Inácio de Carvalho, V., Jara, A., and de Carvalho, M. (2015), "Bayesian Nonparametric Approaches for ROC Curve Inference," In Nonparametric Bayesian Methods in Biostatistics and Bioinformatics," Eds R. Mitra and P. Müller, Cham: Springer.

Bayesian nonparametric approaches for ROC curve inference

Vanda Inácio de Carvalho, Alejandro Jara, and Miguel de Carvalho

Abstract The development of medical diagnostic tests is of great importance in clinical practice, public health, and medical research. The receiver operating characteristic (ROC) curve is a popular tool for evaluating the accuracy of such tests. We review Bayesian nonparametric methods based on Dirichlet process mixtures and the Bayesian bootstrap for ROC curve estimation and regression. The methods are illustrated by means of data concerning diagnosis of lung cancer in women.

1 Introduction

Medical diagnostic tests are designed to discriminate between alternative states of health, generally referred throughout to as diseased and non-diseased/healthy states. Their ability to discriminate between these two states must be rigorously assessed trough statistical analysis before the test is approved for use in practice. In what follows, we assume the existence of a gold standard test, that is, a test that perfectly classifies the individuals as diseased and non-diseased. Compared to the truth one wants to know how well the test being evaluated performs.

The accuracy of a dichotomous test, a test that yields binary results (e.g., positive or negative), can be summarized by its sensitivity and specificity. The sensitivity (Se) is the test-specific probability of correctly detecting diseased subjects, while the specificity (Sp) is the test-specific probability of correctly detecting healthy subjects. In turn, the accuracy of a continuous scale diagnostic test is measured by

Vanda Inácio de Carvalho
Pontifica Universidad Católica de Chile, Chile, e-mail: icalhau@mat.puc.cl

Alejandro Jara
Pontifica Universidad Católica de Chile, Chile, e-mail: ajara@mat.uc.cl

Miguel de Carvalho
Pontifica Universidad Católica de Chile, Chile, e-mail: mdecarvalho@mat.puc.cl

the separation of test outcomes distribution in the diseased and non-diseased populations. The receiver operating characteristic (ROC) curve, which is a plot of Se against $1 - \text{Sp}$ for all cutoff points that can be used to convert continuous test outcomes into dichotomous outcomes, measures exactly such amount of separation and it is probably the most widely used tool to evaluate the accuracy of continuous or ordinal tests.

A critical aspect when developing inference for ROC curves is the specification of a probability distribution for the test outcomes in the diseased and healthy groups. The main issue is that parametric models, such as the binormal model (arising when a normal distribution is assumed for both populations), are often too restrictive to capture nonstandard features of the data, such as skewness and multimodality, potentially leading to unsatisfactory inferences on the ROC curve. In this situations, we would like to relax parametric assumptions in order to gain modeling flexibility and robustness against misspecification of a parametric statistical model. Specifically, we would like to consider flexible modelling approaches that can handle nonstandard features of the data when that is needed, but that do not overfit the data when parametric assumptions are valid.

Moreover, recently, the interest on the subject has moved beyond determining the basic accuracy of a test. It has been recognized that the discriminatory power of a test is often affected by patient-specific characteristics, such as age or gender. In this situations, the parameter of interest is a collection of ROC curves associated with different covariate levels. In this context, understanding the covariate impact on the ROC curve may provide useful information regarding the test accuracy towards different populations or conditions. On the other hand, ignoring the covariate effects may lead to biased inferences about the test accuracy. As in the no-covariate case, here it is also important to consider flexible modelling approaches for assessing the effect of the covariates on test accuracy and, consequently, on the corresponding ROC curves.

In this chapter, we discuss two Bayesian nonparametric (BNP) approaches that are used to obtain data-driven inferences for a single ROC curve, based on mixtures induced by a Dirichlet process (DP) and on the Bayesian bootstrap. We also discuss an approach to model covariate-dependent ROC curves based on mixture models induced by a dependent DP (DDP), which allows for the entire distribution of the test outcomes, in each population, to smoothly change as a function of covariates. The chapter is organized as follows. In Section 2 we provide background material on ROC curves. BNP approaches for single ROC curve estimation are discussed in Section 3. A BNP ROC regression model is discussed in Section 4. In Section 5 we illustrate the methods using data concerning diagnosis of lung cancer in women. We conclude with a short discussion in Section 6.

2 ROC curves

Let Y_0 and Y_1 be two independent random variables denoting the diagnostic test outcomes in the non-diseased and diseased populations, with cumulative distribution function (CDF) F_0 and F_1, respectively. Further, let c be a cutoff value for defining a positive test result and, without loss of generality, we proceed with the assumption that a subject is classified as diseased when the test outcome is greater or equal than c and as non-diseased when it is below c. Then, for each cutoff value c, the sensitivity and specificity associated with such decision criterion are

$$\mathrm{Se}(c) = \Pr(Y_1 \geq c) = 1 - F_1(c), \qquad \mathrm{Sp}(c) = \Pr(Y_0 < c) = F_0(c).$$

Obviously, for each value of c, we obtain a different sensitivity and specificity. The ROC curve summarizes the tradeoffs between Se and 1-Sp (also known as false positive fraction) as the cutoff c is varied and it corresponds to the set of points

$$\{(1 - F_0(c), 1 - F_1(c)) : c \in \mathbb{R}\}.$$

Alternatively, and letting $p = 1 - F_0(c)$, the ROC curve can be expressed as

$$\mathrm{ROC}(p) = 1 - F_1\{F_0^{-1}(1-p)\}, \quad 0 \leq p \leq 1, \tag{1}$$

where $F_0^{-1}(1-p) = \inf\{z : F_0(z) \geq 1-p\}$. ROC curves measure the amount of separation between the distribution of the test outcomes in the diseased and non-diseased populations. Figure 1 illustrates the effect of separation on the resulting ROC curve. When both distributions completely overlap, the ROC curve is the diagonal line of the unit square (that is, $\mathrm{Se}(c) = 1 - \mathrm{Sp}(c)$ for all c), thus indicating an useless test. On the other hand, the more separated the distributions the closer the ROC curve is to the point $(0,1)$ in the unit square. A curve that reaches the point $(0,1)$ has $\mathrm{Sp}(c) = \mathrm{Se}(c) = 1$ for some cutoff c, and hence corresponds to a perfect test. As it is clear from expression (1), estimating the ROC curve is basically a matter of estimating the distribution functions of the diseased and non-diseased populations and, hence, flexible models for estimating such distributions are in order.

