

Nonparametric estimation of the ROC curve based on smoothed empirical distribution functions

Alicja Jokiel-Rokita · Michał Pulit

Received: 15 October 2011 / Accepted: 4 June 2012 / Published online: 28 June 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract The receiver operating characteristic (ROC) curve is a graphical representation of the relationship between false positive and true positive rates. It is a widely used statistical tool for describing the accuracy of a diagnostic test. In this paper we propose a new nonparametric ROC curve estimator based on the smoothed empirical distribution functions. We prove its strong consistency and perform a simulation study to compare it with some other popular nonparametric estimators of the ROC curve. We also apply the proposed method to a real data set.

Keywords Receiver operating characteristic (ROC) curve · Empirical distribution function · Nonparametric estimation

1 Introduction

The receiver operating characteristic (ROC) curve is commonly used to describe the accuracy of a medical or another diagnostic test, which classifies individuals into “healthy” and “diseased” categories. For comprehensive review of the literature, see the books Zhou et al. (2002), Pepe (2003) and Krzanowski and Hand (2009). Suppose that the independent real random variables X and Y denote the test score from healthy and diseased patients, respectively, and for a given cutoff point c , the test result is positive if it is greater than c . Let F and G be completely unknown distribution functions

of the random variables X and Y , respectively. The sensitivity of the test is defined as $SE(c) = 1 - G(c)$, which is the probability that a truly diseased individual has a positive test result. Similarly, the specificity of the test is given by $SP(c) = F(c)$ and describes the probability that a truly non-diseased individual has a negative test result. The receiver operating characteristic (ROC) curve is defined as a plot of $SE(c)$ versus $1 - SP(c)$ for $-\infty \leq c \leq \infty$, or equivalently as a plot of

$$ROC(t) = 1 - G(F^{-1}(1 - t)) \quad (1)$$

against t , for $t \in [0, 1]$.

There exist many different methods of estimating the ROC curve, but most of them are based on parametric or semiparametric models (e.g. Pepe 2000; Qin and Zhang 2003; Davidov and Nov 2012). In this paper we are interested in nonparametric estimation, which seems to be more reliable. Let $\mathbf{X}_m = (X_1, \dots, X_m)$ and $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ be simple independent random samples from the healthy and diseased populations, respectively. There exist several methods of estimating nonparametrically the ROC curve from such data. The commonly used nonparametric estimator is the empirical ROC curve of the form

$$ROC_{m,n}(t) = 1 - G_n(F_m^{-1}(1 - t)), \quad t \in [0, 1], \quad (2)$$

where F_m^{-1} and G_n respectively denote the empirical quantile function and the empirical cumulative distribution function of the samples \mathbf{X}_m and \mathbf{Y}_n , respectively (e.g. Hsieh and Turnbull 1996, Bowyer et al. 2001). Asymptotic properties of this estimator were studied by Hsieh and Turnbull (1996). They showed that, under some basic assumptions for distribution functions F and G ,

$$\sup_{t \in [0,1]} |ROC_{m,n}(t) - ROC(t)| \xrightarrow{n \rightarrow \infty} 0 \quad \text{a.s.}$$

A. Jokiel-Rokita (✉) · M. Pulit
Institute of Mathematics and Computer Science,
Wrocław University of Technology, Wrocław, Poland
e-mail: alicja.jokiel-rokita@pwr.wroc.pl

M. Pulit
e-mail: michal.pulit@pwr.wroc.pl

when $m = m(n)$ is nondecreasing function of n , and $m(n) \rightarrow \infty$ as $n \rightarrow \infty$.

The empirical ROC curve retains many properties of the empirical distribution function. It is uniformly convergent to the theoretical curve (Hsieh and Turnbull 1996), but it is also not continuous and not very accurate for small sample sizes. Other methods are needed to obtain a smooth estimator of the ROC curve.

Lloyd (1998) used the kernel smoothing technique to obtain a smooth ROC curve estimator given by

$$\widetilde{ROC}_{m,n}(t) = 1 - \widetilde{G}_n(\widetilde{F}_m^{-1}(1-t)), \quad t \in [0, 1], \tag{3}$$

where

$$\widetilde{F}_m(x) = \frac{1}{m} \sum_{j=1}^m \mathcal{K}\left(\frac{x - X_j}{h_m}\right),$$

$$\widetilde{G}_n(x) = \frac{1}{n} \sum_{j=1}^n \mathcal{K}\left(\frac{x - Y_j}{h_n}\right)$$

are standard kernel estimators with kernel function K , $\mathcal{K}(v) = \int_{-\infty}^v K(z)dz$ and bandwidth parameters h_n and h_m . Lloyd and Yong (1999) showed that estimator (3) has better mean squared error properties than the empirical ROC curve. In the problem of kernel density estimation, choosing between the many available kernel functions is relatively unimportant as all give comparable results, but more care needs to be taken over the selection of bandwidth. Therefore, in the kernel ROC curve estimation the main emphasis is put on the bandwidth selection (Zhou and Harezlak 2002; Hall and Hyndmann 2003). Unfortunately, to the best of our knowledge, in the case of estimator (3), there is no uniform, but only pointwise convergence to the theoretical ROC curve.

In the problem of kernel distribution function estimation, Zieliński (2007) proposed to replace the standard smoothing parameter h_m by random bandwidth

$$H_m = \min\{X_{j:m} - X_{(j-1):m} : j = 2, \dots, m\}, \tag{4}$$

where $X_{1:m} \leq X_{2:m} \leq \dots \leq X_{m:m}$ are order statistics from the sample \mathbf{X}_m . He obtained a continuous estimator of the unknown distribution function with asymptotic properties similar to the empirical distribution function. Unfortunately, his estimator is not invertible, because it is constant on some subintervals of the real line \mathbb{R} , so it cannot be used to obtain a continuous ROC curve estimator.

In the next section, basing on the idea of Zieliński (2007), we propose a construction of continuous and easily invertible estimator of the distribution function. It leads us to obtain a continuous and strictly increasing nonparametric estimator of the ROC curve, which is in fact the smoothed version of the empirical ROC curve. We prove that proposed

estimator converges uniformly to the theoretical ROC curve, almost surely. In Sect. 3 we report results of simulation studies and compare efficiency of the proposed estimator with some other nonparametric estimators. In Sect. 4 we apply the proposed estimator to a real data set. Section 5 contains conclusions and some prospects.

