Bayesian Estimations for Weibull Competing Risk Model with Masked Causes and Heavily Censored Data

Patrick Pamphile, Gilles Celeux

Bayesian Estimations for Weibull Competing Risk Model with Masked Causes and Heavily Censored Data

Patrick Pamphile[a,b,*], Gilles Celeux[b]

[a] *Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, Bat 407 ,91405 Orsay Cedex 9, France*
[b] *Team CELESTE Inria-Saclay, Batiment Turing 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France*

Abstract

In a reliability or maintenance analysis of a complex system, it is important to be able to identify the main causes of failure. Therefore a Weibull competing risks model is generally used. However, in this framework estimating the model parameters is a difficult ill-posed problem. Indeed, the cause of the system failure may not be identified and the experiment may be censored by the duration of the study. In addition, the other causes are naturally censored by the first one. In this paper, we propose a new method for estimating the parameters of the Weibull competing risks model, with masked causes and heavily censored data. We use a restoration of missing data through a Bayesian sampling of parameters with a weakly informative prior distribution. The mean of the posterior distribution can thus estimated by importance sampling. The proposed method is not an iterative method and therefore can be parallelized. Experiments based on simulated data and a reliability data set show that the prediction performance of the proposed method is superior to the EM algorithm and its variants, as well as the Gibbs sampler, for low to very heavy censoring rates and even in the case of weakly separated components.

Keywords: Competing risks; Weibull distribution; Masked causes; Censored data; EM algorithms; Bayesian sampling, BRM algorithm

---

*Corresponding author
Email addresses: patrick.pamphile@univeriste-paris-saclay.fr (Patrick Pamphile), gilles.celeux@inria.fr (Gilles Celeux)

Acronyms and notations

CR	competing risks
QQ-plot	quantiles-quantiles plot
ML	maximum likelihood
EM	estimation-maximization
S-EM	stochastic EM
MAP	maximum a posteriori
BR-PM	Bayesian restoration- log posterior maximization
BR-LM	Bayesian restoration- log likelihood maximization
RMSE	root mean square error
MTTF	mean time to failure

n	number of systems under study
t_{ik}	i-th system lifetime related to failure mode k
	; $\mathbf{t} = (t_{ik}, i = 1, \dots, n; k = 1, 2)$
$\widetilde{t}_{ik}$	simulated lifetime
t_i	i-th system lifetime without regard
	to failure mode: $t_i = \min_k t_{ik}$;
e_i	censoring indicator:
	$e_i = 1$, if t_i is lifetime
	$e_i = 0$ if t_i is right censored.
	; $\mathbf{e} = (e_1, \dots, e_n)$
y_i	observed data: $y_i = (t_i, e_i)$; $\mathbf{y} = (y_1, \dots, y_n)$
o_i	cause of the i-th failure; $\mathbf{o} = (o_1, \dots, o_n)$
z_i	missing data :
	simple missing ; the cause of failure $z_i = o_i$
	full missing ; the cause of failure and
	the lifetime of each cause $z_i = (o_i; (t_{ik}))$
	; $\mathbf{z} = (z_1, \dots, z_n)$
$\widetilde{z}_i$	restored missing data

f, R and λ	probability density function, reliability and failure rate of a Weibull distribution
$\theta_k = (\beta_k, \eta_k)$	shape and scale Weibull parameters of the k-th cause of failure
$\boldsymbol{\theta}$	set of the CR model parameters $= (\theta_1, \theta_2) = (\beta_1, \eta_1, \beta_2, \eta_2)$
$\widetilde{\boldsymbol{\theta}}$	simulated parameters
$\widehat{\boldsymbol{\theta}}$	parameters estimate
$L(\boldsymbol{\theta}\|\mathbf{y})$	CR model likelihood with the observed data
$L(\boldsymbol{\theta}\|\mathbf{y}, \mathbf{z})$	CR model likelihood with the complete data
$\pi(\boldsymbol{\theta})$	prior distribution
$\pi(\boldsymbol{\theta}\|\mathbf{y})$	posterior distribution knowing the observed data
$\pi(\boldsymbol{\theta}\|\mathbf{y}, \mathbf{z})$	posterior distribution knowing the observed data $\mathbf{y}$ completed by the missing data $\mathbf{z}$
$\mathbf{E}_f[X]$	expectation for $X \sim f$.

1. Introduction

Competing risk models are widely used to model lifetimes in medical or industrial applications. Indeed, they correspond to the lifetime of a patient or a system with multiple causes of failure. In order to have an effective diagnosis, it is essential to identify the distributions related to main causes of failures. Unfortunately, when a failure occurs, the cause is often unidentified. In fact, exact diagnosis of the failure cause could be impossible or too resource-consuming to be performed on every failure. In such cases, the modeler is faced with a competing risk model with hidden causes of failure (see among others [1]).

Statistical inference for competing risks models has been widely discussed by many authors (see, for example [2, 3] or [4]). In these references, different competing failure time distributions such as exponential, gamma and Weibull are considered. When the causes of the failures are hidden, one is faced with an ill-posed problem and the standard maximum likelihood method is no longer satisfactory. The EM algorithm or its stochastic variant S-EM are then used

[5]. But, In the case of real data, it is usual to have heavily censored (e.g. 70% of censored data). The EM and S-EM algorithms are then no longer effective in fitting a competing risk model with masked causes to very heavily censored data. The Bayesian method is then an alternative solution *cf.* [6, 7, 8, 9] or [10].

In addition, the case of competing Weibull distributions is widely used. Indeed, Weibull distributions exhibit a decreasing, constant or increasing hazard function, which makes them suitable for modeling complex failure data (see for instance [11] and reference therein). Unfortunately, the estimator of the shape parameters has no closed-form expression : it is necessary to use iterative numerical approximations sensitive to the starting point and the censoring rate. For Bayesian methods, there is no conjugated prior distribution for Weibull distribution when the two parameters are unknown. The estimation of the posterior distribution or its moments requires, here again, approximations by iterative numerical methods. The Gibbs algorithm then converges very slowly *cf.* [12].

In parametric estimation with missing data, Bacha and *al* [13] have proposed the Bayesian restoration maximization (BRM) method. It is a Bayesian sampling -importance -resampling method [14]: the parameters are simulated from the prior distribution and the missing data are then restored. The expectation of the posterior distribution is then estimated by an importance sampling technic [15]. One of the traps of importance sampling is to use a proposal distribution too far from target distribution. Therefore, Rubin suggests a re-sampling *cf.* [16]. In the case of a Weibull model competing with very heavily censored data, that is not sufficient.

In this work, we first of all propose to improve the BRM method by amending the proposal distribution to make it more effective in terms of variance. In addition a new proposal distribution is obtained from the maximization of the mean of the posterior distribution completed by the missing data. The efficiency of the proposed methods was evaluated by a large number of simulations for different levels of censoring rate. Simulations have shown that the new proposed methods are effective both in terms of relative bias and relative root mean square error.

The article is organized as follows. The Weibull competing risks model with masked failure causes is introduced in Section 2. Section 3 deals with the handling of missing data. The use of the EM and S-EM algorithms is reviewed. Section 4 is devoted to Bayesian estimation. The Gibbs sampler and our further improvements to the BRM method are presented. In Section 5, the different procedures are compared through simulations and a reliability data set. A discussion section and a conclusion end this paper.

2. The Weibull competing risks model

Competing risks (CR) models encompass any failure process in which there is more than one distinct causes of failure. In a reliability framework the objective is then to estimate the failure rate of the main causes of the system failure, when the exact cause of system failure is unknown.

In this article, we assume that there are two dominant failure mechanisms in competition. This assumption is often realistic in practice and allows us to simplify the presentation. The following assumptions are made throughout this paper. Let us consider n identical systems put on a life test and let $T_1, \ldots, T_n$ be the lifetime of these systems. Assume that

$$T_i = \min_{k=1,2} T_{ik} \tag{1}$$

where $(T_{ik}, k = 1, 2)$ are independent latent failure times corresponding to two competing causes of failure; T_{ik} has the Weibull probability density function (pdf)

$$f(t \,|\, \beta_k, \eta_k) = \frac{\beta_k}{\eta_k} \left(\frac{t}{\eta_k} \right)^{\beta_k - 1} \mathrm{e}^{-\left(\frac{t}{\eta_k} \right)^{\beta_k}}, \tag{2}$$

where $\beta_k > 0$ and $\eta_k > 0$ are respectively the shape and scale parameters of the k-th Weibull distribution. Its corresponding reliability function R_W and failure rate λ_W are for $t \geq 0$

$$\begin{aligned} R(t \,|\, \beta_k, \eta_k) &= \mathrm{e}^{-\left(\frac{t}{\eta_k} \right)^{\beta_k}} && (3) \\ \lambda(t \,|\, \beta_k, \eta_k) &= \frac{f(t \,|\, \beta_k, \eta_k)}{R(t \,|\, \beta_k, \eta_k)} = \frac{\beta_k}{\eta_k} \left(\frac{t}{\eta_k} \right)^{\beta_k - 1}. && (4) \end{aligned}$$

2.1. The missing data structure

Masked competing risks model is a missing data structure model. Consequently, statistical inference is to be made on incomplete data and maximum likelihood could resort to EM-like algorithms.

The observed data are as follows: $(\mathbf{t}, \mathbf{e}) = ((t_1, e_1), \ldots, (t_n, e_n)))$ where t_i is a failure time due to cause $o_i \in \{1, 2\}$ if $e_i = 1$ or a right censored time if $e_i = 0$. The cause of failure $o_i \in \{1, 2\}$ is not observed.

The missing data are of different natures. They include the unknown labels $\mathbf{o} = (o_1, \ldots, o_n)$, with $o_i \in \{1, 2\}$ and the failure times of the k-th cause $t_{ik}, k \neq o_i$ *cf.* Eq. (1).