Related to the ROC curve is the notion of placement value (Pepe and Cai, 2004), which is simply a standardization of test outcomes with respect to a reference population. Let $U = 1 - F_0(Y_1)$ be the placement value of diseased subjects with respect to the non-diseased population. This variable quantifies the degree of separation between the two populations. Specifically, if the test outcomes in the two populations are highly separated, the placement of most diseased individuals is at the upper tail of the non-diseased distribution, so that most diseased individuals will have small placement values. In turn, if the populations overlap substantially, U will have a $\mathrm{Uniform}(0,1)$ distribution. Interestingly, the ROC curve turns out to be the cumulative distribution function of U

$$\Pr(U \leq p) = \Pr(1 - F_0(Y_1) \leq p) = 1 - F_1\{F_0^{-1}(1-p)\} = \mathrm{ROC}(p). \tag{2}$$

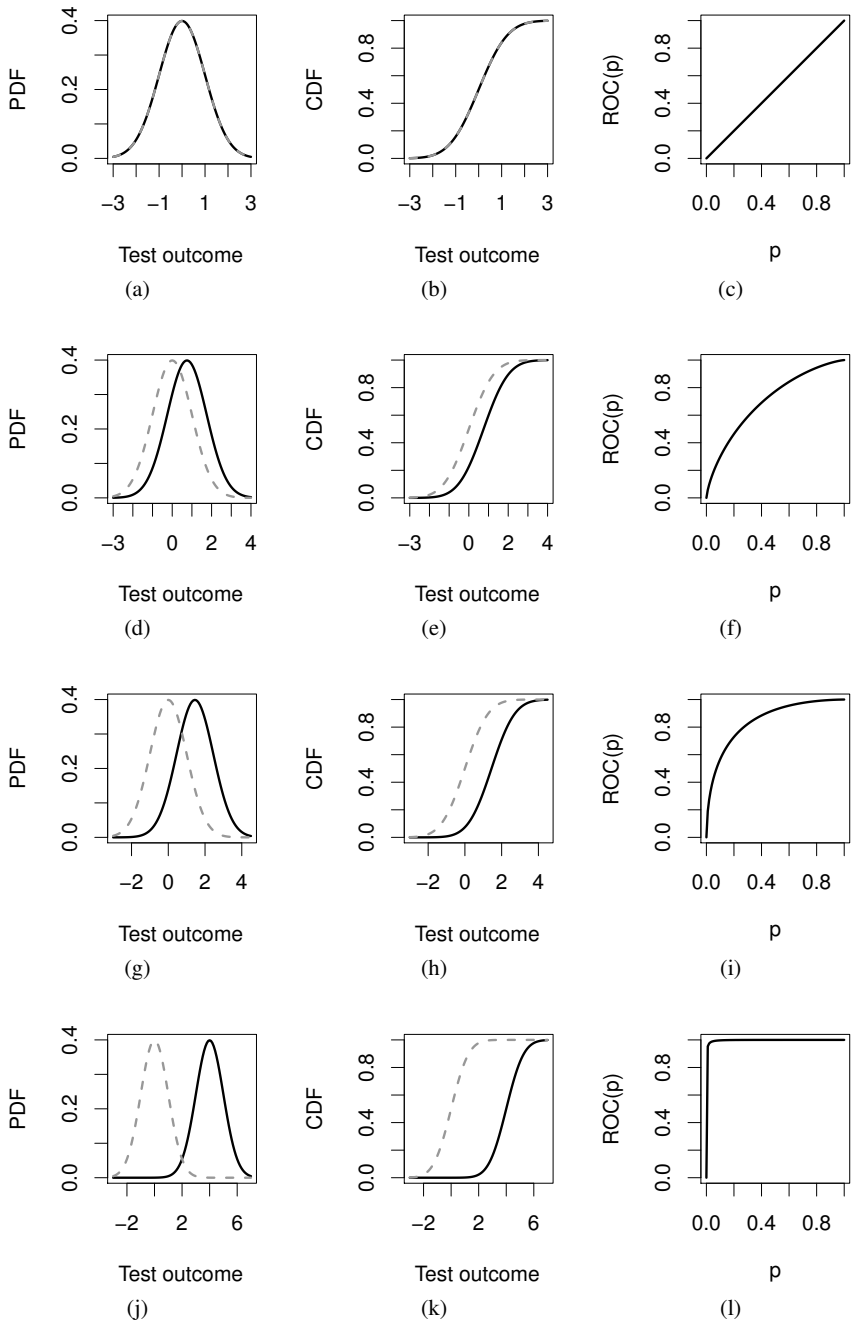


Fig. 1 ROC curve illustrations: The first column displays the densities of the test outcomes for diseased (solid black line) and non-diseased populations (dashed grey line). The second column displays the corresponding distribution functions of the test outcomes for diseased (solid black line) and non-diseased populations (dashed grey line). The third column displays the corresponding ROC curves.

It is common to summarize the information of the ROC curve into a single summary index and the most widely used is the area under the ROC curve (AUC), which is defined as

$$\text{AUC} = \int_0^1 \text{ROC}(p)\mathrm{d}p. \tag{3}$$

The AUC can be interpreted as the probability that an individual chosen from the diseased population exhibits a test outcome greater than the one exhibited by a randomly selected individual from the non-diseased population, that is, $\text{AUC} = \Pr(Y_1 > Y_0)$. A test with a perfect discriminatory ability would have $\text{AUC} = 1$, while a test with no discriminatory power would have $\text{AUC} = 0.5$. Although there are some other summary indices available, such as the Youden index (Fluss et al, 2005), which has the nice feature of providing an optimal cutoff for screening subjects in practice, or the partial AUC (Dodd and Pepe, 2003), which is a meaningful measure for cases where only a specific region of the ROC curve (e.g., high sensitivities or specificities) is of clinical interest, throughout this chapter we use the AUC as the preferred summary measure of diagnostic accuracy.

