Analysis of Strength Distributions of Multi-Modal Failures Using the EM Algorithm

CHANSEOK PARK

Department of Mathematical Sciences, Clemson University, Clemson, SC 29634

W. J. PADGETT

Department of Mathematical Sciences, Clemson University, Clemson, SC 29634

and

Department of Statistics, University of South Carolina, Columbia, SC 29208

Analysis of various multi-modal strength distributions are studied by using competing risks models. This multi-modality may arise due to several kinds of defects in a material. The fracture of a material is controlled by the most severe of all the defects, the so-called "weakest-link theory," which is also commonly referred to as "competing risks" in the statistics literature. These multi-modal problems can also be further complicated due to possible censoring. In practice, censoring is very common because of time and cost considerations on experiments. Moreover, in certain situations, it is observed that the mode of failure is not properly identified due to lack of appropriate diagnostics, expensive and time-consuming autopsy, etc. This is known as the masking problem. Several studies have been carried out, but they have mainly focused on bi-modal Weibull distributions with no censoring or masking considered.

In this paper, we deal with the strength distribution of multi-modal failures when censoring and masking are present. We provide the EM-type parameter estimator for a variety of strength distributions including exponential, Weibull, lognormal and inverse Gaussian distributions, along with useful `R` programs for computation. The applicability of this method is illustrated for several real-data examples.

Key Words: Competing risks, censoring, masking, EM algorithm, MLE, missing data, likelihood function, exponential, Weibull, lognormal, inverse Gaussian (Wald).

---

1 Introduction

Knowledge of the strength of a type of material is required for engineering design of various structures made from such materials in order for the structures to withstand predicted stresses. To determine the strength properties, specimens of the materials are typically tested under laboratory conditions, and appropriate statistical models are investigated in order to predict strengths of specimens or structures of different sizes than those tested. This approach is taken, for example, in the case of modern fibrous composite materials. Due to flaws occurring at random in the material specimens under test, perhaps from various imperfections or other causes, the tensile strength of a single specimen must be considered as a random variable whose probability distribution depends on the various kinds of flaws that are present. Such a probability distribution is used to estimate strengths for the design of larger structures made from the material. Thus, finding appropriate statistical models that fit observed specimen data well is important.

Most statistical analysis of material properties has been studied assuming that the material strength follows a single Weibull distribution which gives a linear Weibull plot. On the other hand, it has been frequently reported by several authors that there are different modes of flaws which determine the fracture of the material. Among them are Johnson and Thorne (1969), Jones and Wilkins (1972), Layden (1973), Boggio and Vingsbo (1976), Beetz (1982), Martineau et al. (1984), Simon and Bunsell (1984), Chi et al. (1984), Goda and Fukunaga (1986), Wagner (1989), Stoner et al. (1994), and Meeker and Escobar (1998), among others.

In the case where there are several potential modes of causes, statistical strength distributions based on "weakest-link theory," which is also commonly referred to as "competing risks" in the statistics literature, have been developed by several authors. Goda and Fukunaga (1986) analyzed the strength distributions of silicon carbide and alumina fibers using a multi-modal Weibull distribution, Wagner (1989) also studied competing risks model, and Taylor (1994) developed a Poisson-Weibull flaw model. In this context, end-effects (or clamp-effects) models were developed by several authors to explain the strengths observed in very small fiber or composite specimens; see Phoenix and Sexsmith (1972), Stoner et al. (1994), and Padgett et al. (1995). They, however, have mainly focused on Weibull distributions and they did not consider censoring or masking problems. Although they stated that their methods extend to general multi-modal Weibull distributions, no explicit illustration was provided. The main reason, we think, is that the parameter estimation under the large number of different failure modes is extremely difficult. It has also been reported that the commonly used Weibull distributions often do not provide good fits to tensile strength data. For example, for carbon fiber or composite tensile strengths, see Durham and Padgett (1997). These motivate the need for developing the highly stable parameter estimation methodology under various distribution models with both censoring and masking considered.

In this paper, we deal with multi-modal problems with censoring and masking under a variety of strength distributions including Weibull, lognormal and inverse Gaussian (Wald) distributions. We provide the EM-type parameter estimator, which is fairly stable in estimation and can handle any number of failure modes. In Section 2, we introduce the competing risks model. We provide the general likelihood method in Section 3. Parameter estimation using the EM algorithm is described in Sections 4 and 5 followed up with real-data examples in Section 6. The `R` source codes are provided in the Appendix.

2 Competing Risks Model

The analysis of lifetime or failure time data has been of considerable interest in many branches of statistical applications such as reliability engineering, electrical engineering, industrial engineering, biological sciences, etc. In an industrial application, a system is made up of multiple components connected in series. In this case, the failure of the whole system is caused by the earliest failure of any of the components, which is commonly referred to as competing risks. In certain situations, it is observed that the determination of the cause of failure may be expensive or may be very difficult to observe due to the lack of appropriate diagnostics. Therefore it might be the case that the failure time of an individual is observed, but its corresponding cause of failure is not fully investigated. This is known as masking. We consider that the cause of the ith system failure may or may not be exactly identified, so the cause-of-failure leads to non-empty subset of labels defining the component in the module. For example, if the ith system with J components fails due to the jth component, then the set of labels is $M_i = \{j\}$ (no masking); if its failure is completely unknown, then $M_i = \{1, 2, \ldots, J\}$ (complete masking); and if its failure is identified by the modes containing more than one failure but not all failures, then $M_i = \{j_1, \ldots, j_i\}$ (partial masking). Moreover, this competing risks problem is further complicated due to possible censoring. In practice, censoring is very common because of time and cost considerations on experiments. The data are said to be censored when, for certain observations, only a lower or upper bound on lifetime is available.

The traditional approach when dealing with competing risks is to consider the hypothetical latent lifetimes corresponding to each cause in the absence of the others (see Moeschberger and David, 1971). We formulate the problem formally using the following notation. A subject is exposed to several potential causes of failure. Let there be a finite number of independent causes of failure indexed by $j = 1, \ldots, J$. Let $T_i^{(j)}$ denote the continuous lifetime of the ith subject due to the jth cause, where $i = 1, \ldots, n$. It is assumed that $T_i^{(j)}$ are independent for all i, j and are *iid* for all i for given j. The corresponding cdf, pdf, survival function, and hazard function of $T_i^{(j)}$ are denoted in general by $F^{(j)}(\cdot|\boldsymbol{\theta}^{(j)})$, $f^{(j)}(\cdot|\boldsymbol{\theta}^{(j)})$, $S^{(j)}(\cdot|\boldsymbol{\theta}^{(j)})$, and $h^{(j)}(\cdot|\boldsymbol{\theta}^{(j)})$, respectively, where $\boldsymbol{\theta}^{(j)}$ is a vector of real valued parameters for each j. Then the observed lifetime of the ith subject is given by the random variable

$$T_i = \min\{T_i^{(1)},\ T_i^{(2)}, \ldots, T_i^{(J)}\}.$$

Typically, in reliability analysis problems, complete observation of T_i may not be possible due to various censoring schemes that can be inherent in data collection. It is further assumed that each T_i can be randomly right-censored by censoring times C_i which are independent of lifetimes T_i for all i. Thus, one observes triplets (X_i, Δ_i, M_i), where $X_i = \min\{T_i, C_i\}$, M_i is the set of labels defining the components that failed, and Δ_i is a censoring indicator variable defined as

$$\Delta_i = \begin{cases} -1 & \text{if masked} \\ j & \text{if failed with } j\text{th cause} \\ 0 & \text{if censored} \end{cases} . \tag{1}$$

We denote a realization of the random variable (X_i, Δ_i) as (x_i, δ_i).

The analysis of exponential data with two causes was studied by Cox (1959), which was extended to multiple causes by Herman and Patell (1971). The parametric estimation problem for the case with two causes with possible missing causes has been discussed by Miyakawa (1984) without censored data. Usher and Hodgson (1988), Usher and Guess (1989), Guess et al. (1991), and Reiser et al. (1995) have considered the masking problem, but they mainly focused on exponential models. They provided closed-form solutions under very restrictive assumptions. Although some authors provided the likelihood function with censored data, no explicit estimates were given. Kundu and Basu (2000) also extended Miyakawa's work to provide the approximate and asymptotic properties of the parameter estimators, confidence intervals, and bootstrap confidence bounds. They provided the exact MLE for the exponential model with only two causes and gave likelihood equations for the Weibull case. However, their exact MLE is applicable only in the complete masking case and the case of censored data was not considered. Although they stated that their solutions extend to the multiple cause case, no explicit expressions were provided. Recently, Park and Kulasekera (2004) extended their work and provided the closed-form MLE for the exponential model with multiple causes, censored data, and completely-masked causes together, but they only considered the case where the lifetime distributions were exponential and Weibull. For the Weibull distribution, the closed-form MLE is available only when the common shape parameter is estimated by the likelihood function. Ishioka and Nonaka (1991) presented a technique to stably estimate the common Weibull shape parameter with two causes using a quasi-Newton method when the data consists only of the system lifetime (the concomitant indicator is unknown). Here, the unknown concomitant indicator is equivalent to the masking problem in our context. Thus, their method can be used for the masking problem, but it is very limited to only two causes and a common shape parameter. Another approach using the EM algorithm was considered by Albert and Baxter (1995). They found the EM sequences for the exponential model with multiple causes, censoring and general masking. However, unless one assumes an exponential distribution for the lifetimes, it is very difficult to apply their idea because it requires that the hazard and survival functions have nice closed forms.

3 Strength Distribution and Likelihood Function

3.1 Strength Distribution

Most multi-modal strength analyses of materials have been studied based on the so-called "weakest link theory" which requires two assumptions (Beetz, 1982; Goda and Fukunaga, 1986):

A1 The material contains inherently many strength-limiting defects, and its strength depends on the weakest defect of all of them.

A2 There are no interactions among the defects.

These assumptions exactly match with the competing risks model under the assumption of the hypothetical latent lifetimes. Using the observed material strengths instead of lifetimes, we can adopt the competing risks model theory in this context. Assume that there are a finite number of independent defects in the material specimen, indexed by $j = 1, \ldots, J$, and let $T_i^{(j)}$ denote the strength of the ith material specimen due to the jth type of defect, where $i = 1, \ldots, n$. Similarly as before, the observed strength of the ith material specimen is given by $T_i = \min\{T_i^{(1)}, \ldots, T_i^{(J)}\}$. Then, we have the following strength distribution of T_i

$$F(t|\mathbf{\Theta}) = 1 - \prod_{j=1}^{J} \left\{ 1 - F^{(j)}(t|\boldsymbol{\theta}^{(j)}) \right\} = 1 - \prod_{j=1}^{J} S^{(j)}(t|\boldsymbol{\theta}^{(j)}),$$

where $\mathbf{\Theta} = \left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \cdots, \boldsymbol{\theta}^{(J)}\right)$. In what follows, we construct the general likelihood function of the parameters. This likelihood function also considers masking and censoring problems.

3.2 Likelihood Function

Let $\mathbb{I}[A]$ be the indicator function of an event A. For convenience, denote $\mathbb{I}_i(j) = \mathbb{I}[\delta_i = j]$. The likelihood function of the censored sample is

$$\begin{aligned} L(\mathbf{\Theta}) &\propto \prod_{i=1}^{n} \Bigg[\Big\{ f^{(1)}(x_i) \prod_{\substack{j=1 \\ j\neq 1}}^{J} S^{(j)}(x_i) \Big\}^{\mathbb{I}_i(1)} \Big\{ S^{(1)}(x_i) \Big\}^{\mathbb{I}_i(0)} \times \\ &\qquad \cdots \times \Big\{ f^{(J)}(x_i) \prod_{\substack{j=1 \\ j\neq J}}^{J} S^{(j)}(x_i) \Big\}^{\mathbb{I}_i(J)} \Big\{ S^{(J)}(x_i) \Big\}^{\mathbb{I}_i(0)} \Bigg] \\ &= \prod_{i=1}^{n} \prod_{j=1}^{J} L_i(\boldsymbol{\theta}^{(j)}), \end{aligned} \tag{2}$$

where

$$L_i(\boldsymbol{\theta}^{(j)}) = \Big\{ f^{(j)}(x_i) \Big\}^{\mathbb{I}_i(j)} \prod_{\substack{\ell=0 \\ \ell\neq j}}^{J} \Big\{ S^{(j)}(x_i) \Big\}^{\mathbb{I}_i(\ell)}. \tag{3}$$

Maximizing $L(\mathbf{\Theta})$ with respect to $\mathbf{\Theta}$ is equivalent to individually maximizing $L(\boldsymbol{\theta}^{(j)})$ for each cause j. Thus we have reduced the joint maximum likelihood problem for a set of J parameters to J separate estimation problems for the single parameter $\boldsymbol{\theta}^{(j)}$. This simplifies the numerical work considerably.

Next, we consider a lifetime of a subject T_i due to an unknown cause of failure (masking), but its cause is known up to being one in a set M_i. We need to find the pdf of T_i and add this into the likelihood function. The cumulative incidence function (CIF) for each jth cause is

$$G(t, j) = \Pr\big\{ T_i \leq t \text{ and } \Delta_i = j \big\} \tag{4}$$

with its corresponding sub-density function

$$g(t, j) = h^{(j)}(t) \prod_{\ell=1}^{J} S^{(\ell)}(t), \quad j = 1, \ldots, J. \tag{5}$$

The pdf of T_i with M_i is given by

$$f^{(M_i)}(t) = \sum_{j\in M_i} g(t, j) = \sum_{j\in M_i} h^{(j)}(t) \prod_{\ell=1}^{J} S^{(\ell)}(t).$$

Denote $\delta_i = -1$ if the cause of failure is unknown. Then the overall likelihood of the censored and masked data is given by

$$L^*(\mathbf{\Theta}) \propto \prod_{i=1}^{n}\prod_{j=1}^{J} L_i(\boldsymbol{\theta}^{(j)}) \times \prod_{i=1}^{n} \left\{ f^{(M_i)}(x_i) \right\}^{\mathbb{I}_i(-1)} = \prod_{i=1}^{n} L_i^*(\mathbf{\Theta}), \tag{6}$$

where

$$L_i^*(\mathbf{\Theta}) = \prod_{j=1}^{J} L_i(\boldsymbol{\theta}^{(j)}) \times \left\{ f^{(M_i)}(x_i) \right\}^{\mathbb{I}_i(-1)}. \tag{7}$$

In general, the closed-form MLE from the likelihood function above is not available and numerical methods are required to maximize $L^*(\mathbf{\Theta})$. One popular method that is often used is the Newton-Raphson method, but a problem with this method is that it can be very sensitive to the choice of starting values and therefore can often fail to converge to a solution. Also, in the case of the likelihood function (7) above, if the number of causes is large, the likelihood can become overparameterized and the Newton-Raphson method becomes totally ineffective. The difficulty with using direct maximization of the likelihood in (7) is overcome through the use of the EM algorithm discussed in the following section.

