Case Studies in Bayesian Data Science

2: BNP Implementation Details

David Draper

Department of Applied Mathematics and Statistics
University of California, Santa Cruz

draper@ucsc.edu

Short Course (Day 5)
University of Reading (UK)

27 Nov 2015

users.soe.ucsc.edu/∼draper/Reading-2015-Day-5.html

© 2015 David Draper (all rights reserved)

1. Introduction

- A Bayesian nonparametric approach to modeling, say, distribution functions requires priors for spaces of distribution functions
- Informally, it requires stochastic processes with sample paths that are distribution functions defined on an appropriate sample space $\mathcal{X}$ (e.g., $\mathcal{X} = R$, or R^+, or R^d), equipped with a σ-field $\mathcal{B}$ of subsets of $\mathcal{X}$ (e.g., the Borel σ-field for $\mathcal{X} \subseteq R^d$)
- The **Dirichlet process** (DP), anticipated in the work of Freedman (1963) and Fabius (1964), and formally developed by Ferguson (1973, 1974), is the first prior developed for spaces of distribution functions
- The DP is, formally, a probability measure on the space of probability measures (distributions) on $(\mathcal{X}, \mathcal{B})$
- Hence, the DP generates random distributions on $(\mathcal{X}, \mathcal{B})$, and thus, for $\mathcal{X} \subseteq R^d$, equivalently, random distribution functions (cdfs) on $\mathcal{X}$

2. The Dirichlet process

- The DP is characterized by two parameters:
  → Q_0 a specified probability measure on $(\mathcal{X}, \mathcal{B})$ (equivalently, G_0 a specified distribution function on $\mathcal{X}$)
  → α a positive scalar parameter

- **DEFINITION** (Ferguson, 1973): The DP generates random probability measures (random distributions) Q on $(\mathcal{X}, \mathcal{B})$ such that for any finite measurable partition B_1,...,B_k of $\mathcal{X}$,

$$(Q(B_1), ..., Q(B_k)) \sim \text{Dirichlet}(\alpha Q_0(B_1), ..., \alpha Q_0(B_k))$$

→ here, $Q(B_i)$ (a random variable) and $Q_0(B_i)$ (a constant) denote the probability of set B_i under Q and Q_0, respectively
→ also, the B_i define a measurable partition if $B_i \in \mathcal{B}$, they are pairwise disjoint, and their union is $\mathcal{X}$

The Dirichlet process

Recall the definition of the Dirichlet distribution

- Start with independent rvs $Z_j \sim \text{gamma}(a_j, 1)$, $j = 1, ..., k$ (with $a_j > 0$)
- Define $Y_j = Z_j / (\sum_{\ell=1}^{k} Z_\ell)$, for $j = 1, ..., k$
- Then $(Y_1, ..., Y_k) \sim \text{Dirichlet}(a_1, ..., a_k)$ (distribution singular w.r.t. Lebesgue measure on R^k, since $\sum_{j=1}^{k} Y_j = 1$)
- $(Y_1, ..., Y_{k-1})$ has density $C(1 - \sum_{j=1}^{k-1} y_j)^{a_k - 1} \prod_{j=1}^{k-1} y_j^{a_j - 1}$, where $C = \Gamma(\sum_{j=1}^{k} a_j) / \{\prod_{j=1}^{k} \Gamma(a_j)\}$
- Moments: $\text{E}(Y_j) = a_j / \sum_{\ell=1}^{k} a_\ell$, $\text{E}(Y_j^2) = a_j(a_j + 1) / \{\sum_{\ell=1}^{k} a_\ell (1 + \sum_{\ell=1}^{k} a_\ell)\}$, and, for $i \neq j$, $\text{E}(Y_i Y_j) = a_i a_j / \{\sum_{\ell=1}^{k} a_\ell (1 + \sum_{\ell=1}^{k} a_\ell)\}$
- Note that for $k = 2$, $\text{Dirichlet}(a_1, a_2) = \text{Beta}(a_1, a_2)$

The Dirichlet process

- For any measurable subset B of $\mathcal{X}$, we have from the definition that $Q(B) \sim \text{Beta}(\alpha Q_0(B), \alpha Q_0(B^c))$, and thus

$$\text{E}(Q(B)) = Q_0(B)$$

and

$$\text{Var}(Q(B)) = \frac{Q_0(B)\{1 - Q_0(B)\}}{\alpha + 1},$$

- Q_0 plays the role of the *center* of the DP (base probability measure, or base distribution)
- α can be viewed as a precision parameter: for large α there is small variability in DP realizations; the larger α is the *closer* we expect a realization Q from the process to be to Q_0
- See Ferguson (1973) for the role of Q_0 on more technical properties of the DP (e.g., Ferguson shows that the support of the DP contains all probability measures on $(\mathcal{X}, \mathcal{B})$ that are absolutely continuous w.r.t. Q_0)

The Dirichlet process

- Analogously, for the random distribution function G on $\mathcal{X}$ generated from a DP with parameters α and G_0, a specified distribution function on $\mathcal{X}$
- For example, with $\mathcal{X} = R$, $B = (-\infty, x]$, $x \in R$, and $Q(B) = G(x)$,

$$G(x) \sim \text{Beta}(\alpha G_0(x), \alpha\{1 - G_0(x)\})$$

and thus

$$\text{E}(G(x)) = G_0(x)$$

and

$$\text{Var}(G(x)) = \frac{G_0(x)\{1 - G_0(x)\}}{\alpha + 1}$$

- **notation**: G will denote either the random probability measure or the random distribution function (depending on the context)
  $G \sim \text{DP}(\alpha G_0)$ will indicate that a DP prior is placed on G

The Dirichlet process

- The definition can be used to simulate sample paths (which are distribution functions) from the DP — this is convenient when $\mathcal{X} \subseteq R$
- Consider any grid of points $x_1 < x_2 < ... < x_k$ in $\mathcal{X}$
- Then, the random vector $(G(x_1), G(x_2) - G(x_1), ..., G(x_k) - G(x_{k-1}), 1 - G(x_k))$ follows a Dirichlet distribution with parameters $(\alpha G_0(x_1), \alpha(G_0(x_2) - G_0(x_1)), ..., \alpha(G_0(x_k) - G_0(x_{k-1})), \alpha(1 - G_0(x_k)))$
- Hence, if $(u_1, u_2, ..., u_k)$ is a draw from this Dirichlet distribution, then $(u_1, ..., \sum_{j=1}^{i} u_j, ..., \sum_{j=1}^{k} u_j)$ is a draw from the distribution of $(G(x_1), ..., G(x_i), ..., G(x_k))$
- Example (Figure 1): $\mathcal{X} = (0, 1)$, $G_0(x) = x$, $x \in (0, 1)$ (Unif(0, 1) base distribution)