Now, suppose that along with Y_0 and Y_1, covariate vectors $\mathbf{x}_0$ and $\mathbf{x}_1$ are also available. Hereafter, we assume that these covariates are the same in both populations. However, this not always have to be the case. For instance, the severity of disease could play an important role on the discriminatory power of the test. As a natural extension of the ROC curve, the conditional or covariate-dependent ROC curve, for a given covariate level $\mathbf{x}$, is defined as

$$\text{ROC}(p \mid \mathbf{x}) = 1 - F_1\{F_0^{-1}(1-p \mid \mathbf{x}) \mid \mathbf{x}\}, \tag{4}$$

where $F_0(\cdot \mid \mathbf{x})$ and $F_1(\cdot \mid \mathbf{x})$ denote the conditional distribution function of Y_0 and Y_1 given covariate $\mathbf{x}$, respectively. For each value of $\mathbf{x}$, we possibly obtain a different ROC curve and, hence, also a possibly different AUC value, which is computed simply by replacing (4) in (3).

There is a vast literature on parametric, semiparametric, and nonparametric frequentist ROC data analysis. The books by Pepe (2003) and Zhou et al (2011) discuss many frequentist approaches to ROC curve estimation and regression. See also the recent surveys by Gonçalves et al (2014) and Pardo-Fernández et al (2014). The amount of existing work in the Bayesian literature is by comparison reduced. This is particularly valid for the BNP literature, which is fairly limited. Recent work on the latter includes the DP mixture (DPM) model-based approach of Erkanli et al (2006), the Bayesian bootstrap ROC curve estimator of Gu et al (2008), and the stochastic ordering approach of Hanson et al (2008b). Moreover, Branscum et al (2008) used mixtures of finite Polya trees to analyze ROC data when the true disease status is unknown (that is, when there is no gold standard), while Hanson et al (2008a) used bivariate mixtures of finite Polya trees to model data from two continuous tests. Additionally, Inácio et al (2011) proposed the use of mixture of finite Polya trees to model the ROC surface for problems where the patients have to be classified into one of three ordered classes. In what respects ROC regression, Inácio de Carvalho et al (2013) proposed to model the conditional ROC curve using DDPs,

whereas Rodríguez and Martínez (2014) used Gaussian process priors to model the mean and variance functions in each population and then computed the corresponding induced ROC curve. Finally, Branscum et al (2014) proposed a method based on mixtures of finite Polya trees to model ROC regression data when there is not a gold standard test available.

3 Modeling approaches for the no covariate case

3.1 DPM models

When seeking for flexible modeling approaches and inferences for the distributions of the test outcomes in each population, mixture models appear as a natural option. More specifically, mixtures of normal distributions are particularly well suited for our purposes. Let $(Y_{01},\ldots,Y_{0n_0})$ and $(Y_{11},\ldots,Y_{1n_1})$ be random samples of sizes n_0 and n_1 from the non-diseased and diseased populations, respectively. It would be natural to assume that

$$Y_{01},\ldots,Y_{0n_0} \mid F_0 \overset{ind.}{\sim} F_0,$$

and

$$Y_{11},\ldots,Y_{1n_1} \mid F_1 \overset{ind.}{\sim} F_1,$$

with

$$F_h(\cdot) = \sum_{k=1}^{K_h} \omega_{hk}\Phi(\cdot \mid \mu_{hk},\sigma^2_{hk}), \quad h \in \{0,1\}, \tag{5}$$

where $\Phi(\cdot \mid \mu,\sigma^2)$ denotes the CDF of the normal distribution with mean μ and variance σ^2. Thus, each test outcome would arise from one of the K_h mixture components, with each component having its own mean and variance. The model in expression (5) can be equivalently written as

$$F_h(\cdot) = \int \Phi(\cdot \mid \mu,\sigma^2)\mathrm{d}G_h(\mu,\sigma^2),$$

where G_h is a discrete mixing distribution given by

$$G_h(\cdot) = \sum_{k=1}^{K_h} \omega_{hk}\delta_{(\mu_{hk},\sigma^2_{hk})}(\cdot),$$

with $\delta_a(\cdot)$ denoting the Dirac measure at a. Usually, the weights $\{\omega_{hk}\}$ are assigned a Dirichlet distribution, while the component specific parameters $\{(\mu_{hk},\sigma^2_{hk})\}$ arise from a prior distribution, say, $G_{0h}(\mu_h,\sigma^2_h)$, typically, a normal-inverse-gamma distribution. Hence, placing a prior on the collection

$$(\{\omega_{hk}\},\{(\mu_{hk},\sigma^2_{hk})\}),$$

is equivalent to placing a prior on the discrete mixture distribution G_h. A drawback of this model specification is that we must choose the number of components K_h, which is not a trivial task in general. Although there are methods available that place an explicit parametric prior on K_h, they tend to be quite difficult to implement efficiently. An alternative is to use a DP prior (Ferguson, 1973, 1974) for G_h, which, on one hand, offers the theoretical advantage of having full weak support on all mixing distributions and, on the other hand, the practical advantage of automatically determining the number of components that best fits a given dataset. We write $G_h \sim \text{DP}(\alpha_h, G_{0h})$ to denote that a DP prior is being assumed for G_h, which is defined in terms of a parametric centering distribution G_{0h} (for which $\text{E}(G_h) = G_{0h}$), and a precision parameter α_h ($\alpha_h > 0$) which controls the uncertainty of G_h about G_{0h}.

Undoubtedly, the most useful definition of the DP is its constructive definition (Sethuraman, 1994), according to which G_h has an almost sure representation of the form

$$G_h(\cdot) = \sum_{k=1}^{\infty} \omega_{hk} \delta_{(\mu_{hk}, \sigma^2_{hk})}(\cdot), \tag{6}$$

where $(\mu_{hk}, \sigma^2_{hk}) \overset{i.i.d.}{\sim} G_{0h}$ and the weights arise from a stick breaking construction $\omega_{h1} = v_{h1}$, and $\omega_{hk} = v_{hk} \prod_{l<k}(1 - v_{hl})$, for $k \geq 2$, with $v_{hk} \overset{i.i.d}{\sim} \text{Beta}(1, \alpha_h)$.

