Received: 6 August 2021 | Revised: 5 April 2022 | Accepted: 6 May 2016

DOI: 10.1002/sim.9480

RESEARCH ARTICLE

Statistics in Medicine WILEY

Bayesian nonparametric inference for the overlap coefficient: With an application to disease diagnosis

Vanda Inácio | **Javier E. Garrido Guillén**

School of Mathematics, University of Edinburgh, Edinburgh, UK

Correspondence
Vanda Inácio, School of Mathematics, University of Edinburgh, The King's Buildings, JCMB, EH9 3FD Edinburgh, UK.
Email: vanda.inacio@ed.ac.uk

Funding information
Fundação para a Ciência e Tecnologia, Grant/Award Numbers: PTDC/MAT-STA/28649/2017, UID/MAT/00006/2019

Diagnostic tests play an important role in medical research and clinical practice. The ultimate goal of a diagnostic test is to distinguish between diseased and nondiseased individuals and before a test is routinely used in practice, it is a pivotal requirement that its ability to discriminate between these two states is thoroughly assessed. The overlap coefficient, which is defined as the proportion of overlap area between two probability density functions, has gained popularity as a summary measure of diagnostic accuracy. We propose two Bayesian nonparametric estimators, based on Dirichlet process mixtures, for estimating the overlap coefficient. We further introduce the covariate-specific overlap coefficient and develop a Bayesian nonparametric approach based on Dirichlet process mixtures of additive normal models for estimating it. A simulation study is conducted to assess the empirical performance of our proposed estimators. Two illustrations are provided: one concerned with the search for biomarkers of ovarian cancer and another one aimed to assess the age-specific accuracy of glucose as a biomarker of diabetes.

KEYWORDS

Bayesian nonparametrics, covariate-adjustment, diagnostic test, Dirichlet process mixtures, overlap coefficient

1 | INTRODUCTION

The assessment of the accuracy of a diagnostic test is an important task in many subfields of medicine and in clinical research. The first and most fundamental task before a test is routinely used in practice is to evaluate its ability to discriminate diseased from nondiseased individuals or, more generally, to distinguish between different disease stages. The most popular summary measures of diagnostic accuracy are the area under the receiver operating characteristic curve (AUC) or the Youden index (YI); these and other summary indices can be found, among many other sources, in Pepe.[1(ch.4)] Recently, the overlap coefficient (OVL), defined as the proportion of area between two density functions, and first introduced by Weitzman[2] in a comparison of income distributions by race, has been proposed as an alternative summary measure of diagnostic accuracy (see, for instance, Samawi et al,[3] Wang and Tian,[4] and Franco-Pereira et al[5]). The coefficient of overlap can directly measure how similar/different are the distributions of the test outcomes in the diseased and nondiseased populations and therefore it can serve as a summary index of the diagnostic accuracy that is sensitive to any differences between such two distributions. Further background about the overlap coefficient is provided in Section 2. Besides its increased use in diagnostic medicine, the overlap coefficient is more broadly quite a popular measure in medical data analysis.[6-9] It also finds applications in ecology,[10,11] in economics,[12] and in psychology.[13]

An advantage of the OVL over the AUC and the YI, besides its intuitive interpretation, is that in addition of taking into account both the location and shape of the distributions of test outcomes in the two populations, it is "non-directional." Being non-directional means that there is no need to assume that larger test outcomes are more indicative of the presence of disease or vice versa. Of course, one can easily change the classification rule behind the AUC/YI but, as we shall see in one of the illustrations provided in Section 6, when we have a large number of candidate tests, it is handy to have a measure that does not need to rely on any classification rule. The OVL is also well-tailored for dealing with bimodal/multimodal distributions, which are common in gene expression data.[4,8] For instance, when both low and high test results are associated with nondisease, with test results in the diseased group lying in between the two modes of the nondiseased test outcomes' distribution, such that there is a perfect separation between the two populations, the AUC would take the value 0.5, thus falsely implying that the test classifies individuals no better than chance.[14] On the other hand, the coefficient of overlap, because it is not based on any classification rule, would correctly recognize that test outcomes of both populations are perfectly separated. At this point, it is fair to mention the work of Martinez-Camblor et al[15] and of de Carvalho et al,[16] who proposed, respectively, a generalization of the receiver operating characteristic curve and an affinity based measure of diagnostic accuracy that also provide meaningful values in such a context. Summing up, the coefficient of overlap is a versatile measure that (i) has an appealing graphical interpretation, and (ii) does not rely on any classification rule. We shall highlight that we do not foresee the use of the overlap coefficient as a replacement of the existing measures of diagnostic accuracy but rather as a companion and/or alternative to these.

In this work, we first develop two Bayesian nonparametric estimators for the coefficient of overlap. At the core of both estimators it is a Dirichlet process mixture of normal distributions, thus leading to two highly flexible estimators that can adapt to intricate distributional features, such as multimodality, skewness, and/or extreme variability, without the need to know them in advance. Hence, our proposed estimators can be used for a wide variety of populations and for a large number of diseases and continuous diagnostic measures. Compared to the existing kernel approaches,[12,17,18] our estimators do not depend on a smoothing parameter, the choice of which is a nontrivial issue in practice and may have a great impact on inference. Further, because we are working under the Bayesian paradigm, point and interval estimates for the overlap coefficient are obtained in a single integrated framework.

Second, we introduce the covariate-specific overlap coefficient, as it is now widely recognized that the performance of a test may be affected by covariates (such as age and gender) and when this is the case, ignoring the information provided by covariates may lead to erroneous inferences about a test's accuracy. The covariate-specific overlap coefficient will help determining the optimal and suboptimal populations, as determined by the covariates values, in which to perform the tests on and for its estimation we propose an estimator based on a Dirichlet process mixture of additive normal models. As in the no-covariate case, the resulting estimator is highly flexible and can be applied in many contexts. It is worth noting that the fact that the overlap coefficient is non-directional is of especially imortance here in the covariate case as the same monotone order between the diseased and nondiseased groups may not apply across all covariates levels.

The remainder of the article is organized as follows. In the next section we introduce background material about the coefficient of overlap. Section 3 presents our proposed Bayesian nonparametric estimators for the coefficient of overlap, while the covariate-specific overlap coefficient and its corresponding estimator are presented in Section 4. The performance of our (unconditional and conditional) estimators is assessed in Section 5 using simulated data. In Section 6, our methods are illustrated using (i) a dataset concerned with the search for biomarkers of ovarian cancer, and (ii) using data from a population based survey where the goal is to infer about the age-specific accuracy of the glucose as a biomarker of diabetes. Concluding remarks are offered in Section 7.

2 | PRELIMINARIES

Throughout we will be assuming that regardless of a single or multiple biomarkers being measured on each individual, a univariate continuous outcome, denoted as Y, is ultimately used for diagnosing purposes. Let D be the binary variable indicating the presence ($D = 1$) or absence ($D = 0$) of disease. We further use the subscripts D and $\overline{D}$ to denote quantities conditional on $D = 1$ and $D = 0$, respectively. For example, Y_D and $Y_{\overline{D}}$ denote the test outcomes in the diseased and nondiseased populations, with probability density functions given by f_D and $f_{\overline{D}}$, respectively. We further assume that a so-called gold standard test, that perfectly classifies all individuals as diseased or nondiseased, is available. Compared to the gold standard test, we want to evaluate how the candidate test, which is possibly less invasive and/or costly, performs.

The overlap coefficient is mathematically defined as

$$\mathrm{OVL} = \int_{-\infty}^{\infty} \min\{f_{\bar{D}}(y), f_D(y)\}\mathrm{d}y, \tag{1}$$

and it can be interpreted as a measure of agreement between the distributions of test outcomes in the diseased and nondiseased populations, taking values between zero and one. Specifically, an overlap of zero means that test outcomes in the two populations are completely separated (perfect diagnostic accuracy), whereas a value of one means that the distributions are identical and thus the test is useless from a diagnostic viewpoint. Values between zero and one, correspond to different degrees of diagnostic accuracy, and the closer to zero the coefficient of overlap is, the better the diagnostic accuracy. If desirable, to make the OVL values more comparable to those of the AUC and YI, one can work instead with 1 − OVL, so that larger values imply better diagnostic accuracy. As it was demonstrated by Schmid and Schimdt,[18] the overlap coefficient is invariant with respect to strictly increasing and differentiable transformations of Y_D and $Y_{\bar{D}}$. The overlap coefficient can be rewritten in various other forms. As it was pointed out by Schmid and Schimdt,[18] using the following equality

$$\min\{f_{\bar{D}}(y), f_D(y)\} = \frac{1}{2}\left\{f_{\bar{D}}(y) + f_D(y)\right\} - \frac{1}{2}\left|f_{\bar{D}}(y) - f_D(y)\right|,$$

where the role of $f_{\bar{D}}$ and f_D is obviously interchangeable, leads to the following expression

$$\mathrm{OVL} = 1 - \frac{1}{2}\int_{-\infty}^{+\infty} \left|f_{\bar{D}}(y) - f_D(y)\right| \mathrm{d}y. \tag{2}$$

As it was shown by the same authors, the overlap coefficient in (1) can also be rewritten as

$$\begin{aligned}
\mathrm{OVL} &= \int_{-\infty}^{\infty} \min\{f_{\bar{D}}(y), f_D(y)\}\mathrm{d}y \\
&= \int_{-\infty}^{\infty} I\{f_{\bar{D}}(y) < f_D(y)\} f_{\bar{D}}(y)\mathrm{d}y + \int_{-\infty}^{\infty} I\{f_{\bar{D}}(y) \geq f_D(y)\} f_D(y)\mathrm{d}y \\
&= \mathbb{E}[I\{f_{\bar{D}}(Y_{\bar{D}}) < f_D(Y_{\bar{D}})\}] + \mathbb{E}[I\{f_{\bar{D}}(Y_D) \geq f_D(Y_D)\}] \\
&= \Pr\{f_{\bar{D}}(Y_{\bar{D}}) < f_D(Y_{\bar{D}})\} + \Pr\{f_{\bar{D}}(Y_D) \geq f_D(Y_D)\}.
\end{aligned} \tag{3}$$

As the authors have mentioned, Equation (3) emphasizes that the overlap coefficient can be written as the sum of two error probabilities. Indeed, $\Pr\{f_{\bar{D}}(Y_{\bar{D}}) < f_D(Y_{\bar{D}})\}$ is the probability of choosing f_D if $f_{\bar{D}}$ is the true density and $\Pr\{f_{\bar{D}}(Y_D) \geq f_D(Y_D)\}$ is the probability of choosing $f_{\bar{D}}$ if f_D is the true density.

3 | PROPOSED (UNCONDITIONAL) ESTIMATORS

The expressions in (1) to (3) suggest a two-step modeling procedure whose first step involves estimating the probability density functions of the test outcomes in each population. In the second step, for the estimators based on (1) and (2), the integral can be computed via numerical integration, while the estimator based on (3) requires estimating the outer probabilities. We found the performance of the estimator based on (2) to be superior or on par to that of the estimator based on (1) and therefore, hereafter, we focus only on the estimator arising from (2). A similar finding was reported by Schmid and Schimdt.[18] In the next subsections we detail the first estimation step, which is common to both estimators we propose, and the specifics of the second step for each estimator.

