DPpackage: Bayesian Semi- and Nonparametric Modeling in R

Alejandro Jara
Universidad de Concepción

Timothy E. Hanson
University of Minnesota

Fernando A. Quintana
Pontificia Universidad Católica de Chile

Peter Müller and **Gary L. Rosner**
University of Texas M.D. Anderson Cancer Center

Abstract

Data analysis sometimes requires the relaxation of parametric assumptions in order to gain modeling flexibility and robustness against mis-specification of the probability model. In the Bayesian context, this is accomplished by placing a prior distribution on a function space, such as the space of all probability distributions or the space of all regression functions. Unfortunately, posterior distributions ranging over function spaces are highly complex and hence sampling methods play a key role. This paper provides an introduction to a simple, yet comprehensive, set of programs for the implementation of some Bayesian non- and semi-parametric models in R, **DPpackage**. Currently **DPpackage** includes models for marginal and conditional density estimation, ROC curve analysis, interval-censored data, binary regression data, item response data, longitudinal and clustered data using generalized linear mixed models, and regression data using generalized additive models. The package also contains functions to compute pseudo-Bayes factors for model comparison, and for eliciting the precision parameter of the Dirichlet process prior. To maximize computational efficiency, the actual sampling for each model is carried out using compiled FORTRAN.

Keywords

Bayesian semiparametric analysis; Random probability measures; Random functions; Markov chain Monte Carlo; R

1. Introduction

In many practical situations, a parametric model cannot be expected to coherently describe the chance mechanism generating an observed dataset. Unrealistic features of some common models (e.g., the thin tails of the normal distribution when compared to the distribution of the observed data) can lead to unsatisfactory inferences. Constraining the analysis to a specific parametric form may limit the scope and type of inferences that can be drawn from such models. In these situations, we would like to relax parametric assumptions in order to gain modeling flexibility and robustness against mis-specification of a parametric statistical model. In the Bayesian context such flexible inference is typically achieved by placing a prior distribution on infinite-dimensional spaces, such as the space of all probability distributions for a random

Alejandro Jara Department of Statistics Universidad de Concepción Avenida Esteban Iturra S/N Barrio Universitario Concepción, Chile Telephone: +56-41-2203163 Fax: +56-41-2251529 ajarav@udec.cl URL: http://www2.udec.cl/~ajarav.

variable of interest. These models are usually referred to as Bayesian nonparametric (BNP) or semiparametric (BSP) models depending on whether all or at least one of the parameters is infinite dimensional (see, e.g. Dey, Müller, and Sinha, 1998; Walker, Damien, Laud, and Smith, 1999; Ghosh and Ramamoorthi, 2003; Müller and Quintana, 2004; Hanson, Branscum, and Johnson, 2005).

BNP is a relatively young research area in statistics. First advances were made in the sixties and seventies, and were primarily mathematical formulations. It was only in the early nineties with the advent of sampling based methods, in particular Markov Chain Monte Carlo (MCMC) methods, that substantial progress has been made. Posterior distributions ranging over function spaces are highly complex and hence sampling methods play a key role. The introduction of MCMC methods in the area began with the work of Escobar (1994) for Dirichlet process mixtures. A number of themes are still undergoing development, including issues in theory, methodology and applications. We refer to Walker *et al*. (1999), Müller and Quintana (2004) and Hanson *et al*. (2005) for recent overviews.

While BNP and BSP are extremely powerful and have a wide range of applicability, they are not as widely used as one might expect. One reason for this has been the gap between the type of software that many applied users would like to have for fitting models and the software that is currently available. The most general programs currently available for Bayesian inference are BUGS (see, e.g. Gilks, Thomas, and Spiegelhalter, 1994) and OpenBugs (Thomas, O'Hara, Ligges, and Sibylle, 2006). BUGS can be accessed from the publicly available R program (R Development Core Team, 2009), using the **R2WinBUGS** package (Strurtz, Ligges, and Gelman, 2005). OpenBugs can run on Windows and Linux, as well as from inside R. In addition, various R packages exist that directly fit particular Bayesian models. We refer to Appendix C in Carlin and Louis (2008), for an extensive list of software for Bayesian modeling. Although the number of fully Bayesian programs continues to burgeon, with many available at little or no cost, they generally do not include semiparametric models. An exception to this rule is the R package **bayesm** (Rossi, Allenby, and McCulloch, 2005; Rossi and McCulloch, 2008), including functions for some models based on Dirichlet process priors (Ferguson, 1973). The range of different Bayesian semiparametric models is huge. It is practically impossible to build flexible and efficient software for the generality of such models.

In this paper we present an up to date introduction to a publicly available R (R Development Core Team, 2009) package designed to help bridging the previously mentioned gap, the **DPpackage**, originally presented in Jara (2007). Although the name of the package is due to the most widely used prior on the space of the probability distributions, the Dirichlet Process (DP) (Ferguson, 1973), the package includes many other priors on function spaces. Currently, **DPpackage** includes models considering DP (Ferguson, 1973), mixtures of DP (MDP) (Antoniak, 1974), DP mixtures (DPM) (Lo, 1984; Escobar and West, 1995), linear dependent DP (LDDP) (De Iorio, Müller, Rosner, and MacEachern, 2004; De Iorio, Johnson, Müller, and Rosner, 2009), weight dependent DP (WDDP) (Müller, Erkanli, and West, 1996), hierarchical mixture of DPM of normals (HDPM) (Müller, Quintana, and Rosner, 2004), centrally standardized DP (CSDP) (Newton, Czado, and Chapell, 1996), Polya Trees (PT) (Ferguson, 1974; Mauldin, Sudderth, and Williams, 1992; Lavine, 1992, 1994), mixtures of Polya trees (MPT) (Lavine, 1992, 1994; Hanson and Johnson, 2002; Hanson, 2006; Christensen, Hanson, and Jara, 2008), mixtures of triangular distributions (Perron and Mengersen, 2001), and random Bernstein polynomials (Petrone, 1999a,b; Petrone and Wasserman, 2002). The package also includes models considering Penalized B-Splines (Lang and Brezger, 2004).

The article is organized as follows. Section 2 reviews the general syntax and design philosophy. Although the material in this section was presented in Jara (2007), its inclusion here is necessary in order to make the paper self-contained. In Section 3 the available functions are described in

detail. In Section 4 the main features and usages of **DPpackage** are illustrated by means of simulated and real life data analyses. We conclude with additional comments and discussion in Section 5.

2. Design philosophy and general syntax

The design philosophy behind **DPpackage** is quite different from the one of a general purpose language. The most important design goal has been the implementation of model-specific MCMC algorithms. A direct benefit of this approach is that the sampling algorithms can be made dramatically more efficient than in a generic environment.

Fitting a model in **DPpackage** begins with a call to an R function, for instance, DPmodel, or PTmodel. Here "model" denotes a descriptive name for the model being fitted. Typically, the model function will take a number of arguments that control the specific MCMC sampling strategy adopted. In addition, the model(s) formula(s), data, and prior parameters are passed to the model function as arguments. The common elements in any model function are:

   i. prior: an object list which includes the values of the prior hyper-parameters.

   ii. mcmc: an object list which must include the integers nburn giving the number of burn-in scans, nskip giving the thinning interval, nsave giving the total number of scans to be saved, and ndisplay giving the number of saved scans to be displayed on screen: the function reports on the screen when every ndisplay iterations have been carried out and returns the process's runtime in seconds. For some specific models, one or more tuning parameters for Metropolis steps may be needed and must be included in this list. The names of these tuning parameters are explained in each specific model description in the associated help files.

   iii. state: an object list giving the current value of the parameters, when the analysis is the continuation of a previous analysis, or giving the starting values for a new Markov chain, which is useful to run multiple chains starting from different points.

   iv. status: a logical variable indicating whether it is a new run (TRUE) or the continuation of a previous analysis (FALSE). In the latter case the current value of the parameters must be specified in the object state.

Inside the R model function the inputs are organized in a more useable form, the MCMC sampling is performed by calling a shared library written in a compiled language, and the posterior sample is summarized, labeled, assigned into an output list, and returned. The output list includes:

   i. state: a list of objects containing the current value of the parameters.

   ii. save.state: a list of objects containing the MCMC samples for the parameters. This list contains two matrices randsave and thetasave which contain the MCMC samples of the variables with random distribution (errors, random effects, etc.) and the parametric part of the model, respectively.

In order to exemplify the extraction of the output elements, consider the abstract model fit:

```
fit <- DPmodel (…, prior, mcmc,
state, status, ….)
```

The lists can be extracted using the following code:

```
fit$state
fit$save.state$randsave
fit$save.state$thetasave
```

Based on these output objects, it is possible to use, for instance, the **boa** (Smith, 2007) or the **coda** (Plummer, Best, Cowles, and Vines, 2006) R packages to perform convergence diagnostics. For illustration, we consider the **coda** package here. It requires a matrix of posterior draws for relevant parameters to be saved as a mcmc object. Assume that we have obtained `fit1`, `fit2`, and `fit3`, by independently running a model function three times, specifying different starting values each. To compute the Gelman-Rubin convergence diagnostic statistic for the first parameter stored in the `thetasave` object, the following commands may be used:

```
library(coda)
coda.obj <- mcmc.list(
chain1=mcmc(fit1$save.state$thetasave[,1]),
chain2=mcmc(fit2$save.state$thetasave[,1]),
chain3=mcmc(fit3$save.state$thetasave[,1]))
gelman.diag(coda.obj, transform = TRUE)
```

Note that the second command line saves the results as a `mcmc.list` object of class and the third command line computes the Gelman-Rubin statistic from these three chains.

Generic R functions such as `print, plot, summary,` and anova have methods to display the results of the **DPpackage** model fit. The function `print` displays the posterior means of the parameters in the model, and `summary` displays posterior summary statistics (mean, median, standard deviation, naive standard errors, and credibility intervals). By default, the function summary computes the 95% HPD intervals using the Monte Carlo method proposed by Chen and Shao (1999). The user can display the order statistic estimator of the 95% credible interval by using the following code,

```
summary(fit, hpd=FALSE)
```

The `plot` function displays the trace plots and a kernel-based estimate of the posterior distribution for the parameters of the model. Similarly to `summary`, the `plot` function displays the 95% HPD regions in the density plot and the posterior mean. The same plot but considering the the 95% credible region can be obtained by using,

```
plot(fit, hpd=FALSE)
```

The anova function computes simultaneous credible regions for a vector of parameters from the MCMC sample using the method described by Besag, Green, Higdon, and Mengersen (1995). The output of the anova function is an anova-like table containing the pseudo-contour probabilities for each of the factors included in the linear part of the model.

3. Implemented Models

In this section we describe in detail the functions available in version 1.0-8 of **DPpackage**.

3.1. Marginal density estimation

DPdensity, PTdensity, TDPdensity, and BDPdensity functions implement models for marginal density estimation using DPM of normals, MPT, triangular-Dirichlet, and a Bernstein-Dirichlet prior, respectively. The first two functions allow the user to fit uni- and multi-variate models. We next introduce the notation used for each model along with the associated computational approaches used to fit the models.

Dirichlet Process Mixtures of Normals—The DPdensity function considers the multivariate extension of the univariate DPM of normals model presented in Escobar and West (1995). Let y_i be a k-dimensional vector of measurements for the ith subject, $i = 1,\ldots, n$. The model assumes

$$\boldsymbol{y}_i | G \stackrel{iid}{\sim} \int N_k \left(\boldsymbol{y}_i | \mu, \boldsymbol{\Sigma} \right) dG \left(\mu, \boldsymbol{\Sigma} \right),$$

and

$$G | \alpha, G_0 \sim DP \left(\alpha G_0 \right),$$

where, the baseline distribution, G_o, corresponds to the conjugate normal-inverted-Wishart distribution

$$G_0 \equiv N_k \left(\mu | \boldsymbol{m}_1, \kappa_0^{-1} \boldsymbol{\Sigma} \right) IW_k \left(\boldsymbol{\Sigma} | \nu_1, \Psi_1 \right).$$

To complete the model specification, the following independent hyper-priors are assumed,

$$\alpha \sim \Gamma \left(a_0, b_0 \right),$$

$$\boldsymbol{m}_1 | \boldsymbol{m}_2, \boldsymbol{S}_2 \sim N_k \left(\boldsymbol{m}_2, \boldsymbol{S}_2 \right),$$

$$\kappa_0 | \tau_1, \tau_2 \sim \Gamma \left(\tau_1/2, \tau_2/2 \right),$$

and

$$\Psi_1 | \nu_2, \Psi_2 \sim IW_k \left(\nu_2, \Psi_2 \right).$$

Note that the inverted-Wishart prior, $W | \nu, \Psi \sim IW_k (\nu, \Psi)$, is parameterized such that $E(W) = \frac{1}{\nu - k - 1} \Psi^{-1}$.

The computation implementation is based on the marginalized version of the model where the random probability measure G is integrated out. Although the baseline distribution, G_0, is a conjugate prior in this model specification, the algorithms with auxiliary parameters described in MacEachern and Müller (1998) and Neal (2000) are adopted. Specifically, the no-gaps

algorithm of MacEachern and Müller (1998) and algorithm 8 of Neal (2000), with $m = 1$, are considered. The default method is algorithm 8 of Neal (2000).

Mixtures of Polya trees—The current implementation of the `PTdensity` function considers a MPT model as in Hanson (2006). As in the previous section, let y_i be a k-dimensional vector of measurements for the ith subject, $i = 1,\ldots, n$. The model assumes

$$y_i | G \overset{iid}{\sim} G,$$

and

$$G | \alpha, \mu, \Sigma, MPT^M \left(\Pi^{\mu,\Sigma}, \mathscr{A}^\alpha \right),$$

where M is the maximum level of the partition to be updated (the default value is $M = \infty$), $\Pi^{\mu,\Sigma} = \{\pi_j\}_{j\ 0}$ is a set of partitions of $\mathbb{R}^k$, indexed by μ and Σ, and A^α is a family of non-negative vectors controlling the variability of the process indexed by α. Following Hanson (2006), the PT is centered around the $N_k (\mu, \Sigma)$ distribution by taking

$$\mathscr{A}^\alpha = \left\{ \gamma^\alpha (j, r) : r \in \left\{1, \ldots, 2^{j-1}\right\}^k, j=1, \ldots \right\},$$

with $\gamma^\alpha(j, r) = aj^2 1_k$, and further taking each level j of the sequence of partitions in $\Pi^{\mu,\Sigma}$, as the sets arising from a location-scale transformation $\mu + \Sigma^{1/2} z$ of the Cartesian products of intervals obtained as quantiles from the standard univariate normal distribution, where $\Sigma^{1/2}$ is the Cholesky decomposition of Σ. Notice that we consider a different parameterization than the one considered by Hanson (2006), were $\Sigma^{1/2}$ is taken to be the unique symmetric square root of Σ. The base sets for level j are given by

$$B_0 (j, p) = \left(\Phi^{-1} \left((p_1 - 1) / 2^j\right), \Phi^{-1} \left((p_1) / 2^j\right) \right] \times \cdots \times \left(\Phi^{-1} \left((p_k - 1) / 2^j\right), \Phi^{-1} \left((p_k) / 2^j\right) \right],$$

for vectors $p = (p1,\ldots,p_k)$ with $p_i \in \{1,\ldots, 2^j\}$, $i = 1,\ldots, k$. The location-scale transformation applied to each base set yields the final sets $B(j, p) = \{\mu + \Sigma^{1/2} z : z \in B_0(j, p)\}$, such that $\pi_j = \{B(j,p) : p \in \{1,\ldots, 2^j\}^k\}$.

