Current status data with two competing risks and missing failure types: Time dependent missing probabilities

Technical Report No. ASU/2020/4
Dated 30 April, 2020

Tamalika Koley
and
Anup Dewanji

Applied Statistics Unit
Indian Statistical Institute
Kolkata 700 108

Current status data with two competing risks and missing failure types: Time dependent missing probabilities

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

Abstract

In competing risks data, in practice, there may be lack of information or uncertainty about the true failure type, termed as 'missing failure type', for some subjects. This type of uncertainty arises mainly in tumorigenicity data where it becomes difficult to determine the cause of death in presence of tumor and the true cause becomes missing in the above sense. In this work, we focus on both parametric and non-parametric estimation based on current status data with two competing risks and missing failure types. Here, the missing probabilities are assumed to be time-dependent, that is, dependent on both failure and monitoring time points. We carry out maximum likelihood estimation and obtain the asymptotic properties of the estimators in both the cases. Simulations studies are conducted to investigate the finite sample properties of the estimators. Finally, the methods are illustrated through a hearing loss data set.

Keywords— Monitoring time, Masking probabilities, Maximum likelihood estimation, Identifiability, Interval hazards.

1 Introduction

Current status data is a special form of interval-censored data where the failure time is not observed exactly but is known to lie in a left-bounded or a right-bounded interval. In this situation, for each individual under study, the status (failure/survival) of the event of interest at the monitoring time is observed. If an individual experiences failure, then it is exposed to the risk of more than one causes called competing risks (See Kalbfleisch and Prentice (2002)). For example, a study examining time to death attributable to cardiovascular causes, or non-cardiovascular causes, consists of two competing risks (See Austin et al. (2016)). Several works on current status data with competing risks are present in the literature. A great deal of attention is focused on both parametric and non-parametric estimation of the sub-distribution functions. Some of them are Hudgens et al. (2001), Jewell et al. (2003), Jewell and Kalbfleisch (2004), Maathuis (2006), Maathuis and Hudgens (2011).

Missing failure type is a common phenomenon in competing risks setup which, although studied in the analysis of right censored survival data with competing risks, has not received much attention in the context of current status data with competing risks. In this work, we consider a general missing pattern in which, if a failure type is not observed, one observes a set of all possible types containing the true one. This type of missing pattern was studied by Flehinger et al. (1998) for estimating cause specific survival functions based on right censored survival data with competing risks under the assumption of proportional hazards due to different types. Dewanji and Sengupta (2003) also considered such missing failure type in right censored survival data with competing risks and proposed a Nelson-Aalen type estimator when additional information on the missing probabilities is available. Mukhopadhyay (2006) considered such missing type on the identity of the failed component in the context of lifetime data from series systems with independent components. Parametric and non-parametric methods of estimation based on current status data with competing risks under the above-mentioned missing pattern and two competing risks were studied by Koley and Dewanji (2018a) and Koley and Dewanji (2019), respectively. In their work, they made a simple assumption on the missing probabilities to be independent of the monitoring time and the failure time.

In this work, we consider current status data with competing risks subject to the above-mentioned missing pattern in failure types and assume that the missing probabilities depend on the difference between the monitoring and the failure time. Both parametric and non-parametric methods of estimation are considered in this work. In Section 2, the data is described along with the construction of the corresponding likelihood. In Section 3, we develop the parametric method for maximum likelihood estimation, while Section 4 considers non-parametric estimation. Section 5 presents a detailed simulation study to investigate the finite sample properties of the estimators. We demonstrate the proposed methods through the analysis of a hearing loss data in Section 6 and finally Section 7 ends with some concluding remarks.

2 The Data and the likelihood

Let T denote the random variable representing the failure time subject to $m(=2)$ competing risks. Let $J \in \{1,2\}$ be the true failure type and $G \in \{\{1\},\{2\},\{1,2\}\}$ the observed set of possible failure types containing the true one. Also, define X as the monitoring time random variable which is assumed to be independent of the random vector (T,J). It is assumed that the monitoring time distribution does not involve any common parameter with the failure time distribution and we are only interested in the set of parameters related to the failure time distribution. If $T \le X$, we observe the set G instead of the true failure type J and, if $T > X$, G is written as the empty set ϕ. Therefore, the support of G, denoted by $\mathcal{G}$, is given by $\{\{1\},\{2\},\{1,2\},\phi\}$. Thus, for an individual under study, we observe (x,g), realization of the random vector (X,G). Let us write the conditional probability of observing g as the possible set of failure types, given the true type $J=j$ and all other related information, by

$$p_{gj}(x,t) = P\left[G = g \mid t = T < X = x, J = j\right],$$

for $t \le x$ and $g \ni j$. If $j \notin g$, this probability is equal to 0, since it is assumed that there is no misspecification in the data in the sense that the observed set of possible types always contains the true one for an individual experiencing the failure. These conditional probabilities are termed as masking probabilities (See Basu (2009)) since the true failure type is masked into a subset of all possible failure types. Note that, if g is a singleton set $\{j\}$, then we observe the true type j with certainty, for $j = 1,2$, and the other case $g = \{1,2\}$ represents complete missing. Also, for $T \le X = x$ (so $g \ne \phi$),

$$\begin{aligned} P\left[T \le x, G = g\right] &= \sum_{j \in g} P\left[T \le x, G = g, J = j\right] \\ &= \left[\sum_{j \in g} \int_0^x p_{gj}(x,t) f_j(t;\underset{\sim}{\theta})\mathrm{d}t\right] h(x), \end{aligned} \tag{2.1}$$

where $f_j(\cdot;\underset{\sim}{\theta})$ is the sub-density function for failure type j with $\underset{\sim}{\theta}$ being the vector of parameters associated with the random variable T and $h(x)$ the density function of X with support $DomX$, say. Let us denote the density function of the observed random vector

(X, G) as $f^*(\cdot, \cdot)$. Then, for $T \leq X = x,\ g = \{1\}$, we have, using (2.1),

$$f^*(x, g = \{1\}) = \left[\int_0^x p_{\{1\}1}(x, t) f_1(t; \underset{\sim}{\theta}) \mathrm{d}t\right] h(x).$$

Similarly, for $T \leq X = x,\ g = \{2\}$, we have

$$f^*(x, g = \{2\}) = \left[\int_0^x p_{\{2\}2}(x, t) f_2(t; \underset{\sim}{\theta}) \mathrm{d}t\right] h(x).$$

Now, for $T \leq X = x,\ g = \{1, 2\}$, we have, using similar arguments,

$$f^*(x, g = \{1, 2\}) = \left[\sum_{j=1}^{2} \int_0^x p_{\{1,2\}j}(x, t) f_j(t; \underset{\sim}{\theta}) \mathrm{d}t\right] h(x).$$

Finally, for $T > X = x, g = \phi$, we have

$$f^*(x, g = \phi) = S(x; \underset{\sim}{\theta}) h(x),$$

where $S(x; \underset{\sim}{\theta})$ is the survival function of T at x given by $\int_x^\infty f(t; \underset{\sim}{\theta}) dt$ with $f(t; \underset{\sim}{\theta}) = \sum_{j=1}^{2} f_j(t; \underset{\sim}{\theta})$ being the density of T. Then, the density $f^*(x, g)$ is given by

$$f^*(x, g) = \begin{cases} \left[\int_0^x p_{\{j\}j}(x, t) f_j(t; \underset{\sim}{\theta}) \mathrm{d}t\right] h(x) & \text{if } g = \{j\}, \text{ for } j = 1, 2 \\ \left[\sum_{j=1}^{2} \int_0^x p_{\{1,2\}j}(x, t) f_j(t; \underset{\sim}{\theta}) \mathrm{d}t\right] h(x) & \text{if } g = \{1, 2\}, \\ \left[\int_x^\infty \sum_{j=1}^{2} f_j(t; \underset{\sim}{\theta}) \mathrm{d}t\right] h(x) & \text{if } g = \{\phi\}, \end{cases} \tag{2.2}$$

with respect to the dominating measure $H \times \mu$, where $H(\cdot)$ and μ are, respectively, the distribution function of X and the counting measure. Note that the data consists of n iid realizations, $\{(x_i, g_i); i = 1, \cdots, n\}$, of the random vector (X, G) from the common density given by $f^*(x, g)$. Then, the likelihood of the data can be written as the product of these densities from (2.2) and is given by $\prod_{i=1}^n f^*(x_i, g_i)$.

3 Parametric estimation

In this section, we assume parametric forms of the sub-density functions $f_j(\cdot; \underset{\sim}{\theta})$, for $j = 1, 2$, involving the unknown parameter(s) in $\underset{\sim}{\theta}$, and consider estimation of $\underset{\sim}{\theta}$ in addition to the parameter(s) associated with the masking probabilities. Two following forms of masking probabilities having different interpretations are considered for analysis.

3.1 Model 1

Let us assume that, when the true failure type is $J = j$ for a failure at $T = t$ and observed at a later time $X = x$, the probability of observing the true type j is given by

$$p_{\{j\}j}(x, t) = e^{-\alpha_j(x-t)}, \quad \text{with } \alpha_j > 0, \text{ for } j = 1, 2. \tag{M1}$$

This assumption is reasonable in the sense that the probability of making the correct diagnosis regarding the failure type reduces with time since failure, depending on the difference $(x - t)$ between the time of failure and the time of monitoring. It conforms with the usual notion that proper diagnosis becomes more difficult as it is carried out further away from the failure time. This, however, implies $p_{\{j\}j}(x = t, t) = 1$; that is, the probability of correct diagnosis is 1 if the failure time and monitoring time coincide which may be somewhat unreasonable. But, in a continuous parametric model, the probability of $X = T$ is 0. Nevertheless, as the failure time lies further from the monitoring time, the probability of correct diagnosis, $p_{\{j\}j}(x, t)$, reduces exponentially from 1 with unknown rate α_j. As in one of our previous work (See Koley and Dewanji (2018a)), we shall require some assumptions on α_1 and α_2 to achieve model identifiability.

Theorem 3.1. *The model $f^*(x, g)$ for $(x, g) \in$ DomX $\times \mathcal{G}$, given by* (2.2) *and* (M1)*, is identifiable under the assumptions $\alpha_2 = c\alpha_1$, for some known constant $c > 0$.*

Proof. Note that the vector of unknown parameters is $(\underset{\sim}{\theta}, \alpha_1)$. Let us first write the model $f^*(x, g)$ as $f^*(x, g; \underset{\sim}{\theta}, \alpha_1)$ to make the dependence on the parameter vector explicit. To check the identifiability of the model parameters, let $(\underset{\sim}{\theta}^{(1)}, \alpha_1^{(1)})$ and $(\underset{\sim}{\theta}^{(2)}, \alpha_1^{(2)})$ be two different choices of the parameter vector $(\underset{\sim}{\theta}, \alpha_1)$ and consider the identity

$$f^*(x, g; \underset{\sim}{\theta}^{(1)}, \alpha_1^{(1)}) = f^*(x, g; \underset{\sim}{\theta}^{(2)}, \alpha_1^{(2)}), \tag{3.1}$$

for all (x, g). Putting $g = \{1\}, \{2\}$ and ϕ on both sides of (3.1), we get

$$\int_0^x e^{-\alpha_1^{(1)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(1)}) \mathrm{d}t = \int_0^x e^{-\alpha_1^{(2)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(2)}) \mathrm{d}t,$$

$$\int_0^x e^{-c\alpha_1^{(1)}(x-t)} f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t = \int_0^x e^{-c\alpha_1^{(2)}(x-t)} f_2(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t$$

and

$$S(x; \underset{\sim}{\theta}^{(1)}) = S(x; \underset{\sim}{\theta}^{(2)}), \tag{3.2}$$

for all $x \in$Dom$X$. The first two equations give

$$\frac{\int_0^x e^{-\alpha_1^{(1)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t}{\int_0^x e^{-c\alpha_1^{(1)}(x-t)} f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t} = \frac{\int_0^x e^{-\alpha_1^{(2)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t}{\int_0^x e^{-c\alpha_1^{(2)}(x-t)} f_2(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t} = k(x), \quad \text{say,}$$

for all $x \in$Dom$X$. Clearly, $k(x) \geq 0$ for all x and positive for at least one x and is possibly a function of $\alpha_1^{(1)}, \alpha_1^{(2)}, \underset{\sim}{\theta}^{(1)}$ and $\underset{\sim}{\theta}^{(2)}$. This gives

$$\int_0^x e^{-\alpha_1^{(1)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t = k(x)\int_0^x e^{-c\alpha_1^{(1)}(x-t)} f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t \tag{3.3}$$

and

$$\int_0^x e^{-\alpha_1^{(2)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t = k(x)\int_0^x e^{-c\alpha_1^{(2)}(x-t)} f_2(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t, \tag{3.4}$$

for all $x \in$Dom$X$. From (3.3) and for all $x \in$Dom$X$,

$$\sum_{l=0}^{\infty}\left[(-1)^l \int_0^x f_1(t; \underset{\sim}{\theta}^{(1)})\frac{(x-t)^l}{l!}\mathrm{d}t\right](\alpha_1^{(1)})^l = k(x)\sum_{l=0}^{\infty}\left[(-1)^l \int_0^x f_2(t; \underset{\sim}{\theta}^{(1)})\frac{c^l(x-t)^l}{l!}\mathrm{d}t\right](\alpha_1^{(1)})^l$$

$$\iff \sum_{l=0}^{\infty} a_l(\alpha_1^{(1)})^l = \sum_{l=0}^{\infty} b_l(\alpha_1^{(1)})^l, \quad \text{say,} \tag{3.5}$$

where $a_l = (-1)^l \int_0^x f_1(t; \underset{\sim}{\theta}^{(1)})\frac{(x-t)^l}{l!}\mathrm{d}t$ and $b_l = k(x)(-1)^l \int_0^x f_2(t; \underset{\sim}{\theta}^{(1)})\frac{c^l(x-t)^l}{l!}\mathrm{d}t$, for $l = 0, 1, \cdots$. Both sides of (3.5) are power series in $\alpha_1^{(1)}$ and, hence, using the property of power series expansion (See Shafarevich, 2012, p234), $a_l = b_l$, for $l = 0, 1, \cdots$. In particular, for $l = 0$, $a_0 = b_0$ and we have

$$\int_0^x f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t = k(x)\int_0^x f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t,$$

for all $x \in$Dom$X$. That is,

$$F_1(x; \underset{\sim}{\theta}^{(1)}) = k(x)F_2(x; \underset{\sim}{\theta}^{(1)}),$$

for all $x \in$Dom$X$. Similarly, from (3.4) and proceeding in the similar way

$$F_1(x; \underset{\sim}{\theta}^{(2)}) = k(x)F_2(x; \underset{\sim}{\theta}^{(2)}),$$

for all $x \in$Dom$X$. Then, using (3.2) corresponding to $g = \phi$,

$$(1 + k(x))F_2(x; \underset{\sim}{\theta}^{(1)}) = (1 + k(x))F_2(x; \underset{\sim}{\theta}^{(2)}),$$

for all $x \in$Dom$X$. This implies $F_2(x; \underset{\sim}{\theta}^{(1)}) = F_2(x; \underset{\sim}{\theta}^{(2)})$, for all $x \in$Dom$X$. Hence, using (3.2), $F_1(x; \underset{\sim}{\theta}^{(1)}) = F_1(x; \underset{\sim}{\theta}^{(2)})$, for all $x \in$Dom$X$. This implies $\underset{\sim}{\theta}^{(1)} = \underset{\sim}{\theta}^{(2)}$. Now differentiating both sides of $\int_0^x e^{-\alpha_1^{(1)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t = \int_0^x e^{-\alpha_1^{(2)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t$ with respect to x, we have

$$f_1(x; \underset{\sim}{\theta}^{(1)}) - \alpha_1^{(1)} \int_0^x e^{-\alpha_1^{(1)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t = f_1(x; \underset{\sim}{\theta}^{(2)}) - \alpha_1^{(2)} \int_0^x e^{-\alpha_1^{(2)}(x-t)} f_1(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t.$$

This, along with $\underset{\sim}{\theta}^{(1)} = \underset{\sim}{\theta}^{(2)}$, gives $\alpha_1^{(1)} = \alpha_1^{(2)}$. Hence, the model is identifiable. □

3.2 Model 2

Let us consider a different form of the masking probabilities as given by

$$p_{\{j\}j}(x, t) = e^{-\{\alpha_j + \beta(x-t)\}}, \tag{M2}$$

where $\alpha_j > 0$, for $j = 1, 2$, and $\beta > 0$ is a known constant. In this case, the probability of correct diagnosis at $x = t$ is $e^{-\alpha_j}$, an unknown quantity, but reduces exponentially with a known rate, unlike that in Model 1, as failure time lies further away from the monitoring time. In the next theorem, we will check for the identifiability of the model.

