boldsymbol{\theta} \in \Theta,\ (\boldsymbol{\theta} - \boldsymbol{\theta_0})^\top \mathbf{D}_n(\boldsymbol{\theta}_o)(\boldsymbol{\theta} - \boldsymbol{\theta_0}) \leq A^2\}. \tag{13}$$

Elements of Observed Information Matrix:

Note that, the observed information matrix as defined in (11) can be written as,

$$\mathbf{F}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathscr{F}_i(\boldsymbol{\theta})$$

where

$$\mathscr{F}_i(\boldsymbol{\theta}) = \begin{bmatrix} f_i^{11}(\boldsymbol{\theta}) & f_i^{12}(\boldsymbol{\theta}) & f_i^{13}(\boldsymbol{\theta}) \\ f_i^{21}(\boldsymbol{\theta}) & f_i^{22}(\boldsymbol{\theta}) & f_i^{23}(\boldsymbol{\theta}) \\ f_i^{31}(\boldsymbol{\theta}) & f_i^{32}(\boldsymbol{\theta}) & f_i^{33}(\boldsymbol{\theta}) \end{bmatrix}. \tag{14}$$

Here,

$$f_i^{11}(\boldsymbol{\theta}) = \frac{z_i(2-\delta_i)}{\lambda_1^2} - \frac{(1-z_i)pt_i^2(1-p)e^{-(\lambda_1+\lambda_2)t_i}}{(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^2} \tag{15}$$

$$f_i^{22}(\boldsymbol{\theta}) = \frac{z_i(\delta_i-1)}{\lambda_2^2} - \frac{(1-z_i)pt_i^2(1-p)e^{-(\lambda_1+\lambda_2)t_i}}{(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^2} \tag{16}$$

$$f_i^{33}(\boldsymbol{\theta}) = \frac{z_i}{p^2} + \frac{(1-z_i)(1-e^{-(\lambda_1+\lambda_2)t_i})^2}{(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^2} \tag{17}$$

$$f_i^{12}(\boldsymbol{\theta}) = f_i^{12}(\boldsymbol{\theta}) = -\frac{(1-z_i)pt_i^2(1-p)e^{-(\lambda_1+\lambda_2)t_i}}{(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^2} \tag{18}$$

$$f_i^{13}(\theta) = f_i^{31}(\boldsymbol{\theta}) = f_i^{23}(\boldsymbol{\theta}) = f_i^{32}(\boldsymbol{\theta}) = \frac{(1-z_i)t_ie^{-(\lambda_1+\lambda_2)t_i}}{(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^2} \tag{19}$$

Fisher Information Matrix:

Note that, we can write, the Fisher information matrix, as defined in (12), as,

$$\mathbf{D}_n(\boldsymbol{\theta}_o) = \sum_{i=1}^{n} \mathscr{D}_i(\boldsymbol{\theta}_o), \tag{20}$$

where, $\mathscr{D}_i(\boldsymbol{\theta}_o) = \begin{bmatrix} d_i^{11}(\boldsymbol{\theta}_o) & d_i^{12}(\boldsymbol{\theta}_o) & d_i^{13}(\boldsymbol{\theta}_o) \\ d_i^{21}(\boldsymbol{\theta}_o) & d_i^{22}(\boldsymbol{\theta}_o) & d_i^{23}(\boldsymbol{\theta}_o) \\ d_i^{31}(\boldsymbol{\theta}_o) & d_i^{32}(\boldsymbol{\theta}_o) & d_i^{33}(\boldsymbol{\theta}_o) \end{bmatrix}$ and $d_i^{kl}(\boldsymbol{\theta}_o) = E[f_i^{kl}(\boldsymbol{\theta}_\mathbf{0})],\ k,l = 1,2,3.$

The expressions for the elements of $\mathscr{D}_i(\boldsymbol{\theta}_\mathbf{0})$ can be obtained using Lemma 6.1.

Lemma 6.1 *Let R(.) be a non-negative measurable function on* $\big[0,\infty\big)$. *Then, for all* $1 \le i \le n,\ \lambda_{10} > 0,\ \lambda_{20} > 0$ *and* $p_0 \in (0,1\big]$,

$$E\big[(1-Z_i)R(T_i)\big] = E\big[\{1 - p_0 + p_0 e^{-(\lambda_{10}+\lambda_{20})C_i}\}R(C_i)\big]$$

provided the expectation on the right exists.

Proof: For any $u > 0$,

$$\begin{aligned} P\{(1-Z_i)R(T_i) > u\} &= p_o P\{(1-Z_i)R(T_i) > u, Z_i = 0 | Y_i = 1\} \\ &+ (1-p_o) P\{(1-Z_i)R(T_i) > u, Z_i = 0 | Y_i = 0\}. \end{aligned}$$

Since Y_i is independent of C_i, we have

$$\begin{aligned} P\{(1-Z_i)R(T_i) > u\} &= p_o P\{R(C_i) > u, T_i^* > C_i | Y_i = 1\} + (1-p_o) P\{R(C_i) > u\} \\ &= \int_{c:R(c)>u} \Big[p_o P\{T_i^* > c | Y_i = 1\} + 1 - p_o\Big] dP\{C_i \le c\} \\ &= \int_{c:R(c)>u} (1 - p_o + p_o e^{-(\lambda_{10}+\lambda_{20})c})\, dP\{C_i \le c\}. \end{aligned}$$

Hence,

$$\begin{aligned} E\{(1-Z_i)R(T_i)\} &= \int_0^\infty P\{(1-Z_i)R(T_i) > u\} du \\ &= \int_0^\infty \int_{c:R(c)>u} (1 - p_o + p_o e^{-(\lambda_{10}+\lambda_{20})c})\, dP\{C_i \le c\}\, du. \end{aligned}$$

By Fubini's theorem, we have

$$\begin{aligned} E\{(1-Z_i)R(T_i)\} &= \int_0^\infty \int_{c:R(c)>u} du(1-p_o+p_o e^{-(\lambda_{10}+\lambda_{20})c})\, dP\{C_i \le c\} \\ &= E\{(1-p_o+p_o e^{-(\lambda_{10}+\lambda_{20})C_i})R(C_i)\}, \end{aligned}$$

provided the expectation on the right-side exists. ∎

The following Lemma 6.2 summarizes the expressions for the elements of $\mathscr{D}_i(\boldsymbol{\theta_0})$.