The model specification is completed by assuming the following hyper-priors

$$p (\mu, \Sigma) \propto |\Sigma|^{-(k+1)/2},$$

and

$$\alpha | a_0, b_0 \Gamma (a_0, b_0).$$

As noticed by Jara, Hanson, and Lesaffre (2009), the PT prior specification is dependent on the square root of the centering covariance matrix considered to define the partitions sets. Indeed, in the $N_k (\mu, \Sigma)$-centered multivariate extension considered by Hanson (2006), the direction of the sets are completely defined through the decomposition of the covariance matrix

by the unique symmetric square root. In the context of multivariate random effects distributions, Jara *et al.* (2009) proposed a novel mixture of PT priors where the effect of the partitions is smoothed over by mixing over the decomposition of the centering covariance matrix (see, Section 3.2). This option will be considered in future version of the package.

For univariate analyses using a finite ($M < \infty$) PT, a full version of the model is considered where the Dirichlet vectors are updated during the MCMC scheme. For univariate analysis with a fully specified PT ($M = \infty$) and for multivariate analyses, a marginalized version of the model is considered, where the random probability measure G is integrated out. The baseline parameters μ and Σ, and the precision parameter α are updated using Metropolis-Hastings (MH) steps (Tierney, 1994).

Bernstein-Dirichlet prior—The function `BDPdensity` consider density estimation using a Bernstein-Dirichlet prior (BDP) proposed by Petrone (1999a,b). For a continuous cdf G on (0,1], the associated Bernstein polynomial (BP) is defined as

$$B(x|k, G) = \sum_{j=0}^{k} G(j/k) \begin{pmatrix} k \\ j \end{pmatrix} x^j (1-x)^{k-j},$$

which is a mixture of beta distributions. Its density is given by

$$b(x|k, G) = \sum_{j=1}^{k} (G(j/k) - G((j-1)/k)) \, \beta(x|j, k-j+1),$$

where $\beta(x|j, k-j+1)$ stands for a beta density with parameters j and $k-j+1$. Petrone (1999a,b) proposed a hierarchical prior, called the Bernstein polynomial prior (BPP), where the random density f($\cdot$) is given by the following mixture of beta densities,

$$f(x) = \sum_{j=1}^{k} w_{j,k} \beta(x|j, k-j+1), \quad (1)$$

where $w_{j,k} = G(j/k) - G((j-1)/k)$, k as probability mass function $\rho(\cdot)$, and given k, $w_k = (w_{1,k}, \ldots, w_{k,k})$ has distribution $H_k(\cdot)$ on the k-dimensional simplex

$$\Delta_k = \left\{ (\omega_1, \ldots, \omega_k) : 0 \le \omega_j \le 1, \ j = 1, \ldots, k, \ \sum_{j=1}^{k} \omega_j = 1 \right\}.$$

Petrone (1999a,b) called expression (1) the Bernstein polynomial density with parameters k and w_k, and shows that to assume $w_k = (w_{1,k}, \ldots, w_{k,k}) \sim$ Dirichlet($\zeta_{1,k}, \ldots, \zeta_{k,k}$), with $\zeta_{j,k} = \alpha(G_0(j/k) - G_0((j-1)/k))$, $j = 1, \ldots, k$, G_0 a probability distribution on (0,1] and α a positive constant, is equivalent to assume that $G \mid \alpha, G_0 \sim DP(\alpha G_0)$. Petrone (1999a,b) refers to this as the Bernstein-Dirichlet prior (BDP) and discussed MCMC algorithms to scan the posterior distribution.

Our MCMC implementation is similar to the one described by Petrone (1999a,b) but adds the resampling step described by Bush and MacEachern (1996) for Dirichlet process mixture models. The function `BDPdensity` considers

$$y_i|G \overset{iid}{\sim} G,$$

and

$$G|kmax, \alpha, G_0 \tilde{} BDP(kmax, \alpha G_0),$$

where y_i is the data transformed to lie in $(0,1]$ and $G_0 = Beta(a_0, b_0)$. It is further assumed that

$$\alpha|aa_0, \ ab_0 \tilde{} Beta(aa_0, ab_0),$$

and

$$k|kmax \tilde{} DU(\{1, \ldots, kmax\}),$$

where $DU(A)$ refers to the discrete uniform distribution on the set A. Although BDP are naturally defined as probability models for distributions on the unit interval $(0,1)$, different measurable mappings could be considered to transform the data when the support is not the unit interval. With this aim we consider the uniform CDF on the range of the data.

Mixtures of triangular distributions—The `TDPdensity` function considers a triangular-Dirichlet prior (TDP) for univariate density estimation. The logic behind the TDP is similar to the BDP construction but replaces the beta kernels in the mixture model by triangular distributions as proposed by Perron and Mengersen (2001). The model is given by

$$y_i|G \overset{iid}{\sim} G,$$

and

$$G|kmax, \alpha, G_0 \tilde{} TDP(kmax, \alpha G_0),$$

$$\alpha|aa_0, \ ab_0 \tilde{} Beta(aa_0, ab_0),$$

and

$$k|kmax \tilde{} DU(\{1, \ldots, kmax\}),$$

where y_i is the data transformed to lie in $(0,1]$, $kmax$ is the upper limit of the discrete uniform prior or the number of components in the mixture of Triangular distributions, α is the total mass parameter of the Dirichlet process component, and G_0 is the centering distribution of the DP. The centering distribution corresponds to a $G_0 = Beta(a_0, b_0)$ distribution.

Our representation is equivalent to the mixture of triangular distributions proposed by Perron and Mengersen (2001), with random weights following a Dirichlet prior. However, in this function we exploit the underlying DP structure, thus avoiding the use of Reversible-Jump algorithms (Green, 1995). In fact, the same MCMC algorithm considered for the BDP prior is implemented in the `TDPdensity` function.

3.2. Nonparametric random effects distributions in mixed effects models

Assume that for each of m experimental units the regression data (Y_{ij}, x_{ij}, z_{ij}), $1 \leq i \leq m$, $1 \leq j \leq n_i$, is recorded, where Y_{ij} is a response variable, and $x_{ij} \in \mathbb{R}^p$ and $z_{ij} \in \mathbb{R}^q$ are vectors of p and q explanatory variables, respectively. Let $Y_i = (Y_{n}1, \ldots, Y_{in_i})^T$, $X_i = (x_i1, \ldots, x_{in_i})^T$, and $Z_i = (z_i1, \ldots, z_{in_i})^T$, $i = 1, \ldots, m$. The observations are assumed to be conditionally independent with exponential family distribution,

$$p\left(Y_{ij}|\vartheta_{ij}, \tau\right) = \exp\left\{\left[Y_{ij}\vartheta_{ij} - b\left(\vartheta_{ij}\right)\right]/\tau\right\} c\left(Y_{ij}, \tau\right).$$

The means $\mu_{ij} = E(Y_{ij} \mid \vartheta_{ij}, \tau)$ and variances $\sigma_{ij}^2 = Var\left(Y_{ij}|\vartheta_{ij}, \tau\right)$ are related to the canonical ϑ_{ij} and dispersion parameter τ via $\mu_{ij} = b'(\vartheta_{ij})$ and $\sigma_{ij}^2 = \tau b''\left(\vartheta_{ij}\right)$, respectively. The means μ_{ij} are related to the p-dimensional and q-dimensional "fixed" effects vectors β^F and β^R, respectively, and the q-dimensional "random" effects vector b_i via the link relation

$$h\left(\mu_{ij}\right) = \eta_{ij} = x'_{ij}\beta^F + z'_{ij}\beta^R + z'_{ij}b_i, \quad (2)$$

where, $h(\cdot)$ is a known monotonic differentiable link function, and η_{ij} is called the linear predictor. Due to software limitations, the analyses are often restricted to the setting in which the random effects follow a multivariate normal distribution, $b_1, \ldots, b_m|\Sigma \overset{iid}{\sim} N_q(0, \Sigma)$. In this context, Bayesian nonparametric extensions incorporate a probability model for the random effects distribution in order to better represent the distributional uncertainty and to avoid the effects of the miss-specification of an arbitrary parametric random effects distribution. Bush and MacEachern (1996) and Kleinman and Ibrahim (1998b) describe Bayesian semiparametric versions of the linear mixed model considering DP prior for the random effects distribution. Under this approach the DP prior is centered at a normal base mesure with zero mean. Similar approaches were considered by Mukhopadhyay and Gelfand (1997) and Kleinman and Ibrahim (1998a) in the context of GLMM. In order to avoid the discrete nature of the DP realizations, Müller and Rosner (1997) consider a DPM of normals model in the context of a normal nonlinear mixed model. Alternatively, Walker and Mallick (1997) and Hanson (2006) consider PT and mixtures of PT priors in random intercept models. Jara *et al.* (2009) propose a novel mixture of multivariate PT priors to define flexible nonparametric models for multivariate distributions that reduces the undesirable sensitivity to the choice of the partitions associated with the PT constructions. Under these approaches, the parametric assumption is relaxed by considering

$$b_1, \ldots, b_m|G \overset{iid}{\sim} G,$$

and

$$G|H \sim H,$$

where H is one of the previously mentioned probability models for probability distributions. We will specify the nonparametric priors in more detail next, but first it is necessary to discuss some important issues regarding the specification of the semiparametric model. Specifically, it is important to stress that under parametrization (2), β^R represents the mean of random effects, and b_i represents the subject-specific deviation from the mean. It follows that fixing the mean of the normal prior distribution for the random effects b at zero in the parametric context

corresponds to an identification restriction for the model parameters (see e.g., Newton, 1994; San Martín, Jara, Rolin, and Mouchart, 2007). Equivalently, the random probability measure must be appropriately restricted in a semiparametric GLMM specification. In our settings, the location of G is "confounded" with the parameters β^R. Although such identification issues present no difficulties to a Bayesian analysis in the sense that a prior is transformed into a posterior using the sampling model and the probability calculus, if the interest focuses on a "confounded" parameter then such formal assurances have little practical value. Furthermore, as more data become available, the posterior mass will not concentrate on a point in the model, making asymptotic analysis difficult. As pointed out by Newton (1994), from a computational point of view, identification problems imply ridges in the posterior distribution and MCMC methods can be difficult to implement in these situations.

Following Jara *et al.* (2009), we consider the following re-parameterization of the model

$$\eta_{ij} = x'_{ij}\beta + z'_{ij}\theta_i,$$

$$\theta_1, \ldots, \theta_m | G \overset{iid}{\sim} G,$$

and

$$G|H \quad H$$

where $\beta = \beta^F$, and $\theta_i = \beta^R + b_i$, and we center the nonparametric priors for G at a $N_q(\mu, \Sigma)$ distribution. Notice that samples under the original parameterization can be obtained in a straightforward manner from MCMC samples as explained in Jara *et al.* (2009) for PT priors. For DP or DPM priors the ϵ–DP approximation proposed by Muliere and Tardella (1998) is considered, with $\epsilon = 0.01$. The latter is similar to the approach proposed by Gelfand and Kottas (2002) who considered a fixed truncation to the DP. When a DP or DPM prior is used to model the random effects distribution, Dunson, Yang, and Baird (2007a) and Li, Müller, and Lin (2007) proposed alternative strategies to avoid the identifiability problem described above but these approaches are not implemented in the current version of **DPpackage**.

The functions DPlmm, DPglmm, and DPolmm implement mixed effects models using a DP prior for G such that

$$G|\alpha, \mu, \Sigma \quad DP\left(\alpha N_q(\mu, \Sigma)\right).$$

The functions DPMlmm, DPMglmm, and DPolmm consider a DPM of normals prior for G such that

$$G|\Sigma_k, H \int N_q(\mu, \Sigma_k)\, dP(\mu),$$

and

$$P|\alpha, \mu, \Sigma \quad DP\left(\alpha N_q(\mu, \Sigma)\right).$$

The functions `PTlmm`, `PTglmm`, and `PTolmm` consider a multivariate PT prior for *G* such that

$$G|\alpha, \mu, \Sigma, O \sim PT\left(\Pi^{\mu,\Sigma,O}, \mathscr{A}^\alpha\right),$$

where *O* is a $q \times q$ orthogonal matrix defining the "direction" of the partition sets. The models are completed by assuming the following prior distributions:

$$\beta \sim N_p\left(\beta_0, S_{\beta_0}\right),$$

$$\tau^{-1}|\tau_1, \tau_2 \sim \Gamma\left(\tau_1/2, \tau_2/2\right),$$

$$\mu|\mu_b, S_b \sim N_q\left(\mu_b, S_b\right),$$

$$\Sigma|\nu_0, T \sim IW_k\left(\nu_0, T\right),$$

$$O \sim \mathrm{Haar}\left(q\right),$$

and

$$\alpha|a_0, b_0 \sim \Gamma\left(a_0, b_0\right),$$

where Γ and *IW* refers to the Gamma and inverted Wishart distributions, respectively. As before, the inverted Wishart prior is parameterized such that $E(\Sigma) = T^{-1}/(\nu_0 - q - 1)$.

The `DPlmm`, `DPMlmm` and `PTlmm` functions consider the normal sampling distribution with an identity link. The `DPglmm`, `DPMglmm`, and `PTglmm` functions include the following sampling distributions (link): binomial (logit and probit), Poisson (log) and gamma (log). The `DPolmm`, `DPMolmm` and `PTolmm` consider a multinomial sampling distribution and an ordered-probit link function.

In all functions, a marginalized version of the semiparametric GLMM is considered where the random probability distribution *G* is integrated out. For the multinomial and probit-binomial models, the latent variable approach of Albert and Chib (1993) is considered.

The computational implementation associated to the functions `DPMlmm` and `DPMolmm`, and to the probit-Bernoulli model included in the `DPMglmm` function, is based on the use of MCMC methods for conjugate priors for a collapsed state of MacEachern (1998). For the poisson, Gamma, and logit-binomial models included in the `DPglmm` and `DPMglmm` functions, MCMC methods for non-conjugate priors are used. Specifically, algorithm 8 of Neal (2000), with *m* = 1, is considered. In this case, a MH step with the iterative weighted least square (IWLS) normal proposal of Gamerman (1997) is used to update fixed and random effects.

For the functions `DPlmm` and `DPolmm`, and the probit-Bernoulli model included in `DPglmm`, the MCMC strategy described by Bush and MacEachern (1996) is employed. Finally, for the `PTlmm`, `PTgmm` and `PTomm` the modified IWLS proposal normal proposal described by Jara *et al.* (2009) is considered for sampling the random effects. In these functions, IWLS normal proposal of Gamerman (1997) is used to update the fixed effects in the nonconjugate case. The PT centering and precision parameters are updated using adaptive MCMC algorithms as described by Jara *et al.* (2009).