The resulting model for the test outcomes in each population is then a DPM of normals and is written as

$$F_h(\cdot) = \int \Phi(\cdot \mid \mu, \sigma^2) \text{d}G_h(\mu, \sigma^2), \quad G_h \sim \text{DP}(\alpha_h, G_{0h}), \tag{7}$$

where the centering distribution G_{0h} is defined on $\mathbb{R} \times \mathbb{R}_+$. More specifically, we take G_{0h} to be the normal-inverse-gamma distribution, that is,

$$G_{0h} \equiv \text{N}(m_h, S_h)\text{IG}(\tau_{h1}/2, \tau_{h2}/2),$$

where $\text{N}(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2 and $\text{IG}(a, b)$ refers to the inverse-gamma distribution with parameters a and b. The stick-breaking representation of the DP given in expression (6) allows us to write expression (7) as the following countably infinite mixture of normals

$$F_h(\cdot) = \sum_{k=1}^{\infty} \omega_{hk} \Phi(\cdot \mid \mu_{hk}, \sigma^2_{hk}).$$

The model specification is completed by assuming the following independent hyper-priors

$$\begin{aligned} \alpha_h &\sim \text{G}(a_h, b_h), \quad \tau_{h2} \sim \text{G}(\tau_{sh1}/2, \tau_{sh2}/2), \\ m_h &\sim \text{N}(\mu_{m_h}, S_{m_h}), \quad S_h \sim \text{IG}(\nu_h, \Psi_h), \end{aligned}$$

where $\text{G}(a, b)$ refers to the gamma distribution with parameters a and b.

Posterior inference can be conducted using two different kinds of Markov chain Monte Carlo (MCMC) strategies: (i) to employ a truncation of the stick-breaking representation (Ishwaran and James, 2001) or (ii) to use a marginal Gibbs sampling where the mixing distributions are integrated out from the model (MacEachern and Müller, 1998; Neal, 2000). Finally, we can plug-in each MCMC realization of F_0 and F_1 in expression (1) and compute the corresponding realization of the ROC curve. Note that the computation of the ROC curve requires the evaluation of the quantile function of F_0, which is done numerically. A model similar to the one described here was proposed by Erkanli et al (2006).

3.2 Bayesian bootstrap

The Bayesian bootstrap (BB) estimator of the ROC curve was proposed by Gu et al (2008) and it is a computationally simple, yet robust, estimator. We start by outlining how the BB works in the one-population setting. Let $(Y_1,\ldots,Y_n)$ be a random sample from an unknown distribution F and suppose that F itself is the parameter of interest. In Efron's frequentist bootstrap (Efron, 1979), estimation and inference about F are obtained by repeatedly generating bootstrap samples, where each sample is drawn with replacement from the original data. In the bth bootstrap replicate, $F^{(b)}$ is computed as

$$F^b(\cdot) = \sum_{i=1}^{n} \pi_i^{(b)} \delta_{Y_i}(\cdot), \tag{8}$$

where $\pi_i^{(b)}$ is the proportion of times Y_i appears in the bth bootstrap sample, with $\pi_i^{(b)}$ taking values on $\{0, 1/n, \ldots, n/n\}$. By contrast, in Rubin's BB (Rubin, 1981), the weights $\pi_i^{(b)}$ in expression (8) are assigned an Dirichlet$_n(1,\ldots,1)$ distribution and thus are smoother than those from the frequentist bootstrap. It is important to stress that in the BB the data is regarded as fixed and so we do not resample from it. The BB has connections with the DP. Specifically, it can be regarded as a non-informative version of the DP, which can be obtained by letting the precision parameter tending to zero (Gasparini, 1995, Theorem 2).

The representation of the ROC curve given in expression (2) provides the rationale for the following two step BB algorithm, which we fully describe due to its simplicity. Let us suppose, again, that $(Y_{01},\ldots,Y_{0n_0})$ and $(Y_{11},\ldots,Y_{1n_1})$ are random samples from the non-diseased and diseased populations and let B be the number of BB resamples.

Bayesian bootstrap algorithm

For $b = 1,\ldots,B$:

Step 1 (**Compute the placement values based on the BB resampling**)

For $j = 1,\ldots,n_1$, compute the placement values

$$U_j = \sum_{i=1}^{n_0} q_i^{(b)} I(Y_{0i} \geq Y_{1j}), \quad (q_1^{(b)}, \ldots, q_{n_0}^{(b)}) \overset{ind.}{\sim} \text{Dirichlet}_{n_0}(1, \ldots, 1).$$

Step 2 (**Generate a random realization of the ROC curve**)
Based on (2), generate a random realization of $\text{ROC}(p)$, the cumulative distribution function of $(U_1, \ldots, U_{n_1})$, where

$$\text{ROC}^{(b)}(p) = \sum_{j=1}^{n_1} r_j^{(b)} I(U_j \leq p), \quad (r_1^{(b)}, \ldots, r_{n_1}^{(b)}) \overset{ind.}{\sim} \text{Dirichlet}_{n_1}(1, \ldots, 1),$$

with $0 \leq p \leq 1$. Compute the AUC associated to $\text{ROC}^{(b)}(p)$, $\text{AUC}^{(b)}$, using numerical integration.

The BB estimate of the ROC curve, denoted as $\widehat{\text{ROC}}^{\text{BB}}(p)$, is then obtained by averaging the random realizations of the ROC curve, that is,

$$\widehat{\text{ROC}}^{\text{BB}}(p) = \frac{1}{B} \sum_{b=1}^{B} \text{ROC}^{(b)}(p), \quad 0 \leq p \leq 1.$$

Similarly,

$$\widehat{\text{AUC}}^{\text{BB}} = \frac{1}{B} \sum_{b=1}^{B} \text{AUC}^{(b)}.$$

4 Modeling approaches for the covariate case

Let $\{(\mathbf{x}_{01}, Y_{01}), \ldots, (\mathbf{x}_{0n_0}, Y_{0n_0})\}$ and $\{(\mathbf{x}_{11}, Y_{11}), \ldots, (\mathbf{x}_{1n_1}, Y_{1n_1})\}$ be regression data for the non-diseased and diseased groups, respectively, where $\mathbf{x}_{0i} \in \mathscr{X} \subseteq \mathbb{R}^p$ and $\mathbf{x}_{1j} \in \mathscr{X} \subseteq \mathbb{R}^p$ are p-dimensional covariate vectors and Y_{0i} and Y_{1j} are test outcomes, $i = 1, \ldots, n_0$, $j = 1, \ldots, n_1$. It is assumed, that given the covariates, the test outcomes in the diseased and non-diseased populations are independent and that

$$Y_{0i} \mid \mathbf{x}_{0i} \overset{ind.}{\sim} F_0(\cdot \mid \mathbf{x}_{0i}), \quad i = 1, \ldots, n_0,$$
$$Y_{1j} \mid \mathbf{x}_{1j} \overset{ind.}{\sim} F_1(\cdot \mid \mathbf{x}_{1j}), \quad j = 1, \ldots, n_1.$$

