ROCnReg: An R Package for Receiver Operating Characteristic Curve Inference With and Without Covariates

by María Xosé Rodríguez-Álvarez and Vanda Inácio

Abstract This paper introduces the package ROCnReg that allows estimating the pooled ROC curve, the covariate-specific ROC curve, and the covariate-adjusted ROC curve by different methods, both from (semi) parametric and nonparametric perspectives and within Bayesian and frequentist paradigms. From the estimated ROC curve (pooled, covariate-specific, or covariate-adjusted), several summary measures of discriminatory accuracy, such as the (partial) area under the ROC curve and the Youden index, can be obtained. The package also provides functions to obtain ROC-based optimal threshold values using several criteria, namely, the Youden index criterion and the criterion that sets a target value for the false positive fraction. For the Bayesian methods, we provide tools for assessing model fit via posterior predictive checks, while the model choice can be carried out via several information criteria. Numerical and graphical outputs are provided for all methods. This is the only package implementing Bayesian procedures for ROC curves.

1 Introduction

The receiver operating characteristic (ROC) curve (Metz, 1978) is, unarguably, the most popular tool used for evaluating the discriminatory ability of continuous-outcome diagnostic tests. The ROC curve displays the false positive fraction (FPF) against the true positive fraction (TPF) for all possible threshold values that can be used to dichotomize the test result. The ROC curve thus provides a global description of the trade-off between the FPF and the TPF of the test as the threshold changes. Plenty of parametric and semi/nonparametric methods are available for estimating the ROC curve, either from frequentist or Bayesian viewpoints, and we refer the interested reader to Pepe (1998, Chapter 5), Zhou et al. (2011, Chapter 4), Inácio et al. (2020), and references therein.

It is known that in many situations, the outcome of a test and, possibly, its discriminatory capacity can be affected by covariates. Two different ROC-based measures that incorporate covariate information have been proposed: the covariate-specific or conditional ROC curve (see, e.g., Pepe, 2003, Chapter 6) and the covariate-adjusted ROC curve (Janes and Pepe, 2009). The formal definition of both curves is given in Section Notation and definitions. Succinctly, a covariate-specific ROC curve is an ROC curve that conditions on a specific covariate value, thus describing the accuracy of the test in the ‘subpopulation’ defined by that covariate value. On the other hand, the covariate-adjusted ROC curve is a weighted average of covariate-specific ROC curves. Regarding estimation, since the seminal paper of Pepe (1998), a plethora of methods have been proposed in the literature for the estimation of the covariate-specific ROC curve and associated summary measures. Without being exhaustive, we mention the work of Faraggi (2003), Rodríguez-Álvarez et al. (2011a,b), Inácio de Carvalho et al. (2013), and Inácio de Carvalho et al. (2017). A detailed review can be found in Rodríguez-Álvarez et al. (2011c), Pardo-Fernández et al. (2014), and Inácio et al. (2020). With respect to the covariate-adjusted ROC curve, estimation has been discussed in Janes and Pepe (2009), Rodríguez-Álvarez et al. (2011a), Guan et al. (2012), and Inácio de Carvalho and Rodríguez-Álvarez (2018).

A few R packages for ROC curve analysis are available on the Comprehensive R Archive Network and, as far as we are aware, all of them implementing frequentist approaches. The package sROC (Wang, 2012) contains functions to perform nonparametric, kernel-based, estimation of ROC curves. pROC (Robin et al., 2011) offers a set of tools to visualize, smooth, and compare ROC curves, and nsROC (Pérez Fernández et al., 2018) also allows estimating ROC curves, building confidence bands as well as comparing several curves both for dependent and independent data (i.e., data arising from paired and unpaired study designs, respectively). However, covariate information cannot be explicitly taken into account in any of these packages. The packages ROCRegression (available at https://bitbucket.org/mxrodriguez/rocregression) and npROCRegression (Rodriguez-Alvarez and Roca-Pardinas, 2017) provide routines to estimate semiparametrically and nonparametrically, under a frequentist framework, the covariate-specific ROC curve. We also mention OptimalCutpoints (López-Ratón et al., 2013) and ThresholdROC (Perez Jaume et al., 2017) that provide a collection of functions for point and interval estimation of optimal thresholds for continuous diagnostic tests. To the best of our knowledge, there is no statistical software package implementing Bayesian inference for ROC curves and associated summary indices and optimal thresholds.

To close this gap, in this paper we introduce the ROCnReg package that allows conducting
### Method Description

<table>
<thead>
<tr>
<th>Pooled ROC curve</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>emp (Frequentist) empirical estimator</td>
<td>(Hsieh and Turnbull, 1996).</td>
</tr>
<tr>
<td>kernel (Frequentist) kernel-based approach</td>
<td>(Zou et al., 1997).</td>
</tr>
<tr>
<td>BB (Bayesian bootstrap method)</td>
<td>(Gu et al., 2008).</td>
</tr>
<tr>
<td>dpm Nonparametric Bayesian approach based on a Dirichlet process mixture of normal distributions</td>
<td>(Erkanli et al., 2006).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariate-specific ROC curve</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp (Frequentist) parametric and semiparametric induced ROC regression approach</td>
<td>(Pepe, 1998; Faraggi, 2003).</td>
</tr>
<tr>
<td>kernel Nonparametric (kernel-based) induced ROC regression approach</td>
<td>(Rodríguez-Álvarez et al., 2011a).</td>
</tr>
<tr>
<td>bnp Nonparametric Bayesian model based on a single-weights dependent Dirichlet process mixture of normal distributions</td>
<td>(Inácio de Carvalho et al., 2013).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariate-adjusted ROC curve</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp (Frequentist) semiparametric method</td>
<td>(Janes and Pepe, 2009).</td>
</tr>
<tr>
<td>kernel Nonparametric (kernel-based) induced ROC regression approach</td>
<td>(Rodríguez-Álvarez et al., 2011a).</td>
</tr>
<tr>
<td>bnp Nonparametric Bayesian model based on a single-weights dependent Dirichlet process mixture of normal distributions and the Bayesian bootstrap</td>
<td>(Inácio de Carvalho and Rodríguez-Álvarez, 2018).</td>
</tr>
</tbody>
</table>

Table 1: Overview of ROC estimation methods included in the ROCnReg package.

Bayesian inference for the (pooled or marginal) ROC curve, the covariate-specific ROC curve, and the covariate-adjusted ROC curve. For the sake of generality, frequentist approaches are also implemented. Specifically, in what concerns estimation of the pooled ROC curve, ROCnReg implements the frequentist empirical estimator described in Hsieh and Turnbull (1996), the kernel-based approach proposed by Zou et al. (1997), the Bayesian Bootstrap method of Gu et al. (2008), and the Bayesian nonparametric method based on a Dirichlet process mixture of normal distributions model proposed by Erkanli et al. (2006). Regarding the covariate-specific ROC curve, ROCnReg implements the frequentist normal method of Faraggi (2003) and its semiparametric counterpart as described in Pepe (1998), the kernel-based approach of Rodríguez-Álvarez et al. (2011a), and the Bayesian nonparametric model based on a single-weights dependent Dirichlet process mixture of normal distributions proposed by Inácio de Carvalho et al. (2013). As for the covariate-adjusted ROC curve, the ROCnReg package allows estimation using the frequentist semiparametric approach of Janes and Pepe (2009), the frequentist nonparametric method discussed in Rodríguez-Álvarez et al. (2011a), and the recently proposed Bayesian nonparametric estimator of Inácio de Carvalho and Rodríguez-Álvarez (2018). Table 1 shows a summary of all methods implemented in the package. In addition, ROCnReg also provides functions to obtain ROC-based optimal thresholds to perform the classification/diagnosis of individuals as, say, diseased or nondiseased, using two different criteria, namely, the Youden index and the criterion that sets a target value for the false positive fraction. These are implemented for both the ROC curve, the covariate-specific, and the covariate-adjusted ROC curve.

The remainder of the paper is organized as follows. In Section Notation and definitions, we formally introduce the (pooled or marginal) ROC curve, the covariate-specific ROC curve, and the covariate-adjusted ROC curve. The description of the Bayesian estimation methods implemented in the ROCnReg package is given in Section Methods. In Section Package presentation and illustration, the usage of the main functions and capabilities of ROCnReg are described and illustrated using a synthetic dataset mimicking endocrine data. The paper concludes with a discussion in Section Summary and future plans.

## 2 Notation and definitions

This section sets out the formal definition of the pooled or marginal ROC curve, the covariate-specific ROC curve, and the covariate-adjusted ROC curve. It also describes the most commonly used summary measures of discriminatory accuracy, namely, the area under the ROC curve, the partial area under the ROC curve, and the Youden Index. For conciseness, we intentionally avoid giving too many details...
and refer the interested reader to Pepe (2003) (and references therein) for an extensive account of many aspects of ROC curves with and without covariates.

In what follows, we denote as $Y$ the outcome of the diagnostic test and as $D$ the binary variable indicating the presence ($D = 1$) or absence ($D = 0$) of disease. We also assume that along with $Y$ and the true disease status $D$, a covariate vector $X$ is also available and that it may encompass both continuous and categorical covariates. For ease of notation, the covariate vector $X$ is assumed to be the same in both the diseased ($D = 1$) and nondiseased ($D = 0$) populations, although this is not always necessarily the case in practice (e.g., disease stage is, obviously, a disease-specific covariate). By a slight abuse of notation, we use the subscripts $D$ and $\bar{D}$ to denote (random) quantities conditional on, respectively, $D = 1$ and $D = 0$. For example, $Y_D$ and $Y_{\bar{D}}$ denote the test outcomes in the diseased and nondiseased populations, respectively.

### Pooled ROC curve

In the case of a continuous-outcome diagnostic test, the classification is usually made by comparing the test result $Y$ against a threshold $c$. If the outcome is equal or above the threshold, $Y \geq c$, the subject will be diagnosed as diseased. On the other hand, if the test result is below the threshold, $Y < c$, he or she will be classified as nondiseased. The ROC curve is then defined as the set of all possible pairs of false positive fractions, $FPF(c) = Pr(Y \geq c | D = 0) = Pr(Y_{\bar{D}} \geq c)$, and true positive fractions, $TPF(c) = Pr(Y \geq c | D = 1) = Pr(Y_D \geq c)$, that can be obtained by varying the threshold value $c$, i.e.,

$$\{(FPF(c), TPF(c)) : c \in \mathbb{R}\}.$$  

It is common to represent the ROC curve as $\{(p, \text{ROCr}(p)) : p \in [0, 1]\}$, where

$$p = FPF(c) = 1 - F_D(c), \quad \text{ROCr}(p) = 1 - F_{\bar{D}} \left\{F_D^{-1}(1 - p)\right\},$$   (1)

with $F_D(y) = Pr(Y_D \leq y)$ and $F_{\bar{D}}(y) = Pr(Y_{\bar{D}} \leq y)$ denoting the cumulative distribution function (CDF) of $Y$ in the nondiseased and diseased groups, respectively. Several indices can be used as global summary measures of the accuracy of a test. The most widely used is the area under the ROC curve (AUC), defined as

$$\text{AUC} = \int_0^1 \text{ROCr}(p) \, dp.$$  (2)

