Received: 17 August 2017 | Revised: 18 November 2017 | Accepted: 20 December 2017

DOI: 10.1002/sim.7611

RESEARCH ARTICLE

WILEY Statistics in Medicine

Discrimination surfaces with application to region-specific brain asymmetry analysis

Gabriel Martos[1] | Miguel de Carvalho[2]

[1]Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires-CONICET, Buenos Aires, Argentina

[2]School of Mathematics, The University of Edinburgh, Edinburgh, UK

Correspondence
Miguel de Carvalho. School of Mathematics, University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Peter Guthrie Tait Road, Edinburgh, EH9 3FD.
Email: miguel.decarvalho@ed.ac.uk.

Funding information
Chilean NSF (Conicyt) , Grant/Award Number: Fondecyt 11121186; FCT (Fundação para a Ciência e a Tecnologia, Portugal) , Grant/Award Number: UID/MAT/00006/2013

Discrimination surfaces are here introduced as a diagnostic tool for localizing brain regions where discrimination between diseased and nondiseased participants is higher. To estimate discrimination surfaces, we introduce a Mann-Whitney type of statistic for random fields and present large-sample results characterizing its asymptotic behavior. Simulation results demonstrate that our estimator accurately recovers the true surface and corresponding interval of maximal discrimination. The empirical analysis suggests that in the anterior region of the brain, schizophrenic patients tend to present lower local asymmetry scores in comparison with participants in the control group.

KEYWORDS

conditional area under the curve, distance to symmetry, Mann-Whitney statistic, neuroanatomical asymmetry, receiver operating characteristic, schizophrenia

1 | INTRODUCTION

The study of brain asymmetries is of fundamental importance to understand the origins of several diseases in humans and animals; see for instance, Galaburda et al,[1] Springer and Deutsch,[2] Toga and Thompson,[3] Rogers et al,[4] and the references therein. Morphological brain asymmetries have also been studied so to assess whether they contain information that can be used to discriminate between diseased and nondiseased participants. For example, in a recent paper, Brignell et al[5] propose a Bayesian registration method for 3-D magnetic resonance images (MRIs) brain images and analyze brain-shape asymmetries by comparing schizophrenia with healthy (or control) participants. To this aim, the authors propose a global measure of symmetry that consists in the ratio between the estimated left brain volume minus the estimated right brain volume, relative to the total volume of the brain, for each patient, and find weak evidence of differences between the 2 groups: schizophrenic vs control patients. Neuroanatomical differences between schizophrenic patients and control patients have been widely discussed in other studies in the medical literature.[3,6-9] Beyond differences at a morphological level, asymmetries of the brain at a functional level have also been examined.[10] Neuroanatomical differences between diseased and nondiseased participants—rather than functional differences—will be the ones of interest in this article.

Despite the results by Brignell et al,[5] the evidence for differences in morphological asymmetries between schizophrenic and control patients is not clear cut, and for instance, in Narr et al,[11](p945) it can be read as follows:

> "Many studies have observed schizophrenia-related reductions (or reversals) in asymmetric perisylvian regions, but negative findings are common . . . ."

With the exception of Brignell et al,[5] most aforementioned studies are mainly descriptive and do not address from a rigorous statistical perspective where the location of structural differences between groups take place with higher probability.

In this article, we propose a novel statistical approach to localize specific regions on the brain where the asymmetry, measured as the difference between left and right sides in the brain, is more likely to be observed when comparing the group of schizophrenic patients with healthy controls participants. To this aim, we propose what we refer to as a *discrimination surface*, and we use it to define 2 sets, namely, *intervals of maximal discrimination* (IMDs) and *regions of over-threshold discrimination* (ROTDs). Such surfaces and corresponding sets allow us to identify regions of the brain at which left-to-right morphological differences between diseased and nondiseased participants are more likely to occur. From a conceptual viewpoint, discrimination surfaces—as formally defined in Section 3—have connections with the area under conditional ROC curves, as discussed for example by Inácio de Carvalho et al. [12,13] However, while in conditional ROC curves, the objective is often on assessing how the discrimination ability of a diagnostic test changes over a predictor; here, the goal is on searching for regions of maximum discrimination. Another object, which has connections with discrimination surfaces, is the so-called free-response receiver operating characteristic (ROC) curve[14]; yet an important distinction between the 2 paradigms is that discrimination surfaces deliver as output a suspected location, whereas free-response ROC curve take as input suspected locations—along with a rating on the level of suspicion. Brain curves are modeled here as a functional data analysis[15-17] object and are used to generate a random field of local asymmetry scores. To estimate discrimination surfaces, we propose what we call as *empirical discrimination surfaces*, which can be regarded as a Mann-Whitney type of statistic for random fields, in the sense that such empirical surfaces consist of pointwise Mann-Whitney estimates. Some large-sample results characterizing the asymptotic behavior of our methods are derived, and in particular, we show that under mild conditions, the empirical surface is strongly uniformly consistent and that the corresponding IMD is weakly consistent. A smoothed version of the estimator is also discussed.

The article is organized as follows. In Section 2, we describe the data, from a schizophrenia study, that motivate the paper. In Section 3, we present the proposed discrimination surface-based methods. In Section 4, we assess the finite-sample performance of the proposed estimator in a simulation study. In Section 5, we analyze the data from the above-mentioned schizophrenia study. We conclude in Section 6. Proofs are included in Appendices A.1 to A.3.

2 | DESCRIPTION OF STUDY DATA

The raw data consist of $n = 68$ 3-D MRI, gathered from a neuroscience study conducted at the University of British Columbia, Canada, and documented in Brignell et al[5]; the study involved $n_D = 30$ schizophrenic patients and $n_{\bar{D}} = 38$ healthy controls. Some comments on the geometry of the data are in order. Following common practice in neuroscience, each brain was registered into the so-called *Talairach space*[18]—so that brains can be compared on the same 3-dimensional referential coordinate space. The x-axis is created through the identification of 2 landmarks: the anterior commissure and the posterior commissure, with the anterior commissure being the origin $(0, 0, 0)$; the remainder axes are the y-axis: inferior → superior (bottom to top) and the z-axis: right → left ("left" = patient's left). Of main interest is the so-called *axial plane*, which corresponds to the x-z plane. In Figure 1A, we show the raw data corresponding to the x-z plane registered by the MRI scanner with $m = 500(x, z)$ coordinates on each participant; the other relevant perspectives, not shown here, are the sagittal plane (x-y) and the coronal plane (y-z). These data are in nature continuous; therefore, the framework of functional data analysis[15-17] is natural to model brain sections on the axial plane. The raw data are available from the R package `shapes`.[19]

Such data have been converted into functional data using a reproducing kernel Hilbert space approach discussed in the Supporting Information. After transforming the raw data into functions, we obtain the brain curves $\tilde{B}_i : [0, 2\pi] \to \mathbb{R}^2$ for $i = 1, \ldots, n$ shown in Figure 1B.

Following Brignell et al,[5] we assume the midline plane of the brain as the line of symmetry; therefore, we decompose the brain curves into left ($\tilde{B}_{\mathrm{L}}$) and right ($\tilde{B}_{\mathrm{R}}$) as follows:

$$\tilde{B}_{\mathrm{L}}(\theta) = \begin{cases} \tilde{B}(\theta), & \text{if } \theta \in (\frac{\pi}{2}, \frac{3\pi}{2}], \\ 0, & \text{otherwise,} \end{cases}$$

$$\tilde{B}_{\mathrm{R}}(\theta) = \begin{cases} \tilde{B}(\theta), & \text{if } \theta \in (-\frac{\pi}{2}, \frac{\pi}{2}], \\ 0, & \text{otherwise.} \end{cases}$$

It is clear that $\tilde{B}(\theta) = \tilde{B}_{\mathrm{L}}(\theta) + \tilde{B}_{\mathrm{R}}(\theta)$ from the previous decomposition. To analyze the brain asymmetries with respect to the midline plane, we rotate and reflect the functions $\tilde{B}_{\mathrm{L}}$ and $\tilde{B}_{\mathrm{R}}$ and also rescale the domain of both functions to the interval $[0, 1]$. In Figure 1C,D, we respectively represent left and right brain curves after these transformations; we denote these curves throughout as $B_{\mathrm{L}} : [0, 1] \to L \subset [0, \infty)$ and $B_{\mathrm{R}} : [0, 1] \to R \subset [0, \infty)$.

