TECHNICAL NOTE

Receiver operating characteristic (ROC) curves: equivalences, beta model, and minimum distance estimation

Tilmann Gneiting[1,2] [ORCID] · **Peter Vogel[3]**

Abstract

Receiver operating characteristic (ROC) curves are used ubiquitously to evaluate scores, features, covariates or markers as potential predictors in binary problems. We characterize ROC curves from a probabilistic perspective and establish an equivalence between ROC curves and cumulative distribution functions (CDFs). These results support a subtle shift of paradigms in the statistical modelling of ROC curves, which we view as curve fitting. We propose the flexible two-parameter beta family for fitting CDFs to empirical ROC curves and derive the large sample distribution of minimum distance estimators in general parametric settings. In a range of empirical examples the beta family fits better than the classical binormal model, particularly under the vital constraint of the fitted curve being concave.

Keywords Binary prediction · Classification · Evaluation of predictive potential

1 Introduction

Through all realms of science and society, the assessment of the predictive ability of scores or features for binary outcomes is of critical importance. To give but a few examples, bio-markers are used to diagnose diseases, weather forecasts serve to anticipate extreme pre-cipitation events, judges need to assess recidivism in convicts, banks use customers' par-ticulars to assess credit risk, and email messages are to be identified as spam or legitimate.

✉ Tilmann Gneiting
tilmann.gneiting@h-its.org

1 Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

2 Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

3 CSL Behring Innovation, Marburg, Germany

In these and myriads of similar settings, receiver operating characteristic (ROC) curves are key tools in the evaluation of predictive potential (Fawcett 2006; Flach 2016).

A ROC curve is simply a plot of the hit rate (HR) against the false alarm rate (FAR) across the range of thresholds for a real-valued score. Specifically, consider the joint distribution $\mathbb{Q}$ of the pair (S, Y), where the score S is real-valued, and the event Y is binary, with the implicit understanding that higher values of S provide stronger support for the event to materialize ($Y = 1$). The joint distribution $\mathbb{Q}$ of (S, Y) is characterized by the *prevalence* $\pi_1 = \mathbb{Q}(Y = 1) \in (0, 1)$ along with the conditional cumulative distribution functions (CDFs)

$$F_1(s) = \mathbb{Q}(S \leq s \mid Y = 1) \qquad \text{and} \qquad F_0(s) = \mathbb{Q}(S \leq s \mid Y = 0).$$

Any threshold value s can be used to predict a positive outcome ($Y = 1$) if $S > s$ and a negative outcome ($Y = 0$) if $S \leq s$, to yield a classifier with *hit rate* (HR),

$$\text{HR}(s) = \mathbb{Q}(S > s \mid Y = 1) = 1 - F_1(s),$$

and *false alarm rate* (FAR),

$$\text{FAR}(s) = \mathbb{Q}(S > s \mid Y = 0) = 1 - F_0(s).$$

Terminologies abound and differ markedly between communities. For example, the hit rate has also been referred to as *true positive rate* (TPR), *sensitivity* or *recall*. The false alarm rate is also known as *false positive rate* (FPR) or *fall-out* and equals one minus the *true negative rate*, *specificity*, or *selectivity*.

The term *raw ROC diagnostic* refers to the set-theoretic union of the points of the form $(\text{FAR}(s), \text{HR}(s))'$ within the unit square.[1] The *ROC curve* is a linearly interpolated raw ROC diagnostic, and therefore it also is a point set that may or may not admit a direct interpretation as a function. However, if F_1 and F_0 are continuous and strictly increasing, the raw ROC diagnostic and the ROC curve can be identified with a function R, where $R(0) = 0$,

$$R(p) = 1 - F_1(F_0^{-1}(1 - p)) \quad \text{for} \quad p \in (0, 1), \tag{1}$$

and $R(1) = 1$. High hit rates and low false alarm rates are desirable, so the area under the ROC curve (AUC) is a widely used, positively oriented measure of the predictive potential of a score or feature. In data analytic practice, the measure $\mathbb{Q}$ is the empirical distribution of a sample $(s_i, y_i)_{i=1}^n$ of real-valued scores s_i and corresponding binary observations y_i. In this setting, it suffices to consider the unique values of $s_1, \ldots, s_n$ to generate the raw ROC diagnostic, and linear interpolation yields the empirical ROC curve, as illustrated in our data examples.

The remainder of our note is organized as follows. Section 2 establishes rigorous versions of fundamental theoretical results that thus far have been available informally or in special cases only. In particular, we demonstrate an equivalence between ROC curves and CDFs. In Sect. 3 we introduce the flexible yet parsimonious two-parameter beta model, which uses the CDFs of beta distributions to model ROC curves, and we discuss estimation and testing in this context. The paper closes with data examples in Sect. 4.

---

[1] We use the notation $(a, b)'$ for a column vector in $\mathbb{R}^2$.