Theorem 3.2. *The model $f^*(x, g)$ for $(x, g) \in$DomX $\times \mathcal{G}$, given by* (2.2) *and* (M2)*, is not identifiable.*

Proof. Suppose $(\underset{\sim}{\theta}^{(1)}, \alpha_1^{(1)}, \alpha_2^{(1)})$ and $(\underset{\sim}{\theta}^{(2)}, \alpha_1^{(2)}, \alpha_2^{(2)})$ are two values of the parameter vector $(\underset{\sim}{\theta}, \alpha_1, \alpha_2)$ and consider the identity

$$f^*(x, g; \underset{\sim}{\theta}^{(1)}, \alpha_1^{(1)}, \alpha_2^{(1)}) = f^*(x, g; \underset{\sim}{\theta}^{(2)}, \alpha_1^{(2)}, \alpha_2^{(2)}), \quad \text{for all} \quad (x, g). \tag{3.6}$$

The model is identifiable iff the identity (3.6) implies $(\underset{\sim}{\theta}^{(1)}, \alpha_1^{(1)}, \alpha_2^{(1)}) = (\underset{\sim}{\theta}^{(2)}, \alpha_1^{(2)}, \alpha_2^{(2)})$.

Putting $g = \{1\}, \{2\}$ and ϕ on both sides of (3.6), we have

$$\int_0^x e^{-\{\alpha_1^{(1)}+\beta(x-t)\}} f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t = \int_0^x e^{-\{\alpha_1^{(2)}+\beta(x-t)\}} f_1(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t,$$

$$\int_0^x e^{-\{\alpha_2^{(1)}+\beta(x-t)\}} f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t = \int_0^x e^{-\{\alpha_2^{(2)}+\beta(x-t)\}} f_2(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t,$$

$$\text{and} \quad S(x; \underset{\sim}{\theta}^{(1)}) = S(x; \underset{\sim}{\theta}^{(2)}), \tag{3.7}$$

for all $x \in$Dom$X$. From the first two equations, we have

$$\frac{\int_0^x e^{-\{\alpha_1^{(1)}+\beta(x-t)\}} f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t}{\int_0^x e^{-\{\alpha_2^{(1)}+\beta(x-t)\}} f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t} = \frac{\int_0^x e^{-\{\alpha_1^{(2)}+\beta(x-t)\}} f_1(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t}{\int_0^x e^{-\{\alpha_2^{(2)}+\beta(x-t)\}} f_2(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t} = K(x), \quad \text{say}, \tag{3.8}$$

where $K(x) \geq 0$ for all x and positive for at least one x, possibly depending on $\underset{\sim}{\theta}^{(1)}, \underset{\sim}{\theta}^{(2)}, \alpha_1^{(1)}, \alpha_1^{(2)}, \alpha_2^{(1)}$ and $\alpha_2^{(2)}$. This gives

$$\int_0^x e^{-\{\alpha_1^{(1)}+\beta(x-t)\}} f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t = K(x)\int_0^x e^{-\{\alpha_2^{(1)}+\beta(x-t)\}} f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t \tag{3.9}$$

$$\int_0^x e^{-\{\alpha_1^{(2)}+\beta(x-t)\}} f_1(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t = K(x)\int_0^x e^{-\{\alpha_2^{(2)}+\beta(x-t)\}} f_2(t; \underset{\sim}{\theta}^{(2)})\mathrm{d}t, \tag{3.10}$$

for all x. From (3.9), we have

$$e^{-\alpha_1^{(1)}} \sum_{l=0}^{\infty} \left[(-1)^l \int_0^x \frac{(x-t)^l}{l!} f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t\right]\beta^l = K(x)e^{-\alpha_2^{(1)}} \sum_{l=0}^{\infty} \left[(-1)^l \int_0^x \frac{(x-t)^l}{l!} f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t\right]\beta^l$$

$$\Longleftrightarrow \sum_{l=0}^{\infty} a_l \beta^l = \sum_{l=0}^{\infty} b_l \beta^l,$$

where $a_l = e^{-\alpha_1^{(1)}}(-1)^l \int_0^x \frac{(x-t)^l}{l!} f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t$ and $b_l = K(x)e^{-\alpha_2^{(1)}}(-1)^l \int_0^x \frac{(x-t)^l}{l!} f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t$. Both sides of the above equation are power series in β and hence, $a_l = b_l$, for $l = 0, 1, \cdots$. Taking $l = 0$, we have

$$e^{-\alpha_1^{(1)}} \int_0^x f_1(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t = K(x)e^{-\alpha_2^{(1)}} \int_0^x f_2(t; \underset{\sim}{\theta}^{(1)})\mathrm{d}t$$

$$\Longleftrightarrow e^{-\alpha_1^{(1)}} F_1(x; \underset{\sim}{\theta}^{(1)}) = K(x)e^{-\alpha_2^{(1)}} F_2(x; \underset{\sim}{\theta}^{(1)})$$

$$\Longleftrightarrow F_1(x; \underset{\sim}{\theta}^{(1)}) = K(x)e^{\alpha_1^{(1)}-\alpha_2^{(1)}} F_2(x; \underset{\sim}{\theta}^{(1)}),$$

for all x. Similarly, from (3.10), we get $F_1(x;\underset{\sim}{\theta}^{(2)}) = K(x)e^{\alpha_1^{(2)}-\alpha_2^{(2)}}F_2(x;\underset{\sim}{\theta}^{(2)})$, for all x. Using these expressions in (3.7), we have

$$(1+K(x)e^{\alpha_1^{(1)}-\alpha_2^{(1)}})F_2(x;\underset{\sim}{\theta}^{(1)}) = (1+K(x)e^{\alpha_1^{(2)}-\alpha_2^{(2)}})F_2(x;\underset{\sim}{\theta}^{(2)})$$

$$\iff K(x) = \frac{F_2(x;\underset{\sim}{\theta}^{(2)}) - F_2(x;\underset{\sim}{\theta}^{(1)})}{e^{\alpha_1^{(1)}-\alpha_2^{(1)}}F_2(x;\underset{\sim}{\theta}^{(1)}) - e^{\alpha_1^{(2)}-\alpha_2^{(2)}}F_2(x;\underset{\sim}{\theta}^{(2)})},$$

for all x. If the model is identifiable, then the above expression of $K(x)$ is of the form $0/0$ which contradicts (3.8) that shows $K(x) > 0$, for at least one x. Hence, the model is not identifiable. □

Hence, we make some additional assumption on α_1 and α_2 to achieve identifiability. The following theorem assumes certain conditions on α_1 and α_2 and check for model identifiability.

Theorem 3.3. *If α_1 and α_2 satisfy (i) $\alpha_2 = \alpha_1 + f$ or (ii) $\alpha_2 = c\,\alpha_1$, where both f and c are known constants and $c > 0$, then the above model is identifiable.*

Proof. (i) To check the identifiability of the model, let $(\underset{\sim}{\theta}^{(1)}, \alpha_1^{(1)})$ and $(\underset{\sim}{\theta}^{(2)}, \alpha_1^{(2)})$ be two values of the parameter vector $(\underset{\sim}{\theta}, \alpha_1)$ and consider the identity (3.6) with $\alpha_2 = \alpha_1 + f$. Proceeding in the similar way as in the previous theorem with α_2 being replaced by $\alpha_1 + f$, we have from (3.9) and (3.10),

$$F_1(x;\underset{\sim}{\theta}^{(1)}) = K(x)e^{-f}F_2(x;\underset{\sim}{\theta}^{(1)}) \quad \text{and} \quad F_1(x;\underset{\sim}{\theta}^{(2)}) = K(x)e^{-f}F_2(x;\underset{\sim}{\theta}^{(2)}),$$

respectively, for all $x \in$Dom$X$. Finally, from (3.7), we get

$$(1+K(x)e^{-f})F_2(x;\underset{\sim}{\theta}^{(1)}) = (1+K(x)e^{-f})F_2(x;\underset{\sim}{\theta}^{(2)}), \quad \text{for all} \quad x \in \text{Dom}X.$$

$$\iff F_2(x;\underset{\sim}{\theta}^{(1)}) = F_2(x;\underset{\sim}{\theta}^{(2)}), \quad \text{for all} \quad x \in \text{Dom}X.$$

This is because $e^{-f} > 0$ and $K(x) > 0$, for at least one x. Using this in (3.7), we get $F_1(x;\underset{\sim}{\theta}^{(1)}) = F_1(x;\underset{\sim}{\theta}^{(2)})$, which gives $\underset{\sim}{\theta}^{(1)} = \underset{\sim}{\theta}^{(2)}$. Putting $g = \{1\}$ on both sides of (3.6) with $\underset{\sim}{\theta}^{(1)} = \underset{\sim}{\theta}^{(2)} = \underset{\sim}{\theta}$, say, we have

$$\int_0^x e^{-\{\alpha_1^{(1)}+\beta(x-t)\}}f_1(t;\underset{\sim}{\theta})\mathrm{d}t = \int_0^x e^{-\{\alpha_1^{(2)}+\beta(x-t)\}}f_1(t;\underset{\sim}{\theta})\mathrm{d}t, \quad \text{for all } x.$$

$$\iff e^{-\alpha_1^{(1)}}\int_0^x e^{-\beta(x-t)}f_1(t;\underset{\sim}{\theta})\mathrm{d}t = e^{-\alpha_1^{(2)}}\int_0^x e^{-\beta(x-t)}f_1(t;\underset{\sim}{\theta})\mathrm{d}t, \quad \text{for all } x.$$

Hence, $\alpha_1^{(1)} = \alpha_1^{(2)}$, since the integral on both sides are equal and positive, for all $x \in \text{Dom}X$. Thus, the model is identifiable. Clearly, the constant f needs to satisfy $\alpha_2 = \alpha_1 + f > 0$, or $f > -\alpha_1$.

(ii) In the second case, $\alpha_2 = c\ \alpha_1$, where $c > 0$ is a known constant. Let $(\underset{\sim}{\theta}^{(1)}, \alpha_1^{(1)})$ and $(\underset{\sim}{\theta}^{(2)}, \alpha_1^{(2)})$ be two values of the parameter vector $(\underset{\sim}{\theta}, \alpha_1)$ and consider the identity (3.6). Here, also proceeding in the similar manner as in the previous theorem with $\alpha_2 = c\ \alpha_1$, we have, from (3.9),

$$e^{-\alpha_1^{(1)}} F_1(x; \underset{\sim}{\theta}^{(1)}) = K(x) e^{-c\ \alpha_1^{(1)}} F_2(x; \underset{\sim}{\theta}^{(1)}).$$

$$\Longleftrightarrow \sum_{l=0}^{\infty} \left[(-1)^l F_1(x; \underset{\sim}{\theta}^{(1)}) \right] (\alpha_1^{(1)})^l = \sum_{l=0}^{\infty} \left[(-c)^l K(x) F_2(x; \underset{\sim}{\theta}^{(1)}) \right] (\alpha_1^{(1)})^l.$$

Both sides of the above equation are power series in $\alpha_1^{(1)}$ and, hence, again using the property of power series for $l = 0$, we have

$$F_1(x; \underset{\sim}{\theta}^{(1)}) = K(x) F_2(x; \underset{\sim}{\theta}^{(1)}), \quad \text{for all } x.$$

Similarly, from (3.10), we get $F_1(x; \underset{\sim}{\theta}^{(2)}) = K(x) F_2(x; \underset{\sim}{\theta}^{(2)})$, for all x. Using these two equations in (3.7), we get

$$(1 + K(x)) F_2(x; \underset{\sim}{\theta}^{(1)}) = (1 + K(x)) F_2(x; \underset{\sim}{\theta}^{(2)}), \quad \text{for all } x.$$

$\Longleftrightarrow F_2(x; \underset{\sim}{\theta}^{(1)}) = F_2(x; \underset{\sim}{\theta}^{(2)})$, since $K(x) > 0$ for at least one x. Using this in (3.7), we have $F_1(x; \underset{\sim}{\theta}^{(1)}) = F_1(x; \underset{\sim}{\theta}^{(2)})$, which implies $\underset{\sim}{\theta}^{(1)} = \underset{\sim}{\theta}^{(2)} = \underset{\sim}{\theta}$, say.

Putting $g = \{1\}$ on both sides of the identity (3.6) we get

$$e^{-\alpha_1^{(1)}} \int_0^x e^{-\beta(x-t)} f_1(t; \underset{\sim}{\theta}) \mathrm{d}t = e^{-\alpha_1^{(2)}} \int_0^x e^{-\beta(x-t)} f_1(t; \underset{\sim}{\theta}) \mathrm{d}t.$$

This gives $\alpha_1^{(1)} = \alpha_1^{(2)}$, since $\int_0^x e^{-\beta(x-t)} f_1(t; \underset{\sim}{\theta}) \mathrm{d}t > 0$, for some $x \in \text{Dom}X$. Thus, the model is identifiable.

□

3.3 Maximum Likelihood Estimation

For a given $g \in \mathcal{G} \setminus \phi$, define the indicator variable $\delta_{gi} = \mathbb{1}\left[T \le X = x_i, G = g\right]$, and $\delta_i = \sum_{\mathcal{G} \setminus \phi} \delta_{gi}$, for $i = 1, \cdots, n$, giving the total number of failures observed at x_i Then the likelihood of the data under (M1) is given by

$$L_I(\underset{\sim}{\theta},\alpha_1) \propto \prod_{i=1}^{n}\Bigg[\left\{\int_0^{x_i} f_1(t;\underset{\sim}{\theta})e^{-\alpha_1(x_i-t)}\mathrm{dt}\right\}^{\delta_{\{1\}i}}\left\{\int_0^{x_i} f_2(t;\underset{\sim}{\theta})e^{-\alpha_2(x_i-t)}\mathrm{dt}\right\}^{\delta_{\{2\}i}}$$
$$\times\left\{\int_0^{x_i}\{(1-e^{-\alpha_1(x_i-t)})f_1(t;\underset{\sim}{\theta})+(1-e^{-\alpha_2(x_i-t)})f_2(t;\underset{\sim}{\theta})\}\mathrm{dt}\right\}^{\delta_{\{1,2\}i}}$$
$$\times S(x_i;\underset{\sim}{\theta})^{1-\delta_i}\Bigg], \tag{3.11}$$

where $\alpha_2 = c\ \alpha_1$ with known $c > 0$. Similarly, the likelihood of the data under (M2) is given by

$$L_{II}(\underset{\sim}{\theta},\alpha_1) \propto \prod_{i=1}^{n}\Bigg[\left\{\int_0^{x_i} f_1(t;\underset{\sim}{\theta})e^{-\alpha_1+\beta(x_i-t)}\mathrm{dt}\right\}^{\delta_{\{1\}i}}\left\{\int_0^{x_i} f_2(t;\underset{\sim}{\theta})e^{-\alpha_2+\beta(x_i-t)}\mathrm{dt}\right\}^{\delta_{\{2\}i}}$$
$$\times\left\{\int_0^{x_i}\{(1-e^{-\alpha_1+\beta(x_i-t)})f_1(t;\underset{\sim}{\theta})+(1-e^{-\alpha_2+\beta(x_i-t)})f_2(t;\underset{\sim}{\theta})\}\mathrm{dt}\right\}^{\delta_{\{1,2\}i}}$$
$$\times S(x_i;\underset{\sim}{\theta})^{1-\delta_i}\Bigg], \tag{3.12}$$

where α_2 is either α_1+f or $c\,\alpha_1$. It is clear from both the likelihood functions that there is no closed form for the maximum likelihood estimators (MLEs) of the model parameters $(\underset{\sim}{\theta},\alpha_1)$. So numerical maximization method is used for estimation of the identifiable parameters. The *optim* function in R software is used to serve the purpose of maximizing the likelihood. In the following, we study the asymptotic results for the MLEs.

