Ann. Inst. Statist. Math.
Vol. 53, No. 4, 865–876 (2001)
©2001 The Institute of Statistical Mathematics

NONPARAMETRIC BAYESIAN MODELING FOR STOCHASTIC ORDER

ALAN E. GELFAND* AND ATHANASIOS KOTTAS**

Department of Statistics, The College of Liberal Arts and Sciences, University of Connecticut,
196 Auditorium Road, U-120, Storrs, CT 06269-3120, U.S.A.

(Received June 21, 1999; revised August 18, 2000)

Abstract. In comparing two populations, sometimes a model incorporating stochastic order is desired. Customarily, such modeling is done parametrically. The objective of this paper is to formulate nonparametric (possibly semiparametric) stochastic order specifications providing richer, more flexible modeling. We adopt a fully Bayesian approach using Dirichlet process mixing. An attractive feature of the Bayesian approach is that full inference is available regarding the population distributions. Prior information can conveniently be incorporated. Also, prior stochastic order is preserved to the posterior analysis. Apart from the two sample setting, the approach handles the matched pairs problem, the k-sample slippage problem, ordered ANOVA and ordered regression models. We illustrate by comparing two rather small samples, one of diabetic men, the other of diabetic women. Measurements are of androstenedione levels. Males are anticipated to produce levels which will tend to be higher than those of females.

Key words and phrases: Dirichlet process mixing, linear functionals, Monte Carlo sampling and integration, semiparametric models.

1. Introduction

In comparing two populations, it is sometimes anticipated that observations from one will tend to be larger than those from the other. One formal way to capture this is to assume that the two populations are stochastically ordered. That is, labeling the population c.d.f.'s as F_1 and F_2, if X is drawn from F_1 and Y is drawn from F_2, we say that Y is stochastic larger than X if $F_1(c) \geq F_2(c)$ for all c.

Customarily, stochastic order is modeled parametrically. A parametric family of distributions is selected with c.d.f. $F(\cdot;\theta)$ such that whenever $\theta_1 < \theta_2$, $F(c;\theta_1) \geq F(c;\theta_2)$. Then, we take $F_j(\cdot) = F(\cdot;\theta_j)$, $j = 1, 2$. For instance, $F(\cdot;\theta)$ might be a one-parameter exponential family therefore having monotone likelihood ratio which implies stochastic order (as discussed in, e.g., Lehmann (1986), Section 3.3). Alternatively, for random variables on R^1, taking a fixed c.d.f. F_0, we might introduce a location parameter θ creating $F(\cdot;\theta) = F_0(\cdot - \theta)$. Now, if $F_j(\cdot) = F_0(\cdot - \theta_j)$, $j = 1, 2$, with $\theta_1 < \theta_2$ we achieve stochastic order. $\eta = \theta_2 - \theta_1$ is referred to as the shift parameter (see, e.g., Randles and Wolfe (1979), Chapter 9, for a full discussion). In the fully parametric context, inference about F_1, F_2 and η is straightforward but the modeling is limited by the required

*The work of the first author was supported in part by NSF grant DMS 96-25383.

**A. Kottas is Visiting Assistant Professor in the Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251, U.S.A.

specification of the parametric family.

Joe ((1997), p. 21) employs the terminology that X is stochastically increasing in θ if the conditional distribution of X given θ is such that $F(x \mid \theta)$ decreases in θ for fixed x. The foregoing examples illustrate this definition.

Classical inference regarding θ_1 and θ_2 (hence F_1 and F_2) is implemented under the *hard* restriction $\theta_1 < \theta_2$. If we adopt a Bayesian approach, the modeling requires θ_1 and θ_2 random and thus a prior confined to the set $\{(\theta_1, \theta_2) : \theta_1 < \theta_2\}$. Attractively, such a prior restriction implies that, after data collection, stochastic order is retained a posteriori. Bayesian inference avoids constrained optimization.

The objective of this paper is to formulate stochastic order and stochastically increasing model specifications which arise nonparametrically, providing a richer, more flexible class of models to work with. A fully Bayesian approach is developed through the use of nonparametric mixture models. We illustrate using Dirichlet processes (Ferguson (1973)) implemented through Dirichlet process mixing, as in Antoniak (1974) and Lo (1984).

Focusing on the two sample problem, again prior stochastic order implies posterior stochastic order. Additionally, full inference regarding F_1 and F_2 is available. By contrast, classical nonparametric approaches to the two sample problem, e.g., the Mann-Whitney test (as in, say, Randles and Wolfe (1979)), assume F_2 is shifted from F_1 but do not specify F_0. Inference is limited to the shift parameter. Comparison of c.d.f.'s requires empirical c.d.f.'s which need not be stochastically ordered. Such order would have to be imposed in an ad hoc fashion.

The parametric Bayesian approach requires a prior on (θ_1, θ_2). As a first step toward nonparametric extension (in the spirit of Lehmann (1986), p. 84), suppose $g(x)$ is a strictly increasing function such that $g(x) \geq x$ and F is a c.d.f. Set $F_1(\cdot) = F(g(\cdot))$, $F_2(\cdot) = F(\cdot)$. Then F_2 is stochastically larger than F_1. Assuming F is unknown, let it be random; then F_1 and F_2 are. However, if g is fixed this class is too limited; this "F" prior specification is such that F_1 determines F_2 and vice versa. The parametric analog is to select θ_1 at random and then set $\theta_2 = \theta_1 + c$, with c a fixed positive constant. To make $g(x)$ an arbitrary random increasing function lying above x is awkward. A simplification would set $g(x) = x + \mu$ where μ is a positive random variable. This "(F, μ)" prior specification is semiparametric. The parametric analog here is to draw θ_1 at random and then μ at random, setting $\theta_2 = \theta_1 + \mu$. This, of course, is equivalent to drawing (θ_1, θ_2) at random on the set $\{(\theta_1, \theta_2) : \theta_1 < \theta_2\}$. In what follows we develop an "(F_1, F_2)" prior specification, a nonparametric prior yielding F_1, F_2 random with F_2 stochastically larger than F_1. In this regard, the only existing Bayesian work we are aware of appears in Arjas and Gasbarra (1996). For survival models, they specify a random pair of piecewise hazard functions obeying a partial ordering which implies stochastic order for the associated distributions.