The EM algorithm [17] or data augmentation methods [18] incorporate the missing data into the likelihood to improve the bias of the maximum likelihood estimator.

For masked Weibull competing risks, it is possible to consider the **full missing** data structure with the label and failure times of each cause, $\mathbf{z} = (o_i : e_i = 1; t_{ik}, k = 1, 2, i = 1, \ldots, n)$; or the **simple missing** data structure with only the labels, $\mathbf{z} = (o_i : e_i = 1, i = 1, \ldots, n)$. In the following, the the methods of estimation that we propose make use of the full or the simple missing structure. This important point will be specified explicitly for all the presented algorithms.

2.2. Identifiability

The question of parameter identifiability is a prerequisite for estimation. A parametric model is identifiable if a model cannot be obtained by two different parameterizations: if $f(t \mid \boldsymbol{\theta}) = f(t \mid \boldsymbol{\theta'})$ then $\boldsymbol{\theta} = \boldsymbol{\theta'}$ with the exception of one permutation. There is a lot of works on identifiability conditions for the competing risks model. For the model under study, [19] have proved that a Weibull masked CR is identifiable if and only if the shape parameters (β_k) are not equal. In the following we will assume that $\beta_1 < \beta_2$.

2.3. Graphical analysis

The presence of two main competing causes of failure can be related to system configuration (*e.g* two critical components connected in series) or using simple graphical analyses of failure data. In particular, Weibull quantiles-quantiles plot (QQ-plot) is a quick and simply confirmatory graphical method *cf.* [3] : for standard single Weibull distribution, the Weibull QQ-plot is lined up; whereas for Weibull CR, the Weibull QQ-plot is convex. In addition, a crude point estimate of parameters can be obtained from the Weibull QQ-plot: the smallest β is the predominant mode for small mission times. On the other hand, the largest β corresponds to the predominant mode for high mission times (see [3]). Therefore, with $\beta_1 < \beta_2$, the tangent on the left of the QQ-plot allows to estimate $(\beta_1 \,;\, \eta_1)$ and the tangent on the right $(\beta_2 \,;\, \eta_2)$, see Figure 1. These estimates are not precise, nor robust but nonetheless can be used as starting values for the various algorithms proposed hereafter.

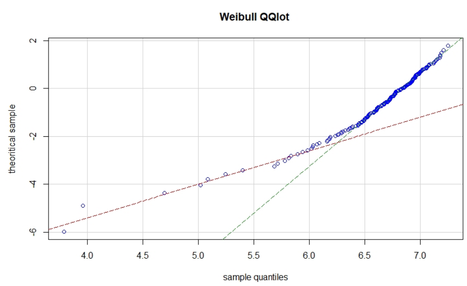


Figure 1: Weibull QQ-plot for competing risks data: the QQ-plot is convex and for small t, $R_{WCR} \simeq R(t \,|\, \beta_1 \,;\, \eta_1)$ and for large t, $R_{WCR} \simeq R(t \,|\, \beta_2 \,;\, \eta_2)$. Thus, the left tangent (right resp.) allows to estimate $(\beta_1 \,;\, \eta_1)$ $((\beta_2 \,;\, \eta_2)$ resp.) *cf.* [3]

3. Missing data and maximum likelihood methods

Let us denote $\boldsymbol{\theta} = (\theta_1, \theta_2)$, with $\theta_1 = (\beta_1, \eta_1)$, $\theta_2 = (\beta_2, \eta_2)$, the parameters of the two competing Weibull distributions and $\mathbf{y} = (\mathbf{t}, \mathbf{e})$, the observed data. When causes of failure are masked, the likelihood function of the competing risks model is

$$
\begin{aligned}
L(\boldsymbol{\theta} \,|\, \mathbf{y}) &= \left[\prod_{i:\, e_i=1} f(t_i \mid \theta_1) R(t_i \mid \theta_2) + f(t_i \mid \theta_2) R(t_i \mid \theta_1) \right] \\
&\cdot \prod_{i:\, e_i=0} R(t_i \mid \theta_1) R(t_i \mid \theta_2) \\
&= \left[\prod_{i:\, e_i=1} \lambda(t_i \mid \theta_1) + \lambda(t_i \mid \theta_2) \right] \\
&\cdot \prod_{i=1}^{n} R(t_i \mid \theta_1) R(t_i \mid \theta_2). \qquad (5)
\end{aligned}
$$

A direct maximization of likelihood is possible in the case of two competing causes, with a quasi Newton algorithm *cf.* [20]. It will be proposed in the numerical experiment section and we will see that the estimator is strongly biased for heavily censored data and weakly separated components which is a pathological case for maximum likelihood estimation.

3.1. The EM algorithm

The Weibull competing risks model at hand is a typical hidden structure data model for which the EM algorithm is appropriate. The EM algorithm consists of maximizing iteratively the expectation of the completed likelihood conditionally on the observed data, *cf.* [17]:

E-step: let $\boldsymbol{\theta}^{(b)}$ being the current value of the CR model parameter, compute:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(b)}) = \mathbf{E}_{f(\mathbf{z} \,|\, \mathbf{y}, \boldsymbol{\theta}^{(b)})} \left[\ln L(\boldsymbol{\theta} \,|\, \mathbf{y}, \mathbf{z}) \right]; \qquad (6)$$

M-step: then

$$\boldsymbol{\theta}^{(b+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(b)}).$$

For example, considering the simple missing data structure, we have $\mathbf{z} = \mathbf{o}$ and

$$\ln L(\boldsymbol{\theta} \,|\, \mathbf{y}, \mathbf{o}) = \sum_{i: e_i=1} \ln \lambda(t_i \mid \theta_{o_i}) + \sum_{i=1}^{n} \ln R(t_i \mid \theta_1) + \ln R(t_i \mid \theta_2).$$

Thus,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(b)}) = \sum_{k=1}^{2} \mathcal{Q}(\theta_k \mid \boldsymbol{\theta}^{(b)})$$

where

$$\mathcal{Q}(\theta_k \mid \boldsymbol{\theta}^{(b)}) = \sum_{i:\, e_i=1} p(o_i = k \mid \mathbf{y}; \theta_k^{(b)}) \ln \lambda(t_i \mid \theta_k^{(b)}) + \sum_{i=1}^{n} \ln R(t_i \mid \theta_k^{(b)}) \tag{7}$$

where $\theta_k^{(b)}$ is the current value of the Weibull parameters of the kth cause at iteration (b) and

$$p(o_i = k \mid \mathbf{y}; \theta_k^{(b)}) = \frac{\lambda(t_i \mid \theta_k^{(b)})}{\sum_{\ell=1}^{2} \lambda(t_i \mid \theta_\ell^{(b)})}, \tag{8}$$

is the posterior probability of the cause of failure. It leads to the maximization of $\mathcal{Q}(\theta_k \mid \boldsymbol{\theta}^{(b)})$ for each cause $k = 1, 2$ *cf.* Eq. (7):

$$\frac{1}{\beta_k^{(b+1)}} + \frac{\sum_{i:\, e_i=1} p(o_i = k \mid \mathbf{y}; \theta_k^{(b)}) \ln t_i}{\sum_{i:\, e_i=1} p(o_i = k \mid \mathbf{y}; \theta_k^{(b)})} - \frac{\sum_{i=1}^{n} t_i^{\beta_k^{(b+1)}} \ln t_i}{\sum_{i=1}^{n} t_i^{\beta_k^{(b+1)}}} = 0 \tag{9}$$

and

$$\eta_j^{(b+1)} = \left[\frac{\sum_{i=1}^{n} t_i^{\beta_k^{(b+1)}}}{\sum_{i=1}^{n} p(o_i = k \mid \mathbf{y}; \theta_k^{(b)})} \right]^{\frac{1}{\beta_k^{(b+1)}}}. \tag{10}$$

Remarks:

The advantages and drawbacks of EM algorithm are well-documented (see for instance [21]). Its main advantage is that the likelihood is increasing at each iteration. The E and M steps lead generally to closed form and simple formulas easy to program: the complexity of the M step is analogous to the complexity of the ML estimation of the parameters of a single Weibull distribution. Its main drawbacks are that it can converge painfully and its solutions can be highly dependent on its initial position, especially in a multivariate context.

3.2. The Stochastic EM algorithm

A stochastic version of EM (S-EM) leads to simpler equations in the M-step and it is by far less sensitive than EM to its starting value *cf.* [22]. The principle of the S-EM algorithm is to complete, at each iteration, the missing data by drawing them according to their conditional distribution knowing the observed data and the current value of the parameters.

Remarks:

The stochastic EM algorithm generates an ergodic Markov chain. Thus a sequence of parameter estimates via S-EM is expected to visit the whole parameter space with long sojourns in the neighborhood of the local maxima of the likelihood function. Therefore, S-EM algorithm offers a relevant starting values for EM algorithm, see [23].

3.3. Missing data restorations

Two versions for restoring missing data are possible according that the simple or the full is considered. In the full version, the origins of the observed failure times and all the censored failure times are simulated at each iteration (full restoration). In the simple version, only the origins of the observed failure times are simulated (simple restoration).

Full restoration

In the full restoration version, the missing data restoration is as follows:

E-step: for $i = 1, \dots, n$ compute the posterior probability of the cause of failure $\left(p(o_i = k \mid \mathbf{y}, \boldsymbol{\theta}^{(b)}),\ k = 1, 2\right)$ *cf.* Eq (8).