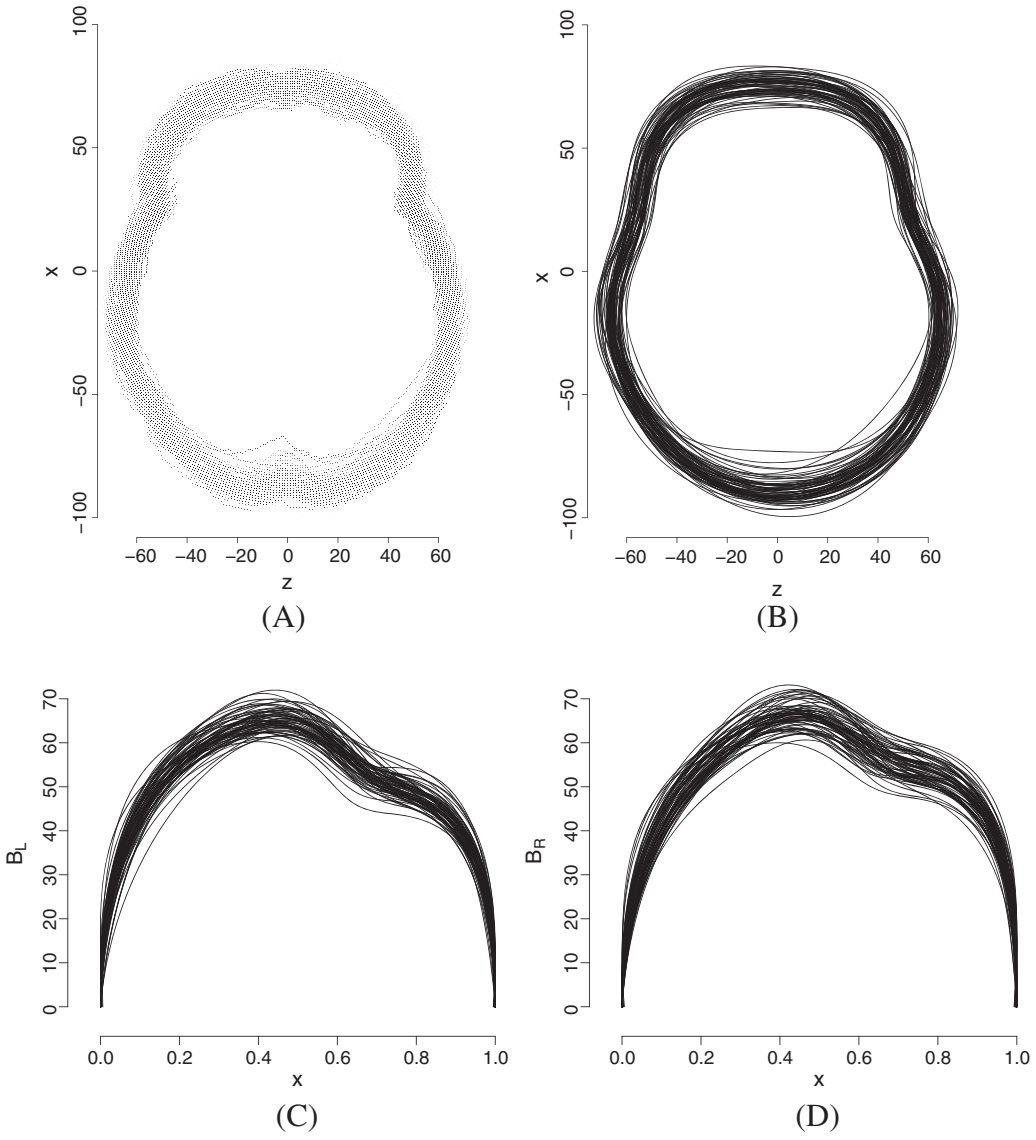

FIGURE 1 A, Brain sections on the axial plane, for healthy and diseased participants; x-axis: posterior → anterior (back to front); z-axis: right → left ("left" = patient's left). B, Brain curves $\tilde{B} : [0, 2\pi] \to \mathbb{R}^2$. C, Left $B_{\mathrm{L}} : [0,1] \to L \subset [0,\infty)$. D, Right $B_{\mathrm{R}} : [0,1] \to R \subset [0,\infty)$ brain curves after suitable rotations, reflections, and domain rescaling

3 | DISCRIMINATION SURFACES

3.1 | Distance to symmetry

Denote by D and $\bar{D}$ the population of diseased and nondiseased participants, respectively, then, for pairs of functional brain curves $(B_{\mathrm{L},D}, B_{\mathrm{R},D})$ and $(B_{\mathrm{L},\bar{D}}, B_{\mathrm{R},\bar{D}})$, we define the distance to local symmetry as follows:

$$Y_D(\boldsymbol{t}) = \int_a^b \{B_{\mathrm{L},D}(x) - B_{\mathrm{R},D}(x)\}^2 \,\mathrm{d}x, \quad Y_{\bar{D}}(\boldsymbol{t}) = \int_a^b \{B_{\mathrm{L},\bar{D}}(x) - B_{\mathrm{R},\bar{D}}(x)\}^2 \,\mathrm{d}x, \tag{1}$$

for $\boldsymbol{t} = (a, b) \in T$, where

$$T = \{(a, b) \in [0,1]^2 : 0 \leqslant a < b \leqslant 1\}. \tag{2}$$

The scores in Equation 1 should be interpreted as measures of distance to symmetry in the region $[a, b]$, and for this reason, we will refer to these as *local asymmetry scores*; indeed, for a completely symmetric brain, $B_{\mathrm{L}} = B_{\mathrm{R}}$, it holds that $Y(\boldsymbol{t}) = 0$, for all $\boldsymbol{t} \in T$. The larger the value of $Y(\boldsymbol{t}) = \int_a^b \{B_{\mathrm{L}}(x) - B_{\mathrm{R}}(x)\}^2 \,\mathrm{d}x$, the less symmetric the brain should be on the region parametrized by $\boldsymbol{t}$. We refer to a and b as *localization parameters*.

3.2 | Discrimination surfaces, IMD, and ROTD

Using the local asymmetry scores $Y_D(\boldsymbol{t})$ and $Y_{\bar{D}}(\boldsymbol{t})$ from Equation 1, we define the discrimination surface as

$$\Lambda(\boldsymbol{t}) = P\{Y_D(\boldsymbol{t}) > Y_{\bar{D}}(\boldsymbol{t})\}, \quad \boldsymbol{t} \in T. \tag{3}$$

Such surfaces can be used for assessing discrimination between the local asymmetry scores of diseased and nondiseased participants by appraising how likely it is for $Y_D(\boldsymbol{t})$ to be larger than $Y_{\bar{D}}(\boldsymbol{t})$, over the regions of the brain corresponding to the interval $(a, b) = \boldsymbol{t} \in T$. For the purpose of consistency with existing literature, the definition in Equation 3 follows the standard convention in medical diagnostic statistics—that a higher value of the "marker" $Y_D(\boldsymbol{t})$ would be more indicative of disease—but all concepts in the paper can be readily adapted to the setting where a higher value of $Y_D(\boldsymbol{t})$ is less indicative of disease.

Another graphical device that can be used to summarize information on the discrimination ability over different intervals is the *discrimination contour*, which is defined as $\lambda_u = \{\boldsymbol{t} : \Lambda(\boldsymbol{t}) = u\}$, for $u \in (0, 1)$. We can interpret λ_u as the sets of intervals with the same discrimination power.

Example 1. (Discrimination surfaces for lognormal-distributed asymmetry scores)
Consider the location and scale functions $\mu_\cdot : [0,1]^2 \to \mathbb{R}$ and $\sigma_\cdot : [0,1]^2 \to [0, \infty)$, with $\mu_D(\boldsymbol{t}) > \mu_{\bar{D}}(\boldsymbol{t})$. Let $Y_D(\boldsymbol{t}) \sim \text{LN}(\mu_D(\boldsymbol{t}), \sigma^2_D(\boldsymbol{t}))$ and $Y_{\bar{D}}(\boldsymbol{t}) \sim \text{LN}(\mu_{\bar{D}}(\boldsymbol{t}), \sigma^2_{\bar{D}}(\boldsymbol{t}))$ be 2 independent and lognormal distributed asymmetry scores. Then,

$$\Lambda(\boldsymbol{t}) = P\{Y_D(\boldsymbol{t}) > Y_{\bar{D}}(\boldsymbol{t})\} = \Phi\left(\frac{\alpha(\boldsymbol{t})}{\{1 + \beta^2(\boldsymbol{t})\}^{1/2}}\right), \quad \boldsymbol{t} \in T,$$

where Φ is the cumulative distribution function of the standard normal distribution, $\alpha(\boldsymbol{t}) = \{\mu_D(\boldsymbol{t}) - \mu_{\bar{D}}(\boldsymbol{t})\}/\sigma_{\bar{D}}(\boldsymbol{t})$ and $\beta^2(\boldsymbol{t}) = \sigma_D(\boldsymbol{t})/\sigma_{\bar{D}}(\boldsymbol{t})$. The so-called *binormal model* is a particular case of the current model, and it corresponds to the case $\mu_D(\boldsymbol{t}) = \mu_D$, $\mu_{\bar{D}}(\boldsymbol{t}) = \mu_{\bar{D}}$, $\sigma_D(\boldsymbol{t}) = \sigma_D$, $\sigma_{\bar{D}}(\boldsymbol{t}) = \sigma_{\bar{D}}$, with $\mu_D(\boldsymbol{t}) > \mu_{\bar{D}}(\boldsymbol{t})$ and $\sigma_D, \sigma_{\bar{D}} > 0$; cf section 4.4 in Pepe.[20] In Figure 2, we plot an example of a bi-lognormal discrimination surface and corresponding discrimination contours with $\sigma_D(\boldsymbol{t}) = \sigma_{\bar{D}}(\boldsymbol{t}) = 1$, $\mu_{\bar{D}}(t) = 0$, and $\mu_D(\boldsymbol{t}) = 1 - 10(b - 3/4)^2 - 10(a - 1/2)^2$, for $0 \leqslant a < b \leqslant 1$; note that in the latter specification, $\alpha(\boldsymbol{t}) = \mu_D(\boldsymbol{t})$ and $\beta(\boldsymbol{t}) = 1$.

In practice, one may also be interested in assessing on what regions discrimination is above a certain threshold and this leads us to define the ROTD:

$$\text{ROTD}_u = \{\boldsymbol{t} : \Lambda(\boldsymbol{t}) \geqslant u \text{ or } 1 - \Lambda(\boldsymbol{t}) \geqslant u\}, \quad u \in (0, 1). \tag{4}$$

Discrimination surfaces—as defined above—have connections with the area under conditional ROC curves,[12,13] defined as $\text{AUC}(\boldsymbol{x}) = \int_0^1 \text{ROC}(p|\boldsymbol{x})\text{d}p$, where $\text{ROC}(p|\boldsymbol{x}) = 1 - F_D(F^{-1}_{\bar{D}}(1 - p|\boldsymbol{x})|\boldsymbol{x})$, with $\boldsymbol{x}$ in $\mathbb{R}^p$ being a covariate. Yet as it can

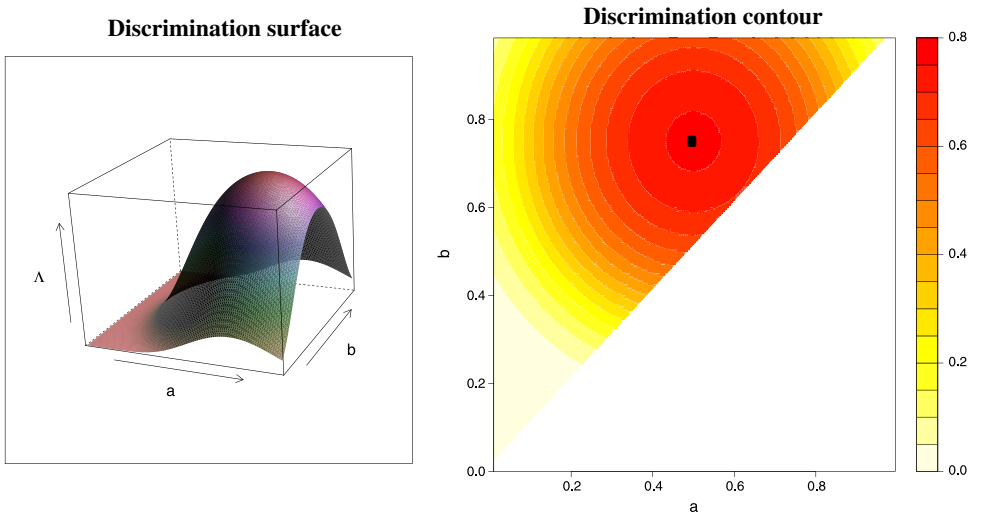

FIGURE 2 (Left) Bi-lognormal discrimination surface and corresponding discrimination contour (right) using the configuration of parameters in Example 1. The solid black dot (•) represents the coordinates $\boldsymbol{t}^* = (a^*, b^*) = (1/2, 3/4)$ underlying the interval of maximal discrimination, IMD = $[a^*, b^*]$ [Colour figure can be viewed at wileyonlinelibrary.com]

be seen from Equation 4 (and as it will be seen from Equation 5), here, the target is on seeking for regions of maximum discrimination and not simply on assessing how discrimination ability changes over a covariate. Also, while in a typical conditional ROC curve setting, data are of the types $\{(X_{D,i}, Y_{D,i})\}_{i=1}^{n_D}$ and $\{(X_{\bar{D},j}, Y_{\bar{D},j})\}_{j=1}^{n_{\bar{D}}}$ in our context data that consist of the random fields of asymmetry scores, that is,

$$\{Y_{D,i}(\boldsymbol{t}) : \boldsymbol{t} \in T\}_{i=1}^{n_D}, \quad \{Y_{\bar{D},j}(\boldsymbol{t}) : \boldsymbol{t} \in T\}_{j=1}^{n_{\bar{D}}},$$

which we compute using $\{(B_{L,D,i}, B_{R,D,i})\}_{i=1}^{n_D}$ and $\{(B_{L,\bar{D},j}, B_{R,\bar{D},j})\}_{j=1}^{n_D}$.

