



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. 2019. hal-02410489v1

HAL Id: hal-02410489

<https://inria.hal.science/hal-02410489v1>

Preprint submitted on 13 Dec 2019 (v1), 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

December 13, 2019

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 Bayesian restoration of missing data through a Bayesian importance sampling of parameters with a weakly informative prior distribution. The mode of the posterior distribution can thus be obtained directly or in an approximate way. The proposed method is not an iterative method and therefore is not costly in terms of computing time. 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

Acronyms & Notation

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
n	number of systems under study
t_{ik}	lifetime of the i -th system related to failure mode k ;
t_i	lifetime of the i -th system without regard to failure mode; $\mathbf{t} = (t_1, \dots, t_n)$
\tilde{t}_i	simulated lifetime
e_i	censoring indicator; $\mathbf{e} = (e_1, \dots, e_n)$ $e_i = 1$, if t_i is lifetime ; $e_i = 0$ if t_i is right censored.
y_i	observed data: $y_i = (t_i, e_i)$
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 the each cause $z_i = (o_i; (t_{ik})$;
\tilde{z}_i	restored missing data
f_W, R_W	probability density function and reliability of a Weibull distribution
f_{CR}, R_{CR}	probability density function and reliability of the CR
β_k, η_k	shape and scale parameter of the k -th cause of failure
$\boldsymbol{\theta} = (\beta_1, \eta_1, \beta_2, \eta_2)$	the set of CR model parameters
$\tilde{\boldsymbol{\theta}}$	simulated parameters
$\hat{\boldsymbol{\theta}}$	parameters estimate
$L(\boldsymbol{\theta} \mathbf{y})$	likelihood of observed data
$L(\boldsymbol{\theta} \mathbf{y}, \mathbf{z})$	the likelihood of observed data completed by missing data
$\pi(\boldsymbol{\theta})$	prior distribution

1 Introduction

Weibull distributions are extensively used for modeling life data in medical or industrial applications. As a matter of fact, 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, a serial type system 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. Indeed, 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 causes of failure (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. And, 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 and 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 masked competing risk models. 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%).

The present article aims to improve statistical inference for Weibull competing risk model in heavily censored settings with masked causes. In order to dramatically reduce the variability of the maximum likelihood estimator, we opt for a Bayesian restoration of missing data through a Bayesian importance sampling of parameters with a weakly informative prior distribution. The mode of the posterior distribution can thus be obtained directly or in an approximate way.

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 masked Weibull competing risk model. 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 non iterative importance sampling procedures to estimate the mode of the posterior distribution. In Section 6, the different procedures are compared using simulated and a reliability data set. These numerical experiments highlight the great interest of initializing the EM algorithm with our estimate of the posterior mode. A discussion section and a conclusion end this paper.

2 Weibull Competing risk models

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. In practice, risk analysis is generally jeopardized by masking: the exact cause of system failure is unknown. Indeed, exact diagnosis of the failure cause could be impossible or too resource-consuming to be performed on every failed system.

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. It is also realistic because

estimating more than two masked competing failure rates from right censored data is an inextricable problem if the data size is not huge. The following assumptions are made throughout this paper

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

where T_{ik} are independent latent failure times for corresponding to two competing causes of failure, $k = 1, 2$; T_{ik} has a 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 and failure rate λ 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 $t_{ik}, k = 1, 2$ cf. Eq. (1).

The EM algorithm [Dempster et al., 1977] or data augmentation methods [Tanner and Wong, 2010] take profit of the missing data to derive estimates of the model parameters.

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 algorithms 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 analyzed with simple graphics 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 a straight line shape; 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.

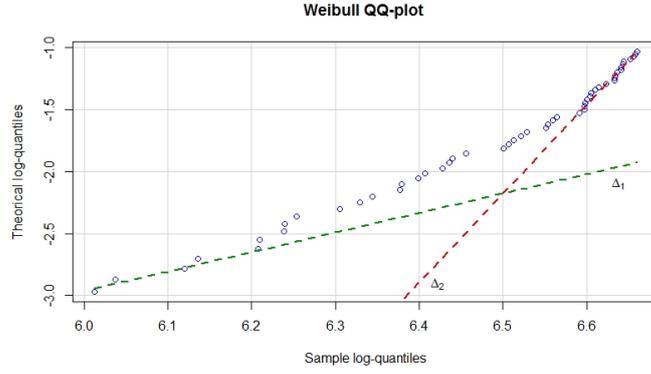


Figure 1: Weibull QQ-plot for competing risk data: the QQ-plot is convex and for small t , $R_{WCR} \simeq R(t; (\beta_1; \eta_1))$ and for large t , $R_{WCR} \simeq R(t; (\beta_2; \eta_2))$. Thus, the left tangent (right resp.) allows to estimate $(\beta_1; \eta_1)$ ($(\beta_2; \eta_2)$ resp.) *cf.*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

3 Maximum likelihood estimation

Let us denote $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, with $\boldsymbol{\theta}_1 = (\beta_1, \eta_1)$, $\boldsymbol{\theta}_2 = (\beta_2, \eta_2)$, 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

$$\begin{aligned} L(\boldsymbol{\theta} | \mathbf{y}) &= \left[\prod_{i: e_i=1} f(t_i | \boldsymbol{\theta}_1) R(t_i | \boldsymbol{\theta}_2) + f(t_i | \boldsymbol{\theta}_2) R(t_i | \boldsymbol{\theta}_1) \right] \prod_{i: e_i=0} R(t_i | \boldsymbol{\theta}_1) R(t_i | \boldsymbol{\theta}_2) \\ &= \left[\prod_{i: e_i=1} \lambda(t_i | \boldsymbol{\theta}_1) + \lambda(t_i | \boldsymbol{\theta}_2) \right] \prod_{i=1}^n R(t_i | \boldsymbol{\theta}_1) R(t_i | \boldsymbol{\theta}_2). \end{aligned} \quad (5)$$

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]:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') = \int_{\mathbf{z}} \ln(L(\boldsymbol{\theta} | \mathbf{y}, \mathbf{z})) f(\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}') d\mathbf{z}. \quad (6)$$

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

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

with

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

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 $Q(\theta_k | \boldsymbol{\theta}^{(b)})$ cf. Eq. (7) for each cause $k = 1, 2$. 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

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

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: $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(b)}) = \mathbb{E}_{f(\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^{(b)})} [\log L(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta})]$

M-step:

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

3. **Estimation:**

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

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 slowly 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 current conditional distribution knowing the observed data and the current value of the parameters.

Algorithm 2: Stochastic EM estimation

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

2. **Iterations:**

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

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

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

3. **M-step:**

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

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.