In addition to its geometric definition, the AUC has also a probabilistic interpretation (see, e.g., Pepe, 2003, p. 78)

$$\text{AUC} = Pr(Y_D \geq Y_{\bar{D}}),$$  (3)

that is, the AUC is the probability that a randomly selected diseased subject has a higher test outcome than that of a randomly selected nondiseased subject. The AUC takes values between 0.5 and 1 in the case of an uninformative test that classifies individuals no better than chance, and 1.0 for a perfect test. We note that an AUC below 0.5 simply means that the classification rule should be reversed. As it is clear from its definition, the AUC integrates the ROC curve over the whole range of FPFs. However, depending on the clinical circumstances, interest might lie only on a relevant interval of FPFs or TPFs, which leads to the notion of the partial area under the ROC curve (pAUC). The pAUC over a range of FPFs $(0, u_1)$, where $u_1$ is typically low and represents the largest acceptable FPF, is defined as

$$\text{pAUC}(u_1) = \int_0^{u_1} \text{ROCr}(p) \, dp.$$  (4)

On the other hand, the pAUC over a range of TPFs $(v_1, 1)$, where $v_1$ is typically large and represents the lowest acceptable TPF, is defined as

$$\text{pAUC}_{TPF}(v_1) = \int_{v_1}^1 \text{ROCr}_TNF(p) \, dp,$$  (5)

where $\text{ROCr}_TNF$ is a $270^\circ$ rotation of the ROC curve, which can be expressed as

$$\text{ROCr}_TNF(p) = F_{\bar{D}} \left\{F_D^{-1}(1 - p)\right\}.$$  (6)

The curve in (6) is referred to as the true negative fraction (TNF) ROC curve since TNF $= 1 - FPF$ is plotted on the $y$-axis. We shall highlight that the argument $p$ in the ROC curve stands for a false positive fraction, whereas in the $\text{ROCr}_TNF$ curve, it stands for a true positive fraction. In Figure 1, we graphically illustrate the two partial areas.
Figure 1: (a) Shaded area in blue represents the partial area under the ROC curve over the interval $(0, u_1)$ of FPFs. (b) Shaded area in blue represents the partial area under the ROC curve over the interval $(v_1, 1)$ of TPFs. (c) The same as in (b) but now represented as an area under the true negative fraction ROC curve.

Another summary index of diagnostic accuracy is the Youden index (Shapiro, 1999; Youden, 1950)

$$YI = \max_c \{TPF(c) - FPF(c)\}$$

$$= \max_c \{F_D(c) - F_D(c)\}$$

$$= \max_p \{ROC(p) - p\}. \tag{9}$$

The YI ranges from 0 to 1, taking the value of 0 in the case of an uninformative test and 1 for a perfect test. As for the AUC, a YI below 0 means that the classification rule should be reversed. The value $c^*$, which maximizes Equation (7) (or, equivalently, Equation (8)), is frequently used in practice to classify subjects as diseased or nondiseased. It should be noted that the Youden index is equivalent to the Kolmogorov–Smirnov measure of distance between the distributions of $Y_D$ and $\bar{Y}_D$ (Pepe, 2003, p. 80).

**Covariate-specific ROC curve**

The conditional or covariate-specific ROC curve, given a covariate value $x$, is defined as

$$ROC(p \mid x) = 1 - F_D\{F_D^{-1}(1 - p) \mid x\}, \tag{10}$$

where $F_D(y \mid x) = Pr(Y_D \leq y \mid X_D = x)$ and $F_D(y \mid x) = Pr(Y_D \leq y \mid X_D = x)$ are the conditional CDFs of the test in the nondiseased and diseased groups, respectively. In this case, a number of possibly different ROC curves (and therefore discriminatory accuracies) may be obtained for different values of $x$. Thus, the covariate-specific ROC curve is an important tool that helps to understand and determine the optimal and suboptimal populations where to apply the tests on. That is, the covariate-specific ROC curve allows determining the populations, defined by or homogeneous with respect to $x$, where the diagnostic test has a ‘good’ or ‘poor’ discriminatory capacity. Similarly to the unconditional case, the covariate-specific TNF-ROC curve is given by

$$ROC_{TNF}(p \mid x) = F_D\{F_D^{-1}(1 - p) \mid x\}, \tag{11}$$
and the covariate-specific AUC, pAUC, and Youden index are

\[
AUC(x) = \int_0^1 \text{ROC}(p \mid x) dp,
\]
(12)

\[
pAUC(u_1 \mid x) = \int_0^{u_1} \text{ROC}(p \mid x) dp,
\]
(13)

\[
pAUC_{\text{TPF}}(v_1 \mid x) = \int_{v_1}^1 \text{ROC}_{\text{TNF}}(p \mid x) dp,
\]
(14)

\[
YI(x) = \max_c \left| \text{TPF}(c \mid x) - \text{FPF}(c \mid x) \right|
= \max_c |F_D(c \mid x) - F_D(c \mid x)|
(15)
= \max_p |\text{ROC}(p \mid x) - p|.
(16)

The value \( c^*_x \) that achieves the maximum in (15) (or (16)) is called the optimal covariate-specific YI threshold and can be used to classify a subject, with covariate value \( x \), as diseased or nondiseased.

**Covariate-adjusted ROC curve**

The covariate-specific ROC curve and associated AUC, pAUCs, and YI described in the previous section depict the accuracy of the test for specific covariate values. However, it would be undoubtedly useful to have a global summary measure that also takes covariate information into account. Such summary measure was developed by Janes and Pepe (2009), who proposed the covariate-adjusted ROC (AROC) curve, defined as

\[
\text{AROC}(p) = \int \text{ROC}(p \mid x) dH_D(x),
\]
(18)

where \( H_D(x) = \Pr(X_D \leq x) \) is the CDF of \( X_D \). That is, the AROC curve is a weighted average of covariate-specific ROC curves, weighted according to the distribution of the covariates in the diseased group. Equivalently, as shown by Janes and Pepe (2009), the AROC curve can also be expressed as

\[
\text{AROC}(p) = \Pr\{Y_D > F_D^{-1}(1 - p \mid X_D)\}
= \Pr\{1 - F_D(Y_D \mid X_D) \leq p\}.
(19)

As will be seen in Section Methods, Expression (19) is very convenient when it comes to estimating the AROC curve. Also, it emphasizes that the AROC curve at an FPF of \( p \) is the overall TPF when the thresholds used for defining a positive test result are covariate-specific and chosen to ensure that the FPF is \( p \) in each subpopulation defined by the covariate values.

In contrast to the pooled ROC curve (see Expressions (1) and (6)) and the covariate-specific ROC curve (see Expressions (10) and (11)), the AROC curve (and its 270° rotation) cannot be expressed in terms of the (conditional) CDFs of the test in each group. This does not, however, preclude the possibility of defining AROC-based summary accuracy measures, yet more care is needed. Thus, for the AROC curve, the area under the AROC, as well as the partial areas and YI, are expressed as follows

\[
\text{AAUC} = \int_0^1 \text{AROC}(p) dp,
\]
(20)

\[
p\text{AUC}(u_1) = \int_0^{u_1} \text{AROC}(p) dp,
\]
(21)

\[
p\text{AUC}_{\text{TPF}}(v_1) = \int_{\text{AROC}^{-1}(v_1)}^1 \text{AROC}(p) dp - \{1 - \text{AROC}^{-1}(v_1)\}v_1,
\]
(22)

\[
YI_{\text{AROC}} = \max_p \{\text{AROC}(p) - p\}.
(23)

Note, in particular, that the expressions for both the partial area under the AROC curve over a range of TPFs (see also Figure 1b) and for the YI are defined in terms of the AROC curve. For the YI, once the value that achieves the maximum in (23) is obtained, say \( p^* \), covariate-specific threshold values can be calculated as follows

\[
c^*_x = F_D^{-1}(1 - p^* \mid X_D = x).
\]

Note that, by construction, these threshold values will ensure that the FPF is \( p^* \) in each subpopulation defined by the covariate values. However, the TPF may vary with the covariate values, i.e.,

\[
\text{TPF}(c^*_x) = 1 - F_D(c^*_x \mid X_D = x).
\]
To finish this part, we mention that when the accuracy of a test is not affected by covariates, this does not necessarily mean that the covariate-specific ROC curve (which, in this case, is the same for all covariate values) coincides with the pooled ROC curve. It does coincide, however, with the AROC curve (see Janes and Pepe, 2009; Pardo-Fernández et al., 2014; Inácio de Carvalho and Rodríguez-Alvarez, 2018, for more details). As such, in all cases where covariates affect the test results, even though they might not affect its discriminatory capacity, inferences based on the pooled ROC curve might be misleading. In such cases, the AROC curve should be used instead. This also applies to the selection of (optimal) threshold values, which might be covariate-specific (i.e., possibly different for different covariate values).

3 Methods

For space reasons, we focus ourselves here on the Bayesian methods for ROC curve inference (with and without covariates) implemented in the ROCnReg package. A detailed description, as well as usage examples, of the frequentist approaches are available as Supplementary Material at https://bitbucket.org/mxrodriguez/rocnreg.

Pooled ROC curve

In what follows, let \( \{y_{D1}^{(n)}\}_{i=1}^{n} \) and \( \{y_{D2}^{(n)}\}_{j=1}^{m} \) be two independent random samples of test outcomes from the nondiseased and diseased groups of size \( n_D \) and \( n_D \), respectively.

Bayesian bootstrap based estimator

The function pooledROC.bb implements the Bayesian bootstrap (BB) approach proposed by Gu et al. (2008). Their estimator relies on the notion of placement value (Pepe, 2003, Chapter 5), which is simply the cumulative distribution function of \( y_{Dj} \) for \( j = 1, \ldots, n_D \). Specifically, \( U_D = 1 - F_D(y_D) \) is to be interpreted as a standardization of a diseased test outcome with respect to the distribution of test results in the nondiseased population. The ROC curve can be regarded as the CDF of \( U_D \)

\[
Pr(U_D \leq p) = Pr(1 - F_D(Y_D) \leq p) = 1 - F_D(F_D^{-1}(1 - p)) = ROC(p), \quad 0 \leq p \leq 1. \tag{24}
\]

The representation of the ROC given in (24) provides the rationale for the two-step algorithm of Gu et al. (2008), which can be described as follows. Let \( S \) be the number of iterations.

Step 1: Computation of the placement value based on the BB.

For \( s = 1, \ldots, S \), let

\[
U_{Dj}^{(s)} = \frac{n_D}{n} \sum_{i=1}^{n_D} \mathbb{I}(y_{Di} \geq y_{Dj}), \quad j = 1, \ldots, n_D,
\]

where \( (q_{11}^{(s)}, \ldots, q_{1n_D}^{(s)}) \sim \text{Dirichlet}(n_D, 1, \ldots, 1) \).

Step 2: Generate a realization of the ROC curve.

Based on (24), generate a realization of \( ROC^{(s)}(p) \), the cumulative distribution function of \( (U_{D1}^{(s)}, \ldots, U_{Dn_D}^{(s)}) \), where

\[
ROC^{(s)}(p) = \sum_{j=1}^{n_D} q_{2j}^{(s)} \mathbb{I}(U_{Dj}^{(s)} \leq p), \quad (q_{21}^{(s)}, \ldots, q_{2n_D}^{(s)}) \sim \text{Dirichlet}(n_D; 1, \ldots, 1).
\]

The BB estimate of the ROC curve is obtained by averaging over the ensemble of ROC curves \( \{ROC^{(1)}(p), \ldots, ROC^{(S)}(p)\} \), that is,

\[
\text{ROC}_{BB}^{BB}(p) = \frac{1}{S} \sum_{s=1}^{S} ROC^{(s)}(p),
\]

and a \((1 - a) \times 100\%\) pointwise credible band can be obtained from the \( a/2 \times 100\%\) and \((1 - a/2) \times 100\%\) percentiles of the same ensemble \((a \in (0,1))\). Note that these pointwise credible bands for the ROC curve are to be interpreted as credible intervals for the corresponding constituents TPFs.

