Modeling Variability Order: A Semiparametric Bayesian Approach

Athanasios Kottas and Alan E. Gelfand*

November 8, 2000

Abstract

In comparing two populations, sometimes a model incorporating a certain probability order is desired. In this setting, Bayesian modeling is attractive since a probability order restriction imposed a priori on the population distributions is retained a posteriori. Extending the work in Gelfand and Kottas (2000) for stochastic order specifications, we formulate modeling for distributions ordered in variability. We work with Dirichlet process mixtures resulting in a fully Bayesian semiparametric approach. The details for simulation-based model fitting and prior specification are provided. An example, based on two small subsets of time intervals between eruptions of the Old Faithful geyser, illustrates the methodology.

KEY WORDS: Dirichlet process mixing; dispersion ordering; Markov chain Monte Carlo; sign changes.

AMS 1991 Subject Classification: 60E05, 62E10, 62G05.

---

[*] A. Kottas is Visiting Assistant Professor in the Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251, USA and A.E. Gelfand is Professor in the Department of Statistics, University of Connecticut, Storrs, CT 06269-3120, USA. This work is part of A. Kottas' Ph.D. thesis completed in the Department of Statistics, University of Connecticut under the supervision of A.E. Gelfand. The work of the second author was supported in part by NSF grant DMS 99-71206.

1 Introduction

Probability order restrictions are often appropriate when comparing two or more populations. Such restrictions presume strong conditions on the population probability measures. However, if the samples are small, the probability order constraint of interest may not hold for the empirical distributions of these samples. The Bayesian paradigm provides a convenient framework for the development of related modeling since any probability ordering postulated a priori is preserved to the posterior analysis.

The most common probability order is stochastic order. We call two distribution functions F_1 and F_2 satisfying $F_1(u) \geq F_2(u)$, for all u, stochastically ordered and denote it by $F_1 \leq_{st} F_2$. Bayesian nonparametric modeling for random stochastically ordered distribution functions has been discussed by Arjas and Gasbarra (1996) and Gelfand and Kottas (2000). For survival models, Arjas and Gasbarra specify a random pair of piecewise hazard functions obeying a partial ordering which implies stochastic order for the associated distributions. Gelfand and Kottas employ Dirichlet process mixtures to model the distribution functions. By imposing certain restrictions on the kernel and the mixing distributions stochastic ordering for the resulting random mixtures is ensured.

Another useful probability order concept is variability order. Roughly, two distributions, F_1 and F_2, are ordered in variability if, given a fixed location, mass accumulates more rapidly as we move away from this location, for say F_1 relative to F_2, in either direction. In fact, there is no unique definition for variability ordering. Shaked and Shanthikumar (1994) note at least three possibilities: convex order, dispersive order and peakedness order. We adopt a natural and flexible definition based upon a sign changes condition. Under this definition we develop semiparametric Dirichlet process mixture models for variability ordering, extending the ideas in Gelfand and Kottas (2000). Full comparative inference can be implemented employing the approach in Gelfand and Kottas (1999). Dirichlet process mixing provides a rich and computationally feasible framework for Bayesian nonparametric inference. We use it as the stochastic mechanism under which variability order is modeled.

The format of the paper is as follows. A brief review of Dirichlet process mixing and inference under such specifications is given in section 2. In section 3 we summarize the literature associated with variability ordering and discuss relations between the existing definitions. In the process we identify a condition, based on the notion of sign changes of a function, which can be used as a natural definition of variability order. Section 4 presents our modeling approach for variability ordered symmetric distributions. Section 5 discusses computational implementation and prior specification for the model. In section 6 we provide a data illustration. Finally, section 7 demonstrates how the methodology can be extended to handle skewed distributions ordered in variability.

2 Dirichlet process mixing

A distribution G on Θ follows a Dirichlet process $DP(\nu G_0)$ if, given an arbitrary finite measurable partition, B_1,...,B_r of Θ, the joint distribution of $(G(B_1),...,G(B_r))$ is $Dirichlet(\nu G_0(B_1),...,\nu G_0(B_r))$ where $G(B_i)$ and $G_0(B_i)$ denote the probability of set B_i under G and G_0, respectively. Here, G_0 is a specified distribution on Θ and $\nu > 0$ is a precision parameter. See Ferguson (1973, 1974) for the formal development of Dirichlet processes along with earlier references.

Let $F(\cdot;\theta)$ be a parametric family of distribution functions, indexed by $\theta \in \Theta$, with associated densities, $f(\cdot;\theta)$. If G is proper we define the mixture distribution

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

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

If G is random say $G \sim DP(\nu G_0)$, then $F(\cdot;G)$ is random. Letting $D = \{Y_i,\, i = 1,...,n\}$ denote a sample from $F(\cdot;G)$ and using the bracket notation of Gelfand and Smith (1990), we write its posterior as $[F(\cdot;G) \mid D]$. Functionals of $F(\cdot;G)$, for which we use the generic notation $H(F(\cdot;G))$, are of interest with posteriors denoted by $[H(F(\cdot;G)) \mid D]$.

In the context of (2.1), suppose for each Y_i, $i = 1,...,n$ we introduce a latent θ_i and assume that the Y_i's are conditionally independent given the θ_i's. Assume further that the θ_i's are conditionally independent and identically distributed given G. As a result the Y_i's are marginally independent, with joint density $\prod_{i=1}^{n} f(y_i;G) = \prod_{i=1}^{n} \int f(y_i;\theta_i)\, G(d\theta_i)$.

Adding $G \sim DP(\nu G_0)$ completes the Bayesian model specification, apart, perhaps, from a hyperprior on ν (see Escobar and West, 1995). Such Dirichlet process mixed models were originally studied by Antoniak (1974) and Lo (1984). In particular, Antoniak (1974) noted that this Bayesian model can be *marginalized* over G to obtain $\prod_{i=1}^{n} f(y_i;\theta_i)[\theta_1,...,\theta_n \mid G_0,\nu]$. After marginalization the θ_i are no longer independent but an MCMC algorithm can be implemented (Escobar, 1994, Escobar and West, 1995 or MacEachern and Müller, 1998) to obtain samples from the posterior $[\theta_1,...,\theta_n \mid D]$.

Gelfand and Mukhopadhyay (1995) describe how to use these samples to infer about linear functionals associated with $F(\cdot;G)$. They show how posterior expectations of linear functionals and products of linear functionals can be computed. Restriction to posterior moments of linear functionals severely limits inference. Gelfand and Kottas (1999) provide a computational approach to obtain the entire posterior distribution for more general functionals. Briefly, note that for a linear functional H, $H(F(\cdot;G)) = \int H(F(\cdot;\theta_0))G(d\theta_0)$. Now, instead of marginalizing over G in $[\theta_0,\theta,G \mid D] \propto [D \mid \theta][\theta_0,\theta \mid G][G]$, observe that this joint posterior is proportional to $[\theta_0 \mid G][G \mid \theta][\theta \mid D]$. Hence given the posterior sample θ_b^*, $b = 1,...,B$, for each θ_b^* draw $G_b^* \sim [G \mid \theta_b^*]$ and then $\theta_{0lb}^* \sim G_b^*$, for $l = 1,...,L$. Finally, $H_b^* = L^{-1}\sum_{l=1}^{L} H(F(\cdot;\theta_{0lb}^*))$ is a Monte Carlo integration for a realization from $[H(F(\cdot;G)) \mid D]$. To obtain an approximate realization from $[G \mid \theta_b^*]$, 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 Dirichlet process 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 (1999) can be applied a priori by approximately sampling $[G]$ rather than $[G \mid D]$. If we write $\theta = (\theta^{(1)},\theta^{(2)})$, we might place a Dirichlet process prior on $\theta^{(1)}$, i.e., $\theta^{(1)} \sim G$ where $G \sim DP(\nu G_0)$ with a parametric prior on $\theta^{(2)}$, yielding $F(\cdot;G,\theta^{(2)}) = \int F(\cdot;\theta^{(1)},\theta^{(2)})G(d\theta^{(1)})$, a semiparametric specification.