Suppose the vector of model parameters is denoted by $\underset{\sim}{\psi}$; that is, $\underset{\sim}{\psi} = (\underset{\sim}{\theta},\alpha_1)$. Let us assume that the failure time distribution satisfies the regularity conditions (See Lehmann and Casella 1998, p449) and all third order partial derivatives of $f_j(\cdot;\underset{\sim}{\theta})$, for $j = 1,2$, exist and are continuous functions of $\underset{\sim}{\theta}$. Also, as noted in Section 2, the data (x_i, g_i), for $i = 1,\cdots,n$, are n iid observations from the common density function $f^*(x,g;\underset{\sim}{\psi})$ and the likelihood function is the product of these density functions corresponding to these observations. Consider the following conditions.

C1 The identifiability of the parameter vector $\underset{\sim}{\psi}$ with respect to the density function $f^*(x,g;\underset{\sim}{\psi})$ follows clearly from Theorems 3.1 and 3.3 in Section 3.

C2 The parameter vector $\underset{\sim}{\psi} = (\underset{\sim}{\theta},\alpha_1)$ is an open set since $\underset{\sim}{\theta}$ is open as the failure time distribution satisfies the regularity conditions and $\alpha_1 > 0$.

C3 The support of the density $f^*(x,g;\underset{\sim}{\psi})$, Dom$X \times \mathcal{G}$, is independent of the parameter vector $\underset{\sim}{\psi}$.

C4 As assumed, the third order partial derivatives of $f_j(\cdot;\underset{\sim}{\theta})$, for $j = 1, 2$, exist and are continuous; also, the masking probabilities $e^{-\alpha_j(x-t)}$ or $e^{-\{\alpha_j+\beta(x-t)\}}$ are continuous functions of α_1 admitting continuous derivatives. Hence, the third order partial derivatives of $f^*(x, g; \underset{\sim}{\psi})$ with respect to $\underset{\sim}{\psi}$ exist and are continuous.

C5 It is assumed that there exists an open neighbourhood of the true value $\underset{\sim}{\psi}_0$ of $\underset{\sim}{\psi}$ within which the third order partial derivatives of $f^*(x, g; \underset{\sim}{\psi})$ are bounded by functions of (x, g) with finite expectations.

C6 The Fisher information matrix is assumed to be positive definite.

Theorem 3.4. *The MLE of the parameter vector $\underset{\sim}{\psi}$, denoted by $\hat{\underset{\sim}{\psi}}$, satisfies*

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

(ii) $\sqrt{n}(\hat{\underset{\sim}{\psi}} - \underset{\sim}{\psi}_0)$ *is asymptotically a mean zero normal random vector with variance covariance matrix estimated by the inverse of the hessian matrix computed at the MLE.*

Proof. The proof directly follows from Theorem 7.5.2 of Lehmann and Casella (1998, p463-465) using the conditions **C1-C6**. □

4 Non-parametric Estimation

In the absence of any modeling assumption regarding the functional forms of the sub-density functions, we consider non-parametric estimation of the same based on current status data with missing failure types. Let us write $F_j(t) = P[T \leq t, J = j]$ as the sub-distribution function for the failure of type j, for $j = 1, 2$. Also, let $\mathcal{F}$ be the collection of all such $\underset{\sim}{F} = (F_1(\cdot), F_2(\cdot))$ with $F_1(\cdot) + F_2(\cdot) \leq 1$. Here we assume the monitoring times to be fixed. Let $0 = \tau_0 < \tau_1 < \cdots < \tau_k$ be the k fixed monitoring time points and n_i individuals are observed at τ_i, for $i = 1, \cdots, k$. Suppose d_{gi} individuals are observed at time τ_i to have failed with g as the observed set of possible causes and $n_i - d_i$ are censored at τ_i with $d_i = \sum_{g\in\mathcal{G}/\phi} d_{gi}$, for $i = 1, \cdots, k$.

The likelihood function of this data under the model $\underset{\sim}{F} = (F_1(\cdot), F_2(\cdot))$ is given by

$$L(\underset{\sim}{F}) = \prod_{i=1}^{k} \Bigg[\prod_{g\in G\setminus\phi} \Bigg(\sum_{j\in g} \int_0^{\tau_i} p_{gj}(\tau_i, u) f_j(u) \mathrm{d}u \Bigg)^{d_{gi}} \times \Bigg(S(\tau_i) \Bigg)^{n_i - d_i} \Bigg], \tag{4.1}$$

where $S(\cdot) = 1 - F_1(\cdot) - F_2(\cdot)$ and $f_j(\cdot)$, for $j = 1, 2$, are the sub-densities corresponding to $F_1(\cdot)$ and $F_2(\cdot)$, respectively. Although $\underset{\sim}{F} \in \mathcal{F}$ is assumed to be arbitrary, we make some modeling assumptions on the masking probabilities, $p_{gj}(x,t)$'s. Unlike the modeling of the $p_{gj}(x,t)$'s in the parametric estimation case, here we assume $p_{gj}(\tau_i,u)$'s to be piecewise constant in u within the intervals $I_l = (\tau_{l-1},\tau_l]$, for $l = 1,\cdots,i$, with $\tau_0 = 0$, to facilitate non-parametric estimation. Let us write $p_{\{j\}j}(\tau_i,u) = p_j(\tau_i,\tau_l)$, if $u \in I_l$, for $l \le i$ and $j = 1,2$. Then, it is easily seen that

$$\begin{aligned}\int_0^{\tau_i} p_{\{j\}j}(\tau_i,u)f_j(u)\mathrm{d}u &= \sum_{l=1}^{i} p_j(\tau_i,\tau_l)\int_{\tau_{l-1}}^{\tau_l} f_j(u)\mathrm{d}u \\ &= \sum_{l=1}^{i} p_j(\tau_i,\tau_l)P\left[T \in I_l, J = j\right] \qquad (4.2)\end{aligned}$$

and, similarly,

$$\int_0^{\tau_i} p_{\{1,2\}j}(\tau_i,u)f_j(u)\mathrm{d}u = \sum_{l=1}^{i}(1 - p_j(\tau_i,\tau_l))P\left[T \in I_l, J = j\right], \qquad (4.3)$$

for $j = 1,2$. Hence, the likelihood function (4.1) can be written, using (4.2) and (4.3), as

$$\begin{aligned}L(\underset{\sim}{F}) = \prod_{i=1}^{k}\Bigg[&\left(\sum_{l=1}^{i} p_1(\tau_i,\tau_l)P\left[T \in I_l, J = 1\right]\right)^{d_{\{1\}i}}\left(\sum_{l=1}^{i} p_2(\tau_i,\tau_l)P\left[T \in I_l, J = 2\right]\right)^{d_{\{2\}i}} \\ &\times\left(\sum_{l=1}^{i}\sum_{j=1}^{2}(1 - p_j(\tau_i,\tau_l))P\left[T \in I_l, J = j\right]\right)^{d_{\{1,2\}i}} S(\tau_i)^{n_i - d_i}\Bigg]. \qquad (4.4)\end{aligned}$$

Note that this likelihood is a function of the probability terms $P\left[T \in I_l, J = j\right]$, for $l = 1,\cdots,k$, $j = 1,2$, and the unknown parameters in the masking probabilities $p_j(\tau_i,\tau_l)$'s. Therefore, these probability terms, as functions of $\underset{\sim}{F}$, are the only quantities which may be estimable. In the following, we consider a reparametrization of these probabilities in terms of "interval hazards", as in Koley and Dewanji (2018b), to facilitate the maximization of likelihood in a less constrained manner.

Write $\lambda_{jl} = P\left[T \in I_l, J = j \mid T > \tau_{l-1}\right]$ and $\lambda_l = \lambda_{1l} + \lambda_{2l} = P\left[T \in I_l \mid T > \tau_{l-1}\right]$, for $l = 1,\cdots,k$. It can be easily seen that $P\left[T \in I_l, J = j\right] = \lambda_{jl}\prod_{l'<l}(1 - \lambda_{l'})$ and $S(\tau_i) = P\left[T > \tau_i\right] = \prod_{l=1}^{i}(1 - \lambda_l)$. Then, the likelihood (4.4) can be written in terms of these λ_{jl}'s and the parameters involved in the $p_j(\tau_i,\tau_l)$'s. In the next two subsections, we consider two modeling scenarios for the $p_j(\tau_i,\tau_l)$'s, both decreasing in $(\tau_i - \tau_l)$ to conform with the usual notion.

4.1 Model 1

In order to mimick the model (M1) for the parametric analysis (See Subsection 3.1), let us assume

$$p_j(\tau_i, \tau_l) = e^{-\{\beta+\alpha_j(\tau_i-\tau_l)\}}, \tag{M3}$$

with $\alpha_j > 0$, for $j = 1, 2$, and $\beta > 0$ is a known constant.Unlike the parametric analysis in Section 3.1, here the probability of τ_l being equal to τ_i is non-zero, for since τ_l effectively represents the interval I_l and the interval I_i may contain the corresponding failure time for an observation at τ_i. This is the reason for adding the term β while modeling the masking probabilities through (M3) so that the probability of correct diagnosis $p_j(\tau_i, \tau_i)$ at time τ_i, when failure occurs in I_i, is $e^{-\beta}$, which is more reasonable than having it as 1. The probability $p_j(\tau_i, \tau_l)$ keeps decreasing with unknown rate α_j as the failure interval I_l is further away from τ_i, for $l \leq i$ and $j = 1, 2$. The likelihood (4.4) can now be written as

$$\begin{aligned} L_I(\underset{\sim}{\lambda}, \alpha_1) = \prod_{i=1}^{k} \Bigg[& \left(\sum_{l=1}^{i} e^{-\{\beta+\alpha_1(\tau_i-\tau_l)\}} \lambda_{1l} \prod_{l'<l} (1-\lambda_{l'}) \right)^{d_{\{1\}i}} \\ & \times \left(\sum_{l=1}^{i} e^{-\{\beta+\alpha_2(\tau_i-\tau_l)\}} \lambda_{2l} \prod_{l'<l} (1-\lambda_{l'}) \right)^{d_{\{2\}i}} \\ & \times \left(\sum_{l=1}^{i} \sum_{j=1}^{2} (1 - e^{-\{\beta+\alpha_j(\tau_i-\tau_l)\}}) \lambda_{jl} \prod_{l'<l} (1-\lambda_{l'}) \right)^{d_{\{1,2\}i}} \\ & \times \left(\prod_{l \leq i} (1-\lambda_l) \right)^{n_i - d_i} \Bigg], \end{aligned} \tag{4.5}$$

where $\underset{\sim}{\lambda}$ denotes the vector of λ_{ji}'s.

Theorem 4.1. *The likelihood* (4.5) *is identifiable under the assumption* $\alpha_2 = c\ \alpha_1$, *for some known constant* $c > 0$.

Proof. To check the identifiability, let $(\underset{\sim}{\lambda}^{(1)}, \alpha_1^{(1)})$ and $(\underset{\sim}{\lambda}^{(2)}, \alpha_1^{(2)})$ be two values of the parameter vector $(\underset{\sim}{\lambda}, \alpha_1)$ and consider the identity

$$\log L_I(\underset{\sim}{\lambda}^{(1)}, \alpha_1^{(1)}) = \log L_I(\underset{\sim}{\lambda}^{(2)}, \alpha_1^{(2)}) \tag{4.6}$$

for all data configurations $\{d_{gi},\ g \in \mathcal{G} \setminus \phi,\ i = 1, \cdots, k\}$. For a given i, equating the coefficients of $d_{\{1\}i}$, $d_{\{2\}i}$, and $n_i - d_i$ from both sides of (4.6), we get

$$\sum_{l=1}^{i} e^{-\{\beta+\alpha_1^{(1)}(\tau_i-\tau_l)\}} \lambda_{1l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)}) = \sum_{l=1}^{i} e^{-\{\beta+\alpha_1^{(2)}(\tau_i-\tau_l)\}} \lambda_{1l}^{(2)} \prod_{l'<l} (1-\lambda_{l'}^{(2)}),$$

$$\sum_{l=1}^{i} e^{-\{\beta+c\alpha_1^{(1)}(\tau_i-\tau_l)\}}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)}) = \sum_{l=1}^{i} e^{-\{\beta+c\alpha_1^{(2)}(\tau_i-\tau_l)\}}\lambda_{2l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)}),$$

$$\text{and} \quad \prod_{l\le i}(1-\lambda_l^{(1)}) = \prod_{l\le i}(1-\lambda_l^{(2)}),$$

respectively. Note that this third equality is equivalent to

$$\sum_{l=1}^{i} \lambda_l^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)}) = \sum_{l=1}^{i} \lambda_l^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)}). \tag{4.7}$$

We have, from the first two equalities,

$$\frac{\sum_{l=1}^{i} e^{-\{\beta+\alpha_1^{(1)}(\tau_i-\tau_l)\}}\lambda_{1l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})}{\sum_{l=1}^{i} e^{-\{\beta+c\alpha_1^{(1)}(\tau_i-\tau_l)\}}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})} = \frac{\sum_{l=1}^{i} e^{-\{\beta+\alpha_1^{(2)}(\tau_i-\tau_l)\}}\lambda_{1l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)})}{\sum_{l=1}^{i} e^{-\{\beta+c\alpha_1^{(2)}(\tau_i-\tau_l)\}}\lambda_{2l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)})}$$

$$\text{Or,} \frac{\sum_{l=1}^{i} e^{-\alpha_1^{(1)}(\tau_i-\tau_l)}\lambda_{1l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})}{\sum_{l=1}^{i} e^{-c\alpha_1^{(1)}(\tau_i-\tau_l)}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})} = \frac{\sum_{l=1}^{i} e^{-\alpha_1^{(2)}(\tau_i-\tau_l)}\lambda_{1l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)})}{\sum_{l=1}^{i} e^{-c\alpha_1^{(2)}(\tau_i-\tau_l)}\lambda_{2l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)})} = k_i, \tag{4.8}$$

say, which is non-negative possibly depending on $\underset{\sim}{\lambda}^{(1)}, \underset{\sim}{\lambda}^{(2)}, \alpha_1^{(1)}$ and $\alpha_1^{(2)}$, for $i = 1, \cdots, k$. This implies

$$\sum_{l=1}^{i} e^{-\alpha_1^{(1)}(\tau_i-\tau_l)}\lambda_{1l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)}) = k_i \sum_{l=1}^{i} e^{-c\alpha_1^{(1)}(\tau_i-\tau_l)}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)}) \tag{4.9}$$

and

$$\sum_{l=1}^{i} e^{-\alpha_1^{(2)}(\tau_i-\tau_l)}\lambda_{1l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)}) = k_i \sum_{l=1}^{i} e^{-c\alpha_1^{(2)}(\tau_i-\tau_l)}\lambda_{2l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)}), \tag{4.10}$$

for $i = 1, \cdots, k$. From (4.9), for a fixed i,

$$\sum_{j=0}^{\infty}\left[(-1)^j \sum_{l=1}^{i} \frac{(\tau_i-\tau_l)^j}{j!}\lambda_{1l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})\right](\alpha_1^{(1)})^j = k_i \sum_{j=0}^{\infty}\left[(-c)^j \sum_{l=1}^{i} \frac{(\tau_i-\tau_l)^j}{j!}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})\right](\alpha_1^{(1)})^j$$

$$\iff \sum_{j=0}^{\infty} a_j^{(i)}(\alpha_1^{(1)})^j = \sum_{j=0}^{\infty} b_j^{(i)}(\alpha_1^{(1)})^j, \quad \text{say}, \tag{4.11}$$

where $a_j^{(i)} = (-1)^j \sum_{l=1}^{i} \frac{(\tau_i-\tau_l)^j}{j!}\lambda_{1l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})$ and $b_j^{(i)} = k_i(-c)^j \sum_{l=1}^{i} \frac{(\tau_i-\tau_l)^j}{j!}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})$, for $j = 0, 1, \cdots$. Both sides of (4.11) are

power series in $\alpha_1^{(1)}$ and hence we have, $a_j^{(i)} = b_j^{(i)}$, for all $j = 0, 1, \cdots$. In particular, for $j = 0$,

$$\sum_{l=1}^{i} \lambda_{1l}^{(1)} \prod_{l'<l} (1 - \lambda_{l'}^{(1)}) = k_i \sum_{l=1}^{i} \lambda_{2l}^{(1)} \prod_{l'<l} (1 - \lambda_{l'}^{(1)}),$$

for all $i = 1, \cdots, k$. Similarly, from (4.10) and for a fixed i, we have

$$\sum_{l=1}^{i} \lambda_{1l}^{(2)} \prod_{l'<l} (1 - \lambda_{l'}^{(2)}) = k_i \sum_{l=1}^{i} \lambda_{2l}^{(2)} \prod_{l'<l} (1 - \lambda_{l'}^{(2)}).$$

