Pablo Martínez-Camblor[1] / Juan Carlos Pardo-Fernández[2]

The Youden Index in the Generalized Receiver Operating Characteristic Curve Context

[1] The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth College, 7 Lebanon Street, Suite 309, Hinman Box 7251, Hanover, NH 03755, USA, E-mail: Pablo.Martinez.Camblor@Dartmouth.edu
[2] Department of Statistics and OR, and Biomedical Research Centre, Universidade de Vigo, Vigo, Spain

Abstract:
The receiver operating characteristic (ROC) curve and their associated summary indices, such as the Youden index, are statistical tools commonly used to analyze the discrimination ability of a (bio)marker to distinguish between two populations. This paper presents the concept of Youden index in the context of the generalized ROC (gROC) curve for non-monotone relationships. The interval estimation of the Youden index and the associated cutoff points in a parametric (binormal) and a non-parametric setting is considered. Monte Carlo simulations and a real-world application illustrate the proposed methodology.

Keywords: (Bio)markers, cutoff point, generalized receiver-operating characteristic (gROC) curve, receiver-operating characteristic (ROC) curve, Youden index

DOI: 10.1515/ijb-2018-0060

Received: June 20, 2018; **Revised:** March 13, 2019; **Accepted:** March 13, 2019

1 Introduction

The problem of classifying subjects as being inside or outside a group using one or more individual characteristics frequently appears in fields such as business intelligence or machine learning. Its connection with both diagnosis and prognosis tasks makes its role fundamental in biomedical sciences. From a statistical point of view, the problem consists in determining regions which minimize the potential classification errors when allocating individuals into the two possible groups, which will be called *positive* (holding the studied characteristic) and *negative* (not holding the studied characteristic). Relevant quantities related to the classification problem are the *specificity*, S_P, – probability of classifying a negative subject as negative, also called *true negative rate* –, and the *sensitivity*, S_E, – probability of classifying a positive subject as positive, also called *true positive rate*. The *receiver operating characteristic curve* (ROC curve) is a graphical tool that depicts the sensitivity against one minus specificity – also called *false positive rate* – for all possible classification regions (see, for example, Zhou et al. [1]). The ROC curve has become popular and is routinely used in classification problems. In addition, the area under the ROC curve (AUC) is commonly used for summarizing the diagnostic capacity by means of a single number [2] and even for comparing the diagnostic quality of two or more diagnostic procedures [3]. Standard ROC curves mainly consider a continuous biomarker, Y, and situations in which larger values of Y are associated with larger probabilities of belonging to the positive group. That is, the classification subsets (those classifying a subject as positive) are of the form $[u, \infty)$ with $u \in \mathbb{R}$. In this context, for each $t \in [0,1]$, ROC curve is defined by

$$\mathscr{R}_r(t) = 1 - F(G^{-1}(1-t)),$$

where, if D determines that an individual belongs ($D = 1$) or does not belong ($D = 0$) to the positive group, $F(\cdot) = \mathscr{P}(Y \leq \cdot | D = 1)$ and $G(\cdot) = \mathscr{P}(Y \leq \cdot | D = 0)$. Martínez-Camblor et al. [4] considered the case in which extreme values of the biomarker are associated with having a higher probability of being positive and introduced the *generalized receiver operating characteristic curve* (gROC curve). In this case, the considered classification subsets are of the form $(-\infty, u_L] \cup [u_U, \infty)$ with $u_L, u_U \in \mathbb{R}$, $u_L \leq u_U$. Martínez-Camblor and Pardo-Fernández [5] studied the parametric estimator of the gROC curve and proved that, under particular assumptions, the area under the gROC curve (gAUC) is *the probability of selecting, randomly and independently, two subjects, one negative and one positive, for which there exists a classification subset such that both subjects are correctly allocated* (see Theorem 1 in [5] for further details on the conditions about this interpretation of the AUC in the gROC context).

Classification processes involve a set of *classification rules* which are the ones actually used for making the final decision. Conventionally, each sensitivity (specificity) value has a particular associated classification subset

Pablo Martínez-Camblor is the corresponding author.
© 2019 Walter de Gruyter GmbH, Berlin/Boston.

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

on which the decision is based. In practice, the choice of one particular classification subset is usually a trade-off between the sensitivity and the specificity values.

In the standard ROC curve context, that is, when subjects with larger values of the biomarker have more probability of belonging to the positive group, different criteria have been proposed for selecting an adequate cutoff point. The North-West corner, the Youden index, the concordance probability or the symmetry point are some of them. A brief review of these approaches is provided in Section 2.

In this paper, we deal with the problem of selecting an adequate classification criterion in the gROC curve context. Particularly, we will extend the Youden index [6] for classification processes based on classification subsets of the form $(-\infty, u_L] \cup [u_U, \infty)$ with $u_L, u_U \in \mathbb{R}$, $u_L \leq u_U$. In Sections 3 and 4, the resulting parametric and non-parametric estimators are studied and the asymptotic distributions for both the Youden index and the associated cutoff points are provided. Via Monte Carlo simulations, in Section 5, we investigate the performance of the derived confidence intervals. In Section 6 a real-world problem is considered; we study the ability of the white blood cells count of identifying the type of disease in patients having either acute viral meningitis or acute bacterial meningitis. Full technical details about the provided asymptotic results are included as appendix. The R (www.r-project.org) code employed in this work is available as supplementary material.

2 The Youden index and its competitors

The area under the ROC curve (AUC) is the most commonly used global index of diagnostic accuracy. One of its main handicaps is that it does not provide an associated decision rule or classification subset. There are a number of indices that, based on some particular criterion, select a particular classification rule, while providing and summarizing the global diagnostic accuracy as well. Next, we briefly revise four of these criteria.

Taking into account that ROC curves obtained from biomarkers that lead to perfect classification reach the point $(0, 1)$ in the unit square, Coffin and Sukhatme [7] proposed to use the point which minimizes the distance between the ROC curve and that point. This is the so-called North-West corner criterion, which is mathematically defined by $N = \min_{t\in[0,1]}\{(\mathscr{R}_r(t) - 1)^2 + t^2\}$. The lack of a probabilistic interpretation is perhaps the main handicap of this procedure.

The Concordance probability index [8] maximizes the probability of making a correct classification of two subjects, one positive and one negative, randomly selected. Therefore, it is defined as $C = \max_{t\in[0,1]}\{\mathscr{R}_r(t) \cdot (1 - t)\}$. The Concordance probability is the area of the largest rectangle that can be plotted below the ROC curve.

The Symmetry point [9] is the point, c_S, such that $\mathscr{R}(c_S) = 1 - c_S$. This point is the one which maximizes simultaneously the probability of correctly classifying positive and negative subjects. Besides, from a Bayesian decision perspective, the Symmetry point coincides with the minimax rule that aims to minimize the maximum of the misclassification error probability.

The Youden index [6] is the most popular of those criteria. It is defined as the maximum of the sensitivity plus the specificity minus 1 or, equivalently, in ROC curve terms,

$$\mathfrak{J}_r = \max_{t\in[0,1]}\{\mathscr{R}_r(t) - t\}.$$

It ranges between 0 and 1. A Youden index of 0 indicates that the biomarker is equally distributed on the positive and the negative populations while a value of 1 indicates completely separate distributions. Notice that, for a given fixed threshold c, the probability that a randomly selected subject is correctly classified based on the classification subset $[c, \infty)$ is

$$\pi \cdot \mathscr{P}\{Y > c|D = 1\} + (1 - \pi) \cdot \mathscr{P}\{Y \leq c|D = 0\},$$

where $\pi = \mathscr{P}\{D = 1\}$. That is, the Youden index is equivalent to the optimal correct classification probability obtained when we use rules based on the biomarker Y and the prevalence of the studied characteristic is $1/2$. Martínez-Camblor [10] generalized the Youden index considering $\pi \in (0, 1)$.

Figure 1 illustrates the above four indices for normally distributed populations. We highlight on the graphs the points of the ROC for which the corresponding indices are achieved, regardless whether they are actually defined as points on the ROC curve (for example, the NW corner) or not (for example, the Youden index). In the left panel, the marker has standard deviation 1 in both positive and negative populations while the means are 0 and 1.19 in the negative and the positive groups, respectively (AUC = 0.8). The four indices are achieved on similar points, in fact, three of them, N, C and S, coincide. In the right panel, the positive population has mean 2.66 and standard deviation 3 while negative population is standard normal (AUC = 0.8). In this case, the four points that lead to the corresponding indices are quite different.

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

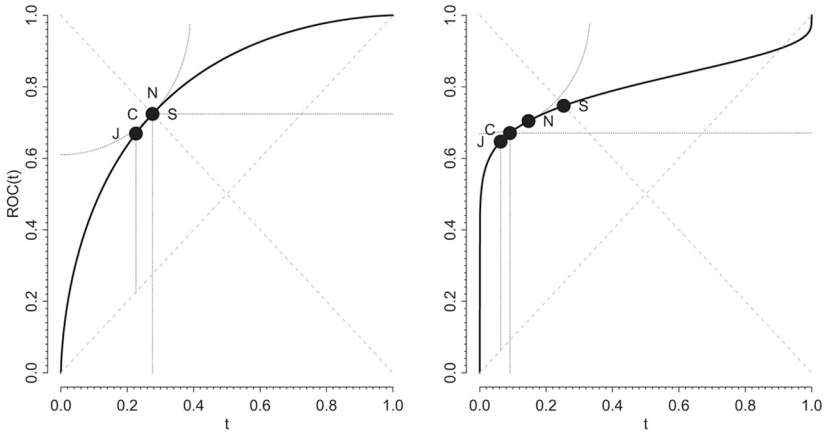

Figure 1: North-West corner, *N*, Concordance probability, *C*, Symmetry point, *S* and Youden index, *J*, for normally distributed populations. Left: homoscedastic model (negative population with mean 0 and standard deviation 1, positive population with mean 1.19 and standard deviation 1, AUC= 0.8). Right: heteroscedastic model (negative with mean 0 and standard deviation 1, positive with mean 2.66 and standard deviation 3, AUC= 0.8).