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

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

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

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 sensible maxima of 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 of restoration are possible according that the simple or the full missing data structure 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 $\left(p(o_i = k | \mathbf{y}, \theta^{(b)}), k = 1, 2 \right)$, *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 $\left(p(o_i = k | \mathbf{y}, \theta^{(b)}), k = 1, 2 \right)$:

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

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 :

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

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 :

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

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 $\left(p(o_i = k | \mathbf{y}, \boldsymbol{\theta}^{(b)}), k = 1, 2\right)$, *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 $\left(p(o_i = k | \mathbf{y}, \boldsymbol{\theta}^{(b)}), k = 1, 2\right)$, *cf.* Eq. (3.3);

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

Bayesian estimation is an alternative to the ML estimation which could be relevant when the information provide by the observed data is poor. It consists of incorporating prior information on the parameter $\boldsymbol{\theta}$ in a prior probability distribution $\pi(\boldsymbol{\theta})$ and to design the statistical inference from the posterior distribution of the parameter knowing the data. This posterior distribution, $\pi(\boldsymbol{\theta} | \mathbf{y})$, is obtained using Bayes formula,

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \frac{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta} | \mathbf{y})}{\int \pi(\boldsymbol{\theta})L(\boldsymbol{\theta} | \mathbf{y})d\boldsymbol{\theta}}. \quad (12)$$

This distribution combines the prior distribution for the parameter $\pi(\boldsymbol{\theta})$ 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 $\pi(\boldsymbol{\theta})$. 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 regularized Bayesian estimators are proper to produce a regularized estimation of the model parameters protected from spurious and insensible local maxima of the EM algorithm. In this perspective, we favor the posterior mode which maximizes the posterior distribution Eq. (12). An other advantage of the posterior mode is that it appears to be easier to compute than the posterior mean which requires a good knowledge of the posterior distribution. Actually, in the present context, computing a good approximation of the posterior distribution $\pi(\boldsymbol{\theta} | \mathbf{y})$ appears to be quite difficult, but estimating the posterior mode is an easier task.

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 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. Barringer [2000] deal with the range of β in various applications. Therefore a beta prior distribution, symmetric on range interval $[b^{min}; b^{max}]$, has been chosen for shape parameters (β_1, β_2) :

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

where

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

and thus

$$\frac{\beta - b_{min}}{b_{max} - b_{min}} \sim \mathcal{B}(a_{shape}, b_{shape}). \quad (13)$$

Hyperparameters (b^{min}, b^{max}) reflect engineer expertise on the parameters (β_k) . They shall be fixed in accordance to the field failure data. According to Barringer [2000], the hyperparameters b^{min} and b^{max} were chosen as follows: $b^{min} = 0.5$ and $b^{max} = 10$. Hyperparameters (a^{shape}, b^{shape}) reflect uncertainty about (β_k) . A vague informative prior was chosen by setting $a^{shape} = b^{shape} = 1.1$.

Scale parameter

For scale parameters (η_1, η_2) , the prior density of η given β is a Generalized Inverse Gamma distribution (GIG):

$$\pi(\eta | \beta) = \frac{\beta a^{\beta b}}{\Gamma(b)} \eta^{-(\beta b + 1)} e^{-\left(\frac{a}{\eta}\right)^\beta}, \quad (14)$$

where (a, b) are hyperparameters. Thus η^β has an inverse gamma distribution.

Notice that The GIG distribution is log-concave, that is a convenient characteristic when using the Gibbs sampler presented hereafter.

As in Ducros and Pamphile [2018], the hyperparameters (a_{scale}, b_{scale}) are obtained by equating the expected value of the inverted gamma distribution to graphical estimation from the Weibull QQ-plot (see Section 2.3). The hyperparameter b^{shape} was chosen large enough to specify a vague informative prior: $b^{shape} = 5$.

Before presenting, the Bayesian restoration of missing data that we advocate, we now sketch the Gibbs Sampling.

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 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 full conditional distributions. Under mild conditions, and after a large number of iterations, resulting samples follow the joint posterior distribution $\pi(\theta | \mathbf{y})$ cf. Roberts and Smith [1994].

Algorithm 3: Gibbs sampler

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

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

$$\left[\begin{array}{l} \tilde{o}_i^{(b+1)} \sim p(o | \mathbf{y}, \tilde{\boldsymbol{\theta}}^{(b)}), \text{ for } i = 1, \dots, n \\ \tilde{\beta}_k^{(b+1)} \sim \pi(\beta | \tilde{\eta}_k^{(b)}, \{t_i : o_i^{(b+1)} = k\}, \mathbf{e}), \text{ for } k = 1, 2 \\ \tilde{\theta}_k^{(b+1)} \sim \pi(\theta | \tilde{\beta}_k^{(b+1)}, \{t_i : o_i^{(b+1)} = k\}, \mathbf{e}), \text{ for } k = 1, 2 \end{array} \right.$$

Gibbs estimator: $(\tilde{\boldsymbol{\theta}}^{(1)}, \dots, \tilde{\boldsymbol{\theta}}^{(B)})$ 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:

$$\hat{\boldsymbol{\theta}}(GIBBS) = \arg \max_{b \in \{M+1, \dots, B\}} \log \pi(\tilde{\boldsymbol{\theta}}^{(b)} | \mathbf{y}). \quad (15)$$

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 adaptative 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 $(\tilde{\boldsymbol{\theta}}^{(1)}, \dots, \tilde{\boldsymbol{\theta}}^{(B)})$, 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.* Casella and George [1992].

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. The non iterative algorithms we now propose are aiming to provide good estimates of the posterior mode without encountering these difficulties. These algorithms take profit of the missing data structure of the Weibull masked competing risk model. They are based on a repeated Bayesian restoration of the missing data in a importance sampling scheme already used in Bacha et al. [1998] for the Weibull CR model and in Ducros and Pamphile [2018] for a mixture of Weibull distributions. This Bayesian sampling leads to repeated approximations of the posterior mode of the model parameters knowing the completed data.

In its principle, the resulting importance sampling scheme is as follows:

Repeat a large number of times B

- Simulate $\theta_1^{(b)}$ and $\theta_2^{(b)}$ according to their prior distribution.
- Restore the full missing data (cause of failure and all the censored failure times) according to their conditional distribution knowing the observed data and the simulated $\theta_1^{(b)}$ and $\theta_2^{(b)}$.
- Derive the posterior modes $\tilde{\theta}_1^{(b)}$ and $\tilde{\theta}_2^{(b)}$ knowing the completed data.

Use the B posterior modes of the previous steps to derive an approximation of the posterior mode of the parameters knowing the observed data.

Hereafter, we detail two algorithms using Bayesian sampling which simply differ in the way to derive an approximation of the posterior mode.