Note that, in the parametric case we might introduce a dispersion parameter σ, extending the parametric family to $F(\cdot; \theta, \sigma)$ which is stochastically increasing in θ for each fixed σ. In the Bayesian framework, the prior must then be extended to σ. We shall similarly enrich the nonparametric case, adding a random σ to F_1 and F_2, yielding a semiparametric prior specification.

The plan of the paper is the following. In Section 2 we briefly review the Dirichlet process (DP), Dirichlet process mixing and inference under a nonparametric Bayesian specification for the distribution of a population. In Section 3 we describe our approach to creating random bivariate distributions with stochastically ordered marginals. The advantage to DP mixing rather than working directly with DP's is revealed. Section 4 presents a range of applications. Apart from the two sample problem we consider the matched pairs case, the k-sample slippage problem, ordered regression models, ordered ANOVA models and partial stochastic order structures. Section 5 addresses computational matters; DP mixing is implemented using Gibbs sampling but the details are a bit different from the usual implementation (see, e.g., Escobar and West (1995)). In Section 6 we indicate how rough prior knowledge can be conveniently utilized in the required prior specifications. Section 7 provides an illustrative analysis with a small dataset.

2. Dirichlet process and Dirichlet process mixing

Following Ferguson (1973), a distribution G on Θ follows a Dirichlet process $DP(\alpha G_0)$ if, given an arbitrary finite measurable partition, $B_1,\ldots,B_r$ of Θ, $(G(B_1),\ldots, G(B_r)) \sim Dirichlet(\alpha G_0(B_1),\ldots,\alpha G_0(B_r))$ where $G(B_i)$ denotes the probability of set B_i under G, and similarly for G_0. Here, G_0 is a specified distribution on Θ and $\alpha > 0$ is a precision parameter.

Let $F(\cdot;\boldsymbol{\theta})$ be a parametric family of distributions (c.d.f.'s), indexed by $\boldsymbol{\theta} \in \Theta$, with associated densities, $f(\cdot;\boldsymbol{\theta})$. If G is proper we define the mixture distribution

$$F(\cdot;G) = \int F(\cdot;\boldsymbol{\theta})G(d\boldsymbol{\theta}). \tag{2.1}$$

In (2.1) it is useful to think of $G(d\boldsymbol{\theta})$ as the conditional distribution of $\boldsymbol{\theta}$ given G. Differentiating both sides of (2.1) with respect to $(\cdot)$ defines $f(\cdot;G) = \int f(\cdot;\boldsymbol{\theta})\ G(d\boldsymbol{\theta})$.

If G is random say $G \sim DP(\alpha G_0)$, then $F(\cdot;G)$ is random. If the data D are $Y_1,\ldots,Y_n$ independent and identically distributed from $F(\cdot;G)$ then, using the convenient bracket notation of Gelfand and Smith (1990), we write the posterior of $F(\cdot;G)$ as $[F(\cdot;G) \mid D]$. Functionals of $F(\cdot;G)$, which we denote by $H(F(\cdot;G))$, are of interest with posteriors denoted by $[H(F(\cdot;G)) \mid D]$.

In the context of (2.1), suppose for each Y_i, $i = 1,\ldots,n$ we introduce a latent $\boldsymbol{\theta}_i$ and assume that the Y_i's are conditionally independent given the $\boldsymbol{\theta}_i$'s. Assume further that the $\boldsymbol{\theta}_i$'s are conditionally independent and identically distributed given G. As a result the Y_i's are marginally independent, with joint density $\prod_{i=1}^n f(y_i;G) = \prod_{i=1}^n \int f(y_i;\boldsymbol{\theta}_i)$ $G(d\boldsymbol{\theta}_i)$. Adding $G \sim DP(\alpha G_0)$ completes the Bayesian model specification, apart, perhaps, from a hyperprior on α (see Escobar and West (1995)). Antoniak (1974) noted that this Bayesian model can be *marginalized* over G to obtain $\prod_{i=1}^n f\ (y_i;\boldsymbol{\theta}_i)[\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n \mid G_0,\alpha]$. After marginalization the $\boldsymbol{\theta}_i$ are no longer independent but a Gibbs sampler can be routinely implemented (Escobar and West (1995)) to obtain samples from the posterior $[\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n \mid D]$.

Gelfand and Mukhopadhyay (1995) describe how to use these samples to compute moments of $[H(F(\cdot;G)) \mid D]$ when H is a linear functional. Such functionals include expectations with respect to $F(\cdot;G)$, enabling the mean, variance and characteristic function functionals to be studied along with the "c.d.f.-at-a-point" and "p.d.f.-at-a-point" functionals. The important quantile functional is not linear.

Restriction to posterior moments of linear functionals necessarily limits inference. In a recent paper, Gelfand and Kottas (2001) show how to obtain the entire posterior distribution for more general functionals. Hence, exact inference is available for many population features and for comparing such features across populations.

Briefly, note that for H, a linear functional, $H(F(\cdot;G)) = \int H(F(\cdot;\boldsymbol{\theta}_0))G(d\boldsymbol{\theta}_0)$. Now, instead of marginalizing over G in $[\boldsymbol{\theta}_0, \boldsymbol{\theta}, G \mid D] \propto [D \mid \boldsymbol{\theta}][\boldsymbol{\theta}_0, \boldsymbol{\theta} \mid G][G]$, observe that this joint posterior is proportional to $[\boldsymbol{\theta}_0 \mid G]\ [G \mid \boldsymbol{\theta}]\ [\theta \mid D]$. Hence given $\boldsymbol{\theta}_b^*$, $b = 1, \ldots, B$ from $[\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n \mid D]$, for each $\boldsymbol{\theta}_b^*$ draw $G_b^* \sim [G \mid \boldsymbol{\theta}_b^*]$ and then $\boldsymbol{\theta}_{0lb}^* \sim G_b^*$, for $l = 1, \ldots, L$. Finally, $H_b^* = L^{-1}\sum_{l=1}^{L} H(F(\cdot;\boldsymbol{\theta}_{0lb}^*))$ is a Monte Carlo integration for a realization from $[H(F(\cdot;G)) \mid D]$. To obtain an approximate realization from $[G \mid \boldsymbol{\theta}_b^*]$, which is an updated Dirichlet process (Ferguson (1973)), we use the constructive definition of Sethuraman (1994). Sampling from the posterior of the "c.d.f.-at-a-point" functional, for a grid of points, we can invert to obtain samples from the posterior of any quantile functional. Other functionals of interest can also be handled.

