MODELLISTICA E CALCOLO SCIENTIFICO

MODELING AND SCIENTIFIC COMPUTING

MOX–Report No. 15/2011

A semiparametric Bayesian generalized linear mixed model for the reliability of Kevlar fibres

Argiento, R.; Guglielmi, A.; Soriano, J.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

A semiparametric Bayesian generalized linear mixed model for the reliability of Kevlar fibres

Raffaele Argiento[a], Alessandra Guglielmi[b] and Jacopo Soriano[b]

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

March 28, 2011

Abstract

We analyze the reliability of NASA composite pressure vessels using a new Bayesian semiparametric model. The dataset consists of lifetimes of pressure vessels, wrapped with a Kevlar fiber, grouped by spool, subject to different stress levels; 10% of data are right censored. The model we consider is a regression on the log-scale for the lifetimes, with fixed (stress) and random (spool) effects. The prior of the spool parameters is nonparametric, namely they are a sample from a normalized generalized gamma process, which encompasses the well-known Dirichlet process. The nonparametric prior is assumed to robustify inferences to mispecification of the parametric prior. Here, this choice of likelihood and prior yields a new Bayesian model in reliability analysis. Via a Bayesian hierarchical approach, it is easy to analyze the reliability of the Kevlar fiber by predicting quantiles of the failure time when a new spool is selected at random from the population of spools. Moreover, for comparative purposes we review the most interesting frequentist and Bayesian models analyzing this dataset. Our credibility intervals of the quantiles of interest for a new random spool are narrower than those derived by previous Bayesian parametric literature. Additionally, the discreteness of the random-effects distribution induces a natural clustering of the spools into three different groups, which is in accordance with the frequentist spool rankings.

Keywords: accelerated failure time regression model; Bayesian clustering; Bayesian nonparametrics; random-effects model; reliability.

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

1 Introduction

The main purpose of this work is to analyze the reliability of NASA composite pressure vessels, which are critical components of the Space Shuttle, using a Bayesian semiparametric methodology. The dataset consists of 108 lifetimes of pressure vessels, wrapped with a Kevlar yarn, coming from 8 different spools, at different levels of pressure. Eleven lifetimes with the lowest level of pressure are right censored. Here we propose a new semiparametric Bayesian prior to fit the data, using a generalized linear model with the batch effect. The dataset have been repeatedly analyzed both from frequentist and Bayesian perspective. Gerstle and Kunz (1983) fitted six separated (frequentist) regression models using data from 6 spools only. Glaser (1983) and Crowder et al. (1991) fitted accelerated failure time (AFT) models for the failure times using Weibull distributions with both stress and spool as fixed effects. Feiveson and Kulkarni (2000) were the first to emphasize the need of assuming the spool effect as random, and provided a frequentist estimate of the parameters of the Weibull distribution for each stress level. Recently Bayesian approaches were proposed to analyze the dataset: Leon et al. (2007) consider a Bayesian Weibull regression model with the spool random effect, while Argiento et al. (2010a) propose a semiparametric Bayesian Weibull regression model, where the random spool effect is grouped via a discrete nonparametric component.

The aim of this paper is twofold. On one hand, we want to predict the reliability of the Kevlar yarn, using a Bayesian hierarchical approach to the problem. Under this framework, it will be easy to predict the failure time (or to monitor its quantiles) when a new spool is selected at random from the population of spools. The model we consider here is an AFT model for the pressure vessel lifetimes, with covariates given by the stress (fixed) and the spool (random). The same likelihood has been considered in Leon et al. (2007), but unlike them, we assume that the spool-effects parameters are a sample from a nonparametric random distribution. Therefore, as a second aim of the paper, a comparison between the inferences given by the parametric prior in Leon et al. (2007) and by our nonparametric prior will be made. Generally, nonparametric Bayesian models are assumed to avoid critical dependence on parametric assumptions, to robustify inferences to mispecification of the parametric prior, or to perform sensitivity analysis for parametric models by embedding them in a larger encompassing nonparametric model. Priors under a nonparametric Bayesian perspective consist in probabilities on probability spaces: instead of considering models that can be indexed by a finite-dimensional parameter, we consider a prior probability for the unknown population distribution P, which, in the case considered here, represents the probability distribution of the spool-effects parameter. For a review of Bayesian nonparametric inference see Müller and Quintana (2004). In this paper, P is a normalized generalized gamma (NGG) random measure, indexed by two parameters (σ, κ), controlling the amount of mass the distribution of P puts on the mean distribution P_0. The Dirichlet process is contained within this family, for $\sigma = 0$. Therefore, our model can be considered as a generalization of the generalized linear mixed model (GLMM) in Kleinman and Ibrahim (1998), who assume the random-effects parameters to be a sample from a Dirichlet process. The GLMMs with semiparametric priors are widespread in biostatistical applications, while new in reliability analysis. Observe that, since the random distribution function we assume selects discrete probabilities, there will be a positive probability of having coincident values among the sampled (from P) random-effects parameters. The number of distinct values among them will represent the number of effectively distinct groups among the 8 spools.

As far as the comparison to the parametric model is concerned, we have monitored the first percentile of the failure time distribution at the lowest stress level (23.4MPa) and the median of the failure time distribution at the (extrapolated) level of stress of 22.5MPa. Those two quantiles provide an insight on the failure behaviour: the first percentile at 23.4MPa can be considered as a low risk index, while the median at 22.5MPa is an an estimate under low loads corresponding to normal operating conditions. Of course, interval estimates are preferable to point ones, since they provide a degree of belief in such estimates. We can compute the quantile estimates for both a given and a new random spools, but, from a practical point of view, we are particularly interested in the latter, since the Kevlar fiber wrapping the Space Shuttle vessels will come from a new and unknown spool. We find that credibility intervals of the two quantiles of interest for a new spool were narrower than the parametric ones. However, they were sensitive to hyperparameters (σ, κ) of the nonparametric component P. To robustify inferences with respect to σ and κ, we have assumed a prior distribution on them. In this case, the credibility intervals of the two quantiles for a new random spool are still narrower than those under the parametric model. The discreteness of the random-effects distribution P induces a clustering of the spools into three different groups (the worst, the average and the best spools), which is coherent with the frequentist spool ranking in Crowder et al. (1991). As far as the reliability of the pressure vessels is concerned, we have obtained better interval estimates of the quantiles for a new random spool than before, but they are still too wide to conclude that the pressure vessels are fully reliable.