2 Fundamental properties of ROC curves

Consider the random vector (S, Y) where S is a real-valued *score*, feature, covariate, or marker, and Y is the binary response. As before, we refer to the joint distribution of (S, Y) as $\mathbb{Q}$. Let $\pi_0 = 1 - \pi_1 = \mathbb{Q}(Y = 0)$, and let

$$F(s) = \mathbb{Q}(S \le s) = \pi_0 F_0(s) + \pi_1 F_1(s)$$

denote the marginal cumulative distribution function (CDF) of the score S. We write $G(s-)$ for the left-hand limit of a function G at $s \in \mathbb{R}$, and we let $(a, b)'_{(1)} = a$ and $(a, b)'_{(2)} = b$ denote coordinate projections.

2.1 Raw ROC diagnostics and ROC curves

In this setting ROC diagnostics concern the points of the form $(\mathrm{FAR}(s), \mathrm{HR}(s))'$ within the unit square. Formally, the *raw ROC diagnostic* for the bivariate distribution $\mathbb{Q}$ is the point set

$$R^* = \left\{ \begin{pmatrix} 1 - F_0(s) \\ 1 - F_1(s) \end{pmatrix} : s \in \mathbb{R} \right\}. \tag{2}$$

The raw ROC diagnostic along with a single marginal does not characterize $\mathbb{Q}$, due to its well known invariance properties (Fawcett 2006; Krzanowski and Hand 2009). However, the raw ROC diagnostic along with both marginal distributions determines $\mathbb{Q}$.

Theorem 1 *The joint distribution $\mathbb{Q}$ of (S, Y) is characterized by the raw ROC diagnostic and the marginal distributions of S and Y.*

Proof The mapping $g : [0, 1]^2 \to [0, 1]$ defined by $(a, b)' \mapsto (1 - a)\pi_0 + (1 - b)\pi_1$ induces a bijection between the raw ROC diagnostic R^* and the range of F. Therefore, it suffices to note that $\mathbb{Q}(S \le s, Y \le y) = 0$ for $y < 0$,

$$\begin{aligned}
\mathbb{Q}(S \le s, Y \le y) &= F_0(s)\, \pi_0 \\
&= F(s) - (1 - \mathrm{HR}(s))\, \pi_1 \\
&= F(s) - (1 - g_{(2)}^{-1}(F(s)))\, \pi_1
\end{aligned}$$

for $y \in [0, 1)$, and $\mathbb{Q}(S \le s, Y \le y) = F(s)$ for $y \ge 1$. $\qquad \square$

Briefly, a ROC curve is obtained from the raw ROC diagnostic by linear interpolation. Formally, the *full ROC diagnostic* or *ROC curve* is the point set

$$R = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \cup R^* \cup \{L_s : s \in \mathbb{R}\} \cup \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \tag{3}$$

within the unit square, where

$$L_s = \left\{ \alpha \begin{pmatrix} 1 - F_0(s) \\ 1 - F_1(s) \end{pmatrix} + (1 - \alpha) \begin{pmatrix} 1 - F_0(s-) \\ 1 - F_1(s-) \end{pmatrix} : \alpha \in [0, 1] \right\}$$

is a possibly degenerate, nondecreasing line segment. The choice of linear interpolation to complete the raw ROC diagnostic into the ROC curve (3) is natural and persuasive, as the

line segment L_s represents randomized combinations of the classifiers associated with its end points.

The raw ROC diagnostic can be recovered from the ROC curve and the two marginal distributions, as the mapping g in the proof of Theorem 1 induces a bijection between the raw ROC diagnostic and the range of F that can be expressed in terms of π_1 and π_0. From this simple fact the following result is immediate.

Corollary 1 *The joint distribution $\mathbb{Q}$ of (S, Y) is characterized by the ROC curve and the marginal distributions of S and Y.*

Given a ROC curve R, an obvious task is to find CDFs F_0 and F_1 that realize R. For a particularly simple and appealing construction, let F_0 be the CDF of the uniform distribution on the unit interval, and define F_1 as $F_1(s) = 0$ for $s \le 0$,

$$F_1(s) = 1 - R_+(1 - s) \quad \text{for} \quad s \in (0, 1), \tag{4}$$

and $F_1(s) = 1$ for $s \ge 1$, where the function $R_+ : (0, 1) \rightarrow [0, 1]$ is induced by the ROC curve at hand, in that $R_+(s) = \inf \{ b : (a, b)' \in R, a \ge s \}$.

2.2 Concave ROC curves

We proceed to elucidate the critical role of concavity in the interpretation and modelling of ROC curves. Its significance is well known and has been reviewed in monographs (Pepe 2003; Zhou et al. 2011). Nevertheless, we are unaware of any rigorous treatment in the extant literature that applies in both continuous and discrete settings, which is what we address now.