S-step:

Simulating the causes of failure,. the i-th failure is assigned to one of the causes; $(o_i^{(b)}, i = 1, \dots, n)$ are simulated according to $\left(p(o_i = k \mid \mathbf{y}, \boldsymbol{\theta}^{(b)}),\right.$ $\left. k = 1, 2\right)$:

$$\widetilde{o}_i^{(b)} \sim p(o \mid \mathbf{y}, \boldsymbol{\theta}^{(b)}). \tag{11}$$

Simulating the failure times for the censored data,. t_i with $e_i = 0$, the failure times $\widetilde{t}_{ik}$ of the two causes *cf.* Eq. (1), $k = 1, 2$, are simulated according to the Weilbull distribution with parameter $\theta_k^{(b)}$, with the constraint of being greater than t_i :

$$\widetilde{t}_{ik}^{(b)} \sim \frac{f(t \mid \theta_k^{(b)})}{R(t_i \mid \theta_k^{(b)})}. \tag{12}$$

Simulating the censored failure times for the observed data,. t_i with $e_i = 1$, it is assigned to the cause $\widetilde{o}_i^{(b)}$, ($i.e$ $\widetilde{t}_{ik} = t_i$ if $\widetilde{o}_i^{(b)} = k$); the failure time of the other competing cause, $(t_{ik}, k \neq \widetilde{o}_i^{(b)})$, is simulated according to the Weibull distribution with parameter $\theta_k^{(b)}, k \neq \widetilde{o}_i^{(b)}$, with the constraint of being greater than t_i:

$$\widetilde{t}_{ik}^{(b)} \begin{cases} = t_i & \text{if } \widetilde{o}_i^{(b)} = k; \\ \sim \dfrac{f(t \mid \theta_k^{(b)})}{R(t_i \mid \theta_k^{(b)})} & \text{if } \widetilde{o}_i^{(b)} \neq k. \end{cases} \tag{13}$$

M-step: it consists of computing the ML estimate of each component, $(\theta_k^{(b+1)}, k = 1, 2)$, from the completed sub-samples $(\widetilde{t}_{ik}^{(b)}, i = 1, \ldots, n; k = 1, 2)$, by a standard procedure, see for instance [11].

Simple restoration

In the simple restoration version, he missing data restoration is as follows:

E-step: for $i = 1, \ldots, n$ compute the distribution $\left(p(o_i = k \mid \mathbf{y}, \boldsymbol{\theta}^{(b)}),\ k = 1, 2\right)$ *cf.* Eq (8);

S-step: Simulating the causes of failure, the i-th failure is assigned to one of the causes; $(o_i^{(b)}, i = 1, \ldots, n)$ are simulated according to $\left(p(o_i = k \mid \mathbf{y}, \boldsymbol{\theta}^{(b)}),\ k = 1, 2\right)$, *cf.* Eq. (11);

M-step: it consists of computing the ML estimates of each component, $(\theta_k^{(b+1)}, k = 1, 2)$, from the censored sub-samples (t_i such that $e_i = 1$ and $\widetilde{o}_i^{(b)} = k$; t_i such that $e_i = 0$) by a standard procedure, see for instance [11].

Remark:

When restoring labels in S-step (full or simple), it can happen that no failure is attributed to a cause. Then, the S-step is rerun until each cause has a minimum number of failures.

4. Bayesian inference

The Bayesian estimation is an alternative to the ML estimation which is relevant when the information provided by the observed data is poor. It is worth recalling that we have four parameters to estimate $\boldsymbol{\theta} = (\beta_1, \eta_1, \beta_2, \eta_2)$ and the censoring rate is about 70%. The Bayesian estimation consists of incorporating prior information to reduce the uncertainty of the parameters. With the aid of a prior probability distribution $\pi(\boldsymbol{\theta})$, statistical inference is made from the posterior distribution of the parameters knowing the data. This posterior distribution, $\pi(\boldsymbol{\theta} \,|\, \mathbf{y})$, is obtained using Bayes formula,

$$
\begin{aligned}
\pi(\boldsymbol{\theta} \,|\, \mathbf{y}) &= \frac{\pi(\boldsymbol{\theta}) L(\boldsymbol{\theta} \,|\, \mathbf{y})}{\int \pi(\boldsymbol{\theta}) L(\boldsymbol{\theta} \,|\, \mathbf{y}) d\boldsymbol{\theta}} \\
&\propto \pi(\boldsymbol{\theta}) L(\boldsymbol{\theta} \,|\, \mathbf{y}). \qquad (14)
\end{aligned}
$$

This distribution combines the prior distribution for the parameters $\pi(\boldsymbol{\theta})$ with the likelihood which contains information on the parameter provided by the data. Thus, Bayesian inference is expected to be useful to produce "regularized" maximum likelihood estimation. It is used in the present article from this point of view.

Prior information may be derived from data of previous studies or expert advices. It quantifies uncertainty about the parameters and it is formalized in the prior distribution $\pi(\boldsymbol{\theta})$. Therefore, we consider Bayesian inference in order to derive a pointwise estimate of $\boldsymbol{\theta}$ which overcomes the difficulties of the EM algorithm for maximizing the likelihood in the challenging context of heavily censored masked Weibull competing risks model. A pointwise estimate of the parameter can be derived from this posterior distribution. Such Bayesian estimators are proper to produce regularized estimation of the model parameters protected from spurious and insensible local maxima of the EM algorithm.

4.1. Prior distribution

The choice of priors distribution is the key problem in the Bayesian inference. The prior distribution expresses both the engineer's expertise on parameter values and statistical uncertainty. The same joint prior has been chosen for the parameters of the two competing Weibull distributions.

Shape parameter

Thanks to its shape parameter, β, Weibull distributions have the ability to assume the characteristics of many different types of lifetimes. The range of β in various applications are addressed in [24] and [25] for instance. The same beta distribution, symmetric on range interval $[b_{min}\,;\,b_{max}]$, has been chosen as prior for shape parameters (β_k):

$$\begin{aligned}\pi(\beta)=&\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\,\frac{(\beta-b_{min})^{p-1}(b_{max}-\beta)^{q-1}}{(b_{max}-b_{min})^{p+q+1}}\\ &\cdot\mathbf{1}_{[0;1]}\left(\frac{\beta-b_{min}}{b_{max}-b_{min}}\right),\end{aligned} \tag{15}$$

where

$$\Gamma(t)=\int_0^{+\infty}t^t\mathrm{e}^{-t}dt.$$

Thus

$$\frac{\beta-b_{min}}{b_{max}-b_{min}}\sim\mathcal{B}(p,q). \tag{16}$$

The hyperparameters (b_{min}, b_{max}) reflect engineer expertise on the parameters (β_k). They shall be fixed in accordance to the field failure data. The hyperparameters b_{min} and b_{max} were chosen according to [24]: $b_{min} = 0.5$ and $b_{max} = 10$. Hyperparameters (p, q) reflect uncertainty about (β_k). A vague informative prior was chosen by setting $p = q = 1.1$.

Scale parameter

For scale parameters (η_k), the prior density of η_k given β_k is a Generalized Inverse Gamma distribution (GIG):

$$\pi(\eta_k \,|\, \beta_k) = \frac{\beta_k \, a_k^{\beta_k \, b_k}}{\Gamma(b_k)} \, \eta_k^{-(\beta_k \, b_k + 1)} \, \mathrm{e}^{-\left(\frac{a_k}{\eta_k}\right)^{\beta_k}}, \tag{17}$$

where $(a_k \,,\, b_k)$ are hyperparameters. Thus ${\eta_k}^{\beta_k}$ has an inverse gamma distribution.

4.2. Gibbs sampling

A standard solution to get an approximation of the full posterior distribution through Monte Carlo Markov Chains (MCMC) algorithms is the Gibbs sampling, see for instance [26] or [12]. The Gibbs algorithm consists of simulating iteratively all unknown quantities (parameters and missing data) according to their full conditional distributions *cf.* [27]:

For $b = 0, \dots B-1$:
$\widetilde{o}_i^{(b+1)} \sim p(o \,|\, \mathbf{y}, \widetilde{\boldsymbol{\theta}}^{(b)})$, for $i = 1, \dots, n$;
$\widetilde{\beta}_k^{(b+1)} \sim \pi(\beta \,|\, \widetilde{\eta}_k^{(b)},\, \{t_i : o_i^{(b+1)} = k\}, \mathbf{e})$, for $k = 1, 2$;
$\widetilde{\theta}_k^{(b+1)} \sim \pi(\theta \,|\, \widetilde{\beta}_k^{(b+1)}, \{t_i : o_i^{(b+1)} = k\}, \mathbf{e})$, for $k = 1, 2$.

Under mild conditions, and after a large number of iterations, $(\widetilde{\boldsymbol{\theta}}^{(1)}, \dots, \widetilde{\boldsymbol{\theta}}^{(B)})$ is a Markov chain of which distribution converges to $\pi(\theta \,|\, \mathbf{y})$, the posterior distribution. Therefore, after a burning period of M iterations the Gibbs estimator is computed from the $B - M$ last terms:

$$\widehat{\boldsymbol{\theta}}(GIBBS) = \arg \max_{b \in \{M+1, \dots, B\}} \log \pi(\widetilde{\boldsymbol{\theta}}^{(b)} \,|\, \mathbf{y}). \tag{18}$$

Remarks:. Note that most of the conditional distributions are quite simple, except for β_k: however, our choice of the prior ensures that the distribution is log-concave and thus the adaptatve rejection method of [28] can be used to simulate the full conditional distribution of $\beta_k, k = 1, 2$. The Gibbs sampler is described more precisely in the appendix A.

The Gibbs sampler produces a Markov chain $(\widetilde{\boldsymbol{\theta}}^{(1)}, \dots, \widetilde{\boldsymbol{\theta}}^{(B)})$, whose stationary law is the posterior distribution. However, the convergence is not assured and it may exhibit poor mixing properties in the case of heavily censored data *cf.* [12].

