Biometrika (2011), **98**, 3, *pp.* 553–566
doi: 10.1093/biomet/asq082

A class of mixtures of dependent tail-free processes

BY A. JARA

Department of Statistics, Pontificia Universidad Católica de Chile, Casilla 306, Correo 22, Santiago, Chile
ajara@mat.puc.cl

AND T. E. HANSON

Department of Statistics, University of South Carolina, Columbia, South Carolina 29208, U.S.A.
hansont@stat.sc.edu

SUMMARY

We propose a class of dependent processes in which density shape is regressed on one or more predictors through conditional tail-free probabilities by using transformed Gaussian processes. A particular linear version of the process is developed in detail. The resulting process is flexible and easy to fit using standard algorithms for generalized linear models. The method is applied to growth curve analysis, evolving univariate random effects distributions in generalized linear mixed models, and median survival modelling with censored data and covariate-dependent errors.

Some key words: Bayesian nonparametrics; Median regression; Partial exchangeability; Polya tree; Related probability distribution.

1. INTRODUCTION

Consider regression data (x_i, y_i) $(i = 1, \ldots, n)$, where $x_i \in \mathcal{X} \subset \mathbb{R}^q$ is a set of predictors, and y_i is a response variable on the real-line $\mathbb{R}$. Typically, parametric or semiparametric regression models specify a location function $m : \mathbb{R}^q \longrightarrow \mathbb{R}$, such as the mean or median, and assume that $y_i = m(x_i) + e_i$, where the errors, e_i, are independent random variables with common distribution G. However, for many data sets, it is not appropriate to assume that the error distribution is constant over the predictor space $\mathcal{X}$. For example, this assumption is typically violated in growth curve modelling (MacEachern, 1999; Kapitula & Bedrick, 2005). Although several parametric and semiparametric generalizations indexing the error distribution by the covariates, G_{x_i}, have been proposed (Kapitula & Bedrick, 2005), they enable only a few aspects of G_x to change with x. Therefore, these approaches do not enable a full investigation of changes in the conditional distribution of the response given the covariates. For instance, as time continues, heterogeneity in timing of biological processes can produce modal separation and changes in skew of the error distribution, simultaneously. The key challenging aspect is to anticipate and enable such distributional evolution without grossly overfitting the data. In the Bayesian nonparametric literature, this problem has been faced through the construction of prior distributions for the collection of related probability distributions $\{G_x : x \in \mathcal{X}\}$.

The problem of defining priors over related probability distributions has recently received increasing attention. To date, much effort has focused on constructions that generalize the widely used class of Dirichlet process priors (Ferguson, 1973, 1974). An exception is Tokdar et al. (2010), who model conditional densities related by continuous predictors based on logistic

Gaussian processes and subspace projection. MacEachern (1999) proposed the dependent Dirichlet process to define a full joint model on the set $\{G_x : x \in \mathcal{X}\}$, where every G_x follows a Dirichlet process marginally. The dependent Dirichlet process generalizes the representation of Sethuraman (1994) of the Dirichlet process, $G(\cdot) = \sum_{i=1}^{\infty} w_i \delta_{\eta_i}(\cdot)$, where the w_is are weights following a stick-breaking construction, $\delta_\eta(\cdot)$ denotes the Dirac measure at η, and the η_is are independent random support points with common distribution G_0. In the basic version of this approach, the random support points $\{\eta_i\}$ are replaced by stochastic processes indexed by x, $\eta_i = \{\eta_i(x) : x \in \mathcal{X}\}$ (Gelfand et al., 2005), or linear regressions $\eta_i = x^{\mathrm{T}}\beta_i$ (De Iorio et al., 2004, 2009; Jara et al., 2010). Alternatively, Griffin & Steel (2006) and Dunson et al. (2007) developed models where the dependence is introduced by making also the weights dependent on the covariates. An earlier approach that is related to the latter reference is Müller et al. (1996). These authors fitted a standard Dirichlet process mixture of multivariate Gaussian distributions to observations $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ and looked at the conditional predictive density $p(y_{n+1} \mid x_{n+1}, \mathcal{D})$. Although Müller et al. (1996) focused on the posterior predictive mean function $\{E(y_{n+1} \mid x, \mathcal{D}) : x \in \mathcal{X}\}$, their method can easily be extended to provide inferences for the conditional density at covariate level x, yielding a density regression model in the spirit of Dunson et al. (2007). Both Müller et al. (1996) and Dunson et al. (2007) produce estimates of $m(x)$ that are essentially an infinite sum of linear regressions more highly weighted by covariates $\{x_i\}$ near to x, similar to loess regression.

All of these approaches first induce dependence among aspects, weights and/or locations, of a Dirichlet process, then convolve the process with a smooth kernel. In contrast to these approaches we propose a natural, but remarkably effective, dependent nonparametric process $\{G_x : x \in \mathcal{X}\}$ obtained by simply modelling tail-free conditional probabilities with transformed Gaussian processes. A particular specification of the process is developed in detail which enables a logistic regression representation of the tail-free conditional probabilities and has the following desirable properties: the prior yields a density with respect to Lebesgue measure at each $x \in \mathcal{X}$, the process is continuous as a function of the predictors, the process has large support on the space of all conditional density functions, the posterior computation is simple, relying on algorithms for fitting generalized linear models, and the process closely matches a conventional Polya tree at each value of the predictor. Polya trees have been studied extensively in the literature and have desirable properties in terms of support and posterior consistency. The dependence of the process on the partitions associated to the tail-free construction is avoided by considering mixtures of the proposed process. The process can easily be constrained to have median zero, facilitating its use in median regression models, survival analysis and generalized mixed models with evolving univariate random effect distributions.

2. A GENERAL CLASS OF DEPENDENT TAIL-FREE PROCESSES

2·1. *Tail-free processes*

Tail-free processes are stochastic processes that can be defined to have trajectories on the space of all probability distributions on a given space (Freedman, 1963; Fabius, 1964; Ferguson, 1974). A tail-free random probability measure G with support on $\mathbb{R}$ is defined by allocations of random probabilities to increasingly refined partitions of $\mathbb{R}$. Let $E = \{0, 1\}$ and E^m be the m-fold Cartesian product $E \times \cdots \times E$. Further, set $E^* = \bigcup_{m=1}^{\infty} E^m$. Consider the sequence of partitions of $\mathbb{R}$ given by $\pi_0 = \{\mathbb{R}\}$, $\pi_1 = \{B_0, B_1\}$, $\pi_2 = \{B_{00}, B_{01}, B_{10}, B_{11}\}, \ldots$, such that $\mathbb{R} = B_0 \cup B_1$ and $B_0 \cap B_1 = \emptyset$, and for each $m \in \mathbb{N}$ and every $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^m$, $B_\epsilon = B_{\epsilon 0} \cup B_{\epsilon 1}$ and $B_{\epsilon 0} \cap B_{\epsilon 1} = \emptyset$. Assume that $B_{\epsilon 0}$ lies below $B_{\epsilon 1}$ and that for all $\epsilon \in E^*$, B_ϵ is a left-open right-closed

interval unless ϵ is a string of ones only. Throughout the paper, we use the convention that $\epsilon = \varepsilon_1 \cdots \varepsilon_{m-1} 0 = 0$ and $\epsilon = \varepsilon_1 \cdots \varepsilon_{m-1} 1 = 1$, if $m = 1$. Let $\Pi = \bigcup_{i=0}^{\infty} \pi_i$ and further assume that the partitions form a rich class in the sense that Π is a generator of the Borel σ-field of $\mathbb{R}$, $\mathcal{B} \equiv \mathcal{B}(\mathbb{R})$.