The Dirichlet process

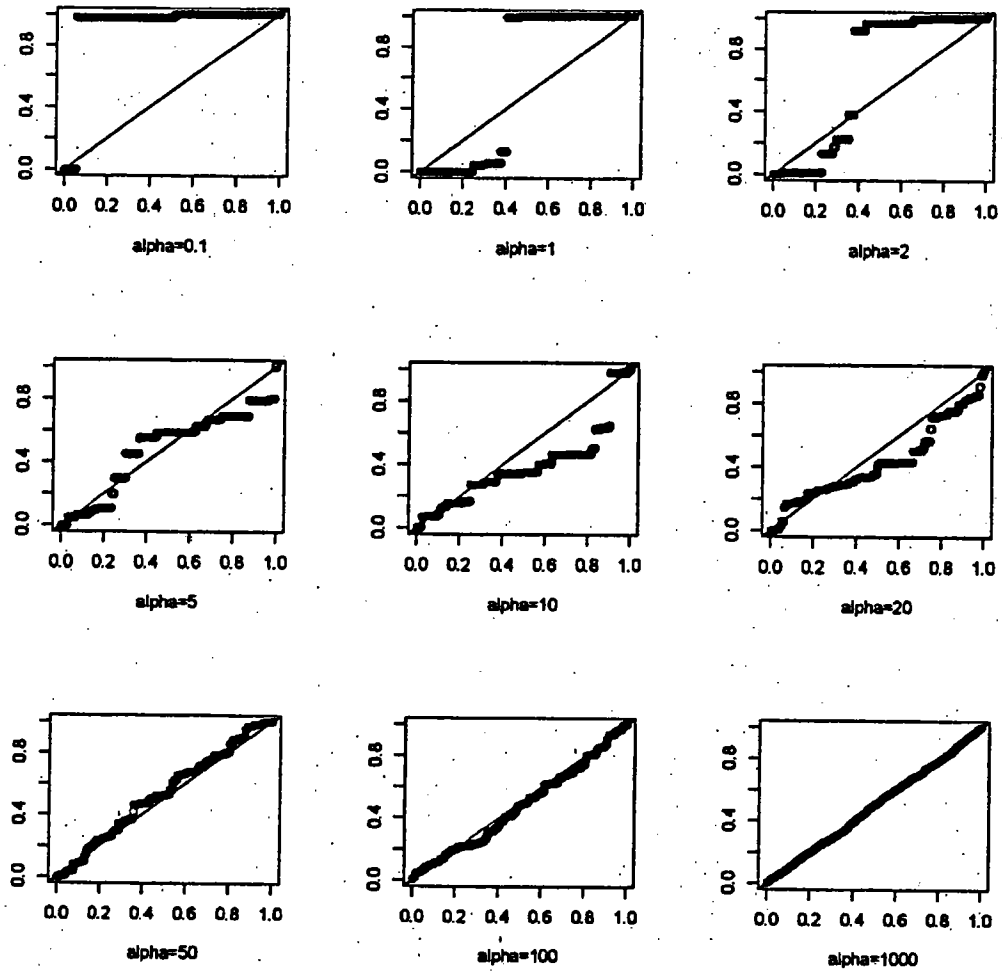

Figure 1: Cdf sample paths from a DP($\alpha G_0 = \text{Unif}(0,1)$) prior, for different values of α. The solid line denotes the cdf of G_0.

The Dirichlet process

- **Constructive definition of the DP**

  (Sethuraman and Tiwari, 1982; Sethuraman, 1994)

  $\rightarrow$ let $\{z_r : r = 1, 2, ...\}$ and $\{\vartheta_\ell : \ell = 1, 2, ...\}$ be independent sequences of i.i.d. random variables:

  * $z_r \sim \text{Beta}(1, \alpha)$, $r = 1, 2, ...$

  * $\vartheta_\ell \sim G_0$, $\ell = 1, 2, ...$

  $\rightarrow$ define $\omega_1 = z_1$, $\omega_\ell = z_\ell \prod_{r=1}^{\ell-1}(1 - z_r)$, $\ell = 2, 3, ...$ (thus, $\sum_{\ell=1}^{\infty} \omega_\ell = 1$)

  $\rightarrow$ then, a realization G from $\text{DP}(\alpha G_0)$ is (almost surely) of the form

$$G = \sum_{\ell=1}^{\infty} \omega_\ell \delta_{\vartheta_\ell}$$

  (here, $\delta_z(\cdot)$ denotes a point mass at z)

- Hence, the DP generates distributions that have an (almost sure) representation as countable mixtures of point masses — the locations ϑ_ℓ are i.i.d. draws from the base distribution — their associated weights ω_ℓ are defined using the *stick-breaking* construction above

The Dirichlet process

- Based on its constructive definition, it becomes evident that the DP generates (almost surely) discrete distributions on $\mathcal{X}$ (this result was proved, using different approaches, by Ferguson, 1973, and Blackwell, 1973)
- The DP constructive definition yields another method to simulate from DP priors — in fact, it provides (up to a truncation approximation) the entire distribution G, not just cdf sample paths — for example, a simple approximation is $G_J = \sum_{j=1}^{J} p_j \delta_{\vartheta_j}$, where
  $\rightarrow p_j = \omega_j$, $j = 1, ..., J-1$, and $p_J = 1 - \sum_{j=1}^{J-1} \omega_j = \prod_{r=1}^{J-1}(1 - z_r)$
  $\rightarrow$ to choose J, note, for example, that

$$\mathrm{E}(\sum_{j=1}^{J} \omega_j) = \mathrm{E}(1 - \prod_{r=1}^{J}(1 - z_r)) = 1 - \prod_{r=1}^{J} \mathrm{E}(1 - z_r) = 1 - \prod_{r=1}^{J} \frac{\alpha}{\alpha+1} = 1 - (\frac{\alpha}{\alpha+1})^J$$

  $\rightarrow$ hence, we can choose J such that $(\alpha/(\alpha+1))^J = \varepsilon$, for small ε

- Moreover, the constructive definition of the DP has motivated several of its extensions in the recent literature, including the ϵ-DP (Muliere and Tardella, 1998), generalized DPs (Hjort, 2000), and dependent DPs (MacEachern, 1999, 2000; De Iorio et al. 2004; Gelfand et al., 2004)

The Dirichlet process

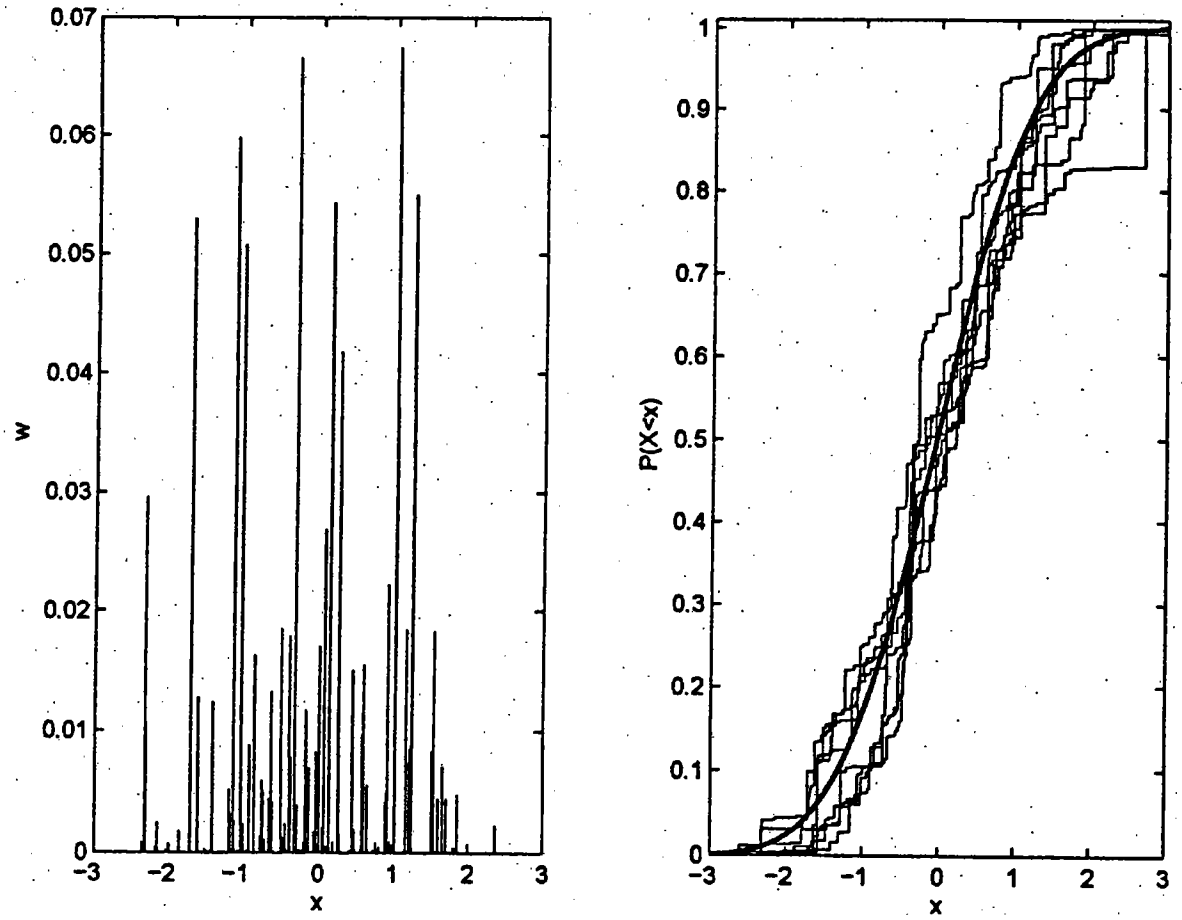

Figure 2: Illustration for a DP with $G_0 = N(0, 1)$ and $\alpha = 20$. In the left panel, the spiked lines are located at 1000 sampled values of x drawn from $N(0, 1)$ with heights given by the weights, ω_ℓ, calculated using the stick-breaking algorithm (a truncated version so that the weights sum to 1). These spikes are then summed from left to right to generate one cdf sample path from the DP. The right panel shows 8 such sample paths indicated by the lighter jagged lines. The heavy smooth line indicates the $N(0, 1)$ cdf.

