A semiparametric Bayesian mixed-effects model
for failure time data

RAFFAELE ARGIENTO[a], ALESSANDRA GUGLIELMI[b,a] and ANTONIO PIEVATOLO[a]

IMATI-CNR[a] and Politecnico di Milano[b]

Abstract

We propose a Bayesian semiparametric accelerated failure time mixed-effects model with an illustrative application to a Kevlar fibre lifetime dataset (with censoring). The error is a shape-scale mixture of Weibull densities, mixed by a normalized generalized gamma random measure, encompassing the Dirichlet process. We implement an MCMC scheme, obtaining posterior credibility intervals for the predictive distributions and for the quantiles of the failure times under different stress levels. Random spool effects are taken up by the nonparametric mixture, where every component accounts for a different spool. Compared to a previous parametric Bayesian analysis, we obtain narrower credibility intervals and a better fit to the data. We also fit a similar semiparametric model, which can be seen a a special case of ours, where the error is a scale mixture of Weibull densities, mixed by a Dirichlet process, whereas the shape parameter has a parametric prior. The adequacy of the two semiparametric models is comparable, but the more general model provides a different evaluation of the posterior variance of the left tail of the lifetime distribution, which is an effect of assuming a nonparametric prior on both parameters of the Weibull density.

Keywords: Accelerated failure time regression models, Mixed-effects models, Bayesian semiparametrics, Mixture models, MCMC algorithm.

AMS 2000 Mathematics Subject Classification: 62F15, 62N01, 62N05.

1 Introduction

We will present a Bayesian semiparametric approach for an Accelerated Failure Time (AFT) model for censored univariate data, on the basis of an application involving pressure vessels, which are critical components of the Space Shuttle. In the survival/lifetime literature, the

accelerated failure time model is usually meant as the multiplicative effect of a fixed p-vector of covariates $x = (x_1, \ldots, x_p)'$ on the failure time T, *i.e.* $\log T = x'\beta + W$, where $\beta = (\beta_1, \ldots, \beta_p)'$ is the vector of regression parameters and W denotes the error. The AFT model is a valuable alternative to the Cox Proportional Hazard model, although far less used. Recently this model has received much attention in the Bayesian community, in particular in papers where the error W (or $\exp(W)$) has been represented hierarchically as a mixture of parametric densities with a Dirichlet process as mixing measure, *i.e.*, the well-known Dirichlet process mixture (DPM) models, introduced by Lo (1984). In Kuo and Mallick (1997), the authors model $\exp(W)$ as a Dirichlet process location mixture of normals. Kottas and Gelfand (2001) and Gelfand and Kottas (2003) propose a flexible semiparametric class of zero median distributions for the error, which essentially consists of a Dirichlet process scale mixture of split 0-mean normals (with skewness handled parametrically). In Ghosh and Ghosal (2006) the distribution of $\exp(W)$ is given as a scale mixture of Weibull distributions with a Dirichlet process as mixing measure, and the consistency of the posterior distribution of the regression parameters is established. Hanson (2006) proposes a DPM model of gamma densities, mixing over both on the scale and the shape of the gammas, for the distribution of $\exp(W)$. Argiento et al. (2009) compare models in the same setting when the mixing measure is either a Dirichlet process or a normalized inverse Gaussian process.

A different (and recent) stream of research considers dependent nonparametric priors, by which is meant that covariates affect the response variable through the mixing measure, while they do not enter the parametric density directly. For an example of this approach applied to survival times and bibliographical references see Jara et al. (2010).

The application considered here is based on a dataset of 108 lifetimes of pressure vessels wrapped with a commercial fibre called Kevlar (originally published by Gerstle and Kunz (1983), obtained from a series of accelerated life tests; the fibre comes from eight different spools and four levels of stress (pressure) are used. Eleven lifetimes with the lowest level of stress are administratively censored at 41000 hours. Crowder et al. (1991) present a frequentist analysis of the data and fit an AFT model with both stress and spool as fixed effects, using Weibull distributions to model survival times. Leon et al. (2007) take a Bayesian parametric approach to the problem, still using Weibull survival times, however considering a mixed-effects model, where the stress and intercept parameters are a priori independent from each other and from the spool parameters, and the spool parameters are exchangeable. This choice follows from the finding (common to all the cited references) that spools have a significant effect on the failure time, so that it is necessary to have a model for predicting the failure time when a new spool is selected at random from the population of spools. However, their interval estimates of two quantiles of interest are too large to make statements about

the reliability of the Space Shuttle.

To overcome this lack of predictive capability and to allow for a greater degree of model flexibility we impose a nonparametric hierarchical mixture on the error term. We will see that a consequence of this assumption is a new representation of the exchangeable spool effects in the model as mixture components. Moreover, this model falls into the family of frailty models where the number of groups is unknown. We consider then $T = \exp(x'\beta) V, \ V := e^W$, where the error distribution is represented as a nonparametric hierarchical mixture of Weibull distributions on both the shape and the scale parameters. The mixing measure G is assumed to be random, namely G is a *normalized generalized gamma random measure*, indexed by two parameters (σ, η), controlling the "amount" of mass the distribution of G puts on the mean distribution G_0. The Dirichlet process is contained within this family, for $\sigma = 0$. The Bayesian semiparametric approach makes it possible to draw inference on quantities lying in functional spaces (such as the predictive distribution of the failure time of a vessel under a given stress condition); in particular, the normalized generalized gamma random measure has a characterization in terms of Poisson processes, which allows the direct MCMC simulation of trajectories of probability distributions. This is particularly useful for our analysis, where we aim to determine point and interval estimates of predictive failure time distributions and of their quantiles. A distinctive feature of our modelling with respect to the parametric Bayesian mixed-effects model is that the grouping of observations is not fixed (as dictated by the spool number), but is random and is inferred from the data. This information however is not lost, but is included in the prior distribution via the hyperparameters. Therefore the membership of a fibre to a group is modelled nonparametrically, thanks to the clustering property of the discrete random probability measure G. This will yield an interpretation of our model in terms of Bayesian mixed effects. Among the papers on mixed-effects models for survival data, we refer to Zeger and Karim (1991), where the random-effects models in a Bayesian context were introduced, Sun et al. (2000), and two recent papers by Cai and Dunson (2008) and Kinney and Dunson (2008), respectively.

Compared to previous analyses, we find that our nonparametric mixture model for the error better follow the log-linear relationship between failure times and covariates. Finally, the interval estimates obtained with our semiparametric shape-scale mixture error model are much narrower than those under the parametric mixed-effects model. A similar study of this dataset appeared as Argiento et al. (2010b). Here we also fit another semiparametric model, which can be seen a a special case of ours, where the error is a scale mixture of Weibull densities, mixed by a Dirichlet process, whereas the shape parameter has a parametric prior. The adequacy of the two semiparametric models is comparable, but the more general model provides a different evaluation of the posterior variance of the left tail of the lifetime

distribution, which is an effect of assuming a nonparametric prior on both parameters of the Weibull density.

The paper is organized as follows. Section 2 presents the AFT model used so far to analyse the dataset considered. In Section 3 we introduce our semiparametric AFT model and give the structure of the nonparametric hierarchical mixture prior for the error term. Computational algorithms are discussed in Section 4 and the application to Kevlar fibres is presented in Section 5. Conclusions and comments are given in Section 6.