4 The EM Algorithm and Likelihood Construction

In this section, we introduce the EM algorithm and develop the likelihood functions which can be conveniently used as inputs in the E-step of the EM algorithm.

4.1 The EM Algorithm

The EM algorithm is a general iterative approach for computing the MLE of parametric models when there are no closed-form ML estimates, or the data are incomplete. The EM algorithm was introduced by Dempster et al. (1977) to overcome the above difficulties. The main references for the EM are Schafer (1997), Little and Rubin (2002), and Tanner (1996).

The EM algorithm consists of an expectation step (E-step) and a maximization step (M-step). The advantage of the EM algorithm is that it solves a difficult incomplete-data problem by constructing two easy steps. The E-step only needs to compute the conditional expectation of the log-likelihood with respect to the incomplete data given the observed data. The M-step needs to find the maximizer of this expected likelihood. An additional advantage of this method compared to other optimization techniques is that it is very simple and it converges reliably. In general, if it converges, it converges to a local maximum. Hence in the case of the unimodal and concave likelihood function, the EM algorithm converges to the global maximizer from any starting value. Below, we provide a short summary of the EM algorithm when it is applied in the missing-data framework.

Let $\boldsymbol{\theta}$ be the vector of unknown parameters. Then the complete-data likelihood is

$$L^C(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^{n} f(x_i).$$

Denote the observed part of $\mathbf{x} = (x_1, \ldots, x_n)$ by $\mathbf{y} = (y_1, \ldots, y_m)$ and the missing part by $\mathbf{z} = (z_{m+1}, \ldots, z_n)$, and denote the estimate at the sth EM sequence by $\boldsymbol{\theta}_s$. The EM algorithm consists of two distinct steps:

- E-step: `Compute` $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_s)$,
  where $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_s) = \int \log L^C(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}) p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}_s) d\mathbf{z}$.

- M-step: `Find` $\boldsymbol{\theta}_{s+1}$
  which maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_s)$ over $\boldsymbol{\theta}$.

4.2 Application of the EM to Competing Risks Model

The question is whether we can apply the EM algorithm to the competing risks problem. When the data are masked, this is equivalent to the cause of failure being missing, so we can construct the complete-data likelihood, $L_i^C(\boldsymbol{\Theta})$, by treating the cause of failure as missing data. Constructing the complete-data likelihood is not difficult once we introduce an indicator variable. Define $U_i^{(j)} = \mathbb{I}[\Delta_i = j | X_i = x_i]$ for $j = 1, \ldots, J$. Then $U_i^{(j)}$ has a Bernoulli distribution with $\Pr\big\{U_i^{(j)} = 1\big\} = \Pr\big\{\Delta_i = j | X_i = x_i\big\}$. It follows that

$$E\big[U_i^{(j)}\big] = \begin{cases} \dfrac{h^{(j)}(x_i)}{\displaystyle\sum_{\ell \in M_i} h^{(\ell)}(x_i)} & \text{if } \ j \in M_i \\ 0 & \text{if } \ j \notin M_i \end{cases}.$$

Replacing $f^{(M_i)}(x_i)$ with $\prod_{j=1}^{J} \big\{f^{(j)}(x_i)\big\}^{U_i^{(j)}} \big\{S^{(j)}(x_i)\big\}^{1-U_i^{(j)}}$ in (7), we have the complete-data likelihood of the censored and masked data as follows:

$$L_i^C(\boldsymbol{\Theta}) = \prod_{j=1}^{J} L_i^C(\boldsymbol{\theta}^{(j)}),$$

where

$$\begin{aligned} L_i^C(\boldsymbol{\theta}^{(j)}) &= \Big\{f^{(j)}(x_i)\Big\}^{\mathbb{I}_i(j)} \prod_{\substack{\ell=0 \\ \ell \neq j}}^{J} \Big\{S^{(j)}(x_i)\Big\}^{\mathbb{I}_i(\ell)} \times \left[\Big\{f^{(j)}(x_i)\Big\}^{U_i^{(j)}} \Big\{S^{(j)}(x_i)\Big\}^{1-U_i^{(j)}}\right]^{\mathbb{I}_i(-1)} \\ &= \Big\{h^{(j)}(x_i)\Big\}^{\mathbb{I}_i(j) + U_i^{(j)} \mathbb{I}_i(-1)} \prod_{\ell=-1}^{J} \Big\{S^{(j)}(x_i)\Big\}^{\mathbb{I}_i(\ell)}. \end{aligned} \tag{8}$$

If $\delta_i = j$, then clearly $M_i = \{j\}$ and thus $E[U_i^{(j)}] = 1$. It follows that

$$\mathbb{I}_i(j) + U_i^{(j)} \mathbb{I}_i(-1) = U_i^{(j)}.$$

Using this and $\sum_{\ell=-1}^{J} \mathbb{I}_i(\ell) = 1$, we can simplify (8) as follows

$$L_i^C(\boldsymbol{\theta}^{(j)}) = \Big\{h^{(j)}(x_i)\Big\}^{U_i^{(j)}} \times S^{(j)}(x_i). \tag{9}$$

Now, because the likelihood $L_i^C(\boldsymbol{\Theta})$ is fully factorized by $L_i^C(\boldsymbol{\theta}^{(j)})$, the estimation problem can be solved individually for each single parameter $\boldsymbol{\theta}^{(j)}$. So, just as we did in (3) and (7), by using this factorized complete-data likelihood instead of $L_i^*(\boldsymbol{\Theta})$, we have reduced the joint maximum likelihood problem for a set of J parameters to J individual estimation problems each with a single parameter $\boldsymbol{\theta}^{(j)}$. So, although the likelihood in (7) is not easy to solve because of numerical difficulties, considering the masked data as missing data and applying an EM framework allows one to obtain a likelihood which is made up of individual likelihoods for each parameter $\boldsymbol{\theta}^{(j)}$. Therefore, the transformation of the problem to a missing-data problem simplifies the numerical difficulties considerably. Nevertheless, it still may not be obvious how the EM algorithm is implemented in the missing-data case and this is discussed in the next section.

4.3 EM Implementation Issues

When the distribution for the lifetimes is assumed to be exponential and the data is censored and masked, we can easily implement an EM algorithm using (9) since the hazard and survival functions are of closed forms. On the other hand, suppose one wants to consider the case where the lifetimes have the normal distribution and the data consist of both censored and masked observations. The application of (9) is clearly not straightforward because the hazard and survival functions do not have closed forms and the overall likelihood cannot be written as a product of individual likelihoods each with a single parameter. Yet, by treating the censored observations as missing data, it is possible to write the complete-data likelihood in (9) as closed-form pdf's. Using this "trick" of treating the censored data as missing data can be thought of as a general "approach" that will allow one to find the closed form independently of the distribution assumed for the lifetimes. Therefore, the EM algorithm can be easily implemented. The approach can be applied to a variety of distributions including the exponential, normal, lognormal and Laplace distributions. Below, we show just how to obtain (9) as closed-form pdf's by treating the censored data as missing data.

Let Z_i be a truncation of X_i at x_i with $Z_i > x_i$. Then we have the complete-data likelihood corresponding to (9)

$$\begin{aligned} L_i^C(\boldsymbol{\theta}^{(j)}) &= \left\{f^{(j)}(x_i)\right\}^{\mathbb{I}_i(j)} \prod_{\substack{\ell=0 \\ \ell\neq j}}^{J} \left\{f^{(j)}(Z_i)\right\}^{\mathbb{I}_i(\ell)} \times \left[\left\{f^{(j)}(x_i)\right\}^{U_i^{(j)}}\left\{f^{(j)}(Z_i)\right\}^{1-U_i^{(j)}}\right]^{\mathbb{I}_i(-1)} \\ &= \left\{f^{(j)}(x_i)\right\}^{U_i^{(j)}}\left\{f^{(j)}(Z_i)\right\}^{1-U_i^{(j)}}, \end{aligned} \tag{10}$$

where the pdf of Z_i is given by

$$f_Z^{(j)}(t|\boldsymbol{\theta}^{(j)}) = \frac{f^{(j)}(t)}{1-F^{(j)}(x_i)}$$

for $t > x_i$.

In the section following, using (9) or (10), we estimate the parameters of a variety of distributions for the material strengths and then show how doing so allows one to obtain simple closed forms in the M-step of the EM algorithm.

5 Parameter Estimation

In this section, we develop the EM-type MLE of the parameters of a variety of strength distributions including exponential, Weibull, normal, lognormal and inverse Gaussian distributions.

5.1 Exponential Distribution Model

In the exponential case, the EM sequences can be obtained by either (9) or (10) without using numerical optimization in the M-step since both the hazard and survival functions are of closed forms.

We assume that $T_i^{(j)}$ is an exponential random variable with the rate parameter $\boldsymbol{\theta}^{(j)} = (\lambda^{(j)})$. Thus, the pdf of $T_i^{(j)}$ is

$$f^{(j)}(t) = \lambda^{(j)} \exp(-\lambda^{(j)} t).$$

First, we obtain an EM sequence using (9). Using $h^{(j)}(x_i) = \lambda^{(j)}$ and $S^{(j)}(x_i) = \exp\big(-\lambda^{(j)} x_i\big)$, we have the complete-data log-likelihood of $\lambda^{(j)}$:

$$\log L_i^C(\lambda^{(j)}) = U_i^{(j)} \log \lambda^{(j)} - \lambda^{(j)} x_i.$$

Let $\boldsymbol{\Theta} = (\lambda^{(1)}, \ldots, \lambda^{(J)})$ and denote the estimate of $\boldsymbol{\Theta}$ at the sth EM sequence by $\boldsymbol{\Theta}_s = (\lambda_s^{(1)}, \ldots, \lambda_s^{(J)})$.

- E-step:

  It follows from $Q_i(\lambda^{(j)}|\boldsymbol{\Theta}_s) = E\big[\log L_i^C(\lambda^{(j)})|\boldsymbol{\Theta}_s\big]$ that

$$Q_i(\lambda^{(j)}|\boldsymbol{\Theta}_s) = \Upsilon_{i,s}^{(j)} \log \lambda^{(j)} - \lambda^{(j)} x_i,$$

where

$$\Upsilon_{i,s}^{(j)} = E[U_i^{(j)}|\mathbf{\Theta}_s] = \begin{cases} \dfrac{\lambda_s^{(j)}}{\sum\limits_{\ell \in M_i} \lambda_s^{(\ell)}} & \text{if } \ j \in M_i \\ 0 & \text{if } \ j \notin M_i \end{cases}.$$

It is worth mentioning that the above $Q_i(\cdot)$ function using (9) coincides with the equation (3.1) of Albert and Baxter (1995).

- M-step:

  Differentiating $Q(\lambda^{(j)}|\mathbf{\Theta}_s) = \sum_{i=1}^n Q_i(\lambda^{(j)}|\mathbf{\Theta}_s)$ with respect to $\lambda^{(j)}$ and setting this to zero, we obtain

$$\sum_{i=1}^n \frac{\partial Q_i(\lambda^{(j)}|\mathbf{\Theta}_s)}{\partial \lambda^{(j)}} = \sum_{i=1}^n \frac{\Upsilon_{i,s}^{(j)}}{\lambda^{(j)}} - \sum_{i=1}^n x_i = 0.$$

  Solving for $\lambda^{(j)}$, we obtain the $(s+1)$th EM sequence in the M-step

$$\lambda_{s+1}^{(j)} = \frac{\sum_{i=1}^n \Upsilon_{i,s}^{(j)}}{\sum_{i=1}^n x_i}. \tag{11}$$

Next, we can obtain a different EM sequence using (10) instead of (9). We have the complete-data log-likelihood of $\lambda^{(j)}$:

$$\log L_i^C(\lambda^{(j)}) = U_i^{(j)}(\log \lambda^{(j)} - \lambda^{(j)} x_i) + (1 - U_i^{(j)})(\log \lambda^{(j)} - \lambda^{(j)} Z_i).$$

- E-step:

  It follows from $E[Z_i|\mathbf{\Theta}_s] = 1/\lambda_s^{(j)} + x_i$ that

$$Q_i(\lambda^{(j)}|\mathbf{\Theta}_s) = \log \lambda^{(j)} - \lambda^{(j)} x_i - (1 - \Upsilon_{i,s}^{(j)}) \frac{\lambda^{(j)}}{\lambda_s^{(j)}}.$$

- M-step:

  Differentiating $Q(\lambda^{(j)}|\mathbf{\Theta}_s) = \sum_{i=1}^n Q_i(\lambda^{(j)}|\mathbf{\Theta}_s)$ with respect to $\lambda^{(j)}$ and setting this to zero, we obtain

$$\sum_{i=1}^n \frac{\partial Q_i(\lambda^{(j)}|\mathbf{\Theta}_s)}{\partial \lambda^{(j)}} = \frac{n}{\lambda^{(j)}} - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{1 - \Upsilon_{i,s}^{(j)}}{\lambda_s^{(j)}} = 0.$$

Solving for $\lambda^{(j)}$, we obtain the $(s+1)$th EM sequence in the M-step

$$\lambda_{s+1}^{(j)} = \frac{n}{\sum_{i=1}^{n} x_i + \frac{1}{\lambda_s^{(j)}} \sum_{i=1}^{n} \left(1 - \Upsilon_{i,s}^{(j)}\right)}. \tag{12}$$

Note that in the limit as $s \to \infty$ the equation (12) becomes

$$\lambda_{\infty}^{(j)} = \frac{n}{\sum_{i=1}^{n} x_i + \frac{1}{\lambda_{\infty}^{(j)}} \sum_{i=1}^{n} \left(1 - \Upsilon_{i,\infty}^{(j)}\right)}.$$

Solving for $\lambda_{\infty}^{(j)}$, we have

$$\lambda_{\infty}^{(j)} = \frac{\sum_{i=1}^{n} \Upsilon_{i,\infty}^{(j)}}{\sum_{i=1}^{n} x_i}.$$

Therefore, although the above EM sequence (12) is different from (11), they give the same limiting estimates.