DEFINITION 1. *Let Π be a sequence of binary partitions as before. A random probability measure G on $(\mathbb{R}, \mathcal{B})$ is said to be a tail-free random probability measure with respect to the set of partitions Π if there exist a collection $\mathcal{Y} = \{Y_\epsilon : \epsilon \in E^*\}$ of $[0, 1]$-valued random variables such that the following hold:*

(i) the vectors $(Y_0, Y_1), (Y_{00}, Y_{01}, Y_{10}, Y_{11}), \ldots$, are mutually independent;
(ii) for every $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$, $Y_{\varepsilon_1 \cdots \varepsilon_{m-1} 0} + Y_{\varepsilon_1 \cdots \varepsilon_{m-1} 1} = 1$ almost surely;
(iii) for every $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$, the random probability measure G is related to $\mathcal{Y}$ through the relations

$$G(B_\epsilon) = \prod_{j=1}^{m} Y_{\varepsilon_1 \cdots \varepsilon_j}.$$

We consider partition sets in Π such that their limits correspond to quantiles of a parametric distribution G_θ, $\theta \in \Theta$, defined on $(\mathbb{R}, \mathcal{B})$. Specifically, we consider sets in Π of the form $B_\epsilon^\theta = (l_\epsilon^\theta, u_\epsilon^\theta]$, where $l_\epsilon^\theta = G_\theta^{-1}(k/2^m)$ and $u_\epsilon^\theta = G_\theta^{-1}\{(k+1)/2^m\}$, with $G_\theta^{-1}(\cdot)$ being the quantile function of G_θ, and k is the decimal representation of $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$. If needed, the notation Π^θ will be used to make the dependence of Π on the parameters of G_θ explicit. Without loss of generality, for the rest of the paper we assume that the sets are constructed based on the quantiles of the $N(0, \theta)$ distribution, which implies that the collection of right endpoints of B_ϵ^θ is dense in $\mathbb{R}$ and that Π^θ is a generator of $\mathcal{B}$.

The class of tail-free processes admits Polya trees and Dirichlet processes as important special cases. Polya trees assign the elements in the collection $\{Y_{\epsilon 0}\}_{\epsilon \in E^*}$-independent beta distributions. The Dirichlet process is tail-free with respect to any sequence of partitions and is the only process in the class that has this property. Tail-free priors satisfy a zero-one law: the random measure generated by a tail-free process is absolutely continuous with respect to a given finite measure with probability zero or one (Ghosh & Ramamoorthi, 2003). Dubins & Freedman (1967), Kraft (1964) and Metivier (1971) gave useful sufficient conditions for the almost sure continuity and absolute continuity of a tail-free process. We discuss these conditions in the context of the dependent tail-free process in the next sections.

2·2. *Dependent tail-free processes*

In order to introduce dependence in random probability measures, we consider transformed Gaussian processes, indexed by predictors $x \in \mathcal{X} \subset \mathbb{R}^q$, where $\mathcal{X}$ is a bounded subset of $\mathbb{R}^q$, to define the tail-free conditional probabilities.

DEFINITION 2. *Let Π be a sequence of binary partitions as before, let $h : \mathbb{R} \longrightarrow [0, 1]$ be a strictly increasing continuous function, let $\mathcal{A} = \{V_{\epsilon 0} : \epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*\}$ be a set of covariance functions, and let $\mathcal{P}(\mathbb{R})$ be the set of Borel probability measures on $(\mathbb{R}, \mathcal{B})$. Let $\{G(x, \omega) : x \in \mathcal{X}\}$ be a $\mathcal{P}(\mathbb{R})$-valued stochastic process on an appropriate probability space $(\Omega, \mathcal{F}, P)$ such that:*

(i) the sets $\{\eta_{\epsilon 0}(x, \omega) : x \in \mathcal{X}\}$, for every $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, are realizations of mutually independent zero-mean Gaussian processes with covariance functions $V_{\epsilon 0}$, respectively;

(ii) for every $x \in \mathcal{X}$ *and every* $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, $Y_{\epsilon 0}(x, \omega) = h\{\eta_{\epsilon 0}(x, \omega)\}$ *and* $Y_{\epsilon 1}(x, \omega) = 1 - Y_{\epsilon 0}(x, \omega)$;

(iii) for every $x \in \mathcal{X}$ *and every* $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$,

$$G(x, \omega)(B_\epsilon) = \prod_{j=1}^{m} Y_{\varepsilon_1 \cdots \varepsilon_j}(x, \omega).$$

Such a process $\{G_x = G(x, \omega) : x \in \mathcal{X}\}$ *will be referred to as a dependent tail-free process with parameters* $(h, \Pi, \mathcal{A})$*, and denoted* DTFP$(h, \Pi, \mathcal{A})$.

The definition of the dependent tail-free process implies that if h is taken to be the cumulative distribution function of any absolutely continuous random variable with distribution symmetric around 0 and the limits of the partition sets are taken such that they coincide with quantiles of G_θ, then the collection $\{G_x : x \in \mathcal{X}\}$ is centred around G_θ, that is, $E(G_x) = G_\theta$, for all $x \in \mathcal{X}$. We refer the reader to Appendix A of the Supplementary Material for a proof. Choices for h that achieve this goal are the standard logistic or normal cumulative distribution functions. Other links can be considered. The logistic link

$$Y_{\epsilon 0}(x, \omega) = \frac{\exp\{\eta_{\epsilon 0}(x, \omega)\}}{1 + \exp\{\eta_{\epsilon 0}(x, \omega)\}},$$

for every $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, is used throughout the paper primarily to simplify and speed up computations in model fitting.

From the definition of the dependent tail-free process, it is easy to see that regardless of the link function h, $G(x, \cdot)$ is a tail-free process for all $x \in \mathcal{X}$. Furthermore, for every $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$ and $\epsilon 1 = \varepsilon_1 \cdots \varepsilon_{m-1} 1 \in E^*$, the vectors $\{Y_{\epsilon 0}(x, \cdot), Y_{\epsilon 1}(x, \cdot)\}$ are mutually independent. This allows us to specify the process in such a way that marginally, $G(x, \cdot)$ closely match Polya tree priors in § 3·2. Similarly to tail-free processes for single probability measures, the dependent tail-free process can easily be constrained in order to have trajectories on the space of the probability distributions on $(\mathbb{R}, \mathcal{B})$ with median zero almost surely, that is, such that all elements in $\{G_x : x \in \mathcal{X}\}$ have median zero with probability one, facilitating the use in location models. This can easily be done by letting $\eta_0(x, \cdot) = 0$ almost surely, for every $x \in \mathcal{X}$.