Here, we detail the approach proposed by Inácio de Carvalho et al (2013) for the conditional ROC curve estimation problem, which extends the no covariate approach of Section 3.1. Specifically, these authors proposed a model for the conditional ROC curves based on the specification of a probability model for the entire collection of distributions $\mathscr{F}_h = \{F_h(\cdot \mid \mathbf{x}) : \mathbf{x} \in \mathscr{X}\}$, $h \in \{0, 1\}$, and they further modeled the conditional distributions in each population using the following

covariate-dependent mixture of normal models

$$F_h(\cdot \mid \mathbf{x}) = \int \Phi(\cdot \mid \mu, \sigma^2) \mathrm{d}G_{h\mathbf{x}}(\mu, \sigma^2), \quad h \in \{0,1\}.$$

The probability model for the conditional distributions is induced by specifying a prior for the collection of mixing distributions

$$G_{h\mathscr{X}} = \{G_{h\mathbf{x}} : \mathbf{x} \in \mathscr{X}\} \sim \mathscr{G}_h,$$

where $G_{h\mathbf{x}}$ denotes the random mixing distribution at covariate $\mathbf{x}$, which is defined on $\mathbb{R} \times \mathbb{R}_+$, and $\mathscr{G}_h$ is the prior for the collection $G_{h\mathscr{X}}$.

One possibility for modelling $\mathscr{G}_h$ is the DDP proposed by MacEachern (2000), which is built upon the constructive definition of the DP in (6), where the atoms and the components of the weights are realizations of a stochastic process over $\mathscr{X}$, and the weights arise from a stick-breaking representation. Justified by results in Barrientos et al (2012), on the full support of MacEachern's DDPs, Inácio de Carvalho et al (2013) considered the 'single weights' DDP (De Iorio et al, 2004, 2009; De la Cruz et al, 2007; Jara et al, 2010), where only the atoms are indexed by the covariates, thus resulting in the following specification for the conditional random mixing distribution

$$G_{h\mathbf{x}}(\cdot) = \sum_{k=1}^{\infty} \omega_{hk} \delta_{\theta_{hk}(\mathbf{x})}(\cdot), \tag{9}$$

where the weights $\{\omega_{hk}\}_{k=1}^{\infty}$ match those from a standard DP and the atoms are given by $\theta_{hk}(\mathbf{x}) = (m_{hk}(\mathbf{x}), \sigma^2_{hk})$, where $\{m_{hk}(\mathbf{x}) : \mathbf{x} \in \mathscr{X}\}_{k=1}^{\infty}$ are i.i.d. Gaussian processes which are independent across h.

Although such formulation leads to a very flexible prior, it implies sampling realizations of the Gaussian processes at each distinct value of the covariate and, thus, inferences could take prohibitively long. This motivated Inácio de Carvalho et al (2013) to elaborate on a linear DDP (LDDP) prior formulation (De Iorio et al, 2004, 2009; Jara et al, 2010), where the Gaussian processes are replaced by sufficiently rich linear (in the coefficients) functions, $m_{hk}(\mathbf{x}) = \mathbf{z}'\beta_{hk}$. Here $\mathbf{z}$ is a q-dimensional design vector possibly including non-linear transformations of the original covariates $\mathbf{x}$. To this end, the authors considered an additive formulation based on B-splines (Eilers and Marx, 1996), referred to as B-splines DDP,

$$m_{hk}(\mathbf{x}) = \beta_{hk0} + \sum_{l=1}^{p} \left(\sum_{n=1}^{K_l} \beta_{hkln} \psi(x_l, d_l) \right),$$

where $\psi_n(x,d)$ corresponds to the nth B–spline basis function of degree d evaluated at x, and $\beta_{hk} = (\beta_{hk0}, \ldots, \beta_{hkpK_p})$. This formulation allows for the inclusion of discrete and continuous predictors.

Thus, under the LDDP formulation, the base stochastic processes are replaced with a group-specific distribution G_{0h} that generates the component specific regression coefficients and variances. Therefore, the B-splines DDP mixture model can be

equivalently formulated as a DP mixture of Gaussian regression models

$$F_h(\cdot \mid \mathbf{x}) = \int \Phi(\cdot \mid \mathbf{z}'\beta, \sigma^2) \mathrm{d}G_h(\beta, \sigma^2), \quad G_h \sim \mathrm{DP}(\alpha_h, G_{0h}). \tag{10}$$

For each group, normal-inverse-gamma distributions were used for the parametric centering distribution,

$$G_{0h} \equiv N_q(\mu_h, \Sigma_h) \times \mathrm{IG}(\tau_{h1}/2, \tau_{h2}/2).$$

The model specification is completed by specifying the following hyper-priors

$$\begin{aligned} &\alpha_h \sim \mathrm{G}(a_h, b_h), \quad \tau_{h2} \sim \mathrm{G}(\tau_{sh1}/2, \tau_{sh2}/2), \\ &\mu_h \sim \mathrm{N}_q(\mathbf{m}_h, \mathbf{S}_h), \quad \Sigma_h \sim \mathrm{IW}_q(\nu_h, \Psi_h). \end{aligned}$$

With regard to posterior inference, the computational strategies for Dirichlet process mixture models referred in Section 3.3.1 apply here in the covariate setup directly. Finally, after obtaining MCMC samples for each of the parameters, we can plug-in, for each covariate $\mathbf{x}$, each MCMC realization of $F_0(\cdot \mid \mathbf{x})$ and $F_1(\cdot \mid \mathbf{x})$ in (4) and compute the corresponding realization of the conditional ROC curve. The model previously described is implemented in the function `LDDProc` of the `R` library `DPpackage` (Jara et al, 2011).

5 Illustration

The accuracy of a soluble isoform of epidermal growth factor receptor (sEFGR), present in blood, as a diagnostic test for lung cancer in women is investigated. How this accuracy may vary with age is also subject of interest. The data were collected from a case-control study conducted at the Mayo clinic in Minnesota between 1998 and 2003. The dataset includes information for 140 non-diseased women and 101 lung cancer cases. This dataset was previously analyzed by Branscum et al (2013).