3 The Youden index in the gROC curve context

In practice, it is not unusual that subjects with higher or lower values of the biomarker have more probability of belonging to the positive group. Sometimes, the lower values are associated with one particular pathology while higher values are associated with another one. For instance, in the intensive care units, critically ill patients with high values of *leukocyte counts* are associated with leukocytosis while those with low values are associated with leukopenia; both are related with bad prognosis [11]. When three populations can be clearly stated and there is natural order between them, three-way ROC curves can be used (see the seminal paper of [12] and, for example, [13] for an analysis of the Youden index in that context). But in general, the main problem is often to classify subjects into healthy and diseased populations and just two groups are defined. The gROC curve [4] deals with this situation. For each $t \in [0, 1]$, the gROC curve is defined by

$$\mathscr{R}_g(t) = \sup_{\gamma\in[0,1]} \{\mathscr{R}_r(\gamma \cdot t) + 1 - \mathscr{R}_r(1 - t - \gamma \cdot t)\}.$$

Notice that for each $t \in [0, 1]$ there exists just one pair (u_L, u_U) such that $\mathscr{R}_g(t) = F(u_L) + 1 - F(u_U)$ and $t = G(u_L) + 1 - G(u_U)$. ROC and gROC curves are equal when the distributions of the markers in the positive and negative groups are stochastically ordered. The gROC curve dominates the ROC curve otherwise.

The area under the gROC curve, gAUC, has been proposed as a summary measure for the classification accuracy [4]. Besides, as already mentioned, under certain conditions the gAUC has an appealing probabilistic interpretation [5]. The main handicap is that the gAUC itself (as the standard AUC) does not provide a classification rule to be used in the diagnostic process. However, all of the criteria considered for the ROC curve can be extended to the gROC context. Particularly, we introduce here the Youden index for the gROC curve, which is defined as

$$\mathfrak{J}_g \quad = \max_{t\in[0,1]}\{\mathscr{R}_g(t) - t\}. \tag{1}$$

This definition extends the different interpretations of the Youden index for the standard ROC curves when the classification rules are based on subsets in the form $(-\infty, u_L] \cup [u_U, \infty)$ with $u_L, u_U \in \mathbb{R}$, $u_L \leq u_U$. In the next two subsections we propose and study parametric binormal and non-parametric Youden index estimators in the gROC curve context.

3.1 Normally distributed biomarker

Let $\mathcal{N}(\mu, \sigma^2)$ denote a normal distribution with mean μ and variance σ^2. Assume that Y follows a $\mathcal{N}(\mu_0, \sigma_0^2)$ distribution in the negative group ($D = 0$) and a $\mathcal{N}(\mu_1, \sigma_1^2)$ distribution in the positive group ($D = 1$). Equivalently, and without loss of generality due to the invariance of the ROC and gROC curves with respect to monotone increasing transformations, we will work with the transformed biomarker $Y^* = (Y - \mu_0)/\sigma_0$, which follows a $\mathcal{N}(0, 1)$ in the negative group and and a $\mathcal{N}(a, b^2)$, with $a = (\mu_1 - \mu_0)/\sigma_0$ and $b = \sigma_1/\sigma_0$, in the positive group. If

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

$b = 1$, the gROC curve and the ROC curve coincide. Since we are assuming that extreme values of the biomarker are associated with subjects having more probability of belonging to the positive group, b is assumed to be above 1. Results provided in Martínez-Camblor et al. [14] imply that the gROC curve for Y^* is based on classification subsets of the form $u_L = a/(1-b^2) - x$, and $u_U = a/(1-b^2) + x$, where $x \in \mathbb{R}$ is a non-negative number. In this context, routine calculations show that the Youden index is

$$\mathfrak{J}_g = \max_{x\in[0,\infty)} \left\{ \Phi\left(\frac{ab^2 - x(1-b^2)}{b(1-b^2)}\right) - \Phi\left(\frac{ab^2 + x(1-b^2)}{b(1-b^2)}\right) + \Phi\left(\frac{a + x(1-b^2)}{1-b^2}\right) - \Phi\left(\frac{a - x(1-b^2)}{1-b^2}\right)\right\},$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a $\mathcal{N}(0,1)$. The solution to the above maximization problem gives that $\mathfrak{J}_g$ occurs at $x = x_J$, where

$$x_J = b(b^2-1)^{-1}\sqrt{2(b^2-1)\log(b) + a^2}. \tag{2}$$

The corresponding classification subsets are determined by the values where the densities of the positive and negative groups cross each other, that is $a/(1-b^2) \pm x_J$.

Let $Y_{01}, \ldots, Y_{0m}$ be a sample of i.i.d. observations from the negative population, with sample mean $\hat{\xi}_0$ and sample standard deviation $\hat{S}_0$, and let $Y_{11}, \ldots, Y_{1n}$ be a sample of i.i.d. observations from the positive population, with sample mean $\hat{\xi}_1$ and sample standard deviation $\hat{S}_1$. We propose the plug-in estimator of the Youden index associated to the binormal gROC curve

$$\widehat{\mathfrak{J}}_g^P = \Phi\left(\frac{ab^2 - x_J(1-b^2)}{b(1-b^2)}\right) - \Phi\left(\frac{ab^2 + x_J(1-b^2)}{b(1-b^2)}\right) + \Phi\left(\frac{a + x_J(1-b^2)}{1-b^2}\right) - \Phi\left(\frac{a - x_J(1-b^2)}{1-b^2}\right)$$

where $\hat{a} = (\hat{\xi}_1 - \hat{\xi}_0)/\hat{S}_0$, $\hat{b} = \hat{S}_1/\hat{S}_0$ and $\hat{x}_J = \hat{b}(\hat{b}^2-1)^{-1}\sqrt{2(\hat{b}^2-1)\log(\hat{b}) + \hat{a}^2}$.

If the sample sizes satisfy that $n/m \to_n \lambda > 0$, the delta method (more details are provided in the Appendix) guarantees the following weak convergence ($\mathscr{L}$)

$$\sqrt{n}\cdot\{\widehat{\mathfrak{J}}_g^P - \mathfrak{J}_g\} \xrightarrow{\mathscr{L}}_n \sigma_{J^P}\cdot\mathscr{Z},$$

where $\mathscr{Z}$ is a standard normal random variable and

$$\sigma^2_{J^P} = \left[\left(\frac{\partial\mathfrak{J}_g(\boldsymbol{\beta})}{\partial\mu_1}\right)^2 + \frac{1}{2}\left(\frac{\partial\mathfrak{J}_g(\boldsymbol{\beta})}{\partial\sigma_1}\right)^2\right]\cdot\sigma_1^2 + \lambda\cdot\left[\left(\frac{\partial\mathfrak{J}_g(\boldsymbol{\beta})}{\partial\mu_0}\right)^2 + \frac{1}{2}\left(\frac{\partial\mathfrak{J}_g(\boldsymbol{\beta})}{\partial\sigma_0}\right)^2\right]\cdot\sigma_0^2, \tag{3}$$

with

$$\widehat{\mathfrak{J}}_g(\boldsymbol{\beta}) = \Phi\left(\frac{ab^2 - x_J(1-b^2)}{b(1-b^2)}\right) - \Phi\left(\frac{ab^2 + x_J(1-b^2)}{b(1-b^2)}\right) + \Phi\left(\frac{a + x_J(1-b^2)}{1-b^2}\right) - \Phi\left(\frac{a - x_J(1-b^2)}{1-b^2}\right)$$

and $\boldsymbol{\beta} = (\mu_1, \mu_0, \sigma_1, \sigma_0)$. This weak convergence and the properties of the normal distribution guarantee that, for a given $\alpha \in (0,1)$, then $(\widehat{\mathfrak{J}}_g - z_{\alpha/2}\cdot\hat{\sigma}_{J^P}\cdot n^{-1/2}, \widehat{\mathfrak{J}}_g + z_{\alpha/2}\cdot\hat{\sigma}_{J^P}\cdot n^{-1/2})$, where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ and $\hat{\sigma}^2_{J^P}$ is the sample version of $\sigma^2_{J^P}$, is a confidence interval for $\mathfrak{J}_g$ with asymptotic level $1-\alpha$.

3.2 Non-parametric assumption on biomarker distribution

If no underlying model is assumed for the biomarker, the empirical (non-parametric) estimator of $\mathfrak{J}_g$ is the resulting of replacing the (unknown) theoretical gROC curve in eq. (1) by its empirical estimator, $\widehat{\mathscr{R}}_g$. The properties of $\widehat{\mathscr{R}}_g$ have been already studied in Martínez-Camblor et al. [4]. Notice that, in practice, the obtention of $\widehat{\mathscr{R}}_g$ implies to numerically compute the values

$$\hat{\gamma}_t = \arg\max_{\gamma\in[0,1]}\left\{\widehat{\mathscr{R}}_r(\gamma\cdot t) + 1 - \widehat{\mathscr{R}}_r(1-t+\gamma\cdot t)\right\}$$

Automatically generated rough PDF by ProofCheck from River Valley Technologies Ltd

where $\hat{\mathscr{R}}_r$ is the empirical estimator of the standard ROC curve. The R package nsROC, freely available in the CRAN (www.r-project.org), provides point and interval estimators for the gROC curve [15]. Under certain regularity conditions of the theoretical gROC curve and assuming that $n/m \to_n \lambda > 0$ where n and m are the sample sizes in the positive and the negative samples, respectively, we have the following weak convergence (Theorem 2 in the Appendix provides full details)

$$\sqrt{n} \cdot \{\hat{\mathfrak{J}}_g - \mathfrak{J}_g\} \xrightarrow{\mathscr{L}}_n \sigma_J \cdot \mathscr{Z},$$

where $\sigma_J^2 = \mathscr{R}_g(t_Y)[1 - \mathscr{R}_g(t_Y)] + \lambda \cdot t_Y \cdot [1 - t_Y]$, with $t_Y = \arg\max_{t\in[0,1]}\{\mathscr{R}_g(t) - t\}$ and $\mathscr{Z}$ is a standard normal random variable.