Since the goal is to localize brain regions where schizophrenic patients differ further from healthy controls. Then, the regions of interest are those where one of the random variables is stochastically greater than the other with maximum probability. Therefore, let $\boldsymbol{t}^+ := \arg\max_{\boldsymbol{t}\in T}\Lambda(\boldsymbol{t})$ and $\boldsymbol{t}^- := \arg\max_{\boldsymbol{t}\in T} 1-\Lambda(\boldsymbol{t}) = \arg\min_{\boldsymbol{t}\in T}\Lambda(\boldsymbol{t})$; both are well defined if we assume $\Lambda(\boldsymbol{t})$ is a uniformly continuous surface (note that T in Equation 2 is a compact set); then, the region of interest is determined by $\boldsymbol{t}^* := \boldsymbol{t}^+$ if $\Lambda(\boldsymbol{t}^+) > 1-\Lambda(\boldsymbol{t}^-)$ or $\boldsymbol{t}^* := \boldsymbol{t}^-$, otherwise. In this way, we define the IMD as

$$\text{IMD} = [a^*, b^*], \tag{5}$$

where the limits of this interval are defined through the elements in the vector $\boldsymbol{t}^* = (a^*, b^*)$.

3.3 | Estimating discrimination surfaces and their functionals

We start by noting that the trajectories of the random fields of asymmetry scores, as defined in Equation 1, live in $\mathscr{Y}$, the space of all nonnegative and differentiable random functions on T such that $\partial Y(\boldsymbol{t})/\partial a \leqslant 0$ and $\partial Y(\boldsymbol{t})/\partial b \geqslant 0$, a.s. Indeed, it holds that

$$\begin{cases} \frac{\partial Y}{\partial a} = \frac{\partial}{\partial a}\left(\int_a^b \{B_{\text{L}}(x) - B_{\text{R}}(x)\}^2 \,\text{d}x\right) = -\{B_{\text{L}}(a) - B_{\text{R}}(a)\}^2 \leqslant 0, \\ \frac{\partial Y}{\partial b} = \frac{\partial}{\partial b}\left(\int_a^b \{B_{\text{L}}(x) - B_{\text{R}}(x)\}^2 \,\text{d}x\right) = \quad \{B_{\text{L}}(b) - B_{\text{R}}(b)\}^2 \geqslant 0. \end{cases} \tag{6}$$

Keeping in mind the applied context under analysis—and to rule out uninteresting cases from a morphological perspective—we further assume that random functions in $\mathscr{Y}$ are "bounded," in the sense that $P\{Y(\boldsymbol{t}) \in S\} = 1$, with $S = [0, M]$, for some possibly large but finite $M > 0$, for all $\boldsymbol{t} \in T$.

Let $F_{\boldsymbol{t}}(y) = P\{Y(\boldsymbol{t}) \leqslant y\}$, and define the marginal empirical distribution function as

$$\hat{F}_{\boldsymbol{t}}(y) = \frac{1}{n}\sum_{i=1}^{n} I\{Y_i(\boldsymbol{t}) \leqslant y\}, \quad y \in S, \quad \boldsymbol{t} = (a, b) \in T,$$

where $I\{\cdot\}$ is the indicator function and $Y_1(\boldsymbol{t}), \ldots, Y_n(\boldsymbol{t})$ is a sequence of independent identically distributed random functions in $\mathscr{Y}$. Given that for each fixed $\boldsymbol{t}$, $F_{\boldsymbol{t}}(y)$ is a distribution function, the standard Glivenko-Cantelli theorem implies that, as $n \to \infty$, it holds that

$$\sup_y |\hat{F}_{\boldsymbol{t}}(y) - F_{\boldsymbol{t}}(y)| = o(1), \quad \text{a.s.} \tag{7}$$

Yet, since the trajectories of asymmetry scores live in $\mathscr{Y}$, a stronger result actually holds for our setting. Indeed, $F_{(\cdot,\cdot)}(y)$ is nondecreasing, and as a consequence of Equation 6, it follows that for $0 \leqslant a_0 \leqslant a_1 < b \leqslant 1$, then,

$$P\{Y(a_1, b) \leqslant y | Y(a_0, b) \leqslant y\} = 1 \text{ and } P\{Y(a_0, b) \leqslant y | Y(a_1, b) \leqslant y\} \leqslant 1. \tag{8}$$

Therefore, by evaluating the ratio of the 2 conditional probabilities in Equation 8, we obtain that $P\{Y(a_0, b) \leqslant y\} \leqslant P\{Y(a_1, b) \leqslant y\}$, that is, $F_{(a,\cdot)}(\cdot)$ is nondecreasing; the same reasoning can be used to verify that $F_{(\cdot,b)}(\cdot)$ is nonincreasing. The same monotonicity properties hold for the marginal empirical distribution function $\hat{F}_{\boldsymbol{t}}(y)$. Thus, monotonicity on each dimension of $F_{\boldsymbol{t}}(y)$ and $\hat{F}_{\boldsymbol{t}}(y)$, along with the extra assumption that the true $F_{\boldsymbol{t}}(y)$ is continuous, allows us to go beyond and state the following generalization of the result presented in Equation 7.

Theorem 1. *Let $Y_1(\boldsymbol{t}), \ldots, Y_n(\boldsymbol{t})$ be a sequence of independent identically distributed random functions in $\mathscr{Y}$. Suppose $F_{\boldsymbol{t}}(y)$ is continuous for all $(y, \boldsymbol{t})$ in $S \times T$. Then, as $n \to \infty$, it holds that*

$$\sup_{(y,\boldsymbol{t})} |\hat{F}_{\boldsymbol{t}}(y) - F_{\boldsymbol{t}}(y)| = o(1), \quad a.s.$$

Given a random sample of size $n = n_D + n_{\bar{D}}$ of random fields of asymmetry scores from diseased and nondiseased participants, define the empirical discrimination surface as

$$\hat{\Lambda}(\boldsymbol{t}) = \frac{1}{n_D n_{\bar{D}}} \sum_{i=1}^{n_D} \sum_{j=1}^{n_{\bar{D}}} I\{Y_{Di}(\boldsymbol{t}) > Y_{\bar{D}j}(\boldsymbol{t})\}. \tag{9}$$

Here, $Y_{Di}(\boldsymbol{t})$ is the distance to symmetry score, as defined in Equation 1, for the ith individual in the sample of diseased patients, and $Y_{\bar{D}j}(\boldsymbol{t})$ is the distance to symmetry score of the jth nondiseased participant in the control group.

Let $\widehat{\boldsymbol{t}}^{+} := \arg\max_{\boldsymbol{t}\in T}\hat{\Lambda}(\boldsymbol{t})$ and $\widehat{\boldsymbol{t}}^{-} := \arg\max_{\boldsymbol{t}\in T} 1 - \hat{\Lambda}(\boldsymbol{t})$; then, we define the estimated optimal localization vector parameter $\widehat{\boldsymbol{t}}^{*} := \widehat{\boldsymbol{t}}^{+}$ if $\hat{\Lambda}(\widehat{\boldsymbol{t}}^{+}) > 1 - \hat{\Lambda}(\widehat{\boldsymbol{t}}^{-})$ and $\widehat{\boldsymbol{t}}^{*} := \widehat{\boldsymbol{t}}^{-}$, otherwise. The corresponding interval of maximum discrimination estimate consists of a random interval, $\widehat{\text{IMD}} = [\hat{a}^{*}, \hat{b}^{*}]$, where the limits of this random interval are defined through the elements in the vector $\widehat{\boldsymbol{t}}^{*} = (\hat{a}^{*}, \hat{b}^{*})$.

Empirical ROTD can be obtained by setting a high level of discrimination $u \in (0, 1)$ and consist of the following random subset of T:

$$\widehat{\text{ROTD}}_u = \{\boldsymbol{t} : \hat{\Lambda}(\boldsymbol{t}) \geqslant u \,\text{or}\, 1 - \hat{\Lambda}(\boldsymbol{t}) \geqslant u\}. \tag{10}$$

Some comments are in order. For each $\boldsymbol{t} \in T$, our estimator in Equation 9 is a Mann-Whitney type of statistic, and thus, it is straightforward to show that it is pointwisely unbiased and pointwisely weakly consistent, ie,

$$E\{\hat{\Lambda}(\boldsymbol{t})\} = \Lambda(\boldsymbol{t}), \quad \hat{\Lambda}(\boldsymbol{t}) - \Lambda(\boldsymbol{t}) = o_{\text{p}}(1), \tag{11}$$

assuming $n_D/n \to \rho_D \in (0, 1)$ and $n_{\bar{D}}/n \to \rho_{\bar{D}} \in (0, 1)$, with $\rho_D + \rho_{\bar{D}} = 1$. Beyond such straightforward statements, more can actually be said by taking advantage of our extended Glivenko-Cantelli theorem (Theorem 1 above).