As a final comment we point out that the problem of the reliability of pressure vessels is still present and relevant nowadays (NASA has recently organized a meeting on composite pressure vessels). On one hand, the classification of spools we have provided suggests to investigate more deeply the different physical features of the groups of spools. On the other hand, our conclusions could be used to elicit an informative prior for new data, when available.

The setup of the paper is as follows. Section 2 briefly describes the experiment and the dataset considered. In Section 3 we review the most interesting literature, from a statistical point of view, of frequentist and Bayesian models analyzing the reliability of the Kevlar yarn in the dataset. Section 4 describes the semiparametric model, the nonparametric component of the model, and the quantiles considered in order to make a comparison between the semiparametric and parametric models. A few details about the MCMC algorithm for computing the posterior distribution are given in Section 4 itself. Results are presented in Section 5. Conclusions and comments are given in Section 6.

2 Accelerated life test on composite pressure vessels

In accelerated life tests (ALTs) a mechanical component is subject to high stress (for instance pressure or temperature) to accelerate its failure time, with the purpose of extrapolating its lifetime characteristics at a lower stress corresponding to normal use conditions. Here we analyze an ALT test on NASA composite pressure vessels, which are critical components of the Space Shuttle. A pressure vessel is a closed container designed to hold gases or liquids at high pressure. Pressure vessels wrapped with composite materials have been largely used since they offer considerable weight savings over steel or titanium alternatives. For instance, the Space-Shuttle has 22 composite pressure vessels containing Helium or Nitrogen at high pressure, and a mission usually lasts two weeks at most.

At the U.S. Department of Energy Lawrence Livermore National Laboratories two experiments on 220 scaled-down replicates of NASA pressure vessels were done, to estimate short-term static-burst pression and creep rupture time at given pressure; the dataset and the description of the experiments are reported in Gerstle and Kunz (1983). All the pressure vessels were wrapped with a Kevlar yarn and the fiber came from 8 different spools. In the first experiment, 29 vessels were used for the burst rupture test and pressurized at rate of 23.3 KPa/s. The application we will consider here is based on the second experiment, i.e. an ALT to estimate creep rupture time. In particular, 191 vessels were tested at different levels of pressure, 17.2, 23.4, 25.5, 27.6, 29.7 MPa, but 37 of them were discarded during the test because they developed leaks on the internal aluminum liner, yielding a dataset of 154 lifetimes. All the 46 vessels at the lowest level of stress and 11 vessels (out of 21) at stress level 23.4MPa were censored at 41000 hours. Gerstle and Kunz (1983) conclude that the short-term static-burst behavior of such vessels is quite predictable (the mean estimate is 34.5MPa, the standard deviation estimate is 1.3MPa). On the other hand, the response of the composite vessels to sustained pressurization cannot be easily forecast, because there is a large variability among the failure times and, in fact, several authors proposed different model to analyze those data; see Section 3.

Gerstle and Kunz (1983) provide a description of the vessels: each one consisted of a 112mm internal diameter aluminum liner, 1mm thick, overwrapped with 1.1mm of Kevlar 49 yarn wetted during winding with epoxy. The liner is essentially made from aluminum hemispheres, electron-beam welded at the equator. Although the liner is nearly as thick as the composite, it is much weaker and therefore contributes little to the strength of the vessel. Moreover, the authors provide further information about the microscopic characteristics of a single fiber and a single yarn of spool 2 and 7, respectively. For ease of reading, Table 1 reports the results of mechanical tests shown in Gerstle and Kunz (1983). One of the findings of all the analysis (see next sections) is that spool 7 has low performance and spool 2 has average behaviour. The bad performance of spool 7 can be explained by the microscopic characteristics of the yarn; in fact it has only 127 filaments instead of the nominal 267 and the diameter of the fiber is 17mm (instead of 11mm). Similarly, we guess that spool 2 has average performance since its properties are closer to the nominal ones.

	Nominal	Spool 2	Spool 7
Fiber strength (GPa)	2.86	3.33	3.31
Yarn failure load (Kg)	7.3	6.6	6.0
Fiber diameter (μm)	11.0	11.5	17.0
Fibers per strand of yarn	267	263	127
Yarn strength (GPa)	2.76	2.59	2.02

Table 1: Mechanical properties of a single fiber and a single yarn of Kevlar.

3 Frequentist and Bayesian approaches to the dataset

Several authors made a statistical analysis of this dataset. Some included the censored times at 23.4MPa, but nobody included the 46 vessels at lowest level of stress since they are all censored; spool 7 was discarded sometimes, since it does not respect the nominal conditions. Here we provide a brief description of the analyses mentioned in the Introduction.

Gerstle and Kunz (1983), who seem the first ones to consider this dataset, omit all the censored times. Using a F-test on transformed failure times, they find that there is variability among the spools; moreover, they conclude that the short-term burst tests do not correlate with the long term results, since the former data are homogeneous with respect to the spool, unlike the latter one. Therefore, their findings are that the short-term results cannot be used to predict long-term lifetime of the pressure vessels. As mentioned in the Introduction, the authors propose to fit 6×2 separated regression models for log-failure times (assuming Gaussian distributions for the responses); for each spool they consider two regression models with the covariate equal to the stress or to the log of the stress. They conclude that spool 4 is the best and spool 3 is the worst in terms of point estimate, while spool 6 has the largest variability. Finally, for each spool, they provide point estimates of the 10^{-6} percentiles of the failure time distribution for a new pressure vessel and the pressure value corresponding to a ten years life with a 0.999999 reliability, concluding that there are guarantees for ten years life. However, since such small percentiles or longest-life reliability are generally not robust, we cannot fully rely on their results.

Glaser (1983) considers neither the censored times nor spool 7, and model the failure times as independent, each marginally Weibull-distributed with parameter (ϑ, λ). By a Weibull(ϑ, λ) random variable we mean the variable with survival function

$$S(v) = \exp\Big\{ -\Big(\frac{v}{\lambda}\Big)^{\vartheta} \Big\}, \quad v \geq 0;$$

note that $\vartheta > 0$ and $\lambda > 0$ represent the shape and the scale parameters, respectively.