The Dirichlet process

- **Pólya urn characterization of the DP**
  (Blackwell and MacQueen, 1973)
  → if, for $i = 1, ..., n$, $x_i \mid G$ are i.i.d. from G, and $G \sim \mathrm{DP}(\alpha G_0)$, then, marginalizing G over its DP prior, the induced joint distribution for the x_i is given by

$$p(x_1, ..., x_n) = G_0(x_1) \prod_{i=2}^{n} \left\{ \frac{\alpha}{\alpha + i - 1} G_0(x_i) + \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{x_j}(x_i) \right\}$$

  → that is, the sequence of the x_i follows a generalized Pólya urn scheme such that
  * $x_1 \sim G_0$, and
  * for any $i = 2, ..., n$, $x_i \mid x_1, ..., x_{i-1}$ follows the mixed distribution that places point mass $(\alpha + i - 1)^{-1}$ at x_j, $j = 1, ..., i - 1$, and continuous mass $\alpha(\alpha + i - 1)^{-1}$ on G_0

The Dirichlet process

- **Prior to posterior updating with DP priors**
  (Ferguson, 1973)
  $\rightarrow$ let G denote the random distribution function for the following results
  $\rightarrow$ if the observations $y_i \mid G$ are i.i.d. from G, $i = 1, ..., n$, and $G \sim \text{DP}(\alpha G_0)$, then the posterior distribution of G is a $\text{DP}(\tilde{\alpha}\tilde{G}_0)$, with $\tilde{\alpha} = \alpha + n$, and

$$\tilde{G}_0(t) = \frac{\alpha}{\alpha + n} G_0(t) + \frac{1}{\alpha + n} \sum_{i=1}^{n} 1_{[y_i, \infty)}(t)$$

- Hence, the DP is a *conjugate* prior — all the results and properties developed for DPs can be used directly for the posterior distribution of G

The Dirichlet process

- For example, the posterior point estimate for $G(t)$

$$\mathrm{E}(G(t) \mid y_1, ..., y_n) = \frac{\alpha}{\alpha + n} G_0(t) + \frac{n}{\alpha + n} G_n(t)$$

where $G_n(t) = n^{-1} \sum_{i=1}^{n} 1_{[y_i, \infty)}(t)$ is the empirical distribution function of the data (the standard classical nonparametric estimator)
$\rightarrow$ for small α relative to n, little weight is placed on the prior guess G_0
$\rightarrow$ for large α relative to n, little weight is placed on the data
$\rightarrow$ α can be viewed as a measure of faith in the prior guess G_0 measured in units of number of observations (thus, $\alpha = 1$ indicates strength of belief in G_0 worth one observation)
$\rightarrow$ note that, $\lim_{\alpha \to 0} \mathrm{E}(G(t) \mid y_1, ..., y_n) = G_n(t)$

- Refer to the Appendix ("References for Dirichlet processes") for references on theoretical and methodological work with DPs, including approaches for other types of inference based on DP priors

1. Introduction

- Mixture models arise naturally as flexible alternatives to standard parametric families
- Continuous mixture models (e.g., t, Beta-binomial and Poisson-gamma models) typically achieve increased heterogeneity but are still limited to unimodality and usually symmetry
- Finite mixture distributions (Titterington et al., 1985; McLachlan and Peel, 2000) provide more flexible modeling – now feasible to implement due to advances in simulation-based model fitting (e.g., Diebolt and Robert, 1994; Richardson and Green, 1997; Stephens, 2000)
- Though it may appear paradoxical, rather than handling the very large number of parameters resulting from finite mixture models with a large number of mixands, it may be easier to work with an infinite dimensional specification by assuming a random mixing distribution which is not restricted to a specified parametric family

2. Model details, examples, hierarchical formulation

- The Dirichlet process (DP) is the most widely used prior for the random mixing distribution (Antoniak, 1974; Lo, 1984; Ferguson, 1983)
  Dirichlet process mixture model

$$F(\cdot; G) = \int K(\cdot; \theta) \mathrm{d}G(\theta), \quad G \sim \mathrm{DP}(\alpha G_0)$$

  with $K(\cdot; \theta)$ a parametric family of distribution functions indexed by θ

- Corresponding mixture density (or probability mass) function,

$$f(\cdot; G) = \int k(\cdot; \theta) \mathrm{d}G(\theta)$$

  where $k(\cdot; \theta)$ is the density (or probability mass) function of $K(\cdot; \theta)$

- Because G is random, the distribution function $F(\cdot; G)$ and the density function $f(\cdot; G)$ are random (Bayesian nonparametric mixture models)

Model details, examples, hierarchical formulation

- Contrary to DP prior models, the DP mixture $F(\cdot; G)$ can model both discrete distributions (e.g., $K(\cdot; \theta)$ might be Poisson or binomial) and continuous distributions, either univariate ($K(\cdot; \theta)$ can be, e.g., normal, gamma, or uniform) or multivariate (with $K(\cdot; \theta)$, say, multivariate normal)

- Several useful results for general mixtures of parametric families, e.g.,
  $\rightarrow$ (discrete) normal location-scale mixtures, $\sum_{j=1}^{M} w_j \mathrm{N}(\cdot \mid \mu_j, \sigma_j^2)$, can approximate arbitrarily well any density on the real line (Lo, 1984; Ferguson, 1983; Escobar and West, 1995) — analogously, for densities on R^d (West et al., 1994; Müller et al., 1996)
  $\rightarrow$ for any non-increasing density $f(t)$ on the positive real line there exists a distribution function G such that f can be represented as a scale mixture of uniform densities, i.e., $f(t) = \int \theta^{-1} 1_{[0,\theta)}(t) \mathrm{d}G(\theta)$ — the result yields flexible DP mixture models for symmetric unimodal densities (Brunner and Lo, 1989; Brunner, 1995) as well as general unimodal densities (Brunner, 1992; Lavine and Mockus, 1995; Kottas and Gelfand, 2001; Kottas and Krnjajić, 2005)

Model details, examples, hierarchical formulation

- Typically, semiparametric DP mixtures are employed

$$\begin{aligned} y_i \mid G, \phi &\overset{i.i.d.}{\sim} f(\cdot; G, \phi) = \int k(\cdot; \theta, \phi) \mathrm{d}G(\theta), \quad i = 1, ..., n \\ G &\sim \mathrm{DP}(\alpha G_0) \end{aligned}$$

with a parametric prior $p(\phi)$ placed on ϕ (and, perhaps, hyperpriors for α and/or the parameters ψ of $G_0 \equiv G_0(\cdot \mid \psi)$)

- **Hierarchical formulation** for DP mixture models: introduce latent mixing parameter θ_i associated with y_i

$$\begin{aligned} y_i \mid \theta_i, \phi &\overset{ind.}{\sim} k(\cdot; \theta_i, \phi), \quad i = 1, ..., n \\ \theta_i \mid G &\overset{i.i.d.}{\sim} G, \quad i = 1, ..., n \\ G \mid \alpha, \psi &\sim \mathrm{DP}(\alpha G_0), \; G_0 = G_0(\cdot \mid \psi) \\ \phi, \alpha, \psi &\sim p(\phi)p(\alpha)p(\psi) \end{aligned}$$

Model details, examples, hierarchical formulation

- In the context of DP mixtures, the (almost sure) discreteness of realizations G from the DP(αG_0) prior is an asset — it allows ties in the θ_i, and thus makes DP mixture models appealing for applications where clustering is anticipated (e.g., density estimation, classification, regression)
- Using the constructive definition of the DP, $G = \sum_{\ell=1}^{\infty} \omega_\ell \delta_{\vartheta_\ell}$ (see session 4), the prior probability model $f(\cdot; G, \phi)$ admits an (almost sure) representation as a countable mixture of parametric densities,

$$f(\cdot; G, \phi) = \sum_{\ell=1}^{\infty} \omega_\ell k(\cdot; \vartheta_\ell, \phi)$$

→ *weights*: $\omega_1 = z_1$, $\omega_\ell = z_\ell \prod_{r=1}^{\ell-1}(1 - z_r)$, $\ell \geq 2$, with z_r i.i.d. Beta(1,α)
→ *locations*: ϑ_ℓ i.i.d. G_0
(and the sequences $\{z_r, r = 1,2,...\}$ and $\{\vartheta_\ell, \ell = 1,2,...\}$ are independent)

- This formulation has motivated study of certain variants of the DP mixture model (e.g., Ishwaran and Zarepour, 2000; Ishwaran and James, 2001; Quintana and Iglesias, 2003)