3 A review of variability orders

Three definitions relating to the general concept of variability ordering, mentioned in the introduction, can be found in the literature. We refer to Shaked and Shanthikumar (1994, chapter 2) for a comprehensive review. Hereafter, let X and Y be two random variables with distribution functions F_1 and F_2, density functions, assuming they exist, f_1 and f_2, and generalized inverses F_1^{-1} and F_2^{-1}, respectively. (Here, $F_i^{-1}(u) = \inf\{x : F_i(x) \geq u \}$, i =1,2.)

We say that Y is larger than X in the convex order (denoted by $X \leq_{cx} Y$, or $F_1 \leq_{cx} F_2$) if $\int \varphi(u)dF_1(u) \leq \int \varphi(u)dF_2(u)$, for every convex function $\varphi : R \to R$. The definition and the convexity of $\varphi_1(u) = u$, $\varphi_2(u) = -u$ yield that if $X \leq_{cx} Y$ then $EX = EY$, provided the expectations exist. Furthermore from the convexity of $\varphi_3(u) = u^2$ it follows that $X \leq_{cx} Y$ implies $Var(X) \leq Var(Y)$, when $Var(Y) < \infty$. See Schweder (1982) for earlier references and a result on convex ordering of certain mixture distributions.

The random variable Y is said to be larger than X in the dispersive order (for which we use the notation $X \leq_{disp} Y$) if $F_1^{-1}(q) - F_1^{-1}(p) \leq F_2^{-1}(q) - F_2^{-1}(p)$, for all $0 \leq p \leq q \leq 1$. Again, it can be proved that $X \leq_{disp} Y$ implies $Var(X) \leq Var(Y)$, provided $Var(Y) \leq \infty$. This ordering was first considered by Saunders and Moran (1978) and Bickel and Lehmann (1979). Lewis and Thompson (1981), Oja (1981) and Shaked (1982) provide further detail.

Finally, if X and Y have symmetric distributions about μ_1 and μ_2, respectively, Y is larger than X in the peakedness order (denoted by $X \leq_{peak} Y$) if $|Y - \mu_2|$ is stochastically larger than $|X - \mu_1|$. Birnbaum (1948) introduced peakedness order and Bickel and Lehmann (1976) used it to define measures of dispersion for symmetric distributions.

The notion of sign changes of a function provides a convenient framework for the study of conditions that imply the above variability orders. Following Karlin (1968, p. 20), the number of sign changes of a function h defined in $U \subseteq R$, is given by $S^-(h) = \sup S^-(h(u_1),...,h(u_m))$ where the supremum is extended over all sets $u_1 \leq ... \leq u_m$, $u_i \in U$, with m arbitrary but finite and $S^-(t_1,...,t_m)$ is the number of sign changes of the indicated sequence, zero terms being discarded. Regarding the convex order, consider the following conditions,

(C1) $EX = EY$ and $S^-(f_2 - f_1) = 2$, the sign sequence being $+, -, +$.
(C2) $EX = EY$ and $S^-(F_2 - F_1) = 1$, the sign sequence being $+, -$.
(C3) $EX = EY$ and $\int_{-\infty}^t (F_2(u) - F_1(u\ ))du \geq 0$, for all t.
(C4) $X \leq_{cx} Y$.
Then we have (C1) $\Rightarrow$ (C2) $\Rightarrow$ (C3) $\Leftrightarrow$ (C4) (see, e.g., Shaked, 1980, Schweder, 1982 and Shaked and Shanthikumar, 1994 for more details and earlier references). With regard to dispersive ordering, Shaked (1982) proved various related conditions. In particular, for F_1 and F_2 which are strictly increasing and continuous on their supports, he established the equivalence of the conditions,
(D1) $S^-(F_1(\cdot - c) - F_2(\cdot)) \leq 1$, for any $c \in R$, the sign sequence being $-, +$ in the case of equality.
(D2) $X \leq_{disp} Y$.
Combining (C2), (C4) and (D1), (D2) it follows that if X and Y have finite first moment, $X \leq_{disp} Y$ implies $X - EX \leq_{cx} Y - EY$. Moreover, Shaked (1982) notes that $X \leq_{disp} Y$ implies $X \leq_{peak} Y$, when X and Y have symmetric distributions about 0. Another relation, between (C2) and the peakedness order, can also be easily obtained. If F_1 and F_2 are symmetric about the same mean, which, without loss of generality, can be taken to be 0, then (C2) $\Leftrightarrow$ $X \leq_{peak} Y$. We omit the straightforward proof.

The foregoing discussion reveals that dispersive order is the strongest among these variability orders and also peakedness order implies convex order within the class of symmetric distributions. Although it is possible to formulate modeling for dispersive ordered distributions, we propose adopting a condition similar to (C2), assuming equal locations (where for us location is either the median or perhaps the mode under unimodality), as the definition of variability order. This condition allows for skewness in the distributions, implies convex order and peakedness order (in the special case of symmetry) and quantifies the notion of variability order in a natural way. Furthermore, dispersive order has been most sucessful in comparing distributions with support on R^+, while our objective is modeling for real valued, possibly asymmetric, variability ordered random variables. In this context it seems natural to assume equal locations for the distributions, which can be obtained possibly after a shift as in the definition of peakedness order. We will therefore say that Y is larger than X in variability order, and write $F_1 \leq_{var} F_2$, if F_1 and F_2 have the same location and $S^-(F_2 - F_1) = 1$, the sign sequence being $+, -$. We could equally well have defined variability order in terms of the stronger condition (C1) but the definition above suffices for our purposes.

4 Semiparametric mixture modeling for variability order

Section 4.1 develops the methodology while section 4.2 discusses applications.

4.1 Modeling approach

We develop a modeling approach for joint distributions over the space $\mathcal{G} = \{(F_1,F_2) : F_1 \leq_{var} F_2\}$ employing mixtures of the generic form (2.1). Here we focus on symmetric distributions, discussing one possible way to introduce skewness in section 7. The following result forms the basis of our method.