The author fits an AFT model with both stress and spool as fixed effects. We can describe the model equivalently both in multiplicative and additive form:

$$T = \mathrm{e}^{\boldsymbol{x}'\boldsymbol{\beta}} \cdot V, \quad V \sim \text{Weibull}(\vartheta, 1), \tag{1}$$

or

$$\log T = \boldsymbol{x}'\boldsymbol{\beta} + \frac{W}{\vartheta}, \quad W \sim \text{Gumbel}(0,1), \tag{2}$$

where

$$\begin{aligned} \boldsymbol{x}'\boldsymbol{\beta} &= \beta_0 + \beta_1 x + \sum_{j=1, j\neq 7}^{8} \alpha_j z_j, \quad \sum_{j=1, j\neq 7}^{8} \alpha_j = 0, \\ \vartheta &= \gamma_0 + \gamma_1 \mathrm{e}^{2x} \end{aligned}$$

and

$$x = \text{stress level}, \quad z_j = \text{effect of spool j (binary)}, \quad j \geq 1 \ .$$

In both specifications we have written the error term as a random variable with a standard distribution, so that the role of ϑ as a scale parameter in (2) or as a shape parameter in (1) 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 $\text{Var}(W) = \pi^2/6$. Then, by maximum likelihood (ML) estimation, Glaser (1983) provides the point estimates of the 10^{-3} percentile of the failure time distribution and the five-year reliability for a new pressure vessel from the *worst*, *mean* and *best* spools, respectively. The author assumes the best spool in the testing to be spool 4, since it has the longest lifetimes, the worst to be spool 3 (spool 7 is not considered there); the *mean* spool effect is obtained without considering the spool adjustment term $\sum_{j=1,j\neq 7}^{8} \alpha_j z_j$, i.e. it is represented by $\beta_0 + \beta_1 x$ at a given stress level x.

Crowder et al. (1991), who omit the censored times as well, present an analysis of variance based on the Weibull distribution with ML fitting, similar to Glaser (1983).

Although they assert that the shape parameter ϑ may be different at each stress level, they assume it constant, while $\boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 \log(x) + \sum_{j=1}^{8} \alpha_j z_j$ and $\sum_{j=1}^{8} \alpha_j = 0$. They provide point and interval estimates of the first percentile at stress level 23.4MPa and of the median at the extrapolated stress value of 22.5MPa of the failure times for the eight spools and for the mean spool at given stress level. Finally, they compare their results to those of Gerstle and Kunz (1983) and observe that they obtain the same spool ranking except for two worst spool 3 and 7, which are switched over.

As mentioned before, Feiveson and Kulkarni (2000) emphasize the necessity of treating the spool effect as random, since it is unknown which spools are used to wrap the pressure vessels of the Space Shuttle. Each vessel may be wound from one of the known spools, or from a spool selected at random from the population of spools. Moreover, they are the first ones who considered the censored times in their analysis, while they omit spool 7, and assume that the lifetimes are Weibull-distributed. First of all, for each stress level, they estimate the scale and the shape parameters of the Weibull distributions using a bootstrap method, requiring an estimate of the censoring probability at each stress level; then, they extrapolate for any stress level. Finally, they estimate the probability R_m that the system of 22 pressure vessels survives m two-weeks missions, and the conditional probability $p_m = R_m/R_{m-1}$ that the system survives m mission given that it survived $m-1$ missions, concluding that there is a considerable amount of uncertainty in both estimates.

Leon et al. (2007) propose an AFT model as Crowder et al. (1991) using a Bayesian parametric approach. They consider data from all the 8 spools and also include censored times for a total of 108 observations. They treat the spool effect as fixed first, only for comparative purposes, and then as random yielding a GLMM in a Bayesian framework. Specifically, the authors take $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_8)$ to be conditionally i.i.d. Gaussian with zero mean and variance σ^2, which is given an inverse gamma prior, i.e.

$$T_i = \mathrm{e}^{\boldsymbol{x}_i'\boldsymbol{\beta}} \cdot V_i, , \quad i = 1, \ldots, n, \tag{3}$$

with

$$\boldsymbol{x}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 \log(x_i) + \sum_{j=1}^{8} \alpha_j z_{j,i}, \tag{4}$$

$$V_i|\vartheta \overset{\text{iid}}{\sim} Weibull(\vartheta, 1), \quad i = 1, \ldots, n, \tag{5}$$

and prior

$$
\begin{aligned}
&\beta_0 \sim N(\mu_0, \sigma_0^2), \quad \beta_1 \sim N(\mu_1, \sigma_1^2)\\
&\alpha_1, \ldots, \alpha_8 | \lambda \overset{\text{iid}}{\sim} N(0, \lambda)\\
&\lambda \sim InvGamma(\frac{\tau_1}{2}, \frac{\tau_2}{2})\\
&\vartheta \sim Gamma(a_0, b_0),\\
&\beta_0,\ \beta_1,\ \boldsymbol{\alpha},\ \vartheta \text{ independent.}
\end{aligned} \tag{6}
$$

Finally, they provide posterior point and interval estimates of the conditional first percentile at stress level 23.4MPa and of the conditional median at 22.5MPa of the failure times for all the 8 spools, for the mean spool and for a new random spool. A precise description of the quantile functions considered is given later in Section 4.3. Their results for all the 8 spools and for the mean spool are comparable to those in Crowder et al. (1991)'s, while the interval estimates for the vessel wrapped with Kevlar from a new random spool are quite larger. They conclude that the interval estimates are too wide to make statements about the reliability of the pressure vessels.

Argiento et al. (2010a) present a Bayesian semiparametric approach to the AFT model for the dataset, observing as the nonparametric component in their model induces a grouping criterion on the observations. As Leon et al. (2007), they consider the failure times from all the 8 spools and include the censored times. The model can be described hierarchically in log-scale as follows:

$$
\begin{aligned}
&\log T_i = \beta_1 x_i + \log \vartheta_{2,i} + \frac{W_i}{\vartheta_{1i}}, \quad i = 1, \ldots, n\\
&W_i \overset{i.i.d.}{\sim} \text{Gumbel}(0, 1)\ ,\\
&\theta_i = (\vartheta_{1i}, \vartheta_{2i}) | P \overset{iid}{\sim} P,
\end{aligned} \tag{7}
$$

where P has a nonparametric distribution, namely P is a NGG. The prior is completed assuming independence between β_1 and P, and a Gaussian distribution for β_1. The authors draw an interesting analogy between the Bayesian mixed-effects model (3)-(5) and model (7). Indeed, the terms $\{\log \vartheta_{2,i}\}$ hold the same place of the α_j's in (4), 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 P. Then, the grouping of observations in Argiento et al. (2010a) is not fixed (as dictated by the spool number), but is random and is inferred from the data. Even if the authors obtain narrow posterior credibility intervals, it should be noted that their grouping criterion cannot exactly preserve the hierarchical nature of the data, since it is not unequivocal and can be influenced by unobserved latent factors.

