

Bayesian semiparametric modeling for stochastic precedence, with applications in epidemiology and survival analysis

ATHANASIOS KOTTAS

Department of Applied Mathematics and Statistics, University of California, 1156 High Street, Santa Cruz, CA 95064 (thanos@ams.ucsc.edu)

Abstract. We propose a prior probability model for two distributions that are ordered according to a stochastic precedence constraint, a weaker restriction than the more commonly utilized stochastic order constraint. The modeling approach is based on structured Dirichlet process mixtures of normal distributions. Full inference for functionals of the stochastic precedence constrained mixture distributions is obtained through a Markov chain Monte Carlo posterior simulation method. A motivating application involves study of the discriminatory ability of continuous diagnostic tests in epidemiologic research. Here, stochastic precedence provides a natural restriction for the distributions of test scores corresponding to the non-infected and infected groups. Inference under the model is illustrated with data from a diagnostic test for Johne's disease in dairy cattle. We also apply the methodology to comparison of survival distributions associated with two distinct conditions, and illustrate with analysis of data on survival time after bone marrow transplantation for treatment of leukemia.

Keywords: Dirichlet process prior; Markov chain Monte Carlo; Mixtures of normal distributions; Receiver operating characteristic curve; Stochastic order; Survival function.

1 Introduction

In certain applications, including problems in the biomedical sciences, that involve comparison of two populations, there is interest in incorporating a stochastic relationship between the corresponding distributions. In this context, the Bayesian paradigm provides an attractive modeling framework, since any probability order constraint incorporated in the prior model is preserved to the posterior analysis. In general, a key argument for forcing a particular order restriction in the model (or estimation technique) for the two distributions is that the order constraint of interest may not hold for the empirical distribution functions, especially, for small to moderate sample sizes. Moreover, incorporation of the order restriction, if appropriate, can improve estimation efficiency and predictive accuracy.

The most extensively studied probability order constraint in the applied probability literature is stochastic ordering. The stochastic order relationship quantifies the notion that random variable X_1 tends to be *smaller* than random variable X_2 by ordering the respective distribution functions F_1 and F_2 . In particular, F_1 is stochastically smaller than F_2 (denoted by $F_1 \leq_{st} F_2$) if $F_1(x) \geq F_2(x)$, for all x . Furthermore, with regard to statistical modeling for data analysis, stochastic ordering is the most commonly utilized type of probability order restriction. In particular, Bayesian testing methods for stochastic ordering among a set of categorical random variables are discussed in Evans et al. (1997), and a Bayesian nonparametric estimation method for stochastically ordered survival functions is developed in Arjas and Gasbarra (1996). The more recent Bayesian nonparametrics literature contains further modeling approaches for stochastic ordering, including Dirichlet process based models (Gelfand and Kottas, 2001; Hoff, 2003; Dunson and Peddada, 2008), and methods based on Pólya tree priors (Karabatsos and Walker, 2007; Hanson, Kottas and Branscum, 2008). See Müller and Quintana (2004) and Hanson, Branscum and Johnson (2005) for reviews of Dirichlet process and Pólya tree priors and their use in Bayesian nonparametric data analysis.

The stochastic order constraint is arguably restrictive for applications where a stochastic rela-

relationship between two distributions is anticipated, but the ordering of the respective distribution functions over their entire support is not a plausible constraint, especially, in the tails of the distributions. Arcones, Kvam and Samaniego (2002) introduced *stochastic precedence*, a weaker constraint than stochastic order, which builds the restriction through $\Pr(X_1 \leq X_2)$ for two independent random variables X_1 and X_2 with distribution functions F_1 and F_2 , respectively. Specifically, X_1 is said to stochastically precede X_2 if $\Pr(X_1 \leq X_2) \geq 0.5$, denoted by $X_1 \leq_{sp} X_2$ or, equivalently, $F_1 \leq_{sp} F_2$. For example, the normal distribution is stochastic precedence constrained with respect to its mean, in particular, $N(\theta_1, \sigma_1^2) \leq_{sp} N(\theta_2, \sigma_2^2)$ if and only if $\theta_1 \leq \theta_2$, with no further restriction needed on the variances σ_1^2 and σ_2^2 . Arcones, Kvam and Samaniego (2002) discussed potential applications for stochastic precedence, and developed classical nonparametric stochastic precedence constrained estimators for F_1 (with F_2 assumed known) and for both F_1 and F_2 .

Note that, under the Bayesian setting, the independence of random variables X_1 and X_2 in the stochastic precedence definition is replaced with conditional independence. The conditioning involves the parameters of distributions F_1 and F_2 under a parametric framework, or the random distributions under a nonparametric prior model. Using the (conditional) independence of X_1 and X_2 , we have $\Pr(X_1 \leq X_2) = E^{F_2}\{F_1(X_2)\} = E^{F_1}\{1 - F_2(X_1)\}$, where the superscript indicates the distribution with respect to which the expectation is taken. Hence, the stochastic precedence assumption, $F_1 \leq_{sp} F_2$, implies $E^{F_2}\{F_1(X_2)\} \geq 0.5$ and $E^{F_1}\{F_2(X_1)\} \leq 0.5$.

With regard to Bayesian estimation methods under stochastic precedence constraints, we are only aware of the approach by Chen and Dunson (2004) for discrete random variables with finite support. The approach is based on a prior for the cumulative probabilities of random variables X_1 and X_2 , which arises from the conjugate “product of independent ordered Dirichlet densities” prior by direct truncation to incorporate the restrictions on the cumulative probabilities induced by constraints $E^{F_2}\{F_1(X_2)\} \geq 0.5$ and $E^{F_1}\{F_2(X_1)\} \leq 0.5$. The resulting estimation method is applied to stochastic precedence constrained survival functions, given right censored data, by partitioning the support of the corresponding survival distributions into a finite number of intervals.

We propose a semiparametric prior probability model, which, to our knowledge, represents the first attempt to Bayesian modeling and inference for continuous distributions subject to the stochastic precedence constraint. We achieve this through structured Dirichlet process mixtures of normal distributions, using a result that yields a sufficient condition to preserve under mixing the stochastic precedence property of the normal distribution. A posterior simulation method is developed to obtain full inference for the stochastic precedence constrained distributions given data that may include censoring.

The methodology is applied to a motivating problem from epidemiologic research involving study of the discriminatory ability of a continuous diagnostic test for a particular infection (or disease). In this application, stochastic precedence can be motivated as a practically important restriction for the distributions of diagnostic test scores associated with the groups of non-infected and infected subjects in the study. In particular, it forces a natural constraint on the values of the area under the receiver operating characteristic (ROC) curve, a widely used graphical measure of the accuracy of the continuous diagnostic test. Moreover, we consider applications to survival analysis problems involving comparison of survival distributions associated with two distinct conditions.

The paper is organized as follows. In Section 2, we present the modeling approach, including methods for posterior inference and prior specification, with technical details provided in the two appendices. The methodology is illustrated in Section 3 with data on survival times after bone marrow transplantation for treatment of leukemia, and on diagnostic test scores for Johne's disease in dairy cattle. Finally, Section 4 concludes with a summary.

2 Methods

Section 2.1 develops the semiparametric prior model for two stochastic precedence constrained distributions. Posterior inference under the resulting Bayesian model is addressed in Section 2.2, and an approach to prior specification is presented in Section 2.3. Section 2.4 discusses the extension

to semiparametric regression modeling under the stochastic precedence restriction.

2.1 The modeling approach

We develop a semiparametric modeling approach for two continuous distributions F_1 and F_2 , associated with random variables X_1 and X_2 , that are ordered according to the stochastic precedence constraint. (We will use F_1 and F_2 to denote, depending on the context, either the distributions or the corresponding distribution functions.) In particular, we seek a structured semiparametric prior probability model for the pair of distribution functions (F_1, F_2) such that prior realizations F_1 and F_2 satisfy the stochastic precedence restriction, $F_1 \leq_{sp} F_2$, that is, $\Pr(X_1 \leq X_2) \geq 0.5$. This will, of course, preserve the restriction to the posterior realizations for F_1 and F_2 . Such an approach is in contrast to working with independent prior models for F_1 and F_2 , and forcing by truncation the $F_1 \leq_{sp} F_2$ restriction in the posterior estimation method for F_1 and F_2 .

We take the support for F_1 and F_2 to be the real line, \mathbb{R} . The application to survival analysis involves random variables T_1 and T_2 on \mathbb{R}^+ corresponding to two survival distributions that we wish to estimate under the stochastic precedence constraint, $\Pr(T_1 \leq T_2) \geq 0.5$. But then, we can apply the model for (F_1, F_2) on \mathbb{R} to random variables $X_\ell = \log(T_\ell)$, $\ell = 1, 2$, and use the straightforward transformation to report inference for the survival functions on the original scale (as in the data example of Section 3.1). Analogously, the illustration with epidemiological data (Section 3.2) involves log-transformed test scores, although, in that case, inference on \mathbb{R} suffices.

The proposed stochastic precedence constrained prior for (F_1, F_2) is based on semiparametric location normal mixtures. Denote by $N(m, s^2)$ the normal distribution with mean m and variance s^2 , and by $F_N(\cdot; m, s^2)$ and $f_N(\cdot; m, s^2)$ the corresponding distribution function and density function, respectively. The key result to building the prior model is given by the following lemma whose proof is included in Appendix A.

Lemma. Consider the representation of F_1 and F_2 in terms of general location normal mixtures,

$$F_\ell(x) \equiv F_\ell(x; H_\ell, \sigma_\ell^2) = \int F_N(x; \theta, \sigma_\ell^2) dH_\ell(\theta), \quad \ell = 1, 2$$

with stochastically ordered mixing distributions H_1 and H_2 , that is, $H_1 \leq_{st} H_2$. Then, $F_1 \leq_{sp} F_2$.

The lemma provides a sufficient condition to preserve under mixing the stochastic precedence property of the normal distribution discussed in the Introduction. Note that, although the result is stated for the class of normal mixtures on which we focus, the proof can be readily extended for general mixtures $\int K(\cdot; \theta, \varphi_\ell) dH_\ell(\theta)$, $\ell = 1, 2$, provided the mixture kernel $K(\cdot; \theta, \varphi)$ satisfies the stochastic precedence restriction in the θ component of its parameter vector (θ, φ) .

The result of the lemma gives rise to a constructive approach to defining a semiparametric stochastic precedence prior model for (F_1, F_2) . Along with parametric priors for σ_1^2 and σ_2^2 , what is required is a nonparametric stochastically ordered prior for the pair of mixing distribution functions (H_1, H_2) . An approach to defining such a prior, which balances model flexibility and computational complexity, is to use latent distribution functions on \mathbb{R} , say G_1 and G_2 , such that $H_1(\cdot) = G_1(\cdot)$ and $H_2(\cdot) = G_1(\cdot)G_2(\cdot)$. Placing nonparametric priors on G_1 and G_2 , taken to be independent for computational convenience, induces a nonparametric prior on the space of stochastically ordered distributions (H_1, H_2) . In fact, it will be useful to think of H_1 and H_2 as the distribution of θ and $\max\{\theta, \phi\}$, respectively, where $\theta \sim G_1$ and, independently, $\phi \sim G_2$. This approach was discussed in Gelfand and Kuo (1991), developed in Gelfand and Kottas (2001), and also used in Kottas, Branco and Gelfand (2002) and Hanson, Kottas and Branscum (2008).