It is also worth noting that if we solve the stationary-point equations $\hat{\lambda}^{(j)} = \lambda_{s+1}^{(j)} = \lambda_s^{(j)}$ using the above results (11) and (12) with only complete masking considered, then both solutions give

$$\hat{\lambda}^{(j)} = \left\{1 + \frac{n^{(-1)}}{\sum_{j=1}^{J} n^{(j)}}\right\} \frac{n^{(j)}}{\sum_{i=1}^{n} x_i},$$

where $n^{(\ell)} = \sum_{i=1}^{n} \mathbb{I}_i(\ell)$ for $\ell = -1, 0, 1, \ldots, J$. As expected, this result is the same as that of Park and Kulasekera (2004) with a single group.

5.2 Weibull Distribution Model

In the case of the Weibull models, the EM sequence can be obtained by either (9) or (10). For this model, we used (9). In the M-step, we need to estimate the shape parameter $\alpha^{(j)}$ numerically, but this is only a one-dimensional root search and the uniqueness of this solution is guaranteed. Lower and upper bounds for the root are explicitly obtained, so with these bounds we can find the root easily. We provide the sketch of the proof of the uniqueness under quite reasonable conditions and give lower and upper bounds of $\alpha^{(j)}$ in the Appendix.

We assume that $T_i^{(j)}$ is a Weibull random variable with the parameter vector $\boldsymbol{\theta}^{(j)} = (\alpha^{(j)}, \lambda^{(j)})$. Thus, the pdf and cdf of $T_i^{(j)}$ are

$$\begin{aligned} f^{(j)}(t) &= \alpha^{(j)}\lambda^{(j)} t^{\alpha^{(j)}-1} \exp(-\lambda^{(j)} t^{\alpha^{(j)}}) \\ F^{(j)}(t) &= 1 - \exp(-\lambda^{(j)} t^{\alpha^{(j)}}). \end{aligned}$$

First, we obtain an EM sequence using (9). Using $h^{(j)}(x_i) = \alpha^{(j)}\lambda^{(j)} x_i^{\alpha^{(j)}-1}$ and $S^{(j)}(x_i) = \exp\big(-\lambda^{(j)} x_i^{\alpha^{(j)}}\big)$, we have the complete-data log-likelihood of $\lambda^{(j)}$:

$$\log L_i^C(\lambda^{(j)}) = U_i^{(j)} \Big\{ \log \alpha^{(j)} + \log \lambda^{(j)} + (\alpha^{(j)} - 1)\log x_i \Big\} - \lambda^{(j)} x_i^{\alpha^{(j)}}.$$

Let $\boldsymbol{\Theta} = (\alpha^{(1)}, \lambda^{(1)}, \ldots, \alpha^{(J)}, \lambda^{(J)})$ and denote the estimate of $\boldsymbol{\Theta}$ at the sth EM sequence by $\boldsymbol{\Theta}_s = (\alpha_s^{(1)}, \lambda_s^{(1)}, \ldots, \alpha_s^{(J)}, \lambda_s^{(J)})$.

- E-step:

  It follows from $Q_i(\lambda^{(j)}|\boldsymbol{\Theta}_s) = E\big[\log L_i^C(\lambda^{(j)})|\boldsymbol{\Theta}_s]$ that

  $$Q_i(\lambda^{(j)}|\boldsymbol{\Theta}_s) = \Upsilon_{i,s}^{(j)} \Big\{ \log \alpha^{(j)} + \log \lambda^{(j)} + (\alpha^{(j)} - 1)\log x_i \Big\} - \lambda^{(j)} x_i^{\alpha^{(j)}},$$

  where

  $$\Upsilon_{i,s}^{(j)} = E[U_i^{(j)}|\boldsymbol{\Theta}_s] = \begin{cases} \dfrac{\alpha_s^{(j)}\lambda_s^{(j)} x_i^{\alpha_s^{(j)}-1}}{\sum\limits_{\ell \in M_i} \alpha_s^{(\ell)}\lambda_s^{(\ell)} x_i^{\alpha_s^{(\ell)}-1}} & \text{if } \ j \in M_i \\ 0 & \text{if } \ j \notin M_i \end{cases}.$$

- M-step:

  Differentiating $Q(\alpha^{(j)}, \lambda^{(j)}|\boldsymbol{\Theta}_s) = \sum_{i=1}^n Q_i(\lambda^{(j)}|\boldsymbol{\Theta}_s)$ with respect to $\alpha^{(j)}$ and $\lambda^{(j)}$, and setting this to zero, we obtain

  $$\begin{aligned} \sum_{i=1}^n \frac{\partial Q_i}{\partial \alpha^{(j)}} &= \sum_{i=1}^n \Upsilon_{i,s}^{(j)} \Big\{ \frac{1}{\alpha^{(j)}} + \log x_i \Big\} - \lambda^{(j)} \sum_{i=1}^n x_i^{\alpha^{(j)}} \log x_i = 0 \\ \sum_{i=1}^n \frac{\partial Q_i}{\partial \lambda^{(j)}} &= \sum_{i=1}^n \frac{\Upsilon_{i,s}^{(j)}}{\lambda^{(j)}} - \sum_{i=1}^n x_i^{\alpha^{(j)}} = 0. \end{aligned}$$

  Rearranging for $\alpha^{(j)}$, we have the equation of $\alpha^{(j)}$ as

  $$\frac{1}{\alpha^{(j)}} \sum_{i=1}^n \Upsilon_{i,s}^{(j)} + \sum_{i=1}^n \Upsilon_{i,s}^{(j)} \log x_i - \sum_{i=1}^n \Upsilon_{i,s}^{(j)} \frac{\sum_{i=1}^n x_i^{\alpha^{(j)}} \log x_i}{\sum_{i=1}^n x_i^{\alpha^{(j)}}} = 0. \tag{13}$$

The $(s+1)$th EM sequence of $\alpha^{(j)}$ is the solution of the above equation. After finding $\alpha_{s+1}^{(j)}$, we obtain the $(s+1)$th EM sequence of $\lambda^{(j)}$ as

$$\lambda_{s+1}^{(j)} = \frac{\sum_{i=1}^{n} \Upsilon_{i,s}^{(j)}}{\sum_{i=1}^{n} x_i^{\alpha_{s+1}^{(j)}}}. \tag{14}$$

5.3 Normal Distribution Model

For the normal distribution, it is extremely difficult or impossible to obtain the EM sequences using (9) because finding the closed-form maximizer is not feasible in the M-step. Using (10), we can avoid these difficulties so that we obtain the EM sequences. This idea can easily be extended to the lognormal case using the fact that the logarithm of a random variable which is lognormally distributed has a normal distribution.

We assume that $T_i^{(j)}$ is a normal random variable with the mean and variance parameter $\boldsymbol{\theta}^{(j)} = (\mu^{(j)}, \sigma^{(j)})$. The pdf of $T_i^{(j)}$ is

$$f^{(j)}(t) = \frac{1}{\sqrt{2\pi}\,\sigma^{(j)}} \exp\Big(-\frac{1}{2}\big(\frac{t-\mu^{(j)}}{\sigma^{(j)}}\big)^2\Big).$$

We have the complete-data log-likelihood of $\boldsymbol{\theta}^{(j)}$:

$$\log L_i^C(\boldsymbol{\theta}^{(j)}) = U_i^{(j)} \log f^{(j)}(x_i) + (1-U_i^{(j)}) \log f^{(j)}(Z_i),$$

where Z_i is the truncated normal random variable with the pdf given by

$$f_Z^{(j)}(t|\boldsymbol{\theta}^{(j)}) = \frac{\frac{1}{\sigma^{(j)}}\phi\big(\frac{t-\mu^{(j)}}{\sigma^{(j)}}\big)}{1-\Phi\big(\frac{x_i-\mu^{(j)}}{\sigma^{(j)}}\big)}, \qquad t > x_i.$$

We denote the estimate of $\boldsymbol{\theta}^{(j)}$ and $\boldsymbol{\Theta}$ at the sth EM sequence by $\boldsymbol{\theta}_s^{(j)}$ and $\boldsymbol{\Theta}_s$, respectively.

- E-step:

  We have

$$\begin{aligned} \log f^{(j)}(Z_i) &= C - \frac{1}{2}\log {\sigma^{(j)}}^2 - \frac{1}{2{\sigma^{(j)}}^2}\big({Z_i}^2 - 2\mu^{(j)} Z_i + {\mu^{(j)}}^2\big) \\ E\big[\log f^{(j)}(Z_i)|\boldsymbol{\theta}_s^{(j)}\big] &= C - \frac{1}{2}\log {\sigma^{(j)}}^2 - \frac{1}{2{\sigma^{(j)}}^2}\big(m_{2i,s}^{(j)} - 2\mu^{(j)} m_{1i,s}^{(j)} + {\mu^{(j)}}^2\big), \end{aligned}$$

where

$$
\begin{aligned}
m_{1i,s}^{(j)} &= E[Z_i|\boldsymbol{\theta}_s^{(j)}] = \mu_s^{(j)} + \sigma_s^{(j)}\omega_{i,s}^{(j)} \\
m_{2i,s}^{(j)} &= E[{Z_i}^2|\boldsymbol{\theta}_s^{(j)}] = {\mu_s^{(j)}}^2 + {\sigma_s^{(j)}}^2 + \sigma_s^{(j)}(\mu_s^{(j)} + x_i)\omega_{i,s}^{(j)} \\
\omega_{i,s}^{(j)} &= \frac{\phi\left(\frac{x_i-\mu_s^{(j)}}{\sigma_s^{(j)}}\right)}{1-\Phi\left(\frac{x_i-\mu_s^{(j)}}{\sigma_s^{(j)}}\right)}.
\end{aligned}
$$

Using the above results, we have

$$
\begin{aligned}
Q_i(\mu^{(j)}, \sigma^{(j)}|\boldsymbol{\Theta}_s) = C &- \frac{\Upsilon_{i,s}^{(j)}}{2}\Big\{ \log {\sigma^{(j)}}^2 + \frac{1}{{\sigma^{(j)}}^2}\big({x_i}^2 - 2\mu^{(j)}x_i + {\mu^{(j)}}^2\big)\Big\} \\
&- \frac{\overline{\Upsilon}_{i,s}^{(j)}}{2}\Big\{ \log {\sigma^{(j)}}^2 + \frac{1}{{\sigma^{(j)}}^2}\big(m_{2i,s}^{(j)} - 2\mu^{(j)}m_{1i,s}^{(j)} + {\mu^{(j)}}^2\big)\Big\},
\end{aligned}
$$

where

$$
\begin{aligned}
\Upsilon_{i,s}^{(j)} &= E[U_i^{(j)}|\boldsymbol{\Theta}_s] = \begin{cases} \dfrac{\omega_{i,s}^{(j)}/\sigma_s^{(j)}}{\displaystyle\sum_{\ell\in M_i}\omega_{i,s}^{(\ell)}/\sigma_s^{(\ell)}} & \text{if } \ j \in M_i \\ \qquad 0 & \text{if } \ j \notin M_i \end{cases} \\
\overline{\Upsilon}_{i,s}^{(j)} &= 1 - \Upsilon_{i,s}^{(j)}.
\end{aligned}
$$

- M-step:

Differentiating $Q_i(\mu^{(j)}, \sigma^{(j)}|\boldsymbol{\Theta}_s)$ with respect to $\mu^{(j)}$, we obtain

$$
\frac{\partial Q_i}{\partial \mu^{(j)}} = \frac{1}{{\sigma^{(j)}}^2}\Big\{\Upsilon_{i,s}^{(j)}x_i + \overline{\Upsilon}_{i,s}^{(j)}m_{1i,s}^{(j)} - \mu^{(j)}\Big\}.
$$

Differentiating $Q_i(\mu^{(j)}, \sigma^{(j)}|\boldsymbol{\Theta}_s)$ again with respect to ${\sigma^{(j)}}^2$, we obtain

$$
\frac{\partial Q_i}{\partial {\sigma^{(j)}}^2} = \frac{1}{2{\sigma^{(j)}}^4}\Big\{\Upsilon_{i,s}^{(j)}(x_i - \mu^{(j)})^2 + \overline{\Upsilon}_{i,s}^{(j)}(m_{2i,s}^{(j)} - 2\mu^{(j)}m_{1i,s}^{(j)} + {\mu^{(j)}}^2) - {\sigma^{(j)}}^2\Big\}.
$$

Solving $\sum_{i=1}^n \partial Q_i/\partial\mu^{(j)} = 0$ and $\sum_{i=1}^n \partial Q_i/\partial{\sigma^{(j)}}^2 = 0$ for $\mu^{(j)}$ and ${\sigma^{(j)}}^2$, we obtain the $(s+1)$th EM sequence in the M-step as follows:

$$
\begin{aligned}
\mu_{s+1}^{(j)} &= \frac{1}{n}\sum_{i=1}^n \big\{\Upsilon_{i,s}^{(j)}x_i + \overline{\Upsilon}_{i,s}^{(j)}m_{1i,s}^{(j)}\big\} \\
{\sigma_{s+1}^{(j)}}^2 &= \frac{1}{n}\sum_{i=1}^n \big\{\Upsilon_{i,s}^{(j)}x_i^2 + \overline{\Upsilon}_{i,s}^{(j)}m_{2i,s}^{(j)}\big\} - \Big\{\mu_{s+1}^{(j)}\Big\}^2.
\end{aligned}
$$

Note that if the data are fully observed, then the $\Upsilon_{i,s}^{(j)} = 1$ so that the EM sequences become simply the MLE of μ and σ^2. It is of interest to look at the role of $\Upsilon_{i,s}^{(j)}$ and $\overline{\Upsilon}_{i,s}^{(j)}$ when an observation is incomplete. If an observation x_i is right-censored, then $\Upsilon_{i,s}^{(j)} = 0$, which results in the full weight (*i.e.*, $\overline{\Upsilon}_{i,s}^{(j)} = 1$) toward $m_{1i,s}^{(j)}$ and $m_{2i,s}^{(j)}$, the expectations of the respective random variables Z_i and Z_i^2 having the pdf truncated at x_i. If an observation x_i is masked, then $\Upsilon_{i,s}^{(j)}$ has a value between 0 and 1 of which the value is determined by the extent of masking. That is, as the number of indices in the set $M_i = \{j_1, \ldots, j_i\}$ gets larger, the value $\Upsilon_{i,s}^{(j)}$ becomes smaller, which results in more weight on $m_{1i,s}^{(j)}$ and $m_{2i,s}^{(j)}$.