4 A Bayesian Semiparametric Mixed-Effects model

4.1 The model

Here we generalize the Bayesian AFT model of Leon et al. (2007), described in (3)-(6), relaxing the normality assumption for the random-effects parameters; in fact, we allow now the distribution for the spool effect to be multi-modal, in order not to over-estimate its variance. The model we propose is similar to that in Kleinmann and Ibrahim (1998), where the authors used a Dirichlet process prior for the random effects, while here we use a NGG prior. As mentioned in the Introduction, the NGG process prior is ruled by two scalar parameters (instead of one of the Dirichlet process) controlling the clustering mechanism, with the aim of obtaining a more flexible prior than the Dirichlet process. The likelihood we consider is as in (3), with

$$\boldsymbol{x_i'\beta} = \beta_1 \log(x_i) + \sum_{j=1}^{8} \alpha_j z_{j,i} \tag{8}$$

and

$$V_i|\vartheta \overset{\text{iid}}{\sim} Weibull(\vartheta, 1), \quad i = 1, \ldots, n, \tag{9}$$

where x_i represents the stress and $z_{j,i} = 1$ if the i-th observation is from spool j (and 0 otherwise). The prior specifications for the parameters are

$$\begin{aligned}
&\vartheta \sim Gamma(a_0, b_0), \quad \beta_1 \sim N(\mu_1, \sigma_1^2) \\
&\alpha_1, \ldots, \alpha_8 | P \overset{\text{iid}}{\sim} P \\
&P \sim NGG(\sigma, \kappa, P_0) \\
&P_0|\mu, \lambda \sim N(\mu, \lambda), \quad \lambda \sim InvGamma(\frac{\tau_1}{2}, \frac{\tau_2}{2}) \quad \mu \sim N(\mu_0, \sigma_0^2), \\
&\beta_1, \ \boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_8), \ \vartheta \text{ independent.}
\end{aligned} \tag{10}$$

By (3), (8)-(10) and the scaling property of the Weibull distribution, the failure times $\{T_i\}$ have conditional distribution

$$T_i|\beta_1, \boldsymbol{\alpha}, \boldsymbol{x_i}, \vartheta \overset{\text{ind}}{\sim} Weibull(\vartheta, \exp\{\beta_1 \log(x_i) + \sum_{j=1}^{8} \alpha_j z_{j,i}\}).$$

Our model can be considered as a straightforward generalization of a GLMM, using the two-parameter Weibull instead of an exponential family distribution. In this framework, the overall mean spool effect is represented by the conditional (random) mean $\mathbb{E}(\alpha_j|P) = \int_{\Re} xP(dx)$; this is clear if we rewrite the linear predictor (8) as

$$\boldsymbol{x_i'\beta} = \mathbb{E}(\alpha_1 \mid P) + \beta_1 \log(x_i) + \sum_{j=1}^{8} (\alpha_j - \mathbb{E}(\alpha_1 \mid P))\, z_{j,i}.$$

In the previous equality the random-effects parameters have zero conditional mean, while $\mathbb{E}(\alpha_1 \mid P)$ takes the place of β_0 in (4). In this paper the focus is not on the estimation of the overall mean effect, and therefore we will not report results on it. However, we refer to Jara et al. (2009), where this quantity is estimated by sampling the trajectories of the posterior mixing distribution P, and to Li, Müller, and Lin (2007), where an approximation of its distribution is given.

4.2 The NGG process

The family of the NGG processes was first introduced as a nonparametric prior by Regazzini et al. (2003). The NGG, indexed by σ, κ and P_0, selects discrete distributions. The parameter σ, which assumes values in (0,1), controls the clustering, while $\kappa > 0$ plays the role of the mass parameter as in the Dirichlet process, which is recovered when $\sigma = 0$. The distribution P_0 represents the mean probability measure of the process, i.e. $\mathbb{E}[P(A)] = P_0(A)$ for all A, and it is usually assumed absolutely continuous. Moreover, both σ and κ control the overall variance of the process, and therefore their prior elicitation is quite difficult. If $X_1, X_2, \ldots, X_n$ is a sample from P, the predictive distribution of X_{n+1}, given $X_1, X_2, \ldots, X_n$, is a mixture of the mean measure P_0 and a weighted version of the empirical distribution, i.e.

$$X_{n+1}|X_1, X_2, \ldots, X_n \sim w_0(n, k; \sigma, \kappa) P_0(\cdot) + w_1(n, k; \sigma, \kappa) \sum_{j=1}^{k} (e_j - \sigma)\delta_{\psi_j}(\cdot), \tag{11}$$

where $\psi_1, \psi_2, \ldots, \psi_k$ are the distinct values among $\{X_1, X_2, \ldots, X_n\}$, $e_j = \#\{X_i : X_i = \psi_j,\ 1 \leq i \leq n\}$,

$$\begin{aligned} w_0(n, k; \sigma, \kappa) &= \frac{\sigma}{n} \frac{\sum_{i=0}^{n} \binom{n}{i} (-1)^i (\frac{\kappa}{\sigma})^{i/\sigma} \Gamma(k + 1 - i/\sigma; \frac{\kappa}{\sigma})}{\sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i (\frac{\kappa}{\sigma})^{i/\sigma} \Gamma(k - i/\sigma; \frac{\kappa}{\sigma})} \\ w_1(n, k; \sigma, \kappa) &= \frac{1}{n} \frac{\sum_{i=0}^{n} \binom{n}{i} (-1)^i (\frac{\kappa}{\sigma})^{i/\sigma} \Gamma(k - i/\sigma; \frac{\kappa}{\sigma})}{\sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i (\frac{\kappa}{\sigma})^{i/\sigma} \Gamma(k - i/\sigma; \frac{\kappa}{\sigma})}, \end{aligned} \tag{12}$$

for any $k = 1, \dots, n$. Here $\Gamma(\alpha, x) := \int_x^{+\infty} \mathrm{e}^{-t} t^{\alpha-1} dt$ denotes the incomplete gamma function. ¿From (11) it is clear that, if $X_1, X_2, \dots, X_n$ is a sample from the NGG process, there is a positive probability of having coincident values among the X_j's. Therefore the number of distinct values among the $(X_1, X_2, \dots, X_n)$ is a random variable, denoted by K_n, with values in $\{1, 2, \dots, n\}$; see its distribution in Lijoi et al. (2007), Sect 3.