The Bayesian bootstrap estimator leads to closed-form expressions for the AUC and pAUC, which
are, respectively, given by

\begin{align*}
\text{AUC}^{(s)}(p) &= \int_0^1 \text{ROC}^{(s)}(p) \, dp = 1 - \sum_{j=1}^{n_s} q_{2j}^{(s)} U_{Dj}^{(s)}, \\
p\text{AUC}^{(s)}(u) &= \int_0^{u_s} \text{ROC}^{(s)}(p) \, dp = u_s - \sum_{j=1}^{n_s} q_{2j}^{(s)} \min\{ u_s, U_{Dj}^{(s)} \}.
\end{align*}

It is easy to show that

\[ p\text{AUC}^{(s)}_{\text{TPF}}(v) = \int_{u_s}^{1} \text{ROC}^{(s)}_{\text{TPF}}(p) \, dp = \sum_{i=1}^{n_s} q_{i2i}^{(s)} \max\{ v_s, U_{Di}^{(s)} \} - v_s, \]

where

\[ U_{Di}^{(s)} = \sum_{j=1}^{n_s} q_{2ji}^{(s)} I \left( y_{Di} \geq y_{Dj} \right), \quad i = 1, \ldots, n_D, \]

and it is also easy to demonstrate that the ROC_{\text{TNF}} curve is the survival function of the placement value \( U_D = 1 - F_D(\gamma_D) \). With respect to the Youden index, it is obtained by maximising, over a grid of possible threshold values, the following expression

\[ \text{YI}^{(s)} = \max_c \left\{ F_D^{(s)}(\gamma_D) - F_D^{(s)}(c) \right\}. \]

As for the ROC curve, point estimates for the AUC, pAUC, pAUC_{\text{TPF}}, YI, and \( c^* \) can be obtained by averaging over the respective ensembles of \( S \) realizations, with credible bands derived from the percentiles of such ensembles.

**Dirichlet process mixture of normal distributions based estimator**

The Bayesian nonparametric approach, based on a Dirichlet process mixture (DPM) of normal distributions, for estimating the pooled ROC curve (Erkanli et al., 2000) is implemented in the pmledROC.dpm function. In this case, as implicit by the name, the CDFs of the test outcomes in each group are, respectively, given by

\begin{align*}
\Phi_D(\mu, \sigma^2) &= \int \Phi(y | \mu, \sigma^2) \, dG_D(\mu, \sigma^2), \\
G_D &\sim \text{DP}(\alpha_D, G_D(\mu, \sigma^2)), \quad \quad (25)
\end{align*}

where \( \Phi(y | \mu, \sigma^2) \) denotes the CDF of the normal distribution with mean \( \mu \) and variance \( \sigma^2 \) evaluated at \( y \). Here, \( G_D \sim \text{DP}(\alpha_D, G_D^*) \) is used to denote that the mixing distribution \( G_D \) follows a Dirichlet process (DP) (Ferguson, 1973) with centering distribution \( G_D^* \) for which \( E(G_D) = G_D^* \), and precision parameter \( \alpha_D \). Usually, due to conjugacy reasons, \( G_D(\mu, \sigma^2) = N(\mu \mid \mu_D, S_D)\Gamma(\sigma^2 \mid \sigma^2_D) \), and this is the centering distribution used by the pmledROC.dpm function. Note that here, \( S_D \) denotes the variance of the normal distribution, and \( \mu_D \) and \( \sigma^2_D \) are, respectively, the shape and rate parameters of the gamma distribution. All hyperparameter values are fixed.

For ease of posterior simulation and because it provides a highly accurate approximation, we make use of the truncated stick-breaking representation of the DP (Ishwaran and James, 2001), according to which \( G_D \) can be written as

\[ G_D(\cdot) = \sum_{l=1}^{L_D} \omega^{(l)} \delta_{(\mu_{Dl}, \sigma_{Dl}^{2})}(\cdot), \]

where \( (\mu_{Dl}, \sigma_{Dl}^{2}) \sim G_D^*(\mu, \sigma^2) \), for \( l = 1, \ldots, L_D \), and the weights follow the so-called (truncated) stick-breaking construction: \( \omega_{D1} = \nu_{D1}, \omega_{Dl} = \nu_{Dl} \prod_{l'<l} (1 - \nu_{Dl'}), \quad l = 2, \ldots, L_D \), and \( \nu_{D1}, \ldots, \nu_{DL_D-1} \sim \text{Beta}(1, \alpha_D) \). Further, one must set \( \nu_{DLD} = 1 \) in order to ensure that the weights add up to one. The CDF in (25) can therefore be written as

\[ F_D(y) = \sum_{l=1}^{L_D} \omega^{(l)} \Phi(y | \mu_{Dl}, \sigma_{Dl}^{2}), \]

where we shall note that \( L_D \) is not the exact number of components expected to be observed, but
rather an upper bound on it, as some of the components may be unoccupied. Some comments are in order regarding the specification of the hyperparameters’ values. In what concerns the centering distribution, \( m_{D0} \) represents the prior belief about the components’ means, and \( S_{D0} \) represents the confidence in such prior belief. Similarly, the values of \( a_D \) and \( b_D \) can be chosen to represent the prior belief about the components’ variance. Of course, when setting these parameters, it is crucial to consider the measurement scale of the data. By default, test outcomes are standardized (so that the resulting mean is zero and the variance is one) in the \( \text{pooledROC.dpm} \) function and the default values are as follows:

\[
m_{D0} = 0, \quad S_{D0} = 10, \quad a_D = 2, \quad b_D = 0.5.
\]

Because test outcomes are standardized, we expect the means of the components to be near zero and hence \( m_{D0} = 0 \). The parameter \( S_{D0} \) then controls where the drawn \( \mu_D \) can lie, and the value of 10 implies that approximately 95% of the values roughly lie within \(-6 \) and \( 6 \). Further, note that \( a_D = 2 \) leads to a prior with an infinite variance that is centered around a finite mean (\( b_D = 0.5 \)) and therefore favors variances less than one. Considering that the standardized data have a variance of one, it is reasonable to expect the within component variance to be smaller than the overall variance. The option of not standardizing the test outcomes is also available in \( \text{pooledROC.dpm} \), and in such a case, the defaults for the centering distribution hyperparameters’ values are as following:

\[
m_{D0} = \bar{y}_D, \quad S_{D0} = 100s_D^2/n_D, \quad a_D = 2, \quad b_D = s_D^2/2,
\]

with \( \bar{y}_D = \frac{1}{n_D} \sum_{j=1}^{n_D} y_{Dj} \) and \( s_D^2 = \frac{1}{n_D-1} \sum_{j=1}^{n_D} (y_{Dj} - \bar{y}_D)^2 \). Regarding the precision parameter of the DP, \( \alpha_D \), it has a direct relationship with the number of occupied mixture components. One possible strategy for specifying \( \alpha_D \) is to fix it to a small value to favor a small number of occupied components relative to the sample size. In the \( \text{pooledROC.dpm} \) function, we set \( \alpha_D = 1 \), a commonly used default value (Gelman et al., 2013, p. 553). Lastly, by default, \( L_D = 10 \). Before proceeding, we shall emphasize that these two configurations of hyperparameters values (for standardized and not standardized test outcomes) have proved to work well for a different range of test outcomes distributions, but it is certainly not our goal to encourage users to use it blindly and indeed thought should be dedicated to this important task. Nevertheless, output from the function \( \text{pooledROC.dpm} \) may be post-processed, and (informal) model fit diagnostics obtained; see more in Section Package presentation and illustration and in the Supplementary Materials.

Because the full conditional distributions for all model parameters are available in closed-form, posterior simulation can be easily conducted through Gibbs sampler (see the details, for instance, in Gelman et al., 2013). (We note that the default values for the centering distribution hyperparameters’ values are as following:

\[
m_{D0} = \bar{y}_D, \quad S_{D0} = 100s_D^2/n_D, \quad a_D = 2, \quad b_D = s_D^2/2,
\]

where the inversion is performed numerically. There is a closed-form expression for the AUC as

\[
AUC = \sum_{k=1}^{L_D} \sum_{l=1}^{L_D} \omega_D^{(s)}(\theta_{Dk}) \Phi \left( \frac{\theta_{Dl}^{(s)} - \mu_D^{(s)}}{\sqrt{\frac{1}{2} + \frac{n_D}{2}}} \right), \quad a_{kl}^{(s)} = \frac{p_{Dk}^{(s)} - p_{Dl}^{(s)}}{c_{Dl}^{(s)}} - \frac{p_{Dl}^{(s)}}{c_{Dl}^{(s)}} \phi \left( \frac{\theta_{Dl}^{(s)} - \mu_D^{(s)}}{\sqrt{\frac{1}{2} + \frac{n_D}{2}}} \right)
\]

Also, when \( L_D = L_D = 1 \), there are closed-form expressions for the pAUC and pAUC\(_{TPF} \) which are used in the package (see Hillis and Metz, 2012). For the pAUC/pAUC\(_{TPF} \), when \( L_D > 1 \) or \( L_D > 1 \), the integrals are approximated numerically using Simpson’s rule. The Youden index/optimal threshold is computed as in the Bayesian bootstrap method, with the obvious difference that here the CDFs are expressed as in (26). At the end of the sampling procedure, we have an ensemble of \( S \) ROC curves and AUCs/pAUCs/pAUC\(_{TPF} \)s/YIs/optimal thresholds, which, as before, allows obtaining point and interval estimates.

### Covariate-specific ROC curve

We now let \( \{(x_{Dij}, y_{Dij})\}_{i=1}^{n_{Dj}} \) and \( \{(x_{Di}, y_{Di})\}_{i=1}^{n_D} \) be two independent random samples of test outcomes and covariates from the nondiseased and diseased groups of size \( n_D \) and \( n_D \), respectively. Further, for all \( i = 1, \ldots, n_D \) and \( j = 1, \ldots, n_{Dj} \), let \( x_{Di} = (x_{Di1}, \ldots, x_{Diq})^\top \) and \( x_{Dij} = (x_{Dij1}, \ldots, x_{Dijq})^\top \) be \( q \)-dimensional vectors of covariates, which can be either continuous or categorical.

The function \( \text{cROC.bnp} \) implements the Bayesian nonparametric approach for conducting inference.
about the covariate-specific ROC curve of Inácio de Carvalho et al. (2013), which is based on a single-
weights dependent Dirichlet process mixture of normal distributions (De Iorio et al., 2009). Specifically,
under this method, the conditional CDF in the diseased group is modeled as follows

\[
F_D(y_{Dl} | x_{Dl}) = \int \Phi(y_{Dl} | \mu_D(x_{Dl}, \beta), \sigma^2)dG_D(\beta, \sigma^2),
\]

with the conditional CDF in the nondiseased group \( \bar{F}_D \) (Geisser and Eddy, 1979), and the widely applicable information criterion (WAIC) (Gelman et al., 2014). In turn, for the location of the interior knots themselves, we follow Rosenberg (1995) and

\[
\mu_D(x_{Dl}, \beta_{Dl}) = \beta_{Dl0} + f_{Dl1}(x_{Dl,1}) + \ldots + f_{Dlq}(x_{Dl,q}), \quad l = 1, \ldots, L_D,
\]

where the smooth functions, \( f_{Dlm} (m = 1, \ldots, q) \), are approximated using a linear combination of cubic B-splines basis functions. To avoid notational burden, we have assumed that all \( q \) covariates are continuous and modeled in a flexible way. However, the function \( \text{cROC.bnp} \) can also deal with categorical covariates, linear effects of continuous covariates, as well as interactions. For the reasons mentioned before, we recommend that all continuous covariates are modeled as in (27). Nonetheless, posterior predictive checks, as illustrated in Section Package presentation and illustration, can also be used to informally validate the fitted model. We write

\[
\mu_D(x_{Dl}, \beta_{Dl}) = z_{Dlj}^\top \beta_{Dlj}, \quad l = 1, \ldots, L_D, \quad j = 1, \ldots, n_D,
\]

where \( z_{Dlj} \) is the jth row of the design matrix that contains the intercept, the continuous covariates that are modeled in a linear way (if any), the cubic B-splines basis representation for those modeled in a flexible way, the categorial covariates (if any), and their interaction(s) (if believed to exist). Also, \( \beta_{Dlj} \) collects, for the lth component, the regression coefficients associated with the aforementioned covariates. For the covariate effects modeled using cubic B-splines, an important issue is the selection of the number and location of the knots at which to anchor the basis functions, as this has the potential to impact inferences, more so for the former than the latter. The selection of the number of knots can be assisted by a model selection criterion, for example, (the adaptation to the case of mixture models of) the deviance information criterion (DIC) (Celeux et al., 2006), the log pseudo marginal likelihood (LPML) (Geisser and Eddy, 1979), and the widely applicable information criterion (WAIC) (Gelman et al., 2014). In turn, for the location of the interior knots themselves, we follow Rosenberg (1995) and use the quantiles of the covariate values.

