An alternative competing risk model to the Weibull distribution in lifetime data analysis

Henri Bertholon, Nicolas Bousquet, Gilles Celeux

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

An alternative competing risk model to the Weibull distribution in lifetime data analysis

Henri Bertholon, Nicolas Bousquet, Gilles Celeux

N° 5265

Juillet 2004

Thème COG

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--5265--FR+ENG

An alternative competing risk model to the Weibull distribution in lifetime data analysis

Henri Bertholon*, Nicolas Bousquet†, Gilles Celeux‡

Thème COG — Systèmes cognitifs
Projet SELECT

Rapport de recherche n° 5265 — Juillet 2004 — 25 pages

Abstract: A simple competing risk distribution as a possible alternative to the Weibull distribution in lifetime analysis is proposed. This distribution corresponds to the minimum between exponential and Weibull distributions. Our motivation is to take account of both accidental and aging failures in lifetime data analysis. First, the main characteristics of this distribution are presented. Then the estimation of its parameters are considered through maximum likelihood and Bayesian inference. Decision tests to choose between an exponential, Weibull and this competing risk distribution are presented. And this alternative model is compared to the Weibull model from numerical experiments on both real and simulated data sets.

Key-words: Failure Time Distribution; Aging; Weibull Distribution; Accidental Failure; Competing risk Model; EM algorithm; Bayesian Inference; Importance sampling; Likelihood ratio Test.

* INRIA Rhône-Alpes, 655 avenue de l'Europe - Montbonnot - 38334 Saint Ismier Cedex, email : henri.bertholon@inria.fr

† INRIA Futurs, Département de Mathématiques, Université Paris-Sud, 91405 Orsay Cedex,email : nicolas.bousquet@math.u-psud.fr

‡ INRIA Futurs, Département de Mathématiques, Université Paris-Sud 91405 Orsay Cedex, email: Gilles.Celeux@inria.fr

Unité de recherche INRIA Futurs
Parc Club Orsay Université, ZAC des Vignes,

Un modèle de durée de vie à risques concurrents comme alternative au modèle de Weibull

Résumé : Un modèle de durée de vie à risques de défaillance concurrents est proposé, comme alternative au modèle classique de Weibull du vieillissement. Ce modèle met en compétition les risques de défaillance par vieillesse et par accident, sa distribution correspondant au minimum entre des distributions exponentielle et de Weibull. Les caractéristiques principales de ce modèle sont présentées. On propose ensuite des procédures d'estimation fréquentielles et bayésiennes du modèle. Une stratégie de tests de choix de modèles entre les modèles exponentiels, Weibull et à risques concurrents est également présentée, ainsi que des analyses comparatives sur des exemples de données simulées et réelles.

Mots-clés : temps de défaillance ; vieillisement ; modèle de Weibull ; modèle exponentiel ; défaillance accidentelle ; modèle à risques concurrents ; algorithme EM ; inférence bayésienne ; échantillonnage pondéré ; test de rapport de vraisemblance.

Contents

1 Introduction

In a reliability context, the most employed lifetime distributions are the exponential and the Weibull distributions (see for instance Meeker and Escobar, 1998). The exponential distribution $\mathcal{E}(\eta)$ whose reliability function is

$$R_E(t) = \exp(-\frac{t}{\eta}), \tag{1}$$

the scale parameter η being the inverse of the constant hazard rate λ, is modelling accidental failure times of a no aging material cleared of infant mortality defects. While the versatile Weibull $\mathcal{W}(\eta, \beta)$ distribution, with reliability function

$$R_W(t) = \exp[-(\frac{t}{\eta})^{\beta}] \tag{2}$$

and hazard rate

$$h_W(t) = \frac{\beta}{\eta}(\frac{t}{\eta})^{\beta-1} \tag{3}$$

can be used for modelling infant mortality defects when the shape parameter $\beta < 1$ or aging when $\beta > 1$. Note that when $\beta = 1$ the Weibull distribution reduces to an exponential distribution with scale parameter η.

Reliability feedback experience data are often modelled with the Weibull distribution and the important question is to decide if $\beta = 1$ versus $\beta < 1$ when concerned with infant mortality or $\beta = 1$ versus $\beta > 1$ when concerned with aging. This question can be solved using likelihood ratio tests (see d'Agostino and Stephens, 1986). For simplicity in the following, we consider that we are interested to model possible aging of a material cleared of infant mortality defects. If, for instance, aging is diagnosed, then further statistical inference is made assuming that the observed failure times arise from a Weibull distribution. Acting in such a way means that the occurrence of accidental failures can be regarded as negligible as compared to the occurrence of failures caused by material aging. This assumption could appear to be reasonable in many circumstances, but there are a lot of situations where neglecting accidental failures can introduce an important bias in statistical inference on material lifetimes. Even when aging is the most frequent cause of failure, accidental failures can remain numerous.

Thus a more realistic way of modelling failure times is to consider a competing risk model which takes into account the fact that a failure can be caused by aging or accidentally. This model is defined as follows. A failure time is the realization of a random variable $B = \min(E, W)$ where E is a random variable with an exponential distribution $\mathcal{E}(\eta_0)$ and W is a random variable with a Weibull distribution $\mathcal{W}(\eta_1, \beta)$ where $\beta > 1$. Consequently, the distribution of B is characterized by three parameters η_0, η_1 and β. It will be denoted $\mathcal{B}(\eta_0, \eta_1, \beta)$. Its failure rate function is the sum of the exponential and Weibull failure rates

$$h_B(x) = \frac{1}{\eta_0} + \frac{\beta}{\eta_1}(\frac{x}{\eta_1})^{(\beta-1)}. \tag{4}$$

The aim of this paper is to analyze the possibility of using a $\mathcal{B}(\eta_0, \eta_1, \beta)$ instead of a $\mathcal{W}(\eta, \beta)$ distribution for modelling aging.

The paper is organized as follows. In Section 2 the main characteristics of the $\mathcal{B}(\eta_0, \eta_1, \beta)$ distribution are presented. In Section 3, the estimation of the $\mathcal{B}(\eta_0, \eta_1, \beta)$ distribution is considered. First maximum likelihood estimation of the three parameters of the $\mathcal{B}$ distribution from possibly right censored data is presented through the EM algorithm. Then, Bayesian estimation of those parameters is presented using an importance sampling approach to approximate the posterior distribution of the parameters. Section 4 is concerned with hypothesis tests. In particular, the important problem of testing a Weibull distribution against a $\mathcal{B}$ distribution is considered. Section 5 is devoted to the presentation of numerical experiments on both simulated and real data sets and a short discussion section ends the paper.

2 Characteristics of the $\mathcal{B}$ distribution

Let a random variable (r.v.) $B = \min(E, W)$, where the r.v. E has an exponential distribution with mean value η_0 and the r.v. W has a Weibull distribution with scale parameter η_1 and shape parameter β, and E and W are independent. The main characteristics of the $\mathcal{B}$ probability distribution are as follows. Its hazard function is

$$h_B(x) = h_E(x) + h_W(x) = \frac{1}{\eta_0} + \frac{\beta}{\eta_1}(\frac{x}{\eta_1})^{(\beta-1)}, \tag{5}$$

its reliability (or survival) function is

$$S_B(x) = S_E(x) \times S_W(x) = \exp[-\frac{1}{\eta_0}x - (\frac{x}{\eta_1})^{\beta}], \tag{6}$$

and its probability density function (pdf) is

$$f_B(x) = \left[\frac{1}{\eta_0} + \frac{\beta}{\eta_1}\left(\frac{x}{\eta_1}\right)^{\beta-1}\right] \exp[-\frac{1}{\eta_0}x - (\frac{x}{\eta_1})^{\beta}]. \tag{7}$$

On Figure 1 are displayed examples of $\mathcal{B}$ pdf's with the corresponding exponential and Weibull pdf for one example.

