



HAL
open science

Estimating parameters of the Weibull Competing Risk model with Masked Causes and Heavily Censored Data

Gilles Celeux, Patrick Pamphile

► **To cite this version:**

Gilles Celeux, Patrick Pamphile. Estimating parameters of the Weibull Competing Risk model with Masked Causes and Heavily Censored Data. 2020. hal-02410489v2

HAL Id: hal-02410489

<https://inria.hal.science/hal-02410489v2>

Preprint submitted on 15 Oct 2020 (v2), last revised 27 Aug 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating parameters of the Weibull Competing Risk model with Masked Causes and Heavily Censored Data

Gilles Celeux² and Patrick Pamphile^{*1,2}

¹Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, Bat 407 ,91405 Orsay Cedex 9, France

²CELESTE Inria-Saclay, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France

October 15, 2020

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 risk 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 may also 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 risk 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 be 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 maximum likelihood method, the standard EM algorithm and the Gibbs sampler, for low to very heavy censoring rates.

*Corresponding author: patrick.pamphile@u-psud.fr

Nomenclature

Acronym

CR	Competing Risks model
QQ-plot	Quantiles-Quantiles plot
ML	Maximum Likelihood
EM	Estimation-Maximization
SEM	Stochastic EM
MCMC	Markov chain Monte Carlo
SIR	Sampling Importance Resampling
BR-PM	Bayesian Restoration- log Posterior Maximization
BR-LM	Bayesian Restoration - log Likelihood Maximization
RMSE	Root Mean Square Error
MTTF	Mean Time To Failure

Notation

n	number of systems under study
t_{ik}	i -th system lifetime related to failure mode k ; $\mathbf{t} = (t_{ik}, i = 1, \dots, n; k = 1, 2)$
\tilde{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 $z_i = (o_i; (t_{ik}))$; $\mathbf{z} = (z_1, \dots, z_n)$
\tilde{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 parameter of the k -th cause of failure
$\boldsymbol{\theta}$	set of the CR model parameters $= (\theta_1, \theta_2) = (\beta_1, \eta_1, \beta_2, \eta_2)$
$\tilde{\boldsymbol{\theta}}$	simulated parameters
$\hat{\boldsymbol{\theta}}$	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

Weibull distributions are extensively used for modeling life data in medical or industrial applications. Indeed, Weibull distributions exhibit a decreasing, constant or increasing hazard function, which makes them suitable for modeling complex failure data (see for instance Murthy et al. [2003] and reference therein). However, possible heterogeneity in the failure data cannot be taken into account with a simple Weibull distribution. In particular the system failure may be due to different competing causes. For example, serial systems for which failure is due to the failure of one of its components (*i.e* Mukhopadhyay and Basu [2007]). Or even, a system exposed to accidental failure or wear due to aging (*i.e* Bertholon et al. [2006]). In practice it is essential to identify the Weibull distributions related to main causes of failures. Unfortunately, the exact cause of a failure is often unknown. In fact, exact diagnosis of the failure cause could be impossible or too resource-consuming to be performed on every failed system. In such cases the modeler is facing Weibull competing risk models with masked failure causes (see among others Crowder [2001] or Murthy et al. [2003]).

The maximization of the likelihood of a single Weibull distribution requires the use of numerical optimization procedures. As failure data are most often right censored, the maximum likelihood estimation is biased. Statistical inference with the Weibull masked competing risks model is intrinsically difficult since the failure time of component not causing the failure is always right censored. Obviously, the estimation task is made more difficult if, in addition, the component causing the failure is unknown. Statistical inference for competing risks models has been widely discussed by many authors (see, for example Ishioka and Nonaka [1991], Berger and Sun [1993], Bacha et al. [1998], Jiang and Murthy [2003], Park [2005], Sarhan et al. [2010], Ranjan and Upadhyay [2016], Kundu et al. [2017], Wang [2018] or Freels et al. [2019]). In these references, different competing failure time distributions such as exponential, gamma and Weibull are considered. Their parameters are estimated with different methods: graphical estimation, maximum likelihood estimation through EM algorithm and variants or Bayesian inference. Moreover, causes of the failures could be known or masked and all lifetimes can be observed or possibly censored with various modes of censorship (right, progressive,...). Thus above mentioned authors have paid effort to propose efficient statistical estimation procedures for competing risk models with masked causes. However, the resulting procedures remain unsatisfactory in ill-posed settings when the observed failures times are rare in regard to the censored data (*e.g.* censoring rate higher than 70%).

We had proposed an estimation method based on the restoration of missing data using Bayesian sampling of parameters see Bacha et al. [1998] for the Weibull CR model and Ducros and Pamphile [2018] for a mixture of Weibull distributions. The mean of the posterior distribution can thus be obtained by an importance sampling technique. We propose an improvement of our method and we apply it to the inference of the parameters of a Weibull competing risk model in heavily censored settings with masked causes.

The proposed approach is not an iterative method and therefore significantly reduces computation times comparing to EM or Gibbs algorithms. Simulations have shown that it is effective both in terms of relative bias and relative root mean square error.

The article is organized as follows. In Section 2, we present the Weibull competing risk model with masked failure causes. In Section 3, its maximum likelihood estimation through the EM algorithm and stochastic variants are presented. Section 4 is devoted to the presentation of Bayesian inference. We present there two the two estimation algorithms of mean of the posterior distribution, using a Bayesian restoration of missing data. In Section 6, 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 risk model

Competing risk (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, \dots, T_n be the lifetime of these systems. Assume that

$$T_i = \min_{k=1,2} T_{ik} \quad (1)$$

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)

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

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 $t \geq 0$

$$R(t | \beta_k, \eta_k) = e^{-\left(\frac{t}{\eta_k} \right)^{\beta_k}} \quad (3)$$

$$\lambda(t | \beta_k, \eta_k) = \frac{f(t | \beta_k, \eta_k)}{R(t | \beta_k, \eta_k)} = \frac{\beta_k}{\eta_k} \left(\frac{t}{\eta_k} \right)^{\beta_k - 1}. \quad (4)$$

2.1 The hidden data structure

Masked competing risk model is an hidden 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: $(\mathbf{t}, \mathbf{e}) = ((t_1, e_1), \dots, (t_n, e_n))$ 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 $\mathbf{o} = (o_1, \dots, o_n)$, 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 Dempster et al. [1977] or data augmentation methods Tanner and Wong [2010] incorporate the missing data into the likelihood to improve the bias of the maximum likelihood estimator.