The regression coefficients and variances associated with each of the \( L_D \) components are sampled from the conjugate centering distribution \((\beta_{Dlj}, \sigma^2_{Dlj}^{-1})^\text{iid} \sim N_Q_0(m_{Dlj}, S_{Dlj})\Gamma(a_{Dlj}, b_{Dlj})\), with conjugate hyperpriors \( m_{Dlj} \sim N(m_{Dlj}, S_{Dlj}) \), and \( S_{Dlj}^{-1} \sim \text{Wishart}(v_{Dlj}, v_{Dlj}^{-1}) \) (a Wishart distribution with degrees of freedom \( v_{Dlj} \) and expectation \( Y_{lj}^{-1} \)), and where \( Q_{lj} \) is the dimension of the vector \( z_{Dlj} \). Hyperparameters \( m_{Dlj} \) and \( Y_{lj} \) must be chosen to represent the prior belief about the regression coefficients associated to each mixture component and about their covariance matrix, respectively, whereas \( S_{lj0} \) and \( v_{lj} \) are chosen to represent the confidence in the prior belief of \( m_{lj0} \) and \( Y_{lj} \), respectively. As in the no-covariate case, by default, in \( \text{cROC.bnp} \), test outcomes and covariates are standardized, which not only facilitates specification of the hyperparameter values but also improves the mixing of the Markov Chain Monte Carlo (MCMC) sampler.
When test outcomes and covariates are not standardized, the defaults are as follows
\[ m_{D0} = 0_{Q_D}, \quad S_{D0} = 10I_{Q_D}, \quad v_D = Q_D + 2, \quad \Psi_D = I_{Q_D}, \quad a_D = 2, \quad b_D = 0.5. \]
When test outcomes and covariates are not standardized, the defaults are the following
\[ m_{D0} = \bar{\beta}_D, \quad S_{D0} = \bar{\Sigma}_D, \quad v_D = Q_D + 2, \quad \Psi_D = 30\bar{\Sigma}_D, \quad a_D = 2, \quad b_D = \bar{\sigma}_D^2/2, \]
where \( \bar{\beta}_D \) and \( \bar{\Sigma}_D \) are the least squares estimates from fitting the linear model \( y_{D_i} = z_{D_i}\beta_D + \sigma_D\varepsilon_{D_i} \), where \( E(\varepsilon_{D_i}) = 0, \text{var}(\varepsilon_{D_i}) = 1, \) and \( \bar{\Sigma}_D \) is the estimated covariance matrix of \( \bar{\beta}_D \). With regard to the specification of \( a_D \) and \( L_D \), as in the DPM model (no-covariate case), we set them, respectively, to 1 and 10. The blocked Gibbs sampler is used to simulate draws from the posterior distribution, and details about it can be found, for instance, in the Supplementary Materials of Inácio de Carvalho et al. (2017).

Similarly to the analogous model for the no-covariate case, at iteration \( s \) of the Gibbs sampler procedure, the covariate-specific ROC curve is computed as
\[ \text{ROC}^{(s)}(p \mid x) = 1 - \frac{1}{\text{B}(1, s)} \left\{ \int_{0}^{1} \text{B}(1, s) (1 - p) \mid x \right\}, \quad s = 1, \ldots, S, \]
with
\[ F_D^{(s)}(y \mid x) = \sum_{l=1}^{L} \omega_D^{(s)} \Phi \left( y \mid z^\top \bar{\beta}_D^{(s)}, \sigma_D^{2(s)} \right), \quad F_D^{(s)}(y \mid x) = \sum_{k=1}^{L} \omega_D^{(s)} \Phi \left( y \mid z^\top \bar{\beta}_D^{(s)}, \sigma_D^{2(s)} \right), \]
and where the inversion is performed numerically. A point estimate for \( \text{ROC}(p \mid x) \) can be obtained by computing the mean of the ensemble \( \{ \text{ROC}^{(1)}(p \mid x), \ldots, \text{ROC}^{(S)}(p \mid x) \} \), with pointwise credible bands derived from the percentiles of the ensemble. Although the results presented in Erkanli et al. (2006) can be extended to derive a closed-form expression for the covariate-specific AUC, for computational reasons, in \text{ROCnReg}, the integral in (12) is approximated using Simpson’s rule, and the same applies for the partial areas. Conditionally on a specific covariate value, the computation of the Youden index and of the optimal threshold proceeds in a similar way as in the DPM model (see Inácio de Carvalho et al., 2017 for details). As for the covariate-specific ROC curve, point and interval estimates can be obtained from the corresponding covariate-specific ensemble of each summary measure.

We finish this section by noting that a particular case of the above estimator arises when the effect of all continuous covariates is assumed to be linear and only one component is considered, i.e.,
\[ F_D^{(s)}(y \mid x) = \Phi \left( y \mid \bar{x}^\top \bar{\beta}_D^{(s)}, \sigma_D^{2(s)} \right), \quad \text{and} \quad F_D^{(s)}(y \mid x) = \Phi \left( y \mid \bar{x}^\top \bar{\beta}_D^{(s)}, \sigma_D^{2(s)} \right), \]
with \( \bar{x}^\top = (1, x^\top) \). In this case, it is easy to show that
\[ \text{ROC}^{(s)}(p \mid x) = 1 - \Phi \left\{ a^{(s)}(x) + b^{(s)}\Phi^{-1}(1 - p) \right\}, \]
where
\[ a^{(s)}(x) = \bar{x}^\top \left( \frac{\bar{\beta}_D^{(s)} - \bar{\beta}_D^{(s)}}{\sigma_D^{2(s)}} \right), \quad \text{and} \quad b^{(s)} = \sigma_D^{2(s)} / \sigma_D^{2(s)}. \]
With this configuration, the model for the covariate-specific ROC curve can be regarded as a Bayesian counterpart of the induced ROC approach proposed by Faraggi (2003) (and detailed in the Supplementary Material). We denote it as the Bayesian normal linear model (for the test outcomes).

**Covariate-adjusted ROC curve**

The estimation of the AROC curve rests on the following three steps:

1. Estimation of the conditional distribution of test outcomes in the nondiseased group, \( F_D(y_{D_i} \mid x_{D_i}) \).
2. Computation of the placement value \( U_D = 1 - F_D(y_D \mid X_D) \), where, by a slight abuse of notation, we are designating it by the same letter used for the unconditional case.
3. Estimation of the cumulative distribution function of \( U_D \).

The approach proposed by Inácio de Carvalho and Rodríguez-Álvarez (2018) for estimating the AROC curve is implemented in function \text{AROC.bnp}, and it combines a single-weights dependent Dirichlet
process mixture of normal distributions in Step 1 and the Bayesian bootstrap in Step 3. Again, here, in Step 1, we also recommend using cubic B-splines transformations of all continuous covariates. Using the same notation as before, we model the conditional density as

\[ F_D(y_{Dj} | x_{Dj}) = \sum_{l=1}^{L_D} \omega_{Dlj} \Phi(y_{Dj} | z_{Dlj}^{T} \beta_{Dlj}, \sigma_{Dlj}^2). \]

The same prior distributions and default values as in the cROC.bnp function are adopted for \( \beta_{Dlj} \) and \( \sigma_{Dlj}^2 \). Once Step 1 has been completed, and given a posterior sample from the parameters of interest, the corresponding realization of the placement value of a diseased subject in the nondiseased population is easily computed as

\[ U_{Dlj}^{(s)} = 1 - F_D^{(s)}(y_{Dlj} | x_{Dlj}) = \sum_{j=1}^{n_D} \omega_{Dlj}^{(s)} \Phi(y_{Dlj} | z_{Dlj}^{T} \beta_{Dlj}^{(s)} , \sigma_{Dlj}^{2(s)}), \quad j = 1, \ldots, n_D, \quad s = 1, \ldots, S. \]

Finally, in Step 3, the cumulative distribution function of \( \{ U_{Dlj}^{(s)} \}_{j=1}^{n_D} \) is estimated through the Bayesian bootstrap

\[ \text{AROC}^{(s)}(p) = \frac{1}{n_D} \sum_{j=1}^{n_D} q_j^{(s)} I(U_{Dlj}^{(s)} \leq p), \quad (q_1^{(s)}, \ldots, q_{n_D}^{(s)}) \sim \text{Dirichlet}(n_D; 1, \ldots, 1). \]

As before, closed-form expressions do exist for the AAUC and pAAUC

\[ \text{AAUC}^{(s)} = \int_0^1 \text{AROC}^{(s)}(p) dp = 1 - \frac{1}{n_D} \sum_{j=1}^{n_D} q_j^{(s)} U_{Dlj}^{(s)}, \]

\[ \text{pAAUC}^{(s)}(u_1) = \int_0^{u_1} \text{AROC}^{(s)}(p) dp = u_1 - \frac{1}{n_D} \sum_{j=1}^{n_D} q_j^{(s)} \min \left\{ u_1, U_{Dlj}^{(s)} \right\}, \]

and the pAAUC-TNF (Equation (22)) is computed using numerical integration methods. With regards to the YI, it is obtained by directly plugging in AROC\(^{(s)}\)(p) in Expression (23).

A point estimate for AROC(p) can be obtained by computing the mean of the ensemble \( \{ \text{AROC}^{(1)}(p), \ldots, \text{AROC}^{(S)}(p) \} \), that is,

\[ \overline{\text{AROC}}(p) = \frac{1}{S} \sum_{s=1}^{S} \text{AROC}^{(s)}(p), \]

and the percentiles of the ensemble can be used to provide pointwise credible bands/credible intervals. The same applies for the AAUC, pAAUC, and YI.

