Appl. Statist. (2008)
57, *Part* 2, *pp.* 207–225

Modelling stochastic order in the analysis of receiver operating characteristic data: Bayesian non-parametric approaches

Timothy E. Hanson,

University of Minnesota, Minneapolis, USA

Athanasios Kottas,

University of California, Santa Cruz, USA

and Adam J. Branscum

University of Kentucky, Lexington, USA

[Received October 2006. Revised October 2007]

Summary. The evaluation of the performance of a continuous diagnostic measure is a commonly encountered task in medical research. We develop Bayesian non-parametric models that use Dirichlet process mixtures and mixtures of Polya trees for the analysis of continuous serologic data. The modelling approach differs from traditional approaches to the analysis of receiver operating characteristic curve data in that it incorporates a stochastic ordering constraint for the distributions of serologic values for the infected and non-infected populations. Biologically such a constraint is virtually always feasible because serologic values from infected individuals tend to be higher than those for non-infected individuals. The models proposed provide data-driven inferences for the infected and non-infected population distributions, and for the receiver operating characteristic curve and corresponding area under the curve. We illustrate and compare the predictive performance of the Dirichlet process mixture and mixture of Polya trees approaches by using serologic data for Johne's disease in dairy cattle.

Keywords: Area under the curve; Dirichlet process mixtures; Johne's disease; Markov chain Monte Carlo methods; Mixtures of Polya trees; Serologic data

1. Introduction

The characterization of the discriminatory ability of a continuous diagnostic measure is a commonly encountered task in both human and veterinary epidemiologic research. Traditional parametric approaches for modelling continuous serologic data, such as the binormal model where serologic values for infected and non-infected individuals are assumed to be normally distributed, are often too restrictive to capture non-standard features such as multimodality and skewness that are often seen in serologic data. The realm of Bayesian non-parametrics allows for the broadening of the class of models under consideration, and hence for the development of a widely applicable approach to the analysis of serologic data that can be used for practically any population and for a vast number of diseases and diagnostic measures. We propose Bayesian non-parametric approaches that use Dirichlet process mixture (DPM)

Address for correspondence: Timothy E. Hanson, Division of Biostatistics, University of Minnesota School of Public Health, A460 Mayo Building, MMC 303, 420 Delaware Street South East, Minneapolis, MN 55455, USA.
E-mail: hanson@biostat.umn.edu

 0035–9254/08/57207

models and mixture of Polya tree (MPT) models for the analysis of continuous serologic data.

The data that we consider are obtained under the following setting. Samples of n_0 and n_1 individuals are drawn from the non-infected and infected populations respectively. A continuous diagnostic test is applied to all sampled individuals, resulting in data that are $n_0 + n_1$ 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 immuno-sorbent assay (ELISA) serological test. A relatively large serology score indicates that the test detected a high concentration of analytes that are suggestive of the presence of disease or infection. A relatively low serology score indicates the absence of such analytes. Let F_0 and F_1 denote the cumulative distribution functions (CDFs) corresponding to the distributions of serology scores for the non-infected and infected populations respectively. Also, let the density functions corresponding to F_0 and F_1 be denoted by f_0 and f_1.

The models that we present are used to obtain data-driven inferences for the distributions of serologic values for the populations of infected and non-infected individuals, and the corresponding receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC). The concepts of ROC curves and AUCs are described in Section 2.1.

A novel feature of our proposed statistical methods is the incorporation of a stochastic ordering constraint for the distributions of serologic values for the infected and non-infected populations, i.e. we impose the constraint that F_0 and F_1 be such that $F_0(t) \geqslant F_1(t)$ for all t. Qualitatively this constraint results in densities f_0 and f_1 where f_1 is shifted to the right of f_0. Biologically such a constraint is essentially always appropriate because serologic values for infected individuals tend to be larger than serologic values for non-infected individuals, provided that the diagnostic test has reasonable discriminatory ability. Modelling stochastic order is in contrast with the vast majority of both Bayesian and frequentist approaches to ROC data analysis where the distributions of serology scores for the infected and non-infected populations are assumed to be unconstrained. We develop a novel approach to modelling dependent Polya tree (PT) processes and compare them with DPMs by using a measure of predictive performance. The MPT model proposed is, to our knowledge, the first example of a PT prior approach for two dependent distributions.

Two illustrations of our non-parametric models are presented in which the operating characteristics of commercially available ELISAs designed to detect antibodies to Johne's disease in dairy cows are evaluated. In the USA, Johne's disease is an endemic, incurable wasting disease that leads to appreciable annual economic loss which is sustained by the dairy industry.

The remainder of the paper is organized as follows. Background on ROC curves, and the Bayesian non-parametric prior models that are used in this work, is provided in Section 2. Section 3 presents a modelling framework for analysing ROC data under stochastic order constraints, using DPMs and MPTs, including discussion of prior specification and model comparison. In Section 4 we present analyses of serologic data for Johne's disease, and Section 5 contains concluding remarks. The two appendices provide technical details for the DPM and MPT models, including methods for posterior simulation.

2. Background

In this section we present background on ROC curves and related parameters of interest such as AUC. We then discuss DPMs and MPT models.

2.1. Receiver operating characteristic curves

The ROC curve is a graphical measure of the accuracy of a continuous diagnostic test. It represents a plot of all possible pairs of true positive response probability *versus* false positive response probability across all cut-off values k that could be used to dichotomize the data into test positive or negative categories, i.e. the ROC curve represents the plot $(1 - F_0(k), 1 - F_1(k))$ for all cut-off values k. Let $S_0 = 1 - F_0$ and $S_1 = 1 - F_1$. Then the ROC curve is given by

$$\mathrm{ROC}(u) = S_1\{S_0^{-1}(u)\} = 1 - F_1\{F_0^{-1}(1-u)\}$$

for $u \in [0, 1]$. In addition to the ROC curve, a parameter of interest is its corresponding $\mathrm{AUC} = \int_0^1 \mathrm{ROC}(u)\,\mathrm{d}u$, which has the interpretation of being the probability that a randomly selected infected individual has a serology score that is greater than that for a randomly selected non-infected individual (Bamber, 1975).

A useful consequence of the stochastic order constraint is that $F_0(t) > F_1(t)$ if and only if $\mathrm{ROC}(u) > u$, which further implies that $\mathrm{AUC} > 0.5$. Stochastic order has a natural interpretation: subjects that are infected are more likely to test positively than non-infected subjects for all k. Similarly, non-infected subjects are more likely to test negatively than infected subjects. Although AUC provides an aggregate summary of test usefulness across all possible cut-offs, the test is ultimately used for classification into disease status. For classification purposes, it is desirable to find a model that maximizes a predictive criterion, i.e. to find the model which, in some sense, provides the best predictive ability given the data.

In practice, the distributions F_1 and F_0 often exhibit non-standard features such as multimodality and skewness. This is especially true for the distribution of serology scores for the infected population where multimodality can occur, for instance, when the infected population is a composite of individuals in different stages of disease. In this case, individuals in an advanced stage of disease are expected to have higher serology scores compared with newly infected or subclinical individuals, although substantial 'overlap' between the densities f_0 and f_1 can occur. Although the distribution of serology scores for non-infected individuals is often unimodal symmetric and a parametric model may adequately characterize F_0, multimodality can arise when, for instance, the distribution of cross-contaminants in the population clusters within subgroups. In general, parametric models will often not be sufficiently flexible to capture the non-standard features of F_0 and F_1. Therefore, we consider non-parametric models that can handle unspecified skewness and multimodality, or any other non-standard features without the need to know in advance of their existence.

There is a vast literature on parametric and non-parametric frequentist ROC data analysis. For instance, the non-parametric approaches of Zou *et al.* (1997) and Lloyd (1998) used kernel density estimates of f_0 and f_1 and a plug-in estimate of the corresponding ROC curve. We refer the reader to Greiner *et al.* (2000) and to Pepe (2003), both of which provide additional background beyond that presented here. In particular, Pepe (2003) discussed many parametric and non-parametric or semiparametric frequentist approaches to ROC curve estimation.

The amount of existing work in the Bayesian literature is, by comparison, limited. (This is particularly true for the Bayesian non-parametric literature.) Recent work includes the parametric approach of Choi *et al.* (2006), and the DPM approach of Erkanli *et al.* (2006), which was, however, implemented by utilizing finite mixtures. Moreover, Branscum *et al.* (2006) used MPTs to analyse ROC data when true infection status is unknown. Their approach is reliant on informative priors or the availability of additional data that are related to infection status, such as results from a second conditionally independent diagnostic test, because the model lacks identifiability.

Again, distinguishing features of our work include the incorporation of the stochastic order constraint for F_0 and F_1, and the use of two distinct Bayesian non-parametric modelling approaches based on DPM and MPT priors. We note that, although these two prior models have dominated the literature on modelling random distributions, little work exists on their comparison in the context of generic settings such as the setting that we study in this paper.

2.2. Dirichlet process mixtures

Ferguson (1973) introduced the Dirichlet process (DP) as a (prior) probability model for distributions (or, equivalently, CDFs) G. The DP is characterized by two parameters: a base distribution G_0 (the centre of the process) and a positive scalar parameter α, which can be interpreted as a precision parameter; larger values of α result in realizations G that are closer to G_0. We shall write $G \sim \mathrm{DP}(\alpha, G_0)$ to indicate that a DP prior is used for the random distribution G. An important constructive definition of the DP was given by Sethuraman (1994). According to 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 by using a *stick breaking* mechanism that is based on independent draws, $\{z_k : k = 1, 2, \ldots\}$, from a beta$(1, \alpha)$ distribution; specifically, the first weight is equal to z_1 and, for $i = 2, 3, \ldots$, the ith weight is given by $z_i \prod_{k=1}^{i-1}(1 - z_k)$.