Lemma 1 Consider a parametric family of distributions $\{F(\cdot;\theta) : \theta \in (\underline{\theta}, \overline{\theta}) \subseteq R\}$, with $F(\cdot;\theta)$ differentiable in θ, such that for any $\theta_1 < \theta_2$, $S^-(F(\cdot;\theta_2) - F(\cdot;\theta_1)) = 1$, the sign sequence being $+, -$, where the crossing point u_0 of $F(\cdot;\theta_2)$ and $F(\cdot;\theta_1)$ is the same for all θ_1 and θ_2. For distributions F_1 and F_2 on $(\underline{\theta}, \overline{\theta})$ define the mixtures $F(\cdot;F_i) = \int_{\underline{\theta}}^{\overline{\theta}} F(\cdot;\theta)F_i(d\theta)$, $i = 1, 2$. Then if $F_1 \leq_{st} F_2$, $S^-(F(\cdot;F_2) - F(\cdot;F_1)) = 1$ and the sign sequence is $+, -$.

Proof: By assumption, $F(u;\theta)$ is non-decreasing in θ for any fixed $u \leq u_0$ and non-increasing in θ for $u > u_0$. Hence for an arbitrary fixed $u \leq u_0$,

$$\begin{aligned} F(u;F_2) - F(u;F_1) &= \int_{\underline{\theta}}^{\overline{\theta}} F(u;\theta) d(F_2(\theta) - F_1(\theta)) \\ &= \int_{\underline{\theta}}^{\overline{\theta}} (F_1(\theta) - F_2(\theta)) \frac{dF(u;\theta)}{d\theta} d\theta \geq 0, \end{aligned}$$

while for any $u > u_0$ the same argument yields $F(u;F_2) - F(u;F_1) \leq 0$. Therefore $F(\cdot;F_2)$ and $F\ (\cdot;F_1)$ have exactly one crossing point, the sign sequence being $+, -$.

Lemma 1 indicates how variability order can be preserved through mixing. The assumption of a common crossing point for the kernel of the mixtures will be satisfied if we place some restriction on the location. In particular, if $F(\cdot;\theta)$ is a symmetric family, with

$\theta^{1/2}$ a scale parameter, u_0 will be the center of symmetry. Moreover, the assumption of differentiability can be relaxed which implies that the kernel of the mixtures can be taken to be any symmetric scale family $\{F(\frac{\cdot}{\theta^{1/2}}) : \theta \in R^+\}$. For this choice, the distributions $F(\cdot; F_i)$, $i =$ 1,2 have medians and means, provided they are finite, equal to 0. Hence from Lemma 1 we get that $(F(\cdot; F_1), F(\cdot; F_2)) \in \mathcal{G}$. Next we need some modeling specification for the mixing distributions F_1 and F_2 belonging in the space $\mathcal{P} = \{(F_1,F_2) : F_1 \leq_{st} F_2\}$. Such a specification for arbitrary members of the generic space $\mathcal{P}$ is very difficult, if at all possible, to obtain. As in Gelfand and Kottas (2000), we consider the easier to work with subspace of $\mathcal{P}$, $\mathcal{P}^{'} = \{(F_1,F_2) : F_1 = G_1,\ F_2 = G_1 G_2\}$ where G_1 and G_2 are distribution functions. 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 $\delta \sim G_2$. Assuming that the pair (F_1,F_2) of mixing distributions arises from $\mathcal{P}^{'}$, Lemma 1 again implies that the resulting class $\mathcal{G}_F = \{(F(\cdot; F_1), F(\cdot; F_2)) : (F_1, F_2) \in \mathcal{P}^{'}\} \subset \mathcal{G}$, where F denotes the choice of the symmetric scale family for the kernel. This class can be enriched by adding a location parameter μ to $F(\frac{\cdot}{\theta^{1/2}})$, retaining variability order. The semiparametric specification is completed by choosing a Gaussian kernel for the mixtures. Denoting by $\Phi(\cdot)$ the standard normal distribution function, we finally have

$$F\left(\cdot; \mu, F_1\right) = \int \Phi((\,\cdot - \mu)/\theta^{1/2}) G_1\left(d\theta\right) \tag{4.1}$$

while

$$F(\cdot; \mu, F_2) = \iint \Phi((\,\cdot - \mu)/\max\left(\theta^{1/2}, \delta^{1/2}\right)) G_1(d\theta) G_2(d\delta), \tag{4.2}$$

with $F(\cdot; \mu, F_1) \leq_{var} F(\cdot; \mu, F_2)$. Full inference with this model, assuming independent Dirichlet process priors for G_1 and G_2, is discussed in section 5 and illustrated in section 6. Note that under (4.1) and (4.2) both $F(\cdot; \mu, F_i)$ are symmetric and unimodal with μ, the common location, being their median and mode.

In view of the general characterization of symmetric unimodal distributions as scale mixtures of symmetric uniforms (see, e.g., Feller, 1971, p. 158), one might suggest the use of mixtures with a symmetric uniform kernel. However, for this choice Lemma 1 is no longer applicable since the parameter space of the kernel depends on the sample space. As an alternative, Gaussian kernels emerge as natural and convenient to work with.

Finally, we provide two additional results, similar in spirit to Lemma 1, that enable modeling variability order through mixtures of the form $F(\cdot; F_i) = \int_{\underline{\theta}}^{\overline{\theta}} F(\cdot; \theta) F_i(d\theta)$, $i = 1, 2$, where again the parameter space of the kernel is free of the sample space. Lemma 2 below generalizes Lemma 1 by considering sign changes of the densities instead of the distribution functions.

Lemma 2 Assume that the parametric family of densities $\{f(\cdot; \theta) : \theta \in (\underline{\theta}, \overline{\theta}) \subseteq R\}$ satisfies for $\theta_1 < \theta_2$, $S^-(f(\cdot; \theta_2) - f(\cdot; \theta_1)) = 2$, with the sign sequence $+, -, +$, where the crossing points u_1 and u_2 of $f(\cdot; \theta_2)$ and $f(\cdot; \theta_1)$ do not depend on θ_1 and θ_2. Moreover, $f(\cdot; \theta)$ is taken to be differentiable in θ. Then for the densities $f(\cdot; F_i) = \int_{\underline{\theta}}^{\overline{\theta}} f(\cdot; \theta) F_i(d\theta)$ of $F(\cdot; F_i)$, $i =$ 1,2 we have $S^-(f(\cdot; F_2) - f(\cdot; F_1)) = 2$, the sign sequence being $+, -, +$, provided $F_1 \leq_{st} F_2$.

Proof: Given the conditions of the lemma, $f(u; \theta)$, as a function of θ, is non-decreasing for any fixed $u \leq u_1$ or $u > u_2$ and non-increasing when $u_1 < u \leq u_2$, where $u_1 < u_2$ are points on the support of the distribution associated with $f(\cdot; \theta)$. Now

$$\begin{aligned} f(u; F_1) - f(u; F_2) &= \int_{\underline{\theta}}^{\overline{\theta}} f(u; \theta) d(F_1(\theta) - F_2(\theta)) \\ &= \int_{\underline{\theta}}^{\overline{\theta}} (F_2(\theta) - F_1(\theta)) \tfrac{df(u;\theta)}{d\theta} d\theta, \end{aligned}$$

which implies that $f(\cdot; F_2)$ and $f(\cdot; F_1)$ have two crossing points and the sign sequence is $+, -, +$.