2 Smoothed empirical ROC curve

Let $\mathbf{X}_m = (X_1, \dots, X_m)$ and $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ be random samples from unknown continuous distribution functions F and G defined on the real line \mathbb{R} , respectively. Let $X_{1:m} \leq X_{2:m} \leq \dots \leq X_{m:m}$ and $Y_{1:n} \leq Y_{2:n} \leq \dots \leq Y_{n:n}$ denote order statistics from the samples \mathbf{X}_m and \mathbf{Y}_n . We set

$$X_{0:m} = 2L - X_{1:m}, \quad X_{(m+1):m} = 2U - X_{m:m},$$

$$Y_{0:n} = 2L - Y_{1:n}, \quad Y_{(n+1):n} = 2U - Y_{n:n},$$

where L, U are random variables such that $L \leq \min\{X_{1:m}, Y_{1:n}\}$ and $U \geq \max\{X_{m:m}, Y_{n:n}\}$ almost surely. Denote

$$Q_j(\mathbf{X}_m) = \frac{X_{(j-1):m} + X_{j:m}}{2}, \quad j = 1, 2, \dots, m + 1,$$

$$R_j(\mathbf{X}_m) = Q_{j+1}(\mathbf{X}_m) - Q_j(\mathbf{X}_m) = \frac{X_{(j+1):m} - X_{(j-1):m}}{2}, \quad j = 1, 2, \dots, m,$$

$$Q_j(\mathbf{Y}_n) = \frac{Y_{(j-1):n} + Y_{j:n}}{2}, \quad j = 1, 2, \dots, n + 1,$$

$$R_j(\mathbf{Y}_n) = Q_{j+1}(\mathbf{Y}_n) - Q_j(\mathbf{Y}_n) = \frac{Y_{(j+1):n} - Y_{(j-1):n}}{2}, \quad j = 1, 2, \dots, n.$$

With this notation we define the distribution functions estimators given by

$$\widehat{F}_m(x) = \frac{1}{m} \sum_{j=1}^m T\left(\frac{x - Q_j(\mathbf{X}_m)}{R_j(\mathbf{X}_m)}\right), \tag{5}$$

$$\widehat{G}_n(x) = \frac{1}{n} \sum_{j=1}^n T\left(\frac{x - Q_j(\mathbf{Y}_n)}{R_j(\mathbf{Y}_n)}\right), \tag{6}$$

where

$$T(x) = \begin{cases} 0, & \text{for } x < 0, \\ r(x), & \text{for } 0 \leq x \leq 1, \\ 1, & \text{for } x > 1, \end{cases} \tag{7}$$

where $r: [0, 1] \rightarrow [0, 1]$ is a continuous, strictly increasing function such that $r(0) = 0, r(1) = 1$, e.g. $r(x) = x$. In

comparison with the kernel estimator proposed by Zieliński (2007), we have replaced the random bandwidth H_m of the form (4) by the differences $R_j(\cdot) = Q_{j+1}(\cdot) - Q_j(\cdot)$ to make the estimators $\widehat{F}_m(x)$ and $\widehat{G}_n(x)$ continuous and strictly increasing on $[L, U]$. The plain order statistics have been replaced by the statistics $Q_j(\mathbf{X}_m)$ and $Q_j(\mathbf{Y}_n)$, indicating centers of the intervals between the consecutive order statistics. The purpose of this change was to avoid a situation in which the smoothed estimators are always below or above the empirical distribution functions.

Lemma 1 For each $x \in \mathbb{R}$

$$|\widehat{F}_m(x) - F_m(x)| \leq \frac{1}{m},$$

$$|\widehat{G}_n(x) - G_n(x)| \leq \frac{1}{n}.$$

Proof We give the proof only of the first inequality. The proof of the second inequality is essentially the same.

Consider arbitrary $x \in \mathbb{R}$. It must be the element of one of the disjoint intervals $(-\infty, Q_1(\mathbf{X}_m)), [Q_j(\mathbf{X}_m), Q_{j+1}(\mathbf{X}_m)), j = 1, 2, \dots, m, [Q_{m+1}(\mathbf{X}_m), \infty)$. It is easy to check that $\widehat{F}_m(x) = F_m(x) = 0$ for $x < Q_1(\mathbf{X}_m)$ and $\widehat{F}_m(x) = F_m(x) = 1$ for $x \geq Q_{m+1}(\mathbf{X}_m)$. Consider one of the other cases. Let $x \in [Q_j(\mathbf{X}_m), Q_{j+1}(\mathbf{X}_m)), j = 1, 2, \dots, m$. Then

$$\begin{aligned} 0 &\leq \frac{x - Q_j(\mathbf{X}_m)}{R_j(\mathbf{X}_m)} < \frac{Q_{j+1}(\mathbf{X}_m) - Q_j(\mathbf{X}_m)}{R_j(\mathbf{X}_m)} \\ &= \frac{R_j(\mathbf{X}_m)}{R_j(\mathbf{X}_m)} = 1. \end{aligned}$$

Moreover, for $i < j$

$$\begin{aligned} \frac{x - Q_i(\mathbf{X}_m)}{R_i(\mathbf{X}_m)} &\geq \frac{Q_j(\mathbf{X}_m) - Q_i(\mathbf{X}_m)}{R_i(\mathbf{X}_m)} \\ &\geq \frac{Q_{j+1}(\mathbf{X}_m) - Q_i(\mathbf{X}_m)}{R_i(\mathbf{X}_m)} = \frac{R_i(\mathbf{X}_m)}{R_i(\mathbf{X}_m)} = 1 \end{aligned}$$

and for $i > j$

$$\begin{aligned} \frac{x - Q_i(\mathbf{X}_m)}{R_i(\mathbf{X}_m)} &< \frac{Q_{j+1}(\mathbf{X}_m) - Q_i(\mathbf{X}_m)}{R_i(\mathbf{X}_m)} \\ &\leq \frac{Q_i(\mathbf{X}_m) - Q_i(\mathbf{X}_m)}{R_i(\mathbf{X}_m)} = 0. \end{aligned}$$

Hence

$$\begin{aligned} \widehat{F}_m(x) &= \frac{1}{m} \sum_{i=1}^m T\left(\frac{x - Q_i(\mathbf{X}_m)}{R_i(\mathbf{X}_m)}\right) \\ &= \frac{j-1}{m} + \frac{1}{m} r\left(\frac{x - Q_j(\mathbf{X}_m)}{R_j(\mathbf{X}_m)}\right) \end{aligned}$$

and consequently

$$\frac{j-1}{m} \leq \widehat{F}_m(x) < \frac{j}{m}.$$

Of course, the same inequality holds in the case of the empirical distribution function $F_m(x)$. It follows that

$$|\widehat{F}_m(x) - F_m(x)| \leq \frac{1}{m},$$

which completes the proof. \square

The following lemma is an immediate consequence of Lemma 1 and the Glivenko-Cantelli theorem.

Lemma 2 We have

$$\lim_{m \rightarrow \infty} \sup_{x \in \mathbb{R}} |\widehat{F}_m(x) - F(x)| = 0,$$

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\widehat{G}_n(x) - G(x)| = 0,$$

with probability one.

The inverse function of $\widehat{F}_m(t)$ on $[L, U]$ can be written as

$$\widehat{F}_m^{-1}(t) = \begin{cases} L, & \text{for } t = 0, \\ r^{-1}(mt - (k-1)R_k(\mathbf{X}_m) + Q_k(\mathbf{X}_m)), & \\ & \text{for } \frac{k-1}{m} < t \leq \frac{k}{m}, k = 1, \dots, m. \end{cases}$$

It is clear that $\widehat{F}_m^{-1}(t)$ is continuous and strictly increasing on $[0, 1]$. Since $\widehat{G}_n(t)$ is continuous and strictly increasing on $[L, U]$, it follows that the composition $\widehat{G}_n(\widehat{F}_m^{-1}(t))$ is continuous and strictly increasing on $[0, 1]$.

