Concentration and ROC Curves, Revisited

Sankhya A manuscript No.
(will be inserted by the editor)

Concentration and ROC curves

Mauro Gasparini · Lidia Sacchetto

Received: date / Accepted: date

Abstract This work is aimed at illustrating the strict relationship between a general definition of concentration function appeared quite some time ago on this journal and a widely used measure of the diagnostic strength of a family of binary classifiers indexed by a threshold parameter, the so-called ROC curve.

The ROC curve is a common work tool in Statistics, Machine Learning and Artificial Intelligence, appearing in many applications where a binary classification (diagnosis) procedure is of interest. Hence, it is worth remarking that diagnostic strength and concentration are two sides of the same coin: the higher the concentration of one probability measure with respect to another, the higher the diagnostic strength of the likelihood ratio classification rule.

Keywords likelihood ratio · Neyman-Pearson lemma · classification

Mathematics Subject Classification (2010) 62H30 · 62H20

1 Introduction

More than a hundred years ago Corrado Gini started his elaboration on the notion of concentration, with particular application to transferable characteristics such as wealth. The products of his work, e.g. the Gini mean difference, the Gini concentration coefficient and the Lorenz-Gini concentration curve, are part of the toolbox of any data analyst.

More recently, on this journal, Cifarelli and Regazzini ([1]) extended the notion of concentration to become a relationship between two probability measures, rather than a one-dimensional concept, giving at the same time solid measure-theoretical justifications.

M. Gasparini
Department of Mathematical Sciences "G.L. Lagrange"
Politecnico di Torino, 10129 Torino, Italy E-mail: mauro.gasparini@polito.it

L. Sacchetto
Department of Mathematical Sciences "G.L. Lagrange"
Politecnico di Torino, 10129 Torino, Italy

Parallely, in a variety of literature scattered across many disciplines such as Signal Processing, Medical Diagnosis and Artifical Intelligence, the ROC (Receiver Operating Characteristic) curve was developed as a tool to measure the diagnostic strength of a family of classification rules indexed by some threshold parameter. The importance of the ROC curve in applied work can not be overstated, since it commonly appears in all applications where a binary classification (or diagnosis) procedure is of interest. See the recent textbooks by Krzanowski [6], Pepe [10] and Zou [14]. It was soon realized in the earlier literature on ROC (see e.g. [3]), that, an optimal decision rule for binary classification exists, as long as the two probability measures compared are completely specified: such optimal rule is based on the likelihood ratio (LR from now on), as proven by the Neyman-Pearson lemma, another milestone in the development of Statistics in the last century.

Remarkably, the ROC curve of the LR classifier can also be viewed as an application of the notion of concentration: the higher the concentration of one probability measure with respect to another, the higher the diagnostic strength of the optimal decision rule. The aim of this work is to illustrate this relationship, clarifying the theoretical and interpretational advantages given by a model based approach.

2 The ROC curve of the likelihood ratio based classification rule

From a mathematical point of view we can describe the binary classification problem as a competition between two populations, represented as probability measures. An object, on which one or more random variables are observed, is to be assigned to one of the two populations, based on some classification rule. We focus here on the situation where the two probability measures are completely known, hence avoiding the statistical problems of estimation or, as they say in the Machine Learning literature, learning.

Assume therefore the two alternative probability measures P_+ and P_- are absolutely continuous with respect to one another and have densities f_+ and f_-, respectively, with respect to a common dominating measure. Without loss of generality, f_- can be taken to be positive, so that the Likelihood Ratio (LR)

$$L = \frac{f_+}{f_-} \tag{1}$$

is a well defined non negative random variable and, as such, has distribution functions under P_- and P_+, which are denoted by H_- and H_+ respectively. More precisely, for each $\ell \in \mathcal{R}$:

$$H_-(\ell) = P_-(L \leq \ell)$$

and

$$H_+(\ell) = P_+(L \leq \ell).$$

Next, define the quantile function associated with H_- in the usual way as follows:

$$q_t = \inf\{y \in \mathcal{R} : H_-(y) \geq t\} \quad 0 < t < 1 \tag{2}$$

and recall that, for any real number ℓ, $q_t \leq \ell$ if and only if $H_-(\ell) \geq t$. For any given value $t \in (0,1)$, it may or may not happen that $t = H_-(q_t)$, depending

on whether t does not correspond or does correspond to a jump of H_-. More specifically, if $t \neq H_-(q_t)$, then $H_-(q_t^-) \leq t < H_-(q_t)$, where the notation $^-$ indicates left limits (nothing to do with P_-), a particularly relevant occurrence for the discussion below.
H_- and H_+ may have jumps, even though P_- and P_+ are absolutely continuous laws on the real line. This happens, for example, if P_+ and P_- have piecewise constant densities as provided by an example in Section 4.
In this paper we focus on the following definition of LR based classification rule:

Definition 1 (LR based classification rule) Given two alternative probability laws P_- or P_+ mutually absolutely continuous with densities f_- and f_+ respectively, define the likelihood ratio $L = f_+/f_-$, its respective distribution functions H_- and H_+ and the following classification rule. For any given $0 < t < 1$:

- if $L > q_t$, declare positive;
- if $L < q_t$, declare negative;
- if $L = q_t$, then perform an auxiliary independent randomization and declare positive with probability

$$r(t) = \frac{H_-(q_t) - t}{H_-(q_t) - H_-(q_t^-)}$$

and negative otherwise.

The LR based classification rule is optimal because it is nothing else than the Neyman-Pearson lemma, enriched with the possibility of randomization (as in [8], for example).

In the classification literature, it is generally recognized that the LR based classification rule is optimal, although for reasons related to the statistical estimation of P_- and P_+ in the presence of data and the computational problems with highly dimensional observations, other kinds of classification rules are often considered. As mentioned above, we focus here on the no-data situation and assume P_- and P_+ are known.

Whatever the classification rule, it is typically indexed by a real-valued threshold parameter $t \in \mathcal{R}$, like the LR based rule is. By varying t, the associated ROC is generated: it is defined as the parametric two-dimensional locus

$$\{(\mathrm{FPR}(t), \mathrm{TPR}(t)), t \in \mathcal{R}\},$$

where the false positive rate FPR is the probability the classification rule assigns the object to population P_+ given the object comes from population P_- and the true positive rate TPR is the probability the classification rule assigns the object to population P_+ given the object comes from population P_+. A variety of other names exist, in particular *sensitivity* for the TPR and *specificity* for 1-FPR.

Theorem 1 *The ROC function of the classification rule of Definition 1 is*

$$\mathrm{ROC}(x) = 1 - H_+(q_{1-x}) + q_{1-x}(H_-(q_{1-x}) - (1-x)), \quad 0 < x < 1. \tag{3}$$

As usual, we can complete the result by setting $ROC(0) = 0$ and $ROC(1) = 1$.

Proof First of all, the FPR and the TPR are computed separately.

$$\begin{aligned}\text{FPR} &= P_-(\text{declare positive}) = P_-(L > q_t) + P_-(L = q_t)r(t)\\ &= 1 - H_-(q_t) + (H_-(q_t) - H_-(q_t^-))r(t)\\ &= 1 - H_-(q_t) + H_-(q_t) - t = 1 - t.\end{aligned}$$

Notice that if $t = H_-(q_t)$ then $H_-(q_t^-) - H_-(q_t) = 0$; in other words the expression simplifies for points which are not H_--atoms.