3.1 | Step 1: Modeling $f_{\bar{D}}$ and f_D

As both expressions (2) and (3) make clear, accurate estimation of the coefficient of overlap requires accurately estimating the densities of the test outcomes in each population. In what follows, let $\{y_{\bar{D}i}\}_{i=1}^{n_{\bar{D}}}$ and $\{y_{Dj}\}_{j=1}^{n_D}$ be two independent random samples of test outcomes of size $n_{\bar{D}}$ and n_D from the nondiseased and diseased populations, respectively, with

$$y_{\bar{D}1}, \ldots, y_{\bar{D}n_{\bar{D}}} | f_{\bar{D}} \overset{\text{iid}}{\sim} f_{\bar{D}}, \quad \text{and} \quad y_{D1}, \ldots, y_{Dn_D} | f_D \overset{\text{iid}}{\sim} f_D.$$

One possible and popular approach is to assume a normal distribution for both f_D and $f_{\bar{D}}$, possibly after some transformation of the Y_D and $Y_{\bar{D}}$ scales (eg, the logarithmic one). However, such model would be unsuitable for test outcomes data exhibiting asymmetry or multiple modes. Although the normal distribution could be extended (by using, for example, a skew normal distribution to account for asymmetry), mixtures of normal distributions can be used to represent a wide variety of density shapes and therefore they are our preferred choice here. In particular, Dirichlet process mixtures[19] of normal distributions have been shown to accurately approximate any smooth density on the real line.[20] Under such a model, the density function of test outcomes in the nondiseased population (the one in the diseased population follows analogously) can be written as

$$f_{\bar{D}}(y_{\bar{D}i}) = \int \phi(y_{\bar{D}i} | \mu, \sigma^2) \mathrm{d}G_{\bar{D}}(\mu, \sigma^2), \qquad G_{\bar{D}} \sim \mathrm{DP}(\alpha_{\bar{D}}, G^*_{\bar{D}}(\mu, \sigma^2)), \tag{4}$$

where $\phi(y | \mu, \sigma^2)$ stands for the probability density function of the normal distribution with mean μ, variance σ^2, and evaluated at y. The mixing distribution $G_{\bar{D}}$ follows a Dirichlet process[21] (DP) with centering distribution $E\{G_{\bar{D}}(\mu, \sigma^2)\} = G^*_{\bar{D}}(\mu, \sigma^2)$ and precision parameter $\alpha_{\bar{D}}$ (> 0). The constructive definition of the Dirichlet process,[22] unarguably its most popular representation, under which $G_{\bar{D}}$ can be written as an infinite sum of weighted point masses

$$G_{\bar{D}}(\cdot) = \sum_{l=1}^{\infty} \omega_{\bar{D}l} \delta_{(\mu_{\bar{D}l}, \sigma^2_{\bar{D}l})}(\cdot), \qquad \omega_{\bar{D}l} = \begin{cases} v_{\bar{D}1}, & \text{if } l = 1, \\ v_{\bar{D}l} \prod_{m<l}(1 - v_{\bar{D}m}), & \text{if } l \geq 2, \end{cases}$$
$$v_{\bar{D}l} \overset{\text{iid}}{\sim} \mathrm{Beta}(1, \alpha_{\bar{D}}), \quad (\mu_{\bar{D}l}, \sigma^2_{\bar{D}l}) \overset{\text{iid}}{\sim} G^*_{\bar{D}}(\mu, \sigma^2), \quad l \geq 1, \tag{5}$$

allows us to write the density in (4) as a countable mixture of normal densities

$$f_{\bar{D}}(y_{\bar{D}i}) = \sum_{l=1}^{\infty} \omega_{\bar{D}l} \phi(y_{\bar{D}i} | \mu_{\bar{D}l}, \sigma^2_{\bar{D}l}). \tag{6}$$

For the ease of posterior inference, a conjugate centering distribution is specified, that is, $G^*_{\bar{D}}(\mu, \sigma^2) \equiv \mathrm{N}(\mu | m_{\bar{D}0}, S_{\bar{D}0}) \Gamma(\sigma^{-2} | a_{\bar{D}}, b_{\bar{D}})$, where $S_{\bar{D}0}$ denotes the variance of the normal distribution and $\Gamma(a, b)$ denotes a gamma distribution with shape parameter a and rate parameter b. The precision parameter of the DP, $\alpha_{\bar{D}}$, has a direct relationship with the number of occupied mixture components and it can be either set to a fixed value or a prior distribution placed on it. Here we set $\alpha_{\bar{D}} = 1$, which is a commonly used default value and which favors a small number of occupied mixture components relative to the sample size.[23(p. 553)]

In order to facilitate posterior simulation, we make use of the truncated stick-breaking representation of the DP,[24] replacing $G_{\bar{D}}$ in (5) with $G^{L_{\bar{D}}}_{\bar{D}}(\cdot) = \sum_{l=1}^{L_{\bar{D}}} \omega_{Dl} \delta_{(\mu_{\bar{D}l}, \sigma^2_{\bar{D}l})}(\cdot)$, and hence the resulting model for the density of test outcomes can be written as

$$f_{\bar{D}}(y_{\bar{D}i}) = \sum_{l=1}^{L_{\bar{D}}} \omega_{\bar{D}l} \phi(y_{\bar{D}i} | \mu_{\bar{D}l}, \sigma^2_{\bar{D}l}), \tag{7}$$

where the weights result from a truncated version of the stick-breaking construction that in order to ensure that they add up to one, sets $v_{\bar{D}L_{\bar{D}}} = 1$, so that $\omega_{\bar{D}L_{\bar{D}}} = 1 - \sum_{l=1}^{L_{\bar{D}}-1} \omega_{\bar{D}l}$. We shall note that $L_{\bar{D}}$ is not the exact number of components expected to be observed but instead an upper bound on it, as some of the components may be unoccupied. Because the weights in the infinite stick-breaking representation in (5) decrease rapidly for typical choices of $\alpha_{\bar{D}}$ (eg, $\alpha_{\bar{D}} = 1$), the model in (7) is still an accurate approximation of (6) for small $L_{\bar{D}}$. Upon introduction of latent variables that identify the mixture component to which each nondiseased individual belongs to, the full conditional distributions for all model's parameters are available in closed form, thus allowing for ready posterior simulation through Gibbs sampling. The details of the Gibbs sampler scheme are provided in the Supplementary Materials.

3.2 | Step 2: Numerical integration/modeling the outer probabilities

Our first proposed estimator, henceforth denoted by $\mathrm{OVL}_{\mathrm{DPM}}$, is based directly on expression (2) and computes the integral numerically through the trapezoidal rule, using a grid, say $(y_0, y_1, \dots, y_N)$, of N equally spaced points that covers the range of test outcomes in the two populations. Specifically, once Step 1 is completed and posterior realizations of the density functions in the two populations are available, a posterior realization of the coefficient of overlap can be computed as

$$
\begin{aligned}
\mathrm{OVL}_{\mathrm{DPM}}^{(s)} &= 1 - \frac{1}{2}\int_{-\infty}^{+\infty} \left| f_{\bar{D}}^{(s)}(y) - f_{D}^{(s)}(y) \right| \mathrm{d}y \\
&\approx 1 - \frac{1}{2}\frac{\Delta y}{2}\left\{ \left| f_{\bar{D}}^{(s)}(y_0) - f_{D}^{(s)}(y_0) \right| + 2\sum_{r=1}^{N-1} \left| f_{\bar{D}}^{(s)}(y_r) - f_{D}^{(s)}(y_r) \right| + \left| f_{\bar{D}}^{(s)}(y_N) - f_{D}^{(s)}(y_N) \right| \right\}, \\
f_{\bar{D}}^{(s)}(y) &= \sum_{l=1}^{L_{\bar{D}}} \omega_{\bar{D}l}\,\phi\left(y \mid \mu_{\bar{D}l}^{(s)}, (\sigma_{\bar{D}l}^{(s)})^2\right), \qquad f_{D}^{(s)}(y) = \sum_{l'=1}^{L_{D}} \omega_{Dl'}\,\phi\left(y \mid \mu_{Dl'}^{(s)}, (\sigma_{Dl'}^{(s)})^2\right),
\end{aligned}
$$

for $s = 1, \dots, S$, where S denotes the number of posterior realizations after burn-in, and $\Delta y = y_1 - y_0 = y_2 - y_1 = \dots = y_N - y_{N-1}$ is the length of each subinterval. A point estimate of the overlap coefficient can be obtained by considering the mean or median of the ensemble $\left\{\mathrm{OVL}_{\mathrm{DPM}}^{(1)}, \dots, \mathrm{OVL}_{\mathrm{DPM}}^{(S)}\right\}$. A 95% credible interval can also be obtained from the 2.5% and 97.5% percentiles of the same ensemble. We shall finish this part mentioning that a similar estimator was proposed by Núñez-Antonio et al,[11] with the difference being that due to the authors' focus in circular data, a projected normal distribution, instead of a normal distribution as here, was used as the kernel of the Dirichlet process mixture model.

Our second proposed estimator requires modeling the following two probabilities: $\Pr\{f_{\bar{D}}(Y_{\bar{D}}) < f_D(Y_{\bar{D}})\}$ and $\Pr\{f_{\bar{D}}(Y_D) \geq f_D(Y_D)\}$. Let us consider the following two random variables

$$
U_{\bar{D}} = f_{\bar{D}}(Y_{\bar{D}}) - f_D(Y_{\bar{D}}), \quad \text{and} \quad U_D = f_{\bar{D}}(Y_D) - f_D(Y_D),
$$

with cumulative distribution functions given by $F_{U_{\bar{D}}}$ and F_{U_D}, respectively. The coefficient of overlap in (3) can therefore be rewritten as

$$
\mathrm{OVL} = F_{U_{\bar{D}}}(0) + 1 - F_{U_D}(0). \tag{8}
$$

For modeling purposes, we assign a DP prior to both $F_{U_{\bar{D}}}$ and F_{U_D}, that is,

$$
\begin{aligned}
u_{\bar{D}1}, \dots, u_{\bar{D}n_{\bar{D}}} \mid F_{U_{\bar{D}}} &\overset{\text{iid}}{\sim} F_{U_{\bar{D}}}, \qquad F_{U_{\bar{D}}} \sim \mathrm{DP}(\alpha_{U_{\bar{D}}}, F_{U_{\bar{D}}}^{*}), \\
u_{D1}, \dots, u_{Dn_D} \mid F_{U_D} &\overset{\text{iid}}{\sim} F_{U_D}, \qquad F_{U_D} \sim \mathrm{DP}(\alpha_{U_D}, F_{U_D}^{*}).
\end{aligned}
$$

Due to the conjugacy property of the DP,[21(theorem 1)] the resulting posterior distributions are again a DP, that is,

$$
F_{u_{\bar{D}}} \mid u_{\bar{D}1}, \dots, u_{\bar{D}n_{\bar{D}}} \sim \mathrm{DP}\left(\alpha_{U_{\bar{D}}} + n_{\bar{D}}, \frac{\alpha_{U_{\bar{D}}}}{\alpha_{U_{\bar{D}}} + n_{\bar{D}}} F_{U_{\bar{D}}}^{*} + \frac{n_{\bar{D}}}{\alpha_{U_{\bar{D}}} + n_{\bar{D}}} \widehat{F}_{U_{\bar{D}}}^{\mathrm{emp}}\right), \quad \widehat{F}_{U_{\bar{D}}}^{\mathrm{emp}}(u) = \frac{1}{n_{\bar{D}}}\sum_{i=1}^{n_{\bar{D}}} I\left(u_{\bar{D}i} \leq u\right),
$$

with an analogous expression holding for the posterior distribution of F_{U_D}. For the sake of computational simplicity, we consider the limiting case where the precision parameter $\alpha_{U_{\bar{D}}}$ (α_{U_D}) approaches zero and the resulting posterior distribution therefore simplifies to a DP with precision parameter equal to $n_{\bar{D}}$ and centering distribution given by $\widehat{F}_{U_{\bar{D}}}^{\mathrm{emp}}$. Note that in this case we do not even need to specify $F_{U_{\bar{D}}}^{*}$ ($F_{U_D}^{*}$). This limiting posterior distribution is also known as the Bayesian bootstrap,[25] whose samples correspond to discrete distributions supported at the observed data points with weights distributed according to a Dirichlet distribution.[23(p. 548)]