Hence we can define the continuous and strictly increasing ROC curve estimator given by

$$\widehat{ROC}_{m,n}(t) = 1 - \widehat{G}_n(\widehat{F}_m^{-1}(1-t)), \quad t \in [0, 1]. \tag{8}$$

An appropriate choice of the function r appearing in formula (7) can guarantee differentiability of the estimator (e.g. if function r is differentiable and $r'_+(0) = r'_-(1) = 0$). Simultaneously, a determination of the estimator (8) remains as easy as in the case of the empirical ROC curve.

Figure 1 shows an example of a true ROC curve corresponding to the normal distributions $\mathcal{N}(0, 1)$ and $\mathcal{N}(1, 1)$, the fitted empirical ROC curve $R_{m,n}$ and the smoothed non-parametric ROC curve $\widehat{ROC}_{m,n}$ for $m = n = 50$.

Lemma 3 Let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of nondecreasing continuous surjective functions such that $f_n: \mathbb{R} \xrightarrow{\text{onto}} [0, 1]$. Let $f_n^{-1}(y) = \inf\{x : f_n(x) = y\}$. Assume that $\sup_{x \in \mathbb{R}} |f_n(x) - f(x)| \xrightarrow{n \rightarrow \infty} 0$, where $f: \mathbb{R} \xrightarrow{\text{onto}} (0, 1)$ is differentiable surjective function such that $f'(x) > 0$ for all $x \in \mathbb{R}$. Then:

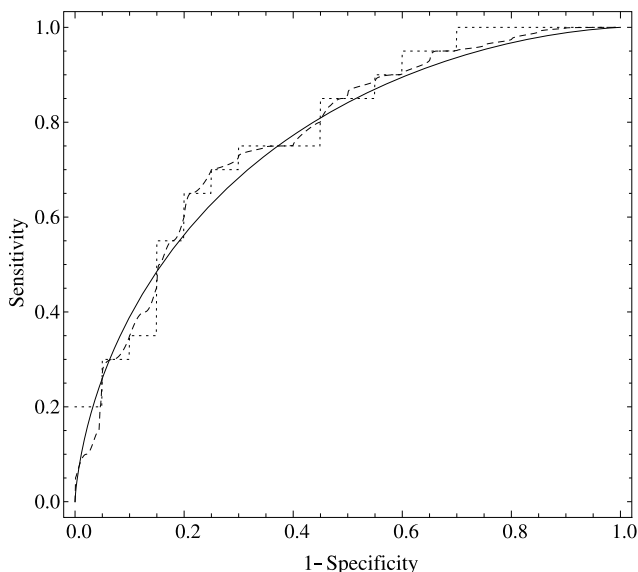


Fig. 1 The fitted empirical ROC curve (dotted curve), the smoothed nonparametric estimator of the ROC curve (dashed curve) and the true ROC curve (solid curve) for a sample data, generated from the normal distributions $\mathcal{N}(0, 1)$ and $\mathcal{N}(1, 1)$

- (1) $f_n^{-1}(y) \xrightarrow{n \rightarrow \infty} f^{-1}(y)$ for all $y \in (0, 1)$,
- (2) $\sup_{y \in [a,b]} |f_n^{-1}(y) - f^{-1}(y)| \xrightarrow{n \rightarrow \infty} 0$, for every $0 < a < b < 1$.

Proof Fix $y \in (0, 1)$ and $\varepsilon > 0$. To prove statement (1) of the lemma we need to show that for sufficiently large n

$$f^{-1}(y) - \varepsilon \leq f_n^{-1}(y) \leq f^{-1}(y) + \varepsilon.$$

Function f_n is nondecreasing, so the above inequality is equivalent to

$$f_n(f^{-1}(y) - \varepsilon) \leq f_n(f_n^{-1}(y)) \leq f_n(f^{-1}(y) + \varepsilon).$$

Let us remark that $f_n(f_n^{-1}(y)) = f_n(\inf\{x : f_n(x) = y\}) = y$. Therefore, it suffices to show that

- (i) $f_n(f^{-1}(y) - \varepsilon) \leq y$,
- (ii) $f_n(f^{-1}(y) + \varepsilon) \geq y$,

for all n greater than same $n_0 \in \mathbb{N}$.

Let us now consider the expressions $f(f^{-1}(y) - \varepsilon)$ and $f(f^{-1}(y) + \varepsilon)$. Lagrange theorem states that there exist $c_1 \in (f^{-1}(y) - \varepsilon, f^{-1}(y))$ and $c_2 \in (f^{-1}(y), f^{-1}(y) + \varepsilon)$ such that

$$f(f^{-1}(y) - \varepsilon) = -f'(c_1)\varepsilon + f(f^{-1}(y)) = y - \varepsilon f'(c_1),$$

$$f(f^{-1}(y) + \varepsilon) = f'(c_2)\varepsilon + f(f^{-1}(y)) = y + \varepsilon f'(c_2).$$

Let $\varepsilon_0 = \min\{\varepsilon f'(c_1), \varepsilon f'(c_2)\} > 0$. The sequence $\{f_n\}_{n \in \mathbb{N}}$ converges uniformly to function f , so there exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ and for every $x \in \mathbb{R}$

$$f(x) - \varepsilon_0 \leq f_n(x) \leq f(x) + \varepsilon_0.$$

Thus, in particular, for $x = f^{-1}(y) \pm \varepsilon$ and for $n \geq n_0$

$$f_n(f^{-1}(y) - \varepsilon) \leq f(f^{-1}(y) - \varepsilon) + \varepsilon_0$$

$$= y - \varepsilon f'(c_1) + \varepsilon_0 \leq y,$$

$$f_n(f^{-1}(y) + \varepsilon) \geq f(f^{-1}(y) + \varepsilon) - \varepsilon_0$$

$$= y + \varepsilon f'(c_2) - \varepsilon_0 \geq y,$$

which proves (i), (ii) and hence statement (1) of the lemma.

Let us remark that the convergence proved above is not uniform, because constant n_0 depends on ε_0 , whose choice depends on constants c_1 and c_2 , so also depends on y . If we consider only $y \in [a, b]$, $0 < a < b < 1$, then we can take $\varepsilon_0 = \varepsilon \inf_{c \in [f^{-1}(a) - \varepsilon, f^{-1}(b) + \varepsilon]} f'(c) > 0$, which is not depend on y . Therefore there exist $n_1 \in \mathbb{N}$, independent of y , such that for $n \geq n_1$ inequalities (i) and (ii) are satisfied. This proves statement (2) of the lemma. \square

We can now formulate the main theorem of the paper.