2 Accelerated life models for Kevlar fibre life data

Crowder et al. (1991) consider 108 Kevlar fibre lifetimes, coming from the combination of eight different spools and four levels of stress (pressure), and fit an AFT model with both stress and spool as fixed effects. We can describe the model equivalently both in additive and multiplicative form:

$$(1) \qquad \log T = x'\beta + \frac{W}{\vartheta_1}, \quad W \sim \mathrm{Gumbel}(0,1)$$

or

$$(2) \qquad T = \mathrm{e}^{x'\beta} \cdot V, \quad V = U^{\frac{1}{\vartheta_1}} \sim \mathrm{Weibull}(\vartheta_1,1) \quad [U \sim \mathrm{gamma}(1,1)]$$

where

$$x'\beta = \beta_0 + \beta_1 x_1 + \sum_{j=2}^{8} \beta_j x_j$$

and

$$x_1 = \log(\text{stress}), \quad x_j = \text{effect of spool j (binary)}, \quad j \geq 2 \ .$$

In both specifications we have written the error term as a random variable with a standard distribution, so that the role of ϑ_1 as a scale parameter in model (1) or as a shape parameter in model (2) is evident. The survival function of W is $\mathrm{e}^{-\mathrm{e}^{w}}$, corresponding to the Gumbel distribution of the smallest extreme, with $\mathbb{E}(W) = -\gamma$ (minus the Euler constant) and $\mathrm{Var}(W) = \pi^2/6$. The survival function of U is e^{-u}, with unit mean and variance.

Crowder et al. (1991) find that the spool effect is very significant in the model and obtain an acceptable fit as far as the plot of residuals are concerned, but with a less satisfactory performance for the lowest stress level. In a recent paper, Leon et al. (2007) argue that the fixed-effects model does not allow to make inference on vessels wrapped with fibre taken from a new spool. Then they fit a Bayesian mixed-effects model and also a Bayesian fixed-effects model, the latter mainly for comparison with the frequentist model.

We will find it convenient to have the Bayesian mixed-effects model in the additive form:

$$(3) \qquad \log T = \beta_0 + \beta_1 x_1 + \sum_{j=1}^{8} \gamma_j s_j + \frac{W}{\vartheta_1}, \quad W \sim \text{Gumbel}(0,1)$$

where $s_j = 1$ if spool j is used (and zero otherwise), and $(\gamma_1, \dots, \gamma_8)$ are the exchangeable random-effects parameters. According to Gelman et al. (2004), the *Bayesian random-effects models* are meant as regression models in which groups of the regression coefficients are exchangeable; on one hand, the simple random-effects model is obtained if all the regression parameters are assumed exchangeable, while we get a mixed-effects model if only a subset of regression parameters are exchangeable and the rest are assigned independent prior distributions with "large" variances (the latter are labelled *fixed effects* in this model). Specifically, Leon et al. (2007) take $(\gamma_1, \dots, \gamma_8)$ to be conditionally i.i.d. Gaussian with zero mean and variance σ^2, which is given an inverse gamma prior. For a pressure vessel wrapped with fibre from a new random spool (not necessarily included in the dataset), the authors calculate point and interval estimates for the first percentile of the failure time distribution when stress is 23.4 MPa (MegaPascal), the lowest value in the dataset, and for the median with stress 22.5 MPa. The latter stress was chosen to represent a stress lower than those in the experiment, but still close enough to provide reasonable estimates. However, the resulting prediction intervals were too wide to make statements about the reliability of the Space Shuttle.

3 The Bayesian nonparametric AFT model

3.1 Normalized generalized gamma process priors

In order to obtain useful prediction intervals for the two quantiles considered by Leon et al. (2007), we model the error term V, in the AFT model $T = e^{x'\beta} \cdot V$, as a nonparametric mixture of parametric densities mixed by a (a.s.) discrete random probability G, namely G is a normalized generalized gamma process. Let Θ be a Borel subset of $\mathbb{R}^s$ for some positive integer s, with its Borel σ-algebra $\mathcal{B}(\Theta)$ and let G_0 be a probability measure on Θ. We say that G is a normalized generalized gamma random probability measure on Θ, and we write $G \sim NGG(\sigma, \kappa\, G_0(\cdot), \omega)$, where $0 \leq \sigma \leq 1$, $\omega \geq 0$, $\kappa \geq 0$, if G can be written as

$$(4) \qquad G = \sum_{i=1}^{+\infty} P_i \delta_{X_i} = \sum_{i=1}^{+\infty} \frac{J_i}{T} \delta_{X_i} \ ,$$

where $P_i := \frac{J_i}{T}$ for any $i = 1, 2, \dots$, $T := \sum_i J_i$, and $J_1 \geq J_2 \geq \dots$ are the ranked values of points in a Poisson process on $(0, +\infty)$ with intensity

$$\rho(ds) = \frac{\kappa}{\Gamma(1-\sigma)} s^{-\sigma-1} e^{-\omega s} \mathbb{I}_{(0,+\infty)}(s)\, ds \ .$$

Moreover, the sequences $(P_i)_i$ and $(X_i)_i$ are independent, $(X_i)_i$ being i.i.d. from G_0. This process includes several well-known stochastic processes, namely the Dirichlet process if $\sigma = 0$, the Normalized Inverse Gaussian process if $\sigma = 1/2$ Lijoi, Mena and Pruenster (2005), the two parameter Poisson-Dirichlet process Pitman and Yor (1997) if $\omega = 0$ and $0 < \sigma < 1$, and it degenerates on G_0 if $\sigma = 1$. See Pitman (1996), Pitman (2003).

Generally, the finite dimensional distributions of P are not available in closed analytic form. However, the distribution G_0 functions as mean distribution of the process, because, for any set B,

$$\mathbb{E}\left(G(B)\right) = G_0(B) \quad \text{and} \quad \text{Var}\left(G(B)\right) = G_0(B)(1 - G_0(B))\mathcal{I}(\sigma, \kappa, \omega)$$

where

$$\mathcal{I}(\sigma, \kappa, \omega) := (1 - \sigma)\left(1 - (\frac{\kappa\omega^\sigma}{\sigma})^{1/\sigma} \exp(\frac{\kappa\omega^\sigma}{\sigma})\Gamma\left(-\frac{1}{\sigma} + 1, \frac{\kappa\omega^\sigma}{\sigma}\right)\right);$$

see James et al. (2006). The factor in the variance depends on κ and ω only through $\eta := \kappa\omega^\sigma/\sigma$, so that we rewrite it as

$$\mathcal{I} = \mathcal{I}(\sigma, \eta) = \left(\frac{1}{\sigma} - 1\right)\eta^{1/\sigma}e^\eta\Gamma\left(-\frac{1}{\sigma}, \eta\right) \ .$$

It can be shown that $\mathcal{I}(\sigma, \eta)$ is a decreasing function of σ for fixed $\eta > 0$ and a decreasing function of η for fixed $\sigma \in (0, 1)$.

This reparameterization can be used for the entire process and not only for an easier assessment of the relationship between the parameter and the variance, thanks to a scaling property of the NGG process, by which $(\sigma, \kappa G_0(\cdot), \omega)$ and $(\sigma, s^\sigma \kappa G_0(\cdot), \omega/s)$ (for any $s > 0$) yield the same distribution for G. This means that we may let $\omega = 1$ and obtain the same class of stochastic processes indexed by the triplet (σ, η, G_0). Thereby we write $G \sim NGG(\sigma, \eta, G_0)$. We remark that triplets (σ, κ, ω) are not in a one-to-one correspondence with pairs (σ, η). However, by the scaling property, (σ, κ, ω) is in the same equivalence class of the triplet $(\sigma, \kappa\omega^\sigma, 1)$, which corresponds one-to-one with the pair (σ, η), when $\eta = \kappa\omega^\sigma/\sigma$.

3.2 The AFT model

We can now specify with some generality our semiparametric AFT model. We will not explicitly include the spools as random-effects parameters in the model as in (3); here the grouping of observations is not fixed (as dictated by the spool number), but is random and is inferred from the data. This will yield a more flexible error distribution; see the next subsection for more thorough discussion. Let $T_1, \ldots, T_n$ denote survival times, and let x_{1i} be the corresponding stress in log-scale, as $i = 1, \ldots, n$. Then we have, hierarchically, for

$i = 1, \ldots, n,$

$$T_i = \mathrm{e}^{\beta_0 + \beta_1 x_{1i}} \cdot V_i$$

$$V_i | \theta_i \overset{ind}{\sim} k(\cdot\,; \theta_i), \quad \text{a parametric family of densities on } \mathbb{R}^+, \quad \Theta \subset \mathbb{R}^s$$

(5)
$$\theta_i | G \overset{iid}{\sim} G,$$

$$G \sim NGG(\sigma, \eta, G_0)$$

$$(\beta_0, \beta_1) \sim \pi, \quad (\beta_0, \beta_1) \perp G.$$

With this specification, the hyperparameters are G_0, η, σ and those within the prior of (β_0, β_1). Observe that, given G, the error term V is distributed as a nonparametric mixture with density

(6)
$$f(v; G) = \int_\Theta k(v; \theta) G(d\theta)$$

which exists if $k(\cdot\,; \cdot)$ is a kernel, even though the NGG process produces discrete distributions only.