Lemma 6.2 *For all* $1 \le i \le n,\ \lambda_{10} > 0,\ \lambda_{20} > 0$ *and* $p_0 \in (0,1\big]$, *the following are finite:*

$$d_i^{11}(\boldsymbol{\theta}_o) = p_0 E\Big[\frac{1-e^{-(\lambda_{10}+\lambda_{20})C_i}}{\lambda_{10}(\lambda_{10}+\lambda_{20})} - \frac{C_i^2(1-p_0)e^{-(\lambda_{10}+\lambda_{20})C_i}}{1-p_0+p_0e^{-(\lambda_{10}+\lambda_{20})C_i}}\Big] \tag{21}$$

$$d_i^{22}(\boldsymbol{\theta}_o) = p_0 E\Big[\frac{1-e^{-(\lambda_{10}+\lambda_{20})C_i}}{\lambda_{20}(\lambda_{10}+\lambda_{20})} - \frac{C_i^2(1-p_0)e^{-(\lambda_{10}+\lambda_{20})C_i}}{1-p_0+p_0e^{-(\lambda_{10}+\lambda_{20})C_i}}\Big] \tag{22}$$

$$d_i^{33}(\boldsymbol{\theta}_o) = \frac{1}{p_0} E\Big[\frac{1-e^{-(\lambda_{10}+\lambda_{20})C_i}}{1-p_0+p_0e^{-(\lambda_{10}+\lambda_{20})C_i}}\Big] \tag{23}$$

$$d_i^{12}(\boldsymbol{\theta}_o) = d_i^{21}(\boldsymbol{\theta}_o) = -\frac{1}{\lambda_{10}+\lambda_{20}} E\Big[\frac{p_0(1-p_0)C_i^2 e^{-(\lambda_{10}+\lambda_{20})C_i}}{\{1-p_0+p_0e^{-(\lambda_{10}+\lambda_{20})C_i}\}^2}\Big] \tag{24}$$

$$d_i^{13}(\boldsymbol{\theta}_o) = d_i^{31}(\boldsymbol{\theta}_o) = d_i^{23}(\boldsymbol{\theta}_o) = d_i^{32}(\boldsymbol{\theta}_o) = \frac{1}{\lambda_{10}+\lambda_{20}} E\Big[\frac{(\lambda_{10}+\lambda_{20})C_i e^{-(\lambda_{10}+\lambda_{20})C_i}}{1-p+pe^{-(\lambda_1+\lambda_2)C_i}}\Big] \tag{25}$$

Proof: The proof follows easily applying the above Lemma 6.1. ∎

Remarks: Type-I censoring (termination of the experiment at time $\tau > 0$) turns out to be a particular case when $P(C_i = \tau) = 1,\ i = 1, 2, \ldots, n$. Then, all the expectations will exist since

$$E\big[(1-Z_i)R(T_i)\big] = E\big[\{1-p_0+p_0e^{-(\lambda_{10}+\lambda_{20})C_i}\}R(C_i)\big] = 1-p_0+p_0e^{-(\lambda_{10}+\lambda_{20})\tau}R(\tau)$$

will always exist.

Lemma 6.3 *If the censored times C_i's for $i = 1, \ldots, n$, are iid random variables whose common distribution does not degenerate at origin, then*

$$\liminf_{n\to\infty} \frac{1}{n}\lambda_{min}\{\mathbf{D}_n(\boldsymbol{\theta_0})\} > 0 \tag{26}$$

will hold i.e there exists some $\eta > 0$ such that $\lambda_{min}\{\mathbf{D}_n(\boldsymbol{\theta_0})\} \geq n\eta$ for n large enough.

Proof: When C_i's are iid random variables, $\lambda_{min}\{\mathbf{D}_n(\boldsymbol{\theta_0})\} = n\lambda_{min}\{\mathscr{D}_1(\boldsymbol{\theta_0})\}$. Therefore, it is enough to show that $\mathscr{D}_1(\boldsymbol{\theta_0})$ is positive definite.

From (21), (22), (23), it is clear that, $d_1^{11}(\boldsymbol{\theta_0})$, $d_1^{22}(\boldsymbol{\theta_0})$, $d_1^{33}(\boldsymbol{\theta_0})$ all are 0 if and only if $C_1 = 0$. This is true, since, for $\lambda_{10} > 0$, $\lambda_{20} > 0$, $p_0 \in (0,1)$, and $C \geq 0$, the following functions of C defined as

$$g_1(C) = 1 - e^{-(\lambda_{10}+\lambda_{20})C} - \frac{C^2(1-p_0)e^{-(\lambda_{10}+\lambda_{20})C}\lambda_{10}(\lambda_{10}+\lambda_{20})}{1-p_0+p_0e^{-(\lambda_{10}+\lambda_{20})C}} \tag{27}$$

$$g_2(C) = 1 - e^{-(\lambda_{10}+\lambda_{20})C} - \frac{C^2(1-p_0)e^{-(\lambda_{10}+\lambda_{20})C}\lambda_{20}(\lambda_{10}+\lambda_{20})}{1-p_0+p_0e^{-(\lambda_{10}+\lambda_{20})C}} \tag{28}$$

$$g_3(C) = \frac{1 - e^{-(\lambda_{10}+\lambda_{20})C}}{1-p_0+p_0e^{-(\lambda_{10}+\lambda_{20})C}} \tag{29}$$

are non-negative, increasing, see for example Ghitany and Maller[9]) and at $C = 0$, $g_i(C) = 0$, $i = 1, 2, 3$. Thus from (12), we have

$$Var(s_{11}(\boldsymbol{\theta_0})) = d_1^{11}(\boldsymbol{\theta_0}) > 0, Var(s_{12}(\boldsymbol{\theta_0})) = d_2^{11}(\boldsymbol{\theta_0}) > 0, Var(s_{13}(\boldsymbol{\theta_0})) = d_3^{11}(\boldsymbol{\theta_0}) > 0,$$

where $s_{ij}(\boldsymbol{\theta_0})'s$ are defined in (3), (4), (5). We will prove the result by contradiction. Let us first assume that the determinant of $\mathscr{D}_1(\boldsymbol{\theta_0})$ is 0. Then, $s_{11}(\boldsymbol{\theta_0})$, $s_{12}(\boldsymbol{\theta_0})$ and $s_{13}(\boldsymbol{\theta_0})$ are almost surely linearly related. We can write

$$a_1 s_{11}(\boldsymbol{\theta_0}) + a_2 s_{12}(\boldsymbol{\theta_0}) + a_3 s_{13}(\boldsymbol{\theta_0}) = 0 \tag{30}$$

for some a_1, a_2, a_3. When $C_1 \neq 0$ almost surely, then $P(Z_1 = 1) > 0$ and $P(\Delta_1 = 1) > 0$. Now plugging in $z_i = 1$ and $\delta_i = 1$ in (30), we obtain $t_{11}^*(a_1 + a_2) = \frac{a_1}{\lambda_{10}} + \frac{a_3}{p_0}$, almost surely on a set say $\mathcal{B}$ with positive probability. This implies that on $\mathcal{B}$ t^*_{11} is degenerate which is impossible, since $\mathcal{B}$ t^*_{11} is exponential. Therefore, $\mathscr{D}_1$ is non singular , hence positive definite. ∎