For masked Weibull competing risk, it is possible to consider the **full missing** data structure with the label and failure times of each cause, $\mathbf{z} = (o_i : e_i = 1; t_{ik}, k = 1, 2, i = 1, \dots, n)$; or the **simple missing** data structure with only the labels, $\mathbf{z} = (o_i : e_i = 1, i = 1, \dots, n)$. 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 $f(t | \boldsymbol{\theta}) = f(t | \boldsymbol{\theta}')$ 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, Liang and Sun [2016] 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.* Jiang and Murthy [2003]: 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 Jiang and Murthy [2003]). 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.

3 Maximum likelihood estimation

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

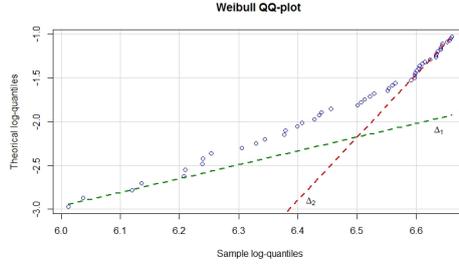


Figure 1: Weibull QQ-plot for competing risk data: the QQ-plot is convex and for small t , $R_{WCR} \simeq R(t; (\beta_1; \eta_1))$ and for large t , $R_{WCR} \simeq R(t; (\beta_2; \eta_2))$. Thus, the left tangent (right resp.) allows to estimate $(\beta_1; \eta_1)$ ($(\beta_2; \eta_2)$ resp.) *cf.* Jiang and Murthy [2003]

of the competing risks model is

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

A direct maximization of likelihood is possible in the case of two competing causes, with a quasi Newton algorithm *cf.* Dennis and Moré [1977]. It will be proposed in the numerical experiment section and we will see that the estimator is strongly biased for heavily censored data.

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 in $\boldsymbol{\theta}$, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ the expectation of the completed likelihood conditionally on the observed data and a current value $\boldsymbol{\theta}'$ of the parameter, *cf.* Dempster et al. [1977]:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathbf{E}_{f(\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^{(b)})} [\ln L(\boldsymbol{\theta} | \mathbf{y}, \mathbf{z})]. \tag{6}$$

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

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

Thus,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{k=1}^2 Q(\theta_k | \boldsymbol{\theta}')$$

where

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

with

$$p(o_i = k | \mathbf{y}; \theta'_k) = \frac{\lambda(t_i | \theta'_k)}{\sum_{\ell=1}^2 \lambda(t_i | \theta'_\ell)}. \tag{8}$$

Calculation of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is the E-step of the EM algorithm and its maximization is the M-step. It leads to the maximization of $Q(\theta_k | \boldsymbol{\theta}^{(b)})$ for each cause $k = 1, 2$ cf. Eq. (7). Let $\theta_k^{(b)}$ being the current value of the Weibull parameters of the k th cause at iteration (b) , $\theta_k^{(b+1)}$ is solution of the equation

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

and

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

Algorithm 1 : EM algorithm

1. **Initialization:** $\boldsymbol{\theta}^{(0)}$

2. **Iterations:**

for $b = 0$ to $B - 1$ do

E-step: compute $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(b)})$ see Eq. (6)

M-step:

$$\boldsymbol{\theta}^{(b+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(b)}) \quad \text{see Eq.(9, 10)}$$

end for

3. **Estimation:**

$$\widehat{\boldsymbol{\theta}}_{EM} = \boldsymbol{\theta}^{(B)}$$

Remarks:

The advantages and drawbacks of EM algorithm are well-documented (see for instance McLachlan and Krishnan [2007]). 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 (SEM) leads to simpler equations in the M-step and it is by far less sensitive than EM to its starting value cf. Celeux et al. [1996]. The principle of the SEM 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.

Algorithm 2: Stochastic EM estimation

1. **Initialization:** $\boldsymbol{\theta}^{(0)}$

2. **Iterations:**

for $b = 0$ to $B - 1$ do

 1. **E-step:** for i such that $e_i = 1$ and $k = 1, 2$ compute $p(o_i = k | \mathbf{y}, \boldsymbol{\theta}^{(b)})$

 2. **Stochastic-step:** $\tilde{\mathbf{z}}^{(b)} \sim f(\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^{(b)})$

 3. **M-step:**

$$\boldsymbol{\theta}^{(b+1)} = \arg \max_{\boldsymbol{\theta}} \widehat{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(b)}) = \arg \max_{\boldsymbol{\theta}} \ln L(\boldsymbol{\theta} | \mathbf{y}, \tilde{\mathbf{z}}^{(b)})$$

end for

3. **Estimation:**

SEM estimator: the sequence $\theta^{(1)}, \dots, \theta^{(B)}$ is a Markov chain and converges to a stationary distribution *cf.* Celeux et al. [1996]. Hence, after a sufficiently long warm-up period of M iterations, the parameters are estimated by the average of the $B - M$ last terms.

$$\hat{\theta}_{SEM} = \frac{1}{B - M} \sum_{b=M+1}^B \theta^{(b)}. \quad (11)$$

SEM-EM estimator: the EM algorithm (*i.e.* algorithm 1) is initialized from SEM iterations after warm-up:

$$\hat{\theta}_{SEM-EM} = \arg \max_{M+1 \leq b \leq B} L(\theta^{(b)} | \mathbf{y})$$

Remarks:

The stochastic EM algorithm generates an ergodic Markov chain. Thus a sequence of parameter estimates via SEM is expected to visit the whole parameter space with long sojourns in the neighborhood of the local maxima of the likelihood function. Therefore, SEM algorithm offers a relevant starting values for EM algorithm, see Biernacki et al. [2003].

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 distribution

$$\left(p(o_i = k | \mathbf{y}, \theta^{(b)}), k = 1, 2 \right) \text{ cf. Eq (8).}$$

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 $\left(p(o_i = k | \mathbf{y}, \theta^{(b)}), k = 1, 2 \right)$:

$$\tilde{o}_i^{(b)} \sim p(o | \mathbf{y}, \theta^{(b)}). \quad (12)$$

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

$$\tilde{t}_{ik}^{(b)} \sim \frac{f(t | \theta_k^{(b)})}{R(t_i | \theta_k^{(b)})}. \quad (13)$$

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

$$\tilde{t}_{ik}^{(b)} \begin{cases} = t_i & \text{if } \tilde{o}_i^{(b)} = k; \\ \sim \frac{f(t | \theta_k^{(b)})}{R(t_i | \theta_k^{(b)})} & \text{if } \tilde{o}_i^{(b)} \neq k. \end{cases} \quad (14)$$

M-step: it consists of computing the ML estimate of each component, $(\theta_k^{(b+1)}, k = 1, 2)$, from the completed sub-samples $(\tilde{t}_{ik}^{(b)}, i = 1, \dots, n; k = 1, 2)$, by a standard procedure, see for instance Murthy et al. [2003].