A natural way to increase the applicability of DP-based modelling is by using the DP as a prior for the mixing distribution in a mixture model with a parametric kernel $K(\cdot; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$. This approach yields the class of DPM models (Antoniak, 1974) which can be generically expressed as $F(\cdot; G) = \int K(\cdot; \boldsymbol{\theta})\, \mathrm{d}G(\boldsymbol{\theta})$, $G \sim \mathrm{DP}(\alpha, G_0)$, with the analogous notation for the random mixture density $f(\cdot; G)$. The kernel can be chosen to be a continuous distribution, thus overcoming the (almost sure) discreteness of DP realizations. The model is, typically, completed with hyperpriors for the parameters of G_0 and α. Moreover, practically important semiparametric versions emerge by working with kernels $K(\cdot; \boldsymbol{\theta}, \phi)$ where the ϕ-portion of the parameter vector is modelled parametrically. Reviews of modelling and Markov chain Monte Carlo (MCMC) inference with DPMs can be found in, for example, MacEachern and Müller (2000), Müller and Quintana (2004) and Hanson *et al.* (2005).

On the basis of the DP constructive definition, the prior probability model $f(\cdot; G)$ can be expressed in the form of a countable mixture. Hence, DPMs generalize finite mixture models, offering practical advantages in modelling and inference for data that arise from non-standard distributions, e.g. data that are clustered as may be the case for serologic data from infected populations with subgroups of the population in different stages of disease.

2.3. Mixtures of finite Polya trees

Lavine (1992, 1994) has provided foundational material on PTs and MPTs and Hanson (2006) has developed computational aspects. As with DPMs, a particular parameterization of PTs and MPTs places probability 1 on absolutely continuous distributions. Roughly speaking, an MPT prior on the random CDF F augments a standard parametric family of CDFs $\{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ with $2^J - 1$ additional parameters $\mathcal{X}$ that stochastically 'adjust' the density $f_{\boldsymbol{\theta}}(\cdot)$ to place additional mass in areas where data are seen more often than expected under $f_{\boldsymbol{\theta}}(\cdot)$. The random MPT density f acts much like a histogram with a fixed number of bins whose end points are obtained as quantiles from the family $\{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$; the PT prior induces a particular correlation between the bin heights.

Specifically, a finite PT prior for a distribution F is constructed by partitioning the sample space into increasingly fine disjoint sets at successive levels of the tree up to level J. At level

$j \leqslant J$ of the tree, the sample space is partitioned into 2^j disjoint sets each with a corresponding branch probability. The product of the branch probabilities that lead to sets at level j provides the marginal probabilities of those sets. Following Hanson (2006), the sets that partition the sample space at the jth level of the tree are denoted by $B_{\boldsymbol{\theta}}(j,k)$, $k = 1, \ldots, 2^j$. The standard parameterization defines

$$B_{\boldsymbol{\theta}}(j,k) = \left[F_{\boldsymbol{\theta}}^{-1}\left(\frac{k-1}{2^j}\right), F_{\boldsymbol{\theta}}^{-1}\left(\frac{k}{2^j}\right) \right),$$

$k = 2, \ldots, 2^j$, for a parametric distribution $F_{\boldsymbol{\theta}}$; define

$$B_{\boldsymbol{\theta}}(j,1) = \left(F_{\boldsymbol{\theta}}^{-1}(0), F_{\boldsymbol{\theta}}^{-1}\left(\frac{1}{2^j}\right) \right).$$

The corresponding branch probabilities, which are denoted by $X_{j,k}$, are modelled according to independent beta distributions, namely $X_{j,k} \sim \text{beta}(\alpha_{j,k}, \alpha_{j,k+1})$ with $X_{j,k+1} = 1 - X_{j,k}$, $j = 1, 2, \ldots, J$, $k = 1, 3, 5, \ldots, 2^j - 1$. Under the standard parameterization, $\alpha_{j,k} = cj^2$ for all j and k, and $c > 0$. We use this specification hereafter. The value of c reflects the researcher's prior degree of uncertainty in the underlying distribution $F_{\boldsymbol{\theta}}$ with smaller values leading to a more 'non-parametric' analysis. Uncertainty about $\boldsymbol{\theta}$ and c are modelled as $\mathrm{d}P(c, \boldsymbol{\theta})$, giving rise to a mixture of finite PTs model for F, which we denote heuristically as $F \sim \int \text{FPT}(c, F_{\boldsymbol{\theta}})\, p(\mathrm{d}c, \mathrm{d}\boldsymbol{\theta})$. See Hanson and Johnson (2002) and Hanson (2006) for additional details.

3. Models and methods

The support for F_0 and F_1 is taken to be the real line (perhaps, after an appropriate transformation). Moreover, let $\mathcal{D} = \{\mathbf{x}_0, \mathbf{x}_1\}$, where $\mathbf{x}_0 = (x_{01}, \ldots, x_{0n_0})$ and $\mathbf{x}_1 = (x_{11}, \ldots, x_{1n_1})$ are the data vectors from the non-infected and infected groups respectively. We propose non-parametric prior probability models under the stochastic order restriction $F_0 \leqslant_{\text{st}} F_1$, i.e. $F_1(t) \leqslant F_0(t)$, for all $t \in \mathbb{R}$. In particular, we consider two modelling approaches: one based on DPM priors (Section 3.1) and one based on MPT priors (Section 3.2). Prior specification is discussed in Section 3.3 and model comparison in Section 3.4.

3.1. Modelling by using Dirichlet process mixtures

We develop location normal mixture models for the distribution functions of the infected and non-infected populations, $F_l(t) \equiv F_l(t; H_l, \sigma^2) = \int N(t; \theta, \sigma^2)\, \mathrm{d}H_l(\theta)$, $l = 0, 1$. (We denote by $N(t; \theta, \sigma^2)$ the CDF or density function, depending on the context, of the normal distribution with mean θ and variance σ^2.) On the basis of a standard result on stochastic ordering for mixtures (see, for example, Shaked and Shanthikumar (1994), page 8), if the mixing distributions are stochastically ordered, $H_0 \leqslant_{\text{st}} H_1$, then $F_0(\cdot; H_0, \sigma^2) \leqslant_{\text{st}} F_1(\cdot; H_1, \sigma^2)$. This result relies on the fact that the $N(\theta, \sigma^2)$ distribution is stochastically ordered in θ for fixed σ^2, i.e., if $\theta_1 \leqslant \theta_2$, then $N(\theta_1, \sigma^2) \leqslant_{\text{st}} N(\theta_2, \sigma^2)$. Therefore, for the normal mixtures above, it suffices to construct a stochastically ordered prior model for the pair of mixing distributions (H_0, H_1). A convenient approach to defining such a prior is through latent distribution functions on $\mathbb{R}$, say H and G, such that $H_0(t) = H(t)$ and $H_1(t) = H(t)\, G(t)$. This approach was briefly discussed in Gelfand and Kuo (1991), was developed in Gelfand and Kottas (2001), and also used in Kottas and Gelfand (2001) and Kottas *et al.* (2002).

We can now use independent DP priors for H and G, obtaining the stochastically ordered DPM model

$$F_0(t; H, \sigma^2) = \int N(t; \theta, \sigma^2)\, \mathrm{d}H(\theta),$$

$$F_1(t; H, G, \sigma^2) = \int\int N\{t; \max(\theta, \phi), \sigma^2\}\, \mathrm{d}H(\theta)\, \mathrm{d}G(\phi),$$

where, given hyperparameters $\psi = (\alpha_H, \mu_H, \tau_H^2, \alpha_G, \mu_G, \tau_G^2)$, $H \sim \mathrm{DP}\{\alpha_H, N(\mu_H, \tau_H^2)\}$, and, independently, $G \sim \mathrm{DP}\{\alpha_G, N(\mu_G, \tau_G^2)\}$. We use an inverse gamma prior for σ^2 and place (independent) priors on the components of ψ, specifically, normal priors for μ_H and μ_G, inverse gamma priors for τ_H^2 and τ_G^2, and gamma priors for α_H and α_G. Further details on the DPM model, including methods for posterior inference, are provided in Appendix A.

3.2. A mixtures of Polya tree modelling approach

The setting here is as in the DPM case. Serology scores, $x_{01}, \ldots, x_{0n_0}$, for the sampled individuals from the subpopulation of non-infected individuals are modelled according to F_0, and serology scores for infected individuals, $x_{11}, \ldots, x_{1n_1}$, are governed by F_1. Again, our prior structure imposes $F_1(t) \leqslant F_0(t)$, $t \in \mathbb{R}$, which we model directly as $F_0(t) = H(t)$ and $F_1(t) = H(t)\, G(t)$ where H and G are assigned mixture of finite PT priors,

$$H \sim \int \mathrm{FPT}(c_H, H_{\boldsymbol{\theta}_H})\, \mathrm{d}P_H(c_H, \boldsymbol{\theta}_H),$$

$$G \sim \int \mathrm{FPT}(c_G, G_{\boldsymbol{\theta}_G})\, \mathrm{d}P_G(c_G, \boldsymbol{\theta}_G).$$

Specifically, we centre random G at $G_{\boldsymbol{\theta}_G} = N(\mu_G, \tau_G^2)$ where $\boldsymbol{\theta}_G = (\mu_G, \tau_G^2)$, and H at $H_{\boldsymbol{\theta}_H} = N(\mu_H, \tau_H^2)$ where $\boldsymbol{\theta}_H = (\mu_H, \tau_H^2)$. The levels of the finite PTs defining H and G are set to J_H and J_G, although in practice we fix $J_H = J_G \equiv J$. The level J determines the 'level of detail' that is accommodated by the MPT model. Let $\mathcal{X}_G = \{X_{j,k}^G : j = 1, \ldots, J, k = 1, \ldots, 2^j\}$ denote the set of branch probabilities for G, with a similar definition for $\mathcal{X}_H$ for H. Increasing J to $J+1$ essentially doubles the number of conditional probabilities in $\mathcal{X}_H$ and $\mathcal{X}_G$, and can capture fine detail in the modelled densities. However, adding levels also increases overall variability and can reduce the predictive ability of the model when not needed. We considered several values for J in the examples of Section 4 and found $J = 4$ to be sufficient. The mixing parameters μ_H and μ_G have independent normal priors; τ_H and τ_G have independent gamma priors. In the data analyses the weight parameters c_H and c_G are given independent gamma priors, allowing for differing prior variability in the shapes of H and G; see the next section. We also considered the flat prior $p(\mu_H, \mu_G, \tau_H, \tau_G) \propto 1$. See Appendix B for a proof of posterior propriety under this prior as well as for details on posterior inference under the MPT model.