### 4 Package presentation and illustration

This section describes the main functions in the ROCnReg package and illustrates their usage, due to confidentiality reasons, a synthetic dataset mimicking endocrine data from a cross-sectional study carried out by the Galician Endocrinology and Nutrition Foundation. A detailed description of the original dataset can be found in Tomé Martínez de Rituerto et al. (2009). The original data have also been previously analyzed in Rodríguez-Álvarez et al. (2011a,b) and Inácio de Carvalho and Rodríguez-Álvarez (2018). The synthetic data can be found in the ROCnReg package under the name endosyn, and a summary of it follows.

```r
R> library("ROCnReg")
R> data("endosyn")
R> summary(endosyn)
```

<table>
<thead>
<tr>
<th>cvd_idf</th>
<th>age</th>
<th>gender</th>
<th>bmi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>:0.0000</td>
<td>Min.</td>
<td>:18.25</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>:0.0000</td>
<td>1st Qu.</td>
<td>:29.57</td>
</tr>
<tr>
<td>Median</td>
<td>:0.0000</td>
<td>Median</td>
<td>:39.28</td>
</tr>
<tr>
<td>Mean</td>
<td>:0.2433</td>
<td>Mean</td>
<td>:41.43</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>:0.0000</td>
<td>3rd Qu.</td>
<td>:50.84</td>
</tr>
<tr>
<td>Max.</td>
<td>:1.0000</td>
<td>Max.</td>
<td>:84.66</td>
</tr>
</tbody>
</table>

The dataset is comprised of 2840 individuals (1317 men and 1523 women, variable gender), with an age range between 18 and 85 years old. Variable bmi contains the body mass index (BMI) values,
and cvd_idf is the variable that indicates the presence (1) or absence (0) of two or more cardiovascular disease (CVD) risk factors. Following previous studies, the CVD risk factors considered include raised triglycerides, reduced HD-cholesterol, raised blood pressure, and raised fasting plasma glucose. Note that from the 2840 individuals, about 24% present two or more CVD risk factors.

Using the ROCnReg package, in the subsequent sections, we will illustrate how to ascertain, through the pooled ROC curve, the discriminatory capacity of the BMI (which acts as our diagnostic test in this example) in differentiating individuals with two or more CVD risk factors (those belonging to the diseased class $D$) from those having none or just one CVD risk factor (and that therefore belong to the nondiseased group $\bar{D}$). We will also show how to evaluate, through the covariate-specific ROC curve, the possible modifying effect of age and gender on the discriminatory capacity of the BMI. Finally, the last part of this section focuses on the covariate-adjusted ROC curve as a global summary measure of the BMI discriminatory ability when taking the age and gender effects into account. In the Supplementary Material, we show the usage of the package for those methods not described here in the main text.

Pooled ROC curve

The ROCnReg package allows estimating the pooled ROC curve by means of the four methods listed in Table 1. Here, we only present the syntax for the functions pooledROC.BB and pooledROC.dpm that correspond, respectively, to the Bayesian bootstrap estimator and the approach based on a Dirichlet process mixture (of normal distributions). The function pooledROC.emp, which implements an empirical estimator, and the function pooledROC.kernel, which is based on kernel methods, are illustrated in the Supplementary Material. The input arguments in the functions are method-specific (details can be found in the manual accompanying the package), but in all cases, numerical and graphical summaries can be obtained by calling the functions print.pooledROC, summary.pooledROC, and plot.pooledROC, which can be abbreviated by print, summary, and plot. Recall that our aim is to ascertain, using the endosyn dataset, the discriminatory capacity of the BMI in differentiating individuals with two or more CVD risk factors from those having just one or none CVD risk factors.

```r
> set.seed(123, "L'Ecuier-CMRG") # for reproducibility
> pROC_dpm <- pooledROC.dpm(marker = "bmi", group = "cvd_idf", tag.h = 0,
+ data = endosyn, standardise = TRUE, p = seq(0, 1, l = 101), ci.level = 0.95,
+ compute.lpm1 = TRUE, compute.WAIC = TRUE, compute.DIC = TRUE,
+ pauc = pauccontrol(compute = TRUE, focus = "FPF", value = 0.1),
+ density = densitycontrol(compute = TRUE),
+ prior.h = priorcontrol.dpm(L = 10), prior.d = priorcontrol.dpm(L = 10),
+ mcmc = mcmccontrol(nsave = 8000, nburn = 2000, nskip = 1),
+ parallel = "snow", ncpus = 2, ci = NULL)
```

Before describing in detail the previous call, we first present the control functions that are used. In particular,

```
> pauccontrol(compute = FALSE, focus = c("FPF", "TPF"), value = 1)
```

can be used to indicate whether the pAUC should be computed (by default it is not computed), and in case it is computed (i.e., compute = TRUE), whether the focus should be placed on restricted FPFs (pAUC; see (4)) or on restricted TPFs (pAUC$_{TP}$; see (5)). In both cases, the upper bound $u_1$ (if focus is the FPF) or the lower bound $v_1$ (if focus is the TPF) should be indicated in the argument value. In addition to the pooled ROC curve, AUC, and pAUC (if required), the function pooledROC.dpm also allows computing the probability density function (PDF) of the test outcomes in both the diseased and nondiseased groups. In order to do so, we use

```
> densitycontrol(compute = FALSE, grid.h = NA, grid.d = NA)
```

By default, PDFs are not returned by the function pooledROC.dpm, but this can be changed by setting compute = TRUE, and through grid.h and grid.d, the user can specify a grid of test results where the PDFs are to be evaluated in, respectively, the nondiseased and diseased groups. Value NA signals auto initialization, with default a vector of length 200 in the range of the test results. Regarding the hyperparameters for the Dirichlet process mixture process of normals model (used for the estimation of the PDFs/CDFs of the test outcomes in each group), they can be controlled using

```
> priorcontrol.dpm(m0 = NA, s0 = NA, a = 2, b = NA, alpha = 1, L = 10)
```

A detailed description of these hyperparameters is found in Section Methods. Finally, to set the various parameters controlling the MCMC procedure (which in our case is simply a Gibbs sampler), we use

```
> mcmccontrol(nsave = 8000, nburn = 2000, nskip = 1)
```
Here, \( n_{\text{save}} \) is an integer value with the total number of scans to be saved, \( n_{\text{burn}} \) is the number of burn-in scans, and \( n_{\text{skip}} \) is the thinning interval. Unless due to memory usage reasons, we recommend not thinning and instead monitoring the effective sample size of the MCMC chain.

Coming back to the \texttt{pooledROC.dpm} function, through \texttt{marker}, the user specifies the name of the variable containing the test results. In our case, these are the values of the BMI. The name of the variable that distinguishes diseased (two or more CVD risk factors, \( D \)) from nondiseased individuals (none or one CVD risk factor, \( \bar{D} \)) is represented by the argument \texttt{group}, and the value coding nondiseased individuals in group is specified by \texttt{tag.h}. The \texttt{data} argument is a data frame containing the data and all needed variables. Setting \texttt{setting=TRUE} (the default) will standardize (i.e., subtract the mean and divide by the standard deviation) the test outcomes. The set of FPFs at which to estimate the pooled ROC curve is specified in the argument \texttt{p}, and argument \texttt{ci.level} allows specifying the level for the credible intervals (by default: 0.95). The LPML, WAIC, and DIC are computed by setting, respectively, the arguments \texttt{compute.lpml}, \texttt{compute.WAIC}, and \texttt{compute.DIC} to \texttt{TRUE}. Argument \texttt{pauc} is an (optional) list of values to replace the default values returned by the function \texttt{pauccontrol}. Here, we ask for the pAUC to be computed, with the focus on restricted FPFs and upper bound \( u_1 = 0.1 \). Similarly, the argument density is an (optional) list of values to replace the default values returned by the function \texttt{densitycontrol}, as it is the argument \texttt{mcmc}. Through \texttt{prior.h} and \texttt{prior.d} arguments, we specify the hyperparameters in the nondiseased and diseased groups, respectively. Again, both arguments are (optional) lists of values to replace the default values returned by the function \texttt{priorcontrol.dpm}. We shall remember that different hyperparameters' default values are set depending on whether test outcomes are standardized or not. Finally, arguments \texttt{parallel}, \texttt{ncpus} and \texttt{cl} allow performing parallel computations (based on the R-package \texttt{parallel}). In particular, through \texttt{parallel}, the user specifies the type of parallel operation: either "no" (default), "multicore" (not available on Microsoft Windows operating systems), or "snow". Argument \texttt{ncpus} is used to indicate the number of processes to be used in a parallel operation (when \texttt{parallel = "multicore"}, or \texttt{parallel = "snow"}), and \texttt{cl} is an optional parallel or snow cluster to be used when \texttt{parallel = "snow"}. If \texttt{cl} is not supplied (as in our example), a cluster on the local machine is created for the duration of the call.

A numerical summary of the fitted model can be obtained by calling the function \texttt{summary}, which provides, among other information, the estimated AUC (posterior mean) and 95\% credible interval (recall that we set in the call to the function \texttt{ci.level = 0.95}) and, if required, the LPML, WAIC, and DIC, separately, in the nondiseased (denoted here as \texttt{Group H}) and diseased (\texttt{Group D}) groups.

\begin{verbatim}
R> summary(pROC_dpm)
Call:
pooledROC.dpm(marker = "bmi", group = "cvd_idf", tag.h = 0, data = endosyn,
standardise = TRUE, p = seq(0, 1, l = 101), ci.level = 0.95,
compute.lpml = TRUE, compute.WAIC = TRUE, compute.DIC = TRUE,
pauc = pauccontrol(compute = TRUE, focus = "FPF", value = 0.1),
density = densitycontrol(compute = TRUE), prior.h = priorcontrol.dpm(L = 10),
prior.d = priorcontrol.dpm(L = 10), mcmc = mcmccontrol(nsave = 8000,
nburn = 2000, nskip = 1), parallel = "snow", ncpus = 2, cl = NULL)

Approach: Pooled ROC curve - Bayesian DPM

Area under the pooled ROC curve: 0.759 (0.74, 0.777)*
Partial area under the pooled ROC curve (FPF = 0.1): 0.168 (0.139, 0.199)*
* Credible level: 0.95

Model selection criteria:

<table>
<thead>
<tr>
<th></th>
<th>Group H</th>
<th>Group D</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAIC</td>
<td>12490.485</td>
<td>4017.063</td>
</tr>
<tr>
<td>WAIC (Penalty)</td>
<td>8.431</td>
<td>5.468</td>
</tr>
<tr>
<td>LPML</td>
<td>-6245.247</td>
<td>-2008.541</td>
</tr>
<tr>
<td>DIC</td>
<td>12490.276</td>
<td>4016.920</td>
</tr>
<tr>
<td>DIC (Penalty)</td>
<td>8.326</td>
<td>5.396</td>
</tr>
</tbody>
</table>

Sample sizes:

<table>
<thead>
<tr>
<th></th>
<th>Group H</th>
<th>Group D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>2149</td>
<td>691</td>
</tr>
<tr>
<td>Number of missing data</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

To complement these numerical results, the \texttt{ROCnReg} package also provides graphical results that
can be used to further explore the fitted model. Specifically, the function `plot` depicts the estimated pooled ROC curve and AUC (posterior means), jointly with `ci.level` × 100% (pointwise) credible intervals (here 95%).

```r
R> plot(pROC_dpm, cex.main = 1.5, cex.lab = 1.5, cex.axis = 1.5, cex = 1.5)
```

The result of the above code is shown in Figure 2.