The restoration of missing data makes easier the maximization of the posterior distribution. Values thus obtained constitute a grid on which the observed posterior distribution is calculated and then can be straightforwardly maximized.

Denoting

$$P(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathbb{E}_{f(\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}')} [\ln \pi(\boldsymbol{\theta} | \mathbf{y}, \mathbf{z})] = \int_{\mathbf{z}} \log(\pi(\boldsymbol{\theta} | \mathbf{y}, \mathbf{z})) \cdot f(\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}') d\mathbf{z} = Q(\boldsymbol{\theta}, \boldsymbol{\theta}') + \ln \pi(\boldsymbol{\theta}), \quad (16)$$

and, if $\tilde{\boldsymbol{\theta}} \sim \pi(\boldsymbol{\theta})$ and $\tilde{\mathbf{z}} \sim f(\mathbf{z} | \mathbf{y}, \tilde{\boldsymbol{\theta}})$,

$$\hat{P}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) = \ln L(\boldsymbol{\theta} | \mathbf{y}, \tilde{\mathbf{z}}) + \log \pi(\boldsymbol{\theta}),$$

more formally we propose the two following algorithms.

Algorithm 4: Bayesian restoration and completed log-posterior maximization (BR-PM)

for $b \in \{1, \dots, B\}$, repeat

1. Bayesian sampling: $\boldsymbol{\theta}^{(b)} = (\theta_1^{(b)}, \theta_2^{(b)}) \sim \pi(\theta_1, \theta_2)$

2. Full restoration: for $i = 1, \dots, n$,

labels $o_i^{(b)} \sim p(o | \mathbf{y}, \boldsymbol{\theta}^{(b)})$;

observed failure times: $e_i = 1$, for $k = 1, 2$,

$$t_{ki}^{(b)} \begin{cases} = t_i & \text{if } o_i^{(b)} = k; \\ \sim \frac{f(t | \theta_{bk})}{R(t_i | \theta_k^{(b)})} & \text{else;} \end{cases}$$

censored failure times: $e_i = 0$, for $k = 1, 2$,

$$t_{ki}^{(b)} \sim \frac{f(t | \theta_k^{(b)})}{R(t_i | \theta_k^{(b)})};$$

3. maximization: based on Eq. (16) the completed log posterior is maximized,

$$\tilde{\theta}_k^{(b)} = \arg \max_{\theta} \log \pi(\theta | (t_{ki}^{(b)}, i = 1, \dots, n)). \quad (17)$$

Let us denote $((\tilde{\theta}_1^{(b)}, \tilde{\theta}_2^{(b)}), b = 1, \dots, B)$ the new sample thus obtained.

4. Estimation: the mode of the posterior distribution is estimated by

$$\hat{\theta}_{BR-PM} = \arg \max_{b \in \{1, \dots, B\}} \log \pi(\tilde{\theta}_1^{(b)}, \tilde{\theta}_2^{(b)} | \mathbf{y}). \quad (18)$$

The solutions of Eq. (17) require numerical calculation detailed in Appendix B. They are used as a grid to maximize the posterior distribution. A simple alternative to obtain the grid is to maximize the completed log-likelihood and then to use a few runs of EM in order to correct the approximation thus made. The target value is the mode of the posterior distribution knowing the observed data.

Algorithm 5: Bayesian restoration and completed log-likelihood maximization (BR-LM)

For $b \in \{1, \dots, B\}$ repeat,

1. Bayesian sampling: $\boldsymbol{\theta}^{(b)} = (\theta_1^{(b)}, \theta_2^{(b)}) \sim \pi(\theta_1, \theta_2)$

2. Full restoration: for $i = 1, \dots, n$,

cause of failure: $o_i^{(b)} \sim p(o | \mathbf{y}, \boldsymbol{\theta}^{(b)})$;

observed failure times: for $k = 1, 2$,

$$t_{ki}^{(b)} \begin{cases} = t_i & \text{if } o_i^{(b)} = k; \\ \sim \frac{f(t | \theta_k^{(b)})}{R(t_i | \theta_k^{(b)})} & \text{else;} \end{cases}$$

censored failure times: for $k = 1, 2$,

$$t_{ki}^{(b)} \sim \frac{f(t | \theta_k^{(b)})}{R(t_i | \theta_k^{(b)})};$$

3. maximisation: based on Eq. (16) the completed log posterior is maximized,

$$\tilde{\theta}_k^{(b)} = \arg \max_{\theta} \ln L(\theta | (t_{ki}^{(b)}, i = 1, \dots, n)) \quad (19)$$

Equation Eq. (19) consists of computing the ML estimates of each component, from the completed sub-samples by a standard procedure.

3. Estimation: The sample $(\tilde{\boldsymbol{\theta}}^{(b)})$ is then used as starting value for EM algorithm: let $(\tilde{\theta}_k^{(b)}(EM), k = 1, 2)$ denote the resulting EM estimates. Therefore, a bootstrap confidence interval of parameters can be obtained. The mode of the posterior distribution can also be estimated by

$$\hat{\theta}_{BR-LM-EM} = \arg \max_{b \in \{1, \dots, B\}} \log \pi(\tilde{\theta}_1^{(b)}(EM), \tilde{\theta}_2^{(b)}(EM) | \mathbf{y}). \quad (20)$$

Remark: The two algorithms presented above can be regarded as approximations of the ideal Bayesian sampling proposed in Smith and Gelfand [1992] to obtain a sample of the posterior distribution. It can be sketched as follows.

Algorithm 6: Bayesian sampling, importance re-sampling (SIR)

Let us denote

$$M = \max_{\boldsymbol{\theta}} L(\mathbf{y} | \boldsymbol{\theta}).$$

B runs are simulated as follows : for $b \in \{1, \dots, B\}$,

1. Bayesian sampling: $\tilde{\boldsymbol{\theta}}_b = (\tilde{\theta}_{b1}, \tilde{\theta}_{b2}) \sim \pi(\theta_1, \theta_2)$

2. Importance resampling: $r \sim \mathcal{U}(0; 1)$

if $r < \frac{1}{M}$ then $\tilde{\boldsymbol{\theta}}_b$ is accepted;
else $\tilde{\boldsymbol{\theta}}_b$ is rejected.

The sub-sample $\tilde{\boldsymbol{\theta}}$ thus obtained is distributed in accordance with the posterior distribution

3. Estimation: the following estimator is then constructed,

$$\hat{\boldsymbol{\theta}}_{BSIR} = \arg \max_{1 \leq b \leq B'} \pi(\tilde{\boldsymbol{\theta}}_b | \mathbf{y}). \quad (21)$$

The algorithm 6 can be regarded as a "weighted bootstrap": a sample from the posterior distribution is obtained by combining a Bayesian sampling and an importance re-sampling. Unfortunately, with masked

Weibull competing risk models the maximum of the likelihood, M , is not available and the Algorithm 6 is not applicable contrary to Algorithms 4 and 5.