4.3. BRM method and improvements

Actually, in the present context, computing a good approximation of the posterior distribution appears to be quite difficult, but estimating the posterior mean or mode is an easier task. We then proposed a non-iterative method that is effective in the case of very heavily censored data, see [13] for the Weibull CR model and [29] for a mixture of Weibull distributions. We sketch the likelihood maximization algorithm with missing data completed by a Bayesian restoration (BR-LM algorithm):

Algorithm BR-LM

For a large number of runs $1 \leq b \leq B$

Step 1 Bayesian sampling: the parameter $\theta_1^{(b)}$ and $\theta_2^{(b)}$ are simulated according to their prior distribution;

Step 2 Restoration: the full missing data $\tilde{z}^{(b)}$ are restored according to their conditional distribution knowing the observed data and the simulated $\theta_1^{(b)}$ and $\theta_2^{(b)}$;

Step 3 Maximization: the completed log-likelihood is maximised :

$$\widehat{\boldsymbol{\theta}}^{(b)} = \arg\max_{\theta} \ln L(\boldsymbol{\theta} \,|\, \mathbf{y}, \widetilde{\mathbf{z}^{(b)}});$$

Step 4 Estimation: the mean of the posterior distribution is then obtained using a importance sampling technique *cf.* [15]: for a proposal distribution ρ

$$\boldsymbol{\mu} = \mathbf{E}_{\pi(\boldsymbol{\theta}|\mathbf{y})}[\boldsymbol{\theta}] = \mathbf{E}_{\rho(\boldsymbol{\theta})}\left[\frac{\pi(\boldsymbol{\theta}|\mathbf{y})}{\rho(\boldsymbol{\theta})}\boldsymbol{\theta}\right].$$

Here $\widehat{\rho}$, an estimate of the distribution of $(\widehat{\boldsymbol{\theta}}^{(b)}, b = 1, \dots, B)$ is used as proposal distribution:

$$\widehat{\rho}(\boldsymbol{\theta}) = \frac{1}{B \cdot h_{\boldsymbol{\theta}}} \sum_{b=1}^{B} K\left(\frac{\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(b)}}{h_{\boldsymbol{\theta}}}\right), \tag{19}$$

where K is a Gaussian kernel. The window size, $h_{\boldsymbol{\theta}}$, minimizing the mean integrated square error is given in [30]. Then assign a weight to each $\boldsymbol{\theta}^{(b)}$:

$$w_b = \frac{\pi(\boldsymbol{\theta}^{(b)})L(\boldsymbol{\theta}^{(b)} \mid \mathbf{y})}{\widehat{\rho}(\boldsymbol{\theta}^{(b)})}.$$

Defining the normalized weights

$$\widetilde{w}_b = w_b / \sum_k w_k.$$

then importance sampling estimate of $\boldsymbol{\mu}$ is

$$\widehat{\boldsymbol{\mu}} = \frac{1}{B} \sum_b \widetilde{w}_b \cdot \boldsymbol{\theta^{(b)}}.$$

Remarks:

One of the traps of importance sampling is to use a proposal distribution too far from target distribution (*i.e.* $\boldsymbol{\theta}\pi(\boldsymbol{\theta}|\mathbf{y}) \neq \boldsymbol{\mu}\,\rho(\boldsymbol{\theta})$ with $\rho(\boldsymbol{\theta}) > 0$). Many weights w_b are then almost zero, with some are very high. Rubin suggests a re-sampling *cf.* [16]. In the case of a Weibull model competing with very heavily censored data, that is not sufficient.

Here we propose a first improvement of the BR-LM algorithm. In order to correct the weight problem, we propose to use the EM algorithm on every $\widehat{\boldsymbol{\theta}}^{(b)}$. The following EM step is added to the BR-LM algorithm

Step 3' EM: The EM algorithm is initialized with $\widehat{\boldsymbol{\theta}}^{(b)}$.

The Step 3', allows to avoid having many almost zero weights and thus to improve the variability of the estimator.

Remarks:

To be as pragmatic as possible, we have chosen a very uninformative prior distribution and consequently the estimation of the expectation of likelihood completed with the Bayesian sampler can be very bad. The use of the EM algorithm then improves the estimation of the expectation of completed likelihood.

In step 2 of the BR-LM algorithm, the completed log-likelihood is maximized. One can also directly maximize the log posterior distribution completed with missing data :

Algorithm BR-PM

Step 1 Bayes: the parameter $\theta_1^{(b)}$ and $\theta_2^{(b)}$ are simulated according to their prior distribution;

Step 2 Restoration: the full missing data $\tilde{z}^{(b)}$ are restored according to their conditional distribution knowing the observed data and the simulated $\theta_1^{(b)}$ and $\theta_2^{(b)}$;

Step 3 Maximization: the completed log-posterior distribution is maximised :

$$\boldsymbol{\theta}^{(b)} = \arg\max_{\boldsymbol{\theta}} \mathbf{E}_{f(\mathbf{z}\,|\,\mathbf{y},\boldsymbol{\theta}^{(m)})}\left[\ln \pi(\boldsymbol{\theta}\,|\,\mathbf{y},\mathbf{z})\right] \tag{20}$$

Details of the maximization can be found in the Appendix.

Step 3' EM-MAP: The EM algorithm is easily modified to produce the maximum *a posteriori* (MAP) *cf.* [21], see the appendix B. Let us denote $\widehat{\boldsymbol{\theta}}^{(b)}_{MAP}$ the result of M iterations of the algorithm EM-MAP with $\widehat{\boldsymbol{\theta}}^{(b)}$ as initialization.

Step 4 Estimation: the mean of the posterior distribution is then estimated using a importance sampling technique with the empirical distribution of $(\widehat{\boldsymbol{\theta}}^{(b)}_{MAP}, b = 1, \cdots, B)$ as proposal distribution.

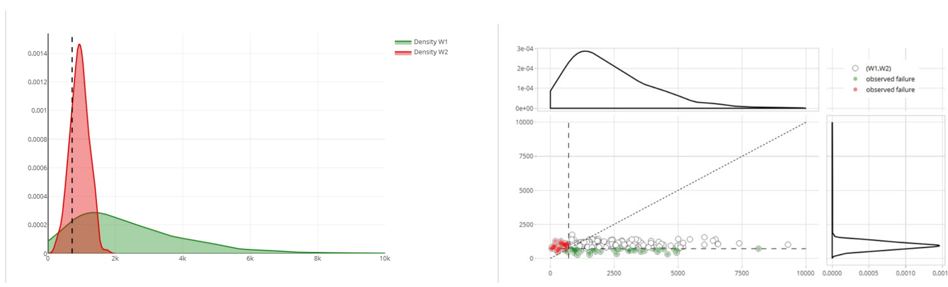
Remarks: The direct maximization of the expectation of the completed log-posterior distribution requires the use of a numerical method. However, we note that this maximization appears as a regularization of the maximization of the completed log-likelihood of BR-LM, with the prior distribution as a penalization function. The EM-MAP algorithm, like the EM algorithm for BR-LM, is used to improve the law proposition used in preferential sampling. As will be seen in the next section, this improves the estimation.

5. Numerical experiments

Numerical experiments are carried out first on simulated data and then on a set of reliability data.

5.1. Simulated data sets

We simulated the operation of a serial system of two independent components. The component lifetimes are distributed according to Weibull's law: $W_1 \sim \mathcal{W}(\beta_1 = 1.5\,;\eta_1 = 2\,500)$ and $W_2 \sim \mathcal{W}(\beta_2 = 5\,;\eta_2 = 1\,000)$. The two Weibull distributions are weakly separated: this is a pathological case for maximum likelihood estimation.



a) Components pdf

$T_1 \sim \mathcal{W}(\beta_1 = 1.5\,,\, \eta_1 = 2\,500)$

$T_2 \sim \mathcal{W}(\beta_2 = 4\,,\, \eta_2 = 1\,000)$

b) synthetic data : (t_1, t_2)

green: $t_1 < t_2 <$ censoring

red : $t_2 < t_1 <$ censoring

white: $t_1, t_2 >$ censoring

Figure 2: Distributions and synthetic data with $n = 200$, and 70% right censoring. The two Weibull distributions are weakly separated: this is a pathological case for maximum likelihood estimation. Moreover with a very high censorship rate, there is very few data for component 1 and it is extremely difficult to estimate its distribution.

The sample size was $n = 200$ and the censorship rates 70% and 20% have been used. It can be seen in Figure 2 that the number of observed lifetimes is not evenly distributed between the two components for the simulated data set. For heavily censored data (*e.g.* 70% censorship), there is very few data for component 1 and it is extremely difficult to estimate its distribution. As a matter of fact, there is few published works with such high censoring rates, although they are common in industrial contexts.

The performance of the estimators is evaluated in terms of bias and root mean square error (RMSE) from 500 replications of the dataset :

$$Bias_{rel}(\widehat{\theta}) = \frac{Bias(\widehat{\theta})}{\theta} \qquad RMSE_{rel}(\widehat{\theta}) = \frac{RMSE(\widehat{\theta})}{\theta}.$$

From these 500 replications, we compare the performance of the estimators provided by (i) the EM algorithm initiated with the S-EM algorithm with $B = 100$ iterations, (ii) the Gibbs sampling estimates with $5\,000$ iterations and $4\,700$ discarded "burn-in" iterations, (iii) the algorithm BR-LM and (iv) the algorithm BR-PM with $B = 50\,000$ runs.

For comparison, a high (70%) and low (20%) level of censorship were tested. The results are shown in Table 1, for a 70% censorship level and in Table refTab20, for a 20% censorship level. The shaded cells correspond to the optimal estimate either in terms of relative bias or RMSE. It can be noted the Bayesian restoration algorithms outperform EM with S-EM algorithms and the Gibbs sampling, even with moderately censored data. Indeed, even when there is no censorship on the right in the data collection process, there is always missing data: on the one hand the cause of the failure is masked, on the other hand the other competing cause is censored by the cause involved in the failure. The algorithms BR-LM and BR-PM provide estimations with similar biases; however, more often the algorithm BR-PM provides estimates with a lower RSME.