In the interest of clarifying Bayesian learning under this DP mixing framework we might wish to summarize prior features associated with $F(\cdot;G)$. The approach of Gelfand and Mukhopadhyay (1995) can be applied to prior expectations in the same fashion as for posterior expectations. Also the foregoing ideas of Gelfand and Kottas (2001) can be applied a priori by approximately sampling $[G]$ rather than $[G \mid D]$.

If we write $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$, we might place a DP prior on $\boldsymbol{\theta}^{(1)}$, i.e., $\boldsymbol{\theta}^{(1)} \sim G$ where $G \sim DP(\alpha G_0)$ with a parametric prior on $\boldsymbol{\theta}^{(2)}$. For instance, $\boldsymbol{\theta}^{(1)}$ might be a location parameter μ and $\boldsymbol{\theta}^{(2)}$ a dispersion parameter σ yielding $F(\cdot;G,\sigma) = \int F(\cdot;\mu,\sigma)G(d\mu)$, a semiparametric specification. In Section 7 we take $F(\cdot;\mu,\sigma) = \Phi((\cdot - \mu)/\sigma)$, where Φ is the standard Normal c.d.f.

3. A nonparametric approach for modeling stochastic order

Following the introduction, we seek joint distributions (F_1, F_2) over the space $\mathcal{P} = \{(F_1, F_2) : F_1 \leq_{st} F_2\}$ where $F_1 \leq_{st} F_2$ means F_2 is stochastically larger than F_1. Again, the elements of $\mathcal{P}$ are a pair of univariate functions. In general, the richness of a family of probability measures over a function space is hard to assess. Special spaces offering analytic characterizations can clarify this assessment, e.g., scale and location mixtures of a continuous symmetric density on R^1 provide all distributions on R^1 (Lo (1984), p. 355).

For $\mathcal{P}$ we only know $0 \leq F_2(c)/F_1(c) \leq 1$ (defining $F_2(c)/F_1(c) = 0$ if $F_1(c) = 0$.) Consider the subspace of $\mathcal{P}, \mathcal{P}' = \{(F_1, F_2) : F_1 = G_1, F_2 = G_1G_2\}$ where G_1 and G_2 are c.d.f.'s. Obviously, on $\mathcal{P}'$, $F_2(c)/F_1(c)$ increases from 0 to 1 in c. $\mathcal{P}'$ provides a constructive characterization for developing classes of probability measures. Any joint distribution over (G_1, G_2) induces a distribution on $\mathcal{P}'$. In fact, it is helpful to think of F_1 as the distribution of θ and F_2 as the distribution of $\max(\theta, \delta)$ where $\theta \sim G_1$ and independently $\delta \sim G_2$.

Customary probability models for G_1 and G_2 would be independent Dirichlet processes or more generally Polya tree processes (see, e.g., Walker, *et al.*, (1999) and references therein). Suppose $G_1 \sim DP(\alpha G_{10})$, $G_2 \sim DP(\beta G_{20})$, then the distribution for the pair of univariate random variables $(F_1(c), F_2(d))$ with, say $c < d$, can be obtained. Letting $B_1 = \{x : x \in (-\infty, c]\}$, $B_2 = \{x : x \in (c, d]\}$, $B_3 = \{x : x \in (d, \infty)\}$, $C_1 = \{z : z \in (-\infty, d]\}$, $C_2 = \{z : z \in (d, \infty)\}$, $U_i = G_1(B_i)$, $i = 1, 2, 3$, $V_j = G_2(C_j)$, $j = 1, 2$, we have $F_1(c) = U_1$, $F_2(d) = (U_1 + U_2)V_1$ so that the joint distribution of $(F_1(c), F_2(d))$ is obtained by simple transformation. We note that $E(F_1(c)) = G_{10}(c)$, i.e., $EF_1 = G_{10}$. But also $E(F_2(c)) = E(G_1(c)G_2(c)) = EG_1(c)EG_2(c) = G_{10}(c)G_{20}(c)$, i.e., $EF_2 = G_{10}G_{20}$. Other moments of $F_1(c)$ and $F_2(c)$ can be readily calculated. Also it is straightforward to show that $\text{Cov}(F_1(c), F_2(d)) > 0$.

In principle we could similarly obtain the joint distribution of any finite collection $(F_1(c_i), i = 1, \ldots, I, F_2(d_j),\ j = 1, \ldots, J)$. However, this distribution will be messy, awkward to sample and only provides a finite approximation to $[F_1, F_2]$. Introducing DP mixing simplifies matters considerably as we now clarify. We begin with a lemma which is a general version of a result in Shaked and Shanthikumar ((1994), p. 8). It indicates how stochastic order can be preserved through mixing.

LEMMA 1. *Suppose $F(\cdot;\theta)$ is a family of c.d.f.'s, $\theta \in (\underline{\theta}, \overline{\theta})$, which is strictly decreasing in θ. Let F_1 and F_2 be two c.d.f.'s on $(\underline{\theta}, \overline{\theta})$ such that $F_1 \leq_{st} F_2$. Let $F(\cdot; F_i) = \int_{\underline{\theta}}^{\overline{\theta}} F(\cdot;\theta) F_i(d\theta)$, $i = 1, 2$ (as in (2.1)). Then $F(\cdot;F_1) \leq_{st} F(\cdot;F_2)$.*

PROOF. Since $F(c;\theta)$ is strictly decreasing in θ, for each c we can define the measure Γ_c on $(\underline{\theta}, \overline{\theta})$ by $\Gamma_c((\underline{\theta}, \theta]) \equiv \Gamma_c(\theta) = F(c;\underline{\theta}) - F(c;\theta)$. Then

$$
\begin{aligned}
0 < \int_{\underline{\theta}}^{\overline{\theta}} (F_1(\theta) - F_2(\theta)) d\Gamma_c(\theta) &= \Gamma_c(\theta)(F_1(\theta) - F_2(\theta)) \,\Big|_{\underline{\theta}}^{\overline{\theta}} - \int_{\underline{\theta}}^{\overline{\theta}} \Gamma_c(\theta) d(F_1(\theta) - F_2(\theta)) \\
&= \int_{\underline{\theta}}^{\overline{\theta}} F(c;\theta) d(F_1(\theta) - F_2(\theta)) \\
&= F(c;F_1) - F(c;F_2).
\end{aligned}
$$

Remark 1. Under the assumptions of Lemma 1, we see that the "c.d.f.-at-a-point" functional is ordered under F_1 and F_2. (By inversion, the quantiles of F_1 and F_2 are ordered.) Such order applies to other linear functionals. If say $H(F(\cdot;\theta))$ increases in θ, then $H(F(\cdot;F_1)) < H(F(\cdot;F_2))$, e.g., provided the expectations exist, $E(Y \mid F_1) < E(Y \mid F_2)$ if $E(Y \mid \theta)$ increases in θ.