The final lemma handles convex ordering of the mixtures $F(\cdot; F_i)$, $i =$ 1,2, providing in this direction a different result than that of Schweder (1982, p. 166).

Lemma 3 For the parametric family of distributions $\{F(\cdot; \theta) : \theta \in (\underline{\theta}, \overline{\theta}) \subseteq R\}$ assume that $F(\cdot; \theta_1) \leq_{cx} F(\cdot; \theta_2)$, when $\theta_1 < \theta_2$, and that $\psi(\theta) = \int \varphi(u) dF(u; \theta)$ is differentiable in θ, where $\varphi : R \rightarrow R$ is convex. Then if $F_1 \leq_{st} F_2$, $F(\cdot; F_1) \leq_{cx} F(\cdot; F_2)$.

Proof: Consider an arbitrary convex function $\varphi : R \rightarrow R$ for which $\int \varphi(u) dF(u; F_i) < \infty$, $i =$ 1,2, a non-restrictive assumption given the definition of convex order. By assumption, the function $\psi(\theta)$ is non-decreasing in θ. Hence, using Fubini's theorem,

$$\begin{aligned}\int \varphi(u)dF(u;F_1) - \int \varphi(u)dF(u;F_2) &= \int_{\underline{\theta}}^{\overline{\theta}} \psi(\theta)d(F_1(\theta) - F_2(\theta)) \\ &= \int_{\underline{\theta}}^{\overline{\theta}} (F_2(\theta) - F_1(\theta))\frac{d\psi(\theta)}{d\theta}d\theta \leq 0.\end{aligned}$$

4.2 Applications

Here, we present some applications of variability ordered distributions modeled as in the previous section.

The simplest setting is the two-sample problem for variability ordered distributions with common location. Here we have samples of size m and n from $F(\cdot;\mu,F_1)$ and $F(\cdot;\mu,F_2)$, respectively, with $F(\cdot;\mu,F_i)$, $i = 1,2$ defined in (4.1) and (4.2) whence $F(\cdot;\mu,F_1)$ $\leq_{var}$ F $(\cdot;\mu,F_2)$. Interest lies on the posterior predictive distribution functions and densities as well as on comparison of dispersion functionals. Since expectations need not exist when working with mixtures, the interquartile range functional seems to be the appropriate choice. Inference for the common location μ can also be obtained. In order to fit the model we add $m + n$ latent variables being i.i.d. G_1 and n latent i.i.d. G_2. All the details are provided in section 5 with an illustrative data analysis appearing in section 6.

The extension to different locations for the two ordered populations is straightforward. We consider $X_i = \mu_1 + \varepsilon_i$, $i = 1,...,m$ and $Y_j = \mu_2 + \epsilon_j$, $j = 1,...,n$ and then model the ε_i's and ϵ_j's as i.i.d. realizations from (4.1) and (4.2), respectively, taking $\mu = 0$ as their common location. More generally, we might consider two distinct datasets, both to be modeled using the same regression model, where we anticipate that the error distributions exhibit variability ordering. In this case we would replace μ_1 by μ_{1i}, $i = 1,...,m$ and μ_2 by μ_{2j}, $j = 1,...,n$.

5 Implementation details

Here, we provide computational details for the two-sample problem with common location. Extensions to handle the other settings discussed in section 4.2 are then straightforward. Assuming $X_1,...,X_m$ i.i.d. from $F(\cdot;\mu,F_1)$ in (4.1) and $Y_1,...,Y_n$ i.i.d. $F(\cdot;\mu,F_2)$ defined in (4.2), with independent Dirichlet process priors for the mixing distributions G_1 and G_2, the following hierarchical model emerges

$$\begin{array}{rcc}
X_i \mid \mu, \theta_i & \overset{ind.}{\sim} & f_N(\cdot \mid \mu, \theta_i),\ i = 1, ..., m \\
Y_j \mid \mu, \theta_{m+j}, \delta_j & \overset{ind.}{\sim} & f_N(\cdot \mid \mu, \max\left(\theta_{m+j}, \delta_j\right)),\ j = 1, ..., n \\
\theta_i \mid G_1 & \overset{i.i.d.}{\sim} & G_1,\ i = 1, ..., m+n \\
\delta_j \mid G_2 & \overset{i.i.d.}{\sim} & G_2,\ j = 1, ..., n \\
G_r & \sim & DP(\nu_r G_{r0})\ ,\ r = 1, 2 \\
\mu & \sim & N(\lambda, \tau^2),
\end{array} \tag{5.1}$$

where $f_N(\cdot \mid \mu, \theta)$ denotes the density of a $N(\mu, \theta)$ distribution and the base distributions G_{r0}, r = 1,2 are taken to be inverse Gamma distributions, $IGamma(a_r, b_r)$ (with means $b_r/(a_r - 1)$, provided $a_r > 1$). The hyperparameters λ, τ^2, ν_r, a_r, b_r, r = 1,2 are all assumed fixed. The introduction of the additional latent variables θ_{m+j}, j = 1,...,n enables the marginalization over G_r, r = 1,2, noted in section 3. The resulting posterior $[\mu, \theta, \delta \mid D]$, where $\theta = (\theta_1,...,\theta_m,\theta_{m+1},...,\theta_{m+n})$, $\delta = (\delta_1,...,\delta_n)$ and $D = \{x_i, y_j, i = 1,...,m, j = 1,...,n\}$, can be sampled employing Gibbs sampling. We next provide the associated full conditional distributions.

Full conditional for μ: A standard calculation yields that $[\mu \mid \theta, \delta, D]$ is an updated normal with variance $\widetilde{\tau}^2 = \{\tau^{-2} + \sum_{i=1}^m \theta_i^{-1} + \sum_{j=1}^n (\max(\theta_{m+j}, \delta_j))^{-1}\}^{-1}$ and mean $\widetilde{\lambda} = \widetilde{\tau}^2\{\lambda\tau^{-2} + \sum_{i=1}^m x_i\theta_i^{-1} + \sum_{j=1}^n y_j(\max(\theta_{m+j}, \delta_j))^{-1}\}$.

Full conditional for θ_i, i = 1,...,m: Following Escobar and West (1995), this is a mixed distribution placing point mass $f_N(x_i \mid \mu, \theta_l)/(\nu_1 A(x_i, \mu) + \sum_{k\neq i} f_N(x_i \mid \mu, \theta_k))$ at $\theta_i = \theta_l$, l = 1,...,$m+n$, $l \neq i$ and continuous mass $\nu_1 A\ (x_i, \mu)/(\nu_1 A(x_i, \mu) + \sum_{k\neq i} f_N(x_i \mid \mu, \theta_k))$ on the $IGamma(a_1 + \frac{1}{2},\ b_1 + \frac{1}{2}(x_i - \mu)^2)$ distribution. Here,

$$A(x_i, \mu) = \frac{b_1^{a_1}\Gamma(a_1 + \frac{1}{2})}{(2\pi)^{1/2}\Gamma(a_1)\{b_1 + \frac{1}{2}(x_i - \mu)^2\}^{a_1 + \frac{1}{2}}},$$

where $\Gamma(\cdot)$ is the Gamma function.