As mentioned previously (see Figure 2), the high rate of censorship makes (β_1, η_1) estimation much more difficult than that of (β_2, η_2). It is important to note this because $\mathcal{W}(\beta_1, \eta_1)$ is the distribution of the most reliable component.

Table 1: 70% censoring rate: the performance of the estimators is evaluated in terms of bias and root mean square error (RMSE) from 500 replications of the data set with the parameters below.

True parameters	β_1	β_2	η_1	η_2	MTTF1	MTTF2
	1.5	**5**	**2 500**	**1 000**	**2 256.86**	**906.40**
ML using BFGS quasi-newton method with WQQPlot starting value						
Average	2.74	3.72	1 857.41	966.00	1 698.53	872.35
Bias$_{rel}$	82.38%	-25.63%	-25.70%	-3.40%	-24.74%	-4.99%
RMSE$_{rel}$	145.58%	23.11%	49.82%	42.49%	54.10%	11.01%
EM with S-EM starting value						
Average	2.25	6.17	1 822.76	1 189.16	1 649.58	1 119.99
Bias$_{rel}$	50.02%	23.39%	-27.09%	18.92%	-26.91%	23.56%
RMSE$_{rel}$	70.06%	43.03%	55.89%	112.82%	59.26%	131.06%
Gibbs sampling						
Average	2.02	5.45	1561.07	895.39	1383.42	826.16
Bias$_{rel}$	34.99%	36.21%	-37.56%	-10.46%	-38.70%	-8.85%
RMSE$_{rel}$	4.82%	3.69%	7.61%	0.67%	7.47%	0.72%
BR-LM estimation with full restoration						
Average	1.86	5.99	2 024.01	1 119.38	1 829.01	1 035.53
Bias$_{rel}$	23.96%	19.85%	-19.04%	11.94%	-18.96%	14.25%
RMSE$_{rel}$	30.35%	28.85%	33.98%	35.57%	35.33%	34.12%
BR-PM estimation with full restoration						
Average	1.88	4.42	2 220.02	1 073.46	1 864.74	978.45
Bias$_{rel}$	25.33%	-19.26%	-19.20%	7.35%	-17.37%	7.97%
RMSE$_{rel}$	25.19%	19.97%	9.64%	25.54%	9.44%	25.68%

Table 2: 20% censoring rate: the performance of the estimators is evaluated in terms of bias and root mean square error (RMSE) from 500 replications of the data set with the parameters below.

True	β_1	β_2	η_1	η_2	MTTF1	MTTF2
parameters	**1.5**	**5**	**2 500**	**1 000**	**2 256.86**	**906.40**
ML using BFGS quasi-newton method with WQQPlot starting value						
Average	2.49	4.47	2 003.10	942.37	1 804.42	859.21
Bias_{rel}	66.06%	-10.65%	-19.88%	-5.76%	-20.05%	-5.22%
RMSE_{rel}	75.10%	17.83%	72.15%	3.92%	74.28%	4.30%
EM with S-EM starting value						
Average	1.70	5.56	2 315.01	1 010.64	2 103.20	933.06
Bias_{rel}	13.36%	11.16%	-7.40%	1.06%	-6.81%	2.94%
RMSE_{rel}	22.90%	15.43%	42.61%	2.98%	45.17%	3.32%
Gibbs sampling						
Average	1.67	5.38	2 277.25	1 041.92	2 102.62	868.46
Bias_{rel}	11.30%	7.64%	-8.91%	4.19%	-6.83%	-4.19%
RMSE_{rel}	32.38%	12.78%	24.67%	3.58%	24.38%	4.64%
BR-LM estimation with full restoration						
Average	1,58	5.14	2391.82	1 018.00	2 111.16	915.23
Bias_{rel}	5,28%	2.80%	-4.33%	1.20%	-6.45%	0.97%
RMSE_{rel}	15.47%	12.57%	13.89%	2.29%	16.85%	2.60%
BR-PM estimation with full restoration						
Average	1.62	4.82	2 371.24	1 008.48	2 094.71	917.40
Bias_{rel}	8.08%	-3.60%	-6.35%	0.85%	-7.18%	1.21%
RMSE_{rel}	13.28%	14.30%	11.74%	2.64%	12.20%	3.46%

5.2. Real data

Data from aircraft windshields maintenance are analyzed (*cf.* [11]). The windshield of an aircraft is a complex equipment, essentially composed of several layers of materials, including a very resistant outer skin with a heating coating just below it. All the layers are laminated at high temperature and under high pressure. Windshield failures are usually due to damage to the outer layer or failure of the heating system. During a maintenance operation, 153 aircraft windshields were inspected: 88 were classified as faulty and the remaining 66 as still operational. Their operating times (in thousands of hours) since they were commissioned were recorded. (*cf.* Table 3). The Weibull QQ-plot suggests two main competing causes of failures (see Figure 3).

Tables 4-6 provide 95% bootstrap or credible intervals for the EM with S-EM algorithms, the Gibbs sampler and the BR-PM algorithm we propose. Simulations have shown us that EM with S-EM provide estimates with high variance. The poor performance of the Gibbs algorithm illustrates the fact that it converges very slowly (*cf.* [12]). By comparing the intervals we have noticed that the algorithm BR-PM provides more accurate intervals for the scale parameter of the first mode.

Table 3: Aircraft windshields: lifetimes ($\times 1\,000\,h$). "-" indicated that the windshied had not failed, the lifetime is then censored. The censorship rate is 57.5%

0.04	−0.046	−0.14	−0.15	−0.248	−0.28	0.301
0.309	−0.313	−0.389	−0.487	0.557	−0.622	−0.9
0.943	−0.952	−0.996	−1.003	−1.01	1.07	−1.085
−1.092	1.124	−1.152	−1.183	−1.244	1.248	−1.249
−1.262	1.281	1.281	1.303	−1.36	1.432	−1.436
1.48	−1.492	1.505	1.506	1.568	−1.58	1.615
1.619	1.652	1.652	−1.719	1.757	−1.794	1.795
1.866	1.876	1.899	1.911	1.912	1.914	−1.915
−1.92	−1.963	−1.978	1.981	2.01	2.038	−2.053
−2.065	2.085	2.089	2.097	−2.117	2.135	−2.137
−2.141	2.154	−2.163	−2.183	2.19	2.194	2.223
2.224	2.229	−2.24	2.3	2.324	−2.341	2.349
2.385	−2.435	−2.464	2.481	−2.543	−2.56	−2.592
−2.6	2.61	2.625	2.632	2.646	2.661	−2.67
2.688	−2.717	−2.819	−2.82	2.823	−2.878	2.89
2.902	2.934	−2.95	2.962	2.964	3.0	−3.003
−3.102	3.103	3.114	3.117	3.166	−3.304	3.344
3.376	3.385	3.443	3.467	3.478	−3.483	−3.5
3.578	3.595	−3.622	−3.665	−3.695	3.699	3.779
3.924	−4.015	4.035	4.121	4.167	4.24	4.255
4.278	4.305	4.376	4.449	4.485	4.57	4.602
−4.628	4.663	4.694	−4.806	−4.881	−5.14	

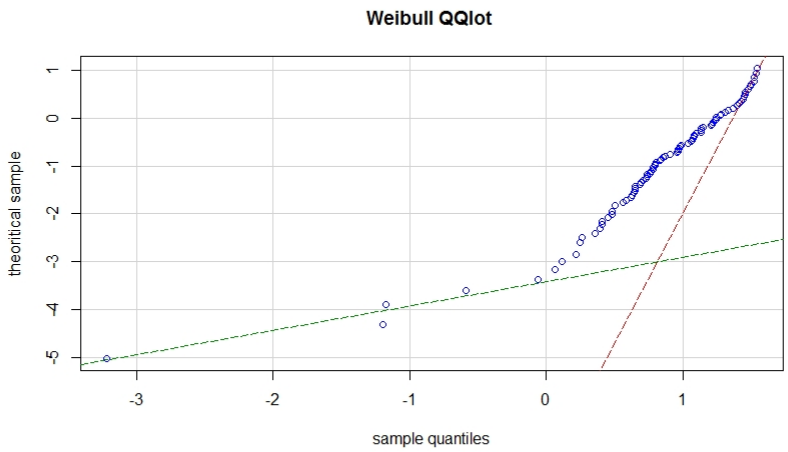


Figure 3: Weibull QQ-plot for aircraft windshield *cf.* [11]; there are two main competing causes of failures.

Table 4: Aircraft windshields: $n = 153$ and the censorship rate is 57.5%. Parameters 95% bootstrap intervals with the Algorithm S-EM and EM

θ	β_1	η_1	β_2	η_2
25%	0.636	393 645.2	2.834	3 526.2
97.5%	0.643	583 570.7	2.838	3 527.8

Table 5: Aircraft windshields: $n = 153$ and the censorship rate is 57.5%. Parameters 95% CI's with the Gibbs sampling

θ	β_1	η_1	β_2	η_2
25%	0.554	511 .0	2.660	3 429.4
97.5%	0.825	395 756.0	3.361 5	3 805.4

Table 6: Aircraft windshields: $n = 153$ and the censorship rate is 57.5%. Parameters 95% credible intervals with the Algorithm BR-PM

θ	β_1	η_1	β_2	η_2
25%	0.642 5	386 701.2	2.838	3 527.7
97.5%	0.670	394 590.4	2.851	3 534.1

6. Discussion

Algorithmic issues

For iterative algorithms, especially for S-EM and Gibbs sampler, the labels of components may be randomly switched. This label switching problem that complicates the estimation must then be addressed. The solution chosen is the one proposed by [32]: algorithm is post-processing by a $k-$means clustering.