Lemma 6.4 *For $\lambda_{10}, \lambda_{20} > 0$ and $p_0 \in (0,1)$, $|f_i^{rs}(\boldsymbol{\theta_0})| < M(\boldsymbol{\theta_0})$ for all $i = 1,\ldots,n$ and $r,s = 1,2,3$ where $f_i^{rs}(\boldsymbol{\theta})'s$ are defined in* (15), (16), (17),(18) *and* (19) *and $M(\boldsymbol{\theta_0})$ is a positive finite quantity depending on $\boldsymbol{\theta_0}$.*

Proof: At $\boldsymbol{\theta} = \boldsymbol{\theta_0}$, $f_i^{rs}(\boldsymbol{\theta_0})$'s are continuous function of t_i for $r,s = 1,2,3$. Again for all $i = 1,\ldots,n$,

$$\begin{aligned}
\lim_{t_i\to 0} f_i^{11}(\boldsymbol{\theta_0}) &= \frac{z_i(2-\delta_i)}{\lambda_{1o}^2}; \lim_{t_i\to 0} f_i^{22}(\boldsymbol{\theta_0}) = \frac{z_i(\delta_i-1)}{\lambda_{2o}^2}; \lim_{t_i\to 0} f_i^{33}(\boldsymbol{\theta_0}) = \frac{z_i}{p_0}; \lim_{t_i\to 0} f_i^{12}(\boldsymbol{\theta_0}) = 0;\\
\lim_{t_i\to 0} f_i^{13}(\boldsymbol{\theta_0}) &= 0 \quad \text{and}\\
\lim_{t_i\to \infty} f_i^{11}(\boldsymbol{\theta_0}) &= \frac{z_i(2-\delta_i)}{\lambda_{1o}^2}; \lim_{t_i\to \infty} f_i^{22}(\boldsymbol{\theta_0}) = \frac{z_i(\delta_i-1)}{\lambda_{2o}^2}; \lim_{t_i\to \infty} f_i^{33}(\boldsymbol{\theta_0}) = \frac{z_i}{p_0} + \frac{(1-z_i)}{(1-p_0)};\\
\lim_{t_i\to \infty} f_i^{12}(\boldsymbol{\theta_0}) &= 0; \lim_{t_i\to \infty} f_i^{13}(\boldsymbol{\theta_0}) = 0.
\end{aligned}$$

z_i and δ_i being binary variable all the $f_i^{rs}(\boldsymbol{\theta_0})$'s have finite limit at 0 and ∞, implies $|f_i^{rs}(\boldsymbol{\theta_0})| < M^{rs}(\boldsymbol{\theta_0})$ where $M^{rs}(\boldsymbol{\theta_0})$ is a positive quantity for $r,s = 1,2,3$.
Let $M(\boldsymbol{\theta_0}) = \max_{r,s} M^{rs}(\boldsymbol{\theta_0})$. Therefore, $|f_i^{rs}(\boldsymbol{\theta_0})| < M(\boldsymbol{\theta_0})$ for all $r,s = 1,2,3$ and $i = 1,\ldots,n$. ∎

Lemma 6.5 *For any $\boldsymbol{\theta} \in \mathbf{N}_n(\boldsymbol{\theta_0})$ and* (26) *holds, $|w_{ij}^{rs}(\boldsymbol{\theta})| < M_1(\boldsymbol{\theta}) < M_2(\boldsymbol{\theta_0})$ for all $i = 1,\ldots,n$ and $j,r,s = 1,2,3,$, where, $w_i^{rs}(\boldsymbol{\theta}) = \frac{\partial f_i^{rs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (w_{i1}^{rs}(\boldsymbol{\theta}), w_{i2}^{rs}(\boldsymbol{\theta}), w_{i3}^{rs}(\boldsymbol{\theta}))$. Here, $M_1(\boldsymbol{\theta})$ is a positive finite quantity depend on $\boldsymbol{\theta}$ and $M_2(\boldsymbol{\theta_0})$ is a positive finite quantity depend on $\boldsymbol{\theta_0}$.*

Proof: For fixed $\boldsymbol{\theta}$,and for all $i = 1,\ldots,n$, $j = 1,2,3$, each of $w_{ij}^{rs}(\boldsymbol{\theta})$ is a continuous function of t_i with finite limit at 0 and ∞. Hence for each $j,r,s = 1,2,3$, there exists a positive quantity $M_j^{rs}(\boldsymbol{\theta})$ such that $|w_{ij}^{rs}(\boldsymbol{\theta})| \leq M_j^{rs}(\boldsymbol{\theta})$. For $\lambda_{10}, \lambda_{20} > 0$ and $p_0 \in (0,1)$, each $M_j^{rs}(\boldsymbol{\theta})$ is continuous function of $\boldsymbol{\theta}$. Again when $\boldsymbol{\theta} \in \mathbf{N}_n(\boldsymbol{\theta_0})$, using (26) we obtain,

$$|\boldsymbol{\theta} - \boldsymbol{\theta_0}|^2 \leq \frac{A^2}{\lambda_{min}\{\mathbf{D}_n(\boldsymbol{\theta}_0)\}} < \frac{A^2}{n\eta} \to 0 \quad \text{as} \quad n \to \infty.$$

where $\eta > 0$ is a suitable constant and this implies, for each $j, r, s = 1, 2, 3$ there exist a quantity $N_j^{rs}(\boldsymbol{\theta_0})$ such that $M_j^{rs}(\boldsymbol{\theta}) \leq N_j^{rs}(\boldsymbol{\theta_0})$. Let $\max_{j,r,s} M_j^{rs}(\boldsymbol{\theta}) = M_1(\boldsymbol{\theta})$ and $\max_{j,r,s} N_j^{rs}(\boldsymbol{\theta_0}) = M_2(\boldsymbol{\theta_0})$. Therefore, for each $i = 1, \ldots, n$ and for each $j, r, s = 1, 2, 3$ $|w_{ij}^{rs}(\boldsymbol{\theta})| < M_1(\boldsymbol{\theta}) < M_2(\boldsymbol{\theta_0})$. ∎

Lemma 6.6 *For $\lambda_{10}, \lambda_{20} > 0$ and $p_0 \in (0,1)$, if the condition (26) holds, $\mathbf{D}_n(\boldsymbol{\theta_0})$ is a positive definite matrix and*

$$\sup_{\theta \in \mathbf{N}_n(\boldsymbol{\theta_0})} \|\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\mathbf{F}_n(\boldsymbol{\theta})\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0}) - \mathbf{I}\| \to 0 \text{ as } n \to \infty \text{ in probability,} \tag{31}$$

will hold along with $\lambda_{min}\{\mathbf{D}_n(\boldsymbol{\theta_0})\} \to \infty$ as $n \to \infty$.