Full conditional for θ_{m+j}, j = 1,...,n: This is again a mixed distribution with point masses $f_N(y_j \mid \mu, \max(\theta_l, \delta_j))/(\nu_1 B_1(y_j, \mu, \delta_j) + \sum_{k\neq m+j} f_N(y_j \mid \mu, \max(\theta_k,\ \delta_j)))$ at $\theta_{m+j} = \theta_l$, l = 1,...,$m+n$, $l \neq m+j$ and the remaining mass on the mixture distribution $(\xi_1^{(1)}(y_j, \mu, \delta_j)\ TIGamma(a_1,\ b_1;\ \theta_{m+j} < \delta_j) + \xi_2^{(1)}(y_j, \mu, \delta_j)\ TIGamma(a_1 + \frac{1}{2},\ b_1 +$

$\frac{1}{2}(y_j - \mu)^2;\ \theta_{m+j} > \delta_j))\ /\ (\xi_1^{(1)}(y_j, \mu, \delta_j) + \xi_2^{(1)}(y_j, \mu, \delta_j))$. Here, $TIGamma(a,\ b;\ \cdot \in V)$ denotes a truncated inverse Gamma distribution over the set V. The weights are given by $\xi_1^{(1)}(y_j, \mu, \delta_j) = \int_{\theta_{m+j} < \delta_j} f_N(y_j \mid \mu, \delta_j) G_{10}(d\theta_{m+j}) = f_N(y_j \mid \mu, \delta_j) F_{IGamma}(\delta_j \mid a_1, b_1)$ and $\xi_2^{(1)}(y_j, \mu, \delta_j) = \int_{\theta_{m+j} > \delta_j} f_N(y_j \mid \mu, \theta_{m+j}) G_{10}(d\theta_{m+j}) = A(y_j, \mu)(1 - F_{IGamma}(\delta_j \mid a_1 + \frac{1}{2},$ $b_1 + \frac{1}{2}(y_j - \mu)^2))$, where $F_{IGamma}(\cdot \mid a, b)$ is the distribution function of an $IGamma(a, b)$ distribution. Finally, $B_1(y_j, \mu, \delta_j) = \xi_1^{(1)}(y_j, \mu, \delta_j) + \xi_2^{(1)}(y_j, \mu, \delta_j)$.

Full conditional for δ_j, $j = 1,...,n$: Again, we have a mixed distribution that places point mass $f_N(y_j \mid \mu, \max(\theta_{m+j}, \delta_l))/(\nu_2 B_2(y_j, \mu, \theta_{m+j}) + \sum_{k \neq j} f_N(y_j \mid \mu, \max(\theta_{m+j},\ \delta_k)))$ at $\delta_j = \delta_l$ for $l = 1,...,n$, $l \neq j$ and the remaining mass on the mixture $(\xi_1^{(2)}(y_j, \mu, \theta_{m+j})$ $TIGamma(a_2,\ b_2;\ \delta_j < \theta_{m+j}) + \xi_2^{(2)}(y_j, \mu, \theta_{m+j})\ TIGamma(a_2 + \frac{1}{2},\ b_2 + \frac{1}{2}(y_j - \mu)^2;$ $\delta_j > \theta_{m+j}))\ /\ (\xi_1^{(2)}(y_j, \mu, \theta_{m+j}) + \xi_2^{(2)}(y_j,\ \mu,\ \theta_{m+j}))$. Here, $\xi_1^{(2)}(y_j, \mu, \theta_{m+j}) = f_N(y_j \mid \mu, \theta_{m+j}) F_{IGamma}(\theta_{m+j} \mid a_2, b_2)$ and

$$\xi_2^{(2)}(y_j, \mu, \theta_{m+j}) = \frac{b_2^{a_2} \Gamma(a_2 + \frac{1}{2})(1\ -\ F_{IGamma}(\theta_{m+j} \mid a_2 + \frac{1}{2}, b_2 + \frac{1}{2}(y_j - \mu)^2))}{(2\pi)^{1/2} \Gamma(a_2) \{b_2 + \frac{1}{2}(y_j - \mu)^2\}^{a_2 + \frac{1}{2}}},$$

with $B_2(y_j, \mu, \theta_{m+j}) = \xi_1^{(2)}(y_j, \mu, \theta_{m+j}) + \xi_2^{(2)}(y_j, \mu, \theta_{m+j})$.

The samples $(\mu_b^*, \theta_b^*, \delta_b^*)$, $b = 1,...,B$ from the posterior of the model, obtained using the Gibbs sampler described above, can be used to infer about functionals of $F(\cdot; \mu, F_i)$, $i = 1,2$ in the spirit of Gelfand and Kottas (1999). Under the model specification (5.1), for any linear functional H we obtain $H(F(\cdot; \mu, F_1)) = \int H(F(\cdot; \mu, \theta_0)) G_1(d\theta_0)$ and

$$H(F(\cdot; \mu, F_2)) = \int \int H(F(\cdot; \mu, \max(\theta_0, \delta_0))) G_1(d\theta_0) G_2(d\delta_0).$$

Observing that $[\mu, \theta, \delta, G_1, G_2 \mid D] \propto [G_1 \mid \theta][G_2 \mid \delta][\mu, \theta, \delta \mid D]$, for each b we draw $G_{1b}^* \sim [G_1 \mid \theta_b^*]$, $G_{2b}^* \sim [G_2 \mid \delta_b^*]$, and next θ_{0lb}^* and δ_{0lb}^* from G_{1b}^* and G_{2b}^*, respectively, for $l = 1,...,L$. Then $L^{-1} \sum_{l=1}^{L} H(F(\cdot; \mu_b^*, \theta_{0lb}^*))$ and $L^{-1} \sum_{l=1}^{L} H(F(\cdot; \mu_b^*, \max(\theta_{0lb}^*, \delta_{0lb}^*)))$ are draws from the posteriors of $H(F(\cdot; \mu, F_1))$ and $H(F(\cdot; \mu, F_2))$, respectively. If, for $b = 1,...,B$, we generate $\mu_b^{**} \sim N(\lambda, \tau^2)$ and $G_{rb}^{**} \sim DP(\nu_r G_{r0})$, $r = 1,2$, draw $\theta_{0lb}^{**} \sim G_{1b}^{**}$, $\delta_{0lb}^{**} \sim G_{2b}^{**}$, $l = 1,...,L$ and replace μ_b^*, θ_{0lb}^*, δ_{0lb}^* with μ_b^{**} , θ_{0lb}^{**}, δ_{0lb}^{**} in the summations above we obtain realizations from the priors of $H(F(\cdot; \mu, F_i))$, $i = 1,2$. Sampling, for each population, from the posterior (prior) of the "distribution function-at-a-point" functional we can invert to get samples from the posterior (prior) of any quantile functional. Hence the posteriors (priors) of the interquartile range functionals $IQR(F(\cdot;\mu,F_i))$, $i = 1{,}2$ can be obtained and compared.