In the *regular setting* we suppose that F_1 and F_0 have continuous, strictly positive Lebesgue densities f_1 and f_0 in the interior of an interval, which is their common support. For every s in the interior of the support, we can define the *likelihood ratio*,

$$\text{LR}(s) = \frac{f_1(s)}{f_0(s)},$$

and the *conditional event probability*,

$$\text{CEP}(s) = \mathbb{Q}(Y = 1 \mid S = s) = \frac{\pi_1 f_1(s)}{\pi_0 f_0(s) + \pi_1 f_1(s)}.$$

We note the equivalence of the following three conditions:

(a)   The ROC curve is concave.
(b)   The likelihood ratio is nondecreasing.
(c)   The conditional event probability is nondecreasing.

Theorem 2 *In the regular setting statements (a), (b), and (c) are equivalent.*

Next we consider the *discrete setting* where we assume that the support of the score S is finite or countably infinite. This setting includes, but is not limited to, the case of empirical ROC curves. For every s in the discrete support of S, we can define the *likelihood ratio*,

$$\text{LR}(s) = \begin{cases} \mathbb{Q}(S = s \mid Y = 1)\big/\mathbb{Q}(S = s \mid Y = 0) & \text{if} \quad \mathbb{Q}(S = s \mid Y = 0) > 0, \\ \infty, & \text{if} \quad \mathbb{Q}(S = s \mid Y = 0) = 0, \end{cases}$$

and the *conditional event probability*,

$$\text{CEP}(s) = \mathbb{Q}(Y = 1 \mid S = s).$$

Theorem 3 *In the discrete setting statements* (*a*), (*b*), *and* (*c*) *are equivalent.*

We skip proofs as Corollary 2.8 of Mösching and Dümbgen (2021) implies the equivalence of (a) and (b) in both settings. Furthermore, $\text{LR}(s) \propto \text{CEP}(s)/(1 - \text{CEP}(s))$, and the function $c \mapsto c/(1 - c)$ is nondecreasing, which yields the equivalence of (b) and (c).

The critical role of concavity in the interpretation and modelling of ROC curves stems from the monotonicity condition (c) on the conditional event probability, which needs to be invoked to justify the thresholding that lies at the heart of ROC analysis. When theoretical models are fit to empirical ROC curves, the model parameters can be restricted to guarantee concavity. Empirical ROC curves typically fail to be concave, but can be morphed into their concave hull, by subjecting a score to the pool-adjacent violators algorithm, thereby converting it into an isotonic, calibrated probabilistic classifier (Fawcett and Niculescu-Mizil 2007).

2.3 An equivalence between ROC curves and probability measures

We move on to provide concise and practically relevant characterizations of ROC curves.

Theorem 4 *There is a one-to-one correspondence between ROC curves and probability measures on the unit interval.*

Proof Given a ROC curve, we remove any vertical line segments, except for the respective upper endpoints, to yield the CDF of a probability measure on the unit interval. Conversely, given the CDF of a probability measure on the unit interval, we interpolate vertically at any jump points to obtain a ROC curve. This mapping is a bijection, and save for the symmetries in (4) it is realized by the construction in the proof of Corollary 1. □

We say that a curve C in the Euclidean plane is nondecreasing if $a_0 \leq a_1$ is equivalent to $b_0 \leq b_1$ for points $(a_0, b_0)', (a_1, b_1)' \in C$. The following result then is immediate.

Corollary 2 *The ROC curves are the nondecreasing curves in the unit square that connect the points* $(0, 0)'$ *and* $(1, 1)'$.

Natural analogues apply under the further constraints of either strict or non-strict concavity. From applied perspectives, these results support a shift of paradigms in the statistical modelling of ROC curves. In extant practice, the emphasis is on modelling the conditional distributions F_0 and F_1. Our characterizations suggest a subtle but important change

of perspective, in that ROC modelling can be approached as an exercise in curve fitting, with any nondecreasing curve that connects $(0, 0)'$ to $(1, 1)'$ being a permissible candidate.

3 Parametric models, estimation, and testing

The binormal model is by far the most frequently used parametric model and "plays a central role in ROC analysis" (Pepe 2003, p. 81). Specifically, the binormal model assumes that F_1 and F_0 are Gaussian with means $\mu_1 \geq \mu_0$ and strictly positive variances σ_0^2 and σ_1^2, respectively. We are in the regular setting of Sect. 2.2, and the resulting ROC curve is represented by the function $R : [0, 1] \to [0, 1]$ with $R(0) = 0$,

$$R(p) = \Phi(\mu + \sigma \, \Phi^{-1}(p)) \quad \text{for} \quad p \in (0, 1), \tag{5}$$

and $R(1) = 1$, where Φ is the CDF of the standard normal distribution, $\mu = (\mu_1 - \mu_0)/\sigma_1 \geq 0$ is a scaled difference in expectations, and $\sigma = \sigma_0/\sigma_1$ is the ratio of the respective standard deviations. The associated area under the curve is $\text{AUC}(\mu, \sigma) = \Phi(\mu/\sqrt{1 + \sigma^2})$. A binormal ROC curve is concave only if $\sigma = 1$ or, equivalently, if F_0 and F_1 differ in location only.