Theorem 1 Let $\mathbf{X}_m = (X_1, \dots, X_m)$ and $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ be simple random samples from continuous cumulative distribution functions F and G , respectively. Let $\widehat{ROC}_{m,n}(t)$ denote the ROC curve estimator given by (8) and let $ROC(t) = 1 - G(F^{-1}(1 - t))$. Then

$$\sup_{t \in [0,1]} |\widehat{ROC}_{m,n}(t) - ROC(t)| \xrightarrow{n \rightarrow \infty} 0,$$

almost surely when $m = m(n)$ is nondecreasing function of n , and $m(n) \rightarrow \infty$ as $n \rightarrow \infty$.

Proof Consider the inequality

$$\begin{aligned} & \sup_{t \in [0,1]} |\widehat{ROC}_{m,n}(t) - ROC(t)| \\ &= \sup_{t \in [0,1]} |G(F^{-1}(t)) - \widehat{G}_n(\widehat{F}_m^{-1}(t))| \\ &\leq \sup_{t \in [0,1]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))| \\ &\quad + \sup_{t \in [0,1]} |G(\widehat{F}_m^{-1}(t)) - \widehat{G}_n(\widehat{F}_m^{-1}(t))|. \end{aligned} \tag{9}$$

It is clear that

$$\begin{aligned} 0 &\leq \sup_{t \in [0,1]} |G(\widehat{F}_m^{-1}(t)) - \widehat{G}_n(\widehat{F}_m^{-1}(t))| \\ &\leq \sup_{x \in \mathbb{R}} |G(x) - \widehat{G}_n(x)|. \end{aligned}$$

By the above inequality and Lemma 2, we have

$$\sup_{t \in [0,1]} |G(\widehat{F}_m^{-1}(t)) - \widehat{G}_n(\widehat{F}_m^{-1}(t))| \xrightarrow{n \rightarrow \infty} 0 \quad \text{a.s.}$$

What is left is to show that the first term on the right side of the inequality (9) converges to zero almost surely as $m \rightarrow \infty$. To prove this, fix $\varepsilon > 0$ and denote $M = \sup_{x \in \mathbb{R}} G'(x) < \infty$. Let $\varepsilon_1 = \varepsilon/2$, $\varepsilon_2 = \varepsilon/2M$ and let $\delta, \eta \in (0, 1)$ be such that $G(F^{-1}(\delta)) \leq \varepsilon_1$ and $G(F^{-1}(\eta)) \geq 1 - \varepsilon_1$.

Let us remark that

$$\begin{aligned} & \sup_{t \in [0,1]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))| \\ &= \max \left\{ \sup_{t \in [0,\delta]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))|, \right. \\ & \quad \sup_{t \in [\delta,\eta]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))|, \\ & \quad \left. \sup_{t \in [\eta,1]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))| \right\}. \end{aligned} \tag{10}$$

With respect to the first element of the maximum appearing in the equality (10), the following inequalities hold:

$$\begin{aligned} & \sup_{t \in [0,\delta]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))| \\ & \leq \sup_{t \in [0,\delta]} \max \{G(F^{-1}(t)), G(\widehat{F}_m^{-1}(t))\} \\ &= \max \{G(F^{-1}(\delta)), G(\widehat{F}_m^{-1}(\delta))\} \\ &= \max \{G(F^{-1}(\delta)), \\ & \quad G(\widehat{F}_m^{-1}(\delta)) - G(F^{-1}(\delta)) + G(F^{-1}(\delta))\} \\ & \leq \max \{G(F^{-1}(\delta)), \\ & \quad |G(\widehat{F}_m^{-1}(\delta)) - G(F^{-1}(\delta))| + G(F^{-1}(\delta))\} \\ & \leq \max \{\varepsilon_1, |G(\widehat{F}_m^{-1}(\delta)) - G(F^{-1}(\delta))| + \varepsilon_1\} \\ &= |G(\widehat{F}_m^{-1}(\delta)) - G(F^{-1}(\delta))| + \varepsilon_1 \\ & \leq \sup_{x \in \mathbb{R}} G'(x) |\widehat{F}_m^{-1}(\delta) - F^{-1}(\delta)| + \varepsilon_1 \\ &= M |\widehat{F}_m^{-1}(\delta) - F^{-1}(\delta)| + \varepsilon_1. \end{aligned}$$

Similarly, we obtain

$$\begin{aligned} & \sup_{t \in [\eta,1]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))| \\ & \leq \sup_{t \in [\eta,1]} \max \{1 - G(F^{-1}(t)), 1 - G(\widehat{F}_m^{-1}(t))\} \\ &= \max \{1 - G(F^{-1}(\eta)), 1 - G(\widehat{F}_m^{-1}(\eta))\} \\ &= \max \{1 - G(F^{-1}(\eta)), \\ & \quad 1 - G(\widehat{F}_m^{-1}(\eta)) + G(F^{-1}(\eta)) - G(F^{-1}(\eta))\} \\ & \leq \max \{1 - G(F^{-1}(\eta)), \end{aligned}$$

$$\begin{aligned} & |G(\widehat{F}_m^{-1}(\eta)) - G(F^{-1}(\eta))| + 1 - G(F^{-1}(\eta))\} \\ & \leq \max \{\varepsilon_1, |G(\widehat{F}_m^{-1}(\eta)) - G(F^{-1}(\eta))| + \varepsilon_1\} \\ &= |G(\widehat{F}_m^{-1}(\eta)) - G(F^{-1}(\eta))| + \varepsilon_1 \\ & \leq \sup_{x \in \mathbb{R}} G'(x) |\widehat{F}_m^{-1}(\eta) - F^{-1}(\eta)| + \varepsilon_1 \\ &= M |\widehat{F}_m^{-1}(\eta) - F^{-1}(\eta)| + \varepsilon_1. \end{aligned}$$

Finally, we have

$$\begin{aligned} & \sup_{t \in [\delta,\eta]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))| \\ & \leq \sup_{t \in [\delta,\eta]} \sup_{x \in \mathbb{R}} G'(x) |\widehat{F}_m^{-1}(\eta) - F^{-1}(\eta)| \\ &= M \sup_{t \in [\delta,\eta]} |\widehat{F}_m^{-1}(\eta) - F^{-1}(\eta)|. \end{aligned}$$

Replacing expressions appearing in formula (10), by their upper estimates obtained above, we get

$$\begin{aligned} & \sup_{t \in [0,1]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))| \\ & \leq \max \{M |\widehat{F}_m^{-1}(\delta) - F^{-1}(\delta)| + \varepsilon_1, \\ & \quad M \sup_{t \in [\delta,\eta]} |\widehat{F}_m^{-1}(\eta) - F^{-1}(\eta)|, \\ & \quad M |\widehat{F}_m^{-1}(\eta) - F^{-1}(\eta)| + \varepsilon_1\} \\ & \leq M \max \left\{ |\widehat{F}_m^{-1}(\delta) - F^{-1}(\delta)|, \right. \\ & \quad \left. \sup_{t \in [\delta,\eta]} |\widehat{F}_m^{-1}(\eta) - F^{-1}(\eta)|, \right. \\ & \quad \left. |\widehat{F}_m^{-1}(\eta) - F^{-1}(\eta)| \right\} + \varepsilon_1. \end{aligned} \tag{11}$$