3.3. Semiparametric IRT-type models

Item response theory (IRT) models are widely used in educational measurement (see e.g., De Boeck and Wilson, 2004). Rasch-type models (Rasch, 1960) are typical examples of this class and can be viewed as a particular case of GLMM (see e.g., De Boeck and Wilson, 2004). In Rasch-type models, the linear predictor η_{ij} depends on two parameters in an additive way $\eta_{ij} = \theta_i - \beta_j$, where $\theta_i \in \mathbb{R}$ corresponds to the ability of subject i, $i = 1, \ldots, m$, and $\beta_j \in \mathbb{R}$ corresponds to the difficulty of probe/item j, $j = 1, \ldots, p$. The difficulty and ability parameters are interpreted as "fixed" and "random" effects, respectively. Two versions of the model are considered here: the Rasch model (RM) and the Rasch Poisson count model (RPCM). In the RM, Y_{ij} represents a binary variable coding the correct answer of individual i to the item j, such that

$$Y_{ij}|\theta_i, \beta_j \overset{ind.}{\sim} \text{Bernoulli}\left(\Psi\left(\theta_i - \beta_j\right)\right),$$

where $\Psi(x) = \exp(x)/(1 + \exp(x))$. In the RPCM the sampling distribution is given by

$$Y_{ij}|\theta_i, \beta_j \overset{ind.}{\sim} \text{Poisson}\left(\exp\left(\theta_i - \beta_j\right)\right),$$

where Y_{ij} is an "unbounded" count variable, typically representing the number of miss-reading / miss-copying for the subject i in the text j. We consider semiparametric versions of the models where the abilities distribution G is modeled using DP, PT and DPM priors. To avoid identification problems in the semiparametric specification of the model (see, San Martín *et al.*, 2007), we fixed the first difficulty parameter at 0 and consider a normal prior for the remaining elements in the vector

$$\beta_{2:p}|\beta_0, \boldsymbol{S}_{\beta_0} \sim N_{p-1}\left(\beta_0, \boldsymbol{S}_{\beta_0}\right).$$

The functions `DPrasch` and `DPraschpoisson` implement semiparametric versions of the RM and RPCM, respectively, where

$$\theta_i|G \overset{iid}{\sim} G,$$

and

$$G|\alpha, G_0 \sim DP\left(\alpha N\left(\mu, \sigma^2\right)\right),$$

In a similar way, the functions `FPTrasch` and `FPTraschpoisson` implement semiparametric versions of the RM and RPCM, respectively, using a finite PT prior,

$$G|\alpha, \mu, \alpha^2 PT^M\left(\Pi^{\mu, \sigma^2}, \mathscr{A}^\alpha\right),$$

where, the PT is centered around a $N(\mu, \sigma^2)$ distribution, by taking each m level of the partition Π^{μ, σ^2} to coincide with the $k/2^m$, $k = 0, \ldots, 2^m$ quantiles of the $N(\mu, \sigma^2)$ distribution. The family $\mathscr{A}^a = \{a_\epsilon : \epsilon \in E^*\}$, where $E^* = \bigcup_{m=1}^{\infty} E^m$ and E^m is the m-fold product of $E = \{0,1\}$, was specified as $a_{\epsilon_1 \ldots \epsilon_m} = am^2$. For the DP and PT priors, the model is completed by assuming

$$\alpha|a_0, b_0\Gamma\left(a_0, b_0\right),$$

$$\mu|\mu_b, S_b \tilde{} N\left(\mu_b, S_b\right),$$

and

$$\alpha^{-2}|\tau_1, \tau_2\Gamma\left(\tau_1/2, \tau_2/2\right).$$

The functions DPMrasch and DPMraschpoisson consider DPM of normals priors for the abilities distribution in a RM and RPCM, respectively, given by

$$\theta_i|G \overset{iid}{\sim} \int N\left(\mu, \sigma^2\right) dG\left(\mu, \sigma^2\right),$$

and

$$G|\alpha, G_0 \tilde{} DP\left(\alpha G_0\right),$$

where $G_0 \equiv N\left(\mu|\mu_b, \sigma_b^2\right) IG\left(\sigma^2|\tau_{k1}, \tau_{k2}\right)$. We further assume that

$$\alpha|a_0, b_0\Gamma\left(a_0, b_0\right),$$

$$\mu_b|m_0, s_0 \tilde{} N\left(m_0, s_0\right),$$

$$\sigma_b^{-2}|\tau_{b1}, \tau_{b2}\Gamma\left(\tau_{b1}/2, \tau_{b2}/2\right),$$

and

$$\tau_{k2}|\tau_{s1}, \tau_{s2}\Gamma\left(\tau_{s1}/2, \tau_{s2}/2\right).$$

In all functions, the difficulty and ability parameters are updated using a MH step with the IWLS normal proposal of Gamerman (1997). The computational implementation in the DPrasch and DPraschpoisson functions is based on the marginalization of the DP and on

the use of algorithm 8 of Neal (2000), with $m = 1$. The DPM implementations of functions `DPMrasch` and `DPMraschpoisson` are based on the finite approximation for DP proposed by Ishwaran and James (2002). Finally, the functions using finite PT priors for the abilities distribution, `FPTrasch` and `FPTraschpoisson`, fit a full version of the models where the PT conditional probabilities are updated during the MCMC scheme. In this case, the abilities, centering and precision parameters are updated using slice sampling (Neal, 2003).

3.4. Semiparametric meta-analysis models

The `DPmeta`, `DPMmeta` and `PTmeta` functions implement random (mixed) effects univariate metaanalysis models using a MDP, DPM of normals, and MPT prior for the random effects, respectively. In this case, the conditional model is given by

$$y_i|\theta_i, \beta, \sigma_i^2 \overset{ind.}{\sim} N\left(\theta_i + \boldsymbol{x}_i'\beta, \sigma_i^2\right),$$

where the variances σ_i^2 are known, $\boldsymbol{X_i}$ is a p-dimensional design vector, excluding an intercept term, and

$$\beta_p|\beta_0, \boldsymbol{S}_{\beta_0} \sim N_p\left(\beta_0, \boldsymbol{S}_{\beta_0}\right).$$

The `DPmeta` function assumes that

$$\theta_i|G \overset{iid}{\sim} G,$$

and

$$G|\alpha, \mu, \sigma^2 DP\left(\alpha N\left(\mu, \sigma^2\right)\right),$$

The `PTmeta` function, replaces the latter assumption by a PT prior,

$$G|\alpha, \mu, \sigma^2 PT\left(\Pi^{\mu,\sigma^2}, \mathscr{A}^\alpha\right),$$

where the PT prior is centered around a $N(\mu, \sigma^2)$ distribution. The `PTmeta` function can also center the PT prior around a $N(0, \sigma^2)$ distribution for the median-0 model described by Branscum and Hanson (2008). This model is fitted if the option `frstlprob` is set equal to TRUE in the model prior object. In this case, the design vector $\boldsymbol{x}_i$ includes an intercept term and the associated regression coefficient represents the median effect. The computational implementation of the `DPmeta` and `PTmeta` functions are based on the marginalization of the DP and PT, respectively. In both cases, the model specification is completed by assuming

$$\alpha|a_0, b_0 \Gamma\left(a_0, b_0\right),$$

$$\mu|\mu_b, S_b N\left(\mu_b, S_b\right),$$

and

$$\alpha^{-2}|\tau_1, \tau_2 \Gamma\left(\tau_1/2, \tau_2/2\right).$$

The the average effect in the `DPmeta` function is sampled using the method of composition and the ϵ–DP approximation proposed by Muliere and Tardella (1998), with $\epsilon = 0.01$. For the `PTmeta` function, the mean effect is sampled using the finite PT approximation described by Jara *et al.* (2009).

The `DPMmeta` function considers a location DPM of normals priors for the study effects

$$\theta_i|\sigma^2, G \stackrel{iid}{\sim} \int N\left(\mu, \sigma^2\right) dG\left(\mu\right),$$

$$\sigma^{-2}|\tau_{01}, \tau_{02} \Gamma\left(\tau_{01}/2, \tau_{02}/2\right),$$

and

$$G|\alpha, G_0 \tilde{} DP\left(\alpha G_0\right),$$

where $G_0 \equiv N\left(\mu|\mu_b, \sigma_b^2\right)$. This function further assumes that

$$\alpha|a_0, b_0 \Gamma\left(a_0, b_0\right),$$

$$\mu_b|m_b, S_b \tilde{} N\left(m_b, S_b\right),$$

and

$$\sigma_b^{-2}|\tau_{11}, \tau_{12} \Gamma\left(\tau_{11}/2, \tau_{12}/2\right).$$

The computational implementation of the model is based on the marginalization of the DP and on the use of MCMC methods for conjugate priors for a collapsed state, as presented in MacEachern (1998). The average effect is also sampled using the method of composition and the ϵ–DP approximation proposed by Muliere and Tardella (1998), with $\epsilon = 0.01$.

The function `DPmultmeta` implements a multivariate extension of the no-covariate model considered in the `DPmeta` function, given by

$$y_i|\theta_i, \boldsymbol{\Sigma}_i \stackrel{ind.}{\sim} N_k\left(\theta_i, \boldsymbol{\Sigma}_i\right),$$

$$\theta_i|G \stackrel{iid}{\sim} G,$$

and

$$G | \alpha, \boldsymbol{m}_1, \boldsymbol{S}_1 \tilde{} DP \left(\alpha N_k \left(\boldsymbol{m}_1, \boldsymbol{S}_1 \right) \right),$$

where the covariance matrices Σ_i are known. To complete the model specification, independent hyperpriors are assumed,

$$\alpha | a_0, b_0 \tilde{} \Gamma \left(a_0, b_0 \right),$$

$$\boldsymbol{m}_1 | \boldsymbol{m}_2, \boldsymbol{S}_2 \tilde{} N_k \left(\boldsymbol{m}_2, \boldsymbol{S}_2 \right),$$

and

$$\boldsymbol{S}_1 | \nu, \boldsymbol{\Psi} \tilde{} IW_k \left(\nu, \boldsymbol{\Psi} \right).$$

The computational implementation is similar to the one employed for the `DPmeta` function.

3.5. Accelerated failure time modeling for interval-censored data

The `DPsurvint` function implements the algorithm described by Hanson and Johnson (2004) for semiparametric accelerated failure time (AFT) models. The AFT regression model is given by

$$T_i \in [l_i, u_i), i=1, \ldots, n,$$

$$T_i = \exp \left(-\boldsymbol{x}_i' \beta \right) V_i,$$

$$\beta | \beta_0, \boldsymbol{S}_{\beta_0} \tilde{} N_p \left(\beta_0, \boldsymbol{S}_{\beta_0} \right),$$

$$V_1, \ldots, V_n | G \overset{iid}{\sim} G,$$

and

$$G | \alpha, \mu, \sigma^2 \tilde{} DP \left(\alpha LN \left(\mu, \sigma^2 \right) \right),$$

where $LN \left(v | \mu, \sigma^2 \right)$ refers to a log-normal distribution with location and scale parameter μ and σ^2, respectively. The model is completed by assuming independent hyperpriors,

$$\alpha \tilde{} \Gamma \left(a_0, b_0 \right),$$

$$\mu | m_0, s_0 \tilde{} N(m_0, s_0),$$

and

$$\sigma^{-2} | \tau_1, \tau_2 \Gamma(\tau_1/2, \tau_2/2).$$

The likelihood in the AFT model for interval censored data involves the product of indicator functions $\prod_{i=1}^{n} I(T_i \in A_i)$, where A_i is an interval in the sample space. This fact gives rise to algorithmic possibilities which are unavailable or very difficult to implement under standard hierarchical models with uncensored data. As described in Hanson and Johnson (2004), the `DPsurvint` function partially sample G, in order to sample $(V_1, \ldots, V_n, V_{n+1}, \beta, \alpha)$ with perfect accuracy. This can be performed by using the properties of DP. Specifically, the following representation of the process is considered

$$G = \sum_{j=1}^{M} G_j G^j,$$

where j indexes the intervals that define a finite partition of the sample space $\{B_1, \ldots, B_m\}$, $G_j = G(B_j)$, and $G^j(\cdot) = G(\cdot \mid B_j)$, with the G_j's being Dirichlet distributed random variables and the G^j's being independent Dirichlet processes. Therefore, G can be updated by first updating $\{G_j\}$ using Ferguson's definition of DP and then by updating each $G^j \mid \{G_j\}, \ldots$ using the Sethuraman (1994) stick-breaking representation of DP (see, e.g. Doss, 1994; Hanson and Johnson, 2004). Based on this, a MH step is used to update the regression coefficients, followed by updates of $V_1, \ldots, V_{n+1}$.

The function `predict.DPsurvint` can be used to extract posterior information about the survival curve based on the MCMC output. Given a sample of the parameters of size J, a sample of the survival curve for a given x is drawn as follows. For the jth MCMC scan of the posterior distribution, $j = 1, \ldots, J$, the survival function evaluated at t is sampled from where

$$S^{(j)}(t|x, \text{data}) \tilde{} Beta\left(a^{(j)}(t), b^{(j)}(t)\right),$$

where

$$a^{(j)}(t) = \alpha^{(j)} G_0^{(j)}\left(\left(t \exp\left(x'\beta^{(j)}\right), +\infty\right)\right) + \sum_{i=1}^{n} \delta_{v_i^{(j)}}\left(\left(t \exp\left(x'\beta^{(j)}\right), +\infty\right)\right),$$

and $b^{(j)}(t) = \alpha^{(j)} + n - a^{(j)}(t)$.

3.6. Binary regression with nonparametric link

Consider binary regression data, (Y_i, x_i), $1 \quad i \quad n$, where Y_i is a binary response variable ($Y_i \in \{0,1\}$) and $x_i \in \mathbb{R}^p$ is a vector of p explanatory variables. Parametric versions of this model are characterized by the following assumption

$$Pr\left(Y_i=1|\boldsymbol{x}_i,\theta\right) = E\left(Y_i=1|\boldsymbol{x}_i,\theta\right)$$
$$= F_\varphi\left(m\left(\beta,\boldsymbol{x}_i\right)\right),$$

where F_φ is a distribution function on $\mathbb{R}$, called the inverse link function in the context of generalized linear models, known up to a Euclidean parameter φ, and $m(\cdot)$ is a known function, called the index function, parameterized by β. Popular parametric versions include a linear index function, $m\left(\beta,\boldsymbol{x}_i\right)=\boldsymbol{x}_i'\beta$, and where F_φ is considered to be a known cumulative distribution function, i.e. with $\varphi = \varphi_0$, thus allowing relatively simple treatment of the finite regression parameters, $\theta = \beta$. The function Pbinary implements parametric versions of this model considering the logit, probit, cloglog, and Cauchy link functions.

The DPbinary, FPTbinary, and CSDPbinary functions replace the parametric inverse link function F_φ by a general distribution G and placing a DP prior,

$$G|\alpha, G_0 \tilde{} DP\left(\alpha G_0\right),$$

a finite PT where the first and second quartiles are fixed (Hanson, 2006),

$$G|\alpha \tilde{} PT^M\left(\Pi, \mathscr{A}^\alpha\right),$$

and a CSDP (Newton *et al.*, 1996),

$$G|p, d, h, G_0 \tilde{} CSDP\left(\alpha G_0, p, d, h\right)$$

on G, respectively. Newton *et al.* (1996) described the CSDP as a prior distribution on the space of the probability distribution with fixed location and scale in order to assure sampling identification. The reasoning behind their construction is presented here for completeness. The following definition is a slight modification of the one given by Newton *et al.* (1996). Let G_0 and H be two probability measures on $\mathbb{R}$ and $(0, d)$, respectively, such that for all $d > 0$, $G_0\left(\left(-\infty, -d\right)\right) > 0$ and $G_0\left(\left(d, \infty\right)\right) > 0$. Let $\theta \tilde{} h$, where h is the density of H with respect to Lebesgue measure. Given θ, define the following partition of the real line, $A_1(\theta) = \left(-\infty, \theta - d\right]$, $A_2(\theta) = \left(\theta - d, 0\right]$, $A_3(\theta) = \left(0, \theta\right]$, and $A_1(\theta) = \left(\theta, \infty\right)$. Finally, suppose that for each $\theta \in (0, d)$, the random probability measures φ_1, φ_2, φ_3, and φ_4 follow conditionally independent DP priors, $\varphi_i|\theta, \alpha, G_0 \overset{ind}{\tilde{}} DP\left(\alpha G_0 I_{\left(A_i(\theta)\right)}\right)$, $i = 1, \ldots, 4$. The random probability measure G on $(\mathbb{R}, \mathcal{B})$ is said to follow CSDP prior with parameter (α, G_0, p, d, h), written $G \sim CSDP(\alpha G_0, p, d, h)$, if,

$$G = \frac{1-p}{2}\left(\varphi_1+\varphi_4\right) + \frac{p}{2}\left(\varphi_2+\varphi_3\right) a.s.$$

In all cases, the functions allows for misclassified binary responses with known misclassification parameters and the model specification is completed by assuming

$$\alpha \tilde{} \Gamma\left(a_0, b_0\right),$$

and

$$\beta | \beta_0, \boldsymbol{S}_{\beta_0} \tilde{} N_p \left(\beta_0, \boldsymbol{S}_{\beta_0} \right).$$

The DPbinary function allows the user to center the DP around a logistic, normal or Cauchy distribution. The CSDPbinary function takes $H \equiv U(0, d)$ distribution and G_0 as the standard logistic distribution. In both functions, a latent variable representation

$$Y_i = I_{\{V_i \le \boldsymbol{x}_i^T \beta\}},$$

and

$$V_1, \ldots, V_n | G \tilde{} G,$$

is used, along with a MH step to update the regression coefficients. In the computational implementation of this model, G is considered as latent data and sampled partially with sufficient accuracy to be able to generate $V_1, \ldots, V_{n+1}$ such that are exactly iid random variables from G, as proposed by Doss (1994). Both Ferguson's definition of DP and the Sethuraman (1994)'s representation of the process are used. As in Bush and MacEachern (1996), an extra step which moves the clusters in such a way that the posterior distribution is still the stationary distribution, is performed in order to improve the mixing of the chain.