Simple restoration

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

E-step: for $i = 1, \dots, n$ compute the distribution

$$\left(p(o_i = k \mid \mathbf{y}, \boldsymbol{\theta}^{(b)}), k = 1, 2 \right) \text{ cf. Eq (8);}$$

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 $\left(p(o_i = k \mid \mathbf{y}, \boldsymbol{\theta}^{(b)}), k = 1, 2 \right)$, cf. Eq. (12);

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 $\tilde{o}_i^{(b)} = k$; t_i such that $e_i = 0$) by a standard procedure, see for instance Murthy et al. [2003].

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, $\pi(\boldsymbol{\theta} \mid \mathbf{y})$, is obtained using Bayes formula,

$$\begin{aligned} \pi(\boldsymbol{\theta} \mid \mathbf{y}) &= \frac{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta} \mid \mathbf{y})}{\int \pi(\boldsymbol{\theta})L(\boldsymbol{\theta} \mid \mathbf{y})d\boldsymbol{\theta}} \\ &\propto \pi(\boldsymbol{\theta})L(\boldsymbol{\theta} \mid \mathbf{y}). \end{aligned} \quad (15)$$

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 risk 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 Barringer [2000]

and ? for instance. The same beta distribution, symmetric on range interval $[b_{min}; b_{max}]$, has been chosen as prior for shape parameters (β_k):

$$\pi(\beta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(\beta - b_{min})^{p-1} (b_{max} - \beta)^{q-1}}{(b_{max} - b_{min})^{p+q+1}} \cdot \mathbf{1}_{[0;1]} \left(\frac{\beta - b_{min}}{b_{max} - b_{min}} \right), \quad (16)$$

where

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

Thus

$$\frac{\beta - b_{min}}{b_{max} - b_{min}} \sim \mathcal{B}(p, q). \quad (17)$$

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 Barringer [2000]: $b_{min} = 0.5$ 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):

$$\pi(\eta_k | \beta_k) = \frac{\beta_k a_k^{\beta_k b_k}}{\Gamma(b_k)} \eta_k^{-(\beta_k b_k + 1)} e^{-\left(\frac{a_k}{\eta_k}\right)^{\beta_k}}, \quad (18)$$

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

Note that the GIG distribution is log-concave, see Lemma 1 in Appendix 8. This is a convenient characteristic when using the Gibbs sampler presented hereafter.

4.2 EM algorithm for the Maximum *a posteriori*

The EM algorithm is easily modified to produce the maximum *a posteriori* (MAP) *cf.* McLachlan and Krishnan [2007].

Algorithm 3 : EM-MAP

1. **Initialization:** $\theta^{(0)}$
2. **Iterations:**
for $b = 0$ to $B - 1$ do
 E-step:

$$\begin{aligned} P(\theta, \theta^{(b)}) &= \mathbf{E}_{f(\mathbf{z} | \mathbf{y}, \theta^{(b)})} [\ln \pi(\theta | \mathbf{y}, \mathbf{z})] \\ &= \mathbf{E}_{f(\mathbf{z} | \mathbf{y}, \theta^{(b)})} [\ln L(\theta | \mathbf{y}, \mathbf{z})] + \ln \pi(\theta). \end{aligned} \quad (19)$$

M-step:

$$\theta^{(b+1)} = \arg \max_{\theta} P(\theta, \theta^{(b)})$$

end for

3. **Estimation:**

$$\hat{\theta}_{MAP} = \theta^{(B)}$$

Remarks: The interest of this algorithm comes from the fact that the presence of the log-prior in Eq. (19), acts as a regularization term.

4.3 Gibbs sampling

A standard solution to get an approximation of the posterior distribution through Monte Carlo Markov Chains (MCMC) algorithms is the Gibbs Sampling, see for instance Tanner and Wong [1987] or Casella and George [1992]. The Gibbs algorithm consists of simulating iteratively all unknown quantities (parameters and missing data) according to their conditional distributions. Under mild conditions, and after a large number of iterations I , resulting samples follow the joint posterior distribution $\pi(\theta | \mathbf{y})$ *cf.* ?.

Algorithm 4 : Gibbs sampler

Initialization: $\tilde{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)})$

for = 0 to $I - 1$ **do**

$\tilde{o}_i^{(b+1)} \sim p(o | \mathbf{y}, \tilde{\theta}^{(b)})$, for $i = 1, \dots, n$

$\tilde{\eta}_k^{(b+1)} \sim \pi(\eta | \tilde{\beta}_k^{(b)}, \{t_i : o_i^{(b+1)} = k\})$, for $k = 1, 2$

$\tilde{\beta}_k^{(b+1)} \sim \pi(\beta | \tilde{\eta}_k^{(b+1)}, \{t_i : o_i^{(b+1)} = k\})$, for $k = 1, 2$

end for

Gibbs estimator: $(\tilde{\theta}^{(1)}, \dots, \tilde{\theta}^{(B)})$ is a Markov chain of which distribution converges to the posterior distribution.

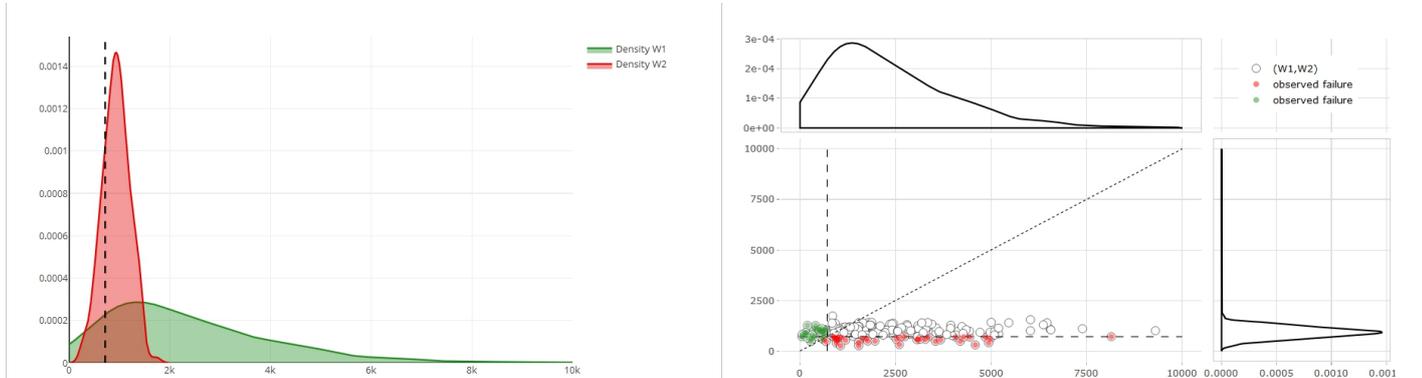
Therefore, after a burning period of M iterations, the Gibbs estimator is computed from the $B - M$ last terms:

$$\hat{\theta}_{GIBBS} = \arg \max_{b \in \{M+1, \dots, B\}} \ln \pi(\tilde{\theta}^{(b)} | \mathbf{y}). \quad (20)$$

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 adaptive rejection method of Gilks and Wild [1992] can be used to simulate the full conditional distribution of $\beta_k, k = 1, 2$. The Gibbs sampler is described more precisely in Appendix A.