We can obtain a wide class of models by choosing the kernel density in the mixture. Here we assume the following Weibull kernel with associated survival function

(7) $\qquad \bar{K}(v; \theta) = \bar{K}(v; \vartheta_1, \vartheta_2) = \mathrm{e}^{-\left(\frac{v}{\vartheta_2}\right)^{\vartheta_1}}, \quad v > 0, \qquad \theta = (\vartheta_1, \vartheta_2), \qquad \vartheta_1, \vartheta_2 > 0\,,$

where both ϑ_1, ϑ_2 are random.

3.3 Random effects interpretation

We may draw an interesting analogy between the Bayesian mixed-effects model (3) and model (5) under (7) when ϑ_1 and ϑ_2 are both unknown and random, as we have assumed. Indeed, from (5), $(\theta_1, \ldots, \theta_n)$ are exchangeable and, conditionally on them and on β, it is easy to see that, for each $i = 1, \ldots, n$,

(8) $\qquad \log T_i = \beta_0 + \beta_1 x_{1i} + \log \vartheta_{2,i} + \dfrac{W_i}{\vartheta_{1i}}, \quad W_i \overset{i.i.d.}{\sim} \text{Gumbel}(0, 1)\,.$

Therefore, the terms $(\log \vartheta_{2,i})$ hold the same place of the γ_j's in model (3) considered by Leon et al. (2007), with the difference that here the number of distinct parameters is random and can vary between one and n, because of the ties induced by the discreteness of G. In this way we seem to have a twofold advantage over the Bayesian mixed-effects model: the distribution of the error term is more flexible due to the nonparametric structure and we need not fix the number of random-effects parameters in advance, because they are inferred along with the other unknown quantities, thanks to the discreteness of G. The prior information about

the actual number of spools can be incorporated into the model through (σ, η), since these hyperparameters induce a prior distribution on the number of clusters in (6). A sensitivity analysis can then follow by varying (σ, η) on a finite grid. When no such prior information on the number of components in the mixture is available, (σ, η) could be selected based on a Bayes factor, as done by Argiento et al. (2010) in the context of mixture density estimation. Moreover, under our approach, the the predictive distribution of the life T_{109} of a new pressure vessel wrapped with fibre from a random spool, subject to a log-stress x, can be computed as:

$$(9) \quad \mathbb{P}(T_{109} > t \mid \text{data}, x) = \int_{\mathbb{R}^2 \times \mathcal{P}} \left(\int_{\Theta} \exp\left(-\left(\frac{t}{\lambda_x} \right)^{\vartheta_1} \right) G(d\theta) \right) \mathcal{L}(d\beta_0, d\beta_1, dG|\text{data})$$

where $\mathcal{P}$ is the space of all probability measures on Θ, $\lambda_x := \beta_0 + \beta_1 x + \log \vartheta_2$, and $\mathcal{L}(d\beta_0, d\beta_1, dG|data)$ denotes the joint posterior of β_0, β_1 and the random probability G.

4   Posterior distributions and MCMC algorithm

Computation of full Bayesian inferences requires the knowledge of the posterior distribution of the random density $f(v; G)$ in (6), or the corresponding distribution function $F(v; G)$, as well as the posterior distribution of the regression parameters β. Here it is possible to build an MCMC algorithm which approximates the posterior distribution of G, so that we will also provide credibility intervals for $F(v; G)$.

The hierarchical structure of model (5), indicates that, conditionally on G, every lifetime T_i is associated with a point $\theta_i = (\vartheta_{1i}, \vartheta_{2i})$ from the support of G. These points are not necessarily distinct. We denote the distinct values within the n-ple $\underline{\theta} = (\theta_1, \ldots, \theta_n)$ by $\underline{\psi} = (\psi_1, \ldots, \psi_{n(\pi)})$, where $n(\pi)$ is their number, $1 \leq n(\pi) \leq n$. The elements of $\underline{\psi}$ are matched to the elements of $\underline{\theta}$ by means of the induced partition of the indexes $\{1, \ldots, n\}$, which we denote by $\pi = \{C_1, \ldots, C_{n(\pi)}\}$, where $C_j = \{i : \theta_i = \psi_j\}$.

Let $t_{98}^{108} = (t_{98}, \ldots, t_{108})$ be the vector of the imputed censored failure times. The state of the Markov chain will be $(G, \underline{\theta}, \beta, t_{98}^{108}) = (G, \underline{\psi}, \pi, \beta, t_{98}^{108})$. In order to build an MCMC sampler, we must be able to sample from all the full conditional posterior distributions, and in particular from the full conditional of G. By a characterization of the posterior distribution of G given in James et al. (2008), sampling from the full conditional of G amounts to sampling the $n(\pi)$ weights assigned to the points in $\underline{\psi}$ and both the (infinite) remaining weights and support points of G (see (10) below). An augmentation of the state space with an auxiliary variable u, allows for an independent conditional sampling of the two groups of random variables. Then the actual state of the chain will be $(G, \underline{\psi}, \pi, \beta, t_{98}^{108}, u)$. For more

explanations and details about the meaning of u, we refer the reader to James et al. (2008), Nieto-Barajas and Prünster (2009) and to Argiento et al. (2010) and simply describe the steps of our algorithm, with the square brackets notation denoting probability distributions. We also omit the indication of the observed failure times among the conditioning random variables for ease of notation.

Sampling G. By conditional independence, G depends on the observed lifetimes only through the vector $\underline{\theta}$. Therefore, using the equivalent representation $(\underline{\psi}, \pi)$ for $\underline{\theta}$,

$$[G \mid \underline{\psi}, \pi, \beta, t_{98}^{108}, u] = [G \mid \psi, \pi, u] \ .$$

As illustrated in the above cited papers, $[G \mid \psi, \pi, u]$ is the law of the following random distribution function

$$(10) \quad G^* = \frac{1}{T_{n(\pi)} + \sum_{j=1}^{n(\pi)} L_j} \sum_{j=1}^{+\infty} J_j \delta_{\tau_j} + \frac{1}{T_{n(\pi)} + \sum_{j=1}^{n(\pi)} L_j} \sum_{j=1}^{n(\pi)} L_j \delta_{\psi_j}, \qquad T_{n(\pi)} = \sum_{j=1}^{\infty} J_j \ ,$$

where the ψ_j's are fixed points in the support of G^* and the remaining weights and support points are random. In detail: the L_j's are independently $gamma(e_j - \sigma, \omega + u)$-distributed, with $e_j = \#C_j$; the τ_j's are a random sample from G_0; the sequence $(J_j)_j$ is generated according to representation (4) of an $NGG(\sigma, \kappa G_0, \omega)$, with $\omega = 1$ and $\kappa = \sigma\eta$.

While the sequence $(J_j)_j$ should be infinite, we use a finite sequence $(J_j)_{1 \le j \le M}$, for M such that $\mathbb{P}\left(\sum_{M+1}^{+\infty} J_j \le \tilde{\eta}\mathbb{E}(T_{n(\pi)})\right) \ge 1 - \varepsilon$, where ϵ and $\tilde{\eta}$ are suitably small. For details about the truncation method and the simulation procedure from the Poisson process we refer to Argiento et al. (2010).