Let us now remark that the sequence $\{\widehat{F}_m\}_{m \in \mathbb{N}}$ of cumulative distribution function estimators satisfies the conditions of Lemma 3 almost surely. Indeed, $\widehat{F}_m : \mathbb{R} \xrightarrow{\text{onto}} [0, 1]$ is nondecreasing continuous surjective function and $\widehat{F}_m(\widehat{F}_m^{-1}(t)) = t$ for all $t \in (0, 1)$. Moreover, by Lemma 2, the sequence $\{\widehat{F}_m\}_{m \in \mathbb{N}}$ converges uniformly to F and $F'(x) = f(x) > 0$ for all $x \in \mathbb{R}$. Therefore, by Lemma 3, statement (1), there exist $m_0 = m_0(\delta, \eta) \in \mathbb{N}$ such that for $m \geq m_0$

$$|\widehat{F}_m^{-1}(\delta) - F^{-1}(\delta)| < \varepsilon_2, \quad |\widehat{F}_m^{-1}(\eta) - F^{-1}(\eta)| \leq \varepsilon_2.$$

By Lemma 3, statement (2), there exist $m_1 \in \mathbb{N}$ such that for $m \geq m_1$

$$\sup_{t \in [\delta,\eta]} |F^{-1}(t) - \widehat{F}_m^{-1}(t)| \leq \varepsilon_2.$$

It follows that for $m \geq \max\{m_0, m_1\}$, we have

$$\begin{aligned} & \sup_{t \in [0,1]} |G(F^{-1}(t)) - G(\widehat{F}_m^{-1}(t))| \\ & \leq M \max \left\{ \left| \widehat{F}_m^{-1}(\delta) - F^{-1}(\delta) \right|, \right. \\ & \quad \left. \sup_{t \in [\delta, \eta]} \left| \widehat{F}_m^{-1}(\eta) - F^{-1}(\eta) \right|, \right. \\ & \quad \left. \left| \widehat{F}_m^{-1}(\eta) - F^{-1}(\eta) \right| \right\} + \varepsilon_1 \\ & \leq M\varepsilon_2 + \varepsilon_1 = M \frac{\varepsilon}{2M} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

This completes the proof. □

3 Simulation study

To investigate the performance of the proposed ROC curve estimator (8) for the realistic sample sizes (m, n) , a small simulation study was conducted. We have considered three different combinations of distributions for X and Y : (1) $X \sim \mathcal{N}(0, 1)$, $Y \sim \mathcal{N}(1, 1)$; (2) $X \sim \mathcal{LG}(0, 1)$, $Y \sim \mathcal{LG}(3, 1)$; (3) $X \sim \mathcal{N}(0, 9)$, $Y \sim \mathcal{LG}(3, 1)$. The corresponding true ROC curves are plotted in Fig. 2. The areas under the considered ROC curves equal 0.76025, 0.88697, 0.80561, respectively.

For each ROC curve we have generated 1000 training data sets with $(m, n) \in \{(16, 24), (20, 20), (24, 16), (40, 60), (50, 50), (60, 40), (100, 150), (125, 125), (150, 100)\}$. Next, for each data set, we have computed the empirical ROC

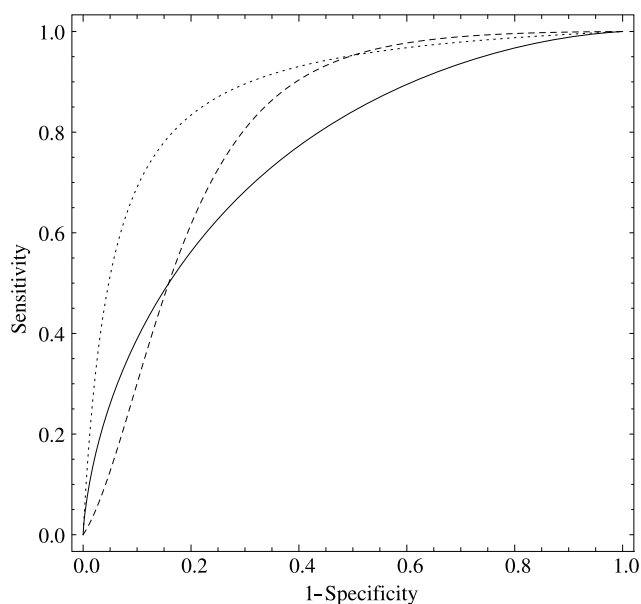


Fig. 2 The ROC curve corresponding to: (1) $X \sim \mathcal{N}(0, 1)$, $Y \sim \mathcal{N}(1, 1)$ (solid curve); (2) $X \sim \mathcal{LG}(0, 1)$, $Y \sim \mathcal{LG}(3, 1)$ (dotted curve); (3) $X \sim \mathcal{N}(0, 9)$, $Y \sim \mathcal{LG}(3, 1)$ (dashed curve)

curve estimator (2), the kernel ROC curve estimator (3) and the ROC curve estimator (8) proposed in this paper. In the problem of bandwidth selection for the kernel estimator, we have used the Normal-reference method proposed by Hall and Hyndmann (2003), which is recommended when the sampled distributions are not far from Normal. They found that in the context of the ROC curve estimation the method proposed gives substantial improvement in the mean integrated squared error (MISE) over other known methods of bandwidth selection.

In constructing the executable computer programs, procedures of the package Mathematica 8.0 were used.

To compare the accuracy of the proposed estimator with the accuracy of the empirical ROC curve estimator and the kernel ROC curve estimator we have estimated their MISE and integrated bias. We also have checked the accuracy of the considered methods in estimating the area under the curve (AUC) by calculating the mean squared error (MSE) of the areas under considered estimators. The estimated values of all mentioned error measures are illustrated in Tables 1, 2 and 3.

Table 1 shows that there are quite big differences in accuracy of the investigated estimators. Both $\widehat{ROC}_{m,n}$ and $\widetilde{ROC}_{m,n}$ provide more accurate estimators than empirical ROC curve. In most of cases, observed improvement is bigger for the kernel estimator, but it becomes really significant only when the total sample size $m + n$ gets larger. Interestingly, the proposed estimator $\widehat{ROC}_{m,n}$ performs better when both X and Y arise from the Logistic distributions. It is due to the fact, that the corresponding ROC curve lies closer to the upper left-hand corner of the graph and in such cases the kernel estimators of the ROC curve may perform worse than the empirical ones.

Comparing the integrated bias of the estimators, we do not consider the empirical ROC curve. Note that the integrated bias of the empirical ROC curve is equal to the difference between the expectation of the area under the empirical ROC curve and the real value of AUC. This is obviously equal to zero as the area under the empirical ROC curve is equal to the Mann-Whitney U-statistic, which is an unbiased estimator of the AUC (see e.g. Bamber 1975). For this reason, in Table 2, we only present the estimated integrated bias of the estimators $\widetilde{ROC}_{m,n}$ and $\widehat{ROC}_{m,n}$. Both of them have a negative bias, which means that they tend to underestimate the real curve. However, the bias of the estimator proposed in this paper is not big and its absolute value is always much smaller than in the case of the kernel estimator.