With regard to prior choice, we need to specify the hyperparameters λ, τ^2, ν_r, a_r, b_r, $r = 1{,}2$. The prior for the common location μ can be fully specified by setting λ equal to a prior guess for the center of the populations and taking large variance τ^2. The parameters of G_{r0}, $r = 1{,}2$ can be determined using available prior information on $F(\cdot;\mu,F_i)$, $i = 1{,}2$ and solving certain integral equations following from their definition in (4.1) and (4.2). In the absence of strong prior information and seeking a simple alternative, we suggest the following approach that requires only a rough range for the values of each population, say $range_i$ corresponding to $F(\cdot;\mu,F_i)$, $i = 1{,}2$. Taking $a_1 = a_2 = 2$, which yields infinite variances for G_{10} and G_{20}, G_{10} is specified by choosing b_1 such that $median(\theta \mid a_1, b_1) \simeq (range_1/4)^2$, where $\theta \sim G_{10}$. Next to specify G_{20}, with a_1, a_2 and b_1 determined, choose b_2 that satisfies $median(\max(\theta,\delta) \mid a_1, b_1, a_2, b_2) \simeq (range_2/4)^2$, where $\delta \sim G_{20}$ independently of θ. Finally, we set $\nu_1 = \nu_2 = 1$ as is customary.

6 Example

We consider a data set from Gibbons (1976, p. 208) consisting of measurements on the time interval between eruptions of the Old Faithful geyser in Yellowstone National Park for two years, 1878 and 1947. As discussed in Gibbons, the geyser has been very stable in average time interval between eruptions since it was discovered in 1870. Hence we assume that time intervals have the same median for the two years. However, the range between the minimum and maximum time interval between eruptions seems to have been increasing. The measurements in minutes are 54.5, 63.1, 64.9, 67.3, 70.8, 71.5 for 1878 and 47.4, 50.1, 61.3, 65.2, 73.1, 73.6, 79.2 for the 1947 year. The sample mean, median, variance and interquartile range for the 1878 data is 65.35, 66.1, 38.863 and 6.37, respectively, with the associated values for the 1947 data being 64.27, 65.2, 147.166 and 17.65, respectively. With these measurements there seems to be no justification for a log transformation to improve symmetry. Gibbons has used this dataset to illustrate a nonparametric test of scale, arguing that any parametric test would be suspect with such small sample sizes.

We employ the methodology of section 4, and in particular model (5.1), for the underlying populations, with $F(\cdot;\mu,F_1)$ in (4.1) corresponding to the 1878 population and $F(\cdot;\mu,F_2)$ in (4.2) to the 1947 population. Following the approach of section 5 for prior specification, we set $\nu_1=\nu_2=1$, $a_1=a_2=2$, $b_1=65.7$ and $b_2=191.5$. Here we have used rough ranges 25 and 45 for the values of the 1878 and 1947 populations, respectively. Finally, we assume that a priori $\mu\sim N(50,(30)^2)$. Setting $\tau=30$ for the standard deviation of μ corresponds to a rather vague prior, as the actual range of the data suggests. Attractively, the resulting posterior inference is essentially identical to that obtained using smaller values, e.g., $\tau=10$.

The predictive distribution functions are compared both a priori and a posteriori in Figure 1. The crossing point becomes evident in the posterior curves. The amount of learning, given the small sample sizes, is very encouraging for the modeling approach. Similar patterns are observed when comparing the prior and posterior predictive densities. (See Figure 2). Note the two crossing points that emerge for the posterior predictive densities. This is a result of the choice of kernel for the mixtures in (4.1) and (4.2); Gaussian kernels satisfy the conditions of Lemma 2 as well as these of Lemma 1.

We next turn to posterior inference regarding the interquartile range functionals. Their posteriors are plotted in Figure 3. Point (posterior medians) and 95% equal-tail interval estimates for $IQR(F(\cdot;\mu,F_1))$, $IQR(F(\cdot;\mu,F_2))$ and $IQR(F(\cdot;\mu,F_2))/(IQR(F\ (\cdot;\mu,F_1)))$ are 8.238 (5.780,12.851), 14.757 (10.675,22.721) and 1.799 (1.114,3.031), respectively. The last interval estimate suggests significantly greater dispersion for the 1947 data relative to that of the 1878 data. Finally, the posterior median and a 95% interval estimate for the common location μ are 65.376 and (60.569,69.974), respectively.

7 Extension to variability ordered skewed distributions

In the literature that discusses variability order for real valued random variables most of the attention has focused on symmetric distributions. Although variability ordering for distributions with different forms of skewness seems inappropriate, adding common asymmetry to $F(\cdot;\mu,F_i)$, $i=1,2$, defined in (4.1) and (4.2), might be of interest in some applications. A convenient way to accomplish this is by introducing skewness parametrically in the kernel of the mixtures.

One possibility arises if we employ the approach of Fernández and Steel (1998) to transform the Gaussian kernel of (4.1) and (4.2) into a skewed normal family, preserving unimodality. For a positive scalar parameter γ consider the density

$$\begin{aligned} f_{SK}(\cdot;\gamma,\mu,\theta) = \ & \tfrac{(2/\pi\theta)^{1/2}}{(\gamma+\gamma^{-1})}\{\exp\left(-\gamma^2(\cdot-\mu)^2/2\theta\right)1_{(-\infty,\mu)}(\cdot) \\ & +\exp\left(-(\cdot-\mu)^2/2\theta\gamma^2\right)1_{[\mu,\infty)}(\cdot)\}, \end{aligned} \tag{7.1}$$

with associated distribution function

$$F_{SK}(\cdot;\gamma,\mu,\theta)=\frac{2}{1+\gamma^2}\{\Phi(\frac{\gamma(\,\cdot-\mu)}{\theta^{1/2}})1_{(-\infty,\mu)}(\cdot)+(\frac{1-\gamma^2}{2}+\gamma^2\Phi(\frac{\cdot-\mu}{\gamma\theta^{1/2}}))1_{[\mu,\infty)}(\cdot)\}. \tag{7.2}$$

For $\gamma>1$ ($\gamma<1$) (7.1) is right (left) skewed, while for $\gamma=1$ it reduces to the normal density we begin with. The mode of (7.1) is again at μ. It can be shown more formally that γ is a skewness parameter by obtaining several skewness functionals of (7.1). In particular, from (7.2), $P(X\geq\mu\mid\gamma,\mu,\theta)/P(X<\mu\mid\gamma,\mu,\theta)=\gamma^2$. It is straightforward to verify that mixing (7.2) with respect to θ preserves the role of γ as a skewness parameter. In particular, if $Y\sim F(\cdot;\gamma,\mu,G)=\int F_{SK}(\cdot;\gamma,\mu,\theta)G(d\theta)$, where G is a distribution on R^+, applying Fubini's theorem we obtain $P(Y\geq\mu\mid\gamma,\mu,G)/P(Y<\mu\mid\gamma,\mu,G)=\gamma^2$. Moreover, using (7.2) we can show that for $\theta_1<\theta_2$, $S^-(F_{SK}(\cdot;\gamma,\mu,\theta_2)-F_{SK}(\cdot;\gamma,\mu,\theta_1))$ $=1$, with the sign sequence being $+,-$. Here, the common crossing point of $F_{SK}(\cdot;\gamma,\mu,\theta_2)$ and $F_{SK}(\cdot;\gamma,\mu,\theta_1)$ is μ. Hence if we define the mixtures

$$F(\cdot;\gamma,\mu,F_i)=\int F_{SK}(\cdot;\gamma,\mu,\theta)F_i(d\theta),\ i=1,2, \tag{7.3}$$

imposing the restriction $F_1\leq_{st}F_2$ on the mixing distributions, Lemma 1 yields that $F(\cdot;\gamma,\mu,F_1)\leq_{var}F(\cdot;\gamma,\mu,F_2)$. In this context, the common location of $F(\cdot;\gamma,\mu,F_i)$, $i=1,2$ is their mode μ. The distributions defined in (7.3) allow for skewness, which however must be of the same form and amount for both since it is controlled by their common parameter γ.