![Figure 2: Graphical results as provided by the `plot.pooledROC` function for an object of class `pooledROC.dpm`. Posterior mean and 95% pointwise credible band for the pooled ROC curve and corresponding posterior mean and 95% credible interval for the AUC.](image)

By means of `density = densitycontrol(compute = TRUE)` in the call to the function, the estimates of the PDFs of the BMI in both groups are to be returned. This information can be accessed through component `dens` in the object `pROC_dpm` (i.e., `pROC_dpm$dens`), which is a list with elements `h` and `d` associated with the nondiseased and diseased groups, respectively. Each of the two elements is itself another list of two components: (1) `grid`, a vector that contains the grid of test results at which the PDFs have been evaluated (estimated); and (2) `dens`, a matrix with the PDFs at each iteration of the MCMC procedure. We can use these results to plot, e.g., the posterior mean (and 95% pointwise credible bands) of the PDF of the BMI in the healthy and diseased populations (see Figure 3a obtained using the R package `ggplot2` by Wickham, 2016). As can be observed, the estimated densities obtained under the DPM method follow very closely the histograms of the data. Further, the estimated densities available in `dens` can be used, as advised by Gelman et al. (2013, p. 553), to monitor convergence of the MCMC chains. The well-known label switching problem often leads to poor mixing of the chains of the component-specific parameters, but this may not impact convergence and mixing of the induced density/distribution of interest. For instance, Figure 4 shows the trace plots of the MCMC iterations (after burn-in) of the PDFs of the BMI in the two groups for different (and randomly selected) values of the BMI, and Figure 5 depicts the corresponding effective sample sizes and Geweke statistics (obtained using the R package `coda` by Plummer et al., 2006). Note that all plots give evidence of a good mixing and do not suggest a lack of convergence. For conciseness, the R code for reproducing Figures 3a, 4, and 5 is not provided here but in the replication code that accompanies the paper.

It is worth noting that the function `pooledROC.dpm` also allows fitting a normal distribution in each group. This is just a particular case (for which $L_D = L_D = 1$) of the more general DPM model. In order to fit such model, one simply needs to set $L = 1$ in the `prior.d` and `prior.h` arguments. The code follows.

```r
R> set.seed(123, "L'Ecuuyer-CMRG") # for reproducibility
R> pROC_normal <- pooledROC.dpm(marker = "bmi", group = "cvd_idf", tag.h = 0,
+ data = endosyn, standardise = TRUE, p = seq(0, 1, l = 101), ci.level = 0.95,
+ compute.lpm1 = TRUE, compute.WAIC = TRUE, compute.DIC = TRUE,
+ pauc = pauccontrol(compute = TRUE, focus = "FPF", value = 0.1),
+ density = densitycontrol(compute = TRUE),
```

The R Journal Vol. 13/1, June 2021 ISSN 2073-4859
Figure 3: Histogram of the (observed) BMI and posterior mean jointly along with 95% pointwise credible bands (red lines) of the PDF of the BMI obtained using (a) a Dirichlet process mixture of normals model (object \texttt{pROC\_dpm}); and (b) a normal model (object \texttt{pROC\_normal}). Left: Nondiseased individuals (none or one CVD risk factor). Right: Diseased individuals (two or more CVD risk factors).

\begin{verbatim}
+ prior.h = priorcontrol.dpm(L = 1), prior.d = priorcontrol.dpm(L = 1),
+ mcmc = mcmccontrol(nsave = 8000, nburn = 2000, nskip = 1),
+ parallel = "snow", ncpus = 2)
\end{verbatim}

For the sake of space, we omit from the summary the call to the function

\begin{verbatim}
R> summary(pROC_normal)
\end{verbatim}

\begin{verbatim}
Call: [...]

Approach: Pooled ROC curve - Bayesian DPM

Area under the pooled ROC curve: 0.748 (0.728, 0.768)*
Partial area under the pooled ROC curve (FPF = 0.1): 0.224 (0.194, 0.253)*
* Credible level: 0.95

Model selection criteria:

<table>
<thead>
<tr>
<th>Group</th>
<th>WAIC</th>
<th>WAIC (Penalty)</th>
<th>LPML</th>
<th>DIC</th>
<th>DIC (Penalty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>12639.952</td>
<td>2.431</td>
<td>-6319.976</td>
<td>12639.505</td>
<td>1.986</td>
</tr>
<tr>
<td>D</td>
<td>4049.004</td>
<td>2.267</td>
<td>-2024.502</td>
<td>4048.714</td>
<td>1.987</td>
</tr>
</tbody>
</table>

Sample sizes:

\end{verbatim}
Figure 4: Trace plots of the MCMC draws (after burn-in) of the PDFs of the BMI based on the model \texttt{pROC\_dpm}. Results are shown separately for the nondiseased and diseased populations and for different values of the BMI.

The fit of the DPM and normal models in each group can be compared on the basis of the WAIC, DIC, and/or the LPML. Remember that for the LPML, the higher its value, the better the model fit, while for the WAIC and DIC, it is the other way around. By comparing these values, provided in the summary of each fitted model, we can conclude that the three criteria favor, in both the diseased and (especially in the) nondiseased groups, the more general DPM model. This is also corroborated by the plot of the fitted densities in each group shown in Figure 3b.

We now estimate the pooled ROC curve using the Bayesian bootstrap estimator (function \texttt{pooledROC\_BB}), and comparisons with the results obtained using the DPM approach are provided.

```r
R> set.seed(123, "L\'Ecuyer-CMRG") # for reproducibility
R> pROC\_BB <- pooledROC\_BB(marker = "bmi", group = "cvd\_idf", tag.h = 0, data = endosyn, + p = seq(0, 1, l = 101), pauc = pauccontrol(compute = TRUE, focus = "TPF", value = 0.8), + B = 5000, ci.level = 0.95, parallel = "snow", ncpus = 2)
R> summary(pROC\_BB)

Call: ...

Approach: Pooled ROC curve - Bayesian bootstrap

Area under the pooled ROC curve: 0.76 (0.74, 0.779)*
Partial area under the pooled ROC curve (FPF = 0.1): 0.17 (0.14, 0.201)*
  * Credible level: 0.95

Sample sizes:

<table>
<thead>
<tr>
<th></th>
<th>Group H</th>
<th>Group D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>2149</td>
<td>691</td>
</tr>
<tr>
<td>Number of missing data</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note that the posterior means for the AUC and pAUC obtained using the DPM method (0.759 and
0.168, respectively) and the Bayesian bootstrap approach (0.760 and 0.170) are almost identical. This can also be observed when plotting the estimated ROC curves under the two methods (Figure 6).

We finish this section by showing how to use `ROCnReg` to obtain an (optimal) threshold value which could be further used to ‘diagnose’ an individual as diseased (two or more CVD risk factors) or healthy/nondiseased (none or only one CVD risk factor). To that aim, and for pooledROC objects (i.e., those obtained using functions `pooledROC.dpm`, `pooledROC.BB`, `pooledROC.emp`, and `pooledROC.kernel`), we use the function `compute.threshold.pooledROC`, which allows obtaining (optimal) threshold values using two criteria: the YI and the one that sets a target value for the FPF. For illustration, we show here the results using the YI criterion.

```R
R> th_pROC_dmp <- compute.threshold.pooledROC(pROC_dpm, criterion = "YI", + ci.level = 0.95, parallel = "snow", ncpus = 2)
R> th_pROC_dmp

$call
compute.threshold.pooledROC(object = pROC_dpm, criterion = "YI", + ci.level = 0.95, parallel = "snow", ncpus = 2)

$thresholds
   est   ql   qh
26.46877 26.07129 26.85029

$YI
   est   ql   qh
0.4045776 0.3721684 0.4366298

$FPF
```

Figure 5: Effective sample size and Geweke statistic of the MCMC chains (after burn-in) and of the PDFs of the BMI based on model `pROC_dpm` in the nondiseased and diseased populations. In both cases, results are shown along BMI values.
Figure 6: Estimated ROC curve using the Bayesian bootstrap approach (in black) and the DPM method (in red). Solid lines represent the posterior means and dashed lines the 95% pointwise credible bands.

<table>
<thead>
<tr>
<th>TPF est</th>
<th>ql</th>
<th>qh</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7854112</td>
<td>0.7528575</td>
<td>0.8161865</td>
</tr>
</tbody>
</table>

The function returns the posterior mean (est) and ci.level×100% (here 95% since ci.level = 0.95) credible interval (lower bound: ql, upper bound: qh) for the TPF associated with this cutoff value. For our example, the (posterior mean of the) TPF is 0.79.

Covariate-specific ROC curve

We now turn our attention to the inclusion of covariates in ROC analysis. As shown in Table 1, with ROCnReg, the user can estimate the covariate-specific ROC curve by means of three approaches. As for the functions in ROCnReg for estimating the pooled ROC curve, the input arguments are method-specific, and we refer the reader to the manual for details. For all methods, numerical and graphical summaries are obtained using functions print.cROC, summary.cROC, and plot.cROC. Here, we describe how to use the function cROC.bnp that implements the Bayesian nonparametric approach for estimating the covariate-specific ROC curve detailed in Section Methods. Also, for objects of this class, ROCnReg provides the function predictive.checks, which implements tools for assessing model fit via posterior predictive checks.

Recall that, when including covariate information in ROC analysis, interest resides in evaluating if and how the discriminatory capacity of the test varies with such covariates. In particular, in our endocrine study, we aim at evaluating the possible effect of both age and gender in the discriminatory capacity of the BMI. In what follows, with this aim in mind, two different models are fitted using the function cROC.bnp. One which considers a normal distribution in each group and that incorporates the age effect in a linear way and a second one which caps the maximum number of mixture components in each group at 10 (i.e., LD = L_D = 10) and that models the age effect using cubic B-splines (and thus allows for a nonlinear effect of age). Following Rodríguez-Álvarez et al. (2011a,b), both models consider the interaction between age and gender. For clarity, we first focus on the code that models the age effect in a linear way and use it to describe in detail the different arguments of the cROC.bnp function.
R> agep <- seq(22, 80, l = 30)
R> endopred <- data.frame(age = rep(agep,2), gender = factor(rep(c("Women", "Men"),
+ each = length(agep))))
R> set.seed(123, "L’Ecuyer-CMRG") # for reproducibility
R> cROC_bp <- cROC.bnp(formula.h = bmi ~ gender*age, formula.d = bmi ~ gender*age,
+ group = "cvd_idf", tag.h = 0, data = endosyn, newdata = endopred,
+ standardise = TRUE, p = seq(0, 1, l = 101), ci.level = 0.95, compute.lpm = TRUE,
+ compute.WAIC = TRUE, compute.DIC = TRUE, pauc = pauccontrol(compute = FALSE),
+ prior.h = priorcontrol.bnp(L = 1), prior.d = priorcontrol.bnp(L = 1),
+ density = densitycontrol(compute = TRUE),
+ mcmc = mcmccontrol(nsave = 8000, nburn = 2000, nskip = 1),
+ parallel = "snow", ncpus = 2)

As can be seen, many arguments coincide with those of the function pooledROC.dpm (described
in the previous section). We thus focus here on those that are specific to cROC.bnp. The arguments
formula.h and formula.d are formula objects specifying the model for the regression function (see
Equation (28)) in, respectively, the nondiseased and diseased groups. They are similar to the formula
used with the glm function, except that nonlinear functions (modeled by means of cubic B-splines) can
be added using function $f$ (an example will follow later in this section). Note that in both cases, the
left-hand side of the formulas should include the name of the test/marker (in our case bmi). In our
application, and for both groups, the model for the component’s means includes, in addition to the
linear effect of age and gender, the (linear) interaction between these two covariates (i.e., gender*age
≡ gender + age + gender:age). Through the newdata argument, the user can specify a new data
frame containing the values of the covariates at which the covariate-specific ROC curve and AUC (and
also pAUC and PDFs, if required) are to be computed. Finally, prior.h (the same holds for prior.d) is
a (optional) list of values to replace the defaults returned by priorcontrol.bnp, which allows setting
the hyperparameters for the single-weights dependent Dirichlet process mixture of normals model
(see Section Methods and the manual accompanying the package for more details)

priorcontrol.bnp(m0 = NA, S0 = NA, nu = NA, Psi = NA, a = 2, b = NA,
alpha = 1, L = 10)

In our example, we only modified the upper bound for the number of components in the mixture
model, which by default is 10, and set it to 1.

In this case, the summary of the fitted model provides the following information.
R> summary(cROC_bp)

Call: [...]

Approach: Conditional ROC curve - Bayesian nonparametric
-------------------------------------------------------------

Parametric coefficients
Group H:

<table>
<thead>
<tr>
<th></th>
<th>Post. mean</th>
<th>Post. quantile 2.5%</th>
<th>Post. quantile 97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>26.1459</td>
<td>25.8765</td>
<td>26.4096</td>
</tr>
<tr>
<td>genderWomen</td>
<td>-0.9160</td>
<td>-1.2726</td>
<td>-0.5680</td>
</tr>
<tr>
<td>age</td>
<td>1.1949</td>
<td>0.9180</td>
<td>1.4690</td>
</tr>
<tr>
<td>genderWomen:age</td>
<td>1.1948</td>
<td>0.8455</td>
<td>1.5394</td>
</tr>
</tbody>
</table>

Group D:

<table>
<thead>
<tr>
<th></th>
<th>Post. mean</th>
<th>Post. quantile 2.5%</th>
<th>Post. quantile 97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>29.1865</td>
<td>28.7625</td>
<td>29.6115</td>
</tr>
<tr>
<td>genderWomen</td>
<td>2.0826</td>
<td>1.3705</td>
<td>2.7665</td>
</tr>
<tr>
<td>age</td>
<td>0.6578</td>
<td>0.2162</td>
<td>1.0904</td>
</tr>
<tr>
<td>genderWomen:age</td>
<td>-0.7711</td>
<td>-1.4655</td>
<td>-0.0956</td>
</tr>
</tbody>
</table>

ROC curve:

<table>
<thead>
<tr>
<th></th>
<th>Post. mean</th>
<th>Post. quantile 2.5%</th>
<th>Post. quantile 97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.6959</td>
<td>-0.8177</td>
<td>-0.5776</td>
</tr>
<tr>
<td>genderWomen</td>
<td>-0.6863</td>
<td>-0.8695</td>
<td>-0.5946</td>
</tr>
<tr>
<td>age</td>
<td>0.1229</td>
<td>0.0045</td>
<td>0.2415</td>
</tr>
</tbody>
</table>
CONTRIBUTED RESEARCH ARTICLE

<table>
<thead>
<tr>
<th>genderWomen:age</th>
<th>0.4499</th>
<th>0.2745</th>
<th>0.6245</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>0.9391</td>
<td>0.8824</td>
<td>0.9975</td>
</tr>
</tbody>
</table>

Model selection criteria:

<table>
<thead>
<tr>
<th></th>
<th>Group H</th>
<th>Group D</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAIC</td>
<td>12174.986</td>
<td>4007.980</td>
</tr>
<tr>
<td>WAIC (Penalty)</td>
<td>6.283</td>
<td>5.646</td>
</tr>
<tr>
<td>LPML</td>
<td>-6087.492</td>
<td>-2003.990</td>
</tr>
<tr>
<td>DIC</td>
<td>12173.664</td>
<td>4007.329</td>
</tr>
<tr>
<td>DIC (Penalty)</td>
<td>4.994</td>
<td>5.053</td>
</tr>
</tbody>
</table>

Sample sizes:

<table>
<thead>
<tr>
<th></th>
<th>Group H</th>
<th>Group D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>2149</td>
<td>691</td>
</tr>
<tr>
<td>Number of missing data</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The first aspect to note is that, in this case, the summary function does not provide the estimated AUC as there is one (possibly different) AUC for each combination of covariate values. Also, given that: (1) only one component has been considered for modeling the CDFs of test results in the diseased and nondiseased groups, and (2) covariate effects have been modeled in a linear way, the summary function provides the posterior mean (and quantiles) of the (parametric) coefficients associated with the regression functions (Equation (30)) and with the covariate-specific ROC curve (Equation (32)). We note that since in the call to the function we have specified standardise = TRUE (and consequently both the test outcomes and covariates are standardized), the regression coefficients are on the scale of the standardized covariates.

If we focus on the coefficients for the covariate-specific ROC curve, it seems that the discriminatory capacity of the BMI decreases with age, with the decrease being more pronounced in women (note that the expression of the covariate-specific ROC curve in Equation (31) implies that positive coefficients correspond to a decrease in discriminatory capacity). These results are possibly better judged by plotting the estimated covariate-specific ROC curves and associated AUCs. This can be done using the plot function. For the covariate-specific ROC curve, the depicted graphics will depend on the number and nature of the covariates included in the analyses. In particular, for our application, we obtain, separately for men and women, the covariate-specific ROC curves (and AUCs) along age. These are shown in Figure 7, obtained using the code