The form of the covariance functions in $\mathcal{A}$ completely defines the process and determines important properties. Its specification determines whether the trajectories of the process have a density with respect to Lebesgue measure. Furthermore, under certain specifications of the covariance functions, the process is stochastically continuous at every $x \in \mathcal{X}$ and can be almost surely a continuous function on $\mathcal{X}$. The following proposition is proved in Appendix B of the Supplementary Material.

PROPOSITION 1. *Let* $\{G_x : x \in \mathcal{X}\} \mid h, \Pi, \mathcal{A} \sim$ DTFP$(h, \Pi, \mathcal{A})$*, and let* c *and* C *be positive constants. If every covariance function* $V \in \mathcal{A}$ *is specified and satisfies the Lipschitz condition, that is,* $V(x_1, x_1) + V(x_2, x_2) - 2V(x_1, x_2) \leqslant C||x_1 - x_2||^c$, $x_1, x_2 \in \mathcal{X}$*, then for all* $\{x_j\}_1^\infty$*, with* $x_j \in \mathcal{X}$ *such that* $\lim_{j \to +\infty} x_j = x$*, and for all* $B \in \Pi$, $G(x_j, \cdot)(B)$ *converges in probability to* $G(x, \cdot)(B)$ *when* $j \to +\infty$*, for all* $x \in \mathcal{X}$*. If, in addition to the Lipschitz condition, the associated Gaussian processes are almost surely separable, then* $G(\cdot, \omega)(B)$ *is almost surely a continuous function on* $\mathcal{X}$*, for all* $B \in \Pi$.

The specification of the covariance functions in $\mathcal{A}$ also determines the support properties of the process. The following proposition is proved in Appendix C of the Supplementary Material.

PROPOSITION 2. *Let* $\{G_x : x \in \mathcal{X}\} \mid h, \Pi, \mathcal{A} \sim \mathrm{DTFP}(h, \Pi, \mathcal{A})$, *where* h *is the logistic transformation and the partition sets in* Π, *created as before, are based on the probability distribution* G_θ *defined on* $(\mathbb{R}, \mathcal{B})$. *Assume that the dependent tail-free process is specified such that for every* $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{j-1} 0 \in E^*$, $\{\eta_{\epsilon 0}(x, \cdot) : x \in \mathcal{X}\}$ *is a Gaussian process with continuous sample paths and continuous covariance kernel such that* $V_{\epsilon 0}(x, x) = 2/\rho(j)$, *for all* $x \in \mathcal{X}$, *where* $\rho : \mathbb{N} \longrightarrow \mathbb{R}^+$ *is an increasing function such that* $\sum_{j=1}^{\infty} \rho(j)^{-1} < \infty$. *Let* $\{G_x^0 : x \in \mathcal{X}\}$ *be a set of probability distributions defined on* $(\mathbb{R}, \mathcal{B})$ *and such that, for every* $x \in \mathcal{X}$, $G_x^0 \ll G_\theta$ *and has a density with respect to Lebesgue measure. If for every* $\epsilon 0 \in E^*$, *the composition of functions* $\eta_{\epsilon 0}^0 = (h^{-1} \; o \; y_{\epsilon 0}^0) : \mathcal{X} \to \mathbb{R}$ *is in the reproducing kernel Hilbert space of* $V_{\epsilon 0}$, *where* $\{y_{\epsilon 0}^0(x) = G_x^0(B_{\epsilon 0} \mid B_\epsilon) : x \in \mathcal{X}\}$ *is the corresponding conditional probability function, then*

$$\mathrm{pr}\left[\omega \in \Omega : \sup_{x \in \mathcal{X}} \mathrm{D_{KL}}\{G_x^0, G(x, \omega)\} < \alpha\right] > 0,$$

for all $\alpha > 0$, *where* $\mathrm{D_{KL}}$ *denotes the Kulback–Leibler divergence.*

For many choices of covariance kernels in $\mathcal{A}$, the associated reproducing kernel Hilbert space is dense in the space of the continuous functions on $\mathcal{X}$, in which case the corresponding Gaussian process assigns positive probability to the uniform neighbourhoods of every continuous function and, therefore, every continuous set of conditional densities will be in the Kulback–Leibler support of the dependent tail-free process. Natural choices for longitudinal and/or spatial modelling, which induce continuous dependent tail-free processes with large support, are the Ornstein–Uhlenbeck process, or an integrated version thereof (Wang & Taylor, 2001). Although flexible, priors such as these require the sampling of an entire tree of conditional probabilities at each distinct covariate value. Inference can take on the order of days to obtain and in many cases is not practically possible. In the next section, we develop an approach that achieves the continuity property, has large support and enables straightforward Markov chain Monte Carlo implementations. The construction is a particular case of a dependent tail-free process and is based on a connection between Gaussian processes and linear models.

3. A TRACTABLE MIXTURE OF DEPENDENT TAIL-FREE PROCESSES

3·1. *Linear dependent tail-free processes*

We now examine a special case of the general model where for all $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, the $\{\eta_{\epsilon 0}(x) : x \in \mathcal{X}\}$ are Gaussian processes such that for any finite sequence of distinct elements $x_1, \ldots, x_l$ of $\mathcal{X}$, the processes have finite-dimensional multivariate normal distributions given by

$$\{\eta_{\epsilon 0}(x_1), \ldots, \eta_{\epsilon 0}(x_l)\}^{\mathrm{T}} \sim N_l\{0, \Sigma(x_1, \ldots, x_l, W_{\epsilon 0})\},$$

where $\Sigma(x_1, \ldots, x_l, W_{\epsilon 0}) = D(x_1, \ldots, x_l) W_{\epsilon 0} D(x_1, \ldots, x_l)^{\mathrm{T}}$, with $D(x_1, \ldots, x_l)$ being a full-rank $l \times r$ design matrix, whose rows are given by r-dimensional vectors $d(x_i)^{\mathrm{T}}$ based on the continuous predictors x_i, and $W_{\epsilon 0}$ is a $r \times r$-dimensional covariance matrix. Clearly, these processes can be induced by the linear, in the coefficients, model given by

$$\eta_{\epsilon 0}(x, \beta_{\epsilon 0}) = d(x)^{\mathrm{T}} \beta_{\epsilon 0}, \quad \beta_{\epsilon 0} \sim N_r(0, W_{\epsilon 0}).$$

It is important to stress that the linear specification can be interpreted as an approximation to any general function, which can be very flexible by including standard nonlinear transformations of the original continuous predictors, such as additive models based on B-splines or series

expansions, as well as linear forms in the continuous predictors themselves and other discrete predictors. From Proposition 2, the greater the flexibility of the linear specification, the larger the support of the dependent tail-free process, which is determined by the reproducing kernel Hilbert space of the covariance functions. As discussed in § 5, the complexity needed by the model can be evaluated using standard model determination techniques.