Using these two expressions in (4.7), we have,

$$(1 + k_i) \sum_{l=1}^{i} \lambda_{2l}^{(1)} \prod_{l'<l} (1 - \lambda_{l'}^{(1)}) = (1 + k_i) \sum_{l=1}^{i} \lambda_{2l}^{(2)} \prod_{l'<l} (1 - \lambda_{l'}^{(2)})$$

$$\iff \sum_{l=1}^{i} \lambda_{2l}^{(1)} \prod_{l'<l} (1 - \lambda_{l'}^{(1)}) = \sum_{l=1}^{i} \lambda_{2l}^{(2)} \prod_{l'<l} (1 - \lambda_{l'}^{(2)}), \quad \text{since} \quad (1 + k_i) > 0.$$

Again, using this in (4.7), we obtain

$$\sum_{l=1}^{i} \lambda_{1l}^{(1)} \prod_{l'<l} (1 - \lambda_{l'}^{(1)}) = \sum_{l=1}^{i} \lambda_{1l}^{(2)} \prod_{l'<l} (1 - \lambda_{l'}^{(2)}).$$

Both the above expressions hold for all $i = 1, \cdots, k$ and, hence, $\underset{\sim}{\lambda}^{(1)} = \underset{\sim}{\lambda}^{(2)}$. Finally, equating the coefficients of $d_{\{1\}i}$ from both sides of the identity (4.6) and noting that $\underset{\sim}{\lambda}^{(1)} = \underset{\sim}{\lambda}^{(2)}$, we obtain $\alpha_1^{(1)} = \alpha_1^{(2)}$. This proves the required identifiability. □

4.2 Model 2

In this subsection also, we have tried to mimick the second model (M2) of the parametric analysis (See Subsection 3.2) by assuming

$$p_j(\tau_i, \tau_l) = e^{-\{\alpha_j + \beta(\tau_i - \tau_l)\}}, \tag{M4}$$

with β known, $\alpha_1, \alpha_2 > 0$. The likelihood (4.4) under this model can be written as

$$
\begin{aligned}
L_{II}(\underset{\sim}{\lambda}, \alpha_1) = \prod_{i=1}^{k} \Bigg[& \left(\sum_{l=1}^{i} e^{-\{\alpha_1 + \beta(\tau_i - \tau_l)\}} \lambda_{1l} \prod_{l' < l} (1 - \lambda_{l'}) \right)^{d_{\{1\}i}} \\
& \times \left(\sum_{l=1}^{i} e^{-\{\alpha_2 + \beta(\tau_i - \tau_l)\}} \lambda_{2l} \prod_{l' < l} (1 - \lambda_{l'}) \right)^{d_{\{2\}i}} \\
& \times \left(\sum_{l=1}^{i} \sum_{j=1}^{2} (1 - e^{-\{\alpha_j + \beta(\tau_i - \tau_l)\}}) \lambda_{jl} \prod_{l' < l} (1 - \lambda_{l'}) \right)^{d_{\{1,2\}i}} \\
& \times \left(\prod_{l \le i} (1 - \lambda_l) \right)^{n_i - d_i} \Bigg].
\end{aligned} \tag{4.12}
$$

Theorem 4.2. *The likelihood* (4.12) *is not identifiable.*

Proof. Let $(\underset{\sim}{\lambda}^{(1)}, \alpha_1^{(1)}, \alpha_2^{(1)})$ and $(\underset{\sim}{\lambda}^{(2)}, \alpha_1^{(2)}, \alpha_2^{(2)})$ be two values of the parameter vector $(\underset{\sim}{\lambda}, \alpha_1, \alpha_2)$ and consider the identity

$$
\log L_{II}(\underset{\sim}{\lambda}^{(1)}, \alpha_1^{(1)}, \alpha_2^{(1)}) = \log L_{II}(\underset{\sim}{\lambda}^{(2)}, \alpha_1^{(2)}, \alpha_2^{(2)}) \tag{4.13}
$$

for all data configurations $\{d_{gi},\ g \in \mathcal{G} \setminus \phi,\ i = 1, \cdots, k\}$. For a fixed i, equating the coefficients of $d_{\{1\}i}$, $d_{\{2\}i}$ and $(n_i - d_i)$ from both sides of (4.13), we get

$$
\sum_{l=1}^{i} e^{-\{\alpha_1^{(1)} + \beta(\tau_i - \tau_l)\}} \lambda_{1l}^{(1)} \prod_{l' < l} (1 - \lambda_{l'}^{(1)}) = \sum_{l=1}^{i} e^{-\{\alpha_1^{(2)} + \beta(\tau_i - \tau_l)\}} \lambda_{1l}^{(2)} \prod_{l' < l} (1 - \lambda_{l'}^{(2)}),
$$

$$
\sum_{l=1}^{i} e^{-\{\alpha_2^{(1)} + \beta(\tau_i - \tau_l)\}} \lambda_{2l}^{(1)} \prod_{l' < l} (1 - \lambda_{l'}^{(1)}) = \sum_{l=1}^{i} e^{-\{\alpha_2^{(2)} + \beta(\tau_i - \tau_l)\}} \lambda_{2l}^{(2)} \prod_{l' < l} (1 - \lambda_{l'}^{(2)}),
$$

$$
\text{and} \quad \prod_{l \le i} (1 - \lambda_l^{(1)}) = \prod_{l \le i} (1 - \lambda_l^{(2)}),
$$

respectively. Using the first two equations, we have

$$
\frac{\sum_{l=1}^{i} e^{-\{\alpha_1^{(1)} + \beta(\tau_i - \tau_l)\}} \lambda_{1l}^{(1)} \prod_{l' < l} (1 - \lambda_{l'}^{(1)})}{\sum_{l=1}^{i} e^{-\{\alpha_2^{(1)} + \beta(\tau_i - \tau_l)\}} \lambda_{2l}^{(1)} \prod_{l' < l} (1 - \lambda_{l'}^{(1)})} = \frac{\sum_{l=1}^{i} e^{-\{\alpha_1^{(2)} + \beta(\tau_i - \tau_l)\}} \lambda_{1l}^{(2)} \prod_{l' < l} (1 - \lambda_{l'}^{(2)})}{\sum_{l=1}^{i} e^{-\{\alpha_2^{(2)} + \beta(\tau_i - \tau_l)\}} \lambda_{2l}^{(2)} \prod_{l' < l} (1 - \lambda_{l'}^{(2)})} = K_i,
$$

say, which is non-negative possibly depending on $\underset{\sim}{\lambda}^{(1)}, \underset{\sim}{\lambda}^{(2)}$, $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$. This gives

$$
\sum_{l=1}^{i} e^{-\{\alpha_1^{(1)} + \beta(\tau_i - \tau_l)\}} \lambda_{1l}^{(1)} \prod_{l' < l} (1 - \lambda_{l'}^{(1)}) = K_i \sum_{l=1}^{i} e^{-\{\alpha_2^{(1)} + \beta(\tau_i - \tau_l)\}} \lambda_{2l}^{(1)} \prod_{l' < l} (1 - \lambda_{l'}^{(1)}) \tag{4.14}
$$

$$\text{and} \quad \sum_{l=1}^{i} e^{-\{\alpha_1^{(2)}+\beta(\tau_i-\tau_l)\}}\lambda_{1l}^{(2)} \prod_{l'<l}(1-\lambda_{l'}^{(2)}) = K_i \sum_{l=1}^{i} e^{-\{\alpha_2^{(2)}+\beta(\tau_i-\tau_l)\}}\lambda_{2l}^{(2)} \prod_{l'<l}(1-\lambda_{l'}^{(2)}), \tag{4.15}$$

for all $i = 1, \cdots, k$. For a given i, from (4.14), we have

$$e^{-\alpha_1^{(1)}} \sum_{l=1}^{i} e^{-\beta(\tau_i-\tau_l)}\lambda_{1l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)}) = K_i e^{-\alpha_2^{(1)}} \sum_{l=1}^{i} e^{-\beta(\tau_i-\tau_l)}\lambda_{2l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)})$$

$$\iff \sum_{j=0}^{\infty}\left[e^{-\alpha_1^{(1)}} \sum_{l=1}^{i}(-1)^j \frac{(\tau_i-\tau_l)^j}{j!}\lambda_{1l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)})\right]\beta^j$$

$$= \sum_{j=0}^{\infty}\left[K_i e^{-\alpha_2^{(1)}} \sum_{l=1}^{i}(-1)^j \frac{(\tau_i-\tau_l)^j}{j!}\lambda_{2l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)})\right]\beta^j$$

$$\iff \sum_{j=0}^{\infty} A_{ij}\beta^j = \sum_{j=0}^{\infty} B_{ij}\beta^j,$$

where $A_{ij} = e^{-\alpha_1^{(1)}} \sum_{l=1}^{i}(-1)^j \frac{(\tau_i-\tau_l)^j}{j!}\lambda_{1l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)})$

$$\text{and} \quad B_{ij} = K_i e^{-\alpha_2^{(1)}} \sum_{l=1}^{i}(-1)^j \frac{(\tau_i-\tau_l)^j}{j!}\lambda_{2l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)}),$$

for all $j = 0, 1, \cdots$. Both sides of the above equation are power series in β and hence, $A_{ij} = B_{ij}$, for all $i = 1, \cdots, k$ and $j = 0, 1, \cdots$. For a given i and $j = 0$, we have

$$e^{-\alpha_1^{(1)}} \sum_{l=1}^{i} \lambda_{1l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)}) = K_i e^{-\alpha_2^{(1)}} \sum_{l=1}^{i} \lambda_{2l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)})$$

$$\iff \sum_{l=1}^{i} \lambda_{1l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)}) = K_i e^{\alpha_1^{(1)}-\alpha_2^{(1)}} \sum_{l=1}^{i} \lambda_{2l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)}),$$

for all $i = 1, \cdots, k$. Similarly, from (4.15), we get

$$\iff \sum_{l=1}^{i} \lambda_{1l}^{(2)} \prod_{l'<l}(1-\lambda_{l'}^{(2)}) = K_i e^{\alpha_1^{(2)}-\alpha_2^{(2)}} \sum_{l=1}^{i} \lambda_{2l}^{(2)} \prod_{l'<l}(1-\lambda_{l'}^{(2)}), \quad \text{for all } i = 1, \cdots, k.$$

Using these equations in $\prod_{l\le i}(1-\lambda_l^{(1)}) = \prod_{l\le i}(1-\lambda_l^{(2)})$, we have

$$\left(1+K_i e^{\alpha_1^{(1)}-\alpha_2^{(1)}}\right)\sum_{l=1}^{i}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)}) = \left(1+K_i e^{\alpha_1^{(2)}-\alpha_2^{(2)}}\right)\sum_{l=1}^{i}\lambda_{2l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)})$$

$$\iff K_i = \frac{\sum_{l=1}^{i}\lambda_{2l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)}) - \sum_{l=1}^{i}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})}{e^{\alpha_1^{(1)}-\alpha_2^{(1)}}\sum_{l=1}^{i}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)}) - e^{\alpha_1^{(2)}-\alpha_2^{(2)}}\sum_{l=1}^{i}\lambda_{2l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)})},$$

for all $i = 1, \cdots, k$. If the likelihood is identifiable, then the above expression of K_i is of the form $0/0$ which contradicts the fact that $K_i > 0$, for at least one i. Thus, the likelihood (4.12) is not identifiable. □

Theorem 4.3. *If α_1 and α_2 satisfy (i) $\alpha_2 = \alpha_1 + f$, or (ii) $\alpha_2 = c\ \alpha_1$, for some known constants f and c with $c > 0$, then the likelihood* (4.12) *is identifiable.*

Proof. Suppose $(\underset{\sim}{\lambda}^{(1)}, \alpha_1^{(1)}, \alpha_2^{(1)})$ and $(\underset{\sim}{\lambda}^{(2)}, \alpha_1^{(2)}, \alpha_2^{(2)})$ are two values of the parameter vector $(\underset{\sim}{\lambda}, \alpha_1, \alpha_2)$.

(i) Here, $\alpha_2 = \alpha_1 + f$, where f is a known constant. This proof is similar to that for the parametric analysis (See Theorem 3.3(i)) with integration replaced by summation and is presented below.

Replacing α_2 by $\alpha_1 + f$ in Theorem 4.2, we have from (4.14) and (4.15),

$$\sum_{l=1}^{i} e^{-\beta(\tau_i-\tau_l)}\lambda_{1l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)}) = K_i\sum_{l=1}^{i} e^{-f-\beta(\tau_i-\tau_l)}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)}) \tag{4.16}$$

and
$$\sum_{l=1}^{i} e^{-\beta(\tau_i-\tau_l)}\lambda_{1l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)}) = K_i\sum_{l=1}^{i} e^{-f-\beta(\tau_i-\tau_l)}\lambda_{2l}^{(2)}\prod_{l'<l}(1-\lambda_{l'}^{(2)}), \tag{4.17}$$

respectively, for all $i = 1, \cdots, k$. From (4.16), we have

$$\sum_{j=0}^{\infty}\left[(-1)^j\sum_{l=1}^{i}\frac{(\tau_i-\tau_l)^j}{j!}\lambda_{1l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})\right]\beta^j = \sum_{j=0}^{\infty}\left[K_i e^{-f}(-1)^j\sum_{l=1}^{i}\frac{(\tau_i-\tau_l)^j}{j!}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})\right]\beta^j$$

$$\iff \sum_{j=0}^{\infty}A_{ij}\beta^j = \sum_{j=0}^{\infty}B_{ij}\beta^j,$$

where $A_{ij} = (-1)^j\sum_{l=1}^{i}\dfrac{(\tau_i-\tau_l)^j}{j!}\lambda_{1l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)})$ and

$$B_{ij} = (-1)^j K_i e^{-f}\sum_{l=1}^{i}\frac{(\tau_i-\tau_l)^j}{j!}\lambda_{2l}^{(1)}\prod_{l'<l}(1-\lambda_{l'}^{(1)}),$$

for all $i = 1, \cdots, k$ and $j = 0, 1, \cdots$. For a given i, both sides of the above equation are power series in β and hence, $A_{ij} = B_{ij}$, for $j = 0, 1, \cdots$. Taking $j = 0$, we have

$$\sum_{l=1}^{i} \lambda_{1l}^{(1)} \prod_{l'<l}(1 - \lambda_{l'}^{(1)}) = K_i e^{-f} \sum_{l=1}^{i} \lambda_{2l}^{(1)} \prod_{l'<l}(1 - \lambda_{l'}^{(1)}),$$

for all $i = 1, \cdots, k$. Similarly, from (4.17), we get

$$\sum_{l=1}^{i} \lambda_{1l}^{(2)} \prod_{l'<l}(1 - \lambda_{l'}^{(2)}) = K_i e^{-f} \sum_{l=1}^{i} \lambda_{2l}^{(2)} \prod_{l'<l}(1 - \lambda_{l'}^{(2)}),$$

for all $i = 1, \cdots, k$. Using these expressions in $\prod_{l \le i}(1 - \lambda_l^{(1)}) = \prod_{l \le i}(1 - \lambda_l^{(2)})$, we have

$$\sum_{l=1}^{i} \left(1 + K_i e^{-f}\right) \lambda_{2l}^{(1)} \prod_{l'<l}(1-\lambda_{l'}^{(1)}) = \sum_{l=1}^{i} \left(1 + K_i e^{-f}\right) \lambda_{2l}^{(2)} \prod_{l'<l}(1-\lambda_{l'}^{(2)}), \quad \text{for all } i = 1, \cdots, k.$$

$$\iff \sum_{l=1}^{i} \lambda_{2l}^{(1)} \prod_{l'<l}(1 - \lambda_{l'}^{(1)}) = \sum_{l=1}^{i} \lambda_{2l}^{(2)} \prod_{l'<l}(1 - \lambda_{l'}^{(2)}), \quad \text{for all } i = 1, \cdots, k.$$

This is because $e^{-f} > 0$ and $K_i > 0$, for all $i = 1, \cdots, k$. Using this in $\prod_{l \le i}(1-\lambda_l^{(1)}) = \prod_{l \le i}(1 - \lambda_l^{(2)})$, we get $\sum_{l=1}^{i} \lambda_{1l}^{(1)} \prod_{l'<l}(1 - \lambda_{l'}^{(1)}) = \sum_{l=1}^{i} \lambda_{1l}^{(2)} \prod_{l'<l}(1 - \lambda_{l'}^{(2)})$, for all $i = 1, \cdots, k$, which gives $\underset{\sim}{\lambda}^{(1)} = \underset{\sim}{\lambda}^{(2)}$. Equating the coefficients of $d_{\{1\}i}$ from both sides of (4.13) with $\underset{\sim}{\lambda}^{(1)} = \underset{\sim}{\lambda}^{(2)} = \underset{\sim}{\lambda}$, say, we have

$$\sum_{l=1}^{i} e^{-\{\alpha_1^{(1)} + \beta(\tau_i - \tau_l)\}} \lambda_{1l} \prod_{l'<l}(1-\lambda_{l'}) = \sum_{l=1}^{i} e^{-\{\alpha_1^{(2)} + \beta(\tau_i - \tau_l)\}} \lambda_{1l} \prod_{l'<l}(1-\lambda_{l'}), \quad \text{for all } i = 1, \cdots, k.$$

$$\iff e^{-\alpha_1^{(1)}} \sum_{l=1}^{i} e^{-\beta(\tau_i - \tau_l)} \lambda_{1l} \prod_{l'<l}(1-\lambda_{l'}) = e^{-\alpha_1^{(2)}} \sum_{l=1}^{i} e^{-\beta(\tau_i - \tau_l)} \lambda_{1l} \prod_{l'<l}(1-\lambda_{l'}), \quad \text{for all } i = 1, \cdots, k.$$

$$\iff \alpha_1^{(1)} = \alpha_1^{(2)}.$$

This proves the identifiability of the likelihood (4.12).