3.3. Prior elicitation

Here, we discuss an approach to prior specification, which ensures that roughly the same amount of prior information is incorporated in each of the models that were developed in Sections 3.1 and 3.2. In particular, we first match prior predictive densities (i.e. the centres of the prior models for the non-infected and infected random densities) and then control the prior variability around the expected densities.

A useful feature of both models is that the prior predictive densities, $E\{f_0(\cdot)\}$ and $E\{f_1(\cdot)\}$, are driven only by the parameters of the centring distributions, i.e. they do not depend on parameters α_H and α_G, and c_H and c_G for the DPM and MPT models respectively. For instance, under

the DPM model, $E\{f_0(\cdot)\}$ arises by averaging an $N(\theta_0, \sigma^2)$ density over θ_0, which is drawn from the centring distribution $N(\mu_H, \tau_H^2)$, and over the priors for μ_H, τ_H^2 and σ^2. Under the MPT model, $E\{f_0(\cdot)\}$ can be obtained by averaging the centring density $N(\mu_H, \tau_H^2)$ over the priors for μ_H and τ_H^2. The first step of the prior elicitation approach involves matching $E\{f_0(\cdot)\}$ and $E\{f_1(\cdot)\}$ under the two models. This is accomplished by specifying the prior hyperparameters for μ_H and τ_H^2 (and σ^2 under the DPM model), working with $E\{f_0(\cdot)\}$ and using rough prior guesses at the centre and range of the data for the non-infected group. Prior predictive densities $E\{f_0(\cdot)\}$ are matched *exactly* when the priors for μ_H are the same under the two models and τ_H^2 under the MPT model is given the prior that is induced by $\tau_H^2 + \sigma^2$ under the DPM model. Having specified the priors for μ_H and τ_H^2, we use the same priors for μ_G and τ_G^2.

Next, we specify the priors for α_H and α_G, and c_H and c_G, using a measure of prior density variability, specifically, the L_1-distance. Focusing first on the non-infected group, the L_1-distance between the prior predictive density and random prior realizations, $f_0(\cdot) = F_0'(\cdot)$, from either model is given by $\| f_0 - E\{f_0\}\|_1 = \int_{\mathbb{R}} |f_0(t) - E\{f_0(t)\}|\, \mathrm{d}t$. On the basis of their structure, it is possible to approximate the prior predictive densities under both models with normal densities. Moreover, the L_1-measure is invariant to location–scale transformations of the prior centring distribution for both the MPT and the DPM models. Hence, without loss of generality for this approach to prior elicitation, we can assume that $E\{f_0(t)\} \approx N(t; 0, 1)$, $t \in \mathbb{R}$, and $E\{F_0(A)\} \approx N(A; 0, 1)$, for all measurable $A \subset \mathbb{R}$. Note that the L_1-distance has a useful interpretation in terms of the total variation norm:

$$\tfrac{1}{2}\| f_0 - N(0,1)\|_1 = \tfrac{1}{2}\int_{\mathbb{R}} |f_0(t) - N(t;0,1)|\, \mathrm{d}t = \sup_{A \subset \mathbb{R}} |F_0(A) - N(A;0,1)|.$$

Through simulation of prior densities under both models we find that $\alpha_H \sim \Gamma(5,1)$ and $c_H \sim \Gamma(10,3)$ produce a prior median and 95% credible interval (CI) of 0.3 and (0.1,0.8) for $\| f_0 - N(0,1)\|_1$. (We denote by $\Gamma(a,b)$ the gamma distribution with mean a/b.) That is, the amount of probability mass that is 'shifted' from the centring density to the random draw is typically 0.15 and ranges roughly over (0.05,0.4). Similarly, $\alpha_H \sim \Gamma(5, 0.5)$ and $c_H \sim \Gamma(15,2)$ produce a prior median and 95% CI of 0.2 and (0.1,0.6) for $\| f_0 - N(0,1)\|_1$, and $\alpha_H \sim \Gamma(2,2)$ and $c_H \sim \Gamma(2, 1.5)$ produce a prior median and 95% CI of 0.5 and (0.2,1.3).

Having matched the prior variability of the DPM and MPT priors for the non-infected population, the same priors for α_G and c_G are used for the random distribution G, inducing the stochastically ordered prior on the infected population.

Finally, we note that, as discussed in Appendix A, the DP precision parameters α_H and α_G control the number of distinct components in the semiparametric mixture models for the non-infected and infected populations that were developed in Section 3.1. This property of the DPM model can be used to obtain an alternative approach to prior elicitation for α_H and α_G. However, there are no analogous results for the MPT weight parameters c_H and c_G, and we thus favour the approach that was discussed above.

3.4. Model comparison

Although it is ill advised to reduce model comparison and selection to a single number summarizing some measure of model fit, such statistics provide useful information when used in conjunction with posterior summaries such as estimated densities, ROC curves and posterior AUC estimates. Ultimately, serology scores are used for classification of animals of unknown disease status. Therefore, a statistic summarizing the predictive utility of a model through the data provides evidence towards a serology test's usefulness in terms of classification.

To compare the two Bayesian non-parametric models that were developed in Sections 3.1 and 3.2 formally, we work on posterior predictive space and compute the log-pseudo-marginal likelihood (LPML) statistic (Geisser and Eddy, 1979). Consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$, which are to be compared on the basis of available data $\mathcal{D} = (y_1, \ldots, y_n)$. Let $p_l(y_{\text{new}}|\mathcal{D})$ denote the posterior predictive density under model $\mathcal{M}_l$, $l = 1, 2$. The conditional predictive ordinate (CPO) for observation y_j under model $\mathcal{M}_l$ is given by $\text{CPO}_{lj} = p_l(y_j|\mathcal{D}_{(j)})$, where $\mathcal{D}_{(j)}$ denotes the remaining data with y_j removed. The ratio $\text{CPO}_{1j}/\text{CPO}_{2j}$ describes how well model $\mathcal{M}_1$ supports the observation y_j relatively to model $\mathcal{M}_2$, on the basis of the remaining data $\mathcal{D}_{(j)}$. CPO statistics can be readily estimated, by using MCMC output, across a wide variety of models (see, for example, Chen *et al.* (2000)). The logarithm of the product of the n CPO statistics under a given model is the LPML statistic for that model, which provides an aggregate 'leave-one-out' measure of model predictive utility.

Turning to our setting for ROC data, let $\mathbf{x}_{0(i)} = \{x_{0i'} : i' \neq i\}$, for $i = 1, \ldots, n_0$, and $\mathbf{x}_{1(j)} = \{x_{1j'} : j' \neq j\}$, for $j = 1, \ldots, n_1$. Estimation of the CPOs, $\text{CPO}_{0i} = p(x_{0i}|\mathbf{x}_{0(i)}, \mathbf{x}_1)$, $i = 1, \ldots, n_0$, and $\text{CPO}_{1j} = p(x_{1j}|\mathbf{x}_0, \mathbf{x}_{1(j)})$, $j = 1, \ldots, n_1$, under each of the models in Sections 3.1 and 3.2, is discussed in Appendix A and B respectively. In Section 4, we compare the performance of the two models in the context of the analysis of serology scores from two ELISA tests. For this, we consider inference results for densities, the ROC curve and AUC, as well as the LPML statistics, $\text{LPML} = \text{LPML}_0 + \text{LPML}_1$, where $\text{LPML}_0 = \sum_{i=1}^{n_0} \log(\text{CPO}_{0i})$ and $\text{LPML}_1 = \sum_{j=1}^{n_1} \log(\text{CPO}_{1j})$.

4. Serology score analyses

Johne's disease (*Mycobacterium avium paratuberculosis* (MAP)) is endemic throughout the USA, affecting multiple species of animals, and successful surveillance and control programmes are dependent on current diagnostic technologies for its detection. For example, decisions to modify herd management practices or to redirect regional vaccination efforts for the control and eradication of the disease are aided by efficient veterinary diagnostics. We consider the evaluation of antibody detection technology from two different ELISA kits for detection of MAP in dairy cattle.

In particular, we analysed serologic data for Johne's disease in dairy cattle from commercially available ELISA kits that were developed by the Synbiotic Corp. in San Diego, California, USA, and the Institut Pourquier in Montpellier, France. We refer to these tests as the Synbiotic and Pourquier ELISAs respectively. The data came from a study that was discussed in detail by Collins *et al.* (2005). Characteristics of the 14 dairy herds that were included in the study were catalogued by Collins *et al.* (2005), and sampling strategies and diagnostic testing procedures that were implemented were also described. Briefly, the non-infected cows came from seven Minnesota herds that satisfied disease freedom criteria that were set out by the 'US voluntary bovine Johne's disease herd status program'. Infected cows came from seven 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 ELISA data that we analysed were from $n_0 = 345$ non-infected and $n_1 = 258$ infected cows. We refer the reader to Collins *et al.* (2005) for additional details pertaining to the data and their collection.

The serology scores were log-transformed to facilitate the use of normal centring distributions of the non-parametric priors. Inferences are made for the distributions of ELISA scores for the non-infected and infected populations, and the ROC curve and AUC by using DPM and MPT approaches.

We followed the approach of Section 3.3 for prior specification. Regarding the hyperpriors for the centring distributions of the DPM and MPT models, we conducted a sensitivity analysis considering priors for μ_H and μ_G, and τ_H^2 and τ_G^2 (and σ^2 under the DPM model) with various levels of dispersion, including the flat prior $p(\mu_H, \mu_G, \tau_H, \tau_G) \propto 1$, under the MPT model. This study demonstrated that, at least with the sample size that is available in our two data examples, for both models, the use of vague or relatively focused priors makes little difference in posterior summaries of interest, i.e. for the ROC curve, AUC and the densities for the non-infected and infected groups. The specific priors that are given in Sections 4.1 and 4.2 yield prior predictive densities under each model that are centred roughly at the data histograms but are an order of magnitude more dispersed. Finally, for both data examples, we discuss results under three different prior choices for the precision parameters under each model.

4.1. Synbiotic Corporation enzyme-linked immuno-sorbent assay

In this section, we evaluate the operating characteristics of the Synbiotic ELISA test for detection of MAP.