Figure 2 shows the histogram of the $-\log(\mathrm{sEFGR})$ in both populations. The minus sign is due to the fact that the values of sEFGR tend to be lower for lung cancer cases than for controls, and so with the minus sign the usual convention that diseased individuals tend to have larger test outcomes than the non-diseased ones applies. As it can be observed, normality does not seem to apply, especially for the non-diseased population, where a bimodality is easily noticed. Figure 2 also displays the estimated densities, in each group of women, under the DPM of normals model, and we can see that the model captures well the bimodality in the non-diseased group, as well as, a certain skewness in the diseased group. The hyper-priors of the DPM of normals model were set to $a_h = 5$, $b_h = 1$, $\tau_{h1} = 2$, $\tau_{sh1} = 2$, $\tau_{sh2} = 10$, $\mu_{mh} = 0$, $S_{mh} = 100$, $\nu_h = 5$, and $\Psi_h = 1$, for $h \in \{0, 1\}$, while the BB estimates were obtained using 5000 resamples. With respect to the estimation of the cumulative distribution functions, which are displayed in Figure 3 (Panels (a) to (f)), it can be observed that

the estimates provided by the DPM of normals and the BB are almost indistinguishable. When superimposing these fits (DP mixture and BB) with the one obtained by the binormal model, a discrepancy can be seen, especially in the non-diseased group. The resulting ROC curves, also presented in Figure 3, are smooth and practically identical (except the one obtained by the binormal fit) and the corresponding posterior means (95% credible interval) of the AUC are $0.792\ (0.728, 0.848)$ under the DPM model and $0.792\ (0.731, 0.848)$ under the BB method. These values reveal a quite good discriminatory ability of the sEFGR to detect lung cancer in women.

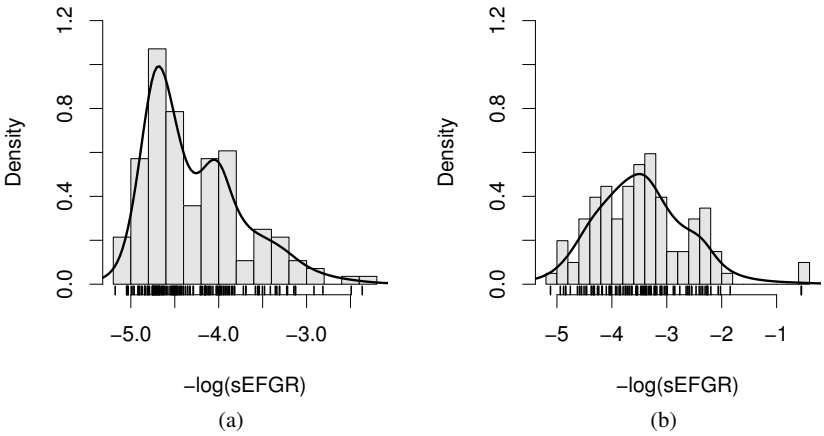
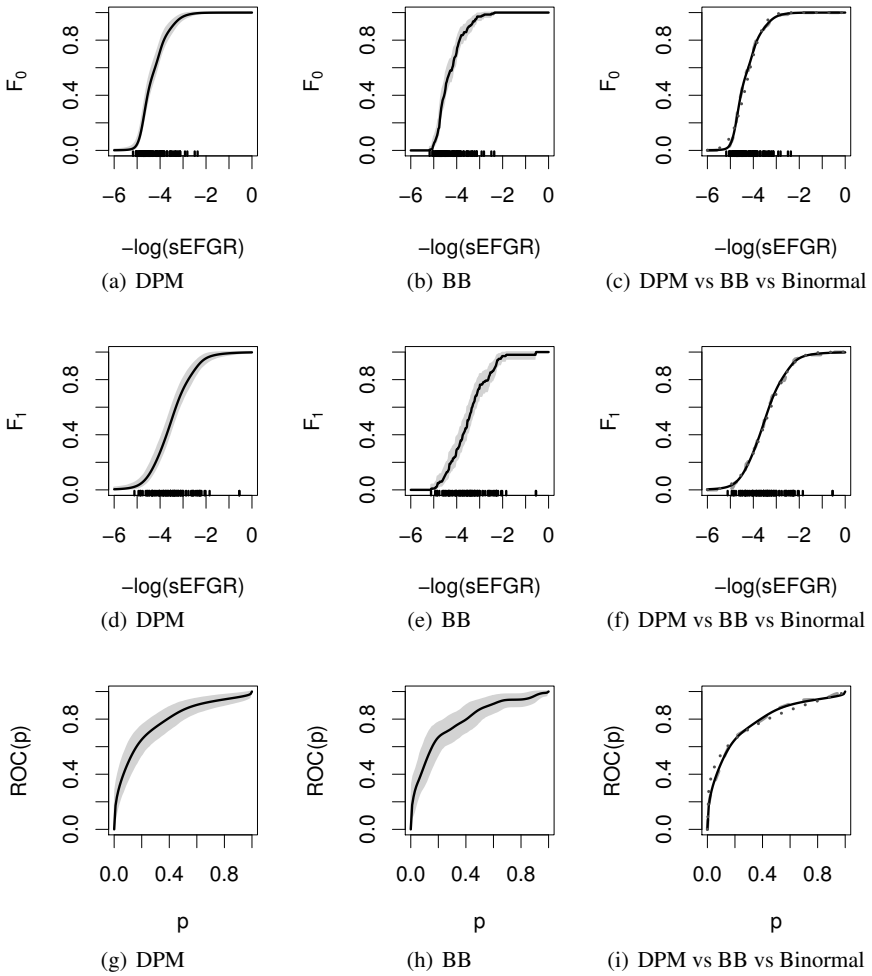


Fig. 2 sEFGR data: Histogram of the $-\log(\text{sEFGR})$ in the nondiseased (Panel a) and diseased (Panel b) populations along with a rug representation of the data. The posterior mean of the density for each population under the DPM of normals models is displayed as a solid line.

We now examine the age effect on the accuracy of the sEFGR. The B-splines dependent DPM of normals model was fit by assuming $K_1 = 3$, $a_h = 5$, $b_h = 1$, $\tau_{h1} = 2$, $\tau_{sh1} = 2$, $\tau_{sh2} = 10$, $\mathbf{m}_h = (0,0,0,0)$, $\mathbf{S}_h = 100 \times \mathbf{I}_4$, $\nu_h = 5$, and $\Psi_h = \mathbf{I}_4$, for $h \in \{0,1\}$. Figure 4 shows the posterior means for the conditional mean functions, along with point-wise 95% credible bands for $-\log(\text{sEFGR})$ levels. These estimates are overlaid on the top of the raw data. This figure suggests that the $-\log(\text{sEFGR})$ levels are more concentrated in the non-diseased than in the diseased women, across age and, further, a slightly nonlinear behavior of the conditional mean function of both groups can be observed.