In the theoretical considerations made below, we work under the following assumptions:

(A1) Suppose $Y_D(\boldsymbol{t})$ and $Y_{\bar{D}}(\boldsymbol{t})$ are in $\mathscr{Y}$, the space of all nonnegative and differentiable random functions on T, such that $\partial Y(\boldsymbol{t})/\partial a \leqslant 0$ and $\partial Y(\boldsymbol{t})/\partial b \geqslant 0$, and which are supported over $S = [0, M]$, for some $M > 0$, for every $\boldsymbol{t} \in T = \{(a, b) \in [0, 1]^2 : 0 \leqslant a < b \leqslant 1\}$.
(A2) Suppose $F_{D\boldsymbol{t}}(y) = P\{Y_D(\boldsymbol{t}) \leqslant y\}$ and $F_{\bar{D}\boldsymbol{t}}(y) = P\{Y_{\bar{D}}(\boldsymbol{t}) \leqslant y\}$ are continuous in $T \times S$ and strictly increasing in $y \in S$.
(A3) Let $Y_{D,1}(\boldsymbol{t}), \ldots, Y_{D,n_D}(\boldsymbol{t})$ and $Y_{\bar{D},1}(\boldsymbol{t}), \ldots, Y_{\bar{D},n_{\bar{D}}}(\boldsymbol{t})$ be 2 independent sequences of independent identically distributed random functions in $\mathscr{Y}$.
(A4) Suppose that as $n \to \infty$, it holds that $n_D/n \to \rho_D \in (0, 1)$ and $n_{\bar{D}}/n \to \rho_{\bar{D}} \in (0, 1)$, with $\rho_D + \rho_{\bar{D}} = 1$.

Here, A1 and A2 are regularity conditions on the true data-generating process, whereas A3 and A4 are conditions about the way we sample from such process. For identification reasons, throughout, we assume that $\Lambda(\boldsymbol{t})$ is uniquely maximized at $\boldsymbol{t}^{*}$; this plays no role on the consistency of $\hat{\Lambda}(\boldsymbol{t})$. The following results hold.

Theorem 2. *Suppose assumptions A1 to A4 hold. Let $\Lambda(\boldsymbol{t})$ be a discrimination surface and $\hat{\Lambda}(\boldsymbol{t})$ be its empirical discrimination surface; let $\boldsymbol{t}^{*} = (a^{*}, b^{*}) \in T$ be the pair of points delimiting the IMD and $\widehat{\boldsymbol{t}}^{*} \in T$ be their corresponding estimates. Then, as $n \to \infty$, it holds that*

1. $\sup_{\boldsymbol{t}}|\hat{\Lambda}(\boldsymbol{t}) - \Lambda(\boldsymbol{t})| = o(1)$, *a.s.*,
2. $\widehat{\boldsymbol{t}}^{*} - \boldsymbol{t}^{*} = o_{\text{p}}(1)$.

In addition to the empirical estimate of the discrimination surface, smooth estimations can be obtained by considering the following kernel-based version of our estimator in Equation 9:

$$\tilde{\Lambda}(\boldsymbol{t}) = \sum_{l=1}^{L} K_H(\boldsymbol{t} - \boldsymbol{t}')\hat{\Lambda}(\boldsymbol{t}), \tag{12}$$

where L is the length of the grid over which we smooth the empirical discrimination surface and H is a positive definite bandwidth matrix. In practice, both L and H act as smoothing parameters in the sense that if we increase L and fix H, we should expect to get smoother estimates.

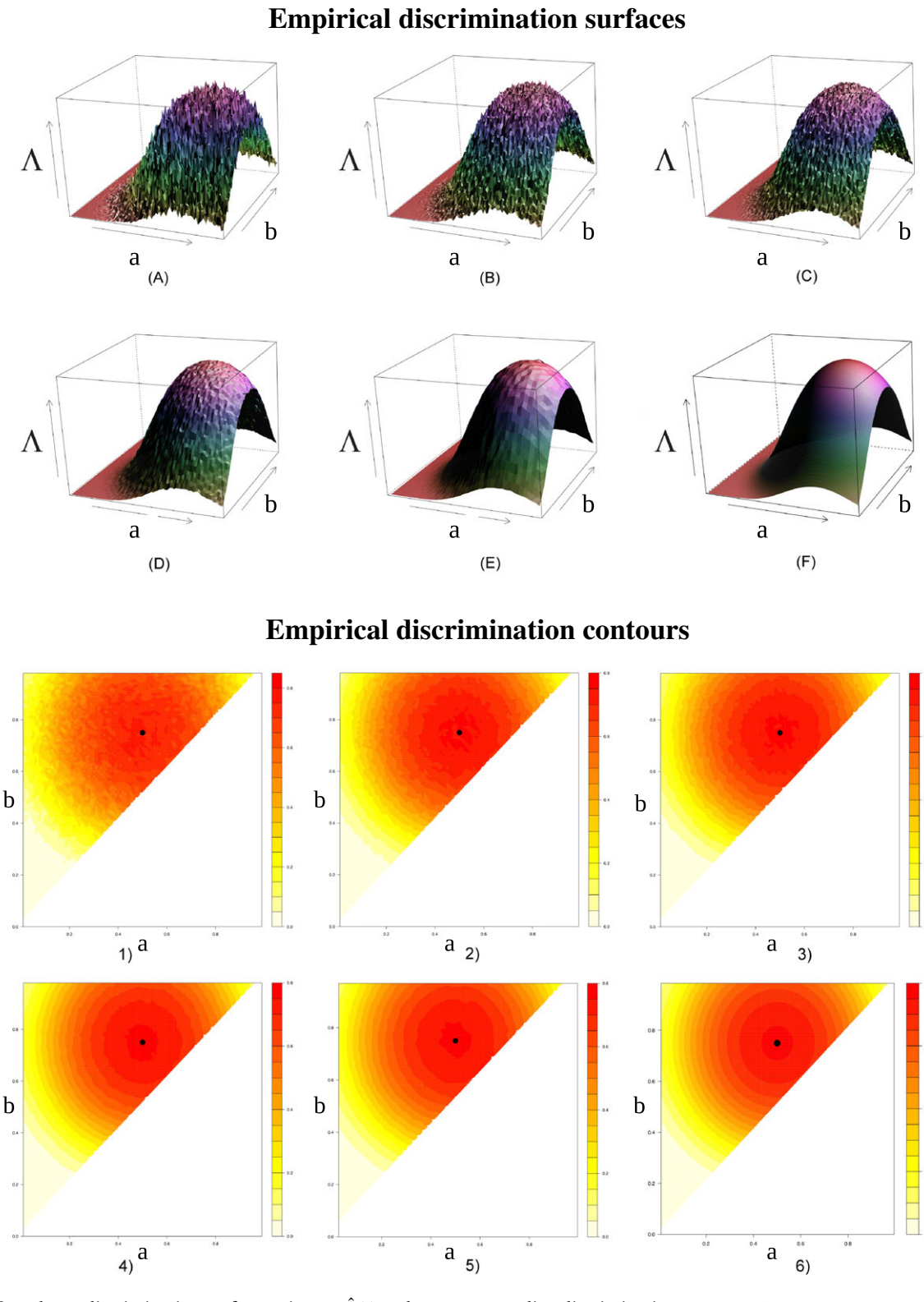

FIGURE 3 Above: discrimination surface estimates $\hat{\Lambda}(\boldsymbol{t})$. Below: corresponding discrimination contours. A, $n = 200$; B, $n = 1000$; C, $n = 2000$; D, $n = 10000$; E, $n = 20000$; F, True $\Lambda(\boldsymbol{t})$ and λ_c. In F, the solid black dot (•) represents the coordinates $\boldsymbol{t}^+ = (a^+, b^+) = (1/2, 3/4)$ underlying the interval of maximal discrimination, IMD $= [a^+, b^+]$, whereas in A to E, solid black dot (•) denotes its corresponding estimates [Colour figure can be viewed at wileyonlinelibrary.com]

4 | SIMULATION STUDY

4.1 | Preliminary experiments and preparations for Monte Carlo study

In this section, we illustrate numerically the proposed methodology to estimate discrimination surfaces and corresponding IMD. To this aim, we use as a reference framework the bi-lognormal model introduced in Example 1. A Monte Carlo study will be reported in Section 4.2.

First, we simulate random samples of Y_D and $Y_{\bar{D}}$, the lognormal-distributed asymmetry scores corresponding to the Example 1, with sample size $n = 200, 1000, 2000, 10000$, and 20000 assuming that $\rho_D = \rho_{\bar{D}} = 0.5$. The estimated discrimination surfaces $\hat{\Lambda}(\boldsymbol{t})$ along with the corresponding discrimination contours $\hat{\lambda}_u$ are presented in Figure 3. The results of this single-run experiment, suggest a fast convergence of the empirical discrimination surface, along with that of the corresponding IMD. To evaluate the uncertainty around the estimation of the IMD in this single-run experiment, we compute a bootstrap confidence region (CR) for the IMD. For that, we use the data relative to the sample size $n = 200$ and generate $B = 1000$ bootstrap samples from the original data. The bootstrap α-CRs, based on $\hat{P}(\text{IMD} \in \text{CR}_\alpha) = 1 - \alpha$ for $\alpha \in [0, 1]$, are reported in Figure 4, and for 95% confidence level, the estimated region contains the true localization parameter, $\boldsymbol{t}^+ = (a^+, b^+) = (1/2, 3/4)$. Coverage of the bootstrap will be assessed in the Monte Carlo study from Section 4.2.

4.2 | Monte Carlo study

In this Monte Carlo study, we investigate the performance of our estimator in Equation 9 over a variety of sample sizes and over 2 data configurations (Scenarios A and B). Specifically, we consider the following data generating processes based on the bi-lognormal model from Example 1, where $\sigma_D(\boldsymbol{t}) = \sigma_{\bar{D}}(\boldsymbol{t}) = 1$ and

$$\mu_{\bar{D}}(\boldsymbol{t}) = 0, \quad \mu_D(\boldsymbol{t}) = 1 - 10(b - 3/4)^2 - 10(a - 1/2)^2, \quad 0 \leqslant a < b \leqslant 1. \tag{13}$$

Namely, we consider the following:

- Scenario A: $Y_D(\boldsymbol{t}) \sim \text{N}(\mu_D(\boldsymbol{t}), 1)$, with $\mu_D(\boldsymbol{t})$ as in (13), and $Y_{\bar{D}}(\boldsymbol{t}) \sim \text{N}(0, 1)$, with $Y_D(\boldsymbol{t})$ independent of $Y_D(\boldsymbol{t}')$ and $Y_{\bar{D}}(\boldsymbol{t})$ independent of $Y_{\bar{D}}(\boldsymbol{t}')$, for $\boldsymbol{t} \neq \boldsymbol{t}'$.
- Scenario B: $Y_D(\boldsymbol{t})$ and $Y_{\bar{D}}(\boldsymbol{t})$ are Gaussian random fields with respective means $\mu_D(\boldsymbol{t})$ as in (13) and $\mu_{\bar{D}}(\boldsymbol{t}) = 0$, and with Matérn covariance function, $\gamma(h) = \sigma^2 2^{1-\kappa}/\Gamma(\kappa)(h/\varphi)^\kappa K_\kappa(h/\varphi)$, for $h > 0$, where $\sigma = 1$, $\varphi = 2$, and $\kappa = 0.5$ are the sill, range, and shape parameters, Γ is the gamma function, and with K_κ denoting the modified Bessel function of order κ.