The proposed empirical non-parametric estimator of $\mathfrak{J}_g$ has a bias of order $n^{-1/6}$, that is $\mathbb{E}[\sqrt{n} \cdot \{\hat{\mathfrak{J}}_g - \mathfrak{J}_g\}] = \mathscr{O}_{\mathscr{P}}\left(n^{-1/6}\right)$ (see the Appendix section for full details). Unfortunately, this bias is not negligible for usual sample sizes and confidence intervals based on the above approximation are extremely anti-conservative. The quality of the approximation depends on the underlying distributions. When the negative subjects are standard normally distributed, $\mathscr{N}(0,1)$, and the positive subjects follow the mixture $(1/2)\cdot\mathscr{N}(-1.61,1/4) + (1/2)\cdot\mathscr{N}(1.61,1/4)$ (for this configuration, the Youden index for the gROC curve is 0.57 and σ_J is 0.56) the values of the mean and the standard deviation (between brackets) of $\sqrt{n} \cdot \{\hat{\mathfrak{J}}_g - \mathfrak{J}_g\}$ are 0.34 (0.53), 0.27 (0.54), 0.26 (0.56) and 0.21 (0.56) for sample sizes ($n = m$) of 500, 1000, 2500 and 5000, respectively (based on 2000 Monte Carlo simulations). Figure 2 shows the histograms for the four considered cases.

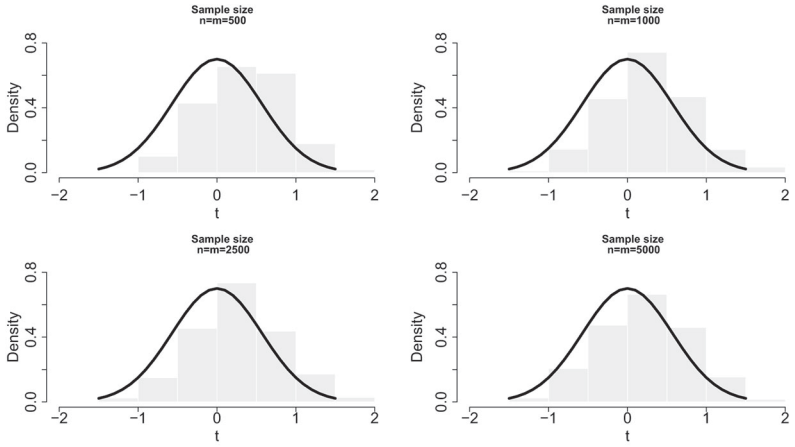

Figure 2: Convergence. Histograms for the values of $\mathbb{E}[\sqrt{n} \cdot \{\hat{\mathfrak{J}}_g - \mathfrak{J}_g\}]$ when the negative subjects are drawn from a standard normal distribution and the positive subjects follow the mixture $(1/2)\cdot\mathscr{N}(-1.61,1/4) + (1/4)\cdot\mathscr{N}(1.61,1/4)$. Black-line represents the asymptotic distribution.

We propose to base those confidence intervals on the next resampling algorithm which includes a procedure for bias correction.

S_1. Compute the empirical estimation of the Youden index, $\hat{\mathfrak{J}}_g$, based on the original samples from negative population, $\mathbf{Y}_0 = \{Y_{01},\cdots,Y_{0m}\}$, and the positive population $\mathbf{Y}_1 = \{Y_{11},\cdots,Y_{1n}\}$. Let $\hat{G}_m$ and $\hat{F}_n$ be the empirical distribution functions based on $\mathbf{Y}_0$ and $\mathbf{Y}_1$, respectively.

S_2. For $b = 1,\ldots,B$ (B large enough), draw two bootstrap random samples, $\mathbf{Y}_0^{*,b}$ of size m from $\hat{G}_m$ and $\mathbf{Y}_1^{*,b}$ of size n from $\hat{F}_n$, and compute the empirical (bootstrap) estimator for the Youden index, $\hat{\mathfrak{J}}_g^{*,b}$.

S_3. Estimate the bias by $\widehat{\text{Bias}} = \overline{\hat{\mathfrak{J}}_g^*} - \hat{\mathfrak{J}}_g$, where $\overline{\hat{\mathfrak{J}}_g^*} = B^{-1}\cdot\sum_{b=1}^B \hat{\mathfrak{J}}_g^{*,b}$. Notice that $\hat{\mathfrak{J}}_g^{BC} = [\hat{\mathfrak{J}}_g - \widehat{\text{Bias}}]$ is a bias-corrected estimator of $\hat{\mathfrak{J}}_g$ and then, for each $b = 1,\ldots,B$, $\hat{\mathfrak{J}}_g^{BC,*,b} = [\hat{\mathfrak{J}}_g^{*,b} - 2\cdot\widehat{\text{Bias}}]$ is a bias-corrected bootstrap replication for $\hat{\mathfrak{J}}_g^{BC}$.

S_4. Let $J_g(p)$ denote the $100\cdot p\%$ empirical percentile of the values $\{\hat{\mathfrak{J}}_g^{BC,*,1},\ldots,\hat{\mathfrak{J}}_g^{BC,*,B}\}$. Then the interval $\left(J_g(\alpha/2), J_g(1-\alpha/2)\right)$ is the $100\cdot(1-\alpha)\%$ bootstrap confidence interval for $\mathfrak{J}_g$.

In practice, for estimating $\mathfrak{J}_g$ is advisable to use the bias-corrected version estimator, $\hat{\mathfrak{J}}_g^{BC}$, instead of the original one, $\hat{\mathfrak{J}}_g$. Although Monte Carlo simulations show that some bias remains (see Table 4), the bias-corrected estimator is, of course, less biased than $\hat{\mathfrak{J}}_g$.

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

4 Associated cutoff point estimation

In the standard ROC curve context, each $t \in (0,1)$ has associated one particular value $x \in \mathbb{R}$ such that $\mathscr{R}_r(t) = 1 - F(x)$. Obviously, $x = G^{-1}(1-t)$ where $G^{-1}(s) = \inf\{z : G(z) \geq s\}$ and $[x, \infty)$ is the classification subset associated with t. In the gROC curve context, the classification subset depends on two points and, in general, the link between both points depends on each particular value of $t \in (0,1)$. In order to avoid the ambiguity of dealing with confidence intervals of a classification subset which depends on two different points, we consider here as associate cutoff point the value of t_Y which leads to the Youden index. We realize that even if we obtain a good approximation of t_Y, the associated classification subset might be not well approximated. However, the Glivenko-Cantelli Theorem guarantees that, for large enough samples, a good approximation of t_Y provides a good approximation for the classification subset.

4.1 Normally distributed biomarker

Under the parametric binormal assumption, the point that leads to the Youden index is

$$t_Y = \Phi\left(\frac{a - x_J(1-b^2)}{1-b^2}\right) + 1 - \Phi\left(\frac{a + x_J(1-b^2)}{1-b^2}\right),$$

where x_J is defined in eq. (2). Hence, the parametric estimator of t_Y is

$$\hat{t}_Y^P = \Phi\left(\frac{\hat{a} - \hat{x}_J(1-\hat{b}^2)}{1-\hat{b}^2}\right) + 1 - \Phi\left(\frac{\hat{a} + \hat{x}_J(1-\hat{b}^2)}{1-\hat{b}^2}\right),$$

where $\hat{a}$, $\hat{b}$ and $\hat{x}_J$ are the sample versions of a, b and x_J, respectively, and were defined above. Applying again the delta method we obtain the weak convergence

$$\sqrt{n} \cdot \{\hat{t}_Y^P - t_Y\} \xrightarrow{\mathscr{L}}_n \sigma_{t_Y^P} \cdot \mathscr{Z},$$

where $\mathscr{Z}$ is a standard normal random variable and

$$\sigma^2_{t_Y^P} = \left[\left(\frac{\partial t_Y}{\partial \mu_1}\right)^2 + \frac{1}{2}\left(\frac{\partial t_Y}{\partial \sigma_1}\right)^2\right] \cdot \sigma_1^2 + \lambda \cdot \left[\left(\frac{\partial t_Y}{\partial \mu_0}\right)^2 + \frac{1}{2}\left(\frac{\partial t_Y}{\partial \sigma_0}\right)^2\right] \cdot \sigma_0^2. \tag{4}$$

More details are provided in Theorem 1 in the Appendix. For a given $\alpha \in (0,1)$, the direct substitution of $\sigma^2_{t_Y^P}$ by its sample version, $\hat{\sigma}^2_{t_Y^P}$, and the properties of the normal distribution guarantee that the interval $(\hat{t}_Y^P - z_{\alpha/2} \cdot \hat{\sigma}_{t_Y^P} \cdot n^{-1/2}, \hat{t}_Y^P + z_{\alpha/2} \cdot \hat{\sigma}_{t_Y^P} \cdot n^{-1/2})$, where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$, is a confidence interval for t_Y with asymptotic level $1-\alpha$.