Figure 5 present the estimated posterior means, along with 95% point-wise credible bands, of the conditional distribution functions in the two groups of women at three selected ages (40, 55, and 70 years old), and a change across age is clearly seen. Obviously, the same is visible in terms of the corresponding estimated ROC curves.



(a) DPM (b) BB (c) DPM vs BB vs Binormal

(d) DPM (e) BB (f) DPM vs BB vs Binormal

(g) DPM (h) BB (i) DPM vs BB vs Binormal

Fig. 3 sEFGR data: Panels (a) and (b) show the estimated posterior mean (solid black line), along with the point-wise 95% credible bands (grey area) of the cumulative distribution function of the non-diseased population, under the DPM of normals model and the Bayesian bootstrap, respectively. In Panel (c) the two estimates are superimposed along with the estimates obtained under the binormal model (solid black line represents the DPM estimate, light grey dashed line represents the BB estimate, and dark grey dotted line is the binormal estimate). Panels (d), (e), and (f) show the analogous figures but in terms of the cumulative distribution function of the diseased population, and panels (g), (h), and (i) in terms of the ROC curve.

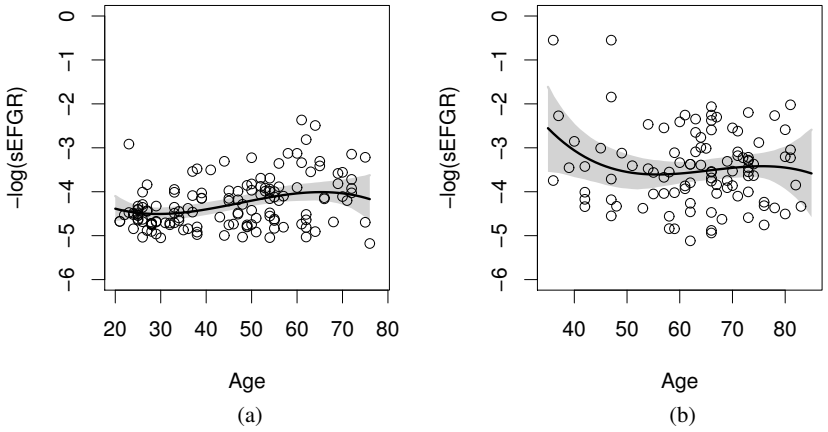


Fig. 4 sEFGR data: Posterior mean (solid line) and 95% point-wise credible band (grey area) for the conditional mean function in the group of non-diseased women (Panel (a)) and in the group of diseased women (Panel (b)).

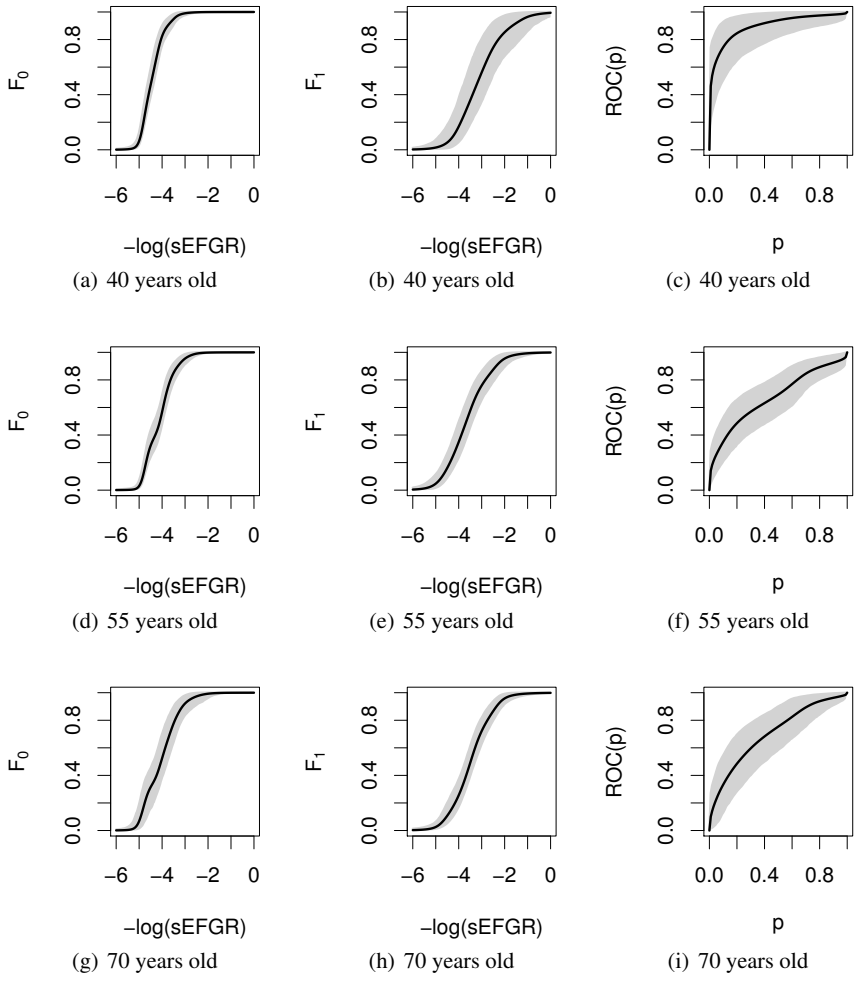


Fig. 5 sEFGR data: Panels (a), (d), and (g) display the estimated posterior mean (solid line), as well as, the 95% point-wise credible bands (grey area) of the conditional distribution function, in the non-diseased group, for ages of 40, 55, and 70 years old. Panels (b), (e), and (h) show the analogous figures but in the diseased group. Panels (c), (f), and (i) show the corresponding ROC curves.

To examine the age effect further, Figure 6 shows the estimated posterior mean, as well as the 95% point-wise credible band, of the AUC as a function of age. This figure suggests a decrease in AUC until an age around 60 years old, and then a slight increase.