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

The competing risk model, *cf.* Eq. (1), has been simulated with the component distributions $W_1 \sim \mathcal{W}(\beta_1 = 1.5; \eta_1 = 2500)$ and $W_2 \sim \mathcal{W}(\beta_2 = 4; \eta_2 = 1000)$. The sample size was $n = 200$ and the right censorship rates 70%, 50% and 10% have been used to better understand the difficulty of estimations for a competing risk model. The performance of the methods are evaluated in terms of bias and root mean square error (RMSE) from 500 replications. 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.

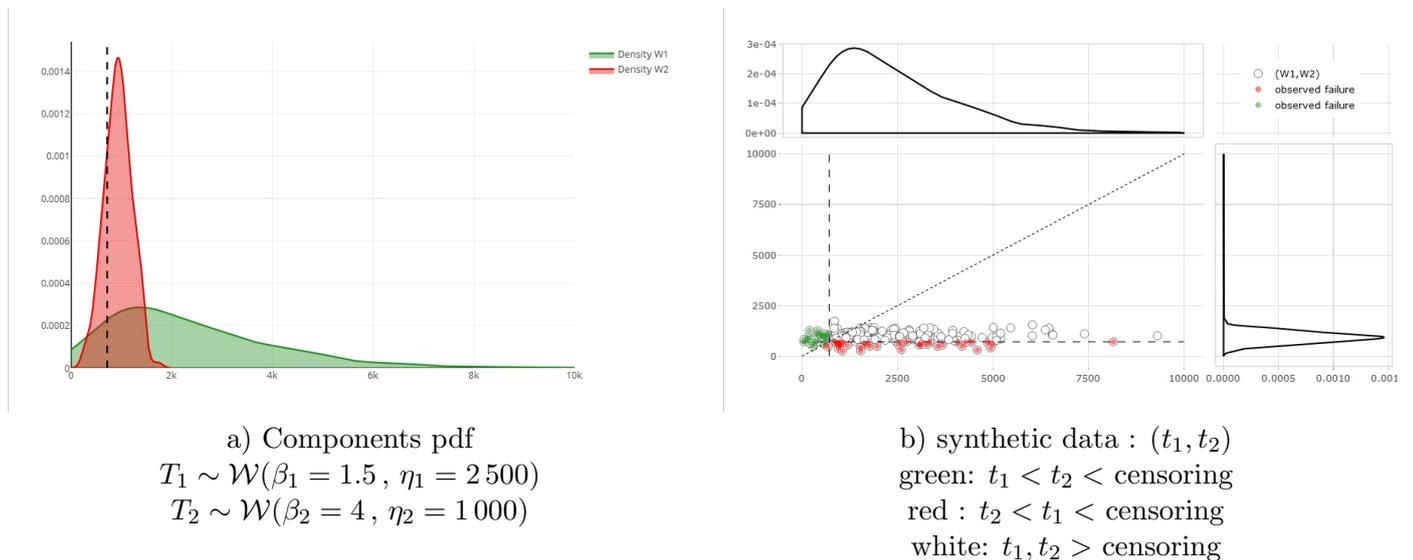


Figure 2: Distributions and synthetic data with $n = 200$, and 70% right censoring

For these data sets, we compared the estimated provided by (i) the directed maximization of the model likelihood with the quasi-Newton BFGS method, *cf.* Dennis and Moré [1977], (ii) the EM algorithm initiated with the SEM algorithm, (iii) the Gibbs sampling in the above described Bayesian setting, (iv) the algorithm BR-PM and (v) the algorithm BR-LM presented in Section 5.

For an easy comparison, ratio bias and RMSE are calculated:

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

All the results for various censoring rate are summarized in Tables 1-15. The shaded cells correspond to the optimal estimate either in terms of relative bias or relative standard error.

For heavily censored data (*i.e.* 70% right censoring), Tables 1-5 show that the BR-LM-EM method (*i.e.* algorithm 6) is optimal both in terms of bias and variance. As mentioned previously (see Figure 2), the high rate of censorship makes η_1 estimation much more difficult than η_2 estimation. On the contrary, it can be noted that it is much more difficult to estimate β_2 than β_1 . Indeed, the initialization of β_1 is based on the large lifetime values, which are, due to censorship, highly biased. Lastly, it can be seen that the GIBBS sampler converges very slowly.

The BR-LM-EM method remains efficient even at lower censorship rates of 50% and 10% (see Tables 6-10 and Tables 11-15). 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.

6.1.1 Comparisons for 70% censorship rate

Table 1: Average ML estimates and Mean Time to Failure (MTTF) for the two components with 70% right censoring data. The likelihood (5) is maximized using BFGS method.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
ML	2.06	6.82	1 587.08	915.73	1 405.89	855.48
Bias _{rel}	37.46%	70.62%	-36.52%	-8.43%	-37.70%	-5.61%

Table 2: Average ML estimates and MTTF for the two components with 70% right censoring data. The likelihood is maximized by EM with SEM starting value and 400 iterations.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.65	5.98	1888.59	974.10	1689.68	903.36
Bias _{rel}	10.17%	49.49%	-24.46%	-2.59%	-25.13%	-0.34%
RMSE _{rel}	3.89%	9.66%	7.02%	0.61%	7.27%	0.24%

Table 3: Average ML estimates and MTTF for the two components with 70% right censoring data. The estimates are computed with **Algorithm 3**, Gibbs sampling estimates with 5 000 iterations and 4 700 discarded "burn-in" iterations.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	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%

Table 4: Average ML estimates and MTTF for the two components with 70% right censoring data. The estimates are computed with **Algorithm 4**, BR-PM estimation with full restoration and 5 000 Bayesian sample size.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.92	5.49	1803.33	889.46	1689.88	821.04
Bias _{rel}	28.14%	37.18%	-27.87%	-11.05%	-25.12%	-9.42%
RMSE _{rel}	37.42%	0.70%	38.31%	0.14%	61.54%	0.15%

Table 5: Average ML estimates and MTTF for the two components with 70% right censoring data. The estimates are computed with **Algorithm 5**, BR-LM estimation with full restoration and 5 000 Bayesian sample size.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.40	4.60	3066.35	999.05	2793.67	912.84
Bias _{rel}	-6.43%	14.92%	22.65%	-0.09%	23.79%	0.71%
RMSE _{rel}	0.28%	0.46%	1.38%	0.04%	1.45%	0.02%

6.1.2 Comparisons for 50% censorship rate

Table 6: Average ML estimates and MTTF for the two components with 50% right censoring data. The likelihood (5) is maximized using BFGS method.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
ML	2.27	4.48	1 921.76	964.77	1 702.31	880.22
Bias _{rel}	51.50%	12.05%	-23.13%	-3.52%	-24.57%	-2.89%