5.4 Inverse Gaussian (Wald) Distribution Model

Also for the inverse Gaussian distribution, it is extremely difficult or impossible to obtain the EM sequences using (9) because finding the closed-form maximizer is not feasible in the M-step. Using (10), we can avoid these difficulties to obtain the EM sequences.

We assume that $T_i^{(j)}$ is an inverse Gaussian random variable with the location and scale parameter $\boldsymbol{\theta}^{(j)} = (\mu^{(j)}, \lambda^{(j)})$. Then, the pdf of $T_i^{(j)}$ is

$$f^{(j)}(t) = \sqrt{\frac{\lambda^{(j)}}{2\pi t^3}} \exp\Big(- \frac{\lambda^{(j)}(t-\mu^{(j)})^2}{2{\mu^{(j)}}^2 t} \Big),$$

and its cdf is

$$F^{(j)}(t) = \Phi\Bigg\{ \sqrt{\frac{\lambda^{(j)}}{t}} \Big(\frac{t-\mu^{(j)}}{\mu^{(j)}} \Big) \Bigg\} + \exp\Big(\frac{2\lambda^{(j)}}{\mu^{(j)}} \Big) \Phi\Bigg\{ - \sqrt{\frac{\lambda^{(j)}}{t}} \Big(\frac{t+\mu^{(j)}}{\mu^{(j)}} \Big) \Bigg\},$$

where $\Phi(\cdot)$ is the standard normal cdf.

We have the complete-data log-likelihood of $\boldsymbol{\theta}^{(j)}$:

$$\log L_i^C(\boldsymbol{\theta}^{(j)}) = U_i^{(j)} \log f^{(j)}(x_i) + (1 - U_i^{(j)}) \log f^{(j)}(Z_i),$$

where Z_i is the truncated inverse Gaussian random variable with the pdf given by

$$f_Z^{(j)}(t|\boldsymbol{\theta}^{(j)}) = \frac{f^{(j)}(t)}{1 - F^{(j)}(x_i)}, \qquad t > x_i.$$

We denote the estimate of $\boldsymbol{\theta}^{(j)}$ and $\boldsymbol{\Theta}$ at the sth EM sequence by $\boldsymbol{\theta}_s^{(j)}$ and $\boldsymbol{\Theta}_s$, respectively.

- E-step:

  We have

$$\log f^{(j)}(Z_i) = C + \frac{1}{2}\log\lambda^{(j)} - \frac{3}{2}\log Z_i - \frac{\lambda^{(j)}}{2{\mu^{(j)}}^2}Z_i + \frac{\lambda^{(j)}}{\mu^{(j)}} - \frac{\lambda^{(j)}}{2}\frac{1}{Z_i}$$
$$E\big[\log f^{(j)}(Z_i)|\boldsymbol{\theta}_s^{(j)}\big] = C + \frac{1}{2}\log\lambda^{(j)} - \frac{3}{2}m_{Ai,s}^{(j)} - \frac{\lambda^{(j)}}{2{\mu^{(j)}}^2}m_{Bi,s}^{(j)} + \frac{\lambda^{(j)}}{\mu^{(j)}} - \frac{\lambda^{(j)}}{2}m_{Ci,s}^{(j)},$$

  where $m_{Ai,s}^{(j)} = E[\log Z_i|\boldsymbol{\theta}_s^{(j)}]$, $m_{Bi,s}^{(j)} = E[Z_i|\boldsymbol{\theta}_s^{(j)}]$, and $m_{Ci,s}^{(j)} = E[1/Z_i|\boldsymbol{\theta}_s^{(j)}]$. Here $m_{Ai,s}^{(j)}$, $m_{Bi,s}^{(j)}$ and $m_{Ci,s}^{(j)}$ can be obtained by numerical integration. Using these results, we have

$$\begin{aligned} Q_i(\mu^{(j)}, \lambda^{(j)}|\boldsymbol{\Theta}_s) = C &+ \frac{\Upsilon_{i,s}^{(j)}}{2}\Big\{\log\lambda^{(j)} - 3\log x_i - \frac{\lambda^{(j)}}{{\mu^{(j)}}^2}x_i + 2\frac{\lambda^{(j)}}{\mu^{(j)}} - \lambda^{(j)}\frac{1}{x_i}\Big\} \\ &+ \frac{\overline{\Upsilon}_{i,s}^{(j)}}{2}\Big\{\log\lambda^{(j)} - 3m_{Ai,s}^{(j)} - \frac{\lambda^{(j)}}{{\mu^{(j)}}^2}m_{Bi,s}^{(j)} + 2\frac{\lambda^{(j)}}{\mu^{(j)}} - \lambda^{(j)}m_{Ci,s}^{(j)}\Big\}, \end{aligned}$$

  where

$$\Upsilon_{i,s}^{(j)} = E[U_i^{(j)}|\boldsymbol{\Theta}_s] = \begin{cases} \dfrac{h^{(j)}(x_i|\boldsymbol{\theta}_s^{(j)})}{\displaystyle\sum_{\ell\in M_i} h^{(\ell)}(x_i|\boldsymbol{\theta}_s^{(\ell)})} = \dfrac{f_Z^{(j)}(x_i|\boldsymbol{\theta}_s^{(j)})}{\displaystyle\sum_{\ell\in M_i} f_Z^{(\ell)}(x_i|\boldsymbol{\theta}_s^{(\ell)})} & \text{if } \ j \in M_i \\ 0 & \text{if } \ j \notin M_i \end{cases}$$
$$\overline{\Upsilon}_{i,s}^{(j)} = 1 - \Upsilon_{i,s}^{(j)}.$$

- M-step:

  Differentiating $Q_i(\mu^{(j)}, \lambda^{(j)}|\boldsymbol{\Theta}_s)$ with respect to $\mu^{(j)}$ and $\lambda^{(j)}$, we obtain

$$\frac{\partial Q_i}{\partial\mu^{(j)}} = \frac{\lambda^{(j)}}{{\mu^{(j)}}^3}\Big\{\Upsilon_{i,s}^{(j)}x_i + \overline{\Upsilon}_{i,s}^{(j)}m_{Bi,s}^{(j)} - \mu^{(j)}\Big\}$$
$$\frac{\partial Q_i}{\partial\lambda^{(j)}} = \frac{1}{\lambda^{(j)}} - \frac{\Upsilon_{i,s}^{(j)}}{2\mu^{(j)}}\Big(x_i + 2\mu^{(j)} + \frac{{\mu^{(j)}}^2}{x_i}\Big) - \frac{\overline{\Upsilon}_{i,s}^{(j)}}{2\mu^{(j)}}\Big(m_{Bi,s}^{(j)} + 2\mu^{(j)} + {\mu^{(j)}}^2 m_{Ci,s}^{(j)}\Big).$$

  Solving $\sum_{i=1}^n \partial Q_i/\partial\mu^{(j)} = 0$ and $\sum_{i=1}^n \partial Q_i/\partial\lambda^{(j)} = 0$ for $\mu^{(j)}$ and $\lambda^{(j)}$, we obtain the $(s+1)$th EM sequence in the M-step as follows:

$$\mu_{s+1}^{(j)} = \frac{1}{n}\sum_{i=1}^n \Big\{\Upsilon_{i,s}^{(j)}x_i + \overline{\Upsilon}_{i,s}^{(j)}m_{Bi,s}^{(j)}\Big\}$$
$$\frac{1}{\lambda_{s+1}^{(j)}} = \frac{1}{n}\sum_{i=1}^n \Big\{\Upsilon_{i,s}^{(j)}\frac{1}{x_i} + \overline{\Upsilon}_{i,s}^{(j)}m_{Ci,s}^{(j)}\Big\} - \frac{1}{\mu_{s+1}^{(j)}}.$$

As with the normal case, the value $\Upsilon_{i,s}^{(j)}$ plays a role of giving a weight on x_i versus $m_{Bi,s}^{(j)}$ and $1/x_i$ versus $m_{Ci,s}^{(j)}$.

In concluding this section, we should stress that in the case of the exponential and Weibull distributions, $h(\cdot)$ and $S(\cdot)$ are of closed forms, so applying the EM algorithm using (9) is straightforward so there is no need to treat the censored data as missing data. On the other hand, for the normal, lognormal, and inverse Gaussian distributions, it is either impossible or very difficult to obtain closed forms for $h(\cdot)$ and $S(\cdot)$, so applying the EM algorithm through the use of (9) is quite difficult. However, (10) involves only the corresponding pdf $f(\cdot)$, so applying the EM algorithm using (10) can be thought of as a straightforward generalized approach to the competing risks problem.

6 Examples

In this section, we illustrate several real-data examples. Some data sets in these examples can be found in the mainstream statistical literature. The data analysis is performed using code in the `R` language, which is an open source software for statistical computing and graphics originally developed by R Core Team (2016). This can be obtained at no cost from `http://www.r-project.org/`. In the Appendix, we provide the `R` functions which were used to analyze the data sets in the examples.

To compare the fits of the strength distribution models, the MSE from the fitted model to the empirical distribution were used. Letting $\hat{F}_n(t_i)$ denote the empirical cdf and $F(t_i;\hat{\boldsymbol{\theta}})$ denote the fitted cdf using the MLE of $\boldsymbol{\theta}$, the MSE for the fitted model is calculated as

$$\mathrm{MSE}\big(F(\cdot;\hat{\boldsymbol{\theta}})\big) = \frac{1}{n}\sum_{i=1}^{n}\big\{F(t_i;\hat{\boldsymbol{\theta}}) - \hat{F}_n(t_i)\big\}^2.$$

If the data are not censored, the empirical cdf $\hat{F}_n(\cdot)$ can be easily calculated. Several versions of these empirical estimates $\hat{F}_n(\cdot)$ have been suggested in the statistics literature, but the most popular one is $(j-1/2)/n$ (also known as *median rank method*) for $n \geq 11$ and $(j-3/8)/(n+1/4)$ for $n \leq 10$, due to Blom (1958) and Wilk and Gnanadesikan (1968). However, if the data set has censoring, then the empirical cdf $\hat{F}_n(t)$ can be estimated by the well-known product limit estimator of $S(t) = 1 - F(t)$ originally attributed to Kaplan and Meier (1958), which is defined here as

$$\hat{S}_n(t) = \begin{cases} 1 & \text{if} \quad 0 \le t \le t_1 \\ \displaystyle\prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1} \right)^{\mathbb{I}(\delta_i > 0)} & \text{if} \quad t_{k-1} < t \le t_k, \quad k = 2, \ldots, n \\ 0 & \text{if} \quad t > t_n \end{cases}$$

so that we have $\hat{F}_n(t) = 1 - \hat{S}_n(t)$. An alternative way to compare the fits of the proposed models for each specific failure mode is to compare the empirical CIF, proposed by Aalen (1978), with the parametric CIF defined by (4). Here, we are interested in the strength distribution of a whole system or a material specimen, not in the distribution due to each specific failure mode, so we do not consider the CIF in this paper. If one is interested in the lifetime distribution due to each failure mode, one should look at the CIF. For applications of the CIF with the masked data, the reader is referred to Park and Kulasekera (2004) and Park (2005).

6.1 Wire Connections

The data in this example were obtained by King (1971) and have since then been often used for illustration in competing risks literature, including Nelson (1972) and Crowder (2001). Table 1 gives breaking strengths in milligrams of 23 wire connections. The wire is bonded at one end to a semiconductor wafer and at the other end to a terminal post. There are two types of failures: breakage at the bonded end and a wire breakage.

Table 1: Breaking Strengths of Wire Connections

Bond	**0**	**0**	550	950	1150	1150	1250	1250	1450	1450	1550	2050	**3150**
Wire	750	950	1150	1150	1150	1350	1450	1550	1550	1850			

The zero values must be faulty bonds and should be eliminated as mentioned by Nelson (1972) who also expressed some doubts about the value 3150. To fit the data into the Weibull model, he simply deleted the value 3150. Rather than this *ad hoc* method of deleting any suspicious values, it is more appropriate to try to fit the data into several different models. We try exponential, lognormal and Wald distributions. We estimate the parameters under the above models, the results of which are summarized in Table 2.

Table 2: Parameter Estimation under the Models Considered

	Exponential	Weibull		Lognormal		Wald	
Mode	$\lambda^{(j)} \times 10^{-4}$	$\lambda^{(j)} \times 10^{-9}$	$\alpha^{(j)}$	$\mu^{(j)}$	$\sigma^{(j)}$	$\mu^{(j)}$	$\lambda^{(j)}$
Bond	3.8128	0.7877	2.7650	7.3909	0.4328	1785.2	8467.5
Wire	3.4662	3.1608	2.5679	7.4116	0.4139	1820.4	9629.8

One of the appropriate model selection procedures is to compare the MSE of the models. We report the MSE of the models considered in Table 3. The lognormal and Wald models are fairly competitive to the others. In Figure 1, we show Weibull plots of the data set along with the parametric cdf of the models. This figure clearly shows that the lognormal and Wald models fit better than the Weibull model.