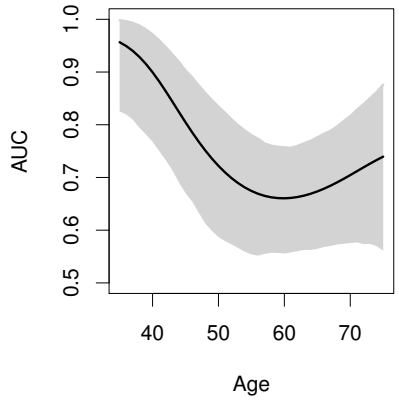


Fig. 6 sEFGR data: Posterior mean (solid line) along with the 95% point-wise credible band (grey area) for the AUC as a function of age.

6 Concluding remarks

ROC curves are a valuable tool for assessing the discriminatory power of continuous diagnostic tests. We have described and illustrated BNP approaches for ROC curve estimation and regression. Specifically, we have discussed DPM models and the BB for ROC curve estimation and an extension for the regression case based on DDP mixture models. A nice feature of the latter model is that the complete distribution of the test outcomes is allowed to smoothly change with the values of the covariates instead of just one or two characteristics (such as the mean and/or variance), as implied for most ROC regression models.

Topics of future research on BNP methods for ROC analysis include, among others, modeling diagnostic tests with mass at zero, optimal combinations of multiple tests, and time-dependent ROC curves. We end remarking that `R` packages for the implementation of ROC analysis tools are of great importance for practitioners.

Acknowledgements The authors thank Adam Branscum for sharing the lung cancer dataset with them. The first authors was supported by Fondecyt grant 11130541. The second authors was supported by Fondecyt grant 1141193. The third author was supported by Fondecyt grant 11121186.

References

Barrientos AF, Jara A, Quintana F (2012) On the support of MacEachern's dependent Dirichlet processes and extensions. Bayesian Analysis 7:277–310

Branscum AJ, Johnson WO, Hanson TE, Gardner IA (2008) Bayesian semiparametric ROC curve estimation and disease diagnosis. Statistics in Medicine 27:2474–2496

Branscum AJ, Johnson WO, Baron AT (2013) Robust medical test evaluation using flexible Bayesian semiparametric regression models. Epidemiology Research International 2103:1–8

Branscum AJ, Johnson WO, Hanson TE, Baron AT (2014) Flexible regression models for ROC and risk analysis, with or without a gold standard. Submitted

De la Cruz R, Quintana FA, Müller P (2007) Semiparametric Bayesian classification with longitudinal markers. Applied Statistics 56:119–137

De Iorio M, Müller P, Rosner GL, MacEachern SN (2004) An ANOVA model for dependent random measures. Journal of the American Statistical Association 99:205–215

De Iorio M, Johnson WO, Müller P, Rosner GL (2009) Bayesian nonparametric non-proportional hazards survival modelling. Biometrics 65:762–771

Dodd LE, Pepe MS (2003) Partial auc estimation and regression. Biometrics 59:614–623

Efron B (1979) Bootstrap methods: another look at the jackknife. The Annals of Statistics 7:1–26

Eilers PHC, Marx BD (1996) Flexible smoothing with B-splines and penalties. Statistical Science 11:89–121

Erkanli A, Sung M, Costello EJ, Angold A (2006) Bayesian semiparametric ROC analysis. Statistics in Medicine 25:3905–3928

Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. Annals of Statistics 1:209–230

Ferguson TS (1974) Prior distribution on the spaces of probability measures. Annals of Statistics 2:615–629

Fluss R, Faraggi D, Reiser B (2005) Estimation of the Youden index and its associated cutoff point. Biometrical Journal 47:458–472

Gasparini M (1995) Exact multivariate Bayesian bootstrap distributions of moments. The Annals of Statistics 23:762–768

Gonçalves L, Subtil A, Oliveira R, Bermúdez PZ (2014) ROC curve estimation: An overview. REVSTAT—Statistical Journal 12:1–20

Gu J, Ghosal S, Roy A (2008) Bayesian bootstrap estimation of ROC curve. Statistics in Medicine 27:5407–5420

Hanson T, Branscum A, Gardner I (2008a) Multivariate mixtures of Polya trees for modelling ROC data. Statistical Modelling 8:81–96

Hanson T, Kottas A, Branscum AJ (2008b) Modelling stochastic order in the analysis of receiver operating characteristic data: Bayesian non-parametric approaches. Journal of the Royal Statistical Society, Series C 57:207–225

Inácio V, Turkman AA, Nakas CT, Alonzo TA (2011) Nonparametric Bayesian estimation of the three-way receiver operating characteristic surface. Biometrical Journal 53:1011–1024

Inácio de Carvalho V, Jara A, Hanson TE, de Carvalho M (2013) Bayesian nonparametric ROC regression modeling. Bayesian Analysis 8:623–646

Ishwaran H, James LF (2001) Gibbs sampling methods for stick-breaking priors. Journal of the American Statistical Association 96:161–173

Jara A, Lesaffre E, De Iorio M, Quintana FA (2010) Bayesian semiparametric inference for multivariate doubly-interval-censored data. Annals of Applied Statistics 4:2126–2149

Jara A, Hanson T, Quintana F, Müller P, Rosner GL (2011) DPpackage: Bayesian semi- and nonparametric modeling in R. Journal of Statistical Software 40:1–30

MacEachern SN (2000) Dependent Dirichlet processes. Tech. rep., Department of Statistics, The Ohio State University

MacEachern SN, Müller P (1998) Estimating mixture of Dirichlet process models. Journal of Computational and Graphical Statistics 7:223–238

Neal R (2000) Markov chain sampling methods for Dirichlet process mixture models. Journal of Computational and Graphical Statistics 9:249–265

Pardo-Fernández JC, Rodríguez-Álvarez MX, Van Keilegom I (2014) A review on ROC curves in the presence of covariates. REVSTAT —Statistical Journal 12:21–41

Pepe MS (2003) The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, New York

Pepe MS, Cai T (2004) The analysis of placement values for evaluating discriminatory measures. Biometrics 60:528–535

Rodríguez A, Martínez JC (2014) Bayesian semiparametric estimation of covariate-dependent ROC curves. Biostatistics 15:353–369

Rubin DB (1981) The Bayesian bootstrap. The Annals of Statistics 9:130–134

Sethuraman J (1994) A constructive definition of Dirichlet priors. Statistica Sinica 2:639–650

Zhou XH, Obuchowski NA, McClish DK (2011) Statistical Methods in Diagnostic Medicine, 2nd Ed. Wiley, New York