$$\begin{aligned}\text{TPR} &= P_+(\text{declare positive}) = P_+(L > q_t) + P_+(L = q_t)r(t)\\ &= 1 - H_+(q_t) + (H_+(q_t) - H_+(q_t^-))\frac{H_-(q_t) - t}{H_-(q_t) - H_-(q_t^-)}\\ &= 1 - H_+(q_t) + q_t(H_-(q_t) - t)\end{aligned}$$

since, P_+ and P_- being mutually absolutely continuous, they will both have or not have an atom in q_t and their LR in q_t will be exactly $(H_+(q_t) - H_+(q_t^-))/(H_-(q_t) - H_-(q_t^-))$, i.e. q_t itself. Next, set FPR $= x$, i.e. $t = 1 - x$, to eliminate the parameter t and obtain the explicit form of the ROC curve:

$$\text{TPR} = 1 - H_+(q_{1-x}) + q_{1-x}(H_-(q_{1-x}) - (1 - x)).$$

Expression (3) may seem cumbersome when compared to simpler expressions for special cases contained in popular textbooks, but one should consider that is the ROC curve of the optimal LR based classification rule, covering discrete, continuous and mixed cases. Still greater generality could be achieved by considering the case where P_- and P_+ are not absolutely continuous with respect to one another, in which case we would have ROC curves starting at $(0, y_0)$, with $y_0 > 0$, or ending at $(x_0, 1)$, with $x_0 < 1$, although that is usually of little interest.

3 Relationship with a general concentration function

Another important reason justifying the generality of expression (3) is that it is strictly related to a definition of concentration function given on this journal by Cifarelli and Regazzini [1]. Such definition was given with the aim of extending the classical concepts of concentration developed by Gini at the beginning of the XX-th century and was further expanded by Regazzini [11].

It is recalled here for the case in which P_+ and P_- are mutually absolutely continuous (see the discussion at the end of last section):

Definition 2 (Concentration function by Regazzini and Cifarelli) Let P_+ and P_- be mutually absolutely continuous probability measures, let f_+ and f_- be their respective derivatives with respect to a common dominating measure μ, let their LR be defined as the real-valued random variable $L = f_+/f_-$, let H_- be its distribution function under P_- and let q_x be its quantile function. Then Cifarelli and Regazzini [1] define the concentration function of P_+ with respect to P_- as $\varphi(0) = 0$, $\varphi(1) = 1$ and

$$\varphi(x) = P_+(L < q_x) + q_x(x - H_-(q_x^-)).$$

The easy connection between this definition and the ROC curve of the LR based classification rule of the previous section is established in the next Theorem.

Theorem 2 *Under the hypotheses described in Definition 1,*

$$\mathrm{ROC}(x) = 1 - \varphi(1-x) \quad \forall 0 \le x \le 1.$$

where $\varphi(\cdot)$ is the concentration function of P_+ with respect to P_-.

Proof The equivalent relationship

$$1 - \mathrm{ROC}(1-x) = \varphi(x) \quad \forall 0 \le x \le 1.$$

can be verified directly for $x = 0, 1$ and as follows for $0 < x < 1$:

$$\begin{aligned}
1 - \mathrm{ROC}(1-x) &= H_+(q_x) - q_x(H_-(q_x) - x) \\
&= H_+(q_x) \pm H_+(q_x^-) + q_x(x - H_-(q_x) \pm H_-(q_x^-)) \\
&= H_+(q_x^-) + q_x(x - H_-(q_x^-)) + \\
&\qquad (H_+(q_x) - H_+(q_x^-)) - q_x(H_-(q_x) - H_-(q_x^-)) \\
&= H_+(q_x^-) + q_x(x - H_-(q_x^-)) + \\
&\qquad (H_-(q_x) - H_-(q_x^-)) \left(\frac{H_+(q_x) - H_+(q_x^-)}{H_-(q_x) - H_-(q_x^-)} - q_x \right) \\
&= P_+(L < q_x) + q_x(x - H_-(q_x^-)) \\
&= \varphi(x).
\end{aligned}$$

Corollary 1 *Under the assumptions described in Definition 1,* ROC$(\cdot)$ *is a non-decreasing, continuous and concave function on* $[0, 1]$*. In particular,* ROC$(\cdot)$ *is proper.*