The independent priors that were used for the DPM analysis were as follows: $\mu_H, \mu_G \sim N(2, 4)$; $\tau_H^{-2}, \tau_G^{-2} \sim \Gamma(2, 5)$; $\sigma^{-2} \sim \Gamma(2, 2.5)$. For the MPT analysis the prior predictive densities were matched to the DPM model, yielding $\mu_H, \mu_G \sim N(2, 1)$ and $\tau_H, \tau_G \sim \Gamma(3, 1)$. The levels of the finite trees were fixed at $J_H = J_G = 4$; increasing the levels to $J_H = J_G = 5$ changed posterior inferences of interest negligibly.

On the basis of considerations in Section 3.3, three priors for the DPM model precision parameters were considered, reflecting different levels of prior variability about the prior predictive density: $\alpha_H, \alpha_G \sim \Gamma(2, 2)$ (prior A), $\alpha_H, \alpha_G \sim \Gamma(5, 1)$ (prior B) and $\alpha_H, \alpha_G \sim \Gamma(5, 0.5)$ (prior C). Similarly, we used three priors for the MPT model precisions c_H and c_G, which were chosen to match prior variability to that of the DPM: $c_H, c_G \sim \Gamma(2, 1.5)$ (prior A), $c_H, c_G \sim \Gamma(10, 3)$ (prior B) and $c_H, c_G \sim \Gamma(15, 2)$ (prior C).

Prior A, allowing for greatest variability in the modelled densities, provided the largest LPML_0 and LPML_1 under the MPT model, which are larger than those which were obtained under the DPM models (Table 1). The predictively 'best' DPM model occurs under prior B; however, as indicated by the LPML values that are given in Table 1, the three prior specifications yield essentially identical posterior inferences under the DPM model. The DPM is, in comparison, more robust to the specification of the precision priors.

The precision prior affects posterior inference, such as AUC, and estimated densities more markedly under the MPT model. This reflects the direct modelling of densities through the

Table 1. Synbiotic ELISA: LPML estimates and posterior summaries (medians and 95% CIs) for AUC under three prior choices for each of the DPM and MPT models†

Prior	*Results for DPMs*			*Results for MPTs*		
	$LPML_0$	$LPML_1$	*AUC*	$LPML_0$	$LPML_1$	*AUC*
A	−277.8	−408.9	0.720 (0.671,0.769)	−269.0	−398.6	0.730 (0.690,0.772)
B	−275.0	−408.1	0.716 (0.675,0.757)	−271.6	−400.3	0.738 (0.696,0.775)
C	−274.9	−408.6	0.720 (0.679,0.760)	−276.7	−407.1	0.743 (0.707,0.781)

†Priors A, B and C are described in Section 4.1.

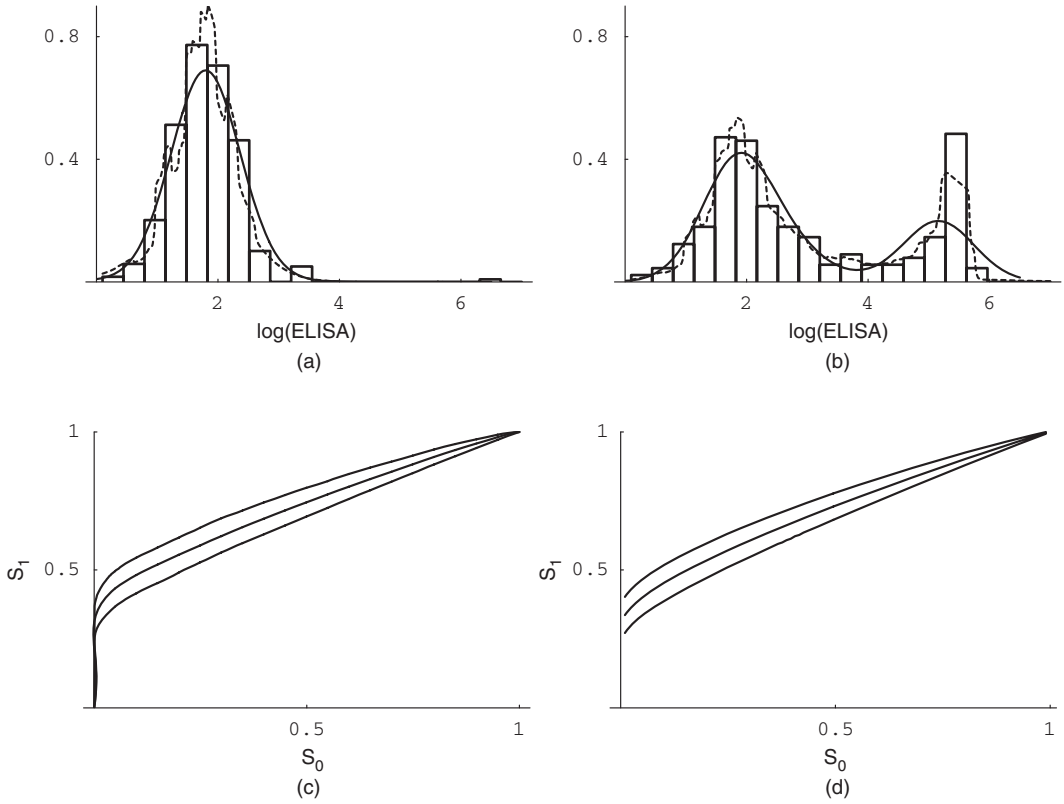

Fig. 1. (a), (b) Data histograms and estimated population densities of log-transformed serology scores (———, DPM results; -------, MPT results) and (c), (d) estimated ROC curves with 95% pointwise CIs for the Synbiotic ELISA test of Section 4.1: (a) non-infected group; (b) infected group; (c) MPT model; (d) DPM model

products $F_0(t) = H(t)$ and $F_1(t) = H(t)\,G(t)$ under the MPT rather than convolving the products with a smooth kernel as in the DPM. Posterior estimates of the densities f_0 and f_1 under the MPT (prior A) and DPM (prior B) models, along with histograms of the data for non-infected and infected cows, are presented in Figs 1(a) and 1(b). Both the DPM and the MPT models identify two modes for the MAP-infected subpopulation. The MPT model captures the rather abrupt change in f_1, tracking the histogram more closely than the DPM, which is comparatively smoother. This partition-based locally adaptive behaviour can be viewed as a relative merit of PTs, not unlike the sort of spatially inhomogeneous adaptation that a wavelet basis can accommodate (Draper, 1999), but also results in more jagged looking density estimates than those obtained from the DPM model. Under priors B and C the estimated densities under the MPT model are significantly smoother, attenuating the spike in f_1 under prior A. This suggests that the priors on c_H and c_G should allow for small values, allowing for greater prior variability, for the MPT to perform predictively well with highly inhomogeneous, e.g. clumped, data. In contrast, the DPM density estimates do not appreciably change across priors A, B and C.

Table 1 includes AUC estimates under the three precision priors for each model. Again, by comparison, the DPM model yields inference that is more robust to the prior choice for the precision parameters. However, it is important to note that the predictively 'best' MPT model (prior A) and DPM model (prior B) result in inferences for the ROC curve and AUC that are

qualitatively similar. In particular, the posterior median and 95% CI for AUC was 0.716 and (0.675, 0.757) on the basis of the DPM (prior B) analysis and was 0.730 and (0.690, 0.772) for the MPT (prior A) analysis. Moreover, posterior estimates for the ROC curve were almost identical under the two models (see Figs 1(c) and 1(d)).

Finally, for comparison with a parametric analysis, we consider a simple model that assumes (independent) normal distributions (on the log-scale) for the non-infected and infected populations. This model can be viewed as the univariate version of the binormal model, which is commonly used for serology scores from two correlated diagnostic tests (see, for example, Choi *et al.* (2006)). The model was fitted by using dispersed normal and inverse gamma priors for the means and variances respectively of the non-infected and infected groups. The posterior median for AUC was 0.763 and the 95% CI was (0.721, 0.802). These estimates are a little more optimistic than the non-parametric-based estimates. Under the parametric model, $\mathrm{LPML}_0 = -283.0$ and $\mathrm{LMPL}_1 = -487.5$, providing considerably worse predictive accuracy than either non-parametric model, especially in the infected subpopulation. Furthermore, fitted density estimates (which are not shown) clearly show the inadequacy of this parametric approach for the log-transformed serology scores. Evidently, more flexible parametric models can be formulated (e.g. by using finite normal mixtures) to obtain for this data set inferences that are comparable with those resulting from the non-parametric models. However, incorporating stochastic order in such models is not straightforward. As importantly, the key feature of the Bayesian non-parametric models proposed is that they can adapt to a range of distributional shapes that might be suggested by the data.

4.2. Institut Pourquier enzyme-linked immuno-sorbent assay

Here, we apply the two models to the serologic data from the Institut Pourquier ELISA.

The independent priors that were used in the DPM analysis were as follows: $\mu_H, \mu_G \sim N(4, 4)$, $\tau_H^{-2}, \tau_G^{-2} \sim \Gamma(2, 5)$ and $\sigma^{-2} \sim \Gamma(2, 2.5)$. The following independent priors were used for the parameters of the MPT model: $\mu_H, \mu_G \sim N(4, 1)$ and $\tau_H, \tau_G \sim \Gamma(3, 1)$. We set $J_H = J_G = 4$.

Moreover, we considered the same three priors for the precision parameters as in Section 4.1. Under these three priors, Table 2 provides results from the LPML-based model comparison as well as posterior estimates for AUC. In this case, the DPM model fares better than the MPT model on the basis of the LPML estimates. Again, the MPT model is more sensitive to the choice of the precision priors compared with the DPM model. However, as with the Synbiotic ELISA data, the predictively 'best' DPM model (prior C) and MPT model (prior A) yield similar posterior results for AUC; posterior medians and 95% CIs were given by 0.706 and (0.658, 0.751) under the DPM analysis, and by 0.709 and (0.666, 0.753) on the basis of the MPT analysis.