Once Step 1 has been completed, we can compute posterior realizations of the variables $U_{\overline{D}}$ and U_D as

$$u^{(s)}_{\overline{D}i} = f^{(s)}_{\overline{D}}(y_{\overline{D}i}) - f^{(s)}_{D}(y_{\overline{D}i}) = \sum_{l=1}^{L_{\overline{D}}} \omega^{(s)}_{\overline{D}l}\phi\left(y_{\overline{D}i}|\mu^{(s)}_{\overline{D}l},(\sigma^{(s)}_{\overline{D}l})^2\right) - \sum_{l'=1}^{L_D} \omega^{(s)}_{Dl'}\phi\left(y_{\overline{D}i}|\mu^{(s)}_{Dl'},(\sigma^{(s)}_{Dl'})^2\right), \quad i = 1, \dots , n_{\overline{D}},$$

$$u^{(s)}_{Dj} = f^{(s)}_{\overline{D}}(y_{Dj}) - f^{(s)}_{D}(y_{Dj}) = \sum_{l=1}^{L_{\overline{D}}} \omega^{(s)}_{\overline{D}l}\phi\left(y_{Dj}|\mu^{(s)}_{\overline{D}l},(\sigma^{(s)}_{\overline{D}l})^2\right) - \sum_{l'=1}^{L_D} \omega^{(s)}_{Dl'}\phi\left(y_{Dj}|\mu^{(s)}_{Dl'},(\sigma^{(s)}_{Dl'})^2\right), \quad j = 1, \dots , n_D,$$

for $s = 1, \dots , S$. Using the aforementioned Bayesian bootstrap to compute the cumulative distribution functions involved in (8), the corresponding realization of the overlap coefficient is given by

$$\begin{aligned}
\text{OVL}^{(s)}_{\text{DPM-BB}} &= F^{(s)}_{U_{\overline{D}}}(0) + 1 - F^{(s)}_{U_D}(0)\\
&= \sum_{i=1}^{n_{\overline{D}}} q^{(s)}_{\overline{D}i} I\left\{u^{(s)}_{\overline{D}i} < 0\right\} + \sum_{j=1}^{n_D} q^{(s)}_{Dj} I\left\{u^{(s)}_{Dj} \geq 0\right\},\\
&(q^{(s)}_{\overline{D}1}, \dots , q^{(s)}_{\overline{D}n_{\overline{D}}}) \sim \text{Dirichlet}(n_{\overline{D}}; 1, \dots , 1), \quad (q^{(s)}_{D1}, \dots , q^{(s)}_{Dn_D}) \sim \text{Dirichlet}(n_D; 1, \dots , 1).
\end{aligned} \tag{9}$$

Note that if there are ties in the test outcomes within each population, thus implying also ties in the realizations of $U_{\overline{D}}$ and/or U_D, then in such a case the parameter vector of the Dirichlet distribution should be adjusted accordingly by adding together the ones for each repeated test outcome. Although we are considering continuous test results, because measurements are made with finite precision, ties can occur. Finally, as for our first proposed estimator, a point estimate of the overlap coefficient can be obtained by considering the mean or median of the ensemble of posterior realizations, with a 95% credible interval obtained from the 2.5% and 97.5% percentiles of the same ensemble.

At last, it is fair to ask why do we use the Bayesian bootstrap to model $F_{U_{\overline{D}}}$ and F_{U_D} and not a Dirichlet process mixture with an appropriate kernel. The reason is simply computational. While employing the Bayesian bootstrap is straightforward from a computational perspective, using a Dirichlet process mixture would be much more intricate as, for each set $\{u^{(s)}_{\overline{D}i}\}_{i=1}^{n_{\overline{D}}}$ and $\{u^{(s)}_{Dj}\}_{j=1}^{n_D}$, we would need to generate another S' realizations from the resulting model parameters to compute the required distribution functions.

4 | COVARIATE-SPECIFIC OVERLAP COEFFICIENT AND PROPOSED ESTIMATOR

Let $\mathbf{X}_{\overline{D}}$ and $\mathbf{X}_D$ be covariate vectors and for simplicity we assume that the covariates of interest are the same in both the diseased and nondiseased populations. The covariate-specific overlap coefficient, for a given covariate value $\mathbf{x}$, is defined as

$$\begin{aligned}
\text{OVL}(\mathbf{x}) &= \int_{-\infty}^{\infty} \min\left\{f_{\overline{D}}(y|\mathbf{X}_{\overline{D}} = \mathbf{x}), f_D(y|\mathbf{X}_D = \mathbf{x})\right\} \mathrm{d}y\\
&= 1 - \frac{1}{2}\int_{-\infty}^{\infty} \left|f_{\overline{D}}(y|\mathbf{X}_{\overline{D}} = \mathbf{x}) - f_D(y|\mathbf{X}_D = \mathbf{x})\right| \mathrm{d}y,
\end{aligned}$$

where $f_{\overline{D}}(\cdot|\mathbf{X}_{\overline{D}} = \mathbf{x})$ denotes the conditional density of $Y_{\overline{D}}$ given $\mathbf{X}_{\overline{D}} = \mathbf{x}$, with $f_D(\cdot|\mathbf{X}_D = \mathbf{x})$ being analogously defined. Note that in this setting, for each possible value $\mathbf{x}$, we might obtain a different OVL and, therefore, also a possible different accuracy.

We now let $\{(\mathbf{x}_{\overline{D}i}, y_{\overline{D}i})\}_{i=1}^{n_{\overline{D}}}$ and $\{(\mathbf{x}_{Dj}, y_{Dj})\}_{j=1}^{n_D}$ be two independent random samples of covariates and test outcomes from the nondiseased and diseased populations, respectively. Further, for all $i = 1, \dots , n_{\overline{D}}$ and $j = 1, \dots , n_D$, let $\mathbf{x}_{\overline{D}i} = (x_{\overline{D}i,1}, \dots , x_{\overline{D}i,p})'$ and $\mathbf{x}_{Dj} = (x_{Dj,1}, \dots , x_{Dj,p})'$ be p-dimensional vectors of covariates. Our proposed estimator for the covariate-specific overlap coefficient is an adaptation to the conditional case of the OVL_{DPM} estimator. Specifically, we rely on a single-weights dependent Dirichlet process mixture of normal distributions[26] for modeling the conditional densities in each population. Under this setup we write the nondiseased conditional density as

$$f_{\bar{D}}(y_{\bar{D}i}|\mathbf{x}_{\bar{D}i}) = \int \phi(y_{\bar{D}i}|\mu(\mathbf{x}_{\bar{D}i},\boldsymbol{\beta}),\sigma^2)\mathrm{d}G_{\bar{D}}(\boldsymbol{\beta},\sigma^2), \quad G_{\bar{D}} \sim \mathrm{DP}(\alpha_{\bar{D}}, G^*_{\bar{D}}(\boldsymbol{\beta},\sigma^2)), \tag{10}$$

with the conditional density in the diseased population similarly defined. Note that the only difference to the model in (4) is that now the mean of each component is covariate-specific. Indeed, the model in (4) is a particular case of (10) by setting $\mu(\mathbf{x}_{\bar{D}i},\boldsymbol{\beta})$ to simply be an intercept. By using again (truncated) Sethuraman's representation, we have that the following expression holds for the conditional density of test outcomes in the nondiseased population

$$f_{\bar{D}}(y_{\bar{D}i}|\mathbf{x}_{\bar{D}i}) = \sum_{l=1}^{L_{\bar{D}}} \omega_{\bar{D}l}\phi(y_{\bar{D}i}|\mu_{\bar{D}}(\mathbf{x}_{\bar{D}i},\boldsymbol{\beta}_{\bar{D}l}),\sigma^2_{\bar{D}l}), \quad (\boldsymbol{\beta}_{\bar{D}l},\sigma^2_{\bar{D}l}) \overset{\text{iid}}{\sim} G^*_{\bar{D}}(\boldsymbol{\beta},\sigma^2),$$

where the weights follow the truncated stick-breaking construction described in Section 3.1. It is worth mentioning that although we are not explicitly modeling the variance of each mixture component as a function of the covariates, the overall variance of the mixture model still depends on covariates as its expression makes clear

$$\mathrm{var}(y_{\bar{D}i}|\mathbf{x}_{\bar{D}i}) = \sum_{l=1}^{L_{\bar{D}}}\omega_{\bar{D}l}\sigma^2_{\bar{D}l} + \sum_{l=1}^{L_{\bar{D}}}\omega_{\bar{D}l}\left\{\mu^2_{\bar{D}}(\mathbf{x}_{\bar{D}i},\boldsymbol{\beta}_{\bar{D}l}) - \left(\sum_{l=1}^{L_{\bar{D}}}\omega_{\bar{D}l}\mu_{\bar{D}}(\mathbf{x}_{\bar{D}i},\boldsymbol{\beta}_{\bar{D}l})\right)^2\right\}.$$

Note that by assuming that the weights $\omega_{\bar{D}l}$ do not vary with the covariates, the model might has limited flexibility in practice[27] and hence a flexible formulation for the mean of each component $\mu(\mathbf{x}_{\bar{D}i},\boldsymbol{\beta}_{\bar{D}l})$ is key to a good performance of the estimator. Following Inácio and Rodríguez-Álvarez,[28] we model the mean of each component as an additive sum of smooth functions, that is,

$$\mu_{\bar{D}}(\mathbf{x}_{\bar{D}i},\boldsymbol{\beta}_{\bar{D}l}) = \beta_{\bar{D}l0} + f_{\bar{D}l1}(x_{\bar{D}i,1}) + \cdots + f_{\bar{D}lp}(x_{\bar{D}i,p}),$$

where $f_{\bar{D}lh}(\cdot)$ is an unknown smooth function, for $l = 1, \ldots, L_{\bar{D}}$ and $h = 1, \ldots, p$. Particularly, we approximate each smooth function $f_{\bar{D}lh}(\cdot)$ by a linear combination of cubic B-splines basis functions defined over a sequence of knots $\xi_{\bar{D}h0} < \xi_{\bar{D}h1} < \cdots < \xi_{\bar{D}hK_{\bar{D}h}} < \xi_{\bar{D}h,K_{\bar{D}h}+1}$, where $\xi_{\bar{D}h0}$ and $\xi_{\bar{D}h,K_{\bar{D}h}+1}$ are boundary knots, while the remaining ones are interior knots. We write

$$f_{\bar{D}lh}(x_{\bar{D}i,h}) = \sum_{k=1}^{K_{\bar{D}h}+3} B_{\bar{D}hk}(x_{\bar{D}i,h})\beta_{\bar{D}lhk} = \mathbf{B}'_{\bar{D}\boldsymbol{\xi}_{\bar{D}h}}(x_{\bar{D}i,h})\boldsymbol{\beta}_{\bar{D}lh},$$

where $\mathbf{B}_{\bar{D}\boldsymbol{\xi}_{\bar{D}h}}(x_{\bar{D}i,h}) = (B_{\bar{D}h1}(x_{\bar{D}i,h}), \ldots, B_{\bar{D}hK_{\bar{D}h}+3}(x_{\bar{D}i,h}))'$ and $B_{\bar{D}hk(x)}$ denotes the kth cubic B-spline basis function in the nondiseased population, evaluated at x, and defined by the sequence of knots $\boldsymbol{\xi}_{\bar{D}h} = (\xi_{\bar{D}h0}, \ldots, \xi_{\bar{D}hk_{\bar{D}h}+1})'$ and, lastly, $\boldsymbol{\beta}_{\bar{D}lh} = (\beta_{\bar{D}lh1}, \ldots, \beta_{\bar{D}lh,K_{\bar{D}h}+3})'$. The mean function of each component is then expressed as

$$\begin{aligned}\mu_{\bar{D}}(x_{\bar{D}i},\boldsymbol{\beta}_{\bar{D}l}) &= \beta_{\bar{D}l0} + \mathbf{B}'_{\bar{D}\boldsymbol{\xi}_{\bar{D}1}}(x_{\bar{D}i,1})\boldsymbol{\beta}_{\bar{D}l1} + \cdots + \mathbf{B}'_{\bar{D}\boldsymbol{\xi}_{\bar{D}p}}(x_{\bar{D}i,p})\boldsymbol{\beta}_{\bar{D}lp} \\ &= \mathbf{z}'_{\bar{D}i}\boldsymbol{\beta}_{\overline{Dl}},\end{aligned}$$