Now, the choice of the Dirichlet process (DP) prior (Ferguson, 1973) for G_1 and G_2 becomes attractive as it yields a DP mixture structure for F_1 and F_2 (Antoniak, 1974; Escobar and West, 1995). We will use the generic notation $G \sim \text{DP}(\alpha, G_0)$ to indicate that a DP prior is assigned to random distribution G , where α is the DP precision parameter and G_0 is the DP centering distribution. For later reference, it is useful to recall the constructive definition of the DP (Sethuraman, 1994). Based on this definition, the DP generates (almost surely) discrete distributions with a countable number of possible values drawn independently from G_0 . The corresponding weights are generated using a *stick-breaking* mechanism based on independent draws, $\{z_r : r = 1, 2, \dots\}$, from

a Beta(1, α) distribution; specifically, the first weight is equal to z_1 and, for $i = 2, 3, \dots$, the i -th weight is given by $z_i \prod_{r=1}^{i-1} (1 - z_r)$.

Hence, assuming independent DP priors for G_1 and G_2 , we obtain the stochastic precedence model for (F_1, F_2) as follows:

$$\begin{aligned}
 F_1(x; G_1, \sigma_1^2) &= \int F_N(x; \theta, \sigma_1^2) dG_1(\theta) \\
 F_2(x; G_1, G_2, \sigma_2^2) &= \iint F_N(x; \max\{\theta, \phi\}, \sigma_2^2) dG_1(\theta) dG_2(\phi) \\
 G_\ell | \boldsymbol{\psi}_\ell &\stackrel{\text{ind.}}{\sim} \text{DP}(\alpha_\ell, \text{N}(\mu_\ell, \tau_\ell^2)), \quad \ell = 1, 2
 \end{aligned} \tag{1}$$

where $\boldsymbol{\psi}_\ell = (\alpha_\ell, \mu_\ell, \tau_\ell^2)$ are the hyperparameters of the DP prior for G_ℓ , $\ell = 1, 2$. We place (independent) priors on the components of $\boldsymbol{\psi}_\ell$, specifically, for $\ell = 1, 2$, $p(\mu_\ell) = \text{N}(c_\ell, d_\ell)$, $p(\tau_\ell^2) = \text{inv-gamma}(w_\ell, e_\ell)$, and $p(\alpha_\ell) = \text{gamma}(a_{\alpha_\ell}, b_{\alpha_\ell})$. Here, $\text{inv-gamma}(a, b)$ denotes an inverse gamma distribution with mean $b/(a - 1)$ (provided $a > 1$), and $\text{gamma}(a, b)$ stands for the gamma distribution with mean a/b . Finally, we take $p(\sigma_\ell^2) = \text{inv-gamma}(a_{\sigma_\ell}, b_{\sigma_\ell})$, $\ell = 1, 2$.

We note that, although its motivation and development is different, the final form of model (1) is related to the DP mixture models from Kottas and Gelfand (2001) and Gelfand and Kottas (2001). The former utilizes scale normal DP mixtures, with a common location parameter and stochastically ordered mixing distributions, to model a particular form of variability order. The latter develops a model for $F_1 \leq_{st} F_2$ under which F_1 and F_2 are represented as in (1) albeit with $\sigma_1^2 = \sigma_2^2 \equiv \sigma^2$. Recall that the normal distribution satisfies the stochastic precedence constraint in its mean with no further restriction on the variance parameters, i.e., $\text{N}(\theta_1, \sigma_1^2) \leq_{sp} \text{N}(\theta_2, \sigma_2^2)$ if and only if $\theta_1 \leq \theta_2$. However, stochastic ordering requires a common variance parameter, i.e., $\text{N}(\theta_1, \sigma_1^2) \leq_{st} \text{N}(\theta_2, \sigma_2^2)$ if and only if $\theta_1 \leq \theta_2$ and $\sigma_1^2 = \sigma_2^2$. Interestingly, location mixing with stochastically ordered mixing distributions preserves this structure of the normal mixture kernel to the resulting mixture distributions.

2.2 Posterior inference

Let data = $\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \{x_{1i} : i = 1, \dots, n_1\}$ and $\mathbf{x}_2 = \{x_{2j} : j = 1, \dots, n_2\}$, be the data vectors from distributions F_1 and F_2 . The model for the data can be expressed in hierarchical form by introducing latent mixing parameters $\boldsymbol{\theta} = \{\theta_i : i = 1, \dots, n_1, n_1 + 1, \dots, n_1 + n_2\}$, which, given G_1 , are i.i.d. from G_1 , and $\boldsymbol{\phi} = \{\phi_j : j = 1, \dots, n_2\}$, with the ϕ_j , given G_2 , i.i.d. from G_2 . Specifically,

$$\begin{aligned}
 x_{1i} \mid \theta_i, \sigma_1^2 &\stackrel{\text{i.i.d.}}{\sim} f_N(x_{1i}; \theta_i, \sigma_1^2), \quad i = 1, \dots, n_1 \\
 x_{2j} \mid \theta_{n_1+j}, \phi_j, \sigma_2^2 &\stackrel{\text{i.i.d.}}{\sim} f_N(x_{2j}; \max\{\theta_{n_1+j}, \phi_j\}, \sigma_2^2), \quad j = 1, \dots, n_2 \\
 \theta_i \mid G_1 &\stackrel{\text{i.i.d.}}{\sim} G_1, \quad i = 1, \dots, n_1 + n_2 \\
 \phi_j \mid G_2 &\stackrel{\text{i.i.d.}}{\sim} G_2, \quad j = 1, \dots, n_2 \\
 G_\ell \mid \boldsymbol{\psi}_\ell &\stackrel{\text{i.i.d.}}{\sim} \text{DP}(\alpha_\ell, \text{N}(\mu_\ell, \tau_\ell^2)), \quad \ell = 1, 2
 \end{aligned} \tag{2}$$

with priors for $\boldsymbol{\psi}_\ell = (\alpha_\ell, \mu_\ell, \tau_\ell^2)$ and σ_ℓ^2 , for $\ell = 1, 2$, as discussed in Section 2.1. The introduction of the additional mixing parameters θ_{n_1+j} , $j = 1, \dots, n_2$, is key for implementation of posterior simulation, since the augmented vector $\boldsymbol{\theta}$ preserves the first stage conditionally independent specification in the hierarchical model after marginalizing in (2) the random distributions G_1 and G_2 over their DP priors.

The hierarchical model formulation in (2) assumes fully observed realizations from distributions F_1 and F_2 . To handle censoring, the normal density in the first stage specification of the model is replaced by appropriate functions of the corresponding distribution function. For instance, assume that F_1 and F_2 model survival times, on the logarithmic scale, for two distinct groups of subjects. Consider data, on the original scale, for the first group that comprise observed survival times $\{t_{1i} : i = 1, \dots, n_{1o}\}$ and right censored survival times $\{t_{1k}^+ : k = 1, \dots, n_{1c}\}$, and thus $n_1 = n_{1o} + n_{1c}$. Then, the first stage of model (2) becomes $\prod_{i=1}^{n_{1o}} f_N(x_{1i}; \theta_i, \sigma_1^2) \prod_{k=1}^{n_{1c}} \{1 - F_N(x_{1k}^+; \theta_k, \sigma_1^2)\}$, where $x_{1i} = \log(t_{1i})$ and $x_{1k}^+ = \log(t_{1k}^+)$. A similar modification is applied to the second stage of model (2) if the data vector from the second group includes right censored survival times, and the approach is analogous for left or interval censored observations.

The full posterior distribution corresponding to model (2) can be written as

$$p(G_1, G_2, \boldsymbol{\theta}, \boldsymbol{\phi}, \sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \mid \text{data}) = p(G_1 \mid \boldsymbol{\theta}, \boldsymbol{\psi}_1)p(G_2 \mid \boldsymbol{\phi}, \boldsymbol{\psi}_2)p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \mid \text{data})$$

where $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \mid \text{data})$ is the marginal posterior that arises from model (2) by integrating out G_1 and G_2 over their DP priors. Appendix B provides details on Markov chain Monte Carlo (MCMC) sampling from this posterior distribution given data that may include censored observations.

To obtain inference for G_ℓ , $\ell = 1, 2$, we use the draws $\{\boldsymbol{\theta}_b, \boldsymbol{\phi}_b, \sigma_{1,b}^2, \sigma_{2,b}^2, \boldsymbol{\psi}_{1,b}, \boldsymbol{\psi}_{2,b} : b = 1, \dots, B\}$ from $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \mid \text{data})$ to sample from $p(G_1 \mid \boldsymbol{\theta}, \boldsymbol{\psi}_1)$ and $p(G_2 \mid \boldsymbol{\phi}, \boldsymbol{\psi}_2)$. Based on results from Antoniak (1974), the former is a DP distribution with precision parameter $\alpha_1 + n_1 + n_2$ and centering distribution $G'_{1,0}(\cdot; \boldsymbol{\theta}, \boldsymbol{\psi}_1) = \alpha_1(\alpha_1 + n_1 + n_2)^{-1}N(\cdot; \mu_1, \tau_1^2) + (\alpha_1 + n_1 + n_2)^{-1} \sum_{i=1}^{n_1+n_2} \delta_{\theta_i}(\cdot)$; the latter denotes a DP with precision parameter $\alpha_2 + n_2$ and centering distribution $G'_{2,0}(\cdot; \boldsymbol{\phi}, \boldsymbol{\psi}_2) = \alpha_2(\alpha_2 + n_2)^{-1}N(\cdot; \mu_2, \tau_2^2) + (\alpha_2 + n_2)^{-1} \sum_{j=1}^{n_2} \delta_{\phi_j}(\cdot)$. (Here, $\delta_y(\cdot)$ denotes a point mass at y .) To sample from $p(G_1 \mid \boldsymbol{\theta}, \boldsymbol{\psi}_1)$ and $p(G_2 \mid \boldsymbol{\phi}, \boldsymbol{\psi}_2)$, we employ the DP stick-breaking representation, discussed in Section 2.1, with a truncation approximation (e.g., Gelfand and Kottas, 2002; Kottas, 2006). In particular, the posterior samples for G_ℓ , $\ell = 1, 2$, are of the form $G_{\ell,b}(\cdot) = \sum_{k=1}^{K_\ell} \omega_{k,b}^{(\ell)} \delta_{y_{k,b}^{(\ell)}}(\cdot)$. Here, the $y_{k,b}^{(1)}$, $k = 1, \dots, K_1$, are i.i.d. from $G'_{1,0}(\cdot; \boldsymbol{\theta}_b, \boldsymbol{\psi}_{1,b})$, the $y_{k,b}^{(2)}$, $k = 1, \dots, K_2$, are i.i.d. from $G'_{2,0}(\cdot; \boldsymbol{\phi}_b, \boldsymbol{\psi}_{2,b})$, and the stick-breaking weights $\{\omega_{k,b}^{(1)} : k = 1, \dots, K_1\}$, and $\{\omega_{k,b}^{(2)} : k = 1, \dots, K_2\}$, are built from i.i.d. $\text{Beta}(1, \alpha_{1,b} + n_1 + n_2)$ and $\text{Beta}(1, \alpha_{2,b} + n_2)$ draws, respectively. The truncation levels can be chosen such that the DP weights mass is covered up to any desired tolerance ϵ ; for example, K_1 and K_2 can be specified from $\{(n_1 + n_2 + \max_b \alpha_{1,b}) / (n_1 + n_2 + 1 + \max_b \alpha_{1,b})\}^{K_1} = \{(n_2 + \max_b \alpha_{2,b}) / (n_2 + 1 + \max_b \alpha_{2,b})\}^{K_2} = \epsilon$.