Proof: Condition (26) implies that, there exist a $\eta > 0$ such that $\lambda_{min}\{\mathbf{D}_n(\boldsymbol{\theta_0})\} \geq n\eta$ for n large enough. Therefore all the eigenvalues of $\mathbf{D}_n(\boldsymbol{\theta_0})$ are positive for large n which implies $\mathbf{D}_n(\boldsymbol{\theta_0})$ is a positive definite matrix for large n. Now, we can define $\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})$, the symmetric square root of $\mathbf{D}_n^{-1}(\boldsymbol{\theta_0})$.

Now to show that (31) will hold, we decompose $\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\mathbf{F}_n(\boldsymbol{\theta})\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})$ as

$$\begin{aligned} \mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\mathbf{F}_n(\boldsymbol{\theta})\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0}) = \mathbf{I} &+ \mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\{\mathbf{F}_n(\boldsymbol{\theta_0}) - \mathbf{D}_n(\boldsymbol{\theta_0})\}\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0}) \\ &+ \mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\{\mathbf{F}_n(\boldsymbol{\theta}) - \mathbf{F}_n(\boldsymbol{\theta_0})\}\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0}). \end{aligned}$$

Hence, it is sufficient to prove that for any unit vector $\mathbf{u}$ and for any $\boldsymbol{\theta} \in \mathbf{N}_n(\boldsymbol{\theta_0})$,

$$e_n^{(1)}(\boldsymbol{\theta_0}) = \mathbf{u}^\top\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\{\mathbf{F}_n(\boldsymbol{\theta_0}) - \mathbf{D}_n(\boldsymbol{\theta_0})\}\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\mathbf{u} = o_p(1), \tag{32}$$

$$e_n^{(2)}(\boldsymbol{\theta}) = \mathbf{u}^\top\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\{\mathbf{F}_n(\boldsymbol{\theta}) - \mathbf{F}_n(\boldsymbol{\theta_0})\}\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\mathbf{u} = o_p(1). \tag{33}$$

To show $e_n^{(1)}(\boldsymbol{\theta_0}) = o_p(1)$, we need the results of Lemma 6.4. Now, $f_i^{rs}(\boldsymbol{\theta_0})$ is a sequence of independent random variables and based on Lemma 6.4, we have $|f_i^{rs}(\boldsymbol{\theta_0})| < M(\boldsymbol{\theta_0})$, $r, s = 1, 2, 3$. Hence,

$$\frac{1}{n^2}\sum_{i=1}^{n} Var\{f_i^{rs}(\boldsymbol{\theta_0})\} \leq \frac{M^2(\boldsymbol{\theta_0})}{n} \to 0 \quad \text{as } n \to \infty.$$

Therefore, by the weak law of large numbers, for all $r, s = 1, 2, 3$,

$$\frac{1}{n}\sum_{i=1}^{n}\{f_i^{rs}(\boldsymbol{\theta_0}) - d_i^{rs}(\boldsymbol{\theta_0})\} = o_p(1).$$

For any unit vector $\mathbf{u}$, let $\mathbf{u}_n = \frac{\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\mathbf{u}}{\sqrt{\mathbf{u}^\top \mathbf{D}_n^{-1}(\boldsymbol{\theta_0})\mathbf{u}}}$. Then $\mathbf{u}_n$ is also a unit vector and hence

$$|\mathbf{u}_n^\top M_n \mathbf{u}_n| \leq \sum_{r,s=1,2,3} |m_n^{rs}|$$

for any matrix $M_n = (m_n^{rs}),\ r, s = 1, 2, 3$. Thus, we have

$$|\mathbf{u}_n^\top\{\mathbf{F}_n(\boldsymbol{\theta_0}) - \mathbf{D}_n(\boldsymbol{\theta_0})\}\mathbf{u}_n| \leq \sum_{r,s=1,2,3} |\sum_{i=1}^{n}\{f_i^{rs}(\boldsymbol{\theta_0}) - d_i^{rs}(\boldsymbol{\theta_0})\}|.$$

Using the result in (26), we have for $\eta > 0$, $\mathbf{u}^\top\mathbf{D}_n^{-1}(\boldsymbol{\theta_0})\mathbf{u} \leq \lambda_{max}\{\mathbf{D}_n^{-1}(\boldsymbol{\theta_0})\} \leq \frac{1}{n\eta}$. Then

$$|e_n^{(1)}(\boldsymbol{\theta_0}| = |\mathbf{u}_n^\top(\mathbf{F}_n(\boldsymbol{\theta_0}) - \mathbf{D}_n(\boldsymbol{\theta_0}))\mathbf{u}_n||\mathbf{u}^\top\mathbf{D}_n^{-1}(\boldsymbol{\theta_0})\mathbf{u}| \leq \frac{1}{n\eta}\sum_{r,s=1,2,3}|\sum_{i=1}^{n}(f_i^{rs}(\boldsymbol{\theta_0}) - d_i^{rs}(\boldsymbol{\theta_0}))| = o_p(1).$$

To show $e_n^{(2)}(\boldsymbol{\theta}) \xrightarrow{P} 0$, we need the following Taylor series expansion of $f_i^{rs}(\boldsymbol{\theta})$ about $\boldsymbol{\theta_0}$.

$$f_i^{rs}(\boldsymbol{\theta}) - f_i^{rs}(\boldsymbol{\theta_0}) = (\boldsymbol{\theta} - \boldsymbol{\theta_0})^\top w_i^{rs}(\tilde{\boldsymbol{\theta}}),$$

where $\tilde{\boldsymbol{\theta}}$ is any point between $\boldsymbol{\theta}$ and $\boldsymbol{\theta_0}$, $w_i^{rs}(\boldsymbol{\theta}) = \frac{\partial f_i^{rs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (w_{i1}^{rs}(\boldsymbol{\theta}), w_{i2}^{rs}(\boldsymbol{\theta}), w_{i3}^{rs}(\boldsymbol{\theta}))$, $i = 1, 2, \ldots, n,\ r, s = 1, 2, 3$. For fixed (r, s), $w_i^{rs}(\boldsymbol{\theta})$ is a sequence of independent random vectors. The expression of the elements of $w_i^{rs}(\boldsymbol{\theta})$ are as follows.