If no failure or less than five are assigned to a cause, then the algorithm is shut down. This has an impact on the RMSE.

Full versus simple restoration

Table 7 provides comparisons between Full and Simple restoration for the BR-PM algorithm (*cf.* section 3.3). It can be seen that the Simple restoration version gives poorer results in terms of bias.

Table 7: Comparison Full versus Simple restoration for **BR-PM algorithm**. Synthetic data with 70% right censoring

	β_1	β_2	η_1	η_2	MTTF1	MTTF2
	1.5	**4**	**2 500**	**1 000**	**2256.863**	**906.4025**
Full restoration (masked causes and censored lifetimes)						
Average	1.88	4.42	2 220.02	1 073.46	1 864.74	978.45
Bias$_{rel}$	25.33%	-19.26%	-19.20%	7.35%	-17.37%	7.97%
RMSE$_{rel}$	25.19%	19.97%	9.64%	25.54%	9.44%	25.68%
Simple restoration (only masked causes)						
Mean	2.01	7.16	1373.51	1389.99	1217.71	1301.65
Bias$_{rel}$	34.24%	43.14%	-45.06%	39.00%	-46.04%	41.77%
RMSE$_{rel}$	18.05%	16.64%	11.57%	11.00%	11.58%	11.25%

Bayesian restoration

The EM algorithm is rather stable but tends to be biased in the case of heavily censored data. As can be seen in Figure 4, the initialization with the S-EM algorithm does not correct the bias. With the help of Figure 5 we can observe the principle of the Bayesian restoration algorithms we propose. The first step is Bayesian sampling. The completed likelihood maximization step then acts as an importance re-sampling. The EM step then acts as a regularization step.

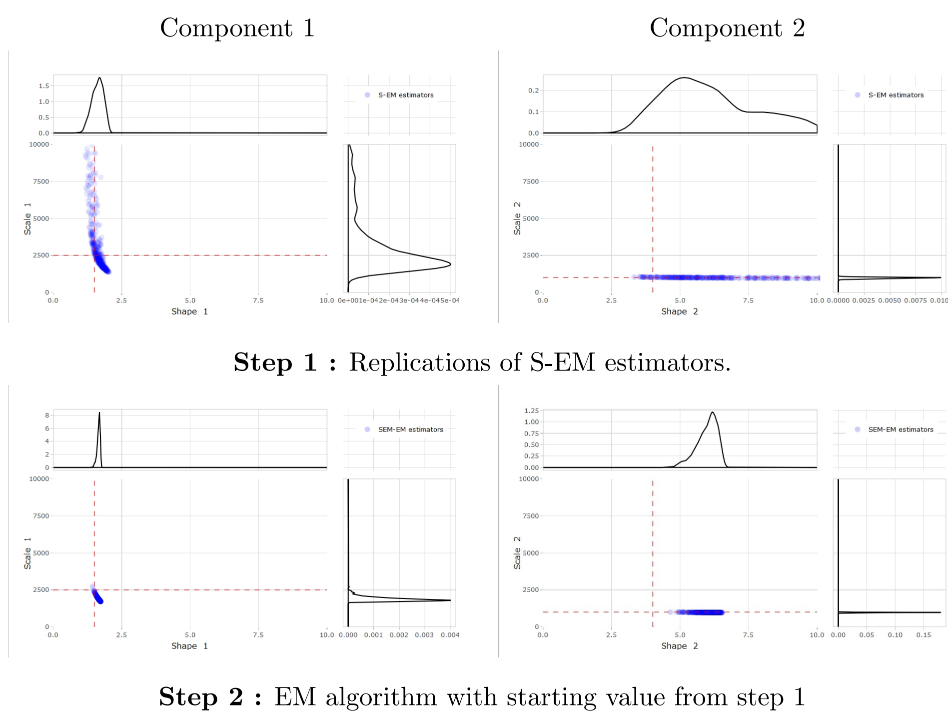
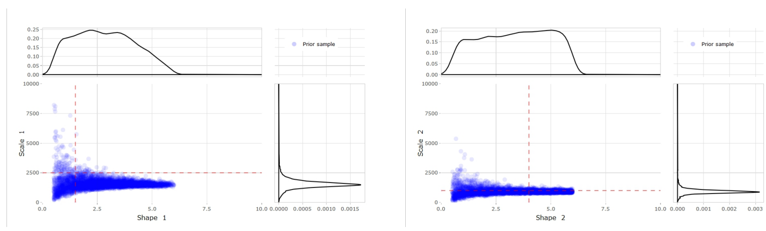
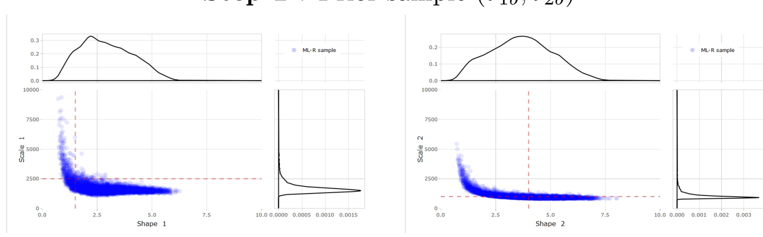


Figure 4: EM algorithm with S-EM estimate as starting values.

Component 1 Component 2

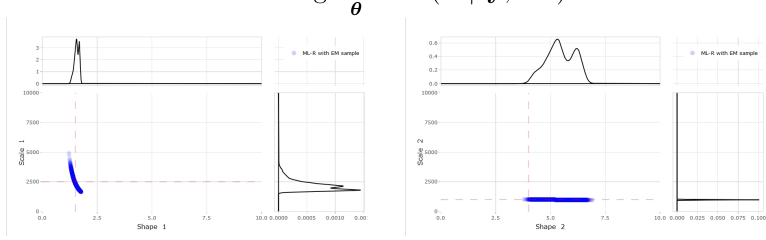


Step 1 : Prior sample $(\widetilde{\theta}_{1b}, \widetilde{\theta}_{2b})$



Step 2 : Maximization of the completed likelihood estimator after a full Bayesian restoration of missing data from step 1

$$\widehat{\widetilde{\boldsymbol{\theta}}}^{(b)} = \arg\max_{\boldsymbol{\theta}} \ln L(\boldsymbol{\theta} \mid \mathbf{y}, \widetilde{\mathbf{z}}^{(b)})$$



Step 3' : EM algorithm with starting value from step 2

Figure 5: BR-LM with EM algorithm

7. Conclusion

Estimating the parameters of a competing risks model is a ill-posed problem when the causes of failures are masked. The problem is then even more tricky in the framework of competing Weibull distributions with heavily censored data. The EM algorithm and its variants are usually used in the case of missing data. But they are not satisfactory in case of data heavily censored and particularly in the pathological case for the likelihood where the distributions are weakly separated. The Gibbs sampler have difficulty to converge when they are initialized far from the true values of the parameters. The BR-LM algorithm is suitable for parametric estimation with missing data. It is based on the restoration of missing data using a Bayesian sampling and the use of a importance sampling technic. In this work, we first show how to improve the proposal distribution of the importance sampling. We then present a regularized version called BR-PM. Parameters estimation is improved by directly maximizing the expectation of the posterior distribution completed by the missing data. The proposed methods are not an iterative method and therefore can be parallelized. The numerous simulations carried out show the effectiveness of the proposed improvements in terms of bias and variance for low to very heavy censoring rates and even in the case of weakly separated components.

8. Acknowledgements

Appendices

Appendix A: Gibbs sampler for the Weibull CR model

In order to implement the Gibbs sampler, the full conditional distribution of all parameters must be derived.

1. Cause of failure

$$\pi(o_i = k \mid \mathbf{y}, \mathbf{o}^{(b)}_{(-i)}, \boldsymbol{\beta}^{(b)}, \boldsymbol{\eta}^{(b)}) = \frac{\lambda(t_i \mid \beta_k^{(b)}, \eta_k^{(b)})}{\lambda(t_i \mid \beta_1^{(b)}, \eta_1^{(b)}) + \lambda(t_i \mid \beta_2^{(b)}, \eta_2^{(b)})} \quad (21)$$

2. Shape

$$\pi(\beta_k \mid \mathbf{y}, \mathbf{o}^{(b)}, \boldsymbol{\beta}^{(b)}_{(-k)}, \boldsymbol{\eta}^{(b)})$$

$$\propto \pi(\beta_k) \cdot \pi(\eta_k^{(b)} \mid \beta_k) \cdot \prod_{i:\, o_i^{(b)}=k} \lambda(t_i \mid \beta_k, \eta_k^{(b)}) \cdot \prod_{i=1}^{n} R(t_i \mid \beta_k, \eta_k^{(b)})$$

$$\propto f(\beta_k \mid \mathbf{t}, \mathbf{o}^{(b)}, \eta_k^{(b)})$$

Lemma 1: the posterior marginal distribution $\pi(\beta_k \mid \mathbf{y}, \mathbf{o}^{(b)}, \boldsymbol{\beta}^{(b)}_{(-k)}, \boldsymbol{\eta}^{(b)})$ is log-concave.