Skewness is more naturally measured with respect to the median. It might therefore be preferable to work with mixtures similar to (7.3) but with their median η as the common location. Such a model can be developed if we use the split normal family for the kernel of the mixtures (see Geweke, 1989, where the split normal density is employed as an importance sampling density and Kottas and Gelfand, 1999, for an application to semiparametric median regression modeling). The split normal distribution function is defined by

$$F_{SP}(\cdot;\gamma,\eta,\theta) = \Phi(\frac{\cdot - \eta}{\gamma\theta^{1/2}})1_{(-\infty,\eta)}(\cdot) + \Phi(\frac{\gamma(\cdot - \eta)}{\theta^{1/2}})1_{[\eta,\infty)}(\cdot), \tag{7.4}$$

where η is the median, $\theta^{1/2}$ the scale parameter and $\gamma > 0$ a skewness parameter with $\gamma < 1$ (> 1) corresponding to right (left) skewness. In Kottas and Gelfand (1999) it is shown that if we mix (7.4) with respect to θ, η and γ are still the median and a skewness parameter, respectively, for the resulting mixture. Furthermore, the conditions of Lemma 1 are satisfied by (7.4) where now the common crossing point is the median η. Therefore defining $F(\cdot;\gamma,\eta,F_i) = \int F_{SP}(\cdot;\gamma,\eta,\theta)F_i(d\theta)$, $i = 1{,}2$, for distributions F_1 and F_2 on R^+ with $F_1 \leq_{st} F_2$, we get $F(\cdot;\gamma,\eta,F_1) \leq_{var} F(\cdot;\gamma,\eta,F_2)$.

References

Antoniak, C.E. (1974), "Mixtures of Dirichlet Processes With Applications to Nonparametric Problems," *The Annals of Statistics*, 2, 1152-1174.

Arjas, E., and Gasbarra, D. (1996), "Bayesian Inference of Survival Probabilities, Under Stochastic Ordering Constraints," *Journal of the American Statistical Association*, 91, 1101-1109.

Bickel, P.J., and Lehmann, E.L. (1976), "Descriptive Statistics for Nonparametric Models. III. Dispersion," *The Annals of Statistics*, 4, 1139-1158.

Bickel, P.J., and Lehmann, E.L. (1979), "Descriptive Statistics for Nonparametric Models. IV. Spread," in *Contributions to Statistics, Jaroslav Hájek Memorial Volume*, ed. J. Jureckova, Dordrecht: Riedel, pp. 33-40.

Birnbaum, Z.W. (1948), "On Random Variables With Comparable Peakedness," *The Annals of Mathematical Statistics*, 19, 76-81.

Escobar, M.D. (1994), "Estimating Normal Means With a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89, 268-277.

Escobar, M.D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association,* 90, 577-588.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications Vol. II*, 2nd ed., New York: Wiley.

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.

Fernández, C., and Steel, M.F.J. (1998), "On Bayesian Modeling of Fat Tails and Skewness," *Journal of the American Statistical Association*, 93, 359-371.

Gelfand, A.E., and Kottas, A. (1999), "A Computational Approach for Full Nonparametric Bayesian Inference in Single and Multiple Sample Problems," Technical Report 99-08, University of Connecticut, Department of Statistics.

Gelfand, A.E., and Kottas, A. (2000), "Nonparametric Bayesian Modeling for Stochastic Order," to appear in the *The Annals of the Institute of Statistical Mathematics.*

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 Smith, A.F.M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association,* 85, 398-409.

Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317-1339.

Gibbons, J.D. (1976), *Nonparametric Methods for Quantitative Analysis*, New York: Holt, Rinehart and Winston.

Karlin, S. (1968), *Total Positivity. Vol. I*, Stanford: Stanford University Press.

Kottas, A., and Gelfand, A.E. (1999), "Bayesian Semiparametric Median Regression Modeling," Technical Report 99-20, University of Connecticut, Department of Statistics.

Lewis, T., and Thompson, J.W. (1981), "Dispersive Distributions, and the Connection Between Dispersivity and Strong Unimodality," *Journal of Applied Probability*, 18, 76-90.

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.

Oja, H. (1981), "On Location, Scale, Skewness and Kurtosis of Univariate Distributions," *Scandinavian Journal of Statistics*, 8, 154-168.

Saunders, I.W., and Moran, P.A.P. (1978), "On the Quantiles of the Gamma and F Distributions," *Journal of Applied Probability*, 15, 426-432.

Schweder, T. (1982), "On the Dispersion of Mixtures," *Scandinavian Journal of Statistics*, 9, 165-169.

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

Shaked, M. (1980), "On Mixtures from Exponential Families," *Journal of the Royal Statistical Society*, Ser. B, 42, 192-198.

Shaked, M. (1982), "Dispersive Ordering of Distributions," *Journal of Applied Probability*, 19, 310-320.