$$
\begin{aligned}
w_{i1}^{11}(\boldsymbol{\theta}) &= -\frac{2}{\lambda_1^3}\Big\{z_i(2-\delta_i) - \frac{(1-z_i)p(1-p)(1-p-pe^{-(\lambda_1+\lambda_2)t_i}\lambda_1^3 t_i^3 e{-}(\lambda_1+\lambda_2)t_i}{2(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^3}\Big\},\\
w_{i2}^{11}(\boldsymbol{\theta}) &= \frac{(1-z_i)p(1-p)t_i^3(1-p-pe^{-(\lambda_1+\lambda_2)t_i}e^{-(\lambda_1+\lambda_2)t_i}}{(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^3},\\
w_{i3}^{11}(\boldsymbol{\theta}) &= -\frac{(1-z_i)t_i^2 e^{-(\lambda_1+\lambda_2)t_i}(1-p-pe^{-(\lambda_1+\lambda_2)t_i})}{(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^3},\\
w_{i2}^{22}(\boldsymbol{\theta}) &= -\frac{2}{\lambda_2^3}\Big\{z_i(\delta_i-1) - \frac{(1-z_i)p(1-p)(1-p-pe^{-(\lambda_1+\lambda_2)t_i}\lambda_2^3 t_i^3 e^{-(\lambda_1+\lambda_2)t_i}}{2(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^3}\Big\},
\end{aligned}
$$

$$
\begin{aligned}
w_{i3}^{13}(\boldsymbol{\theta}) &= \frac{2(1-z_i)t_i e^{-(\lambda_1+\lambda_2)t_i}(1-e^{-(\lambda_1+\lambda_2)t_i})}{(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^3},\\
w_{i3}^{33}(\boldsymbol{\theta}) &= \frac{z_i}{p^2} + \frac{(1-z_i)(1-e^{-(\lambda_1+\lambda_2)t_i})^2}{(1-p+pe^{-(\lambda_1+\lambda_2)t_i})^2},\\
w_{i2}^{11}(\boldsymbol{\theta}) &= w_{i1}^{22}(\boldsymbol{\theta}) = w_{i1}^{12}(\boldsymbol{\theta}) = w_{i2}^{12}(\boldsymbol{\theta}) = w_{i3}^{12}(\boldsymbol{\theta}) = w_{i2}^{21}(\boldsymbol{\theta}) = w_{i3}^{21}(\boldsymbol{\theta}) = w_{i1}^{21}(\boldsymbol{\theta}),\\
w_{i3}^{11}(\boldsymbol{\theta}) &= w_{i3}^{22}(\boldsymbol{\theta}) = w_{i1}^{13}(\boldsymbol{\theta}) = w_{i2}^{13}(\boldsymbol{\theta}) = w_{i1}^{23}(\boldsymbol{\theta}) = w_{i1}^{31}(\boldsymbol{\theta}) = w_{i1}^{32}(\boldsymbol{\theta}) = w_{i2}^{32}(\boldsymbol{\theta}) = w_{i2}^{31}(\boldsymbol{\theta}) = w_{i2}^{23}(\boldsymbol{\theta}),\\
w_{i3}^{13}(\boldsymbol{\theta}) &= w_{i3}^{31}(\boldsymbol{\theta}) = w_{i3}^{32}(\boldsymbol{\theta}) = w_{i3}^{23}(\boldsymbol{\theta}) = w_{i2}^{33}(\boldsymbol{\theta}) = w_{i1}^{33}(\boldsymbol{\theta}).
\end{aligned}
$$

For further development we need the result of Lemma 6.5. As a result of Lemma 6.5,

$$\sup_{\boldsymbol{\theta}\in\mathbf{N}_n(\boldsymbol{\theta_0})} \|\frac{1}{n}\sum_{i=1}^n w_i^{rs}(\tilde{\boldsymbol{\theta}})\| = O_p(1),$$

where $\|.\|$ denotes the norm of a vector.

Also, when $\boldsymbol{\theta}\in\mathbf{N}_n(\boldsymbol{\theta_0})$, $|\boldsymbol{\theta}-\boldsymbol{\theta_0}|^2 \le \frac{A^2}{\lambda_{min}\{\mathbf{D}_n(\boldsymbol{\theta_0})\}} \le \frac{A^2}{n\eta} \to 0$ as $n\to\infty$.

Since, $\tilde{\boldsymbol{\theta}}\in\mathbf{N}_n(\boldsymbol{\theta_0})$, it follows that

$$
\begin{aligned}
\sup_{\boldsymbol{\theta}\in\mathbf{N}_n(\boldsymbol{\theta_0})} |\frac{1}{n}(\boldsymbol{\theta}-\boldsymbol{\theta_0})^\top \sum_{i=1}^n w_i^{rs}(\tilde{\boldsymbol{\theta}})| &\le \sup_{\boldsymbol{\theta}\in\mathbf{N}_n(\boldsymbol{\theta_0})} \|(\boldsymbol{\theta}-\boldsymbol{\theta_0})\|\|\frac{1}{n}\sum_{i=1}^n w_i^{rs}(\tilde{\boldsymbol{\theta}})\|\\
&\le \frac{A}{\sqrt{n\eta}} \sup_{\boldsymbol{\theta}\in\mathbf{N}_n(\boldsymbol{\theta_0})} \|\frac{1}{n}\sum_{i=1}^n w_i^{rs}(\tilde{\boldsymbol{\theta}})\|\\
&= o_p(1)
\end{aligned}
$$

Therefore,

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\mathbf{N}_n(\theta_0)} |e_n^{(2)}(\boldsymbol{\theta})| &= \sup_{\theta\in\mathbf{N}_n(\boldsymbol{\theta_0})} |\mathbf{u_n}^\top\{\mathbf{F}_n(\boldsymbol{\theta})-\mathbf{F}_n(\boldsymbol{\theta_0})\}\mathbf{u_n}||\mathbf{u}^\top\mathbf{D}_n^{-1}\mathbf{u}| \\
&\leq \frac{1}{n\eta}\sup_{\boldsymbol{\theta}\in\mathbf{N}_n(\boldsymbol{\theta_0})}\sum_{r,s=1,2,3}|\sum_{i=1}^{n}\{f_i^{rs}(\boldsymbol{\theta})-f_i^{rs}(\boldsymbol{\theta_0})\}| \\
&\leq \frac{1}{n\eta}\sum_{r,s=1,2,3}\sup_{\boldsymbol{\theta}\in\mathbf{N}_n(\boldsymbol{\theta_0})}|(\boldsymbol{\theta}-\boldsymbol{\theta_0})^\top\sum_{i=1}^{n}w_i^{rs}(\tilde{\boldsymbol{\theta}})| \\
&= o_p(1)
\end{aligned}$$

This concludes the proof of Lemma (6.6). ∎

Lemma 6.7 *For $\lambda_{10}, \lambda_{20} > 0$ and $p_0 \in (0,1)$, if condition (26) holds, a unique MLE of $\boldsymbol{\theta}_o$, say $\widehat{\boldsymbol{\theta_n}}$, will exist in $\mathbf{N}_n(\boldsymbol{\theta}_o)$ with probability 1 as $n \to \infty$.*

Proof: If the condition (26) holds, from Lemma 6.6, $\mathbf{D}_n(\boldsymbol{\theta_0})$ is positive definite matrix and a symmetric square root of $\mathbf{D}_n^{-1}(\boldsymbol{\theta_0})$ will always exist. Let $\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})$ be the symmetric square root of $\mathbf{D}_n^{-1}(\boldsymbol{\theta_0})$.