In view of analyzing the roles of scale parameters η_0 and η_1 in failure time data analysis, it is of interest to calculate the probability that a failure arises from E (accidental failure), that is the probability that $B = E$. It is $P(B = E) = P(E \leq W)$. Assuming a shape parameter $\beta = 2$, it leads to

$$P(B = E) = \frac{\eta_1}{\eta_0}\frac{\sqrt{\pi}}{2}\,\mathrm{erfcx}(\frac{\eta_1}{2\eta_0})$$

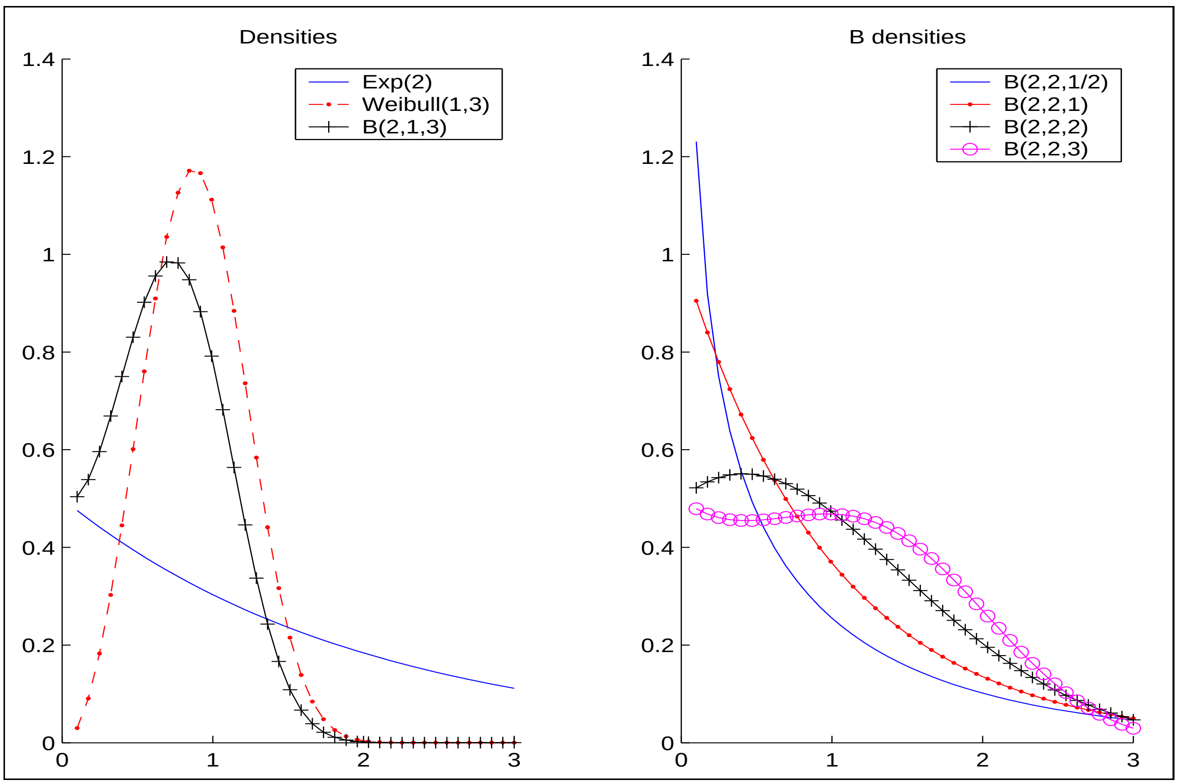

Figure 1: Examples of $\mathcal{B}$ pdf's.

η_0/η_1	0.1	0.2	0.5	1	1.5	2	5	10
$P(X=E)=P(E\leq W)$	0.98	0.93	0.75	0.54	0.42	0.34	0.15	0.08

Table 1: Probability of an accidental failure as a function of ratio $\frac{\eta_0}{\eta_1}$ in a $\mathcal{B}(\eta_0,\eta_1,2)$ distribution.

where the function erfcx is

$$\text{erfcx}(x) = e^{x^2}\frac{2}{\sqrt{\pi}}\int_x^{+\infty} e^{-u^2}du.$$

Table 1 displays the evolution of this probability as a function of ratio $\frac{\eta_0}{\eta_1}$. No surprisingly, $\mathcal{B}(\eta_0,\eta_1,\beta) \approx \mathcal{W}(\eta_1,\beta)$ as $\eta_0 >> \eta_1$ and $\mathcal{B}(\eta_0,\eta_1,\beta) \approx \mathcal{E}(\eta_0)$ as $\eta_1 >> \eta_0$. Since we are presumably concerned with situations where aging can be sensitive, it is reasonable to assume that $\eta_0 \geq \eta_1$, because $\eta_1 > \eta_0$ implies a predominant frequency of accidental failure. In the following, this a priori assumption will be made in the Bayesian framework.

The Laplace transform of $\mathcal{B}$ is

$$\begin{aligned} G_B(u) &= \int_0^{+\infty} e^{ux} f_B(x)dx, \\ &= \int_0^{+\infty}\left[\frac{1}{\eta_0}+\frac{\beta}{\eta_1}\left(\frac{x}{\eta_1}\right)^{\beta-1}\right]e^{-\left(\frac{1}{\eta_0}-u\right)x}e^{-\left(\frac{x}{\eta_1}\right)^{\beta}}dx, \\ &= \int_0^{+\infty}\left[\frac{\eta_1}{\eta_0}+\beta y^{\beta-1}\right]e^{-\left(\frac{1}{\eta_0}-u\right)(\eta_1 y)}e^{-y^{\beta}}dy, \end{aligned}$$

with $y=\frac{x}{\eta_1}$.

Laplace transform of $\mathcal{B}$ cannot be calculated in a closed formed. For $\beta = 2$, using the function

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{+\infty} e^{-u^2}du,$$

it can be written (see Bertholon 2001 for details)

$$G(u) = 1+\exp\left[\frac{1}{2}\eta_1\left(\frac{1}{\eta_0}-u\right)^2\right]\left[\eta_1 u\frac{\sqrt{\pi}}{2}\text{erfc}\left(\frac{1}{2}\eta_1\left(\frac{1}{\eta_0}-u\right)\right)\right]. \tag{8}$$

From this expression, it is possible to derive the mean and the variance of the $\mathcal{B}$ distribution for $\beta = 2$ (Bertholon 2001):

$$E[X] = \eta_1 e^{\frac{\eta_1^2}{4\eta_0^2}}\frac{\sqrt{\pi}}{2}\text{erfc}(\frac{\eta_1}{2\eta_0})$$

and

$$Var[X] = 2\left[\eta_1 e^{-(\frac{\eta_1}{2\eta_0})^2}\left(-\frac{\eta_1^2}{2\eta_0}\frac{\sqrt{\pi}}{2}\,\mathrm{erfc}(\frac{\eta_1}{2\eta_0}) + \frac{1}{2}\eta_1 e^{-(\frac{\eta_1}{2\eta_0})^2}\right)\right] - \left[\eta_1 e^{\frac{\eta_1^2}{4\eta_0^2}}\frac{\sqrt{\pi}}{2}\,\mathrm{erfc}(\frac{\eta_1}{2\eta_0})\right]^2.$$

It is worthwhile to give some limit values of those statistics.

- When $\eta_0 \rightarrow +\infty$,

  $$E[X] \rightarrow \eta_1 \frac{\sqrt{\pi}}{2}$$

  because $\mathrm{erfc}(x) \rightarrow 1$ when $x \rightarrow 0$, and

  $$Var[X] \rightarrow \eta_1^2\left(1 - \left(\frac{\sqrt{\pi}}{2}\right)^2\right).$$

  As expected, the mean and variance of the $\mathcal{B}$ distribution tend to the mean and variance of the corresponding Weibull distribution

  $$E[X] = \eta_1 \Gamma\left(1 + \frac{1}{\beta}\right)$$

  and

  $$Var[X] = \eta_1^2\left(\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right).$$

- When $\eta_1 \rightarrow +\infty$

  $$E[X] \rightarrow \eta_0$$

  because $\frac{\sqrt{\pi}}{2}\mathrm{erfc}(x) \sim e^{-x^2}\left(\frac{1}{2x} - \frac{1}{4x^3}\right)$ when $x \rightarrow +\infty$, and

  $$Var[X] \rightarrow \eta_0^2,$$

  namely the mean and variance of the corresponding exponential distribution.

3 Estimating the parameter of the $\mathcal{B}$ distribution

The $\mathcal{B}$ distribution is a three-parameter distribution and estimating its parameters can be thought of as more difficult than estimating the two parameters of a Weibull distribution. However, it is possible to take profit of the fact that the $\mathcal{B}$ distribution can be regarded as an incomplete data model to use efficient estimation algorithms as the EM algorithm for maximum likelihood (ml) estimation(Dempster, Laird and Rubin, 1977) or Data augmentation algorithms for Bayesian inference (Tanner and Wong, 1987).

Before going into the details of the EM algorithm to derive the ml estimates of the $\mathcal{B}$ distribution parameters, we first prove that the likelihood equations have a unique consistent root.

3.1 Existence of a consistent root of the likelihood equations

The aim of this section is to prove that the likelihood equations for the $\mathcal{B}$ distribution have a root which is consistent and asymptotically normally distributed. The proof uses a Chanda's theorem that is recalled first.