Table 7: Average ML estimates and MTTF for the two components with 50% right censoring data. The likelihood is maximized by EM with SEM starting value and 400 iterations.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.402
Mean	1.63	5.15	2 036.79	1 038.69	1 828.99	955.00
Bias _{rel}	8.35%	28.71%	-18.53%	3.87%	-18.96%	5.36%
RMSE _{rel}	9.06%	14.00%	14.69%	0.74%	15.20%	1.37%

Table 8: Average ML estimates and MTTF for the two components with 50% right censoring data. The estimates are computed with **Algorithm 3**, Gibbs sampling estimates with 5 000 iterations and 4 700 discarded "burn-in" iterations.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.86	5.41	1 844.90	946.43	1 638.55	872.87
Bias _{rel}	24.19%	35.19%	-26.20%	-5.36%	-27.40%	-3.70%
RMSE _{rel}	4.18%	3.97%	12.18%	0.69%	11.99%	0.74%

Table 9: Average ML estimates and MTTF for the two components with 50% right censoring data. The estimates are computed with **Algorithm 4**, BR-PM estimation with full restoration and 5 000 Bayesian sample size.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.75	5.49	2 119.40	937.18	1 988.28	865.09
Bias _{rel}	16.55%	37.17%	-15.22%	-6.28%	-11.90%	-4.56%
RMSE _{rel}	30.67%	0.75%	39.28%	0.15%	55.11%	0.16%

Table 10: Average ML estimates and MTTF for the two components with 50% right censoring data. The estimates are computed with **Algorithm 5**, BR-LM estimation with full restoration and 5 000 Bayesian sample size.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.51	4.71	2 323.40	1 032.34	2 095.28	944.61
Bias _{rel}	0.87%	17.80%	-7.06%	3.23%	-7.16%	4.22%
RMSE _{rel}	0.785%	0.675%	4.725%	0.065%	5.135%	0.105%

6.1.3 Comparisons for 10% censorship rate

Table 11: Average ML estimates and MTTF for the two components with 10% right censoring data. The likelihood (5) is maximized using BFGS method.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
ML	2.10	3.59	1 598.79	963	1 416.01	867.75
Bias _{rel}	40.41%	-10.13%	-36.05%	-3.69%	-37.26%	-4.26%

Table 12: Average ML estimates and MTTF for the two components with 10% right censoring data. The likelihood is maximized by EM with SEM starting value and 400 iterations.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.72	5.51	1798.75	1043.55	1607.05	963.19
Bias _{rel}	14.34%	37.83%	-28.05%	4.36%	-28.79%	6.27%
RMSE _{rel}	8.58%	15.23%	10.67%	0.64%	10.93%	1.28%

Table 13: Average ML estimates and MTTF for the two components with 10% right censoring data. The estimates are computed with **Algorithm 3**, Gibbs sampling estimates with 5 000 iterations and 4 700 discarded "burn-in" iterations.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.89	4.87	1773.95	1053.68	1574.53	965.96
Bias _{rel}	25.88%	21.74%	-29.04%	5.37%	-30.23%	6.57%
RMSE _{rel}	2.13%	1.57%	0.81%	0.28%	0.78%	0.25%

Table 14: Average ML estimates and MTTF for the two components with 10% right censoring data. The estimates are computed with **Algorithm 4**, BR-PM estimation with full restoration and 5 000 Bayesian sample size.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.65	4.27	2 457.63	978.67	2 372.37	890.44
Bias _{rel}	10.11%	6.85%	-1.69%	-2.13%	5.11%	-1.76%
RMSE _{rel}	36.10%	3.28%	52.49%	0.51%	74.51%	0.68%

Table 15: Average ML estimates and MTTF for the two components with 10% right censoring data. The estimates are computed with **Algorithm 5**, BR-LM estimation with full restoration and 5 000 Bayesian sample size.

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.86	MTTF2 906.40
Mean	1.23	3.74	2587.42	1024.54	4285.70	925.17
Bias _{rel}	-17.70%	-6.47%	3.49%	2.45%	89.90%	2.07%
RMSE _{rel}	0.82%	0.26%	6.69%	0.12%	7.41%	0.13%

Comments: It appears clearly that the best estimates are obtained by the algorithm 5 and that its advantage over the other algorithms increases with the censoring rate as expected. The superiority of algorithm 5 on algorithm 4 is noteworthy. Otherwise, the two importance Bayesian sampling algorithms outperformed the ml algorithms, while Gibbs sampling did not improve so much the performance of ml algorithms.

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 16). The Weibull QQ-plot suggests two main competing causes of failures (see Figure 3).

Tables 17-19 provide 95% bootstrap or credible intervals for the SEM-EM algorithm, the Gibbs sampler and the BR-LM-EM 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-ML-EM provides more accurate intervals, especially for scale parameters.

Table 16: Aircraft windshields: lifetimes ($\times 1000 h$). "-" indicated that the windshield 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							

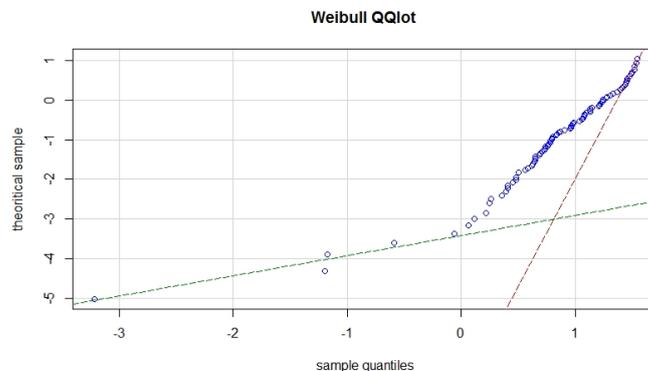


Figure 3: Weibull QQ-plot for aircraft windshield *cf.* Murthy et al. [2003]; there are two main competing causes of failures.

Table 17: 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 18: Aircraft windshields: $n = 153$ and the censorship rate is 57.5%. Parameters 95% CI's with the Algorithm 3, Gibbs sampler with

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 19: Aircraft 2.838 windshields: $n = 153$ and the censorship rate is 57.5%. Parameters 95% CI's with the Algorithm 5, BR-LM

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

As can be seen in Figure 4, the EM algorithm is rather stable but tends to be biased in the case of heavily censored data. It can be seen that initialization with the SEM algorithm does not correct the bias. Therefore, our Bayesian regularization is useful: a Bayesian sample is obtained from the prior distribution (see Figure 5)-Step 1). Hence, the missing data are restored from this Bayesian sample. Thus, it is easier to maximize the completed posterior distribution *cf.* Algorithm 4 or the completed likelihood *cf.* Algorithm 5 (see Figure 5)-Step 2). The values obtained are then used as starting values for the EM algorithm: the bias is then much better better attenuated (see Figure 5 - Step 3).