DEFINITION 3. *Let* Π *be a sequence of binary partitions as before, let* $h : \mathbb{R} \longrightarrow [0, 1]$ *be a strictly increasing continuous function, let* $d : \mathbb{R}^q \to \mathbb{R}^r$ *be a design-vector generating one-to-one mapping, let* $\mathcal{A} = \{W_{\epsilon 0} : \epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*\}$ *be a set of* $r \times r$ *covariance matrices, and let* $\mathcal{P}(\mathbb{R})$ *be the set of Borel probability measures on* $(\mathbb{R}, \mathcal{B})$. *Let* $\{G(x, \omega) : x \in \mathcal{X}\}$ *be a* $\mathcal{P}(\mathbb{R})$*-valued stochastic process on an appropriate probability space* $(\Omega, \mathcal{F}, P)$ *such that:*

(i) the r*-dimensional vectors* $\beta_{\epsilon 0}(\omega)$*, for every* $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$*, are realisations of mutually independent and normally distributed random vectors with zero-mean and covariance matrix* $W_{\epsilon 0}$*, respectively*;

(ii) for every $x \in \mathcal{X}$ *and every* $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, $Y_{\epsilon 0}(x, \omega) = h\{d(x)^{\mathrm{T}} \beta_{\epsilon 0}(\omega)\}$ *and* $Y_{\epsilon 1}(x, \omega) = 1 - Y_{\epsilon 0}(x, \omega)$;

(iii) for every $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$,

$$G(x, \omega)(B_\epsilon) = \prod_{j=1}^{m} Y_{\varepsilon_1 \ldots \varepsilon_j}(x, \omega).$$

Such a process $\{G_x = G(x, \omega) : x \in \mathcal{X}\}$ *will be referred to as a linear dependent tail-free process with parameters* $(h, \Pi, \mathcal{A})$*, and denoted* $\mathrm{LDTFP}(h, \Pi, \mathcal{A})$.

3·2. *Prior specification*

If the dimension of the design vectors is smaller than the sample size, the linear dependent tail-free process represents a reduction of the dimensionality of the inferential problem. In fact, the problem is reduced to updating independent binary regression coefficients $\beta_{\epsilon 0}$, for all $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, instead of the finite dimensional realisations of the Gaussian processes in each distinct predictor value. However, there are still countably many parameters $\{W_{\epsilon 0}\}$ that need to be specified. A default prior specification is clearly needed. A reasonable linear dependent tail-free process specification would take the scale, location and correlation among the predictor variables into account in an effort to standardize and bound prior variability across representative predictor values. In the context of binary regression models, Gelman et al. (2008) suggest standardizing the data and then placing independent Cauchy prior distributions on the regression coefficients. This is reasonable, and produces good results, but does not take into account correlation among the predictor variables. A prior that is location-scale invariant in terms of the predictors is a suitably modified version of Zellner's g-prior (Zellner, 1983), originally developed as a reference informative prior for Gaussian linear models. The basic idea extends to the linear dependent tail-free setting by noting that, as shown below, $\mathrm{beta}(a, a) \approx \mathrm{logit}\, N(0, 2/a)$ and taking $\beta_{\epsilon 0} \sim N_r(0, g_{\epsilon 0} \Sigma^{-1})$, where $\Sigma = D^{\mathrm{T}} D$, $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$ and the $g_{\epsilon 0}$s are positive constants, that is, $W_{\epsilon 0} = g_{\epsilon 0} (D^{\mathrm{T}} D)^{-1}$. Thus the logit tail-free probabilities at the observed design points satisfy

$$D\beta_{\epsilon 0} \sim N_n(0, g_{\epsilon 0} D \Sigma^{-1} D^{\mathrm{T}}).$$

This specification implies positive correlation among probabilities for design points that are near each other. The minimal variance occurs at $\bar{d}$, and is $\text{var}(\bar{d}^{\mathrm{T}}\beta_{\epsilon 0}) = g_{\epsilon 0}/n$.

Christensen (2002) points out that the perpendicular projection matrix $M = D(D^{\mathrm{T}}D)^{-1}D^{\mathrm{T}}$ has diagonal leverage elements $m_{ii} = n^{-1} + \hat{\Delta}_i^2/(n-1)$, where $\hat{\Delta}_i^2 = (d_i^* - \bar{d}^*)^{\mathrm{T}}S^{-1}(d_i^* - \bar{d}^*)$ is the sample Mahalanobis distance of the predictors without the intercept, that is, $d_i = (1, d_i^*)$ and $(n-1)S^{-1} = \sum_{i=1}^n (d_i^* - \bar{d}^*)(d_i^* - \bar{d}_i^*)'$. Also note $\sum_{i=1}^n m_{ii} = r$ and $n^{-1} \leqslant m_{ii} \leqslant 1$. If the logistic transformation is considered, the further d_i^* is from $\bar{d}^*$, the greater the variability of $Y_{\epsilon 0}(x_i, \cdot)$ around 0·5 under the prior. Zellner (1983) suggests taking $g_{\epsilon 0} = f(n)$ to grow with n, inducing prior precision to grow with n, that is, learning by doing. We suggest choosing $g_{\epsilon 0} = 2n/c\rho(l_{\epsilon 0})$, where $l_{\epsilon 0}$ is the length of $\epsilon 0$, that is, $\epsilon 0 \in E^{l_{\epsilon 0}}$, and $\rho : \mathbb{N} \longrightarrow \mathbb{R}^+$ is an increasing function. This has the effect of standardizing the prior variability at $\bar{x}$, that is, $\text{var}\{Y_{\epsilon 0}(\bar{x}, \cdot)\}$, to be approximately that of traditional Polya tree priors with precision parameter c.

The similarity between the marginal realizations of the linear-dependent tail-free process under this specification and Polya trees is explained by the similarity of the symmetric logistic normal and symmetric beta distributions. In fact, although the logistic normal distribution, introduced by Johnson (1949) as his type $\mathcal{S}_B$ class, and beta distributions can look very similar, a stronger statement can be made. The following proposition is proved in Appendix D of the Supplementary Material.

PROPOSITION 3. *Let $y_a = e^{\eta_a}/(1+e^{\eta_a})$, $\eta_a \sim N(0, 2/a)$ and $b_a \sim \text{beta}(a, a)$. Then y_a and b_a are asymptotically equivalent as $a \to \infty$.*

For a given value of c, the function $\rho(\cdot)$ determines the type of trajectories of the process. In fact, for certain functions $\rho(\cdot)$, the linear dependent tail-free process has a density with respect to Lebesgue measure. The following proposition is proved in Appendix E of the Supplementary Material.