Proof This is a consequence of Theorem 2.3 in Cifarelli [1]. In particular, $\varphi(x)$ is always convex over its domain, i.e. $\forall x_1, x_2$ and $\nu \in [0,1], \varphi(\nu x_1 + (1-\nu)x_2) \le \nu\varphi(x_1) + (1-\nu)\varphi(x_2)$. By Theorem 2:

$$1 - \mathrm{ROC}(1 - (\nu x_1 + (1-\nu)x_2)) \le \nu(1 - \mathrm{ROC}(1-x_1)) + (1-\nu)(1 - \mathrm{ROC}(1 - \mathrm{x}_2)).$$

The left hand side of the previous equality becomes:

$$\begin{aligned}
1 - \mathrm{ROC}(1 - (\nu x_1 + (1-\nu)x_2)) &= 1 - \mathrm{ROC}(\nu + (1-\nu) - \nu x_1 - (1-\nu)x_2) \\
&= 1 - \mathrm{ROC}(\nu(1-x_1) + (1-\nu)(1-x_2)),
\end{aligned}$$

while the right hand side can be rewritten as:

$$\begin{aligned}
&\nu(1 - \mathrm{ROC}(1-x_1)) + (1-\nu)(1 - \mathrm{ROC}(1-x_2)) = \\
&\nu - \nu\mathrm{ROC}(1-x_1) + 1 - \nu - (1-\nu)\mathrm{ROC}(1-x_2) = \\
&1 - \nu\mathrm{ROC}(1-x_1) - (1-\nu)\mathrm{ROC}(1-x_2).
\end{aligned}$$

Therefore:

$$\mathrm{ROC}(\nu t_1 + (1-\nu)t_2) \ge \nu\mathrm{ROC}(t_1) + (1-\nu)\mathrm{ROC}(t_2), \quad \forall t_1, t_2, \nu \in [0,1]$$

where $t_1 = 1 - x_1, t_2 = 1 - x_2$.

It is not the first time a relationship between ROC curves and Lorenz-Gini concentration curves is noticed (see for example [7] and [12]), but Definitions 1 and 2 allow for great generality and clarify many misunderstandings present in the literature. In particular, we emphasize that, unless the classification rule is based on the LR or on a monotone transformation of it, it can not sensibly be related to a concentration curve.

Properness of ROC curves has been discussed since the early literature, since there are cases where classification rules which are not based on the LR are used and may give rise to non-concave ROC functions. The binormal heteroschedastic case is a notable well-known example. Suboptimal rules are often used since the LR rule may not be easily calculated or not even estimated, but the fact that the LR based classification rule gives automatically a proper ROC curve is an important foundational result.

At the same foundational level, it finally appears clear that the diagnostic strength of the LR based classification rule is equivalent to the mutual concentration of the two competing probability laws: the more P_+ is concentrated with respect to P_- the higher the diagnostic strength of the LR based classification rule, since one can tell the two probability laws better apart.

4 Examples

This section contains some examples. The first one is a well known case in which both P_+ and P_- are discrete, since they are given by the empirical frequencies of an ordinal diagnosis. The usual practice of connecting the dots obtained at the thresholds in order to draw a concave ROC curve is fully justified by Definition 1. The second example is instead based on an observable variable which is absolutely continuous both under P_+ and under P_- and has nonetheless LR distributions with discrete components, resulting in a mixed ROC curve (partially linear and partially curvilinear). The third and fourth examples are multivariate and they are based on the normality assumption, leading to a rediscovery of Fisher's discriminant functions.

4.1 Example 1: ordinal diagnosis

The following example is taken from the Encyclopedia of Biostatistics [2]. Suppose 109 patients have been classified as diseased (P_+) or not diseased (P_-), based on a gold standard such as biopsy or autopsy. On the basis of radiological exams, they have also been classified over five ordinal levels

$$\begin{aligned} -- &= \text{very mild} \\ - &= \text{mild} \\ +- &= \text{neutral} \\ + &= \text{serious} \\ ++ &= \text{very serious} \end{aligned}$$

Here are the results:

	$--$	$-$	$+-$	$+$	$++$	total
P_-	33	6	6	11	2	58
P_+	3	2	2	11	33	51

In particular, P_+ and P_- are two empirical measures, relative to the diseased and not diseased population respectively, derived from data. There are four possible values for the LR:

$$L = \begin{cases} \frac{58}{561} & \text{if } -- \\ \frac{58}{153} & \text{if } - \text{ or } +- \\ \frac{58}{51} & \text{if } + \\ \frac{319}{17} & \text{if } ++ \end{cases}$$

which give rise to four empirical ROC points { (25/58, 48/51); (19/58, 46/51); (13/58, 44/51); (2/58, 33/51)}, shown in Figure 1. Now, thanks to the randomization device, it is possible to ... connect the dots! This is so since the distribution functions of L under P_- and P_+ are

$$H_-(\ell) = \begin{cases} 0 & \text{if } 0 \le \ell < \frac{58}{561} \\ \frac{33}{58} & \text{if } \frac{58}{561} \le \ell < \frac{58}{153} \\ \frac{45}{58} & \text{if } \frac{58}{153} \le \ell < \frac{58}{51} \\ \frac{56}{58} & \text{if } \frac{58}{51} \le \ell < \frac{319}{17} \\ 1 & \text{if } \frac{319}{17} \le \ell \end{cases}$$

and

$$H_+(\ell) = \begin{cases} 0 & \text{if } 0 \le \ell < \frac{58}{561} \\ \frac{3}{51} & \text{if } \frac{58}{561} \le \ell < \frac{58}{153} \\ \frac{7}{51} & \text{if } \frac{58}{153} \le \ell < \frac{58}{51} \\ \frac{18}{51} & \text{if } \frac{58}{51} \le \ell < \frac{319}{17} \\ 1 & \text{if } \frac{319}{17} \le \ell. \end{cases}$$

Therefore, the ROC curve can be calculated using equation (3):

$$\mathrm{ROC}(x) = \begin{cases} \frac{319}{17}x & \text{if } 0 \le x < \frac{2}{58} \\ \frac{31}{51} + \frac{58}{51}x & \text{if } \frac{2}{58} \le x < \frac{13}{58} \\ \frac{7}{9} + \frac{58}{153}x & \text{if } \frac{13}{58} \le x < \frac{25}{58} \\ \frac{503}{561} + \frac{58}{561}x & \text{if } \frac{25}{58} \le x < 1 \end{cases}$$

The continuous ROC curve interpolates the empirical ROC points, as shown in Figure 1.

4.2 Example 2: absolutely continuous measures with discrete/continuous LR

Let P_- be uniform between 0 and 3 (an absolutely continuous probability measure on the real line) and let P_+ have density f_+ defined as follows:

$$f_+(s) = \begin{cases} \frac{2}{9} + \frac{2}{3}(1-s)^5 & \text{if } 0 < s \le 1 \\ \frac{2}{9} & \text{if } 1 < s \le 2 \\ \frac{4}{9} & \text{if } 2 < s \le 3 \\ 0 & \text{otherwise.} \end{cases}$$

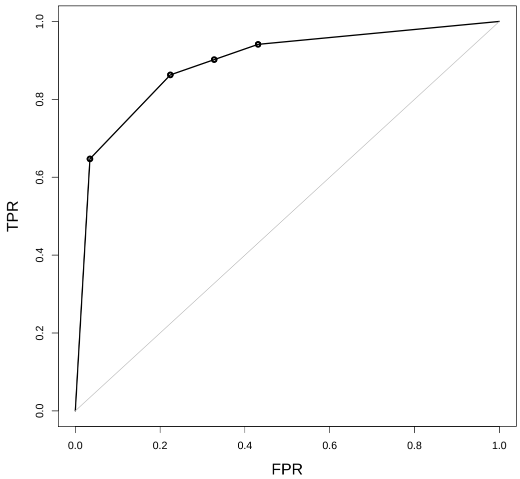

Fig. 1 Example 1: the ROC curve based on the LR interpolates the empirical ROC points.