The Gibbs sampler produces a Markov chain $(\tilde{\theta}^{(1)}, \dots, \tilde{\theta}^{(B)})$, 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.* Casella and George [1992].



a) Components pdf

$$T_1 \sim \mathcal{W}(\beta_1 = 1.5, \eta_1 = 2500)$$

$$T_2 \sim \mathcal{W}(\beta_2 = 4, \eta_2 = 1000)$$

b) synthetic data : (t_1, t_2)

green: $t_1 < t_2 < \text{censoring}$

red : $t_2 < t_1 < \text{censoring}$

white: $t_1, t_2 > \text{censoring}$

Figure 2: Distributions and synthetic data with $n = 200$, and 70% right censoring

5 Bayesian restoration of missing data

In ill-posed situations, the EM algorithm and the Gibbs sampler are jeopardized and can often perform poorly. The EM algorithm can converge dramatically slowly or converge to insensible local maxima while

the Gibbs sampler can get stuck in local modes. 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 Bacha et al. [1998] for the Weibull CR model and Ducros and Pamphile [2018] for a mixture of Weibull distributions. We sketch the likelihood maximization algorithm with missing data completed by a Bayesian restoration (BR-LM algorithm):

Algorithm 5 : 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 : $\hat{\theta}^{(b)} = \arg \max_{\theta} \ln L(\theta | \mathbf{y}, \tilde{\mathbf{z}}^{(b)})$;
- Step 4 Estimation: the mean of the posterior distribution is then obtained using a importance sampling technique *cf.* ? : 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 $\hat{\rho}$, an estimate of the distribution of $(\hat{\boldsymbol{\theta}}^{(b)}, b = 1, \dots, B)$ is used as proposal distribution:

$$\hat{\rho}(\boldsymbol{\theta}) = \frac{1}{B \cdot h_{\boldsymbol{\theta}}} \sum_{b=1}^B K \left(\frac{\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{(b)}}{h_{\boldsymbol{\theta}}} \right), \tag{21}$$

where K is a Gaussian kernel. Then assign a weight to each $\boldsymbol{\theta}^{(b)}$:

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

Defining the normalized weights

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

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

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

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.* Rubin [1988]. 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 $\hat{\boldsymbol{\theta}}^{(b)}$. The following EM step is added to the BR-LM algorithm

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

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

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

Algorithm 6 : 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 :

$$\boldsymbol{\theta}^{(b)} = \arg \max_{\boldsymbol{\theta}} \mathbf{E}_{f(\mathbf{z}|\mathbf{y},\boldsymbol{\theta}^{(m)})} [\ln \pi(\boldsymbol{\theta} | \mathbf{y}, \mathbf{z})] \quad (22)$$

Details of the maximization can be found in the Appendix B.

Step 3' EM-MAP: Let us denote $\hat{\boldsymbol{\theta}}_{MAP}^{(b)}$ the result of M iteration of the algorithm EM-MAP with $\hat{\boldsymbol{\theta}}^{(b)}$ as initialization.

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

Remarks: The number of iterations of EM algorithm do not need to be very high. We have set it at $M = 50$.

6 Numerical experiments

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

6.1 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: $W_1 \sim \mathcal{W}(\beta_1 = 1.5; \eta_1 = 2500)$ and $W_2 \sim \mathcal{W}(\beta_2 = 4; \eta_2 = 1000)$. The sample size was $n = 200$ and the censorship rates 70% and 20% have been used to better understand the difficulty induced by censorship in the case of competing risk model. 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 :

$$Bias_{rel}(\hat{\theta}) = \frac{Bias(\hat{\theta})}{\theta} \quad RMSE_{rel}(\hat{\theta}) = \frac{RMSE(\hat{\theta})}{\theta}.$$

From these 500 replications, we compare the performance of the estimators provided by (i) the graphical estimation, (ii) the directed maximization of the model likelihood with the quasi-Newton BFGS method, cf. Dennis and Moré [1977], (iii) the EM algorithm initiated with the SEM algorithm with $B = 100$ iterations, (iv) the Gibbs sampling in the above described Bayesian setting with $B = 50000$ iterations, (v) the algorithm BR-LM and (vi) the algorithm BR-PM with $B = 50000$ runs.

All the results for various censoring rate are summarized in Tables 1-12. 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 the ML and Gibbs algorithms, 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. Lastly, it can be seen that the Gibbs sampler converges very slowly.

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.

6.1.1 Comparisons for 70% censorship rate

Table 1: **WQQPlot** Average parameters estimates and MTTFs for both components.

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2 256.86	MTTF2 906.40
Average	2.12	3.81	1 857.52	966.08	1 698.15	871.72
Bias _{rel}	41.36%	-23.74%	-25.70%	-3.39%	-24.76%	-5.06%
RMSE _{rel}	51.61%	23.46%	49.82%	42.46%	54.35%	11.35%

Table 2: **ML** Average parameters estimates and MTTFs for both components. The likelihood is maximized using BFGS quasi-newton method with WQQPlot starting value

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2 256.86	MTTF2 906.40
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%

Table 3: **EM with SEM starting value.** Average parameters estimates and MTTFs for both components.

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2 256.86	MTTF2 906.40
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%

Table 4: **Algorithm 3, Gibbs sampling estimate.** Average parameters estimates and MTTFs for both components.

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2 256.86	MTTF2 906.40
Average	2.23	6.10	2 073.81	1 137.80	1 800.23	1 061.53
Bias _{rel}	48.48%	22.03%	-17.05%	13.78%	-20.25%	14.7%
RMSE _{rel}	73.27%	35.96%	60.77%	34.83%	66.83%	35.80%

Table 5: **Algorithm 4, BR-LM estimation with full restoration.** Average parameters estimates and MTTFs for both components.