Table 3: MSE under the Models Considered

Model	Exponential	Weibull	Lognormal	Wald
MSE $\times 10^3$	45.531940	8.104004	3.851452	4.023914

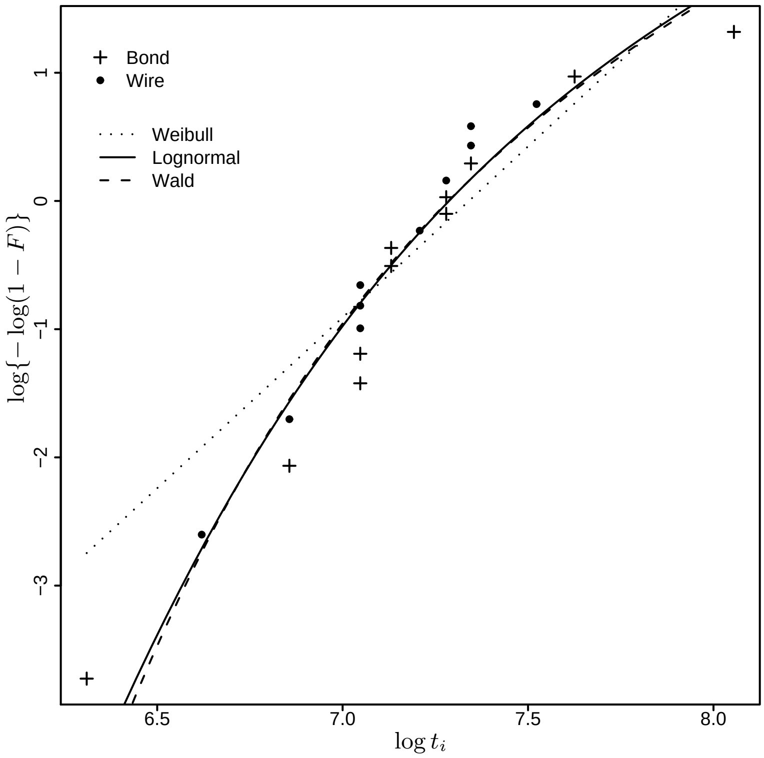

Figure 1: Weibull probability plot.

6.2 Time to Failure of Mechanical Devices

The data set in this example is given by Michael (1979). The data are from the lifetime testing of a batch of 40 mechanical devices which give the time to failure measured in millions of operations. Each device has two switches, A or B, either failure of which can cause the breakdown of the whole device. The data set is provided again in Table 4, where the failure modes are coded as follows: $\delta_i = 1$ (mode A), $\delta_i = 2$ (mode B), and $\delta_i = 0$ (censored).

Table 4: Time to Failure of Mechanical Devices

Time	Mode	Time	Mode	Time	Mode	Time	Mode
1.151	2	1.667	1	2.119	2	2.547	1
1.170	2	1.695	1	2.135	1	2.548	1
1.248	2	1.710	1	2.197	1	2.738	2
1.331	2	1.955	2	2.199	2	2.794	1
1.381	2	1.965	1	2.227	1	2.883	0
1.499	1	2.012	2	2.250	2	2.883	0
1.508	2	2.051	2	2.254	1	2.910	1
1.538	2	2.076	2	2.261	2	3.015	1
1.577	2	2.109	1	2.349	2	3.017	1
1.584	2	2.116	2	2.369	1	3.793	0

This data set has been analyzed by Chambers et al. (1983), who also explicitly provide the raw data set. With the emphasis on the interest in the separate failure modes, they simply split the data set into two groups, (modes A and B). They could not consider competing failure modes because they simply classified the data into two groups. To better analyze this problem, it is more appropriate to compare the CIF under the competing risks model. For this kind of problem, the reader is referred to Park (2005) who analyzed the competing risks problem with both censoring and masking considered.

In this example, we analyze the data with emphasis on the lifetime of the whole device. Hence, we are interested in the appropriate lifetime model of the whole system rather than each failure mode. We summarize the parameter estimation in Table 5 and the MSE in

Table 5: Parameter Estimation under the Models Considered

	Exponential	Weibull		Lognormal		Wald	
Mode	$\lambda^{(j)}$	$\lambda^{(j)} \times 10^{-2}$	$\alpha^{(j)}$	$\mu^{(j)}$	$\sigma^{(j)}$	$\mu^{(j)}$	$\lambda^{(j)}$
A	0.2004	0.7248	4.6526	0.9455	0.2611	2.6651	37.730
B	0.2358	4.5945	2.9117	0.9031	0.4335	2.7308	13.088

Table 6. By comparing the MSE of the models, the lognormal and Wald models fit much better than the others. Figure 2 also provides the same conclusion (the exponential model is not shown). The curves of the lognormal and Wald models in the Weibull plots are almost overlapping. This also shows that these two models are fairly good.

Table 6: MSE under the Models Considered

Model	Exponential	Weibull	Lognormal	Wald
MSE $\times 10^3$	41.326541	3.282886	1.923135	1.937779

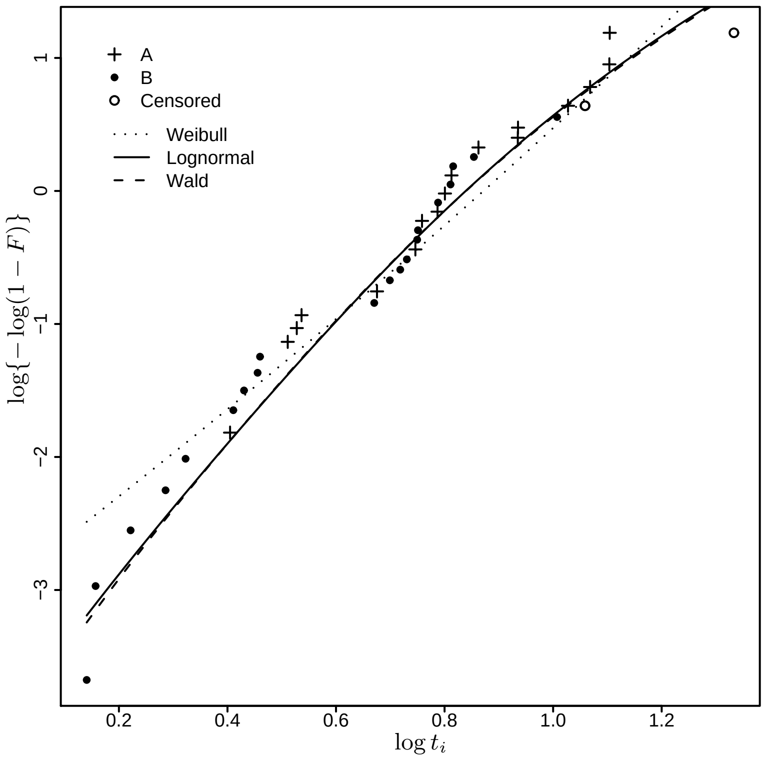

Figure 2: Weibull probability plot.

6.3 Device-G from a Field-Tracking Study

The data set in Table 7 was illustrated by Meeker and Escobar (1998). They analyzed the data using the Weibull model. The data set gives times of failure and running times for a sample of devices from a field-tracking study of a larger system. At a certain point in time, 30 units were installed in normal service conditions. Failure cause was investigated for each unit that failed. Mode S denotes the failure caused by an accumulation of randomly occurring damage from power-line voltage spikes during electric storms. Mode W denotes the failure caused by normal product wear. Here, the failure modes are coded as follows: $\delta_i = 1$ (mode S), $\delta_i = 2$ (mode W), and $\delta_i = 0$ (censored).

From our results summarized in Table 9, the lognormal model outperforms the Weibull model. For the Wald model, the EM algorithm is very slow because of the large proportion of censoring. In the E-step of the algorithm with the Wald model, we have to use numerical

Table 7: Device-G Failure Times and Cause of Failure

Cycle$\times 10^{-3}$	Mode	Cycle$\times 10^{-3}$	Mode	Cycle$\times 10^{-3}$	Mode	Cycle$\times 10^{-3}$	Mode
275	2	300	0	300	0	300	0
13	1	173	1	2	1	23	1
147	2	106	1	261	1	300	0
23	1	300	0	293	2	80	1
181	2	300	0	88	1	245	2
30	1	212	2	247	1	266	2
65	1	300	0	28	1		
10	1	300	0	143	1		

Table 8: Parameter Estimation under the Models Considered

	Exponential	Weibull		Lognormal		Wald	
Mode	$\lambda^{(j)} \times 10^{-3}$	$\lambda^{(j)}$	$\alpha^{(j)}$	$\mu^{(j)}$	$\sigma^{(j)}$	$\mu^{(j)}$	$\lambda^{(j)}$
S	2.8243	1.6598×10^{-2}	0.6710	5.5728	2.1830	10886.6	32.6
W	1.3180	1.0426×10^{-11}	4.3373	5.7706	0.3760	346.6	2209.3

integrations which cause slowdown in the estimation. It is a challenging future work to speed up the EM of the Wald model. But for other models, the EM algorithm works fairly fast. The parameter estimates and MSE of the models are summarized in Tables 8 and their fits using these estimates are superimposed on the Weibull plots in Figure 3.

Table 9: MSE under the Models Considered

Model	Exponential	Weibull	Lognormal	Wald
MSE $\times 10^3$	1.8584566	0.8314317	0.5407669	24.1763249

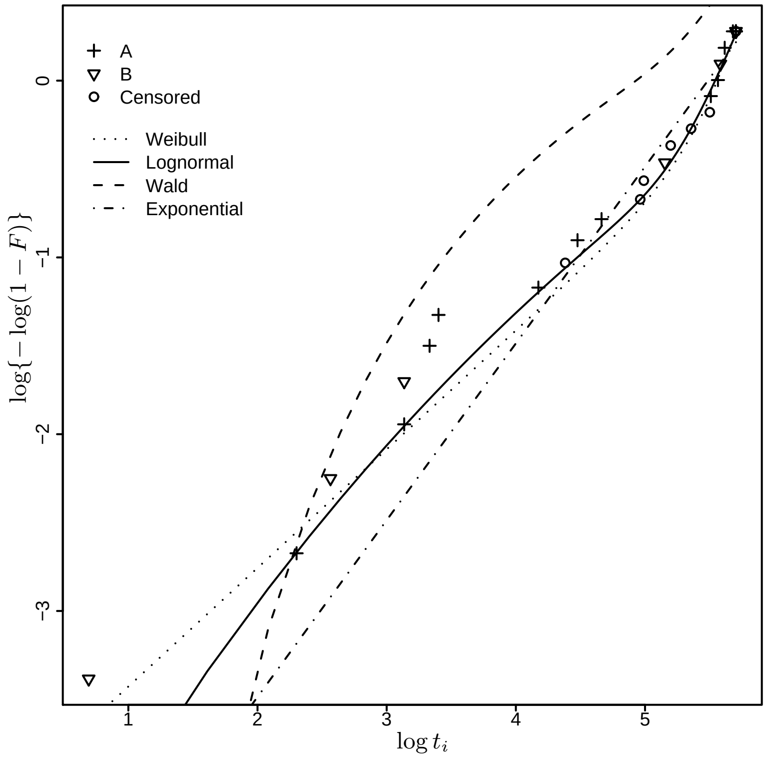

Figure 3: Weibull probability plot.

6.4 Microbond Testing of Pitch-based Fibers

An experiment was performed at Clemson University by Harwell (1995) to study the strength of the interfacial bond of a carbon fiber (whose average diameter is approximately $8 \sim 12$ μm) and matrix material. Ribbon fibers, *i.e.*, flat-shaped rather than round-shaped fibers, were used in Harwell's "microbond tests." In the experiment, a droplet of the epoxy resin was placed on a fiber and cured by heat treatment. The fibers were coated with SiC since it was thought that such a coating would improve the interfacial bond. Microbond tests were performed on "uncoated" fibers and on fibers with either a "thin" SiC coating or a "thick" SiC coating. The fiber-in-droplet specimen was then placed in a "micro-vise," and the fiber was placed under tensile load in an attempt to force it to debond from the matrix droplet. The applied stress required to debond the fiber from the droplet was recorded by a load cell. However, for some of the specimens tested, due to inherent flaws in the fiber which tend to decrease its tensile strength, the fiber broke before debonding occurred. Kuhn and Padgett (1997) analyzed this data set using the kernel density estimation with the main interest in comparing the debonding strengths of ribbon fibers. Hence the breakdowns due to the internal flaws were treated as right-censored. In this example, we consider all kinds of flaws to estimate the strength distribution of the specimen. So the internal flaws, debonding and coating are considered to be causes of failure. In what follows, we analyzed the data for the uncoated, thin coated and thick coated fibers separately.

6.4.1 No Coating

The data in Table 10 show the tensile strength (in Newton) of fibers without SiC coating. In this case, there are only two causes of failure — fiber breakdown (denoted by B) and debonding (denoted by D). Table 11 summarizes the parameter estimates of the models under consideration. Based on the MSE criterion, the lognormal model is the best, but the MSE of the Wald model is also very small and close to that of the lognormal. As also shown in the Weibull plot in Figure 4, the lognormal and Wald models are very close and better than the Weibull model.

Table 10: Strength of the Interfacial Bond (No Coating)

Strength	Mode	Strength	Mode	Strength	Mode	Strength	Mode
0.198	B	0.268	D	0.320	D	0.282	D
0.212	D	0.219	D	0.275	D	0.246	D
0.330	B	0.211	D	0.298	D	0.181	D
0.321	D	0.206	D	0.334	D	0.183	D
0.371	D	0.253	D	0.295	D	0.283	D
0.216	D	0.264	D	0.281	D	0.244	D
0.285	D	0.266	D	0.222	D	0.224	D
0.259	D	0.247	D	0.199	D	0.286	D
0.356	B	0.234	D	0.283	D		
0.338	D	0.285	D	0.217	D		

Table 11: Parameter Estimation under the Models Considered

	Exponential	Weibull		Lognormal		Wald	
Mode	$\lambda^{(j)}$	$\lambda^{(j)}$	$\alpha^{(j)}$	$\mu^{(j)}$	$\sigma^{(j)}$	$\mu^{(j)}$	$\lambda^{(j)}$
Debonding	3.5028	1230.2	5.6968	−1.3402	0.1900	0.2666	7.2711
Break	0.3002	2357.2	8.3785	−0.8421	0.2901	0.4500	5.0159

Table 12: MSE under the Models Considered

Model	Exponential	Weibull	Lognormal	Wald
MSE $\times 10^3$	65.280163	1.721003	1.457419	1.461700

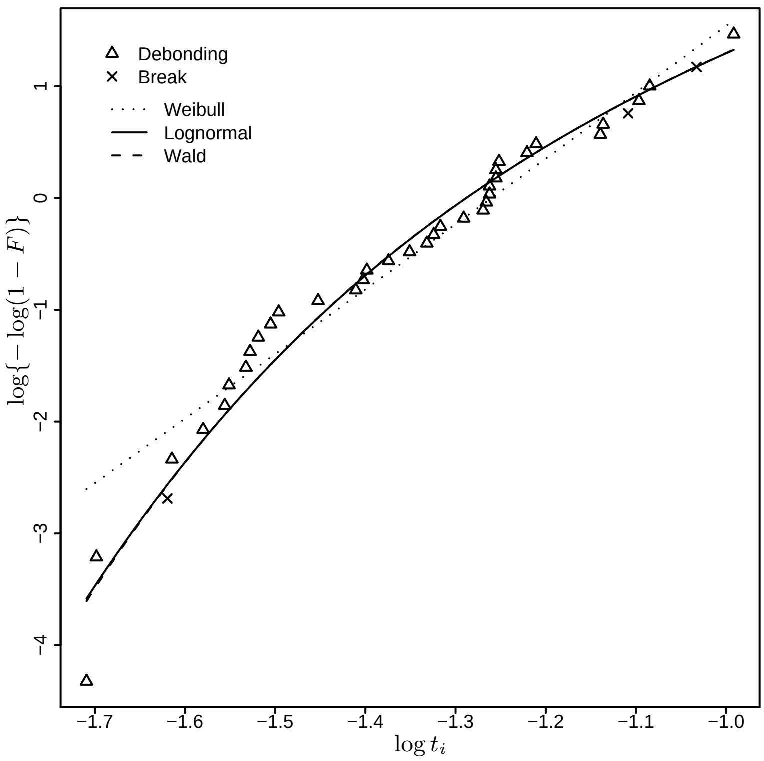

Figure 4: Weibull probability plot.