Suppose S is a real random variable with density f_- under P_- and f_+ under P_+. It is easy to see that the LR $L = f_+/f_-$ has mixed components and it is not monotone in S, being:

$$L = \begin{cases} \frac{2}{3} + 2(1-s)^5 & \text{if } 0 < s \leq 1 \\ \frac{2}{3} & \text{if } 1 < s \leq 2 \\ \frac{4}{3} & \text{if } 2 < s \leq 3. \end{cases}$$

A naive classification rule based solely on S gives rise to the ROC curve

$$\mathrm{ROC}_S(x) = \begin{cases} \frac{4}{3}x & \text{if } 0 \leq x < \frac{1}{3} \\ \frac{2}{9} + \frac{2}{3}x & \text{if } \frac{1}{3} \leq x < \frac{2}{3} \\ \frac{2}{9} + \frac{2}{3}x + \frac{(3x-2)^6}{9} & \text{if } \frac{2}{3} \leq x < 1 \end{cases}$$

which is not concave, shown as dashed line in Figure 2. Using instead the LR based classification rule, the ROC curve is:

$$\mathrm{ROC}_{\mathrm{LR}}(x) = \begin{cases} \frac{1}{9} + \frac{2}{3}x - \frac{1}{9}(1-3x)^6 & \text{if } 0 \leq x < x_1 \\ -\frac{1}{9} + \frac{4}{3}x + \frac{2}{9}(\frac{1}{3})^{1/5} - \frac{1}{9}(\frac{1}{3})^{6/5} & \text{if } x_1 \leq x < x_2 \\ \frac{1}{3} + \frac{2}{3}x - \frac{1}{9}(2-3x)^6 & \text{if } x_2 \leq x < \frac{2}{3} \\ \frac{1}{3} + \frac{2}{3}x & \text{if } \frac{2}{3} \leq x < 1 \end{cases}$$

where $x_1 = \frac{1}{3} - \frac{1}{3}(\frac{1}{3})^{1/5}$ and $x_2 = \frac{2}{3} - \frac{1}{3}(\frac{1}{3})^{1/5}$. This curve is concave and dominates the previous one as shown in Figure 2.
This example deals with absolutely continuous densities which, nonetheless, have a likelihood ratio - often called score in classification - with mixed components (partly discrete and partly continuous): it is an exquisitely theoretical exercise, but it addresses a case particularly difficult for the usual approach to ROC curves (which emphasizes a continuous score is necessary).

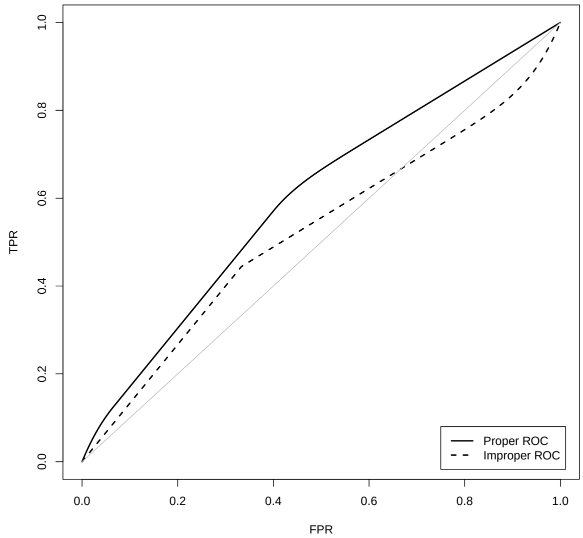

Fig. 2 Example 2: LR base ROC curve (solid line) versus the improper ROC curve of a naive test based on S alone (dashed line)