3. Prior specification

- Taking expectation over G with respect to its DP prior $\mathrm{DP}(\alpha G_0)$, $\mathrm{E}(F(\cdot; G, \phi)) = F(\cdot; G_0, \phi)$, and $\mathrm{E}(f(\cdot; G, \phi)) = f(\cdot; G_0, \phi)$ — these expressions facilitate prior specification for the parameters ψ of $G_0(\cdot \mid \psi)$
- Recall that for the $\mathrm{DP}(\alpha G_0)$ prior, α controls how *close* a realization G is to G_0 — in the DP mixture model, α controls the distribution of the number of distinct elements n^* of the vector $\theta = (\theta_1, ..., \theta_n)$, and hence the number of distinct components of the mixture (Antoniak, 1974; Escobar and West, 1995; Liu, 1996)
- In particular,

$$P(n^* = m \mid \alpha) = c_n(m) n! \alpha^m \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)}, \quad m = 1, ..., n,$$

where the factors $c_n(m) = P(n^* = m \mid \alpha = 1)$ can be computed using certain recurrence formulas (Escobar and West, 1995)

Prior specification

- Moreover, for moderately large n,

$$\mathrm{E}(n^*|\alpha) \approx \alpha \log\left(\frac{\alpha+n}{\alpha}\right)$$

and

$$\mathrm{Var}(n^*|\alpha) \approx \alpha\left\{\log\left(\frac{\alpha+n}{\alpha}\right) - 1\right\}$$

- If α is assigned a prior $p(\alpha)$, $P(n^* = m) = \int P(n^* = m \mid \alpha)p(\alpha)\mathrm{d}\alpha$,
  $\rightarrow$ in particular, if the prior for α is gamma(a_α, b_α) (with $\mathrm{E}(\alpha) = a_\alpha/b_\alpha$, and $\mathrm{Var}(\alpha) = a_\alpha/b_\alpha^2$), using an additional Taylor series approximation,

$$\mathrm{E}(n^*) \approx \frac{a_\alpha}{b_\alpha}\log\left(1+\frac{nb_\alpha}{a_\alpha}\right)$$

and

$$\mathrm{Var}(n^*) \approx \frac{a_\alpha}{b_\alpha}\log\left(1+\frac{nb_\alpha}{a_\alpha}\right) - \frac{a_\alpha}{b_\alpha} + \left\{\log\left(1+\frac{nb_\alpha}{a_\alpha}\right) - \frac{nb_\alpha}{a_\alpha+nb_\alpha}\right\}^2 \frac{a_\alpha}{b_\alpha^2}$$

Prior specification

- Two *limiting* special cases of the DP mixture model

- One cluster, when $\alpha \rightarrow 0^+$

$$\begin{aligned} y_i \mid \theta, \phi &\overset{ind.}{\sim} k(\cdot; \theta, \phi), \quad i = 1, ..., n \\ \theta \mid \psi &\sim G_0(\cdot \mid \psi) \\ \phi, \psi &\sim p(\phi)p(\psi) \end{aligned}$$

- n clusters (one associated with each observation), when $\alpha \rightarrow \infty$

$$\begin{aligned} y_i \mid \theta_i, \phi &\overset{ind.}{\sim} k(\cdot; \theta_i, \phi), \quad i = 1, ..., n \\ \theta_i \mid \psi &\overset{i.i.d.}{\sim} G_0(\cdot \mid \psi), \quad i = 1, ..., n \\ \phi, \psi &\sim p(\phi)p(\psi) \end{aligned}$$

4. Methods for posterior inference

- Data $= \{y_i, i = 1, ..., n\}$, i.i.d., conditionally on G and ϕ, from $f(\cdot; G, \phi)$ (if the model includes a regression component, the data also include the covariate vectors $\boldsymbol{x}_i$, and, in such cases, ϕ, typically, includes the vector of regression coefficients)

- Interest in inference for the latent mixing parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)$, for ϕ (and the hyperparameters α, ψ), for $f(y_0; G, \phi)$, and, in general, for functionals $H(F(\cdot; G, \phi))$ of the random mixture $F(\cdot; G, \phi)$ (e.g., cdf function, hazard function, mean and variance functionals, percentile functionals)

- Full and exact inference, given the data, for all these random quantities is based on the joint posterior of the DP mixture model

$$p(G, \phi, \boldsymbol{\theta}, \alpha, \psi \mid \text{data})$$

Methods for posterior inference

- Key result: representation of the joint posterior (Antoniak, 1974)

$$p(G, \phi, \theta, \alpha, \psi \mid \text{data}) = p(G \mid \theta, \alpha, \psi) p(\theta, \phi, \alpha, \psi \mid \text{data})$$

where

$\rightarrow p(\theta, \phi, \alpha, \psi \mid \text{data})$ is the marginal posterior for the finite-dimensional portion of the full *parameter vector* $(G, \phi, \theta, \alpha, \psi)$

$\rightarrow G \mid \theta, \alpha, \psi \sim \text{DP}(\tilde{\alpha}\tilde{G}_0)$, where $\tilde{\alpha} = \alpha + n$, and

$$\tilde{G}_0(\cdot) = \frac{\alpha}{\alpha + n} G_0(\cdot \mid \psi) + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{\theta_i}$$

(hence, $\tilde{G}_0(t) = \frac{\alpha}{\alpha+n} G_0(t \mid \psi) + \frac{1}{\alpha+n} \sum_{i=1}^{n} 1_{[\theta_i, \infty)}(t)$ for the distribution function)

- Sampling from DPs is possible using one of its definitions — thus, we can obtain full posterior inference under DP mixture models if we can sample from the marginal posterior $p(\theta, \phi, \alpha, \psi \mid \text{data})$

Methods for posterior inference

- The marginal posterior $p(\theta, \phi, \alpha, \psi \mid \text{data})$ corresponds to the marginalized version of the DP mixture model, obtained after integrating G over its DP prior (Blackwell and MacQueen, 1973),

$$\begin{aligned} y_i \mid \theta_i, \phi &\overset{ind.}{\sim} k(\cdot; \theta_i, \phi), \quad i = 1, ..., n \\ \theta = (\theta_1, ..., \theta_n) \mid \alpha, \psi &\sim p(\theta \mid \alpha, \psi) \\ \phi, \alpha, \psi &\sim p(\phi)p(\alpha)p(\psi) \end{aligned}$$

- The induced prior distribution $p(\theta \mid \alpha, \psi)$ for the mixing parameters θ_i can be developed by exploiting the Pólya urn characterization of the DP,

$$p(\theta \mid \alpha, \psi) = G_0(\theta_1 \mid \psi) \prod_{i=2}^{n} \left\{ \frac{\alpha}{\alpha + i - 1} G_0(\theta_i \mid \psi) + \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{\theta_j}(\theta_i) \right\}$$

$\rightarrow$ for increasing sample sizes, the joint prior $p(\theta \mid \alpha, \psi)$ gets increasingly complex to work with

Methods for posterior inference

- Therefore, the marginal posterior

$$p(\theta, \phi, \alpha, \psi \mid \text{data}) \propto p(\theta \mid \alpha, \psi) p(\phi) p(\alpha) p(\psi) \prod_{i=1}^{n} k(y_i; \theta_i, \phi)$$

  is difficult to work with — even point estimates practically impossible to compute for moderate to large sample sizes

- Early work for posterior inference:
  → some results for certain problems in density estimation, i.e., expressions for Bayes point estimates of $f(y_0; G)$ (Lo, 1984; Brunner and Lo, 1989)
  → approximations for special cases, e.g., for binomial DP mixtures (Berry and Christensen, 1979)
  → Monte Carlo integration algorithms to obtain point estimates for the θ_i (Ferguson, 1983; Kuo, 1986a,b)

Methods for posterior inference

Simulation-based model fitting

- Note that, although the joint prior $p(\theta \mid \alpha, \psi)$ has an awkward expression for samples of realistic size n, the prior full conditionals have convenient expressions

$$p(\theta_i \mid \{\theta_j : j \neq i\}, \alpha, \psi) = \frac{\alpha}{\alpha + n - 1} G_0(\theta_i \mid \psi) + \frac{1}{\alpha + n - 1} \sum_{j=1}^{n-1} \delta_{\theta_j}(\theta_i)$$

- **Key idea:** (Escobar, 1988; 1994) setup a Markov chain to explore the posterior $p(\theta, \phi, \alpha, \psi \mid \text{data})$ by simulating only from posterior full conditional distributions, which arise by combining the likelihood terms with the corresponding prior full conditionals
  (in fact, Escobar's algorithm is essentially a Gibbs sampler developed for a specific class of models!)
- Several other Markov chain Monte Carlo (MCMC) methods following the work of Escobar (e.g., West et al., 1994; Escobar and West, 1995; Bush and MacEachern, 1996; MacEachern and Müller, 1998; Neal, 2000; Ishwaran and James, 2001)