Theorem 1 (Chanda) Let $f(x;\theta)$ be a probability density function, $\theta = (\theta_1, ..., \theta_k)$ being a vector parameter belonging to the parameter space Ω, and $x_1, ..., x_n$ be independent observations of a random variable X with density $f(x;\theta)$. The likelihood equations are given by $\frac{\partial \ln L}{\partial \theta} = 0$, where $\ln L = \sum_{i=1}^{n} \ln f(x_i, \theta)$. Let θ_0 denote the true value of θ. It is assumed that θ_0 lies at some point in Ω. Then, if Conditions 1-3 below hold, there exists a unique consistent estimator θ_n, solution of the likelihood equations. Furthermore, $\sqrt{n}(\theta_n - \theta_0)$ is asymptotically normally distributed with mean zero and covariance matrix $\mathrm{I}(\theta_0)^{-1}$, where $\mathrm{I}(\theta_0)$ is the Fisher information matrix.

- Condition 1 For almost all x and for all $\theta \in \overline{\Omega}$, $\frac{\partial \ln f}{\partial \theta_r}$, $\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}$ and $\frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t}$ exist for all $r, s, t = 1, ..., k$.
- Condition 2 For almost all x and for all $\theta \in \overline{\Omega}$, $\left|\frac{\partial f}{\partial \theta_r}\right| < F_r(x)$, $\left|\frac{\partial^2 f}{\partial \theta_r \partial \theta_s}\right| < F_{rs}(x)$ and $\left|\frac{\partial^3 f}{\partial \theta_r \partial \theta_s \partial \theta_t}\right| < H_{rst}(x)$, where H_{rst} is such that $\int_{-\infty}^{+\infty} H_{rst}(x) f(x) dx \leq M < \infty$, and $F_r(x)$ and $F_{rs}(x)$ are bounded for all $r, s, t = 1, ..., k$.
- Condition 3 For all $\theta \in \overline{\Omega}$, the matrix $I(\theta) = \int_{-\infty}^{+\infty} \left(\frac{\partial \ln f}{\partial \theta}\right) \left(\frac{\partial \ln f}{\partial \theta}\right)' f dx$ is positive definite.

The three conditions of the Chanda theorem are now checked for the likelihood equations of the $\mathcal{B}$ distribution. First, the partial derivatives of the pdf f are of the following form

$$\frac{\partial \ln f}{\partial \theta_r} = \frac{\partial f}{\partial \theta_r} \frac{1}{f}$$

$$\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s} = \frac{1}{f} \frac{\partial^2 f}{\partial \theta_r \partial \theta_s} - \frac{\partial f}{\partial \theta_r} \frac{\partial f}{\partial \theta_s} \frac{1}{f^2}$$

$$\frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t} = 2 \frac{\partial f}{\partial \theta_r} \frac{\partial f}{\partial \theta_s} \frac{\partial f}{\partial \theta_t} \frac{1}{f^3} - \frac{\partial^2 f}{\partial \theta_r \partial \theta_t} \frac{\partial f}{\partial \theta_s} \frac{1}{f^2} - \frac{\partial f}{\partial \theta_r} \frac{\partial^2 f}{\partial \theta_s \partial \theta_t} \frac{1}{f^2} - \frac{\partial f}{\partial \theta_t} \frac{\partial^2 f}{\partial \theta_r \partial \theta_s} \frac{1}{f^2} + \frac{\partial^3 f}{\partial \theta_r \partial \theta_s \partial \theta_t} \frac{1}{f},$$

where θ is the vector parameter $\theta = (\eta_0, \eta_1, \beta)$.

By induction, it can be proved that f and its partial derivatives of any order (denoted below by g) can be written in the following way:

$$e^{-\frac{1}{\eta_0}x - \left(\frac{x}{\eta_1}\right)^\beta} \left[P(\frac{1}{\eta_0}, \frac{1}{\eta_1}, \beta) + \sum_{k_1=0}^{M_1} \sum_{k_2=1}^{M_2} \sum_{k_3=0}^{1} Q_{k_1 k_2 k_3}(\frac{1}{\eta_0}, \frac{1}{\eta_1}, \beta) \left(\ln \left(\frac{x}{\eta_1} \right) \right)^{k_1} \left(\frac{x}{\eta_1} \right)^{k_2 \beta - k_3} \right]$$

where $P(\frac{1}{\eta_0}, \frac{1}{\eta_1}, \beta)$ and $Q_{k_1 k_2 k_3}(\frac{1}{\eta_0}, \frac{1}{\eta_1}, \beta)$ are polynomials in $\frac{1}{\eta_0}, \frac{1}{\eta_1}$ and β. k_1 and k_2 take respectively their values in discrete sets $\{0, 1, \ldots, M_1\}$ and $\{0, 1, \ldots, M_2\}$. Consequently, Condition 1 is satisfied.

Now any partial derivative g is a continuous function in x and θ. Thus g is bounded for $\theta \in \overline{\Omega}$ and x in any closed interval. Therefore to check Condition 2, it suffices to consider its behavior for large values of x. It is easily seen that there exist positive numbers A and B such that g is inferior to $e^{-Bx} \times x^A$ for sufficiently large x and $\theta \in \overline{\Omega}$. Since $e^{-Bx} \times x^A$ is bounded, Condition 2 is satisfied.
As for Condition 3, $I(\theta)$, which is a covariance matrix, is positive definite unless it exists a, b, c not all equal to zero such that $a\frac{\partial \ln f}{\partial \eta_0} + b\frac{\partial \ln f}{\partial \eta_1} + c\frac{\partial \ln f}{\partial \beta} = 0$. A simple examination of the derivatives shows straightforwardly that they are not collinear. Thus the three conditions of Chanda theorem are verified.

3.2 The EM algorithm

Let $\mathbf{y} = (y_1, \ldots, y_n)$ be a sample from a $\mathcal{B}$ distribution which can contain type I right censored data (fixed censoring time) . Each y_i can be written $y_i = (t_i, \delta_i)$, where $\delta_i = \begin{cases} 0 & \text{if } t_i \text{ is a censoring time,} \\ 1 & \text{if } t_i \text{ is a failure time.} \end{cases}$

Thus the observed likelihood which is

$$\begin{aligned} L(\eta_0, \eta_1, \beta | \mathbf{y}) &= \prod_{i=1}^{n} f_B(t_i)^{\delta_i} S_B(t_i)^{1-\delta_i} \\ &= \prod_{i=1}^{n} h_B(t_i)^{\delta_i} S_B(t_i), \end{aligned}$$

can be written

$$L(\eta_0, \eta_1, \beta | \mathbf{y}) = \left[\prod_{i=1}^{n} \left(\frac{1}{\eta_0} + \frac{\beta}{\eta_1} \left(\frac{t_i}{\eta_1} \right)^{\beta-1} \right)^{\delta_i} \right] \exp \left[-\frac{1}{\eta_0} \sum_{i=1}^{n} t_i - \sum_{i=1}^{n} \left(\frac{t_i}{\eta_1} \right)^{\beta} \right]. \quad (9)$$

Maximizing (9) is typically a difficult problem for which the use of the EM algorithm can be recommended. As a matter of fact, the $\mathcal{B}$ distribution is the distribution of a competing risk model with missing data. The missing data are binary indicator values associated to the failure times. If t_i is a failure time, we define $z_i = (z_i^E, z_i^W)$ where $z_i^E = 1$ and $z_i^W = 0$ if the failure time t_i arose from the exponential distribution and $z_i^W = 1$ and $z_i^E = 0$ if t_i arose from the Weibull distribution. By convention, if t_i is a censoring time ($\delta_i = 0$), $z_i^E = 0$ and $z_i^W = 0$. Thus, the complete data set can be written $\mathbf{x} = (x_i = (y_i, z_i), i = 1, \ldots, n) = (\mathbf{y}, \mathbf{z})$. The density of a complete observation x_i is

$$f(x_i) = (f_E(t_i))^{z_i^E} S_E(t_i)^{1-z_i^E} (f_W(t_i))^{z_i^W} S_W(x_i)^{1-z_i^W},$$

$$f(x_i) = (h_E(t_i))^{z_i^E} (h_W(t_i))^{z_i^W} S_E(t_i) S_W(t_i).$$

And, the complete loglikelihood can be written

$$l(\theta|\mathbf{x}) = \sum_{i=1}^{n} \left[z_i^E \ln(h_E(t_i)) + z_i^W \ln(h_W(t_i)) + \ln(S_E(t_i)) + \ln(S_W(t_i)) \right]. \tag{10}$$

The EM algorithm consists of maximizing the conditional expectation of the complete likelihood knowing the observed data and a current value $\tilde{\theta}$ of the parameter in an iterative two steps algorithm (Dempster, Laird and Rubin 1977, McLachlan and Krishnam 1997). The E step is calculating this conditional expectation, denoted $Q(\theta|\tilde{\theta})$, and the M step is maximizing $Q(\theta|\tilde{\theta})$ with respect to θ.