4.3 Example 3: two multivariate normal measures

Assume P_- is multivariate normal with mean μ_- and variance Σ_- and P_+ is multivariate normal with mean μ_+ and variance Σ_+ and both densities exist. By taking the logarithmic transformation of the LR, it can easily be seen that for the normal case the LR based classification rule in Definition 1 declares positive if the quadratic score

$$(X - \mu_-)^{\mathrm{T}} \Sigma_-^{-1} (X - \mu_-) - (X - \mu_+)^{\mathrm{T}} \Sigma_+^{-1} (X - \mu_+) \tag{4}$$

is large. This is the well known Fisher's Quadratic Discriminant Analysis (QDA) rule [4], which reduces to linear – hence the corresponding Linear Discriminant Analysis (LDA) – in the case $\Sigma_- = \Sigma_+$ (homoschedasticity). The original work by Fisher did not actually focus on the normality assumption, but QDA and LDA are well established terminology in the literature. Being based on the LR, QDA has a proper ROC curve: the score in equation (4) is a continuous random variable and no randomization device is needed.

4.4 Example 4: Fisher versus best linear rules

In the multivariate normal case, insisting on a *linear* classifier leads to suboptimal procedures in the case of heteroschedasticity. The classifier which is optimal within the class of linear classifiers is considered in Su and Liu[13] and it declares positive if

$$(\mu_+ - \mu_-)^{\mathrm{T}} (\Sigma_- + \Sigma_+)^{-1} X \tag{5}$$

is large. If $\Sigma_- \neq \Sigma_+$ it gives an improper ROC curve, which has a "hook" and is dominated by the ROC curve of the corresponding quadratic score in Expression (4). It is worth considering a numerical example in this case, since the optimality of the quadratic score in the normal case is being continuously rediscovered(see e.g. [9] and [5]), but it actually boils down to Fisher [4].

Consider a bivariate normal vector (X, Y) which in population P_- has a bivariate standard normal distribution, whereas in population P_+ has independent components X distributed normally with mean $\mu_x > 0$ and variance σ_x^2 and Y distributed normally with mean $\mu_y > 0$ and variance $\sigma_y^2 \neq \sigma_x^2$. According to equation (4), the QDA classifier declares positive if

$$\left(\frac{X-\mu_x}{\sigma_x}\right)^2 + \left(\frac{Y-\mu_y}{\sigma_y}\right)^2 - X^2 - Y^2 < c$$

where c is an arbitrary threshold. By varying c and calculating the appropriate probabilities under P_- and P_+, we can obtain the ROC curve, by simulation or, if greater precision is needed, by using non-central chi-square distributions. The ROC curve for the case $\mu_x = 1$, $\mu_y = 2$, $\sigma_x = 2$, $\sigma_y = 4$ is plotted as a solid line in Figure 3.

The best linear classifier according to Expression (5) is instead

$$S = \frac{\mu_x}{1+\sigma_x^2} X + \frac{\mu_y}{1+\sigma_y^2} Y.$$

S has normal distributions under P_- and P_+ and by a well-known result its ROC is

$$\mathrm{ROC}(t) = \phi(A + \phi^{-1}(t)B) \tag{6}$$

where $\phi(\cdot)$ is the standard normal distribution function,

$$A = \frac{\mu_x^2(1+\sigma_y^2) + \mu_y^2(1+\sigma_x^2)}{\sqrt{\mu_x^2\sigma_x^2(1+\sigma_y^2)^2 + \mu_y^2\sigma_y^2(1+\sigma_x^2)^2}}$$

and

$$B = \frac{\sqrt{\mu_x^2(1+\sigma_y^2)^2 + \mu_y^2(1+\sigma_x^2)^2}}{\sqrt{\mu_x^2\sigma_x^2(1+\sigma_y^2)^2 + \mu_y^2\sigma_y^2(1+\sigma_x^2)^2}}.$$

This ROC curve for the case $\mu_x = 1$, $\mu_y = 2$, $\sigma_x = 2$, $\sigma_y = 4$ is plotted as a dashed line in Figure 3. We can easily see that the QDA ROC curve is concave and dominates the best linear ROC curve.

5 Conclusions

The brief historical overview given in the Introduction can be completed with a look into the present and future times.