with $\mathbf{z}'_{\bar{D}i} = (1, \mathbf{B}'_{\bar{D}\boldsymbol{\xi}_{\bar{D}1}}(x_{\bar{D}i,1}), \ldots, \mathbf{B}'_{\bar{D}\boldsymbol{\xi}_{\bar{D}p}}(x_{\bar{D}i,p}))$ and $\boldsymbol{\beta}_{\overline{Dl}} = (\beta_{\bar{D}l0}, \boldsymbol{\beta}'_{\bar{D}l1}, \ldots, \boldsymbol{\beta}'_{\bar{D}lp})'$. As it is widely acknowledged, the number of interior knots, and to a less extent, their location along the covariate's range, may strongly impact the inferences. To assist the selection of the number of knots, we use two model selection criteria, namely the log pseudo marginal likelihood[29] (LPML) and the widely applicable information criterion[30] (WAIC). The empirical performance of these two criteria in selecting an appropriate number of knots in a similar context was investigated by Inácio and Rodríguez-Álvarez.[28] Quantile residuals[31] and posterior predictive checks[32] can also aid to assess the overall fit of the model. Regarding the location of the $K_{\bar{D}h}$, $h = 1, \ldots, p$, interior knots, and to ensure an approximately equal number of observations at each interval defined by the knots, we follow Rosenberg[33] and $\xi_{\bar{D}hk}$ is set equal to the $k/(K_{\bar{D}h} + 1)$, $k = 1, \ldots, K_{\bar{D}h}$ quantile of $\mathbf{x}_{\bar{D},h} = (x_{\bar{D}1,h}, \ldots, x_{\bar{D}n_{\bar{D}},h})'$. In addition, we set the boundary knots $\xi_{\bar{D}h0}$ and $\xi_{\bar{D}h,K_{\bar{D}h}+1}$ to the minimum and maximum of $\mathbf{x}_{\bar{D},h}$, respectively. Before proceeding we shall note that although for notational simplicity we have assumed that all p

covariates are continuous, our model's formulation can easily include factor variables, interactions between factor variables, and interactions between a smooth function and a factor variable. The model for the conditional density is thus a mixture of normal distributions where the components' mean vary differentially and nonlinearly with the covariates and this model can also be regarded as a Dirichlet process mixture of additive normal models, that is,

$$f_{\overline{D}}(y_{\overline{D}i}|\mathbf{x}_{\overline{D}i}) = \sum_{l=1}^{L_{\overline{D}}} \omega_{\overline{D}l}\phi(y_{\overline{D}i}|\mathbf{z}'_{\overline{D}i}\boldsymbol{\beta}_{\overline{D}l}, \sigma^2_{\overline{D}l}), \quad (\boldsymbol{\beta}_{\overline{D}l}, \sigma^2_{\overline{D}l}) \overset{\text{iid}}{\sim} G^*_{\overline{D}}(\boldsymbol{\beta}, \sigma^2).$$

Our model is completed with the specification of the centering distribution. As in the no covariate case, we specify a conditionally conjugate centering distribution, that is,

$$G^*_{\overline{D}}(\boldsymbol{\beta}, \sigma^2) \equiv N_{Q_{\overline{D}}}(\boldsymbol{\beta}|\mathbf{m}_{\overline{D}}, \mathbf{S}_{\overline{D}})\Gamma(\sigma^{-2}|a_{\overline{D}}, b_{\overline{D}}),$$

with conjugate hyperpriors $\mathbf{m}_{\overline{D}} \sim \text{N}(\mathbf{m}_{\overline{D}0}, \mathbf{S}_{\overline{D}0})$ and $\mathbf{S}^{-1}_{\overline{D}} \sim \text{Wishart}(\nu_{\overline{D}}, (\nu_{\overline{D}}\boldsymbol{\Psi}_{\overline{D}})^{-1})$ (a Wishart distribution with degrees of freedom $\nu_{\overline{D}}$ and expectation $\Psi^{-1}_{\overline{D}}$) and where $Q_{\overline{D}}$ is the dimension of the vector $\mathbf{z}_{\overline{D}i}$. Hyperparameters $\mathbf{m}_{\overline{D}0}$ and $\Psi_{\overline{D}}$ must be chosen to represent the prior belief about the regression coefficients associated to each mixture component and about their covariance matrix, respectively, whereas $\mathbf{S}_{\overline{D0}}$ and $\nu_{\overline{D}}$ are chosen to represent the confidence in the prior belief of $\mathbf{m}_{\overline{D}0}$ and $\Psi_{\overline{D}}$, respectively. The full conditional distributions for all model parameters are also available in closed form (details in the Supplementary Materials), again allowing for ready posterior simulation through Gibbs sampling. Similarly, to the unconditional case, we have

$$\begin{aligned}
\text{OVL}^{(s)}(\mathbf{x}) &= 1 - \frac{1}{2}\int_{-\infty}^{+\infty} \left|f^{(s)}_{\overline{D}}(y|\mathbf{x}) - f^{(s)}_{D}(y|\mathbf{x})\right| \mathrm{d}y \\
&\approx 1 - \frac{1}{2}\frac{\Delta y}{2}\left\{\left|f^{(s)}_{\overline{D}}(y_0|\mathbf{x}) - f^{(s)}_{D}(y_0|\mathbf{x})\right| + 2\sum_{r=1}^{N-1}\left|f^{(s)}_{\overline{D}}(y_r|\mathbf{x}) - f^{(s)}_{D}(y_r|\mathbf{x})\right| + \left|f^{(s)}_{\overline{D}}(y_N|\mathbf{x}) - f^{(s)}_{D}(y_N|\mathbf{x})\right|\right\},
\end{aligned}$$

$$f^{(s)}_{\overline{D}}(y|\mathbf{x}) = \sum_{l=1}^{L_{\overline{D}}}\omega_{\overline{D}l}\phi\left(y|\mathbf{z}'\boldsymbol{\beta}^{(s)}_{\overline{D}l}, (\sigma^{(s)}_{\overline{D}l})^2\right), \qquad f^{(s)}_{D}(y|\mathbf{x}) = \sum_{l'=1}^{L_{D}}\omega_{Dl'}\phi\left(y|\mathbf{z}'\boldsymbol{\beta}^{(s)}_{Dl'}, (\sigma^{(s)}_{Dl'})^2\right), \quad s = 1, \dots, S.$$

For a given $\mathbf{x}$, a point estimate can be obtained by considering the mean or median over the ensemble of covariate-specific overlap coefficients $\left\{\text{OVL}^{(1)}(\mathbf{x}), \dots, \text{OVL}^{(S)}(\mathbf{x})\right\}$. Similarly, a pointwise credible interval can be obtained from the percentiles of the same ensemble.

5 SIMULATION STUDY

5.1 Unconditional case

5.1.1 Simulation scenarios

We considered three scenarios as listed in Table 1. Scenario I corresponds to the case where test outcomes in the two populations follow normal distributions. In Scenario II, test outcomes in both groups arise from different and non-normal distributions, namely a gamma distribution and a skew normal distribution. Lastly, Scenario III considers mixtures of normal distributions in each of the two populations. For each simulation scenario, three different parameters' configurations, that lead to small, intermediate, and large overlaps, were used and 100 datasets were generated using sample sizes of $(n_{\overline{D}}, n_D) \in \{(50, 50), (100, 100), (200, 200), (500, 500), (1000, 1000)\}$.

5.1.2 Models

For Step 1 of our proposed estimators and to facilitate prior specification, test outcomes were standardized so that the resulting mean is zero and the variance is one and we transformed back to the original scale when calculating the relevant quantities. We have considered $m_{d0} = 0$, $S_{d0} = 10$, $a_d = 2$, and $b_d = 0.5$, where $d \in \{\overline{D}, D\}$. Our reasoning behind this

TABLE 1 Different distributional assumptions for $Y_{\bar{D}}$ and Y_D under Scenario I, II, and III

Scenario	$Y_{\bar{D}}$	Y_D	OVL
I	$\mathrm{N}(-0.75, 1^2)$	$\mathrm{N}(2.5, 1^2)$	0.104
	$\mathrm{N}(1.1, 1^2)$	$\mathrm{N}(2.5, 1^2)$	0.484
	$\mathrm{N}(2.2, 1^2)$	$\mathrm{N}(2.5, 1^2)$	0.880
II	$\Gamma(3, 1)$	$\mathrm{SN}(5, 2, 5)$	0.172
	$\Gamma(3, 1)$	$\mathrm{SN}(3, 2, 5)$	0.482
	$\Gamma(3, 1)$	$\mathrm{SN}(1.25, 2, 5)$	0.860
III	$0.5\mathrm{N}(-2.5, 1^2) + 0.5\mathrm{N}(0.5, 1^2)$	$0.5\mathrm{N}(2.5, 1^2) + 0.5\mathrm{N}(5, 1^2)$	0.159
	$0.5\mathrm{N}(-1.15, 1^2) + 0.5\mathrm{N}(1.5, 1^2)$	$0.5\mathrm{N}(1.5, 1^2) + 0.5\mathrm{N}(3.5, 1^2)$	0.510
	$0.5\mathrm{N}(0, 1^2) + 0.5\mathrm{N}(3, 1^2)$	$0.5\mathrm{N}(0.5, 1^2) + 0.5\mathrm{N}(3.25, 1^2)$	0.897

Note: Note that $\mathrm{SN}(\nu, \eta, \lambda)$ denotes a skew normal distribution with location ν, scale η, and skewness parameter λ.

hyperparameters' choice is as follows. Because test outcomes are standardized, we expect the means of the mixture' components to be close to zero and hence $m_{d0} = 0$. The variance S_{d0} then controls where the sampled μ_{dl} may lie and considering $S_{d0} = 10$ implies that about 95% of the drawn values roughly lie within -6 and 6. We shall also note that $a_d = 2$ implies a prior for σ_{dl}^2 with infinite variance that is centered around a finite mean ($b_d = 0.5$) and therefore favors variances less than one. If we recall that the standardized test outcomes have unit variance, it makes sense to expect the within component variance to be smaller than the overall variance. As we mentioned in Section 3, we considered $\alpha_d = 1$ which favors a small number of occupied mixture components relative to the sample size. Finally, we have considered $L_d = 10$, thus capping the maximum number of mixture components at ten. Posterior inference was based on 5000 iterations after a burn-in of 2000 iterations of the Gibbs sampler was discarded. For the $\mathrm{OVL}_{\mathrm{DPM}}$ estimator a grid of $N = 1001$ points was considered.

We also compared the performance of our Bayesian nonparametric estimators against that of those obtained by in (2) and (3) modeling f_D and $f_{\bar{D}}$ using kernel estimation techniques. Specifically, a normal kernel is used, such that the resulting density function estimate in the nondiseased population can be written as

$$\hat{f}_{\bar{D}}(y) = \frac{1}{n_{\bar{D}} h_{\bar{D}}} \sum_{i=1}^{n_{\bar{D}}} \phi\left(\frac{y - y_{\bar{D}i}}{h_{\bar{D}}}\right),$$

where $h_{\bar{D}}$ is the bandwidth/smoothing parameter and following Schmid and Schmidt[18] we selected it using Silverman's rule of thumb.[34(ch.3)] Trivially, a similar expression holds for the estimate of the density function of test outcomes in the diseased population. The corresponding kernel counterpart of our $\mathrm{OVL}_{\mathrm{DPM}}$ estimator takes the form

$$\widehat{\mathrm{OVL}} \approx 1 - \frac{1}{2}\frac{\Delta y}{2}\left\{\left|\hat{f}_{\bar{D}}(y_0) - \hat{f}_D(y_0)\right| + 2\sum_{r=1}^{N-1}\left|\hat{f}_{\bar{D}}(y_r) - \hat{f}_D(y_r)\right| + \left|\hat{f}_{\bar{D}}(y_N) - \hat{f}_D(y_N)\right|\right\}, \tag{11}$$

where as before $N = 1001$ was considered. The estimator that arises from (3), and which can be regarded as the frequentist counterpart of our $\mathrm{OVL}_{\mathrm{DPM\text{-}BB}}$ estimator,[18] is given by

$$\widehat{\mathrm{OVL}} = \frac{1}{n_{\bar{D}}}\sum_{i=1}^{n_{\bar{D}}} I\left\{\hat{f}_{\bar{D}}(y_{\bar{D}i}) < \hat{f}_D(y_{\bar{D}i})\right\} + \frac{1}{n_D}\sum_{j=1}^{n_D} I\left\{\hat{f}_{\bar{D}}(y_{Dj}) \geq \hat{f}_D(y_{Dj})\right\}. \tag{12}$$

5.1.3 | Results

The boxplot of the estimated overlap coefficients, across the 100 datasets, for all scenarios, distribution parameters' configurations, and sample sizes considered are presented in Figures 1–3. We can observe that our $\mathrm{OVL}_{\mathrm{DPM}}$ estimator is able