E step
It consists of calculating $Q(\theta|\tilde{\theta})$, $\tilde{\theta}$ being the current parameter value.

$$\begin{aligned} Q(\theta|\tilde{\theta}) &= E\left(l(\theta|\mathbf{x})|\mathbf{y},\tilde{\theta}\right) \\ &= \sum_{i=1}^{n} \left[E\left(z_i^E|\mathbf{y},\tilde{\theta}\right) \ln(h_E(t_i)) + E\left(z_i^W|\mathbf{y},\tilde{\theta}\right) \ln(h_W(t_i)) + \ln(S_E(t_i)) + \ln(S_W(t_i)) \right] \\ &= \sum_{i=1}^{n} \left[\widetilde{p}_E(y_i) \ln(h_E(t_i)) + \widetilde{p}_W(y_i) \ln(h_W(t_i)) + \ln(S_E(t_i)) + \ln(S_W(y_i)) \right] \end{aligned}$$

where

$$\widetilde{p}_E(y_i) = \begin{cases} 0 & \text{if } \delta_i = 0 \\ \frac{h_E(t_i)}{h_E(t_i)+h_W(t_i)} & \text{if } \delta_i = 1, \end{cases}$$

and

$$\widetilde{p}_W(y_i) = \begin{cases} 0 & \text{if } \delta_i = 0 \\ \frac{h_W(t_i)}{h_E(t_i)+h_W(t_i)} & \text{if } \delta_i = 1. \end{cases}$$

$Q(\theta|\tilde{\theta})$ can be written

$$Q(\theta|\tilde{\theta}) = Q_E(\eta_0|\tilde{\theta}) + Q_W(\eta_1,\beta|\tilde{\theta}) \tag{11}$$

where

$$Q_E(\eta_0|\tilde{\theta}) = \sum_{i=1}^{n} \left[\widetilde{p}_E(x_i) \ln(h_E(x_i)) + \ln(S_E(x_i)) \right]$$

and

$$Q_W(\eta_1,\beta|\tilde{\theta}) = \sum_{i=1}^{n} \left[\widetilde{p}_W(x_i) \ln(h_W(x_i)) + \ln(S_W(x_i)) \right].$$

This additive decomposition of $Q(\theta|\tilde{\theta})$ between the contribution of the exponential and the Weibull distributions will facilitate the M step which is now described.

M step
It consists of deriving $\hat{\theta} = \arg\max_\theta Q(\theta|\tilde{\theta})$. From (11) it leads to derive $\hat{\eta_0} = \arg\max_{\eta_0} Q_E(\eta_0|\tilde{\theta})$, and $(\hat{\eta}_1, \hat{\beta}) = \arg\max_{(\eta_1,\beta)} Q(\eta_1, \beta|\tilde{\theta})$. Thus the following equations are straightforwardly obtained

$$\hat{\eta_0} = \frac{\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} \widetilde{p}_E(y_i)},$$

$$\frac{1}{\hat{\beta}} + \frac{\sum_{i=1}^{n} \widetilde{p}_W(y_i) \ln(t_i)}{\sum_{i=1}^{n} \widetilde{p}_W(y_i)} - \frac{\sum_{i=1}^{n} (t_i)^{\hat{\beta}} \ln(t_i)}{\sum_{i=1}^{n} (t_i)^{\hat{\beta}}} = 0,$$

and

$$\hat{\eta}_1 = \left(\frac{\sum_{i=1}^{n} (t_i)^{\hat{\beta}}}{\sum_{i=1}^{n} \widetilde{p}_W(y_i)} \right)^{\frac{1}{\hat{\beta}}}.$$

And the resulting $\hat{\theta}$ becomes the current parameter value.

The EM algorithm is increasing the observed likelihood $L(\eta_0, \eta_1, \beta|\mathbf{y})$ at each iteration and is expected to converge toward the ml estimate of η_0, η_1 and β (Dempster, Laird and Rubin 1977, McLachlan and Krishnam 1997).

It can be noticed that, as for the ml estimation of Exponential and Weibull distribution parameters, type II (random censoring times) censoring times would lead to the same ml equations for the $\mathcal{B}$ distribution parameters.

3.3 Bayesian inference through importance sampling

In many cases, the estimation of reliability distribution functions has to be done from small and highly censored samples. In such a situation, ml estimation turns out to be imprecise or even unreliable. Bayesian inference can be expected to be useful since, in many circumstances including engineering applications, there exists some expert knowledge on the underlying failure mechanism which can be translated into prior information on the failure distribution parameters. Bayesian inference concerning competing risk models involving Weibull distributions has been considered by several authors including Berger and Sun (1993), Bacha *et al.* (1998) and Basu *et al.* (2003).

In Bayesian inference, a prior probability distribution $\pi(\theta)$ is specified for the parameter to be estimated and leads to the posterior distribution $\pi(\theta|\mathbf{y}) \propto L(\theta|\mathbf{y})\pi(\theta)$ from which the inference is based. Approximating the posterior distribution of the parameters $\theta = (\eta_0, \eta_1, \beta)$ of a $\mathcal{B}$ distribution or more generally of a competing risk model is a difficult task for which Monte Carlo approximation using Monte Carlo Markov Chains (MCMC) methods (a good reference on MCMC methods is Robert and Casella, 1999) as Gibbs sampling or Hasting-Metropolis algorithm can be used (see Berger and Sun, 1993). However, It has been noticed in Bacha *et al.* (1998) that MCMC methods can encounter prohibitively slow convergence situation especially when there is a small amount of observed failure times and it can present some advantages to use importance sampling techniques instead. This is the approach considered in the present article.

Importance sampling (see Robert and Casella 1999, chapter 3) is based on the simulation of θ_i's $(i = 1, \ldots, M)$ from an instrumental distribution $\rho(\theta)$. The difference between the distribution of interest $\pi(\theta|\mathbf{y})$ and the instrumental distribution $\rho(\theta)$ is corrected using importance weights

$$\omega_i = \frac{\pi(\theta_i|\mathbf{y})/\rho(\theta_i)}{\sum_{j=1}^{M} \pi(\theta_j|\mathbf{y})/\rho(\theta_j)}$$

to preserve that, for any function of interest $h(\theta)$,

$$\sum_{j=1}^{M} \omega_i h(\theta_i) \approx \int h(\theta)\pi(\theta|\mathbf{y})d\theta. \tag{12}$$

The art of importance sampling lies in choosing a good importance function ρ. This choice is paramount to ensure that the convergence to the posterior distribution π occurs at the right rate, the minimum requirement being that the variance of the importance weights ω_i is finite (Robert and Casella 1999, chapter 3). Following Celeux *et al.* (2003), in the spirit of data augmentation methods (Tanner and Wong 1987, Tanner 1991, Robert and Casella 1999), it can be taking profit of the missing data structure of the problem to produce a simple and feasible importance function by simulating missing data. It leads to propose two adaptative importance sampling schemes that are now presented.

The SRE scheme Using the ideas developed in Steele, Raftery and Emond (2003) a first scheme, denoted SRE in the following, is now described. Denoting $L(\theta|\mathbf{y}, \mathbf{z})$ the completed likelihood and $k(\mathbf{z}|\mathbf{y}, \theta) = L(\theta|\mathbf{y}, \mathbf{z})/L(\theta|\mathbf{y})$ being the conditional density of the missing data knowing the observed data,

- Compute $\hat{\theta} = \arg\max_\theta L(\theta|\mathbf{y})$;

- For $j = 1, \ldots, M$

  Generate $\mathbf{z}^{(j)}$ from the conditional distribution of the missing data $k(\mathbf{z}|\mathbf{y}, \hat{\theta})$;

Generate $\theta^{(j)}$ from $\pi\left(\theta|\mathbf{y},\mathbf{z}^{(j)}\right)$;

Compute $r^{(j)} = \frac{L\left(\theta^{(j)}|\mathbf{y},\mathbf{z}^{(j)}\right)\pi\left(\theta^{(j)}\right)}{k\left(\mathbf{z}^{(j)}|\mathbf{y},\hat{\theta}\right)\pi\left(\theta^{(j)}|\mathbf{y},\mathbf{z}^{(j)}\right)}$ and $\omega^{(j)} = r^{(j)}/\sum_{s=1}^{M} r^{(s)}$;

- Generate K realizations $\left(\hat{\theta}^{(1)},\ldots,\hat{\theta}^{(K)}\right)$ from $\left(\theta^{(1)},\ldots,\theta^{(M)}\right)$ using a multinomial distribution with probabilities $\left(\omega^{(1)},\ldots,\omega^{(M)}\right)$.