Our next result demonstrates that this restriction is unavoidable if location-scale families are used to model the class conditional distributions.

Proposition 1 *Given any strictly increasing CDF F on $\mathbb{R}$, let $F_0(s) = F((s - \mu_0)/\sigma_0)$ and $F_1(s) = F((s - \mu_1)/\sigma_1)$ for some $\mu_0, \mu_1 \in \mathbb{R}$ and $\sigma_0, \sigma_1 > 0$. Then the ROC curve associated with the conditional CDFs F_0 and F_1 is non-concave whenever $\sigma_0 \neq \sigma_1$.*

Proof If $\text{HR}(s) < \text{FAR}(s)$ for some $s \in \mathbb{R}$, the ROC curve extends below the diagonal and thus is non-concave. We note that $\text{HR}(s) < \text{FAR}(s) \iff F_0(s) < F_1(s) \iff (s - \mu_0)/\sigma_0 < (s - \mu_1)/\sigma_1$ and consider two cases. If $\sigma_1 > \sigma_0$ then $F_0(s) < F_1(s)$ for $s < (\mu_0\sigma_1 - \mu_1\sigma_0)/(\sigma_1 - \sigma_0)$; if $\sigma_0 > \sigma_1$ then $F_0(s) < F_1(s)$ for $s > (\mu_0\sigma_1 - \mu_1\sigma_0)/(\sigma_0 - \sigma_1)$. □

3.1 The beta model

Motivated and supported by the characterizations in Sect. 2, we propose a curve fitting approach to the statistical modelling of ROC curves, with the two-parameter family of the cumulative distribution functions (CDFs) of beta distributions being a particularly attractive model. Specifically, consider beta family with ROC curves represented by the function

$$R(p) = B_{\alpha,\beta}(p) = \int_0^p b_{\alpha,\beta}(q) \, dq \quad \text{for} \quad p \in [0, 1], \tag{6}$$

where $b_{\alpha,\beta}(q) \propto q^{\alpha-1}(1 - q)^{\beta-1}$ is the density of the beta distribution with parameter values $\alpha > 0$ and $\beta > 0$. This type of ROC curve arises as a special case of the setting in the case study by Zou et al. (2004, p. 1263), which models the class conditional distributions via beta densities. A beta ROC curve is concave if $\alpha \leq 1$ and $\beta \geq 2 - \alpha$, and its AUC value is $\text{AUC}(\alpha, \beta) = \beta/(\alpha + \beta)$ (Vogel 2019, Appendix 3.C). The condition for concavity is much less stringent than for the binormal family, where it constrains the admissible parameter space to a single dimension.

If further flexibility is desired, mixtures of beta CDFs, i.e., functions of the form

$$R_n(p) = \sum_{k=1}^{n} w_k B_{\alpha_k, \beta_k}(p) \quad \text{for} \quad p \in [0, 1],$$

where $w_1, \ldots, w_n > 0$ with $w_1 + \cdots + w_n = 1$, $\alpha_1, \ldots, \alpha_k > 0$, and $\beta_1, \ldots, \beta_k > 0$, approximate any regular ROC curve to any desired accuracy, as demonstrated by the following result. Recall from Sect. 2.2 that in the regular setting the ROC curve can be identified with the function R in (1), where F_1 and F_0 have continuous, strictly positive Lebesgue densities f_1 and f_0 in the interior of an interval, which is their common support. A ROC curve is *regular* if it arises in this way and *strongly regular* if furthermore the derivative R' is bounded.

Theorem 5 *For every strongly regular ROC curve R there is a sequence of mixtures of beta CDFs with integer parameters that converges uniformly to R.*

Proof We apply the construction in the proof of Corollary 1 and define F_1 as in (4). Due to the assumption of strong regularity, F_1 admits a density on $(0, 1)$ that can be extended to a continuous function f_1 on $[0, 1]$. The arguments in Bernstein's probabilistic proof of the Weierstrass approximation theorem (Levasseur 1984) show that as $n \to \infty$ the sequence

$$m_n(q) = \frac{1}{n+1} \sum_{k=0}^{n} f_1\left(\frac{k}{n}\right) b_{k+1, n-k+1}(q)$$

converges to $f_1(q)$ uniformly in $q \in [0, 1]$. Furthermore, $a_n = \int_0^1 m_n(q) \, dq \to \int_0^1 f_1(q) \, dq = 1$ as $n \to \infty$, and for $n = 1, 2, \ldots$ the mapping $p \mapsto M_n(p) = \int_0^p m_n(q) \, dq / a_n$ respresents a mixture of beta CDFs. The uniform convergence of m_n to f_1 implies that for every $\epsilon > 0$ there exists an n' such that