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2 256.86	MTTF2 906.40
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%

Table 6: **Algorithm 5, BR-PM estimation with full restoration.** Average parameters estimates and MTTFs for both components.

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2 256.86	MTTF2 906.40
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%

6.1.2 Comparisons for 20% censorship rate

Table 7: **WQQPlot** Average parameters estimates and MTTFs for both components.

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Average	2.09	4.44	2 003.07	942.52	1 881.98	858.58
Bias _{rel}	39.05%	-11.20%	-19.88%	-5.75%	-16.61%	-6.49%
RMSE _{rel}	48.20%	19.79%	72.15%	3.79%	87.99%	3.93%

Table 8: **ML** Average parameters estimates and MTTFs for both components. The likelihood is maximized using BFGS quasi-newton method with WQQPlot starting value

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
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%

Table 9: **EM with SEM starting value.** Average parameters estimates and MTTFs for both components.

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
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%

Table 10: **Algorithm 3, Gibbs sampling.** Average parameters estimates and MTTFs for both components.

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
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%

Table 11: **Algorithm 4, BR-LM estimation with full restoration.** Average parameters estimates and MTTFs for both components.

	β_1 1.5	β_2 5	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
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%

Table 12: **Algorithm 5, BR-PM estimation with full restoration.** Average parameters estimates and MTTFs for both components.

	β_1	β_2	η_1	η_2	MTTF1	MTTF2
	1.5	5	2 500	1 000	2256.86	906.40
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%

6.2 Real data

Data from aircraft windshields maintenance are analyzed (*cf.* Murthy et al. [2003]). 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 13). The Weibull QQ-plot suggests two main competing causes of failures (see Figure 3).

Tables 14-16 provide 95% bootstrap or credible intervals for the SEM-EM algorithm, the Gibbs sampler and the BR-PM algorithm we propose. Simulations have shown us that SEM-EM and Gibbs sampler algorithms provide estimates with high variance. The poor performance of the Gibbs algorithm illustrates the fact that it converges very slowly (*cf.* Casella and George [1992]). By comparing the intervals we have noticed that the algorithm BR-PM provides more accurate intervals, especially for scale parameters.

Table 13: Aircraft windshields: lifetimes ($\times 1000 h$). "-" 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	

Weibull QQlot

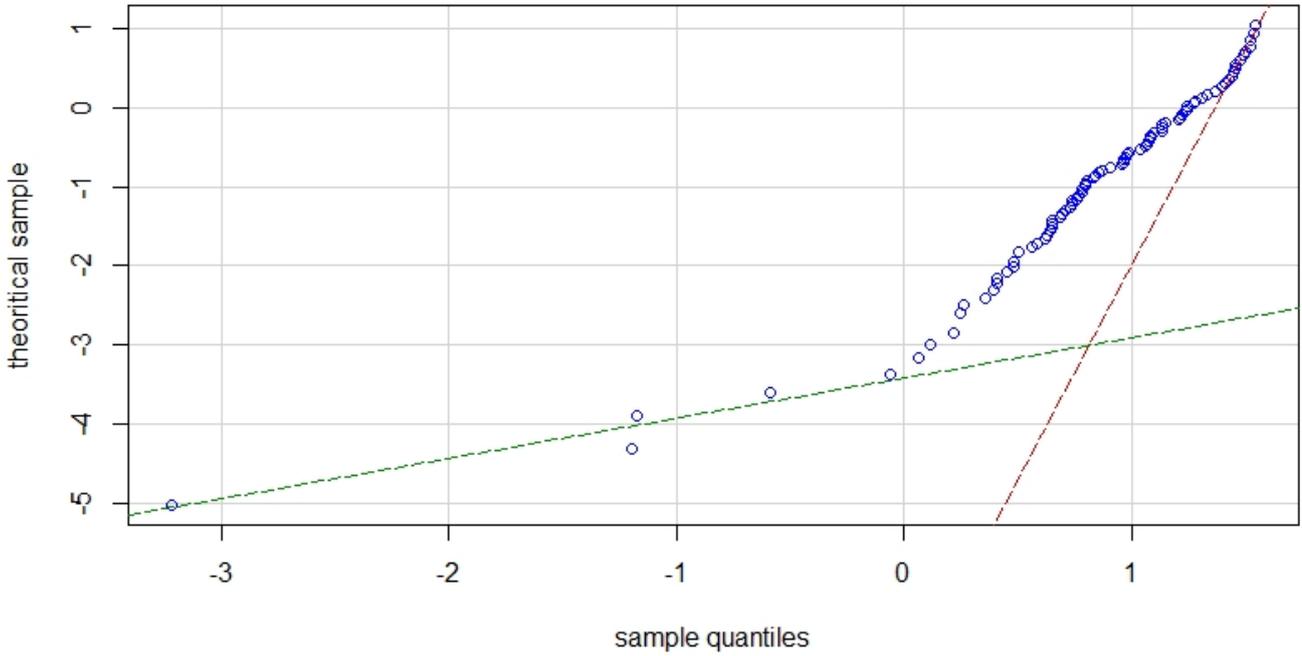


Figure 3: Weibull QQ-plot for aircraft windshield *cf.* Murthy et al. [2003]; there are two main competing causes of failures.

Table 14: Aircraft windshields: $n = 153$ and the censorship rate is 57.5%. Parameters 95% CI's with the Algorithm 2 SEM-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 15: Aircraft windshields: $n = 153$ and the censorship rate is 57.5%. Parameters 95% CI's with the Algorithm 4, Gibbs sampler

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 16: Aircraft windshields: $n = 153$ and the censorship rate is 57.5%. Parameters 95% CI's with the Algorithm 6, 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

7 Discussion

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 SEM 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 important re-sampling. The EM step then acts as a regularization step.