Remark 2. We can add a dispersion parameter σ to the model, as at the end of Section 2, and extend Lemma 1 to conclude that $F(\cdot;F_1,\sigma) \leq_{st} F(\cdot;F_2,\sigma)$. This applies to all of the examples below.

To clarify, Lemma 1 is applicable to any $(F_1, F_2) \in \mathcal{P}$ with common support $(\underline{\theta}, \overline{\theta})$ and by such mixing we obtain the class $\mathcal{P}_F = \{(F(\cdot;F_1), F(\cdot;F_2)) : (F_1, F_2) \in \mathcal{P}\}$ where the subscript F denotes the choice of kernel $F(\cdot;\theta)$ in (2.1). The Lemma asserts that $\mathcal{P}_F \subset \mathcal{P}$. Hence, a probability model over $\mathcal{P}$ induces a probability model over $\mathcal{P}_F$. Applied to $(F_1, F_2) \in \mathcal{P}'$ we note that $F(\cdot;F_1) = \int F(\cdot;\theta) G_1(d\theta)$ and $F(\cdot;F_2) = \iint F(\cdot;\max(\theta,\delta)) G_1(d\theta) G_2(d\delta)$. Attractively, if G_1 and G_2 are from independent Dirichlet processes, we can perform the marginalization over F_1 and F_2 noted in Section 2.

For instance, if $G_1 \sim DP(\alpha G_{10})$ and $G_2 \sim DP(\beta G_{20})$ and H is a linear functional then straightforwardly, $E(H(F(\cdot;F_1))) = H(F(\cdot;G_{10}))$. However, $E(H(F(\cdot;F_2))) = E_{G_1,G_2} \iint H(F(\cdot;\max(\theta,\delta))) G_1(d\theta) G_2(d\delta)$. Building from indicator functions, this latter expectation is $\iint H(F(\cdot;\max(\theta,\delta))) G_{10}(d\theta) G_{20}(d\delta)$. Prior and posterior inference for more general $H(F(\cdot;F_i))$ was discussed in Section 2.

In summary, working with location mixing of a continuous kernel (as we do below), we model the c.d.f.'s directly, the resultant c.d.f.'s are continuous, the location mixing provides flexible distributions and ready interpretation, the computation is relatively straightforward (Section 5 below), finite approximation is avoided and full prior and posterior inference is available.

4. Examples

We now apply our approach to a range of stochastic order problems which are usually addressed parametrically.

- The two-sample problem. Here we seek $X_1, \ldots, X_m$ i.i.d. $F(\cdot; F_1)$, $Y_1, \ldots, Y_n$ i.i.d. $F(\cdot; F_2)$, X_i, Y_j independent for all i and j, with say $F(\cdot; F_1) \leq_{st} F(\cdot; F_2)$. Suppose we introduce $\theta_1, \theta_2, \ldots, \theta_m$, $\theta_{m+1}, \ldots, \theta_{m+n}$ i.i.d. G_1, $\delta_1, \ldots, \delta_n$ i.i.d. G_2. Then we assume $X_i \mid \theta_i \sim F(\cdot; \theta_i)$, $i = 1, \ldots, m$ so that marginally $X_1, \ldots, X_m$ are i.i.d. $F(\cdot; F_1)$, where $F_1 = G_1$. Letting $\eta_j = \max(\theta_{m+j}, \delta_j)$, $j = 1, \ldots, n$, we assume $Y_j \mid \eta_j \sim F(\cdot; \eta_j)$ so that marginally $Y_1, \ldots, Y_n$ are i.i.d. $F(\cdot; F_2)$ where $F_2 = G_1 G_2$. Note that X_i and Y_j are conditionally independent given F_1 and F_2. $F(\cdot; F_1)$ and $F(\cdot; F_2)$ are not independent since F_1 and F_2 are not.

- The paired comparison or matched pairs problem. Here we seek (X_i, Y_i), i.i.d. pairs, $i = 1, \ldots, n$ such that, marginally, $X_i \sim F(\cdot; F_1)$, $Y_i \sim F(\cdot; F_2)$ with again, $F(\cdot; F_2)$ stochastically greater than $F(\cdot; F_1)$. Now we introduce only $\theta_1, \ldots, \theta_n$ i.i.d. G_1 with $\delta_1, \ldots, \delta_n$ i.i.d. G_2 and $\eta_i = \max(\theta_i, \delta_i)$. Then, given F_1 and F_2, (X_i, Y_i) are jointly distributed as $F(x, y; F_1, F_2) = \iint F(x; \theta) F(y; \max(\theta, \delta))\ G_1(d\theta)\ G_2(d\delta)$; as in the previous example, marginally, X_i and Y_i are stochastically ordered.

- The k-sample slippage problem. Here we seek X_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, independent such that $X_{ij} \sim F(\cdot; F_i)$ and the $F(\cdot; F_i)$ are, say stochastically increasing. Using Lemma 1, it suffices that the F_i be stochastically increasing. Most generally, we could set $F_i = G_1 \cdots G_i$ with the G_l independent $\sim DP(\alpha_l G_{l0})$, $l = 1, \ldots, k$. We can draw a parallel with the nonparametric version of the one way ANOVA model in Akritas and Arnold (1994). Though they do not consider stochastic order, they do represent F_i as an additive form. We represent $\log F_i$ as an additive form. In practice, specifying k prior Dirichlet processes may be too much to ask. We may instead set $F_i = G^i$, i.e., F_i is the distribution of $\max(\theta^{(l)} : l = 1, \ldots, i)$, where the $\theta^{(l)}$ are i.i.d. from G, and $G \sim DP(\alpha G_0)$. Following the two sample problem above, implementation requires $\theta_1^{(1)}, \ldots, \theta_{n_1}^{(1)}$ for sample 1, $(\theta_1^{(2)}, \ldots, \theta_{n_2}^{(2)}, \theta_{n_2+1}^{(2)}, \ldots, \theta_{2n_2}^{(2)})$ for sample 2 with $\eta_l = \max(\theta_l^{(2)}, \theta_{n_2+l}^{(2)})$, etc.