4.3 Quantile functions

In our analysis we use the same data as in Leon et al. (2007), including failure times from all the 8 spools and censored times at 23.4MPa. We are interested in the prediction, in a Bayesian framework, of two quantiles of the failure time distribution at two different stress levels for a new and a given spool. The interest is motivated by the comparison of the quantiles estimates in Leon et al. (2007) under the parametric model (3)-(6) on one hand, and the semiparametric one (8)-(10) on the other. In particular, if α_{new} denotes the parameter corresponding to a *new* spool generated from the population of spools, the functions we will consider are:

(i) the first percentile of the conditional failure time distribution at the lowest stress level (23.4MPa), given the parameters $\theta, \beta_1, \alpha_{\text{new}}$,

$$F^{-1}_{T^{\text{new}}_{n+1}|\theta,\beta_1,\alpha_{\text{new}},x=23.4}(0.01);$$

(ii) the median of the conditional failure time distribution at the (extrapolated) level of stress of 22.5MPa, given the parameters $\theta, \beta_1, \alpha_{\text{new}}$,

$$F^{-1}_{T^{\text{new}}_{n+1}|\theta,\beta_1,\alpha_{\text{new}},x=22.5}(0.5).$$

We can easily obtain the conditional quantile functions for a given spool j, replacing α_{new} with α_j in the above quantities. All these conditional quantile functions can be expressed as deterministic transformations of the parameter vector, appearing in the conditioning, by inverting the Weibull distribution function; for instance, the quantiles in (i) and (ii) are obtained as the solution of the relationships

$$1 - \exp\left(- \left(\frac{t}{\mathrm{e}^{\beta_1 \log(x) + \alpha_{\text{new}}}} \right)^{\theta} \right) = 0.01 \text{ or } 0.5, \text{ when } x = 23.4 \text{ or } 22.5.$$

We will compute the whole posterior distributions of these quantities, in order to give not only point-, but also interval-estimates for them. To this aim, and to make a comparison among the spools, we will build a Gibbs sampler that converges to the posterior distribution of the whole parameter vector $(\beta_1, \alpha_1, \ldots, \alpha_8, \alpha_{\mathrm{new}}, \vartheta, \mu, \lambda)$; see the prior in (10).

We point out that the quantile functions we are considering do depend on the nonparametric component P only through the parameter α_{new}. Usually, when there is a nonparametric component in the model, the quantile functions of interest are those depending directly on the nonparametric component P. For instance, in Argiento et al. (2010a), the authors considered the quantile obtained by numerical inversion of $F(t;P) = \int_\Theta \left(1 - \mathrm{e}^{-(t/\vartheta_2)^{\vartheta_1}}\right) P(d\vartheta_1, d\vartheta_2)$, where P is the random probability distribution in (7), which is given a *NGG* prior; we refer to Gelfand and Kottas (2002) or Argiento *et al.* (2010b) for a detailed discussion on estimation of nonparametric functional under Dirichlet process mixture models or NGG-mixture model, respectively. However, in this paper, we integrate out P, which parametrize the spool population distribution, in order to compare the estimates of the same function of the parameter under the parametric and the semiparametric GLMMs.

As far as the Gibbs sampler is concerned, only few information on it is provided here. Recall that the Gibbs sampler algorithm samples sequentially from each full-conditional, i.e. the law of a parameter given all the others and the data. In this case, to sample from the full-conditionals of each α_j, we must sample from the predictive distribution of a sample from a NGG process, as described in (11), yielding a generalized Pólya urn scheme, and adapt Algorithm 8 of Neal (2000) for non-conjugate priors to the NGG case. The full-conditionals of β_1 and ϑ are log-concave, so that an adaptive rejection sampling method (see Robert and Casella, 2004) can be adopted. A final step of the algorithm for imputing the right-censored times and include them into the Markov Chain is required. Analytic expressions of the full-conditionals, as well as all the details on this Gibbs sampler, are provided in Soriano (2010).

5 Results

In this section we analyse the data and compare our results to those of Leon et al. (2007); for ease of reading, some of their results are reported in Table 2 below. In particular, we provide interval estimates for the quantiles described in Section 4.3

for spool 2, 4 and 7 and for a new random spool, as well as a classification of the 8 spools.

Table 2: Interval estimates of the quantiles under the parametric model of Leon et al. (2007).

Spool	2.5%	50%	97.5%
2	153.2	362.4	732.6
4	1773	4524	10060
7	56	131	305
new	22.0	671	19290

(a) 1st percentile failure time in hours at 23.4MPa.

Spool	2.5%	50%	97.5%
2	17.4	28.56	46.42
4	185.8	356.6	682.8
7	4.72	9.19	17.9
new	1.9	53.7	1479

(b) Median failure time in thousands of hours at 22.5MPa.

First, some details on computation and prior elicitation are given. We run the algorithm for 100000 iterations, while the first 80000 iterations were discarded, using a thinning of 4 to reduce the autocorrelation of the Markov chain. The final sample size was 5000. We run longer chains but we do not obtain any relevant reduction of the Monte Carlo error, and some diagnostic convergence tests were done. Concerning hyperparameters specifications (see (10)), we chose fairly vague priors for β_1, μ and λ, fixing $\mu_1 = \mu_0 = 0$, $\sigma_1^2 = \sigma_0^2 = 1000$, $\tau_1 = \tau_2 = 0.2$. Moreover, we assumed $a_0 = 1.5$. $b_0 = 5$ such that the marginal distribution of $\log V$ has finite first moment. A robustness analysis (Soriano, 2010) showed that the inferences are not particularly sensitive to the choice of those hyperparameters, while they are sensitive to the NGG processes hyperparameters σ and κ.