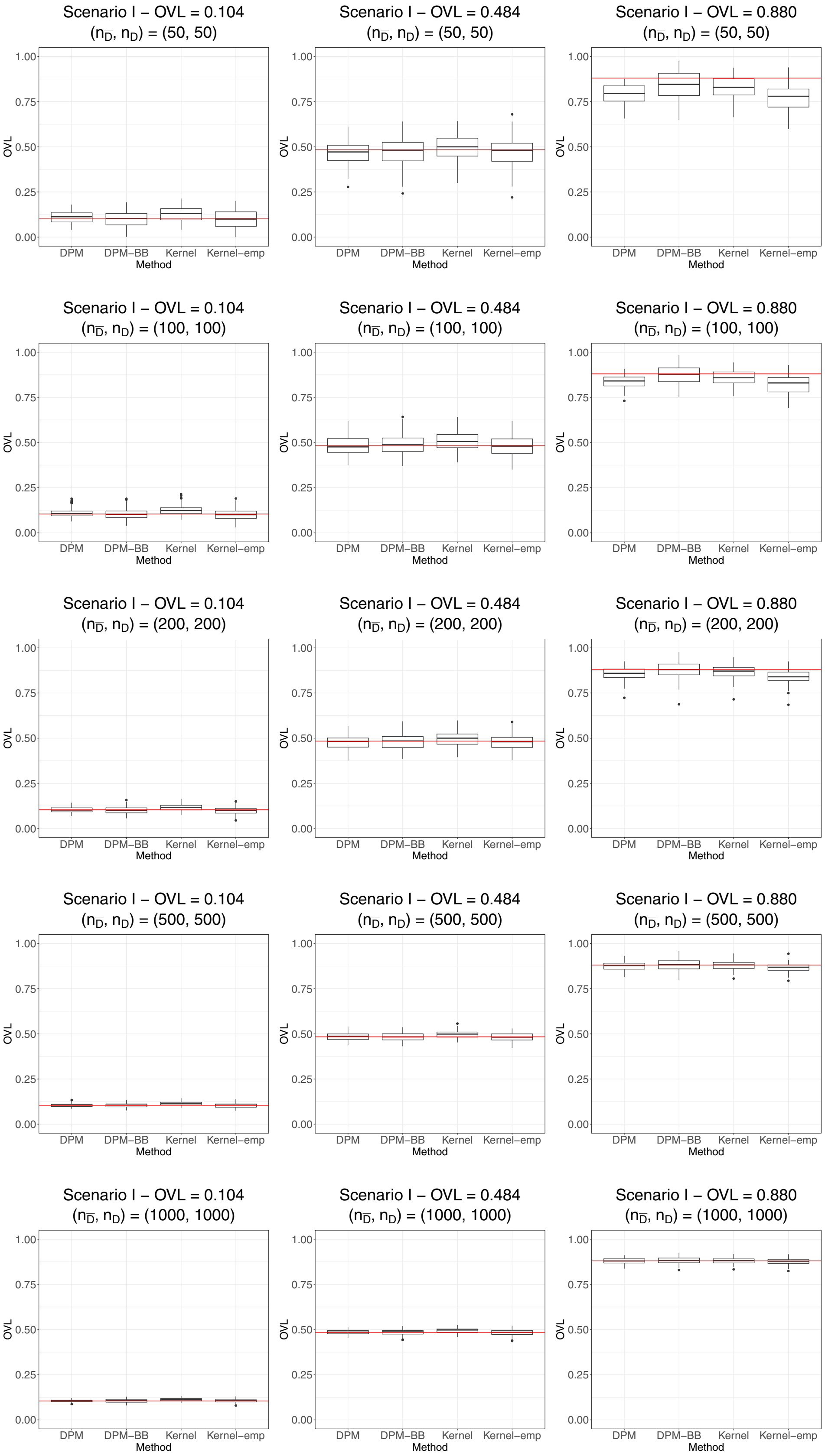

FIGURE 1 Scenario I. Boxplot of the estimates of the coefficient of overlap across the 100 simulated datasets and for the different parameter's configurations and sample sizes considered. The solid red line represents the true OVL. For the Bayesian estimators, what is reported, for each dataset, is the posterior median. Kernel: kernel estimator based on expression (11); Kernel-emp: kernel estimator based on expression (12)

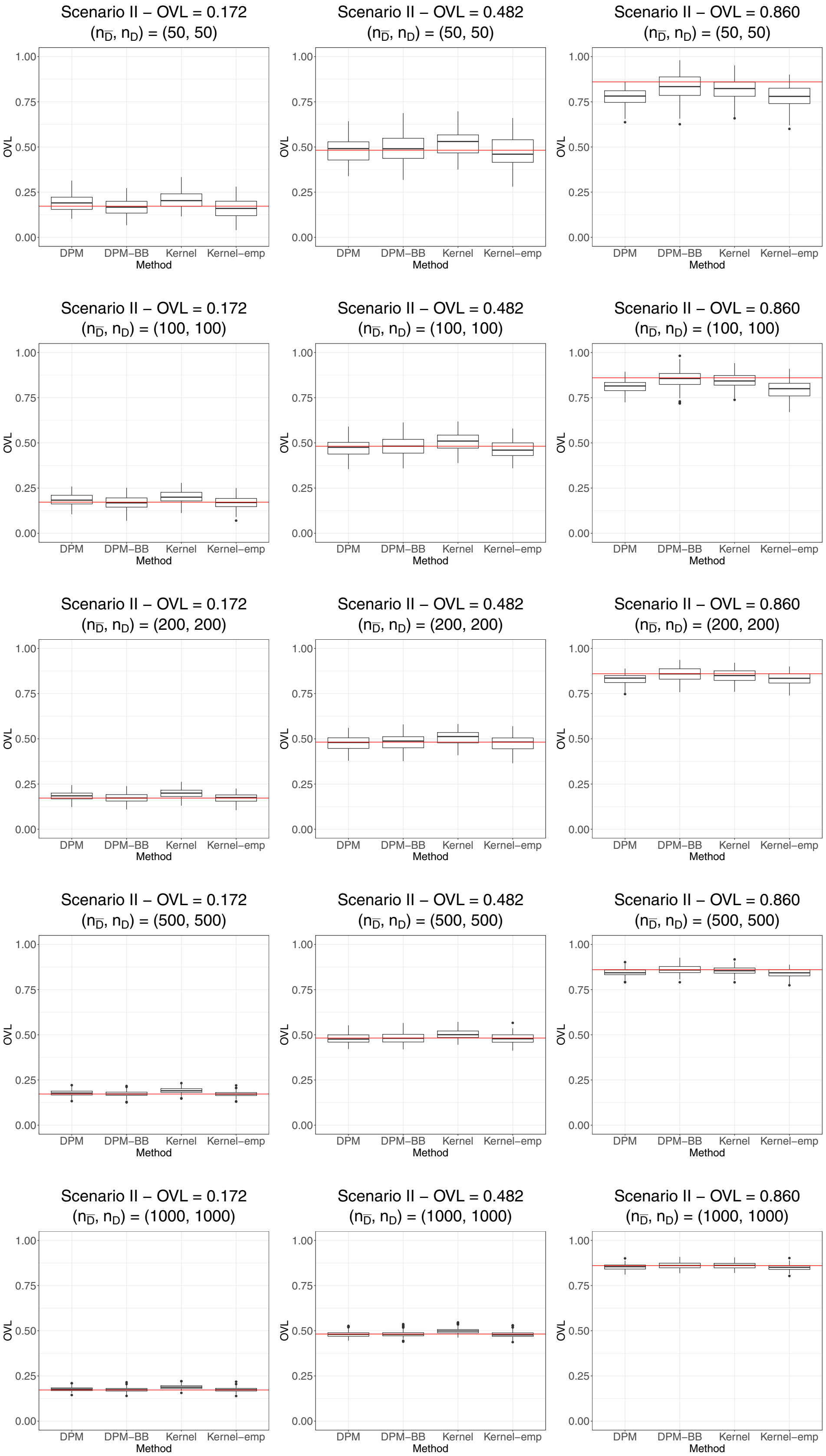

FIGURE 2 Scenario II. Boxplot of the estimates of the coefficient of overlap across the 100 simulated datasets and for the different parameter's configurations and sample sizes considered. The solid red line represents the true OVL. For the Bayesian estimators, what is reported, for each dataset, is the posterior median. Kernel: kernel estimator based on expression (11); Kernel-emp: kernel estimator based on expression (12)

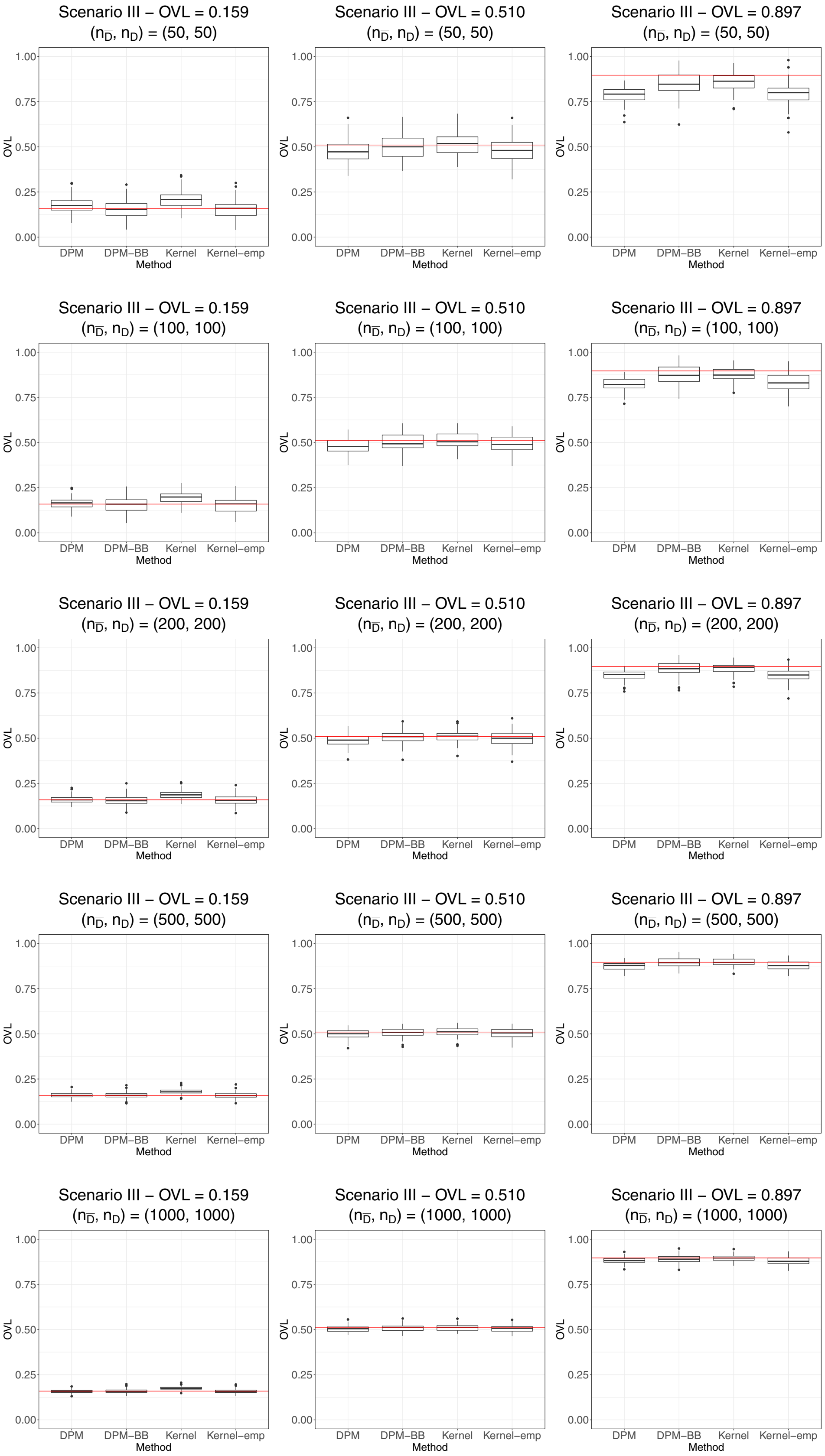

FIGURE 3 Scenario III. Boxplot of the estimates of the coefficient of overlap across the 100 simulated datasets and for the different parameter's configurations and sample sizes considered. The solid red line represents the true OVL. For the Bayesian estimators, what is reported, for each dataset, is the posterior median. Kernel: kernel estimator based on expression (11); Kernel-emp: kernel estimator based on expression (12)

to provide unbiased estimates except when the true value of the overlap coefficient is very high (about 0.8 or above). Such a bias decreases, however, when sample size increases. To investigate whether this bias could be due to the prior information used, we have conducted a sensitivity analysis with a data-driven prior[35(p. 532)] and the results, not shown, were unchanged. In turn, the $\text{OVL}_{\text{DPM-BB}}$ estimator provides unbiased estimates in almost all cases considered; the only exception is in Scenario III when the underlying true value of the overlap coefficient is very high and the sample size is small (eg, $(n_{\bar{D}}, n_D) = (50, 50)$). In addition, the performance of our OVL_{DPM} estimator is only inferior to the corresponding kernel counterpart (see Equation 11) when the overlap between the densities in the two groups is very high, with this being especially true for the smaller sample sizes (eg, a sample size below 200 in each population). On the other hand, the $\text{OVL}_{\text{DPM-BB}}$ estimator has a better performance than the corresponding kernel counterpart (expression 12) regardless of the amount of overlap between the distributions of test outcomes. It can also be noticed that the range of the estimates, across the 100 datasets, is wider for the estimators arising from the representation in (3), Bayesian or frequentist, than those arising from the representation in (2). As expected, for all methods, estimates get more concentrated around the true value of the overlap coefficient as the sample size increases.