(ii) In the second case, $\alpha_2 = c\ \alpha_1$, where $c > 0$ is a known constant. Let $(\underset{\sim}{\lambda}^{(1)}, \alpha_1^{(1)})$ and $(\underset{\sim}{\lambda}^{(2)}, \alpha_1^{(2)})$ be two values of the parameter vector $(\underset{\sim}{\lambda}, \alpha_1)$ and consider the identity (4.13). Proceeding in the similar manner as in the previous theorem with $\alpha_2 = c\ \alpha_1$, we have from (4.14),

$$\sum_{l=1}^{i} e^{-\alpha_1^{(1)}} \lambda_{1l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)}) = K_i \sum_{l=1}^{i} e^{-c\ \alpha_1^{(1)}} \lambda_{2l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)})$$

$$\sum_{j=0}^{\infty} \left[\frac{(-1)^j}{j!} \sum_{l=0}^{\infty} \lambda_{1l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)}) \right] (\alpha_1^{(1)})^j = \sum_{j=0}^{\infty} \left[K_i c^j \frac{(-1)^j}{j!} \sum_{l=0}^{\infty} \lambda_{2l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)}) \right] (\alpha_1^{(1)})^j$$

Both sides of the above equation are power series in $\alpha_1^{(1)}$ and, hence, again using the property of power series for $j = 0$, we have

$$\sum_{l=0}^{i} \lambda_{1l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)}) = K_i \sum_{l=0}^{i} \lambda_{2l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)}), \quad \text{for all } i = 1, \cdots, k.$$

Similarly, from (4.15), we get

$$\sum_{l=0}^{\infty} \lambda_{1l}^{(2)} \prod_{l'<l} (1-\lambda_{l'}^{(2)}) = K_i \sum_{l=0}^{\infty} \lambda_{2l}^{(2)} \prod_{l'<l} (1-\lambda_{l'}^{(2)}), \quad \text{for all } i = 1, \cdots, k.$$

Using these two equations in $\prod_{l\leq i}(1-\lambda_l^{(1)}) = \prod_{l\leq i}(1-\lambda_l^{(2)})$, we get

$$(1+K_i) \sum_{l=0}^{i} \lambda_{2l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)}) = (1+K_i) \sum_{l=0}^{i} \lambda_{2l}^{(2)} \prod_{l'<l} (1-\lambda_{l'}^{(2)}) \quad \text{for all } i = 1, \cdots, k.$$

$$\Longleftrightarrow \sum_{l=0}^{i} \lambda_{2l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)}) = \sum_{l=0}^{i} \lambda_{2l}^{(2)} \prod_{l'<l} (1-\lambda_{l'}^{(2)}),$$

since $K_i \geq 0$, for all $i = 1, \cdots, k$. Using this in $\prod_{l\leq i}(1-\lambda_l^{(1)}) = \prod_{l\leq i}(1-\lambda_l^{(2)})$, we have

$$\sum_{l=0}^{i} \lambda_{1l}^{(1)} \prod_{l'<l} (1-\lambda_{l'}^{(1)}) = \sum_{l=0}^{i} \lambda_{1l}^{(2)} \prod_{l'<l} (1-\lambda_{l'}^{(2)}),$$

for all $i = 1, \cdots, k$, which implies $\underset{\sim}{\lambda}^{(1)} = \underset{\sim}{\lambda}^{(2)}$. Finally equating the coefficients of $d_{\{1\}i}$ from both sides of (4.13) with $\underset{\sim}{\lambda}^{(1)} = \underset{\sim}{\lambda}^{(2)} = \underset{\sim}{\lambda}$, say, we have

$$\sum_{l=1}^{i} e^{-\{\alpha_1^{(1)}+\beta(\tau_i-\tau_l)\}} \lambda_{1l} \prod_{l'<l} (1-\lambda_{l'}) = \sum_{l=1}^{i} e^{-\{\alpha_1^{(2)}+\beta(\tau_i-\tau_l)\}} \lambda_{1l} \prod_{l'<l} (1-\lambda_{l'}), \quad \text{for all } i = 1, \cdots, k.$$

$$\Longleftrightarrow \alpha_1^{(1)} = \alpha_1^{(2)} \quad \text{and hence, the identifiability is proved.}$$

□

4.3 Maximum Likelihood Estimation

From the likelihood functions (4.5) and (4.12), it is clear that no closed form expressions for the maximum likelihood estimators of $(\underset{\sim}{\lambda}, \alpha_1)$ exist. Hence, for the estimation of the parameters, numerical maximization method is used. To serve this purpose, the *optim* function in R is used. Using the invariance property of the MLE, the sub-distribution functions at the monitoring time τ_i is estimated as $\hat{F}_j(\tau_i) = \sum_{l\leq i} \hat{\lambda}_{jl} \prod_{l'<l}(1-\hat{\lambda}_{l'})$, for $j = 1, 2$ and $i = 1, \cdots, k$. The corresponding standard errors can be obtained by using the delta method.

Since the monitoring times are fixed taking finite values, the number of parameters to be estimated is also fixed no matter how large the sample size is. The dimension of $\underset{\sim}{\lambda}$ is $2k$ and the masking probabilities involve a single parameter α_1. In order to study the limiting distribution of the vector $\underset{\sim}{\hat{\lambda}}$ and $\hat{\alpha_1}$, it becomes a problem of parametric analysis with the vector of parameters being denoted by $\underset{\sim}{\eta} = (\underset{\sim}{\lambda}, \alpha_1)$ and the MLE by $\underset{\sim}{\hat{\eta}}$. At each monitoring time τ_i, there are n_i independent realizations of G from the identical distribution given by

$$f^i(g \mid \underset{\sim}{\eta}) = \begin{cases} \sum_{l\leq i}\left[p_{\{1\}1}(\tau_i, \tau_l; \alpha_1)\lambda_{1l}\prod_{l'<l}(1-\lambda_{l'})\right], & \text{if } g = \{1\} \\ \\ \sum_{l\leq i}\left[p_{\{2\}2}(\tau_i, \tau_l; \alpha_1)\lambda_{2l}\prod_{l'<l}(1-\lambda_{l'})\right], & \text{if } g = \{2\} \\ \\ \sum_{l\leq i}\left[\sum_{j=1}^{2}(1-p_{\{j\}j}(\tau_i, \tau_l; \alpha_1))\lambda_{1l}\prod_{l'<l}(1-\lambda_{l'})\right], & \text{if } g = \{1, 2\} \\ \\ \prod_{l\leq i}(1-\lambda_l), & \text{if } g = \phi. \end{cases}$$

Let us denote the information matrix corresponding to the monitoring time τ_i by

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

where $E_{i,\underset{\sim}{\eta}}\left[\cdot\right]$ is the expectation with respect to the density $f^i(g \mid \underset{\sim}{\eta})$, for $i = 1, \cdots, k$. Let us now consider the following remarks.

R1 For a given j, $\lambda_{ji} = 0$, for all $i = 1, \cdots, k$, implies there is no failure due to cause j which is not of interest. Similarly, $\underset{\sim}{\lambda} = \underset{\sim}{1}$ means $\lambda_{ji} = 1$, for all j and i, which is not possible. Also, the range of the parameter α_1 is an open set given by $(0,\infty)$. So, the true value of the parameter vector $\underset{\sim}{\eta}$, denoted by $\underset{\sim}{\eta}_0$, lies in an open set.

R2 Suppose $\underset{\sim}{\eta}'$ and $\underset{\sim}{\eta}''$ be two distinct values of the parameter vector $\underset{\sim}{\eta}$ with $\underset{\sim}{\eta}' \neq \underset{\sim}{\eta}''$. Suppose there exists atleast one component of the two vectors which are not equal, say $\lambda_{ji}' \neq \lambda_{ji}''$, for some $i = 1, \cdots, k$ and $j = 1, 2$. Then the densities $f^{i_1}(\cdot \mid \underset{\sim}{\eta})$, for $i_1 \geq i$ that involve the component λ_{ji} will differ at two values, $\underset{\sim}{\eta}'$ and $\underset{\sim}{\eta}''$. Similar argument can be given for $\alpha_1' \neq \alpha_1''$.

R3 Note that $E_{i,\underset{\sim}{\eta}}\left[\frac{\partial}{\partial \underset{\sim}{\eta}} f^i(G \mid \underset{\sim}{\eta})\right] = 0$ and the second order partial derivative matrix $E_{i,\underset{\sim}{\eta}}\left[\left(\frac{\partial}{\partial \underset{\sim}{\eta}} f^i(G \mid \underset{\sim}{\eta})\right)^2\right] = E\left[-\frac{\partial^2}{\partial \underset{\sim}{\eta} \partial \underset{\sim}{\eta}^T} f^{i_1}(G \mid \underset{\sim}{\eta})\right] = \mathscr{I}_i(\underset{\sim}{\eta})$, which is assumed to be positive definite, for all $i = 1, \cdots, k$.

R4 For a given i, the density function $f^i(g \mid \underset{\sim}{\eta})$ is a polynomial function of the components of the parameter vector $\underset{\sim}{\lambda}$ and an exponential function of the parameter α_1. So it is continuous in each element λ_{ji}, for $j = 1, 2$ and $i = 1, \cdots, k$, and in α_1 as well. Therefore, it admits third order partial derivatives with respect to the parameter vector $\underset{\sim}{\eta}$. Also, $f^i(g \mid \underset{\sim}{\eta})$ being a polynomial function of λ_{ji}'s which are bounded and the term involving α_1 is also bounded, so it can be concluded that the third order partial derivatives with respect to the parameter vector $\underset{\sim}{\eta}$ are also bounded by some constant in the neighbourhood of the true value of $\underset{\sim}{\eta}$.

Theorem 4.4. *Under the assumption $\frac{n_i}{n} \longrightarrow w_i$, where w_i's are positive constants for all $i = 1, \cdots, k$, such that $\sum_{i=1}^k w_i = 1$, and using the remarks **R1-R4**, we have*

(i) $\hat{\underset{\sim}{\eta}} \xrightarrow{P} \underset{\sim}{\eta}_0$,

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

Proof. (i) The proof directly follows from Lehmann and Casella (1998, p463-465). We give only a sketch of the proof. For a given a, sufficiently small, consider a sphere Q_a, with center at the true value $\underset{\sim}{\eta}_0$ and radius a. We will show for any $\underset{\sim}{\eta}$ on the surface of Q_a, $\log L(\underset{\sim}{\eta}) < \log L(\underset{\sim}{\eta_0})$ with probability tending to 1 as $n \to \infty$. Note that

$$\begin{aligned}\frac{1}{n}(\log L(\underset{\sim}{\eta}) - \log L(\underset{\sim}{\eta_0})) =& \frac{1}{n}\sum_{i=1}^{k}\sum_{l=1}^{n_i}\left[\log f^i(g_l \mid \underset{\sim}{\eta}) - \log f^i(g_l \mid \underset{\sim}{\eta_0})\right] \\ =& \sum_{i=1}^{k}\frac{n_i}{n} \times \frac{1}{n_i}\sum_{i=1}^{n_i}\log\frac{f^i(g_l \mid \underset{\sim}{\eta})}{f^i(g_l \mid \underset{\sim}{\eta_0})}. \qquad (4.18)\end{aligned}$$

Then, using the remarks **R1-R4** and Theorem 5.1 of Lehmann and Casella (1998, p463-465), it can be shown that the maximum of (4.18) over all values of $\underset{\sim}{\eta}$ on the surface of Q_a is less than zero.

(ii) Using Taylor's series expansion of the likelihood equation around the true value $\underset{\sim}{\eta_0}$, we have

$$\begin{aligned}0 =& \frac{1}{\sqrt{n}}\sum_{i=1}^{k}\sum_{l=1}^{n_i}\log f^i(g_l \mid \underset{\sim}{\eta}) \mid_{\underset{\sim}{\eta} = \underset{\sim}{\hat\eta}} \\ =& \sum_{i=1}^{k}\sqrt{\frac{n_i}{n}} \times \frac{1}{\sqrt{n_i}}\sum_{l=1}^{n_i}\log f^i(g_l \mid \underset{\sim}{\eta}) \mid_{\underset{\sim}{\eta} = \underset{\sim}{\eta_0}} + \\ & \sum_{i=1}^{k}\frac{n_i}{n} \times \frac{1}{n_i}\sum_{l=1}^{n_i}\frac{\partial^2}{\partial\underset{\sim}{\eta}\partial\underset{\sim}{\eta}^T}\log f^i(g_l \mid \underset{\sim}{\eta}) \mid_{\underset{\sim}{\eta} = \underset{\sim}{\eta_0}} \times\sqrt{n}(\underset{\sim}{\hat\eta} - \underset{\sim}{\eta_0}) + o_p(1). \qquad (4.19)\end{aligned}$$

For a given i, by Weak Law of Large Numbers, we have

$$-\frac{1}{n_i}\sum_{l=1}^{n_i}\frac{\partial^2}{\partial\underset{\sim}{\eta}\partial\underset{\sim}{\eta}^T}\log f^i(g_l \mid \underset{\sim}{\eta}) \mid_{\underset{\sim}{\eta} = \underset{\sim}{\eta_0}} \xrightarrow{P} E\left[-\frac{\partial^2}{\partial\underset{\sim}{\eta}\partial\underset{\sim}{\eta_0}^T}\log f^i(g \mid \underset{\sim}{\eta}) \mid_{\underset{\sim}{\eta} = \underset{\sim}{\eta_0}}\right] = \mathscr{I}_i(\underset{\sim}{\eta_0}).$$

Also, using Central Limit Theorem, we have

$$\frac{1}{\sqrt{n_i}}\sum_{l=1}^{n_i}\log f^i(g_l \mid \underset{\sim}{\eta}) \mid_{\underset{\sim}{\eta} = \underset{\sim}{\eta_0}} \xrightarrow{d} N\left(\underset{\sim}{0}, \mathscr{I}_i(\underset{\sim}{\eta_0})\right)$$

This holds for all $i = 1, \cdots, k$. Using these results along with Slutsky's Theorem, we have

$$\sqrt{n}(\underset{\sim}{\hat\eta} - \underset{\sim}{\eta_0}) \xrightarrow{d} N\left(\underset{\sim}{0}, \left(\sum_{i=1}^{k} w_i\mathscr{I}_i(\underset{\sim}{\eta_0})\right)^{-1}\right).$$

□

Note that, $-\frac{1}{n}\frac{\partial^2}{\partial\underset{\sim}{\eta}\partial\underset{\sim}{\eta}^T}\log L(\underset{\sim}{\eta}) = -\sum_{i=1}^{k}\frac{n_i}{n} \times \frac{1}{n_i}\sum_{l=1}^{n_i}\frac{\partial^2}{\partial\underset{\sim}{\eta}\partial\underset{\sim}{\eta}^T}\log f^i(g_l \mid \underset{\sim}{\eta})$,
evaluated at $\eta = \hat\eta$, consistently estimates $\sum_{i=1}^{k} w_i\mathscr{I}_i(\underset{\sim}{\eta_0})$. To estimate the variance of

the sub-distribution functions $F_j(\tau_i)$'s, for $j = 1, 2$ and $i = 1, \cdots, k$, note that, $F_j(\tau_i) = \lambda_{ji} \prod_{l'<l}(1 - \lambda_{l'})$, a continuous function of $\underset{\sim}{\lambda}$. Therefore, using delta method (See Oehlert, 1992), the variance of $\hat{F}_j(\tau_i)$ is $D_{ji}^T \left[\sum_{i=1}^K w_i \mathscr{I}_i(\underset{\sim}{\eta})\right]^{-1} D_{ji}$, where D_{ji} is the derivative of $F_j(\tau_i)$ with respect to $\underset{\sim}{\lambda}$. This variance is estimated by evaluating it at $\underset{\sim}{\lambda} = \underset{\sim}{\hat{\lambda}}$.