$$
\begin{aligned}
|F_1(p) - M_n(p)| &\leq \int_0^p \left| f_1(q) - \frac{m_n(q)}{a_n} \right| dq \\
&\leq \int_0^p \left| f_1(q) - \frac{f_1(q)}{a_n} \right| dq + \frac{1}{a_n} \int_0^p \left| f_1(q) - m_n(q) \right| dq \\
&\leq \left| 1 - \frac{1}{a_n} \right| + \frac{1}{a_n} \int_0^p \left| f_1(q) - m_n(q) \right| dq < \epsilon
\end{aligned}
$$

for all integers $n > n'$ uniformly in $p \in [0, 1]$. The statement of the theorem follows. $\qquad\square$

3.2 Minimum distance estimation

For the parametric estimation of ROC curves for continuous scores various methods have been proposed, including maximum likelihood, approaches based on generalized linear models, and minimum distance estimation (Pepe 2003; Zhou et al. 2011). Here we pursue the minimum distance estimator, which is much in line with our curve fitting approach.

We assume a parametric model in the regular setting of Sect. 2.2, where now the ROC curve depends on a parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. Specifically, we suppose that for each $\theta \in \Theta$ the ROC curve is represented by a smooth function

$$R(p;\theta) = 1 - F_{1,\theta}(F_{0,\theta}^{-1}(1-p)) \quad \text{for} \quad p \in (0,1),$$

where $F_{1,\theta}$ and $F_{0,\theta}$ admit continuous, strictly positive densities $f_{1,\theta}$ and $f_{0,\theta}$ in the interior of an interval, which is their common support. We also require that the true parameter value θ_0 is in the interior of the parameter space Θ, where the derivative

$$R'(p;\theta) = \frac{\partial R(p;\theta)}{\partial p} = \frac{f_{1,\theta}(F_{0,\theta}^{-1}(1-p))}{f_{0,\theta}(F_{0,\theta}^{-1}(1-p))}$$

exists and is finite for $p \in (0,1)$, and the partial derivative $R_{(i)}(p;\theta)$ with respect to component i of the parameter vector $\theta = (\theta_1, \ldots, \theta_k)$ exists and is continuous for $i = 1, \ldots, k$ and $p \in (0,1)$.

We adopt the asymptotic scenario of Hsieh and Turnbull (1996), where at sample size n there are n_0 and $n_1 = n - n_0$ independent draws from $F_{0,\theta}$ and $F_{1,\theta}$ with corresponding binary outcomes of zero and one, respectively, where $\lambda_n = n_0/n_1$ converges to some $\lambda \in (0, \infty)$ as $n \to \infty$. For $\theta \in \Theta$ we define the difference process $\xi_n(p;\theta) = \hat{R}_n(p) - R(p;\theta)$, where the function $\hat{R}_n(p)$ represents the empirical ROC curve. The minimum distance estimator $\hat{\theta}_n = (\hat{\theta}_1, \ldots, \hat{\theta}_k)_n$ then satisfies

$$\|\xi_n(\cdot\,;\hat{\theta}_n)\| = \min_{\theta \in \Theta} \|\xi_n(\cdot\,;\theta)\|,$$

where $\|\xi_n(\cdot\,;\theta)\| = (\int_0^1 \xi_n(p;\theta)^2 \, dp)^{1/2}$ is the standard L_2-norm. If n is large, $\hat{\theta}_n$ exists and is unique with probability approaching one (Millar 1984), and so we follow the extant literature in ignoring issues of existence and uniqueness.

The minimum distance estimator has a multivariate normal limit distribution in this setting, as suggested by the asymptotic result of Hsieh and Turnbull (1996) that under the usual $\sqrt{n}$ scaling the difference process $\xi_n(p;\theta)$ has limit

$$W(p;\theta) = \sqrt{\lambda}\, B_1(R(p;\theta)) + R'(p;\theta)B_2(p) \tag{7}$$

at $\theta = \theta_0$, where B_1 and B_2 are independent copies of a Brownian bridge. In contrast to the results in Section 4 of Hsieh and Turnbull (1996), which concern minimum distance estimation for the binormal model and ordinal dominance curves, the following theorem applies to general parametric families and ROC curves.