Sampling $\underline{\theta}$. Using the truncated distribution G^* sampled at the previous step, we have, as $i = 1, \ldots, n$, $[\theta_i \mid G^*, \underline{\theta}_{-i}, \beta, t_{98}^{108}, u] = [\theta_i \mid G^*, \beta, t_{98}^{108}]$, where the dependence on the failure times and β is retained, because θ_i is the parameter of a kernel density (see (5)). Using Bayes theorem it can be shown that, for $i = 1, \ldots, n$,

$$[\theta_i \mid G^*, \beta, t_{98}^{108}] \propto \sum_{j=1}^{M} J_j k(v_i; \theta_i)\delta_{\tau_j}(d\theta_i) + \sum_{j=1}^{n(\pi)} L_j k(v_i; \theta_i)\delta_{\psi_j}(d\theta_i)$$

where $v_i = e^{-x_i'\beta}t_i$.

This update may change both the partition π and the $\underline{\psi}$ vector. However, as is now well-known, the ψ_j value associated with a group of failure times can only change after changing the membership of such failure times one at a time, thus slowing down convergence to stationarity dramatically. Then an "acceleration step", by which the ψ_j's are updated for all the failure times indexed through C_j, is usually introduced.

Acceleration step. This step is done by sampling from the full conditional distribution of ψ_j, as $j = 1, \ldots, n(\pi)$, which is proportional to

$$G_0(d\psi_j) \prod_{i \in C_j} k(v_i; \psi_j) .$$

Recall that our specific kernel density is Weibull, with unknown shape and scale. Since there are no known conjugate prior distributions for these parameters, a Metropolis-within-Gibbs algorithm is required in our update scheme.

Sampling β. If the components of β are taken a priori independent, the full conditional of β_j is proportional to the posterior distribution of the vector β:

$$\pi(\beta) \prod_{i=1}^{n} v_i^{\vartheta_{1i}} \exp\left\{ -\left(\frac{v_i}{\vartheta_{2i}}\right)^{\vartheta_{1i}} \right\} \propto \exp\left\{ -\beta_j \sum_{i=1}^{n} \vartheta_{1i} x_{ij} - \sum_{i=1}^{n} \left(\frac{v_i}{\vartheta_{2i}}\right)^{\vartheta_{1i}} \right\} ,$$

where $v_i = \mathrm{e}^{-x_i'\beta} t_i$. We do not develop this expression further to avoid tedious calculations. The full conditional of β_j is log-concave and is amenable to adaptive rejection sampling. The actual implementation of the adaptive rejection sampling, at least with our specific dataset, proved nontrivial, because for some values of the shape parameters the full conditional is very asymmetric and decreasing at a fantastic speed on one side of the mode with respect to the other side. In this situation, the points for the approximating envelope on the side of the mode where the decrease is faster must be selected very carefully, so that they do not all lie where the density is negligible, causing numerical instability.

Sampling u. It is shown in James et al. (2008) that the full conditional of u is proportional to

$$\left(\frac{u}{u+\omega}\right)^n \frac{(u+\omega)^{n(\pi)\sigma}}{u} \mathrm{e}^{-\frac{\kappa}{\sigma}(u+\omega)^\sigma}$$

where again we meet a non-standard distribution and a Metropolis step must be designed.

Sampling t_{98}^{108}. All the censored failure times have the same censoring point at 41000 hours. The failure times are conditionally independent given $\underline{\theta}$, then the full conditional of t_i, $i = 98, \ldots, 108$ is a left-truncated Weibull distribution with shape ϑ_{1i} and scale ϑ_{2i}, with survival function

$$\frac{\mathrm{e}^{-\left(\frac{t_i}{\vartheta_{2i}}\right)^{\vartheta_{1i}}}}{\mathrm{e}^{-\left(\frac{41000}{\vartheta_{2i}}\right)^{\vartheta_{1i}}}}, \qquad t_i > 41000 ,$$

which can be inverted exactly. This concludes the MCMC sweep.

5 Data analysis

As an illustration of our methodology, in this section we analyse the Kevlar fibre failure data and compare our results to those obtained previously by Crowder et al. (1991) and Leon et al. (2007).

Model (5) under (7) needs to be completed with the specification of G_0 and the prior distribution for β. First, observe that, since the log-survival time of the Kevlar fibres is modeled as

$$\log(T) = \beta_0 + \beta_1 x_1 + \log(\vartheta_2) + \frac{W}{\vartheta_1} = \beta_1 x_1 + \log(V),$$

we let $\beta_0 = 0$, because otherwise it would be confounded with $\mathbb{E}(\log(\vartheta_2) \mid G)$. We take the only component β_1 of β to be normal distributed with mean zero and variance 10^4. Concerning G_0 we choose the product of two independent gamma distributions, defining the prior density of $\theta = (\vartheta_1, \vartheta_2)$ as

$$g_0(\vartheta_1, \vartheta_2) = \frac{d^c}{\Gamma(c)} \vartheta_1^{c-1} \mathrm{e}^{-d\vartheta_1} \times \frac{b^a}{\Gamma(a)} \vartheta_2^{a-1} \mathrm{e}^{-b\vartheta_2}, \quad \vartheta_1 > 0, \ \vartheta_2 > 0.$$

Then we choose the hyperparameters a, b, c, d such that $\mathbb{E}(\log(V))$ yields any desired prior mean for the intercept. As commonly done, we impose $\mathbb{E}(\log(V)) = 0$, provided hyperparameters are such that it exists. Details on the computation of the first two moments of $\log(V)$ are given in the Appendix.

We fixed seven quadruplets of hyperparameters, with $c/d = 1$, representing the prior expected value of the shape parameter ϑ_1, whereas (a, b) is selected to have $\mathbb{E}(\log(V)) = 0$ when it is finite. In particular we choose $c = d = 1$ so that $\log(V)$ has no finite moments (and $a = b = 1$); $c = d = 2$ and $(a, b) = (0.5, 0.044)$ such that only the first moment is finite; finally we choose $c = d = 3$ and $(a, b) = (0.5, 0.059)$ to ensure $\log(V)$ has finite variance. As the values of a increase, the prior variance of ϑ_2 decreases. We also let η take values $0.1, 1$ and 10 and σ take 0.1 and 0.3. The combination of different values of a, b, c, d and (σ, η) give a feel of the robustness of the inference to hyperparameters.

In Table 1 and 2 we report point and interval estimates of the quantiles of the predictive distributions (introduced at the end of Section 2) for our mixture model and for the parametric Bayesian mixed-effects model. The distributions of the quantiles themselves are estimated numerically inverting the interpolant of the function $\int_\Theta \exp\left(-\left(t/(\beta_1 x + \log \vartheta_2)\right)^{\vartheta_1}\right) G(d\theta)$ (see formula (9)) evaluated on a fine enough grid of t-values, where G is the sampled value of the random distribution in the MCMC algorithm previously described. Notice how much narrower are the interval estimates compared to those under the parametric Bayesian mixed-effects model in Leon et al. (2007), and how the predictive median survival times corresponding to an extrapolated log-stress level equal to $\log(22.5) = 3.11$ are much larger than those under the parametric Bayesian model.

Table 1: Median predictive failure time (in thousands of hours) at an extrapolated level of stress (22.5 MPa) for the semiparametric mixture model: posterior medians with 95% credibility intervals.