- The ordered two way layout. Here we seek X_{ijk}, $k = 1, \ldots, n_{ij}$ independent such that $X_{ijk} \sim F(\cdot; F_{ij})$ and the $F(\cdot; F_{ij})$ are stochastically increasing in i for fixed j and in j for fixed i. Again, using Lemma 1 it suffices that F_{ij} be stochastically increasing in i for each j and in j for each i. Here, we might set $F_{ij} = G_1^i G_2^{j-1}$, i.e., F_{ij} is the distribution of $\max(\theta_1, \ldots, \theta_i, \delta_1, \ldots, \delta_{j-1})$ where the θ_l are i.i.d. from G_1, the δ_m are i.i.d. from G_2, with $G_1 \sim DP(\alpha G_{10})$ and $G_2 \sim DP(\beta G_{20})$, G_1 and G_2 independent. LogF_{ij} is again additive with a component for each factor. Also, the n_{ij} can be 1 since we have only two unknown G's. Implementation parallels the one-way ANOVA slippage problem.

The previous two examples reveal how, within ANOVA modeling, any parametric order restrictions can be replaced with nonparametric stochastic order restrictions.

- Ordered regression models. To illustrate possibilities in a regression context, we recall the Dirichlet process mixed generalized linear models introduced in Mukhopadhyay and Gelfand (1997). In an attempt to enrich the class of generalized linear models beyond the rather restrictive one-parameter exponential family of stochastic mechanisms, they consider $F(\cdot;\boldsymbol{x}^T\boldsymbol{\beta},G) = \int F(\cdot;\alpha+\boldsymbol{x}^T\boldsymbol{\beta})G(d\alpha)$, where $F(\cdot;\alpha+\boldsymbol{x}^T\boldsymbol{\beta})$ is the c.d.f. of a customary generalized linear model. But then, if $F_1 \leq_{st} F_2$, $F(\cdot;\boldsymbol{x}^T\boldsymbol{\beta},F_1) \leq_{st} F(\cdot;\boldsymbol{x}^T\boldsymbol{\beta},F_2)$. Remark 1 shows that regressions remain ordered, e.g., $med(Y \mid \boldsymbol{x}^T\boldsymbol{\beta},F_1) < med(Y \mid \boldsymbol{x}^T\boldsymbol{\beta},F_2)$ and, provided they exist, $E(Y \mid \boldsymbol{x}^T\boldsymbol{\beta},F_1) < E(Y \mid \boldsymbol{x}^T\boldsymbol{\beta},F_2)$.

5. Computational details

With regard to fitting DP mixed models using Markov chain Monte Carlo, the path is well laid out in Escobar and West (1995). However, our context raises some novel wrinkles. We present the details in the case of the two-sample problem. In fact, we assume Gaussian mixands, mixing on the mean.

Formally then, our model is: $X_1,\ldots,X_m$ i.i.d. $F(\cdot;F_1,\sigma^2)$ and $Y_1,\ldots,Y_n$ i.i.d. $F(\cdot;F_2,\sigma^2)$, where $F(\cdot;F_i,\sigma^2) = \int \Phi(\frac{\cdot-\theta}{\sigma})F_i(d\theta)$, $i=1,2; F_1 = G_1 \sim DP(\alpha G_{10})$, $F_2 = G_1G_2$, $G_2 \sim DP(\beta G_{20})$ independent of G_1; $\sigma^2 \sim IG(a,b)$ (an inverse Gamma with mean $b/(a-1)$, provided $a>1$), $G_{10} = N(\mu_1,\tau_1^2)$, $G_{20} = N(\mu_2,\tau_2^2)$; a, b, μ_1, τ_1^2, μ_2, τ_2^2 all known. Also, we assume that α and β are known. If α and β are assumed unknown, Escobar and West show how the inclusion of latent Beta-distributed variables, one associated with α, one with β, simplifies the MCMC sampling.

The Bayesian model is now fully specified. Recall from the previous section that we introduce latent $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_m,\theta_{m+1},\ldots,\theta_{m+n})$ i.i.d. G_1 and $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_n)$ i.i.d. G_2 in order to marginalize over G_1 and G_2. The MCMC algorithm is implemented as a Gibbs sampler to obtain draws from the posterior, $[\boldsymbol{\theta},\boldsymbol{\delta},\sigma^2 \mid D]$ where D denotes the data $X_1,\ldots,X_m,Y_1,\ldots,Y_n$. The required full conditional distributions $[\theta_i \mid \theta_l,\, l\neq i,\, \boldsymbol{\delta},\, \sigma^2,\, D]$, $[\delta_j \mid \delta_l, l\neq j, \boldsymbol{\theta}, \sigma^2, D]$ and $[\sigma^2 \mid \boldsymbol{\theta},\boldsymbol{\delta},D]$ are given in the Appendix.

6. Prior specification

Illustrating with the two-sample model of the previous section, prior specification requires G_{10}, G_{20} and $IG(a,b)$. A sensible, vague prior for σ^2 can usually be gotten by taking $a=2$ (to provide infinite variance) and a mean based upon σ being one-sixth of the anticipated range of the data.

Turning to G_{10} and G_{20} we recall that $EF_1 = G_{10}$ and $EF_2 = G_{10}G_{20}$. Since F_1 generates the θ's, with two elicited features for this distribution, say two quantiles or a center and a range, we obtain two equations to solve for μ_1 and τ_1^2, hence determining G_{10}. But then since F_2 generates the η's, with again two elicited features for the distribution of the η's and G_{10} determined, we can determine G_{20}. For instance, with two quantiles say q_{p_1} and q_{p_2} we set $G_{10}(q_{p_1})G_{20}(q_{p_1}) = p_1$, $G_{10}(q_{p_2})G_{20}(q_{p_2}) = p_2$ which gives $G_{20}(q_{p_1})$ and $G_{20}(q_{p_2})$ hence μ_2 and τ_2^2 and therefore G_{20}. Similarly, a shift relative to G_{10} and a range will again determine G_{20}.

We note that a proper, rather noninformative, specification should not place essentially all of its mass near the boundary, $F_1 = F_2$, as this is, in fact, quite informative. Recalling the parametric analog using the notation of Section 1, a noninformative prior over $-\infty < \theta_1 < \theta_2 < \infty$ would be rather flat over this range and not concentrated near the line $\theta_1 = \theta_2$. Next, though F_1 follows a Dirichlet process, F_2 does not so it will be easier to provide priors for G_1 and G_2. If we start with a fixed c, $G_1(c) \sim Be(\alpha G_{10}(c), \alpha(1-G_{10}(c)))$ and $G_2(c) \sim Be(\beta G_{20}(c), \beta(1-G_{20}(c)))$. Jeffreys'

Table 1. Androstenedione levels for a sample of diabetic men and a sample of diabetic women.