4.2 Non-parametric assumption on biomarker distribution

Regarding the non-parametric approach, the direct empirical estimator for t_Y is

$$\hat{t}_Y = \arg\max_{t \in [0,1]} \{\hat{\mathscr{R}}_g(t) - t\}.$$

When the underlying gROC curve satisfies certain regularity assumptions and $n/m \to_n \lambda > 0$, it is proved the next weak convergence (see Theorem 2 in the Appendix for full details):

$$\left(\frac{r_g'(t_Y)^2}{4 \cdot (1+\lambda)}\right)^{1/3} \cdot n^{1/3} \cdot \{\hat{t}_Y - t_Y\} \xrightarrow{\mathscr{L}}_n \arg\max_{z \in \mathbb{R}} \{\mathscr{Z}^{(n)}(z) - z^2\},$$

where r_g' is the second derivative of $\mathscr{R}_g$ and $\{\mathscr{Z}^{(n)}(z)\}_{\{-\infty < z < \infty\}}$ is a two-sided Brownian motion. In practice, using this expression for computing asymptotic confidence intervals requires not just the approximation of the distribution of $\arg\max_{z \in \mathbb{R}} \{\mathscr{Z}^{(n)}(z) - z^2\}$ but the approximation of r_g'. The estimation of derivatives is a complex problem which usually involves the previous selection of a tuning parameter that makes the inference process difficult [16]. Alternatively, the bootstrap algorithm S_1-S_4 can be adapted to approximate the distribution of $\hat{t}_Y$. Particularly, for each $b \in \{1,\ldots,B\}$, in S_2 we should also save the values where the bootstrap samples achieve the Youden index, $\hat{t}_Y^{*,b}$, and then, in S_4 compute also a $100 \cdot (1-\alpha)\%$ confidence interval based on $\{\hat{t}_Y^{*,1}, \cdots, \hat{t}_Y^{*,B}\}$.

Automatically generated rough PDF by ProofCheck from River Valley Technologies Ltd

5 Monte Carlo simulations

The finite-sample behavior of both the parametric binormal and the bootstrap-based empirical interval estimators of $\mathfrak{J}_g$ and t_Y are studied by means of Monte Carlo simulations. We consider three different scenarios. In each scenario, the distributions in the positive and negative groups follow particular parametric models. The values of the corresponding parameters are chosen in order to obtain certain values of gAUC, indicated as $\mathscr{A}_g$ in the tables. Three different $\mathscr{A}_g$ values (0.7, 0.8 and 0.9) are considered for each scenario. Values of $a = (\mu_1 - \mu_0)/\sigma_0$ and $b = \sigma_1/\sigma_0$ are also provided. For each studied situation we consider different sample sizes. We report the observed coverage percentages (C) and the average length (L) for the 95% confidence intervals obtained using the two proposed procedures and estimated from 2000 Monte Carlo iterations. For the non-parametric procedure, the confidence intervals are based on 200 bootstrap replications ($B = 200$). This number is large enough in order to estimate averages and percentages. The provided results are similar for other choices of B (data not shown). However, when we deal with a particular real-problem, to consider a larger number of bootstrap replications is advisable in order to have a more precise approximation.

In the first scenario (Normal-Normal), we take normal distributions in both negative and positive populations. Specifically, the observations are drawn from a standard normal distribution in the negative population (sample size m) and from a normal distribution with mean a and standard deviation b in the positive population (sample size n). The considered values for (a,b) are (0.53,1.85), (1.45,2.55) and (4.27,3.69). In this scenario the values of the Youden index are 0.31, 0.49 and 0.71. The points, t_Y, that lead to these values are 0.178, 0.115 and 0.053, respectively.

In the second scenario (Normal-Gamma), the m negative subjects are drawn from a standard normal distribution. The n positive subjects are drawn from a re-centered gamma distribution, specifically, $(Y|D = 1) + 4$ follows a gamma distribution with parameters (δ, β). The chosen values for (δ, β) are (5.50,1.59), (3.12,0.50), and (3.13,0.32). The values of the Youden index are 0.31, 0.49 and 0.71. The points, t_Y, that lead to these values are 0.230, 0.107 and 0.045, respectively.

In the third scenario (Normal-Bimodal), the m negative subjects are drawn from a standard normal distribution. The n positive subjects are drawn from the mixture of the form $(1/2) \cdot \mathcal{N}(-a, 1/4) + (1/4) \cdot \mathcal{N}(a, 1/4)$. Values of a are 1.15, 1.43 and 1.84. The values of the Youden index are 0.34, 0.49 and 0.67. The points, t_Y, that lead to these values are 0.441, 0.340 and 0.220, respectively.

Figure 3 shows the shape of the considered gROC curves (left) and the function $(\mathscr{R}_g(t) - t)$ (right) for the three different situations considered in the three scenarios described above. In all cases the value where the Youden index is achieved is highlighted.

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

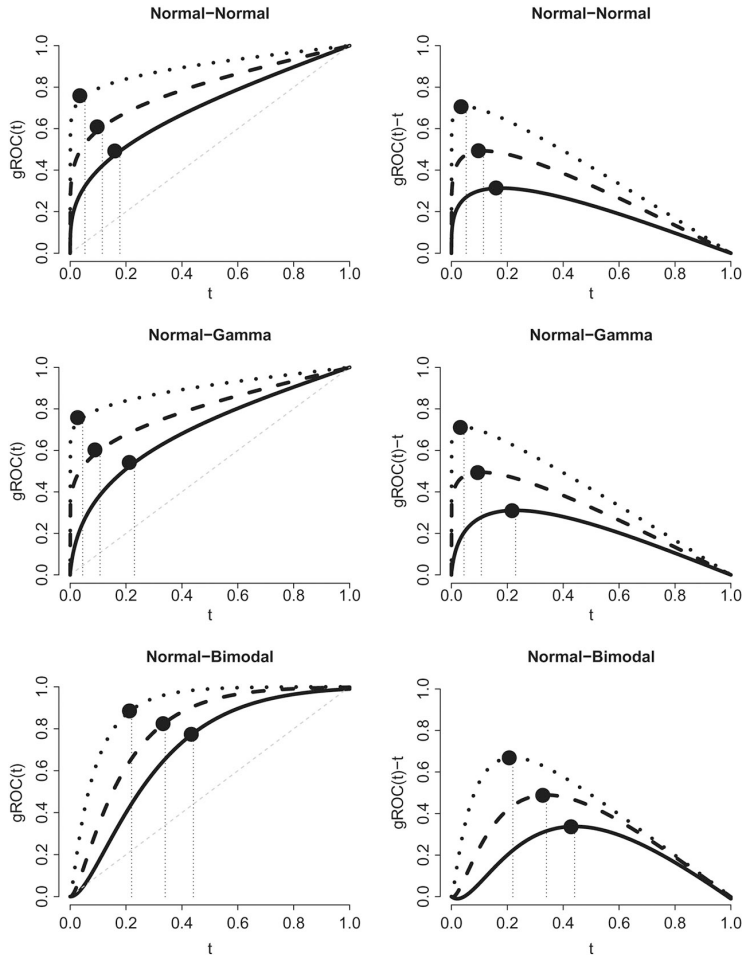

Figure 3: Left: gROC curves for the three considered scenarios. Right: functions $\mathscr{R}_g(t) - t$ ($t \in [0,1]$ for the three considered scenarios. In all cases, the point where the Youden index is is achieved is highlighted.

Table 1 shows the results regarding the first considered scenario (Normal-Normal). The binormal parametric estimators of both the Youden index and the associated threshold (1−specificity) report good results although a little bit conservative. The non-parametric bootstrap approximations obtain coverage percentage around the 90% for the Youden index and around 98% for its associated cutoff point. The lengths of the non-parametric intervals for t_Y are extremely large and even less informative for gROC curves with lower $\mathscr{A}_g$. It should be noted that, when the studied marker has a limited accuracy, there is a wide range of thresholds reporting values of $(\mathscr{R}_g(t) - t)$ close to the Youden index. Particularly, in the first considered case ($\mathscr{A}_g = 0.7$), values of t between 0.099 and 0.287 achieve values of $(\mathscr{R}_g(t) - t)$ above 0.30 (the Youden index is 0.31). That makes difficult to have precise non-parametric estimations of t_Y.

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

Table 1: Normal-Normal. Coverage proportions (C) and length mean (L) for 95% confidence intervals based on the asymptotic parametric binormal and the bootstrap (200 replications) non-parametric approaches for both the Youden index and the associated t_Y (1−specificity) value. The sample sizes are n positive and m negative subjects. Values of a and b are reported.

					Youden index				1-Specificity			
					Binormal		Bootstrap		Binormal		Bootstrap	
n	m	a	b	$\mathcal{A}_g$	C	L	C	L	C	L	C	L
100	100	0.53	1.85	0.7	0.960	0.18	0.884	0.21	0.969	0.07	0.999	0.34
	300				0.961	0.14	0.892	0.19	0.966	0.06	0.996	0.31
200	200				0.958	0.13	0.914	0.16	0.957	0.05	0.997	0.27
	600				0.957	0.10	0.904	0.14	0.955	0.04	0.995	0.25
100	100	1.45	2.55	0.8	0.970	0.17	0.883	0.20	0.957	0.05	0.987	0.21
	300				0.967	0.13	0.904	0.18	0.964	0.04	0.991	0.20
200	200				0.971	0.12	0.900	0.15	0.961	0.04	0.989	0.18
	600				0.968	0.09	0.916	0.13	0.963	0.03	0.992	0.15
100	100	3.69	4.27	0.9	0.982	0.13	0.886	0.17	0.958	0.03	0.970	0.11
	300				0.973	0.10	0.904	0.16	0.974	0.02	0.983	0.10
200	200				0.984	0.09	0.910	0.12	0.980	0.02	0.982	0.09
	600				0.972	0.07	0.920	0.11	0.970	0.02	0.984	0.08

Table 2 reports the results obtained in the second scenario (Normal-Gamma). Although the resulting gROC curves are quite similar to those considered in the Scenario 1 (see Figure 3) the lack of the binormality in the underlying distributions makes that the parametric binormal estimators do not work. They obtain erratic results in the different models. The accuracy of the results depends on how close the real underlying quantities and the ones estimated by the parametric binormal model are. Non-parametric procedure obtains similar results to those reported in the previous scenario: interval confidence intervals are anti-conservative for the Youden index and conservative for its associated cutoff point.