PROPOSITION 4. *Let $\{G_x : x \in \mathcal{X}\} \sim \text{LDTFP}(h, \Pi, \mathcal{A}^{c,\rho})$, where h is the logistic transformation and the set $\mathcal{A}^{c,\rho}$ is specified using the g-prior as before with $\sum_{l=1}^{\infty} \rho(l)^{-1} < \infty$. Then the process has trajectories on the space of the absolutely continuous probability distributions almost surely, that is, G_x has a density with respect to Lebesgue measure for all $x \in \mathcal{X}$.*

3·3. *Partially specified linear-dependent tail-free process*

Exact calculations cannot be performed with a fully specified linear-dependent tail-free process, since the predictive distribution at each value of the predictor x does not have a closed form. This justifies the use of a partially specified or finite version we consider here, where the tail-free process is terminated at level J, which we denote $\{G_x : x \in \mathcal{X}\} \sim \text{LDTFP}^J(h, \Pi, \mathcal{A}^{c,\rho})$. We typically consider $J \approx \log_2(n/N)$, where n is the sample size and N is 5–10 (Hanson, 2006). This rule of thumb is even more conservative for the linear-dependent tail-free process than for regular Polya trees. If the true data generating mechanism is similar to a member of the centering family, then on average two observations will determine the coefficients $\beta_{\epsilon 0}$, $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{J-1} 0 \in E^J$, beyond the intercept at level J when $N = 1$. This amount of data is essentially negligible relative to information in the $\text{logit}[N\{0, 2/(cJ^2)\}] \approx \text{beta}(cJ^2, cJ^2)$ prior at this level.

As is usually done for partially specified Polya tree priors, on sets in the finest partition π_J^θ the random G_x follows the parametric distribution G_θ (Hanson, 2006). It follows that, under the logistic transformation, the conditional density sample path of a partially specified linear

dependent tail-free process is given by

$$g(\cdot,\omega)(e) = 2^J\phi_\theta(e)\prod_{j=1}^{J} Y_{\epsilon_\theta(e,j)}(\cdot,\omega),$$

$$= 2^J\phi_\theta(e)\prod_{j=1}^{J}\frac{\exp\{d(\cdot)^{\mathrm{T}}\beta_{\epsilon_\theta(e,j-1)0}(\omega)\}^{I\{e\in B^\theta_{\epsilon_\theta(e,j-1)0}\}}}{1+\exp\{d(\cdot)^{\mathrm{T}}\beta_{\epsilon_\theta(e,j-1)0}(\omega)\}}, \tag{1}$$

where $e\in\mathbb{R}$, $I\{A\}$ is the indicator function for A, $\epsilon_\theta(e,j)=\varepsilon_1\varepsilon_2\cdots\varepsilon_j$ is the set in π^j_θ that e is in, and $\phi_\theta(\cdot)$ is the density of a $N(0,\theta^2)$ variate. Since expression (1) is a continuous function of x, the following continuity property of the process can be proved. We refer the reader to Appendix F of the Supplementary Material for a proof.

PROPOSITION 5. *Let* $\{G_x : x\in\mathcal{X}\}\sim \mathrm{LDTFP}^J(h,\Pi^\theta,\mathcal{A}^{c,\rho})$, *where* h *is the logistic transformation. Then for all* $\{x_j\}_1^\infty$, *with* $x_j\in\mathcal{X}$, *such that* $\lim_{j\to+\infty}x_j = x\in\mathbb{R}^q$,

$$\lim_{j\to+\infty}\sup_{B\in\mathcal{B}(\mathbb{R})}|G(x_j,\cdot)(B)-G(x,\cdot)(B)|=0,$$

for all $x\in\mathcal{X}$, *that is,* $G(x_j,\cdot)$ *converges in total variation norm to* $G(x,\cdot)$ *when* $x_j\longrightarrow x$.

3·4. *Mixtures of linear-dependent tail-free process*

In practice, it may be difficult to specify a single centring $N(0,\theta^2)$ distribution with which to centre the linear-dependent tail-free process; and once specified, a single centring distribution may affect inference unduly. One way to mitigate the dependence of the process on the partitioning sets is to specify a mixture of prior distributions. A mixture of linear-dependent tail-free processes is induced for $\{G_x : x\in\mathcal{X}\}$ by allowing parameters of the centring distribution G_θ and/or the precision parameters c to be random, that is,

$$\{G_x : x\in\mathcal{X}\}\sim\mathrm{LDTFP}^J(h,\Pi^\theta,\mathcal{A}^{c,\rho}),\quad (\theta,c)\sim Q,$$

where Q refers to the joint prior for θ and c. Smoothness properties in terms of continuity and differentiability of the densities for G_x at a fixed x under the mixture of partially specified linear-dependent tail-free process carry over from Hanson (2006). One important property is posterior propriety under improper priors on the mixing parameter θ, following a simple application of Tonelli's theorem.

3·5. *Posterior inference*

In this section we present results useful for the computation of posterior distributions involving related random distributions modelled using partially specified linear-dependent tail-free processes or mixtures of them. Under independent sampling, the joint density for an n-dimensional vector of errors is given by

$$p(e_1,\ldots,e_n|\theta,\beta)=\left\{\prod_{i=1}^n 2^J\phi_\theta(e_i)\right\}\prod_{\epsilon 0\in E^J}\prod_{i=1}^n\frac{\exp\{d(x_i)^{\mathrm{T}}\beta_{\epsilon 0}\}^{I\{e_i\in B_\theta(\epsilon 0)\}}}{[1+\exp\{d(x_i)^{\mathrm{T}}\beta_{\epsilon 0}\}]^{I\{e_i\in B_\theta(\epsilon)\}}}.$$

This expression has the form of 2^J-1 logistic regression kernels, one for each $\epsilon 0$, times the likelihood for θ obtained from fitting the standard parametric family $N(0,\theta^2)$ to data

$\mathrm{e} = (e_1, \ldots, e_n)$. This forms the basis of an efficient Markov chain Monte Carlo scheme for obtaining posterior inference. In fact, conditionally on θ, any algorithm for fitting binary regression models can be used to update the $\beta_{\epsilon 0}$ parameters (Gamerman, 1997; Holmes & Held, 2006). Metropolis–Hastings steps with normal proposals based on one step of the Newton–Raphson algorithm (Gamerman, 1997) are used here.

Placing a prior on c increases flexibility, essentially allowing data to estimate $g_{\epsilon 0}$, and the g-prior provides a conjugate update. Let $S = 0{\cdot}25n^{-1} \sum_{\epsilon 0 \in E^J} \rho(l_{\epsilon 0}) || D\beta_{\epsilon 0}||^2$. If $c \sim \Gamma(a_c, b_c)$, then the full conditional distribution for c is also a gamma distribution $c \sim \Gamma\{a_c + p(2^{J-1} - t), b_c + S\}$, where $t = 1$ for median a median zero specification and $t = 0{\cdot}5$ otherwise. The parameter θ can be updated via a random walk Metropolis–Hastings step. Assume, for instance, the noninformative prior $p(\theta) \propto 1/\theta$ and say $\log(\theta^*) \sim N\{\log(\theta), v\}$ for some v. Alternatively, a slice sampler can be used to update the centring parameter θ.