$$
\begin{aligned}
&\ln f(\beta_k \,|\, \mathbf{t}, \mathbf{o}^{(b)}, \eta_k^{(b)}) \\
&\qquad = (p-1)\ln(\beta_k - b_{min}) + (q-1)\ln(b_{max} - \beta_k) \\
&\qquad\quad + \ln(\beta_k) + \beta_k\, b_k\, \ln(a_k) - \beta_k\, b_k\, \ln(\eta_k^{(b)}) - \left(\frac{a_k}{\eta_k^{(b)}}\right)^{\beta_k} \\
&\qquad\quad + \#\{o_i^{(b)} = k\}\, \ln(\beta_k) + \beta_k \sum_{i:\, o_i = k} \ln\left(\frac{t_i}{\eta_k^{(b)}}\right) \\
&\qquad\quad - \sum_i^n \left(\frac{t_i}{\eta_k^{(b)}}\right)^{\beta_k} \\
&\qquad\quad = (p-1)\ln(\beta_k - b_{min}) + (q-1)\ln(b_{max} - \beta_k) \\
&\qquad\quad + (\#\{o_i^{(b)} = k\} + 1)\ln(\beta_k) \\
&\qquad\quad + \beta_k \left[b_k\, \ln\left(\frac{a_k}{\eta_k^{(b)}}\right) + \sum_{i:\, o_i = k} \ln\left(\frac{t_i}{\eta_k^{(b)}}\right) \right] \\
&\qquad\quad - \left(\frac{a_k}{\eta_k^{(b)}}\right)^{\beta_k} - \sum_i^n \left(\frac{t_i}{\eta_k^{(b)}}\right)^{\beta_k}
\end{aligned}
\tag{22}
$$

$$
\begin{aligned}
&\frac{\partial \ln f}{\partial \beta_k}(\beta_k \,|\, \mathbf{t}, \mathbf{o}^{(b)}, \eta_k^{(b)}) \\
&\qquad = \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} + \frac{1 + \#\{o_i^{(b)} = k\}}{\beta_k} \\
&\qquad\qquad + b_k\, \ln\left(\frac{a_k}{\eta_k}\right) + \sum_{i: o_i = k} \ln\left(\frac{t_i}{\eta_k}\right) \\
&\qquad\qquad\quad - \ln\left(\frac{a_k}{\eta_k}\right)\left(\frac{a_k}{\eta_k}\right)^{\beta_k} - \sum_i^n \ln\left(\frac{t_i}{\eta_k}\right)\left(\frac{t_i}{\eta_k}\right)^{\beta_k}
\end{aligned}
\tag{23}
$$

$$\frac{\partial^2 \ln f}{\partial \beta_k^2}(\beta_k \,|\, \mathbf{t}, \mathbf{o}^{(b)}, \eta_k^{(b)})$$
$$= -\frac{p-1}{(\beta_k - b_{min})^2} - \frac{q-1}{(b_{max}-\beta_k)^2}$$
$$- \frac{\#\{o_i^{(b)} = k\} + 1}{\beta_k^2} - \left[\ln\left(\frac{a_k}{\eta_k^{(b)}}\right)\right]^2 \left(\frac{a_k}{\eta_k^{(b)}}\right)^{\beta_k}$$
$$- \sum_i \left[\ln\left(\frac{t_i}{\eta_k^{(b)}}\right)\right]^2 \left(\frac{t_i}{\eta_k^{(b)}}\right)^{\beta_k}$$

Hence, the adaptative rejection method proposed by [28] to sample univariate log-concave probability density function can be used here.

3. Scale

$$\pi(\eta_k \,|\, \mathbf{y}, \mathbf{o}^{(b)}, \boldsymbol{\beta}^{(b)}, \boldsymbol{\eta}_{(-k)}^{(b)})$$
$$= \pi(\eta_k \,|\, \beta_k^{(b)}) \cdot \left(\prod_{i:\, o_i^{(b)}=k} \lambda(t_i \,|\, \beta_k^{(b)}, \eta_k) \right)$$
$$\cdot \prod_{i=1}^{n} R(t_i \,|\, \beta_k^{(b)}, \eta_k)$$
$$\propto \beta_k^{(b)}\, a_k^{\beta_k^{(b)}\, b_k}\, \eta_k^{-(\beta_k^{(b)}\, b_k + 1)}\, \mathrm{e}^{-\left(\frac{a_k}{\eta_k}\right)^{\beta_k^{(b)}}}$$
$$\cdot \left(\prod_{i:\, o_i^{(b)}=k} \beta_k^{(b)} \eta_k{}^{-\beta_k^{(b)}}\, t_i^{\beta_k^{(b)}-1} \right) \cdot \exp\left(\sum_i \left(\frac{t_i}{\eta_k}\right)^{\beta_k^{(b)}} \right)$$
$$\propto \beta_k^{(b)}\, \beta_k^{(b)\,\#\{o_i^{(b)}=k\}}\, a_k^{\beta_k^{(b)}\, b_k}$$
$$\cdot \prod_{i:\, o_i^{(b)}=k} t_i^{\beta_k^{(b)}-1}\, \eta_k^{-(\beta_k^{(b)}[b_k + \#\{o_i^{(b)}=k\}]+1)}$$
$$\cdot\ \exp - \frac{\sum_i t_i^{\beta_k^{(b)}} + a_k^{\beta_k^{(b)}}}{\eta_k^{\beta_k^{(b)}}}.$$

Thus, $\pi(\eta_k \,|\, \mathbf{y}, \mathbf{o}^{(b)}, \boldsymbol{\beta}^{(b)}, \boldsymbol{\eta}_{(-k)}^{(b)})$ is the generalized inverse gamma

$$GIG\left(\#\{o_i^{(b)} = k\} + b_k; \left[\sum_i t_i^{\beta_k^{(b)}} + a_k^{\beta_k^{(b)}} \right]^{1/\beta_k^{(b)}}; \beta_k^{(b)} \right).$$

Appendix B: Maximization step of EM for the posterior distribution

In Algorithm 4, the maximization step is performed by equating to zero the partial derivatives of conditionnel expectation of the completed log posterior with respect to each parameter *cf.* Eq. (20):

$$\mathbf{E}_{f(\mathbf{z}\,|\,\mathbf{y},\boldsymbol{\theta}^{(m)})}\left[\ln L(\boldsymbol{\theta}\,|\,\mathbf{y},\mathbf{z})\right] + \ln \pi(\boldsymbol{\theta}).$$

For a simple restoration we get :

$$\begin{aligned}
&\frac{\partial \mathbf{E}_{f(\mathbf{o}\,|\,\mathbf{y},\boldsymbol{\theta}^{(m)})}\left[\ln L(\boldsymbol{\theta}\,|\,\mathbf{y},\mathbf{o})\right]}{\partial \eta_k}(\eta_k) + \frac{\partial \ln \pi}{\partial \eta_k}(\eta_k\,|\,\beta_k) \\
=& \sum_{i:e_i=1} p(o_i = k \mid \mathbf{y};\boldsymbol{\theta}^{(m)}) \frac{\partial \ln \lambda}{\partial \eta_k}(t_i \mid \theta_k) \\
&+ \sum_{i=1}^{n} \frac{\partial \ln R}{\partial \eta_k}(t_i \mid \theta_k) + \frac{\partial \ln \pi}{\partial \eta_k}(\eta_k\,|\,\beta_k) \\
=& -\frac{\beta_k}{\eta_k} \sum_{i:e_i=1} p(o_i = k \mid \mathbf{y};\boldsymbol{\theta}^{(m)}) + \frac{\beta_k}{\eta_k^{\beta_k+1}} \sum_{i=1}^{n} t_i^{\beta_k} \\
&- \frac{b_k \cdot \beta_k + 1}{\eta_k} + \frac{\beta_k}{\eta_k^{\beta_k+1}} a_k^{\beta_k} \\
=& \frac{\beta_k}{\eta_k}\left(\frac{\sum_{i=1}^{n} t_i^{\beta_k} + a_k^{\beta_k}}{\eta_k^{\beta_k}} - \sum_{i:e_i=1} p(o_i = k \mid \mathbf{y};\boldsymbol{\theta}^{(m)}) - b_k - \frac{1}{\beta_k}\right)
\end{aligned} \tag{24}$$

Thus

$$\begin{aligned}
&\frac{\partial \mathbf{E}_{f(\mathbf{o}\,|\,\mathbf{y},\boldsymbol{\theta}^{(m)})}\left[\ln L(\boldsymbol{\theta}\,|\,\mathbf{y},\mathbf{o})\right]}{\partial \eta_k}(\eta_k) + \frac{\partial \ln \pi}{\partial \eta_k}(\eta_k\,|\,\beta_k) = 0 \\
\Leftrightarrow & \sum_{i=1}^{n} t_i^{\beta_k} + a_k^{\beta_k} = \eta_k^{\beta_k}\left(\sum_{i:e_i=1} p(o_i = k \mid \mathbf{y};\boldsymbol{\theta}^{(m)}) + b_k + \frac{1}{\beta_k}\right).
\end{aligned}$$