In Table 3, we provide credibility intervals of the two quantiles for a new random spool for different choices of σ and κ. Since K_8, the number of distinct values in $(\alpha_1, \ldots, \alpha_8)$ is a random variable with support $\{1, \ldots, 8\}$, and its distribution depends on σ and κ, we fixed σ and κ such that the prior expected value of K_8 is around 1, 2 or 4. The value 2 was fixed to translate the prior information that there are at least two groups of spools, those that are close to the nominal characteristics like spool 2, and those that do not respect the nominal conditions like spool 7; values 1 and 4 were considered to test the robustness of the NGG prior. Table 3 displays the posterior expected values of K_8 as well. Observe that K_8 represents the number of distinctive groups among $(\alpha_1, \ldots, \alpha_8)$. Both quantiles of interest are sensitive to the values of the hyperparameters: in particular the intervals get larger and the medians get higher

Table 3: Interval estimates of the quantiles for a new random spool for different values of hyperparameters in the NGG process prior.

σ	κ	$\mathbb{E}[K_J]$	$\mathbb{E}[K_J \| \boldsymbol{T}]$	2.5%	50%	97.5%
0.01	0.01	1.1	2.8	58.7	382.5	4401.1
0.01	0.5	2.0	4.0	52.6	454.4	5620.8
0.1	0.3	2.0	4.1	55.4	472.1	6077.9
0.3	0.09	2.0	5.0	46.9	525.8	7447.5
0.01	2.5	4.0	5.4	49.0	570.8	9357.2
0.1	2	4.0	5.4	39.3	570.6	8529.3
0.3	1.2	4.1	5.7	41.3	577.5	9179.0

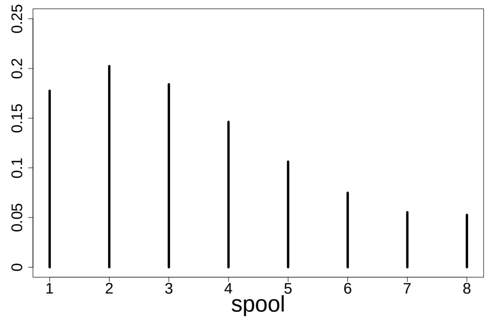
(a) 1st percentile failure time in hours at 23.4MPa.

σ	κ	$\mathbb{E}[K_J]$	$\mathbb{E}[K_J \| \boldsymbol{T}]$	2.5%	50%	97.5%
0.01	0.01	1.1	2.8	8.6	46.6	509.2
0.01	0.5	2.0	4.0	7.5	48.5	558.8
0.1	0.3	2.0	4.1	7.8	49.0	598.3
0.3	0.09	2.0	5.0	6.5	51.0	700.8
0.01	2.5	4.0	5.4	6.1	55.3	838.0
0.1	2	4.0	5.4	5.0	54.7	771.3
0.3	1.2	4.1	5.7	5.2	55.2	892.0

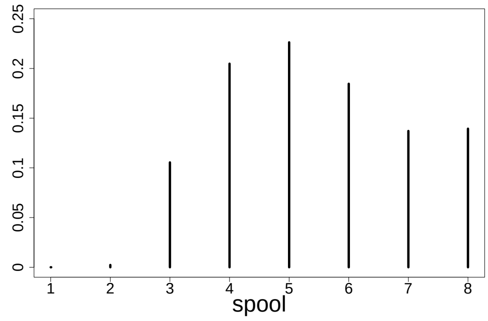
(b) Median failure time in thousands of hours at 22.5MPa.

under a larger a priori expected number of clusters. Notice that all the intervals are narrower than those obtained by the parametric model of Leon et al. (2007) (see the bottom row in Table 2).

To robustify inferences with respect to σ and κ, we have assumed a prior distribution on $(\sigma, \kappa) \in (0,1) \times (0, +\infty)$. In particular, we adopted a discrete prior on a grid of 30×30 points in $(0,1) \times (0,33)$, which assigns high probability to those (σ, κ) such that the prior expected number of clusters K_8 is around 3. Of course, an upper bound for κ different from 33 could be chosen. Figure 1a displays the prior distribution of K_8, marginalized with respect to the prior for (σ, κ). Notice that the prior distribution is rather diffuse, still reflecting our prior information about the different types of spool. On the other hand, the posterior distribution of K_8 in Figure 1b puts no mass to 1 and



(a) Prior



(b) Posterior

Figure 1: Probability distribution of the number of clusters K_8 when (σ, κ) has prior distribution.

2, but the mode is 5. The posterior distribution of (σ, κ) was found to be significantly different from the prior, and does not concentrate mass on $\sigma = 0$. This means that in this case, a better fit is obtained with an NGG prior with σ considerably larger than 0, i.e. *far from* the Dirichlet prior. Table 4 displays credibility intervals of the two quantiles for a new random spool and spool 2, 4, 7, under the prior described above, while Figure 2 shows the marginal posterior distributions of the 8 spool parameters and of a new random spool. The credibility intervals for a new random spool we obtain here are still narrower than those of Leon et al. (2007); the intervals for spool 2, 4, 7 are slightly different: in particular, the interval estimates for the median at 22.5MPa are a bit higher and wider.

Moreover, we provide a classification of the *behaviour* of the different spools from Figure 2: with small uncertainty, it is clear that we have three different groups of spools, the *worst* (spool 3 and 7), the *average* (spool 2, 5 and 6), and the *best* (spool 1, 4 and, probably, 8). The clustering into three groups is more evident here than in the past parametric analysis and coherent with Crowder et al. (1991) spool ranking. We can guess that spool 3 has the same microscopic characteristics of spool 7 (see Section 2), and that spool 5, 6 have the nominal ones, as spool 2. To support the conclusions from Figure 2, as in Medvedovic and Sivaganesan (2002), we used an hierarchical agglomerative clustering based on a similarity matrix $\hat{\pi}$ formed using the observed clusterings in the Gibbs sampler. Namely, the matrix $\hat{\pi}$ is the mean (over all the MCMC iterations) of the incidence matrices describing the partition of $\{1, \ldots, 8\}$ induced by the $(\alpha_1, \cdots, \alpha_8)$ MCMC samples. In Figure 3 we show the dendogram from the agglomerative clustering algorithm with single linkage and similarity $\hat{\pi}$. Cutting the dendogram where the gap between two successive combination similarities is largest provides the "natural" grouping (e.g., see Johnson and Wichern, 1988). In this case, the dendogram confirms our conclusion about the clusterization of the spools into three groups.

Table 4: Interval estimates of the quantiles for the semiparametric model when (σ, κ) has prior distribution.