Even if the proposal distribution is a distribution on $\theta, \mathbf{z}$ given $\mathbf{y}$ and not a marginal distribution on θ given $\mathbf{y}$, the resulting sample $\hat{\theta}^{(1)},\ldots,\hat{\theta}^{(K)}$ can be approximatively regarded as a sample from the posterior distribution $\pi(\theta|\mathbf{y})$ and a standard estimator of $E_\pi(h(\theta))$ is $\sum_{j=1}^{M}\omega^{(j)}h\left(\theta^{(j)}\right)$.

As Tanner (1991), the authors of this scheme justify it by saying that $\pi\left(\mathbf{z}|\mathbf{y},\hat{\theta}\right)$ is a good surrogate for the predictive density $\pi(\mathbf{z}|\mathbf{y})$. A possible problem of this scheme is that, typically, Bayesian inference can be useful in small sample setting for which maximum likelihood can provide unreliable estimates (see Bacha *et al.* 1998). Thus in such cases it is doubtful that initiating the sampling scheme from $\hat{\theta}$ is a good choice. A more Bayesian criticism is that this choice does not take into account the intrinsic variability in θ due to the prior distribution, so the importance function could be too concentrated.

The Population Monte Carlo scheme The Population Monte Carlo scheme proposed by Cappé *et al.* (2004) is an iterated scheme that produces, at each iteration, a sample approximately simulated from $\pi(\theta|\mathbf{y})$ and some approximately unbiased estimators of integrals under that distribution. The novelty of the method is that the iterated call to importance sampling based on the current importance sampling sample allows for a progressive selection of the most relevant points of the sample. The most general version of this procedure is as follows:

- For $j = 1,\ldots,M$, choice of $\theta_0^{(j)}$;
- Step i. $(i = 1,\ldots)$

a) For $j = 1,\ldots,M$

Generate $\theta_i^{(j)}$ from $q_{ij}\left(\theta|\theta_{i-1}\right)$ $(\theta_{i-1} = \left(\theta_{i-1}^{(1)},\ldots,\theta_{i-1}^{(M)}\right))$;

Compute $r^{(j)} = \frac{f\left(y|\theta_i^{(j)}\right)\pi\left(\theta_i^{(j)}\right)}{q_{ij}\left(\theta_i^{(j)}|\theta_{i-1}\right)}$ and $\omega^{(j)} = r^{(j)}/\sum_{s=1}^{M} r^{(s)}$.

By an importance sampling argument (see Cappé *et al.*, 2004), the choice of q_{ij} is fairly unrestricted and this proposal distribution can depend on the previous sample or even on the whole sequence of samples simulated so far.

A specific version of this procedure, denoted PMCH in the following, which takes profit of the missing data structure of the problem is as follows:

- For $j = 1, \ldots, M$, choice of $\theta_0^{(j)}$;
- Step i. $(i = 1, \ldots)$

a) For $j = 1, \ldots, M$

Generate $\mathbf{z}^{(j)}$ from $k\left(\mathbf{z}|\mathbf{y}, \theta_{i-1}^{(j)}\right)$;

Generate $\theta_i^{(j)}$ from $\pi\left(\theta|\mathbf{y}, \mathbf{z}^{(j)}\right)$;

Compute $r^{(j)} = \frac{L\left(\theta_i^{(j)}|\mathbf{y},\mathbf{z}^{(j)}\right)\pi\left(\theta_i^{(j)}\right)}{k\left(\mathbf{z}^{(j)}|\mathbf{y},\theta_{i-1}^{(j)}\right)\pi\left(\theta_i^{(j)}|\mathbf{y},\mathbf{z}^{(j)}\right)}$ and $\omega^{(j)} = r^{(j)} / \sum_{s=1}^{M} r^{(s)}$;

b) Resample the $\left(\theta_i^{(j)}\right)$ using the weights $w^{(j)}$.

Choosing the prior distribution In this article, we are concerned with Bayesian inference in an informative context where experts are expected to be able to give prior information on the parameters (η_0, η_1, β) of the $\mathcal{B}$ distribution. Following Erto (1982), Berger and Sun (1993) and Bacha *et al.* (1998) and owing to our own experience, it is assumed that the shape parameter β is supposed to be in an interval $[\beta_\ell, \beta_r]$. Since, here, we are interested in aging, it is assumed that $\beta_\ell = 1$ and a typical value for β_r is $\beta_r = 5$. The prior density chosen for β is a uniform distribution on this interval. Now, in order to use conjugate Gamma prior distributions for the scale parameters η_0 and η_1, the chosen prior distribution is as follows. Putting $\lambda_0 = 1/\eta_0$ and $\lambda_1 = 1/\eta_1$

$$\pi(\beta, \lambda_1, \lambda_0) = \pi(\beta)\pi(\lambda_1|\beta)\pi(\lambda_0|\beta, \lambda_1) \tag{13}$$

where $\pi(\beta)$ is the uniform distribution on $[\beta_\ell, \beta_r]$,

$$\pi(\lambda_1|\beta) = \frac{\lambda_1^{c\beta-1} d^c \beta \exp\left(-d\lambda_1^\beta\right)}{\Gamma(c)} \mathbf{I}_{]0,+\infty[}(\lambda_1)$$

and

$$\pi(\lambda_0|\beta, \lambda_1) \propto \frac{\lambda_0^{a-1} b^a \exp\left(-b\lambda_0\right)}{\Gamma(a)} \mathbf{I}_{]0,\lambda_1]}(\lambda_0).$$

It means that the prior distribution of λ_0 is a Gamma distribution $\mathcal{G}(a, b)$ truncated in λ_1. And, it can easily be proved that the conditional distribution of λ_1 knowing β is such that the prior distribution of λ_1^β knowing β is a Gamma distribution $\mathcal{G}(c, d)$.

Hyperparameters a, b, c and d are for instance chosen in the following way. Experts are asked to give intervals of possible values for η_0 and η_1. Denoting $[\eta_{0\ell}, \eta_{0r}]$ respectively $[\eta_{1\ell}, \eta_{1r}]$ those intervals, it can lead to

$$\begin{aligned} a &= \alpha(\frac{(1/\eta_{0\ell} + 1/\eta_{0r})^2}{(1/\eta_{0\ell} - 1/\eta_{0r})^2} \\ b &= 2\alpha(\frac{(1/\eta_{0\ell} + 1/\eta_{0r})}{(1/\eta_{0\ell} - 1/\eta_{0r})^2} \end{aligned}$$

where α is chosen small enough to ensure a large variance of the Gamma distribution. And

$$c = \tau(\frac{\left(1/\eta_{1\ell}^{\beta_\ell} + 1/\eta_{1r}^{\beta_r}\right)^2}{\left(1/\eta_{1\ell}^{\beta_\ell} - 1/\eta_{1r}^{\beta_r}\right)^2} \quad (14)$$

$$d = 2\tau(\frac{\left(1/\eta_{1\ell}^{\beta_\ell} + 1/\eta_{1r}^{\beta_r}\right)}{\left(1/\eta_{1\ell}^{\beta_\ell} - 1/\eta_{1r}^{\beta_r}\right)^2} \quad (15)$$

where τ can be chosen in the same manner as α.

Implementation of the SRE and PMCH schemes for the $\mathcal{B}$ distribution The implementation of the two above described importance sampling schemes does not involve difficulties and is not detailed here. However some comments are to be made.

First, it is important to build an importance function ρ with heavier tails than the posterior to be approximated (Robert and Casella 1999, chapter 3). In that purpose, it is beneficial to enlarge the missing data space. Thus, the missing data we considered are not reduced to be the binary vectors z_i indicating the distribution (exponential or Weibull) from which failure times occur. They include the failure time of the alternative distribution (exponential or Weibull) not assigned to the observed failure time, and also the failure times beyond the censoring times for both the exponential and the Weibull distributions. Then, the completed likelihood takes the form

$$L(\eta_0, \eta_1, \beta|\mathbf{y}, \mathbf{z}) = L(\eta_0|e_1, ..., e_n) \times L(\eta_1, \beta|w_1, ..., w_n)$$

where e_i (resp. w_i) are either the observed failure times t_i assigned to the exponential (resp. Weibull) distribution or the simulated failure times according to the exponential distribution $\mathcal{E}(\eta_0)$ (resp. $\mathcal{W}(\eta_1, \beta)$).