No moments $(a, b, c, d) = (1, 1, 1, 1)$	$\sigma = 0.1$	$\sigma = 0.3$
$\eta = 10$	59.21 130.59 250.97	62.47 124.92 230.09
$\eta = 1$	49.86 108.92 232.16	57.94 126.98 251.57
$\eta = 0.1$	47.05 92.50 200.35	49.56 106.93 234.50
Finite first moment $(a, b, c, d) = (0.5, 0.044, 2, 2)$		
$\eta = 10$	100.22 207.24 399.39	124.51 236.90 463.76
$\eta = 1$	61.68 163.09 358.02	88.44 199.03 397.31
$\eta = 0.1$	47.03 102.74 279.75	60.67 168.72 381.15
Finite second moment $(a, b, c, d) = (0.5, 0.059, 3, 3)$		
$\eta = 10$	98.61 199.21 389.99	117.85 223.72 421.20
$\eta = 1$	64.88 168.62 368.19	88.01 190.82 388.97
$\eta = 0.1$	47.86 104.04 289.20	62.76 166.67 371.41
Parametric mixed-effects	1.867 53.68 1479	

Table 2: 0.01 quantile of the predictive distribution of failure time (in hours) at the lowest stress (23.4 MPa) for the semiparametric mixture model: posterior medians with 95% credibility intervals.

| No moments $(a, b, c, d) = (1, 1, 1, 1)$	$\sigma = 0.1$	$\sigma = 0.3$
$\eta = 10$	5.57 254.72 1252.06	0.96 16.46 321.76
$\eta = 1$	20.32 130.91 1381.86	4.43 211.41 1224.56
$\eta = 0.1$	18.01 68.98 296.47	17.90 100.10 1021.70
Finite first moment $(a, b, c, d) = (0.5, 0.044, 2, 2)$		
$\eta = 10$	80.428 850.24 2287.99	5.97 309.57 1337.00
$\eta = 1$	34.54 432.03 2178.31	61.80 721.50 2291.46
$\eta = 0.1$	20.51 81.38 689.43	33.22 469.99 2243.22
Finite second moment $(a, b, c, d) = (0.5, 0.059, 3, 3)$		
$\eta = 10$	107.67 788.25 2266.19	21.89 365.56 1271.24
$\eta = 1$	37.34 498.97 2206.55	72.17 669.45 2145.24
$\eta = 0.1$	19.75 84.99 730.24	34.88 461.84 2036.61
Parametric mixed-effects	21.96 671 19290	
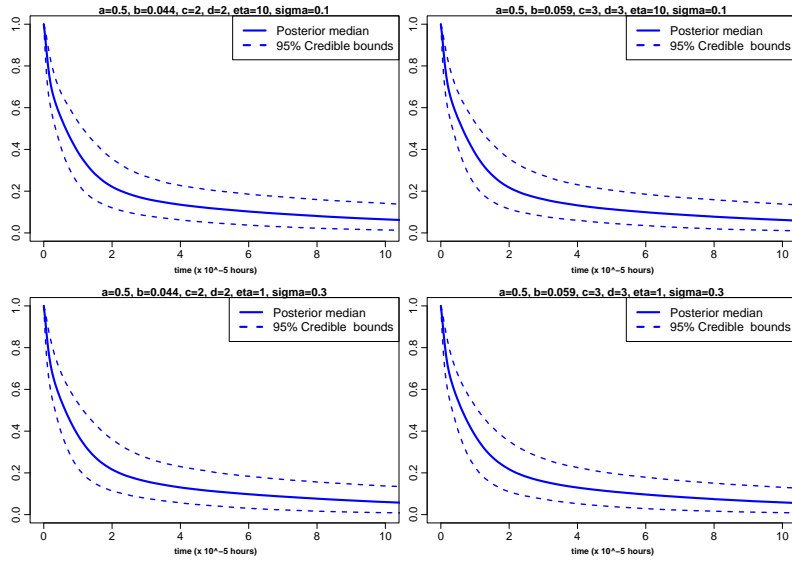
Figure 1: Predictive survival functions for the semiparametric AFT model, with credibility intervals, for different hyperparameter values; the stress covariate is $x_1 = \log(22.5)$

The estimates tend to increase with η and σ, and are not insensitive to hyperparameter variations; we also computed the credible intervals of the functionals of interest for $\sigma = 0.6$, but they are not reported in Table 1 and 2 to avoid overwhelming the reader. We suggest to choose η and σ such that the expected number of components in the mixture (also called *clusters*) is 8, in order to use the prior information on the number of spools. Among the values of our grid, $(\eta, \sigma) = (10, 0.1)$ and $(\eta, \sigma) = (1, 0.3)$ are more suitable, because the expected number of clusters is 6.9 and 7.4, respectively, whereas in other cases it is too small or too large: as high as 50.2 for $(\eta, \sigma) = (10, 0.6)$ and as low as 1.7 for $(\eta, \sigma) = (0.1, 0.1)$. The plots of the posterior distributions of the number of clusters for (η, σ) equal to $(10, 0.1)$ or $(1, 0.3)$ (see Figure 2) have their mode in 5, 6 or 7, suggesting a more parsimonious modelling, with 8 still being an a posteriori credible value. As a sensible choice for the hyperparameters of G_0, we suggest to elicit them fairly vaguely, for instance assuming that only the first moment of $\log(V)$ is finite, letting $(a, b) = (0.5, 0.044)$ and $c = d = 2$. The predictive survival functions (9) with useful credibility intervals are displayed in Figure 1.

In Figure 3 a scatterplot of the log survival-time against the covariate (log-stress) is shown together with the estimates of the median survival time under the parametric and semiparametric AFT model; the hyperparameters for the latter model are those previously suggested. The regression line obtained under the frequentist AFT regression with Weibull errors and without spool effect on one hand, and the empirical medians, on the other, agree with our interval estimate. We notice that the semiparametric structure of the distribution
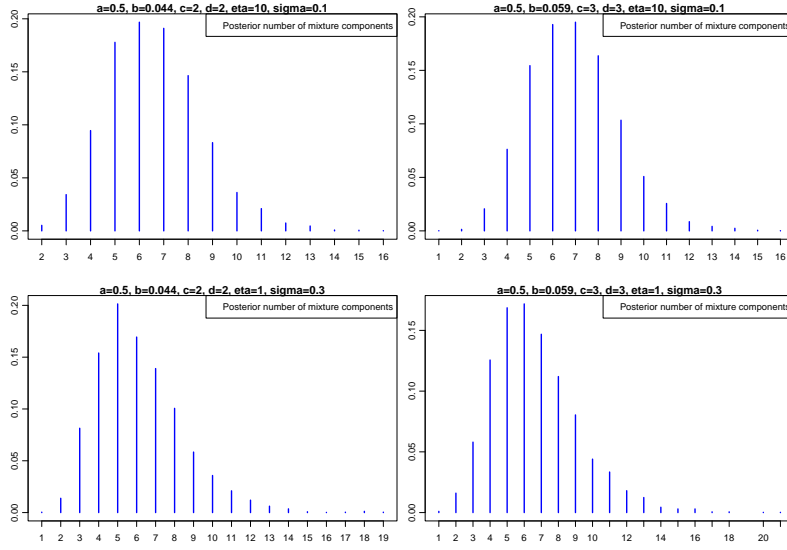
Figure 2: Posterior distributions of $n(\pi)$ for different hyperparameter values; the stresscovariate is $x_1 = \log(22.5)$

of $\log(V)$ better follows the log-linear relationship between survival time and stress.

Finally, we consider the goodness-of-fit of the model through Bayesian residuals as in Chaloner (1991). For our specific model, these are based on seeing that, by equation (8),

$$W_i = \vartheta_{i1}(\log T_i - \lambda_{x_i})$$

has a standard Gumbel distribution, conditionally on the parameter values. Thus, a priori, we expect the corresponding qq-plot of the residuals to be straight. Then we can examine the qq-plot of the posterior means of the realized residuals

$$\varepsilon_i = \vartheta_{i1}(\log t_i - \lambda_{x_i})$$

for indications of possible departures from the assumed model. When $\eta = 1$, $\sigma = 0.3$ and only the first moment of $\log(V)$ is finite, there is a clear improvement in the residuals of non-censored observations at the lowest level of stress (see Figure 4, left panel) with respect to the frequentist fixed-effects model (right panel), as meant by Crowder et al. (1991), Section 4.10.