Methods for posterior inference

- A key property for the implementation of the Gibbs sampler is the discreteness of G, which induces a clustering of the θ_i
  $\rightarrow n^*$: number of clusters (distinct elements) in the vector $(\theta_1, ..., \theta_n)$
  $\rightarrow \theta_j^*$, $j = 1,...,n^*$: the distinct θ_i
  $\rightarrow w = (w_1, ..., w_n)$: vector of configuration indicators, defined by $w_i = j$ if and only if $\theta_i = \theta_j^*$, $i = 1,...,n$
  $\rightarrow n_j$: size of j-th cluster, i.e., $n_j = |\{i : w_i = j\}|$, $j = 1,...,n^*$

- Evidently, $(n^*, w, (\theta_1^*, ..., \theta_{n^*}^*))$ yields an equivalent representation for $(\theta_1, ..., \theta_n)$

- Standard Gibbs sampler to draw from $p(\theta, \phi, \alpha, \psi \mid \text{data})$ (Escobar, 1994; Escobar and West, 1995) is based on the following full conditionals:
  (a) $p(\theta_i \mid \{\theta_{i'} : i' \neq i\}, \alpha, \psi, \phi, \text{data})$, for $i = 1,...,n$
  (b) $p(\alpha \mid n^*, \text{data})$ and $p(\psi \mid \{\theta_j^*, j = 1, ..., n^*\}, n^*)$
  (c) $p(\phi \mid \{\theta_i : i = 1, ..., n\}, \text{data})$
  (the expressions include conditioning only on the relevant variables, exploiting the conditional independence structure of the model and properties of the DP)

Methods for posterior inference

- (a) For each $i = 1, ..., n$

$$p(\theta_i \mid \{\theta_{i'} : i' \neq i\}, \alpha, \psi, \phi, \text{data}) = \frac{q_0 h(\theta_i \mid \psi, \phi, y_i) + \sum_{j=1}^{n^{*-}} n_j^- q_j \delta_{\theta_j^{*-}}(\theta_i)}{q_0 + \sum_{j=1}^{n^{*-}} n_j^- q_j}$$

$\rightarrow q_j = k(y_i; \theta_j^{*-}, \phi)$
$\rightarrow q_0 = \alpha \int k(y_i; \theta, \phi) g_0(\theta \mid \psi) \mathrm{d}\theta$
$\rightarrow h(\theta_i \mid \psi, \phi, y_i) \propto k(y_i; \theta_i, \phi) g_0(\theta_i \mid \psi)$
$\rightarrow g_0$ is the density of G_0
$\rightarrow$ superscript "$-$" denotes all relevant quantities when θ_i is removed from the vector $(\theta_1, ..., \theta_n)$, e.g., n^{*-} is the number of clusters in $\{\theta_{i'} : i' \neq i\}$

- Note that updating θ_i implicitly updates w_i, $i = 1,...,n$ — before updating θ_{i+1}, we redefine n^*, θ_j^*, $j = 1,...,n^*$, w_i, $i = 1,...,n$, and n_j, $j = 1,...,n^*$

Methods for posterior inference

- (b) Although the posterior full conditional for α is not of a standard form, an augmentation method facilitates sampling provided the prior for α is a gamma distribution (say, with mean a_α/b_α) (Escobar and West, 1995),

$$\begin{aligned}[\alpha \mid n^*, \text{data}] &\propto p(\alpha)\alpha^{n^*}\tfrac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \\ &\propto p(\alpha)\alpha^{n^*-1}(\alpha+n)\text{Beta}(\alpha+1,n) \\ &\propto p(\alpha)\alpha^{n^*-1}(\alpha+n)\textstyle\int_0^1 x^\alpha(1-x)^{n-1}dx\end{aligned}$$

$\rightarrow$ introduce an auxiliary variable η such that

$$p(\alpha,\eta \mid n^*, \text{data}) \propto p(\alpha)\alpha^{n^*-1}(\alpha+n)\eta^\alpha(1-\eta)^{n-1},$$

$\rightarrow$ extend the Gibbs sampler to draw from $p(\eta \mid \alpha, \text{data}) = \text{Beta}(\alpha+1, n)$, and $p(\alpha \mid \eta, n^*, \text{data})$, which is given by the two-component mixture

$$p\text{gamma}(a_\alpha+n^*, b_\alpha-\log(\eta)) + (1-p)\text{gamma}(a_\alpha+n^*-1, b_\alpha-\log(\eta))$$

where $p = (a_\alpha+n^*-1)/\{n(b_\alpha-\log(\eta))+a_\alpha+n^*-1\}$

Methods for posterior inference

- Regarding the parameters ψ of G_0,

$$p(\psi \mid \{\theta_j^*, j = 1, ..., n^*\}, n^*) \propto p(\psi) \prod_{j=1}^{n^*} g_0(\theta_j^* \mid \psi)$$

  leading, typically, to standard updates

- (c) The posterior full conditional for ϕ does not involve the nonparametric part of the DP mixture model,

$$p(\phi \mid \{\theta_i : i = 1, ..., n\}, \text{data}) \propto p(\phi) \prod_{i=1}^{n} k(y_i; \theta_i, \phi)$$

Methods for posterior inference

- **Improved Gibbs sampler** (West et al., 1994; Bush and MacEachern, 1996): adds one more step where the cluster locations θ_j^* are resampled at each iteration to improve the mixing of the chain
  $\rightarrow$ at each iteration, once step (a) is completed, we obtain a specific number of clusters n^*, and a specific configuration $\boldsymbol{w} = (w_1, ..., w_n)$
  $\rightarrow$ after the marginalization over G, the prior for the θ_j^*, given the partition $(n^*, \boldsymbol{w})$, is given by

$$p(\theta_j^* : j = 1, ..., n^* \mid n^*, \boldsymbol{w}, \psi) = \prod_{j=1}^{n^*} g_0(\theta_j^* \mid \psi)$$

  i.e., given n^* and $\boldsymbol{w}$, the θ_j^* are i.i.d. from G_0

  $\rightarrow$ hence, for each $j = 1,...,n^*$, the posterior full conditional

$$p(\theta_j^* \mid \boldsymbol{w}, n^*, \psi, \phi, \text{data}) \propto g_0(\theta_j^* \mid \psi) \prod_{\{i:w_i=j\}} k(y_i; \theta_j^*, \phi)$$

Methods for posterior inference

- **Note:** the Gibbs sampler can be difficult or inefficient to implement if
  $\rightarrow$ the integral $\int k(y;\theta,\phi)g_0(\theta \mid \psi)\mathrm{d}\theta$ is not available in closed form (and numerical integration is not feasible or reliable)
  and/or
  $\rightarrow$ random generation from $h(\theta \mid \psi,\phi,y) \propto k(y;\theta,\phi)g_0(\theta \mid \psi)$ is not readily available
- For such cases, alternative MCMC algorithms have been proposed in the literature (e.g., MacEachern and Müller, 1998; Neal, 2000)
- Practically important extensions to data structures that include missing or censored observations are also possible (Kuo and Smith, 1992; Kuo and Mallick, 1997; Kottas, 2003; Kottas and Krnjajić, 2005)
- Alternative fitting techniques that do not involve sampling from a Markov chain have also been suggested (e.g., Liu, 1996; MacEachern et al., 1999; Newton and Zhang, 1999)

Methods for posterior inference

Posterior predictive distributions

- Implementing one of the available MCMC algorithms for DP mixture models, we obtain B posterior samples

$$\{\theta_b = (\theta_{ib} : i = 1, ..., n), \alpha_b, \psi_b, \phi_b\}, \quad b = 1, ..., B,$$

from $p(\theta, \phi, \alpha, \psi \mid \text{data})$, equivalently, posterior samples

$$\left\{n_b^*, w_b, \theta_b^* = (\theta_{jb}^* : j = 1, ..., n_b^*), \alpha_b, \psi_b, \phi_b\right\}, \quad b = 1, ..., B,$$

from $p(n^*, w, \theta^* = (\theta_j^* : j = 1, ..., n^*), \phi, \alpha, \psi \mid \text{data})$

- Bayesian *density estimate* is based on the posterior predictive density $p(y_0 \mid \text{data})$ corresponding to a *new* y_0 with associated mixing parameter θ_0
- Using, again, the Pólya urn structure for the DP,

$$p(\theta_0 \mid n^*, w, \theta^*, \alpha, \psi) = \frac{\alpha}{\alpha + n} G_0(\theta_0 \mid \psi) + \frac{1}{\alpha + n} \sum_{j=1}^{n^*} n_j \delta_{\theta_j^*}(\theta_0)$$