This was also the case for the estimated densities, f_0 and f_1, which are presented in Figs 2(a)

Table 2. Pourquier ELISA: LPML estimates and posterior summaries (medians and 95% CIs) for AUC under three prior choices (see Section 4.1) for each of the DPM and MPT models

Prior	*Results for DPMs*			*Results for MPTs*		
	$LPML_0$	$LPML_1$	*AUC*	$LPML_0$	$LPML_1$	*AUC*
A	−313.3	−366.9	0.701 (0.650,0.750)	−314.5	−369.7	0.709 (0.666,0.753)
B	−315.9	−365.3	0.706 (0.660,0.751)	−315.0	−379.0	0.723 (0.683,0.760)
C	−311.7	−368.1	0.706 (0.658,0.751)	−315.0	−386.0	0.733 (0.698,0.770)

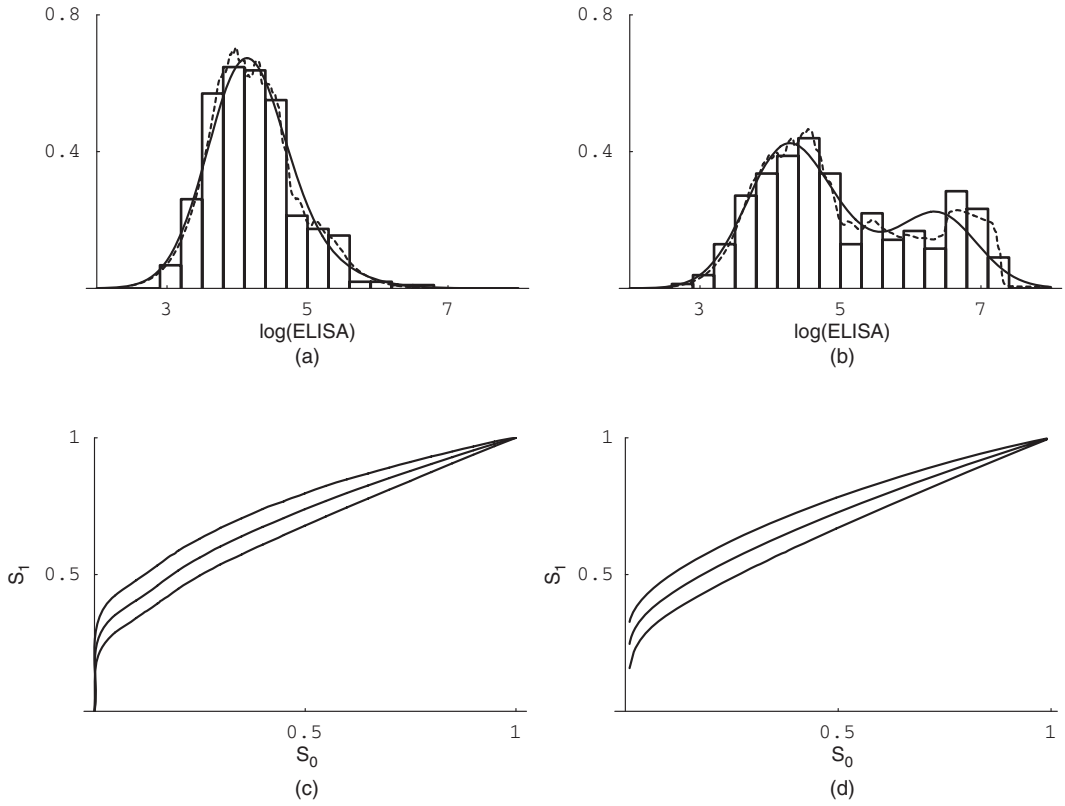

Fig. 2. (a), (b) Data histograms and estimated population densities of log-transformed serology scores (———, DPM results; -------, MPT results) and (c), (d) estimated ROC curves with 95% pointwise CIs for the Pourquier ELISA test of Section 4.2: (a) non-infected group; (b) infected group; (c) MPT model; (d) DPM model

and 2(b) (under priors C and A for the DPM and MPT models respectively), together with histograms of the log-transformed ELISA data from the non-infected and infected cows. Similarly to the Synbiotic ELISA test, the distribution of serology scores for infected cows is bimodal, which could indicate a cluster of infected cows in a more advanced stage of Johne's disease and a cluster in a less advanced stage of infection. For non-infected cows the distribution of serology scores has a slight right skew, which is captured well by both models. Figs 2(c) and 2(d) plot posterior point (posterior means) and 95% CI estimates for the ROC curve. The two models yield similar estimated ROC curves and uncertainty bands around the point estimates.

Regarding results under the parametric model that was considered in Section 4.1, the posterior median for AUC was 0.735 with 95% CI (0.692, 0.774). Again, the particular parametric model overestimates the test's diagnostic ability. Moreover, $\text{LPML}_0 = -320.6$ and $\text{LMPL}_1 = -401.9$, under the parametric analysis, again indicating considerably worse predictive ability than either non-parametric model in the infected subpopulation.

5. Conclusions

Stochastic ordering imposes a natural constraint on ROC curves that any diagnostic test must

satisfy, namely that AUC > 0.5. We developed a novel method for non-parametric analysis of ROC data that incorporates a stochastic order constraint. We considered two flexible modelling approaches using the two prominent Bayesian non-parametric prior models for distributions, i.e. DPM priors and MPT priors. We discussed model comparison based on predictive performance of the two models, using LPML statistics. Two applications were presented for the analysis of ROC data, which illustrated the utility of the non-parametric models proposed.

Although the class of all PT distributions includes the DP as a special case, the two prior processes that we have considered here are quite different. The DPM model behaves like a finite mixture model with an unknown number of components, and this fact has been exploited for modelling multimodal densities and data where clumping naturally occurs. In contrast, the MPT prior model is centred at a parametric family and can inherit the overall shape of the centring density when sample sizes are small or the MPT weight parameter is relatively large.

Regarding the data analyses, the MPT approach modelled the abrupt spike in the infected population density very well for the Synbiotic test (Section 4.1) and outperformed the DPM in this regard. The DPM model outperformed the MPT for the Institut Pourquier ELISA data (Section 4.2), parsimoniously capturing the smoother nature of the densities. Density estimates based on MPT models were comparatively rougher for both data sets. It might seem that choosing between models for a given serology data set is data dependent. For the Institut Pourquier serology scores, where data appear to be comprised of two or three homogeneous components, the DPM model is predictively superior; for the Synbiotic data, exhibiting more localized structure, the MPT wins out. However, although difficult to calibrate, the differences in the values of the LPML statistics for the two models were certainly not substantial. Moreover, for both data examples, the best fitting DPM and MPT models yielded similar posterior inference for estimated densities, the ROC curve and AUC.

Hence, looking beyond any particular model comparison criterion and the structural differences of the DPM and MPT priors (which were discussed in earlier sections), it is encouraging that the two distinct Bayesian non-parametric prior models result in inferences that are qualitatively very similar. In particular, using the posteriors, say, for the ROC curve and AUC, a practitioner would, arguably, draw the same conclusions from the two models regarding the accuracy of the two ELISA tests that were considered in Sections 4.1 and 4.2. Although not typically done in the literature, we argue that it is valuable to 'validate' posterior inference results under one non-parametric prior model with a different prior of comparable flexibility.

We also note that, with the Bayesian approaches that we employ, full inference is available. Once models have been fitted, various criteria can be used to obtain a posterior for an optimal test cut-off k for diagnostic purposes. For example, one well-argued approach is to maximize Youden's index (Le, 2006), which is given in our context by $\text{sensitivity}(k) + \text{specificity}(k) - 1 = F_0(k) - F_1(k)$. Note that stochastic order implies that Youden's index, which is a common measure of test performance, is positive for all cut-offs k. Another approach is to minimize the distance between the point $(0, 1)$ (representing a gold standard test) and the ROC curve. In both cases a posterior for k is easily computed.

Finally, the application of the Bayesian non-parametric stochastic ordering methodology that we developed extends beyond the scope of ROC data. Indeed, implicit in many common semiparametric and parametric models is a stochastic order constraint within subgroups of populations. For example, given a baseline survival function S_0, the proportional hazards model assumes that $S_{\mathbf{x}}(t) = S_0(t)^{\exp(\mathbf{x}'\beta)}$ and the accelerated failure time model assumes that $S_{\mathbf{x}}(t) = S_0\{\exp(\mathbf{x}'\beta)t\}$, so survival times from two different subgroups (say $\mathbf{x}_a$ and $\mathbf{x}_b$) are sto-

chastically ordered under both models. Simple analysis-of-variance and regression models also have this assumption built in.

Acknowledgements

The authors thank the Joint Editor and two referees for helpful comments. The work of the second author was supported in part by National Science Foundation grant DMS-0505085. The work of the third author was supported in part by National Center for Research Resources grant P20 RR0201045.

Appendix A: Posterior inference under the Dirichlet process mixture model

Here, we provide details on modelling and inference under the stochastically ordered DPM prior of Section 3.1.

A.1. Hierarchical model formulation

The model can be expressed in hierarchical form by introducing latent mixing parameters $\boldsymbol{\theta} = \{\theta_i : i = 1, \ldots, n_0, n_0+1, \ldots, n_0+n_1\}$, which, given H, are independent and identically distributed (IID) from H, and $\boldsymbol{\phi} = \{\phi_j : j = 1, \ldots, n_1\}$, with the ϕ_j, given G, IID from G. Specifically,

$$\left.\begin{array}{rl} x_{0i}|\theta_i, \sigma^2 \overset{\text{ind}}{\sim} N(\theta_i, \sigma^2), & i = 1, \ldots, n_0, \\ x_{1j}|\theta_{n_0+j}, \phi_j, \sigma^2 \overset{\text{ind}}{\sim} N\{\max(\theta_{n_0+j}, \phi_j), \sigma^2\}, & j = 1, \ldots, n_1, \\ \theta_i|H \overset{\text{IID}}{\sim} H, & i = 1, \ldots, n_0+n_1, \\ \phi_j|G \overset{\text{IID}}{\sim} G, & j = 1, \ldots, n_1, \\ H, G|\alpha_H, \mu_H, \tau_H^2, \alpha_G, \mu_G, \tau_G^2 \sim \text{DP}\{\alpha_H, N(\mu_H, \tau_H^2)\}\,\text{DP}\{\alpha_G, N(\mu_G, \tau_G^2)\} & \end{array}\right\} \quad (1)$$

with priors for $\alpha_H, \mu_H, \tau_H^2, \alpha_G, \mu_G, \tau_G^2$ and σ^2 as discussed in Section 3.1. The introduction of the additional mixing parameters θ_{n_0+j}, $j = 1, \ldots, n_1$, allows us to retain the first-stage conditionally independent specification in the hierarchical model after marginalizing in formulation (1) the random distribution functions H and G over their DP priors. The priors $p(\boldsymbol{\theta}|\alpha_H, \mu_H, \tau_H^2)$ and $p(\boldsymbol{\phi}|\alpha_G, \mu_G, \tau_G^2)$ resulting after this marginalization can be defined according to a generalized Polya urn scheme (Blackwell and MacQueen, 1973). For instance, for $\boldsymbol{\theta}$, θ_1 follows an $N(\mu_H, \tau_H^2)$ distribution and, for any $i = 2, \ldots, n_0+n_1$, θ_i, conditionally on $\theta_1, \ldots, \theta_{i-1}$, follows a mixed distribution with point masses $(\alpha_H + i - 1)^{-1}$ at θ_l, $l = 1, \ldots, i-1$, and continuous mass $\alpha_H(\alpha_H + i - 1)^{-1}$ on the $N(\mu_H, \tau_H^2)$ distribution. The expression for $p(\boldsymbol{\phi}|\alpha_G, \mu_G, \tau_G^2)$ is obtained in an analogous fashion.