```R
R> op <- par(mfrow = c(2,2))
R> plot(cROC_sp, ask = FALSE)
R> par(op)
```

Although in this example we have modeled the age effect linearly and only one mixture component was considered, ROCnReg also allows for modeling the effect of continuous covariates in a nonlinear way, either using cubic B-spline basis expansions (through the function cROC_cnp) or kernel-based smoothers (via the function cROC_kernel which is described in the Supplementary Material). Also, as noted before, using only one mixture component for the single-weights dependent Dirichlet process mixture of normals model (function cROC_cnp) is equivalent to considering a (Bayesian) normal model, which might be too restrictive for most data applications. In what follows, we provide more flexibility to the model for the covariate-specific ROC curve by means of (1) increasing the number of mixture components and (2) modeling the age effect in a nonlinear way (recall our considerations in Section Covariate-specific ROC curve about the lack of flexibility of the single-weights dependent Dirichlet process mixture of normals model when covariates effects on the components' means are modeled linearly). The former is done by modifying the value of L in the arguments prior.h and prior.d, with 10 being the default value. Regarding the latter, this is done by making use of the function f when specifying the component's mean functions through formula.h and formula.d. In particular, in our application we are interested in modeling the factor-by-curve interaction between age and gender (i.e., we model the age effect 'separately' for men and women). This is done using, e.g., `bmi ~ gender + f(age,by = gender,K = c(3,5))`. Through argument K, we indicate the number of internal knots used for constructing the cubic B-spline basis used to approximate the nonlinear effect of age (with the quantities of age used to anchor the knots). Note that we can specify a different number of internal knots for men and women (K = c(3,5)), where the order of vector K should match the ordering of the levels of the factor gender. We also note that to assist in the selection of the number of interior knots (in ROCnReg, the location is always based on the quantiles of the corresponding covariates), the user can make use of the WAIC, DIC, and/or LPML. For instance, for this application, we fitted different models with a different number of internal knots, and we have chosen the model that provided the lowest WAIC (this was done in both the nondiseased and diseased populations, and we remark that...
Figure 7: Graphical results as provided by the plot.cROC function for an object of class cROC.bnp. Results for the model that includes the linear interaction between age and gender and one mixture component. Top row: Posterior mean of the covariate-specific ROC curve along age, separately for men and women. Bottom row: Posterior mean and 95% pointwise credible band for the covariate-specific AUC along age, separately for men and women.

the number of knots does not need to be the same in the two populations). The final model is shown below.

```r
R> # Levels of gender, and its ordering.
R> # Needed if we want to specify different
R> # number of knots for men and women
R> levels(endosyn$gender)

[1] "Men"  "Women"
```

```r
R> set.seed(123, "L'Ecuyer-CMRG") # for reproducibility
R> cROC_bnp <- cROC.bnp(
+  formula.h = bmi ~ gender + f(age, by = gender, K = c(0,0)),
+  formula.d = bmi ~ gender + f(age, by = gender, K = c(4,4)),
+  group = "cvd_idf", tag.h = 0, data = endosyn, newdata = endopred,
+  standardise = TRUE, p = seq(0, 1, l = 101), ci.level = 0.95, compute.lpml = TRUE,
+  compute.WAIC = TRUE, compute.DIC = TRUE, pauc = pauccontrol(compute = FALSE),
+  prior.h = priorcontrol.bnp(L = 10), prior.d = priorcontrol.bnp(L = 10),
+  density = densitycontrol(compute = TRUE),
+  mcmc = mcmccontrol(nsave = 8000, nburn = 2000, nskip = 1),
)
+ parallel = "snow", ncpus = 2)
R> summary(cROC_bnp)

Call: [...] 

Approach: Conditional ROC curve - Bayesian nonparametric

Model selection criteria:

<table>
<thead>
<tr>
<th>Group</th>
<th>WAIC</th>
<th>WAIC (Penalty)</th>
<th>LPML</th>
<th>DIC</th>
<th>DIC (Penalty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group H</td>
<td>11833.000</td>
<td>31.236</td>
<td>-5916.766</td>
<td>11829.750</td>
<td>29.611</td>
</tr>
<tr>
<td>Group D</td>
<td>3909.828</td>
<td>38.583</td>
<td>-1955.449</td>
<td>3904.532</td>
<td>35.934</td>
</tr>
</tbody>
</table>

Sample sizes:

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of observations</th>
<th>Number of missing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group H</td>
<td>2149</td>
<td>0</td>
</tr>
<tr>
<td>Group D</td>
<td>691</td>
<td>0</td>
</tr>
</tbody>
</table>

R> op <- par(mfrow = c(2,2))
R> plot(cROC_sp, ask = FALSE)
R> par(op)

The graphical results are shown in Figure 8. Note that, especially for women, age displays a marked nonlinear effect. Recall that for objects of class cROCR.bnp, and if required in the call to the function, the summary function provides, separately for the diseased and nondiseased/healthy groups, the WAIC, LPML, and DIC. Note that, in both cases, the three criteria support the use of the more flexible model that uses cubic B-splines and 10 mixture components for modeling the distribution of the BMI (model cROCR_bnp) over the more restrictive Bayesian normal linear model (model cROCR_bp). Because the WAIC, LPML, and DIC are relative criteria, posterior predictive checks are also available in ROCnReg through the function predictive.checks. Specifically, the function generates replicated datasets from the posterior predictive distribution in the two groups $D$ and $\bar{D}$ and compares them to the test values (BMI values in our application) using specific statistics. For the choice of such statistics, we follow Gabry et al. (2019), who suggest choosing statistics that are ‘orthogonal’ to the model parameters. Since we are using a location-scale mixture of normals model for the test outcomes, we use the skewness and kurtosis here and check how well the posterior predictive distribution captures these two quantities.

R> op <- par(mfrow = c(2,3))
R> pc_cROC_bp <- predictive.checks(cROC_bp, + statistics = c("kurtosis", "skewness"), devnew = FALSE)
R> par(op)
R> pc_cROC_bnp <- predictive.checks(cROC_bnp, + statistics = c("kurtosis", "skewness"), devnew = FALSE)
R> par(op)

Results are shown in Figure 9. As can be seen, the model that includes the factor-by-curve interaction between age and gender and 10 mixture components performs quite well in capturing both quantities, while the Bayesian normal linear model fails to do so. Also shown in Figure 9 (and provided by function predictive.checks) are the kernel density estimates of 500 randomly selected datasets drawn from the posterior predictive distribution, in each group, compared to the kernel density estimate of the observed BMI (in each group). Again, the more flexible model, as opposed to the Bayesian normal linear model, is able to simulate data that are very much similar to the observed BMI values. As for the pooled ROC curve, ROCnReg also provides a function that allows obtaining (optimal) threshold values for the covariate-specific ROC curve. For illustration, instead of the threshold values based on the Youden index, we now use the criterion that sets a target value for the FPF. The code for model cROCR_bnp, when setting the FPF = 0.3, is as follows.