For assessing normal ranges in growth chart applications, the quantile function is needed. Generalizing Hanson (2006), given $\Delta = \{\beta_{\epsilon 0} : \epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*, m \leqslant J\}$, let

$$p_x(k \mid \Delta, \theta) = G_x\{B_\theta(\epsilon)\} = \prod_{j=1}^{J} Y_{\varepsilon_1 \cdots \varepsilon_j}(x, \beta_{\varepsilon_1, \ldots, \varepsilon_{j-1} 0}),$$

where $\epsilon = \varepsilon_1, \ldots, \varepsilon_J$ is the decimal representation of $k - 1$, $k = 1, \ldots, 2^J$. For median zero specification, $Y_0(x, \beta_0) = Y_1(x, \beta_0) = 0.5$, that is, $\beta_0 = 0$. The lth quantile, $0 < l < 1$, associated with G_x is given by

$$G_x^{-1}(l \mid \Delta, \theta) = G_\theta^{-1} \left\{ \frac{l - \sum_{k=1}^{N} p_x(k \mid \Delta, \theta) + N p_x(N \mid \Delta, \theta)}{2^J p_x(N \mid \Delta, \theta)} \right\}, \tag{2}$$

where N is such that $\sum_{k=1}^{N-1} p_x(k \mid \Delta, \theta) < l \leqslant \sum_{k=1}^{N} p_x(k \mid \Delta, \theta)$. Obtaining probability intervals for posterior quantiles from Markov chain Monte Carlo output is straightforward based on this formula. For the median regression model discussed in the next section, covariates x add the term $m(x)$ to (2).

4. Application

We illustrate the potential use of the dependent tail-free process by means of the analysis of simulated and real-life data sets. In order to save space, the analysis of a growth curve estimation problem is presented in this section only. The analyses of simulated and other real-life data analyses are given in the Supplementary Material. Specifically, the results of the analyses of simulated data are given in Appendix G in the Supplementary Material. Examples of the use of the dependent tail-free process in accelerated failure time models and generalized linear mixed models are given in Appendices H and I in the Supplementary Material, respectively. User-friendly functions implementing Markov chain Monte Carlo algorithms to fit the models considered were written in compiled language and incorporated into the R library DPpackage (Jara, 2007).

Isaacs et al. (1983) considered serum immunoglobulin G concentrations from $n = 298$ children aged 6 months to 6 years old. These data were further analysed by Royston & Wright (1998) and Kapitula & Bedrick (2005) using the parametric exponential normal family, which includes parameters for skew and kurtosis that can be functions of covariates. Like these authors we consider the log-transformation of the data y_i. The data are plotted in Fig. 1.

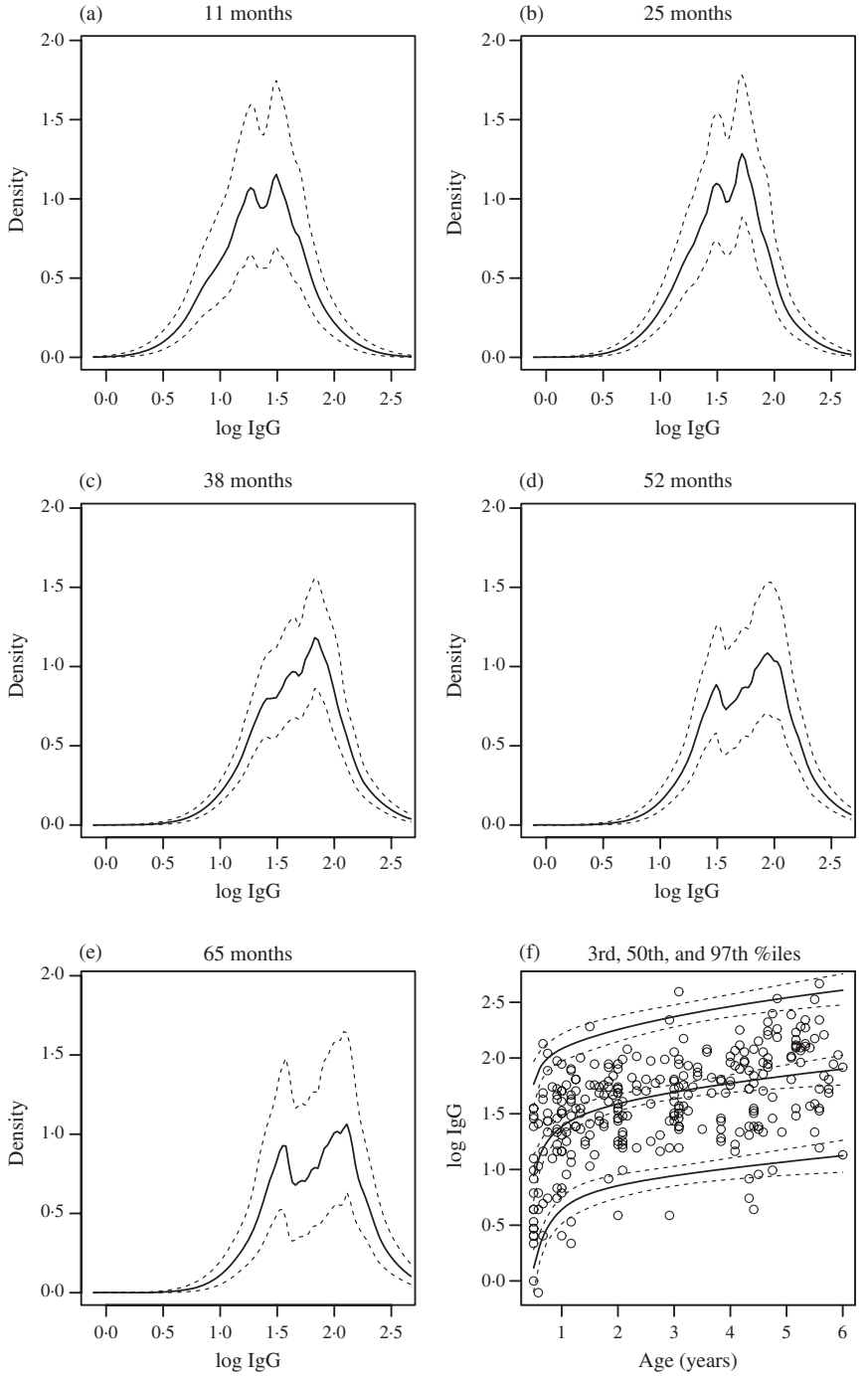

Fig. 1. Serum immunoglobulin G data. The posterior mean (solid) and 95% pointwise credibility interval (dots) of the conditional density of the logarithm of the concentration of serum immunoglobulin G at 11, 25, 38, 52 and 65 months are shown in panels (a), (b), (c), (d) and (e), respectively. Panel (f): the posterior means of the 3rd, 50th and 97th percentiles, with the corresponding 95% pointwise credibility intervals and the data.

The density for an observation with age x, in years, from the model fitted by Kapitula & Bedrick (2005), is given by

$$f_x(y \mid \gamma, \xi, \tau) = \frac{(2\pi)^{-1/2}}{\Phi(|\tau|^{-1})\sigma(x)} \exp\left\{-\frac{(e^{\tau z}-1)^2}{2\tau^2} - \tau z\right\},$$

where $z = \{y - \mu(x)\}/\sigma(x)$, $\mu(x) = \gamma_0 + \gamma_1 x^2 + \gamma_2 x^{-2}$, and $\sigma(x) = \xi_0 + \xi_1 x^{-2}$. For comparison, this model was fitted in WinBUGS assuming independent $N(0, 10^3)$ priors on all six regression effects $(\gamma_0, \gamma_1, \gamma_2, \xi_0, \xi_1, \tau)$. We obtained posterior estimates for model parameters consistent with Kapitula & Bedrick (2005).