Theorem 6 *In the above setting the minimum distance estimator $\hat{\theta}_n$ satisfies*

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \to \mathcal{N}(0, C^{-1}AC^{-1}) \tag{8}$$

as $n \to \infty$, where the matrices A and C have entries

$$A_{ij} = \int_0^1 \int_0^1 R_{(i)}(s;\theta_0)K(s,t;\theta_0)R_{(j)}(t;\theta_0)\, ds\, dt, \quad C_{ij} = \int_0^1 R_{(i)}(s;\theta_0)R_{(j)}(s;\theta_0)\, ds \tag{9}$$

for $i, j = 1, \ldots, k$, respectively, and where

$$K(s,t;\theta_0) = \lambda\left(\min\{R(s;\theta_0), R(t;\theta_0)\} - R(s;\theta_0)R(t;\theta_0)\right) + R'(s;\theta_0)R'(t;\theta_0)(\min\{s,t\} - st). \tag{10}$$

is the covariance function of the process $W(p;\theta)$ in (7) at $\theta = \theta_0$.

Proof We are in the setting of Theorem 2.2 of Hsieh and Turnbull (1996), according to which there exists a probability space with sequences $(B_{1,n})$ and $(B_{2,n})$ of independent versions of Brownian bridges such that

$$\sqrt{n}\,\xi_n(p;\theta_0) = \sqrt{\lambda}\,B_{1,n}(R(p;\theta_0)) + R'(p;\theta_0)B_{2,n}(p) + o\big(n^{-1/2}(\log n)^2\big) \qquad (11)$$

almost surely, and uniformly in p on every interval $[a,b] \subset (0,1)$. We proceed to verify the regularity conditions for Theorem 3.6 of Millar (1984). As regards the identifiability condition (3.2) and the differentiability condition (3.5) it suffices to note that $\xi_n(p;\theta) - \xi_n(p;\theta_0) = R(p;\theta_0) - R(p;\theta)$ is nonrandom, continuously differentiable with respect to p and the components of the parameter vector θ, and independent of n. The boundedness condition (3.3) is trivially satisfied and the convergence condition (3.4) is implied by (11). Finally, we apply (2.17), (2.18), (2.19), and (2.20) in Section II of Millar (1984) to yield (8) and (9), where the covariance function of the process in (7) is

$$
\begin{aligned}
K(s,t;\theta) &= \mathrm{Cov}(W(s;\theta), W(t;\theta)) \\
&= \lambda\,\mathrm{Cov}(B_1(R(s;\theta)), B_1(R(t;\theta))) + R'(s;\theta)R'(t;\theta)\,\mathrm{Cov}(B_2(s), B_2(t)) \\
&= \lambda \quad (\min\{R(s;\theta), R(t;\theta)\} - R(s;\theta)R(t;\theta)) + R'(s;\theta)R'(t;\theta)(\min\{s,t\} - st),
\end{aligned}
$$

whence $K(s,t;\theta_0)$ is as stated in (10). $\qquad\qquad\square$

This result allows for asymptotic inference about model parameters, by plugging in $\hat{\theta}_n$ for θ_0 in the expression for the asymptotic covariance. For the binormal model (5) we have $\theta = (\mu,\sigma)$, $R_{(\mu)}(p;\theta) = \varphi(\mu + \sigma\,\Phi^{-1}(p))$, $R_{(\sigma)}(p;\theta) = \Phi^{-1}(p)\varphi(\mu + \sigma\,\Phi^{-1}(p))$, and $R'(p;\theta) = \sigma\,\varphi(\mu + \sigma\,\Phi^{-1}(p))/\varphi(\Phi^{-1}(p))$, where φ is the standard normal density, so that the integrals in (9) can readily be evaluated numerically. Under the beta model (6) we have $\theta = (\alpha,\beta)$ and $R'(p;\theta) = b_{\alpha,\beta}(p)$. While closed form expressions for the partial derivatives of $R(p;\theta)$ with respect to α and β exist, they are difficult to evaluate, and we approximate them with finite differences.

3.3 Testing goodness-of-fit

A natural question is whether a given parametric model fits the data at hand, and for doing this we propose a simple Monte Carlo test that applies to any parametric model class $\mathcal{C}$. Specifically, given a dataset of size n with n_0 instances where the binary outcome is zero and $n_1 = n - n_0$ instances where it is one, our goodness-of-fit test proceeds as follows. We use the notation of Sect. 3.2 and denote the number of Monte Carlo replicates by M.

1. Fit a model from class $\mathcal{C}$ to the empirical ROC curve, to yield the minimum distance estimate θ_{data}. Compute d_{data} as the L_2-distance between the fitted and the empirical ROC curve.
2. For $m = 1, \dots, M$,

    (a) draw a sample of size n under θ_{data}, with n_0 and n_1 instances from $F_{0,\theta_{\text{data}}}$ and $F_{1,\theta_{\text{data}}}$ and associated binary outcomes of zero and one, respectively,

    (b) fit a model from class $\mathcal{C}$ to the empirical ROC curve, to yield the minimum distance estimate, and

    (c) compute d_m as the L_2-distance between the fitted and the empirical ROC curve.