The FPTbinary function creates the partition sets based on the logistic distribution. In the computational implementation of the model, MH steps are used to update the regression coefficients and the precision parameter, as described in Hanson (2006).

3.7. ROC curve estimation

The DProc function performs a ROC curve analysis based on DPM of normals models for density estimation. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ be the diagnostic marker measurements for the healthy and diseased subjects, respectively. The model is given by

$$\boldsymbol{x}_i | G_{\boldsymbol{x}} \overset{iid}{\tilde{}} \int N \left(\mu_{\boldsymbol{x}}, \Sigma_{\boldsymbol{x}} \right) dG_{\boldsymbol{x}} \left(\mu_{\boldsymbol{x}}, \Sigma_{\boldsymbol{x}} \right),$$

$$\boldsymbol{y}_i | G_{\boldsymbol{y}} \overset{iid}{\tilde{}} \int N \left(\mu_{\boldsymbol{y}}, \Sigma_{\boldsymbol{y}} \right) dG_{\boldsymbol{y}} \left(\mu_{\boldsymbol{y}}, \Sigma_{\boldsymbol{y}} \right),$$

$$G_{\boldsymbol{x}} | \alpha_{\boldsymbol{x}}, G_{\boldsymbol{x}_0} \tilde{} DP \left(\alpha_{\boldsymbol{x}} G_{\boldsymbol{x}_0} \right),$$

and

$$G_{\boldsymbol{y}} | \alpha_{\boldsymbol{y}}, G_{\boldsymbol{y}_0} \tilde{} DP \left(\alpha_{\boldsymbol{y}} G_{\boldsymbol{y}_0} \right),$$

where, the baseline distributions, G_{z0}, $z = \{\boldsymbol{x}, \boldsymbol{y}\}$, correspond to the conjugate normal-inverted-Wishart distribution

$$G_{\boldsymbol{z}_0} \equiv N_k \left(\boldsymbol{\mu}_{\boldsymbol{z}} | \boldsymbol{m}_{\boldsymbol{z}1}, k_{\boldsymbol{z}0}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{z}} \right) IW_k \left(\boldsymbol{\Sigma}_{\boldsymbol{z}} | \nu_{\boldsymbol{z}1}, \boldsymbol{\Psi}_{\boldsymbol{z}1} \right).$$

To complete the model specification, the model is extended by assuming independent hyper-priors,

$$\alpha_{\boldsymbol{x}} \Gamma \left(a_{\boldsymbol{x}0}, b_{\boldsymbol{x}0} \right), \quad \alpha_{\boldsymbol{y}} \Gamma \left(a_{\boldsymbol{y}0}, b_{\boldsymbol{y}0} \right),$$

$$\boldsymbol{m}_{\boldsymbol{x}1} | \boldsymbol{m}_{\boldsymbol{x}2}, \boldsymbol{S}_{\boldsymbol{x}2} \tilde{\ } N_k \left(\boldsymbol{m}_{\boldsymbol{x}2}, \boldsymbol{S}_{\boldsymbol{x}2} \right), \quad \boldsymbol{m}_{\boldsymbol{y}1} | \boldsymbol{m}_{\boldsymbol{y}2}, \boldsymbol{S}_{\boldsymbol{y}2} \tilde{\ } N_k \left(\boldsymbol{m}_{\boldsymbol{y}2}, \boldsymbol{S}_{\boldsymbol{y}2} \right),$$

$$\kappa_{\boldsymbol{x}0} | \tau_{\boldsymbol{x}1}, \tau_{\boldsymbol{x}2} \Gamma \left(\tau_{\boldsymbol{x}1}/2, \tau_{\boldsymbol{x}2}/2 \right), \quad \kappa_{\boldsymbol{y}0} | \tau_{\boldsymbol{y}1}, \tau_{\boldsymbol{y}2} \Gamma \left(\tau_{\boldsymbol{y}1}/2, \tau_{\boldsymbol{y}2}/2 \right),$$

$$\boldsymbol{\Psi}_{\boldsymbol{x}1} | \nu_{\boldsymbol{x}2}, \boldsymbol{\Psi}_{\boldsymbol{x}2} \tilde{\ } IW_k \left(\nu_{\boldsymbol{x}2}, \boldsymbol{\Psi}_{\boldsymbol{x}2} \right), \quad \text{and} \quad \boldsymbol{\Psi}_{\boldsymbol{y}1} | \nu_{\boldsymbol{y}2}, \boldsymbol{\Psi}_{\boldsymbol{y}2} \tilde{\ } IW_k \left(\nu_{\boldsymbol{y}2}, \boldsymbol{\Psi}_{\boldsymbol{y}2} \right).$$

The survival and ROC curves are estimated by using a Monte Carlo approximation to the posterior means $E(G_{\mathbf{x}} | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ and $E(G_{\mathbf{y}} | \boldsymbol{y}_1 \ldots, \boldsymbol{y}_m)$, which is based on MCMC samples from posterior predictive distribution for a future observation. The optimal cut-off point is based on the efficiency of the test and is built on Cohen's kappa as defined in Kraemer (1992).

3.8. Median regression modeling

Consider regression data $(y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, where y_i is the response and x_i is a p-dimensional vector of predictors. By default, the `PTlm` function fits a median regression model using a scale MPT prior for the distribution of the errors (Hanson and Johnson, 2002),

$$Y_i = \boldsymbol{x}_i' \beta + V_i,$$

$$\beta \tilde{\ } N_p \left(\beta_0, \boldsymbol{S}_{\beta_0} \right),$$

$$V_i | G \overset{iid}{\sim} G,$$

and

$$G | \alpha, \sigma^2 \tilde{\ } PT \left(\Pi^{\sigma^2}, \mathscr{A}^\alpha \right),$$

where, the PT is centered around a $N(0, \sigma^2)$ distribution, by taking each m level of the partition Π^{σ^2} to coincide with the $k/2^m$, $k = 0, \ldots, 2^m$ quantiles of the $N(0, \sigma^2)$ distribution. The family

$\mathscr{A}^a = \{a_\epsilon : \epsilon \in E^*\}$, where $E^* = \bigcup_{m=1}^{\infty} E^m$ and E^m is the m-fold product of $E = \{0,1\}$, was specified as $a_{\epsilon_1 \ldots \epsilon_m} = am^2$. To complete the model specification, independent hyperpriors are assumed,

$$\alpha | a_0, b_0 \Gamma(a_0, b_0),$$

$$\sigma^{-2} | \tau_1, \tau_2 \Gamma(\tau_1/2, \tau_2/2).$$

Optionally, if frstlprob=FALSE (the default value is TRUE) is specified, a mean regression model is considered. In this case, the following PT prior is considered

$$G | \alpha, \mu, \sigma^2 PT\left(\Pi^{\mu, \sigma^2}, \mathscr{A}^\alpha\right),$$

where, the PT is centered around a $N(\mu, \sigma^2)$ distribution. In this case, the intercept term is automatically excluded from the model and the hyperparameters for the normal prior for μ must be specified. The normal prior is given by

$$\mu | \mu_b, S_b \tilde{N}(\mu_b, S_b).$$

In the computational implementation of the model, random-walk Metropolis steps are used to update the regression coefficients and hyperparameters.

3.9. Models for related distributions

The current version of **DPpackage** considers models for related random probability distributions based on particular implementations of the dependent DP (DDP) proposed by MacEachern (1999, 2000), a natural generalization of the approach discussed by Müller *et al.* (1996) for nonparametric regression to the context of conditional density estimation, and the hierarchical mixture of DPM models (HDPM) proposed by Müller *et al.* (2004). These approaches and the associated functions are described next.

Linear dependent Dirichlet process—MacEachern (1999, 2000), proposes the DDP as an approach to define a prior model for an uncountable set of random measures indexed by a single continuous covariate, say x, $\{G_x : x \in \chi \subset \mathbb{R}\}$. The key idea behind the DDP is to create an uncountable set of DPs (Ferguson, 1973) and to introduce dependence by modifying the Sethuraman (1994)'s stick-breaking representation of each element in the set. If G follows a DP prior with precision parameter α and base measure G_0, denoted by $G \sim DP(\alpha G_0)$, then the stick-breaking representation of G is

$$G(B) = \sum_{l=1}^{\infty} \omega_l \delta_{\theta_l}(B), \quad (3)$$

where B is a measurable set, $\delta_a(\cdot)$ is the Dirac measure at a, $\theta_l | G_0 \overset{iid}{\sim} G_0$ and $\omega_l = V_l \Pi_{j<l}(1 - V_j)$, with $V_l | \alpha \overset{iid}{\sim} Beta(1, \alpha)$. MacEachern (1999, 2000) generalizes (3) by assuming the point masses $\theta(x)_l$, $l = 1, \ldots$, to be dependent across different levels of x, but independent across l.

De Iorio *et al.* (2004) and De Iorio *et al.* (2009) proposed a particular version of the DDP where the component of the atoms defining the location in a DDP mixture model follows a linear regression model $\theta_l(x) = \left(x'\beta_l, \sigma_l^2\right)$, where x is a p-dimensional design vector. An advantage of this model for related random probability measures, referred to as the Linear DDP (LDDP), is that it can be represented as DPM of linear (in the coefficients) regression models. This approach is implemented in the LDDPdensity function, where for the regression data (y_i, x_i), $i = 1, \ldots, n$, the following model is considered

$$y_i | G \overset{ind.}{\sim} \int N\left(y_i | x_i'\beta, \sigma^2\right) dG\left(\beta, \sigma^2\right),$$

and

$$G | \alpha, G_0 \tilde{} DP\left(\alpha G_0\right),$$

where $G_0 \equiv N_p(\beta | \mu_b, S_b) \ \Gamma\left(\sigma^{-2} | \tau_1/2, \ \tau_2/2\right)$. The LDDP model specification is completed with the following hyper-priors

$$\alpha | a_0, b_0 \Gamma\left(a_0, b_0\right)$$

$$\tau_2 | \tau_{s_1}, \tau_{s_2} \Gamma\left(\tau_{s_1}/2, \tau_{2_2}/2\right),$$

$$\mu_b | m_0, S_0 \tilde{} N_p\left(m_0, S_0\right),$$

and

$$S_b | \nu, \Psi \mathit{I} W_p\left(\nu, \Psi\right).$$

The LDDPsurvival function implements this model in the context of survival data. Now let y_i the time to event for the ith subject. The LDDP mixture of survival models is given by

$$\log y_i | G \overset{ind.}{\sim} \int N\left(\log y_i | x_i'\beta, \sigma^2\right) dG\left(\beta, \sigma^2\right),$$

with the same hierarchical specification given above for the LDDPdensity function. Note that this function can deal with censored observations by using a data-augmented approach.

Finally, the LDDPrach and LDDPrachpoisson functions consider this modeling strategy in a Rasch and Rasch Poisson model context, respectively, as in Fariña, Quintana, San Martín, and Jara (2009). Here the linear predictor is given by $\eta_{ij} = \theta_i - \beta_j$, where the abilities follow a LDDP mixture of normals model based on subject-specific covariates included in x_i,

$$\theta_i | G \overset{ind.}{\sim} \int N\left(\theta_i | x_i'\beta, \sigma^2\right) dG\left(\beta, \sigma^2\right).$$

These functions fit a marginalized version of the models where the random probability measure *G* is integrated out. Full inference on the conditional density, and survival and hazard functions in the case of the LDDPsurvival function, at covariate level are obtained using the ϵ–DP approximation proposed by Muliere and Tardella (1998), with $\epsilon = 0.01$.

Weight dependent Dirichlet process—Let $x_i = \left(1, z_i'\right)'$, where z_i is a *p*-dimensional vector of continuous predictors. The LDDP of the previous section defines a mixture model where the weights are independent of the predictors *z*, given by

$$f_z(\cdot) = \sum_{l=1}^{\infty} \omega_l N\left(\cdot|\beta_{0l} + z'\beta_l, \sigma_l^2\right),$$

where the weights ω_l follow a stick-breaking construction and $\left(\beta_{0l}, \beta_l, \sigma_l^2\right) \overset{iid}{\sim} G_0$. Motivated by regression problems with continuous predictors different extensions have been proposed by making the weights dependent on covariates (see, e.g. Griffin and Steel, 2006; Duan, Guindani, and Gelfand, 2007; Dunson, Pillai, and Park, 2007b; Dunson and Park, 2008), such that

$$f_z(\cdot) = \sum_{l=1}^{\infty} \omega_l(z) N\left(\cdot|\beta_{0l} + z'\beta_l, \sigma_l^2\right). \quad (4)$$

An earlier approach that is related to the latter references and that also induces a weight-dependent DP model, as in expression (4), was discussed by Müller *et al.* (1996). These authors fitted a "standard" DPM of multivariate Gaussian distributions to the complete data $d_i = (y_i, z_i)'$, $i = 1, \ldots, n$, and looked at the induced conditional distributions. Although Müller *et al.* (1996) focused on the mean function only, $m(z) = E(y|z)$, their method can be easily extended to provide inferences for the conditional density at covariate level *z*, i.e. a "density regression" model in the spirit of Dunson *et al.* (2007b). The extension of the approach of Müller *et al.* (1996) for related probability measures is implemented in the DPcdensity function, where the model is given by

$$d_i|G \overset{ind.}{\sim} \int N_k(d_i|\mu, \Sigma)\, dG(\mu, \Sigma),$$

and

$$G|\alpha, G_0 \sim DP(\alpha G_0),$$

where $k = p + 1$ is the dimension of the vector of complete data d_i, the baseline distribution G_0 is the conjugate normal-inverted-Wishart (IW) distribution $G_0 \equiv N_2\left(\mu|m_1, k_0^{-1}\Sigma\right)$. To complete the model specification, the following hyper-priors are assumed

$$\alpha|a_0, b_0 \sim \Gamma(a_0, b_0),$$

$$m_1|m_2, S_2 \sim N_2(m_2, S_2),$$

$$\kappa_0|\tau_1, \tau_2 \Gamma(\tau_1/2, \tau_2/2),$$

and

$$\boldsymbol{\Psi}_1|\nu_2, \boldsymbol{\Psi}_2 \tilde{} IW_2(\nu_2, 2).$$

This model induce a weight dependent mixture models, as in expression (4), where the components are given by

$$\omega_l(z) = \frac{\omega_l N_p(z|\mu_{2l}, \boldsymbol{\Sigma}_{22l})}{\sum_{j=1}^{\infty} \omega_j N_p(z|\mu_{2j}, \boldsymbol{\Sigma}_{22j})}$$

$$\beta_{0l} = \mu_{1l} - \boldsymbol{\Sigma}_{12l} \boldsymbol{\Sigma}_{22l}^{-1} \mu_{2l},$$

$$\beta_l = \boldsymbol{\Sigma}_{12l} \boldsymbol{\Sigma}_{22l}^{-1},$$

and

$$\sigma_l^2 = \sigma_{11l}^2 - \boldsymbol{\Sigma}_{12l} \boldsymbol{\Sigma}_{22l}^{-1} \boldsymbol{\Sigma}_{21l},$$

where the weights ω_l follow a DP stick-breaking construction and the remaining elements arise from the standard partition of the vectors of means and (co)variance matrices given by

$$\mu_l = \begin{pmatrix} \mu_{1l} \\ \mu_{2l} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Sigma}_l = \begin{pmatrix} \sigma_{11l}^2 & \boldsymbol{\Sigma}_{12l} \\ \boldsymbol{\Sigma}_{21l} & \boldsymbol{\Sigma}_{22l} \end{pmatrix},$$

respectively.