We have further investigated the empirical coverage probability of the 95% credible/confidence intervals and the corresponding width of the intervals. For the kernel methods, the intervals were obtained using 500 bootstrap resamples from each simulated dataset of test outcomes. Results are presented in Figures 1-3 and Tables 1-3 of the Supplementary Materials. As can be observed from such figures, for all sample sizes and scenarios, the width of the 95% credible intervals associated to the $\text{OVL}_{\text{DPM-BB}}$ estimator is larger than that of the interval associated with the OVL_{DPM} estimator, with this being more marked for the smaller sample sizes and for the case of a high overlap between the two distributions. The same is true for the corresponding kernel counterparts. Further, the width of the 95% credible intervals associated with the OVL_{DPM} estimator is of the same magnitude as that of the 95% bootstrap confidence intervals associated to (11). However, the width of the 95% credible intervals associated with the $\text{OVL}_{\text{DPM-BB}}$ estimator tend to be larger than those of the 95% bootstrap confidence intervals corresponding to (12), with this being markedly the case when the sample size in the two groups is small. We found the empirical coverage probabilities of the 95% credible intervals associated to the $\text{OVL}_{\text{DPM-BB}}$ estimator to be close to the nominal value for all scenarios and sample sizes under consideration. In turn, for the OVL_{DPM} estimator, the empirical coverage probabilities of the 95% credible intervals were also close to the nominal value, with the exception of the case where the true value of the coefficient of overlap is very high, with this being even more marked for smaller sample sizes. For instance, in Scenario III, in the case where the true value of the overlap coefficient is 0.896, even for the sample size of $(n_{\bar{D}}, n_D) = (1000, 1000)$, the coverage probability is only 87%. This should come at no surprise, as it can be observed from Figure 3, there is still some bias in the estimates and the width of the intervals tend to be smaller for such a large sample size. Also, the coverage of the intervals associated to the Bayesian approaches, in almost all cases considered, tend to be closer to the nominal value than the corresponding kernel counterparts.

Before ending this part, it is worth mentioning that for the kernel estimators, Silverman's rule of thumb is expected to provide a reasonable bandwidth's value when test outcomes are close to be normally distributed.[36(p. 61)] We have also considered a bandwidth selected by unbiased least-squares cross-validation,[36(ch.3)] as implemented in the `R` function `bw.ucv`, and for the scenarios listed here, the results (not shown), in terms point estimates of the overlap coefficient, are basically indistinguishable from those obtained using Silverman's rule of thumb. Of course, depending on the test outcomes distributions, this may not always be the case and our code also has this bandwidth selector option available.

5.2 | Covariate-specific case

5.2.1 | Simulation scenarios and implementation details

We consider three simulation scenarios and for each of them we simulate 100 datasets considering the following sample sizes: $(n_{\bar{D}}, n_D) \in \{(100, 100), (200, 200), (500, 500), (1000, 1000)\}$. In Scenario I, we consider a homoscedastic regression model with a linear covariate effect in the diseased and nondiseased populations

$$y_{\bar{D}i} | x_{\bar{D}i} \overset{\text{ind}}{\sim} \text{N}(0.5 + x_{\bar{D}i}, 1.5^2), \quad y_{Dj} | x_{Dj} \overset{\text{ind}}{\sim} \text{N}(2 + 4x_{Dj}, 2^2).$$

It is interesting to note that although the effect of the covariate in the underlying regression models is linear, the resulting functional form of the covariate-specific coefficient of overlap is nonlinear. In Scenario II, the effect of the covariate on the mean of each regression model is nonlinear in the two groups

$$y_{\overline{D}i}|x_{\overline{D}i} \overset{\text{ind}}{\sim} \text{N}(\sin(\pi(x_{\overline{D}i}+1)), 0.5^2), \quad y_{Dj}|x_{Dj} \overset{\text{ind}}{\sim} \text{N}(x_{Dj}^2, 1^2).$$

Finally, Scenario III is the most challenging scenario, involving a heteroscedastic regression model with a nonlinear covariate effect in the nondiseased population and in the diseased population it involves a two-component mixture of normal distributions whose weights depend on the covariate and whose covariate effect on the mean of one of the components is nonlinear

$$y_{\overline{D}i}|x_{\overline{D}i} \overset{\text{ind}}{\sim} \text{N}\left(\sin(\pi x_{\overline{D}i}), \sqrt{\exp(x_{\overline{D}i})}^{2}\right), \quad y_{Dj}|x_{Dj} \overset{\text{ind}}{\sim} \frac{\exp(x_{Dj})}{1+\exp(x_{Dj})}\text{N}(x_{Dj}, 0.5^2) + \frac{1}{1+\exp(x_{Dj})}\text{N}(x_{Dj}^2, 0.75^2).$$

In all three scenarios, $x_{\overline{D}i}, x_{Dj} \overset{\text{iid}}{\sim} \text{U}(-1, 1)$, for $i = 1, \ldots, n_{\overline{D}}$ and $j = 1, \ldots, n_D$.

For each simulated dataset, we fit our Bayesian nonparametric estimator introduced in Section 4 and posterior inferences are based on 8000 iterations after a burn-in period of 2000 iterations. As in the unconditional case, a grid of $N = 1001$ points was considered when applying the trapezoidal rule. Again, to facilitate prior specification, both test results and covariates were standardized, and we used $\mathbf{m}_{d0} = \mathbf{0}_{Q_d}, \mathbf{S}_{d0} = 10I_{Q_d}, \nu_d = Q_d + 2, \Psi_d = I_{Q_d}, \alpha_d = 1$, and $L_d = 10$, $d \in \{\overline{D}, D\}$. Further, we have considered a cubic B-splines formulation with no interior knots, that is, $K_{\overline{D}} = K_D = 0$ (and therefore $Q_{\overline{D}} = Q_D = 4$), for the mean of each normal mixture component.

5.2.2 | Results

In Figure 4, for each scenario and sample size considered, we depict the mean of the posterior medians across the 100 simulated datasets, along with the 2.5% and 97.5% simulation quantiles of these 100 posterior medians. As it can be observed, our Bayesian nonparametric estimator is able to recover the true functional form of the covariate-specific overlap coefficient in all three scenarios. It should be noted, nonetheless, that in Scenario III there is some bias, more notorious closer to the boundaries of the covariate support. To evaluate whether the internal number of knots would have an impact in such a bias in this scenario, we have also considered a model where three interior knots were used to model the mean of each of the ten components and following the rule discussed in Section 4, these were located at the 0.25, 0.5, and 0.75 quantiles of the covariate. Results are shown in Figure 5 (top row) of the Supplementary Materials and as it can be noticed, the bias still persists. We have also investigated the frequentist coverage probability of the 95% credible intervals associated to our approach and the results are presented in Figure 4 of the Supplementary Materials. From this figure we can see that in Scenarios I and II, for most sample sizes and covariate levels considered, the empirical coverage probabilities are close to the nominal value of 0.95. On the other hand, in Scenario III, for the sample sizes of 500 and 1000 in each of the populations and for a few covariate levels, the empirical coverage probability is substantial below 0.95. This is no surprise, as the covariate levels associated with such a low coverage probability are the same where bias in the estimation occurs (and due to the large sample sizes, the pointwise 95% credible intervals are narrower and therefore do not include the true conditional overlap coefficient, by opposition to what occurs at smaller sample sizes where the credible intervals are wider).

6 | ILLUSTRATIVE APPLICATIONS

6.1 | Search for biomarkers of ovarian cancer

We analyze the microarray dataset from Pepe et al,[37] which is concerned with the search for biomarkers of ovarian cancer that could be used in population screening. The dataset contains mRNA expression of 1536 clones of genes and comprises ovarian tissue from 30 subjects with ovarian cancer and 23 control subjects. The goal is to identify genes that are differentially expressed in ovarian cancer tissue compared with normal ovarian tissue. A gene would be an ideal candidate for being a marker if its values in the cancer tissue are completely different than those in the normal tissue. As mentioned in Pepe et al,[37(pp. 134-135)] in case there is some overlap between the two distributions gene expression (cancer

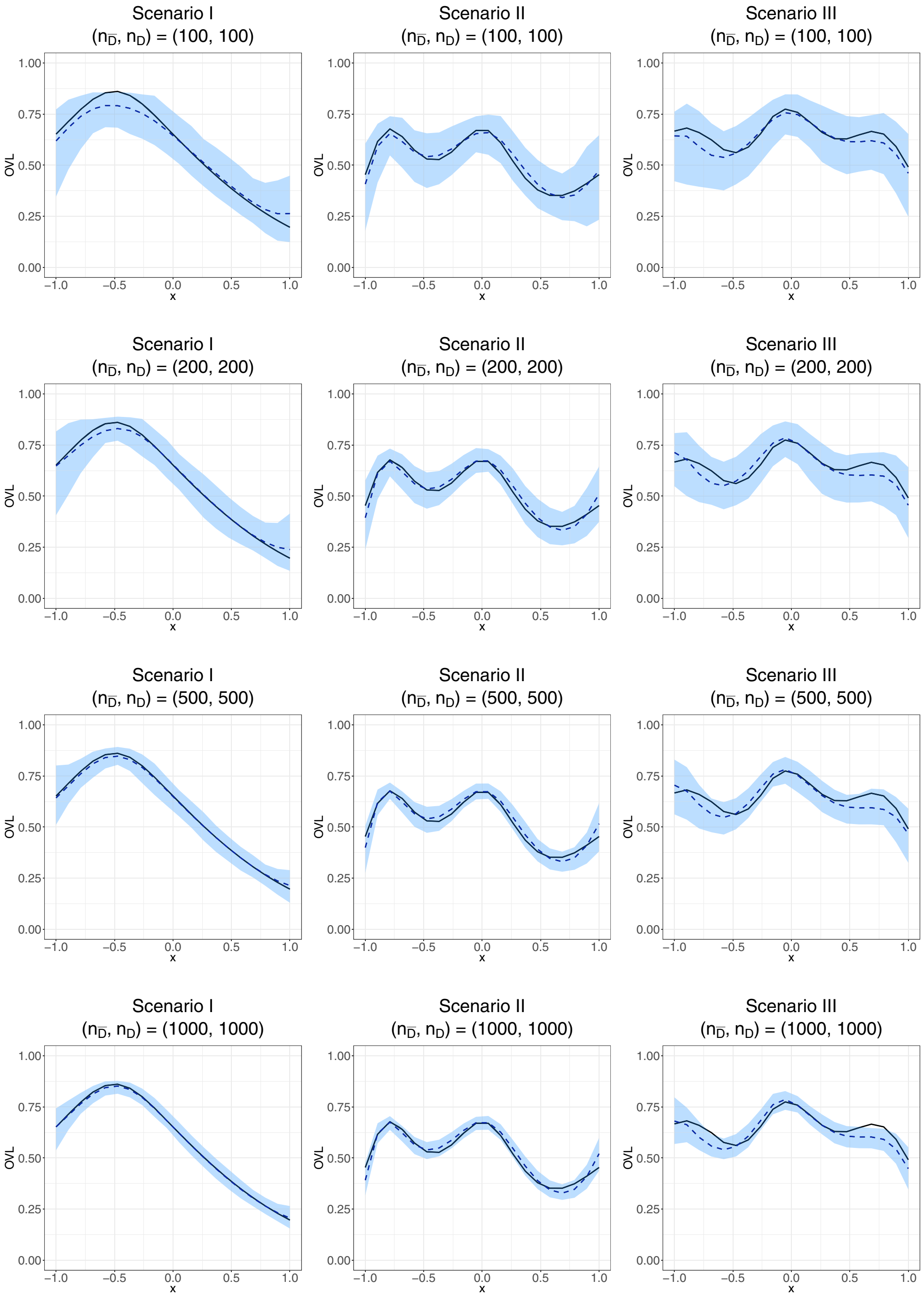

FIGURE 4 True conditional overlap (solid black line) and average value across the 100 simulated datasets (dashed line) of the posterior median of the covariate-specific overlap coefficient. The shaded area are bands constructed using the 2.5% and 97.5% quantiles across of such 100 posterior medians