Having collected posterior samples for G_ℓ , $\ell = 1, 2$, we can obtain full inference for $F_1(\cdot; G_1, \sigma_1^2)$ and $F_2(\cdot; G_1, G_2, \sigma_2^2)$, and for any of their functionals that may be of interest. For instance, for any specified point x_0 in \mathbb{R} , $\{f_{1,b}(x_0) = \int f_N(x_0; \theta, \sigma_{1,b}^2) dG_{1,b}(\theta) : b = 1, \dots, B\}$ and $\{f_{2,b}(x_0) = \iint f_N(x_0; \max\{\theta, \phi\}, \sigma_{2,b}^2) dG_{1,b}(\theta) dG_{2,b}(\phi) : b = 1, \dots, B\}$ are samples from the posteriors of the

mixture densities $f_1(x_0; G_1, \sigma_1^2)$ and $f_2(x_0; G_1, G_2, \sigma_2^2)$ at x_0 . Sampling from these posterior distributions over a grid of x_0 values, yields B posterior realizations for the random density functions $f_1(\cdot; G_1, \sigma_1^2)$ and $f_2(\cdot; G_1, G_2, \sigma_2^2)$, which can be summarized with point and interval estimates.

Furthermore, consider the survival analysis application involving comparison of two groups with associated random variables T_1 and T_2 on \mathbb{R}^+ , which are modeled through F_1 and F_2 on the logarithmic scale. Then, $S_{1,b}(t_0) = 1 - F_{1,b}(\log(t_0)) = 1 - \int F_N(\log(t_0); \theta, \sigma_{1,b}^2) dG_{1,b}(\theta)$, for $b = 1, \dots, B$, yields the posterior distribution for the survival function of T_1 at any specified point t_0 in \mathbb{R}^+ . Repeating over a grid of t_0 values, produces B posterior realizations for the random survival function of the first group, which can be further inverted (with interpolation) to provide posterior samples for the corresponding median survival time, or, more generally, for any percentile survival time of interest. Inference for the survival function and median survival time of the second group is obtained in the same fashion.

Of interest will also be inference for the probability that forms the basis of the stochastic precedence constraint definition. Again, let X_1 and X_2 be random variables with distributions $F_1(\cdot; G_1, \sigma_1^2)$ and $F_2(\cdot; G_1, G_2, \sigma_2^2)$. Then, using the same derivation as in Appendix A,

$$\Pr(X_1 \leq X_2; G_1, G_2, \sigma_1^2, \sigma_2^2) = \int \int \mathbb{E}^{F_N(\cdot; \max\{\theta, \phi\}, \sigma_2^2)} \{F_N(U; \theta, \sigma_1^2)\} dG_1(\theta) dG_2(\phi), \quad (3)$$

where the expectation is taken with respect to random variable U with distribution $N(\max\{\theta, \phi\}, \sigma_2^2)$. With $F_N^{-1}(\cdot; m, s^2)$ denoting the inverse distribution function of the $N(m, s^2)$ distribution, we can write $A(\theta, \phi, \sigma_1^2, \sigma_2^2) \equiv \mathbb{E}^{F_N(\cdot; \max\{\theta, \phi\}, \sigma_2^2)} \{F_N(U; \theta, \sigma_1^2)\} = \int_0^1 F_N(F_N^{-1}(1-z; \max\{\theta, \phi\}, \sigma_2^2); \theta, \sigma_1^2) dz$, which thus allows efficient numerical integration for $A(\theta, \phi, \sigma_1^2, \sigma_2^2)$ over a bounded interval. Finally, using the posterior samples for (G_1, G_2) and (σ_1^2, σ_2^2) , we obtain posterior realizations $\{\Pr(X_1 \leq X_2; G_{1,b}, G_{2,b}, \sigma_{1,b}^2, \sigma_{2,b}^2) : b = 1, \dots, B\}$ through either direct evaluation or, more efficiently, Monte Carlo integration of (3).

The prior distribution for $\Pr(X_1 \leq X_2; G_1, G_2, \sigma_1^2, \sigma_2^2)$ can be obtained in a similar fashion by sampling, with a truncation approximation, from the DP prior distributions for G_1 and G_2 . Anal-

ogously, we can sample from the prior distribution for any functional of the mixture distributions $F_1(\cdot; G_1, \sigma_1^2)$ and $F_2(\cdot; G_1, G_2, \sigma_2^2)$.

2.3 Prior specification

Prior specification for model (1) requires choosing the prior parameter values for the normal kernel variances σ_1^2 and σ_2^2 , the DP precision parameters α_1 and α_2 , and the normal centering distributions means, μ_1 and μ_2 , and variances, τ_1^2 and τ_2^2 . Relatively vague inverse gamma priors for σ_1^2 and σ_2^2 are obtained by setting their means equal to, say, $(R/6)^2$ using a guess, R , at the range of the data from both populations. In general, we set to 2 the shape parameters, a_{σ_ℓ} and w_ℓ , $\ell = 1, 2$, of the inverse gamma priors, resulting in infinite variances (thus, dispersed priors) and prior means given by the rate parameters, b_{σ_ℓ} and e_ℓ , $\ell = 1, 2$.

The DP precision parameters α_1 and α_2 control the prior distribution for the number of distinct components n_θ^* and n_ϕ^* in vectors θ and ϕ (e.g., Escobar and West, 1995). For instance, for moderately large n_1 and n_2 , $E(n_\theta^* | \alpha_1) \approx \alpha_1 \log\{(\alpha_1 + n_1 + n_2)/\alpha_1\}$ and $E(n_\phi^* | \alpha_2) \approx \alpha_2 \log\{(\alpha_2 + n_2)/\alpha_2\}$. Then, the choice of the parameters for the gamma priors on α_1 and α_2 can be guided by $E(n_\theta^*)$ and $E(n_\phi^*)$ the values of which can be approximated by averaging $E(n_\theta^* | \alpha_1)$ and $E(n_\phi^* | \alpha_2)$ over the particular gamma priors.

Finally, to specify the priors for (μ_1, τ_1^2) and (μ_2, τ_2^2) , we work with the prior predictive densities, $E(f_1(\cdot; G_1, \sigma_1^2))$ and $E(f_2(\cdot; G_1, G_2, \sigma_2^2))$, which depend on parameters of the DP centering distributions, but not on the DP precision parameters. Specifically,

$$E(f_1(x_0; G_1, \sigma_1^2)) = \int f_N(x_0; \theta_0, \sigma_1^2) f_N(\theta_0; \mu_1, \tau_1^2) p(\mu_1) p(\tau_1^2) p(\sigma_1^2) d\theta_0 d\sigma_1^2 d\mu_1 d\tau_1^2$$

and

$$\begin{aligned} E(f_2(x_0; G_1, G_2, \sigma_2^2)) &= \iint f_N(x_0; \max\{\theta_0, \phi_0\}, \sigma_2^2) f_N(\theta_0; \mu_1, \tau_1^2) f_N(\phi_0; \mu_2, \tau_2^2) \\ &\quad p(\mu_1) p(\tau_1^2) p(\mu_2) p(\tau_2^2) p(\sigma_1^2) p(\sigma_2^2) d\theta_0 d\phi_0 d\sigma_1^2 d\sigma_2^2 d\mu_1 d\tau_1^2 d\mu_2 d\tau_2^2. \end{aligned}$$

Now, having chosen the prior for σ_1^2 , the priors for (μ_1, τ_1^2) can be specified by matching $E(f_1(\cdot; G_1, \sigma_1^2))$ with rough prior guesses at the center and range of the data from the first population. After the

priors for (μ_1, τ_1^2) are determined, a similar approach can be used for the priors of (μ_2, τ_2^2) . For a less informative, and simpler, specification, we replace $E(f_2(\cdot; G_1, G_2, \sigma_2^2))$ with the corresponding prior predictive density under independent DP mixture prior models for F_1 and F_2 , i.e., with $E(f_2(\cdot; G_2, \sigma_2^2))$ that has the same form with $E(f_1(\cdot; G_1, \sigma_1^2))$. Then, the same priors for (μ_1, τ_1^2) and (μ_2, τ_2^2) can be used based on a proxy for the center and range of the data from both populations.

2.4 Regression modeling under stochastic precedence constraints

Here, we discuss a possible extension of the modeling approach for two stochastic precedence constrained distributions to survival regression settings, where the stochastic precedence restriction for the response $X = \log(T)$ is with respect to the values of a binary covariate, for instance, corresponding to different treatments. Let z_0 denote this covariate (with values $z_0 = 1, 2$) and \mathbf{z} be the vector of other covariates, which are common to both groups of responses induced by z_0 .

It is useful to recall that the stochastic order restriction can be readily incorporated in traditional survival regression models, e.g., accelerated failure time or proportional hazards models. For instance, under the accelerated failure time regression setting, $X = z_0\beta_0 + \mathbf{z}'\boldsymbol{\beta} + \varepsilon$, where the response distribution is defined through a baseline survival function, $S_0(t)$, for $\exp(\varepsilon)$. (The vector \mathbf{z} can be augmented with a vector of ones so that $\boldsymbol{\beta}$ includes an intercept term.) Then, the survival function for responses associated with the first group (where $z_0 = 1$) is given by $S_1(t) = S_0(t \exp(-\beta_0 - \mathbf{z}'\boldsymbol{\beta}))$, whereas for $z_0 = 2$, $S_2(t) = S_0(t \exp(-2\beta_0 - \mathbf{z}'\boldsymbol{\beta}))$. Hence, the prior restriction $\beta_0 > 0$ implies stochastic ordering, $X_1 \leq_{st} X_2$, for the response random variables $X_\ell = \log(T_\ell)$, $\ell = 1, 2$, corresponding to the two groups. In particular, if a $N(0, \sigma^2)$ distribution is assumed for ε (implied by a lognormal baseline survival distribution), we obtain $N(\theta_\ell + \mathbf{z}'\boldsymbol{\beta}, \sigma^2)$ distributions for X_ℓ , $\ell = 1, 2$, where $\theta_1 = \beta_0$ and $\theta_2 = 2\beta_0$, which are stochastically ordered when $\beta_0 > 0$. The more general assumption of a $N(0, \sigma_{z_0}^2)$ distribution for ε results in the stochastic precedence order constraint $X_1 \leq_{sp} X_2$, again, under the $\beta_0 > 0$ prior restriction.