The `DPcdensity` function fits a marginalized version of the model where the random probability measure G is integrated out. Full inference on the conditional density at covariate level z is obtained using the ϵ–DP approximation proposed by Muliere and Tardella (1998), with $\epsilon = 0.01$.

Hierarchical mixture of Dirichlet process mixture of normals—The `HDPMdensity` function considers the hierarchical mixture of DPM of normal models for density estimation presented in Müller *et al.* (2004). Let y_{ij} be the q-dimensional vector of responses for the jth observation, $j = 1, \ldots, n_i$, for the ith group, $i = 1, \ldots, I$. The model assumes that

$$\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in_i}|F_i \overset{iid}{\sim} F_i,$$

where F_i is assumed to arise as a mixture model $F_i = \epsilon H_0 + (1 - \epsilon)H_i$ of one common distribution H_0 and a distribution H_i that is specific or idiosyncratic to the ith group. The random probability measures H_i, $i = 0,1, \ldots, I$ in turn are given a DPM of normal prior,

$$H_i(\boldsymbol{y}) = \int N_q(\boldsymbol{y}|\mu, \boldsymbol{\Sigma}) \, dG_i(\mu),$$

with

$$G_i|\alpha_i, \mu_b, \boldsymbol{\Sigma}_b \tilde{\ } DP\left(\alpha N_q(\mu_b, \boldsymbol{\Sigma}_b)\right).$$

The model specification is completed by assuming the following hyper-priors,

$$\boldsymbol{\Sigma}|\nu, \boldsymbol{T} \tilde{\ } IW_q(\nu, \boldsymbol{T}),$$

$$\alpha_i|a_{0i}, b_{0i} \tilde{\ } \Gamma(a_{0i}, b_{0i}),$$

$$\mu_b|\boldsymbol{m}_0, \boldsymbol{S}_0 \tilde{\ } N_q(\boldsymbol{m}_0, \boldsymbol{S}_0),$$

$$\boldsymbol{\Sigma}_b|\nu_b, \boldsymbol{T}_b \tilde{\ } IW_q(\nu_b, \boldsymbol{T}_b),$$

and

$$\epsilon|\pi_0, \pi_1, a_\epsilon, b_\epsilon \tilde{\ } \pi_0 \delta_0 + \pi_1 \delta_1 + (1 - \pi_0 - \pi_1)\beta(a_\epsilon, b_\epsilon),$$

where δ_c represents the Dirac measure at c, and $\beta(a, b)$ represents the beta distribution with parameters a and b.

The `HDPMcdensity` function considers the extension of the previously described approach to the inclusion of continuos predictors z. This functions fits the HDPM model to the complete data $d_i = (y_i, z_i)'$, $i = 1, \ldots, n$, and reports the induced conditional distributions.

3.10. Generalized additive models

The `PSgam` function fits a generalized additive model (see, e.g. Hastie and Tibshirani, 1990) using Penalized splines (see e.g., Eilers and Marx, 1996; Lang and Brezger, 2004). The linear predictors η_i, $i = 1, \ldots, n$, are modeled in an additive way. Let x_i be a p-dimensional design vector and z_i be a q-dimensional vector of continuous predictors. Then, the model is given by

$$\eta_i = \boldsymbol{x}_i' \beta + \sum_{j=1}^{q} f_j\left(z_{ij}\right)$$

where the effect f_j of the a covariate z_j is approximated by a polynomial spline with equally spaced knots, written in terms of a linear combination of B-spline basis functions. Specifically,

the function f_j is aproximated by a spline of degree l with r equally spaced knots within the domain of z_j,

$$f_j\left(z_j\right) = \sum_{m=1}^{l+r} b_{jm} B_{jm}^l\left(z_j\right),$$

where $B_{jm}^l\left(\cdot\right)$ are B-spline basis function of degree l, and b_{jm} represents the associated B-spline coefficients. For the parametric component of the model, a normal prior distribution is assumed,

$$\beta N_p\left(\beta_0, S_{\beta_0}\right).$$

For the vector of basis coefficients $b_j = (b_{j1}, \ldots, b_{j(l+r)})^T$, independent Gaussian smoothness priors (Lang and Brezger, 2004) are assumed

$$p\left(b_j | \sigma_{bj}^2\right) \propto \exp\left(-\frac{1}{2\sigma_{bj}^2} b_j' K_j b_j\right).$$

The precision matrix acts as a penalty matrix to enforce smoothness and is defined through $K_j = D_j^T D_j$, where D_j is a first or second order difference matrix for adjacent B-spline coefficients. The variance (or inverse smoothing) parameter σ_{bj}^2 controls the amount of smoothness. Note that the log-penalty corresponds exactly to the penalty term introduced by Eilers and Marx (1996) in a frequentist penalized likelihood setting. For the variance parameters, we assume independent inverse gamma priors

$$\sigma_{bj}^{-2} | \tau_{b1}, \tau_{b2} \Gamma\left(\tau_{b1}/2, \tau_{b2}/2\right).$$

Finally, for the gamma and Gaussian models, an inverse gamma prior is assumed for the dispersion parameter σ^2,

$$\sigma^{-2} | \tau_1, \tau_2 \Gamma\left(\tau_1/2, \tau_2/2\right).$$

The computational implementation of the model is model-specific. For the Poisson, gamma, and binomial (logit) models, fixed and random effects are updated using MH steps with a IWLS normal proposal (see, West, 1985; Gamerman, 1997). For the probit-Bernoulli model, the latent variable representation of the binary responses is used, leading to conjugate normal updates.

3.11. Additional tools

Additional functions included in the package are DPelicit and PsBF. The DPelicit function implements methods for eliciting the DP prior using exact and approximated formulas for the mean and variance of the number of clusters given the total mass parameter and the number of subjects (see, e.g. Jara, García-Zattera, and Lesaffre, 2007). The PsBF function computes pseudo-Bayes factors for model comparison.

The practical implementation of models based on DP priors with a random precision parameter requires adopting values for the hyperparameters a_0 and b_0. The discrete nature of the DP realizations leads to their well-known clustering properties. The choice of a_0 and b_0 needs some

careful thoughts, as the parameter α directly controls the number of distinct components. Kottas, Müller, and Quintana (2005), referred to as the KMQ approach, and Jara *et al.* (2007), referred to as the JGL approach, proposed strategies for the specification of these hyperparameters.

The KMQ approach is based on approximations of the conditional mean and conditional variance of the number of clusters, given the precision parameter α (see e.g., Liu, 1996). Specifically, denoting by n the number of elements associated to the DP prior, and n^* the number of resulting clusters, their approach relies on

$$E\left(n^*|\alpha\right) = \sum_{i=1}^{n} \frac{\alpha}{\alpha+i-1} \approx \alpha \log\left(\frac{\alpha+n}{\alpha}\right), \quad (5)$$

$$Var\left(n^*|\alpha\right) = \sum_{i=1}^{n} \frac{\alpha\left(i-1\right)}{\left(\alpha+i-1\right)^2} \approx \alpha \left\{\log\left(\frac{\alpha+n}{\alpha}\right) - 1\right\}, \quad (6)$$

Using the fact that a priori $E\left(\alpha|a_0, b_0\right) = \frac{a_0}{b_0}$ and $Var\left(\alpha|a_0, b_0\right) = \frac{a_0}{b_0^2}$, the resulting expressions for the prior mean and variance of n^* are

$$E\left(n^*\right) \approx \frac{a_0}{b_0} \log\left(1 + \frac{n b_0}{a_0}\right), \quad (7)$$

and

$$Var\left(n^*\right) \approx \frac{a_0}{b_0} \log\left(1 + \frac{n b_0}{b_0}\right) - \frac{n b_0}{a_0} + \left\{\log\left(1 + \frac{n b_0}{a_0}\right) - \frac{n b_0}{a_0 + n b_0}\right\}^2 \frac{a_0}{b_0^2}. \quad (8)$$

On the other hand, the JGL approach is based on the exact value of conditional mean and conditional variance of the number of clusters given the precision parameter α. They noted that the approximations given by the expression (5) and expression (6) may be dangerous when α is considered a function of n. For instance, (5) gives 0 instead of 1 with $\alpha = \frac{1}{n}$. Better approximations may be obtained by noticing that

$$E\left(n^*|\alpha\right) = \sum_{i=1}^{n} \frac{\alpha}{\alpha+i-1} = \alpha \left\{\psi_0\left(\alpha+n\right) - \psi_0\left(\alpha\right)\right\},$$

and

$$Var\left(n^*|\alpha\right) = \sum_{i=1}^{n} \frac{\alpha\left(i-1\right)}{\left(\alpha+i-1\right)^2} = \alpha \left\{\psi_0\left(\alpha+n\right) - \psi_0\left(\alpha\right)\right\} + \alpha^2 \left\{\psi_1\left(\alpha+n\right) - \psi_1\left(\alpha\right),\right\}$$

where $\psi^0(\cdot)$ and $\psi i(\cdot)$ represents the digamma and trigamma function, respectively. Using these results, an approximation based on a first-order Taylor series expansion, and the fact that a priori $E\left(\alpha|a_0, b_0\right) = \frac{a_0}{b_0}$ and $Var\left(\alpha|a_0, b_0\right) = \frac{a_0}{b_0^2}$ we get

$$E\left(n^{*}\right) \approx \frac{a_0}{b_0}\left\{\psi_0\left(\frac{a_0+nb_0}{b_0}\right)-\psi_0\left(\frac{a_0}{b_0}\right)\right\}$$

and

$$Var\left(n^{*}\right) \approx \quad \frac{a_0}{b_0}\left\{\psi_0\left(\frac{a_0+nb_0}{b_0}\right)-\psi_0\left(\frac{a_0}{b_0}\right)\right\}+\frac{a_0^2}{b_0^2}\left\{\psi_1\left(\frac{a_0+nb_0}{b_0}\right)-\psi_1\left(\frac{a_0}{b_0}\right)\right\}+$$
$$\left\{\frac{a_0}{b_0}\left[\psi_1\left(\frac{a_0+nb_0}{b_0}\right)-\psi_1\left(\frac{a_0}{b_0}\right)\right]+\psi_0\left(\frac{a_0+nb_0}{b_0}\right)-\psi_0\left(\frac{a_0}{b_0}\right)\right\}^2\frac{a_0}{b_0^2}.$$

These expressions could be used in order to evaluate the robustness of the model to the specification of prior distribution for the precision parameter. The function DPelicit computes either the expected value and the standard deviation of the number of clusters, given the values of the parameters of the Gamma prior for the precision parameter, a_0 and b_0, or the value of the parameters a_0 and b_0 of the Gamma prior distribution for the precision parameter, a, given the prior judgement for the expected number and the standard deviation of the number of clusters. With this objective in mind, the Newton-Raphson algorithm and the forward-difference approximation to Jacobian are used.

4. Examples

In this section we consider the analyses of simulated and real-life data in order to illustrate the usage of **DPpackage**.

4.1. Bayesian density regression

We illustrate the **DPcdensity** and **LDDPdensity** functions by means of simulated data. We replicate the results reported by Dunson *et al.* (2007b), where a different approach is proposed. Following Dunson *et al.* (2007b), we simulate $n = 500$ observations from from a mixture of two normal linear regression models, with the mixture weights depending on the predictor, with different error variances and with a non-linear mean function for the second component,

$$y_i \mid x_i \overset{ind.}{\sim} \exp\{-2x_i\}\, N\left(y_i|x_i, 0.01\right)+\left(1-\exp\{-2x_i\}\right) N\left(y_i|x_i^4, 0.04\right), \quad i=1,\ldots,n,$$

where the predictor values x_i are simulated from a uniform distribution, $x_i \overset{iid}{\sim} U(0,1)$. The data was simulated using the following piece of code

```
################################################
# true conditional densities,
# mean function and
# simulation of the data.
################################################
dtrue <- function(grid,x)
{
exp(-2*x)*dnorm(grid,mean=x,sd=sqrt(0.01))+
(1-exp(-2 *x))*dnorm(grid,mean=x^4,sd=sqrt(0.04))
}
mtrue <- function(x)
```

```
{
exp(-2*x)*x+(1-exp(-2*x))*x^4
}
set.seed(0)
nrec <- 500
x <- runif(nrec)
y1 <- x + rnorm(nrec, 0, sqrt(0.01))
y2 <- x^4 + rnorm(nrec, 0, sqrt(0.04))
u <- runif(nrec)
prob <- exp(-2*x)
y <- ifelse(u<prob,y1,y2)
```

The extension of the DPM of normals approach of Müller *et al.* (1996) considered by the DPcdensity function, was fitted using the following hyper-parameters: $a_0 = 10$, $b_0 = 1$, $\nu_1 = \nu_2 = 4$, $m_2 = (\bar{y}, \bar{x})'$, $\tau_1 = 6.01$, $\tau_2 = 3.01$, and $S_2 = \Psi_2^{-1}$, where S is the sample covariance matrix for the response and predictor. A total number of 25,000 scans of the Markov chain cycle implemented in the DPcdensity function were completed. A burn-in period of 5,000 samples was considered and the chain was subsampled every 4 iterates to get a final sample size of 5,000. The following commands were used to fit the model, where the conditional density estimates were evaluated on a grid of 100 points on the range of the response,

```
################################################
# prior information
################################################
w <- cbind(y,x)
wbar <- apply(w,2,mean)
wcov <- var(w)
prior <- list(a0=10,
b0=1,
nu1=4,
nu2=4,
s2=0.5*wcov,
m2=wbar,
psiinv2=2*solve(wcov),
tau1=6.01,
tau2=3.01)
################################################
# mcmc specification
################################################
mcmc <- list(nburn=5000,
nsave=5000,
nskip=3,
ndisplay=1000)
################################################
# covariate values where the density
# and mean function is evaluated
################################################
xpred <- seq(0,1,0.02)
################################################
# fitting the model
################################################
```

```
fitWDDP <- DPcdensity(y=y,x=x,
xpred=xpred,
ngrid=100,
prior=prior,
mcmc=mcmc,
state=NULL,
status=TRUE)
```

Using the same MCMC specification, the LDDP model was also fitted to the data. The LDDPdensity function was used to fit a a mixture of B-splines models with

$x'\beta=\beta_0+\Sigma_{j=1}^{6}\psi_j(x)\beta_j$, where $\psi_k(x)$ corresponds to the kth B-spline basis function evaluated at x as implemented in the bs function of the **splines** R package. The LDDP model was fitted using Zellner's g-prior (Zellner, 1983), with $g = 10^3$. The following values for the hyper-parameters were considered: $a_0 = 10$, $b_0 = 1$, $m_0 = (X'X)^{-1}X'y$, $S_0 = g(X'X)^{-1}$, $\tau_1 = 6.01$, $\tau_{s1} = 6.01$, $\tau_{s2} = 2.01$, $\nu = 9$, and $\Psi^{-1} = S_0$. The following piece of code was used to fit the model:

```
################################################
# prior information
################################################
library(splines)
W <- cbind(rep(1,nrec),bs(x,df=6))
S0 <- 1000*solve(t(W)%*%W)
m0 <- solve(t(W)%*%W)%*%t(W)%*%y
prior<-list(a0=10,
b0=1,
m0=m0,
S0=S0,
tau1=6.01,
taus1=6.01,
taus2=2.01,
nu=9,
psiinv=solve(S0))
################################################
# covariate values where the density
# and mean function is evaluated
################################################
xpred <- seq(0,1,0.02)
Wpred <- cbind(rep(1,length(xpred)),bs(xpred,df=6))
################################################
# fitting the model
################################################
fitLDDP <- LDDPdensity(formula=y~W-1,zpred=Wpred,
ngrid=100,
prior=prior,
mcmc=mcmc,
state=NULL,
status=TRUE)
```

Figures 1 and 2 show the true density, the estimated density and point-wise 95% HPD intervals for a range of values of the predictor for the WDDP and LDDP model, respectively. The

estimates correspond approximately to the true densities in each case. The figures also display the plot of the data along with the estimated mean function, which is very close to the true one under both models.