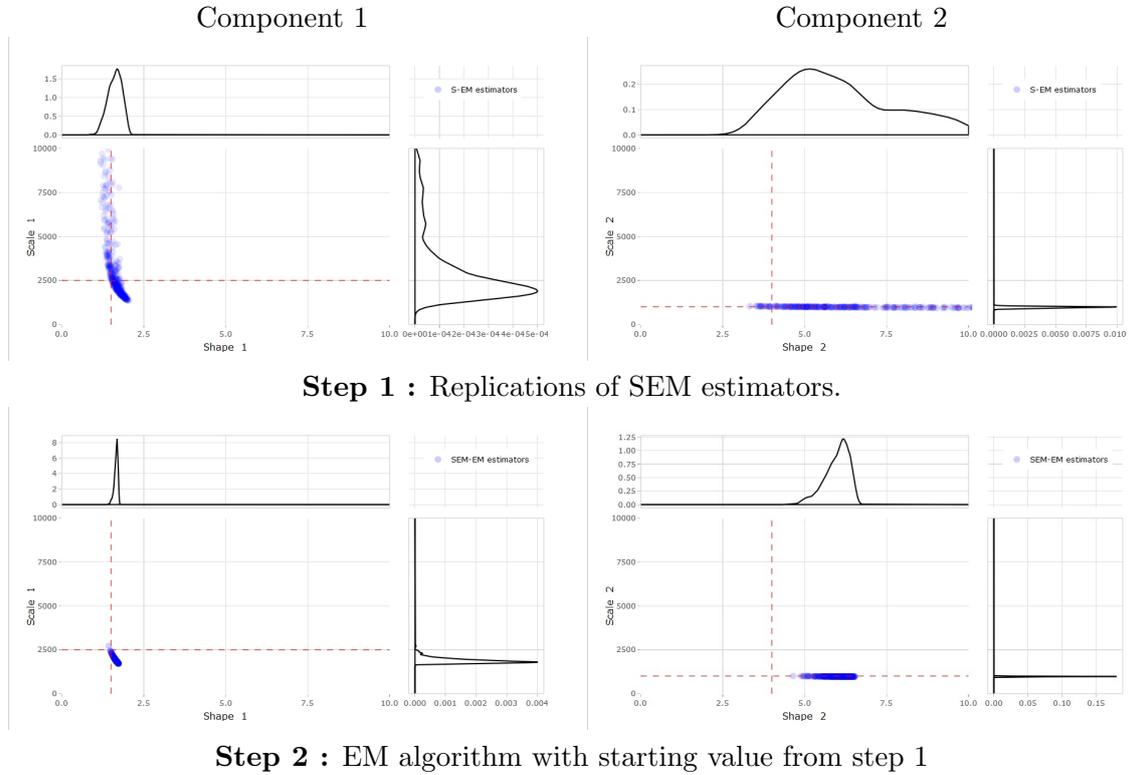


Figure 4: EM with SEM as starting values: the steps of algorithm ??

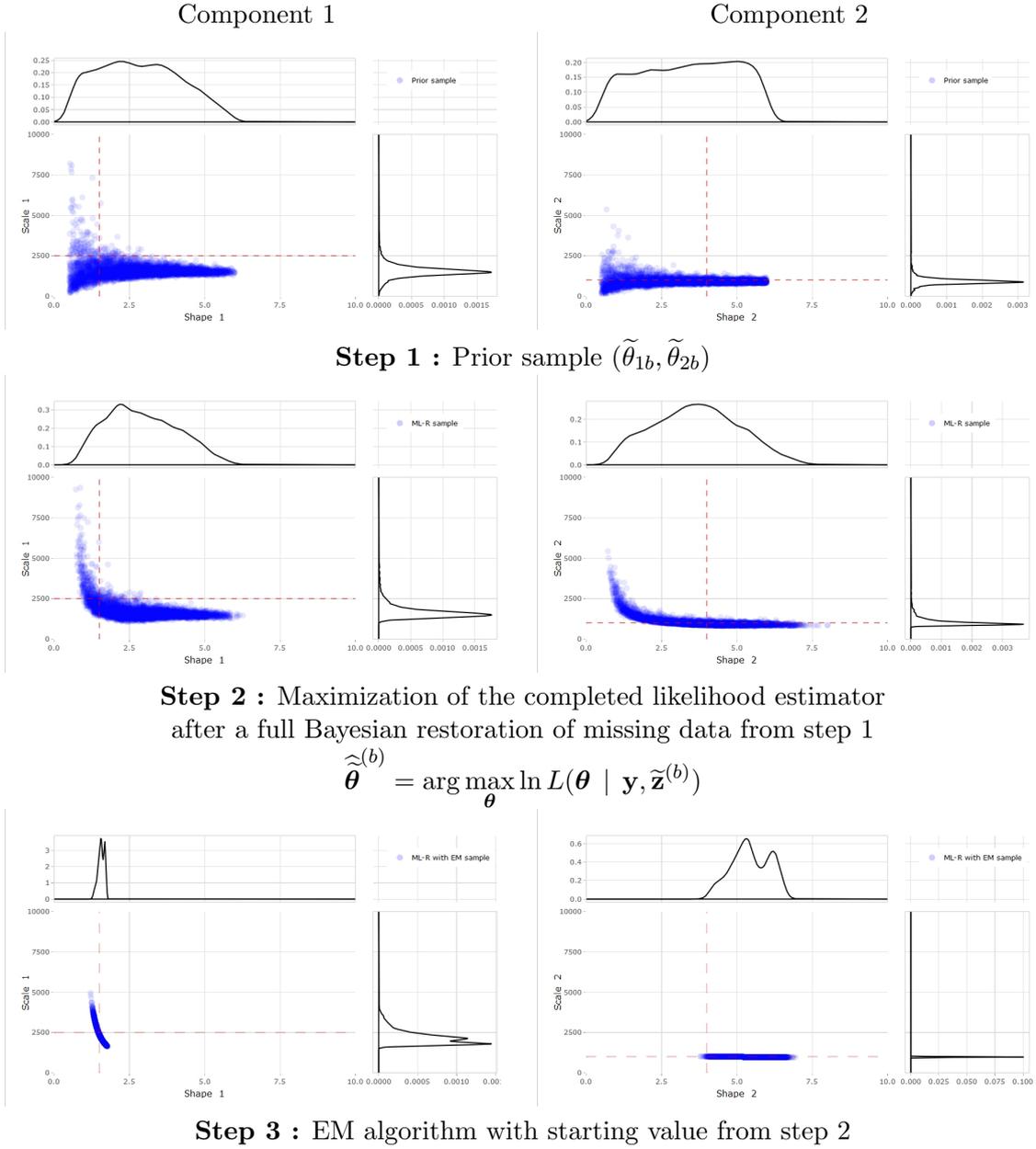


Figure 5: Bayesian restoration: the steps of algorithm ??

Table 17 provides comparisons between Full and Simple restoration for algorithm 6 (*cf.* section 3.3). It can be seen that the Simple restoration version gives poorer results in terms of bias.

Table 17: Comparison Full versus Simple restoration for **Algorithm 6**, BR-PM estimation. 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%

Algorithmic issues

It is worth recalling that the EM, SEM and Gibbs sampler algorithms are iterative algorithms while the Bayesian sampling we use may be parallelizable. Therefore parameters estimation proposed is faster and less time-consuming than standard maximum likelihood estimation or Gibbs sampler.