Males		Females		
117	126	84	55	80
123	70	87	77	101
80	63	77	73	66
140	147	84	56	84
115	122	73	112	
135	108	66	56	
49	70	70	134	

Table 2. A descriptive summary of the data.

	Mean	Median	StDev	First quartile	Third quartile
Males	104.64	116.00	31.83	70.00	128.25
Females	79.72	77.00	20.09	66.00	84.75

prior for $G_1(c)$ and $G_2(c)$ is $Be(1/2, 1/2)$ yielding $\alpha = \beta = 1$, $G_{10}(c) = G_{20}(c) = 1/2$. Of course, $G_{i0}(c)$, $i = 1, 2$ can not equal $1/2$ for all c but if we take G_{10} and G_{20} to be normal with common mean roughly at the center of the data and a large variance, then for c's in the range of interest $G_i(c) \dot{\sim} Be(1/2, 1/2)$. We adopt this approach in the example of the next section.

Prior information may more naturally arise on $F(\cdot; F_1, \sigma^2)$ and $F(\cdot; F_2, \sigma^2)$. This may be handled but at a bit of additional computational expense. Suppose we have two quantiles for $F(\cdot; F_1, \sigma^2)$, say $\gamma_{p_1}^{(1)}$ and $\gamma_{p_2}^{(1)}$. With a prior guess for σ say $\tilde{\sigma}$ we can set $p_i = \int \Phi((\gamma_{p_i}^{(1)} - \theta)/\tilde{\sigma})\ G_{10}(d\theta)$, $i = 1, 2$ yielding two integral equations in μ_1 and τ_1^2. These may be solved using analytic methods or perhaps stochastic approximation. Similarly, with two quantiles for $F(\cdot; F_2, \sigma^2)$, say $\gamma_{p_1}^{(2)}$, $\gamma_{p_2}^{(2)}$ we obtain $p_i = \iint \Phi((\gamma_{p_i}^{(2)} - \max(\theta, \delta))/\tilde{\sigma}) G_{10}(d\theta) G_{20}(d\delta)$, $i = 1, 2$ yielding, with μ_1 and τ_1^2 determined, two integral equations in μ_2 and τ_2^2.

7. Data illustration

To illustrate our methodology we consider a dataset providing androstenedione levels for a sample of 14 diabetic men and a sample of 18 diabetic women. The data from Koopmans (1987) appears, slightly modified, in Table 1. The modification is benign. We changed the smallest male measurement to make it immediately evident that stochastic order would not be seen in the pair of empirical c.d.f.'s. It is anticipated that males will produce levels which will tend to be larger than those of females. Descriptive summary of the data appears in Table 2.

The sample sizes are small, encouraging nonparametric modeling. We model the distribution for each population as a location mixture of normals with common variance. That is, we assume

$$F(\cdot; F_i, \sigma^2) = \int \Phi((\cdot - \theta)/\sigma) F_i(d\theta), \qquad i = 1, 2, \tag{7.1}$$

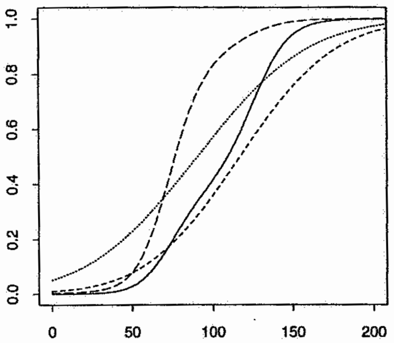

Fig. 1. Prior and posterior means for the c.d.f. functionals $F(c; F_1, \sigma^2)$ and $F(c; F_2, \sigma^2)$. $\hat{E}(F(c; F_1, \sigma^2))$ denoted by the dotted line, $\hat{E}(F(c; F_2, \sigma^2))$ denoted by the smaller dashed line, $\hat{E}(F(c; F_1, \sigma^2) \mid D)$ denoted by the dashed line and $\hat{E}(F(c; F_2, \sigma^2) \mid D)$ denoted by the solid line.

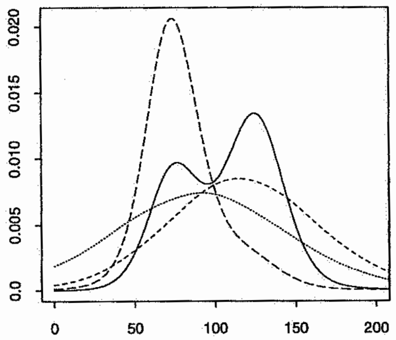

Fig. 2. Prior and posterior means for the p.d.f. functionals $f(c; F_1, \sigma^2)$ and $f(c; F_2, \sigma^2)$. $\hat{E}(f(c; F_1, \sigma^2))$ and $\hat{E}(f(c; F_2, \sigma^2))$ denoted by the dotted and the smaller dashed line, respectively. $\hat{E}(f(c; F_1, \sigma^2) \mid D)$ and $\hat{E}(f(c; F_2, \sigma^2) \mid D)$ denoted by the dashed and the solid line, respectively.

where $i = 1$ denotes the female population, $i = 2$ the males. Under (7.1), σ is not a scale parameter and the two distributions need not have a common variance. As above we set $F_1 = G_1$ and $F_2 = G_1G_2$. Following the suggestions in Section 6, we take $G_1 \sim DP(\alpha G_{10})$, $G_2 \sim DP(\beta G_{20})$ where $G_{10} = N(90, (50)^2)$, $G_{20} = N(90, (50)^2)$ and $\alpha = \beta = 1$. Standard deviations larger than 50 were experimented with revealing negligible change in the resultant posteriors. Also, $\sigma^2 \sim IG(2, 900)$ suggesting a prior mean for σ^2 of $(30)^2$ with infinite variance.