Note that the discreteness of the DP priors for H and G induces a clustering of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ in their prior (and, hence, also in their posterior). Let n_θ^* $(\leqslant n_0+n_1)$ and $\{\theta_l^* : l = 1, \ldots, n_\theta^*\}$, and n_ϕ^* $(\leqslant n_1)$ and $\{\phi_l^* : l = 1, \ldots, n_\phi^*\}$ be the number of and values of the distinct components in $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ respectively. The DP precision parameters α_H and α_G play a key role in the induced priors for n_θ^* and n_ϕ^* (see, for example, Escobar and West (1995)). For instance, with $\Gamma(a_H, b_H)$ and $\Gamma(a_G, b_G)$ priors for α_H and α_G respectively, we have $E(n_\theta^*) \approx a_H b_H^{-1} \log\{1 + (n_0+n_1)a_H^{-1}b_H\}$ and $E(n_\phi^*) \approx a_G b_G^{-1} \log(1 + n_1 a_G^{-1} b_G)$, for moderately large n_0 and n_1.

A.2. Markov chain Monte Carlo posterior simulation

After integrating H and G over their DP priors, the resulting posterior $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \psi|\mathcal{D})$ is proportional to

$$\prod_{i=1}^{n_0} N(x_{0i}; \theta_i, \sigma^2) \prod_{j=1}^{n_1} N\left\{x_{1j}; \max(\theta_{n_0+j}, \phi_j), \sigma^2\right\} p(\sigma^2)\, p(\boldsymbol{\theta}|\alpha_H, \mu_H, \tau_H^2)\, p(\boldsymbol{\phi}|\alpha_G, \mu_G, \tau_G^2)\, p(\psi),$$

where $p(\sigma^2)$ and $p(\psi)$ are the priors of σ^2 and ψ. We sample from $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \psi|\mathcal{D})$ by using an MCMC algorithm that combines ideas from Escobar and West (1995) and Neal (2000).

First, note that, on the basis of the DP Polya urn representation, the prior full conditional for each θ_i, $p(\theta_i|\{\theta_l : l \neq i\}, \alpha_H, \mu_H, \tau_H^2)$, $i = 1, \ldots, n_0 + n_1$, has point masses $(\alpha_H + n_0 + n_1 - 1)^{-1}$ at θ_l, $l \neq i$, and continuous mass $\alpha_H(\alpha_H + n_0 + n_1 - 1)^{-1}$ on the $N(\mu_H, \tau_H^2)$ distribution. Analogously, each ϕ_j, $j = 1, \ldots, n_1$, has a mixed prior full conditional distribution, $p(\phi_j|\{\phi_l : l \neq j\}, \alpha_G, \mu_G, \tau_G^2)$, with point masses $(\alpha_G + n_1 - 1)^{-1}$ at ϕ_l, $l \neq j$, and continuous mass $\alpha_G(\alpha_G + n_1 - 1)^{-1}$ on the $N(\mu_G, \tau_G^2)$ distribution. Therefore, it is straightforward to sample directly from the posterior full conditional for each θ_i, $i = 1, \ldots, n_0$, since it is a mixed distribution with point masses at the distinct values among the θ_l, $l \neq i$, and continuous mass on a normal distribution; the parameters of this normal distribution and the weights of the mixed distribution can be readily obtained by combining $p(\theta_i|\{\theta_l : l \neq i\}, \alpha_H, \mu_H, \tau_H^2)$ with the $N(x_{0i}; \theta_i, \sigma^2)$ term. For each $j = 1, \ldots, n_1$, the posterior full conditional for the pair of latent mixing parameters (θ_{n_0+j}, ϕ_j) is proportional to

$$N\{x_{1j}; \max(\theta_{n_0+j}, \phi_j), \sigma^2\}\, p(\theta_{n_0+j}|\{\theta_l : l \neq n_0 + j\}, \alpha_H, \mu_H, \tau_H^2)\, p(\phi_j|\{\phi_l : l \neq j\}, \alpha_G, \mu_G, \tau_G^2).$$

We update each pair (θ_{n_0+j}, ϕ_j) with a Metropolis step, which involves proposed draws $(\tilde{\theta}_{n_0+j}, \tilde{\phi}_j)$ from $p(\cdot|\{\theta_l : l \neq n_0 + j\}, \alpha_H, \mu_H, \tau_H^2)\, p(\cdot|\{\phi_l : l \neq j\}, \alpha_G, \mu_G, \tau_G^2)$ that are accepted with probability $\min[1, N(x_{1j}; \max(\tilde{\theta}_{n_0+j}, \tilde{\phi}_j), \sigma^2\}/N\{x_{1j}; \max(\theta_{n_0+j}^{(\mathrm{old})}, \phi_j^{(\mathrm{old})}), \sigma^2\}]$, where $(\theta_{n_0+j}^{(\mathrm{old})}, \phi_j^{(\mathrm{old})})$ is the current state of the chain. The full conditional posteriors for μ_H and μ_G are normal, and the full conditionals for τ_H^2, τ_G^2 and σ^2 are inverse gamma. Finally, we update α_H and α_G by using the data augmentation technique from Escobar and West (1995).

We note that this is an improved MCMC algorithm (in terms of its mixing properties) compared with the Gibbs sampler that was described in Gelfand and Kottas (2001) where each of the θ_{n_0+j} and the ϕ_j were updated in separate steps. The superiority of this new MCMC method was particularly pronounced in the analysis of the data that was discussed in Section 4.2 for which the Gibbs sampler from Gelfand and Kottas (2001) failed to produce stable inference results.

A.3. Posterior predictive inference and conditional predictive ordinate estimation

The posterior predictive density for a *new* realization $x_{0,\mathrm{new}}$ that is associated with the non-infected population, with corresponding latent mixing parameter θ_{new}, can be expressed as

$$p(x_{0,\mathrm{new}}|\mathcal{D}) = \int\int N(x_{0,\mathrm{new}}; \theta_{\mathrm{new}}, \sigma^2)\, p(\theta_{\mathrm{new}}|\boldsymbol{\theta}, \alpha_H, \mu_H, \tau_H^2)\, p(\boldsymbol{\theta}, \sigma^2, \alpha_H, \mu_H, \tau_H^2|\mathcal{D}), \tag{2}$$

with

$$p(\theta_{\mathrm{new}}|\boldsymbol{\theta}, \alpha_H, \mu_H, \tau_H^2) = \frac{\alpha_H}{\alpha_H + n_0 + n_1} N(\theta_{\mathrm{new}}; \mu_H, \tau_H^2) + \frac{1}{\alpha_H + n_0 + n_1} \sum_{l=1}^{n_\theta^*} n_l\, \delta_{\theta_l^*}(\theta_{\mathrm{new}}), \tag{3}$$

where n_l is the size of the lth cluster of $\boldsymbol{\theta}$, $l = 1, \ldots, n_\theta^*$. Analogously, the posterior predictive distribution for a new $x_{1,\mathrm{new}}$ from the infected population,

$$\begin{aligned} p(x_{1,\mathrm{new}}|\mathcal{D}) = \int\int\int & N\{x_{1,\mathrm{new}}; \max(\theta_{\mathrm{new}}, \phi_{\mathrm{new}}), \sigma^2\}\, p(\theta_{\mathrm{new}}|\boldsymbol{\theta}, \alpha_H, \mu_H, \tau_H^2) \\ & \times p(\phi_{\mathrm{new}}|\boldsymbol{\phi}, \alpha_G, \mu_G, \tau_G^2)\, p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \boldsymbol{\psi}|\mathcal{D}), \end{aligned} \tag{4}$$

where

$$p(\phi_{\mathrm{new}}|\boldsymbol{\phi}, \alpha_G, \mu_G, \tau_G^2) = \frac{\alpha_G}{\alpha_G + n_1} N(\phi_{\mathrm{new}}; \mu_G, \tau_G^2) + \frac{1}{\alpha_G + n_1} \sum_{l=1}^{n_\phi^*} m_l\, \delta_{\phi_l^*}(\phi_{\mathrm{new}}), \tag{5}$$

with m_l denoting the size of the lth cluster of $\boldsymbol{\phi}$, $l = 1, \ldots, n_\phi^*$. Therefore, using the posterior samples from $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \boldsymbol{\psi}|\mathcal{D})$ and expressions (2)–(5), we can obtain posterior predictive estimates for the densities and distribution functions corresponding to the non-infected and infected populations.