6.4.2 Thin Coating

Next, we consider the data set with "thin" SiC coating. The data are provided in Table 13 In this case, there are three causes of failure — fiber breakdown (denoted by B), debonding (denoted by D), and coating (denoted by C). Table 14 summarizes the parameter estimates of the models under consideration. Based on the MSE criterion, unlike the no-coating case, the Weibull model is the best. It seems that the failure mode of thin coating is dominant and its strength distribution has a Weibull distribution. Figure 5 shows that all the models seem good.

Table 13: Strength of the Interfacial Bond (Thin Coating)

Strength	Mode	Strength	Mode	Strength	Mode	Strength	Mode
0.261	B	0.108	C	0.315	D	0.065	C
0.230	D	0.417	D	0.317	D	0.077	C
0.270	C	0.355	B	0.375	B	0.091	C
0.328	D	0.346	D	0.323	D	0.117	C
0.185	C	0.215	D	0.345	D	0.171	C
0.391	D	0.401	D	0.114	C	0.352	D
0.206	B	0.361	D	0.155	C	0.174	C
0.408	D	0.172	B	0.412	B	0.407	D
0.256	D	0.232	D	0.171	C		
0.088	C	0.493	B	0.223	C		

Table 14: Parameter Estimation under the Models Considered

	Exponential	Weibull		Lognormal		Wald	
Mode	$\lambda^{(j)}$	$\lambda^{(j)}$	$\alpha^{(j)}$	$\mu^{(j)}$	$\sigma^{(j)}$	$\mu^{(j)}$	$\lambda^{(j)}$
Debonding	1.7125	143.7561	5.2613	−1.0445	0.2427	0.3622	5.9759
Break	0.7051	27.2209	4.4400	−0.7626	0.4266	0.5165	2.4633
Coating	1.4103	1.8422	1.2191	−0.7988	1.1431	4.0949	0.2912

Table 15: MSE under the Models Considered

Model	Exponential	Weibull	Lognormal	Wald
MSE $\times 10^3$	24.291772	2.814685	3.449413	3.049166

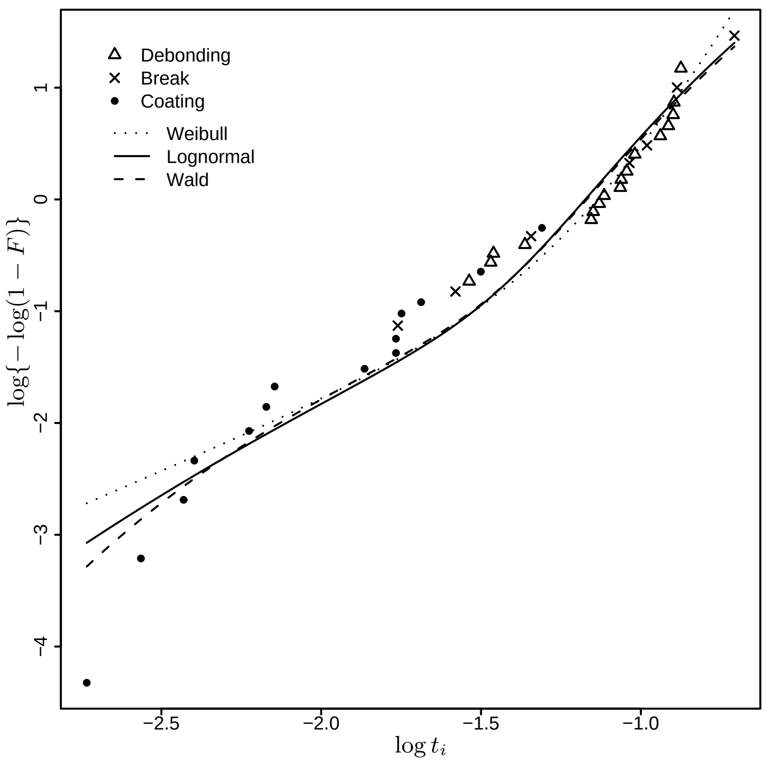

Figure 5: Weibull probability plot.

6.4.3 Thick Coating

Next, we consider the data set with "thick" SiC coating. The data are provided in Table 16, and Table 17 summarizes the parameter estimates. Based on the MSE criterion, the Weibull model is the best. It seems that the failure mode of thick coating is more dominant and its strength distribution is very closely modeled by a Weibull distribution. Figure 6 indicates that all the models are relatively good.

Table 16: Strength of the Interfacial Bond (Thick Coating)

Strength	Mode	Strength	Mode	Strength	Mode	Strength	Mode
0.232	C	0.412	C	0.275	C	0.171	C
0.335	C	0.355	C	0.242	C	0.392	B
0.120	C	0.425	D	0.075	C	0.080	C
0.073	C	0.332	D	0.319	C	0.121	C
0.134	C	0.399	D	0.231	C	0.347	C
0.176	C	0.110	C	0.173	C	0.263	D
0.276	C	0.492	D	0.356	D	0.353	D
0.065	C	0.257	C	0.373	C	0.332	C
0.414	D	0.114	C	0.273	C	0.190	C
0.573	B	0.044	C	0.035	C	0.066	C
				0.289	C	0.397	B

Table 17: Parameter Estimation under the Models Considered

	Exponential	Weibull		Lognormal		Wald	
Mode	$\lambda^{(j)}$	$\lambda^{(j)}$	$\alpha^{(j)}$	$\mu^{(j)}$	$\sigma^{(j)}$	$\mu^{(j)}$	$\lambda^{(j)}$
Debonding	0.7483	58.5486	5.5510	-0.8345	0.2227	0.4453	8.7143
Break	0.2806	190.4963	8.4643	-0.6959	0.1636	0.5042	18.9087
Coating	2.8996	5.21057	1.5031	-1.4169	0.8789	0.3636	0.3034

Table 18: MSE under the Models Considered

Model	Exponential	Weibull	Lognormal	Wald
MSE $\times 10^3$	15.923565	0.992133	1.311246	1.970773

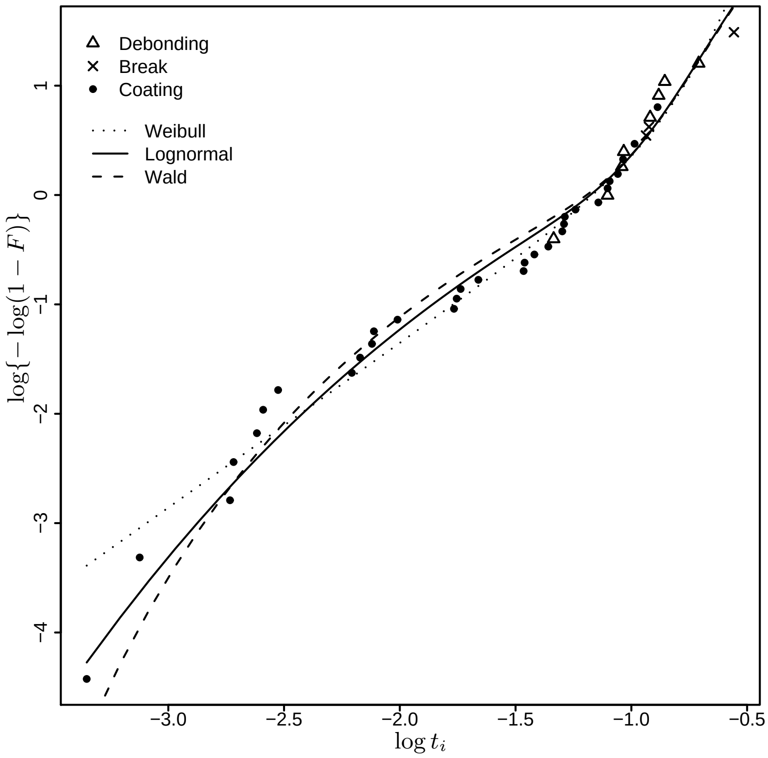

Figure 6: Weibull probability plot.

6.5 Strength Data with Censoring and Masking

In many tensile strength experiments, specimens tested are broken down due to several causes with the cause of fracture not properly identified along with censoring due to time and cost considerations on experiments.

The strength data in Table 19 were obtained using the lognormal random number generator of `R` language to illustrate the use of the proposed method. Here, we assume that the fracture causes are due to a surface defect (mode 1), an inner defect (mode 2), and an end effect at the clamp to hold the specimen (mode 3). The censored observations are denoted by 0. To illustrate the applicability of the partial or complete masking along with censoring, the data were censored at 150 and 10% of the data were randomly masked.

The parameter estimates and MSEs of the models are summarized in Tables 20 and 21 and their fits using these estimates are superimposed on the Weibull plots in Figure 7. Note that unlike the preceding examples, the cause of fracture can not be classified separately due to masking. Thus, we only mark the data with either 'failure' or 'censored' in the figure. Comparing the MSEs comes to the conclusion that the strength distribution can be modeled by a lognormal distribution.

Table 19: Simulated Strength Data for Three Fracture Causes with Censoring and Masking

Strength	Modes	Strength	Modes	Strength	Modes	Strength	Modes
54	$\{3\}$	7	$\{1,2,3\}$	86	$\{2\}$	104	$\{1\}$
143	$\{2\}$	81	$\{3\}$	141	$\{1\}$	89	$\{3\}$
97	$\{3\}$	52	$\{3\}$	79	$\{3\}$	9	$\{3\}$
104	$\{3\}$	40	$\{3\}$	23	$\{3\}$	111	$\{1,2,3\}$
71	$\{1,2\}$	82	$\{2\}$	8	$\{3\}$	150	0
98	$\{1\}$	3	$\{3\}$	17	$\{3\}$	79	$\{2\}$
24	$\{2\}$	130	$\{2\}$	41	$\{2\}$	94	$\{2\}$
138	$\{3\}$	5	$\{3\}$	43	$\{2,3\}$	150	0
38	$\{3\}$	32	$\{2\}$	9	$\{3\}$	77	$\{2\}$
78	$\{3\}$	16	$\{3\}$	92	$\{2\}$	76	$\{3\}$
150	0	33	$\{3\}$	80	$\{2\}$	100	$\{2\}$
46	$\{3\}$	137	$\{1,2\}$	92	$\{3\}$	108	$\{2\}$
109	$\{1\}$	71	$\{1\}$	60	$\{2\}$	88	$\{1\}$
7	$\{3\}$	11	$\{3\}$	150	0	150	0
42	$\{2\}$	6	$\{3\}$	43	$\{3\}$	124	$\{1,2\}$

Table 20: Parameter Estimation under the Models Considered

	Weibull		Lognormal		Wald	
Mode	$\lambda^{(j)}$	$\alpha^{(j)}$	$\mu^{(j)}$	$\sigma^{(j)}$	$\mu^{(j)}$	$\lambda^{(j)}$
Surface defect	2.6150×10^{-10}	4.3078	5.0390	0.3599	165.4	1172.5
Inner defect	8.9161×10^{-6}	2.3295	4.8525	0.6732	169.4	261.6
End effect	1.1309×10^{-2}	0.8802	4.6931	1.7034	7478.7	30.2

Table 21: MSEs under the Models Considered

Model	Weibull	Lognormal	Wald
MSE	0.1322911	0.1314757	0.1396442

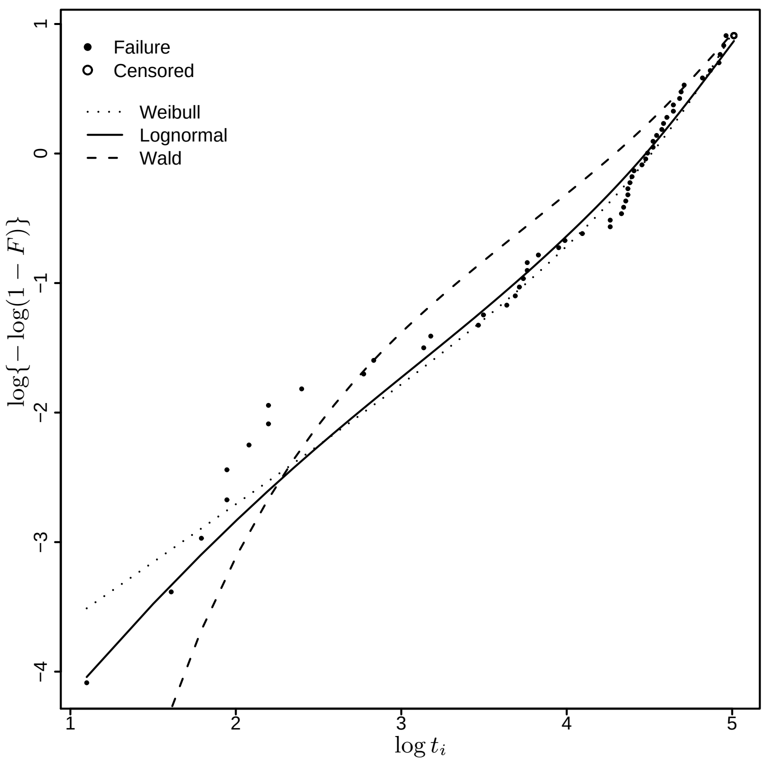

Figure 7: Weibull probability plot.