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

Table 2: Normal-Gamma. Coverage proportions (C) and length mean (L) for 95% confidence intervals based on the asymptotic parametric binormal and the bootstrap (200 replications) non-parametric approaches for both the Youden index and the associated t_Y (1-specificity) value. The sample sizes are n positive and m negative subjects. Values of a and b are reported.

					Youden index				1-Specificity			
					Binormal		Bootstrap		Binormal		Bootstrap	
n	m	a	b	$\mathcal{A}_g$	C	L	C	L	C	L	C	L
100	100	−0.15	1.47	0.7	0.833	0.20	0.894	0.22	0.846	0.10	0.994	0.36
	300				0.691	0.16	0.898	0.19	0.834	0.08	0.990	0.32
200	200				0.584	0.14	0.906	0.16	0.782	0.06	0.990	0.29
	600				0.334	0.11	0.919	0.14	0.685	0.05	0.991	0.25
100	100	2.24	3.53	0.8	0.166	0.14	0.882	0.20	0.258	0.04	0.985	0.21
	300				0.058	0.12	0.909	0.18	0.122	0.03	0.992	0.19
200	200				0.018	0.10	0.915	0.14	0.058	0.03	0.987	0.17
	600				0.005	0.08	0.915	0.13	0.020	0.02	0.994	0.16
100	100	5.79	5.53	0.9	0.140	0.11	0.898	0.16	0.420	0.02	0.974	0.11
	300				0.045	0.09	0.911	0.15	0.247	0.02	0.988	0.10
200	200				0.011	0.08	0.903	0.12	0.163	0.02	0.984	0.09
	600				0.001	0.06	0.923	0.11	0.041	0.01	0.982	0.08

Table 3 depicts the observed results in the third scenario (Normal-Bimodal). Again, the results show the strong dependency of the parametric binormal procedure on the underlying distributional assumption. The results are again erratic and the coverage percentage is close to zero in most of the considered situations. The procedure does not estimate the real parameters but what these parameters would be if the underlying distributions were normal. Both situations can be very different, and in this scenario, they actually are. Non-parametric procedure obtains similar results to the observed in the two previous scenarios. In this case, the confidence intervals are anti-conservative for the Youden index (coverage percentages around 89%) and slightly anti-conservative in some of considered cases when we estimate its associated cutoff point (coverage percentages around 93%).

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

Table 3: Normal-Bimodal. Coverage proportions (C) and length mean (L) for 95% confidence intervals based on the asymptotic parametric binormal and the bootstrap (200 replications) non-parametric approaches for both the Youden index and the associated t_Y (1-specificity) value. The sample sizes are n positive and m negative subjects. Values of a are reported.

				Youden index				1-Specificity			
				Binormal		Bootstrap		Binormal		Bootstrap	
n	m	a	$\mathcal{A}_g$	C	L	C	L	C	L	C	L
100	100	1.15	0.7	0.010	0.19	0.892	0.22	0.027	0.11	0.896	0.33
	300			0.001	0.15	0.881	0.17	0.002	0.07	0.947	0.27
200	200			0.001	0.13	0.904	0.16	0.002	0.06	0.927	0.26
	600			0.001	0.11	0.906	0.14	0.001	0.05	0.935	0.24
100	100	1.43	0.8	0.001	0.18	0.884	0.21	0.001	0.08	0.903	0.26
	300			0.001	0.15	0.905	0.15	0.001	0.06	0.938	0.21
200	200			0.001	0.13	0.910	0.15	0.001	0.06	0.941	0.21
	600			0.001	0.10	0.917	0.12	0.001	0.05	0.960	0.17
100	100	1.84	0.9	0.001	0.17	0.881	0.17	0.094	0.07	0.895	0.19
	300			0.001	0.14	0.879	0.13	0.106	0.06	0.939	0.15
200	200			0.001	0.12	0.892	0.13	0.077	0.05	0.929	0.15
	600			0.001	0.10	0.899	0.10	0.002	0.04	0.957	0.12

Finally, Table 4 reports the bias (B) and the mean-square error (MSE) observed in the Monte Carlo simulations for: $\sqrt{n} \cdot \{\widehat{\mathfrak{Y}}_g - \mathfrak{Y}_g\}$ (Uncorrected), $\sqrt{n} \cdot \{\widehat{\mathfrak{Y}}_g^{BC} - \mathfrak{Y}_g\}$ (BC) and for the points leading to the Youden index, u_L and u_U, for the three considered models. Results confirm the presence of some remaining bias in the bias-corrected version and how both the bias and the mean-square error decrease when the sample size increases and with the value of the Youden index as well. Besides, both of them are smaller for the bias-corrected version (BC). The estimation of the associated points are correct and almost unbiased. Mean-square errors below 0.01 are reported by this number.

Table 4: Additional information. Bias (B) and the mean-square error (MSE) observed in the Monte Carlo simulations for: $\sqrt{n} \cdot \{\widehat{\mathfrak{Y}}_g - \mathfrak{Y}_g\}$ (Uncorrected), $\sqrt{n} \cdot \{\widehat{\mathfrak{Y}}_g^{BC} - \mathfrak{Y}_g\}$ (BC) and for the points leading to the Youden index, u_L and u_U. Values below 0.01 are reported as 0.01.

		Youden index		Interval	
		Uncorrected	BC	u_L	u_U
$n = m$	$\mathscr{A}_g$	B [MSE]	B [MSE]	B [MSE]	B [MSE]
Normal-Normal					
100	0.7	0.623 [0.387]	0.273 [0.075]	−0.061 [0.039]	0.079 [0.063]
200		0.513 [0.262]	0.176 [0.034]	−0.046 [0.023]	0.074 [0.052]
100	0.8	0.445 [0.196]	0.185 [0.033]	−0.086 [0.014]	0.089 [0.014]
200		0.403 [0.163]	0.152 [0.019]	−0.048 [0.010]	0.088 [0.013]
100	0.9	0.315 [0.009]	0.146 [0.021]	−0.058 [0.013]	0.060 [0.008]
200		0.271 [0.073]	0.100 [0.011]	−0.054 [0.011]	0.069 [0.006]
Normal-Gamma					
100	0.7	0.580 [0.341]	0.247 [0.057]	0.077 [0.013]	0.080 [0.013]
200		0.481 [0.232]	0.165 [0.031]	0.002 [0.011]	0.052 [0.012]
100	0.8	0.486 [0.236]	0.220 [0.053]	0.049 [0.012]	0.124 [0.019]
200		0.423 [0.181]	0.169 [0.028]	−0.033 [0.009]	0.118 [0.014]
100	0.9	0.277 [0.076]	0.115 [0.012]	−0.032 [0.006]	−0.111 [0.008]
200		0.237 [0.063]	0.079 [0.007]	−0.018 [0.005]	0.015 [0.006]
Normal-Bimodal					
100	0.7	0.583 [0.336]	0.269 [0.067]	−0.092 [0.012]	0.093 [0.013]
200		0.457 [0.212]	0.149 [0.023]	−0.056 [0.011]	0.074 [0.011]
100	0.8	0.476 [0.227]	0.210 [0.041]	−0.074 [0.009]	0.097 [0.010]
200		0.443 [0.201]	0.181 [0.032]	−0.047 [0.008]	0.069 [0.009]
100	0.9	0.364 [0.134]	0.153 [0.019]	−0.086 [0.006]	0.083 [0.008]
200		0.303 [0.085]	0.090 [0.006]	−0.014 [0.005]	0.074 [0.006]

6 Real-world example

With the objective of knowing the ability of the white blood cells count (WBC) for identifying the type of disease in patients having either acute viral meningitis (AVM) or acute bacterial meningitis (ABM), we consider the 367 (183 AVM and 184 ABM) subjects with this information out of the 581 total included in the original study [17]. The used dataset is freely available at http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets and reports information related to the study made at the Duke University Medical Center and whose main original goal was to establish a multivariate predictive model for classifying these patients. The WBC distribution in both the AVM and ABM patients is depicted in Figure 4. The observed differences are transferred to both the means and standards deviations: 10.17 ± 3.7 for AVM and 16.15 ± 10.6 for ABM patients.

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

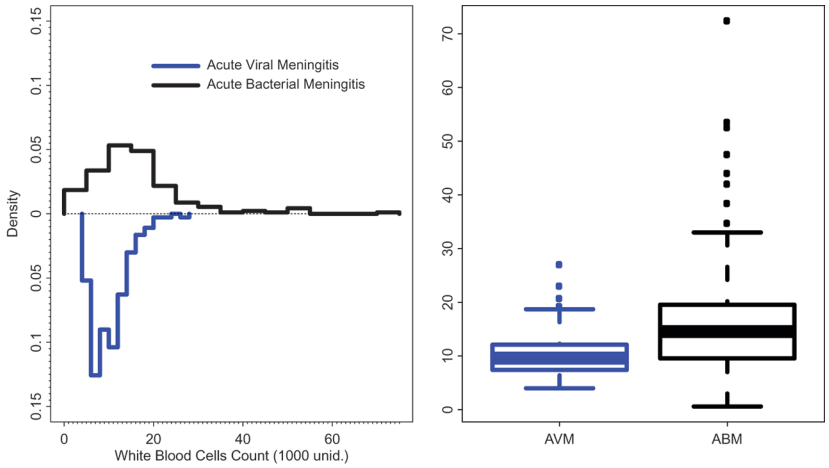

Figure 4: Meningitis dataset. White blood cells count (WBC) in type of meningitis diagnosis. Left: histograms of WBC by type of meningitis. Right: boxplots by type of meningitis.