Table 1 Basic information about the datasets, and minimum distance estimates under the unrestricted and concave binormal and beta models for the ROC curves in Fig. 1

Dataset	Etzioni et al. (1999)	Sing et al. (2005)	Robin et al. (2011)	Vogel et al. (2018)
Binary outcome	Prostate cancer	Coreceptor usage	Clinical outcome	Precipitation
Score/feature	Antigen ratio	SVM predictor	$S100\beta$ concentr.	NWP forecast
Sample size	116	3450	113	5449
Binormal model				
unrestricted (μ, σ)	(1.05, 0.78)	(1.58, 0.65)	(0.75, 0.72)	(1.13, 1.22)
fit	0.043	0.019	0.033	0.008
p-value	0.109	0.001	0.593	0.031
concave (μ, σ)	(1.22, 1.00)	(2.05, 1.00)	(0.91, 1.00)	(0.99, 1.00)
fit	0.056	0.039	0.060	0.031
p-value	0.125	0.001	0.114	0.001
Beta model				
unrestricted (α, β)	(0.34, 1.32)	(0.15, 1.44)	(0.36, 0.96)	(0.79, 2.57)
fit	0.042	0.023	0.032	0.006
p-value	0.121	0.001	0.625	0.157
concave (α, β)	(0.38, 1.61)	(0.17, 1.83)	(0.52, 1.48)	(0.79, 2.57)
fit	0.045	0.025	0.050	0.006
p-value	0.212	0.001	0.203	0.157

Fit is in terms of the L_2-distance to the empirical ROC curve, and the *p*-value is from the goodness-of-fit test of Sect. 3.3

3. Find a *p*-value based on the rank of d_{data} when pooled with $d_1, \ldots, d_M$. Specifically, $p = (\#\{i = 1, \ldots, M : d_{\text{data}} \leq d_i\} + 1)/(M + 1)$.

Under the null hypothesis of the ROC curve being generated by a random sample within class $\mathcal{C}$ the Monte Carlo *p*-value is very nearly uniform, as is readily seen in simulation experiments.

4 Empirical examples

Basic information about the datasets in our empirical examples is given in Table 1. In the dataset from Etzioni et al. (1999), the ratio of free to total prostate-specific antigen (PSA) two years prior to diagnosis in serum from patients later found to have prostate cancer is compared to age-matched controls. The datasets from Sing et al. (2005, Figure1a) and Robin et al. (2011, Figure 1) are prominent examples in the widely used ROCR and pROC packages in R. They concern a score from a linear support vector machine (SVM) trained to predict the usage of HIV coreceptors, and the $S100\beta$ biomarker as it relates to a binary clinical outcome. The dataset from Vogel et al. (2018, Figure 6d) considers probability of precipitation forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) for the binary event of precipitation occurrence in the West Sahel region.

Figure 1 shows binormal and beta ROC curves fitted to the empirical ROC curves, both in the unrestricted case and under the constraint of concavity. The respective unrestricted and restricted minimum distance estimates, the fit in terms of the L_2-distance to
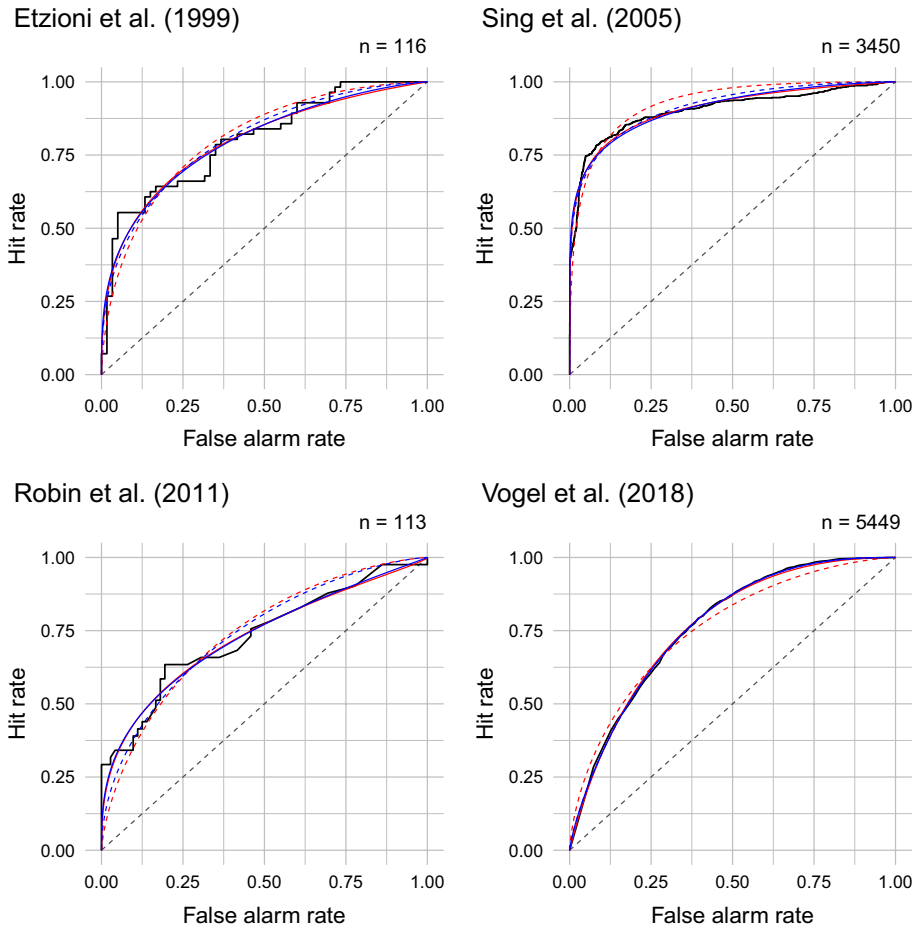
Fig. 1 Empirical (black), fitted binormal (red) and fitted beta (blue) ROC curves in the unrestricted (solid) and concave (dashed) case for the datasets from Table 1 (Color figure online)