We computed also posterior predictive p-values, as meant by Gelman et al. (2004), for the parametric and semiparametric mixed-effects models; more specifically, for all non-censored observations, we computed the predicted distribution of the i-th "replicated data" under the same value of the parameters that "produced" the i-th observation:

$$\min\left(\mathbb{P}(T_i^{new} > t_i | data, x_i, parameter_i), \mathbb{P}(T_i^{new} < t_i | data, x_i, parameter_i)\right), \quad i = 1, \ldots, 97$$
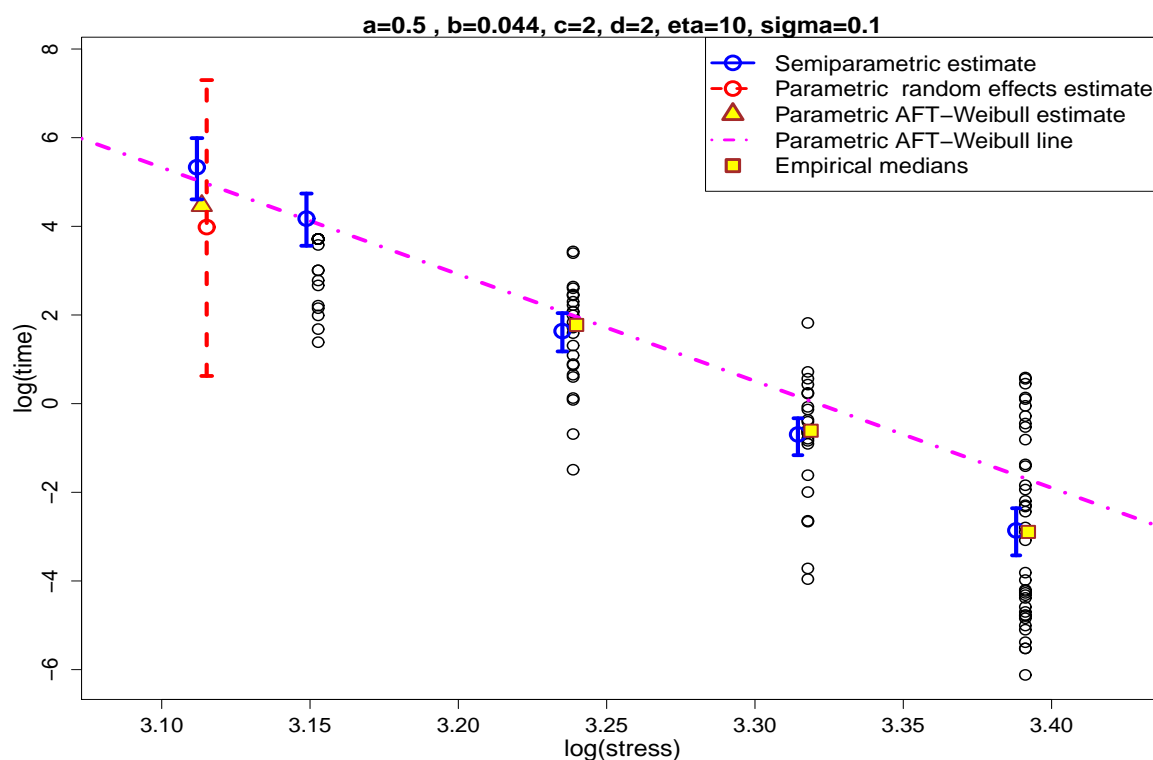
Figure 3: Scatterplot of the log-survival time against the log-stress. The credible intervals of the predictive survival median times are shown (dashed is parametric, solid is semiparametric). The dotted line is the parametric AFT regression line under Weibull errors, the diamond is the corresponding estimate of the median, and the squares are the empirical medians of the non-censored data for each stress level.

under the parametric model, using the WinBUGS code available in Leon's paper (with a $gamma(1, 0.2)$ prior for the Weibull shape parameter), and under our model. In the above formula, $parameter_i$ means the spool effect parameter γ_i in the parametric mixed-effects model, and the latent θ_i associated to t_i in our model (of course, it includes the regression parameters in both cases). To simplify, observations with posterior predictive p-values less than 0.1 (say) were classified as "unusual": we found 18 unusual observations under the parametric mixed-effects model, and only 4 under ours.

Figure 5 displays predicted survival functions (and 95% CI) for the parametric and semiparametric models, together with the Kaplan-Meier estimator, at the lowest stress level. The predictive bands for the parametric model are completely useless, while the semiparametric model provides usable credibility intervals both for the survival function and its quantiles. The values of hyperparameters are those suggested, which implies that $\log(V)$ has finite first moment (only), and produce a posterior predictive median of the survival function close to
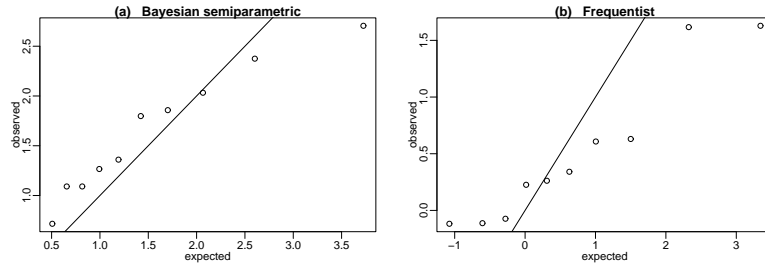
Figure 4: Bayesian semiparametric (a) and frequentist (b) residuals at stress level 23.4 MPa. The hyperparameter values for the nonparametric mixture are $\sigma = 0.3$, $\eta = 1$, $a = 0.5$, $b = 0.044$, $c = d = 2$

Kaplan-Meier. Hyperparameter values such that $\log(V)$ has no finite moments give similar estimates, but with some departure from Kaplan-Meier in the lower tail and more closeness in the upper tail.

Since our model presents more similarities with that in Ghosh and Ghosal (2006) than with the other cited Bayesian semiparametric models, we have made a numerical comparison. Briefly, Ghosh and Ghosal's assumes the same conditional likelihood than ours, i.e. $V_i|(\alpha, \mu_i)$ is Weibull-distributed with parameter (α, μ_i), but the shape parameter α, unlike our case, is parametrically distributed as a gamma(c, d) random variable, while, conditionally on G, $\mu_1, \ldots, \mu_n$ are i.i.d. from G, and G is a Dirichlet process with mean probability measure gamma(a, b) and total mass parameter κ. As in Ghosh and Ghosal's paper, we computed the posterior distribution under their model resorting to the N-finite approximation of Dirichlet processes in Ishwaran and Zarepour (2002). In particular, we fixed $N = 60$, since for a smaller N there were relevant differences between the two posterior distributions of $n(\pi)$; κ was fixed equal to 2 in order that $E(n(\pi)) = 8.5$, a value very close to the actual number of spools. Moreover, to compare the performances of the two semiparametric mixtures, we fitted our model for $\sigma = 0$ (the NGG process is a Dirichlet process), and the same values of a, b, c, d.

In Table 3 we report a comparison when $(a, b, c, d) = (1, 1, 1, 1)$ and $(a, b, c, d) = (0.5, 0.044, 2, 2)$, as specified in Table 1 and 2. The estimates for the median predictive failure time at 22.5 MPa are very similar, while there is much sensitivity of the 0.01 quantile at 23.4 MPa with respect to the mixture model considered. In particular, the left endpoint of the credible interval of the 0.01 quantile at 23.4 MPa is higher under Ghosh and Ghosal's mixture model. To quantitatively analyze these differences, we compared the posterior distributions of the shape parameters α, $\{\vartheta_{1i}\}$ of both semiparametric mixture models in Figure 6 (with $a = 0.5$, $b = 0.044$, $c = d = 2$).