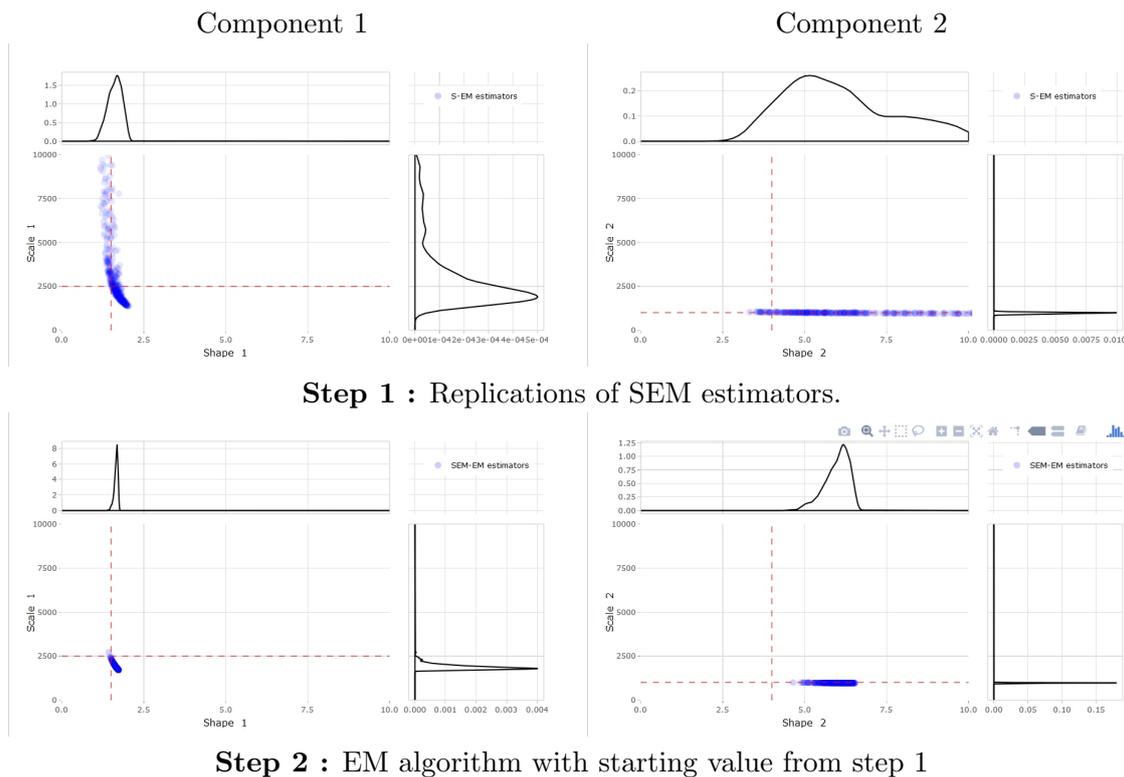
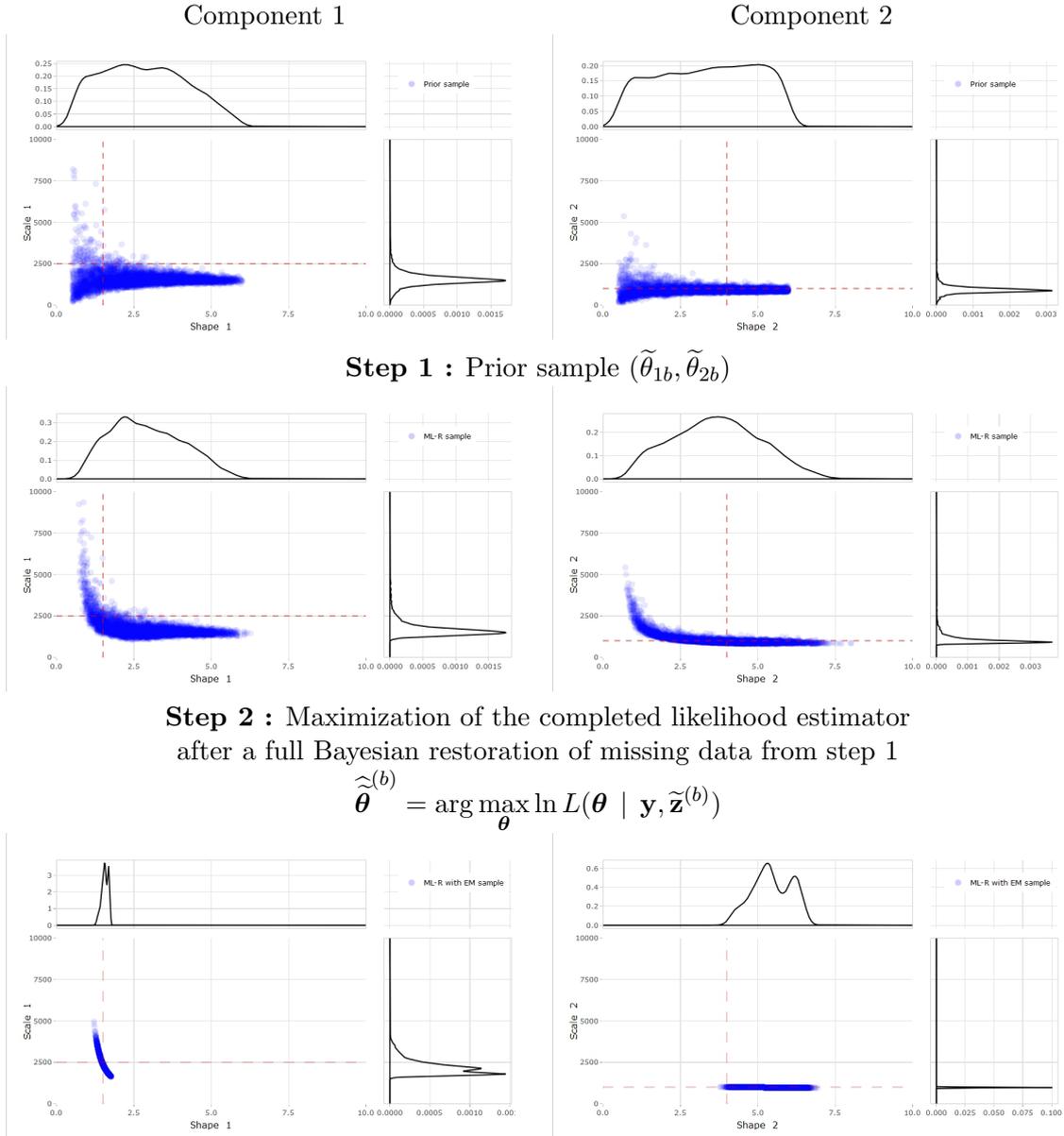


Figure 4: EM with SEM as starting values: the steps of algorithm 2



Step 3 : EM algorithm with starting value from step 2

Figure 5: Bayesian restoration: the steps of algorithm 5

Table 20 provides comparisons between Full and Simple restoration for algorithm 5 (*cf.* section 3.3). It can be seen that the Full restoration version offers better results in terms of relative RSME.