vs non-cancer), if the gene is able to distinguish a subset of cancers from non-cancers, then such a gene should be of more practical interest than a gene whose expression levels in cancer tissues are entirely within the range of those for non-cancer tissues. As also referred in Pepe et al,[37(p. 134)] ovarian tissue cannot, obviously, be directly used as a screening test. However, if a gene turns out to be differentially expressed in cancer tissue, then the corresponding protein product may be detectable in blood or urine and therefore it could constitute the basis for a screening test. In the analysis that follows, we have removed 12 genes that had a missing value each, leaving us with 1524 genes. Further, because for some genes the (relative) expression intensities were very close to zero, we applied the log transformation, so that our Dirichlet process mixture of normal distributions in Step 1 (or the frequentist methods with a normal kernel), can be applied appropriately, that is, without placing (too much) mass in the negative values. This is equivalent to a Dirichlet process mixture of lognormal distributions on the original data scale.[26(p. 765)]

For the sake of illustration, we selected genes 9, 93, and 1033 in the dataset. The reasons behind our choice of genes are as follows. Gene 9 is an example where the relative gene expression intensities tend to be lower in the ovarian cancer group than in the noncancer group. In turn, for gene 93, the relative gene expression intensities tend to be lower for those in the noncancer group. Finally, for gene 1033, the distribution of the relative gene expression intensities shows a bimodality in each of the groups, thus rendering the use of the popular model that assumes a normal distribution, possibly after transformation of the scales, in each population, inadequate. Further, while for genes 9 and 93 there is little overlap between the two distributions of the relative gene expression intensities, for gene 1033 there is a considerable overlap (see Figure 5). Posterior inference was obtained using 10 000 iterates after 2000 iterations were discarded as burn-in period. The same prior information used in the simulation study was applied here. The point estimates (posterior median) of the coefficients of overlap and corresponding 95% credible intervals using our proposed estimators are shown in Table 2. Combining all this information, and as already expected from Figure 5, we can conclude that while genes 9 and 93 are good candidates to form the basis of a screening test, gene 1033 possibly presents too much of an overlap between the two distributions of relative expression intensities to be considered as a candidate. In Figures 6, 9, and 12 of the Supplementary Materials we present, for each of the three genes considered, the corresponding QQ-plots of the quantile residuals resulting from fitting the Dirichlet process mixture of normal distributions model in each of the two groups and, as already evidenced by the estimated densities in Figure 5 which follow quite nicely the histograms of the data, the fit is quite good. In the first step of our estimator, to monitor convergence of the MCMC chains, traceplots and the Geweke statistic were used. Note that the well-known label switching problem often leads to poor mixing of the chains of the component-specific parameters and therefore we have monitored the corresponding MCMC chain of the induced density in each group.[23(p. 553)] The traceplots, shown in Figures 7, 10, and 13, of the Supplementary Materials, for three randomly selected relative gene expression intensities, and the Geweke statistics, shown in Figures 8, 11, and 14 of the Supplementary Materials, do not suggest any lack of convergence of the (density) chains. The effective sample sizes are also reasonably high enough (Figures 8, 11, and 14 in the Supplementary Materials). For comparison purposes, we also include the results from the kernel approaches introduced in the simulation study in Section 5, with the 95% bootstrap confidence intervals were obtained using 1000 resamples of the relative expression intensities in the ovarian and non-ovarian cancer groups. There are no substantial differences between the results obtained under the different methods and the overall message is the same under the four approaches.

Finally, and only for illustrative purposes, the 1524 genes were ranked according the overlap coefficient. The top ten ranked genes selected by the different methods are presented in Table 4 of the Supplementary Materials and, as it can be concluded, the four approaches largely agree with respect to their ranking of the top ten genes.

6.2 | Age-specific accuracy of glucose as a biomarker of diabetes

We applied our covariate-specific Bayesian nonparametric approach to data obtained from a population based survey of diabetes in Cairo, Egypt.[38] Postprandial blood glucose levels were obtained on 286 individuals and according to the World Health Organization criteria at the time for diagnosing diabetes, 88 individuals were classified as diabetic and 198 as nondiabetic. The age of the subjects was included as a covariate as according to Smith and Thompson,[38] the ageing process may be associated with relative insulin resistance or deficiency among nondiabetic individuals and therefore postprandial glucose levels are expected to be higher for older subjects who do not suffer from diabetes. The goal of the analysis is to evaluate how the discriminatory ability of the glucose levels as a marker of diabetes may change with age. In Figure 6 (left and middle panels) we show the scatter plots for the nondiabetic and diabetic populations along with the marginal histograms for the glucose levels and the age and we can observe that in the former group the glucose levels get indeed slightly elevated as age increases, although we have far less older individuals than younger ones.

The age effect was modeled through the mean of each mixture component using a B-splines formulation with no interior knots and this choice was informed by both the LPML and WAIC criteria, with the two favoring the simplest model (when compared to a model with one, two, and three interior knots). Posterior inference was based on 20000 Gibbs sampler iterates after a burn-in of the first 5000 realizations was discarded and the same prior information used in Section 5.2 of the simulation study was applied (on the standardized data). Inspection of traceplots and Geweke criterion do not suggest any lack of convergence of the (density) chains (Figures 15 and 16 of the Supplementary Materials). The

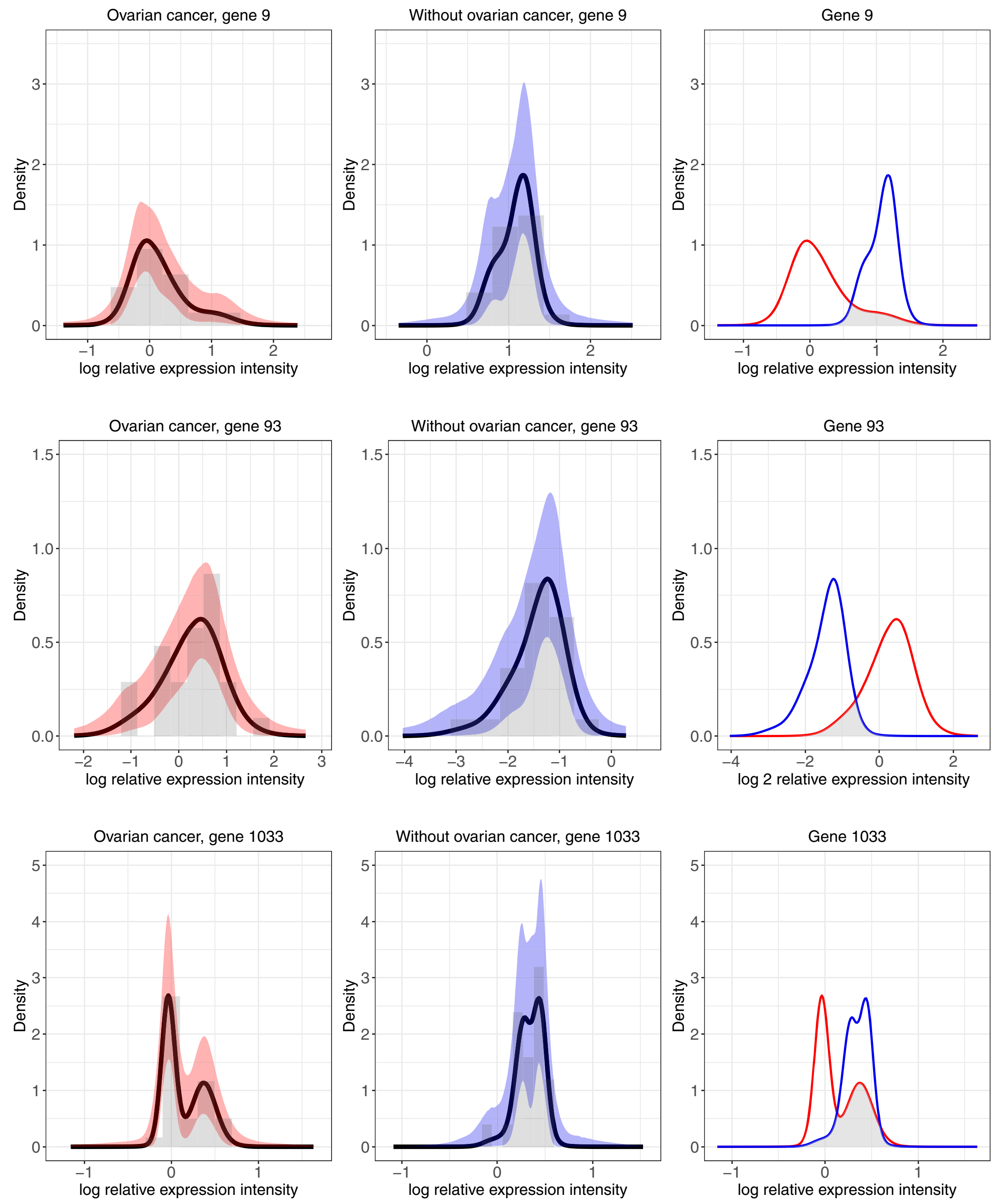

FIGURE 5 Histograms and density estimates (posterior medians and 95% pointwise credible bands) of the (log) relative gene expression intensities in both groups (with and without ovarian cancer) for gene 9 (top row), gene 93 (middle row), and gene 1033 (bottom row). Density estimates were obtained using a Dirichlet process mixture of normal distributions

TABLE 2 Ovarian cancer application

	DPM	DPM-BB	Kernel	Kernel-emp
Gene 9	0.181 (0.082, 0.327)	0.152 (0.046, 0.363)	0.193 (0.063, 0.323)	0.133 (0.033, 0.277)
Gene 93	0.159 (0.070, 0.305)	0.108 (0.026, 0.302)	0.169 (0.047, 0.271)	0.100 (0.000, 0.233)
Gene 1033	0.471 (0.311, 0.653)	0.479 (0.276, 0.753)	0.527 (0.322, 0.690)	0.487 (0.277, 0.674)

Note: Estimates of the coefficient of overlap for gene 9, 93, and 1033. For the Bayesian estimates what is reported is the posterior median and the 95% credible interval (in parentheses). For the kernel estimates, 95% bootstrap confidence intervals are reported. Kernel: kernel estimator based on expression (11); Kernel-emp: kernel estimator based on expression (12).