Scenario A is thus the one we already used to produce the normal distributed asymmetry scores for the 1-shot experiment in Section 4.1. We continue assuming $\rho_D = 0.5$ and consider again the sample size $n = 200, 1000, 2000, 10000$, and 20 000. For each sample size, we simulate $M = 1000$ datasets for each of the Scenarios A and B above. With each dataset, we estimate the discrimination surface and the corresponding IMD; then, we compute the mean integrated squared error (MISE) and mean squared error (MSE) as follows:

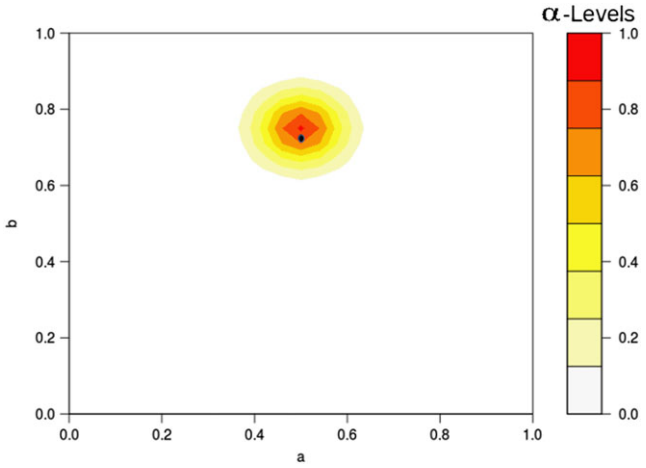

FIGURE 4 Bootstrap confidence regions along with true localization parameter $\boldsymbol{t}^+ = (a^+, b^+) = (1/2, 3/4)$, in a solid dot (•), at different confidence levels [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Mean integrated squared error (MISE) estimates and mean squared error (MSE) (both on a scale of 10^3), respectively for Λ and $\boldsymbol{t}$, over different sample sizes for Scenarios A and B as defined in Section 4.2. The standard-error is reported in parenthesis

		Sample size				
Scenario	**Parameter**	**200**	**1000**	**2000**	**10000**	**20000**
A	Λ	3.833	0.765	0.380	0.076	0.038
		(0.230)	(0.047)	(0.024)	(0.005)	(0.002)
	$\boldsymbol{t}$	5.766	2.761	2.014	0.963	0.739
		(5.320)	(2.482)	(1.787)	(0.874)	(0.680)
B	Λ	4.239	0.790	0.396	0.076	0.040
		(4.666)	(0.843)	(0.429)	(0.088)	(0.044)
	$\boldsymbol{t}$	2.137	0.761	0.493	0.217	0.191
		(2.002)	(0.730)	(0.463)	(0.135)	(0.088)

$$\mathrm{MISE}^n_{\Lambda} = \frac{1}{M}\sum_{m=1}^{M}\int_T \{\widehat{\Lambda}_m(\boldsymbol{t}) - \Lambda(\boldsymbol{t})\}^2 \mathrm{d}\boldsymbol{t}, \quad \mathrm{MSE}^n_{\boldsymbol{t}^+} = \frac{1}{M}\sum_{m=1}^{M}\{(\hat{a}^+_m - a^+)^2 + (\hat{b}^+_m - b^+)^2\},$$

where $\widehat{\Lambda}_m(\boldsymbol{t})$ is the estimate produced using the mth simulated dataset, with $\hat{\boldsymbol{t}}^+_m = (\hat{a}^+_m, \hat{b}^+_m)$ denoting its corresponding estimated localization parameters.

Table 1 summarizes our Monte Carlo simulation study. As expected, MISE decreases as the sample size increases, both in terms of the MISE associated with $\hat{\Lambda}$ and $\hat{\boldsymbol{t}}^+$. However, it can be seen that in Scenario B, the estimates of $\boldsymbol{t}^+$ are more accurate for all sample sizes. This is due to the dependence structure introduced in the data-generating process in Scenario B, that leads to a reduction of the variability associated to $\hat{\boldsymbol{t}}^+$. Note that $\mathrm{MISE}^n_{\Lambda}$ looks proportional to $1/n$, while $\mathrm{MSE}^n_{\boldsymbol{t}^+}$ seems to converge more slowly. This due to the fact the estimation of IMD is a challenging one, as the empirical discrimination surface is only smooth in the limit. We have also conducted some Monte Carlo experiments, with $n = 200$, so to assess coverage probability of the bootstrap CR for $\boldsymbol{t}^+$ in Scenarios A and B. The 99% and 95% coverage probabilities in Scenario A are .999 and .955, respectively. In Scenario B, the 99% and 95% coverage probabilities are 0.987 and 0.945, respectively. Thus, bootstrap CRs achieve nominal coverage in both scenarios.

5 | BRAIN ASYMMETRY ANALYSIS IN SCHIZOPHRENIC PATIENTS

5.1 | Context underlying the analysis

We now apply our methods to the brain curves from the study in Section 2. It has been suggested that schizophrenic patients may tend to have more symmetric brains than controls in the anterior and perisilvian regions.[11,21] This result is controversial,[3,7,8] and we show in this section that our methods can be used to localize brain regions where with maximum probability schizophrenic patients may differ further from healthy controls.

5.2 | Discrimination surfaces and regions of maximal discrimination: region-specific analysis

We estimate the empirical discrimination surface by using the asymmetry scores introduced in Equation 1. In Figure 5, we show the discrimination surfaces estimated using the methods proposed in Equations 9 and 12: At the top of the figure, we show $\hat{\Lambda}(\boldsymbol{t})$ on the left and $\tilde{\Lambda}(\boldsymbol{t})$ on the right, and $1 - \hat{\Lambda}(\boldsymbol{t})$ and $1 - \tilde{\Lambda}(\boldsymbol{t})$ on the bottom (left and right, respectively). The smoothed empirical discrimination surfaces corresponds to the smoothing parameters $L = 2775$ and $H = 0.1^3 I_{2\times 2}$. Following Marron,[22](p533) we conducted inference over a wide range of bandwidth matrices and grid sizes as a way to assess the sensitivity and reliability of the inference to the smoothing parameters L and H. The results do not differ substantially from the ones documented here for L and H close to the values selected to conduct the analysis.

In most regions of the brain, as parametrized by $\boldsymbol{t}$, one finds that $\hat{\Lambda}(\boldsymbol{t})$ is lower than $1 - \hat{\Lambda}(\boldsymbol{t})$; cf Figure 5; in particular, $\max_{\boldsymbol{t}\in T}\hat{\Lambda}(\boldsymbol{t}) = \hat{\Lambda}(\hat{\boldsymbol{t}}^+) \leq \max_{\boldsymbol{t}\in T} 1 - \hat{\Lambda}(\boldsymbol{t}) = \hat{\Lambda}(\hat{\boldsymbol{t}}^-)$. Indeed, evidence provided by empirical discrimination surfaces suggests that for the study data, one is more likely to observe symmetric brains in the schizophrenia group of patients than in the healthy control group. Thus, we concentrate in $\boldsymbol{t}^* = \boldsymbol{t}^-$ as the estimation target in what follows.

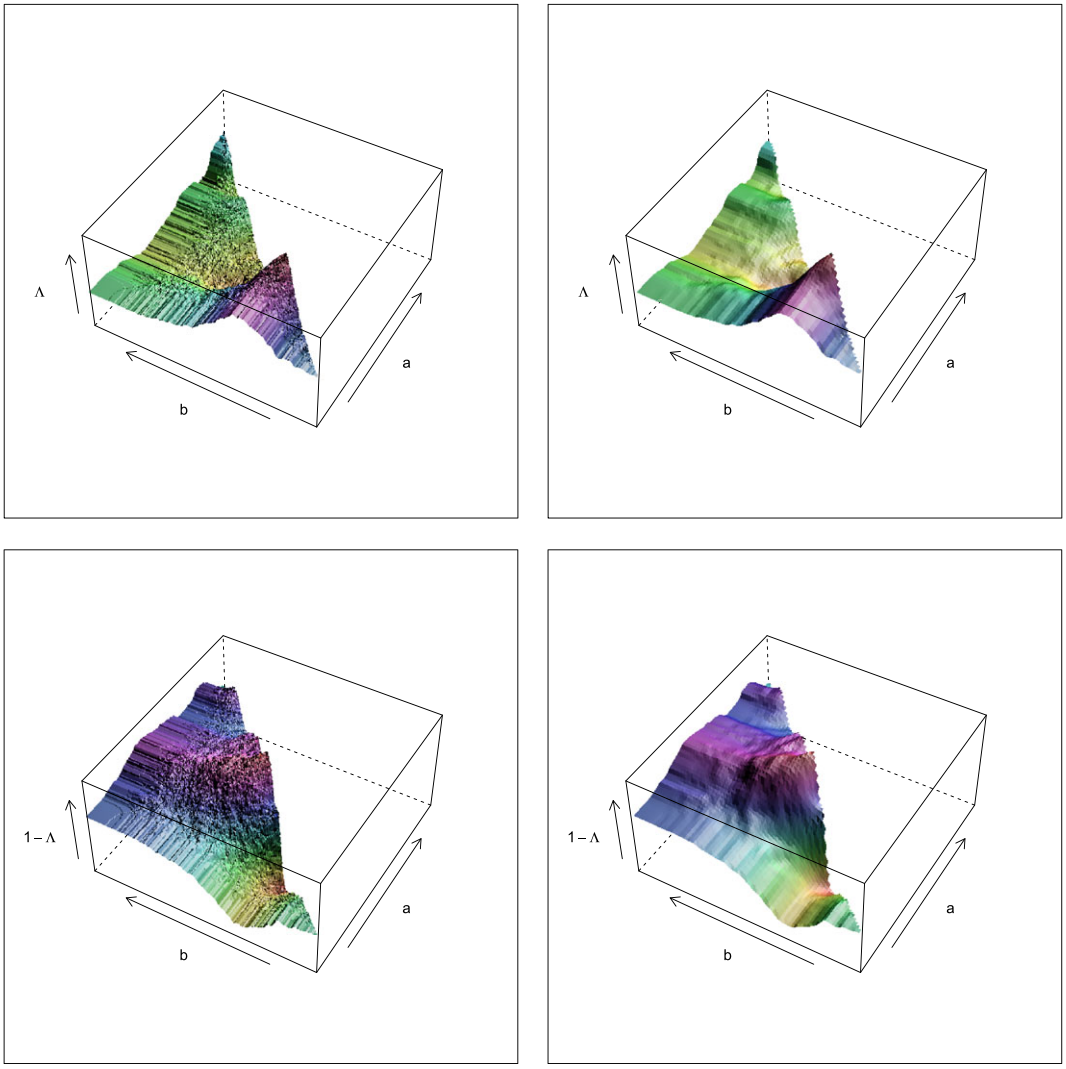

FIGURE 5 Discrimination surface estimates. Above: empirical discrimination surface $\hat{\Lambda}(\boldsymbol{t})$ and its smoothed version $\tilde{\Lambda}(\boldsymbol{t})$. Below: reversed empirical discrimination surface $1 - \hat{\Lambda}(\boldsymbol{t})$ and its smoothed version $1 - \tilde{\Lambda}(\boldsymbol{t})$ [Colour figure can be viewed at wileyonlinelibrary.com]