Spool	2.5%	50%	97.5%
2	130.2	356.2	789.2
4	1247.9	3645.6	9855.6
7	38.8	111.4	275.8
new	40.8	546.3	9073.2

(a) 1st percentile failure time in hours at 23.4MPa.

Spool	2.5%	50%	97.5%
2	20.0	37.0	65.7
4	185.5	370.0	851.3
7	5.5	11.6	24.6
new	5.5	51.8	856.8

(b) Median failure time in thousands of hours at 22.5MPa.

We computed the same estimates under a different prior on (σ, κ), such that the prior distribution of K_8 is almost uniform on the support $\{1, \ldots, 8\}$, ignoring, in a sense, the prior information from the other analyses. However, the posterior inferences on the quantiles and the eight spools were very similar to those in Table 4 and Figure 2. Moreover, the marginal posterior distributions of the spools, as well as their classification, remain essentially unchanged if we do not assume the prior distribution on $(\sigma,\ \kappa)$, but we fix their values as in Table 4.

6 Conclusions

We have presented a new Bayesian semiparametric model to fit a dataset of 108 lifetimes of Kevlar pressure vessels with two covariates (spool and stress). The likelihood has the form of an AFT model, where the stress is a fixed effect and the spool is random. This model follows the framework of generalized linear mixed models, where the (conditional) distribution of the random-effects parameters $\alpha_1 \ldots, \alpha_J$ is modelled nonparametrically. In particular, we assumed a prior for the random-effects parameters which is more flexible than the Dirichlet process prior, namely $(\alpha_1 \ldots, \alpha_J)$ is a sample from a NGG process.

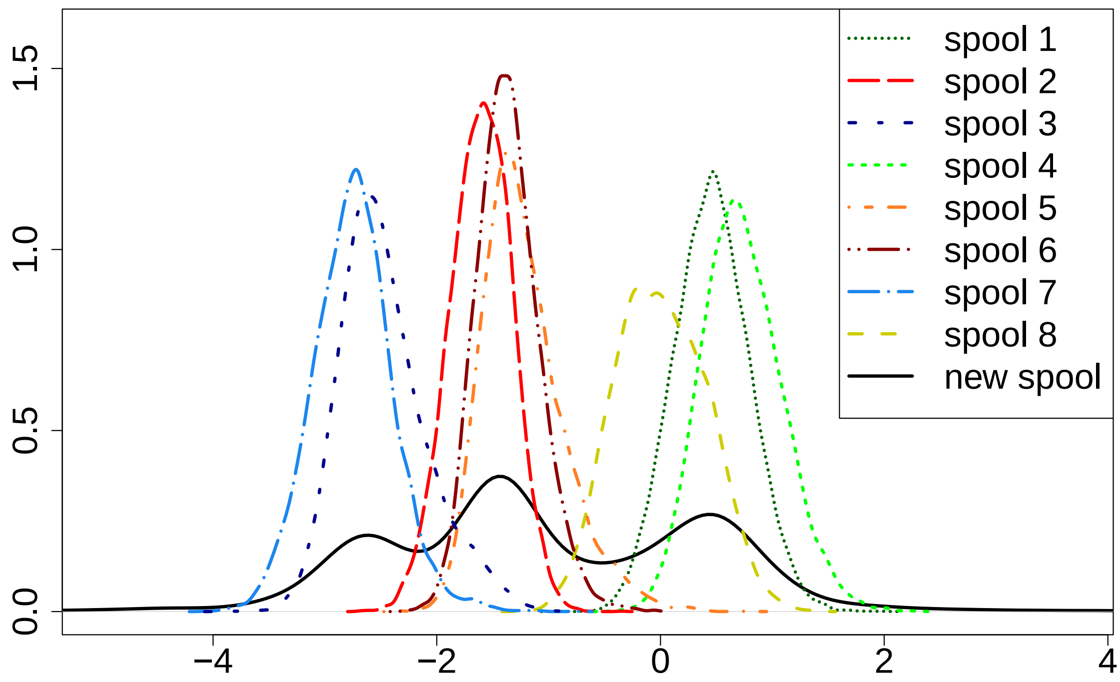


Figure 2: Marginal posterior distributions of the random effect parameters when (σ, κ) has prior distribution.

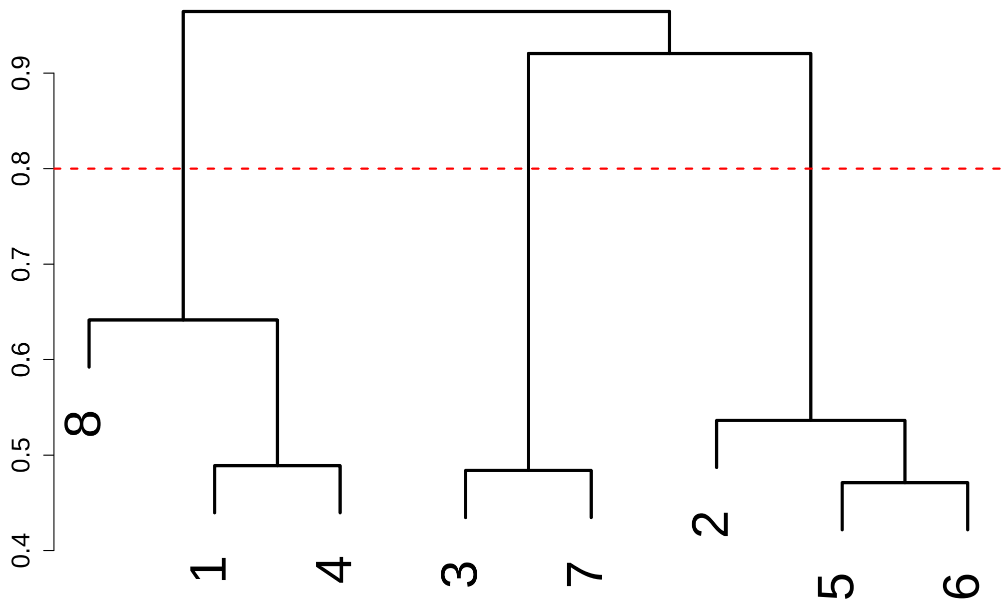


Figure 3: Dendogram from agglomerative hierarchical clustering with single linkage.