A Appendix

A.1 Sketch Proof of the Uniqueness and Bounds

The uniqueness of the MLE of the Weibull is proved by Farnum and Booth (1997) in the case of either complete failure data or right-censored data of Type-I or Type-II. Here, we provide the sketch of the proof of the unique solution of (13). Our proof is similar to theirs. For convenience, omitting the failure mode index (j) and the step index s, and letting

$$g(\alpha) = \frac{1}{\alpha}\sum_{i=1}^{n}\Upsilon_i$$

$$h(\alpha) = \sum_{i=1}^{n}\Upsilon_i \cdot \frac{\sum_{i=1}^{n} x_i^{\alpha}\log x_i}{\sum_{i=1}^{n} x_i^{\alpha}} - \sum_{i=1}^{n}\Upsilon_i \log x_i,$$

we rewrite (13) by $g(\alpha) = h(\alpha)$. The function $g(\alpha)$ is strictly decreasing from ∞ to 0 on $\alpha \in [0, \infty]$ unless $\Upsilon_i = 0$ for all i, while $h(\alpha)$ is increasing because it follows from the Jensen's inequality that

$$\frac{\partial h(\alpha)}{\partial \alpha} = \frac{\sum_{i=1}^{n}\Upsilon_i}{\left\{\sum_{i=1}^{n} x_i^{\alpha}\right\}^2}\Big\{\sum_{i=1}^{n} x_i^{\alpha}\log^2 x_i \sum_{i=1}^{n} x_i^{\alpha} - \big(\sum_{i=1}^{n} x_i^{\alpha}\log x_i\big)^2\Big\} \geq 0.$$

Now, it suffices to show that $h(\alpha) > 0$ for some α. Since

$$\lim_{\alpha\to\infty} h(\alpha) = \sum_{i=1}^{n}\Upsilon_i(\log x_{\max} - \log x_i),$$

we have $h(\alpha) > 0$ for some α unless $\Upsilon_i = 0$ or $x_i = x_{\max}$ for all i. This condition is unrealistic in practice.

Next, we provide upper and lower bounds of α. These bounds guarantee the solution in the interval and enable the root search algorithm to find the solution very stably and easily. Since $h(\alpha)$ is increasing, we have $g(\alpha) \leq \lim_{\alpha\to\infty} h(\alpha)$, that is,

$$\alpha \geq \frac{\sum_{i=1}^{n}\Upsilon_i}{\sum_{i=1}^{n}\Upsilon_i(\log x_{\max} - \log x_i)}.$$

Denote the above lower bound by α_L. Then, since $h(\alpha)$ is again increasing, we have $g(\alpha) = h(\alpha) \geq h(\alpha_L)$. If $h(\alpha_L) > 0$, then we have an upper bound

$$\alpha \leq \frac{\sum_{i=1}^{n}\Upsilon_i}{h(\alpha_L)}.$$

If $h(\alpha_L) < 0$ (it is extremely rare in practice though), then an upper bound can be obtained by

$$k \cdot \max\Big(\alpha_L, \sum_{i=1}^{n} \Upsilon_i / |h(\alpha_L)|\Big),$$

for some large k. This can be easily found by increasing k, say, $k = 2, 3, \ldots$. Since $h(\alpha)$ is increasing and $h(\alpha) > 0$ for some α, this method guarantees to find an upper bound.

A.2 R Program Usages

Here, we provide the usage of the R programs developed for the parameter estimation of the exponential, normal, lognormal, inverse Gaussian (Wald) and Weibull distributions. In the following subsection, we also provide R source codes. These R functions find the MLEs using the EM algorithm. To use the program, we need to call the R functions and type the observations with failure modes. The variable M is a list and the variable d is a vector. In the failure mode variables M and d, '-1' means a complete masking and '0' means censoring. If there are three failure modes, we can code either -1 or c(1,2,3) for complete masking. As an illustration, if there are three failure modes indexed by 1, 2 and 3, one can code the data as follows. As shown below, we obtain the following estimates under the exponential model,

$$\hat{\lambda}^{(1)} = 0.1518975, \quad \hat{\lambda}^{(2)} = 0.1018091, \quad \hat{\lambda}^{(3)} = 0.1018489.$$

```
> source("https://raw.githubusercontent.com/statpnu/R/master/Rpp5.R")
> X = c(1.9, 2.1, 3.2, 1.1, 2.1, 1.0, 2.0, 6.1, 3)      # lifetime observation
> M = list(1, 1, 1, 2, 2, 3, 3, 0, c(1,2,3))           # failure modes
> d =    c(1, 1, 1, 2, 2, 3, 3, 0, -1)                 # same as M
> # for exponential model
> expo.cm.EM(X,M)                                      # same as expo.cm.EM(X,d)
$lam
[1] 0.1518975 0.1018091 0.1018489
$iter
[1] 2
$conv
[1] TRUE
```

For the other models, use the following functions.

`> norm.cm.EM(X,M)`	for the normal model
`> norm.cm.EM(log(X),M)`	for the lognormal model
`> wald.cm.EM(X,M)`	for the inverse Gaussian model
`> weibull.cm.EM(X,M)`	for the Weibull model

If there are partial maskings, we have to use `list`() function for failure modes. For example, one can code as follows.

```
> M = list(1, 1, 0, c(2,3), 2, 3, 3, c(1,2), c(1,2,3))
```

The auxiliary argument `maxits` in each function is for the maximum number of iterations and the argument `eps` is for the error in stopping criterion. The default values for `maxits` and `eps` are `maxits=100` and `eps=1.0E-3`. The EM algorithm stops if the changes are all relatively small, *i.e.*, $\left|\lambda_{s+1}^{(j)} - \lambda_s^{(j)}\right| < \epsilon\lambda_{s+1}^{(j)}$, $j = 1, \ldots, J$, for the exponential distribution model. The auxiliary argument `lam0` is the initial value for the exponential model; the arguments `mu0` and `sd0` are the initial values for the normal model; `mu0` and `lam0` are for the inverse Gaussian model; and `alpha0` (shape) and `lam0` (scale) are for the Weibull model. The initial values can be specified manually by setting these arguments. In our `R` programs, by default, the initial values are determined by considering the observed failures and censored observations after the masked observations were deleted. If all the data are masked, then the initial values are given by the MLE of the failure observations by ignoring failure modes.