Methods for posterior inference

- The posterior predictive distribution for y_0 is given by

$$\begin{aligned} p(y_0 \mid \text{data}) &= \int p(y_0 \mid n^*, \boldsymbol{w}, \boldsymbol{\theta}^*, \alpha, \psi, \phi) p(n^*, \boldsymbol{w}, \boldsymbol{\theta}^*, \alpha, \psi, \phi \mid \text{data}) \\ &= \int\int p(y_0 \mid \theta_0, \phi) p(\theta_0 \mid n^*, \boldsymbol{w}, \boldsymbol{\theta}^*, \alpha, \psi) p(n^*, \boldsymbol{w}, \boldsymbol{\theta}^*, \alpha, \psi, \phi \mid \text{data}) \end{aligned}$$

  $\rightarrow$ hence, a sample $(y_{0,b} : b = 1, ..., B)$ from $p(y_0 \mid \text{data})$ can be obtained using the MCMC output: for each $b = 1, ..., B$, we first draw $\theta_{0,b}$ from $p(\theta_0 \mid n_b^*, \boldsymbol{w}_b, \boldsymbol{\theta}_b^*, \alpha_b, \psi_b)$, and then draw $y_{0,b}$ from $p(y_0 \mid \theta_{0,b}, \phi_b) = K(\cdot; \theta_{0,b}, \phi_b)$

- To further highlight the mixture structure, note that we can also write

$$p(y_0 \mid \text{data}) = \int \left(\frac{\alpha}{\alpha+n} \int k(y_0 \mid \theta, \phi) g_0(\theta \mid \psi) \mathrm{d}\theta + \frac{1}{\alpha+n} \sum_{j=1}^{n^*} n_j k(y_0; \theta_j^*, \phi) \right) p(n^*, \boldsymbol{w}, \boldsymbol{\theta}^*, \alpha, \psi, \phi \mid \text{data})$$

  $\rightarrow$ the integrand above is a mixture with $n^* + 1$ components — the last n^* components (that dominate when α is small relative to n) yield a discrete mixture (in θ) of $k(\cdot; \theta, \phi)$ with the mixture parameters defined by the distinct θ_j^* — the posterior predictive density for y_0 is obtained by averaging this mixture with respect to the posterior of n^*, $\boldsymbol{w}$, $\boldsymbol{\theta}^*$ and all other parameters

Inference for general functionals of the random mixture

- Note that $p(y_0 \mid \text{data})$ is the posterior point estimate for the density functional $f(y_0; G, \phi)$ (at point y_0), i.e., $p(y_0 \mid \text{data}) = \mathrm{E}(f(y_0; G, \phi) \mid \text{data})$ (Bayesian density estimate can be obtained without sampling from the posterior of G)

- Analogously, we can obtain posterior moments for linear functionals $H(F(\cdot; G, \phi)) = \int H(K(\cdot; \theta, \phi))\mathrm{d}G(\theta)$ (Gelfand and Mukhopadhyay, 1995) — for linear functionals, the functional of the mixture is the mixture of the functionals applied to the parametric kernel (e.g., density and cdf functionals, mean functional)

- How about more general inference for functionals?
  → interval estimates for $F(y_0; G, \phi)$ for specified y_0, and, therefore, (pointwise) uncertainty bands for $F(\cdot; G, \phi)$?
  → inference for derived functions from $F(\cdot; G, \phi)$, e.g., cumulative hazard, $-\log(1 - F(\cdot; G, \phi))$, or hazard, $f(\cdot; G, \phi)/(1 - F(\cdot; G, \phi))$, functions?
  → inference for non-linear functionals, e.g., for median, and general percentiles?

Methods for posterior inference

- Such inferences require the posterior of G — recall,

$$p(G, \phi, \theta, \alpha, \psi \mid \text{data}) = p(G \mid \theta, \alpha, \psi) p(\theta, \phi, \alpha, \psi \mid \text{data})$$

and $G \mid \theta, \alpha, \psi \sim \text{DP}(\alpha + n, \tilde{G}_0 = \alpha(\alpha + n)^{-1} G_0(\cdot \mid \psi) + (\alpha + n)^{-1} \sum_{i=1}^{n} \delta_{\theta_i})$,

- Hence, after fitting the marginalized version of the DP mixture model (i.e., with G integrated out over its DP prior), using one of the available MCMC methods, we can obtain draws from the posterior of G

- For each posterior sample $(\theta_b, \alpha_b, \psi_b, \phi_b)$, $b = 1, ..., B$, we can draw G_b from $p(G \mid \theta_b, \alpha_b, \psi_b)$ using:
  → the constructive definition of the DP with a truncation approximation (Gelfand and Kottas, 2002; Kottas, 2003)
  → the original definition of the DP (through its fdds) if we only need posterior sample paths for the cdf of the mixture (and y is univariate)

- Finally, the posterior samples G_b yield posterior samples $\{H(F(\cdot; G_b, \phi_b)) : b = 1, ..., B\}$ from any functional $H(F(\cdot; G, \phi))$ of interest

1. Summary and references

Dirichlet process (DP) mixture models have largely dominated methodological and applied Bayesian nonparametric work in the past ten years, after the technology for their simulation-based model fitting was introduced (see session 5 for related references)

References categorized by methodological/application area include:

- Regression models for binary and univariate ordinal data (Erkanli et al., 1993; Basu and Mukhopadhyay, 2000)
- Hierarchical regression and density estimation (West et al., 1994; Escobar and West, 1995; Müller et al., 1996)
- Isotonic regression (Lavine and Mockus, 1995; Dunson, 2005)
- Randomised block experiments (Bush and MacEachern, 1996)
- Generalized linear, and linear mixed, models (Kleinman and Ibrahim, 1998; Mukhopadhyay and Gelfand, 1997; Müller and Rosner, 1997)

Summary and references

- Regression models for survival data (Kuo and Mallick, 1997; Kottas and Gelfand, 2001b; Gelfand and Kottas, 2003; Hanson, 2004)
- Errors-in-variables models (Müller and Roeder, 1997)
- Non-linear autoregressive time series models (Müller et al., 1997)
- Multiple comparisons problems (Gopalan and Berry, 1998)
- Contingency tables analysis (Quintana, 1998)
- Clinical trials design (Vlachos and Gelfand, 1998)
- Analysis of selection models (Lee and Berger, 1999)
- Applications in meta-analysis (Mallick and Walker, 1997; Müller et al., 1999)
- Inference under probability order constraints (Gelfand and Kottas, 2001; Kottas and Gelfand, 2001a)
- Nonparametric ANOVA models (Tomlinson and Escobar, 1999; De Iorio et al., 2004)

References

- Gelfand, A.E., Kottas, A., and MacEachern, S.N. (2004), "Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing," AMS Technical Report 2004-05, University of California, Santa Cruz. (To appear in *Journal of the American Statistical Association.*)
- Hjort, N.L. (2000), "Bayesian Analysis for a Generalised Dirichlet Process Prior," Statistical Research Report, Department of Mathematics, University of Oslo.
- Kottas, A., Branco, M.D., and Gelfand, A.E. (2002), "A Nonparametric Bayesian Modeling Approach for Cytogenetic Dosimetry," *Biometrics*, 58, 593-600.
- Kuo, L. (1983), "Bayesian Bioassay Design," *The Annals of Statistics*, 11, 886-895.
- Kuo, L. (1988), "Linear Bayes Estimators of the Potency Curve in Bioassay," *Biometrika*, 75, 91-96.
- MacEachern, S.N. (1999), "Dependent Nonparametric Processes," in *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association, pp. 50-55.
- MacEachern, S.N. (2000), "Dependent Dirichlet Processes," Technical Report, Department of Statistics, The Ohio State University.
- Madruga, M.R., Ochi-Lohmann, T.H., Okazaki, K., Pereira, C.A., de B., and Rabello-Gay, M.N. (1996), "Bayesian dosimetry. II: Credibility intervals for radiation dose," *Environmetrics*, 7, 325-331.
- Mukhopadhyay, S. (2000), "Bayesian Nonparametric Inference on the Dose Level With Specified Response Rate," *Biometrics*, 56, 220-226.
- Muliere, P., and Tardella, L. (1998), "Approximating Distributions of Random Functionals of Ferguson-Dirichlet Priors," *The Canadian Journal of Statistics*, 26, 283-297.
- Sethuraman, J., and Tiwari, R.C. (1982), "Convergence of Dirichlet Measures and the Interpretation of their Parameter," in *Statistical Decision Theory and Related Topics III*, eds. S. Gupta and J.O. Berger, New York: Springer-Verlag, 2, pp. 305-315.
- Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639-650.