We now turn our attention to the IMD as defined in (5). The corresponding estimate is $\widehat{\text{IMD}} = [0.44, 0.52]$, which corresponds partially to the perisilvian region of the brain. In the region parametrized by $\hat{\boldsymbol{t}}^*$, the estimated probability $\hat{P}\{Y_D(\hat{\boldsymbol{t}}^*) \leqslant Y_{\bar{D}}(\hat{\boldsymbol{t}}^*)\} = .66$ can be interpreted as the estimated proportion of schizophrenic patients who present lower local asymmetry scores than a healthy patient in the region parametrized by $\hat{\boldsymbol{t}}^*$. In Figure 6, we show the estimated average brain curves $(\bar{B}_{\text{L},D}, \bar{B}_{\text{R},D})$ and $(\bar{B}_{\text{L},\bar{D}}, \bar{B}_{\text{R},\bar{D}})$ and, highlighted in red, the region that corresponds to $\hat{\boldsymbol{t}}^* = (0.44, 0.52)$. As can be seen, the "average patient" in the control group have a more asymmetric brain than the average patient in the disease group (notice that the area highlighted in red on the left of the figure is greater that the same brain area on the right of the figure) in the local region parametrized by $\hat{\boldsymbol{t}}^*$.

The ROTD_u as defined in (4), was estimated using the plug-in estimator in (10); the result is presented in Figure 7 for $u = 0.60$ (on the left) and $u = 0.65$ (on the right). For the dataset, we are considering here $\hat{\Lambda}(\boldsymbol{t}) < 0.60$ for all $\boldsymbol{t} \in T$; therefore, the estimated ROTD_u only highlights regions where $\hat{P}\{Y_D(\boldsymbol{t}) \leq Y_{\bar{D}}(\boldsymbol{t})\} \geq u$. In Figure 7 (left), we can see that schizophrenic patients present more symmetric brains than healthy patients in the regions of the brain corresponding to the anterior zone. Moreover, in Figure 7 (right), we observe that this difference in terms of symmetry is more likely to be observed in the perisylvian region of the brain, evidence which corroborated by the IMD estimate, which localizes the region in Equation 10.

Therefore, the empirical evidence obtained in this local-brain asymmetry study agrees with the results obtained by Narr et al[11] and Oertel-Knöchel and Linden.[21] A sensitivity analysis was conducted using other distances to local symmetry

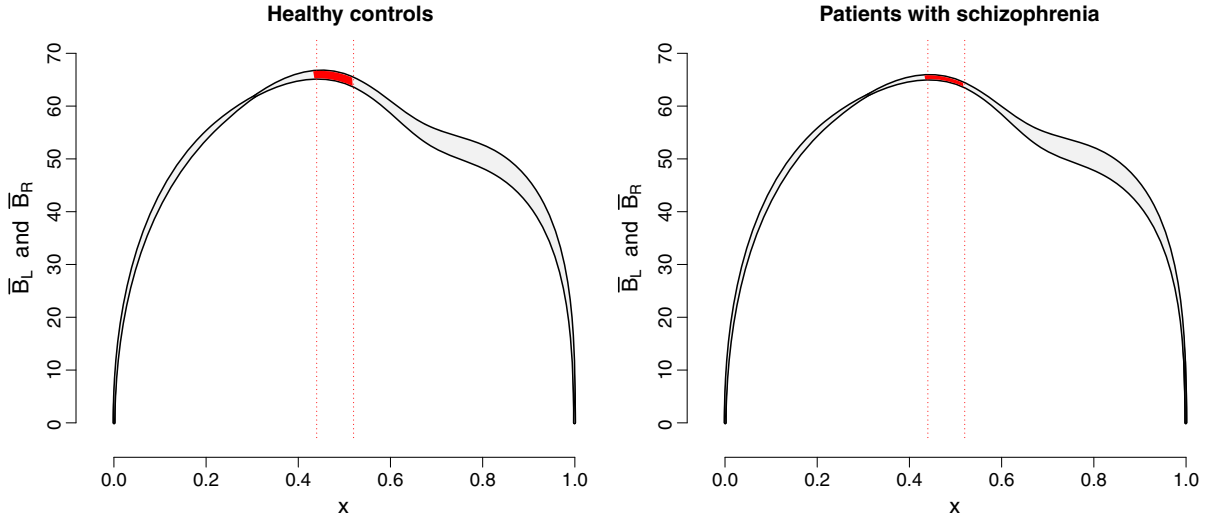

FIGURE 6 Average brain curves and the local region $\widehat{t}^*$ corresponding to estimated interval of maximal discrimination, as defined in Section 3.3, which corresponds to the point where the empirical discrimination surface attain its maximum [Colour figure can be viewed at wileyonlinelibrary.com]

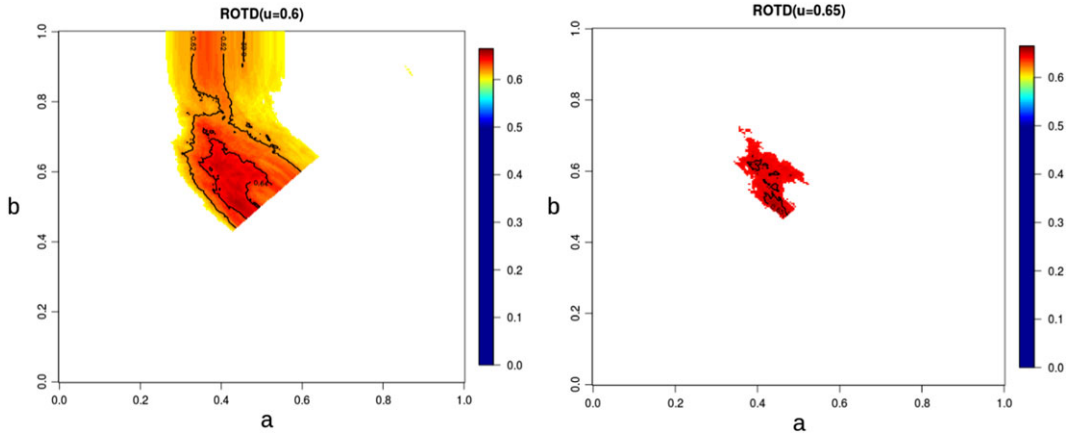

FIGURE 7 Regions over-threshold discrimination (ROTDs) estimate as defined in Equation 10 with $u = 0.60$ (left) and $u = 0.65$ (right), for brain asymmetry analysis data [Colour figure can be viewed at wileyonlinelibrary.com]

instead of Equation 1. The choice of the distance does not appear to affect our main findings; in particular, the IMDs corresponding to the L_1 and L_∞ distances to local asymmetry are $\widehat{\text{IMD}}_{L_1} = [0.45, 0.51]$ and $\widehat{\text{IMD}}_{L_\infty} = [0.47, 0.50]$. To assess the significance of the identified region of local asymmetry, we used Hsieh and Turnbull,[23](Theorem 2.3) which implies that asymptotically $\sqrt{n_D}\{\widehat{\Lambda}(\boldsymbol{t}) - \Lambda(\boldsymbol{t})\} \sim N(0, \sigma^2) + o_p(1)$, with $\sigma^2 = 1/12(1/n_D + 1/n_{\bar{D}})$—under the null $H_0 : \Lambda(\boldsymbol{t}) = 1/2$. The null is rejected at the estimated IMD, and more importantly, it is consistently rejected on a neighborhood of $(\hat{a}, \hat{b})$. See Section S3 in the Supporting Information.

To describe the uncertainty around the IMD, we use the bootstrap to build a empirical CR for this parameter. The study was conducted using $B = 1000$ bootstrap samples from the original data. In Figure 8, we show the empirical α-CRs—computed from $\hat{P}\{\text{IMD} \in \text{CR}_\alpha\} = 1 - \alpha$ for $\alpha \in [0, 1]$—that support the evidence of morphological differences in the brains between groups, being more likely to observe symmetric brains in the perisylvian and anterior regions in schizophrenic patients.

In addition, in Figure 9, we show the bootstrap-based 95% percentile functional CR for $P\{Y_D(\boldsymbol{t}) \leqslant Y_{\bar{D}}(\boldsymbol{t})\}$ and $1 - P\{Y_D(\boldsymbol{t}) \leqslant Y_{\bar{D}}(\boldsymbol{t})\}$. These empirical functional CRs can be used to assess the variance regarding the estimated discrimination surfaces and in particular the uncertainty around the estimated IMD.