The dataset we have analyzed was recorded from an ALT test on NASA composite pressure vessels, which are critical components of the Space Shuttle. The reliability of the pressure vessels is still relevant: in 2009 NASA organized the Composite Pressure Vessel and Structure Summit (see the webpage of the meeting `http://www.nasa.gov/centers/wstf/news/safetysummit2009.html`) to discuss the state of the art on this subject. Several authors have made a statistical analysis of the dataset. For this reason we have also provided a short review of the most interesting papers analyzing the dataset. In particular, the model we propose presents an analogy with the Bayesian parametric mixed-effects model in Leon et al. (2007). The advantages of our model consist in a more flexible prior for the random-effects parameters, which includes the Dirichlet process prior suitably specifying one of the parameters ($\sigma = 0$) of the NGG process.

The credibility intervals of the two quantiles of interest for a new spool were narrower than those obtained by Leon et al. (2007). However, they were sensitive to hyperparameters σ and κ of the NGG process prior. To robustify inferences with respect to σ and κ, we have assumed an weakly informative prior distribution on them. The credibility intervals of the two quantiles for a new random spool are still narrower than those under the parametric model. The plot of the marginal posterior distributions of the random-effects parameters shows a very clear clustering of the spools into three different groups (the worst, the average and the best spools), which is coherent with the frequentist spool ranking in Crowder et al. (1991). As far as the reliability of the pressure vessels is concerned, we have obtained better interval estimates of the quantiles for a new random spool than before, but they are still too wide to conclude that the pressure vessels are fully reliable. Based on the results of our analysis, as well as on all the analyses mentioned here, we think that the log-linear relationship between survival time and stress level is not linear for lower stress values. This issue could be addressed by assuming a model with a non-linear link function or a covariates-dependent error prior. Anyhow, the analysis of such models is outside the scope of this paper and could be left as future work.

References

Argiento, R., Guglielmi, A. and Pievatolo, A. (2010a). Mixed-effects modelling of Kevlar fibre failure times through Bayesian nonparametrics. In *Complex data* *modeling and computationally intensive statistical methods*, eds. Mantovan, P. and Secchi, P., Springer, New York.

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

Chaloner, K. (1991). Bayesian residuals analysis in presence of censoring. *Biometrika*, **31**, 651-659.

Crowder, M.J., Kimber, A.C., Smith, R.L. and Sweeting, T.J. (1991). *Statistical analysis of reliability data*, Chapman and Hall, London.

Feiveson, A.H. and Kulkarni, P.M. (2000). Reliability of Space-Shuttle pressure vessels with random batch effects. *Technometrics*, **42**, 332–344.

Gelfand, A.E. and Kottas, A. (2002). A Computational Approach for Full Nonparametric Bayesian Inference Under Dirichlet Process Mixture Models *Journal of Computational and Graphical Statistics*, **11**, 289–305.

Gerstle, F.P. and Kunz, S.C. (1983). Prediction of Long-Term Failure in Kevlar 49 Composites. In long behaviour composites. *American Society for Testing and Materials*, Philadelphia, 263–292.

Glaser, R.E. (1983). *Statistical Analysis of Kevlar 49 Epoxy Composite Stress-Rupture Data*. Report UCID-19849, Lawrence Livermore National Laboratory.

Kleinman, K.P. and Ibrahim, J.G. (1998). A semi-parametric Bayesian approach to generalized linear mixed models. *Statistics in Medicine*, **17**, 2579–2596.

Jara, A., Hanson, T. and Lesaffre, E. (2009). Robustifying generalized linear mixed models using a new class of mixture of multivariate Polya trees. *Journal of Computational and Graphical Statistics*, **18**, 838–860.

Johnson, R.A. and Wichern, D.W. (1988). *Multivariate statistics, a practical approach*, Chapman and Hall, London.

Leon, R.V., Ashby, A.J. and Thyagarajan, J. (2007). Bayesian modeling of accelerated life tests with random effects. *Journal of quality technology*, **39**, 3–13.

Li, Y., Müller, P. and Lin, X. (2011). Center-adjusted inference for a nonparametric Bayesian random effect distribution. *Statistica Sinica*, in press.

Lijoi, A., Mena, R. H. and Prünster, I. (2007). Controlling the reinforcement in Bayesian nonparametric mixture models. *Journal of the Royal Statistical Society B*, **69**, 715–740.

Medvedovic M. and Sivaganesan, S. (2002). Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics*, **18**, 1194–1206.

Müller, P. and Quintana, F.A. (2004). Nonparametric Bayesian Data Analysis. *Statistical Science*, **19**, 95–110.

Neal, R.M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, **9**, 249–265.

Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *Annals of Statistics*, **31**, 560–585.

Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods.* Springer-Verlag, New York.

Soriano, J. (2010). Bayesian semiparametric approaches to mixed-effects models with an application in reliability analysis. *Master's thesis*, Politecnico di Milano.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

14/2011 Antonietti, P.F.; Mazzieri, I.; Quarteroni, A.; Rapetti, F.
Non-Conforming High Order Approximations for the Elastic Wave Equation

13/2011 Lombardi, M.; Parolini, N.; Quarteroni, A.; Rozza, G.
Numerical simulation of sailing boats: dynamics, FSI, and shape optimization

12/2011 Miglio E.; Cuffaro M.
Tectonic evolution at mid-ocean ridges: geodynamics and numerical modeling

11/2011 Formaggia, L.; Minisini, S.; Zunino, P.
Stent a rilascio di farmaco: una storia di successo per la matematica applicata

10/2011 Zunino, P.; Vesentini, S.; Porpora, A.; Soares, J.S.; Gautieri, A.; Redaelli, A.
Multiscale computational analysis of degradable polymers

09/2011 Pigoli, D.; Sangalli, L.
Wavelets in Functional Data Analysis: estimation of multidimensional curves and their derivatives

08/2011 Ieva, F.; Paganoni, A.M.; Secchi, P.
Mining Administrative Health Databases for epidemiological purposes: a case study on Acute Myocardial Infarctions diagnoses

07/2011 Arioli, G.; Gamba, M.
Automatic computation of Chebyshev polynomials for the study of parameter dependence for hyperbolic systems

06/2011 Secchi, P.; Stamm, A.; Vantini, S.
Large p Small n Data: Inference for the Mean

05/2011 Arioli, G.; Gamba, M.
An algorithm for the study of parameter dependence for hyperbolic systems