Table 20: Comparison Full versus Simple restoration for **Algorithm 5**, BR-LM estimation. Synthetic data with 70% right censoring

	β_1 1.5	β_2 4	η_1 2 500	η_2 1 000	MTTF1 2256.863	MTTF2 906.4025
Full restoration (masked causes and censored lifetimes)						
Mean	1.40	4.60	3066.35	999.05	2793.67	912.84
Bias _{rel}	-6.43%	14.92%	22.65%	-0.09%	23.79%	0.61%
RMSE _{rel}	0.28%	0.46%	1.38%	0.04%	1.45%	0.02%
Simple restoration (only masked causes)						
Mean	1.43	4.71	2877.02	996.60	2614.11	911.91
Bias _{rel}	-4.74%	17.83%	15.08%	-0.34%	15.83%	0.71%
RMSE _{rel}	0.37%	0.67%	1.45%	0.05%	1.52%	0.02%

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 (see Figure 6). This label switching problem that complicates the estimation must then be addressed. The solution chosen is the one proposed by Celeux [1998].

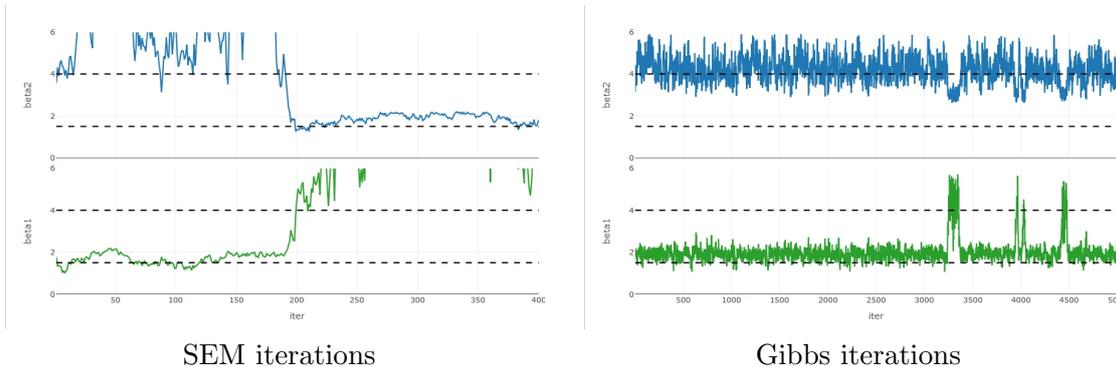


Figure 6: Label switching for the SEM algorithm and the Gibbs sampler

For SEM and the two Bayesian restoration algorithms, if no failure or less than three are assigned to a cause, then the algorithm is shut down. It can be seen that while this is frequently observed for the SEM algorithm, it is not observed with both Bayesian restoration algorithms.

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 is not costly in terms of computing time. 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.

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

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

2. Shape

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

where $f(\beta_k | \mathbf{t}, \mathbf{e}, \mathbf{o}^{(b)}, \boldsymbol{\beta}_{(-k)}^{(b)}, \boldsymbol{\eta}^{(b)})$ is a log concave distribution :

$$\begin{aligned} \log f(\beta_k | \mathbf{y}, \mathbf{o}^{(b)}, \boldsymbol{\beta}_{(-k)}^{(b)}, \boldsymbol{\eta}^{(b)}) & = (p-1) \log(\beta_k - b_{min}) + (q-1) \log(b_{max} - \beta_k) \\ & + \log(\beta_k) + \beta_k b_k \log(a_k) - \beta_k b_k \log(\eta_k^{(b)}) - \left(\frac{a_k}{\eta_k^{(b)}}\right)^{\beta_k} \\ & + \#\{o_i^{(b)} = k\} \log(\beta_k) + \beta_k \sum_{i: o_i=k} \log\left(\frac{t_i}{\eta_k^{(b)}}\right) \\ & - \sum_i^n \left(\frac{t_i}{\eta_k^{(b)}}\right)^{\beta_k} \\ & = (p-1) \log(\beta_k - b_{min}) + (q-1) \log(b_{max} - \beta_k) + (\#\{o_i^{(b)} = k\} + 1) \log(\beta_k) \\ & + \beta_k \left(b_k \log\left(\frac{a_k}{\eta_k^{(b)}}\right) + \sum_{i: o_i=k} \log\left(\frac{t_i}{\eta_k^{(b)}}\right) \right) - \left(\frac{a_k}{\eta_k^{(b)}}\right)^{\beta_k} - \sum_i^n \left(\frac{t_i}{\eta_k^{(b)}}\right)^{\beta_k} \quad (22) \end{aligned}$$

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

$$\begin{aligned} \frac{\partial^2 \log f}{\partial \beta_k^2}(\beta_k | \mathbf{y}, \mathbf{o}^{(b)}, \boldsymbol{\beta}_{(-k)}^{(b)}, \boldsymbol{\eta}^{(b)}) & = -\frac{p-1}{(\beta_k - b_{min})^2} - \frac{q-1}{(b_{max} - \beta_k)^2} - \frac{\#\{o_i^{(b)} = k\} + 1}{\beta_k^2} \\ & - \left[\log\left(\frac{a_k}{\eta_k^{(b)}}\right) \right]^2 \left(\frac{a_k}{\eta_k^{(b)}}\right)^{\beta_k} - \sum_i \left[\log\left(\frac{t_i}{\eta_k^{(b)}}\right) \right]^2 \left(\frac{t_i}{\eta_k^{(b)}}\right)^{\beta_k} \quad (24) \end{aligned}$$

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

3. Scale

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

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

$$GIG\left(\#\{o_i^{(b)} = k\} + b_k; \left[\sum_i t_i^{\beta_k^{(b)}} + a_k^{\beta_k^{(b)}}\right]^{1/\beta_k^{(b)}}; \beta_k^{(b)}\right). \quad (25)$$