A semiparametric DP mixture extension of the parametric model above can be developed fol-

lowing an approach similar to the one in Section 2.1. In this case, the mixture model for X_ℓ assumes the form $F_\ell(x; H_\ell, \sigma_\ell^2, \boldsymbol{\beta}) = \int F_N(x; \theta + \mathbf{z}'\boldsymbol{\beta}, \sigma_\ell^2) dH_\ell(\theta)$, $\ell = 1, 2$, with $H_1 \leq_{st} H_2$. The proof of the lemma in Appendix A can be extended to show that $X_1 \leq_{sp} X_2$ (and thus $T_1 \leq_{sp} T_2$), that is, we obtain stochastic precedence constrained regressions under the two groups defined by covariate z_0 . The methods of Section 2.2 (and Section 2.3) can also be readily modified to develop predictive inference under this regression setting.

3 Data Illustrations

3.1 A survival analysis data example

For an application of the model to a survival analysis setting, we consider a data set involving survival time in days after bone marrow transplantation for treatment of acute leukemia. The data samples three populations, an ALL (*acute lymphoblastic leukemia*) group, an AML (*acute myelocytic leukemia*) low-risk group, and an AML high-risk group. The data set is provided in Klein and Moeschberger (1997, Section 1.3) where further details can be found. Here, we focus on the ALL group and the AML high-risk group. Both sample sizes are fairly small, 38 for the ALL group, including 14 right censored survival times, and 45 for the AML high-risk group, including 11 right censored observations. The Kaplan-Meier estimates for the corresponding survival functions are plotted in Figure 1, with “+” denoting the censored survival times for each group.

The data suggest larger survival times for the ALL group compared to the AML high-risk group. Hence, to illustrate inference for two survival distributions subject to the stochastic precedence constraint, we apply model (1) with $F_1(\cdot; G_1, \sigma_1^2)$ and $F_2(\cdot; G_1, G_2, \sigma_2^2)$ representing the survival distributions on the log scale for the AML high-risk group and ALL group, respectively.

We adopt priors following the approach of Section 2.3. In particular, we take: $N(5, 4)$ priors for μ_1 and μ_2 ; inv-gamma(2, 5) priors for τ_1^2 and τ_2^2 ; inv-gamma(2, 4) priors for σ_1^2 and σ_2^2 ; and gamma(2, 2) priors for α_1 and α_2 . Given the small sample sizes in this example, there was limited

learning for some of the DP prior hyperparameters, in particular, for the τ_ℓ^2 and α_ℓ , $\ell = 1, 2$. However, posterior inference for functionals of the mixture distributions was fairly robust to more dispersed priors that were studied for all hyperparameters.

Figure 1 plots point (posterior mean) and 95% interval estimates for the survival functions. The uncertainty bands are compatible with the small sample sizes and the level of censoring. In fact, in the AML high-risk group, the censored times are larger than all but one of the observed survival times. This suggests the possibility of a heavy tail for this population, which is supported by the posterior mean estimate of the AML high-risk group survival function. Figure 1 shows also the posterior densities for median survival time under the two groups.

Finally, with regard to analysis of the data without an order restriction, note that Gelfand and Kottas (2002) studied this data set also working on the log scale, but with unrestricted location-scale DP mixtures of normals. Based on that model, point (posterior median) and 95% interval estimates for the difference of median survival times between the ALL and AML high-risk group were 271.7 and $(-52.8, 1123.3)$. The corresponding posterior estimates under the stochastic precedence DP mixture model are 153.2 and $(-1.4, 914.1)$. The reduced uncertainty in the posterior distribution for the contrast of median survival times indicates that the incorporation of the stochastic precedence restriction in the modeling has improved estimation efficiency.

3.2 ROC data application

The evaluation of the discriminatory ability of a continuous diagnostic measure is an important task in both human and veterinary epidemiologic research. Here, we consider the *gold standard* setting under which infection (or disease) status is assumed known, and thus the data comprise samples of n_1 and n_2 individuals drawn from the non-infected and infected populations, respectively. A continuous diagnostic test is applied to all sampled individuals, resulting in $n_1 + n_2$ test outcomes. We generically refer to the data as *serology scores*. Serology scores measure the concentration of antigen-specific antibodies in serum. Commonly used continuous diagnostic measures result in

an optical density value or a serum-to-positive ratio for an enzyme linked immunosorbent assay (ELISA) serological test. A relatively large serology score indicates that the test detected a high concentration of analytes that are suggestive of infection presence. A relatively low serology score indicates the absence of such analytes. Let X_1 and X_2 be the random variables that represent serology scores of the non-infected and infected populations, respectively, and denote by F_1 and F_2 the corresponding distribution functions.

The ROC curve is a commonly utilized graphical measure of the accuracy of a continuous diagnostic test. It represents a plot of all possible pairs of true positive probability versus false positive probability across all cutoff values k that could be used to dichotomize the data into test positive or negative categories. That is, the ROC curve represents the plot $(1 - F_1(k), 1 - F_2(k))$ for all cutoff values k , and is thus defined by $\text{ROC}(u) = 1 - F_2(F_1^{-1}(1 - u))$, $u \in (0, 1)$. A standard summary performance measure based on the ROC curve is the area under the curve, $\text{AUC} = \int_0^1 \text{ROC}(u) du$. The AUC has a useful interpretation as the probability that a randomly selected infected individual has a serology score that is greater than that for a randomly selected non-infected individual, i.e., $\text{AUC} = \Pr(X_1 \leq X_2)$ (e.g., Bamber, 1975).

In practice, distributions F_1 and F_2 often exhibit non-standard features such as multimodality and skewness. This is especially true for the distribution of serology scores for the infected population, which is typically a composite of individuals in different stages of infection. In this case, individuals in an advanced infection stage are expected to have higher serology scores as compared to newly infected individuals. In general, parametric distributions will not be sufficiently flexible to model F_1 and F_2 . Indeed, there is a vast literature on nonparametric frequentist techniques (e.g., Pepe, 2003), whereas the amount of existing Bayesian work is, by comparison, limited. Bayesian nonparametric methods include Erkanli et al. (2006), based on normal DP mixtures, and Hanson, Kottas and Branscum (2008) where both DP mixture prior and mixture of Pólya tree prior models were used subject to the stochastic order restriction $F_1 \leq_{st} F_2$.

From a biological point of view, incorporating some form of stochastic relationship in the model

for F_1 and F_2 is essentially always appropriate, since serologic values for infected individuals tend to be larger than serologic values for non-infected individuals, provided the diagnostic test has reasonable discriminatory ability. In fact, $AUC \geq 0.5$ specifies a natural constraint on ROC curves that effectively any diagnostic test must satisfy. But then, stochastic precedence, $F_1 \leq_{sp} F_2$, emerges as a key model restriction that can be studied alternative to, or in conjunction with, the stochastic order constraint.

To illustrate, we consider one of the data sets analyzed in Hanson, Kottas and Branscum (2008) involving a commercially available ELISA kit (developed by the Synbiotic Corp. in San Diego, California) designed to detect antibodies to Johne’s disease (*Mycobacterium avium paratuberculosis*, MAP) in dairy cattle. In the U.S., Johne’s disease is an endemic, incurable wasting disease that leads to appreciable annual economic loss sustained by the dairy industry. The data set comprises log-transformed serology scores from $n_1 = 345$ non-infected and $n_2 = 258$ infected cows. The non-infected cows came from 7 Minnesota herds that satisfied certain disease freedom criteria. Infected cows came from 7 Wisconsin herds with positive Johne’s disease herd level prevalence; individual cows from infected herds were defined to be cases if MAP organisms were identified through fecal culture. The data is part of a study conducted by Collins et al. (2005), where details on data collection and diagnostic testing procedures can be found.

We employ the stochastic precedence constrained DP mixture model (1), where $F_1(\cdot; G_1, \sigma_1^2)$ and $F_2(\cdot; G_1, G_2, \sigma_2^2)$ correspond to the serology score distributions (on the log scale) for the non-infected and infected groups, respectively. Following again the approach of Section 2.3, the priors used in the analysis were as follows: $\mu_1, \mu_2 \sim N(2, 4)$, $\tau_1^2, \tau_2^2 \sim \text{inv-gamma}(2, 5)$, $\alpha_1, \alpha_2 \sim \text{gamma}(5, 0.5)$, and $\sigma_1^2, \sigma_2^2 \sim \text{inv-gamma}(2, 2.5)$. There was prior to posterior learning for all model hyperparameters, the more sensitive to the prior choice being τ_1^2 and τ_2^2 . Moreover, posterior inference for the mixture distributions, including all the results discussed below, was essentially unaffected from choices involving higher levels of prior dispersion for $\mu_1, \mu_2, \tau_1^2, \tau_2^2$ and σ_1^2, σ_2^2 , as well as less dispersed priors for α_1, α_2 .

In Figure 2 we plot posterior mean and 95% interval estimates for the non-infected and infected group density functions. These are obtained as discussed in Section 2.2, in particular, the solid lines correspond to the estimates for $E(f_1(x_0; G_1, \sigma_1^2) \mid \text{data})$ and $E(f_2(x_0; G_1, G_2, \sigma_2^2) \mid \text{data})$ over a grid of x_0 values in $(-1.5, 6.5)$. Contrasting with the prior predictive densities indicates the amount of prior to posterior learning, as well as a relatively non-informative prior specification. The model captures the bimodal shape of the infected group density, and, consistent with the available sample sizes, yields narrower uncertainty bands for the non-infected group density. The top panels of Figure 3 show the analogous inference for the non-infected and infected group distribution functions. Also included in Figure 3 are posterior mean and 95% interval estimates for the ROC curve, and the posterior density for the AUC, both suggesting a moderately accurate ELISA test. Comparison of the prior and posterior AUC densities indicates again a fair amount of learning from the data.

We also consider comparison with the stochastically ordered DP mixture model discussed at the end of Section 2.1, and applied to this data set in Hanson, Kottas and Branscum (2008). Under this model, the serology score distribution (on the log scale) for the non-infected group is assumed stochastically smaller than the one for the infected group. The priors for μ_1 , μ_2 , τ_1^2 , τ_2^2 , and α_1 , α_2 were the same with the ones above for the stochastic precedence constrained model, and the common σ^2 parameter was assigned an inv-gamma(2, 2.5) prior. Figure 4 compares the posterior mean estimates for the non-infected and infected group distribution functions under the two models. By construction, the stochastic order model forces uniform domination of one distribution function by the other, whereas under the stochastic precedence model, the estimated distribution functions cross each other in their left tails, i.e., for $\log(\text{serology score})$ values up to about 1. Noting that such values are within the range of the data for both groups, suggests that stochastic precedence may be a more appropriate constraint than stochastic order for this data set.