In the above methods, it is assumed that the monitoring times are fixed with fixed number of individuals being observed at each of them. However, this method can be readily extended to the situation when the number of individuals observed at each monitoring time is random. Equivalently, the individuals are observed randomly at any of the τ_i's. Let us assume that the monitoring time X is independent of the random vector (T, J), then the observations are n independent realizations from the common density function

$$p(\tau, g) = \begin{cases} \sum_{l=1}^i p_j(\tau_i, \tau_l)\lambda_{jl} \prod_{l'<l}(1 - \lambda_{l'}) P\left[X = \tau_i\right], & \text{if } \tau = \tau_i, g = \{j\}, \\ \sum_{l=1}^i \sum_{j=1}^2 (1 - p_j(\tau_i, \tau_l))\lambda_{jl} \prod_{l'<l}(1 - \lambda_{l'}) P\left[X = \tau_i\right], & \text{if } \tau = \tau_i, g = \{1, 2\}, \\ \prod_{l \le i}(1 - \lambda_l) P\left[T = \tau_i\right], & \text{if } \tau = \tau_i, g = \phi, \end{cases}$$

for $j = 1, 2$ and $i = 1, \cdots, k$, with the dominating measure as the kronecker product of the discrete measure and the counting measure. The observed likelihood, written as the product of these n densities, is proportional to the likelihood (4.5) or (4.12). So, the estimation procedure remains the same as that with fixed monitoring times. Also, the results on consistency and asymptotic normality of the MLE are similar to those of Theorem 4.4 with w_i being replaced by $P\left[X = \tau_i\right]$, for all $i = 1, \cdots, k$.

5 Simulation Studies

In this section, we carry out several simulation studies to investigate the finite sample properties of MLEs of the model parameters. In the following two subsections, we study the parametric and the non-parametric estimates, respectively.

5.1 Parametric Method

We first simulate n observations on failure time T from the assumed distribution $F(\cdot; \underset{\sim}{\theta})$ and then the corresponding monitoring times (X's) from the distribution $G(\cdot)$. For an individual with failure time T less than the corresponding monitoring time X, the observed set of possible failure types is generated by using the probability distribution

$$P\left[G = g \mid T \le X = x\right] = \frac{\sum_{j \in g}\left[\int_0^x p_{gj}(x, t) f_j(t; \underset{\sim}{\theta}) \mathrm{dt}\right]}{F(x; \underset{\sim}{\theta})},$$ for $g \in \{\{1\}, \{2\}, \{1, 2\}\}$, with the

$p_{gj}(x,t)$'s as modeled in Section 3. By generating (X,G) for n times, we have a simulated data set $\{(x_i,g_i); i=1,\cdots,n\}$, which is repeated 10000 times to get 10000 such data sets.

For each simulated data set, we obtain the maximum likelihood estimates of the model parameters $(\underset{\sim}{\theta},\alpha_1)$ along with the corresponding standard errors using the methods of Section 3.3. Sample standard errors, denoted by SSE, are obtained from the sample variances of the estimates obtained from the 10000 simulations. Average of the 10000 standard errors over 10000 simulations, denoted by ASE, are also computed. For each simulated dataset, 95% confidence interval for each of the model parameters is obtained by using normal approximation of MLE. Cover percentage, denoted by CP, is estimated by the proportion of times these intervals contain the true value of the corresponding parameter.

First the failure time distribution is assumed to be Exponential with rate parameter 1 and the two types of failures occurring with rate ratio 3: 2. Next we consider Weibull distribution with two different choices of the shape parameter as $\theta=0.8$ (decreasing failure rate) and $\theta=1.2$ (increasing failure rate) and the scale parameter $(\lambda_1^\theta+\lambda_2^\theta)^{-1}$, where $\lambda_1=0.6,\lambda_2=0.4$. Different choices of sample sizes $n=50,150$ and 250 are considered to study the behaviour of the maximum likelihood estimates with increasing sample size. For analysis with $\alpha_2=c\,\alpha_1$ under (M1), two different choices of c (1 and 2) are considered with $\alpha_1=0.1$ (See Tables 1-3 for the three distributions, respectively). The same analysis with $\alpha_2=c\,\alpha_1$ is carried out under (M2) with $\alpha_1=0.05,\ \beta=0.03$ and two different choices of c as 0.1 and 1 (See Tables 4-6 for the three distributions, respectively). It appears that bias of the estimates decrease with increasing sample size. Also, both ASE and SSE come closer to each other as the sample size increases. The estimated coverage probabilities tend to 0.95 with increasing sample size thus providing some evidence in favour of asymptotic normality of the MLEs. Note that, in all these analyses, the value of c (and also β in (M2)) is assumed known. We have carried out some analysis with moderately mis-specified values of c and the results seem to be rather insensitive.

5.2 Non-parametric Method

We first fix $\tau_1,\cdots,\tau_k$, the k monitoring times. After generating $n=n_1+\cdots+n_k$ failure times from $F(\cdot;\underset{\sim}{\theta})$, the assumed distribution of T, the first n_1 observations are monitored at time τ_1, next n_2 observations at time τ_2 and so on. If an observation monitored at time τ_i is observed to have failed prior to τ_i, then the observed set of possible failure types, g, is generated as in Section 5.1 by using the $p_j(\tau_i,\tau_l)$'s, as modeled in Section 4, for $g\in\{\{1\},\{2\},\{1,2\}\}$. Thus, we get a simulated data set $\{(d_{gi},\ g\in\mathcal{G}/\phi),\ i=1,\cdots,k\}$. We repeat this procedure 10000 times to get 10000 such data sets. The monitoring times τ_i's are chosen as the $10^{th},25^{th},50^{th},75^{th}$ and 90^{th} percentiles of the assumed failure time distribution with $k = 5$ and we take $n_1 = n_2 = \cdots = n_k$ in our study. The failure times of n individuals are simulated from Exponential distribution with rate parameter 1 and the two types of failure occurring with rate ratio 3: 2, as before. Again, we take $\alpha_2 = c\alpha_1$ with two different choices of c (0.1 and 1) and $\beta = 0.01$ for both (M3) and (M4) (See Tables 7-10). The simulation is carried out with three common choices of n_i values given by $50, 150$ and 250. For each simulated data set, we obtain the MLEs of the sub-distribution functions at the monitoring times, that is, $\hat{F}_j(\tau_i)$, for $j = 1, 2$ and $i = 1, \cdots, k$. The corresponding ASE's and the SSE's are also obtained along with the respective CP's. All the results are presented in the Tables 7-10.

As expected bias, SSE and ASE of the estimates decrease with increase in sample size. In fact, both SSE and ASE come closer to each other and the estimated coverage probabilities tend to 0.95 as the sample size increases thus providing evidence in favour of asymptotic normality of the MLEs. Similar to the parametric analysis, here also different choices of c are used in the model and analysis. The results are found to be qualitatively similar to those in the parametric analysis and, hence, not reported here.

Table 1: Simulation results on $\left\{ \underset{\sim}{\hat{\eta}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\alpha}_1) \right\}$ assuming Exponential failure time distribution with $p_j(x,t) = e^{-\alpha_j(x-t)}$, $\alpha_j \geq 0$, for $j = 1, 2$, $t \leq x$, $\alpha_2 = c\ \alpha_1$ and $\lambda_1 = 0.6, \lambda_2 = 0.4, \alpha_1 = 0.1$.

c	n	Para-meters	Value of c used in the analysis							
			$c = 1$				$c = 1.1$			
			MLEs	ASE	SSE	CP	MLE	ASE	SSE	CP
$c = 1$	50	λ_1	0.625	0.177	0.176	0.947	0.622	0.176	0.176	0.946
		λ_2	0.422	0.143	0.147	0.921	0.425	0.144	0.147	0.942
		α_1	0.126	0.070	0.071	1.000	0.122	0.067	0.068	0.912
	150	λ_1	0.597	0.096	0.095	0.933	0.594	0.096	0.095	0.950
		λ_2	0.397	0.077	0.079	0.960	0.400	0.078	0.080	0.970
		α_1	0.115	0.036	0.037	0.999	0.110	0.035	0.036	0.918
	250	λ_1	0.601	0.074	0.075	0.945	0.598	0.074	0.075	0.938
		λ_2	0.398	0.059	0.061	0.953	0.401	0.060	0.061	0.962
		α_1	0.113	0.028	0.029	0.997	0.109	0.027	0.028	0.912
			$c = 1.9$				$c = 2$			
$c = 2$	50	λ_1	0.612	0.176	0.171	0.944	0.609	0.175	0.171	0.939
		λ_2	0.423	0.144	0.148	0.938	0.425	0.145	0.149	0.940
		α_1	0.121	0.061	0.067	0.932	0.117	0.059	0.065	1.000
	150	λ_1	0.598	0.097	0.093	0.970	0.596	0.097	0.093	0.939
		λ_2	0.406	0.079	0.078	0.924	0.408	0.079	0.078	0.943
		α_1	0.109	0.031	0.033	0.958	0.106	0.030	0.032	0.998
	250	λ_1	0.601	0.075	0.073	0.944	0.598	0.074	0.072	0.943
		λ_2	0.407	0.061	0.060	0.964	0.409	0.062	0.060	0.944
		α_1	0.111	0.024	0.026	0.934	0.107	0.023	0.025	0.992

Table 2: Simulation results on $\left\{ \underset{\sim}{\hat{\eta}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta}, \hat{\alpha}_1) \right\}$ assuming Weibull failure time distribution (shape parameter $\theta = 0.8$) with $p_j(x,t) = e^{-\alpha_j(x-t)}$, $\alpha_j \geq 0$, for $j = 1, 2$, $t \leq x$, $\alpha_2 = c\alpha_1$ and $\lambda_1 = 0.6, \lambda_2 = 0.4, \alpha_1 = 0.1$.

Value of c used in both simulation and analysis	n	parameters	Bias	ASE	SSE	CP
$c = 1$	50	λ_1	0.078	0.231	0.231	0.944
		λ_2	0.081	0.194	0.203	0.932
		θ	0.183	0.302	0.312	0.980
		α_1	0.015	0.064	0.059	0.938
	150	λ_1	0.025	0.127	0.125	0.956
		λ_2	0.033	0.111	0.112	0.946
		θ	0.062	0.149	0.145	0.968
		α_1	0.006	0.034	0.032	0.954
	250	λ_1	0.007	0.098	0.097	0.952
		λ_2	0.016	0.085	0.087	0.948
		θ	0.031	0.111	0.106	0.960
		α_1	0.003	0.026	0.025	0.948
$c = 2$	50	λ_1	0.060	0.219	0.217	0.932
		λ_2	0.088	0.196	0.195	0.930
		θ	0.192	0.301	0.299	0.978
		α_1	0.011	0.057	0.055	0.916
	150	λ_1	0.025	0.128	0.129	0.936
		λ_2	0.034	0.112	0.112	0.936
		θ	0.057	0.147	0.150	0.960
		α_1	0.003	0.029	0.031	0.932
	250	λ_1	0.019	0.099	0.099	0.950
		λ_2	0.016	0.086	0.086	0.948
		θ	0.031	0.110	0.110	0.948
		α_1	0.001	0.022	0.022	0.938

Table 3: Simulation results on $\left\{ \underset{\sim}{\hat{\eta}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta}, \hat{\alpha}_1) \right\}$ assuming Weibull failure time distribution (shape parameter $\theta = 1.2$) with $p_j(x,t) = e^{-\alpha_j(x-t)}$, $\alpha_j \geq 0$, for $j = 1, 2$, $t \leq x$, $\alpha_2 = c\alpha_1$ and $\lambda_1 = 0.6, \lambda_2 = 0.4, \alpha_1 = 0.1$.

Value of c used in both simulation and analysis	n	parameters	Bias	ASE	SSE	CP
$c = 1$	50	λ_1	0.044	0.162	0.172	0.932
		λ_2	0.045	0.152	0.157	0.936
		θ	0.173	0.401	0.385	0.978
		α_1	0.015	0.070	0.065	0.914
	150	λ_1	0.016	0.090	0.090	0.948
		λ_2	0.015	0.084	0.085	0.938
		θ	0.059	0.201	0.215	0.960
		α_1	0.002	0.036	0.038	0.934
	250	λ_1	0.009	0.070	0.070	0.950
		λ_2	0.012	0.065	0.065	0.954
		θ	0.033	0.150	0.157	0.952
		α_1	0.001	0.027	0.027	0.940
$c = 2$	50	λ_1	0.028	0.158	0.168	0.928
		λ_2	0.048	0.150	0.157	0.924
		θ	0.201	0.425	0.465	0.978
		α_1	0.010	0.062	0.062	0.898
	150	λ_1	0.009	0.091	0.098	0.940
		λ_2	0.010	0.085	0.089	0.926
		θ	0.046	0.198	0.210	0.960
		α_1	0.005	0.033	0.030	0.948
	250	λ_1	0.002	0.070	0.072	0.944
		λ_2	0.007	0.066	0.066	0.952
		θ	0.017	0.148	0.149	0.954
		α_1	0.002	0.024	0.024	0.952

Table 4: Simulation results on $\left\{ \underset{\sim}{\hat{\eta}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\alpha}_1) \right\}$ assuming Exponential(1) failure time distribution with $p_j(x,t) = e^{-\{\alpha_j + \beta(x-t)\}}$, $\alpha_j \geq 0$, for $j = 1, 2, t \leq x, \alpha_2 = c\alpha_1$ and $\lambda_1 = 0.6, \lambda_2 = 0.4, \alpha_1 = 0.05, \beta = 0.03$

c	n	Para-meters	Value of c used in the analysis							
			$c = 1$				$c = 1.1$			
			Bias	ASE	SSE	CP	Bias	ASE	SSE	CP
$c = 1$	50	λ_1	0.623	0.176	0.178	0.969	0.628	0.177	0.178	0.940
		λ_2	0.420	0.141	0.141	0.902	0.421	0.141	0.141	0.923
		α_1	0.055	0.068	0.051	0.981	0.053	0.066	0.049	0.942
	150	λ_1	0.610	0.096	0.095	0.963	0.608	0.096	0.095	0.956
		λ_2	0.408	0.077	0.073	0.900	0.409	0.077	0.074	0.938
		α_1	0.052	0.033	0.031	0.920	0.050	0.032	0.030	0.928
	250	λ_1	0.609	0.074	0.075	0.961	0.608	0.074	0.075	0.956
		λ_2	0.402	0.059	0.058	0.945	0.403	0.059	0.058	0.952
		α_1	0.051	0.025	0.025	0.939	0.049	0.024	0.024	0.922
			$c = 0.09$				$c = 0.1$			
$c = 0.1$	50	λ_1	0.639	0.177	0.178	0.956	0.638	0.177	0.178	0.957
		λ_2	0.420	0.140	0.145	0.946	0.421	0.141	0.146	0.959
		α_1	0.069	0.106	0.075	0.886	0.068	0.105	0.074	0.861
	150	λ_1	0.611	0.096	0.094	0.942	0.610	0.096	0.094	0.943
		λ_2	0.401	0.076	0.077	0.938	0.402	0.077	0.078	0.956
		α_1	0.053	0.052	0.042	0.994	0.052	0.051	0.042	0.989
	250	λ_1	0.598	0.073	0.072	0.936	0.597	0.072	0.072	0.950
		λ_2	0.400	0.057	0.058	0.944	0.401	0.058	0.058	0.954
		α_1	0.051	0.035	0.031	0.944	0.050	0.034	0.030	0.945

Table 5: Simulation results on $\left\{ \underset{\sim}{\hat{\eta}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta}, \hat{\alpha}_1) \right\}$ assuming Weibull failure time distribution (shape parameter $\theta = 0.8$) with $p_j(x,t) = e^{-\alpha_j + \beta(x-t)}$, $\alpha_j \geq 0$, for $j = 1, 2$, $t \leq x$, $\alpha_2 = c\alpha_1$ and $\lambda_1 = 0.6, \lambda_2 = 0.4, \alpha_1 = 0.05, \beta = 0.03$.