The posterior of α is bimodal with its modes smaller and larger than one, respectively. In good agreement with such bimodality, the posteriors of some ϑ_{1i}'s place most mass below
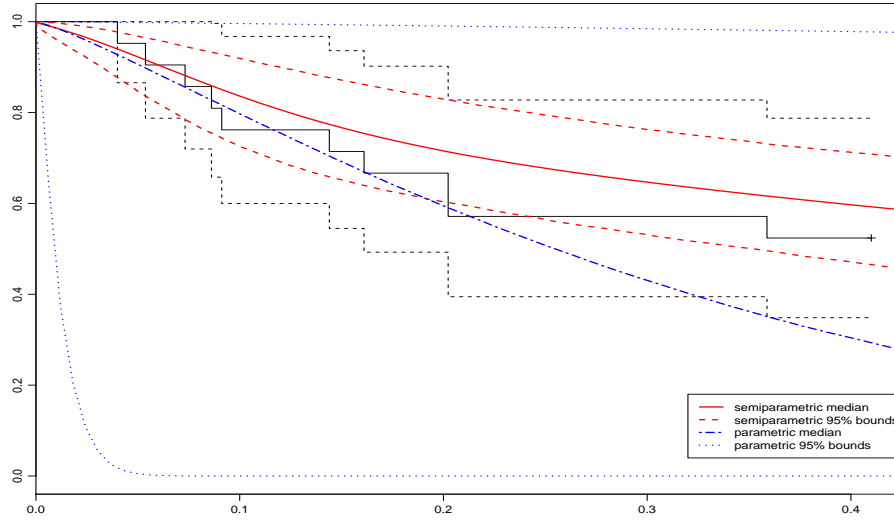
Figure 5: Predicted survival functions at stress level 23.4 MPa for the parametric and semiparametric models, together with the Kaplan-Meier estimator. The Bayesian estimates are medians and 95% credibility intervals. The step functions are the Kaplan-Meier estimator and the corresponding 95% confidence bounds. Time on the x-axis is in hundreds of thousands of hours; the cross $(+)$ at 0.41 is the censoring time. The hyperparameter values used to compute the semiparametric estimates are $\sigma = 0.1$, $\eta = 10$, $a = 0.5$, $b = 0.044$, $\bar{c} = 2$, $d = 2$.

one, whereas some others place most mass above one. It is known that the first percentile of a Weibull distribution with a small shape parameter is close to zero, and decreases not linearly (but faster) than the shape parameter when this is decreasing to zero. Therefore, when many Weibull distribution are mixed during an MCMC iteration for the computation of (9), those corresponding to smaller shape parameters tend to pull down the first percentile. The stronger heaviness of the tails of ϑ_{1i} with respect to the tails of α enhances this behaviour.

6   Concluding remarks and discussion

We have fitted a dataset of 108 lifetimes of Kevlar pressure vessels with two covariates (spool and stress) to an AFT model. The error term was modeled as a hierarchical mixture of Weibull distributions on both the shape and the scale. The mixing measure was assumed to be a normalized generalized gamma random measure, indexed by two parameters. As mentioned in the paper, this model presents an analogy with the Bayesian parametric mixed-effects model; anyhow, the advantages of our model consist in a more flexible distribution of the error term and a random number of random effect parameters induced by the nonparametric model itself. We also found that the nonparametric mixture model for the error better follows the log-linear relationship between failure times and covariates. We obtained usable interval

Table 3: Median predictive failure time (in thousands of hours) at an extrapolated level of stress (22.5 MPa) and 0.01 quantile of the predictive distribution of failure time (in hours) at the lowest stress (23.4 MPa) for Ghosh and Ghosal's and our semiparametric mixture models: posterior medians with 95% credibility intervals.

	Ghosh-Goshal's mixture	NGG-mixture
$(a, b, c, d) = (1, 1, 1, 1)$		
median22.5	60.48 111.78 196.56	63.10 131.27 238.56
0.01perc23.4	84.31 603.85 1540.86	1.67 212.65 1177.37
$(a, b, c, d) = (0.5, 0.044, 2, 2)$		
median22.5	114.13 221.65 409.28	108.31 210.16 409.64
0.01perc23.4	457.05 1478.81 2966.13	54.17 787.70 2185.08

estimates of the quantiles considered (unlike the Bayesian mixed-effects estimates); useful credibility intervals were also obtained for the predictive survival functions. As far as the goodness-of-fit is concerned, the residual plot for the lowest level of stress and the predictive *p*-values show an improvement with respect to previous analyses.

We also compared our semiparametric mixture model to a simpler one by Ghosh and Ghosal (2006). Numerical evaluation of predictive survival functions (see formula (9)) and predictive quantiles is expensive under both models: as mentioned in Section 5, at every step of the MCMC algorithm we have to evaluate the function $\int_{\Theta} \exp\left(-\left(t/(\beta_1 x + \log \vartheta_2)\right)^{\vartheta_1}\right) G(d\theta)$ on a grid of *t*-values and the inverse of its piecewise linear interpolation (were $\vartheta_1 = \alpha$ and $\vartheta_2 = \mu$ in Ghosh and Goshal's model). The MCMC sampler for their joint posterior distribution can be designed more quickly, even using a WinBUGS code, however our model is more general, since both parameters of the Weibull kernel are modeled nonparametrically, and the nonparametric NGG random probability measure includes the Dirichlet process they use (for $\sigma = 0$). The estimates of the median survival time at 22.5 MPa are equivalent, but our model better quantifies the variability of the first percentile of the survival time at 23.4 MPa. This feature, along with the narrowness of the credibility intervals, is brought about by the embedment of the random effect of spools in the nonparametric error term.

Of course, semiparametric mixtures models require a higher computational effort than the Bayesian parametric mixed-effects model (3). On the other hand, we run different simulations using the WinBUGS code of the parametric model, and we found that, in spite of large values of the number of iterations (more that 1 million) and thinning (100), the autocorrelation of β_0, as well as of the spool parameters γ_i's, are still high. This is because the marginal posterior distributions of (β_0, γ_i) concentrate around straight lines. For our model, we coded the algorithm in C, using GSL libraries when necessary, and obtained very-fast-decaying
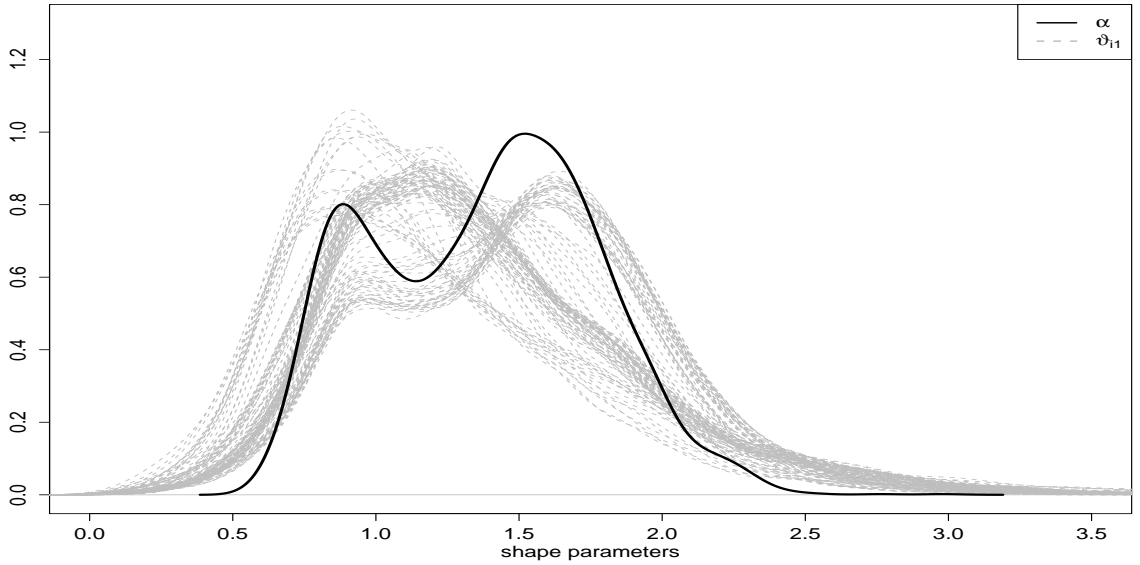
Figure 6: Posterior distributions of the shape parameters α and ϑ_{1i}, $i = 1, \ldots, 108$, for Gosh and Goshal's and NGG- mixtures when $a = 0.5$, $b = 0.044$, $c = d = 2$.

autocorrelations of the kernel density parameters, using one every four iterations. However, in our experiments some mixing problems arose when both η and σ are small. In fact, it is well known (see for instance Argiento et al., 2010, or Lijoi et al., 2007) that small values of these parameters favour samples from the NGG process with few jumps, forcing small values of $n(\pi)$ and therefore a small number of random effects in model (8). In this case, there are not enough distinct pairs $(\vartheta_{1i}, \vartheta_{2i})$ to account for the clustering of the failure times, so that multimodality occurs in the Markov sequence of such pairs.