This is further supported by formal model comparison between the stochastic order and stochastic precedence models, using the posterior predictive loss criterion from Gelfand and Ghosh (1998). In general, this model comparison approach is based on the mean, $E^{(\mathcal{M})}(x_i^* \mid \text{data})$, and variance,

$\text{Var}^{(\mathcal{M})}(x_i^* \mid \text{data})$, under model \mathcal{M} , of the posterior predictive distribution for *replicate* responses x_i^* corresponding to observed responses x_i , $i = 1, \dots, n$. The criterion favors the model, \mathcal{M} , that minimizes the predictive loss measure, $D(\mathcal{M}) = P(\mathcal{M}) + G(\mathcal{M})$, where $P(\mathcal{M}) = \sum_{i=1}^n \text{Var}^{(\mathcal{M})}(x_i^* \mid \text{data})$ is a penalty term for model complexity, and $G(\mathcal{M}) = \sum_{i=1}^n \{x_i - \text{E}^{(\mathcal{M})}(x_i^* \mid \text{data})\}^2$ is a goodness-of-fit term. Under our setting, since both models induce structured dependence in the distributions for the non-infected and infected groups, it seems appropriate to consider the two components of the criterion for all responses from both groups. Hence, the goodness-of-fit term comprises $G(\mathcal{M}) = \sum_{i=1}^{n_1} \{x_{1i} - \text{E}^{(\mathcal{M})}(x_{1i}^* \mid \text{data})\}^2 + \sum_{j=1}^{n_2} \{x_{2j} - \text{E}^{(\mathcal{M})}(x_{2j}^* \mid \text{data})\}^2$, and the penalty term $P(\mathcal{M}) = \sum_{i=1}^{n_1} \text{Var}^{(\mathcal{M})}(x_{1i}^* \mid \text{data}) + \sum_{j=1}^{n_2} \text{Var}^{(\mathcal{M})}(x_{2j}^* \mid \text{data})$. All the required expressions are readily estimated under both models by sampling from the posterior predictive distributions $P(x_{1i}^* \mid \text{data})$, for $i = 1, \dots, n_1$, and $P(x_{2j}^* \mid \text{data})$, for $j = 1, \dots, n_2$. In particular, under the stochastic precedence model, $P(x_{1i}^* \mid \text{data}) = \int F_N(x_{1i}^*; \theta_i, \sigma_1^2) p(\theta_i, \sigma_1^2 \mid \text{data}) d\theta_i d\sigma_1^2$, and $P(x_{2j}^* \mid \text{data}) = \int F_N(x_{2j}^*; \max\{\theta_{n_1+j}, \phi_j\}, \sigma_2^2) p(\theta_{n_1+j}, \phi_j, \sigma_2^2 \mid \text{data}) d\theta_{n_1+j} d\phi_j d\sigma_2^2$. Based on the results, reported in Table 1, the stochastic order model performs slightly better with regard to the penalty term, whereas the stochastic precedence model fares better with the goodness-of-fit term; overall, the criterion favors the stochastic precedence DP mixture model.

4 Summary

Stochastic precedence relaxes the restriction of the familiar stochastic ordering constraint, and thus, provides a practically useful setting for comparison of two distributions that are anticipated to be ordered in a stochastic fashion. We have developed a semiparametric Bayesian model for two stochastic precedence constrained continuous distributions, along with the corresponding inference framework. The modeling approach is based on location normal Dirichlet process mixtures, which are appropriately structured to ensure that all realizations from the prior probability model satisfy the stochastic precedence restriction. We have presented applications of the methodology to

problems from survival analysis and epidemiologic research, including two data illustrations, one comprising survival times after bone marrow transplantation for treatment of leukemia, and one involving diagnostic test scores for Johne's disease in dairy cattle.

Acknowledgements

This work was supported in part by the National Science Foundation under award DEB 0727543.

Appendix A: Proof of the lemma of Section 2.1

Let X_ℓ , $\ell = 1, 2$, be random variables, defined on a common probability space, with distributions $F_\ell(\cdot) \equiv F_\ell(\cdot; H_\ell, \sigma_\ell^2) = \int F_N(\cdot; \theta, \sigma_\ell^2) dH_\ell(\theta)$, $\ell = 1, 2$. Assuming that $H_1 \leq_{st} H_2$, we need to prove that $F_1(\cdot; H_1, \sigma_1^2) \leq_{sp} F_2(\cdot; H_2, \sigma_2^2)$, i.e., that $Q = \Pr(X_1 \leq X_2; H_1, H_2, \sigma_1^2, \sigma_2^2) \geq 0.5$.

The probability of interest can be expressed as follows:

$$\begin{aligned}
Q &= \mathbb{E}^{F_2(\cdot; H_2, \sigma_2^2)} \{F_1(X_2; H_1, \sigma_1^2)\} \\
&= \int_{-\infty}^{\infty} F_1(u; H_1, \sigma_1^2) f_2(u; H_2, \sigma_2^2) du \\
&= \int_{-\infty}^{\infty} \left\{ \int_{\theta_1 \in \mathbb{R}} F_N(u; \theta_1, \sigma_1^2) dH_1(\theta_1) \right\} \left\{ \int_{\theta_2 \in \mathbb{R}} f_N(u; \theta_2, \sigma_2^2) dH_2(\theta_2) \right\} du \\
&= \int_{\theta_1 \in \mathbb{R}} \int_{\theta_2 \in \mathbb{R}} \left\{ \int_{-\infty}^{\infty} F_N(u; \theta_1, \sigma_1^2) f_N(u; \theta_2, \sigma_2^2) du \right\} dH_1(\theta_1) dH_2(\theta_2) \\
&= \int_{\theta_1 \in \mathbb{R}} \int_{\theta_2 \in \mathbb{R}} \Pr(Y_1 \leq Y_2; \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) dH_1(\theta_1) dH_2(\theta_2)
\end{aligned}$$

where Y_1 and Y_2 are random variables, defined on the same probability space, which are conditionally independent, given θ_1 , θ_2 , σ_1^2 and σ_2^2 , with distributions $N(\theta_1, \sigma_1^2)$ and $N(\theta_2, \sigma_2^2)$, respectively. Therefore, we will have $\Pr(Y_1 \leq Y_2; \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) \geq 0.5$ if and only if $\theta_1 \leq \theta_2$ with probability 1.

Next, consider $\theta_1 \sim H_1$ and $\theta_2 \sim H_2$, and recall the characterization of the stochastic order restriction $H_1 \leq_{st} H_2$, which is assumed for the mixing distributions H_1 and H_2 . Based on this characterization, $H_1 \leq_{st} H_2$ if and only if there exist copies θ'_1 and θ'_2 of θ_1 and θ_2 (i.e., random variable θ'_ℓ has the same distribution with θ_ℓ , $\ell = 1, 2$), which are defined on the same probability

space, and $\theta'_1 \leq \theta'_2$ with probability 1 (see, e.g., Shaked and Shanthikumar, 1994, Theorem 1.A.1).

Hence, using the stochastic precedence property of the normal distribution, we finally obtain

$$\begin{aligned} \Pr(X_1 \leq X_2; H_1, H_2, \sigma_1^2, \sigma_2^2) &= \int_{\theta_1 \in \mathbb{R}} \int_{\theta_2 \in \mathbb{R}} \Pr(Y_1 \leq Y_2; \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) dH_1(\theta_1) dH_2(\theta_2) \\ &\geq 0.5 \int_{\theta_1 \in \mathbb{R}} \int_{\theta_2 \in \mathbb{R}} dH_1(\theta_1) dH_2(\theta_2) = 0.5. \end{aligned}$$

Appendix B: MCMC posterior simulation methods

Here, we provide details on the MCMC posterior simulation method from the marginal posterior $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \mid \text{data})$ of model (2) developed in Sections 2.1 and 2.2.

MCMC sampling for fully observed responses. Consider first the case where there are no censored observations among the data. Then, we have

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \mid \text{data}) &\propto p(\boldsymbol{\psi}_1) p(\boldsymbol{\psi}_2) p(\sigma_1^2) p(\sigma_2^2) p(\boldsymbol{\theta} \mid \boldsymbol{\psi}_1) p(\boldsymbol{\phi} \mid \boldsymbol{\psi}_2) \\ &\quad \prod_{i=1}^{n_1} f_N(x_{1i}; \theta_i, \sigma_1^2) \prod_{j=1}^{n_2} f_N(x_{2j}; \max\{\theta_{n_1+j}, \phi_j\}, \sigma_2^2) \end{aligned}$$

where $p(\boldsymbol{\psi}_\ell)$, $\ell = 1, 2$, are the priors for the DP hyperparameters discussed in Section 2.1, and $p(\sigma_1^2)$, $p(\sigma_2^2)$ are the inverse gamma priors for σ_1^2 , σ_2^2 . Moreover, $p(\boldsymbol{\theta} \mid \boldsymbol{\psi}_1)$ and $p(\boldsymbol{\phi} \mid \boldsymbol{\psi}_2)$ denote the priors for the vectors of mixing parameters induced by the DP priors after marginalizing G_1 and G_2 in model (2). These prior distributions are built from a generalized Pólya urn scheme (Blackwell and MacQueen, 1973). In particular, for $\boldsymbol{\theta}$, θ_1 follows a $N(\mu_1, \tau_1^2)$ distribution, and for any $i = 2, \dots, n_1 + n_2$, θ_i , conditionally on $\theta_1, \dots, \theta_{i-1}$, follows a mixed distribution with point masses $(\alpha_1 + i - 1)^{-1}$ at θ_r , $r = 1, \dots, i - 1$, and continuous mass $\alpha_1(\alpha_1 + i - 1)^{-1}$ on the $N(\mu_1, \tau_1^2)$ distribution. Hence,

$$p(\boldsymbol{\theta} \mid \boldsymbol{\psi}_1) = f_N(\theta_1; \mu_1, \tau_1^2) \prod_{i=2}^{n_1+n_2} \left\{ \alpha_1(\alpha_1 + i - 1)^{-1} f_N(\theta_i; \mu_1, \tau_1^2) + (\alpha_1 + i - 1)^{-1} \sum_{r=1}^{i-1} \delta_{\theta_r}(\theta_i) \right\} \quad (\text{B.1})$$

and, analogously,

$$p(\boldsymbol{\phi} \mid \boldsymbol{\psi}_2) = f_N(\phi_1; \mu_2, \tau_2^2) \prod_{i=2}^{n_2} \left\{ \alpha_2(\alpha_2 + i - 1)^{-1} f_N(\phi_i; \mu_2, \tau_2^2) + (\alpha_2 + i - 1)^{-1} \sum_{r=1}^{i-1} \delta_{\phi_r}(\phi_i) \right\}.$$

To sample from $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \mid \text{data})$, we use an MCMC algorithm that combines techniques from Escobar and West (1995) and Neal (2000). Regarding the MCMC updates for $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, note that based on (B.1), the prior full conditional for each θ_i , $p(\theta_i \mid \{\theta_r : r \neq i\}, \boldsymbol{\psi}_1)$, $i = 1, \dots, n_1 + n_2$, has point masses $(\alpha_1 + n_1 + n_2 - 1)^{-1}$ at θ_r , $r \neq i$, and continuous mass $\alpha_1(\alpha_1 + n_1 + n_2 - 1)^{-1}$ on the $N(\mu_1, \tau_1^2)$ distribution. Analogously, each ϕ_j , $j = 1, \dots, n_2$, has a mixed prior full conditional distribution, $p(\phi_j \mid \{\phi_r : r \neq j\}, \boldsymbol{\psi}_2)$, with point masses $(\alpha_2 + n_2 - 1)^{-1}$ at ϕ_r , $r \neq j$, and continuous mass $\alpha_2(\alpha_2 + n_2 - 1)^{-1}$ on the $N(\mu_2, \tau_2^2)$ distribution. Therefore, it is straightforward to sample directly from the posterior full conditional for each θ_i , $i = 1, \dots, n_1$, since it is a mixed distribution with point masses at (the distinct values among) the θ_r , $r \neq i$, and continuous mass on a normal distribution with mean $(x_{1i}\tau_1^2 + \mu_1\sigma_1^2)/(\sigma_1^2 + \tau_1^2)$ and variance $\tau_1^2\sigma_1^2/(\sigma_1^2 + \tau_1^2)$. The weight associated with this normal distribution is proportional to $\alpha_1\{2\pi(\sigma_1^2 + \tau_1^2)\}^{-1/2} \exp(-0.5(x_{1i} - \mu_1)^2/(\sigma_1^2 + \tau_1^2))$; the weights corresponding to the θ_r , $r \neq i$, are proportional to $f_N(x_{1i}; \theta_r, \sigma_1^2)$.