Value of c used in both simulation and analysis	n	parameters	Bias	ASE	SSE	CP
$c = 1$	50	λ_1	0.076	0.220	0.257	0.897
		λ_2	0.085	0.191	0.209	0.864
		θ	0.237	0.370	0.524	0.941
		α_1	0.004	0.065	0.053	0.926
	150	λ_1	0.014	0.125	0.129	0.986
		λ_2	0.041	0.111	0.101	0.897
		θ	0.056	0.148	0.135	0.953
		α_1	0.003	0.032	0.031	0.971
	250	λ_1	0.013	0.098	0.097	0.935
		λ_2	0.031	0.086	0.084	0.954
		θ	0.042	0.112	0.113	0.951
		α_1	0.002	0.024	0.023	0.949
$c = 0.1$	50	λ_1	0.074	0.222	0.217	0.959
		λ_2	0.065	0.185	0.183	0.934
		θ	0.208	0.311	0.319	0.982
		α_1	0.009	0.100	0.082	0.835
	150	λ_1	0.018	0.124	0.128	0.956
		λ_2	0.029	0.108	0.110	0.937
		θ	0.069	0.149	0.155	0.974
		α_1	0.006	0.045	0.039	0.900
	250	λ_1	0.005	0.099	0.097	0.954
		λ_2	0.004	0.085	0.084	0.940
		θ	0.002	0.108	0.106	0.969
		α_1	0.006	0.034	0.029	0.925

Table 6: Simulation results on $\left\{ \underset{\sim}{\hat{\eta}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta}, \hat{\alpha}_1) \right\}$ assuming Weibull failure time distribution (shape parameter $\theta = 1.2$) with $p_j(x,t) = e^{-\alpha_j + \beta(x-t)}$, $\alpha_j \geq 0$, for $j = 1, 2$, $t \leq x$, $\alpha_2 = c\alpha_1$ and $\lambda_1 = 0.6, \lambda_2 = 0.4, \alpha_1 = 0.05, \beta = 0.03$.

Value of c used in both simulation and analysis	n	parameters	Bias	ASE	SSE	CP
$c = 1$	50	λ_1	0.061	0.158	0.174	0.929
		λ_2	0.019	0.145	0.159	0.919
		θ	0.607	0.429	0.454	0.978
		α_1	0.006	0.067	0.055	0.901
	150	λ_1	0.013	0.090	0.091	0.956
		λ_2	0.008	0.084	0.091	0.926
		θ	0.458	0.202	0.216	0.971
		α_1	0.005	0.036	0.031	0.964
	250	λ_1	0.006	0.070	0.071	0.954
		λ_2	0.003	0.065	0.070	0.934
		θ	0.422	0.151	0.159	0.945
		α_1	0.005	0.026	0.025	0.957
$c = 0.1$	50	λ_1	0.030	0.152	0.162	0.909
		λ_2	0.056	0.144	0.169	0.897
		θ	0.687	0.557	0.703	0.979
		α_1	0.025	0.106	0.081	0.848
	150	λ_1	0.018	0.089	0.100	0.916
		λ_2	0.014	0.083	0.085	0.947
		θ	0.483	0.206	0.222	0.976
		α_1	0.008	0.065	0.065	0.100
	250	λ_1	0.011	0.069	0.074	0.925
		λ_2	0.008	0.064	0.065	0.950
		θ	0.444	0.152	0.149	0.956
		α_1	0.004	0.036	0.036	0.975

Table 7: Simulation results on on $\hat{\alpha}_1$ and the $\hat{F}_j(\tau_i)$'s assuming Exponential(1) failure time distribution with $\lambda_1 = 0.6, \lambda_2 = 0.4$ and $p_j(\tau_i, \tau_l) = e^{-\{\beta+\alpha_j(\tau_i-\tau_l)\}}$, for $j = 1, 2,\ l \leq i, \beta = 0.01$ (known), $\alpha_2 = c\ \alpha_1$ with $c = 0.1$ (known), $\alpha_1 = 0.05$.

n_i	Para-meters		Monitoring Time (τ_i)				
			0.105	0.288	0.693	1.386	2.303
50	F_1	$10^2 \times$ |Bias|	0.215	0.144	0.073	0.095	0.448
		SSE	0.030	0.048	0.064	0.062	0.060
		ASE	0.034	0.050	0.065	0.074	0.083
		CP	0.878	0.944	0.943	0.974	0.986
	F_2	$10^2 \times$ |Bias|	0.451	0.154	0.190	0.487	0.663
		SSE	0.022	0.039	0.053	0.053	0.056
		ASE	0.029	0.044	0.057	0.068	0.078
		CP	0.993	0.938	0.961	0.975	0.991
	α_1		|Bias| = 0.509	SSE= 0.036	ASE= 0.039	CP= 0.929	
150	F_1	$10^2 \times$ |Bias|	0.125	0.139	0.022	0.026	0.185
		SSE	0.019	0.028	0.036	0.041	0.037
		ASE	0.020	0.029	0.038	0.042	0.047
		CP	0.953	0.947	0.957	0.966	0.986
	F_2	$10^2 \times$ |Bias|	0.034	0.054	0.189	0.295	0.187
		SSE	0.016	0.025	0.032	0.036	0.037
		ASE	0.016	0.024	0.033	0.039	0.044
		CP	0.922	0.938	0.953	0.961	0.982
	α_1		|Bias| = 0.020	SSE= 0.020	ASE= 0.021	CP= 0.934	
250	F_1	$10^2 \times$ |Bias|	0.024	0.093	0.021	0.013	0.148
		SSE	0.015	0.022	0.029	0.030	0.030
		ASE	0.015	0.022	0.029	0.031	0.036
		CP	0.951	0.949	0.956	0.960	0.981
	F_2	$10^2 \times$ |Bias|	0.026	0.034	0.135	0.134	0.155
		SSE	0.012	0.019	0.024	0.028	0.029
		ASE	0.012	0.019	0.025	0.030	0.034
		CP	0.935	0.941	0.950	0.958	0.981
	α_1		|Bias| = 0.004	SSE= 0.016	ASE= 0.016	CP= 0.949	

Table 8: Simulation results on on $\hat{\alpha}_1$ and the $\hat{F}_j(\tau_i)$'s assuming Exponential(1) failure time distribution with $\lambda_1 = 0.6, \lambda_2 = 0.4$ and $p_j(\tau_i, \tau_l) = e^{-\{\beta+\alpha_j(\tau_i-\tau_l)\}}$, for $j = 1, 2,\ l \leq i, \beta = 0.01$ (known), $\alpha_2 = c\alpha_1$ with $c = 1$ (known), $\alpha_1 = 0.05$.

n_i	Para-meters		Monitoring Time (τ_i) 0.105	0.288	0.693	1.386	2.303
50	F_1	$10^2 \times$ |Bias|	0.168	0.467	0.157	0.351	0.251
		SSE	0.030	0.049	0.062	0.059	0.060
		ASE	0.034	0.050	0.065	0.074	0.081
		CP	0.858	0.934	0.940	0.982	0.992
	F_2	$10^2 \times$ |Bias|	0.497	0.045	0.033	0.114	0.525
		SSE	0.024	0.039	0.052	0.052	0.055
		ASE	0.029	0.044	0.057	0.067	0.075
		CP	0.993	0.937	0.940	0.982	0.985
	α_1		|Bias| = 0.375	SSE= 0.024	ASE= 0.027	CP= 0.919	
150	F_1	$10^2 \times$ |Bias|	0.162	0.057	0.058	0.238	0.229
		SSE	0.020	0.030	0.037	0.039	0.038
		ASE	0.019	0.029	0.038	0.042	0.046
		CP	0.947	0.934	0.945	0.965	0.983
	F_2	$10^2 \times$ |Bias|	0.046	0.033	0.026	0.049	0.104
		SSE	0.017	0.025	0.033	0.034	0.036
		ASE	0.016	0.024	0.032	0.038	0.043
		CP	0.928	0.942	0.946	0.973	0.973
	α_1		|Bias| = 0.062	SSE= 0.014	ASE= 0.016	CP= 0.975	
250	F_1	$10^2 \times$ |Bias|	0.146	0.051	0.019	0.226	0.223
		SSE	0.014	0.023	0.029	0.031	0.030
		ASE	0.015	0.023	0.029	0.033	0.035
		CP	0.948	0.934	0.947	0.965	0.977
	F_2	$10^2 \times$ |Bias|	0.027	0.027	0.026	0.043	0.024
		SSE	0.013	0.018	0.025	0.027	0.028
		ASE	0.012	0.018	0.025	0.030	0.033
		CP	0.935	0.955	0.953	0.970	0.971
	α_1		|Bias| = 0.048	SSE= 0.011	ASE= 0.012	CP= 0.960	

Table 9: Simulation results on on $\hat{\alpha}_1$ and the $\hat{F}_j(\tau_i)$'s assuming Exponential(1) failure time distribution with $\lambda_1 = 0.6, \lambda_2 = 0.4$ and $p_j(\tau_i, \tau_l) = e^{-\{\alpha_j + (\tau_i - \tau_l)\beta\}}$, for $j = 1, 2,\ l \le i, \beta = 0.01$ (known), with $\alpha_2 = c\alpha_1$, $c = 0.1$ (known), $\alpha_1 = 0.05$.

n_i	Para-meters		Monitoring Time (τ_i)				
			0.105	0.288	0.693	1.386	2.303
50	F_1	$10^2 \times$ |Bias|	3.167	1.352	4.100	1.905	2.561
		SSE	0.031	0.049	0.061	0.059	0.059
		ASE	0.034	0.051	0.066	0.075	0.083
		CP	0.843	0.951	0.956	0.981	0.996
	F_2	$10^2 \times$ |Bias|	3.454	1.009	6.551	7.995	1.304
		SSE	0.022	0.039	0.050	0.053	0.057
		ASE	0.028	0.043	0.056	0.067	0.077
		CP	0.993	0.924	0.956	0.977	0.985
	α_1		|Bias|=0.007	SSE= 0.029	ASE= 0.028	CP= 0.931	
150	F_1	$10^2 \times$ |Bias|	0.573	0.808	1.591	1.550	1.049
		SSE	0.019	0.029	0.037	0.039	0.038
		ASE	0.019	0.029	0.038	0.043	0.047
		CP	0.939	0.943	0.947	0.960	0.990
	F_2	$10^2 \times$ |Bias|	0.167	0.188	0.125	2.520	1.106
		SSE	0.016	0.024	0.032	0.035	0.035
		ASE	0.016	0.025	0.033	0.039	0.044
		CP	0.941	0.932	0.953	0.970	0.989
	α_1		|Bias|= 0.005	SSE= 0.017	ASE= 0.016	CP= 0.961	
250	F_1	$10^2 \times$ |Bias|	0.417	0.407	1.243	1.068	0.897
		SSE	0.015	0.023	0.029	0.031	0.030
		ASE	0.015	0.023	0.029	0.033	0.037
		CP	0.942	0.947	0.957	0.962	0.986
	F_2	$10^2 \times$ |Bias|	0.038	0.053	0.088	1.383	0.694
		SSE	0.012	0.019	0.026	0.027	0.029
		ASE	0.012	0.019	0.026	0.030	0.034
		CP	0.919	0.942	0.940	0.973	0.985
	α_1		|Bias|= 0.005	SSE= 0.013	ASE= 0.013	CP= 0.949	

Table 10: Simulation results on on $\hat{\alpha}_1$ and the $\hat{F}_j(\tau_i)$'s assuming Exponential(1) failure time distribution with $\lambda_1 = 0.6, \lambda_2 = 0.4$ and $p_j(\tau_i, \tau_l) = e^{-\{\alpha_j + (\tau_i - \tau_l)\beta\}}$, for $j = 1, 2,\ l \leq i, \beta = 0.01$ (known), with $\alpha_2 = c\alpha_1$, $c = 1$ (known), $\alpha_1 = 0.05$.

n_i	Para-meters		Monitoring Time (τ_i)				
			0.105	0.288	0.693	1.386	2.303
50	F_1	$10^2 \times$ |Bias|	3.126	1.581	1.112	4.029	1.874
		SSE	0.032	0.049	0.061	0.064	0.065
		ASE	0.034	0.051	0.066	0.076	0.084
		CP	0.863	0.954	0.958	0.980	0.993
	F_2	$10^2 \times$ |Bias|	3.894	0.638	4.736	1.607	3.307
		SSE	0.022	0.039	0.052	0.055	0.063
		ASE	0.029	0.045	0.057	0.069	0.079
		CP	0.992	0.956	0.950	0.983	0.985
	α_1		|Bias| = 0.004	SSE= 0.022	ASE= 0.024	CP= 0.931	
150	F_1	$10^2 \times$ |Bias|	0.140	0.659	1.030	1.035	1.713
		SSE	0.020	0.029	0.037	0.040	0.038
		ASE	0.019	0.030	0.038	0.044	0.048
		CP	0.922	0.943	0.953	0.965	0.989
	F_2	$10^2 \times$ |Bias|	0.094	0.289	1.030	0.558	1.215
		SSE	0.017	0.025	0.033	0.034	0.038
		ASE	0.016	0.024	0.033	0.040	0.045
		CP	0.915	0.938	0.947	0.977	0.987
	α_1		|Bias| = 0.003	SSE= 0.013	ASE= 0.013	CP= 0.938	
250	F_1	$10^2 \times$ |Bias|	0.008	0.396	0.572	0.558	0.580
		SSE	0.015	0.022	0.029	0.032	0.032
		ASE	0.015	0.022	0.030	0.034	0.037
		CP	0.939	0.952	0.954	0.966	0.977
	F_2	$10^2 \times$ |Bias|	0.008	0.066	0.006	0.056	0.580
		SSE	0.012	0.019	0.026	0.029	0.030
		ASE	0.012	0.019	0.026	0.031	0.035
		CP	0.931	0.947	0.955	0.962	0.980
	α_1		|Bias| = 0.003	SSE= 0.010	ASE= 0.010	CP= 0.952	

6 Data Analysis

In this section, we illustrate the proposed methodology, both parametric and non-parametric, using a real data set on hearing loss. This data on hearing loss consists of 795 respondents collected from the Department of Speech and Hearing Disabilities (Divyangjan), Eastern Regional Centre, who appeared for the diagnosis of hearing loss during the months of January and February, 2017 (See Banik et al. (2018)). The monitoring time is taken to be the age at which a patient appears for diagnosis. There are two main types of hearing loss depending on which part of the ear is affected: i) Sensorineural hearing loss (SNHL) and ii) Conductive hearing loss. A third type, named mixed hearing loss, is a combination of SNHL and Conductive hearing loss which can be interpreted as the missing type. The event of interest is defined as hearing loss due to the type occurring first.