Assuming that both the negative and the positive subjects are normally distributed ($\mathscr{A}_g = 0.82$), the Youden index is 0.533 ($= \hat{\mathfrak{J}}_g^P$) and it is achieved at $\hat{t}_Y^P = 0.101$. The classification subset associated with this point is $(0, 3.18] \cup [15.53, \infty)$. The non-parametric estimation ($\mathscr{A}_g = 0.79$) reports a Youden index estimation of 0.477 ($= \hat{\mathfrak{J}}_g$), 0.465 after bias correction, achieved at $\hat{t}_Y = 0.115$ and with an associated classification subset $(0, 4.50] \cup [14.10, \infty)$. Figure 5 shows both the parametric binormal and the non-parametric gROC curves and the classification subsets on which they are based on. The Youden indices and the respective associated classification subsets are highlighted. Notice that the x-axis in the classification subsets plot ranges from 0 to 30, covering all negative subjects, however the maximum value of WBC observed in the positive patients is 72.4. Fourteen positive patients (7.6%) have WBC values above 30.

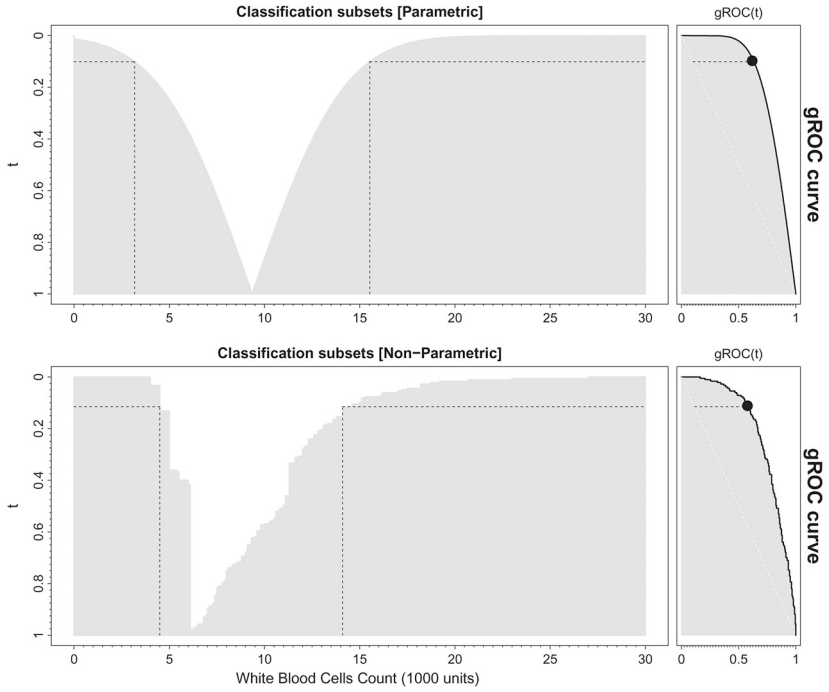

Figure 5: Meningitis dataset. Classification subset (left) and gROC curves (right) for both parametric binormal (top) and non-parametric (bottom) estimations. Youden index and associated classification subsets are highlighted.

The 95% confidence interval based on the parametric binormal distribution are (0.474, 0.591) and (0.083, 0.119) for the Youden index and the associated 1−specificity, respectively. In the non-parametric setting, based on 5000 bootstrap replications, the 95% confidence interval for the Youden index is (0.419, 0.511) and (0.061, 0.225) for the associated 1−specificity. Figure 6 depicts the parametric and the non-parametric kernel density estimations for both the positive and the negative groups (top) and the parametric and non-parametric estimations for $\text{gROC}(t) - t$ (bottom). The Youden index and 95% confidence intervals for both parametric binormal and non-parametric approaches are highlighted.

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

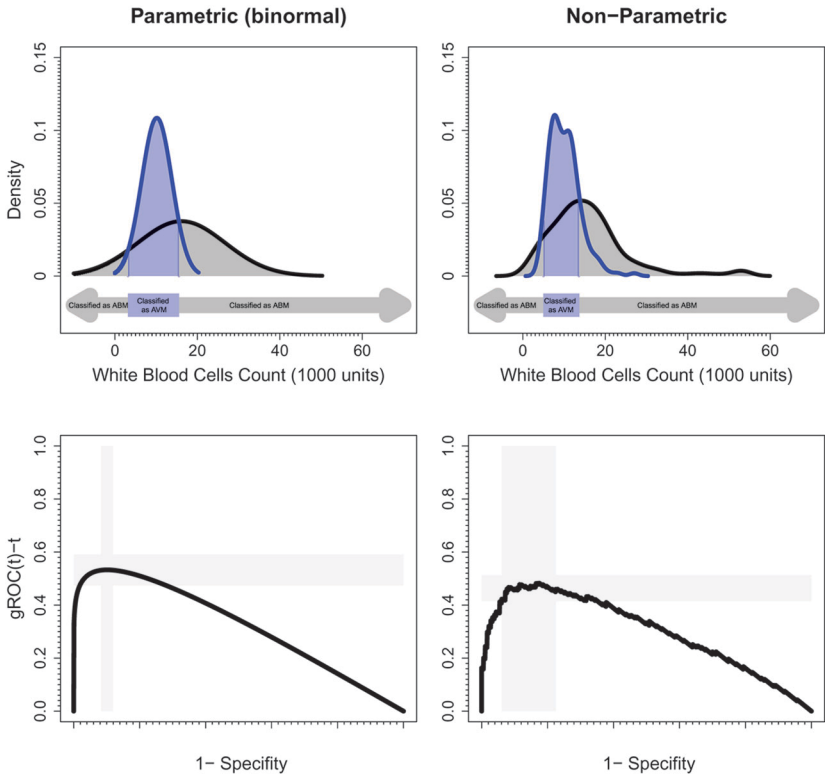

Figure 6: Parametric binormal and non-parametric kernel density estimations for both the positive and the negative groups (top) and the binormal parametric and non-parametric estimations for $\text{gROC}(t) - t$ (bottom). Youden index and 95% confidence intervals for both parametric and non-parametric approaches are highlighted (gray bands).

7 Discussion

Choosing a a classification rule is a crucial task in the diagnosis processes. Although particular problems have special requirements (for instance, screenings test are supposed to have large sensitivities and diagnosis implying aggressive treatments should need large specificities), the Youden index is a useful tool frequently employed for both the automatic selection of the classification rule and to measure the classification accuracy of the studied marker.

The estimation of both the Youden index and its associated cutoff point is a complex task. That complexity justifies the number of recent papers which still concern on their estimation and other related problems such as testing or comparisons [18]. In this paper, we contribute to the literature by working in the gROC curve context. We have studied both parametric binormal and empirical non-parametric Youden estimators and their associated cutoff point when the relationship between the studied characteristic and the marker is not monotone. Stricollaby speaking, in this context, we do not have a cutoff point but a classification subset which leads to the Youden index. Since each t (1−specificity) value has associated only one of those subsets, for simplicity, we consider the estimation of $t_Y = \arg\max_{t\in[0,1]}\{\mathscr{R}_g(t) - t\}$.

The parametric binormal estimator is strongly connected with its ROC curve context counterpart. Both the Youden index and the associated cutoff point can be directly estimated from the means and standard deviations of the underlying normal distributions and therefore, their accuracy does not depend on the shape of the resulting gROC curve but on the plausibility of the binormal assumption. Monte Carlo simulations show that the proposed parametric procedures obtain good results for the binormal models, but they should not be used when this assumption is violated. Even in situations where the obtained gROC curve is similar to the one derived from a particular binormal model (second simulated scenario), the observed results were erratic, and the parametric procedure did not work satisfactorily. We should note that, following the work of Lai, Tian and Schisterman [19], the current binormal approximation can potentially be improved in order to deal with smaller sample sizes.

The direct empirical non-parametric estimator has a non-negligible bias for small and moderate sample sizes. The proposed bootstrap algorithm improves its behavior, but observed coverage percentages are around 90%, suggesting that some bias remains. An interesting future work can be the application of different methodologies proposed for the computation of non-parametric confidence intervals for the Youden index in the stan-

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

dard ROC curve contexts to the non-monotone relationship case [20, 21]. Besides, the implementation of other traditional procedures for bias-correction, such as the BCa (explicitly considered in the ROC context in [1]), could be explored as well. On the other hand, reported confidence intervals for the associated threshold are, depending on the underlying gROC curve, not very informative and always conservative, with coverage percentages around 98%. The first problem is intrinsic to the computation of the Youden index for markers with low classification capacity. In this case, a wide range of thresholds achieve similar values of $\mathscr{R}_g(t) - t$ and that makes that two different and good approximations to the real gROC curve can report similar Youden index but quite different thresholds. The second problem can be mitigated by making a specific smoothed bootstrap algorithm for computing the confidence intervals for the cutoff point. Notice that the asymptotic distribution of $\hat{t}_Y$ depends on the second derivative of the real underlying gROC curve and hence the smoothed bootstrap might be advisable as a resampling plan [22]. In this work, we have preferred to keep the same resampling plan for both estimators in order to gain coherence and simplicity.

Our real-world application considers a problem in which the distributions of both the positive and the negative subjects seem to be not far from the normal distribution (although Shapiro-Wilk test for normality rejects the normality in both cases with p-values below 0.001). However, from what we have learned from the simulation study, we know that the parametric binormal approximation might yield too optimistic results. It seems that violation of the normality assumption is clear enough for not being confident on the parametric binormal estimators and therefore preferring the more conservative non-parametric procedures.