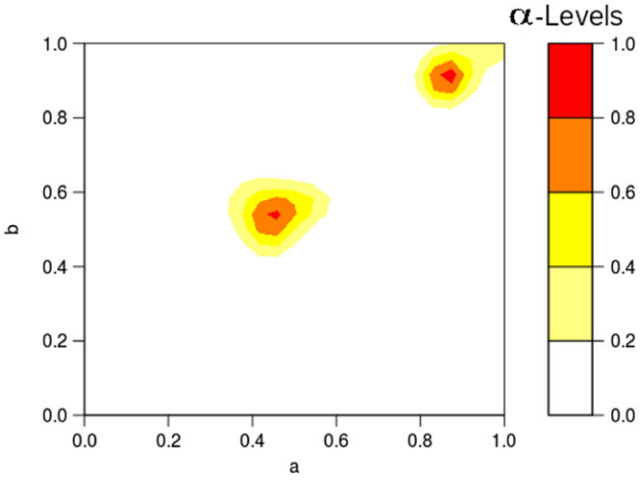

FIGURE 8 Bootstrap estimation of the confidence region for different confidence levels $(1-\alpha)$ [Colour figure can be viewed at wileyonlinelibrary.com]

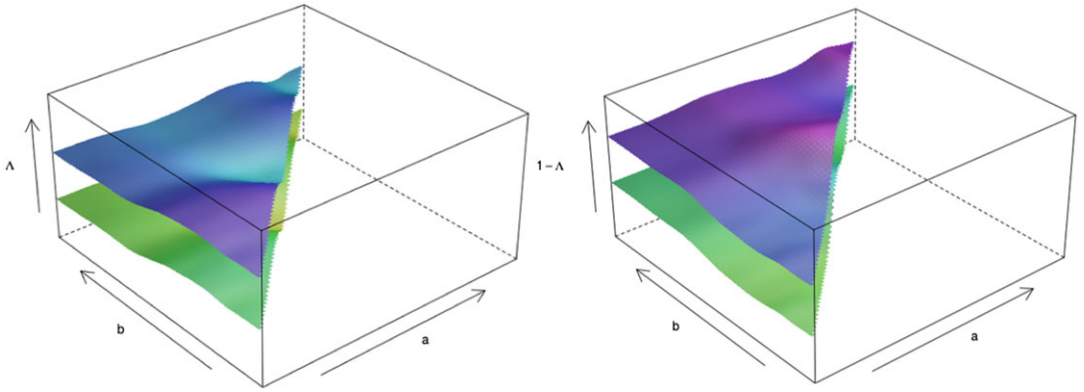

FIGURE 9 Empirical functional 95% confidence interval for the discrimination surfaces: $\Lambda(\boldsymbol{t})$ on the left and $1-\Lambda(\boldsymbol{t})$ on the right [Colour figure can be viewed at wileyonlinelibrary.com]

6 | DISCUSSION

Discrimination surfaces are introduced here as a diagnostic tool for localizing brain regions where discrimination between diseased and nondiseased participants is higher. To estimate discrimination surfaces and associated regions of maximum discrimination, we introduce a Mann-Whitney type of statistic for random fields and present large-sample results characterizing its asymptotic behavior. We have analyzed the data documented in Brignell et al[5] and found evidence of region-specific difference in terms of symmetry between the groups of healthy and schizophrenic patients. The empirical analysis suggests that for this study data, schizophrenic patients are more likely to present symmetric brains in the anterior region. In particular, with maximum probability, this difference in shape is observed in the perisylvian region as is also documented in Narr et al [11] and Oertel-Knöchel and Linden[21] and the reference therein. An obvious limitation with empirical discrimination surface stems from the lack of ability to borrow strength across values of $\boldsymbol{t}$, which motivates the need for also considering the smoothed version in (12).

A natural possibility for future work entails building alternative IMDs from other measures assessing discrimination between groups such as the Youden index[24]—which has links with the Kolmogorov-Smirnov statistic—or a standardized log-rank statistic[25-27]—which has links with the Wilcoxon rank statistic. Specifically, in a similar way that we argue here that discrimination surfaces have connections with the area under conditional ROC curves (cf Section 3.2), it would be naturally modeling our applied setting of interest with an analogue of the covariate-adjusted Youden index,[28] $\mathrm{YI}_{\boldsymbol{t}} = \max_{c_{\boldsymbol{t}}}\{F_{D\boldsymbol{t}}(c_{\boldsymbol{t}}) - F_{\bar{D}\boldsymbol{t}}(c_{\boldsymbol{t}})\}$, with $c_{\boldsymbol{t}}$ being a function of $\boldsymbol{t}$; the corresponding IMD would in this case result from maximizing $\mathrm{YI}_{\boldsymbol{t}}$. In addition, a similar approach would entail developing standardized log-rank statistics for random fields, $|S_{c_{\boldsymbol{t}}}|$, whose corresponding IMD would result from maximizing the discrimination surface $|S_{c_{\boldsymbol{t}}}|$ over $\boldsymbol{t}$ in T.

Another natural avenue for future work is on extending the local asymmetry analysis to 3-D functional data so to estimate 3-D regions of maximal discrimination. Accompanying the latest trends in the study of medical images,[29,30] a natural starting point for this would be to model the brain surface $\tilde{B}(x, y, z)$ as a Riemann manifold, $\mathcal{M}$, embedded in a normed vector space $(\mathbb{R}^3, \|\cdot\|)$. After a suitable partition and reparametrization of the brain manifold into left and right brain submanifolds, $B_L(x, y, z)$ and $B_R(x, y, z)$, one could consider the set C of closed smooth curves $\gamma : [0, 1] \to \mathcal{M}$, that is, for all $\gamma \in C$ then $\gamma(0) = \gamma(1)$ and $\gamma'(0) = \gamma'(1)$; a curve belonging to this set determines a 3-D closed region in the brain surface. Asymmetry scores in the local region determined by the closed curve $\gamma \in C$ could then be defined as

$$Y_D(\gamma) = \int_0^1 \{B_{\mathrm{L},D}(\gamma(s)) - B_{\mathrm{R},D}(\gamma(s))\}^2 \mathrm{d}s, \quad Y_{\bar{D}}(\gamma) = \int_0^1 \{B_{\mathrm{L},\bar{D}}(\gamma(s)) - B_{\mathrm{R},\bar{D}}(\gamma(s))\}^2 \mathrm{d}s,$$

along with the discrimination functional for such 3-D setting,

$$\Lambda(\gamma) = P\{Y_D(\gamma) > Y_{\bar{D}}(\gamma)\}, \quad \gamma \in C.$$

The variational problems associated to the local region of maximal discrimination,

$$\gamma^+ =: \arg\max_{\gamma \in C} \Lambda(\gamma) \quad \text{and} \quad \gamma^- =: \arg\max_{\gamma \in C} 1 - \Lambda(\gamma),$$

would entail estimation over an infinite-dimensional parameter space C and thus, would require the need of developing inference methods and asymptotics tailored for that setting.

ACKNOWLEDGEMENTS

We thank 2 referees, the associate editor, and the editor for constructive comments and insighful recommendations. This work was partially supported by the grant Fondecyt 11121186 from the Chilean NSF (Conicyt) and by FCT (Fundação para a Ciência e a Tecnologia, Portugal), through the project UID/MAT/00006/2013. The authors thank participants of Latent Variables 2016, held at University of South Carolina, for insightful discussions. We extend our thanks to Ian Dryden for making the brain data available and for useful recommendations and comments on the case study and to Vanda Inácio de Carvalho for discussions and comments on an earlier version of the manuscript.

ORCID

Miguel de Carvalho http://orcid.org/0000-0003-3248-6984

REFERENCES

1. Galaburda AM, LeMay M, Kemper TL, Geschwind N. Right-left asymmetrics in the brain. *Science*. 1978;199:852-856.
2. Springer SP, Deutsch G. *Left Brain, Right Brain: Perspectives from Cognitive Neuroscience*. 5th ed. New York: Freeman; 1998.
3. Toga AW, Thompson PM. Mapping brain asymmetry. *Nat Rev Neurosci*. 2003;4:37-48.
4. Rogers LJ, Vallortigara G, Andrew RJ. *Divided Brains: The Biology and Behaviour of Brain Asymmetries*. Cambridge UK: Cambridge University Press; 2013.
5. Brignell CJ, Dryden IL, Gattone SA, Park B, Leask S, Browne WJ, Flynn S. Surface shape analysis with an application to brain surface asymmetry in schizophrenia. *Biostatistics*. 2010;11:609-630.
6. Luchins DJ, Weinberger DR, Wyatt RJ. Schizophrenia: evidence of a subgroup with reversed cerebral asymmetry. *Arch Gen Psychiatry*. 1979;36:1309-1311.
7. Harrison PJ. The neuropathology of schizophrenia: a critical review of the data and their interpretation. *Brain*. 1999;122:593-624.
8. Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. *Schizophrenia Res*. 2001;49:1-52.
9. Csernansky JG, Gillespie SK, Dierker DL, Anticevic A, Wang L, Barch DM, Van Essen DC. Symmetric abnormalities in sulcal patterning in schizophrenia. *Neuroimage*. 2008;43:440-446.
10. Ribolsi M, Daskalakis ZJ, Siracusano A, Koch G. Abnormal asymmetry of brain connectivity in schizophrenia. *Front Hum Neurosci*. 2014;8:1-11.
11. Narr KL, Bilder RM, Luders E, Thompson PM, Woods RP, Robinson D, Szeszko PR, Dimtcheva T, Gurbani M, Toga AW. Asymmetries of cortical shape: effects of handedness, sex and schizophrenia. *Neuroimage*. 2007;34:939-948.
12. Inácio de Carvalho V, Jara A, Hanson TE, de Carvalho M. Bayesian nonparametric ROC regression modeling. *Bayesian Anal*. 2013;8:623-646.
13. Inácio de Carvalho V, de Carvalho M, Alonzo TA, González-Manteiga W. Functional covariate-adjusted partial area under the specificity-ROC curve with an application to metabolic syndrome diagnosis. *Ann Appl Stat*. 2016;10:1472-1495.
14. Bandos A, Rockette HE, Song T, Gur D. Area under the free-response ROC curve (FROC) and a related summary index. *Biometrics*. 2009;65:247-256.