For each $j = 1, \dots, n_2$, the posterior full conditional for the pair of latent mixing parameters (θ_{n_1+j}, ϕ_j) is proportional to

$$f_N(x_{2j}; \max\{\theta_{n_1+j}, \phi_j\}, \sigma_2^2)p(\theta_{n_1+j} \mid \{\theta_r : r \neq n_1 + j\}, \boldsymbol{\psi}_1)p(\phi_j \mid \{\phi_r : r \neq j\}, \boldsymbol{\psi}_2).$$

We update each pair (θ_{n_1+j}, ϕ_j) with a Metropolis-Hastings (M-H) step, which involves proposed draws $(\tilde{\theta}_{n_1+j}, \tilde{\phi}_j)$ from $p(\cdot \mid \{\theta_r : r \neq n_1 + j\}, \boldsymbol{\psi}_1) \times p(\cdot \mid \{\phi_r : r \neq j\}, \boldsymbol{\psi}_2)$ that are accepted with probability $\min\{1, f_N(x_{2j}; \max\{\tilde{\theta}_{n_1+j}, \tilde{\phi}_j\}, \sigma_2^2)/f_N(x_{2j}; \max\{\theta_{n_1+j}^{(\text{old})}, \phi_j^{(\text{old})}\}, \sigma_2^2)\}$, where $(\theta_{n_1+j}^{(\text{old})}, \phi_j^{(\text{old})})$ is the current state of the chain.

Note that the discreteness of the DP priors for G_1 and G_2 induces a clustering of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ in their prior, and thus also in their posterior. In particular, once all the updates above for the θ_i , $i = 1, \dots, n_1$, and for the (θ_{n_1+j}, ϕ_j) , $j = 1, \dots, n_2$, are completed, we obtain the number of and values of the distinct components in $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. Denote these by n_θ^* ($\leq n_1 + n_2$) and $\{\theta_k^* : k = 1, \dots, n_\theta^*\}$ for vector $\boldsymbol{\theta}$, and by n_ϕ^* ($\leq n_2$) and $\{\phi_k^* : k = 1, \dots, n_\phi^*\}$ for vector $\boldsymbol{\phi}$.

The posterior full conditional for μ_1 is proportional to $p(\mu_1) \prod_{k=1}^{n_\theta^*} f_N(\theta_k^*; \mu_1, \tau_1^2)$ resulting in a normal distribution with mean $(c_1\tau_1^2 + d_1 \sum_{k=1}^{n_\theta^*} \theta_k^*) / (d_1 n_\theta^* + \tau_1^2)$ and variance $d_1\tau_1^2 / (d_1 n_\theta^* + \tau_1^2)$. Similarly, the full conditional for μ_2 is normal with mean $(c_2\tau_2^2 + d_2 \sum_{k=1}^{n_\phi^*} \phi_k^*) / (d_2 n_\phi^* + \tau_2^2)$ and variance $d_2\tau_2^2 / (d_2 n_\phi^* + \tau_2^2)$. The posterior full conditional for τ_1^2 is proportional to $p(\tau_1^2) \prod_{k=1}^{n_\theta^*} f_N(\theta_k^*; \mu_1, \tau_1^2)$, which yields an inverse gamma distribution with shape parameter $w_1 + 0.5n_\theta^*$ and rate parameter $e_1 + 0.5 \sum_{k=1}^{n_\theta^*} (\theta_k^* - \mu_1)^2$. Similarly, the full conditional for τ_2^2 is an inv-gamma($w_2 + 0.5n_\phi^*, e_2 + 0.5 \sum_{k=1}^{n_\phi^*} (\phi_k^* - \mu_2)^2$) distribution. The DP precision parameters α_1 and α_2 are updated using the data augmentation technique from Escobar and West (1995).

Finally, σ_1^2 has an inv-gamma($a_{\sigma_1} + 0.5n_1, b_{\sigma_1} + 0.5 \sum_{i=1}^{n_1} (x_{1i} - \theta_i)^2$) posterior full conditional distribution, and σ_2^2 an inv-gamma($a_{\sigma_2} + 0.5n_2, b_{\sigma_2} + 0.5 \sum_{j=1}^{n_2} (x_{2j} - \max\{\theta_{n_1+j}, \phi_j\})^2$) posterior full conditional.

MCMC method for censored data. The approach discussed above can be extended to handle censoring in a relatively straightforward fashion. We provide details for (fixed) right censoring, although the method is similar for left or interval censored observations.

Consider, as in Section 2.2, log transformed data from distribution F_1 that comprise observed survival times $x_{1i} = \log(t_{1i})$, $i = 1, \dots, n_{1o}$, and right censored survival times $x_{1k}^+ = \log(t_{1k}^+)$, $k = 1, \dots, n_{1c}$, with $n_1 = n_{1o} + n_{1c}$. Similarly, the data vector, on the log scale, from distribution F_2 contains observed survival times $x_{2j} = \log(t_{2j})$, $j = 1, \dots, n_{2o}$, and right censored survival times $x_{2m}^+ = \log(t_{2m}^+)$, $m = 1, \dots, n_{2c}$, with $n_2 = n_{2o} + n_{2c}$. In this case, the marginal posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \mid \text{data})$ is proportional to

$$p(\boldsymbol{\psi}_1)p(\boldsymbol{\psi}_2)p(\sigma_1^2)p(\sigma_2^2)p(\boldsymbol{\theta} \mid \boldsymbol{\psi}_1)p(\boldsymbol{\phi} \mid \boldsymbol{\psi}_2) \prod_{i=1}^{n_{1o}} f_N(x_{1i}; \theta_i, \sigma_1^2) \prod_{k=1}^{n_{1c}} \{1 - F_N(x_{1k}^+; \theta_k, \sigma_1^2)\} \\ \prod_{j=1}^{n_{2o}} f_N(x_{2j}; \max\{\theta_{n_1+j}, \phi_j\}, \sigma_2^2) \prod_{m=1}^{n_{2c}} \{1 - F_N(x_{2m}^+; \max\{\theta_{n_1+m}, \phi_m\}, \sigma_2^2)\}$$

with all the priors as before.

The updates for $\alpha_1, \alpha_2, \mu_1, \mu_2, \tau_1^2, \tau_2^2$, for the θ_i , $i = 1, \dots, n_{1o}$, and for the (θ_{n_1+j}, ϕ_j) ,

$j = 1, \dots, n_{2o}$, remain the same as in the case without censoring. The updates for each (θ_{n_1+m}, ϕ_m) , $m = 1, \dots, n_{2c}$, involve a similar M-H step with the one used without censoring replacing the normal density in the acceptance probability with $1 - F_N(x_{2m}^+; \max\{\theta_{n_1+m}, \phi_m\}, \sigma_2^2)$.

For each $k = 1, \dots, n_{1c}$, the posterior full conditional distribution for θ_k is proportional to $\{1 - F_N(x_{1k}^+; \theta_k, \sigma_1^2)\}p(\theta_k \mid \{\theta_r : r \neq k\}, \boldsymbol{\psi}_1)$, and thus, no longer easy to sample directly. We utilize a M-H step based on proposed draws $\tilde{\theta}_k$ from $p(\cdot \mid \{\theta_r : r \neq k\}, \boldsymbol{\psi}_1)$, which are accepted with probability $\min\{1, (1 - F_N(x_{1k}^+; \tilde{\theta}_k, \sigma_1^2))/(1 - F_N(x_{1k}^+; \theta_k^{(\text{old})}, \sigma_1^2))\}$, where $\theta_k^{(\text{old})}$ is the current state of the chain.

Finally, modifications are also needed in the updates for σ_1^2 and σ_2^2 . For instance, the posterior full conditional for σ_1^2 is proportional to

$$\text{inv-gamma}(\sigma_1^2; a_{\sigma_1} + 0.5n_{1o}, b_{\sigma_1} + 0.5 \sum_{i=1}^{n_{1o}} (x_{1i} - \theta_i)^2) \times \prod_{k=1}^{n_{1c}} \{1 - F_N(x_{1k}^+; \theta_k, \sigma_1^2)\}.$$

Although this full conditional is no longer available in a form that can be sampled directly, an efficient M-H step emerges by using the inverse gamma distribution above as the proposal distribution. Hence, the proposed draw $\tilde{\sigma}_1^2$ from $\text{inv-gamma}(a_{\sigma_1} + 0.5n_{1o}, b_{\sigma_1} + 0.5 \sum_{i=1}^{n_{1o}} (x_{1i} - \theta_i)^2)$ is accepted with probability $\min\{1, \prod_{k=1}^{n_{1c}} (1 - F_N(x_{1k}^+; \theta_k, \tilde{\sigma}_1^2)) / \prod_{k=1}^{n_{1c}} (1 - F_N(x_{1k}^+; \theta_k, \sigma_1^{2(\text{old})}))\}$, where $\sigma_1^{2(\text{old})}$ is the current state of the chain. For σ_2^2 , the M-H step involves proposing from $\text{inv-gamma}(a_{\sigma_2} + 0.5n_{2o}, b_{\sigma_2} + 0.5 \sum_{j=1}^{n_{2o}} (x_{2j} - \max\{\theta_{n_1+j}, \phi_j\})^2)$, with the proposed value, $\tilde{\sigma}_2^2$, accepted with probability $\min\{1, \prod_{m=1}^{n_{2c}} (1 - F_N(x_{2m}^+; \max\{\theta_{n_1+m}, \phi_m\}, \tilde{\sigma}_2^2)) / \prod_{m=1}^{n_{2c}} (1 - F_N(x_{2m}^+; \max\{\theta_{n_1+m}, \phi_m\}, \sigma_2^{2(\text{old})}))\}$, where $\sigma_2^{2(\text{old})}$ is the current state of the chain.