In both functions, the posterior mean estimates and the limits of point-wise 95% HPD intervals for the conditional density for each value of the predictors are stored in the model objects densp.m, and densp.l and densp.h, respectively. The following piece of code illustrates how these objects can be used in order to get the posterior estimates for $x = 0.1$ in the LDDP model. This code was used to draw the plots displayed in Figures 1 and 2.

```
par(cex=1.5,mar=c(4.1, 4.1, 1, 1))
plot(fitLDDP$grid,fitLDDP$densp.h[6,],lwd=3,type="l",lty=2,
main="",xlab="y",ylab="f(y|x)",ylim=c(0,4))
lines(fitLDDP$grid,fitLDDP$densp.l[6,],lwd=3,type="l",lty=2)
lines(fitLDDP$grid,fitLDDP$densp.m[6,],lwd=3,type="l",lty=1)
lines(fitLDDP$grid,dtrue(fitLDDP$grid,xpred[6]),lwd=3,
type="l",lty=1,col="red")
```

Finally, both functions return the posterior mean estimates and the limits of point-wise 95% HPD intervals for the mean function in the model objects meanfp.m, and meanfp.l and meanfp.h, respectively. The following pice of code was used to obtain the estimated mean function under the LDDP model along with the true function.

```
par(cex=1.5,mar=c(4.1, 4.1, 1, 1))
plot(x,y,xlab="x",ylab="y",main="")
lines(xpred,fitLDDP$meanfp.m,type="l",lwd=3,lty=1)
lines(xpred,fitLDDP$meanfp.l,type="l",lwd=3,lty=2)
lines(xpred,fitLDDP$meanfp.h,type="l",lwd=3,lty=2)
lines(xpred,mtrue(xpred),col="red",lwd=3)
```

4.2. Dependent random effects distributions

We consider data from the Chilean system for educational quality measurement (Sistema de Medi-ción de la Calidad de la Educación, SIMCE). The Chilean education system is subject to several performance evaluations regularly at the school, teacher and student level. In the last case, SIMCE has developed mandatory census-type tests to regularly assess the educational progress at three stages: 4th and 8th grades in primary school (9 and 13 years old children, respectively), and 2nd grade in secondary school (16 years old children). The SIMCE instruments are designed to assess the achievement of fundamental goals and minimal contents of the curricular frame in different areas of knowledge, currently Spanish, mathematics and science. Here we focus on data from the math test applied in 2004 to 8 grader examinees in primary school. The test consists of 45 multiple choice items questions with 4 alternatives. The response $y_{ij} \in \{0,1\}$ is a binary variable indicating whether the individual i answers item j correctly.

The main purpose of collecting these data is to monitor standards and progress of educational systems, focusing on characterizing the population (and its evolution) rather than individual examinees. It is of particular interest to understand the way in which some factors at individual and/or school level could explain systematic differences in the performance of students in order to establish policies to improve the education system. For instance, a significant characteristic of the Chilean elementary and secondary education system is a variety of different school types.

These are grouped as Public I, financed by the state and administered by county governments; Public II, financed by the state and administered by county corporations; Private I, financed by the state and administered by the private sector; Private II, fee-paying schools that operate solely on payments from parents and administered by the private sector.

In order to evaluate the effect of the type of school and gender on the student performance we consider the LDDP mixture of normals prior for the ablities in a Rasch model as in Fariña *et al.* (2009). For illustration purposes, we consider a subset of 500 children. We refer to Fariña *et al.* (2009) for a full analysis of the complete data. The model is given by

$$y_{ij}|\pi_{ij} \overset{ind}{\sim} \text{Bernoulli}\left(\pi_{ij}\right),$$

$$\text{logit}\left(\pi_{ij}\right) = \theta_i - \beta_j,$$

$$\theta_i|G \overset{ind.}{\sim} \int N\left(\theta_i|x_i'\beta, \sigma^2\right) dG\left(\beta, \sigma^2\right).$$

Here, x_i includes an intercept term, three dummy variables for the type of school and the gender indicator. The LDDP Rasch model was fitted using the `LDDPrasch` function and assuming $\beta \sim N_{44}(0, 10^3 I44)$, $a = 1$, $\mu_0 = 0_5$, $S_0 = 100I_5$, $\tau_1 = 6.01$, $\tau_{s1} = 6.01$, $\tau_{s2} = 2.01$, $\nu = 8$, $\Psi = I_5$. A single Markov chain cycle of length 25, 000 was completed. The full chain was sub-sampled every 4 steps after a burn in period of 5,000 samples, to give a reduced chain of length 5,000. For each gender and type of school the density of the abilities distribution was evaluated on a grid of 100 equally spaced points in the range $(-3, 8)$. The following commands were used to fit the model,

```
################################################
# prediction's design matrix.
# columns: - intercept term.
# - 3 dummies for type of school.
# - gender indicator (1 = girl).
################################################
zpred <- matrix(c(1,0,0,0,0,
1,1,0,0,0,
1,0,1,0,0,
1,0,0,1,0,
1,0,0,0,1,
1,1,0,0,1,
1,0,1,0,1,
1,0,0,1,1),
nrow=8,ncol=5,byrow=T)
################################################
# prior information
################################################
prior <- list(alpha=1,
beta0=rep(0,44),
```

```
Sbeta0=diag(1000,44),
mu0=rep(0,5),
S0=diag(100,5),
tau1=6.01,
taus1=6.01,
taus2=2.01,
nu=8,
psiinv=diag(1,5))
################################################
# mcmc
################################################
mcmc <- list(nburn=5000,
nskip=3,
ndisplay=1000,
nsave=5000)
################################################
# fitting the model
################################################
fitLDDP <- LDDPrasch(formula=y ~ types+gender,
prior=prior,
mcmc=mcmc,
state=NULL,
status=TRUE,
zpred=zpred,
grid=seq(-3,8,len=100),
compute.band=TRUE)
```

Different shapes in the resulting posterior densities were observed. Figure 3 displays the posterior mean and point wise 95% HPD interval for the random effects distribution for different combinations of the predictors. The density estimates show a clear departure from the commonly assumed normality of the random effects distributions. We found no important differences in the behavior of boys and girls. Children in Public I and II schools showed a similar skewed to the right random effects distribution. The estimated abilities distributions for children in private schools were shifted to the right in comparison with the distribution observed for children from public schools. This shift was more pronounced for children in fee-paying schools that operate solely on payments from parents and administered by the private sector (Private II) than those from schools financed by the state and administered by the private sector (Private I). A bimodal random effects distribution was observed in the abilities distributions from private schools.

4.3. Proportional hazards regression with nonparametric frailties

Consider right censored survival data where failure times are repeatedly observed within a group or subject. Let $i = 1,\ldots, n$ denote the strata over which repeated times-to-event are recorded, and $j = 1,\ldots, n_i$ denote the repeated observations within stratum i. The data are denoted $\{(w_{ij}, t_{ij}, \delta_{ij}) : i = 1,\ldots, n; j = 1,\ldots, n_i\}$ where t_{ij} is the recorded event time, $\delta_i = 1$ if t_{ij} is an observed failure time and $\delta_{ij} = 0$ if the failure time is right censored at t_{ij}, and w_{ij} is a p-dimensional vector of covariates.

Functions fitting generalized linear mixed models (PTglmm, DPglmm, and DPMglmm) can be used to fit the Cox proportional hazards model (Cox, 1972) with nonparametric, multivariate frailties. Briefly, the baseline hazard function λ_0 (t) corresponds to an individual with covariates $w = 0$ and survival time T_0. Given that the baseline individual has made it up to t,

$T_0 \quad t$, the baseline hazard is how the probability of expiring in the next instant is changing. In terms of the baseline survival function $S_0(t) = P(T_0 > t)$ and density $f_0(t)$, this is given by

$$\lambda_0(t) = \lim_{\epsilon \to 0^+} \frac{P(t \le T_0 < t + \epsilon | T_0 \ge t)}{\epsilon} = \frac{f_0(t)}{S_0(t)}.$$

The conditionally proportional hazards assumption stipulates that

$$\lambda\left(t_{ij} | \boldsymbol{z}_{ij}\right) = \lambda_0(t) \, \exp\left(\boldsymbol{w}_{ij}' \gamma + \theta_i\right),$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$ are random effects, termed *frailties* in the survival literature. Often the frailties θ_i, or exponentiated frailties e^{θ_i}, are assumed to arise *iid* from some parametric distribution such as $N(0, \sigma^2)$, gamma, positive stable, etc. We consider a nonparametric MPT prior on the frailties below.

The specification is conditional because proportionality only holds for survival times within a given strata i, not across strata unless the distribution of θ_i is positive stable (see, e.g. Qiou, Ravishanker, and Dey, 1999). Precisely, for individuals j_1 and j_2 within strata i,

$$\frac{\lambda\left(t_{ij_1} | \boldsymbol{w}_{ij_1}\right)}{\lambda\left(t_{ij_2} | \boldsymbol{w}_{ij_2}\right)} = \exp\left\{\left(\boldsymbol{w}_{ij_1} - \boldsymbol{w}_{ij_2}\right)' \gamma\right\}.$$

Often the baseline hazard is assumed to be piecewise constant on a partition of R^+ comprised of K intervals, yielding the piecewise exponential model. References are too numerous to list; but see Walker and Mallick (1997), Aslanidou, Dey, and Sinha (1998), and Qiou *et al.* (1999). Assume

$$\lambda_0(t) = \sum_{k=1}^{K} \lambda_k I\{a_{k-1} < t \le a_k\},$$

where $a_0 = 0$ and $a_K = \infty$, although in practice $a_K = \max\{t_{ij}\}$ is sufficient. The prior hazard is specified by cutpoints $\{a_k\}_{k=0}^{K}$ and hazard values $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)'$. If the prior on $\boldsymbol{\lambda}$ is taken to be independent gamma distributions, the model can approximate the gamma process on a fine mesh (Kalbfleisch, 1978). Regardless, the resulting model implies a Poisson likelihood for "data" y_{ijk} taking values $y_{ijk} = 0$ when $t_{ij} \notin (a_{k-1}, a_k]$ or $\delta_{ij} = 0$, and $y_{ijk} = 1$ when $t_{ij} \in (a_{k-1}, a_k]$ and $\delta_{ij} = 1$, for $k = 1, \ldots, K(t_{ij})$, where $K(t) = \max\{k : a_k \quad t\}$. The likelihood for $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma})$ is

$$
\begin{aligned}
\mathcal{L}(\beta, \lambda, \gamma) &= \prod_{i=1}^{n} \prod_{j=1}^{n_i} \left[\prod_{k=1}^{K(t_{ij})} e^{-\exp\left\{\log(\lambda_k) + \boldsymbol{w}_{ij}'\beta + \gamma_i\right\} \Delta_{ijk}} \right] \left[e^{-\log\left\{\lambda_{K(t_{ij})}\right\} + \boldsymbol{w}_{ij}'\beta + \gamma_i} \right]^{\delta_{ij}}, \\
&\propto \prod_{i=1}^{n} \prod_{j=1}^{n_i} \prod_{k=1}^{K(t_{ij})} p\left(y_{ijk} | \mu_{ijk}\right),
\end{aligned}
$$

where $p(y|\mu)$ is the probability mass function for a Poisson(μ) random variable, $\mu_{ijk} = \exp\left\{\log(\lambda_k) + \boldsymbol{w}_{ij}'\beta + \gamma_i\right\} \Delta_{ijk}$, and $\Delta_{ijk} = \min\{a_k, t_{ij}\} - a_{k-1}$. Thus, the Cox model assuming

a piecewise constant baseline hazard can be fitted in any software allowing for Poisson regression. Note that if covariates are time dependent as well, and change only at values included in $\{a_k\}_{k=0}^{K}$, the likelihood is trivially extended to include w_{ijk} above for $k = 1,\ldots, K$ (t_{ij}) rather than $\mathbf{w}_{ij}$.

We consider data on $n = 38$ kidney patients discussed by McGilchrist and Aisbett (1991). Each of the patients provides $n_i = 2$ infection times, some of which are right censored. McGilchrist and Aisbett (1991) found that only gender was significant, and so we follow Aslanidou *et al.* (1998), Walker and Mallick (1997), Qiou *et al.* (1999), and Hemming and Shaw (2005) in considering only this covariate in what follows. We fitted the semiparametric proportional hazards regression model using a nonparametric prior for the frailties distribution. The following commands were used to prepare the data to fit the model. The original dataset, $d[i, j]$, is a 38 by 6 matrix, which for each row (from left to right) contains the subject indicator, t_{i1}, δ_{i1}, t_{i2}, δ_{i2}, and the gender indicator. Ten intervals were considered with cutpoints $\{a_1,\ldots, a_{10}\}$ taken from the empirical distribution of the data.

```
#################################################
# function to make a row with '1' at ind
#################################################
onv <- function(ind,len)
{
onv <- rep(0,len)
onv[ind] <- 1
return(onv)
}
#################################################
# Create data to fit Cox model using
# Poisson likelihood for piecewise
# exponential model.
#################################################
newdat <- matrix(1:(38*2*2),nrow=38*2,ncol=2)
tt <- rep(0,38*2)
delta <- tt
for(i in 1:38)
{
newdat[i*2-1,1] <- d[i,1]
newdat[i*2-1,2] <- d[i,6]
newdat[i*2 ,1] <- d[i,1]
newdat[i*2 ,2] <- d[i,6]
tt[i*2-1] <- d[i,2]
delta[i*2-1] <- d[i,3]
tt[i*2] <- d[i,4]
delta[i*2] <- d[i,5]
}
y <- NULL
mat <- NULL
tot <- 0
p <- ncol(newdat)
off <- NULL
n <- length(tt)
intervals <- 10
```

```
cutpoint <- quantile(tt,(1:intervals)/intervals,names=FALSE)
for(i in 1:n)
{
tot <- tot+1
mat <- matrix(append(mat,c(newdat[i,1:p],onv(1,intervals))),
c(p+intervals,tot))
off <- append(off,min(cutpoint[1],tt[i]))
if(tt[i]<=cutpoint[1] && delta[i]==1)
{
y <- append(y,1)
}
else
{
y <- append(y,0)
}
for(j in 1:(intervals-1))
{
if(tt[i]>cutpoint[j])
{
off <- append(off,min(cutpoint[j+1],
tt[i])-cutpoint[j])
tot <- tot+1
mat <- matrix(append(mat,c(newdat[i,1:p],
onv(j+1,intervals))),
c(p+intervals,tot))
if(tt[i] <= cutpoint[j+1] && delta[i]==1)
{
y <- append(y,1)
}
else
{
y <- append(y,0)
}
}
}
}
mat <- t(mat)
id <- mat[,1]
gender <- mat[,2]
loghazard <- mat[,3:12]
```