In addition, for iterative algorithms, especially for SEM 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 Celeux [1998]: algorithm is post-processing by a k-means clustering.

For SEM and the two Bayesian restoration algorithms, 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.

8 Conclusion

The estimation of the parameters of the components of a competing risk model is a ill-defined problem. This problem is all the more difficult with the presence of heavily censored data. We present here a new estimation method for a Weibull competing risks model. The estimation by the maximum likelihood method is then very strongly biased. The EM algorithm and Gibbs sampler have difficulty to converge when they are initialized far from the true values of the parameters. 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 maximum likelihood method, the standard EM algorithm and the Gibbs sampler, for moderate to very heavy censoring rates.

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

$$\begin{aligned} \pi(o_i = k | \mathbf{y}, \mathbf{o}_{(-i)}^{(b)}, \boldsymbol{\beta}^{(b)}, \boldsymbol{\eta}^{(b)}) \\ = \frac{\lambda(t_i | \beta_k^{(b)}, \eta_k^{(b)})}{\lambda(t_i | \beta_1^{(b)}, \eta_1^{(b)}) + \lambda(t_i | \beta_2^{(b)}, \eta_2^{(b)})} \end{aligned} \quad (23)$$

2. Shape

$$\begin{aligned} \pi(\beta_k | \mathbf{y}, \mathbf{o}^{(b)}, \boldsymbol{\beta}_{(-k)}^{(b)}, \boldsymbol{\eta}^{(b)}) \\ \propto \pi(\beta_k) \cdot \pi(\eta_k^{(b)} | \beta_k) \cdot \prod_{i: o_i^{(b)}=k} \lambda(t_i | \beta_k, \eta_k^{(b)}) \\ \cdot \prod_{i=1}^n R(t_i | \beta_k, \eta_k^{(b)}) \\ \propto f(\beta_k | \mathbf{t}, \mathbf{o}^{(b)}, \eta_k^{(b)}) \end{aligned}$$

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

Proof.

$$\begin{aligned} \ln f(\beta_k | \mathbf{t}, \mathbf{o}^{(b)}, \eta_k^{(b)}) \\ = (p-1) \ln(\beta_k - b_{min}) + (q-1) \ln(b_{max} - \beta_k) \\ + \ln(\beta_k) + \beta_k b_k \ln(a_k) - \beta_k b_k \ln(\eta_k^{(b)}) - \left(\frac{a_k}{\eta_k^{(b)}}\right)^{\beta_k} \\ + \#\{o_i^{(b)} = k\} \ln(\beta_k) + \beta_k \sum_{i: o_i=k} \ln\left(\frac{t_i}{\eta_k^{(b)}}\right) \\ - \sum_i^n \left(\frac{t_i}{\eta_k^{(b)}}\right)^{\beta_k} \\ = (p-1) \ln(\beta_k - b_{min}) + (q-1) \ln(b_{max} - \beta_k) \\ + (\#\{o_i^{(b)} = k\} + 1) \ln(\beta_k) \\ + \beta_k \left[b_k \ln\left(\frac{a_k}{\eta_k^{(b)}}\right) + \sum_{i: o_i=k} \ln\left(\frac{t_i}{\eta_k^{(b)}}\right) \right] \\ - \left(\frac{a_k}{\eta_k^{(b)}}\right)^{\beta_k} - \sum_i^n \left(\frac{t_i}{\eta_k^{(b)}}\right)^{\beta_k} \end{aligned} \quad (24)$$

$$\begin{aligned} \frac{\partial \ln f}{\partial \beta_k}(\beta_k | \mathbf{t}, \mathbf{o}^{(b)}, \eta_k^{(b)}) \\ = \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} + \frac{1 + \#\{o_i^{(b)} = k\}}{\beta_k} \\ + b_k \ln\left(\frac{a_k}{\eta_k}\right) + \sum_{i: o_i=k} \ln\left(\frac{t_i}{\eta_k}\right) \\ - \ln\left(\frac{a_k}{\eta_k}\right) \left(\frac{a_k}{\eta_k}\right)^{\beta_k} - \sum_i^n \ln\left(\frac{t_i}{\eta_k}\right) \left(\frac{t_i}{\eta_k}\right)^{\beta_k} \end{aligned} \quad (25)$$

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

□

Hence, the adaptative rejection method proposed by Gilks and Wild [1992] to sample univariate log-concave probability density function can be used here.

3. Scale

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

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. (22):

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

For a simple restoration we get :

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

Thus

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

Therefore

$$\eta_k = \left(\frac{\sum_{i=1}^n t_i^{\beta_k} + a_k^{\beta_k}}{\sum_{i:e_i=1} p(o_i = k | \mathbf{y}; \boldsymbol{\theta}^{(m)}) + b_k + \frac{1}{\beta_k}} \right)^{1/\beta_k}. \quad (27)$$

$$\begin{aligned}
& \frac{\partial \mathbf{E}_{f(\mathbf{o}|\mathbf{y},\boldsymbol{\theta}^{(m)})} [\ln L(\boldsymbol{\theta}|\mathbf{y},\mathbf{o})]}{\partial \beta_k}(\beta_k) + \frac{\partial \ln \pi}{\partial \beta_k}(\beta_k) + \frac{\partial \ln \pi}{\partial \beta_k}(\eta_k|\beta_k) \\
&= \sum_{i:e_i=1} p(o_i = k|\mathbf{y};\boldsymbol{\theta}^{(m)}) \frac{\partial \ln \lambda}{\partial \beta_k}(t_i|\theta_k) \\
&+ \sum_{i=1}^n \frac{\partial \ln R}{\partial \beta_k}(t_i|\theta_k) + \frac{\partial \ln \pi(\beta_k)}{\partial \beta_k} + \frac{\partial \ln \pi}{\partial \beta_k}(\eta_k|\beta_k) \\
&= \sum_{i:e_i=1} p(o_i = k|\mathbf{y};\boldsymbol{\theta}^{(m)}) \left[\frac{1}{\beta_k} + \ln(t_i) - \ln(\eta_k) \right] \\
&- \frac{1}{\eta_k^{\beta_k}} \sum_{i=1}^n [\ln(t_i) - \ln(\eta_k)] t_i^{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\
&+ \frac{1}{\beta_k} + b_k \ln(a_k) - b_k \ln(\eta_k) + \frac{1}{\eta_k^{\beta_k}} [\ln(\eta_k) - \ln(a_k)] a_k^{\beta_k} \\
&= \frac{\sum_{i:e_i=1} p(o_i = k|\mathbf{y};\boldsymbol{\theta}^{(m)}) + 1}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\
&+ \ln(\eta_k) \left[\frac{1}{\eta_k^{\beta_k}} \left(\sum_{i=1}^n t_i^{\beta_k} + a_k^{\beta_k} \right) - b_k - \sum_{i:e_i=1} p(o_i = k|\mathbf{y};\boldsymbol{\theta}^{(m)}) \right] \\
&- \frac{1}{\eta_k^{\beta_k}} \left(\sum_{i=1}^n \ln(t_i) t_i^{\beta_k} + \ln(a_k) a_k^{\beta_k} \right) \\
&+ \sum_{i:e_i=1} p(o_i = k|\mathbf{y};\boldsymbol{\theta}^{(m)}) \ln(t_i) + b_k \ln(a_k) \tag{28}
\end{aligned}$$