A.3 R Source Codes

```
# ==============================================================================
# File name : fnpp5.R
# Authors   : Chanseok Park and W. J. Padgett
# Version   : 1.1,   January 5, 2005
#             based on R Version 2.0.1 (2004-11-15)
# This progam can be freely distributed for non-commercial use.
# ==============================================================================

# ----------------------------------
# pdf, cdf, MLE of Wald distribution
# ----------------------------------

# pdf of Wald
#
dwald <- function (x, location = 1, scale = 1) {
    k <- max(lx <- length(x), lloc <- length(location), lscale <- length(scale))
    if (lx < k)
        x = rep(x, length = k)
    if (lloc < k)
        location = rep(location, length = k)
    if (lscale < k)
        scale = rep(scale, length = k)
    y = (log(abs(scale)) - log(2 * pi))/2 - 1.5 * (log(x)) -
        scale/(2 * location^2) * ((x - location)^2/x)
    if (!is.null(Names <- names(x)))
        names(y) <- rep(Names, length = length(y))
    return(exp(y))
}

# cdf of Wald
#
pwald <- function (x, location = 1, scale = 1)
{
    k <- max(lx <- length(x), lloc <- length(location), lscale <- length(scale))
    if (lx < k)
        x <- rep(x, length = k)
    if (lloc < k)
        location <- rep(location, length = k)
    if (lscale < k)
        scale <- rep(scale, length = k)
    y = pnorm( sqrt(scale/x) * (x/location - 1)) + exp(2 * scale/location) *
        pnorm(-sqrt(scale/x) * (x/location + 1))
    if (!is.null(Names <- names(x)))
        names(y) <- rep(Names, length = length(y))
    return(y)
}
#
# MLE of Wald (Inverse Gaussian)
#
wald.MLE <- function(x) {
  if ( any (x <= 0) ) stop("The data should be positive")
  mu = mean(x) ; n = length(x)
```

```
  lam = 1 / ( mean(1/x) - 1/mu)
  list ( location=mu, scale=lam )
}
#
# MLE of  Weibull
#
weibull.MLE <- function(x, interval) {
  if ( any (x <= 0) ) stop("The data should be positive")
  if (missing(interval)) {
     meanlog = mean(log(x))
     lower = 1 / ( log(max(x)) - meanlog )
     upper = sum( (x^lower)*log(x) ) / sum( x^lower ) - meanlog
     interval = c(lower,1/upper)
  }
  EE = function(alpha,x) {
     xalpha = x^alpha
     sum(log(x)*(xalpha)) / sum(xalpha) - 1/alpha - mean(log(x))
  }
  tmp = uniroot(EE, interval=interval, x=x)
  alpha = tmp$root
  list ( alpha=alpha, lam=1/mean(x^alpha) )
}

# -----------------------------------
# EM: Exponential Distribution Model
# -----------------------------------
expo.cm.EM <-
function(X, M, lam0, maxits=100, eps=1.0E-3)  {
   nk = length(X)
   J  = max(unlist(M))
   idx = unique( unlist(M) )
   jj  = idx[ idx>0 ]
   if (is.vector(M)) M <- as.list(M)
   for ( i in 1:nk )  if ( any(M[[i]]<0) )  M[[i]] = jj
 #
 # Setting the inital values
   if ( !missing(lam0) && length(lam0) != J ) lam0 = rep(lam0,l=J)
   X0 = as.list(NULL); length(X0) = J
   n1 = numeric(J)
   if ( missing(lam0) ) {
      for ( i in 1:nk ) {
          idx = M[[i]]
          if (length(idx) == 1) {
              if ( idx > 0 ) {
                      X0[[idx]] = c(X0[[idx]], X[i])
                       n1[idx]  = n1[idx] + 1
              } else if (idx == 0) for (j in jj) X0[[j]] = c(X0[[j]], X[i])
          }
      }
      lam0 = rep(NA, l=J)
      for ( j in jj ) {
          if ( is.null(X0[[j]]) ) {
                    lam0[j] = 1/mean(X)
          } else  { lam0[j] = 1/mean(X0[[j]]) }
```

```
      }
   }
   # end of initial value setting

   newlam =  rep(NA, l=J)
   lam = lam0
 #
 # Start the EM algorithm
   iter <- 0
   converged <- FALSE
   sumx  = sum( X )

   while ((iter<maxits)&(!converged)){
      for ( j in jj ) {
          a = rep(0,nk)
          for ( i in 1:nk ) {
              if (any(M[[i]]==j))  a[i] = lam[j] / sum(lam[M[[i]]])
          }
          newlam[j] = sum(a) /  sumx
      }
      converged = all ( abs(newlam[jj]-lam[jj]) < eps*abs(newlam[jj]) )
      iter = iter + 1
      lam = newlam
   }
   list ( lam=newlam, iter=iter, conv=converged )
}
#
# ------------------------------
# EM: Normal Distribution Model
# ------------------------------
norm.cm.EM <-
function(X, M, mu0, sd0, maxits=100, eps=1.0E-3)  {
   nk = length(X)
   J  = max(unlist(M))
   idx = unique( unlist(M) )
   jj  = idx[ idx>0 ]
   if (is.vector(M)) M <- as.list(M)
   for ( i in 1:nk )  if ( any(M[[i]]<0) )  M[[i]] = jj
 #
 # Setting the inital values
   if ( !missing(mu0) && length(mu0) != J ) mu0 = rep(mu0, l=J)
   if ( !missing(sd0) && length(sd0) != J ) sd0 = rep(sd0, l=J)
   X0 = as.list(NULL); length(X0) = J
   if ( missing(mu0) ||  missing(sd0) ) {
      for ( i in 1:nk ) {
          idx = M[[i]]
          if (length(idx) == 1) {
              if ( idx > 0 ) { X0[[idx]]=c(X0[[idx]], X[i])
              } else if ( idx == 0 ) for (j in jj) X0[[j]] = c(X0[[j]], X[i])
          }
      }
   }
   if (  missing(mu0) ) {
      mu0 = rep(NA,J)
```

```
     for ( j in jj ) {
         if ( is.null(X0[[j]]) ) {
                   mu0[j] = mean(X)
         } else  { mu0[j] = mean(X0[[j]]) }
     }
  }
  if (  missing(sd0) ) {
     sd0 = rep(NA,J)
     for ( j in jj ) {
          if ( is.null(X0[[j]]) ) {
              sd0[j] = sqrt( var(X) )
          } else {
              tmp =  var(X0[[j]])
              sd0[j] = ifelse( tmp >0, sqrt(tmp), 1)
          }
     }
  }
  # end of initial value setting

  m1 <- m2 <- array( dim=c(nk,J) )
  newmu <- newsd <- rep(NA,l=J)
  mu = mu0; sd = sd0
#
# Start the EM algorithm
  iter = 0
  converged = FALSE
  while ((iter<maxits)&&(!converged)){
     for ( i in 1:nk ) {
       z = (X[i]-mu[jj])/sd[jj]
       w = exp( dnorm(z, log=TRUE) - pnorm(-z, log.p=TRUE) )
    ## w = dnorm((X[i]-mu[jj])/sd[jj]) / (1-pnorm((X[i]-mu[jj])/sd[jj]))
       m1[i,jj]= mu[jj] + sd[jj] * w
       m2[i,jj]= mu[jj]^2 + sd[jj]^2 + sd[jj]*(mu[jj]+X[i])*w
     }
     w1 = rep(NA,l=J)
     for ( j in jj ) {
         U = rep(0,nk)
         for ( i in 1:nk ) {
             if (any(M[[i]]==j)) {
                z = (X[i]-mu[jj])/sd[jj]
                w1[jj] = exp( dnorm(z, log=TRUE) - pnorm(-z, log.p=TRUE) )
             ## w1[jj] = dnorm((X[i]-mu[jj])/sd[jj]) /
             ##            ( 1-pnorm((X[i]-mu[jj])/sd[jj]) )
                U[i] = (w1[j]/sd[j]) / sum(w1[M[[i]]] /sd[M[[i]]] )
             }
         }
         newmu[j] = mean( U*X + (1-U)*m1[,j] )
         tmp      = mean( U*X^2 + (1-U)*m2[,j] ) - newmu[j]^2
         newsd[j] = sqrt(tmp)
     }
     iter = iter + 1
     conv1 = all ( abs(newmu[jj]-mu[jj]) < eps*abs(newmu[jj]) )
     conv2 = all ( abs(newsd[jj]-sd[jj]) < eps*abs(newsd[jj]) )
     converged = conv1 && conv2
```

```
      mu = newmu; sd = newsd
   }
   list ( mu=newmu, sd=newsd, iter=iter, conv=converged )
}

# ----------------------------
# EM: Wald Distribution Model
# ----------------------------
wald.cm.EM <-
function(X, M, mu0, lam0, maxits=100, eps=1.0E-3)  {
   nk = length(X)
   J  = max(unlist(M))
   idx = unique( unlist(M) )
   jj  = idx[ idx>0 ]
   if (is.vector(M)) M <- as.list(M)
   for ( i in 1:nk )  if ( any(M[[i]]<0) )  M[[i]] = jj
 #
 # Setting the inital values
   if ( !missing(mu0)  && length(mu0)  != J )  mu0 = rep( mu0, l=J)
   if ( !missing(lam0) && length(lam0) != J ) lam0 = rep(lam0, l=J)
   X0 = as.list(NULL); length(X0) = J
   if ( missing(mu0) ||  missing(lam0) ) {
      for ( i in 1:nk ) {
          idx = M[[i]]
          if (length(idx) == 1) {
              if ( idx > 0 ) { X0[[idx]]=c(X0[[idx]], X[i])
              } else if ( idx == 0 ) {
                     for ( j in jj ) X0[[j]] = c(X0[[j]], X[i]) }
          }
      }
   }
   if  (  missing(mu0) ||  missing(lam0) ) {
       mu00 = lam00 = rep(NA,J)
       for ( j in jj ) {
           if ( is.null( X0[[j]] ) ) {
                    tmp = wald.MLE ( X )
           } else { tmp = wald.MLE ( X0[[j]] ) }
           mu00[j] = ifelse(tmp$location>0, tmp$location, 1)
          lam00[j] = ifelse(tmp$scale>0,    tmp$scale,    1)
       }
   }
   if (  missing(mu0) )   mu0 = mu00
   if (  missing(lam0) ) lam0 = lam00
   # end of initial value setting

   newmu <- newlam <- rep(NA,l=J)
   mu = mu0; lam = lam0
 #
 # Start the EM algorithm
   iter = 0
   converged = FALSE
   fz <- function(z, location, scale, xx) {
            dwald(z, location, scale) / (1-pwald(xx, location,scale))
   }
```

```
   fB <- function(z, location, scale, xx) {
        z * dwald(z, location, scale) / (1-pwald(xx, location,scale))
   }
   fC <- function(z, location, scale, xx) {
        1/z*dwald(z, location, scale) / (1-pwald(xx, location,scale))
   }
   while ((iter<maxits)&&(!converged)){
      U = array(0, dim=c(nk,J) )
      for ( j in jj ) {
          for ( i in 1:nk ) {
              if (any(M[[i]]==j)) {
                 tmp =  fz(X[i],location=mu[M[[i]]],scale=lam[M[[i]]],xx=X[i])
                 U[i,j]= fz(X[i],location=mu[j],scale=lam[j],xx=X[i])/sum(tmp)
              }
          }
      }
      mB <- mC <- array(0, dim=c(nk,J) )
      for ( i in 1:nk ) {
         for ( j in jj ) {
             if (U[i,j]  < 1) {
               mB[i,j]=integrate(fB, lower=X[i], upper=Inf, stop.on.error=FALSE,
                                 location=mu[j], scale=lam[j], xx=X[i])$value
               mC[i,j]=integrate(fC, lower=X[i], upper=Inf, stop.on.error=FALSE,
                                 location=mu[j], scale=lam[j], xx=X[i])$value
             }
         }
      }
      for ( j in jj ) {
          newmu[j] = mean( U[,j]*X + (1-U[,j])*mB[,j] )
          tmp      = mean( U[,j]/X + (1-U[,j])*mC[,j] ) - 1/newmu[j]
          newlam[j] = 1 / tmp
      }
      iter = iter + 1
      conv1 = all ( abs(newmu[jj] -mu[jj])  < eps*abs(newmu[jj]) )
      conv2 = all ( abs(newlam[jj]-lam[jj]) < eps*abs(newlam[jj]) )
      converged = conv1 && conv2
      mu = newmu; lam = newlam
      cat(".")
   }
   cat("\n * Done *\n\n")
   list ( mu=newmu, lam=newlam, iter=iter, conv=converged )
}

# ------------------------------
# EM: Weibull Distribution Model
# ------------------------------
weibull.cm.EM <-
function(X, M, alpha0, lam0, maxits=100, eps=1.0E-3)  {
   nk = length(X)
   J  = max(unlist(M))
   idx = unique( unlist(M) )
   jj  = idx[ idx>0 ]
   if (is.vector(M)) M <- as.list(M)
   for ( i in 1:nk )  if ( any(M[[i]]<0) )  M[[i]] = jj
```

```
#
# Setting the inital values
  if ( !missing(alpha0)&&length(alpha0)!=J ) alpha0 = rep(alpha0,l=J)
  if ( !missing(lam0)  && length(lam0) !=J )   lam0 = rep(lam0,  l=J)
  X0 = as.list(NULL); length(X0) = J
  n1 = numeric(J)
  if ( missing(alpha0) || missing(lam0) ) {
     for ( i in 1:nk ) {
         idx = M[[i]]
         if (length(idx) == 1) {
             if ( idx > 0 ) {
                     X0[[idx]] = c(X0[[idx]], X[i])
                      n1[idx]  = n1[idx] + 1
             } else if (idx == 0) for (j in jj) X0[[j]] = c(X0[[j]], X[i])
         }
     }
   }
   if  (  missing(alpha0) ||  missing(lam0) ) {
       alpha00 = lam00 = numeric(J)
       for ( j in jj ) {
          if ( is.null( X0[[j]] ) ) {
                   tmp = weibull.MLE ( X )
          } else { tmp = weibull.MLE ( X0[[j]] ) }
           alpha00[j] = tmp$alpha
             lam00[j] = tmp$lam
       }
  }
  if (  missing(alpha0) ) alpha0 = alpha00
  if (  missing(lam0) )     lam0 = lam00
  # end of initial value setting

  newlam <- newalpha <- rep(NA, l=J)
  lam   = lam0 ;  alpha = alpha0
#
# Start the EM algorithm
  iter <- 0
  converged <- FALSE
  sumx  = sum( X )
  EE2 = function(alpha,x,U) {
      xalpha = x^alpha
      sumU = sum(U)
      sumU*sum(xalpha*log(x))/sum(xalpha)-sumU/alpha-sum(U*log(x))
  }
  while ((iter<maxits)&(!converged)){
     for ( j in jj ) {
         U = rep(0,nk)
         for ( i in 1:nk ) {
             if (any(M[[i]]==j)) {
                tmp  = alpha[M[[i]]]*lam[M[[i]]]*X[i]^(alpha[M[[i]]]-1)
                U[i] = alpha[j]*lam[j]*X[i]^(alpha[j]-1) / sum(tmp)
             }
         }
          sumU = sum(U)
         lower = sumU / sum( U*(log(max(X))-log(X)) )
```

```
          tmp = sumU*sum(X^lower * log(X))/sum(X^lower) - sum(U*log(X))
         upper = max( abs(sumU/tmp), lower )
         while ( EE2(lower,X,U)*EE2(upper,X,U) > 0 ) {
            upper = 2 * upper
         }
         tmp = uniroot(EE2, interval=c(lower,upper),x=X, U=U)
         newalpha[j] = tmp$root
         newlam[j]   = sumU / sum(X^newalpha[j])
      }
      iter = iter + 1
      conv1 = all ( abs(newalpha[jj]-alpha[jj]) < eps*abs(newalpha[jj]) )
      conv2 = all ( abs(newlam[jj]  -  lam[jj]) < eps*abs(newlam[jj]) )
      converged = conv1 && conv2
      lam = newlam ;  alpha = newalpha
   }
   list ( alpha=newalpha, lam=newlam, iter=iter, conv=converged )
}
```

Acknowledgment

The first author's work was partially supported by Clemson RGC Award. The second author's work was partially supported by the National Science Foundation grant DMS–0243594 to the University of South Carolina. The authors thank Michael Harwell of Chemical Engineering at Clemson University for providing the bond strength data set.

References

Aalen, O. O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Annals of Statistics*, 6:534–545.

Albert, J. R. G. and Baxter, L. A. (1995). Applications of the EM algorithm to the analysis of life length data. *Applied Statistics*, 44:323–341.

Beetz, C. P. (1982). The analysis of carbon fibre strength distributions exhibiting multiple modes of failure. *Fibre Science Technology*, 16:45–59.

Blom, G. (1958). *Statistical Estimates and Transformed Beta Variates.* Wiley, New York.

Boggio, J. V. and Vingsbo, O. (1976). Tensile strength and crack nucleation in boron fibres. *Journal of Materials Science*, 11:273–282.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983). *Graphical Methods for Data Analysis.* Duxbury, Boston.

Chi, Z., Chou, T.-W., and Shen, G. (1984). Determination of single fibre strength distribution from fibre bundle testings. *Journal of Materials Science*, 19:3319–3324.

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

Crowder, M. J. (2001). *Classical Competing Risks.* Chapman & Hall.

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

Durham, S. D. and Padgett, W. J. (1997). A cumulative damage model for system failure with application to carbon fibers and composites. *Technometrics*, 39:34–44.

Farnum, N. R. and Booth, P. (1997). Uniqueness of maximum likelihood estimators of the 2-parameter Weibull distribution. *IEEE Transactions on Reliability*, 46:523–525.

Goda, K. and Fukunaga, H. (1986). The evaluation of the strength distribution of silicon carbide and alumina fibres by a multi-modal Weibull distribution. *Journal of Materials Science*, 21:4475–4480.

Guess, F. M., Usher, J. S., and Hodgson, T. J. (1991). Estimating system and component reliabilities under partial information of the cause of failure. *Journal of Statistical Planning and Inference*, 29:75–85.

Harwell, M. (1995). Microbond tests for ribbon fibers. M.S. thesis, Clemson University.

Herman, R. J. and Patell, R. K. N. (1971). Maximum likelihood estimation for multi-risk model. *Technometrics*, 13:385–396.

Ishioka, T. and Nonaka, Y. (1991). Maximum likelihood estimation of Weibull parameters for two independent competing risks. *IEEE Transactions on Reliability*, 40:71–74.

Johnson, J. W. and Thorne, D. J. (1969). Effect of internal polymer flaws on strength of carbon fibres prepared from an acrylic precursor. *Carbon*, 7:659–660.

Jones, B. F. and Wilkins, J. S. (1972). A technique for the analysis of fracture strength data for carbon fibres. *Fibre Science and Technology*, 5:315–320.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481.

King, J. R. (1971). *Probability Charts for Decision Making.* Industrial Press, New York.

Kuhn, J. W. and Padgett, W. J. (1997). Local bandwidth selection from kernel density estimation from right-censored data based on asymptotic mean absolute error. *Nonlinear Analysis, Theory, Methods & Applications*, 30:4375–4384.

Kundu, D. and Basu, S. (2000). Analysis of incomplete data in presence of competing risks. *Journal of Statistical Planning and Inference*, 87:221–239.

Layden, G. K. (1973). Fracture behaviour of boron filaments. *Journal of Materials Science*, 8:1581–1589.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons, New York, 2nd edition.

Martineau, P., Lahaye, M., Pailler, R., Naslain, R., Couzi, M., and Cruege, F. (1984). SiC filament/titanium matrix composites regraded as model composites: Part 1 filament microanalysis and strength characterization. *Journal of Materials Science*, 19:2731–2748.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons, New York.

Michael, J. R. (1979). Fundamentals of probability plotting with applications to censored data. Bell Lab. Memorandum.

Miyakawa, M. (1984). Analysis of incomplete data in competing risks model. *IEEE Transactions on Reliability*, 33:293–296.

Moeschberger, M. L. and David, H. A. (1971). Life tests under competing causes of failure and the theory of competing risks. *Biometrics*, 27:909–933.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14:945–966.

Padgett, W. J., Durham, S. D., and Mason, A. M. (1995). Weibull analysis of the strength of carbon fibers using linear and power law models for the length effect. *Journal of Composite Materials*, 29:1873–1884.

Park, C. (2005). Parameter estimation of incomplete data in competing risks using the EM algorithm. *IEEE Transactions on Reliability*, 54:282–290.

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

Phoenix, S. L. and Sexsmith, R. G. (1972). Clamp effects in fiber testing. *Journal of Composite Materials*, 29:1873–1884.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reiser, B., Guttman, I., Lin, D. K. J., Guess, F. M., and Usher, H. S. (1995). Bayesian inference for masked system lifetime data. *Applied Statistics*, 44:79–90.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall, Boca Raton, FL.

Simon, G. and Bunsell, A. R. (1984). Mechanical and structural characterization of the nicalon silicon carbide fibre. *Journal of Materials Science*, 19:3649–3657.

Stoner, E. G., Edie, D. D., and Durham, S. D. (1994). An end-effect model for the single-filament tensile test. *Journal of Materials Science*, 29:6561–6574.

Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer-Verlag, New York.

Taylor, H. M. (1994). The Poisson-Weibull flaw model for brittle fiber strength. In Galambos, J., Lechner, J., and Simiu, E., editors, *Extreme Value Theory*, pages 43–59. Kluwer, Amsterdam.

Usher, J. S. and Guess, F. M. (1989). An iterative approach for estimating component reliability from masked system life data. *Quality and Reliability Engineering International*, 5:257–261.

Usher, J. S. and Hodgson, T. J. (1988). Maximum likelihood analysis of component reliability using masked system life-test data. *IEEE Transactions on Reliability*, 37:550–555.

Wagner, D. H. (1989). Stochastic concepts in the study of size effects in the mechanical strength of highly oriented polymeric materials. *Journal of Polymer Science*, B-27:115–148.

Wilk, M. B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55:1–17.