R> th_fpf_cROC_bnp <- compute.threshold.cROC(cROCR_bnp, criterion = "FPF", FPF = 0.3, + newdata = endopred, ci.level = 0.95, parallel = "snow", ncpus = 2)
Figure 8: Graphical results as provided by the `plot.cROC` function for an object of class `cROC.bnp`. Results for the model that includes the factor-by-curve interaction between age and gender and 10 mixture components. Top row: Posterior mean of the covariate-specific ROC curve along age, separately for men and women. Bottom row: Posterior mean and 95% pointwise credible band for the covariate-specific AUC along age, separately for men and women.

R> names(th_fpf_cROC_bnp)
[1] "newdata" "thresholds" "TPF" "FPF" "call"

In addition to the data frame `newdata` containing the covariate values at which the thresholds are computed, the function `compute.threshold.cROC` also returns the covariate-specific thresholds corresponding to the specified FPF as well as the covariate-specific TPF attached to these thresholds. In both cases, the function returns the posterior mean and the `ci.level` × 100% (here 95%) pointwise credible intervals. Although `ROCnReg` does not provide a function for plotting the results obtained using `compute.threshold.cROC`, graphical results can be easily obtained. For simplicity, we only show here the code for the covariate-specific threshold values (thresholds), but a similar code can be used to plot the covariate-specific TPFs. Both plots are shown in Figure 10. As can be observed, for an FPF of 0.3, the BMI age-specific thresholds tend to increase with age both for men and women, although for the latter, there is a slight decrease after the age of about 70 years old. The age-specific TPFs corresponding to the thresholds for which the FPF is 0.3 show a nonlinear behavior, and these are in general higher for women than for men (of the same age).

R> df <- data.frame(age = th_fpf_cROC_bnp$newdata$age, gender = th_fpf_cROC_bnP$newdata$gender, y = th_fpf_cROC_bnp$thresholds[[1]]$"est", + ql = th_fpf_cROC_bnp$thresholds[[1]]$"ql"),
Figure 9: Graphical results as provided by the predictive.checks function for an object of class cROC.bnp. Histograms of the statistics skewness and kurtosis computed from 8000 draws from the posterior predictive distribution in the diseased and nondiseased groups. The red line is the estimated statistic from the observed BMI values. The right-hand side plots show the kernel density estimate of the observed BMI (solid black line), jointly with the kernel density estimates for 500 simulated datasets drawn from the posterior predictive distributions.

(a) Model including the linear interaction between age and gender and one component

(b) Model including the factor-by-curve interaction between age and gender and 10 mixture components

For conciseness, we have not shown here how to perform convergence diagnostics of the MCMC
chains for models fitted using the function cROC.bnp. In very much the same way as shown in the previous section for the object pROC_dpm, using the information contained in component dens in the list of returned values (if required), one can produce trace plots of the conditional densities at some sampled values, as well as obtain the corresponding effective sample sizes and Geweke statistics. Some results are provided in the Supplementary Material, and the associated code can be found in the replication code that accompanies this paper.

Covariate-adjusted ROC curve

In this section, we illustrate how to conduct inference about the covariate-adjusted ROC curve using ROCnReg. Similar to the covariate-specific ROC curve, three approaches are available for estimating the AROC curve. The function AROC.bnp is illustrated below, while AROC.sp and AROC.kernel are exemplified in the Supplementary Material.

Recall that the AROC curve is a global summary measure of diagnostic accuracy that takes covariate information into account. In the context of our endocrine application, we seek to study the overall discriminatory capacity of the BMI for detecting the presence of CVD risk factors when adjusting for age and gender. Here, we focus on how to estimate the AROC curve using the AROC.bnp function. The function syntax is exactly similar to the one of cROC.bnp, with the only difference being that we only need to specify the arguments related to the nondiseased population. The code and respective summary follow.

```R
set.seed(123, "L’Ecuyer-CMRG") # for reproducibility
```
CONTRIBUTED RESEARCH ARTICLE

\[
\text{AROC}_{\text{bnp}} <- \text{AROC.bnp}( \\
+ \text{formula.h} = \text{bmi} - \text{gender} + f(\text{age}, \text{by} = \text{gender}, K = c(0,0)) \\
+ \text{group} = "\text{cvd.idf}", \text{tag.h} = 0, \text{data} = \text{endosyn, standardise} = \text{TRUE}, \\
+ \text{p} = \text{seq}(0, 1, l = 101), \text{ci.level} = 0.95, \text{compute.lpm} = \text{TRUE}, \text{compute.WAIC} = \text{TRUE}, \\
+ \text{compute.DIC} = \text{TRUE}, \text{pauc} = \text{pauccontrol(compute} = \text{FALSE}), \\
+ \text{prior.h} = \text{priorcontrol.bnp(L} = 10), \text{density} = \text{densitycontrol(compute} = \text{TRUE}), \\
+ \text{mcmc} = \text{mcmccontrol(nsav}e = 8000, \text{nburn} = 2000, \text{nskip} = 1), \\
+ \text{parallel} = "\text{snow}, \text{n} \text{cpus} = 2)
\]

\[
\text{R> summary(AROC_bnp)}
\]

\[
\text{Call: [...]}
\]

\[
\text{Approach: AROC Bayesian nonparametric}
\]

\[
\text{Area under the covariate-adjusted ROC curve: 0.656 (0.629, 0.684)*}
\]

* Credible level: 0.95

\[
\text{Model selection criteria:}
\]

\[
\begin{align*}
\text{Group H} & \\
\text{WAIC} & 11833.000 \\
\text{WAIC (Penalty)} & 31.236 \\
\text{LPML} & -5916.766 \\
\text{DIC} & 11829.750 \\
\text{DIC (Penalty)} & 29.611
\end{align*}
\]

\[
\text{Sample sizes:}
\]

\[
\begin{array}{ll}
\text{Number of observations} & \text{Group H} \quad \text{Group D} \\
\text{Number of missing data} & 0 \quad 0
\end{array}
\]

The area under the AROC curve is 0.656 (95% credible interval: (0.629, 0.684)) thus revealing a reasonable good ability of the BMI to detect the presence of CVD risk factors when teasing out the age and gender effects. As for the pooled ROC curve and the covariate-specific ROC curve, a plot function is also available (result in Figure 11a).

\[
\text{R> plot(AROC_bnp, cex.main} = 1.5, \text{cex.lab} = 1.5, \text{cex.axis} = 1.5, \text{cex} = 1.3)
\]

Finally, we compare the AROC curve with the pooled ROC curve that was obtained earlier by using a DPM model with 10 components in each group. In Figure 11b, we show the plots of the two curves, and, as can be noticed, the pooled ROC curve lies well above the AROC curve, thus evidencing the need for incorporating covariate information into the analysis.

\[
\text{R> plot(AROC_bnp$p, AROC_bnp$ROC[,1], type} = "1", \text{xlim} = c(0,1), \text{ylim} = c(0,1), \\
+ \text{xlab} = "\text{FPF}, ylab} = "\text{TPF", main} = "\text{Pooled ROC curve vs AROC curve", cex.main} = 1.5, \\
+ \text{cex.lab} = 1.5, \text{cex.axis} = 1.5, \text{cex} = 1.5
\]

Computational aspects

We finish this section with some comments on computational aspects. In our experience, the methods with the largest computing times are those implemented in `cROC.bnp` when \(L = 1\) and in `cROC.kernel` when confidence bands are to be constructed. In the first case, the main reason behind the computational burden is the need to invert \(F_0(\cdot | x)\) in order to obtain the covariate-specific ROC curve (see equation (10)). Note that when \(L > 1\), the conditional distribution function in the nondiseased group is given by a mixture of normal distributions, and the corresponding quantile function needs to be computed for each covariate(s) value we might be interested in and for each iteration of the Gibbs sampler procedure. Regarding the `cROC.kernel` function, the computing time is mainly driven by the number of bootstrap samples used for constructing the confidence bands. In Table 2, we show the
time, in seconds, needed for fitting the pooled, the covariate-specific, and the covariate-adjusted ROC curve using the Bayesian nonparametric and the kernel approaches for the synthetic endocrine data and when both parallel (with 2 and 4 processes) and no parallel options are used. We note that for the Bayesian approaches, we also computed the densities/conditional densities, as well as the WAIC, LPML, and DIC, which further increase the computing time (in the case of the AROC curve, these were only computed in the nondiseased population). With respect to the kernel-based approach (in this case, the fit is done separately for men and women and the corresponding results are presented in the Supplementary Material), we have used 500 bootstrap samples to construct the confidence bands. As it can be appreciated, for these two intense tasks, using 4 processes drastically improves the computation time. All computations were performed in a iMac with 3.6GHz quad core Intel i7 processor and 32GB RAM running under a macOS Catalina 10.15.5 operating system.

<table>
<thead>
<tr>
<th></th>
<th>No parallel</th>
<th>Snow (2 cores)</th>
<th>Snow (4 cores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pooledROC.dpm</td>
<td>138</td>
<td>118</td>
<td>111</td>
</tr>
<tr>
<td>pooledROC.kernel</td>
<td>376</td>
<td>196</td>
<td>105</td>
</tr>
<tr>
<td>cROC.bnp</td>
<td>2052</td>
<td>1117</td>
<td>680</td>
</tr>
<tr>
<td>cROC.kernel</td>
<td>Men: 1159</td>
<td>Men: 528</td>
<td>Men: 279</td>
</tr>
<tr>
<td></td>
<td>Women: 1885</td>
<td>Women: 916</td>
<td>Women: 466</td>
</tr>
<tr>
<td>AROC.bnp</td>
<td>126</td>
<td>115</td>
<td>112</td>
</tr>
<tr>
<td>AROC.kernel</td>
<td>Men: 847</td>
<td>Men: 404</td>
<td>Men: 214</td>
</tr>
<tr>
<td></td>
<td>Women: 1707</td>
<td>Women: 833</td>
<td>Women: 438</td>
</tr>
</tbody>
</table>

Table 2: Time in seconds (rounded to the nearest second) needed to fit the pooled, the covariate-specific, and the covariate-adjusted ROC curve for the Bayesian nonparametric and kernel approaches.

5 Summary and future plans

In this paper, we have introduced the capabilities of the R package ROCnReg for conducting inference about the pooled ROC curve, the covariate-specific ROC curve, and the covariate-adjusted ROC curve and their associated summary indices. As we have illustrated, the current version of the package provides several options for estimating ROC curves, both under frequentist and Bayesian paradigms, either parametrically, semiparametrically, or nonparametrically. To the best of our knowledge, this is the first software package implementing Bayesian inference for ROC curves. Several additions/extensions are planned in the future, and these, among others, include:
• Implement the most time-consuming parts in C or C++.
• Incorporate methods for non-binary disease status (e.g., no disease, mild disease, severe disease). That is, implement ROC surface models.
• Implement new (optimal) threshold criteria (e.g., YI including costs).

6 Computational details

The results in this paper were obtained using R 4.0.3 with the ROCnReg 1.0-5 package. The ROCnReg package has several dependencies: graphics, grDevices, parallel, splines, stats, moments (Komsta and Novomestky, 2015), nor1mix (Maechler, 2019), Matrix (Bates and Maechler, 2019), spatstat (Baddeley and Turner, 2005), np (Hayfield and Racine, 2008), lattice (Sarkar, 2008), MASS (Venables and Ripley, 2002), and pbivnorm (Genz and Kenkel, 2015). R itself and all packages used are available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/.

7 Acknowledgements

We acknowledge the reviewer for their constructive comments that led to an improved version of the article. MX Rodríguez-Álvarez was funded by project MTM2017-82379-R (AEI/FEDER, UE), by the Basque Government through the BERC 2018-2021 program and Elkartek project 3KIA (KK-2020/00049) and by the Spanish Ministry of Science, Innovation, and Universities (BCAM Severo Ochoa accreditation SEV-2017-0718).

Bibliography