References

- Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to nonparametric problems. *Annals of Statistics*, 2, 1152-1174.
- Arcones, M.A., Kvam, P.H. and Samaniego, F.J. (2002). Nonparametric estimation of a distribution subject to a stochastic precedence constraint. *Journal of the American Statistical Association*, 97, 170-182.
- Arjas, E. and Gasbarra, D. (1996). Bayesian Inference of Survival Probabilities, Under Stochastic Ordering Constraints. *Journal of the American Statistical Association*, 91, 1101-1109.
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12, 387-415.
- Blackwell, D., and MacQueen, J.B. (1973). Ferguson distributions via Pólya urn schemes. *Annals of Statistics*, 1, 353-355.
- Chen, Z. and Dunson, D.B. (2004). Bayesian estimation of survival functions under stochastic precedence. *Lifetime Data Analysis*, 10, 159-173.
- Collins, M.T., Wells, S.J., Petrini, K.R., Collins, J.E., Schultz, R.D. and Whitlock, R.H. (2005). Evaluation of five antibody detection tests for diagnosis of bovine paratuberculosis. *Clinical and Diagnostic Laboratory Immunology*, 12, 685-692.
- Dunson, D.B. and Peddada, S.D. (2008). Bayesian nonparametric inference on stochastic ordering. *Biometrika*, 95, 859-874.
- Erkanli, A., Sung, M., Costello, E.J. and Angold, A. (2006). Bayesian semi-parametric ROC analysis. *Statistics in Medicine*, 25, 3905-3928.
- Escobar, M.D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 577-588.

- Evans, M., Gilula, Z., Guttman, I. and Swartz, T. (1997). Bayesian analysis of stochastically ordered distributions of categorical variables. *Journal of the American Statistical Association*, 92, 208-214.
- Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1, 209-230.
- Gelfand, A. and Ghosh, S. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85, 1-11.
- 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. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 11, 289-305.
- Gelfand, A.E., and Kuo, L. (1991). Nonparametric Bayesian bioassay including ordered polytomous response. *Biometrika*, 78, 657-666.
- Hanson, T.E., Branscum, A.J. and Johnson, W.O. (2005). Bayesian nonparametric modeling and data analysis: An introduction. In *Bayesian Thinking: Modeling and Computation* (Handbook of Statistics, volume 25), pp. 245-278. Edited by D.K. Dey and C.R. Rao. Amsterdam: Elsevier.
- Hanson, T.E., Kottas, A. and Branscum, A.J. (2008). Modeling stochastic order in the analysis of receiver operating characteristic data: Bayesian non-parametric approaches. *Applied Statistics*, 57, 207-225.
- Hoff, P.D. (2003). Bayesian Methods for Partial Stochastic Orderings. *Biometrika*, 90, 303-317.

- Karabatsos, G. and Walker, S.G. (2007). Bayesian nonparametric inference of stochastically ordered distributions, with Pólya trees and Bernstein polynomials. *Statistics & Probability Letters*, 77, 907-913.
- Klein, J.P. and Moeschberger, M.L. (1997). *Survival Analysis. Techniques for Censored and Truncated Data*. New York: Springer-Verlag.
- Kottas, A. (2006). Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *Journal of Statistical Planning and Inference*, 136, 578-596.
- Kottas, A., and Gelfand, A.E. (2001). Modeling variability order: A semiparametric Bayesian approach. *Methodology and Computing in Applied Probability*, 3, 427-442.
- Kottas, A., Branco, M.D., and Gelfand, A.E. (2002). A nonparametric Bayesian modeling approach for cytogenetic dosimetry. *Biometrics*, 58, 593-600.
- Müller, P. and Quintana, F.A. (2004). Nonparametric Bayesian data analysis. *Statistical Science*, 19, 95-110.
- Neal, R.M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9, 249-265.
- Pepe, M.S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford Statistical Science Series: Oxford University Press.
- Sethuraman, J. (1994). A constructive definition of the Dirichlet prior. *Statistica Sinica*, 4, 639-650.
- Shaked, M. and Shanthikumar, J.G. (1994). *Stochastic Orders and Their Applications*. Boston: Academic Press.

Table 1: Synbiotic ELISA test data. Results from the posterior predictive loss criterion, $D(\mathcal{M}) = G(\mathcal{M}) + P(\mathcal{M})$, for comparison of the stochastic precedence and stochastic order models. Here, the goodness-of-fit term, $G(\mathcal{M}) = G_1(\mathcal{M}) + G_2(\mathcal{M})$, where $G_1(\mathcal{M}) = \sum_{i=1}^{n_1} \{x_{1i} - E^{(\mathcal{M})}(x_{1i}^* | \text{data})\}^2$ and $G_2(\mathcal{M}) = \sum_{j=1}^{n_2} \{x_{2j} - E^{(\mathcal{M})}(x_{2j}^* | \text{data})\}^2$. Moreover, the penalty term, $P(\mathcal{M}) = P_1(\mathcal{M}) + P_2(\mathcal{M})$, where $P_1(\mathcal{M}) = \sum_{i=1}^{n_1} \text{Var}^{(\mathcal{M})}(x_{1i}^* | \text{data})$ and $P_2(\mathcal{M}) = \sum_{j=1}^{n_2} \text{Var}^{(\mathcal{M})}(x_{2j}^* | \text{data})$.

	$G(\mathcal{M})$		$P(\mathcal{M})$		$D(\mathcal{M})$
	$G_1(\mathcal{M})$	$G_2(\mathcal{M})$	$P_1(\mathcal{M})$	$P_2(\mathcal{M})$	
Stochastic precedence model	60.55	51.12	82.81	129.47	323.95
Stochastic order model	73.17	46.76	117.10	93.15	330.18

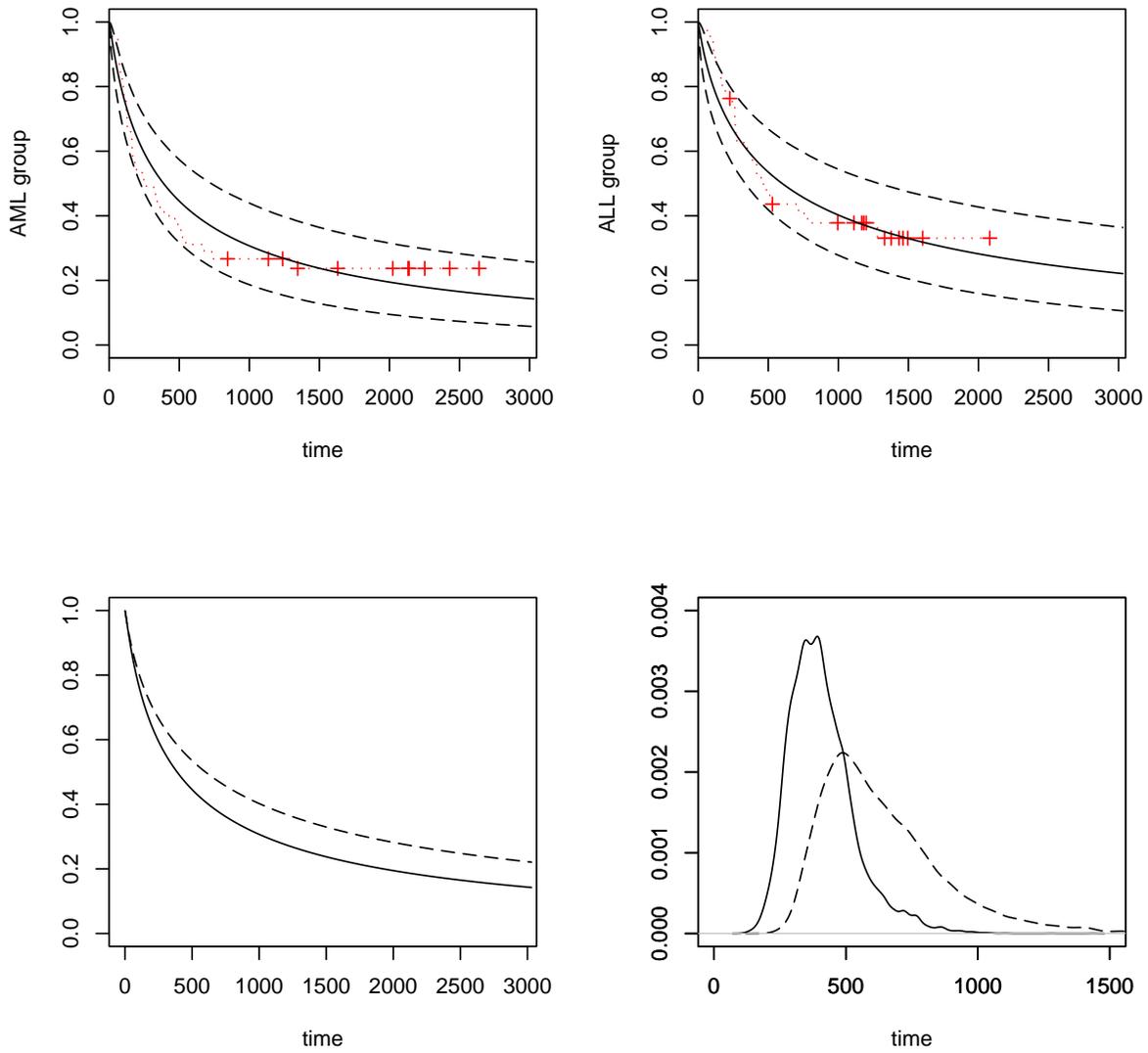


Figure 1: Bone marrow transplantation data. The upper panels include posterior mean estimates (solid lines) and 95% interval estimates (dashed lines) for the survival function of the AML high-risk and ALL groups. Also plotted are the corresponding Kaplan-Meier estimates (red dotted lines). The bottom panels compare the survival function posterior mean estimates (left panel) and the median survival time posterior densities (right panel); in both panels, the solid line corresponds to the AML high-risk group and the dashed line to the ALL group.