Upon substituting Eq. (27) in Eq. (28), it follows

$$\begin{aligned}
& \frac{\partial \mathbf{E}_{f(\mathbf{o}|\mathbf{y},\boldsymbol{\theta}^{(m)})} [\ln L(\boldsymbol{\theta}|\mathbf{y},\mathbf{o})]}{\partial \beta_k}(\beta_k) + \frac{\partial \ln \pi}{\partial \beta_k}(\beta_k) + \frac{\partial \ln \pi}{\partial \beta_k}(\eta_k|\beta_k) = 0 \\
&\Leftrightarrow \frac{\sum_{i:e_i=1} p(o_i = k|\mathbf{y};\boldsymbol{\theta}^{(m)}) + 1}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\
&- \left(\sum_{i:e_i=1} p(o_i = k|\mathbf{y};\boldsymbol{\theta}^{(m)}) + b_k + \frac{1}{\beta_k} \right) \left[\frac{\sum_{i=1}^n \ln(t_i) t_i^{\beta_k} + \ln(a_k) a_k^{\beta_k}}{\sum_{i=1}^n t_i^{\beta_k} + a_k^{\beta_k}} \right] \\
&+ \frac{1}{\beta_k^2} \ln \left(\frac{\sum_{i=1}^n t_i^{\beta_k} + a_k^{\beta_k}}{\sum_{i:e_i=1} p(o_i = k|\mathbf{y};\boldsymbol{\theta}^{(m)}) + b_k + \frac{1}{\beta_k}} \right) \\
&+ \sum_{i:e_i=1} p(o_i = k|\mathbf{y};\boldsymbol{\theta}^{(m)}) \ln(t_i) + b_k \ln(a_k) = 0 \tag{29}
\end{aligned}$$

References

M. Bacha, G. Celeux, E. Idée, A. Lannoy, and D. Vasseur. *Estimation de durées de vie fortement censurées*. Eyrolles, 1998.

H Paul Barringer. Weibull databases, 2000. URL <http://www.barringer1.com/wdbase.htm>.

- J. O. Berger and D. Sun. Bayesian analysis for the poly-Weibull distribution. *Journal of the American Statistical Association*, 88(424):1412–1418, 1993.
- H. Bertholon, N. Bousquet, and G. Celeux. An alternative competing risk model to the Weibull distribution for modelling aging in lifetime data analysis. *Lifetime Data Analysis*, 12:481–504, 2006.
- C. Biernacki, G. Celeux, and 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):561–575, 2003.
- G. Casella and E. I. George. Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.
- G. Celeux. Bayesian inference for mixture: The label switching problem. In *Compstat*, pages 227–232. Springer, 1998.
- G. Celeux, D. Chauveau, and Diebolt J. Stochastic versions of the EM-algorithm: an experimental study in the mixture case. *Journal of Statistical Computation and Simulation*, 55:287–314, 1996.
- M. J. Crowder. *Classical competing risks*. Chapman and Hall/CRC, 2001.
- A. D. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society: Series B*, 39:1–38, 1977.
- J. E. Dennis, Jr and J. J. Moré. Quasi-Newton methods, motivation and theory. *SIAM review*, 19(1):46–89, 1977.
- F. Ducros and P. Pamphile. Bayesian estimation of Weibull mixture in heavily censored data setting. *Reliability Engineering & System Safety*, 180:453–462, 2018.
- J. K. Freels, D. A. Timme, J. J. Pignatiello, R. L. Warr, and R. R Hill. Maximum likelihood estimation for the poly-Weibull distribution. *Quality Engineering*, pages 1–8, 2019.
- W. R. Gilks and P. Wild. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(2):337–348, 1992.
- T. Ishioka and Y. Nonaka. Maximum likelihood estimation of Weibull parameters for two independent competing risk. *IEEE Transactions on Reliability*, 40(1):71–74, 1991.
- R. Jiang and D.N.P. Murthy. Study of n -fold Weibull competing risk model. *Mathematical and computer modelling*, 38(11-13):1259–1273, 2003.
- D. Kundu, D. Mitra, and A. Ganguly. Analysis of left truncated and right censored competing risks data. *Computational Statistics & Data Analysis*, 108:12–26, 2017.
- Y. Liang and D. Sun. Identifiability of masking probabilities in competing risks models with emphasis on Weibull models. *Communications in Statistics-Theory and Methods*, 45(7):2143–2157, 2016.
- G. McLachlan and T. Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.
- Chiranjit Mukhopadhyay and Sanjib Basu. Bayesian analysis of masked series system lifetime data. *Communications in Statistics—Theory and Methods*, 36(2):329–348, 2007.
- D. N. P. Murthy, M. . Xie, and R. Jiang. *Weibull Models*. Wiley, 2003.
- C. Park. Parameter estimation of incomplete data in competing risks using the EM algorithm. *IEEE Transactions on Reliability*, 54:282–290, 2005.
- R. Ranjan and 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):1522–1535, 2016.
- D. B. Rubin. Using the SIR algorithm to simulate posterior distributions. *Bayesian statistics*, 3:395–402, 1988.

- A.M. Sarhan, D.C. Hamilton, and B. Smith. Statistical analysis of competing risks models. *Reliability Engineering & System Safety*, 2010.
- M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540, 1987.
- M. A. Tanner and W. H. Wong. From EM to data augmentation: the emergence of MCMC Bayesian computation in the 1980s. *Statistical science*, 25(4):506–516, 2010.
- L. Wang. Inference for Weibull competing risks data under generalized progressive hybrid censoring. *IEEE Transactions on Reliability*, 67(3):998–1007, 2018.