The implications of this prior specification can be determined using the discussion of Section 2. Taking $B = 1,000$, in all Monte Carlo integrations and using a grid of c values, the prior expected c.d.f. of $F(\cdot; F_1, \sigma^2)$ and the prior expected c.d.f. of $F(\cdot; F_2, \sigma^2)$ are plotted in Fig. 1. Also, for the median functional, denoted by $\eta(F)$, a priori, a point and interval estimate for $\eta(F(\cdot; F_1, \sigma^2))$ is 91.566 (13.879,165.968), for $\eta(F(\cdot; F_2, \sigma^2))$ we obtain 116.111 (59.233,178.956). For the interquartile range functional $IQR(F)$, a priori, a point and interval estimate for $IQR(F(\cdot; F_1, \sigma^2))$ is 49.271 (21.776,125.203), for $IQR(F(\cdot; F_2, \sigma^2))$ we obtain 47.810 (22.956,106.018). For the difference $\eta(F(\cdot; F_2, \sigma^2)) - \eta(F(\cdot; F_1, \sigma^2))$ we obtain 15.757 (0.042,104.042).

Turning to the posterior analysis, we can compare the posteriors for the c.d.f.'s $F(c; F_1, \sigma^2)$ and $F(c; F_2, \sigma^2)$. In fact we overlay plots of $\hat{E}(F(c; F_1, \sigma^2) \mid D)$ and $\hat{E}(F(c; F_2, \sigma^2) \mid D)$ on Fig. 1. We can see the Bayesian learning relative to the prior expectation. The posterior curves increase more rapidly implying more concentrated distributions and the separation between the curves becomes noticeably greater on the interval (75,125). Note that such stochastic order will not emerge from the empirical c.d.f.'s (possibly smoothed) for the two samples. Also, were we to adopt stochastic order through a location model with say $F_1(c) = F_0(c - \theta_1)$ and $F_2(c) = F_0(c - \theta_2)$ with F_0 given, this implies the constraint $F_0^{-1}(F_1(c)) - F_0^{-1}(F_2(c)) = \theta_2 - \theta_1$ regardless of c.

The posteriors for the p.d.f.'s, $f(c; F_1, \sigma^2)$, $f(c; F_2, \sigma^2)$ can also be compared again both a priori and a posteriori. We do this in Fig. 2 where we again see the Bayesian learning and the emergence of a bimodality for the males. This latter feature is not surprising upon reexamination of the data but it can not be captured with standard parametric shift models.

Finally, using the median functional, point and interval estimates for the posteriors of $\eta(F(\cdot; F_1, \sigma^2))$, $\eta(F(\cdot; F_2, \sigma^2))$ and $\eta(F(\cdot; F_2, \sigma^2)) - \eta(F(\cdot; F_1, \sigma^2))$ become respectively 76.785 (68.125,87.844), 108.437 (83.863,127.491) and 31.203 (6.002,52.925). We find evidence that median androstenedione level for men is roughly 1.4 times higher than for women.

Appendix

We present the full conditional distributions needed in Section 5. For $i = 1, \ldots, m$, $[\theta_i \mid \theta_l, l \neq i, \boldsymbol{\delta}, \sigma^2, D]$ is a mixed distribution placing point mass $\sigma^{-1}\phi((x_i - \theta_l)/\sigma)/(\sum_{l \neq i} \sigma^{-1}\phi((x_i - \theta_l)/\sigma) + \alpha A(x_i, \sigma^2))$ at $\theta_i = \theta_l$, $l = 1, \ldots, m+n, l \neq i$ and continuous mass $\alpha A(x_i, \sigma^2)/(\sum_{l \neq i} \sigma^{-1}\phi((x_i - \theta_l)/\sigma) + \alpha A(x_i, \sigma^2))$ on the normal distribution $N(\mu_1(x_i, \sigma^2),\ v_1(\sigma^2))$. Here, ϕ denotes the unit normal p.d.f., $\mu_1(x, \sigma^2) = (\sigma^2 \mu_1 + \tau_1^2 x)/(\sigma^2 + \tau_1^2)$, $v_1(\sigma^2) = \sigma^2 \tau_1^2/(\sigma^2 + \tau_1^2)$ and

$$A(x_i, \sigma^2) = \int \sigma^{-1}\phi((x_i - \theta)/\sigma) G_{10}(d\theta) = (\sigma^2 + \tau_1^2)^{-1/2}\phi((x_i - \mu_1)/(\sigma^2 + \tau_1^2)^{1/2}).$$

For θ_{m+j}, $j = 1, \ldots, n$, $[\theta_{m+j} \mid \theta_l, l \neq m+j, \boldsymbol{\delta}, \sigma^2, D]$ is again a mixed distribution placing point mass $\sigma^{-1}\phi((y_j - \max(\theta_l, \delta_j))/\sigma)/(\sum_{l \neq m+j} \sigma^{-1}\phi((y_j - \max(\theta_l, \delta_j))/\sigma) + \alpha B_1(y_j, \delta_j, \sigma^2))$ at $\theta_{m+j} = \theta_l$ and continuous mass $\alpha B_1(y_j, \delta_j, \sigma^2)/(\sum_{l \neq m+j} \sigma^{-1}\phi((y_j - \max(\theta_l, \delta_j))/\sigma) + \alpha B_1(y_j, \delta_j, \sigma^2))$ on the mixture distribution $\{k_1^{(1)}(y_j, \delta_j, \sigma^2) TN(\mu_1, \tau_1^2; \theta_{m+j} < \delta_j) + k_2^{(1)}(y_j, \delta_j, \sigma^2) TN(\mu_1(y_j, \sigma^2), v_1(\sigma^2); \theta_{m+j} > \delta_j)\}/(k_1^{(1)}(y_j, \delta_j, \sigma^2) + k_2^{(1)}(y_j, \delta_j, \sigma^2))$. Here, TN denotes a truncated normal distribution, in particular for θ_{m+j} over the indicated range,

$$\begin{aligned} k_1^{(1)}(y_j, \delta_j, \sigma^2) &= \int_{\theta_{m+j} < \delta_j} \sigma^{-1}\phi((y_j - \delta_j)/\sigma) G_{10}(d\theta_{m+j}) \\ &= \sigma^{-1}\phi((y_j - \delta_j)/\sigma)\Phi((\delta_j - \mu_1)/\tau_1), \qquad \text{and} \\ k_2^{(1)}(y_j, \delta_j, \sigma^2) &= \int_{\theta_{m+j} > \delta_j} \sigma^{-1}\phi((y_j - \theta_{m+j})/\sigma) G_{10}(d\theta_{m+j}) \\ &= (\sigma^2 + \tau_1^2)^{-1/2}\phi((y_j - \mu_1)/(\sigma^2 + \tau_1^2)^{1/2}) \\ &\quad \cdot (1 - \Phi((\delta_j - \mu_1(y_j, \sigma^2))/v_1^{1/2}(\sigma^2))) \end{aligned}$$

and

$$B_1(y_j,\delta_j,\sigma^2) = k_1^{(1)}(y_j,\delta_j,\sigma^2) + k_2^{(1)}(y_j,\delta_j,\sigma^2).$$