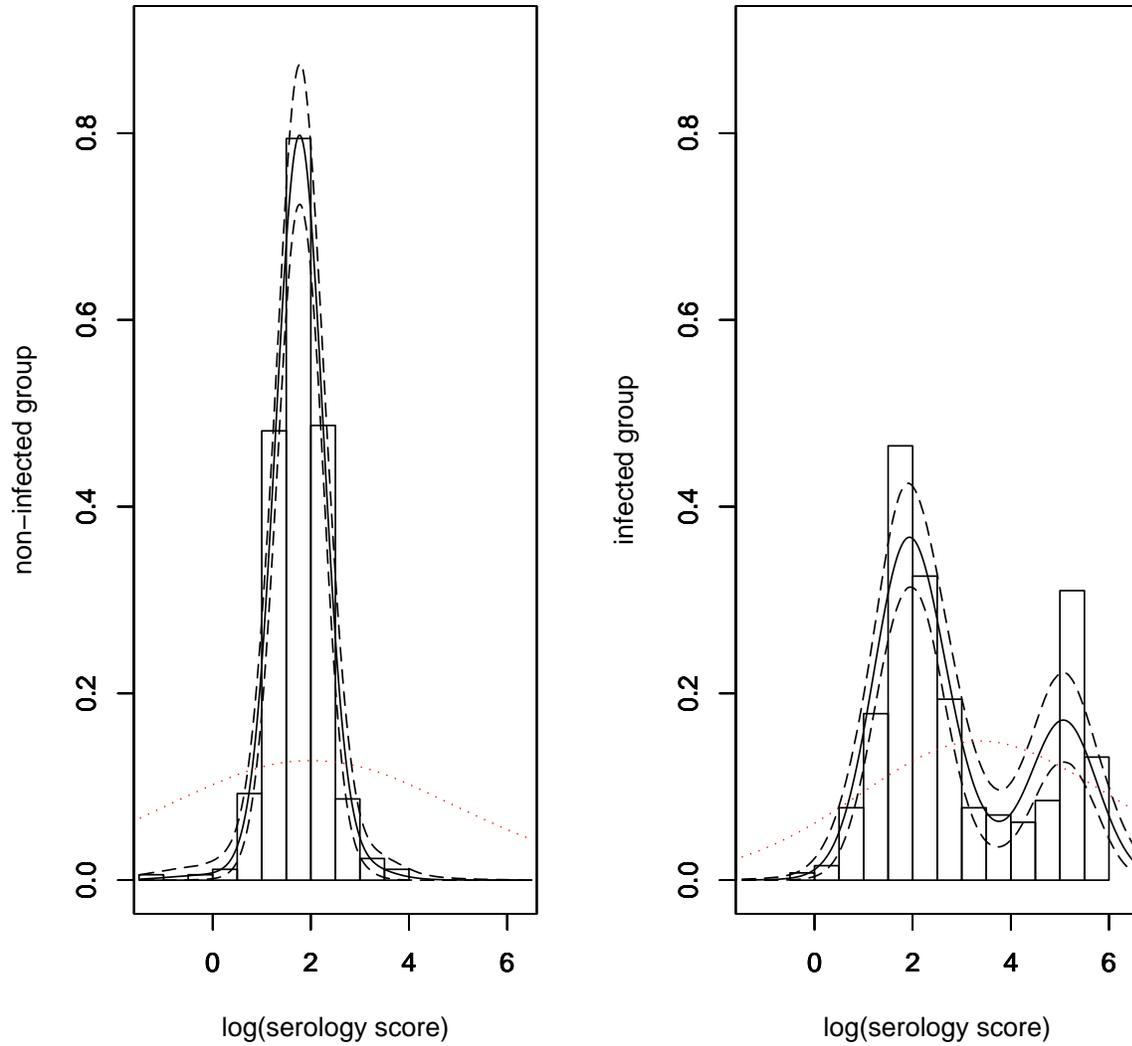


Figure 2: Synbiotic ELISA test data. Posterior mean estimates (solid lines) and 95% interval estimates (dashed lines) for the non-infected and infected group density functions, overlaid on corresponding data histograms. The red dotted lines denote the prior predictive densities.

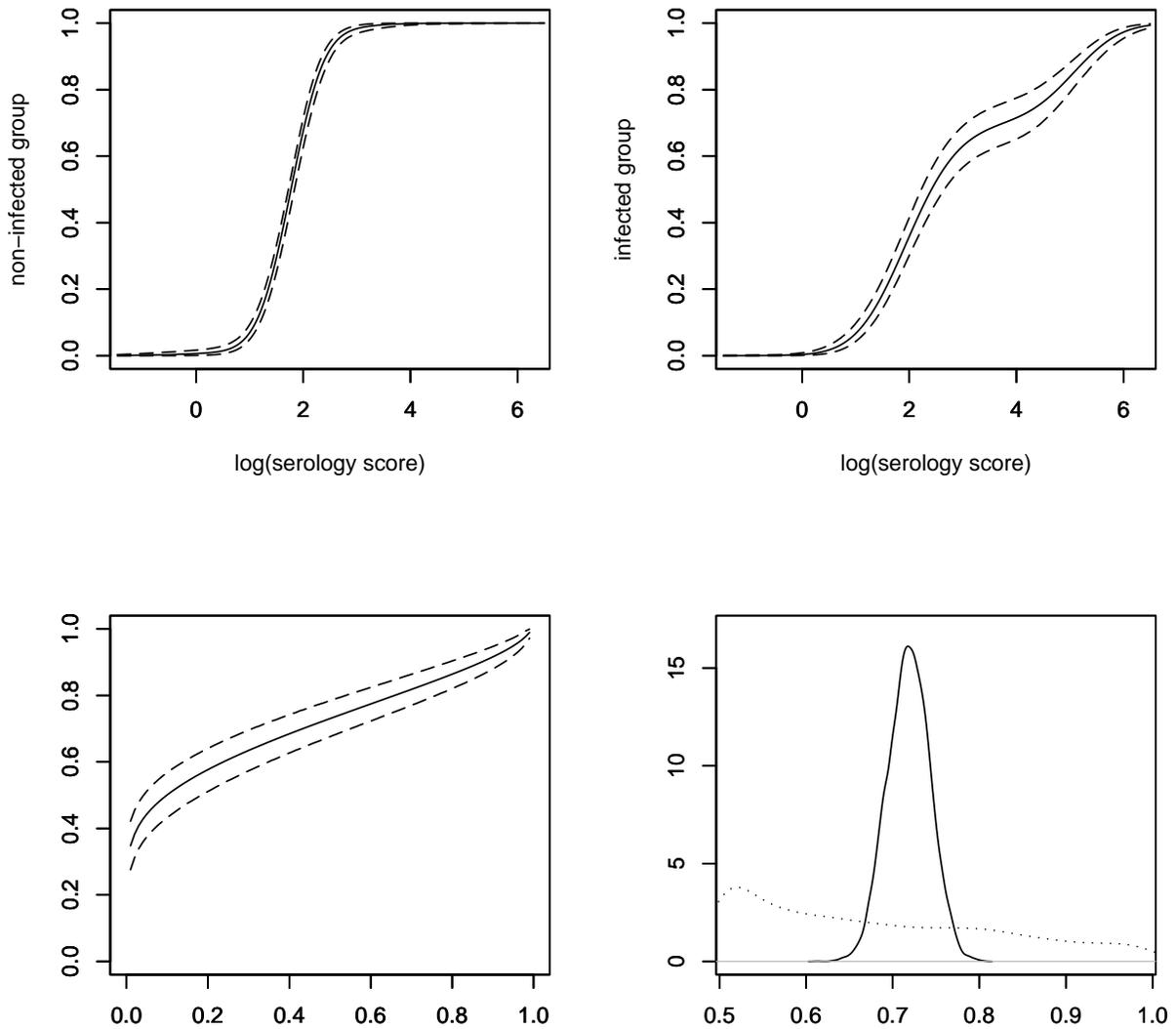


Figure 3: Synbiotic ELISA test data. Posterior mean estimates (solid lines) and 95% interval estimates (dashed lines) for the distribution function of the non-infected group (upper left panel), the distribution function of the infected group (upper right panel), and the ROC curve (lower left panel). The lower right panel shows the prior and posterior density of the AUC denoted by the dotted and solid line, respectively.

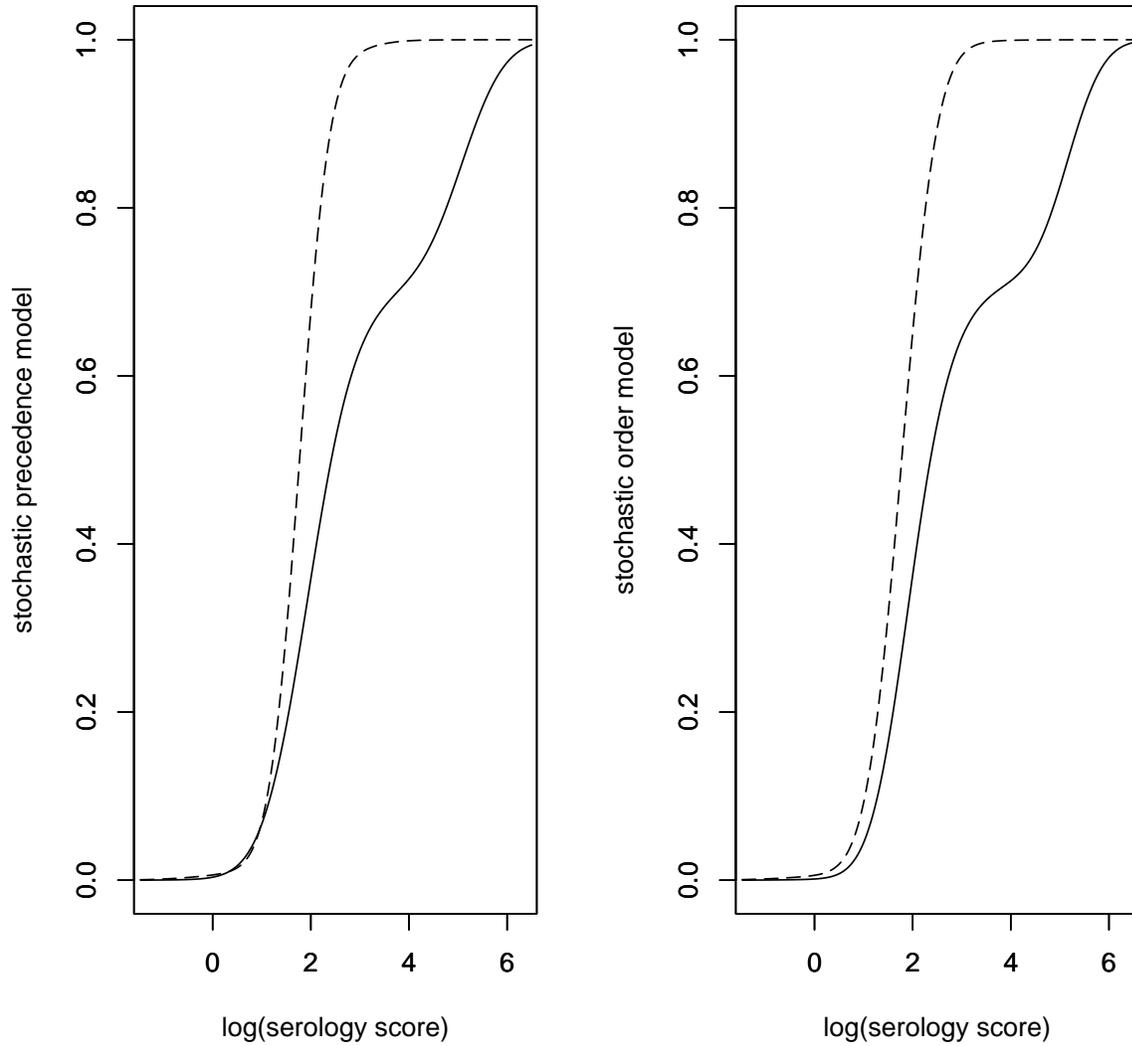


Figure 4: Synbiotic ELISA test data. Comparison of posterior mean estimates for the distribution functions of the non-infected group (dashed lines) and the infected group (solid lines), under the stochastic precedence and stochastic order models.