In short, the obtained results suggest that the Youden index can be a good approach for summarizing the accuracy of a biomarker with a non-monotone relationship with the studied characteristic. The adaptation of other approximations related to the Youden index in the standard ROC curve context [19–21] or other criteria for selecting thresholds [9] to the gROC curve setting and the comparison with the proposed Youden index are interesting goals for future research.

Acknowledgements

This work is financially supported by the Grants MTM2014-55966-P and MTM2017-89422-P Spanish Ministry of Economy, Industry and Competitiveness; State Research Agency; and FEDER funds. J.C. Pardo-Fernández also acknowledges funding from Banco Santander and Complutense University of Madrid (project PR26/16-5B-1). P-MC is also supported by the Grant FC-GRUPIN- IDI/2018/000132 of the Asturies Goverment. The authors thank too anonymous reviewers for their constructive comments and suggestions.

Appendix

In this Appendix we formalize the proofs of the anticipated asymptotic distributions of both the parametric binormal and the non-parametric (empirical) estimators of the Youden index and its associated cutoff points in the gROC curve context. Results for the parametric estimator (Theorem 1) are similar to those derived for the Youden index in the standard ROC curve context [23]. Theorem 2 deals with the empirical estimator. The structure of the proof is similar to the one used by Hsieh and Turnbull [24] in the standard ROC curve context.

Theorem 1.

Let Y_0 and Y_1 denote the biomarker in the negative population and in the positive population, respectively. Assume that Y_i is normally distributed with mean μ_i and standard deviation σ_i, for $i \in \{0,1\}$. Assume that two independent samples of m of i.i.d. observations from Y_0 and n i.i.d. observations from Y_1 are available. The sample sizes satisfy

A_1.-$n/m \to_n \lambda > 0$.

Then, we have the following weak convergences,

i. $\sqrt{n} \cdot \{\hat{\mathfrak{J}}_g^P - \mathfrak{J}_g\} \xrightarrow{\mathscr{L}}_n \sigma_{J^P} \cdot \mathscr{Z}$ and,

ii. $\sqrt{n} \cdot \{\hat{t}_Y^P - t_Y\} \xrightarrow{\mathscr{L}}_n \sigma_{t_Y^P} \cdot \mathscr{Z}$,

where $\mathscr{Z}$ is a standard normal random variable and σ_{J^P} and $\sigma_{t_Y^P}$ are given in eqs. (3) and (4), respectively.

Proof.

Results in [14] allow to derive that the binormal estimation of the gROC curve are based on intervals of the form $(-\infty, a/(1-b^2) - x] \cup [a/(1-b^2) + x, \infty)$, where x is a non-negative real number, $a = (\mu_1 - \mu_0)/\sigma_0$ and

Automatically generated rough PDF by ProofCheck from River Valley Technologies Ltd

$b = \sigma_1/\sigma_0$. Then, direct calculations lead to deduce that the interval leading to the the value of 1−specificity for $\mathfrak{J}_g$ is achieved is of the form

$$(1-b^2)^{-1}\cdot(a+b\sqrt{2(b^2-1)\log(b)+a^2},\ a-b\sqrt{2(b^2-1)\log(b)+a^2}).$$

These points are the ones where the normal densities associated to the positive and negative populations cross each other. Therefore, the Youden index is

$$\begin{aligned}\mathfrak{J}_g &= \left\{\Phi\left(\frac{ab^2-x_J(1-b^2)}{b(1-b^2)}\right)-\Phi\left(\frac{ab^2+x_J(1-b^2)}{\hat{b}(1-b^2)}\right)+\Phi\left(\frac{a+x_J(1-b^2)}{1-b^2}\right)-\Phi\left(\frac{a-x_J(1-b^2)}{1-b^2}\right)\right\}\\ &= \mathfrak{J}_g(\mu_1,\sigma_1,\mu_0,\sigma_0).\end{aligned}$$

Let $\hat{\xi}_i$ and $\hat{S}_i$ be the sample mean and the sample standard deviation, which estimate μ_i and σ_i, respectively ($i \in \{0,1\}$). Then, we directly have that $\hat{\mathfrak{J}}_g^P = \mathfrak{J}_g(\hat{\xi}_1,\hat{S}_1,\hat{\xi}_0,\hat{S}_0)$. In addition, we have the following convergences in distribution:

$$\begin{aligned}&\sqrt{n}\cdot(\hat{S}_1-\sigma_1)\xrightarrow[n]{\mathscr{L}}\mathcal{N}(0,\sigma_1^2/2),\\ &\sqrt{n}\cdot(\hat{S}_0-\sigma_0)\xrightarrow[n]{\mathscr{L}}\mathcal{N}(0,\sigma_0^2\cdot\lambda/2),\\ &\sqrt{n}\cdot(\hat{\xi}_1-\mu_1)\xrightarrow[n]{\mathscr{L}}\mathcal{N}(0,\sigma_1^2),\\ &\sqrt{n}\cdot(\hat{\xi}_0-\mu_0\xrightarrow[n]{\mathscr{L}}\mathcal{N}(0,\sigma_0^2\cdot\lambda).\end{aligned}$$

Besides, the independence among the four random variables is also well-known [25]. Hence, the multivariate delta method ensures the weak convergence stated in *(i)*.

In order to prove (*ii*), we just consider the value of 1−specificity for which the point associated to the Youden index is achieved,

$$t_Y = \Phi\left(\frac{a-x_J(1-b^2)}{1-b^2}\right)+1-\Phi\left(\frac{a+x_J(1-b^2)}{1-b^2}\right) = t_Y(\mu_1,\mu_0,\sigma_1,\sigma_0).$$

Arguing as above, and taking into account that $\hat{t}_Y^P = t_Y(\hat{\xi}_1,\hat{S}_1,\hat{\xi}_0,\hat{S}_0)$, the result can be obtained again by applying the multivariate delta method. □

Theorem 2.

Let Y_0 and Y_1 be the random variables modelling the biomarker behavior in the positive and negative and positive populations, respectively. Assume that the resulting gROC curve, $\mathscr{R}_g$, satisfies

A_2.-$\mathscr{R}_g$ has two continuous derivatives on some subinterval $(a_0,b_0)\subset[0,1]$, such as $t_Y\in(a_0,b_0)$, where $t_Y = \arg\max_{t\in[0,1]}\{\mathscr{R}_g(t)-t\}$.

A_3.-$|r_g'(t_Y)| = a > 0$, where $r_g'(t) = \partial r_g(t)/\partial t$ and $r_g(t) = \partial\mathscr{R}_g(t)/\partial t$.

Assume that two independent samples of m of i.i.d. observations from Y_0 and n i.i.d. observations from Y_1 are available.

Then, if the sample sizes satisfy A_1, the following weak convergences hold

i. $\sqrt{n}\cdot\{\hat{\mathfrak{J}}_g-\mathfrak{J}_g\}\xrightarrow[n]{\mathscr{L}}\sigma_J\cdot\mathcal{Z}$ and

ii. $\left(\frac{r'(t_Y)^2}{4\cdot(1+\lambda)}\right)^{1/3}\cdot n^{1/3}\cdot\{\hat{t}_Y-t_Y\}\xrightarrow[n]{\mathscr{L}}\arg\max_{z\in\mathbb{R}}\{\mathcal{Z}^{(n)}(z)-z^2\}$,

where $\mathcal{Z}$ is a standard normal random variable, $\{\mathcal{Z}^{(n)}(z)\}_{\{-\infty<z<\infty\}}$ is a two-sided Brownian motion and $\sigma_J^2 = \mathscr{R}_g(t_Y)[1-\mathscr{R}_g(t_Y)]+\lambda\cdot t_Y\cdot[1-t_Y]$.

Proof

Let $\mathscr{C}[\mathscr{R}_g]$ the subset containing all the pairs (u_L,u_U) such that there exists $t\in[0,1]$ satisfying that $\mathscr{R}_g(t) = F(u_L)+1-F(u_U)$. Then,

$$\hat{\mathfrak{J}}_g = \max_{(u_L,u_U)\in\mathscr{C}[\hat{\mathscr{R}}_g]}\{\hat{F}_n(u_L)-\hat{F}_n(u_U)+\hat{G}_m(u_U)-\hat{G}_m(u_L)\}.$$

Let (x_L,x_U) be the pair of points that leads to $\mathfrak{J}_g$, that is, $\mathfrak{J}_g = F(x_L)-F(x_U)+G(x_U)-G(x_L)$. Therefore,

$$\begin{aligned}&\{\hat{\mathfrak{J}}_g-\mathfrak{J}_g\}\\ &\quad=\max_{(u_L,u_U)\in\mathscr{C}[\hat{\mathscr{R}}_g]}\{[\hat{F}_n(u_L)-\hat{F}_n(u_U)+\hat{G}_m(u_U)-\hat{G}_m(u_L)]-[F(x_L)-F(x_U)+G(x_U)-G(x_L)]\}\\ &\quad=\hat{\mathfrak{Z}}_n(x_L,x_U)+\max_{(u_L,u_U)\in\mathscr{C}[\hat{\mathscr{R}}_g]}\{\hat{\mathscr{H}}_n(u_L,u_U,x_L,x_U)-\mathscr{H}(u_L,u_U,x_L,x_U)+\mathscr{H}(u_L,u_U,x_L,x_U)\},\end{aligned}\tag{5}$$

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

where

$$\hat{\mathfrak{Z}}_n(u,v) = [\hat{F}_n(u) - \hat{F}_n(v) + \hat{G}_m(v) - \hat{G}_m(v)] - [F(u) - F(v) + G(v) - G(u)]$$
$$\hat{\mathscr{H}}_n(v,w,x,z) = [\hat{F}_n(v) - \hat{F}_n(w)] - [\hat{F}_n(x) - \hat{F}_n(z)] + [\hat{G}_m(w) - \hat{G}_m(v)] - [\hat{G}_m(z) - \hat{G}_m(x)]$$
$$\mathscr{H}(v,w,x,z) = [F(v) - F(w)] - [F(x) - F(z)] + [G(w) - G(v)] - [G(z) - G(x)].$$

On the one hand, assumption A_2 allows to apply the Hungarian embedding [26] to derive that the random variables $\hat{\mathfrak{Z}}_n(x_L, x_U)$ and

$$n^{-1/2} \cdot \{\mathscr{B}_1^{(n)}(F(x_L)) - \mathscr{B}_1^{(n)}(F(x_U))\} + m^{-1/2} \cdot \{\mathscr{B}_2^{(m)}(G(x_L)) - \mathscr{B}_2^{(m)}(G(x_U))\}, \tag{6}$$

where $\{\mathscr{B}_1^{(n)}(t)\}_{\{0\le t\le 1\}}$ and $\{\mathscr{B}_2^{(m)}(t)\}_{\{0\le t\le 1\}}$ are two independent Brownian bridges, have the same asymptotic distribution. Taking into account basic properties of the Brownian bridge and A_1, it can easily be shown that the random variable eq. (6) and

$$n^{-1/2} \cdot [\mathscr{R}_g(t_Y)(1 - \mathscr{R}_g(t_Y)) + \lambda \cdot t_Y(1 - t_Y)]^{1/2} \cdot \mathscr{Z},$$

where $\mathscr{Z}$ is a standard normal, also coincide in distribution.