Having developed the posterior predictive distributions, $\mathrm{CPO}_{0i} = p(x_{0i}|\mathbf{x}_{0(i)}, \mathbf{x}_1)$, $i = 1, \ldots, n_0$, and $\mathrm{CPO}_{1j} = p(x_{1j}|\mathbf{x}_0, \mathbf{x}_{1(j)})$, $j = 1, \ldots, n_1$, can be computed by using an extension of the approach in Mukhopadhyay and Gelfand (1997). (Recall that $\mathbf{x}_{0(i)} = \{x_{0i'} : i' \neq i\}$, $i = 1, \ldots, n_0$, and $\mathbf{x}_{1(j)} = \{x_{1j'} : j' \neq j\}$, $j = 1, \ldots, n_1$.) In particular, it can be shown that

$$\text{CPO}_{0i} = A_{0i}^{-1} E\left\{ \frac{N(x_{0i}; \theta_{\text{new}}, \sigma^2)}{N(x_{0i}; \theta_i, \sigma^2)} p(\theta_{\text{new}} | \boldsymbol{\theta}, \alpha_H, \mu_H, \tau_H^2) \right\}$$

with $A_{0i} = E\{N(x_{0i}; \theta_i, \sigma^2)^{-1}\}$ and $p(\theta_{\text{new}} | \boldsymbol{\theta}, \alpha_H, \mu_H, \tau_H^2)$ given in equation (3). Analogously,

$$\text{CPO}_{1j} = A_{1j}^{-1} E\left[\frac{N\{x_{1j}; \max(\theta_{\text{new}}, \phi_{\text{new}}), \sigma^2\}}{N\{x_{1j}; \max(\theta_{n_0+j}, \phi_j), \sigma^2\}} p(\theta_{\text{new}} | \boldsymbol{\theta}, \alpha_H, \mu_H, \tau_H^2)\, p(\phi_{\text{new}} | \boldsymbol{\phi}, \alpha_G, \mu_G, \tau_G^2) \right],$$

where $A_{1j} = E([N\{x_{1j}; \max(\theta_{n_0+j}, \phi_j), \sigma^2\}]^{-1})$ and $p(\phi_{\text{new}} | \boldsymbol{\phi}, \alpha_G, \mu_G, \tau_G^2)$ is given in equation (5). All the above expectations are with respect to $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \psi | \mathcal{D})$, which can be sampled as discussed previously, resulting in straightforward Monte Carlo estimation of the CPOs.

We note here that, in principle, it is not required to introduce the new latent mixing parameters θ_{new} and ϕ_{new} in the derivation of the expressions for the CPO_{0i} and CPO_{1j}; in fact, the simpler expressions are given through A_{0i}^{-1} and A_{1j}^{-1} respectively. However, in practice, the formulation above seems preferable, since it yields CPO estimates that stabilize numerically with a substantially smaller number of posterior samples.

A.4. Full inference for the mixture densities and the receiver operating characteristic curve

The posterior predictive distributions, which were developed in expressions (2) and (4), yield only posterior point estimates for the mixture distribution functions $F_0(t; H, \sigma^2)$ and $F_1(t; H, G, \sigma^2)$ (or the corresponding densities) at any point $t \in \mathbb{R}$. For our application, of direct interest is full posterior inference for $F_0(t; H, \sigma^2)$ and $F_1(t; H, G, \sigma^2)$, and for ROC(u) (which will also result in the posterior of AUC). Such general inference under model (1) requires the posteriors of H and G. Samples from these posteriors can be obtained by augmenting the MCMC algorithm that was described above with steps that involve sampling from DPs. The key result is the representation of the full posterior corresponding to model (1), specifically, based on Antoniak (1974),

$$p(H, G, \boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \psi | \mathcal{D}) = p(H | \boldsymbol{\theta}, \alpha_H, \mu_H, \tau_H^2)\, p(G | \boldsymbol{\phi}, \alpha_G, \mu_G, \tau_G^2)\, p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \psi | \mathcal{D})$$

where $p(H | \boldsymbol{\theta}, \alpha_H, \mu_H, \tau_H^2)$ denotes a DP distribution with precision parameter $\alpha_H + n_0 + n_1$ and base distribution

$$\alpha_H (\alpha_H + n_0 + n_1)^{-1} N(\cdot; \mu_H, \tau_H^2) + (\alpha_H + n_0 + n_1)^{-1} \sum_{i=1}^{n_0+n_1} \delta_{\theta_i}(\cdot),$$

and $p(G | \boldsymbol{\phi}, \alpha_G, \mu_G, \tau_G^2)$ denotes a DP with precision parameter $\alpha_G + n_1$ and base distribution

$$\alpha_G (\alpha_G + n_1)^{-1} N(\cdot; \mu_G, \tau_G^2) + (\alpha_G + n_1)^{-1} \sum_{j=1}^{n_1} \delta_{\phi_j}(\cdot).$$

To sample from these two DPs, the DP stick breaking representation (which was discussed in Section 2.2) can be used with a truncation approximation (see, for example, Kottas (2006)). Then, posterior samples H_b, G_b (and σ_b^2), $b = 1, \ldots, B$, for H, G (and σ^2) yield directly samples $\{F_{0b}(t_m) = \int N(t_m; \theta, \sigma_b^2)\, \mathrm{d}H_b(\theta) : b = 1, \ldots, B\}$ and $\{F_{1b}(t_m) = \int \int N\{t_m; \max(\theta, \phi), \sigma_b^2\}\, \mathrm{d}H_b(\theta)\, \mathrm{d}G_b(\phi) : b = 1, \ldots, B\}$ from the posteriors of $F_0(t_m; H, \sigma^2)$ and $F_1(t_m; H, G, \sigma^2)$ respectively, for any set of grid points $\{t_m : m = 1, \ldots, M\}$ over a range of interest for the non-infected and infected populations. We can analogously sample from the posteriors of the mixture densities $f_0(t_m; H, \sigma^2)$ and $f_1(t_m; H, G, \sigma^2)$.

Inference for the ROC curve, over a grid of points u in (0,1), can be obtained as follows. First, we invert (with interpolation) each posterior realization $\{F_{0b}(t_m) : m = 1, \ldots, M\}$ from the random distribution function $F_0(\cdot; H, \sigma^2)$ to obtain samples $\{\eta_b(u) : b = 1, \ldots, B\}$ from the posterior of $F_0^{-1}(1 - u; H, \sigma^2)$, for any specified grid point u in (0, 1). Next, computing

$$1 - F_1\{\eta_b(u); H_b, G_b, \sigma_b^2\} = 1 - \int \int N\{\eta_b(u); \max(\theta, \phi), \sigma_b^2\}\, \mathrm{d}H_b(\theta)\, \mathrm{d}G_b(\phi)$$

for each $b = 1, \ldots, B$ yields the posterior of ROC(u), for the fixed u in (0, 1). Repeating for all grid points u results in posterior realizations (up to the finite grid approximation) for the random ROC curve

$\mathrm{ROC}(\cdot) \equiv \mathrm{ROC}(\cdot; H, G, \sigma^2) = 1 - F_1\{F_0^{-1}(1-u; H, \sigma^2); H, G, \sigma^2\}$. A posterior sample for AUC emerges through direct numerical integration of these posterior realizations; they also yield posterior point and (pointwise) interval estimates for the ROC curve.

Appendix B: Posterior inference under the mixture of Polya tree model

We describe here the computational approach to inference under the stochastically ordered MPT model of Section 3.2. We also show propriety of the posterior under the flat prior $p(\mu_G, \tau_G, \mu_H, \tau_H) \propto 1$.

The likelihood, $\mathcal{L}(\mathcal{D}) \equiv \mathcal{L}(\mathcal{D}|\mathcal{X}_H, \mathcal{X}_G, \boldsymbol{\theta}_H, \boldsymbol{\theta}_G, c_H, c_G)$, can be obtained by differentiating $F_1(t) = H(t)G(t)$. Specifically,

$$\begin{aligned}\mathcal{L}(\mathcal{D}) = &\prod_{i=1}^{n_1} \{h(x_{1i}|\mathcal{X}_H, \boldsymbol{\theta}_H, c_H) G(x_{1i}|\mathcal{X}_G, \boldsymbol{\theta}_G, c_G) + H(x_{1i}|\mathcal{X}_H, \boldsymbol{\theta}_H, c_H)\, g(x_{1i}|\mathcal{X}_G, \boldsymbol{\theta}_G, c_G)\} \\ &\times \prod_{i=1}^{n_0} h(x_{0i}|\mathcal{X}_H, \boldsymbol{\theta}_H, c_H), \end{aligned} \tag{6}$$

where h denotes the density corresponding to H. The CDF $G(x|\mathcal{X}_G, \boldsymbol{\theta}_G, c_G)$ is given by

$$G(x|\mathcal{X}_G, \boldsymbol{\theta}_G, c_G) = \sum_{k=1}^{k_{\boldsymbol{\theta}_G}(J,x)-1} p_{\mathcal{X}_G}(k) + p_{\mathcal{X}_G}\{k_{\boldsymbol{\theta}_G}(J,x)\}\{2^J G_{\boldsymbol{\theta}_G}(x) - k_{\boldsymbol{\theta}_G}(J,x) + 1\}, \tag{7}$$

where $k_{\boldsymbol{\theta}_G}(J,x) = \mathrm{Int}[2^J \Phi\{(x-\mu_G)/\tau_G\} + 1]$ and $p_{\mathcal{X}_G}(k) = \Pi_{i=1}^{J} X^G_{i,\mathrm{Int}\{(k-1)2^{i-J}+1\}}$. Here, $\Phi(\cdot)$ denotes the standard normal CDF, and $\mathrm{Int}(z)$ denotes the integer part of z. The corresponding density is given by

$$g(x|\mathcal{X}_G, \boldsymbol{\theta}_G, c_G) = 2^J\, p_{\mathcal{X}_G}\{k_{\boldsymbol{\theta}_G}(J,x)\}\, N(x; \mu_G, \tau_G^2). \tag{8}$$

Similar expressions hold for the CDF and probability distribution function of H.