Further consider the Taylor-series expansion of the log likelihood function $\mathbf{l}_n(\boldsymbol{\theta})$ about the true parameter value $\boldsymbol{\theta_0}$ as follows

$$\mathbf{l}_n(\boldsymbol{\theta}) - \mathbf{l}_n(\boldsymbol{\theta_0}) = (\boldsymbol{\theta}-\boldsymbol{\theta_0})^\top\mathbf{S}_n(\boldsymbol{\theta_0}) - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta_0})^\top\mathbf{F}_n(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta_0}), \tag{34}$$

where $\bar{\boldsymbol{\theta}}$ is a any point on the line segment joining $\boldsymbol{\theta_0}$ and $\boldsymbol{\theta}$.
Let $\boldsymbol{\theta_n}$ be on the boundary, $\partial\mathbf{N}_n(\boldsymbol{\theta_0})$ of $\mathbf{N}_n(\boldsymbol{\theta_0})$. We define $v_n = \dfrac{\mathbf{D}_n^{\frac{1}{2}}(\boldsymbol{\theta_0})(\boldsymbol{\theta_n}-\boldsymbol{\theta_0})}{A}$. Then for $\boldsymbol{\theta_n} \in \partial\mathbf{N}_n(\boldsymbol{\theta_0})$, v_n is an unit vector.

For any arbitrary $\epsilon \in (0,1)$,

$$\begin{aligned}
P(\frac{1}{2}(\boldsymbol{\theta_n}-\boldsymbol{\theta_0})^\top\mathbf{F}_n(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta_n}-\boldsymbol{\theta_0}) > \frac{\epsilon A^2}{2}) &= P(v_n^\top\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\mathbf{F}_n(\bar{\boldsymbol{\theta}})\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})v_n > \epsilon) \\
&\geq P(\lambda_{min}\{\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\mathbf{F}_n(\bar{\boldsymbol{\theta}})\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\} > \epsilon)
\end{aligned}$$

Again if the condition (26) holds, from Lemma 6.6, we get the eigenvalues of $\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})\mathbf{F}_n(\bar{\boldsymbol{\theta}})\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta_0})$ converges to 1 in probability. Therefore, $P(\frac{1}{2}(\boldsymbol{\theta_n} - \boldsymbol{\theta_0})^\top \mathbf{F}_n(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta_n} - \boldsymbol{\theta_0} \leq \frac{\epsilon A^2}{2}) = o(1)$.

As $(\boldsymbol{\theta_n} - \boldsymbol{\theta_0})^\top \mathbf{S}_n(\boldsymbol{\theta_0})$ has mean 0, applying Chebyshev's inequality we obtain,

$$P((\boldsymbol{\theta_n} - \boldsymbol{\theta_0})^\top \mathbf{S}_n(\boldsymbol{\theta_0}) > \frac{\epsilon A}{2}) \leq \frac{4(\boldsymbol{\theta_n} - \boldsymbol{\theta_0})^\top \mathbf{D}_n(\boldsymbol{\theta_0})(\boldsymbol{\theta_n} - \boldsymbol{\theta_0})}{\epsilon^2 A^2} = \frac{4}{\epsilon^2}$$

Thus

$$\begin{aligned} & P\big((\boldsymbol{\theta_n} - \boldsymbol{\theta_0})^\top \mathbf{S}_n(\boldsymbol{\theta_0}) > \frac{1}{2}(\boldsymbol{\theta_n} - \boldsymbol{\theta_0})^\top \mathbf{F}_n(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta_n} - \boldsymbol{\theta_0})\big) \\ \leq \;& P\big((\boldsymbol{\theta_n} - \boldsymbol{\theta_0})^\top \mathbf{S}_n(\boldsymbol{\theta_0}) > \frac{\epsilon A}{2}\big) + P\big(\frac{1}{2}(\boldsymbol{\theta_n} - \boldsymbol{\theta_0})^\top \mathbf{F}_n(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta_n} - \boldsymbol{\theta_0}) \leq \frac{\epsilon A^2}{2}\big) \\ \leq \;& \frac{4}{\epsilon^2} + o(1). \end{aligned}$$

Therefore for A and n large enough, $\mathbf{l}_n(\boldsymbol{\theta}) < \mathbf{l}_n(\boldsymbol{\theta_0})$ for any $\boldsymbol{\theta}$ lies on $\partial \mathbf{N}_n(\boldsymbol{\theta_0})$. As $\mathbf{D}_n(\boldsymbol{\theta_0})$ is positive definite, by (31), $\mathbf{F}_n(\boldsymbol{\theta})$ is positive definite. Therefore, $\mathbf{l}_n(\boldsymbol{\theta})$ is a concave function in $\mathbf{N}_n(\boldsymbol{\theta_0})$ which is an ellipsoid. $\mathbf{l}_n(\boldsymbol{\theta})$ does not have maximum on boundary of $\mathbf{N}_n(\boldsymbol{\theta_0})$ and hence it has a local maximum inside $\mathbf{N}_n(\boldsymbol{\theta_0})$ which is a global maximum because of its concavity. Therefore, an MLE $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$, of $\boldsymbol{\theta_0}$ exists inside the ellipsoid $\mathbf{N}_n(\boldsymbol{\theta_0})$ with probability 1 as $n \to \infty$ and it is unique.

Proof of Theorem 4.1: Based on Lemma 6.3 and Lemma 6.7, the proof immediately follows. ∎

Lemma 6.8 *For* $\lambda_{10}, \lambda_{20} > 0$ *and* $p_0 \in (0, 1)$, *if the condition* (26) *holds,* $\lambda_{min}\{\mathbf{D}_n\} \to \infty$ *as* $n \to \infty$, $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ *is a consistent estimator of* $\boldsymbol{\theta_o}$.