Parametric analysis of the data is done by first fitting exponential distribution and then Weibull distribution for the failure time using the methods of Section 3. The performance of the models is assessed by a goodness of fit test based on a modified χ^2 statistic defined as $\chi^2_M = \sum_{i=1}^{n} \sum_{g\in\mathcal{G}} \left[\dfrac{(\delta_{gi} - p^*_{gi})^2}{p^*_{gi}(1-p^*_{gi})}\right]$, where $p^*_{gi} = \displaystyle\int_0^{x_i} \mathrm{e}^{-\alpha_j(x_i-t)} f_j(t;\underset{\sim}{\theta})\mathrm{d}t$, if $g = \{j\}$, for $j = 1,2$, $p^*_{gi} = \displaystyle\int_0^{x_i} \left[(1-\mathrm{e}^{-\alpha_1(x_i-t)})f_1(t;\underset{\sim}{\theta}) + (1-\mathrm{e}^{-\alpha_2(x_i-t)})f_2(t;\underset{\sim}{\theta})\right]\mathrm{d}t$, if $g = \{1,2\}$ and $p^*_{gi} = S(x_i;\underset{\sim}{\theta})$, if $g = \phi$, all evaluated at the MLEs of the model parameters. Note that the δ_{gi}'s are as defined in Section 3.3 with corresponding expectations estimated by the p^*_{gi}'s. The p-values for the test is computed using Monte Carlo simulation method discussed in the following. First we estimate the model parameters using the proposed method and the observed data. We then compute the value of the modified χ^2_M statistic using the observed data. Next we simulate a large number of data sets of the same sample size n from the assumed model with the estimated model parameters and compute χ^2_M for each simulated data set. This gives an estimate of the null distribution of the χ^2_M statistic. Then, the corresponding p-value is computed as the proportion of times the computed value of χ^2_M in each simulation exceeds its observed value. To compare the models, we also compute the values of the Akaike Information criterion (AIC) and the Akaike Information criterion corrected (AICc) as

$$AIC = (2*p) - 2\ln(\hat{L}) \quad \text{and} \quad AICc = AIC + \frac{2p^2+2p}{n-p-1},$$

respectively, where p is the number of parameters to be estimated and $\hat{L}$ is the value of the likelihood function computed at the MLEs (See Burnham and Anderson (2002)). We have considered the two models (M1) and (M2) for the masking probabilities (See Subsections 3.1 and 3.2) with $\alpha_2 = c\ \alpha_1$ and analyzed the data under both Exponential and Weibull distributions. We have taken $c = 1$ and 2 for both the models and $\beta = 0.02$ and 0.05 for (M2). The results of such model fitting are presented in Tables 11 and 12 for (M1) and (M2), respectively. For both Exponential and Weibull models, the p-values are nearly 0, much less than 0.05, the level of significance. Thus, both the models fail to fit the data. Note that the time-independent masking probability is a special case of (M2) with $\beta = 0$. We have tried very small $\beta = 0.0001$ and $\alpha_2 = \alpha_1 + f$, so that

$$p_2 \approx e^{-\alpha_2} = e^{-(\alpha_1+f)} = c^{'} e^{-\alpha_1} \approx c^{'} p_1,$$

where $c^{'} = e^{-f}$. The constant $f = -\log 0.8$ is chosen in a way that $c^{'}$ is equal to the value of $c(= 0.8)$ giving the best fit for time-independent masking probabilities under Weibull model (See Koley and Dewanji (2018a)). The results are presented in Table 13. As expected, only the Weibull model gives reasonable fit to this data.

In the non-parametric analysis, we first split the range of monitoring times into five sub-intervals as [0, 5], [5, 15], [15, 30], [30, 50] and [50, 90]. Thus, the new set of discrete, finite monitoring time points, at which the sub-distribution functions are to be estimated, are 5, 15, 30, 50 and 90. Using these intervals, the data is summarized as $\{n_i, d_{gi}, g \in \mathcal{G}, i = 1, \cdots, 5\}$. The MLEs of the sub-distribution functions along with the standard errors are obtained at these monitoring time points using the methods of Section 4 with both (M3) and (M4) and $\alpha_2 = c\ \alpha_1$. Also, different choices of (c, β) are considered and the corresponding AIC values are computed to determine the value of (c, β) that fits the data best among the choices considered. The results are presented in Tables 14 and 15 for (M3) and (M4), respectively, with the best fit in bold.

Table 11: Parametric Analysis of Hearing Loss Data under (M1) with $\alpha_2 = c\ \alpha_1$.

Model	c	MLE of parameters with standard errors in parentheses							
		λ_1	λ_2	θ	α_1	χ^2_M	AIC	AICc	p-value
Exponential	1	0.089 (0.005)	0.006 (0.001)	-	0.006 (0.001)	6101.964	1375.443	1375.473	≈ 0
	2	0.088 (0.005)	0.007 (0.001)	-	0.006 (0.001)	6134.888	1380.62	1380.671	≈ 0
Weibull	1	0.097 (0.007)	0.002 (0.001)	0.704 (0.042)	0.006 (0.001)	2686.148	1340.75	1340.801	≈ 0
	2	0.096 (0.007)	0.003 (0.001)	0.695 (0.044)	0.006 (0.001)	2697.763	1343.743	1343.794	≈ 0

Table 12: Parametric Analysis of Hearing Loss Data under (M2) with $\alpha_2 = c\ \alpha_1$.

Model	(c, β)	MLE of parameters with standard errors in parentheses							
		λ_1	λ_2	θ	α_1	χ^2_M	AIC	AICc	p-value
Expo-nential	(1,0.02)	0.087 (0.005)	0.006 (0.001)	-	0 (0.009)	5334.811	1389.119	1389.149	≈ 0
	(2,0.02)	0.087 (0.005)	0.006 (0.001)	-	0 (0.009)	5357.333	1389.121	1389.151	≈ 0
	(1,0.05)	0.064 (0.004)	0.005 (0.001)	-	0 (0.324)	11506.03	2770.772	2770.802	≈ 0
	(2,0.05)	0.064 (0.004)	0.005 (0.001)	-	0 (0.336)	20784.28	2834.132	2834.162	≈ 0
Wei-bull	(1,0.02)	0.094 (0.007)	0.002 ($\approx$ 0.0)	0.705 (0.041)	0 (0.011)	2425.161	1357.377	1357.428	≈ 0
	(2,0.02)	0.090 (0.007)	0.003 (0.001)	0.744 (0.043)	0 (0.020)	2577.684	1463.196	1463.247	≈ 0
	(1,0.05)	0.066 (0.004)	0.003 (0.001)	0.814 (0.040)	0 (0.049)	4167.853	1847.774	1847.825	≈ 0
	(2,0.05)	0.066 (0.004)	0.003 (0.001)	0.814 (0.040)	0 (0.046)	4167.944	1847.78	1847.831	≈ 0

Table 13: Parametric Analysis of Hearing Loss Data under (M2) with $\alpha_2 = \alpha_1 + f$ and $f = -\log 0.8 = 0.223,\ \beta = 0.0001$.

Model	MLE of parameters with							
	λ_1	λ_2	θ	α_1	χ^2_M	AIC	AICc	p-value
Expo-nential	0.087 (0.005)	0.008 (0.001)	-	0.101 (0.045)	5327.703	1429.281	1429.311	≈ 0
Weibull	0.095 (0.007)	0.003 (0.001)	0.695 (0.045)	0.100 (0.016)	2257.575	1393.912	1393.963	0.890

Table 14: Non-Parametric Analysis of Hearing Loss Data under (M3) with $\alpha_2 = c\ \alpha_1$.

(c, β)	F_j	MLE with standard errors in parantheses, Monitoring Times (τ_i): 5	15	30	50	90	α_1	*AIC*
(0.09,0.01)	F_1	0.216	0.659	0.814	0.814	0.850	0.009	**1343.64**
		(0.028)	(0.028)	(0.042)	(0.046)	(0.038)	(0.006)	
	F_2	0.005	0.023	0.045	0.100	0.100		
		(0.005)	(0.009)	(0.021)	(0.025)	(0.023)		
(0.09,0.03)	F_1	0.162	0.607	0.727	0.738	0.887	0.006	1551.612
		(0.022)	(0.028)	(0.042)	(0.031)	(0.024)	(0.004)	
	F_2	0.004	0.021	0.034	0.081	0.086		
		(0.003)	(0.008)	(0.013)	(0.016)	(0.012)		
(0.09,0.05)	F_1	0.133	0.557	0.649	0.726	0.896	0.005	1694.75
		(0.018)	(0.027)	(0.031)	(0.023)	(0.021)	(0.004)	
	F_2	0.003	0.019	0.028	0.066	0.080		
		(0.003)	(0.007)	(0.009)	(0.012)	(0.010)		
(1,0.01)	F_1	0.216	0.660	0.815	0.815	0.850	0.009	1343.818
		(0.028)	(0.028)	(0.041)	(0.046)	(0.038)	(0.006)	
	F_2	0.005	0.023	0.046	0.099	0.099		
		(0.005)	(0.009)	(0.021)	(0.025)	(0.022)		
(1,0.03)	F_1	0.162	0.607	0.724	0.738	0.887	0.005	1551.804
		(0.022)	(0.028)	(0.042)	(0.031)	(0.024)	(0.004)	
	F_2	0.004	0.021	0.033	0.080	0.086		
		(0.004)	(0.008)	(0.012)	(0.016)	(0.012)		
(1,0.05)	F_1	0.132	0.556	0.650	0.726	0.897	0.005	1694.924
		(0.018)	(0.027)	(0.031)	(0.023)	(0.021)	(0.003)	
	F_2	0.003	0.019	0.028	0.065	0.080		
		(0.003)	(0.007)	(0.009)	(0.012)	(0.010)		
(1.2,0.01)	F_1	0.216	0.659	0.814	0.814	0.850	0.008	1343.854
		(0.028)	(0.028)	(0.041)	(0.045)	(0.038)	(0.006)	
	F_2	0.005	0.023	0.047	0.099	0.099		
		(0.005)	(0.009)	(0.021)	(0.025)	(0.023)		
(1.2,0.03)	F_1	0.162	0.606	0.727	0.738	0.887	0.005	1551.832
		(0.022)	(0.028)	(0.042)	(0.031)	(0.024)	(0.004)	
	F_2	0.004	0.021	0.034	0.080	0.086		
		(0.004)	(0.008)	(0.013)	(0.016)	(0.012)		
(1.2,0.05)	F_1	0.132	0.556	0.651	0.727	0.897	0.005	1694.958
		(0.018)	(0.027)	(0.031)	(0.023)	(0.021)	(0.003)	
	F_2	0.003	0.019	0.028	0.066	0.080		
		(0.003)	(0.007)	(0.009)	(0.012)	(0.010)		

Table 15: Non-Parametric Analysis of Hearing Loss Data under (M4) with $\alpha_2 = c\,\alpha_1$.

(c, β)	F_j	MLE with standard errors in parantheses Monitoring Times (τ_i) 5	15	30	50	90	α_1	AIC
(2.623,0.001)	F_1	0.194	0.405	0.508	0.512	0.515	0.005	4242.674
		(0.025)	(0.023)	(0.041)	(0.043)	(0.036)	(0.0007)	-
	F_2	0.005	0.014	0.031	0.093	0.093		
		(0.004)	(0.005)	(0.013)	(0.021)	(0.019)		
(2.623,0.005)	F_1	0.194	0.401	0.522	0.548	0.561	0.006	4237.528
		(0.025)	(0.022)	(0.038)	(0.041)	(0.037)	(0.0008)	
	F_2	0.005	0.014	0.027	0.062	0.062		
		(0.005)	(0.005)	(0.013)	(0.017)	(0.023)		
(1,0.001)	F_1	0.190	0.402	0.528	0.539	0.543	0.005	4245.234
		(0.025)	(0.022)	(0.039)	(0.042)	(0.036)	(0.0007)	
	F_2	0.004	0.014	0.028	0.075	0.075		
		(0.004)	(0.005)	(0.013)	(0.019)	(0.021)		
(1,0.005)	F_1	0.189	0.414	0.499	0.527	0.549	0.005	**4235.642**
		(0.025)	(0.023)	(0.039)	(0.041)	(0.036)	(0.0007)	
	F_2	0.005	0.014	0.029	0.074	0.074		
		(0.004)	(0.005)	(0.014)	(0.019)	(0.020)		

7 Concluding Remarks

In this work, we have considered the situation when the masking probabilities are dependent on both monitoring and failure time points through their difference. This representation of the conditional probabilities has the usual interpretation that, as the difference between monitoring and failure time points increases, the probability of observing the true type decreases. Both parametric and non-parametric methods for maximum likelihood estimation are discussed in this paper. The proposed methods are illustrated using a real data on hearing loss. In parametric estimation, both exponential and Weibull distribution fail to fit the data and, hence, the sub-distribution functions are estimated non-parametrically at some pre-defined monitoring time points obtained by splitting the diagnosis age range of the patients into sub-intervals. In non-parametric analysis, different choices of c and β are considered and the corresponding AIC values are computed.

We have considered only two models describing the relationships between α_1 and α_2. Different other relationships between α_1 and α_2 ensuring identifiability of the model can also be explored. One can also look for other models for the masking probabilities apart from those discussed in this work. Non-parametric analysis with individual monitoring time is diffcult and remains a challenge. In this work, only two competing risks are considered. A more general case may have to consider $m > 2$ competing risks. Sometimes availability of additional information may alleviate the identifiability problem and no further model assumption may then be required. Koley and Dewanji (2020) have considered utilization of two kinds of additional information with time-independent masking probabilities.

References

Austin, P. C., Lee, D. S., and Fine, J. P. (2016). Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, 133(6):601–609.

Banik, P., Jaiswal, J., Koley, T., and Chatterjee., I. (2018). Explanatory analysis of tertiary audiological data. *Asia Pacific Journal of Research.*

Basu, S. (2009). Inference about the masking probabilities in the competing risks model. *Communications in Statistics—Theory and Methods*, 38(16-17):2677–2690.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer-Verlag.

Dewanji, A. and Sengupta, D. (2003). Estimation of competing risks with general missing pattern in failure types. *Biometrics*, 59(4):1063–1070.

Flehinger, B. J., Reiser, B., and Yashchin, E. (1998). Survival with competing risks and masked causes of failures. *Biometrika*, 85(1):151–164.

Hudgens, M. G., Satten, G. A., and Longini Jr, I. M. (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics*, 57(1):74–80.

Jewell, N. P. and Kalbfleisch, J. D. (2004). Maximum likelihood estimation of ordered multinomial parameters. *Biostatistics*, 5(2):291–306.

Jewell, N. P., Van Der Laan, M., and Henneman, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika*, 90(1):183–197.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data, second edition*. John Wiley & Sons.

Koley, T. and Dewanji, A. (2018a). Current status data with two competing risks and missing failure types. `https://www.isical.ac.in/~asu/TR/TechRepASU201802.pdf`.

Koley, T. and Dewanji, A. (2018b). Revisiting non-parametric maximum likelihood estimation of current status data with competing risks. *Sankhya B*, pages 1–21.

Koley, T. and Dewanji, A. (2019). Current status data with two competing risks and missing failure types: A non-parametric approach. `https://www.isical.ac.in/~asu/TR/TechRepASU201901.pdf`.

Koley, T. and Dewanji, A. (2020). Use of additional information for current status data with two competing risks and missing failure types. `https://asu.isical.ac.in/sites/default/files/TechRepASU202003.pdf`.

Lehmann, E. C. and Casella, G. (1998). *Theory of Point Estimation*. Springer.

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

Maathuis, M. H. and Hudgens, M. G. (2011). Nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times. *Biometrika*, pages 325–340.

Mukhopadhyay, C. (2006). Maximum likelihood analysis of masked series system lifetime data. *Journal of statistical planning and inference*, 136(3):803–838.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, 46(1):27–29.

Shafarevich, I. R. (2012). *Discourses on algebra.* Springer Science & Business Media.