We performed the analysis using the PTglmm function to the responses

$$
\boldsymbol{y}_i = \left(y_{i11}, \ldots, y_{i1K(t_{i1})}, \ldots, y_{i21}, \ldots, y_{i2K(t_{i2})} \right),
$$

and where $\boldsymbol{x}_{ij}$ is a 11-dimensional design vector containing the gender indicator and the indicator for the interval associated to the corresponding response. Finally, we set $\boldsymbol{\beta} = (\boldsymbol{\gamma}', \lambda')'$, and assume

$$
\beta | \beta_0, \boldsymbol{S}_{\beta_0} \sim \mathcal{N}_p \left(\beta_0, \boldsymbol{S}_{\beta_0} \right).
$$

$$\theta_1, \ldots, \theta_n | G \overset{iid}{\sim} G,$$

and

$$GPT^M \left(\Pi^{\sigma^2}, \mathscr{A}^\alpha \right).$$

We consider a $M = 5$ finite PT prior which was centered around a $N(0, \sigma^2)$ distribution and constrained to have median-0 (frstlprob=TRUE in the prior object below). The values for the hyper-parameters β_0 and S_{β_0} were obtained from a penalized quasi-likelihood (PQL) fit using the glmmPQL function available from the MASS pakage (Venables and Ripley, 2002). The matrix S_{β_0} was inflated by a factor of 100. The remaining hyper-parameters were $a_0 = b_0 = 1$, $\nu_0 = 3$, and $T = I_1$. Starting values for the model parameters were obtained from the PQL fit. A single Markov chain cycle of length 25,000 was completed. The full chain was sub-sampled every 4 steps after a burn in period of 5,000 samples, to give a reduced chain of length 5,000. The code for fitting the model using PTglmm was

```
###############################################
# PQL estimation
###############################################
library(MASS)
fit0 <- glmmPQL(fixed=y~gender+loghazard-1+
offset(log(off)),
random=~1|id,family=poisson(log))
###############################################
# prior
###############################################
beta0 <- fit0$coefficients$fixed
Sbeta0 <- vcov(fit0)
prior <- list(M=5,
a0=1,
b0=1,
nu0=3,
tinv=diag(1,1),
mu=rep(0,1),
beta0=beta0,
Sbeta0=Sbeta0,
frstlprob=TRUE)
###############################################
# starting values from PQL estimation
###############################################
beta <- fit0$coefficients$fixed
b <- as.vector(fit0$coefficients$random$id)
mu <- rep(0,1)
sigma <- getVarCov(fit0)[1,1]
state <- list(alpha=1,
beta=beta,
b=b,
mu=mu,
```

```
sigma=sigma)
###############################################
# mcmc
###############################################
mcmc <- list(nburn=5000,
nsave=5000,
nskip=19,
ndisplay=1000,
tune3=1.5)
###############################################
# fitting the model
###############################################
fitPT <- PTglmm(fixed=y~gender+loghazard,
offset=log(off),
random=~1|id,
family=poisson(log),
prior=prior,
mcmc=mcmc,
state=state,
status=FALSE)
###############################################
# posterior inferences
###############################################
summary(fitPT)
###############################################
# frailties density estimate
###############################################
predPT <- PTrandom(fitPT,predictive=TRUE,
gridl=c(-2.3,2.3))
plot(predPT)
```

The abridged output is given below. The output lists the estimated effect for gender $\hat{\beta}_1 = -1.13$ followed by $K = 10$ estimated log-hazard values. Notice that the intercept term in the posterior information for the "fixed" effects (regression coefficients in the output), corresponds to the mean of the frailties distribution G. The posterior median estimate of the centering variance was $\hat{\sigma}^2 = 0.35$ and close to the posterior median of the frailties variance (0.33). Further, the posterior median (95% credible interval) for α was 0.75 (0.04; 3.77). The trace plots of the parameters (not shown) indicate a good mixing of the chain. The acceptance rates for the MH steps associated to the regression coefficients, frailties, centering variance and precision parameter was 36, 61, 43 and 0.46%, respectively. Notice that the 0 values for the acceptance rates in the output corresponds to the centering mean, which is sampled, and the decomposition of the centering covariance matrix. The latter is only sampled for dimensions greater than or equal to 2.

Walker and Mallick (1997) analyzed these data with piecewise exponential model and frailties following a Polya tree with fixed centering variance, $PT^8(\Pi^{100}, A^{0.1})$ and find $\hat{\beta}_1 = -1.0$. McGilchrist and Aisbett (1991) obtain $\hat{\beta}_1 = -1.8$, but with other nonsignificant covariates included. Aslanidou *et al*. (1998) also reportes $\hat{\beta} = -1.0$. Hemming and Shaw (2005) obtain $\hat{\beta} = -1.7$ and Qiou *et al*. (1999) obtain $\hat{\beta} = -1.1$ under the positive stable and $\hat{\beta} = -1.6$ under gamma frailties, respectively. The the deviance information criterion (DIC), as presented by Spiegelhalter, Best, Carlin, and Van der Linde (2002), was 398 for either PT or normal model

(not shown), so the normal model does about the same from a predictive standpoint based on the DIC.

```
Bayesian semiparametric generalized linear mixed effect model
Call:
PTglmm.default(fixed = y ~ gender + loghazard, random = ~1 |
id, family = poisson(log), offset = log(off), prior = prior,
mcmc = mcmc, state = state, status = FALSE)
Posterior Predictive Distributions (log):
Min. 1st Qu. Median Mean 3rd Qu. Max.
-5.99200 -0.22250 -0.10970 -0.48500 -0.05714 -0.01381
Model's performance:
Dbar Dhat pD DIC LPML
379.21 360.63 18.58 397.79 -200.29
Regression coefficients:
Mean Median Std. Dev. Naive Std.Error 95%CI-Low 95%CI-Upp
(Intercept) -0.0004443 0.0015210 0.0960076 0.0013578 -0.2066125 0.2021371
gender -1.1321281 -1.1296717 0.3219508 0.0045531 -1.7762785 -0.5117994
loghazard1 -4.2608268 -4.2375512 0.4412274 0.0062399 -5.1598904 -3.4611046
loghazard2 -3.7898628 -3.7638395 0.5018976 0.0070979 -4.8383288 -2.8794989
loghazard3 -3.9792281 -3.9691425 0.4556631 0.0064440 -4.9028932 -3.1213276
loghazard4 -3.0627136 -3.0526713 0.4526581 0.0064016 -4.0124879 -2.2353213
loghazard5 -3.2581084 -3.2477986 0.4219626 0.0059675 -4.1039312 -2.4603991
loghazard6 -3.9951390 -3.9805448 0.4544001 0.0064262 -4.9103962 -3.1403702
loghazard7 -4.9343777 -4.9183270 0.5365962 0.0075886 -6.0496817 -3.9150135
loghazard8 -3.6883152 -3.6845014 0.4479935 0.0063356 -4.5692123 -2.8232222
loghazard9 -3.6723423 -3.6673231 0.4810002 0.0068024 -4.6112294 -2.7315973
loghazard10 -4.1246955 -4.1272752 0.4966618 0.0070239 -5.0749243 -3.1886274
Baseline distribution:
Mean Median Std. Dev. Naive Std.Error 95%CI-Low 95%CI-Upp
mu-(Intercept) 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
sigma-(Intercept) 0.430385 0.354618 0.294752 0.004168 0.119319 1.212674
Precision parameter:
Mean Median Std. Dev. Naive Std.Error 95%CI-Low 95%CI-Upp
alpha 1.05875 0.75117 1.02204 0.01445 0.04448 3.76967
Random effects variance:
Mean Median Std. Dev. Naive Std.Error 95%CI-Low 95%CI-Upp
R.E.Cov-(Intercept) 0.378637 0.331281 0.222121 0.003141 0.096121 0.948495
Acceptance Rate for Metropolis Steps = 0.3570935 0.6072718 0 0.428972 0.463486
0
Number of Observations: 413
Number of Groups: 38
```

Figure 4 shows the estimated frailty distribution from these data along with the posterior mean of the frailty term for each patient. The distribution is remarkably Gaussian-shaped, in contrast to the analysis presented in Walker and Mallick (1997), which showed two well defined density modes corresponding to men and women. We were unable to duplicate this result across several sets of hyper-prior values, including the consideration of PT^8 (Π^{100}, $A^{0.1}$). In retrospect, this is not surprising. Two well separated modes would typically indicate an omitted covariate, yet gender was included as a risk factor in the model.

Finally, Figure 5 show the posterior median and 95% credible interval for survival curves for males and females, taking the individual-level heterogeneity modeled through the frailty distribution into account.

5. Concluding remarks

Because the main obstacle for the practical use of BSP and BNP methods has been the lack of estimation tools, we presented an R package for fitting some frequently used models. Until the release of **DPpackage**, the two options for researchers who wished to fit a BSP or BNP model were to write their own code or to rely heavily on particular parametric approximations to some specific processes using the BUGS code given in Peter Congdon's books (see e.g., Congdon, 2001). **DPpackage** is geared primarily towards users who are not willing to bear the costs associated with both of these options.

Chambers (2000) conceptualized statistical software as a set of tools to organize, analyze and visualize data. Data organization and visualization of results is based on R capabilities. Chambers (2000) also proposed requirements and guidelines for developing and assessing statistical softwares. These requirements may be discussed with respect to **DPpackage**:

1. `Easy specification of simple tasks`: The documentation contains examples, and similar problems can be analyzed by moderate modifications of the model description files. The examples have been chosen so that they demonstrate the functionality of **DPpackage** with well-known data sets.

2. `Gradual refinement of the tasks`: The user can enhance a nonparametric model by adding covariates, and by fixing part of the baseline distributions and the precision parameters.

3. `Arbitrarily extensive programming`: **DPpackage** has a programming environment for implementing sophisticated proposal distributions, if the default proposals are not sufficient.

4. `Implementing high-quality computations`: Also, because the source code in a compiled language is available, new procedures can be added and the old ones modified to improve performance and flexibility.

5. `Embedding the results of items 2-4 as new simple tools`: **DPpackage** has the capability of continuing a Markov chain from the last value of the parameters of a previous analysis. As the MCMC samples are saved in matrix objects, both parts of the Markov chain can be easily merged.

Many improvements to the current status of the package can be made. For example, all **DPpackage** modeling functions compute CPOs for model comparison. However, only some of them compute the effective number of parameters pD and DIC, as presented by Spiegelhalter *et al.* (2002). These and other model comparison criterion will be included for all functions in future versions of **DPpackage**.

The implementation of more models, the development of general-purpose sampling functions, realtime visualization of simulation progress, and the ability to handle large dataset problems, through the use of sparse matrix techniques (George and Liu, 1981), are the topic of further improvements.

Acknowledgments

References

Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association. 1993; 88:669–679.

Antoniak CE. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. The Annals of Statistics. 1974; 2:1152–1174.

Aslanidou H, Dey DK, Sinha D. Bayesian analysis of multivariate survival data using Monte Carlo methods. Canadian Journal of Statistics. 1998; 26:33–48.

Besag J, Green P, Higdon D, Mengersen K. Bayesian computation and stochastic systems (with Discussion). Statistical Science. 1995; 10:3–66.

Branscum A, Hanson T. Bayesian nonparametric meta-analysis using Polya tree mixture models. Biometrics. 2008; 64:825–833. [PubMed: 18162115]

Bush CA, MacEachern SN. A semiparametric Bayesian model for randomised block designs. Biometrika. 1996; 83:275–285.

Carlin, BP.; Louis, TA. Bayesian methods for data analysis. 3rd Ed.. Chapman and Hall/CRC; New York, USA: 2008.

Chambers JM. Users, programmers, and statistical software. Journal of Computational and Graphical Statistics. 2000; 9(3):402–422.

Chen MH, Shao QM. Monte Carlo estimation of Bayesian credible and HPD intervals. Journal of Computational and Graphical Statistics. 1999; 8:69–92.

Christensen R, Hanson T, Jara A. Parametric nonparametric statistics: An introduction to mixtures of finite Polya trees. The American Statistician. 2008; 62:296–306.

Congdon, P. Bayesian statistical modelling. John Wiley and Sons; New York, USA: 2001.

Cox DR. Regression models and life-tables (with Discussion). Journal of the Royal Statistical Society, Series B. 1972; 34:187–220.

De Boeck, P.; Wilson, M. Explanatory item response models. A generalized linear and nonlinear approach. Springer; New York, USA: 2004.

De Iorio M, Johnson WO, Müller P, Rosner GL. Bayesian nonparametric non-proportional hazards survival modelling. Biometrics. 2009; 65:762–771. [PubMed: 19210742]

De Iorio M, Müller P, Rosner GL, MacEachern SN. An ANOVA model for dependent random measures. Journal of the American Statistical Association. 2004; 99:205–215.

Dey, D.; Müller, P.; Sinha, D. Practical nonparametric and semiparametric Bayesian statistics. Springer; New York, USA: 1998.

Doss H. Bayesian nonparametric estimation for incomplete data via successive substitution sampling. The Annals of Statistics. 1994; 22:1763–1786.

Duan JA, Guindani M, Gelfand AE. Generalized spatial Dirichlet process models. Biometrika. 2007; 94:809–825.

Dunson, D.; Yang, M.; Baird, D. Technical report. Department of Statistical Science, Duke University; 2007a. Semiparametric Bayes hierarchical models with mean and variance constraints.

Dunson DB, Park JH. Kernel stick-breaking processes. Biometrika. 2008; 95:307–323. [PubMed: 18800173]

Dunson DB, Pillai N, Park JH. Bayesian density regression. Journal of the Royal Statistical Society, Series B. 2007b; 69:163–183.

Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. Statistical Science. 1996; 11(2): 89–121.

Escobar MD. Estimating normal means with a Dirichlet process prior. Journal of the American Statistical Association. 1994; 89:268–277.

Escobar MD, West M. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association. 1995; 90:577–588.

Fariña, P.; Quintana, FA.; San, Martín E.; Jara, A. Technical report. Department of Statistics, Pontificia Universidad Católica de Chile; 2009. A dependent semiparametric Rasch model for the analysis of Chilean educational data.

Ferguson TS. A Bayesian analysis of some nonparametric problems. Annals of Statistics. 1973; 1:209–230.

Ferguson TS. Prior distribution on the spaces of probability measures. Annals of Statistics. 1974; 2:615–629.

Gamerman D. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing. 1997; 7:57–68.

Gelfand AE, Kottas A. A computational approach for full nonparametric Bayesian inference under Dirichlet Process Mixture models. Journal of Computational and Graphical Statistics. 2002; 11:289–304.

George, A.; Liu, JW. Computer solution of large sparse positive definite systems. Prentice-Hall; New York, USA: 1981.

Ghosh, JK.; Ramamoorthi, RV. Bayesian nonparametrics. Springer; New York, USA: 2003.

Gilks WR, Thomas A, Spiegelhalter DJ. A language and program for complex Bayesian modelling. The Statistician. 1994; 43:169–178.

Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika. 1995; 82:711–732.

Griffin JE, Steel MFJ. Order-based dependent Dirichlet processes. Journal of the American Statistical Association. 2006; 101:179–194.

Hanson T. Inference for mixtures of finite Polya tree models. Journal of the American Statistical Association. 2006; 101:1548–1565.

Hanson, T.; Branscum, A.; Johnson, W. Bayesian nonparametric modeling and data analysis: an introduction. In: Dey, DK.; Rao, CR., editors. Bayesian Thinking: Modeling and Computation (Handbook of Statistics, volume 25). Elsevier; Amsterdam, The Netherlands: 2005. p. 245-278.

Hanson T, Johnson WO. Modeling regression error with a mixture of Polya trees. Journal of the American Statistical Association. 2002; 97:1020–1033.

Hanson T, Johnson WO. A Bayesian semiparametric AFT model for interval-censored data. Journal of Computational and Graphical Statistics. 2004; 13:341–361.

Hastie, T.; Tibshirani, R. Generalized additive models. Chapman and Hall; New York, USA: 1990.

Hemming K, Shaw JEH. A class of parametric dynamic survival models. Lifetime Data Analysis. 2005; 11:81–98. [PubMed: 15747591]

Ishwaran H, James LF. Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. Journal of Computational and Graphical Statistics. 2002; 11:508–532.

Jara A. Applied Bayesian non- and semi-parametric inference using DPpackage. Rnews. 2007; 7:17–26.