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

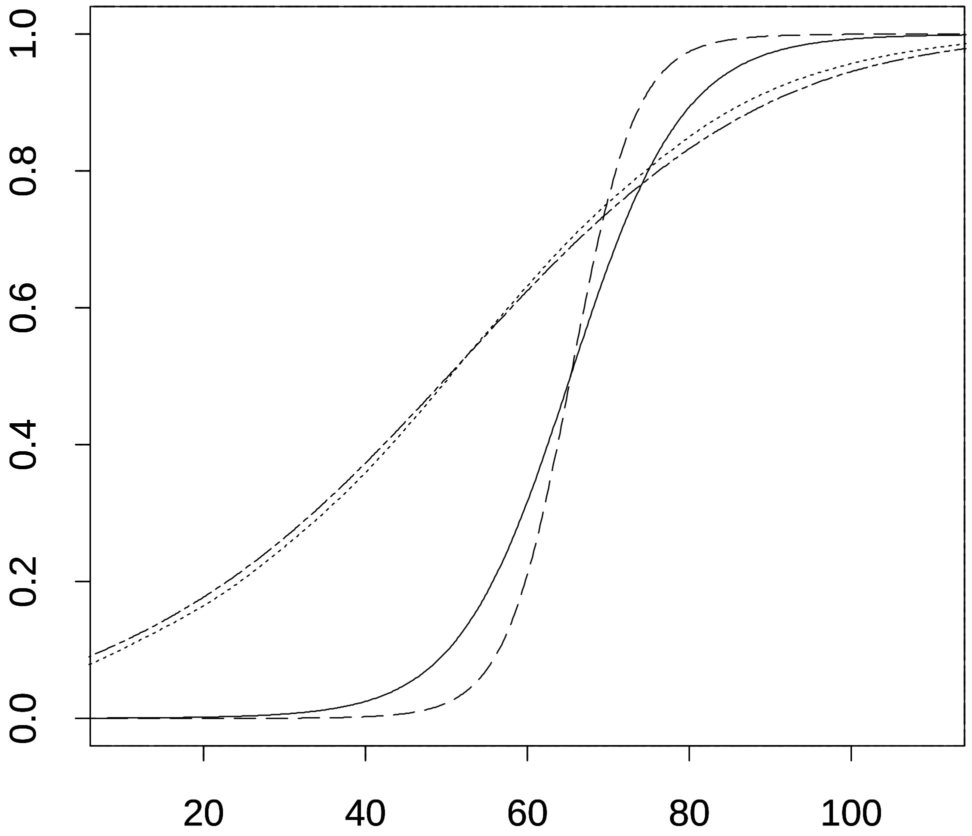

Figure 1: Prior and posterior predictive distribution functions. The prior predictive curve corresponding to $F(\cdot;\mu,F_1)$ and $F(\cdot;\mu,F_2)$ is denoted by the dotted and the dashed-dotted line, respectively, with the associated posteriors denoted by the dashed and the solid line, respectively.

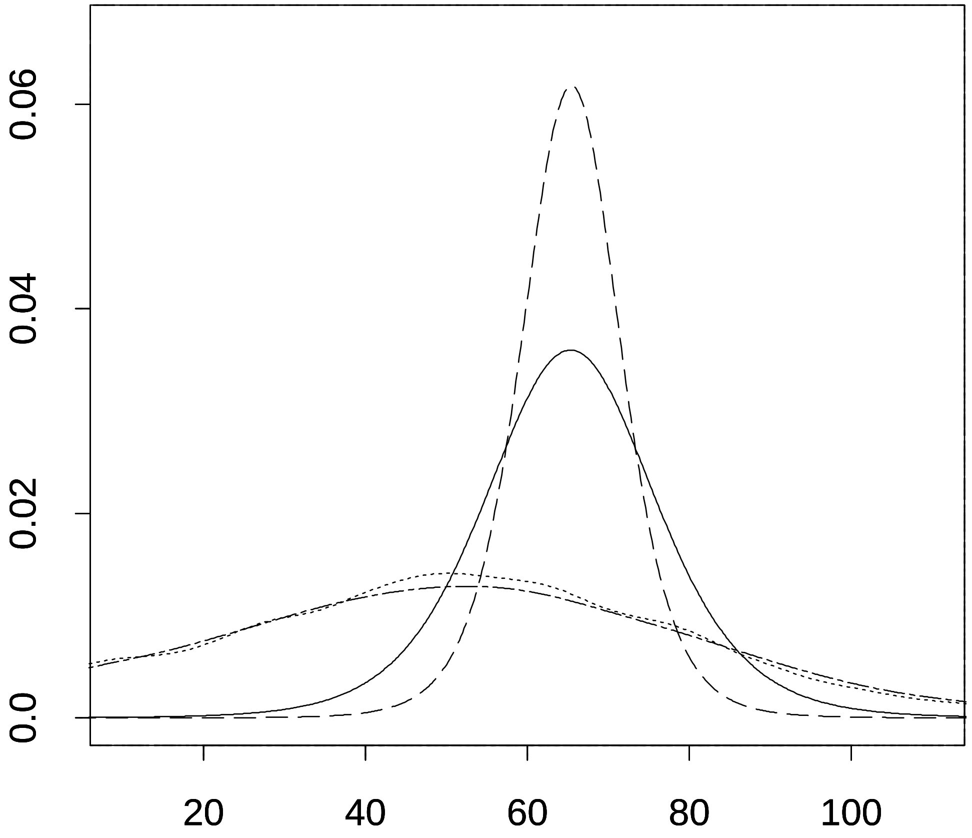

Figure 2: Prior and posterior predictive densities. The dotted and the dashed line denote the prior and posterior predictive density, respectively, for $F(\cdot;\mu,F_1)$ and the dashed-dotted and solid line the corresponding densities for $F(\cdot;\mu,F_2)$.

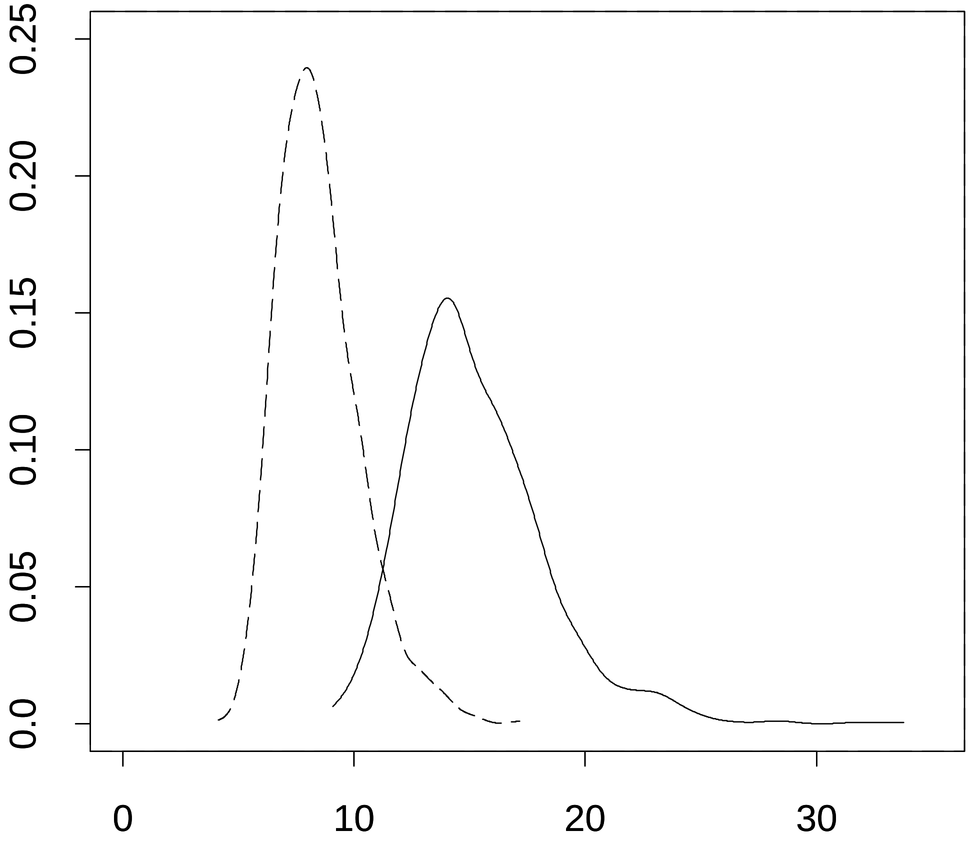

Figure 3: Posteriors for the interquartile range functionals. The dashed line corresponds to $[IQR(F(\cdot;\mu,F_1))|\ D]$ and the solid line to $[IQR(F(\cdot;\mu,F_2))|\ D]$.