We fitted the linear-dependent tail-free regression model to the log-transformed immunoglobulin G values, by considering

$$y_i = \gamma_0 + \gamma_1 x_i^2 + \gamma_2 x_i^{-2} + e_i, \qquad (3)$$
$$e_i \mid G_{x_i} \overset{\text{ind.}}{\sim} G_{x_i},$$

and

$$\{G_x : x \in \mathcal{X}\} \mid h, \theta, c, \rho \sim \text{LDTFP}^J(h, \Pi^\theta, \mathcal{A}^{c,\rho}). \qquad (4)$$

In order to avoid the identification problem associated with the confounding between the location of the error distributions and γ_0, the median-zero linear-dependent tail-free specification was considered. The tail-free conditional probabilities were modelled using the same specification as for the median function in (3) and a logistic link. The linear-dependent tail-free regression parameters were assigned g-priors, parameterized by the precision parameter c, and $J = 4$. As for the parametric model, the median function parameters γ_0 and γ_1 were given independent $N(0, 10^3)$ priors. Finally, the following priors were considered for the centring variance and the precision parameter of the linear-dependent tail-free process: $\theta^{-2} \sim \Gamma(10^{-4}, 10^{-4})$ and $c \sim \Gamma(5, 1)$. A conservative Markov chain Monte Carlo specification was considered for making posterior inferences. A burn-in of 20 000 iterates was followed by a run of 800 000 thinned down to 20 000 iterates. Time series plots of the stored output, not shown, suggest a good mixing of the chain, even with shorter runs.

The linear-dependent tail-free model was also compared with the Dirichlet process mixture of normals approach of Müller et al. (1996) and to the linear-dependent Dirichlet process approach of De Iorio et al. (2004, 2009). The competing models were compared in terms of the log pseudo marginal likelihood (Geisser & Eddy, 1979). For the approach of Müller et al. (1996), we consider the multivariate extension of the univariate Dirichlet process mixture of normals model of Escobar & West (1995) to fit the complete data $w_i = (y_i, x_i)^{\mathrm{T}}$, and focus on the mean function, as in Müller et al. (1996), and conditional densities $f(y|x)$ arising from the model. The Dirichlet process mixture model is given by

$$w_i \mid \mu_i, \Sigma_i \overset{\text{ind.}}{\sim} N_2(\mu_i, \Sigma_i), \quad (\mu_i, \Sigma_i) \mid G \overset{\text{iid}}{\sim} G, \quad G \mid \alpha, G_0 \sim DP(MG_0),$$

where the baseline distribution G_0 is the conjugate normal-inverted-Wishart distribution $G_0 \equiv N_2(\mu \mid m_1, \kappa_0^{-1}\Sigma) IW_2(\Sigma \mid \nu_1, \Psi_1)$. To complete the model specification, the following hyper-priors were assumed: $M \mid a_0, b_0 \sim \Gamma(a_0, b_0)$, $m_1 \mid m_2, S_2 \sim N_2(m_2, S_2)$, $\kappa_0 \mid \tau_1, \tau_2 \sim \Gamma(\tau_1/2, \tau_2/2)$, and $\Psi_1 \mid \nu_2, \Psi_2 \sim IW_2(\nu_2, \Psi_2)$. The linear-dependent Dirichlet process model

can be represented as Dirichlet process mixture of linear, in the coefficients, regression models

$$y_i \mid \beta_i, \sigma_i^2 \overset{\text{ind.}}{\sim} N(z_i^{\mathrm{T}}\beta_i, \sigma_i^2), \quad (\beta_i, \sigma_i^2) \mid G \overset{\text{iid}}{\sim} G, \quad G \mid \alpha, G_0 \sim DP(MG_0),$$

where $G_0 \equiv N_p(\beta \mid \mu_\beta, \Sigma_\beta)\Gamma(\sigma^{-2} \mid s_1/2, s_2/2)$. The linear-dependent Dirichlet process was fitted using the same regression functions as considered for the linear-dependent tail-free model, that is, $p = 3$ and $z_i^{\mathrm{T}} = (1, x_i^2, x_i^{-2})$. The model specification was completed with the following hyper-priors: $M \mid a_0, b_0 \sim \Gamma(a_0, b_0)$, $s_2 \mid \tau_{s_1}, \tau_{s_2} \sim \Gamma(\tau_{s_1}/2, \tau_{s_2}/2)$, $\mu_\beta \mid a, A \sim N_3(a, A)$, and $\Sigma_\beta \mid \nu_\beta, \Psi_\beta \sim IW_p(\nu_\beta, \Psi_\beta)$.

We fitted marginalized versions of Dirichlet process-based models where G is integrated out and standard algorithms to fit these models. To obtain credible intervals for the conditional densities and mean function, we used the ϵ-Dirichlet process approach proposed by Muliere & Tardella (1998), with $\epsilon = 0{\cdot}01$ and based on the Markov chain Monte Carlo samples, similar to Gelfand & Kottas (2002). The inferences were obtained using the Markov chain Monte Carlo specification described above for linear-dependent tail-free models and the following hyper-parameters, $a_0 = 5$, $b_0 = 1$, $\nu_1 = \nu_2 = 4$, $m_2 = (0, 0)^{\mathrm{T}}$, $S_2 = \mathrm{diag}(10^3, 10^3)$, $\Psi_2 = \mathrm{diag}(1, 1)$, $\tau_1 = 2{\cdot}01$, $\tau_2 = 0{\cdot}01$, $s_1 = 6$, $\tau_{s_1} = 6$, $\tau_{s_2} = 2$, $a = (0, 0, 0)^{\mathrm{T}}$, $A = \mathrm{diag}(10^3, 10^3, 10^3)$, $\nu_\beta = 5$, and $\Psi_\beta = \mathrm{diag}(1, 1, 1)$.

Figure 1 shows plots of the estimated 3rd, 50th, and 97th percentiles along with 95% credible intervals. These estimates are similar to those obtained by Royston & Wright (1998). The rest of Fig. 1 shows the evolution of the estimated conditional densities over time. There seems to be a more or less static clump of mass with mode around 1·5 in all panels, and an additional mode that continuously increases from 11 to 65 months; this mass shifting is necessarily absent from the Royston & Wright (1998) fit. The log pseudo-marginal likelihood under the linear-dependent tail-free model was -121, significantly better than one obtained from fitting a normal errors model on the log-scale, -143, the exponential normal model of Kapitula & Bedrick (2005), -136, the Dirichlet process mixture model, -143, and the linear-dependent Dirichlet process model, -139.