References

References

- Antoniak, C.E. (1974), "Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems," *The Annals of Statistics*, 2, 1152-1174.
- Bhattacharya, P.K. (1981), "Posterior Distribution of a Dirichlet Process from Quantal Response Data," *The Annals of Statistics*, 9, 803-811.
- Blackwell, D. (1973), "Discreteness of Ferguson Selections," *The Annals of Statistics*, 1, 356-358.
- Blackwell, D., and MacQueen, J.B. (1973), "Ferguson Distributions via Pólya Urn Schemes," *The Annals of Statistics*, 1, 353-355.
- De Iorio, M., Müller, P., Rosner, G.L., and MacEachern, S.N. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205-215.
- Disch, D. (1981), "Bayesian Nonparametric Inference for Effective Doses in a Quantal-Response Experiment," *Biometrics*, 37, 713-722.
- Fabius, J. (1964), "Asymptotic Behavior of Bayes' Estimates," *The Annals of Mathematical Statistics*, 35, 846-856.
- Ferguson, T.S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209-230.
- Ferguson, T.S. (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615-629.
- Freedman, D.A. (1963), "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case," *The Annals of Mathematical Statistics*, 34, 1386-1403.
- Gelfand, A.E., and Kuo, L. (1991), "Nonparametric Bayesian Bioassay Including Ordered Polytomous Response," *Biometrika*, 78, 657-666.

References

- Antoniak, C.E. (1974), "Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems," *The Annals of Statistics*, 2, 1152-1174.
- Berry, D.A., and Christensen, R. (1979), "Empirical Bayes Estimation of a Binomial Parameter via Mixtures of a Dirichlet Process," *The Annals of Statistics*, 7, 558-568.
- Blackwell, D., and MacQueen, J.B. (1973), "Ferguson Distributions via Pólya Urn Schemes," *The Annals of Statistics*, 1, 353-355.
- Brunner, L.J. (1992), "Bayesian Nonparametric Methods for Data from a Unimodal Density," *Statistics and Probability Letters*, 14, 195-199.
- Brunner, L.J. (1995), "Bayesian Linear Regression With Error Terms That Have Symmetric Unimodal Densities," *Journal of Nonparametric Statistics*, 4, 335-348.
- Brunner, L.J., and Lo, A.Y. (1989), "Bayes Methods for a Symmetric Unimodal Density and its Mode," *The Annals of Statistics*, 17, 1550-1566.
- Bush, C.A., and MacEachern, S.N. (1996), "A Semiparametric Bayesian Model for Randomised Block Designs," *Biometrika*, 83, 275-285.
- Diebolt, J., and Robert, C.P. (1994), "Estimation of Finite Mixture Distributions through Bayesian Sampling," *Journal of the Royal Statistical Society*, Ser. B, 56, 363-375.
- Escobar, M.D. (1988), "Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means," unpublished Ph.D. dissertation, Yale University, Department of Statistics.
- Escobar, M.D. (1994), "Estimating Normal Means With a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89, 268-277.

References

- Escobar, M.D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577-588.
- Ferguson, T.S. (1983), "Bayesian Density Estimation by Mixtures of Normal Distributions," in *Recent Advances in Statistics*, eds. M.H. Rizvi, J.S. Rustagi and D. Siegmund, New York: Academic Press, pp. 287-302.
- Gelfand, A.E., and Mukhopadhyay, S. (1995), "On Nonparametric Bayesian Inference for the Distribution of a Random Sample," *The Canadian Journal of Statistics*, 23, 411-420.
- Gelfand, A.E., and Kottas, A. (2002), "A Computational Approach for Full Nonparametric Bayesian Inference under Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 11, 289-305.
- Ishwaran, H., and James, L.F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96, 161-173.
- Ishwaran, H., and Zarepour, M. (2000), "Markov chain Monte Carlo in approximate Dirichlet and Beta two-parameter process hierarchical models," *Biometrika*, 87, 371-390.
- Kottas, A. (2003), "Nonparametric Bayesian Survival Analysis using Mixtures of Weibull Distributions," to appear in *Journal of Statistical Planning and Inference* (Technical Report AMS 2003-01, UC Santa Cruz, available at: http://www.ams.ucsc.edu/reports/trview.php)
- Kottas, A., and Gelfand, A.E. (2001), "Bayesian Semiparametric Median Regression Modeling," *Journal of the American Statistical Association*, 96, 1458-1468.
- Kottas, A., and Krnjajić, M. (2005), "Bayesian Nonparametric Modeling in Quantile Regression," Technical Report AMS 2005-06, UC Santa Cruz, available at: http://www.ams.ucsc.edu/reports/trview.php
- Kuo, L. (1986a), "A Note on Bayes Empirical Bayes Estimation by Means of Dirichlet Processes," *Statistics and Probability Letters*, 4, 145-150.

References

- Kuo, L. (1986b), "Computations of Mixtures of Dirichlet Processes," *SIAM Journal on Scientific and Statistical Computing*, 7, 60-71.
- Kuo, L., and Mallick, B. (1997), "Bayesian Semiparametric Inference for the Accelerated Failure-Time Model," *The Canadian Journal of Statistics*, 25, 457-472.
- Kuo, L., and Smith, A.F.M. (1992), "Bayesian Computations in Survival Models via the Gibbs Sampler," in *Survival Analysis: State of the Art*, eds. J.P. Klein and P.K. Goel, Dodrecht: Kluwer Academic Publishers, pp. 11-24.
- Lavine, M., and Mockus, A. (1995), "A Nonparametric Bayes Method for Isotonic Regression," *Journal of Statistical Planning and Inference*, 46, 235-248.
- Liu, J.S. (1996), "Nonparametric Hierarchical Bayes via Sequential Imputations," *The Annals of Statistics*, 24, 911-930.
- Lo, A.Y. (1984), "On a Class of Bayesian Nonparametric Estimates: I. Density Estimates," *The Annals of Statistics*, 12, 351-357.
- MacEachern, S.N., and Müller, P. (1998), "Estimating Mixture of Dirichlet Process Models," *Journal of Computational and Graphical Statistics*, 7, 223-238.
- MacEachern, S.N., Clyde, M., and Liu, J.S. (1999), "Sequential Importance Sampling for Nonparametric Bayes Models: The Next Generation," *The Canadian Journal of Statistics*, 27, 251-267.
- McLachlan, G., and Peel, D. (2000), *Finite Mixture Models*, New York: Wiley.
- Müller, P., Erkanli, A., and West, M. (1996), "Bayesian Curve Fitting Using Multivariate Normal Mixtures," *Biometrika*, 83, 67-79.
- Neal, R.M. (2000), "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249-265.

References

- Newton, M.A., and Zhang, Y. (1999), "A Recursive Algorithm for Nonparametric Analysis With Missing Data," *Biometrika*, 86, 15-26.
- Quintana, F.A., and Iglesias, P.L. (2003), "Bayesian clustering and product partition models," *Journal of the Royal Statistical Society*, Ser. B, 65, 557-574.
- Richardson, S., and Green, P.J. (1997), "On Bayesian Analysis of Mixtures With an Unknown Number of Components" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 59, 731-792.
- Stephens, M. (2000), "Bayesian Analysis of Mixture Models With an Unknown Number of Components - An Alternative to Reversible Jump Methods," *The Annals of Statistics*, 28, 40-74.
- Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985), *Statistical Analysis of Finite Mixture Distributions*, Chichester: John Wiley.
- West, M., Müller, P., and Escobar, M.D. (1994), "Hierarchical Priors and Mixture Models, With Application in Regression and Density Estimation," in *Aspects of Uncertainty: A Tribute to D.V. Lindley*, eds. A.F.M. Smith and P. Freeman, New York: Wiley, pp. 363-386.