Therefore

$$\eta_k = \left(\frac{\sum_{i=1}^{n} t_i^{\beta_k} + a_k^{\beta_k}}{\sum_{i:e_i=1} p(o_i = k \mid \mathbf{y};\boldsymbol{\theta}^{(m)}) + b_k + \frac{1}{\beta_k}}\right)^{1/\beta_k}. \tag{25}$$

$$
\begin{aligned}
&\frac{\partial \mathbf{E}_{f(\mathbf{o}\,|\,\mathbf{y},\boldsymbol{\theta}^{(m)})}\left[\ln L(\boldsymbol{\theta}\,|\,\mathbf{y},\mathbf{o})\right]}{\partial \beta_k}(\beta_k) + \frac{\partial \ln \pi}{\partial \beta_k}(\beta_k) + \frac{\partial \ln \pi}{\partial \beta_k}(\eta_k \mid \beta_k) \\
&= \sum_{i:e_i=1} p(o_i = k \mid \mathbf{y}; \boldsymbol{\theta}^{(m)}) \frac{\partial \ln \lambda}{\partial \beta_k}(t_i \mid \theta_k) \\
&\quad + \sum_{i=1}^{n} \frac{\partial \ln R}{\partial \beta_k}(t_i \mid \theta_k) + \frac{\partial \ln \pi(\beta_k)}{\partial \beta_k} + \frac{\partial \ln \pi}{\partial \beta_k}(\eta_k \mid \beta_k) \\
&= \sum_{i:e_i=1} p(o_i = k \mid \mathbf{y}; \boldsymbol{\theta}^{(m)}) \left[\frac{1}{\beta_k} + \ln(t_i) - \ln(\eta_k)\right] \\
&\quad - \frac{1}{\eta_k^{\beta_k}} \sum_{i=1}^{n} \left[\ln(t_i) - \ln(\eta_k)\right] t_i^{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\
&\quad + \frac{1}{\beta_k} + b_k \ln(a_k) - b_k \ln(\eta_k) + \frac{1}{\eta_k^{\beta_k}} \left[\ln(\eta_k) - \ln(a_k)\right] a_k^{\beta_k} \\
&= \frac{\sum_{i:e_i=1} p(o_i = k \mid \mathbf{y}; \boldsymbol{\theta}^{(m)}) + 1}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\
&\quad + \ln(\eta_k) \left[\frac{1}{\eta_k^{\beta_k}} \left(\sum_{i=1}^{n} t_i^{\beta_k} + a_k^{\beta_k}\right) - b_k - \sum_{i:e_i=1} p(o_i = k \mid \mathbf{y}; \boldsymbol{\theta}^{(m)})\right] \\
&\quad - \frac{1}{\eta_k^{\beta_k}} \left(\sum_{i=1}^{n} \ln(t_i) t_i^{\beta_k} + \ln(a_k) a_k^{\beta_k}\right) \\
&\quad + \sum_{i:e_i=1} p(o_i = k \mid \mathbf{y}; \boldsymbol{\theta}^{(m)}) \ln(t_i) + b_k \ln(a_k)
\end{aligned} \tag{26}
$$

Upon substituting Eq. (25) in Eq. (26), it follows

$$\frac{\partial \mathbf{E}_{f(\mathbf{o}\,|\,\mathbf{y},\boldsymbol{\theta}^{(m)})}\left[\ln L(\boldsymbol{\theta}\,|\,\mathbf{y},\mathbf{o})\right]}{\partial \beta_k}(\beta_k) + \frac{\partial \ln \pi}{\partial \beta_k}(\beta_k) + \frac{\partial \ln \pi}{\partial \beta_k}(\eta_k\,|\,\beta_k) = 0$$

$$\Leftrightarrow \frac{\sum_{i:e_i=1} p(o_i = k \mid \mathbf{y}; \boldsymbol{\theta}^{(m)}) + 1}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k}$$

$$-\left(\sum_{i:e_i=1} p(o_i = k \mid \mathbf{y}; \boldsymbol{\theta}^{(m)}) + b_k + \frac{1}{\beta_k}\right)\left[\frac{\sum_{i=1}^{n} \ln(t_i)\, t_i^{\beta_k} + \ln(a_k)\, a^{\beta_k}}{\sum_{i=1}^{n} t_i^{\beta_k} + a_k^{\beta_k}}\right]$$

$$+\frac{1}{\beta_k^2}\ln\left(\frac{\sum_{i=1}^{n} t_i^{\beta_k} + a_k^{\beta_k}}{\sum_{i:e_i=1} p(o_i = k \mid \mathbf{y}; \boldsymbol{\theta}^{(m)}) + b_k + \frac{1}{\beta_k}}\right)$$

$$+\sum_{i:e_i=1} p(o_i = k \mid \mathbf{y}; \boldsymbol{\theta}^{(m)}) \ln(t_i) + b_k \ln(a_k) = 0 \tag{27}$$

References

References

[1] M. J. Crowder, Classical competing risks, Chapman and Hall/CRC, 2001. 4

[2] T. Ishioka, Y. Nonaka, Maximum likelihood estimation of Weibull parameters for two independent competing risk, IEEE Transactions on Reliability 40 (1) (1991) 71–74. doi:10.1109/24.75338. 4

[3] R. Jiang, D. Murthy, Study of n-fold Weibull competing risk model, Mathematical and computer modelling 38 (11-13) (2003) 1259–1273. doi:10.1016/s0895-7177(03)90128-2. 4, 8

[4] A. Sarhan, D. Hamilton, B. Smith, Statistical analysis of competing risks models, Reliability Engineering & System Safety 95 (2010) 953–962. doi:10.1016/j.ress.2010.04.006. 4

[5] C. Park, Parameter estimation of incomplete data in competing risks using the EM algorithm, IEEE Transactions on Reliability 54 (2005) 282–290. doi:10.1109/tr.2005.846360. 5

[6] J. O. Berger, D. Sun, Bayesian analysis for the poly-Weibull distribution, Journal of the American Statistical Association 88 (424) (1993) 1412–1418. doi:10.1080/01621459.1993.10476426. 5

[7] R. Ranjan, S. K. Upadhyay, Classical and Bayesian estimation for the parameters of a competing risk model based on minimum of exponential and gamma failures, IEEE Transactions on Reliability 65 (3) (2016) 1522–1535. doi:10.1109/tr.2016.2575439. 5

[8] D. Kundu, D. Mitra, A. Ganguly, Analysis of left truncated and right censored competing risks data, Computational Statistics & Data Analysis 108 (2017) 12–26. doi:10.1016/j.csda.2016.10.020. 5

[9] L. Wang, Inference for Weibull competing risks data under generalized progressive hybrid censoring, IEEE Transactions on Reliability 67 (3) (2018) 998–1007. doi:10.1109/tr.2018.2828436. 5

[10] F. Azizi, F. Haghighi, N. Tabibi Gilani, Statistical inference for competing risks model under progressive interval censored weibull data, Communications in Statistics-Simulation and Computation 49 (7) (2020) 1931–1944. doi:10.1080/03610918.2018.1510522. 5

[11] D. N. P. Murthy, M. . Xie, R. Jiang, Weibull Models, Wiley, 2003. 5, 12, 23, 25

[12] G. Casella, E. I. George, Explaining the Gibbs sampler, The American Statistician 46 (3) (1992) 167–174. doi:10.2307/2685208. 5, 15, 16, 23

[13] M. Bacha, G. Celeux, E. Idée, A. Lannoy, D. Vasseur, Estimation de durées de vie fortement censurées, Eyrolles, 1998. 5, 16

[14] A. E. Gelfand, A. F. Smith, Sampling-based approaches to calculating marginal densities, Journal of the American statistical association 85 (410) (1990) 398–409. doi:10.21236/ada208388. 5

[15] C. P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer-Verlag, 2013, second Edition. 5, 16

[16] D. B. Rubin, A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm, Journal of the American Statistical Association 82 (1987) 543–546. 5, 17

[17] A. D. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM-algorithm, Journal of the Royal Statistical Society: Series B 39 (1977) 1–38. doi:10.1111/j.2517-6161.1977.tb01600.x. 7, 9

[18] M. A. Tanner, W. H. Wong, From EM to data augmentation: the emergence of MCMC Bayesian computation in the 1980s, Statistical science 25 (4) (2010) 506–516. doi:10.1214/10-sts341. 7

[19] Y. Liang, D. Sun, Identifiability of masking probabilities in competing risks models with emphasis on Weibull models, Communications in Statistics-Theory and Methods 45 (7) (2016) 2143–2157. doi:10.1080/03610926.2013.878358. 7

[20] J. E. Dennis, Jr, J. J. Moré, Quasi-Newton methods, motivation and theory, SIAM review 19 (1) (1977) 46–89. doi:10.1137/1019005. 9

[21] G. McLachlan, T. Krishnan, The EM algorithm and extensions, Vol. 382, John Wiley & Sons, 2007. 10, 18

[22] G. Celeux, D. Chauveau, J. Diebolt, Stochastic versions of the EM-algorithm: an experimental study in the mixture case, Journal of Statistical Computation and Simulation 55 (1996) 287–314. doi:10.1080/00949659608811772. 11

[23] C. Biernacki, G. Celeux, G. Govaert, Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate gaussian mixture models, Computational Statistics & Data Analysis 41 (3-4) (2003) 561–575. doi:10.1016/s0167-9473(02)00163-9. 11

[24] H. P. Barringer, Weibull databases (2000).
URL http://www.barringer1.com/wdbase.htm 14

[25] R. Jiang, D. Murthy, A study of weibull shape parameter: Properties and significance, Reliability Engineering & System Safety 96 (12) (2011) 1619–1626. doi:10.1016/j.ress.2011.09.003. 14

[26] M. A. Tanner, W. H. Wong, The calculation of posterior distributions by data augmentation, Journal of the American statistical Association 82 (398) (1987) 528–540. doi:10.1080/01621459.1987.10478458. 15

[27] G. O. Roberts, A. F. Smith, Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms, Stochastic processes and their applications 49 (2) (1994) 207–216. doi:10.1016/0304-4149(94)90134-1. 15

[28] W. R. Gilks, P. Wild, Adaptive rejection sampling for Gibbs sampling, Journal of the Royal Statistical Society: Series C (Applied Statistics) 41 (2) (1992) 337–348. doi:10.2307/2347565. 15, 32

[29] F. Ducros, P. Pamphile, Bayesian estimation of Weibull mixture in heavily censored data setting, Reliability Engineering & System Safety 180 (2018) 453–462. doi:10.1016/j.ress.2018.08.008. 16

[30] D. W. Scott, Multivariate density estimation: theory, practice, and visualization, John Wiley and Sons, 2015. 17

[31] D. B. Rubin, Using the SIR algorithm to simulate posterior distributions, Bayesian statistics 3 (1988) 395–402.

[32] G. Celeux, Bayesian inference for mixture: The label switching problem, in: Compstat, Springer, 1998, pp. 227–232. doi:10.1007/978-3-662-01131-7_26. 26