Generating $\lambda_0^{(j)}$ and $\lambda_1^{(j)}$ is easy since by conjugate properties, the conditional distribution of λ_0^j is a $\mathcal{G}(a+n, b+\sum_{i=1}^n e_i^{(j)})$ and the conditional distribution of $(\lambda_1^j)^{\beta^{(j)}}$ is a $\mathcal{G}(c+n, d+\sum_{i=1}^n (w_i^{(j)})^{\beta^{(j)}})$.

Since, no conjugate prior exists for the shape parameter of a Weibull distribution, generating $\beta^{(j)}$ is carried out with an accept-reject algorithm:

1. Generate β from $\pi(\beta)$.
2. Generate u from a uniform distribution on $[0, 1]$.
3. Put $\beta^{(j)} = \beta$ if $u \leq \frac{\pi(\beta|\mathbf{y},\mathbf{z}^{(j)},\eta_0^{(j-1)},\eta_1^{(j-1)})}{\pi(\beta)\max_\beta L(\beta,\eta_0^{(j-1)},\eta_1^{(j-1)}|\mathbf{y},\mathbf{z}^{(j)})}$, otherwise goto 1.

4 Assessing the failure distribution

The objective of this section is to give a procedure for choosing the most relevant model among the three following ones, given in the increasing order of complexity : the exponential, Weibull and $\mathcal{B}$ model. In a reliability context, the exponential model is often the first proposed distribution. But, if aging is suspected, the Weibull model can be expected to be more appropriate. As for the $\mathcal{B}$ model, it has already been seen that it is even more accurate than the Weibull model.

Towards that end, the standard Likelihood Ratio Test (denoted LRT) is an adequate goodness-of-fit test, in an asymptotic framework. More precisely, as these three models are embedded, the two-step following procedure is proposed :

- First the standard exponential versus Weibull test is applied (see for example Lawless (1982), p173-174). As a first step, this enables to detect aging. Under the null hypothesis $H_0 : \beta = 1$, the LRT statistic $2 \ln \left(\frac{L_W(\hat{\eta}, \hat{\beta})}{L_E(\hat{\eta_0})} \right) \rightsquigarrow \chi^2(1)$.

- Secondly, to go further in the case where the previous test leads to choose a Weibull model, the LRT is now applied to discriminate between Weibull and $\mathcal{B}$ models. In other words, here, the question is to know whether it is worth taking into account a possible accidental cause of failure. To be more precise, for this second test, denoting $\lambda_0 = 1/\eta_0$, the null hypothesis (corresponding to the Weibull model) is defined as $H_0 \ : \lambda_0 = 0 \ (\eta_0 = +\infty)$ (which means that the accidental component vanishes). Since the parameter λ_0 is constrained to be ≥ 0 under the alternative hypothesis H_1, the distribution of the LRT statistic (with notations of § 3.2)

$$\begin{aligned} 2 \ \ln \left(\frac{L_{\mathcal{B}}(\hat{\lambda_0}, \hat{\lambda_1}, \hat{\beta})}{L_W(\tilde{\lambda}, \tilde{\beta})} \right) \ &= \ 2 \ \left[\sum_{i=1}^n \delta_i \ln \left(\hat{\lambda_0} + \hat{\lambda_1}^{\hat{\beta}} \ x_i^{\hat{\beta}-1} \right) - \sum_{i=1}^n \left(\hat{\lambda_0} \ x_i + \hat{\lambda_1}^{\hat{\beta}} \ x_i^{\hat{\beta}} \right) \right] - \\ & \quad 2 \ \left[\sum_{i=1}^n \delta_i \left(\tilde{\beta} \ln \tilde{\lambda} + \ln \tilde{\beta} \right) + \left(\tilde{\beta} - 1 \right) \sum_{i=1}^n \delta_i \ln x_i - \sum_{i=1}^n \left(\tilde{\lambda}^{\tilde{\beta}} \ x_i^{\tilde{\beta}} \right) \right] \end{aligned} .$$

is $\frac{1}{2}\chi_1^2 + \frac{1}{2} \ \delta_0$, where δ_0 denotes the Dirac delta distribution in 0. A general presentation of this result can be found in Gourieroux and Monfort (1996, chapter 21). Beyond standard regularity conditions that are fulfilled by the $\mathcal{B}$ distribution, their line of proof needs that the Taylor expansion of $\ln L_{\mathcal{B}}$ around $\lambda_0 = 0$ is possible, which is true since $\ln L_{\mathcal{B}}$ is well defined in a neighborhood of $\lambda_0 = 0$.

Thus the application of the LRT is straightforward. Proceeding in this manner, it is possible to examine gradually the complexity of models. Notice that it is also possible to make use of the BIC criterion to select one of the three models. The BIC criterion associated to the model $\mathcal{X}$ with k parameters and the estimation $\hat{\theta}$ can be written $\ln L_{\mathcal{X}}(\hat{\theta}) - \frac{k}{2} \ln n > 0$ where n is the size of the sample. The selected model is the one which presents higher BIC criterion.

5 Numerical experiments

5.1 Estimations and tests on simulated data

In this section estimations of exponential, Weibull and $\mathcal{B}$ distributions on simulated data from a $\mathcal{B}(200, 100, 2)$ distribution are considered. Note that for a $\mathcal{B}(200, 100, 2)$ distribution, the probability for a failure to be accidental is 0.34. Two sample sizes, $n = 50$ and $n = 500$, and two censoring times, 200 (about 0% of censored data) and 100 (about 20% of censored data) have been considered. For each of the four situations a sample has been simulated.

In next tables "ML" denomination embodies the use of EM algorithm for the $\mathcal{B}$ distribution, a Newton-Raphson descent algorithm applied to the Weibull distribution and the direct likelihood maximisation for the exponential distribution. In the Bayesian setting, prior distributions have been designed from the equations given in Section 3.3 from the following prior informations. The shape parameter β is in $[\beta_l, \beta_r] = [1, 5]$, the scale parameter η_0 of the exponential component is concentrated on $[\eta_{0l}, \eta_{0r}] = [1, 300]$ and the scale parameter η_1 of the Weibull component is concentrated on $[\eta_{1l}, \eta_{1r}] = [1, 200]$.

		ML	PMCH
Exponential	η	67.21	71.40
Weibull	η	73.80	78.56
	β	1.74	1.69
	η_0	224.60	210.87
$\mathcal{B}$ **model**	η_1	98.30	101.60
	β	2.02	2.04

Table 2: ML and Bayesian estimations of a 500-sized $\mathcal{B}(200, 100, 2)$ sample (0% of censored data).

		ML	PMCH
Exponential	η	83.38	82.40
Weibull	η	75.20	83.56
	β	1.81	1.59
	η_0	565.45	258.45
$\mathcal{B}$ **model**	η_1	76.78	82.34
	β	1.72	1.89

Table 3: ML and Bayesian estimations of a 500-sized $\mathcal{B}(200, 100, 2)$ sample (20% of censored data).

		ML	PMCH
Exponential	η	70.05	70.12
Weibull	η	78.10	83.10
	β	1.71	1.65
	η_0	355.20	245.31
$\mathcal{B}$ **model**	η_1	90.73	94.78
	β	2.05	2.08

Table 4: ML and Bayesian estimations of a 50-sized $\mathcal{B}(200, 100, 2)$ sample (0% of censored data).

		ML	PMCH
Exponential	η	81.15	79.56
Weibull	η	75.89	81.23
	β	1.84	1.71
	η_0	$\sim 10^{25}$	$\sim 10^{12}$
$\mathcal{B}$ **model**	η_1	81.54	82.30
	β	1.69	1.81

Table 5: ML and Bayesian estimations of a 50-sized $\mathcal{B}(200, 100, 2)$ sample (20% of censored data).

It is to be noticed that the estimation of parameter η_0 seems to be more sensitive to censoring time and sample size than the two other ones. But the model selection procedure allows to prefer the right distribution in most cases, as well by LRT test as by BIC criterion (5%-level tests presented on Tables 6 and 7). In the small sample size and severely censored case, the model choice appears to be obvious since the estimation of η_0 in $\mathcal{B}$ modelling tends to infinity, so that the $\mathcal{B}$ distribution is well approximated with a Weibull distribution. On Figure 2 the theoretical and estimated pdf's are displayed.