Jara A, García-Zattera MJ, Lesaffre E. A Dirichlet process mixture model for the analysis of correlated binary responses. Computational Statistics and Data Analysis. 2007; 51:5402–5415.

Jara A, Hanson T, Lesaffre E. Robustifying generalized linear mixed models using a new class of mixture of multivariate Polya trees. Journal of Computational and Graphical Statistics. 2009 To appear.

Kalbfleisch JD. Nonparametric Bayesian analysis of survival time data. Journal of the Royal Statistical Society, Series B. 1978; 40:214–221.

Kleinman KP, Ibrahim JG. A semi-parametric Bayesian approach to generalized linear mixed models. Statistics in Medicine. 1998a; 17:2579–2596. [PubMed: 9839349]

Kleinman KP, Ibrahim JG. A semiparametric Bayesian approach to the random effects model. Biometrics. 1998b; 54:921–938. [PubMed: 9750242]

Kottas A, Müller P, Quintana F. Nonparametric Bayesian modeling for multivariate ordinal data. Journal of Computational and Graphical Statistics. 2005; 14:610–625.

Kraemer, HC. Evaluating medical tests. Sage Publications; New York, USA: 1992.

Lang S, Brezger A. Bayesian P-splines. Journal of Computational and Graphical Statistics. 2004; 13:183–212.

Lavine M. Some aspects of Polya tree distributions for statistical modeling. The Annals of Statistics. 1992; 20:1222–1235.

Lavine M. More aspects of Polya tree distributions for statistical modeling. The Annals of Statistics. 1994; 22:1161–1176.

Li, Y.; Müller, P.; Lin, X. Technical report. Department of Biostatistics, The MD Anderson Cancer Center; 2007. Center-adjusted inference for a nonparametric Bayesian random effect distribution.

Liu JS. Nonparametric hierarchical Bayes via sequential imputations. The Annals of Statistics. 1996; 24:911–930.

Lo AY. On a class of Bayesian nonparametric estimates I: Density estimates. The Annals of Statistics. 1984; 12:351–357.

MacEachern, SN. Computational methods for mixture of Dirichlet process models. In: Dey, D.; Müller, P.; Sinha, D., editors. Practical Nonparametric and Semiparametric Bayesian Statistics. Springer; 1998. p. 1-22.

MacEachern, SN. ASA Proceedings of the Section on Bayesian Statistical Science. American Statistical Association; Alexandria, VA: 1999. Dependent nonparametric processes.

MacEachern, SN. Technical report. Department of Statistics, The Ohio State University; 2000. Dependent Dirichlet processes.

MacEachern SN, Müller P. Estimating mixture of Dirichlet Process models. Journal of Computational and Graphical Statistics. 1998; 7(7(2)):223–338.

Mauldin RD, Sudderth WD, Williams SC. Polya trees and random distributions. Annals of Statistics. 1992; 20:1203–1221.

McGilchrist CA, Aisbett CW. Regression with frailty in survival analysis. Biometrics. 1991; 47:461–466. [PubMed: 1912255]

Mukhopadhyay S, Gelfand AE. Dirichlet process mixed generalized linear models. Journal of the American Statistical Association. 1997; 92:633–647.

Muliere P, Tardella L. Approximating distributions of random functionals of Ferguson-Dirichlet priors. The Canadian Journal of Statistics. 1998; 26:283–297.

Müller P, Erkanli A, West M. Bayesian curve fitting using multivariate normal mixtures. Biometrika. 1996; 83:67–79.

Müller P, Quintana FA. Nonparametric Bayesian data analysis. Statistical Science. 2004; 19:95–110.

Müller P, Quintana FA, Rosner G. A method for combining inference across related nonpara-metric Bayesian models. Journal of the Royal Statistical Society, Series B. 2004; 66:735–749.

Müller P, Rosner GL. A Bayesian population model with hierarchical mixture priors applied to blood count data. Journal of the American Statistical Association. 1997; 92:1279–1292.

Neal R. Markov chain sampling methods for Dirichlet process mixture models. Journal of Computational and Graphical Statistics. 2000; 9:249–265.

Neal R. Slice sampling. The Annals of Statistics. 2003; 31:705–767.

Newton, MA. Technical report, N. 905. University of Wisconsin-Madison, Department of Statistics; 1994. Computing with priors that support identifiable semiparametric models.

Newton MA, Czado C, Chapell R. Bayesian inference for semiparametric binary regression. Journal of the American Statistical Association. 1996; 91:142–153.

Perron F, Mengersen K. Bayesian nonparametric modeling using mixtures of triangular distributions. Biometrics. 2001; 57:518–528. [PubMed: 11414578]

Petrone S. Bayesian density estimation using Bernstein polynomials. The Canadian Journal of Statistics. 1999a; 27:105–126.

Petrone S. Random Bernstein polynomials. Scandinavian Journal of Statistics. 1999b; 26:373–393.

Petrone S, Wasserman L. Consistency of Bernstein polynomial posterior. Journal of the Royal Statistical Society, Series B. 2002; 64:79–100.

Plummer M, Best N, Cowles K, Vines K. CODA: Output analysis and diagnostics for MCMC. 2006 R package version 0.10-7.

Qiou Z, Ravishanker N, Dey DK. Multivariate survival analysis with positive stable frailties. Biometrics. 1999; 55:81–88.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2009. ISBN 3-900051-07-0, URL http://www.R-project.org

Rasch, G. Probabilistic models for some intelligence and attainment tests. The Danish Institute for Educational Research (Expanded Edition, 1980, The University Chicago Press); Chicago, USA: 1960.

Rossi, P.; Allenby, G.; McCulloch, R. Bayesian statistics and marketing. John Wiley and Sons; New York, USA: 2005.

Rossi, P.; McCulloch, R. bayesm: Bayesian inference for marketing/micro-econometrics. 2008. R package version 2.2-2, URL http://faculty.chicagogsb.edu/peter.rossi/research/bsm.html

San Martín E, Jara A, Rolin JM, Mouchart M. On the analysis of Bayesian semiparametric IRT-type models. 2007 p. (Submitted).

Sethuraman J. A constructive definition of Dirichlet prior. Statistica Sinica. 1994; 2:639–650.

Smith BJ. BOA: An R package for MCMC output convergence assessment and posterior inference. Journal of Statistical Software. 2007; 21:1–37.

Spiegelhalter SD, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B. 2002; 64:583–639.

Strurtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. Journal of Statistical Software. 2005; 12:1–16.

Thomas A, O'Hara B, Ligges U, Sibylle S. Making BUGS open. Rnews. 2006; 6:12–17.

Tierney L. Markov chains for exploring posterior distributions. The Annals of Statistics. 1994; 22:1701–1762.

Venables, WN.; Ripley, BD. Modern applied statistics with S. fourth edition. Springer; New York: 2002. ISBN 0-387-95457-0, URL http://www.stats.ox.ac.uk/pub/MASS4

Walker SG, Damien P, Laud PW, Smith AFM. Bayesian nonparametric inference for random distributions and related functions (with discussion). Journal of the Royal Statistical Society, Series B. 1999; 61:485–527.

Walker SG, Mallick BK. Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. Journal of the Royal Statistical Society, Series B. 1997; 59:845–860.

West, M. Generalized linear models: outlier accomodation, scale parameter and prior distributions. In: Bernardo, JM.; DeGroot, MH.; Lindley, DV.; Smith, AFM., editors. Proceedings of the Second Valencia International Meeting. North Holland, Amsterdam: 1985.

Zellner A. Applications of Bayesian analysis in econometrics. The Statistician. 1983; 32:23–34.
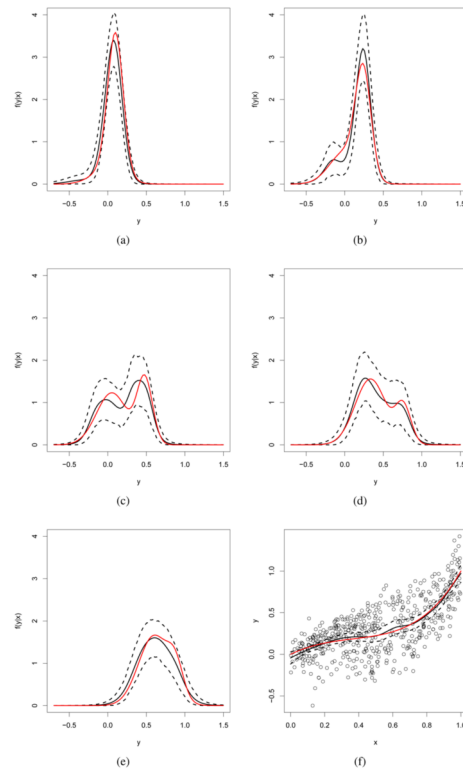
Figure 1.
Simulated data - WDDP model: True conditional densities of $y|x$ (in red), posterior mean estimates (black continuos line) and point-wise 95% HPD intervals (black dashed lines) for: (a) $x = 0.1$, (b) $x = 0.25$, (c) $x = 0.48$, (d) $x = 0.76$, and (e) x = 0.88. Panel (f) shows the data, along with the true and estimated mean regression curves.
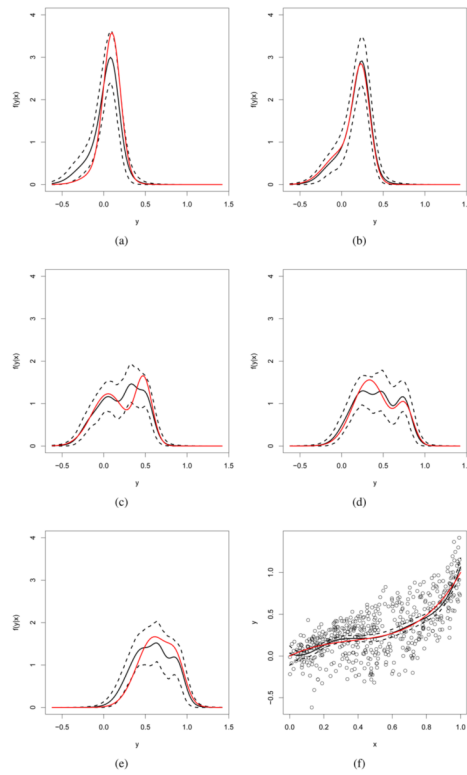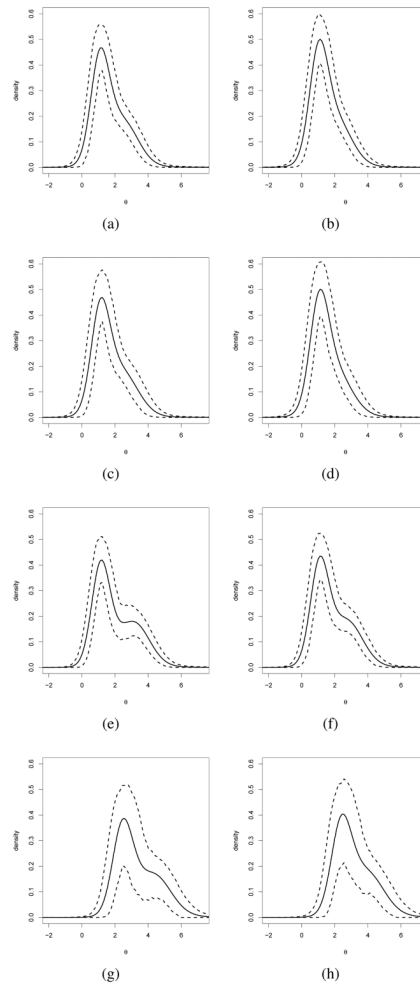
Figure 2.
Simulated data - LDDP model: True conditional densities of $y|x$ (in red), posterior mean estimates (black continuos line) and point-wise 95% HPD intervals (black dashed lines) for: (a) $x = 0.1$, (b) $x = 0.25$, (c) $x = 0.48$, (d) $x = 0.76$, and (e) $x = 0.88$. Panel (f) shows the data, along with the true and estimated mean regression curves.

Figure 3.
SIMCE data: Posterior estimates (mean and point-wise 95% HPD intervals) for the ability distribution for type of school and gender. The results for boys are shown in panels (a), (c), (e) and (g) for type of school Public I, Public II, Private I, and Private II, respectively. The results for girls are shown in panels (b), (d), (f) and (h) for type of school Public I, Public II, Private I, and Private II, respectively.
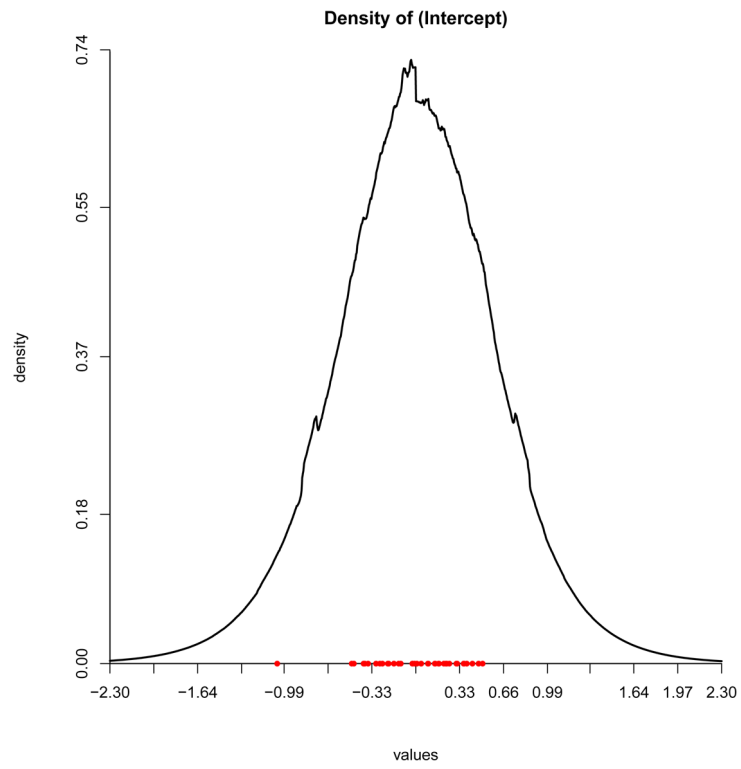
Figure 4.
Kidney data: Posterior mean of the frailty distribution. The density is overlaid on a plot of the posterior mean of the individuals frailty terms.
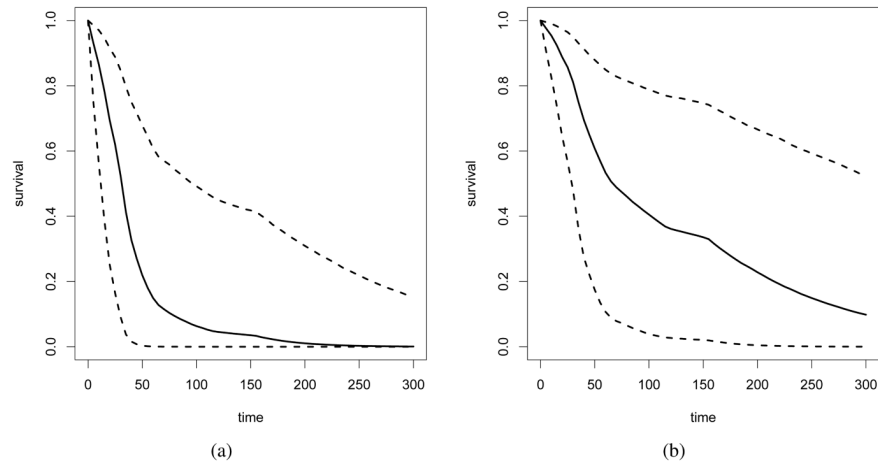
Figure 5.
Kidney data: Posterior estimates (median and point-wise 95% credible intervals) for the survival function for time to infection. The results for males and females are shown in panels (a) and (b), respectively.