Proof:

Based on Lemma 6.7, as the MLE $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} \in \mathbf{N}_n(\boldsymbol{\theta_0})$, $(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0})^\top \mathbf{D}_n(\boldsymbol{\theta}_o)\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}\big) \leq A^2$. Again $(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0})^2 \lambda_{min}\{\mathbf{D}_n(\boldsymbol{\theta}_o)\} \leq (\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0})' \mathbf{D}_n(\boldsymbol{\theta}_o)\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}\big)$. If in addition, $\lambda_{min}\{\mathbf{D}_n\} \to \infty$ as $n \to \infty$, $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} \to \boldsymbol{\theta_0}$ in probability. ∎

Proof of Theorem 4.2: Based on Lemma 6.3 and Lemma 6.8, the proof immediately follows. ∎

Lemma 6.9 *For* $\lambda_{10}, \lambda_{20} > 0$ *and* $p_0 \in (0,1)$ *there exists a finite positive quantity* $G(\boldsymbol{\theta_0})$ *such that* $E|s_i(\boldsymbol{\theta_0})|^4 \leq G(\boldsymbol{\theta_0})$.

Proof: Along the same lines as the previous two lemmas, Lemma 6.4 and Lemma 6.5), we can show that $s_i(\boldsymbol{\theta_0})$ is also bounded, hence, the moments are bounded. ∎

Lemma 6.10 *For* $\lambda_{10}, \lambda_{20} > 0$ *and* $p_0 \in (0,1)$, *if the condition* (26) *holds,* $\mathbf{D}_n^{1/2}(\boldsymbol{\theta}_o)\big(\widehat{\boldsymbol{\theta_n}} - \boldsymbol{\theta}_o\big)$ *is asymptotically normally distributed with mean vector* $\mathbf{0}$ *and covariance matrix* $\mathbf{I}$.

Proof: Considering the element wise Taylor series expansion of $\mathbf{S}_n(\boldsymbol{\theta_0})$, about $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$, we obtain

$$\mathbf{S}_n(\boldsymbol{\theta_0}) - \mathbf{S}_n(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}) = \mathbf{K}_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}})(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}),$$

where, $\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}} = (\boldsymbol{\theta}^*_{\boldsymbol{n}}, \boldsymbol{\theta}^{**}_{\boldsymbol{n}}, \boldsymbol{\theta}^{***}_{\boldsymbol{n}})$ where $\boldsymbol{\theta}^*_{\boldsymbol{n}}$, $\boldsymbol{\theta}^{**}_{\boldsymbol{n}}$ and $\boldsymbol{\theta}^{***}_{\boldsymbol{n}}$ are three points on the line segment joining $\boldsymbol{\theta_0}$ and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ and

$$\mathbf{K}_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) = \begin{bmatrix} \sum_{i=1}^n f_i^{11}(\boldsymbol{\theta}^*_{\boldsymbol{n}}) & \sum_{i=1}^n f_i^{12}(\boldsymbol{\theta}^*_{\boldsymbol{n}}) & \sum_{i=1}^n f_i^{13}(\boldsymbol{\theta}^*_{\boldsymbol{n}}) \\ \sum_{i=1}^n f_i^{21}(\boldsymbol{\theta}^{**}_{\boldsymbol{n}}) & \sum_{i=1}^n f_i^{22}(\boldsymbol{\theta}^{**}_{\boldsymbol{n}}) & \sum_{i=1}^n f_i^{23}(\boldsymbol{\theta}^{**}_{\boldsymbol{n}}) \\ \sum_{i=1}^n f_i^{31}(\boldsymbol{\theta}^{***}_{\boldsymbol{n}}) & \sum_{i=1}^n f_i^{32}(\boldsymbol{\theta}^{***}_{\boldsymbol{n}}) & \sum_{i=1}^n f_i^{33}(\boldsymbol{\theta}^{***}_{\boldsymbol{n}}) \end{bmatrix}.$$

As the likelihood function is maximized at MLE $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$, $\mathbf{S}_n(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}$, and we can write,

$$\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\mathbf{S}_n(\boldsymbol{\theta_0}) = \mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\mathbf{K}_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}})(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}) = A_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}})\mathbf{D}_n^{\frac{1}{2}}(\boldsymbol{\theta}_o)(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0})$$

where $A_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) = \mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\mathbf{K}_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}})\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)$. To prove the asymptotic normality of $\mathbf{D}_n^{\frac{1}{2}}(\boldsymbol{\theta}_o)(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0})$, it is enough to show that for any $\boldsymbol{\theta}^*_{\boldsymbol{n}}$, $\boldsymbol{\theta}^{**}_{\boldsymbol{n}}$ and $\boldsymbol{\theta}^{***}_{\boldsymbol{n}} \in \mathbf{N}_n(\boldsymbol{\theta_0})$,

$A_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) \xrightarrow{P} \mathbf{I}$ and $\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\mathbf{S}_n(\boldsymbol{\theta}_{\mathbf{0}})$ converges in distribution to a standard normal random variable.

$$\begin{aligned} A_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) = \mathbf{I} + \mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\big\{\mathbf{F}_n(\boldsymbol{\theta}_{\mathbf{0}}) - \mathbf{D}_n(\boldsymbol{\theta}_o)\big\}\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o) \\ +\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\big\{\mathbf{K}_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) - \mathbf{F}_n(\boldsymbol{\theta}_{\mathbf{0}}\big\})\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o) \end{aligned}$$

Applying (32), $\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\big\{\mathbf{F}_n(\boldsymbol{\theta}_{\mathbf{0}}) - \mathbf{D}_n(\boldsymbol{\theta}_o)\big\}\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)$ converges in probability to the zero matrix.

Also for any unit vector $\mathbf{u}$ and $\boldsymbol{\theta}_{\boldsymbol{n}}^{*}, \boldsymbol{\theta}_{\boldsymbol{n}}^{**}, \boldsymbol{\theta}_{\boldsymbol{n}}^{***} \in \mathbf{N}_n(\boldsymbol{\theta}_{\mathbf{0}})$, we can prove that $\mathbf{u}^{'}\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\big\{\mathbf{K}_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) - \mathbf{F}_n(\boldsymbol{\theta}_{\mathbf{0}}\big\})\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\mathbf{u} = o_p(1)$. The proof follows exactly along the same lines as the proof of (33). Hence, $A_n(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) \xrightarrow{P} \mathbf{I}$.