Nowadays the model based approach - namely the centrality of two competing probability measures P_+ and P_- - is considered out fashioned by many researchers, solely interested in algorithms for classification. The reason is a certain degree of success obtained in the presence of high dimensional data by methods apparently unrelated to probability: Support Vector Machines, Deep Learning and alike. Focus is on obtaining viable computational methods addressed to minimizing the

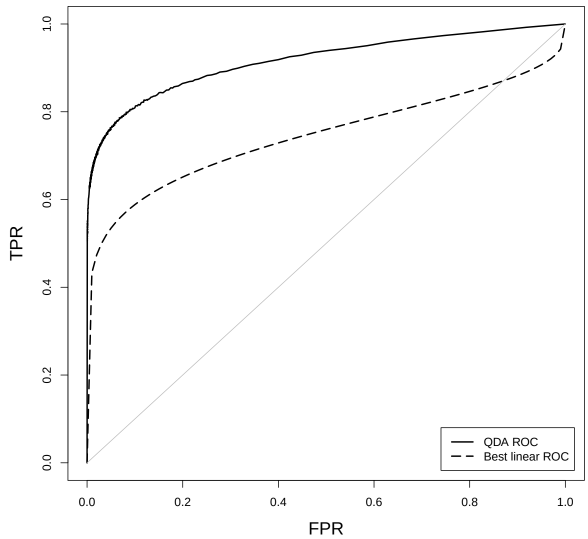

Fig. 3 Example 4: QDA ROC curve (solid) and best linear ROC curve (dashed) for the bi-bivariate normal case, assuming $\mu_x = 1$, $\mu_y = 2$, $\sigma_x = 2$, $\sigma_y = 4$.

empirical risk or prediction error; the resulting ROC curves are often not proper and the distinction between population and sample estimates of many objects - including ROC curves themselves - is often blurred or simply ignored.

Contrarily to this trend - which has undoubtedly many advantages, including challenging statisticians with new computing intensive ideas - we have shown that the diagnostic strength of the theoretically optimal LR based classification rule is strictly related to how concentrated a certain probability measure is with respect to another. We have also demonstrated how this relationship is deeply rooted in a few fundamental concepts from classical Statistics.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Cifarelli, D.M., Regazzini, E.: On a general definition of concentration function. Sankhya: The Indian Journal of Statistics, Series B **49**(3), 307–319 (1987)
2. Colton, T., Armitage, P.: Encyclopedia of biostatistics. John Wiley & Sons (2005)
3. Egan, J.P.: Signal Detection Theory and ROC Analysis. Academic Press, London (1975)
4. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Eugenics **7**(2), 179–188 (1936)
5. Hillis, S.L.: Equivalence of binormal likelihood-ratio and bi-chi-squared ROC curve models. Statistics in Medicine **35**(12), 2031–2057 (2016)
6. Krzanowski, W.J., Hand, D.J.: ROC curves for continuous data. Chapman and Hall/CRC (2009)

7. Lee, W.C.: Probabilistic analysis of global performances of diagnostic tests: interpreting the lorenz curve-based summary measures. Statistics in Medicine **18**(4), 455–471 (1999)
8. Lehmann, E.L.: Testing statistical hypotheses. Springer (1986)
9. Metz, C.E., Pan, X.: "Proper" binormal ROC curves: theory and maximum-likelihood estimation. Journal of Mathematical Psychology **43**(1), 1–33 (1999)
10. Pepe, M.S.: The statistical evaluation of medical tests for classification and prediction. Oxford University Press (2003)
11. Regazzini, E.: Concentration comparisons between probability measures. Sankhya: The Indian Journal of Statistics, Series B **54**(2), 129–149 (1992)
12. Schechtman, E., Schechtman, G.: The relationship between gini terminology and the roc curve. Metron pp. 129–149 (2019)
13. Su, J.Q., Liu, J.S.: Linear combinations of multiple diagnostic markers. Journal of the American Statistical Association **88**(424), 1350–1355 (1993)
14. Zou, K.H., Liu, A., Bandos, A.I., Ohno-Machado, L., Rockette, H.E.: Statistical evaluation of diagnostic performance: topics in ROC analysis. CRC Press (2011)