The priors $p(\mathcal{X}_G|c_G)$ and $p(\mathcal{X}_H|c_H)$ are each the product of $2^J - 1$ beta densities, e.g.

$$p(\mathcal{X}_G|c_G) = \prod_{j=1}^{J} \prod_{k=1}^{2^{j-1}} \mathrm{beta}(X^G_{j,2k-1}; c_G j^2, c_G j^2)$$

where $\mathrm{beta}(x; a, b)$ is the probability distribution function of a $\mathrm{beta}(a, b)$ distribution. Considering the general version of the model with random c_H and c_G, and assuming that $\boldsymbol{\theta}_H$, $\boldsymbol{\theta}_G$, c_H and c_G are *a priori* independent, then the posterior density is proportional to

$$\mathcal{L}(\mathcal{D}|\mathcal{X}_H, \mathcal{X}_G, \boldsymbol{\theta}_H, \boldsymbol{\theta}_G, c_H, c_G)\, p(\boldsymbol{\theta}_H)\, p(\boldsymbol{\theta}_G)\, p(c_H) p(c_G)\, p(\mathcal{X}_H|c_H)\, p(\mathcal{X}_G|c_G),$$

and the posterior can be approximated numerically by using, for instance, Gibbs sampling.

The conjugacy result for PTs (Ferguson, 1974) that is typically used to sample from the full conditionals for $\mathcal{X}_H$ and $\mathcal{X}_G$ does not apply here. So an alternative method of simulation from these distributions is required, such as application of the Metropolis–Hastings algorithm as developed in Hanson (2006), section 3, to this setting. Sampling the remaining full conditionals can also be accomplished by using Metropolis–Hastings steps.

In practice, for a given data set, we used block updates for $(\mu_G, \tau_G, \mu_H, \tau_H)$ and block updates for (c_G, c_H) and sampled the elements of $\mathcal{X}_G$ and $\mathcal{X}_H$ one at a time, all using random-walk Metropolis–Hastings steps. The initial covariance matrices for block updating $(\mu_G, \tau_G, \mu_H, \tau_H)$ were obtained from consideration of the parametric model without stochastic constraints, and by 'refining' it through several initial runs.

Note that given G and H (i.e. given $(\mathcal{X}_G, \mu_G, \tau_G, \mathcal{X}_H, \mu_H, \tau_H)$) the ROC curve can be expressed as

$$\begin{aligned}\mathrm{ROC}(u) &= S_1\{S_0^{-1}(u)\} \\ &= 1 - G\{H^{-1}(1-u)\}\, H\{H^{-1}(1-u)\} \\ &= 1 - G\{H^{-1}(1-u)\}(1-u).\end{aligned}$$

The CDF G is given by equation (7) whereas the quantile function that is associated with H is given by

$$H^{-1}(q|\mathcal{X}_H, \boldsymbol{\theta}_H) = \Phi^{-1}_{\theta_H}\left\{\frac{q - \sum_{k=1}^{N} p_{\mathcal{X}_H}(k) + N\, p_{\mathcal{X}_H}(N)}{2^J\, p_{\mathcal{X}_H}(N)}\right\}, \tag{9}$$

where N is such that $\Sigma_{k=1}^{N-1} p_{\mathcal{X}_H}(k) < q \leqslant \Sigma_{k=1}^{N} p_{\mathcal{X}_H}(k)$ and $\Phi^{-1}_{\mu,\tau}(q) = \mu + \tau\,\Phi^{-1}(q)$. Given the MCMC iterates $\{(\mathcal{X}_{Hb}, \mathcal{X}_{Gb}, \mu_{Hb}, \mu_{Gb}, \tau_{Hb}, \tau_{Gb}) : b = 1, \ldots, B\}$ we can find the posterior median ROC(t_m) over a grid of points $t_m \in \{t_1, \ldots, t_M\}$ and corresponding posterior quantiles.

Regarding the computation of the CPOs, we have

$$\frac{1}{\mathrm{CPO}_{0i}} = E\left\{\frac{1}{h(x_{0i}|\mathcal{X}_H, \boldsymbol{\theta}_H, c_H)}\right\},$$

and

$$\frac{1}{\mathrm{CPO}_{1j}} = E\left\{\frac{1}{h(x_{1j}|\mathcal{X}_H, \boldsymbol{\theta}_H, c_H)\, G(x_{1j}|\mathcal{X}_G, \boldsymbol{\theta}_G, c_G) + H(x_{1j}|\mathcal{X}_H, \boldsymbol{\theta}_H, c_H)\, g(x_{1j}|\mathcal{X}_G, \boldsymbol{\theta}_G, c_G)}\right\},$$

where the expectations are with respect to the posterior of the full model. Hence, the CPOs can be readily computed through Monte Carlo integrations using the samples from $p(\mathcal{X}_H, \mathcal{X}_G, \boldsymbol{\theta}_H, \boldsymbol{\theta}_G, c_H, c_G|\mathcal{D})$.

Finally, we note that an improper prior on $(\mu_G, \tau_G, \mu_H, \tau_H)$ does not necessarily lead to an improper posterior distribution, since the likelihood is bounded. In particular,

$$\begin{aligned}\mathcal{L}(\mathcal{D}) &\leqslant \left\{\prod_{i=1}^{n_1} h(x_{1i}|\mathcal{X}_H, \boldsymbol{\theta}_H, c_H) + g(x_{1i}|\mathcal{X}_G, \boldsymbol{\theta}_G, c_G)\right\} \prod_{i=1}^{n_0} h(x_{0i}|\mathcal{X}_H, \boldsymbol{\theta}_H, c_H) \\ &\leqslant 2^{(n_0+n_1)J}\left\{\prod_{i=1}^{n_1} N(x_{1i}; \mu_H, \tau_H^2) + N(x_{1i}; \mu_G, \tau_G^2)\right\} \prod_{i=1}^{n_0} N(x_{0i}; \mu_H, \tau_H^2).\end{aligned}$$

Clearly, if n_0 and n_1 are greater than 2, the flat prior $p(\mu_G, \tau_G, \mu_H, \tau_H) \propto 1$ leads to a proper posterior.

References

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

Bamber, D. (1975) The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.*, **12**, 387–415.

Blackwell, D. and MacQueen, J. B. (1973) Ferguson distributions via Pólya urn schemes. *Ann. Statist.*, **1**, 353–355.

Branscum, A. J., Johnson, W. O., Hanson, T. E. and Gardner, I. A. (2006) Bayesian semiparametric ROC curve estimation and disease diagnosis. *Technical Report 001*. Department of Biostatistics, University of Kentucky, Lexington.

Chen, M.-H., Shao, Q.-M. and Ibrahim, J. G. (2000) *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

Choi, Y.-K., Johnson, W. O., Collins, M. T. and Gardner, I. A. (2006) Bayesian inferences for Receiver Operating Characteristic curves in the absence of a gold standard. *J. Agric. Biol. Environ. Statist.*, **11**, 210–229.

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. *Clin. Diag. Lab. Immun.*, **12**, 685–692.

Draper, D. (1999) Discussion on 'Bayesian nonparametric inference for random distributions and related functions' (by S. G. Walker, P. Damien, P. W. Laud and A. F. M. Smith). *J. R. Statist. Soc.* B, **61**, 510–513.

Erkanli, A., Sung, M., Costello, E. J. and Angold, A. (2006) Bayesian semi-parametric ROC analysis. *Statist. Med.*, **25**, 3905–3928.

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

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

Ferguson, T. S. (1974) Prior distributions on spaces of probability measures. *Ann. Statist.*, **2**, 615–629.

Geisser, S. and Eddy, W. F. (1979) A predictive approach to model selection. *J. Am. Statist. Ass.*, **74**, 153–160.

Gelfand, A. E. and Kottas, A. (2001) Nonparametric Bayesian modeling for stochastic order. *Ann. Inst. Statist. Math.*, **53**, 865–876.

Gelfand, A. E. and Kuo, L. (1991) Nonparametric Bayesian bioassay including ordered polytomous response. *Biometrika*, **78**, 657–666.

Greiner, M., Pfeiffer, D. and Smith, R. D. (2000) Principles and practical application of receiver operating characteristic analysis for diagnostic tests. *Prev. Veter. Med.*, **45**, 23–41.
Hanson, T. E. (2006) Inference for mixtures of finite Polya tree models. *J. Am. Statist. Ass.*, **101**, 1548–1565.
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* (eds D. K. Dey and C. R. Rao), pp. 245–278. Amsterdam: Elsevier.
Hanson, T. and Johnson, W. O. (2002) Modeling regression error with a mixture of Polya trees. *J. Am. Statist. Ass.*, **97**, 1020–1033.
Kottas, A. (2006) Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *J. Statist. Planng Inf.*, **136**, 578–596.
Kottas, A., Branco, M. D. and Gelfand, A. E. (2002) A nonparametric Bayesian modeling approach for cytogenetic dosimetry. *Biometrics*, **58**, 593–600.
Kottas, A. and Gelfand, A. E. (2001) Modeling variability order: a semiparametric Bayesian approach. *Methodol. Comput. Appl. Probab.*, **3**, 427–442.
Lavine, M. (1992) Some aspects of Polya tree distributions for statistical modeling. *Ann. Statist.*, **20**, 1222–1235.
Lavine, M. (1994) More aspects of Polya tree distributions for statistical modeling. *Ann. Statist.*, **22**, 1161–1176.
Le, C. T. (2006) A solution to the most basic optimization problem associated with an ROC curve. *Statist. Meth. Med. Res.*, **15**, 571–584.
Lloyd, C. J. (1998) Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *J. Am. Statist. Ass.*, **93**, 1356–1364.
MacEachern, S. N. and Müller, P. (2000) Efficient MCMC schemes for robust model extensions using encompassing Dirichlet process mixture models. In *Robust Bayesian Analysis* (eds F. Ruggeri and D. Rios-Insua), pp. 295–315. New York: Springer.
Mukhopadhyay, S. and Gelfand, A. E. (1997) Dirichlet process mixed generalized linear models. *J. Am. Statist. Ass.*, **92**, 633–639.
Müller, P. and Quintana, F. A. (2004) Nonparametric Bayesian data analysis. *Statist. Sci.*, **19**, 95–110.
Neal, R. M. (2000) Markov chain sampling methods for Dirichlet process mixture models. *J. Computnl Graph. Statist.*, **9**, 249–265.
Pepe, M. S. (2003) *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press.
Sethuraman, J. (1994) A constructive definition of the Dirichlet prior. *Statist. Sin.*, **4**, 639–650.
Shaked, M. and Shanthikumar, J. G. (1994) *Stochastic Orders and Their Applications*. Boston: Academic Press.
Zou, K. H., Hall, W. J. and Shapiro, D. E. (1997) Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statist. Med.*, **16**, 2143–2156.