Results from Table 3 indicate that the proposed estimator, after integration, provides the most accurate estimator of AUC (in sense of MSE). The kernel estimator performs even worse than the empirical ROC curve, which is probably caused by the largest bias related to this estimator. The estimator of AUC obtained from the empirical ROC curve is unbiased, so its MSE is in fact equal to its variance.

Table 1 Estimated values of MISE and percentage reduction in MISE from using the kernel ROC curve estimator $\widetilde{ROC}_{m,n}$ and the proposed estimator $\widehat{ROC}_{m,n}$, instead of the empirical ROC curve $ROC_{m,n}$

	$m = 16$ $n = 24$	$m = 20$ $n = 20$	$m = 24$ $n = 16$	$m = 40$ $n = 60$	$m = 50$ $n = 50$	$m = 60$ $n = 40$	$m = 100$ $n = 150$	$m = 125$ $n = 125$	$m = 150$ $n = 100$
$X \sim \mathcal{N}(0, 1) \quad Y \sim \mathcal{N}(1, 1)$									
$ROC_{m,n}$	0.01472	0.01370	0.01402	0.00567	0.00547	0.00577	0.00237	0.00230	0.00228
	–	–	–	–	–	–	–	–	–
$\widetilde{ROC}_{m,n}$	0.01058	0.00951	0.00972	0.00424	0.00416	0.00439	0.00198	0.00184	0.00184
	28.1 %	30.6 %	30.7 %	25.2 %	23.9 %	23.9 %	16.5 %	20.0 %	19.3 %
$\widehat{ROC}_{m,n}$	0.01132	0.01063	0.01089	0.00500	0.00492	0.00512	0.00225	0.00220	0.00218
	23.1 %	22.4 %	22.3 %	11.8 %	10.1 %	11.2 %	5.1 %	4.3 %	4.4 %
$X \sim \mathcal{LG}(0, 1) \quad Y \sim \mathcal{LG}(3, 1)$									
$ROC_{m,n}$	0.01219	0.01165	0.01096	0.00526	0.00481	0.00458	0.00211	0.00194	0.00179
	–	–	–	–	–	–	–	–	–
$\widetilde{ROC}_{m,n}$	0.00967	0.00941	0.00884	0.00442	0.00406	0.00386	0.00181	0.00169	0.00156
	20.7 %	19.2 %	19.3 %	16.0 %	15.6 %	15.7 %	14.2 %	12.9 %	12.8 %
$\widehat{ROC}_{m,n}$	0.00792	0.00741	0.00705	0.00395	0.00371	0.00362	0.00180	0.00171	0.00160
	35.0 %	36.4 %	35.7 %	24.9 %	22.9 %	21.0 %	14.7 %	11.9 %	10.6 %
$X \sim \mathcal{N}(0, 9) \quad Y \sim \mathcal{LG}(3, 1)$									
$ROC_{m,n}$	0.01991	0.01637	0.01549	0.00788	0.00697	0.00639	0.00336	0.00288	0.00260
	–	–	–	–	–	–	–	–	–
$\widetilde{ROC}_{m,n}$	0.01465	0.01262	0.01191	0.00629	0.00568	0.00508	0.00280	0.00244	0.00218
	26.4 %	22.9 %	23.1 %	20.2 %	18.5 %	20.5 %	16.7 %	15.3 %	16.2 %
$\widehat{ROC}_{m,n}$	0.01502	0.01301	0.01261	0.00707	0.00638	0.00588	0.00323	0.00278	0.00252
	24.6 %	20.5 %	18.6 %	10.3 %	8.5 %	8.0 %	3.9 %	3.5 %	3.1 %

Table 2 Estimated values of the integrated bias of the kernel ROC curve estimator $\widetilde{ROC}_{m,n}$ and the proposed estimator $\widehat{ROC}_{m,n}$

	$m = 16$ $n = 24$	$m = 20$ $n = 20$	$m = 24$ $n = 16$	$m = 40$ $n = 60$	$m = 50$ $n = 50$	$m = 60$ $n = 40$	$m = 100$ $n = 150$	$m = 125$ $n = 125$	$m = 150$ $n = 100$
$X \sim \mathcal{N}(0, 1) \quad Y \sim \mathcal{N}(1, 1)$									
$\widetilde{ROC}_{m,n}$	-0.0299	-0.0262	-0.0271	-0.0150	-0.0167	-0.0185	-0.0116	-0.0089	-0.0101
$\widehat{ROC}_{m,n}$	-0.0132	-0.0094	-0.0104	-0.0037	-0.0027	-0.0045	-0.0013	0.0003	-0.0008
$X \sim \mathcal{LG}(0, 1) \quad Y \sim \mathcal{LG}(3, 1)$									
$\widetilde{ROC}_{m,n}$	-0.0297	-0.0278	-0.0284	-0.0181	-0.0169	-0.0168	-0.0098	-0.0098	-0.0097
$\widehat{ROC}_{m,n}$	-0.0247	-0.0229	-0.0242	-0.0087	-0.0074	-0.0075	-0.0022	-0.0022	-0.0022
$X \sim \mathcal{N}(0, 9) \quad Y \sim \mathcal{LG}(3, 1)$									
$\widetilde{ROC}_{m,n}$	-0.0282	-0.0280	-0.0277	-0.0187	-0.0173	-0.0151	-0.0103	-0.0100	-0.0086
$\widehat{ROC}_{m,n}$	-0.0130	-0.0145	-0.0185	-0.0043	-0.0049	-0.0043	-0.0008	-0.0016	-0.0009

4 Real data analysis

To illustrate our method, we apply it to the set of real data which comes from a clinical study performed from November 2008 to August 2011 by a research team led by Dr. Krzysztof Tupikowski from Department of Urology and Oncological Urology, Wroclaw Medical University (article in

press). One investigated the effectiveness of combined treatment of interferon alpha and metronomic cyclophosphamide in patients with metastatic kidney cancer not eligible for tyrosine kinase inhibitors treatment with various negative prognostic factors for survival. It has been approved by an independent local bioethics committee. One of the secondary goals of the study was to assess if there are any

Table 3 Estimated values of MSE in the problem of AUC estimation and percentage reduction (+) or percentage increase (–) in MSE from using the kernel ROC curve estimator $\widehat{ROC}_{m,n}$ and the proposed estimator $\widetilde{ROC}_{m,n}$, instead of the empirical ROC curve $ROC_{m,n}$