5. DISCUSSION

Recently two approaches to variable selection in density regression models have been proposed for finite mixture models. Villani et al. (2009), favouring a few components with complex structure approach, tailor Smith & Kohn's (1996) method for use within Gamerman's (1997) posterior updating scheme in a hierarchical mixture of experts model where component locations, precisions and weights all vary smoothly with predictors according to a spline representation. Alternatively, Chung & Dunson (2009) consider a many-but-simple approach to the hierarchical mixture of experts model, using a modification of Smith & Kohn's (1996) idea to encourage predictors to be included or removed from all components simultaneously. These approaches can be readily incorporated into the current proposal. A simpler, similar approach due to Kuo & Mallick (1998) uses Bernoulli predictor inclusion variables, but does not use a hierarchical prior on the $\beta_{\epsilon 0}$ as do the other approaches.

Another extension of our approach is the analysis of dependent multivariate distributions. Although multivariate outcomes can be handled in theory, the number of parameters needed grows exponentially with the dimension. Conditional density approaches that accommodate both continuous and categorical predictors based on dependent mixtures of Gaussians (De Iorio et al., 2004, 2009; Chung & Dunson, 2009; Villani et al., 2009) have a clear advantage here. We have

developed a marginalized tail-free process that enables multidimensional outcomes, but necessarily has a quite different structure from the dependent tail-free processes proposed here. Our experience with the linear-dependent tail-free prior for univariate conditional densities is that it is considerably faster in providing similar inferences than convolutions of dependent stick-breaking processes (Müller et al., 1996; Dunson et al., 2007; Griffin & Steel, 2010).

Acknowledgement

The work of the first author was supported by a grant of the Fondo Nacional de Desarrollo Científico y Tecnológico, from the Chilean Government. The second author was supported by grants from the U.S. National Institutes of Health and National Science Foundation.

Supplementary material

Supplementary material available at *Biometrika* online includes proofs and further examples.

References

Christensen, R. (2002). *Plane Answers to Complex Questions: The Theory of Linear Models*, 3rd ed. New York: Springer.

Chung, Y. & Dunson, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *J. Am. Statist. Assoc.* **104**, 1646–60.

De Iorio, M., Johnson, W. O., Müller, P. & Rosner, G. L. (2009). Bayesian nonparametric non-proportional hazards survival modelling. *Biometrics* **65**, 762–71.

De Iorio, M., Müller, P., Rosner, G. L. & MacEachern, S. N. (2004). An ANOVA model for dependent random measures. *J. Am. Statist. Assoc.* **99**, 205–15.

Dubins, L. E. & Freedman, D. A. (1967). Random distribution functions. In *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, Statistical Laboratory, University of California, Berkeley, USA, between June 21 and July 18, 1965, and between December 27, 1965, and January 7, 1966, Vol 2, pp. 183–214.

Dunson, D. B., Pillai, N. & Park, J. H. (2007). Bayesian density regression. *J. R. Statist. Soc.* B **69**, 163–183.

Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Am. Statist. Assoc.* **90**, 577–88.

Fabius, J. (1964). Asymptotic behavior of Bayes' estimates. *Ann. Math. Statist.* **35**, 846–56.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–30.

Ferguson, T. S. (1974). Prior distribution on the spaces of probability measures. *Ann. Statist.* **2**, 615–29.

Freedman, D. (1963). On the asymptotic distribution of Bayes' estimates in the discrete case. *Ann. Math. Statist.* **34**, 1386–403.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statist. Comput.* **7**, 57–68.

Geisser, S. & Eddy, W. (1979). A predictive approach to model selection. *J. Am. Statist. Assoc.* **74**, 153–60.

Gelfand, A. E. & Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *J. Comput. Graph. Statist.* **11**, 289–304.

Gelfand, A. E., Kottas, A. MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J. Am. Statist. Assoc.* **100**, 1021–35.

Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y. S. (2008). A default prior distribution for logistic and other regression models. *Ann. Appl. Statist.* **2**, 1360–83.

Ghosh, J. K. & Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. New York, USA: Springer.

Griffin, J. E. & Steel, M. F. J. (2006). Order-based dependent Dirichlet processes. *J. Am. Statist. Assoc.* **101**, 179–94.

Griffin, J. E. & Steel, M. F. J. (2010). Bayesian nonparametric modelling with the Dirichlet process regression smoother. *Statist. Sinica* **20**, 1507–27.

Hanson, T. (2006). Inference for mixtures of finite Polya tree models. *J. Am. Statist. Assoc.* **101**, 1548–65.

Holmes, C. C. & Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Anal.* **1**, 145–68.

Isaacs, D., Altman, D. G., Tidmarsh, C. E., Valman, H. B. & Webster, A. D. B. (1983). Serum immunoglobulin concentrations in preschool children measured by laser nephelometry: reference ranges for IgG, IgA, IgM. *J. Clin. Pathol.* **36**, 1193–1196.

JARA, A. (2007). Applied Bayesian non- and semi-parametric inference using DPpackage. *Rnews* **7**, 17–26.
JARA, A., LESAFFRE, E., DE IORIO, M. & QUINTANA, F. A. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *Ann. Appl. Statist.* **4**, 2126–49.
JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–76.
KAPITULA, L. R. & BEDRICK, E. J. (2005). Diagnostics for the exponential normal growth curve model. *Statist. Med.* **24**, 95–108.
KRAFT, C. M. (1964). A class of distribution function processes which have derivatives. *J. Appl. Probabil.* **1**, 385–88.
KUO, L. & MALLICK, B. (1998). Variable selection for regression models. *Sankya* B **60**, 65–81.
MACEACHERN, S. N. (1999). Dependent nonparametric processes. In *ASA Proc. Section on Bayesian Statist. Sci.* Alexandria, VA. American Statistical Association.
METIVIER, M. (1971). Sur la construction de mesures aléatoires presque sûrement absolument continues par rapport à une mesure donnée. *Z. Wahrs. verw. Geb.* **20**, 332–4.
MULIERE, P. & TARDELLA, L. (1998). Approximating distributions of random functionals of Ferguson–Dirichlet priors. *Can. J. Statist.* **26**, 283–97.
MÜLLER, P., ERKANLI, A. & WEST, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* **83**, 67–79.
ROYSTON, P. & WRIGHT, E. M. (1998). A method for estimation age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. *J. R. Statist. Soc.* A **161**, 79–101.
SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **2**, 639–50.
SMITH, M. S. & KOHN, R. (1996). Nonparametric regression using Bayesian variable selection. *J. Economet.* **75**, 317–44.
TOKDAR, S. T., ZHU, Y. M. & GHOSH, J. K. (2010). Bayesian density regression with logistic Gaussian process and subspace projection. *Bayesian Anal.* **5**, 1–26.
VILLANI, M., KOHN, R. & GIORDANI, P. (2009). Regression density estimation using smooth adaptive Gaussian mixtures. *J. Economet.* **153**, 155–173.
WANG, Y. & TAYLOR, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *J. Am. Statist. Assoc.* **96**, 895–905.
ZELLNER, A. (1983). Applications of Bayesian analysis in econometrics. *Statistician* **32**, 23–34.

[*Received June* 2009. *Revised November* 2010]