Censoring percentage	**Exponential**	**Weibull**	$\mathcal{B}$ **model**	Model Choice
0%	-2610.66	-2540.04	-2525.69	$\mathcal{B}$ *model*
20%	-2172.69	-2155.14	-2150.18	$\mathcal{B}$ *model*

Table 6: Log-likelihood values and model choice for samples of size 500 (5%-level tests).

Censoring percentage	**Exponential**	**Weibull**	$\mathcal{B}$ **model**	Model Choice
0%	-262.86	-250.89	-248.95	$\mathcal{B}$ *model*
20%	-287.69	-264.56	-279.21	*Weibull*

Table 7: Log-likelihood values and model choice for samples of size 50 (5%-level tests).

5.2 Estimations and tests on real data

The following data are sampled from a (prospective) table of mortality given by INSEE[1] for $100,000$ people born in 1993. This mortality table was presented in Bertholon, (2001). In this work the EM estimation of the $\mathcal{B}$ distribution parameters were $\hat{\eta_0} = 1662$ years, $\hat{\eta_1} = 97.4$ years, and $\hat{\beta} = 12.4$. The value of shape parameter β is unusual. It indicates that aging is accelerating after some time, which is not really surprising concerning human life. If aging is only taken into account (as it is generally the case) with a Weibull distribution the mean lifetime remains close to 97 years for people born in 1993. Note that according to these results the mean lifetime should be 1662 years if humans would not be submitted to aging !

From these data two sets of size 1,000 and 100 respectively are randomly sampled, for estimating the parameters of the three considered lifetime models and analyzing the results of hypothesis tests. Both samples are considered uncensored. Because of relatively large sample sizes ($n \geq 100$), maximum likelihood was used to estimate the parameters of the models. The estimates are presented in Tables 8 and 9.

From Figure 3, it clearly appears that the $\mathcal{B}$ distribution provides a better fit to the data than the Weibull distribution. Neglecting the accidental cause of dying produces an unre-

[1] National Institute for Statistics and Economic Studies, 195, rue de Bercy - Tour Gamma A - 75582 Paris cedex 12, France *http://www.insee.fr*

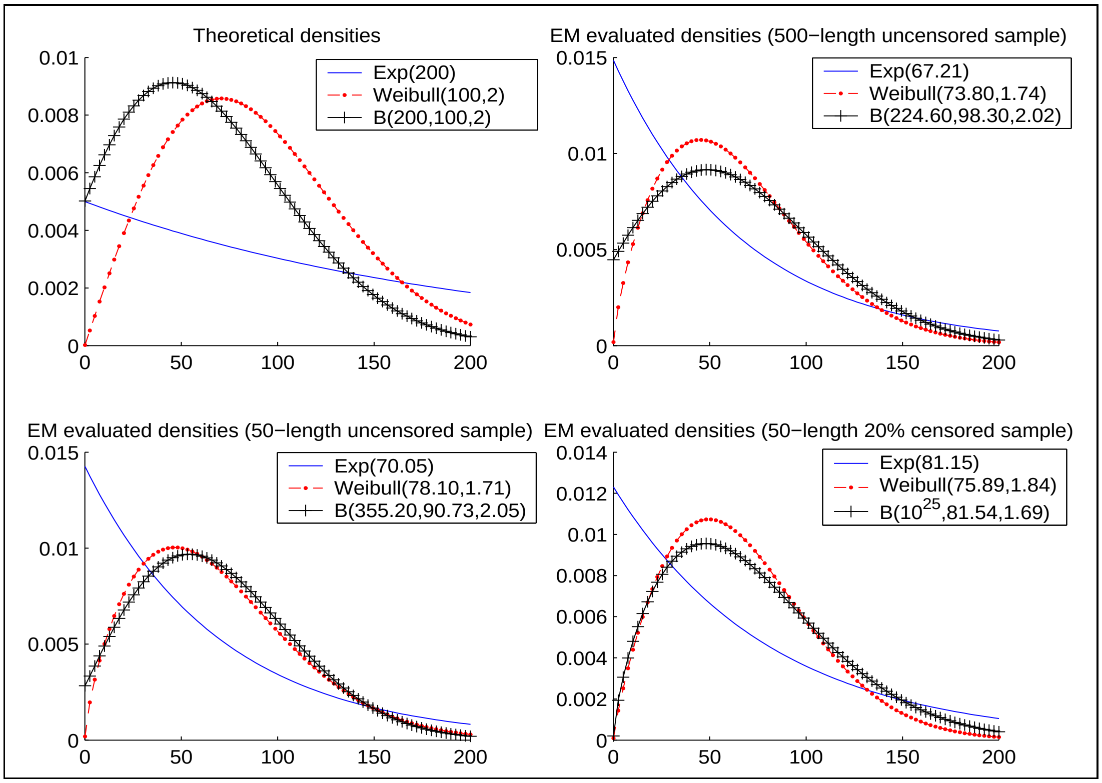

Figure 2: Densities of lifetime distributions obtained with ml estimated parameters with data from a $\mathcal{B}(200, 100, 2)$ distribution.

		EM
Exponential	η	83.35
Weibull	η	87.47
	β	7.49
$\mathcal{B}$ **model**	η_0	685.41
	η_1	90.18
	β	21.40

Table 8: Estimations with a sample of size 1,000 coming from the mortality table (5%-level tests).

		EM
Exponential	η	84.36
Weibull	η	88.81
	β	10.18
$\mathcal{B}$ **model**	η_0	740.94
	η_1	90.80
	β	18.77

Table 9: Estimations with a sample of size 100 coming from the mortality table (5%-level tests).

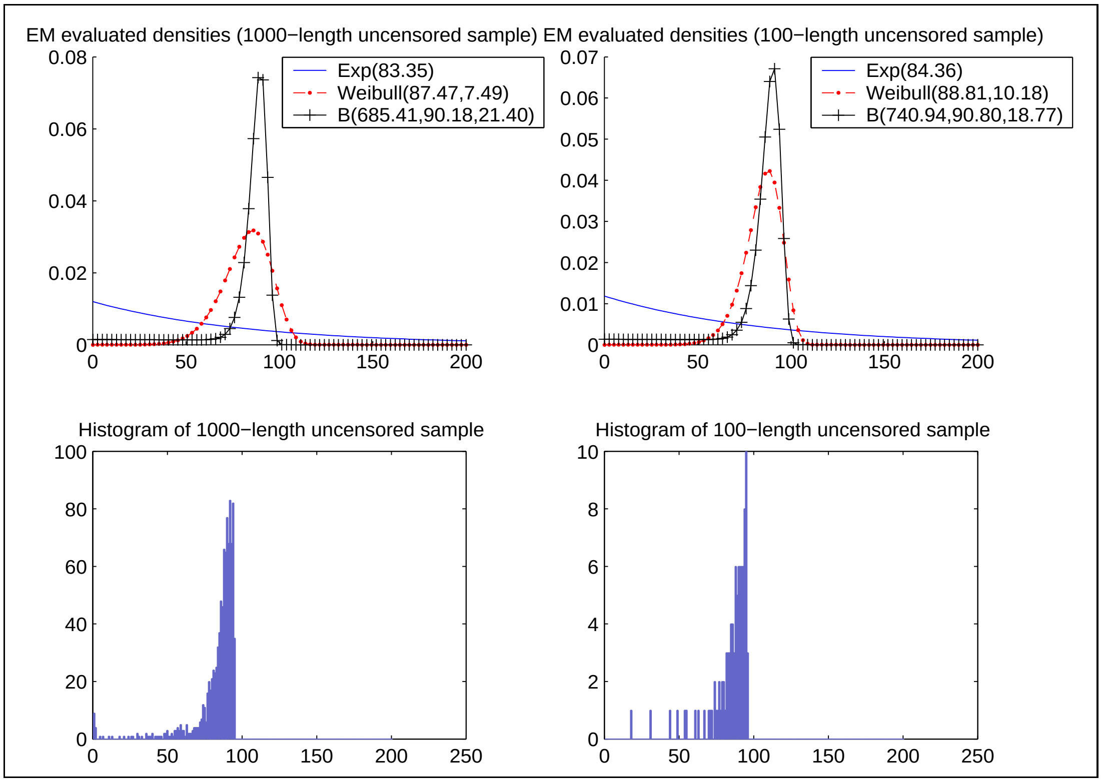

Figure 3: Densities of lifetime distributions obtained with ml estimated parameters from the prospective mortality table of INSEE

liable analysis of prospective human lifetime. Extensive differences of estimated reliability for the Weibull and the $\mathcal{B}$ distributions appear in Figure 5.