15. Ferraty F, Vieu P. *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer; 2006.
16. Ramsay JO. *Functional Data Analysis*. New York: Wiley; 2006.
17. Sørensen H, Goldsmith J, Sangalli LM. An introduction with medical applications to functional data analysis. *Stat Med*. 2013;32:5222-5240.
18. Talairach P, Tournoux J. *A Stereotactic Coplanar Atlas of the Human Brain*. New York: Thieme; 1988.
19. Dryden IL. *Shapes Package*. Vienna: R Foundation for Statistical Computing; 2017.
20. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford UK: Oxford University Press; 2003.
21. Oertel-Knöchel V, Linden DEJ. Cerebral asymmetry in schizophrenia. *Neuroscientist*. 2011;17:456-467.
22. Marron JS. Discussion of "Inference for Density Families using Functional Principal Component Analysis". *J Am Stat Assoc*. 2001;96:532-533.
23. Hsieh F, Turnbull BW. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann Stat*. 1996;24:25-40.
24. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32-35.
25. Lausen B, Schumacher M. Maximally selected rank statistics. *Biometrics*. 1992;48:73-85.
26. Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. *Comput Stat Data Anal*. 2003;43:121-137.
27. Hothorn T, Zeileis A. Generally maximally selected statistics. *Biometrics*. 2008;64:1263-1269.
28. Inácio de Carvalho V, de Carvalho M, Branscum A. Nonparametric Bayesian covariate-adjusted estimation of the Youden index. *Biometrics*. 2017;73:1279-1288.
29. Miller M, Banerjee A, Christensen G, Joshi S, Khaneja N, Grenander U, Matejic L. Statistical methods in computational anatomy. *Stat Meth Med Res*. 1997;6:267-299.
30. Pennec X. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J Math Imag Vis*. 2006;25:127-154.
31. Embrechts P, Klüppelberg C, Mikosch T. *Modelling Extremal Events for Insurance and Finance*. New York: Springer; 1997.
32. Newey WK, McFadden D. Large sample estimation and hypothesis testing. *Handbook of Econometrics*, Vol. 4. Amsterdam: Elsevier; 1994:2111-2245.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Martos G, de Carvalho M. Discrimination surfaces with application to region-specific brain asymmetry analysis. *Statistics in Medicine*. 2018;1–15. https://doi.org/10.1002/sim.7611

APPENDIX A In these appendices, we use $F(y, a, b)$ to denote $F_{\boldsymbol{t}}(y) = P\{Y(\boldsymbol{t}) \leqslant y\}$ when the context so requires and use the notation $F_{D\boldsymbol{t}}(y) = P\{Y_D(\boldsymbol{t}) \leqslant y\}$ and $F_{\bar{D}\boldsymbol{t}}(y) = P\{Y_{\bar{D}}(\boldsymbol{t}) \leqslant y\}$, for $y \in S$ and $\boldsymbol{t} \in T$. Also, $F^{-1}_{\boldsymbol{t}}(p) = \inf\{y : F_{\boldsymbol{t}}(y) \geqslant p\}$ are the marginal quantiles, $\hat{F}^{-1}_{\boldsymbol{t}}(p) = \inf\{y : \hat{F}_{\boldsymbol{t}}(y) \geqslant p\}$ are the marginal empirical quantiles, with $\hat{F}_{\boldsymbol{t}}(y) = n^{-1}\sum_{i=1}^{n} I\{Y_i(\boldsymbol{t}) \leqslant y\}$ denoting the marginal empirical distribution function.

A.1 | Proof of Theorem 1

Our line of attack is similar to that of Embrechts et al.[31](p62) Let $C_{i,j,k} = [y_i, y_{i+1}] \times [a_j, a_{j+1}] \times [b_k, b_{k+1}]$, with

$$0 = y_0 < y_1 < \cdots < y_{I-1} < y_I = M, \quad 0 = a_0 < a_1 < \cdots < a_{J-1} < a_J = 1, \quad 0 = b_0 < b_1 < \cdots < b_{K-1} < b_K = 1,$$

be such

$$|F(y_{i+1}, a_{j+1}, b_k) - F(y_i, a_j, b_{k+1})| < \epsilon,$$

for a given $\epsilon > 0$, for $i = 0, \ldots, I$, $j = 0, \ldots, J$, and $k = 0, \ldots, K$. By the monotonicity properties of $F_{\boldsymbol{t}}(y)$ and $\hat{F}_{\boldsymbol{t}}(y)$ on $S \times [0,1]^2$, namely, $F_{(\cdot,\cdot)}(y)$ and $F_{(a,\cdot)}(\cdot)$ are nondecreasing, $F_{(\cdot,b)}(\cdot)$ is nonincreasing, and analogous properties hold for $\hat{F}_{\boldsymbol{t}}(y)$, it follows that

$$\begin{aligned}
\Delta_n \equiv \sup_{(y,\boldsymbol{t}) \in S \times T} |\hat{F}_{\boldsymbol{t}}(y) - F_{\boldsymbol{t}}(y)| &\leqslant \sup_{(y,\boldsymbol{t}) \in S \times [0,1]^2} |\hat{F}_{\boldsymbol{t}}(y) - F_{\boldsymbol{t}}(y)| \\
&= \max_{i,j,k} \sup_{(y,\boldsymbol{t}) \in C_{i,j,k}} |\hat{F}_{\boldsymbol{t}}(y) - F_{\boldsymbol{t}}(y)| \\
&\leqslant \max_{i,j,k} \max\{|\hat{F}(y_{i+1}, a_{j+1}, b_k) - F(y_i, a_j, b_{k+1})|, |F(y_{i+1}, a_{j+1}, b_k) - \hat{F}(y_i, a_j, b_{k+1})|\},
\end{aligned}$$

and thus, by taking the limit and using the strong law of large numbers,

$$\limsup_{n\to\infty} \quad \Delta_n = \max_{i,j,k}\{|F(y_{i+1}, a_{j+1}, b_k) - F(y_i, a_j, b_{k+1})|\} < \epsilon, \quad \text{a.s.}$$

□

A.2 | Auxiliary results

This appendix includes auxiliary lemmas, which streamline the proof of Theorem 2 in Appendix A.3.

Lemma 1. *Let $\boldsymbol{t} \in T$, and let $g(\boldsymbol{t})$ be a function. Suppose that (1) $g(\boldsymbol{t})$ is uniquely maximized at $\boldsymbol{t}^*$, (2) T is compact, (3) $g(\boldsymbol{t})$ is continuous, and (4) $\hat{g}(\boldsymbol{t})$ converges uniformly in probability to $g(\boldsymbol{t})$, and $\hat{\boldsymbol{t}} = \arg\max_{\boldsymbol{t}} \hat{g}(\boldsymbol{t})$. Then, as $n \to \infty$, it holds that*

$$\hat{\boldsymbol{t}} - \boldsymbol{t}^* = o_{\mathrm{p}}(1).$$

Proof. See Newey and McFadden.[32](p2121) □

Lemma 2. *Let $Y_1(\boldsymbol{t}), \ldots, Y_n(\boldsymbol{t})$ be a sequence of independent identically distributed random functions in $\mathcal{Y}$. Suppose $F_{\boldsymbol{t}}(y)$ is continuous for all $(y, \boldsymbol{t})$ in $S \times T$ and strictly increasing for all y in S. Then, as $n \to \infty$, it holds that*

$$\sup_{(p,\boldsymbol{t})} |\hat{F}^{-1}_{\boldsymbol{t}}(p) - F^{-1}_{\boldsymbol{t}}(p)| = o(1), \quad \text{a.s.}$$

Proof. The proof follows the same reasoning as that of Theorem 1 and can be found in Supporting Information. □

Lemma 3. *Suppose assumptions A1 to A4 of Section 3.2 hold. Then, as $n \to \infty$, it holds that*

$$\sup_{(p,\boldsymbol{t})} |F_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\} - F_{D\boldsymbol{t}}\{F^{-1}_{\bar{D}\boldsymbol{t}}(p)\}| = o(1), \quad \text{a.s.}$$

Proof. Since $F_{\bar{D}\boldsymbol{t}}(y)$ is a bounded and continuous function for all $(y, \boldsymbol{t})$ in $S \times T$, it is uniformly continuous, and thus, for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\sup_{(p,\boldsymbol{t})\in[0,1]\times T} |\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p) - F^{-1}_{\bar{D}\boldsymbol{t}}(p)| < \delta \Rightarrow \sup_{(p,\boldsymbol{t})\in[0,1]\times T} |F_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\} - F_{D\boldsymbol{t}}\{F^{-1}_{\bar{D}\boldsymbol{t}}(p)\}| < \epsilon, \quad \text{a.s.} \tag{A1}$$

Then, by the uniform convergence of $\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)$(Lemma 2), we can always find an $n > N(\epsilon)$ such that the right-hand side of (A1) holds, from where the final result follows. □

A.3 | Proof of Theorem 2

Keeping in mind that the area under the curve coincides with the area under the ordinal dominance curve,[23](p27) it follows that

$$\begin{aligned}
|\hat{\Lambda}(\boldsymbol{t}) - \Lambda(\boldsymbol{t})| &= \left| \int_0^1 \hat{F}_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\}\,\mathrm{d}p - \int_0^1 F_{D\boldsymbol{t}}\{F^{-1}_{\bar{D}\boldsymbol{t}}(p)\}\,\mathrm{d}p \right| \\
&\leqslant \int_0^1 |\hat{F}_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\} - F_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\}|\,\mathrm{d}p + \int_0^1 |F_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\} - F_{D\boldsymbol{t}}\{F^{-1}_{\bar{D}\boldsymbol{t}}\{p)\}|\,\mathrm{d}p \\
&\leqslant \sup_{p\in[0,1]} |\hat{F}_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\} - F_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\}| + \sup_{p\in[0,1]} |F_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\} - F_{D\boldsymbol{t}}\{F^{-1}_{\bar{D}\boldsymbol{t}}(p)\}|,
\end{aligned}$$

and thus,

$$\sup_{\boldsymbol{t}} |\hat{\Lambda}(\boldsymbol{t}) - \Lambda(\boldsymbol{t})| \leqslant \sup_{(p,\boldsymbol{t})\in[0,1]\times T} |\hat{F}_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\} - F_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\}| + \sup_{(p,\boldsymbol{t})\in[0,1]\times T} |F_{D\boldsymbol{t}}\{\hat{F}^{-1}_{\bar{D}\boldsymbol{t}}(p)\} - F_{D\boldsymbol{t}}\{F^{-1}_{\bar{D}\boldsymbol{t}}(p)\}|.$$

Strong uniform consistency of $\hat{\Lambda}$ follows by observing that a.s. convergence to zero of the left-hand term follows by Theorem 1, whereas a.s. convergence to zero of the right-hand term follows by Lemma 3.

Weak consistency of $\hat{\boldsymbol{t}}$ then follows from Lemma 1 and by noting that $\Lambda(\boldsymbol{t})$ is continuous. To justify the latter claim note that by assumption $F_{D\boldsymbol{t}}\{F^{-1}_{\bar{D}\boldsymbol{t}}(p)\}$ is a bounded and continuous function for $(p, \boldsymbol{t})$ in $[0, 1] \times T$, and thus, it is uniformly continuous, so that for every $\epsilon > 0$, there exists a $\delta > 0$, such that if $\max\{|p - p'|, |a - a'|, |b - b'|\} < \delta$, then

$$|F_{D\boldsymbol{t}}\{F^{-1}_{\bar{D}\boldsymbol{t}}(p)\} - F_{D\boldsymbol{t}'}\{F^{-1}_{\bar{D}\boldsymbol{t}'}(p')\}| < \epsilon. \tag{A2}$$

Continuity of $\Lambda(\boldsymbol{t})$ then follows from (A2) and the fact that to ensure $|\Lambda(\boldsymbol{t}) - \Lambda(\boldsymbol{t}')| < \epsilon$, we only need to set $\delta > \max\{|a - a'|, |b - b'|\}$.