	$m = 16$ $n = 24$	$m = 20$ $n = 20$	$m = 24$ $n = 16$	$m = 40$ $n = 60$	$m = 50$ $n = 50$	$m = 60$ $n = 40$	$m = 100$ $n = 150$	$m = 125$ $n = 125$	$m = 150$ $n = 100$
$X \sim \mathcal{N}(0, 1) \quad Y \sim \mathcal{N}(1, 1)$									
$ROC_{m,n}$	0.00615	0.00559	0.00607	0.00225	0.00220	0.00245	0.00095	0.00096	0.00094
	–	–	–	–	–	–	–	–	–
$\widetilde{ROC}_{m,n}$	0.00647	0.00577	0.00614	0.00239	0.00241	0.00266	0.00107	0.00102	0.00102
	–4.9 %	–3.1 %	–1.1 %	–5.9 %	–8.7 %	–7.9 %	–11.2 %	–5.9 %	–7.8 %
$\widehat{ROC}_{m,n}$	0.00549	0.00493	0.00527	0.00213	0.00212	0.00235	0.00094	0.00094	0.00092
	10.7 %	11.8 %	13.2 %	5.3 %	3.6 %	4.1 %	1.1 %	2.1 %	2.1 %
$X \sim \mathcal{LG}(0, 1) \quad Y \sim \mathcal{LG}(3, 1)$									
$ROC_{m,n}$	0.00277	0.00296	0.00297	0.00116	0.00119	0.00116	0.00044	0.00045	0.00045
	–	–	–	–	–	–	–	–	–
$\widetilde{ROC}_{m,n}$	0.00402	0.00411	0.00411	0.00161	0.00158	0.00155	0.00056	0.00057	0.00057
	–31.1 %	–28.0 %	–27.7 %	–28.0 %	–24.7 %	–25.2 %	–21.4 %	–21.1 %	–21.1 %
$\widehat{ROC}_{m,n}$	0.00255	0.00262	0.00269	0.00109	0.00110	0.00106	0.00043	0.00043	0.00044
	7.9 %	11.5 %	9.4 %	6.0 %	7.6 %	8.6 %	2.3 %	4.4 %	2.2 %
$X \sim \mathcal{N}(0, 9) \quad Y \sim \mathcal{LG}(3, 1)$									
$ROC_{m,n}$	0.00592	0.00504	0.00469	0.00226	0.00209	0.00187	0.00098	0.00084	0.00073
	–	–	–	–	–	–	–	–	–
$\widetilde{ROC}_{m,n}$	0.00635	0.00561	0.00524	0.00257	0.00237	0.00206	0.00108	0.00094	0.00079
	–6.8 %	–10.2 %	–10.5 %	–12.1 %	–11.8 %	–9.2 %	–9.3 %	–10.6 %	–7.6 %
$\widehat{ROC}_{m,n}$	0.00511	0.00446	0.00422	0.00216	0.00199	0.00176	0.00096	0.00083	0.00071
	13.7 %	11.5 %	10.0 %	4.4 %	4.8 %	5.9 %	2.0 %	1.2 %	2.7 %

Table 4 The real data in the form (CR, HL, FC)

(0, 8.6, 7.8)	(0, 12.9, x)	(0, 11.5, 8.3)	(1, 11.8, 4.1)	(0, 14.7, 5)	(0, 11.9, 6.7)	(0, 9.5, 9.6)
(1, 13.3, 4.2)	(0, 14.1, x)	(0, 11.1, 6)	(1, 12.2, x)	(1, 15, 3.4)	(1, 17.2, 5.1)	(1, 13.9, 3.1)
(1, 13.1, 9.5)	(0, 9.1, 9.7)	(0, 7.2, 9)	(1, 15.4, 6.8)	(0, 11.7, 6.2)	(0, 10.9, 6.18)	(1, 12.1, 7.6)
(1, 13, x)	(0, 12.7, x)	(1, 10.3, 6)	(0, 10.9, 4.1)	(1, 11.9, 5.9)	(0, 13.9, 4.8)	(1, 14.2, 6)
(1, 11.5, 4.4)	(0, 14.5, 4.5)	(0, 11.9, 5.9)				

predictive factors for response to this novel combination treatment. Table 4 contains presence (1) (or absence—0) of clinical response (CR) observed at 24-th week of treatment, hemoglobin level (HL) and serum fibrinogen concentration (FC) of 31 patients treated per protocol. Missing data are denoted by x . Low HL has been previously associated with short survival and poor response to treatment in disseminated disease (Tonini et al. 2011). High FC is examined as a negative predictor of response to treatment in metastatic kidney cancer patients for the first time.

The estimators of the ROC curves for HL (left) and FC (right) as the predictive factors (positive and negative, respectively) are plotted in Fig. 3.

Figure 3 shows that in the case considered when sample sizes are small ($m = 17, n = 14$ for HL, and $m = 12, n = 14$

for FC) the proposed estimator seems to be better fitted to a (continuous) ROC curve than the empirical ROC curve.

The estimated values of the AUC, obtained using the empirical ROC curve estimator and the proposed estimator, are 0.710107 and 0.69524 for HL and 0.678643 and 0.680438 for FC, respectively, and they do not differ significantly.

5 Concluding remarks

In this article we have proposed a nonparametric estimator of the ROC curve. The new procedure is simple and constructs a strong consistent estimator of the unknown ROC curve. The strong consistency of this estimator was established in Theorem 1. Simulation study show that the proposed method of estimation has satisfactory finite sample

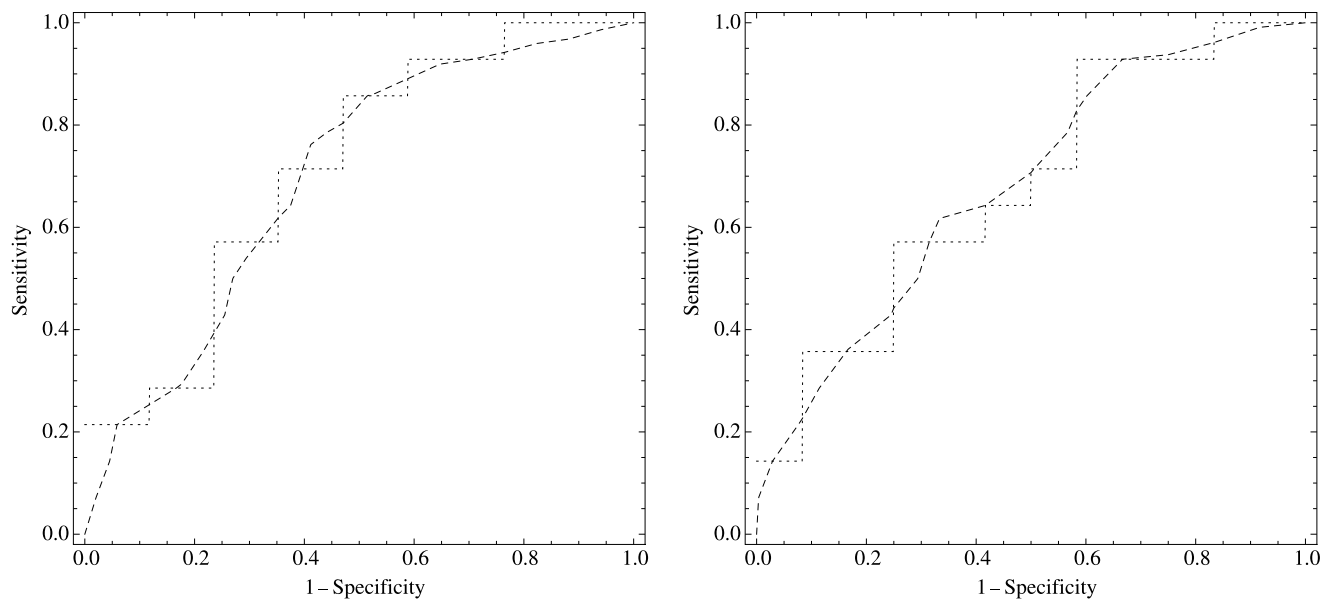


Fig. 3 The fitted empirical ROC curve (*dotted curve*), the smoothed nonparametric estimator of the ROC curve (*dashed curve*) for HL (*left*) and FC (*right*)