On the other hand, from A_2 and A_3, and since $r_g(t_Y) = 1$, for t close to t_Y, $\mathscr{R}_g(t) - \mathscr{R}_g(t_Y)$ can be approximated by $t - t_Y$. The Brownian bridge and the two-sided Brownian motion properties [27] guarantee that, for t close to t_Y and under A_3, the random variable $\{\hat{\mathscr{H}}_n(u_L, u_U, x_L, x_U) - \mathscr{H}(u_L, u_U, x_L, x_U)\}$ has the same asymptotic distribution of

$$n^{-1/2} \cdot \mathscr{Z}_1^{(n)}(t - t_Y) - (n/\lambda)^{-1/2} \cdot \mathscr{Z}_2^{(m)}(t - t_Y) = \sqrt{(1+\lambda)/n} \cdot \mathscr{Z}^{(n)}(t - t_Y), \tag{7}$$

where $\{\mathscr{Z}_1^{(n)}(z)\}_{\{-\infty<z<\infty\}}$ and $\{\mathscr{Z}_2^{(m)}(z)\}_{\{-\infty<z<\infty\}}$ are independent two-sided Brownian motions and $\{\mathscr{Z}^{(n)}(z)\}_{\{-\infty<z<\infty\}}$ is the weighted sum of those independent two-sided Brownian motions and, therefore, a two-sided Brownian motion as well.

Also, by assumptions A_2 and A_3, we have the approximation

$$\mathscr{H}(u_L, u_U, x_L, x_U) = (\mathscr{R}_g(t_Y) - t_Y)) - (\mathscr{R}_g(t) - t)) \approx -(1/2) r_g'(t_Y)(t - t_Y)^2. \tag{8}$$

Therefore, from eqs. (7) and (8), we have that the random variable

$$\max_{(u_L,u_U)\in\mathscr{C}[\hat{\mathscr{R}}_g]} \{\hat{\mathscr{H}}_n(u_L, u_U, x_L, x_U) - \mathscr{H}(u_L, u_U, x_L, x_U) + \mathscr{H}(u_L, u_U, x_L, x_U)\}$$

weakly converges to the distribution of the random variable

$$\max_{t\in[0,1]} \{\sqrt{(1+\lambda)/n} \cdot \mathscr{Z}^{(n)}(t - t_Y) - (1/2) \cdot r_g'(t_Y) \cdot (t - t_Y)^2\} = \kappa \cdot n^{-2/3} \cdot \max_{z\in\mathbb{R}} \{\mathscr{Z}^{(n)}(z) - z^2\} \tag{9}$$

where $z = (t - t_Y)/\gamma$ with $\gamma = (4 \cdot (1+\lambda)/r_g'(t_Y)^2)^{1/3} \cdot n^{-1/3}$, $\kappa = (2(1+\lambda)^2/r_g'(t_Y))^{1/3}$ and $\{\mathscr{Z}^{(n)}(z)\}_{\{-\infty<z<\infty\}}$ is a two-sided Brownian motion. We obtain *i*) directly from the equality eq. (5) and the convergences in eqs. (6) and (9).

On the other hand, if $\hat{t}_Y$ is the point which maximizes $(\hat{\mathscr{R}}_g(t) - t)$, then from eqs. (5) and (9), $(\hat{t}_Y - t_Y)/\gamma$ has the same asymptotic distribution that $\arg\max_{z\in\mathbb{R}}\{\mathscr{Z}^{(n)}(z) - z^2\}$. Result in *ii*) is derived from the equality

$$\frac{\hat{t}_Y - t_Y}{\gamma} = \left(\frac{r'(t_Y)^2}{4 \cdot (1+\lambda)}\right)^{1/3} \cdot n^{1/3} \cdot \{\hat{t}_Y - t_Y\}.$$

Remark. It is worth noting that, from eqs. (5) and (9), is easy to derive that

$$\mathbb{E}[\sqrt{n} \cdot \{\hat{\mathfrak{F}}_g - \mathfrak{F}_g\}] \approx \kappa \cdot n^{-1/6} \cdot \mathbb{E}[\max_{z\in\mathbb{R}} \{\mathscr{Z}^{(n)}(z) - z^2\}].$$

This bias, although asymptotically negligible, is relevant for small and moderate sample sizes.

References

[1] Zhou XH, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. New York: Wiley Blackwell, 2002.
[2] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd

[3] Martínez-Camblor P, Carleos C, Corral N. General nonparametric ROC curve comparison. J Korean Stat Soc. 2013;42:71–81.
[4] Martínez-Camblor P, Corral N, Rey C, Pascual J, Cernuda-Morollón E. Receiver operating characteristic curve generalization for non-monotone relationships. Stat Meth Med Res. 2017;26:113–23.
[5] Martínez-Camblor P, Pardo-Fernández JC.. Parametric estimates for the receiver-operating characteristic curve generalization for non-monotone relationships. Stat Meth Med Res 2017 DOI: 10.1177/0962280217747009
[6] Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32–5.
[7] Coffin M, Sukhatme S. Receiver operating characteristic studies and measurement errors. Biometrics. 1997;53:823–37.
[8] Liu X. Classification accuracy and cut points selections. Stat Med. 2012;31:2676–86.
[9] López-Ratón M, Cadarso-Suárez C, Molanes-López EM, Letón E. Confidence intervals for the symmetry point: an optimal cutpoint in continuous diagnostic tests. Pharma Stat. 2016;15:178–92.
[10] Martínez-Camblor P. Nonparametric cutoff point estimation for diagnostic decisions with weighted errors. Revista Colombiana de Estadóstica 2011;34:133–46.
[11] Khanafer N, Sicot N, Vanhems P, Dumitrescu O, Meyssonier V, Tristan A, Bes M, Lina G, Vandenesch F, Gillet Y, Etienne J. Severe leukopenia in staphylococcus aureus-necrotizing, community-acquired pneumonia: risk factors and impact on survival. BMC Infect Dis. 2013;159:1471–2334.
[12] Mossman D. Three-way ROCs. Med Decis Making 1999;19:78–89.
[13] Nakas CT, Alonzo TA, Yiannoutsos CT. Accuracy and cut-off point selection in three-class classification problems using a generalization of the youden index. Stat Med. 29:2946–55.
[14] Martínez-Camblor P, Pérez-Fernández S, Díaz-Coto S. Improving the biomarker diagnostic capacity via functional transformations. J Appl Stat. 2018; In press.
[15] Pérez-Fernández S, Martónez-Camblor P, Filzmoser P, Corral N. nsROC: An R package for non-standard ROC curve analysis. R J. 2018.
[16] Martínez-Camblor P, de Uña Álvarez J. Studying the bandwidth in k-sample smooth tests. Comput Stat. 2013;28:875–92.
[17] Spanos A, Harrell FE, Durack DT. Differential diagnosis of acute meningitis: an analysis of the predictive value of initial observations. J Am Med Assoc. 1989;262:2700–07.
[18] Zhou H, Qin G. Confidence intervals for the difference in paired Youden indices. Pharma Stat. 2013;12:17–27.
[19] Lai CY, Tian L, Schisterman EF. Exact confidence interval estimation for the Youden index and its corresponding optimal cut-point. Comput Stat Data Anal. 2012;56:1103–14.
[20] Zhou H, Qin G. New nonparametric confidence intervals for the Youden index. J Biopharma Stat. 2012;22:1244–57.
[21] Shan G. Improved confidence intervals for the Youden index. PLOS ONE 2015;10:1–19.
[22] Hall P, DiCiccio TJ, Romano JP. On smoothing and the bootstrap. Annal Stat. 1989;17:692–704.
[23] Schisterman EF, Perkins N. Confidence intervals for the Youden index and corresponding optimal cut-point. Commun Stat - Simul Comput. 2007;36:549–63.
[24] Hsieh F, Turnbull BW. Nonparametric methods for evaluating diagnostic tests. Stat Sin. 1996;6:47–62.
[25] DasGupta A. Asymptotic theory of statistics and probability. New York: Springer, 2008.
[26] van der Vaart AW. Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.
[27] Borodin AN, Palminen P. Handbook of Brownian motion–facts and formulae. Probability and its Applications. Birkhäuser Verlag, Basel, second edition, 2002. MR-1912205.

Supplementary Material: The online version of this article offers supplementary material (DOI:https://doi.org/10.1515/ijb-2018-0060).

Automatically generated rough PDF by *ProofCheck* from River Valley Technologies Ltd