For δ_j, $j = 1,\ldots,n$, $[\delta_j \mid \delta_l, l \neq j, \boldsymbol{\theta}, \sigma^2, D]$ is again a mixed distribution placing point mass $\sigma^{-1}\phi((y_j - \max(\theta_{m+j},\delta_l))/\sigma)/(\sum_{l\neq j}\sigma^{-1}\phi((y_j - \max(\theta_{m+j},\delta_l))/\sigma) + \beta B_2(y_j,\theta_{m+j},\sigma^2))$ at $\delta_j = \delta_l$ and continuous mass, $\beta B_2(y_j,\theta_{m+j},\sigma^2)/(\sum_{l\neq j}\sigma^{-1}\phi((y_j - \max(\theta_{m+j},\delta_l))/\sigma) + \beta B_2(y_j,\theta_{m+j},\sigma^2))$ on the mixture distribution

$$\begin{aligned}&\{k_1^{(2)}(y_j,\theta_{m+j},\sigma^2)TN(\mu_2,\tau_2^2;\delta_j < \theta_{m+j})\\ &\quad + k_2^{(2)}(y_j,\theta_{m+j},\sigma^2)TN(\mu_2(y_j,\sigma^2),v_2(\sigma^2);\delta_j > \theta_{m+j})\}\\ &\qquad /(k_1^{(2)}(y_j,\theta_{m+j},\sigma^2) + k_2^{(2)}(y_j,\theta_{m+j},\sigma^2)).\end{aligned}$$

Here,

$$k_1^{(2)} = \sigma^{-1}\phi((y_j - \theta_{m+j})/\sigma)\Phi((\theta_{m+j} - \mu_2)/\tau_2),$$

$$k_2^{(2)} = (\sigma^2+\tau_2^2)^{-1/2}\phi((y_j-\mu_2)/(\sigma^2+\tau_2^2)^{1/2})(1-\Phi((\theta_{m+j}-\mu_2(y_j,\sigma^2))/v_2^{1/2}(\sigma^2)))$$

and

$$B_2(y_j,\theta_{m+j},\sigma^2) = k_1^{(2)}(y_j,\theta_{m+j},\sigma^2) + k_2^{(2)}(y_j,\theta_{m+j},\sigma^2)$$

with $\mu_2(y,\sigma^2) = (\sigma^2\mu_2 + \tau_2^2 y) \;/\; (\sigma^2 + \tau_2^2)$ and $v_2(\sigma^2) = \sigma^2\tau_2^2 \;/\; (\sigma^2 + \tau_2^2)$.

Finally $[\sigma^2 \mid \boldsymbol{\theta}, \boldsymbol{\delta}, D]$ is an update inverse gamma $IG(a + \frac{m+n}{2},\, b + \frac{1}{2}(\sum_i (x_i - \theta_i)^2 + \sum_j (y_j - \max(\theta_{m+j},\delta_j))^2))$.

In practice the above truncated normal distributions are efficiently sampled using the suggestion of Devroye ((1986), p. 38). A moment's reflection reveals that, even if we introduce more than two ordered populations, the full conditional distributions for the DP parameters will still be mixed distributions and the continuous mass will still involve a mixture of at most two distributions. Computation for DP parameters does not worsen with an increasing number of ordered populations. Lastly, note that the Gaussian mixands with the Gaussian base measures provide convenient conjugacies in the foregoing calculations. If $F(\cdot;\theta,\sigma)$ is not normal, Monte Carlo integration may be required. Alternatively, the generic approaches of MacEachern and Müller (1998) or Walker and Damien (1998) can be tried.

References

Akritas, M. G. and Arnold, S. F. (1994). Fully nonparametric hypotheses for factorial designs I: Multivariate repeated measures designs, *J. Amer. Statist. Assoc.*, **89**, 336–343.

Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to nonparametric problems, *Ann. Statist.*, **2**, 1152–1174.

Arjas, E., and Gasbarra, D. (1996). Bayesian inference of survival probabilities, under stochastic ordering constraints, *J. Amer. Statist. Assoc.*, **91**, 1101–1109.

Devroye, L. (1986). *Non-uniform Random Variate Generation*, Springer, New York.

Escobar, M. D., and West, M. (1995). Bayesian density estimation and inference using mixtures, *J. Amer. Statist. Assoc.*, **90**, 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *Ann. Statist.*, **1**, 209–230.

Gelfand, A. E., and Kottas, A. (2001). A computational approach for full nonparametric Bayesian inference in single and multiple sample problems, *J. Comput. Graph. Statist.* (to appear).

Gelfand, A. E., and Mukhopadhyay, S. (1995). On nonparametric Bayesian inference for the distribution of a random sample, *Canad. J. Statist.*, **23**, 411–420.

Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.*, **85**, 398–409.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.

Koopmans, L. H. (1987). *Introduction to Contemporary Statistical Methods*, Duxbury, Belmont, California.

Lehmann, E. (1986). *Testing Statistical Hypotheses*, 2nd ed., Wiley, New York.

Lo, A. Y. (1984). On a class of bayesian nonparametric estimates: I. Density estimates, *Ann. Statist.*, **12**, 351–357.

MacEachern, S. N., and Müller, P. (1998). Estimating mixture of Dirichlet process Models, *J. Comput. Graph. Statist.*, **7**, 223–238.

Mukhopadhyay, S., and Gelfand, A. E. (1997). Dirichlet process mixed generalized linear models, *J. Amer. Statist. Assoc.*, **92**, 633–639.

Randles, R. H., and Wolfe, D. A. (1979). *Introduction to The Theory of Nonparametric Statistics*, Wiley, New York.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors, *Statistica Sinica*, **4**, 639–650.

Shaked, M., and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications*, Academic Press, Boston.

Walker, S. G., and Damien, P. (1998). Sampling Methods for Bayesian Nonparametric Inference Involving Stochastic Processes, *Practical Nonparametric and Semiparametric Bayesian Statistics* (eds. D. Dey, P. Müller and D. Sinha), 243–254, Springer, New York.

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), *J. Roy. Statist. Soc. Ser. B*, **61**, 485–527.