performance. The new estimators of the ROC curve are more accurate (with respect to the MISE) than the empirical ROC and more accurate on average for small sample sizes than the kernel estimators. Applying the proposed method of estimating the ROC curve to estimation of the AUC leads to more accurate (with respect to the MSE) estimators of the AUC than applying the empirical ROC curve in all cases considered and significantly more accurate than applying the kernel estimators. Using the kernel estimators leads even to less accurate estimators of the AUC than using empirical ROC curve. If the sample sizes are small and the distribution functions are completely unknown, we recommend using the proposed method of estimating the ROC curve. In the future research, we are going to use the proposed estimator to construct the confidence region for the ROC curve in the case of small sample sizes.

In this paper we have applied nonparametric approach to the problem of ROC estimation. Another approach, often used in the ROC estimation, is to assume that some unknown transformation converts both populations to a specific form, for example normal, logistic or Weibull. Such approach is often referred to either as a semiparametric or a parametric distribution-free approach. It was considered, for example, by Hsieh and Turnbull (1996), Zou and Hall (2000), Cai and Moskowitz (2004), Davidov and Nov (2012). Hsieh and Turnbull (1996) considered a problem of estimation of the ordinal dominance curve (ODC) given by $F(G^{-1}(t))$, $0 \leq t \leq 1$, which is the plot of $SP(c)$ versus $1 - SE(c)$, or equivalently $F(c)$ versus $G(c)$ for $-\infty \leq c \leq \infty$. They assumed that some unspecified monotonic transformation of the measurement scale simultaneously converts the F and

G distributions to normal ones. In this case the ODC has the known parametric form

$$F(G^{-1}(t)) = \Phi(\mu + \sigma \Phi^{-1}(t)), \quad t \in (0, 1),$$

where Φ denotes the standard normal cumulative distribution function. They suggested to estimate the parameters μ and σ applying minimum distance estimators (MDEs) (see Wolfowitz 1957) by the values that minimize the L_2 distance between the empirical and the theoretical ODCs, that is, by the values that solve

$$\text{minimize}_{\mu, \sigma} \int_0^1 [F_m(G_n^{-1}(t)) - \Phi(\mu + \sigma \Phi^{-1}(t))]^2 dt. \tag{12}$$

They studied the asymptotic properties of the MDEs, but did not provide any concrete procedure to compute them. In a remark they proposed, as an object for future research, to modify (12) by applying a Φ^{-1} transformation on both the empirical ODC and the true ODC. Davidov and Nov (2012) followed up on this suggestion and studied in detail the estimators

$$(\hat{\mu}, \hat{\sigma}) = \operatorname{argmin}_{\mu, \sigma} \int_a^b [\Phi^{-1}(F_m(G_n^{-1}(t))) - (\mu + \sigma \Phi^{-1}(t))]^2 dt.$$

They used $0 < a < b < 1$ as the integration endpoints rather than 0 and 1, as $F_m(G_n^{-1}(t)) = 0$, and hence $\Phi^{-1}(F_m(G_n^{-1}(t))) = -\infty$, for values of t too close to 0. A similar problem arises with values t too close to 1. In

future research, we are going to compare the performance of the nonparametric estimator proposed in this paper with semiparametric estimators considered by Hsieh and Turnbull (1996) and Davidov and Nov (2012), and also with the semiparametric and rank-based estimator considered by Zou and Hall (2000), the maximum profile likelihood estimator and pseudo maximum likelihood estimator proposed by Cai and Moskowitz (2004).

Acknowledgements The authors wish to thank the referees whose constructive comments and valuable suggestions led to considerable improvements in the presentation.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Bamber, D.: The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* **12**, 387–415 (1975)
- Bowyer, K., Kranenburg, C., Dougherty, S.: Edge detector evaluation using empirical ROC curves. *Comput. Vis. Image Underst.* **84**(1), 77–103 (2001)
- Cai, T., Moskowitz, C.S.: Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test. *Biostatistics* **5**(4), 573–586 (2004)
- Davidov, O., Nov, Y.: Improving an estimator of Hsieh and Turnbull for the binormal ROC curve. *J. Stat. Plan. Inference* **142**, 872–877 (2012)
- Hall, P.G., Hyndmann, R.J.: Improved methods for bandwidth selection when estimating ROC curves. *Stat. Probab. Lett.* **64**, 181–189 (2003)
- Hsieh, F., Turnbull, B.W.: Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Stat.* **24**, 25–40 (1996)
- Krzanowski, W.J., Hand, D.J.: *ROC Curves for Continuous Data*. Chapman and Hall/CRC, London (2009)
- Lloyd, C.J.: Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *J. Am. Stat. Assoc.* **93**, 1356–1364 (1998)
- Lloyd, C.J., Yong, Z.: Kernel estimators of the ROC curve are better than empirical. *Stat. Probab. Lett.* **44**, 221–228 (1999)
- Pepe, M.S.: An interpretation for the ROC curve and inference using GLM procedures. *Biometrics* **56**, 352–359 (2000)
- Pepe, M.S.: *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, Oxford (2003)
- Qin, J., Zhang, B.: Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika* **90**, 585–596 (2003)
- Tonini, G., Fratto, M.E., Imperatori, M., Pantano, F., Vincenzi, B., Santini, D.: Predictive factors of response to treatment in patients with metastatic renal cell carcinoma: new evidence. *Expert Rev. Anticancer Ther.* **11**(6), 921–930 (2011)
- Wolfowitz, J.: The minimum distance method. *Ann. Math. Stat.* **28**(1), 75–88 (1957)
- Zhou, X.H., Harezlak, J.: Comparison of bandwidth selection methods for kernel smoothing of ROC curves. *Stat. Med.* **21**, 2045–2055 (2002)
- Zhou, X.H., Obuchowski, N.A., McClish, D.K.: *Statistical Methods in Diagnostic Medicine*. Wiley, New York (2002)
- Zieliński, R.: Kernel estimators and the Dvoretzky-Kiefer-Wolfowitz inequality. *Appl. Math.* **34**(3), 401–404 (2007)
- Zou, K.H., Hall, W.J.: Two transformation models for estimating an ROC curve derived from continuous data. *J. Appl. Stat.* **27**(5), 621–631 (2000)