the empirical ROC curve, and the *p*-value from the goodness-of-fit test of Sect. 3.3 with $M = 999$ Monte Carlo replicates, are given in Table 1. In the unrestricted case, the binormal and beta fits are visually nearly indistinguishable. The fitted binormal ROC curves fail to be concave and change markedly when concavity is enforced. For the beta ROC curves, the differences between restricted and unrestricted fits are less pronounced, and in the example from Vogel et al. (2018) the unrestricted fit is concave. For this dataset, our goodness-of-fit test rejects both the unrestricted and the concave binormal model, but does not reject the beta model—so the use of the more flexible beta family is of relevance and import. Generally, in the constrained case the improvement in the fit under the more flexible beta model as compared to the classical binormal model is substantial.

For more detailed analyses of these datasets, which include a demonstration of the use of our asymptotic results to generate confidence bands, as well as four-parameter extensions of the beta family, we refer to Vogel (2019, Section 3.4). Omar and Ivrissimtzis (2019) discuss the use of parametric families of ROC curves specifically in machine

learning. Datasets and code in R (R Core Team 2021) for replicating our results and implementing the proposed estimators and tests are available online (Vogel and Jordan 2021).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Etzioni, R., Pepe, M., Longton, G., Hu, C., & Goodman, G. (1999). Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer. *Medical Decision Making, 19*, 242–251.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874.

Fawcett, T., & Niculescu-Mizil, A. (2007). PAV and the ROC convex hull. *Machine Learning, 68*, 97–106.

Flach, P. A. (2016). ROC analysis. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining*. Boston: Springer.

Hsieh, F., & Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics, 24*, 25–40.

Krzanowski, D. J., & Hand, D. (2009). *ROC Curves for Continuous Data*. Boca Raton: CRC Press.

Levasseur, K. M. (1984). A probabilistic proof of the Weierstrass approximation theorem. *American Mathematical Monthly, 91*, 249–250.

Millar, P. W. (1984). A general approach to the optimality of minimum distance estimators. *Transactions of the American Mathematical Society, 286*, 377–418.

Mösching, A., & Dümbgen, L. (2021). Estimation of a likelihood ratio ordered family of distributions — with a connection to total positivity. Preprint, arXiv:2007.11521v2.

Omar, L., & Ivrissimtzis, I. (2019). Using theoretical ROC curves for analysing machine learning binary classifiers. *Pattern Recognition Letters, 128*, 447–451.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press.

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12*, 77.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: Visualizing classifier performance in R. *Bioinformatics, 21*, 3940–3941.

Vogel, P. (2019). Assessing Predictive Performance: From Precipitation Forecasts Over the Tropics to Receiver Operating Characteristic Curves and Back. PhD Thesis, Karlsruhe Institute of Technology, Faculty of Mathematics, available online at https://publikationen.bibliothek.kit.edu/1000091649.

Vogel, P., & Jordan, A. I. (2021). Replication package for "Receiver operating characteristic (ROC) curves: Equivalences, beta model, and minimum distance estimation" (version v0.1.0) [data set]. *Zenodo*. https://doi.org/10.5281/zenodo.4681331

Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A., & Gneiting, T. (2018). Skill of global raw and post-processed ensemble predictions of rainfall over northern tropical Africa. *Weather and Forecasting, 33*, 369–388.

Zhou, X.-H., Obuchowski, N. A., & McClish, D. K. (2011). *Statistical Methods in Diagnostic Medicine* (2nd ed.). Hoboken: Wiley.

Zou, K. H., Wells, W. M., III., Kikinis, R., & Warfield, S. K. (2004). Three validation metrics for automated probabilistic image segmentation of brain tumours. *Statistics in Medicine, 23*, 1259–1282.