$\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\mathbf{S}_n(\boldsymbol{\theta}_{\mathbf{0}})$ can be written as

$$\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\mathbf{S}_n(\boldsymbol{\theta}_{\mathbf{0}}) = \mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\sum_{i=1}^{n}(\mathbf{s}_i(\boldsymbol{\theta}_{\mathbf{0}})$$

Let $X_{in}(\boldsymbol{\theta}_{\mathbf{0}}) = \mathbf{u}^{\top}\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\mathbf{s}_i(\boldsymbol{\theta}_{\mathbf{0}}),\ i = 1, 2, \ldots, n$, where $\mathbf{u}$ is a unit vector.

Then, $E(X_{in}(\boldsymbol{\theta}_{\mathbf{0}})) = 0$ and $\sum_{i=1}^{n} Var(X_{in}(\boldsymbol{\theta}_{\mathbf{0}})) = 1$. To prove that $\mathbf{D}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_o)\mathbf{S}_n(\boldsymbol{\theta}_{\mathbf{0}})$ converges to the standard normal distribution, it is enough to show that the sequence $\{X_{in}(\boldsymbol{\theta}_{\mathbf{0}})\}$ follows Lindberg condition , i.e. for every $\xi > 0$,

$$\sum_{i=1}^{n}\int\limits_{|x|\geq\xi} x^2 dP(X_{in}(\boldsymbol{\theta}_{\mathbf{0}}) \leq x) \to 0 \quad \text{as} \quad n \to \infty$$

We need Lemma 6.9, to prove the Lindberg condition. Based on Lemma 6.9,

$$\begin{aligned} E|X_{in}(\boldsymbol{\theta}_{\mathbf{0}})|^4 &\leq E|s_i(\boldsymbol{\theta}_{\mathbf{0}})|^4|\mathbf{u}^{\top}\mathbf{D}_n^{-1}(\boldsymbol{\theta}_o)\mathbf{u}|^2 \\ &\leq \frac{1}{n^2\eta^2}E|s_i(\boldsymbol{\theta}_{\mathbf{0}})|^4 \\ &\leq \frac{G^2}{n^2\eta^2} \end{aligned} \tag{35}$$

Let $I(\mathscr{A})$ denote the indicator variable on a set $\mathscr{A}$.

By (35), Cauchy Schwartz's inequality and Chebyshev's inequality we obtain,

$$
\begin{aligned}
\sum_{i=1}^{n} \int_{|x| \geq \xi} x^2 dP(X_{in}(\boldsymbol{\theta_0}) \leq x) &= \sum_{i-1}^{n} E\big\{X_{in}^2(\boldsymbol{\theta_0}) I(|X_{in}(\boldsymbol{\theta_0})| \geq \xi)\big\} \\
&\leq \frac{G}{n\eta} \sum_{i=1}^{n} \{P(|X_{in}(\boldsymbol{\theta_0})| \geq \xi)\}^{\frac{1}{2}} \\
&\leq \frac{G}{n\eta\xi} \sum_{i=1}^{n} \{Var(X_{in}(\boldsymbol{\theta_0}))\}^{\frac{1}{2}} \to 0 \quad \text{as} \quad n \to \infty
\end{aligned}
$$

∎

Proof of Theorem 4.3: Based on Lemma 6.3 and Lemma 6.10, the proof immediately follows. ∎

References

[1] Balakrishnan, N., & Pal, S., (2013), "COM -Poisson Cure Rate Models and Associated Likelihood - based Inference with Exponential and Weibull Lifetimes". *Applied Reliability Engineering and Risk Analysis: Probabilistic Models and Statistical Inference,* 308-348.

[2] Berkson, J., & Gage, R. P., (1952), "Survival curve for cancer patients following treatment". *Journal of the American Statistical Association*, vol 47(259), 501-515.

[3] Boag, J. W., (1949), "Maximum likelihood estimates of the proportion of patients cured by cancer therapy". *Journal of the Royal Statistical Society. Series B (Methodological)* , vol 11(1), 15-53.

[4] Casella, G., & Berger, R. L., (2002), "Statistical inference" *Pacific Grove, CA: Duxbury* Vol. 2.

[5] Crowder, M.(2001) Classical Competing Risks Model. Chapman & Hall/CRC, New York.

[6] Cox, D.R. (1959), "The analysis of exponentially distributed lifetimes with two types of failures", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol 21(2), 411 - 421.

[7] Dunsmuir, W., Tweedie, R., Flack, L., & Mengersen, K., (1989), "Modelling of transitions between employment states for young Australians". *Australian & New Zealand Journal of Statistics,* vol 31(1), 165-196.

[8] Farewell, V. T., (1982). "The use of mixture models for the analysis of survival data with long-term survivors". *Biometrics*, 1041-1046.

[9] Ghitany, M. E. , & Maller, R. A., (1992),"Asymptotic results for exponential mixture models with long-term survivors", *Statistics: A Journal of Theoretical and Applied Statistics* , vol 23.4 , 321-336.

[10] Ghitany, M. E., Maller, R. A., & Zhou, S., (1994), "Exponential mixture models with long-term survivors and covariates." *Journal of Multivariate Analysis* , vol 49.2, 218-241

[11] Gallardo, D. I., Gómez, Y. M., & de Castro, M., (2018), "A flexible cure rate model based on the polylogarithm distribution". *Journal of Statistical Computation and Simulation,* 1-13.

[12] Goldman, A. I. (1984). "Survivorship analysis when cure is a possibility: a Monte Carlo study". *Statistics in Medicine,* vol 3(2), 153-163.

[13] Greenhouse, J. B., & Wolfe, R. A., (1984), "A competing risks derivation of a mixture model for the analysis of survival data." *Communications in Statistics-Theory and Methods*, 13(25), 3133-3154.

[14] Gulland, J. A., (1955), "On the estimation of population parameters from marked members." *Biometrika,* vol 42(1-2), 269-270.

[15] Kannan, N., Kundu, D., Nair, P., & Tripathi, R. C., (2010), "The generalized exponential cure rate model with covariates" *Journal of Applied Statistics*, vol 37(10) 1625-1636.

[16] Kundu, D., (2004). "Parameter estimation for partially complete time and type of failure data". *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 46(2), pp.165-179.

[17] Lawless, J. F., (1982), *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

[18] Maltz, D. M. and McCleary, R., (1977), "The mathematics of behavioral change: recidivism and construct validity", *Evaluation Quarterly*, vol 1, 421-38.

[19] Maltz, D. M., (1984), *Recidivism*, Academic Press, Florida.

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

[21] Self, S. G., & Liang, K. L., (1987), "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions". *Journal of the American Statistical Association*, vol 82, 605-610.

[22] Wiangnak, P., & Pal, S., (2018), "Gamma lifetimes and associated inference for interval-censored cure rate model with COM –Poisson competing cause". *Communications in Statistics-Theory and Methods*, vol 47(6), 1491-1509.