Censoring percentage	**Exponential**	**Weibull**	$\mathcal{B}$ **model**	Model Choice
0%	-5423.14	-4164.77	-3477.28	$\mathcal{B}$ *model*

Table 10: Log-likelihood values and model choice for samples of size 1,000 from the mortality table (5%-level tests)

Censoring percentage	**Exponential**	**Weibull**	$\mathcal{B}$ **model**	Model Choice
0%	-543.51	-382.54	-356.86	$\mathcal{B}$ *model*

Table 11: Log-likelihood values and model choice for samples of size 100 from the mortality table (5%-level tests)

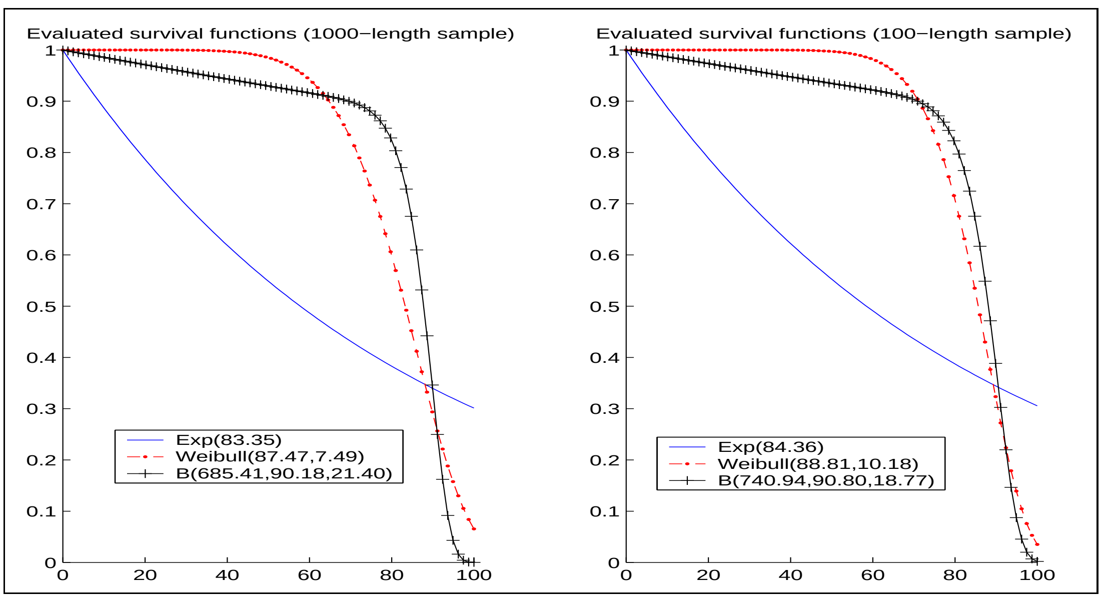

Figure 4: Survival functions obtained with ml estimated parameters from the prospective mortality table of INSEE

6 Discussion

The $\mathcal{B}$ distribution is a natural and simple model to analyse various competing risks of death or failure. In the present paper it has been considered as an alternative to the Weibull distribution for modelling aging.

In some circumstances the $\mathcal{B}$ distribution can be thought of as a more realistic model to describe aging that the Weibull distribution. And a clear strategy of testing has been designed to choose between exponential, Weibull and $\mathcal{B}$ distributions.

The versatile Weibull distribution notably appears to have difficulty in taking into account all aspects of lifetime data. An illustration is highlighted by the life table data. Figure 5 shows that the best estimated Weibull distribution is unable to fit the empirical survival function given by the entire table. On the contrary, the $\mathcal{B}$ distribution gives a perfect adjustment to this curve.

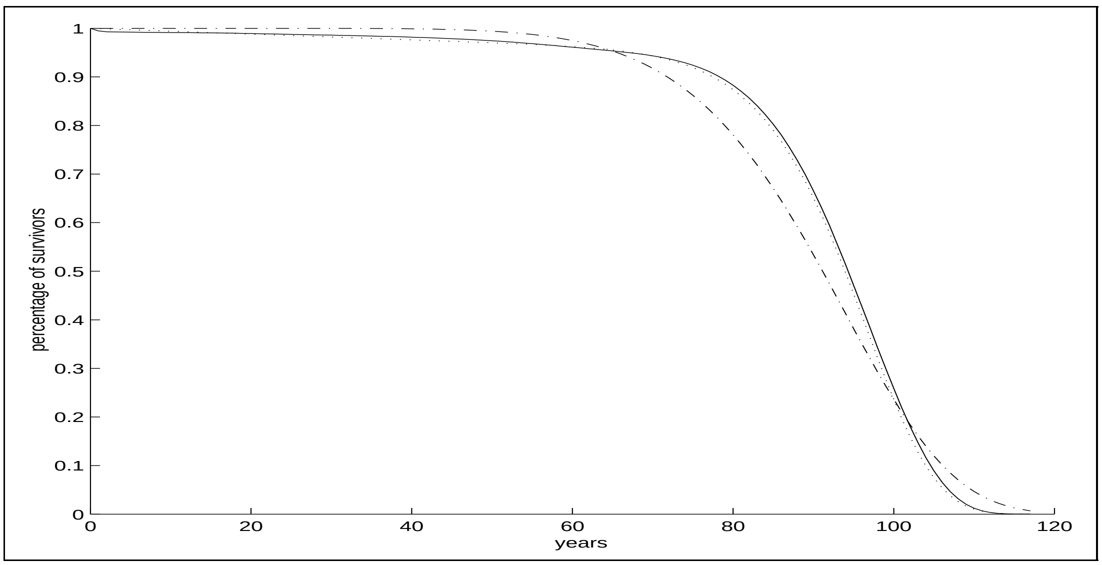

Figure 5: Survival functions for the 100,000-sized life table [*full line: empirical survival function; dotted line: B estimated survival function; dashed line: Weibull estimated survival function*].

Nevertheless a possible drawback of this model is that it involves three parameters. Indeed it can be difficult to estimate those parameters for small-sized and/or censored samples. For this very reason, Bayesian inference is desirable to get reliable estimations. We have proposed an efficient importance sampling method to approximate the posterior and the $\mathcal{B}$ distributions. One of our research perspectives is to design simple and reliable elicitation procedures for taking into account prior information regarding the $\mathcal{B}$ distribution in a proper way.

References

d'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-Fit Techniques.* New York: Marcel Dekker.

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

Basu, S., Sen A. and Banerjee, M. (2003). Bayesian analysis of competing risks with partially masked cause of failure. *Applied Statistics*, **52**, 77-93.

Berger, J. and Sun, D.(1993). Bayesian analysis for the Poly-Weibull distribution. *Journal of the American Statistical Association*, **88**, 1412-1418.

Bertholon, H. (2001). *Une modélisation du vieillissement.* Ph D. Thesis, Université Joseph Fourier, Grenoble.

Cappé, O., Guillin, A., Marin, J.-M. and Robert, C. P. (2004). Population Monte Carlo. *Journal of Computational and Graphical Statistics.* (Forthcoming)

Celeux, G., Marin, J.-M. and Robert, C. P. (2003). Iterated Importance sampling in Missing Data Problems. *Cahiers du Ceremade* 326.

Chanda, K. C. (1954). A note on the Consistency and Maxima of the Roots of Likelihood Equations. *Biometrika*, **41**, 56-61.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal. Roy. Statist. Soc. (Ser. B)*, **39**, 1-38.

Erto, P. (1982). New practical Bayes estimators for the 2-parameter Weibull distribution. *IEEE Transactions on Reliability*, **31**, 194-197.

Gourerioux, Ch., Monfort A. (1996). Statistique et modèles économétriques Vol.2. Paris: Economica.

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data.* Wiley series in probability and mathematical statistics.

Lehmann, E.L. (1986). Testing Statistical Hypotheses. (second edition). New York: John Wiley and Sons

McLachlan, G. J. and Krishnam, T. (1997). *The EM algorithm and Extensions.* New York: Wiley.

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

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods.* New York: Springer-Verlag.

Steele, R., Raftery, A. E. and Emond, M. (2003). "Computing Normalizing Constants for Finite Mixture Models via Incremental Mixture Importance Sampling (IMIS)" Technical Report no. 436, Department of Statistics, University of Washington.

Tanner, M. A. and Wong, W. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal Amer. Statis. Assoc.*, **82**, 528-550.

Tanner, M. A. (1991). *Tools for Statistical Inference, Observed Data and Data Augmentation Methods.* Lectures Notes in Statistics, Springer-Verlag.

Unité de recherche INRIA Futurs
Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399