Appendix: Computation of prior moments

Let V be a random variable defined by

$$\log(V) = \log(\vartheta_2) + \frac{W}{\vartheta_1}$$

where ϑ_1, ϑ_2 and W are independent random variables, $\vartheta_1 \sim \text{gamma}(c, d)$, $\vartheta_2 \sim \text{gamma}(a, b)$ and $W \sim \text{Gumbel}(0, 1)$, then

$$(11) \quad \begin{aligned} \mathbb{E}(\log(V)) &= \psi(a) - \log(b) - \gamma \frac{d}{c - 1} \\ \text{Var}(\log(V)) &= \frac{d^2}{(c - 1)^2} \left[\left(\frac{\pi^2}{6} + \gamma \right) \frac{1}{c - 2} + \frac{\pi^2}{6} \right] + \psi'(a), \end{aligned}$$

where $\psi(\cdot)$ denotes the digamma function defined by

$$\psi(x) := \frac{d}{dx}\Gamma(x) = \frac{d}{dx}\int_0^\infty t^{x-1}e^{-t}\,dt, \quad x > 0.$$

To prove (11) let first observe that if ϑ_2 is a random variable with finite m-th moment, then by the Dominated Convergence Theorem

(12) $$\mathbb{E}\left(\log^m(\vartheta_2)\right) = \lim_{\lambda \to 0}\mathbb{E}\left(\left(\frac{\vartheta_2^\lambda - 1}{\lambda}\right)^m\right) \quad m = 1, 2, \dots.$$

In particular if $\vartheta_2 \sim \text{gamma}(a, b)$ then

$$\mathbb{E}(\vartheta_2^\lambda) = \frac{\Gamma(a + \lambda)}{\Gamma(a) \cdot b^\lambda},$$

and using (12) we obtain

$$\mathbb{E}(\log(\vartheta_2)) = \lim_{\lambda \to 0}\frac{1}{\lambda}\cdot\left(\frac{\Gamma(a + \lambda)}{\Gamma(a) \cdot b^\lambda} - 1\right) = \psi(a) - \log(b).$$

Analogously, from (12) with $m = 2$,

$$\mathbb{E}\left(\log^2 \vartheta_2\right) = \lim_{\lambda \to 0}\frac{\Gamma(a + 2\lambda) + \Gamma(a)b^{2\lambda} - 2b^\lambda\Gamma(a + \lambda)}{\lambda^2 b^{2\lambda}\Gamma(a)} = \psi'(a) + (\psi(a) - \log(b))^2.$$

Simple calculations shows (11).

References

Argiento, R., Guglielmi, A. and Pievatolo, A. (2009), 'A comparison of nonparametric priors in hierarchical mixture modelling for AFT regression', *Journal of Statistical Planning and Inference*, **139**, 3989–4005.

Argiento, R., Guglielmi, A. and Pievatolo, A. (2010), 'Bayesian density estimation and model selection using nonparametric hierarchical mixtures', *Computational Statistics & Data Analysis*, **54**, 816–832.

Argiento, R., Guglielmi, A. and Pievatolo, A. (2010b), 'Mixed-effects modelling of Kevlar fibre failure times through Bayesian non-parametrics', in *Complex Data Modelling and Computationally Intensive Statistical Mehtods*, ed. P. Mantovan and P. Secchi, Milano: Springer-Verlag Italia.

Cai, B. and Dunson, D.B. (2008), 'Bayesian variable selection in generalized linear mixed models', in *Random Effect and Latent Variable Model Selection*, ed. D.B. Dunson, New York: Springer.

Chaloner, K. (1991), 'Bayesian residual analysis in the presence of censoring', *Biometrika*, **78**, 637–644.

Crowder, M.J., Kimber, A.C., Smith, R.L. and Sweeting T.J. (1991), *Statistical Analysis of Reliability Data*, Chapman & Hall

Gelfand, A.E. and Kottas, A. (2003), 'Bayesian semiparametric regression for median residual life', *Scandinavian Journal of Statistics*, **30**, 651–665.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004), *Bayesian Data Analysis*, Second Edition, Chapman & Hall.

Gerstle, F.P. Jr and Kunz, S.C. (1983), 'Prediction of long-term failure in kevlar 49 composites in long-term behavior of composites', in *American Society for Testing and Materials*, Philadelphia pp 263–292.

Ghosh, S.K. and Ghosal, S. (2006), 'Semiparametric accelerated failure time models for censored data', in *Bayesian Statistics and its Applications*, ed. S.K. Upadhyay et al, New Delhi: Anamaya Publishers, 213–219.

Hanson, T.E. (2006), 'Modeling censored lifetime data using a mixture of gammas baseline', *Bayesian Analysis*, **1**, 575–594.

Ishwaran, H. and Zarepour, M. (2002), 'Dirichlet prior sieves in finite normal mixtures', *Statistica Sinica*, **12**, 941-963.

James, L., Lijoi, A. and Prünster, I. (2006), 'Conjugacy as a distinctive feature of the Dirichlet process', *Scandinavian Journal of Statistics*, **33**, 105–120.

James, L., Lijoi, A. and Prünster, I. (2008), 'Posterior analysis for normalized random measures with independent increments', *Scandinavian Journal of Statistics*, **36**, 76–97.

Jara, A., Lesaffre, E., de Iorio, M. and Quintana, F. (2010), 'Bayesian semiparametric inference for multivariate doubly-interval-censored data', *Annals of Applied Statistics*, to appear.

Kinney, S.K. and Dunson, D.B. (2008), 'Bayesian model uncertainty in mixed effects models', in *Random Effect and Latent Variable Model Selection, Lecture Notes in Statistics, 192*, ed. D.B. Dunson, New York: Springer.

Kottas, A. and Gelfand, A.E. (2001), 'Bayesian semiparametric median regression modeling', *Journal of the American Statistical Association*, **96**, 1458–1468.

Kuo, L. and Mallick, B. (1997), 'Bayesian semiparametric inference for the accelerated failure-time model', *The Canadian Journal of Statistics*, **25**, 457–472.

León, R.V., Ramachandran, R., Ashby, A. and Thyagarajan, J. (2007), 'Bayesian modeling of accelerated life tests with random effects', *Journal of Quality Technology*, **39**, 3–13.

Lijoi, A., Mena, R.H. and Prünster, I. (2005), 'Hierarchical mixture modeling with normalized inverse-Gaussian prior', *Journal of the American Statistical Association*, **100**, 1278–1291.

Lijoi, A., Mena, R.H. and Prünster, I. (2007), 'Controlling the reinforcement in Bayesian nonparametric mixture models', *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 715–740.

Lo, A.Y. (1984), 'On a class of Bayesian nonparametric estimates: I. Density estimates', *The Annals of Statistics*, **12**, 351–357.

Nieto-Barajas, L.E. and Prünster, I. (2009), 'A sensitivity analysis for Bayesian nonparametric density estimators', *Statistica Sinica*, **19**, 685–705.

Pitman, J. (1996), 'Some developments of the Blackwell-MacQueen urn scheme', in *Statistic, Probability and Game theory, Papers in Honor of David Blackwell. Lectures Notes - Monograph Series, vol 30*, ed. T.S. Ferguson, Hayward, California: Institute of Mathematical Statistics,245–267.

Pitman, J. (2003), 'Poisson-Kingman partitions', in *Science and Statistics: a Festschrift for Terry Speed. Lectures Notes - Monograph Series., vol 40*, ed. D.R. Goldstein, Hayward,California: Institute of Mathematical Statistics, 1–34.

Pitman, J. and Yor, M. (1997), 'The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator', *The Annals of Probability*, **25**, 855–900.

Sun, D., Speckman, P.L. and Tsutakawa, R.K. (2000), 'Random effects in gereralized linear mixed models (GLMMs)', in *Generalized Linear Models - a Bayesian perspective*, ed. D.K. Dey et al., Boca Raton: CRC - Taylor & Francis.

Zeger, S.L. and Karim, M.R. (1991), 'Generalized linear models with random effects: a Gibbs Sampling approach', *Journal of the American Statistical Association*, **86**, 79–86.