Appendix B: Weibull CR completed posterior distribution

In algorithm 4, after a full restoration, the maximization step is performed by equating to zero the partial derivatives of completed log posterior, $\log \pi(\theta | \mathbf{t})$, with respect to each parameter :

$$\begin{aligned}
 \frac{\partial \log \pi(\theta | \mathbf{t}, \mathbf{o})}{\partial \eta_k} &= \frac{\partial \log L(\mathbf{t}, \mathbf{o} | \theta)}{\partial \eta_k} + \frac{\partial \log \pi(\eta_k | \beta_k)}{\partial \eta_k} \\
 &= \frac{\beta_k}{\eta_k^{\beta_k+1}} \sum_{i=1}^n t_i^{\beta_k} - n \cdot \frac{\beta_k}{\eta_k} \\
 &\quad - \frac{b \cdot \beta_k + 1}{\eta_k} + \frac{\beta_k}{\eta_k^{\beta_k+1}} a^{\beta_k} \\
 &= \frac{\beta_k}{\eta_k} \left(\frac{\sum_{i=1}^n t_i^{\beta_k} + a^{\beta_k}}{\eta_k^{\beta_k}} - n - b - \frac{1}{\beta_k} \right)
 \end{aligned} \tag{26}$$

$$\begin{aligned}
 \frac{\partial \log \pi(\theta | \mathbf{t}, \mathbf{o})}{\partial \beta_k} &= \frac{\partial \log L(\mathbf{t}, \mathbf{o} | \theta)}{\partial \beta_k} + \frac{\partial \log \pi(\beta_k)}{\partial \beta_k} + \frac{\partial \log \pi(\eta_k | \beta_k)}{\partial \beta_k} \\
 &= - \sum_{i=1}^n \log \left(\frac{t_i}{\eta_k} \right) \left(\frac{t_i}{\eta_k} \right)^{\beta_k} + n \cdot \frac{1}{\beta_k} + \sum_{i=1}^n \log(t_i) - n \cdot \log \eta_k \\
 &\quad + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\
 &\quad + \frac{1}{\beta_k} + b \log a - b \log(\eta_k) - \log \left(\frac{a}{\eta_k} \right) \left(\frac{a}{\eta_k} \right)^{\beta_k} \\
 &= - \sum_{i=1}^n \log \left(\frac{t_i}{\eta_k} \right) \left(\frac{t_i}{\eta_k} \right)^{\beta_k} - \log \left(\frac{a}{\eta_k} \right) \left(\frac{a}{\eta_k} \right)^{\beta_k} \\
 &\quad - (n+b) \cdot \log \eta_k \\
 &\quad + \frac{n+1}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\
 &\quad + \sum_{i: o_i=k} \log(t_i) + b \log a. \\
 &= - \frac{\sum_{i=1}^n \log(t_i) t_i^{\beta_k} + \log(a) a^{\beta_k} - \log(\eta_k) \left(\sum_{i=1}^n t_i^{\beta_k} + a^{\beta_k} \right)}{\eta_k^{\beta_k}} \\
 &\quad - (n+b) \cdot \log \eta_k \\
 &\quad + \frac{n+1}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\
 &\quad + \sum_{i=1}^n \log(t_i) + b \log a.
 \end{aligned} \tag{27}$$

Thus

$$\frac{\partial \log \pi(\theta | \mathbf{t}, \mathbf{o})}{\partial \eta_k} = 0 \Leftrightarrow \sum_{i=1}^n t_i^{\beta_k} + a^{\beta_k} = \eta_k^{\beta_k} \left(n + b + \frac{1}{\beta_k} \right). \quad (28)$$

Therefore

$$\eta_k = \left(\frac{\sum_{i=1}^n t_i^{\beta_k} + a^{\beta_k}}{n + b + \frac{1}{\beta_k}} \right)^{1/\beta_k}. \quad (29)$$

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

$$\begin{aligned} & \frac{\partial \log \pi(\theta | \mathbf{t}, \mathbf{o})}{\partial \beta_k} = 0 \\ \Leftrightarrow & \frac{\sum_{i=1}^n t_i^{\beta_k} + a^{\beta_k}}{\eta_k^{\beta_k}} \left[- \frac{\sum_{i=1}^n \log(t_i) t_i^{\beta_k} + \log(a) a^{\beta_k}}{\sum_{i=1}^n t_i^{\beta_k} + a^{\beta_k}} + \log(\eta_k) \right] \\ & - \left(n + b \right) \log(\eta_k) \\ & + \frac{n+1}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\ & + \sum_{i=1}^n \log(t_i) + b \log a = 0 \\ \Leftrightarrow & \left(n + b + \frac{1}{\beta_k} \right) \left[- \frac{\sum_{i=1}^n \log(t_i) t_i^{\beta_k} + \log(b) b^{\beta_k}}{\sum_{i=1}^n t_i^{\beta_k} + b^{\beta_k}} + \log(\eta_k) \right] \\ & - (n + b) \log(\eta_k) \\ & + \frac{n+1}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\ & + \sum_{i=1}^n \log(t_i) + b \log a = 0 \\ \Leftrightarrow & \left(n + b + \frac{1}{\beta_k} \right) \left[- \frac{\sum_{i=1}^n \log(t_i) t_i^{\beta_k} + \log(b) b^{\beta_k}}{\sum_{i=1}^n t_i^{\beta_k} + b^{\beta_k}} \right] \\ & + \frac{1}{\beta_k} \log(\eta_k) \\ & + \frac{n}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\ & + \sum_{i=1}^n \log(t_i) + b \log a = 0 \end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow - \left(n + b + \frac{1}{\beta_k} \right) \left[\frac{\sum_{i=1}^n \log(t_i) t_i^{\beta_k} + \log(a) a^{\beta_k}}{\sum_{i=1}^n t_i^{\beta_k} + a^{\beta_k}} \right] \\
&\quad - \frac{1}{\beta_k^2} \log \left(\frac{\sum_{i=1}^n t_i^{\beta_k} + a^{\beta_k}}{n + b + \frac{1}{\beta_k}} \right) \\
&\quad + \frac{n+1}{\beta_k} + \frac{p-1}{\beta_k - b_{min}} - \frac{q-1}{b_{max} - \beta_k} \\
&\quad + \sum_{i=1}^n \log(t_i) + b \log a = 0
\end{aligned} \tag{30}$$

References

- M. Bacha, G. Celeux, E. Ide, A. Lannoy, and D. Vasseur. *Estimation de dures de vie fortement censures*. 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.
- G. O. Roberts and A. F.M. Smith. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic processes and their applications*, 49(2):207–216, 1994.
- A.M. Sarhan, D.C. Hamilton, and B. Smith. Statistical analysis of competing risks models. *Reliability Engineering & System Safety*, 2010.
- A.F.M. Smith and A. E. Gelfand. Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician*, 46(2):84–88, 1992.
- 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.