References

- Basu, S., and Mukhopadhyay, S. (2000), "Bayesian analysis of binary regression using symmetric and asymmetric links," *Sankhya, Series B, Indian Journal of Statistics*, 62, 372-387.
- Brunner, L.J. (1995), "Bayesian Linear Regression With Error Terms That Have Symmetric Unimodal Densities," *Journal of Nonparametric Statistics*, 4, 335-348.
- Brunner, L.J., and Lo, A.Y. (1989), "Bayes Methods for a Symmetric Unimodal Density and its Mode," *The Annals of Statistics*, 17, 1550-1566.
- Bush, C.A., and MacEachern, S.N. (1996), "A Semiparametric Bayesian Model for Randomised Block Designs," *Biometrika*, 83, 275-285.
- De Iorio, M., Müller, P., Rosner, G.L., and MacEachern, S.N. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205-215.
- Dunson, D.B. (2005), "Bayesian semiparametric isotonic regression for count data," *Journal of the American Statistical Association*, 100, 618-627.
- Erkanli, A., Stangl, D. and Müller, P. (1993), "A Bayesian analysis of ordinal data using mixtures," *ASA Proceedings of the Section on Bayesian Statistical Science*, 51-56, American Statistical Association (Alexandria, VA).
- Escobar, M.D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577-588.
- Gelfand, A.E., and Kottas, A. (2001), "Nonparametric Bayesian Modeling for Stochastic Order," *Annals of the Institute of Statistical Mathematics*, 53, 865-876.
- Gelfand, A.E., and Kottas, A. (2003), "Bayesian Semiparametric Regression for Median Residual Life," *Scandinavian Journal of Statistics*, 30, 651-665.

References

- Gopalan, R., and Berry, D.A. (1998), "Bayesian Multiple Comparisons Using Dirichlet Process Priors," *Journal of the American Statistical Association*, 93, 1130-1139.
- Hanson, T.E. (2004), "Modeling censored lifetime data using a mixture of gammas baseline," Technical Report, Department of Mathematics and Statistics, University of New Mexico.
- Ibrahim, J.G., Chen, M-H., and Sinha, D. (2001), *Bayesian survival analysis*, New York: Springer.
- Kleinman, K.P., and Ibrahim, J.G. (1998), "A semi-parametric Bayesian approach to generalized linear mixed models," *Statistics in Medicine*, 17, 2579-2596.
- Kottas, A. (2003), "Nonparametric Bayesian Survival Analysis using Mixtures of Weibull Distributions," to appear in *Journal of Statistical Planning and Inference* (Technical Report AMS 2003-01, UC Santa Cruz, available at: http://www.ams.ucsc.edu/reports/trview.php)
- Kottas, A., and Gelfand, A.E. (2001a), "Modeling Variability Order: A Semiparametric Bayesian Approach," *Methodology and Computing in Applied Probability*, 3, 427-442.
- Kottas, A., and Gelfand, A.E. (2001b), "Bayesian Semiparametric Median Regression Modeling," *Journal of the American Statistical Association*, 96, 1458-1468.
- Kottas, A., and Krnjajić, M. (2005), "Bayesian Nonparametric Modeling in Quantile Regression," Technical Report AMS 2005-06, UC Santa Cruz, available at: http://www.ams.ucsc.edu/reports/trview.php
- Kottas, A., Müller, P., and Quintana, F. (2003), "Nonparametric Bayesian Modeling for Multivariate Ordinal Data," to appear in *Journal of Computattional and Graphical Statistics* (Technical Report AMS 2003-07, UC Santa Cruz, available at: http://www.ams.ucsc.edu/reports/trview.php)
- Kuo, L., and Mallick, B. (1997), "Bayesian Semiparametric Inference for the Accelerated Failure-Time Model," *The Canadian Journal of Statistics*, 25, 457-472.

References

- Lavine, M., and Mockus, A. (1995), "A Nonparametric Bayes Method for Isotonic Regression," *Journal of Statistical Planning and Inference*, 46, 235-248.
- Lawless, J.F. (1982), *Statistical models and methods for lifetime data*, New York: Wiley.
- Lee, J., and Berger, J.O. (1999), "Semiparametric Bayesian Analysis of Selection Models," *Journal of the American Statistical Association*, 96, 1397-1409.
- Mallick, B.K., and Walker, S.G. (1997), "Combining Information from Several Experiments With Nonparametric Priors," *Biometrika*, 84, 697-706.
- Müller, P., Erkanli, A., and West, M. (1996), "Bayesian Curve Fitting Using Multivariate Normal Mixtures," *Biometrika*, 83, 67-79.
- Müller, P., and Roeder, K. (1997), "A Bayesian Semiparametric Model for Case-control Studies With Errors in Variables," *Biometrika*, 84, 523-537.
- Müller, P., and Rosner, G.L. (1997), "A Bayesian Population Model With Hierarchical Mixture Priors Applied to Blood Count Data," *Journal of the American Statistical Association*, 92, 1279-1292.
- Müller, P., Quintana, F., and Rosner, G. (1999), "Hierarchical Meta-analysis over Related Non-parametric Bayesian Models," Discussion paper 99-22, ISDS, Duke University.
- Müller, P., West, M., and MacEachern, S. (1997), "Bayesian Models for Non-linear Autoregressions," *Journal of Time Series Analysis*, 18, 593-614.
- Quintana, F.A. (1998), "Nonparametric Bayesian Analysis for Assessing Homogeneity in $k \times l$ Contingency Tables With Fixed Right Margin Totals," *Journal of the American Statistical Association*, 93, 1140-1149.
- Tomlinson, G., and Escobar, M. (1999), "Analysis of Densities," Research Report, Department of Public Health Sciences, University of Toronto.

References

- Vlachos, P.K., and Gelfand, A.E. (1998), "Nonparametric Bayesian Group Sequential Design," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Müller and D. Sinha, New York: Springer, pp. 115-132.
- Walker, S.G., Damien, P., Laud, P.W., and Smith, A.F.M. (1999), "Bayesian Nonparametric Inference for Random Distributions and Related Functions" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 61, 485-527.
- West, M., Müller, P., and Escobar, M.D. (1994), "Hierarchical Priors and Mixture Models, With Application in Regression and Density Estimation," in *Aspects of Uncertainty: A Tribute to D.V. Lindley*, eds. A.F.M. Smith and P. Freeman, New York: Wiley, pp. 363-386.

References

References

- Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2003), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall.
- Brix, A., and Diggle, P.J. (2001), "Spatiotemporal Prediction for Log-Gaussian Cox Processes," *Journal of the Royal Statistical Society, Series B*, 63, 823-841.
- Cressie, N.A.C. (1993), *Statistics for Spatial Data* (Revised Edition), New York: Wiley.
- Damian, D., Sampson, P.D., and Guttorp, P. (2001), "Bayesian Estimation of Semi-parametric Non-stationary Spatial Covariance Structures," *Environmetrics*, 12, 161-178.
- De Iorio, M., Müller, P., Rosner, G.L., and MacEachern, S.N. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205-215.
- Diggle, P.J., Mateu, J., and Clough, H.E. (2000), "A Comparison Between Parametric and Non-Parametric Approaches to the Analysis of Replicated Spatial Point Patterns," *Advances in Applied Probability*, 32, 331-343.
- Diggle, P.J., Tawn, J.A., and Moyeed, R.A. (1998), "Model-based Geostatistics (with discussion)," *Applied Statistics*, 47, 299-350.
- Fuentes, M., and Smith, R.L. (2001), "A New Class of Nonstationary Spatial Models," Technical Report, Department of Statistics, North Carolina State University.
- Gelfand, A.E., Kottas, A., and MacEachern, S.N. (2004), "Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing", to appear in *Journal of the American Statistical Association* (Technical Report AMS 2004-05, UC Santa Cruz, available at: http://www.ams.ucsc.edu/reports/trview.php)
- Higdon, D. (2002), "Space and Space-time Modeling Using Process Convolutions," in Quantitative Methods for Current Environmental Issues, eds. C.W. Anderson, V. Barnett, P.C. Chatwin and A.H. El-Shaarawi, London: Springer Verlag.

References

- Kottas, A., Duan, J., and Gelfand, A.E. (2005), "Spatial and Spatio-temporal Dirichlet Process Mixtures with Applications to Modeling Disease Incidence Data," (manuscript in preparation)
- MacEachern, S.N. (1999), "Dependent Nonparametric Processes," in *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association, pp. 50-55.
- MacEachern, S.N. (2000), "Dependent Dirichlet Processes," Technical Report, Department of Statistics, The Ohio State University.
- Møller, J., Syversveen, A.R., and Waagepetersen, R.P. (1998), "Log Gaussian Cox Processes," *Scandinavian Journal of Statistics*, 25, 451-482.
- Sampson, P.D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108-119.
- Schmidt, A.M., and O'Hagan, A. (2003), "Bayesian Inference for Nonstationary Spatial Covariance Structure via Spatial Deformations," *Journal of the Royal Statistical Society, Series B*, 65, 743-758.
- Stein, M.L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer Verlag.
- Wolpert, R.L., and Ickstadt, K. (1998), "Poisson/Gamma Random Field Models for Spatial Statistics," *Biometrika*, 85, 251-267.