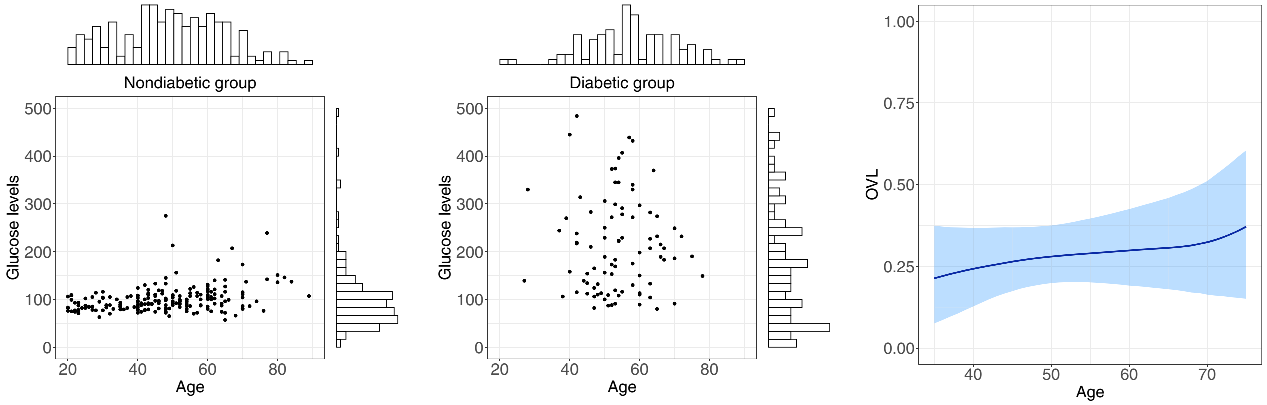

FIGURE 6 Diabetes application. Left and middle panels: scatter plots of the data along with marginal histograms. Right panel: age-specific coefficient of overlap. The solid line is the posterior median and the shaded area represent the 95% pointwise credible band

effective sample sizes, shown in Figure 17 of the Supplementary Materials, are also reasonably high enough. The density estimates (posterior median) in each population, for six different ages, along with their overlap are shown in Figure 18 of the Supplementary Materials and we can observe that the overlap is higher for the older ages. To inspect the age effect further, Figure 6 (right panel) shows a plot of the age-specific overlap coefficient for ages between 35 and 75 years old and we can observe that, in fact, the capacity of the glucose levels to distinguish between diabetic and nondiabetic individuals decreases with age. Inferences are less precise for older ages. Similar conclusions using different diagnostic measures were found by Faraggi,[39] González-Manteiga et al,[40] and Inácio et al.[41,42] The posterior median (95% credible interval) of the coefficient of overlap when ignoring the age effect, obtained for a fair comparison using the OVL_{DPM} estimator, was 0.308 (0.229, 0.402) and so by pooling the glucose levels irrespectively of the age of the individuals, for those younger than 60 years old, one would be underestimating the discriminatory ability of the glucose levels. It is worth mentioning that the 95% credible interval of the unconditional overlap coefficient, for ages younger than 55 years old, is not even entirely contained within the pointwise 95% credible band of the age-specific overlap coefficient. We remark that for the numerical integration step, both for the unconditional and conditional overlap coefficient estimators, $N = 3501$ was used.

To check the fit of the underlying Dirichlet process mixture of additive normal models in each population, Figure 19 (top row) in the Supplementary Materials show the QQ-plots of the quantile residuals and as can be observed, the resulting fit is very good. We have also generated 20 000 replicate datasets from the posterior predictive distribution in each group and the kernel density estimates of 500 of these replicate datasets (randomly selected from the 20 000 available) are compared to the kernel density estimate of the glucose levels. These posterior predictive checks are shown in Figure 19 (bottom row) of the Supplementary Materials and it is evident that our model for each group is able to simulate data that is very much similar to the observed glucose levels.

7 | CONCLUDING REMARKS

We have developed Bayesian nonparametric approaches for conducting inference about the coefficient of overlap and about its covariate-specific counterpart. In addition to providing point and interval estimates in a single and integrated framework, free of restrictive parametric assumptions, our methods are computationally easy to implement and reasonably fast for the sample sizes commonly encountered in diagnostic studies. For the unconditional estimators we have proposed, the results from the simulation study demonstrated that they are a viable alternative to the current nonparametric, kernel-based, estimators of the coefficient of overlap. To the best of our knowledge, we have proposed the first estimator of the covariate-specific overlap coefficient and the simulation study also shown that our proposed estimator has the ability to recover the true functional form of the overlap coefficient even in complex data distribution scenarios.

Although there are settings where one may only be interested in ranking different diagnostic tests (as, for instance, in our data application concerned with the search of potential biomarkers for ovarian cancer) usually, once a diagnostic test proves to be accurate enough to be used in practice, a cutoff value(s) to screen subjects in practice must be determined according to some criterion. Measures based on the receiver operating characteristic curve, or its generalization, are possibly better suited for this task and therefore we regard both measures important in practice and, as we already stated in the Introduction, they should be regarded as complementary rather than competitors.

Lastly, although in this article our focus was on evaluating a test's accuracy when disease status is binary, in clinical practice, physicians often face situations that require discriminating between three or more disease stages. Usually, an intermediate transitional stage exists prior to full disease onset, with this being especially true for neurological disorders where, for instance, three common disease categories are normal cognition, mild impairment, and severe impairment or dementia. Depending on the clinical context, if pairwise analyses are of interest, then our methods can be directly applied. Such pairwise analyses would enable understanding the accuracy of a test to discriminate between each pair of disease categories. However, if interest lies in ascertaining about the discriminatory ability of a test for the three categories simultaneously, then the definition of the coefficient of overlap needs to be adapted/extended to this case. This constitutes an interesting avenue for future research.

ACKNOWLEDGEMENTS

We acknowledge the two reviewers whose comments led to an improved version of the manuscript. We are also grateful to María Xosé Rodríguez-Álvarez and Miguel de Carvalho for helpful comments and discussions.

DATA AVAILABILITY STATEMENT

The R code and the ovarian cancer and diabetes datasets are available at https://github.com/vandainacio/overlapcoefficient/.

ORCID

Vanda Inácio https://orcid.org/0000-0001-8084-1616

REFERENCES

1. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press; 2003.
2. Weitzman MS. Measures of overlap of income distributions of white and Negro families in the United States; Vol. 22, 1970; US Bureau of the Census.
3. Samawi HM, Yin J, Rochani H, Panchal V. Notes on the overlap measure as an alternative to the Youden index: how are they related? *Stat Med*. 2017;36(26):4230-4240.
4. Wang D, Tian L. Parametric methods for confidence interval estimation of overlap coefficients. *Comput Stat Data Anal*. 2017;106:12-26.
5. Franco-Pereira AM, Nakas CT, Reiser B, Carmen PM. Inference on the overlap coefficient: the binormal approach and alternatives. *Stat Methods Med Res*. 2021;30(12):2672-2684.
6. Mizuno S, Yamaguchi T, Fukushima A, Matsuyama Y, Ohashi Y. Overlap coefficient for assessing the similarity of pharmacokinetic data between ethnically different populations. *Clin Trials*. 2005;2(2):174-181.
7. Lei L, Olson K. Evaluating statistical methods to establish clinical similarity of two biologics. *J Biopharm Stat*. 2009;20(1):62-74.
8. Silva-Fortes C, Turkman MAA, Sousa L. Arrow plot: a new graphical tool for selecting up and down regulated genes and genes differentially expressed on sample subgroups. *BMC Bioinform*. 2012;13(1):1-15.
9. Giacoletti KE, Heyse J. Using proportion of similar response to evaluate correlates of protection for vaccine efficacy. *Stat Methods Med Res*. 2015;24(2):273-286.
10. Ridout MS, Linkie M. Estimating overlap of daily activity patterns from camera trap data. *J Agric Biol Environ Stat*. 2009;14(3):322-337.
11. Núñez-Antonio G, Mendoza M, Contreras-Cristán A, Gutiérrez-Peña E, Mendoza E. Bayesian nonparametric inference for the overlap of daily animal activity patterns. *Environ Ecol Stat*. 2018;25(4):471-494.
12. Anderson G, Linton O, Whang YJ. Nonparametric estimation and inference about the overlap of two distributions. *J Econometr*. 2012;171(1):1-23.
13. Pastore M, Calcagnı` A. Measuring distribution similarities between samples: a distribution-free overlapping index. *Front Psychol*. 2019;10:1089.
14. Lee WC, Hsiao CK. Alternative summary indices for the receiver operating characteristic curve. *Epidemiology*. 1996;7(6):605-611.
15. Martinez-Camblor P, Corral N, Rey C, Pascual J, Cernuda-Morollón E. Receiver operating characteristic curve generalization for non-monotone relationships. *Stat Methods Med Res*. 2017;26(1):113-123.
16. de Carvalho M, Barney BJ, Page GL. Affinity-based measures of biomarker performance evaluation. *Stat Methods Med Res*. 2020;29(3):837-853.
17. Clemons TE, Bradley EL Jr. A nonparametric measure of the overlapping coefficient. *Comput Stat Data Anal*. 2000;34(1):51-61.
18. Schmid F, Schmidt A. Nonparametric estimation of the coefficient of overlapping—Theory and empirical application. *Comput Stat Data Anal*. 2006;50(6):1583-1596.
19. Escobar MD, West M. Bayesian density estimation and inference using mixtures. *J Am Stat Assoc*. 1995;90(430):577-588.

20. Lo AY. On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann Stat*. 1984;12(1):351-357.
21. Ferguson TS. A Bayesian analysis of some nonparametric problems. *Ann Stat*. 1973;1(2):209-230.
22. Sethuraman J. A constructive definition of Dirichlet priors. *Stat Sin*. 1994;4(2):639-650.
23. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3rd ed. Boca Raton: Chapman & Hall/CRC Press; 2013.
24. Ishwaran H, James LF. Gibbs sampling methods for stick-breaking priors. *J Am Stat Assoc*. 2001;96(453):161-173.
25. Rubin DB. The Bayesian bootstrap. *Ann Stat*. 1981;9(1):130-134.
26. De Iorio M, Johnson WO, Müller P, Rosner GL. Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics*. 2009;65(3):762-771.
27. MacEachern SN. Dependent Dirichlet Processes. Unpublished Manuscript, Department of Statistics; Vol. 5, 2000; The Ohio State University.
28. Inácio V, Rodríguez-Álvarez MX. The covariate-adjusted ROC curve: the concept and its importance, review of inferential methods, and a new Bayesian estimator. *Stat Sci*. 2022.
29. Geisser S, Eddy WF. A predictive approach to model selection. *J Am Stat Assoc*. 1979;74(365):153-160.
30. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput*. 2014;24(6):997-1016.
31. Dunn PK, Smyth GK. Randomized quantile residuals. *J Comput Graph Stat*. 1996;5(3):236-244.
32. Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. *J Royal Stat Soc Ser A (Stat Soc)*. 2019;182(2):389-402.
33. Rosenberg PS. Hazard function estimation using B-splines. *Biometrics*. 1995;51:874-887.
34. Silverman BW. *Density Estimation for Statistics and Data Analysis*. Boca Raton: Chapman and Hall/CRC Press; 1986.
35. Rodríguez-Álvarez MX, Inácio V. ROCnReg: an R package for receiver operating characteristic curve inference with and without covariates. *R J*. 2021;13(1):525-555. doi:10.32614/RJ-2021-066
36. Wand MP, Jones MC. *Kernel Smoothing*. New York: Chapman & Hall; 1995.
37. Pepe MS, Longton G, Anderson GL, Schummer M. Selecting differentially expressed genes from microarray experiments. *Biometrics*. 2003;59(1):133-142.
38. Smith P, Thompson T. Correcting for confounding in analyzing receiver operating characteristic curves. *Biometr J*. 1996;38(7):857-863.
39. Faraggi D. Adjusting receiver operating characteristic curves and related indices for covariates. *J Royal Stat Soc Ser D*. 2003;52(2):179-192.
40. González-Manteiga W, Pardo-Fernández JC, Keilegom IV. ROC curves in non-parametric location-scale regression models. *Scand J Stat*. 2011;38(1):169-184.
41. de Carvalho VI, Jara A, Hanson TE, de Carvalho M. Bayesian nonparametric ROC regression modeling. *Bayesian Anal*. 2013;8(3):623-646.
42. de Carvalho VI, Jara A, Hanson TE, de Carvalho M. Nonparametric Bayesian covariate-adjusted estimation of the Youden index. *Biometrics*. 2017;73(4):1279-1288.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Inácio V, Garrido Guillén JE. Bayesian nonparametric inference for the overlap coefficient: With an application to disease diagnosis. *Statistics in Medicine*. 2022;41(20):3879-3898. doi: 10.1002/sim.9480